MARTHA GUISELA GAITAN GARAVITO

DISEÑO DEL CURSO
INFERENCIA ESTADISTICA
ENSEÑANZA MODULARIZADA

ASESOR:
ING. JORGE PELAEZ CASTELLANOS

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE HUMANIDADES

PROGRAMA DE MAESTRIA EN DOCENCIA UNIVERSITARIA

GUATEMALA, MAYO 1993

PROPIEDAD DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
Biblioteca Central

DL
07
T(549)c

Este estudio fue presentado por la autora como trabajo de Tesis, requisito previo a su graduación en el Programa de Maestría en Docencia Universitaria.

Guatemala, mayo de 1993

INDICE

INTRODUCCION

La Facultad de Odontología tiene como meta formar recursos humanos técnica y científicamente preparados para tratar los problemas de salud bucal del guatemalteco, la estrategia a seguir en el plan de estudios para su consecución, se encamina a lograr que el estudiante comprenda e interprete eficientemente los resultados del enfoque científico en el análisis y solución de los problemas relativos a sistema estomatognático.

Es evidente el papel de la investigación en esta estrategia, por lo que se hace necesario procurar el aprendizaje efectivo de la Metodología de la Investigación.

No obstante, algunas áreas de estudio en la facultad han orientado sus esfuerzos para facilitar al estudiante el aprendizaje sobre Técnicas de Investigación, el tema relacionado con Análisis Estadístico no ha sido enfocado en forma particular, por lo que se han manifestado deficiencias en la realización de las investigaciones planeadas. Este trabajo pretende superar relativamente esas deficiencias, diseñando un curso básico de Inferencia Estadística que ofrezca los conocimientos fundamentales aplicables a la investigación estomatológica.

El curso está orientado a los estudiantes del último año de la carrera, por ser éstos los más involucrados en trabajos de investigación y cuentan con los conocimientos y el criterio profesional para solucionar problemas de salud bucal en las comunidades donde realizan su Ejercicio Profesional Supervisado, situación que permite la aplicabilidad inmediata de las técnicas inferenciales e incentiva al practicante para el aprendizaje.

El recurso didáctico utilizado es la Enseñanza Modularizada, desarrollando el contenido en siete módulos de instrucción: Probabilidad, Distribuciones de Probabilidad, Distribuciones Muestrales, Estimación, Hipótesis, Regresión y Correlación y Métodos No Paramétricos; en cada uno de ellos se presentan los objetivos específicos, asi como la evaluación, a fin de hacer posible la retroalimentación del lector en su formación. Finalmente se incluye una Prueba para la evaluación del curso y la bibliografía de referencia y consulta que permite al interesado profundizar en los temas tratados.

Se hace constar que los ejercicios de aplicación expuestos han sido seleccionados y adaptados de varios textos de Estadística Aplicada especialmente de Bioestadística de Wayne W. Daniel (11) y Bioestadística de Weintraub, Douglass y Gillings (12).

OBJETIVO DE LA TESIS

Elaborar un intrumento de aprendizaje de la Inferencia Estadística adecuado a los requerimientos y potencialidades de los estudiantes del último año de la carrera de Odontología en la Universidad de San Carlos de Guatemala.

PLANIFICACION DEL CURSO

IDENTIFICACION

Curso:	INFERENCIA ESTADISTICA
Recurso Didáctico:	Enseñanza Modularizada
Ciclo:	6o. año de la carrera de Odontología.
Prerrequisito:	Fase I, Primer año Unidad Estadística.

DESCRIPCION

El curso presenta los conceptos y técnicas básicas de Inferencia Estadística con el propósito de introducir al lector en el tema y darle la información fundamental que le capacite para efectuar el análisis estadístico en sus investigaciones de Fin de Carrera, así como para continuar estudios profundos de la materia según lo requieran sus trabajos profesionales.

Especialmente se dirige a los estudiantes del último año, que es el ciclo que dá más énfasis a la investigación en los proyectos del Ejercicio Profesional Supervisado y Tesis, que los enfrenta al análisis y solución de problemas en las diferentes comunidades de la República, situación que les permite aplicar inmediatemente los conocimientos adquiridos.

En la selección de los contenidos y su profundidad se tomó en cuenta los conocimientos estadísticos y la habilidad matemática que poseen los estudiantes, por lo que el fundamento teorico-matemático no es presentado ampliamente, sino por el contrario, se ha tratado de minimizar, dándo mayor énfasis a la aplicación de las técnicas y la descirpción de sus procedimientos.

Los temas son presentados en forma modularizada, al considerar que ésta facilita la comprensión de los conceptos y hace viable el aprendizaje individualizado, tomando en cuenta que a éste nivel, la mayor parte del tiempo de estudio lo absorven las actividades de Ejercicio Profesional Supervisado y no se tiene la oportunidad de realizar constantemente actividades de aprendizaje grupal.

En el estudio de cada módulo se da la oportunidad al lector de ejercitarse continuamente y la ejemplificación con casos relacionados con aspectos de la salud facilitan la comprensión de la problemática, al mismo tiempo que se desarrollan los procedimientos.

Al iniciar las actividades de aprendizaje, se recomienda al participante que en el estudio de los módulos se detenga a analizar cada concepto para garantizar su comprensión, si ésto le presenta mayores dificultades no dude en solicitar tutoría en el Departamento a cargo de este curso, que oportunamente se le ofrecerá.

ESTRUCTURA

Introducción a la Inferencia Estadística

Módulo 1	Probabilidad
Módulo 2	Distribuciones de Probabilidad
Módulo 3	Distribuciones Muestrales
Módulo 4	Estimación
Módulo 5	Hipótesis
Módulo 6	Regresión y Correlación
Módulo 7	Métodos no Paramétricos.

FINALIDADES

El lector al concluir el estudio de los Módulos estará en capacidad de:

1. Identificar las aplicaciones de las Técnicas Estadísticas expuestas.
2. Utilizar correctamente las Técnicas de Inferencia Estadística estudiadas.
3. Planificar el análisis estadístico que es oportuno realizar en el desarrollo de investigaciones particulares.
4. Efectuar el análisis de los datos recolectados en sus investigaciones.
5. Apoyar voluntariamente, sus trabajos de investigación en un marco de análisis estadístico.

METODOLOGIA

El recurso didáctico utilizado es la Enseñanza Modularizada y como complemento se contempla el aprendizaje tutorial por parte del claustro de profesores del departamento a cargo de éste curso, a tal efecto serán calendarizadas sesiones de trabajo de acuerdo con el programa de actividades de los estudiantes del último ciclo.

EVALUACION

Evaluación formativa: hojas trabajo en cada uno de los módulos de instrucción, calificadas por los tutores y remitidas al estudiante.

Evaluación Sumativa:

Hojas de trabajo (10 puntos cada una)....................70
Evaluación Final..30
Total..............100

TABLA DE ESPECIFICACIONES

CONTENIDO PROGRAMATICO	OBJETIVOS (NIVELES DEL DOMINIO CONGNOSCITIVO)						%	No. DE PREGUN.
	CONOCIMIENTO	COMPRENSION	APLICACION	ANALISIS	SINTESIS	EVALUACION		
PROBABILIDAD	1		1				10%	2
DISTRIBUCIONES DE PROBABILIDAD		1	1				10%	2
DISTRIBUCIONES MUESTRALES	1	1	1				15%	3
ESTIMACION	1		1	1			15%	3
HIPOTESIS		1	2	1	1	1	30%	6
REGRESION Y CORRELACION		1	1				10%	2
METODOS NO PARAMETRICOS		1		1			10%	2
PONDERACION %	17%	26%	33%	14%	5%	5%	100%	
No. DE PREG.	3	5	7	3	1	1		20

DESARROLLO DEL CURSO

PROLOGO

El presente material bibliográfico ha sido diseñado especialmente para emplearse en el desarrollo de un curso básico de Inferencia Estadística; presentándose con énfasis la aplicación de las Técnicas Inferenciales de Estimación y Prueba de Hipótesis en el área de la salud bucodental.

Para su desarrollo es prerrequisito el conocimiento de la Estadística Descriptiva: medición, organización de datos, gráficas, medidas de tendencia central y de dispersión, así como de los métodos combinatorios.

La Estadística no tiene gran dificultad de comprensión, solo requiere para su aprendizaje estudiarla progresivamente de tal manera que para pasar de un punto a otro debe ser comprendido a satisfacción el primero, por lo que se recomien da que en el estudio del material el lector se detenga a analizar cada concepto y ejecute los ejercicios relacionados con el tema para garantizar su aprendizaje.

A fin de poder autocontrolar el avance en el mismo, se incluye el cuadro Control de Avance, en el cual e l estudiante puede anotar diversas observaciones y recordar aspectos importantes a consultar en las sesiones de tutoría que se encomiendan a los docentes encargados del curso.

CONTROL DE AVANCE

CONTENIDO TEMATICO	APRENDIZAJE Regular	Bueno	Muy Bueno	OBSERVACIONES
MODULO 1: Probabilidad				
Experimentos, Experimentos Aleatorios.				
Probabilidad, Definiciones, Teoremas de Probabilidad.				
Frecuencia Relativa, Eventos Equiprobables.				
Suceso Independientes.				
Probabilidad Condicional. Otros teoremas de Probabilidad Condicional				
Evaluación				
MODULO 2: DISTRIBUCIONES DE PROBABILIDAD				
Variables Aleatorias Recorrido, Clasificacion de las Variables Aleatorias.				
Variables Aleatorias Discretas. Función de distribución acumulada. Parámetros de las Variables Aleatorias.				
Distribución Binomial				

CONTROL DE AVANCE

CONTENIDO TEMATICO	APRENDIZAJE Regular	Bueno	Muy Bueno	OBSERVACIONES
Variables Aleatorias Continuas.				
Distribución Normal				
Distribución Normal Estándar.				
Distribuciones usadas en pruebas Estadísticas.				
Evaluación.				
MODULO 3: DISTRIBUCIONES MUESTRALES				
Conceptos Fundamentales.				
Selección de la muestra.				
Teoría del Muestreo.				
Distribuciones Muestrales.				
Distribución muestral de medias				
Distribución muestral de diferencias de medias.				
Distribución muestral de proporciones				
Distribución muestral de diferencia de proporciones.				
Distribución muestral de Varianzas.				
Distribución muestral de relaciones de varianzas.				

CONTROL DE AVANCE

CONTENIDO TEMATICO	APRENDIZAJE Regular	Bueno	Muy Bueno	OBSERVACIONES
Evaluación				
MODULO 4 : ESTIMACION				
Inferencia Estadística y Estimación, Conceptos Básicos. Tipos de Estimación. Intervalos de Confianza.				
Intervalos para la Media de una población Normal con Varianza conocida.				
Intervalos para la media.				
Intervalo para la Varianza de una población Normal.				
Intervalos para Proporciones.				
Selección del tamaño de la muestra.				
Evaluación.				
MODULO 6 : HIPOTESIS				
Hipótesis Hipótesis Nula y Alternativa. Comprobación de Hipótesis.				
Prueba de Hipótesis para la Media de una población Normal con Varianza conocida.				
Planteamiento general de las Pruebas de Hipótesis.				

CONTROL DE AVANCE

CONTENIDO TEMATICO	APRENDIZAJE Regular	Bueno	Muy Bueno	OBSERVACIONES
Otras pruebas de Hipótesis.				
Riesgos de hacer falsas decisiones.				
Elección del tamaño de la muestra.				
Análisis de Varianza, Análisis de Varianza de una factor.				
Análisis de frecuencias y la Prueba de Chicuadrado.				
Evaluación				
MODULO 6: REGRESION Y CORRELACION				
Regresión Suposiciones fundamentales de la Regresión.				
Estimación de la recta por el Método de Mínimos Cuadrados. Observaciones Adicionales. Correlación.				
Coeficiente de Determinación. Coeficiente de correlación Producto Momento de Pearson.				
Inferencia relativa al Coeficiente de Correlación.				
Evaluación.				

CONTROL DE AVANCE

CONTENIDO TEMATICO	APRENDIZAJE Regular	Bueno	Muy Bueno	OBSERVACIONES
MODULO 7: METODOS NO PARAMETRICOS.				
Estadísticos no paramétricos. Prueba del Rango con Signo. Prueba U de Wilcoxon. Prueba H de Kruskal y Wallis. Prueba de la Mediana. Coeficiente de Correlación de Spearman. **Evaluación.**				
EVALUACION: PRUEBA FINAL				

MODULO 1

PROBABILIDAD

INTRODUCCION

La teoría de probabilidad proporciona la base para la Inferencia Estadística, por lo que en éste módulo se presentan al lector los conceptos fundamentales de la misma para facilitarles posteriormente la comprensión de los procedimientos de Inferencia.

OBJETIVOS

Al finalizar el estudio del módulo el lector estará en capacidad de:

1. Definir el concepto de probabilidad.

2. Utilizar los Teoremas de Probabilidad para determinar las probabilidades de ocurrencia de los sucesos.

3. Explicar los conceptos de equiprobabilidad, frecuencia relativa, probabilidad condicional e independencia.

4. Calcular probabilidades de sucesos aplicando los conceptos de: Equiprobabilidad, Frecuencia Relativa, Probabilidad Condicional e Independencia.

1.1 EXPERIMENTO

Se utiliza la palabra experimento para describir un proceso que genera un conjunto de datos; en algunos de ellos las consideraciones físicas de las condiciones bajo las cuales se realizan permiten determinar el resultado con suficiente exactitud, algunos ejemplos son los experimentos químicos; en otros las condiciones bajo las cuales se realizan no determinan su resultado porque están influenciados por el azar, por factores no controlables, a éstos se les llaman experimentos aleatorios.

1.2 EXPERIMENTO ALEATORIO

Un experimento aleatorio es cualquier operación o proceso físico cuyo resultado no puede predecirse con exactitud.

Los resultados del experimento aleatorio están influenciados por circunstancias que no pueden ser controladas por el experimentador. Por ejemplo, no se puede predecir con exactitud si una persona bajo determinado tratamiento sanará en los próximos 2, 3, 5 días o no sanará.

Cuando las condiciones bajo las cuales se realiza el experimento no determinan completamente su resultado, se está ante un EXPERIMENTO ALEATORIO. En el experimento aleatorio no existe un solo resultado posible, sino un conjunto de resultados asociados a él. El concepto de probabilidad es necesario en éstos casos para evaluar la posibilidad de ocurrencia de cada uno de los resultados.

Complete las siguientes oraciones:
Otro ejemplo de experimento aleatorio es: ______________________________

____

____

Porque éste experimento no tiene ______________________________
Si no un ______________________________*posibles.*

1.3 PROBABILIDAD

Es una medida matemática de la posibilidad que se produzca un hecho, un acontecimiento o un resultado en una serie de ensayos (experimentos) repetidos en condiciones similares.

Por ejemplo: puede escucharse decir a un médico que un paciente tiene un 50% de probabilidad de sobrevivir a cierta operación, porque esto se ha evidenciado con la experiencia de otras operaciones similares. En otros casos que está 95 % seguro que el paciente tiene determinada enfermedad.

Se suele medir la probabilidad de un suceso por medio de un número entre cero y uno. Cuando más probable sea, más próximo será a uno y cuando menos probable sea más próximo a cero.

1.4 DEFINICIONES

Un experimento aleatorio tiene un conjunto de resultados posibles al que se le llama **ESPACIO MUESTRAL**; a una parte del espacio muestral se le llama **EVENTO**.

Por ejemplo al seleccionar una persona al azar y clasificarla de acuerdo a su grupo sanguineo, se puede ubicar en el grupo O, A, B, ó AB. El conjunto de todas las posibles clasificaciones es el Espacio Muestral (S) del experimento.

S= {A,B,O, AB}

Un evento de éste espacio es la clasificación O:
El evento X = {O}
Otro evento es la clasificación A:
El evento Z = {A}

Dos eventos de un espacio muestral que no pueden ocurrir juntos se denominan **EXCLUYENTES**

En el ejemplo anterior, un persona puede clasificarse en el grupo sanguíneo O ó clasificarse en el grupo A, pero no puede clasificarse en los dos grupos al mismo tiempo. Los eventos X={O} y Z={A} son excluyentes porque no pueden ocurrir juntos.

Si dos sucesos pueden ocurrir juntos no son excluyentes. Para éstos sucesos no excluyentes se puede definir otro suceso denominado INTESECCION (∩) que implica la ocurrencia de los dos al mimo tiempo.

Suponga que se clasifica a un conjunto de personas por su grupo sanguineo y por su sexo.
El conjunto de resultados, especio muestral es:
S= { MO,MA,MB,MAB,FO,FA,FB,FAB}
donde M es Masculino y F femenino.

Considere los sucesos:
"X" que representa al conjunto de personas con grupo sanguineo O; X = {MO, FO}
"Z" que representa al cunjunto de personas de sexo masculino;
Z = {MO, MA,MB, MAB}

El suceso Intersección X ∩ Z = {MO} representa la categoria que incluye a las personas de sexo masculino y con grupo sanguíneo O.
Los sucesos X & Z pueden ocurrir juntos, X & Z no son excluyentes.

En un espacio muestral al definir un evento E se puede determinar un evento complementario al mismo (E^),que implica su no ocurrencia.

Así al clasificar a las personas por grupo sanguineo e identificar el suceso E como: la clasificación en el grupo B, el suceso "complemento de E" (E^), es la clasificación en un grupo diferente de B, la no ocurrenciade B, que incluye al grupo O,A y AB.

S ={O,A,B, AB} E= {B} E^= {O,A,AB}

Resuelva:

A ocho ratones de laboratorio se les ha administrado cierta droga, tres son ratas blancas, dos son ratones grises, dos son ratones pardos y una es rata parda.
Si se selecciona uno de estos animales al azar y se clasifica de acuerdo a su color y sexo, el espacio muestral del experimento es:
S= {   }
El suceso A está definido como la selección de una rata parda, el suceso A^ se define como: ____________________
*Los Sucesos A= un animal pardo y B= una rata, son sucesos:*__________
*porque:*____________________
Los sucesos C= un ratón y B= un animal blanco, son sucesos:
*porque:*____________________

1.5 TEOREMAS DE PROBABILIDAD

Dado un experimento aleatorio, con un espacio muestral S formado por n resultados mutuamente excluyentes (E1, E2, E3, E4, ..., En). A cada uno de los resultados se asigna un número positivo denotado por P(Ei), i=1,2,3,4,...n , que se llama probabilidad de Ei y mide la posibilidad de ocurrencia del resultado.

P(Ei) satisface las siguientes condiciones:

I. P(Ei) para i = 1,2,.....n es un número positivo entre cero y uno.

$$0 <= P(Ei) <= 1$$

II. La suma de las probabilidades de todos los resultados mutuamente excluyentes de un espacio muestral es igual a uno.

$$P(E1) + P(E2) + P(E3) +.........+ P(En)= 1$$

La probabilidad del espacio muestral es 1.

$$P(S)=1$$

III. Si se consideran dos eventos cualesquiera A y B, mutuamente excluyentes, la probabilidad que ocurra el evento A o el evento B es igual a la suma de las probabilidades individuales de A y de B

$$P(A \text{ ó } B)= P(A) + P(B)$$

IV. Si A, B y C son suceso excluyentes, la probabilidad que ocurra A ó B ó C es la suma de las probabilidades individuales.

$$P(A \text{ ó } B \text{ ó } C)= P(A) + P(B) + P(C)$$

V. Si se consideran los eventos A y A^ mutuamente excluyentes, la probabilidad de que A ó A^ ocurran es:

$$P(A \text{ ó } A^{\wedge}) = P(A) + P(A^{\wedge}) = 1$$

Entonces: $P(A^{\wedge}) = 1-P(A)$

VI. Si D y E son sucesos no excluyentes, la probabilidad de que D o E o ambos ocurran es la suma de las probabilidades menos la probabilidad que ocurran simultáneamente.

$$P(D \text{ ó } E \text{ ó ambos}) = P(D) + P(E) - P(D \cap E)$$

Por ejemplo:

En una población dada la probabilidad de que una persona tenga oclusión clase I es 0.7, clase II es 0.25, clase III es 0.05. Las tres formas de oculsión son mutuamente excluyentes.

A= Oclusión clase I $P(A)= 0.70$
B= Oclusión clase II $P(B)= 0.25$
C= Oclusión clase III $P(C)= 0.05$

Si se elige una persona al azar de esta población, la probabilidad que tenga oclusión clase II o Clase III es:
$P(A \text{ ó } B)= P(A) + P(B) = 0.25 + 0.05 = 0.30$

Además : A^ ={ Oclusión clase II o Clase II }
por lo que también podría calcularse la probabilidad así:
$P(A^{\wedge}) = 1- P(A) = 1- 0.7 = 0.30$

Un grupo de adultos se clasifica de acuerdo a si su primeros molares están cariados (A), Obturados (B) o sanos (C), con las siguientes probabilidades asociadas:

$P(A) = 0.14$
$P(B) = 0.61$
$P(C) = 0.29$
$P(A \cap B) = 0.04$

La probabilidad de que un primer molar seleccionado al azar:

I. Tenga caries pero no esté obturado es:
D = Unicamente cariado
E = Cariado y obturado, $A \cap B$
$A = (D \text{ ó } E)$ entonces $P(A)= P(D) + P(E)$
$P(D) = P(A) - P(E) = 0.14 - 0.04 = 0.10$

II. Tenga caries o esté sano:
$P(A \text{ ó } C) = P(A) + P(C) = 0.14 + 0.29 = 0.43$

III. Esté cariado u obturado o ambas cosas.
$P(A \text{ ó } B \text{ ó Ambos}) = P(A)+P(B)-P(A\cap B)= 0.14+0.61-0.04=0.71$

Resuelva:

Para un matrimonio la probabilidad de que el esposo acuda a la consulta anual con el odontólogo es 0.21, de que la esposa acuda es 0.28 y que ambos acudan es 0.15, los eventos que se identifican en el experimento son:

A= ______________________________ *P(A)=* ______________
B= ______________________________ *P(B)=* ______________
C= ______________________________ *= A∩B P(C)=* __________

La probabilidad de:

I. Cuando menos uno de los esposos acuda a la consulta es :

II. La esposa no acuda a la consulta:

III. Solo uno de los esposos o ninguno acuda a la consulta:

1.6 FRECUENCIA RELATIVA

Para medir la posibilidad de ocurrencia de un evento particular en un experimento, existe una determinación práctica que consiste en repetir el experimento un número grande de veces (n), calcular la frecuencia con que se dió el evento y su frecuencia relativa.

Si algún proceso se repite un gran número de veces (n) y si un evento resultante E, con una característica particular, ocurre m de las n veces, la frecuencia relativa de E es (fr(E)) m/n y será aproximadamente igual a la probabilidad de ocurrencia de E (P(E)) en tanto n tienda al infinito.

La dificultad de utilizar la frecuencia relativa, como método para estimar probabilidades, es que su cálculo debe hacerse después de haber realizado n veces el experimento.
¿ Es posible realizar n veces el experimento ?.....
¿ Qué tan grande puede ser n ?.....
Estas preguntas debe contestarlas el experimentador de acuerdo a su situación particular.

Por ejemplo:

En un grupo de 502 personas se determinó la distribución de los grupos sanguíneos de la siguiente forma:

Grupo sanguíneo	Frecuencia
O	226
A	206
B	50
AB	20

Si se selecciona al azar una persona de este grupo, la probabailidad que tenga sangre tipo A es estimada por:
fr(A)= 206/502 = 0.41

Resuelva:

Entre 78 médicos que trabajan en un hospital 64 poseen seguro contra negligencias, 36 son cirujanos y 34 de los cirujanos tienen seguro contra negligencias. Los eventos que se pueden formar son:

A=______________________________fr(A) =________
A^ =____________________________fr(A^) =________
B=______________________________fr(B) =________
B^=_____________________________fr(B^) =________
A∩B=____________________________fr(A∩B)=________

Si se elige uno de estos médicos al azar, cuál es la probabilidad de elegir:
I. Uno que no sea cirujano?

II. Qué posea seguro contra negligencias?

III. No sea cirujano y no posea seguro contra negligencias?

1.7 EVENTOS EQUIPROBABLES

Si un experimento tiene n resultados posibles (a1, a2, a3 a4,...an), mutuamente excluyentes e igualmente probables, entonces la probabilidad de cada uno de los resultados es 1/n;

$$P(a_i)=1/n \qquad i= 1,2,3,....n$$

además si m de estos resultados forman un evento E, la probabilidad de ocurrencia de E es m/n.

$$P(E)= m/n$$

Por ejemplo:

Un educador en asuntos de sanidad tiene tres carteles, para exhibirlos uno junto a otro en la pared del vestíbulo en el Centro de Salud, colocándolos de manera aleatoria. Las formas distintas como puede colocar los tres carteles son:

I Cartel 1 ; cartel 2 ; cartel 3
II Cartel 2 ; cartel 1 ; cartel 3
III Cartel 2 ; cartel 3 ; cartel 1
IV Cartel 1 ; Cartel 3 ; cartel 2
V Cartel 3 ; cartel 2 ; cartel 1
VI Cartel 3 ; cartel 1 ; cartel 2

n= 6 formas

La probabilidad de colocar el cartel tres en la posición central es:

Al ser la colocación de manera aleatoria, cada arreglo tiene igual probabilidad de seleccionarse, los eventos son equiprobables. La probabilidad de cada arreglo es 1/n=1/6.

El evento E que se define como el cartel 3 en la posición central incluye los arreglos III y IV;

$$m= 2 \quad y \quad P(E)=m/n=2/6$$

Resuelva:

Dos tabletas contra el resfrio se colocan accidentalmente en una caja que contiene 2 analgésicos, las cuatro tabletas son idénticas. Si un paciente selecciona 2 tabletas al azar para tomarlas:

El número de alternativas de selección es: ____________
La probabilildad para cada una de las alternativas es: _______
La probabilidad que al seleccionar las dos tabletas y tomarlas, exactamente tome una tableta contra el resfrio es:______________
La probabilidad que tome al menos un analgésico es:___________

1.8 SUCESOS INDEPENDIENTES

Si la aparición o no aparición de un hecho no afecta en forma alguna a otro hecho particular, se considera a estos como hechos o SUCESOS INDEPENDIENTES.

Para determinar la probabilidad de que se produzca la aparición de dos sucesos independientes, se multiplican las probabilidades de que suceda cada uno de ellos en forma individual

Si A y B son Independientes $P(A \cap B) = P(A)*P(B)$

Por ejemplo:

La probabilidad de que una enfermera de salud pública encuentre a un paciente en su casa es 0.7. Si se considera, que para cada paciente potencial la probabilidad es la misma y que cada paciente es independiente de los otros.

La probabilidad de que después de visitar dos casas, encuentre en ellas a ambos pacientes es:

A= se encuentra al paciente de la primera casa.
B= se encuentra al paciente de la segunda casa.

$P(A \cap B)$= probabilidad que se encuentren ambos pacientes
$P(A \cap B) = P(A)*(PB) = 0.7*0.7 = 0.49$

La probabilidad que encuentre al menos uno de los dos pacientes es:
$P(A \text{ ó } B \text{ ó Ambos}) = P(A)+P(B)-P(A \cap B) = 07.+0.7-0.7*0.7=0.91$
(recuerde que se trata de sucesos no excluyentes)

Resuelva:

Un sistema para detectar humo utiliza dos dispositivos A y B; si hay humo la probabilidad que sea detectado por A es 0.95, y que sea detectado por B es 0.98. Si los dispositivos operan en forma independiente.

¿ Cuál es la probabilidad, si hay humo, que sea detectado por al menos uno de los dispositivos?

¿Cuál es la probabilidad que no sea detectado?

1.9 PROBABILILDAD CONDICIONAL

En algunas ocasiones dos sucesos se relacionan de manera que la probabilidad de ocurrencia de uno aumenta o disminuye dependiendo si el otro a ocurrido o no, en éstos casos es necesario definir la probabilidad condicional, P(A/B)

La probabilidad condicional P(A/B) determina la probabilidad que ocurra el suceso A dado que ya ocurrió el suceso B.

Suponga el experimento que consiste en observar la condición del tiempo en un día específico. Sea A el evento, observar un día lluvioso, y B el evento un día nublado. Si consideramos unicamente los días en que está nublado, la fracción de esas observaciones en las que sucede A (lluvia) es la probabilidad condicional de A dado B, P(A/B).

$$P(A/B)= P(A \cap B) \ / \ P(B)$$

Por ejemplo:

En un estudio del comportamiento después de tratamiento de un gran número de drogadictos, se sugiere que las probabilidades de reincidencia dentro de los dos años siguientes al tratamiento depende de la educación de los transgresores.
Las proporciones del número de casos que caen en cuatro categorías de educación se presentan a continuación.

Condiciones dentro del periodo de dos años
de tratamiento

Educación	Reincidencia Si (B)	No (D)	
10 o más años (A)	0.10	0.30	0.40
9 o menos años (C)	0.27	0.33	0.60
	0.37	0.63	1.00

Si selecciona un transgresor del grupo en tratamiento:

La probabilildad de que tenga 10 años o más de educación (A) dado que reincidió dentro de los 2 años (B) es:

$$P(A/B) = P(A \cap B)/P(B)= 0.10/0.37= 0.27$$

Que representa la fracción de reincidentes que tienen 10 o más años de educación.

La probabilidad que reincida (B) si el transgresor tiene menos de 10 años de educación (C) es:

$$P(B/C) = P(B \cap C)/P(C)= 0.27 /0.60 =0.45$$

Resuelva:

En un experimento para estudiar la relación entre la hipertensión y el hábito de fumar se reúnen los siguientes datos de 180 individuos.

	No fumadores	*Fumador*	*Fumador Empedernido*
Hipertensos	*21*	*36*	*30*
No Hipertensos	*48*	*26*	*19*
total	*69*	*62*	*49*

Si se selecciona aleatoriamente uno de esos individuos, encuentre la probabilidad de que la persona:

I. Experimente hipertensión dado que es un fumador.
P(/)

II. Sea un no fumador, dado que no ha presentado problemas de hipertensión.
P(/)

1.10 OTROS TEOREMAS DE PROBABILIDAD CONDICIONAL

Un teorema útil para calcular la probabilidad de un evento es el de MULTIPLICACION DE PROBABILIDADES, que se define a partir de la probabilidad condicional.

$$P(A \cap B) = P(B)*P(A/B)$$

El teorema de multiplicación de probabilidades permite calcular la probabilidad de la ocurrencia de dos sucesos simultaneamente, si ellos no son independientes.

Por ejemplo.

En una ciudad el 45% son hombres (H) y el 55 % mujeres (M). El 80% de los hombres y el 90% de las mujeres está en contra del aborto (A), si se entrevista a una persona seleccionada al azar, la probabilidad que sea un hombre que esté a favor del aborto (A^) es:

$P(H)= 0.45$ $\quad$ $P(M)=0.55$
$P(A/H)= 0.80$ $\quad$ $P(A^\wedge/H)= 1-PA/H)= 0.20$
$P(A/M)= 0.90$ $\quad$ $P(A^\wedge/M)= 1 - P(A/M) = 0.10$

$H \cap A^\wedge$= se selecciona un hombre y está a favor del aborto.
$P(H \cap A^\wedge)= P(H)*P(A^\wedge/H)= 0.45*0.20= 0.09$

Resuelva:

Cuál es la proababilidad que se seleccione una mujer que esté en contra del aborto?

Otros teoremas importantes para el cálculo de probabilidades es el Teorema de Probabilidad Total y el Teorema de Bayes.

El TEOREMA DE LA PROBABILIDAD TOTAL es útil cuando el interés es calcular la probabilidad de él o los efectos (E) ocasionados por varias causas (Ai)

Si A1, A2, A3,.....Ai.....An, son causas que pueden conducir al mismo efecto E, entonces

P(E)=P(A1)*P(E/A1)+P(A2)*P(E/A2)+...+P(Ai)P(E/Ai)+P(An)P(E/An)

Por otra parte dado la ocurrencia de un efecto (E), el TEOREMA DE BAYES permite calcular la probabilidad que éste se haya debido a cierta causa (Ai).

P(Ai/E)= P(Ai)P(E/Ai)/ P(E)

donde P(E) está dado por el teorema de probabilidad total.

Por ejemplo:

En un laberinto en forma de T se le da comida a una rata (A1) si vira a la izquierda y recibe un choque eléctrico (A2) si lo hace a la derecha. En el primer ensayo hay una probabilidad 50-50 de que la rata vire en una dirección o en otra.

Si recibe comida en el primer ensayo, la probabilidad que vire a la izquierda en el siguiente es 0.72, P(I/A1). Si recibe un choque eléctrico en el primer intento la probabilidad es 0.88 de que vire a la izquierda en el siguiente intento (P(I/A2).

La rata puede virar a la izquierda en el segundo ensayo (I) debido a que en el primero recibió comida o por haber recibido un choque eléctrico.

Entonces la probabilidad total del evento virar a la izquierda en el segundo ensayo (I) es:

P(I)= P(A1)P(I/A1) + P(A2)P(I/A2)= 0.5*0.72+0.5*0.88=0.80

Ahora si se supone que la rata vira a la izquierda en el segundo ensayo (I) la probabilidad que haya recibido comida en el primer intento es:

P(A1/I)= P(A1) P(I/A1)/ P(I) = 0.5*0.72/0.80 = 0.445

Resuelva:

Suponga que de 1200 admisiones a un hospital durante cierto periodo, 750 son admisiones de menores de edad, si se selecciona un paciente al azar la probabilidad que:

Sea adulto (A1) es:____________
No sea adulto (A2) es: _________
Si se sabe que el 25 % de las admisiones de adultos y el 15 % de las admisiones de menores de edad son a la sala de cuidados intensivos (B), identifique:
P(B/A1)
P(B/A2)
P(B^/A1)
P(B^/A2)
Si se selecciona un paciente al azar de los que se encuentran en cuidados intensivos, la probabilidad que sea un adulto es:
P (/   ) =

1.11 EVALUACION

I. Contestes las siguientes preguntas:

1. *Cuál de los siguientes pares de eventos son mutuamente excluyentes:*
A. *Un contratista del departamento de defensa pierde un importante contrato y el mismo contratista aumenta su fueza de trabajo en 50%.*
B. *Un hombre es mayor que su tio pero más joven que sus primos.*
C. *Un equipo de Beisbol pierde el último juego de la temporada y gana la Serie Mundial.*
D. *Un gerente de un banco descubre que su cajero ha cometido fraude y pese a ello le da un ascenso.*

2. *Cuál de los siguientes pares de eventos son estadísticamente independientes:*
A. *Los tiempos antes de falla de una calculadora y de una segunda calculadora comercializadas por una firma diferente.*
B. *La esperanza de vida del actual presidente Norteamericano y del actual presidente de Panamá.*
C. *La proporción de arreglos en los casos de envenenamiento por asbesto de Marylan y Nueva York.*
D. *La adquisición de una compañia y un incremento en el precio de sus acciones.*
E. *La frecuencia de donaciones de órganos en una comunidad y la orientación religiosa que predomina en ella.*

3. *Las probabiliades de que llueva hoy son 80%. Cuál de los siguientes enunciados expresa mejor esa afirmación?*

A. *Hoy lloverá el 80% del día.*

B. *Lloverá en el 80% del area a la cual se aplica este pronóstico.*

C. *En el pasado las condiciones climatológicas han producido lluvia en esta area el 80% de las veces.*

II. Resuelva los siguientes problemas

1. Un profesor ha ordenado libros de una editorial para su nuevo curso. Una tercera parte de esos libros se imprimen en la planta situada en New Jersey, las otras dos terceras partes se imprimen en la planta situada en Delavere. Ambas plantas tienen probabilidades respectivas de errores de impresión de 0.08 y 0.04. Si el profesor selecciona un ejemplar del embarque y descubre que tiene un error de impresión. Qué planta tiene mayores probabilidades de haber impreso el libro?

2. La Sociedad Americana de Cáncer está planeando enviar cuestionarios referentes al Cáncer Mamario, la experiencia con cuestionarios le ha enseñado que sólo un 12% de los que reciben responderán. También sabe que el 1% de los cuestionarios contendrán un error en el domicilio del destinatario y nunca serán entregados, que un 3% se extraviará o será destruido accidentalmente en la oficina postal, que el 22% será enviado a personas que han cambiado de dirección y que apenas el 52% de los que cambian de domicilio dejan una dirección para que se les envien su correspondencia. Cuál es la probabilidad que la sociedad reciba una respuesta al cuestionario?

3. La agencia de Protección del Medio Ambiente está tratando de evaluar el efecto de una fábrica de papel que será contruida cerca de Spokane, en estudios de 6 plantas similares ese organismo determinó los siguientes factores contaminantes:

Planta	*1*	*2*	*3*	*4*	*5*	*6*
Emisión de dioxido de Azufre ppm	*15*	*12*	*18*	*16*	*11*	*19*

La agencia de proterección define la contaminación excesiva como una emisión de dióxido de azufre de 18 ppm en adelante. Estime la probabilidad de que la nueva planta contamine en exceso con el dióxido de azufre.

4. La probabilidad que un paciente se recupere de una delicada operación del corazón es 0.8. Cuál es la probabilidad que sobrevivan exactamente dos de los siguientes tres pacientes que han tenido esta intervención?

MODULO 2

DISTRIBUCIONES DE PROBABILIDAD

INTRODUCCION

En éste modulo se estudian los conceptos relacionados con variables aleatorias, complementando los principios de probabilidad expuestos anteriormente. En forma especial se tratan las distribuciones Binomial y Normal con sus aplicaciones, al considerar que son los modelos más utilizados en la teoría de muestras que será expuesta en los siguientes capitulos. Finaliza con la presentación de las distribuciones Chi-cuadrado, T de Student, F y el manejo de las tablas de probabilidad respectivas.

OBJETIVOS

Al finalizar el estudio del Módulo el lector estará en capacidad de:

1. Identificar la o las variables aleatorias asociadas a un experimento.

2. Calcular las distribuciones de probabilidad de las variables aleatorias.

3. Interpretar los parámetros: Esperanza, Varianza y Desviación Estándar de las variables aleatorias.

4. Identificar las aplicaciones de las distribuciones Normal y Binomial.

5. Calcular la probabilidad de los sucesos, utilizando la distribución Normal y Binomial.

6. Manejar adecuadamente las tablas de las siguientes distribucines: Chi-cuadrado, T de Student y F.

2.1 VARIABLES ALEATORIAS

Al estudiar un fenómeno, la mayoría de las veces se está interesado en un aspecto o resultado en particular. Por ejemplo: Una enfermera de Salud Pública tiene a su cargo 50 familias y está interesada en estudiar el númereo de niños por familia; la característica en particular a estudiar "X" es el número de niños por familia. Puede estar interesada también, en estudiar la edad cronológica que tiene el o la jefe de familia; la característica "Y" a estudiar es la edad en años del jefe de familia. Supónga ahora que la enfermera selecciona una familia al azar de este grupo, puede ser la familia Gómez, Perez, Juárez; cuánto vale "X" para ésta familia? cuánto vale "Y" para ésta familia?; los resultados no estan determinados, la selección de la familia es un Experimento Aleatorio que tiene un conjunto de resultados posbles por lo que X & Y también tienen un conjunto de resultados posibles, X & Y son variables Aleatorias.

Una variable aleatoria es una función (un planteamiento) que asocia números a cada uno de los elementos del espacio muestral de un experimento.

Así, si la familia Gómez tiene 3 hijos y el jefe de familia tine 42 años, al ser seleccionada dentro del grupo (elemento), las variables:
X= Número de hijos, asignará al elemento el número 3 (x=3)
Y= Edad del jefe de familia asignará al elemento el número 42 (y=42).

Complete lo siguiente:

*Qué otros aspectos puede estudiar la enfermera en el grupo de familias?*____________________________________

____

____

Qué nuevas variables aleatorias se pueden asignar al experimento?
1. ____
2. ____

2.2 RECORRIDO DE LA VARIABLE ALEATORIA

Si se considera que en el grupo de 50 familias hay algunas que no tienen hijos y otras que tienen hasta cinco hijos, los posibles valores que X puede tomar son: 0, 1, 2, 3, 4, 5. A este conjunto de resultados se le denomina recorrido de X.

$$Rx = \{0,1,2,3,4,5\}$$

Se denomina Recorrido al conjunto de todos los valores posibles que puede tomar una variable aleatoria.

Los Recorridos de las variables que usted asignó son:
*1.*____
*2.*____

2.3 CLASIFICACION DE LAS VARIABLES ALEATORIAS

De acuerdo a su recorrido las variables aleatorias se pueden clasificar en Discretas y Continuas.

X es una Variable Aleatoria Discreta si solo puede asumir un conjunto numerable de valores, puede asumir n valores.

Rx={x1, x2, x3, x4,....xn.}

Por ejemplo:

De un conjunto de escolares se seleccionan 3 al azar, si X es el Número de varones seleccionados, X puede tomar cuatro valores, Rx= { 0,1,2,3 }
X es una variable aleatoria Discreta.

Otro ejemplo de variable Discreta es:
X= ____
Rx= ____

X es una Variable Aleatoria Continua si el conjunto de sus posibles valores es infinito no numerable.
La forma idonea de representar una variable continua es a través de un intervalo en el conjunto de los números reales.
Rx: a<= X <=b (a,b, pertenecen al conjunto de los números reales).

Por ejemplo:

Se ha observado, en un experimento, que el tiempo que tarda una reacción química de ciertos compuestos está comprendido en el intervalo de 0.1 a 2 segundos.
Si se realiza uno de estos experimentos y X es la variable "tiempo de reacción", el recorrido de X es:

Rx: 0.1 <= X <= 2

El modelo de variables aleatorias continuas puede ser utilizados en experimentos que tengan asociados resultados discretos, pero siendo el número de elementos muy grande se idealizan como continuos.

Por ejemplo:

El tiempo de permanencia en un hospital de enfermedades crónicas para cierto tipo de pacientes es de 45, 46, 47 .... 72, 73, 74, 75 días.
Si X es la variable "número de días de permanencia en el hospital", Rx puede idealizarse como un recorrido continuo

Rx: 45 <= X <=74

Otro ejemplo de Variable Continua es:
*X=*____
*Rx=*____

2.4 VARIABLES ALEATORIAS DISCRETAS

Si X es una variable aleatorias definida en el espacio muestral de un experimento aleatorio, se dice que X es discreta si su recorrido es un conjunto numerable.

$$Rx = \{x1, x2, x3 ....xn\}$$

y a cada uno de los posibles valores de X se le puede asociar un número real positivo P(xi) i=1,2,3....n que es llamado probabilidad (puntual) de xi, que satisface las siguientes condiciones.

$$0 <= P(xi) <= 1$$

$$\sum_{i=1}^{n} P(xi) = 1$$

Por ejemplo:

Considere que la probabilidad de nacimientos de un niño (M) es igual a la probabilidad de nacimientos de una niña (F) y es 0.5, además cada nacimiento es independiente del anterior. Si de un grupo de familia con 2 hijos se selecciona una al azar (experimento), el espacio muestral puede representarse por:

$$S= \{(M,F), (M,M), (F,M), (F,F)\}$$

Si X es la variable número de niños (M) en la familia,

$$Rx= \{0,1,2\}$$

Las probabilidades asociadas a cada elemento son:
P(X=0)= P(F) * P(F)= 0.5*0.5=0.25
P(X=1)= P(F) * P(M) + P(M) * P(F)= 0.50*0.5 + 0.5*0.5 =0.25
P(X=2)= P(M) * P(M) = 0.5 * 0.5 = 0.25

$$\sum_{i=1}^{3} P(xi) = 0.25 + 0.5 + 0.25 = 1$$

La probabilidad de X sea mayor o igual a uno se calcula:
P(X>=1)= P(X=1) + P(X=2) = 0.75

La distribución de probabilidades de la variable X es

X	P(X)
0	0.25
1	0.50
2	0.25
	1.00

Una fórmula, tabla o gráfica que muestre el recorrido de la variable aleatoria asociado a sus probabilidades se denomina Distribución de Probabilidades.

Resuelva:

En una reunión de trabajo se encuentran un Odontólogo, dos Médicos y un Químico Biólogo, si se seleccionan tres personas al azar. Por qué profesionales puede estar formado el grupo?

$S = \{ \quad \}$

Si X es la variable "Número de Odontólogos que forman el grupo", el rescorrido de la variable X es:

Rx=

La distribución de probabilidades de X es

X	*P(X)*

La probabilidad que X sea menor que uno es:

P(X) =

2.5 FUNCION DE DISTRIBUCION ACUMULADA

La probabilidad de que la variable aleatoria X tome un valor menor o igual a un xj especifico se define como Función de Distribución Acumulada F(X)

$$F(X=x_j)= P(X <= x_j) = \sum_{i=1}^{j} P(x_i)$$

$$j=1,2,\ldots\ldots n$$

La función de distribución acumulada de la variable X número de niños del ejemplo precedente es:

$F(X=0)= P(X <= 0)= 0.25$
$F(X=1)= P(X <= 1)= 0.25 + 0.5 = 0.75$
$F(X=2)= P(X <= 2)= 0.25 + 0.5 + 0.25 = 1$

X	F(X)
0	0.25
1	0.75
2	1.00

Resuelva:

La función de distribución acumulada del número de odontólogos que forman el grupo del ejercicio anterior es:

X	*F(X)*

2.6 PARAMETROS DE LAS VARIABLES ALEATORIAS

En la descripción de las variables aleatorias, los parámetros dan una información valiosa acerca de la distribución de probabilidades y algunas veces se usan para señalarlas. Estos parámetros son la Esperanza, la Varianza y la Desviación Estándar.

E(X) la Esperanza, valor esperado o valor promedio de X, es un promedio ponderado de los posibles valores de X, teniendo como medida de ponderación sus probabilidades de ocurrencia.

$$E(X)= \sum_{i=1}^{n} xi\ P(xi)$$

Por ejemplo:

En una clínica de detección del Cáncer Bucal, no se puede conocer con exactitud cuántas personas serán atendidas en un día cualquiera, por tanto, el número de pacientes que serán atendidos en una mañana es una variable aleatoria. Sin embargo algunas decisiones administrativas requerirán conocer el número de pacientes (variable) que se espera que lleguen.
Si los expedientes diarios de la clínica indican la variable fluctúa entre 10 y 15. Qué se espera que suceda? Será una cifra cercana a 15? o cercana a 10?

El valor esperado es un concepto fundamental en éste análisis; que es un promedio ponderado de los resultados que se esperan en el futuro; si el valor esperado es, por ejemplo, 8.02, el Director de la Clinica basará sus decisiones en función de ese valor esperado, así hará una estimación del tiempo que estará ocupado el laboratorio, la cantidad de material a utilizar, etc.

Qué significa esa cifra? Significa que durante un largo periodo el número de exámenes diarios deberá promediar cerca de 8.02. No olvide que es un valor esperado, no significa que mañana visitarán la clinica 8.02 personas.

V(X) la Varianza de la variable aleatoria, es una medida de dispersión de la distribución de probabilidades, es el promedio ponderado o esperanza de las desviaciones cuadradas de cada uno de los valores posibles de X en relación a E(X)

$$V(X) = E\ (\ xi - E(X)\ )^2$$
$$V(X) = E\ (\ X^2) - [\ E(X)\ ]^2$$
$$\text{donde: } E\ (X^2) = \sum xi^2\ P(xi)$$

La raíz cuadrada de la varianza es la Desviación Estándard $\sigma(X)$, da información sobre que tan alejados en promedio pueden situarse los valores observados con respecto a E(X) y una idea del recorrido de la distribución si este es desconocido.

$$\sigma(X)= \sqrt{V(X)}$$

Por ejemplo:

Un Centro de Salud tiene capacidad para internar 4 pacientes, de cierto tipo en emergencia, durante la noche.
Si X es la variable "número de pacientes que llegan durante la noche a la emergencia" y tiene una distribución de probabiliad

$$P(X)= 1/6 \text{ para } X= 0,1,2,3,4,5$$

El valor esperado de X es:

xi	P(xi)	xi P(xi)	xi²P(xi)
0	1/6	0	0
1	1/6	1/6	1/6
2	1/6	2/6	4/6
3	1/6	3/6	9/6
4	1/6	4/6	16/6
6	1/6	5/6	25/6
1		15/6= E(X)	55/6 = E(x²)

E(X)= 15/6= 2.5
En promedio se esperan 2.5 pacientes por la noche.

La varianza y la desviación estándar de X son respectivamente.

$V(X)= E(X^2) - [E(X)]^2 = 55/6 - (15/6)^2 = 2.9166$ pacientes²

$\sigma(X) = \sqrt{V(X)} = \sqrt{2.9166}= 1.708$ pacientes

El número de pacientes que llegan a la emergencia se alejan en promedio de E(X) en ± 1.708

Resuelva:
Un médico se ha dado cuenta, que el número de horas en que cierto medicamento hace efecto contra una enfermedad, es una variable aleatoria con la siguiente distribución de probabilidades
X= horas que transcurren antes de hacer efecto el medicamento

xi	*P(xi)*	*xiP(xi)*	*xi² P(xi)*
12	*0.05*		
14	*0.10*		
16	*0.20*		
18	*0.30*		
20	*0.20*		
22	*0.10*		
24	*0.05*		

El valor esperado de X es ______________
Que representa: ____

La varianza de X es ____
La desviación estándar de X es ____

2.7 DISTRIBUCION BINOMIAL

La distribución Binomial una de las distribuciones que se encuentran con más frecuencia en la Estadística Aplicada, se obtiene a partir del proceso denominado Prueba de Bernoulli.

Un experimento que puede conducir solo a un conjunto de dos posibles resultados, mutuamente excluyentes, tales como enfermo o sano, masculino o femenino, se conoce como PRUEBA DE BERNOULLI.

Los resultados de la prueba de Bernoulli se denominan Exito y Fracaso, su distribución de probabilidades está dada por:

$$P(\text{Exito}) = p \qquad P(\text{fracaso}) = q = 1-p$$

Considere la siguiente situación:

Se repite n veces una prueba de Bernoulli, cada una independiente de las anteriores, es decir el resultado de una prueba no es afectado por el resultado de cualquier otra y la probabilidad de éxito permanece contante en todos los ensayos. Si X es la variable aleatoria definida como el " Número de Exitos" encontrados en las n pruebas, X tiene una Distribución Binomial con recorrido:

$$R_x = \{ 0,1,2,3,4,\ldots\ldots n \}$$

y la probabilidad de que X tome un valor k específico esta dada por:

$$P(X=k) = (n!/ (k!*(n-k)!) p^k * q^{(n-k)}$$
$$k = 0,1,2,3\ldots n$$

donde $k! = 1*2*3*\ldots k$
$n! = 1*2*3*\ldots n$
$(n-k)! = 1*2*3*\ldots(n-k)$

El valor esperado de la variable Binomial, que representa el número de éxitos esperado en las n puebas es:

$$E(X) = np$$

La varianza de la variable Binomial es:

$$V(X) = npq$$

La desviación estándar de la variable Binomial es:

$$\sigma(X) = \sqrt{npq}$$

Por ejemplo:

En un país en vía de desarrollo el 30% de los niños está desnutrido. Si una muestra al azar de 25 niños se extrae de ésta área, la probabilidad de que el número de niños desnutridos en la muestra sea exactamente 10 es:

*Planteamiento:

Cada niño seleccionado al azar (experimento) puede caer en cualquiera de dos categorías,

Exito = desnutrido

Fracaso = No desnutrido

Por lo que en cada caso se tiene una prueba de Bernoulli

$p = 0.3 \quad q = 1-p = 0.7$

*Variable Aleatoria:

X= Número de niños desnutridos en una muestra de 25 (se repite la prueba de Bernoulli n=25 veces)

X es una variable Binomial.

*Probabilidad:

La probabilidad que en una muestra de 25 niños 10 sean desnutridos es:

P(X=10)=(25)!/(10!* (25-10)!)*0.3^10*0.7^15=0.091

La probabilidad de que a lo más dos sean desnutridos es:

P(x=0)=25!/(0!*25!)*0.3^0*0.7^25= 0.000134

P(x=1)=25!/(1!*24!)*0.3^1*0.7^24= 0.00143

P(x=2)=25!/(2!*23!)*0.3^2*0.7^23= 0.00738

P(X <= 2)=P(x=0)+P(x=1)+P(x=2)= 0.0089

El número de niños que se espera que estén desnutridos en la muestra de 25 es:

E(X)= np = 25 * 0.30 = 7.5 niños

La Desviación Estándar de la distribución es:

$\sigma(x) = \sqrt{npq} = \sqrt{25*0.3*0.7} = 2.29$ niños

Los resultados de la variable se alejan en promedio de 7.5 en 2.29 niños

Resuelva:

Después de evaluar una radiografía panorámica el odontólogo informó al paciente que tenía cuatro terceros molares retenidos, le sugirió que concertara una cita con un cirujano bucal para que se los extrajese. El paciente decide analizar la situación más a fondo tomando en consideración lo siguiente: La probabilidad de que uno de estos molares provoque dolor o infección en el futuro es 0.25 y la probabilidad que permanezca no afectado es 0.75. Si se supone que lo que pase con uno de los molares es independiente de lo que pase con los otros.

La prueba de Bernulli asociada a este experimento es:

____

____

____

*p=*________ *q=*________ *n=*________

La variable de estudio en esta situación es:

____

____

*Cuántos molares se espera que causen problemas?*____________

*Cuál es la desviación estandar de la distribución?*__________

Cuál es la probabilidad que a lo más uno de los molares se vean afectados?
P(X   ) =

Cuál es la probabilidad que todos los molares se vean afectados?
P(x   )=

2.8 VARIABLES ALEATORIAS CONTINUAS

Anteriormente se definió una Variable Aleatoria Continua como una variable que puede asumir un número infinito de valores, que corresponden a los puntos sobre un intervalo en el conjunto de los números reales,

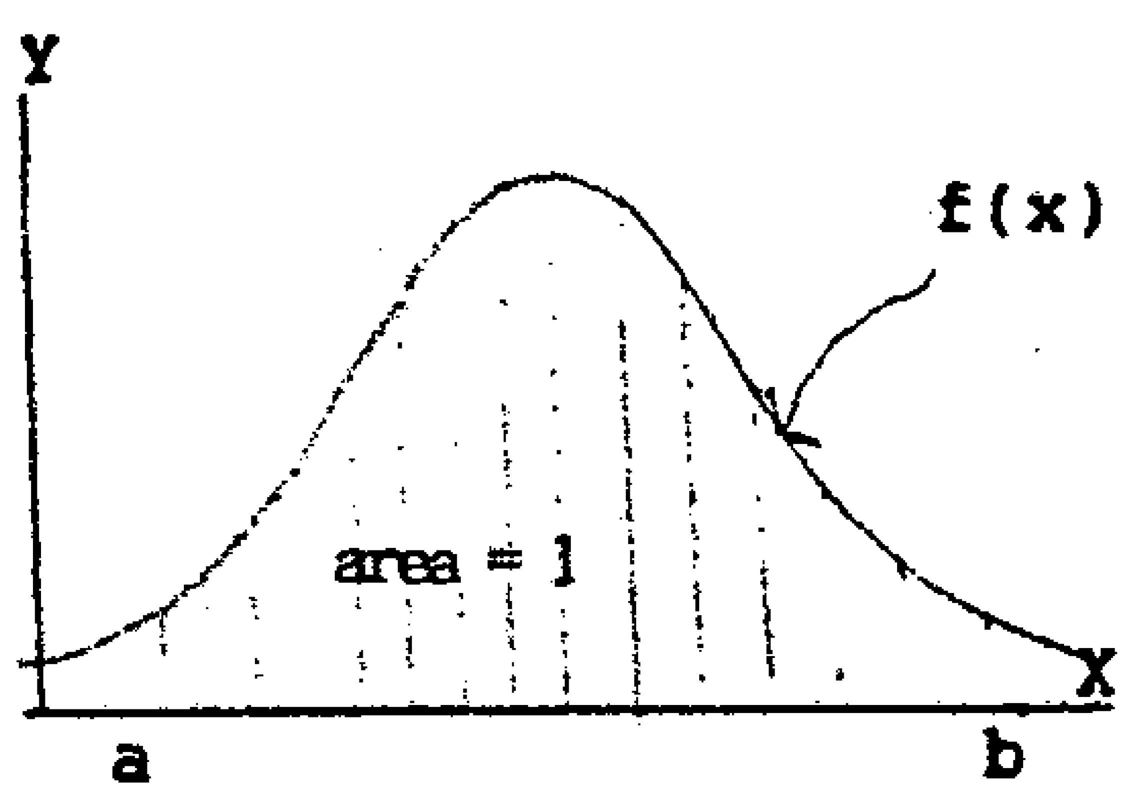
$$Rx: \quad a <= x <= b$$

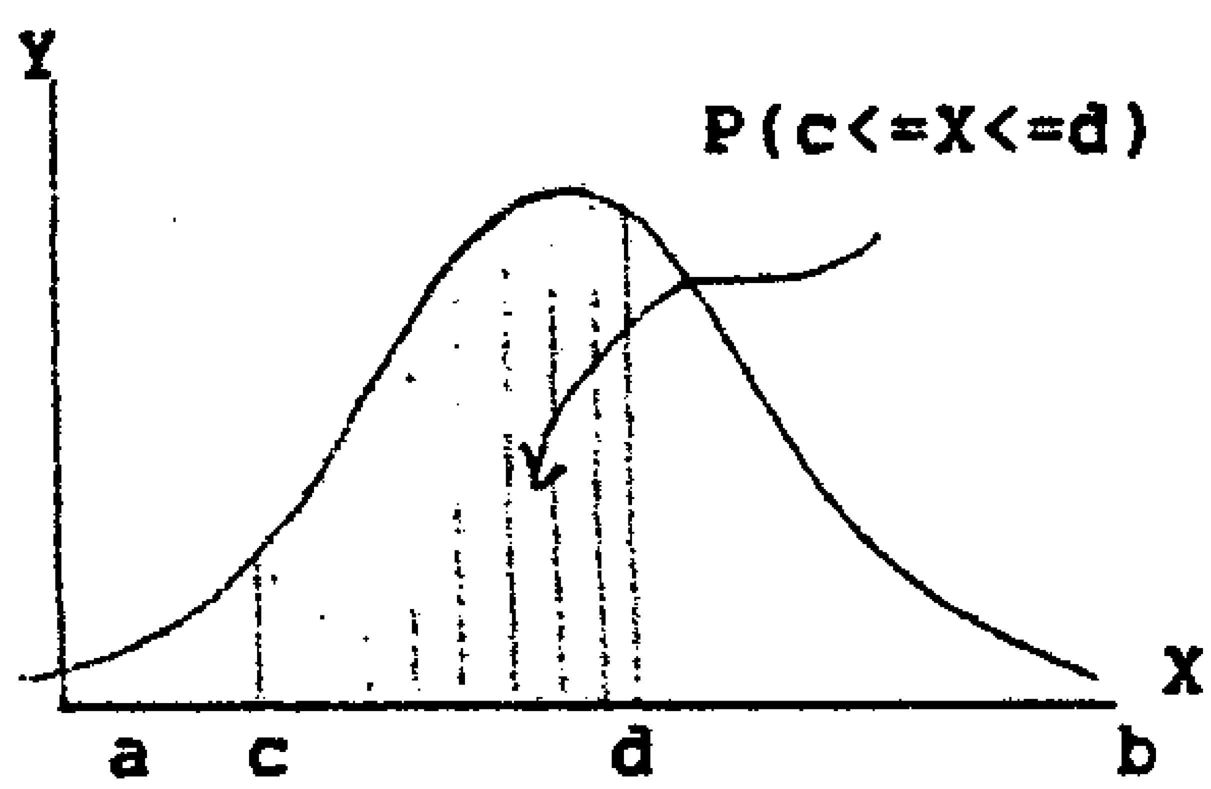
En consecuencia, entre cualquiera de dos valores asumidos en este intervalo existe un número infinito de resultados, para éstos la P(x=xi) pierde significado y es necesario definir la probabilidad a través de una función de Densidad de Probabilidad.

La función de Densidad de Probabilidad f(x) es una función no negativa que describe una curva, en términos de la variable X, en el eje "Y" del plano R^2, y satisface las siguientes condiciones:

- El area total delimitada por la curva f(x) y el eje "X" es igual a uno, por lo que se dice que describe un AREA DE PROBABILIDAD.

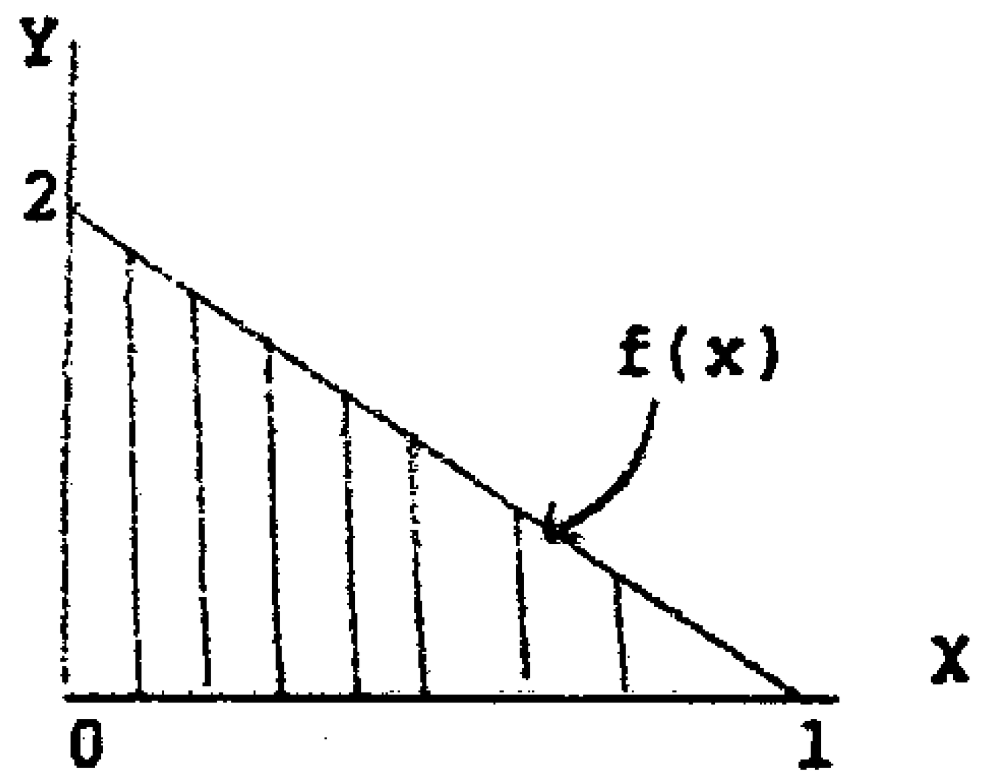
- La subarea bajo la curva limitada por ella, el eje "X" y las perpendiculares trazadas en dos puntos c y d cualquiera, que pertenezcan al intervalo [a,b] constituyen la probabilidad de que la variable X tome los valores entre c y d P(c <= x <= d)

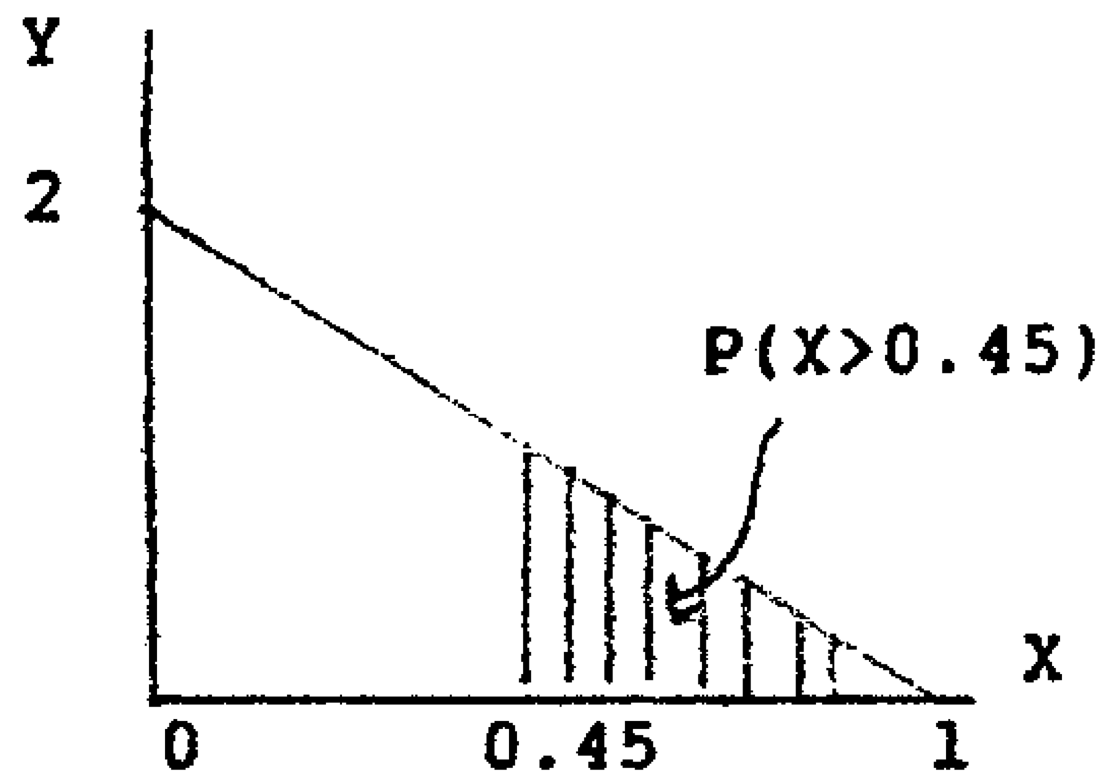




Por ejemplo:
El tiempo que transcurre entre una llamada de emergencia y otra en un hospital, puede considerarse como una variable aleatoria X, continua, en el intervalo de cero a una hora ($0 <= x <= 1$), por lo que puede asumir cualquiera de los infinitos puntos cronometrables en ese intervalo. Así entre una y otra llamada pueden pasar: 30 minutos, 15 segundos, 9 milésimas de segundo; o pueden pasar 30 minutos, 16 segundos una décima de segundo; etc.

Si la función de densidad de probabilidad de esta variable es:
$f(x) = 2 - 2x$, la curva la curva que describe es la siguiente
si x=0 f(x)= 2
si x=0.1 f(x)= 1.9
si x=0.3 f(x)= 1.4
si x=0.45 f(x)= 1.1
si x=1.0 f(x)= 0





Ya que f(x) limita con el eje X el Area de un triángulo rectángulo con base una unidad y altura dos unidades, tenemos:
area = 1/2 base * altura = 1/2*1*2=1 Representa un area de probabilidad.

Para calcular la probabilidad que el tiempo entre una llamada y otra sea mayor de 27 minutos (0.45 horas), se calcula el area bajo la curva como lo indica la figura.
area = 1/2 base * altura = 1/2 (1-0.45) * 1.1 = 0.3025
P(X > 0.45 horas) = 0.3025

Resuelva
*Cuál es la probabilidad que el tiempo entre una llamada y otra sea menor que 6 minutos?*______________________________
*Cuál es la probabilidad que sea mayor que 6 minutos pero menor que 18?*_____
____

2.9 DISTRIBUCION NORMAL

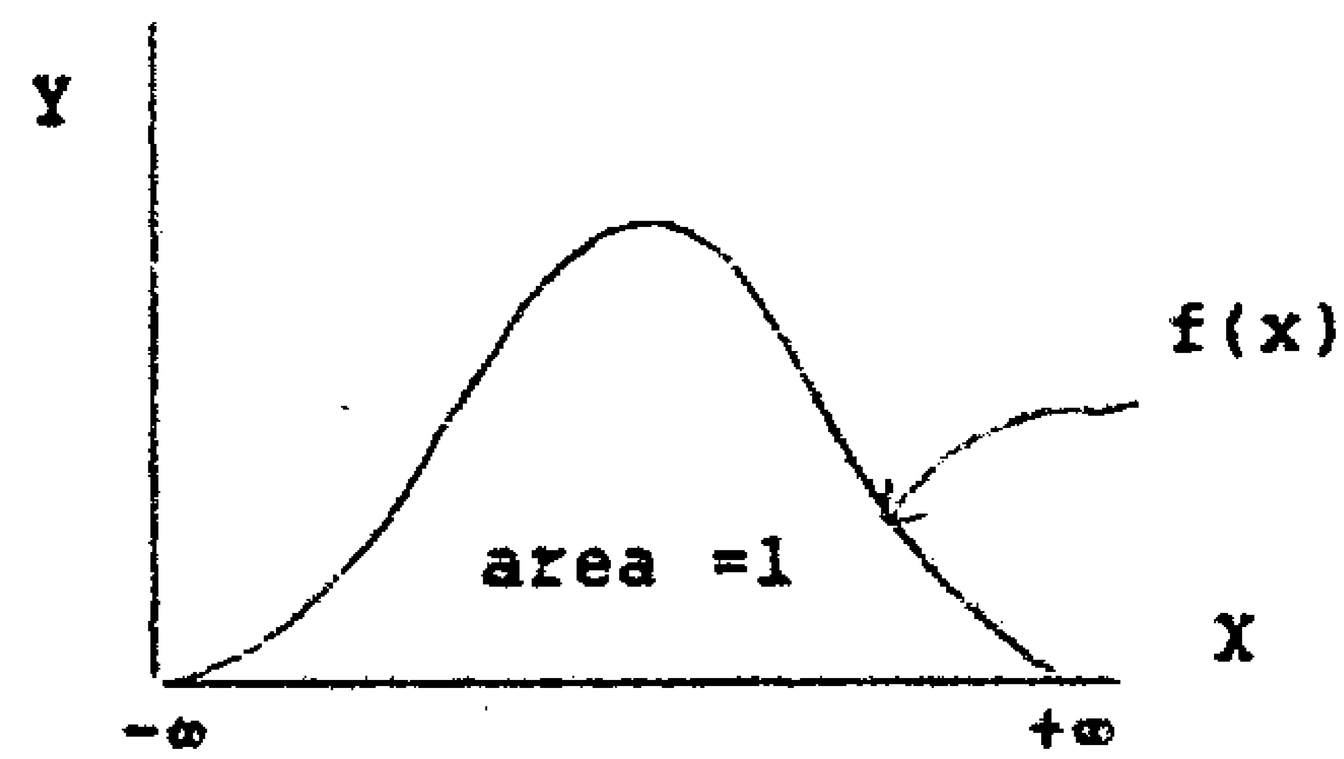
La distribución normal tiene una función de densidad:

$$f(x)= \quad (1\ /(\sqrt{(2\pi)}\sigma))*\ \exp((-1/2)*(x-\mu)/\sigma)$$

donde μ y σ se conocen como parámetros de la distribución, son la media y la desvición estándar de la variable.

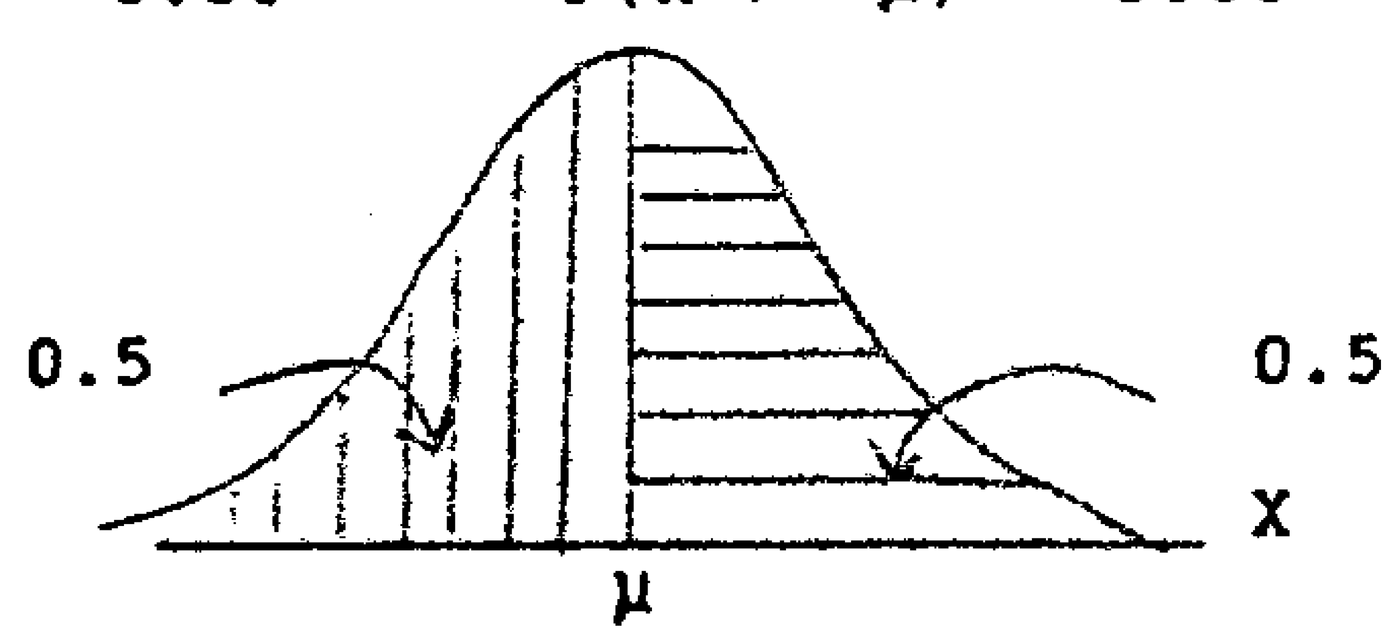
Recorrido de la distribución:

$$Rx:\ -\infty < x < \infty$$



CARACTERISTICAS

- Es simétrica en torno a la media μ, debido a ello el 50% del área está a la derecha de una perpendicular trazada en la media y el 50% restante hacia la izquierda.
  $P(x <= \mu) = 0.50 \qquad P(x >= \mu) = 0.50$

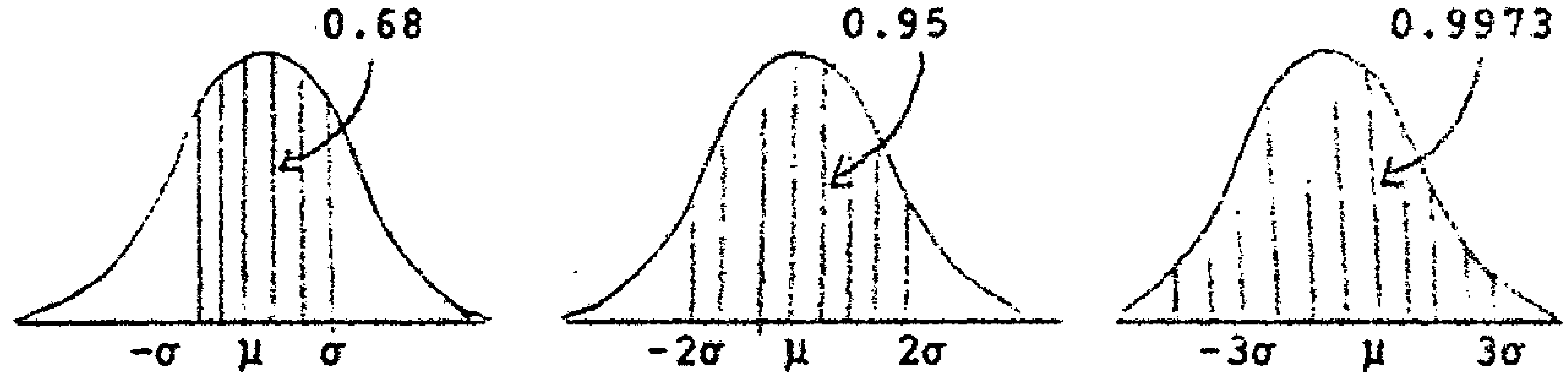


- La media, la mediana y la moda son iguales

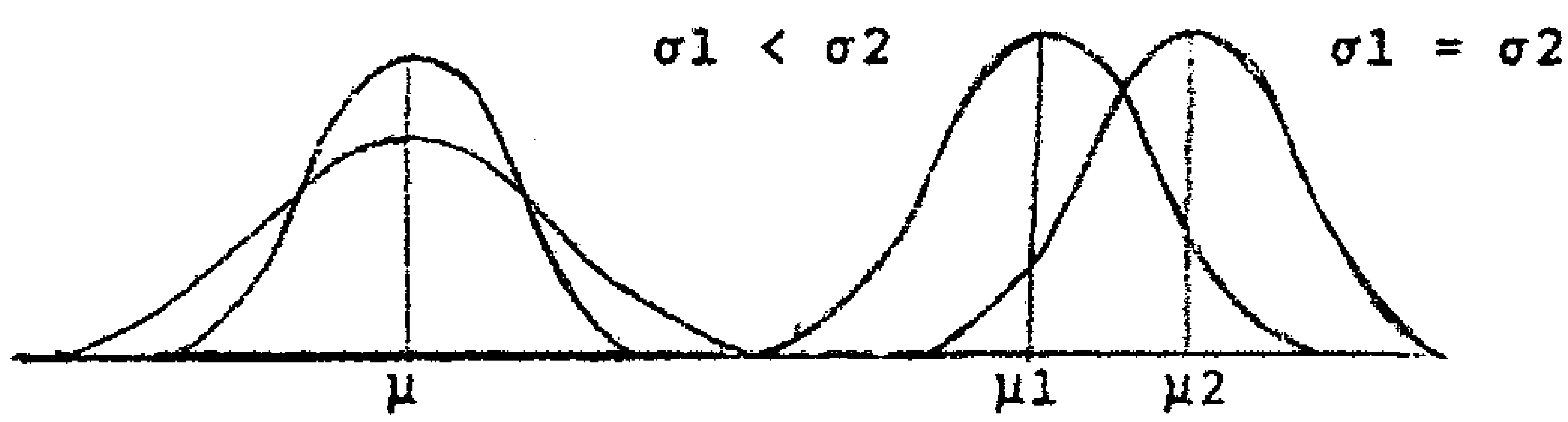
- Si se trazan perpendiculares a una distancia de 1σ a partir de la media en ambas direcciones, el área que encierra es aproximadamente del 68%.
  $P(\mu-\sigma <= x <= \mu+\sigma\ ) = 0.68$

- Si se trazan perpendicularea a una distancia 2σ a partir de μ se encierra aproximadamente el 95% del área.
  $P(\ \mu-2\sigma <= x <= \mu+2\sigma\ ) = 0.95$

- Extendiéndose a una distancia 3 σ se logrará encerrar aproximadamente el 99.73 % del área,
  $P(\ \mu-3\sigma <= x <= \mu+3\sigma\ ) = 0.9973$



- La distribución Normal esta determinada por los parámetros μ y σ, se especifica una distribución distinta para cada valor posible de μ y σ
  $-\infty < \mu < +\infty \qquad \sigma > 0$

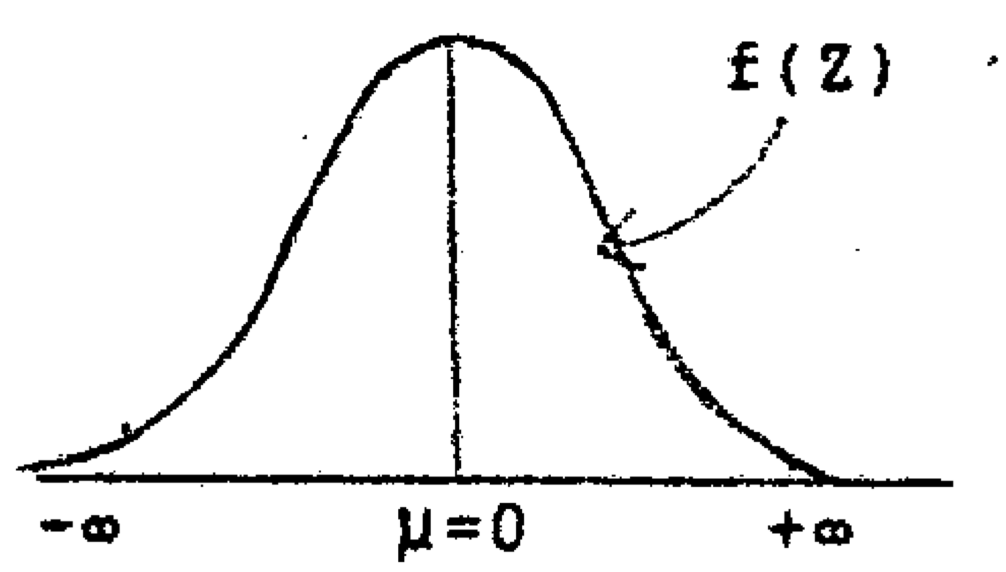


La distribución Normal es la distribución continua más importantes y se utiliza en una amplia gama de problemas. Su importancia radica en que hay un gran número de variables aleatorias que se observan en la naturaleza que poseen una distribución de frecuencias aproximadamente normal, además hay otras variables que por medio de transformaciones pueden convertirse en variables normales. Es fundamental en inferencia estadística ya que ciertas variables, para justificar pruebas, deben estar distribuidas en forma normal.

2.10 DISTRIBUCION NORMAL ESTANDAR

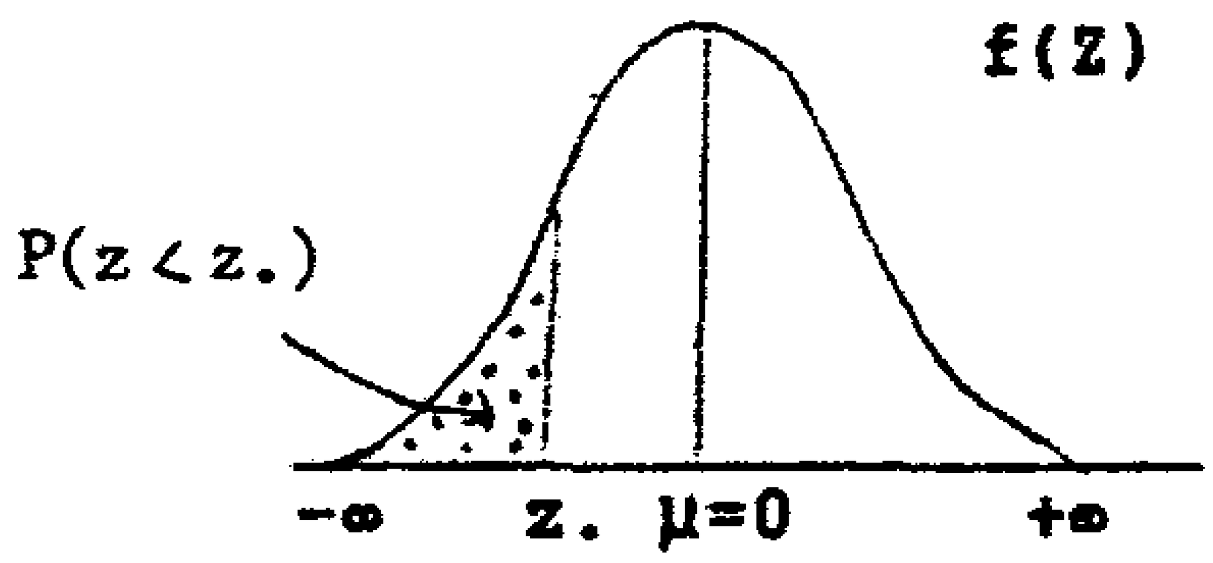
Z es una variable normal estandar si $\mu=0$ y $\sigma=1$.
Su función de densidad de probabilidad es:

$$f(z) = (1/\sqrt{(2\pi)}) \exp(-z^2/2)$$



Para calcular probabilidades de ocurrencia de una variable aleatoria normal estándar, como en cualquier variable continua, se debe encontrar el area bajo la curva de la función de densidad, este procedimiento se hace sencillo a través de las Tablas de la Distribución Normal Estándar, las cuales presentan las Areas Acumuladas (P(z<=z.)) para diversos valores de la variable.

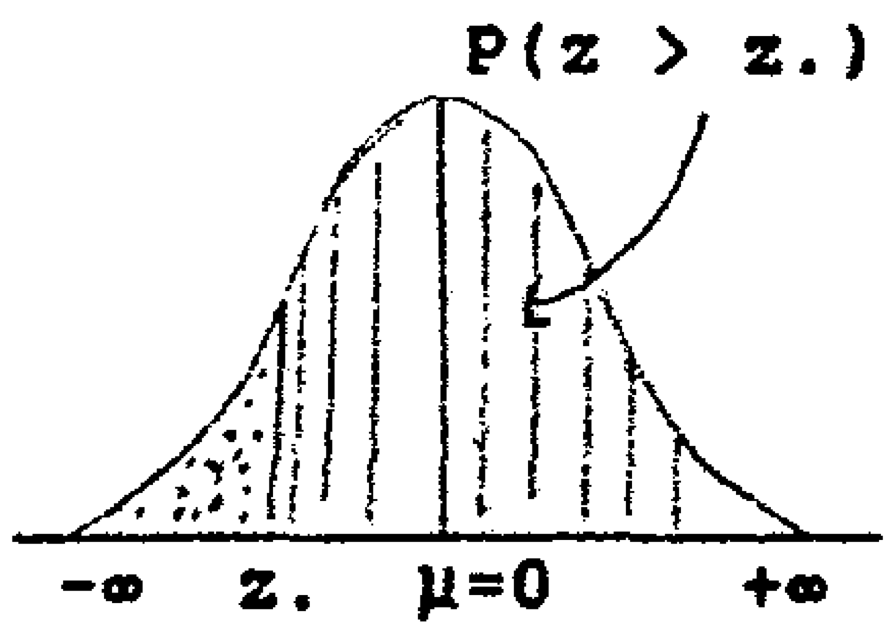
La tabla de la distribución Normal Estándar muestra la P(z <= z.) (z. es el valor tabulado).

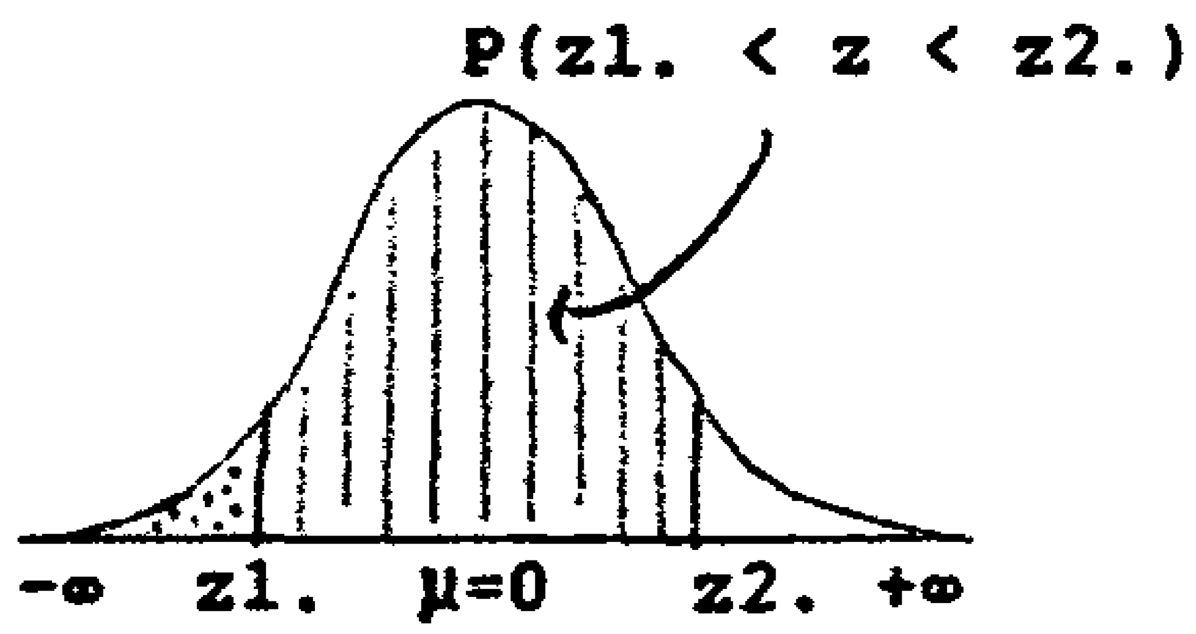


Se hace uso de esa probabilidad para calcular cualquier otro caso, asi:

P(z >= z.) = 1 - P(z <= z.)

P(z1. <= z <= z2.) = P(z <= z1.) - P(z <= z2.)





En el Apendice A se muestra una tabla de la distribucion Normal Estandar.

Areas bajo la curva Normal

Z.	00	01	02	03	04	05	06	07	08	09
-3.4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0003
.										
-2.1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
.										
-1.6	0548	0537	0526	0516	0505*	0495*	0485	0475	0465	0455
.										
-1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
.	.	.	.	.	.	.	.	.	.	.
0.5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
.										
1.9	9713	9719	9726	9732	9738	9744	9750*	9756	9761	9767
.										
2.0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Por ejemplo:

Si Z es una variable aleatoria con distribución Normal Estándar.

P(Z <= 0.5) = 0.695
P(Z >= -2,15) = 1- P(Z <= - 2.15) = 1 - 0.0158 = 0.9842
P(-1.02<=Z<=2.05)=P(Z<=2.50)-P(Z<=-1.02)= 0.9798-0.1539=0.8259

Resuelva:
P(Z <= 2.08)=
P(Z < - 0.29)=
P(Z >= 2.29)=
P(Z > -1.00)=
P(Z > 0)=
P(0.8 <= Z <= 3.00)=
P(-2.89 <= Z <= 0.25)=

Ahora si:

P(Z <= Z.) = 0.05   Z.= - 1.645
P(Z <= Z.) = 0.957   Z. = 1.96

Resuelva:

P(Z < Z.) = 0.75   Z.=
P(-Z. <= Z <= +Z.) = 0.98   ± Z.=

La distribución Normal Estándar es primordial en el cálculo de probabilidades de las variables aleatorias normales en general, ya que cualquier distribución normal X, con media μ y desviación estándar σ conocidas, pueden transformarse en una distribución Normal Estándar Z.

La transformación se hace a través de la función Z

$$Z = \frac{X - \mu}{\sigma}$$

que convierte los valores de X a valores de Z, al encontrar a cuántas desviaciones estándar se localiza el valor X respecto a su promedio.

Por ejemplo:

Un fisioterapeuta nota que las calificaciones obtenidas en cierta prueba de destreza manual están distribuidas en forma aproximadamente normal con media 10 y desviación estándar 2.5

(X = calificación de la prueba).
Si un individuo seleccionado al azar realiza la prueba, la probabilidad que obtenga una calificación mayor a 15 es:

Transformando la variable X a Variable Normal Estandar Z

$$Z = \frac{X - \mu}{\sigma} = \frac{15 - 10}{2.5} = 2$$

X = 15 se encuentra a 2σ a partir de μ

$P(X > 15) = P(Z > 2) = 1 - P(Z < 2) =$
$1 - 0.9772 = 0.0228$

Puede decirse que el 2.28% de las calificaciones de la prueba son superiores a 15 puntos.

Suponga que se sabe que los pesos de cierto grupo de individuos están normalmente distribuidos con media 70 kg. y desviación estándar de 12.5 kg (X=peso de los individuos).
Si se selecciona una persona al azar de este grupo; la probabilidad que su peso esté entre 50 y 85 kg. es

Transformando a la variable z,
si X = 50 $\quad Z = \dfrac{50 - 70}{12.5} = -1.6$

si X = 85 $\quad Z = \dfrac{85 - 70}{12.5} = 1.2$

entonces $P(50 <= X <= 85) = P (-1.6 <= Z <= 1.2) =$
$P(Z <= 1.2) - P(Z <= -1.6) = 0.8849 - 0.0548 = 0.8301$

Se puede decir que la proporción de personas que pesan entre 50 y 85 kg, es de 83.01%

Resuelva:
En un artículo de reciente publicación se ha informado que el ancho de los incisivos centrales superiores de los individuos de cierta población tienen una distribución normal con media 8.5 mm y una desviación estándar de 0.5mm
La variable X es ______________________________
Que es una variable normal con μ= ________ y σ = __________
Qué fracción de incisivos centrales alcanzan un ancho de más de 9.5 mm?

P(X)=

2.11 DISTRIBUCIONES USADAS EN PRUEBAS ESTADISTICAS

Las siguientes distribuciones son aplicadas en diferentes pruebas de Inferencia Estadistica que se expondrán posteriormente, por el momento se centra la atención en el manejo de las tablas para obtener áreas de probabilidad.

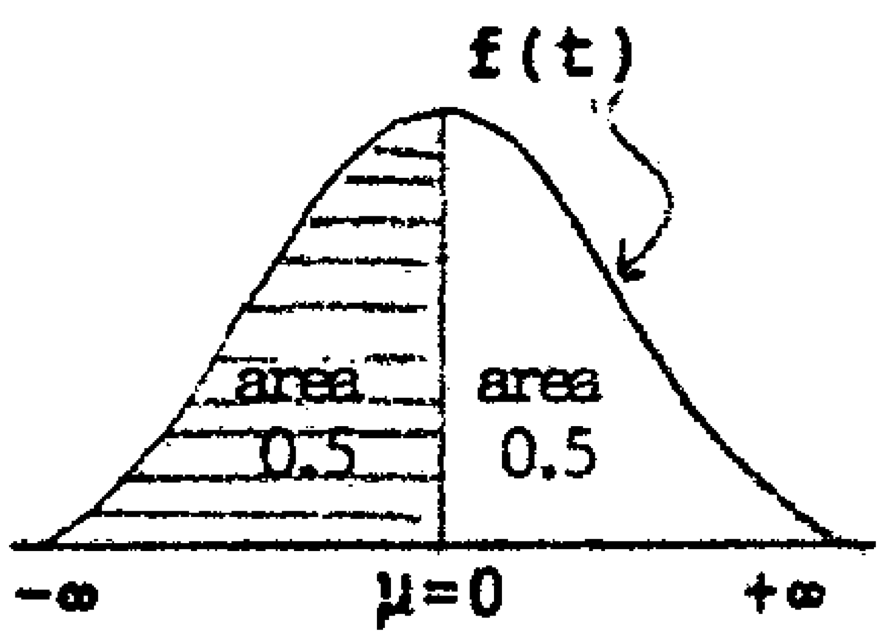
DISTRIBUCION T DE STUDENT

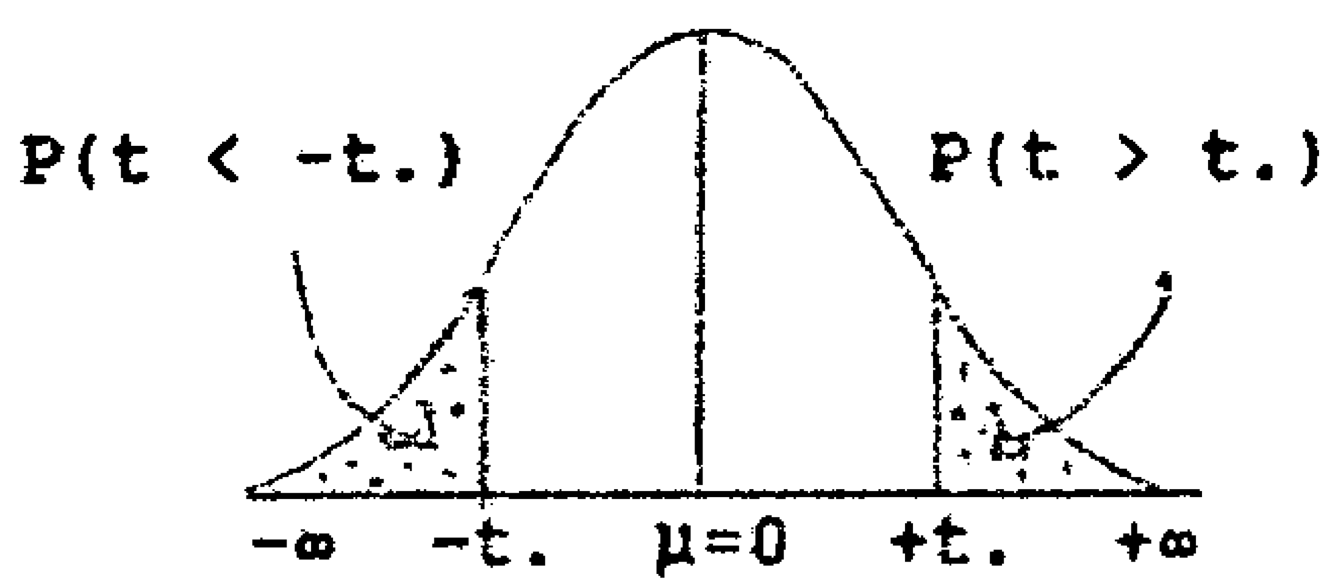
Una variable t con distribución de Student tiene una función de densidad de probabilidad

f(t) = Γ ((v+1)/2) /(Γ (v/2)* √(πv)) * (1-t^2/v)^((v+1)/2)
Γ= función gamma de x

v se conoce como grados de libertad y debe ser definido para determinar completamente la distribución ya que esta es diferente para cada valor posible de v (v>0).

Recorrido de t: $- \infty < t < + \infty$





CARACTERISTICA

t es simétrica a partir de cero

P (t <= 0) = 0.5

P (t >= 0) = 0.5

P (t <= -t.) = P(t >= t.)

P (t > t.) = 1 - P (t <= t.)

La tabla B del apéndice muestra los valores de t. para diferentes áreas acumuladas y un conjunto de valores de v, que permiten calcular probabilidades de sucesos específicos.

Percentiles de la distribucion t
P(t <= t.)

g.l	0.90	0.95	0.975	0.99	0.995
1	3.078	6.3138	12.7060	31.821	63.6570
2	1.886	2.9200	4.3027	6.965	9.9248
.	.	.	.	.	.
5	1.476	2.0150	2.5706	3.365	4.0321
.	.	.	.	.	.
10	1.372	1.8125	2.2281	2.769	3.1690
.	.	.	.	.	.
.	.	.	.	.	.

Por ejemplo

Si v = 5 grados de libertad

$P(t <= 2.015) = 0.95$

$P(t > 3.365) = 1 - P(t < 3.365) = 1 - 0.99 = 0.01$

$P(t <= -1.47) = P(t >= +1.47) = 1 - P(t < 1.47) =$
$1 - 0.9 = 0.10$

$P(-2.57<t<2.57) = P(t<=2.57)-P(t<=-2.57)=P(t<=2.57)-P(t>=2.57)=$
$P(t<=2.57)-(1 -P(<=2.57))=0.975-(1-0.975)=$
$0.975 - 0.025 = 0.95$

Resuelva:

Cuáles son las probabilidades de los siguientes suceso si v=15 grados de libertad.

$P(t> 1.345)=$

$P(t< -1.7613)=$

Además el valor de t. tal que:
$P(t < t.) = 0.95$ con v= 2 es t= 2.92

Y si se desea encontrar los valores de $\pm$ t. tales que:
$P(-t. <= t.) = 0.90$ y $P(t > t.) = \bar{P}(t < t.)$ con v=10,
$P(t>t.) = 0.05$ y $P(t <t.) = 0.95$
Entonces, $t.= \pm 1.8125$

Resuelva:
Cuál es el valor de t si:

$P(t < t.) = 0.90$ si v = 23

$P(t >t.) = 0.01$ si v = 8

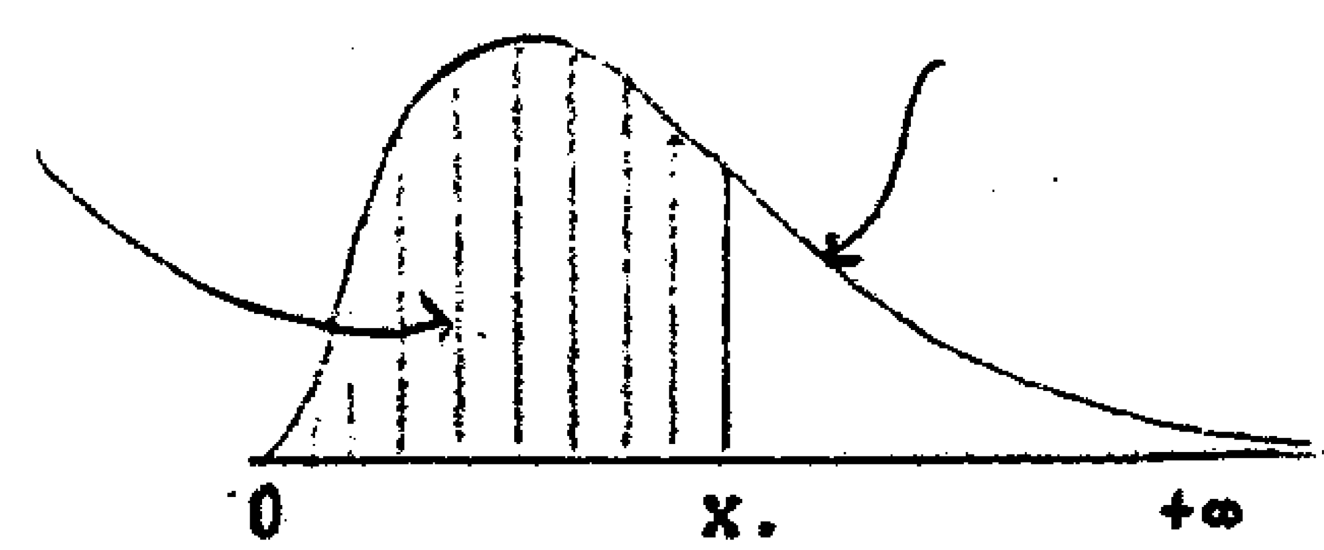
DISTRIBUCION CHICUADRADO

Una variable aleatoria X tiene una distribución Chi-cuadrado χ^2 si su función de densidad de probabilidad es:

$f(x) = (1/ (2^{(v/2)} * \Gamma(v/2-1)) * x^{(v/2-1)} * \exp(-x/2)$
Γ= función gamma de x

v seconoce como grados de libertad ydebe ser definido para determinar completamentela distribución ya queésta es diferente para cada uno de los valores posibles (v>0).

Recorrido de X: x>0



La tabla C del apéndice muestra las áreas acumuladas bajo la curva para un conjunto de valores de v, y pemite calcular probabilidades de sucesos especificas.

Percentiles de la distribución Chicuadrado

g.l	.005	.025	.05	.90	.95	.97	.99	.999
1	.00003	.0009	.0039	2.70	3.84	5.02	6.63	7.87
.	.	.	.	.	.	.	.	.
12	3.07	4.40	5.22	18.54	21.03	23.34	26.22	28.30
.	.	.	.	.	.	.	.	.
18	6.265	8.231	9.390	25.99	28.87	31.53	34.81	37.16
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

Por ejemplo:

Si v= 12 grados de libertad

$P(x \le 4.404) = 0.025$

$P(X \le 18.549) = 0.90$

$P(X > 26.217) = 1 - P(x < 26.217) = 1 - 0.99 = 0.01$

$P(5.22 \le X \le 21.026) = P(X \le 21.026) - P(X \le 5.22) = 0.95 - 0.05 = 0.90$

Resuelva:

Si v = 25 grados de libertad

$P(X \ge 13.12)$

$P(X \le 37.65)$

$P(14.61 \le X \le 44.31)$

Ahora si P(X <= X.) =0.99 con v= 18 grados de libertad, entonces X. = 34.805

Si P(X1. <= X <= X2.) = 0.90 y P(X <= X1.) = P(X >= X2.)
X1. = 9.39 X2.= 28.869

Resuelva

Si v= 11 grados de libertad

X.= ________ si P(X> X) = 0.99

X.= ________ si P(X < X.) = 0.975

X1.= ________ X2.= ________ si P(X1. <= X <= X2.) = 0.99 y P(X<= X1.) = P(X >= X2.)

DISTRIBUCION F

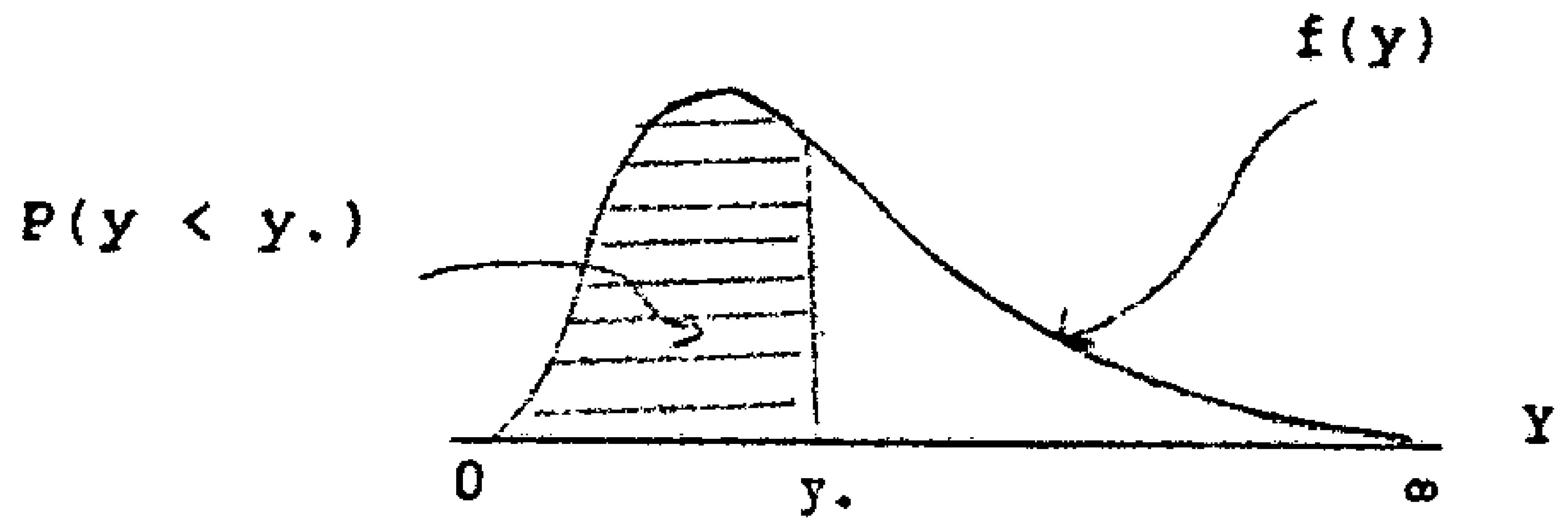
Una variable aleatoria Y tiene una distribución F si su función de densidad de probabilidad es

$$f(y) = \frac{[\,\Gamma((m+n)/2)^{*}\ m^{\wedge}(m/2)]\ ^{*}\ n\ ^{\wedge}(n/2)}{\Gamma(m/2)\ +\ \Gamma(n/2)}\ ^{*}\ \frac{y^{\wedge}\ ((m-2)/2)}{(n+my)^{\wedge}(m+n)/2)}$$

Γ= función Gama

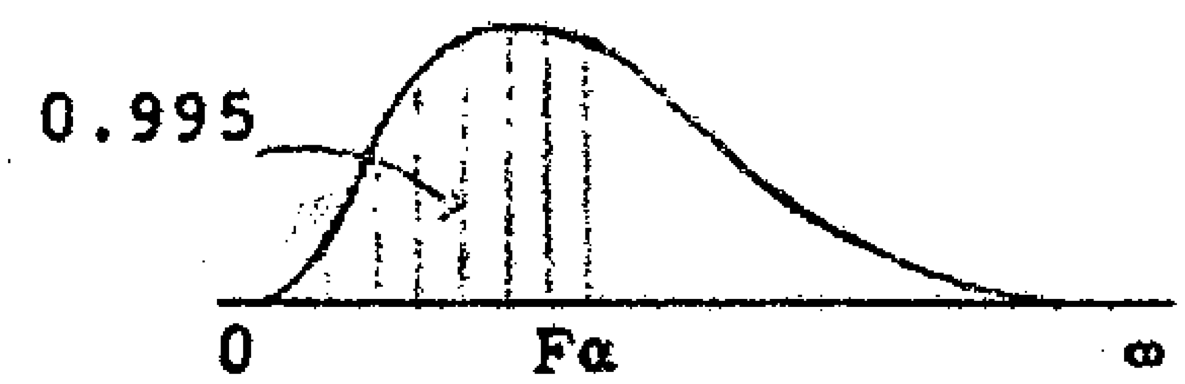
(m,n) se conocen como grados de libertad y deben ser definidos para determinar completamente la distribución.

Recorrido de Y: Y > 0



Las tablas del apéndice D muestran los valores de F para diferentes areas acumuladas.

Percentiles de la Distribución F
F(.995)



Grados de libertad denominador — Grados de libertad del numerador (m)

n	1	2	3	4	5	6	7	8	9
1	16212	20000	21615	22500	23056	23437	23715	23925	
.	.	.	.	.	.	.	.	.....	
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	.....	
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	.....	
.	.	.	.	.	.	.	.	.....	
.	.	.	.	.	.	.	.	.....	

Por ejemplo:

Si (m,n) = (4,4) grados de libertad

P(Y <= 23.15) = 0.995

P(Y <= 15.98) = 0.99

Resuelva:

P(Y <= 4.44) ______________________________

Ahora si P(Y <= Y.) = 0.975 con (m,n) = (5,6)
Y. = 5.99

Resuelva:

Si P(Y <= Y.) = 0.99
Y. = ____________________

2.12 EVALUACION

I. *De cada una de las siguientes Variables Aleatorias liste tres ejemplos:*

1. *Discretas:*
   a.______________________________________
   b.______________________________________
   c.______________________________________

2. *Continuas:*
   a.______________________________________
   b.______________________________________
   c.______________________________________

3. *Con distribución Binomial:*
   a.______________________________________
   b.______________________________________
   c.______________________________________

II. *Resuelva los siguientes problemas:*

1. *Una encuesta de Odontólogos demostró que el Número de Horas Semanales en contacto con pacientes estaba normalemente distribuido con una media de 33.5 horas y una desviación estandar de 2.5 horas. Qué proporción de esta población de Odontólogos dedica entre 30 y 35 horas a la semana a atender pacientes?*

2 *Suponga que se sabe que la probabilidad de recuperación de cierta enfermedad es 0.4. Si 15 personas contraen la enfermedad, cuál es la probabilidad que cuatro o más se recuperen?*

3. *El Presidente Regional de una Sociedad de Distrofia Muscular está interesado en enstimar la cantidad que cada simpatizante donará durante la reunión anual. Empleando los datos recolectados durante los últimos diez años ha calculado las siguientes probabilidades.*

Q prometidos	25	50	75	100	125
Probabilidad	0.45	0.25	0.15	0.10	0.05

Determine:
a. *La probabilidad que las donaciones sean superiores a Q75*
b. *En promedio, cuánto se espera que cada socio done?*
c. *Qué valor tiene la Varianza y la Desviación Estándar de la variable "Cantidad Donada por un Socio"?*

4. *Un embarque de cinco unidades odontológicas, incluye dos que tienen pequeñas manchas de pintura. Si una Agencia recibe tres de éstas unidades, indique los elementos del espacio muestral utilizando las letras B y N para Manchado y No Manchado. Asigne entonces a cada punto muestral un valor x de la variable aleatoria X que representa el Número de Unidades con manchas de pintura compradas por la Agencia.*

MODULO 3

DISTRIBUCIONES MUESTRALES

INTRODUCCION

En éste módulo se presentan los conceptos fundamentales de muestreo y se identifican los modelos que describen el comportamiento de los estadísticos, especialmente: Medias, Proporciones y Varianzas; modelos que servirán de base en la presentación de los conceptos de Estimación e Hipótesis en los capítulos siguientes.

OBJETIVOS

Al finalizar el módulo el lector estará en capacidad de:

1. Definir los términos población, muestra, parámetro y estadístico.

2. Explicar a que se refiere el término muestreo aleatorio simple.

3. Explicar que es una distribución muestral,

4. Identificar la media y el error estándar en la distribución muestral de: Medias, Proporciones, Varianzas.

5. Calcular la probabilidad de que un estadístico se incluya en un intervalo de valores dado.

3.1 CONCEPTOS FUNDAMENTALES

Se conoce como teoría del muestreo al estudio de las relaciones entre una población y una muestra extraída de la misma. Este concepto involucra dos definiciones importantes:

1. **POBLACION: es el total de observaciones concebibles de un tipo particular en el desarrollo de un experimento o investigación (N identifica al número de elementos de la población).**

Por ejemplo:

La población, en un lugar determinado, de individuos de 18 años, todos los individuos concebibles de esta edad en el lugar; la población de estudiantes fumadores de cierta universidad, todos los estudiantes de esa universidad que fuman al menos un cigarrillo diario.

En el desarrollo de una investigación se está interesado en estudiar una o mas características de la población, a esas características se les llaman VARIABLES.

Por ejemplo:

En la población de estudiantes fumadores, interesa conocer el número de cigarrillos que fuman diariamente y su relación con el sexo, las variables son el sexo y el número de cigarrillos.

Para cada variable existe una distribución de probabilidad discreta o continua que determina las posibilidades de ocurrencia de cada uno de sus posibles valores.

Se llaman PARAMETROS a los números que son utilizados para resumir la (s) distribución (es) de la (s) variable (es) en la población. Los parámetros más frecuentemente usados son:
μ Media,
σ^2 Varianza
σ Desviación Estándar,
p Proporción de casos favorables,
q Proporción de casos desfavorables.

Por ejemplo:

La poblacion de estudiantes fumadores se encuentran distribuidos en 80% varones y 20% mujeres (variable sexo), además en promedio fuman 7 cigarrillos al día (variable número de cigarrillos fumados diariamente).

2. **MUESTRA: Es un número limitado de observaciones de una población (el número de elementos en la muestra se representa por n).**

Por ejemplo:

De 1000 escolares inscritos en una escuela se seleccionaron 50 para evaluar la prevalencia de afecciones periodontales.
N= 1000 escolares y n = 50 escolares.

De la población de lechugas producidas en una hectaria determinada se seleccionaron 200 para identificar la presencia de cierto tipo de bacteria.
N= no definida, se puede asumir muy grande ($\rightarrow \infty$)
n= 200 lechugas.

Cualquier función de los elementos de una muestra se conoce como **ESTADISTICO.** Los estadísticos son números que describen el comportamiento de la muestra, entre ellos:

X' = Media aritmética,
S^2= Varianza,
S = Desviación Estándar,
P = Proporción de casos favorables,
Q = Proporción de casos desfavorables.

$X' = \Sigma xi/n$ i=1,2,3,4.....n xi= valor de la varible en cada elemento de la muestra

$S^2 = \Sigma((xi - X')^2/(n-1))$

$S = \sqrt{S^2}$

P= Número de casos favorables/n

Q= Número de casos desfavorables/n

3.2 SELECCION DE LA MUESTRA:

La selección de la muestra debe hacerse de tal modo que garantice que esta sea representativa de la población.
Hay diversas formas de extraer una muestra de la población, para los conceptos de muestreo que se estudiarán en éste curso se considera que son seleccionadas por el Método Muestreo Aleatorio Simple o al Azar.

Si se extrae una muestra de tamaño n de una población de tamaño N , de tal manera que cada elemento de la población tiene igual probabilidad de ser seleccionado para formar parte de la muestra, se dice que el procedimiento usado es el **MUESTREO ALEATORIO SIMPLE.**

Una manera de proceder para seleccionar una muestra aleatoria simple, consiste en enumerar todos los elementos de la población en órden correlativo; utilizando una secuencia de n números de la tabla de Números Aleatorio, se seleccionan como elementos de la muestra los que corresponden a éstos números.

El Apéndice E muestra una tabla de Numeros Aleatorios.

Por ejemplo

Si tenemos una población de 900 individuos y es necesario seleccionar una muestra de 10. De la tabla se toma una columana al azar, en este caso los 3 primeros dígitos de la misma por ser el número mayor 900, y los 10 primeros números que aparecen en ella menores o iguales que 900 serán los elementos que forman parte de la muestra. Si en este caso seleccionamos la columna 3, los elementos que forman la muestra, son los identificados con los números:

145, 116, 614, 221, 168, 094, 897, 246, 869, 476

Recuerde que la población está numerada de 1 a 900.

3.3 TEORIA DEL MUESTREO

La teoría del muestreo involucra los siguientes aspectos:

1. Determinar el probable comportamiento de una muestra, si se conoce la distribución de la población de donde fué extraida, esto es, permite hacer predicciones sobre los valores que se esperan para los estadísticos en las muestras observadas.

2. Inferir acerca del comportamiento una población basándose en la información contenida en la muestra.

3. Identificar si las diferencias entre dos muestras son debidas a la aleatoriedad o son diferencias que tienen causas asignables, comunmente llamadas diferencias significativas.

4. Explicar la relación entre variables y sus repercuciones en la población.

El primero de los aspectos se trata en el siguiente punto de Distribuciones Muestrales.

3.4 DISTIBUCIONES MUESTRALES

Los Estadísticos son cantidades cuyos valores se calculan una vez que se ha tomado la muestra, dado que las muestras son seleccionadas al azar, los estadísticos son distintos para cada una de ellas, pues sus elementos son diferentes, por tanto se comportan como una variable aleatoria que tiene su propia distribución de probabilidades.

La distribución de probabilidad de los valores que puede tomar un Estadístico, calculado a partir de muestras del mismo tamaño y extraidas al azar de la misma población, se conoce como Distribución Muestral del Estadístico.

La aplicación más frecuente de la distribución muestral es calcular la probababilidad de obtener una muestra con un estadístico de magnitud determinada. En las Aplicaciones de las distribuciones de estadísticos específicos se tratará este problema.

PLANTEAMIENTO GENERAL DE LAS DISTRIBUCIONES MUESTRALES

Considere todas las posibles muestras de tamaño n que pueden extraerse de una población con parámetro ϕ ($\phi = \mu, \sigma, p$, etc.). Si en cada una de éstas mustras se calcula un estadístico E (E=X,S,P, etc), que varía de una muestra a otra, se tendrá el conjunto de todos los posibles valores de E:

Muestra	Observaciones				Estadístico
1	X11	X12	X13	X1n	E1
2	X21	X22	X23	X2n	E2
i	Xi1	Xi2	Xi3	Xin	Ei
j	Xj1	Xj2	Xj3	Xj4	Ej
.	...	...	...	...	..
∞	...	...	...	...	..

donde Ei = Ej ó Ei $\neq$ Ej por lo que pueden existir k valores diferentes para E.

Con éste conjunto se puede construir una distribución muestral el estadístico y determinar sus parámetros, media, varianza,y la raíz cuadrada de la varianza conocida como Error Estándar.

DISTRIBUCION MUESTRAL DE E:

Estadístico	Frecuencia
Ea	fa
Eb	fb
..	..
..	..
Ek	fk

$E' = \mu(E)$ = media del estadístico

$V(E) = \sigma^2(E)$ = varianza del estadístico

$\sqrt{V(E)} = \sigma(E)$ = Error Estándar del Estadístico.

Por ejemplo:

Suponga una población de 5 elementos que consta de las edades de 5 niños pacientes externos de un centro de atención odontológica. Las edades son las siguientes: 6,8,10,12,14

La media de la población $\mu= 10$
$\mu= (6+8+10+12+14)/5$

La varianza de la población es:
$\sigma^2=((6-8)^2+(8-8)^2+(10-8)^2+(12-8)^2+(14-8)^2)/5= 8$

La desviacón Estándar de la población es $\sigma=2.828$

Si de esta población se extraen todas las muestras posibles de tamaño n=2, y se calcula el estadístico Media Aritmética (X')

Muestras Posibles

	6	8	10	12	14
6	-	6,8	6,10	6,12	6,14
8	8,6	-	8,10	8,12	8,14
10	10,6	10,8	-	10,12	10,14
12	12,6	12,8	12,10	-	12,14
14	14,6	14,8	14,10	14,12	-

Muestra	Estadístico X'
6,8	7
6,10	8
6,12	9
6,14	10
8,10	9
8,12	10
8,14	11
10,12	11
10,14	12
12,14	13

Distribución de frecuencias del Estadístico

Xi'	frecuencia fi	Xi'*fi	$(Xi'-\mu)^2$	$(Xi'-\mu)^2*fi$
7	1	7	9	9
8	1	8	4	4
9	2	18	1	2
10	2	20	0	0
11	2	22	1	2
12	1	12	4	9
13	1	13	9	9
	Σ 10	100		30

$\mu (X') =(\Sigma Xi*fi)/n = 100/10 = 10$

$\sigma(X') =\sqrt{((\Sigma(Xi-\mu)^2*fi)/10)} = \sqrt{(30/10)} = \sqrt{3} = 1.73$

Se puede decir que para el estadístico Media Aritmética (X') se tiene una distribución muestral con parámetros:

$\mu(X') = 10$ $\sigma(X') = 1.73$

Al seleccionar una muestra de esta poblacion, no se sabe con exactitud que valor tendrá el estadístico X', puede ser cualquier valor del 7 al 13, X' es una variable aleatoria, pero se tiene una probabilidad 2/10 que tenga un valor de 11 o de 1/10 que tenga un valor de 12, etc.

3.5 DISTRIBUCION MUESTRAL DE MEDIAS

Si de una población con media μ y varianza σ^2, se seleccionan todas las posibles muestras de tamaño n y se calcula el estadístico promedio aritmético (X') se obtiene una:

DISTRIBUCION MUESTRAL DE MEDIAS

Media $\mu(X') = \mu$

Varianza $\sigma^2(X') = \sigma^2/n$ si la población es finita
$\sigma^2(X') = (\sigma^2/n)*((N-n)/(N-1))$ si la población es infinita

Error Estándar $\sigma(X') = \sigma/\sqrt{n}$ si la población es finita
$\sigma(X') = (\sigma/\sqrt{n})*\sqrt{((N-n)/(N-1))}$ si la poblacion es infinita.

Al seleccionar una muestra de tamaño n de esta población el estadístico resultante es una Variable Aleatoria que cuando la población tiene una distribución Normal, la forma de la distribución muestral también es normal. Cuando la población no es normal, el Teorema de Límite Central (10) garantiza que la distribución muestral de medias se acerca a la distribución normal a medida que crece el tamaño de la muestra, por lo que generalmente se aplica ésta distribución cuando n >= 30.

Para calcular la probabilidad de que la variable Media Aritmética (X') esté en un intervalo de valores determinado, se utiliza la transformación a la variable Z:

$$Z = \frac{X' - \mu}{\sigma(X')}$$

que tiene una distribución normal estándar.

PROPIEDAD DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
Biblioteca Central

APLICACIONES:

Supóngase que en una gran población formada por personas adultas, la longitud craneal está distribuida en forma casi normal con media 185.6 mm y desviación estándar 12.7 mm.
La probabilidad que al seleccionar una muestra de tamaño 10 de esta población, el promedio de la longitud craneal en la muestra tenga una media mayor que 190 mm es:

Datos:
$n=10 \qquad \mu= 185.6 \qquad \sigma= 12.7$

Planteamiento:
La variable aleatoria X' es el promedio de la longitud craneal en una muestra de 10 individuos.
La variable tiene una distribución muestral con media:
$\mu(X')= 185.6$ y $\sigma(X')= 12.7/\sqrt{10} = 4.02$

Solución:

$$P(X'>190)= P\left(Z > \frac{190 - 185.6}{4.02} \right) = P(Z >= 1.09) = 0.1379$$

La probabilidad que la media de la muestra sea superior a 190 mm es 0.1379

La media y la desviación estándar de la concentración de hierro en el suero de hombres sanos, son respectivamente 120 y 15 microgramos por 100 ml. La probabilidad de que una muestra seleccionada al azar de 50 hombres proporcione una media entre 115 y 125 microgramos por 100 ml es:

Nota: No se especifica la forma de la distribución de la población, pero dado a que se toma una muestra grande (n>30) se considera la Distribución Muestral de Medias aproximadamente Normal.

Datos:
$n= 50 \qquad \mu=120 \qquad \sigma= 15$

Planteamiento:
La variable aleatoria (X') es el promedio de las concentraciones de hierro en el suero en una muestra de 50 hombres sanos y tiene una distribución muestral con:
$\mu(X')= 120 \qquad \sigma(X') = 15/\sqrt{50} = 2.12$

Solución:

$$P(115 <= X'<= 125) = P\left(\frac{115 - 120}{2.12} <= Z <= \frac{125 - 120}{2.12} \right) =$$

$$P(-2.36 <= Z <= 2.36) = P(Z<= 2.30) - P(Z<= -2.36)= 0.9909 - 0.0091 = 0.9818$$

La probabilidad que la concentración de hierro promedio de la muestra se encuentre dentro del intervalo especificado es 0.9818

En cierto hospital trabajan 300 empleados de cierto tipo, los salarios por hora están distribuidos en forma casi normal con media y desviación estándar Q22.5 y Q2.25 respectivamente. Si se selecciona una muestra de tamaño 16 de esta población de empleados, la probabilidad que el salario medio por hora de la muestra sea menor que Q21.00 es:

Datos:
$n = 16$ $N = 300$ $\mu = 22.5$ $\sigma = 2.25$

Planteamiento;
La variable aleatoria X' es el salario promedio por hora en una muestra de 16 empleados y tiene una distribución muestral con media:
$\mu(X') = 22.5$
$\sigma(X') = (2.25/\sqrt{16})*(\sqrt{((300-16)/(300-1))} = 0.5482$

Solución:
$P(X' < 21) = P\left(Z < \frac{21 - 22.5}{0.5482}\right) = P(Z < 2.736) = 0.0031$
La probabilidad que el salario medio por hora sea menor que Q21.00 es 0.0031

Resuelva:

Supónga que se sabe que el tiempo de respuesta de un estímulo particular en individuos sanos, es una variable con distribución normal con media 15 segundos y varianza 16 segundos². ¿Cuál es la probabilidad que una muestra al azar de 16 individuos tenga un tiempo medio de respuesta de 12 segundos o más?

Datos:
n= *μ=* *σ=* *Población=*

Planteamiento:
X'=

Solución:
P(X')=

Para cierta población con lesiones musculares, el número medio de días de incapacidad es de 5.4 con una desviación estándar de 2.8 días. ¿Cuál es la probabilidad que una muestra al azar de tamaño 49 de dicha población tenga un promedio de días de incapacidad entre 4 y 6.

Datos:
n= *μ=* *σ^2=* *σ=*

Población:
Planteamiento:
X'=

Solución:
P(X')=

3.6 DISTRIBUCION MUESTRAL DE DIFERENCIA DE MEDIAS (X1'-X2')

Si de dos poblaciones con parámetros $\mu 1$, $\sigma 1$, $\mu 2$, $\sigma 2$, se toman muestras de tamaño n1 para la primera y n2 para la segunda, luego se computan los estadísticos X1' y X2'. Al tomar todas las posibles combinaciones de estas muestras y obtener las diferencias (X1'-X2') se obtiene una:

DISTRIBUCION MUESTRAL DE DIFERENCIA DE MEDIAS

Media: $\mu(X1' - X2') = \mu 1 - \mu 2$

Varianza: $\sigma^2(X1' - X2') = (\sigma 1^2/n1) + (\sigma 2^2/n2)$

Error Estándar: $\sigma(X1' - X2') = \sqrt{((\sigma 1^2/n1)+(\sigma 2^2/n2))}$

Al seleccionar muestras de ambas poblaciones de tamaño n1 y n2, el estadístico resultante d = (X1'-X2') es una variable aleatoria que, cuando la población está distribuida normalmente o cuando a pesar de tener una distribución desconocida o no normal las muestras son grandes (n>30), tiene una distribución de forma aproximadamente normal.

Para calcular la probabilidad que el estadístico d, (X1'-X2') esté en un intervalo de valores determinado se utiliza la transformación a la variable Z:

$$Z = \frac{(X1' - X2') - (\mu 1 - \mu 2)}{\sqrt{((\sigma 1^2/n1)+(\sigma 2^2/n2))}}$$

que tiene una distribución normal estándar.

APLICACIONES

Se ha establecido que, para cierto tipo de paciente, la duración media de una visita al odontólogo es de 45 minutos con una desviación estándar de 15 minutos, para un segundo tipo de pacientes, la media es de 30 minutos con una desviación de 20 minutos. Si un investigador selecciona al azar 35 pacientes del primer tipo y 40 del segundo, la probabilidad que la diferencia de las medias del tiempo de visita entre los dos grupos sea de 20 minutos o más es:

Datos
$\mu 1 = 45$ $\sigma 1 = 15$ $n1 = 35$ $\mu 2 = 30$ $\sigma 2 = 20$ $n2 = 40$

Planteamiento:
(X1'-X2') es la diferencia entre los promedios del tiempo de visitas en las muestras n1 y n2 y tiene una distribución muestral con:

$\mu(X1'-X2') = 45-30 = 15$

$\sigma(X1'-X2') = \sqrt{((15^2/35)+(20^2/40))} = 4.05$

Solución:

$P(X1'-X2' >= 20) = P(Z >= \frac{20 - 15}{4.05}) = P(Z >= 1.23) =$

$1- P(Z <= 1.23) = 1- 0.8907= 0.1093$

La probabilidad que la diferencia entre los tiempos promedios de visita de las dos muestras sea al menos 20 minutos es 0.1093.

Un inveestigador cree que los niveles de vitamina A en el hígado de dos poblaciones de personas muestran una distribución normal. Las varianzas de las dos poblaciones son las siguientes $\sigma 1^2= 19600$ $\sigma 2^2= 8100$
La probabilidad de que una muestra al azar de tamaño 15 para la primera población y de 10 para la segunda, proporcionen un valor de (X1'-X2') menor o igual a 50, si no existe diferencia entre las medias poblacionales es.

Datos
$\mu 1-\mu 2=0$ No hay diferencia entre las medias poblacionales
$\sigma 1^2= 19600$ n1= 15
$\sigma 2^2= 8100$ n2= 10

Planteamiento:
(X1'-X2') es la diferencia entre los niveles de vitamina A entre las dos muestras y tiene una distribución muestral con:
$\mu(X1'-X2')=0$
$\sigma(X1'-X2') = \sqrt{((19600/15)+(8100/10))} = 46.007$

Solución:

$P((X1'-X2') <50)=P(Z <= \frac{50 - 0}{46}) = P(Z <= 1.09)= 0.8621$

La probabilidad que la diferencia entre los niveles de vitamina A entre las muestras de las dos poblaciones sea menor o igual a 50 es 0.8621.

Resuelva:

En un estudio de gastos anuales por familia en el cuidado bucal, se investigan dos poblaciones con los siguientes resultados:
Población 1: n1=40 X1'=346
Población 2: n2=35 X2'=300
Si se sabe que las dos poblaciones tienen varianzas $\sigma 1^2=2800$ y $\sigma 2^2=3250$ ¿Cuál es la probabilidad de obtener resultados de las muestras tan extremos como los dados anteriormente, si se supone que no existe diferencia entre las medias de ambas poblaciones?

Datos:
$\mu 1-\mu 2=$
$\sigma 1^2=$ *$\sigma 2^2=$*
n1= *n2=*
X1'= *X2'=*

Planteamiento:
(X1'-X2') es

con $\mu(X1'-X2')=$
$\sigma(X1'-X2')=$

Solución:
P((X1'-X2'))=

3.7 DISTRIBUCION MUESTRAL DE PROPORCIONES

Considere una población que tiene probabilidad de ocurrencia de un suceso (conocido como éxito) igual a p y la probabilidad de no ocurrencia del suceso (conocido como fracaso) igual a q (q=1-p), al seleccionar todas las posibles muestras de tamaño n de esa población, contar el número de éxitos que aparecen y calcular el estadístico P, proporción de éxitos en la muestra, se obtiene una:

DISTRIBUCION MUESTRAL DE PROPORCIONES

Media $\mu(P)= p$

Varianza $\sigma^2(P)= pq/n$ para poblaciones infinitas
$\sigma^2(P)= (pq/n)((N-n)/(N-1))$ poblaciones finitas

Error Estándar $\sigma(P)= \surd(pq/n)$ para poblaciones finitas
$\sigma(P)= \surd((pq/n)((N-n)/(N-1)))$ poblaciones infinitas.

Al seleccionar una muestra de tamaño n de ésta población el estadístico P resultante es una variable aleatoria binomial (ver módulo anterior), pero cuando n es grande (n>30) y además np y nq son mayores que cinco, la distribución se aproxima a la distribución Normal.

En este caso para calcular la probabilidad de que P esté en un intervalo de valores definido, se utiliza la transformación a la variable Z:

$$Z = \frac{P - p}{\sigma(P)}$$

que tiene una distribución Normal Estandar.

APLICACIONES:

Supónga que se sabe que en cierta población de mujeres que se encuentran en su tercer trimestre de embarazo, el 90% ha tenido algún cuidado prenatal. Si se extrae de esta población una muestra de tamaño 200, la probabilidad de que la proporción de la muestra que haya tenido algún cuidado prenatal (P) sea mayor que 85% es:

Datos:
n= 200 p= 0.90 (probabilidad de cuidado prenatal)
q= 0.10 (pobabilidad de no cuidado prenatal)

Planteamiento:
La variable aleatoria P es la proporción de la muestra que ha tenido cuidado prenatal y tiene una distribución muestral con media:
$\mu(P) = 0.90$ $\sigma(P) = \sqrt{(0.9*0.1/200)} = 0.0212$

Solución
$$P(P>0.85) = P\left(Z > \frac{0.85 - 0.90}{0.0212}\right) = P(Z > -2{,}36) = 0.91$$
La probabilidad de la muestra presente una proporción de mujeres que han tenido cuidad prental superior al 85% es 0.91.

Resuelva:

Si en una población de adultos el 15 % está sometidos a algún tipo de dieta, ¿ Cuál es la probabilidad de que en una muestra al azar de tamaño 100 se encuentre una proporción de aquellos que se encuentran a dieta entre 0.10 y 0.20 ?

Datos:
p= *q=* *n=*

Planteamiento:
P es :

μ(P)= *σ(P)=*

Solución:
P(P) =

3.8 DISTRIBUCION MUESTRAL DE DIFERENCIA DE PROPORCIONES (P1 - P2)

Si de dos poblaciones distribuidas binomialmente con parámetros p1, q1, p2, q2; se seleccionan todas las posibles muestras de tamaño n1,n2 para la primera y segunda población respectivamente y se computan los estadísticos P1 y P2, al tomar todas las posibles combinaciones de éstas muestras y obtener las diferencias (P1-P2) se obtiene una:

DISTRIBUCION MUESTRAL DE DIFERENCIAS DE PROPORCIONES

Media: $\mu(P1-P2) = p1-p2$
Varianza: $\sigma^2(P1-P2) = (p1q1/n1) + (p1q2/n2)$
Error Estándar: $\sigma(P1-P2) = \sqrt{((p1q1/n1)+(p2q2/n2))}$

Al seleccionar muestras aleatorias de ambas poblaciones de tamaño n1 y n2, el estadístico resultante (P1-P2) es una variable aleatoria que cuando n1 y n2 son mayores que 30 tiene una distribución aproximadamente normal.

Para calcular la probabilidad de que el estadístico (P1-P2) esté dentro de un intervalo de valores especificado se utiliza la transformación a la variable Z:

$$Z = \frac{(P1-P2) - (p1-p2)}{\sqrt{(p1q1/n1) + (p2q2/n2)}}$$

que tiene una distribución normal estándar.

APLICACIONES:

Suponga que la proporción de personas que consumen drogas ilegales en una población es de 50%; mientras que en otra población la proporción es 33% . La probabilidad de que en muestras de tamaño 100, extraídas de cada una de las poblaciones, la diferencia entre P1 y P2 sea mayor que 0.30 es:

Datos:

p1= 0.50	q1= 0.50	n1= 100
p2= 0.33	q2= 0.67	n2= 100

Planteamiento:

(P1-P2) es la variable, diferencia de la proporción de personas que consumen drogas ilegales en las muestras tomadas de las poblaciones 1 y 2 y tiene una distribución muestral con:

$\mu(P1-P2) = p1 - p2 = 0.17$

$\sigma(P1-P2) = \sqrt{(0.33*0.67/100)+(0.50*0.50/100)} = 0.0686$

Solución:

$$P((P1-P2)>0.3) = P\left(Z > \frac{0.3 - 0.17}{0.0686}\right) = P(Z > 1.89) = 0.0294$$

La probabilidad que la diferencia entre las proporciones muestrales sea mayor que 0.30 es 0.0294.

Resuelva:

En cierta población de niños con retraso mental se se sabe que la proporción de lo que son hiperactivos es de 0.40; si se extrajo una muestra al azar de tamaño 120 de esta población y otra de tamaño 100 de una población diferente de niños con el mismo problema; si la porporción de niños hiperactivos es la misma en ambas poblaciones, ¿ Cuál es la probabilidad de que las muestras proporcionen una diferencia (P1-P2) de 0.16 o menos?

Datos:

p1 =	*q1=*	*n1=*
p2=	*q2=*	*n2=*

Planteamiento:
(P1-P2) es:

y tiene una distribución muestral con:

μ (P1-P2) =
σ (P1-P2) =

Solución:
P((P1-P2))

3.9 DISTRIBUCION MUESTRAL DE VARIANZAS (S^2)

Si de una población con media μ y Varianza σ^2 se selaccionan todas las posibles muestras de tamaño n y se calcula el estadístico Varianza Muestral S^2 se obtiene una:

DISTRIBUCION MUESTRAL DE VARIANZAS: que cuando la población es normal:

$$\mu(S^2) = \sigma^2 ((n-1)/n)$$

$$\sigma^2(S^2) = \sigma^2 \surd(2/n)$$

Y para calcular probabilidades de que la variable S^2 de la muestra esté en un intervalo de valores dado se utiliza el estadístico:

$$£^2 = (S^2/\sigma^2)*(n-1)$$

que tiene una distribución Chi-cuadrado con (n-1) grados de libertad.

Nota:
$S^2 = \Sigma ((Xi - X')^2/(n-1))$

APLICACIONES

La desviación estándar de los pesos de una población muy grande de estudiantes es de 10 libras. Si se extrae una muestra al azar de 100 estudiantes, la probabilidad de que la varianza de la muestra sea menor de 77.44 libras^2 es:

Datos:
n= 100 σ= 10 σ^2= 100 S^2= 77.44

Planteamiento:
La variable S^2 es la varianza de los pesos en una muestra de 100 estudiantes.

Solución:
$P(S^2 < 77.44) = P(£^2<(S^2/\sigma^2)(n-1))=P(£^2<77.44*99/100)=$
$P(£^2 < 76.67) \approx$ aproximadamente 0.05
De la tabla de la distribución Chicuadrado con 99 grados de libertad.
La probabilidad que la varianza sea menor de 77.44 libras² es aproximadamente 0.05

Resuelva:

Suponga que se sabe que en cierta población de adictos a las drogas la duración media del abuso es de 5 años, con una desviación estándar de 3 años. ¿Cuál es la probabilidad que una muestra al azar de 36 individuos de dicha población, proporciona una varianza en el tiempo de abuso de drogas superior a 12 años?

Datos:
$n=$ $\sigma=$ $\sigma^2=$ $S^2=$ $S=$

Planteamiento:
La variable S^2 es:

Solución:
$P(S^2 \quad)$

3.10 DISTRIBUCION MUESTRAL DE RELACIONES DE VARIANZAS ($S1^2/S2^2$)

Si de dos poblaciones distribuidas normalmente, con varianzas $\sigma 1^2$ y $\sigma 2^2$, se toman muestras de tamaño n1 y n2 respectivamente y se computan los estadísticos $S1^2$ y $S2^2$; al tomar todas las posibles combinaciones de estas muestras y calcular la razón $S1^2/S2^2$ se obtiene una:

DISTRIBUCION MUESTRAL DE RELACIONES DE VARIANZAS

El estadístico:

$F = (S1^2/ \sigma 1^2) / (S2^2/\sigma 2^2)$ tiene una

distribución F con ((n1-1), (n2-1)) grados de libertad.
y auxilia al cálculo de la probabilidad de que la razón de varianzas esté en un intervalo determinado.

APLICACIONES:

Se registraron las respuestas a la glucosa oral para 8 pacientes con enfermedad de Huntington (grupo 1) y 10 individuos de control (grupo 2). La probabilidad que en el análisis estadístico se encuentre una relación de varianzas ($S1^2/S2^2$) mayor que dos, si las poblaciones de pacientes tienen varianzas de 20 y 36 respectivamente es:

Datos:
$\sigma1^2 = 20$ $\sigma2^2 = 36$ $n1 = 8$ $n2 = 10$

Planteamiento:
La variable $S1^2/S2^2$ es la relación de las varianzas de las dos muestras estudiadas

Solución:
$P((S1^2/S2^2) > 2) = P(F > (S1^2 * \sigma2^2 / S2^2 * \sigma1^2)) =$
$P(F > 2 * (36/20)) = P(F > 3.6) \approx$ aproximadamente 0.05
De la tabla F con (7,9) grados de libertad se concluye que la probabilidad es menor que 5% pero mayor de 1%.

3.11 EVALUACION

I. Señale si las siguientes afirmaciones son verdaderas o falsas:

1. *El Error Estándar de un Estadístico es la Desviación Estándar de su Distribución Muestral.*

2. *Un Estadístico es una característica de la Población.*

3. *La distribución de probabilidades de las medias de todas las posibles muestras, del mismo tamaño, de una población se llama Distribución Muestral de Medias.*

4. *p es un Parámetro y σ es un Estadístico.*

5. *Un Método de Muestreo en el que los elementos se seleccionan de una población a intervalos uniformes se llama Muestreo Aleatorio Simple.*

II. Resuelva los Siguientes Problemas:

1. *Suponga que el consumo medio diario de proteína para cierta población es 125 g. y para otra población la media es 100 g. Si los valores diarios de consumo de proteínas en ambas poblaciones presentan una distribución Normal con una desviación estándar de 15 g. Cuál es la probabilidad que muestras al azar e independientes de 25 de cada población proporciones una diferencia entre sus medias de 12 o menos?*

2. *El 25% de los empleados de una Empresa faltaron a su trabajo debido a una lesión bucal tres o más días en el último año. Si se extrae una muestra aleatoria de 150 de dichos empleados. Cuál es la probabilidad de que la proporción muestral de los que faltaron a su trabajo tres o más días debido a una lesión bucal esté entre 0.20 y 0.30?*

MODULO 4

ESTIMACION

INTRODUCCION

La Inferencia Estadística es el procedimiento mediante el cual se llegan a deducir conclusiones acerca de una población con base en los resultados obtenidos del estudio de una muestra extraída de la misma. La Estimación es una de las áreas de la Inferencia Estadística e implica el cálculo a partir de los datos de una muestra de algún estadístico que se ofrece como una aproximación de un parámetro desconocido.

En éste módulo se presentarán los conceptos fundamentales relacionados con la Estimación de Parámetros, especificando los procedimientos para el cálculo de los intervalos de confianza en la estimación de la media, la varianza y la proporción de éxitos poblacionales.

OBJETIVOS

Al finalizar el estudio del módulo el lector estará en capacidad de:

1. Explicar a que se refieren los términos Estimación, Estimación puntual, Estimación por Intervalo.

2. Definir los términos: Estimador y Estima.

3. Calcular los Intervalos de Confianza para la Media, la Varianza y la Proporción de Exitos de una población.

4. Realice Inferencias utilizando los Intervalos de Confianza.

5. Interpretar el significado del término Error Típico de Estima.

6. Seleccione el Nivel de Confianza oportuno para realizar Análisis Inferenciales particulares.

7. Determinar el tamaño de muestra adecuado, para realizar estudios inferenciales bajo el criterio de Muestreo Aleatorio Simple.

4.1 INFERENCIA ESTADISTICA Y ESTIMACION

La Inferencia es la rama de la Estadística que se ocupa del uso de los conceptos de probabilidad para afrontar la incertidumre en la toma de decisiones, es el procedimiento por medio del cual se infiere respecto al comportamiento de una población con base a los resultados observados en una muestra. Los métodos de Inferencia se clasifican en dos categorías:

- Se puede tomar decisiones respecto al valor del parámetro (Hipótesis).

- Se puede estimar o predecir el valor del Parámetro (Estimación).

La estimación trata sobre como predecir los valores de los parámetros de una población a partir de los resultados mostrados por los Estadísticos en una muestra.

Sus métodos dan una información sobre el error posible que acompaña a la estimación, permitiendo así, aplicar ciertos controles para evitar en lo posible éste error.

4.2 CONCEPTOS BASICOS

Con frecuencia al tratar de describir una población partiendo de lo que se conoce de una muestra , se debe decidir cual es la función estadística que es una estimación adecuada para los valores paramétricos desconocidos.

A la estadística que sirve para estimar el valor del parámetro se llama ESTIMADOR.

Por ejemplo:

Un odontólogo puede estar interesado en saber que proporción de cierto tipo de individuos, tratado con un medicamento, sufre efectos secundarios indeseables.
Sin duda, su población consta de todas aquellas personas que alguna vez han sido o serán tratadas con ese medicamento, para conocer la proporción deseada, decide estimarla a partir del análisis de una parte de esa población (muestra), esto es, calcular la proporción de la muestra que resulta afectada, así el estimador es el estadístico P proporción de afectados en la muestra.

Puede existir, para un mismo parámetro ϕ más de una estimador E y la selección del mejor implica la comparación de varios de ellos, algunos criterios para seleccionar el estimador más adecuado son los siguientes:

1. Debe ser insesgado. Un estimador es insesgado cuando la

media de la distribución muestral del mismo es el parámetro de la población. $\mu(E)=\phi$.

2. Debe ser eficiente: Se dice que un estimador es eficiente si tiene una varianza finita y además no existe otro estimador para el parámetro cuya varianza sea menor que la de éste. $\sigma^2(E)$ es la menor comparada con la de otros estimadores de ϕ.

3. Debe ser congruente. Al aumentar el tamaño de la muestra se logra seguridad, casi absoluta, que el valor del estadístico se acerca al valor del parámetro.

4. Debe ser suficiente. Un estimador es suficiente si utiliza la información contenida en la muestra a tal punto que ningún otro estimador podría extraer de ésta más información referente al parámetro.

ESTIMA es el valor correspondiente al estadístico E que se obtiene a partir de la muestra. Es el valor especifico observado de un estimador.

Suponga que en el ejemplo anterior, el Odontólogo tomó una muestra de 50 pacientes y observó que en seis de ellos se manifestaron efectos secundarios. La proporción de la muestra es 6/50=0.12. Entonces la Estima es 0.12.

4.3 TIPOS DE ESTIMACION

Los procedimientos de estimación se pueden clasificar en:

Estimación puntual: Asigna un número al Parámetro y se puede asociar con un punto en la línea recta.

A menudo la estimación puntual es insuficiente, puesto que acierta o se equivoca, será mucho más útil si se acompaña de una estimación del error en que se puede incurrir.

ESTIMAS PUNTUALES

Parámetro a Estimar	Estadístico Estimador
Media μ	**Media Muestral X** **$X' = \Sigma\, xi/n$**
Varianza σ^2	**Varianza Muestral S^2** **$S^2 = \Sigma\,(xi-X')^2/(n-1)$**
Proporción de Exitos p	**Proporción de Exitos de la muestra P.** $P = \dfrac{\text{Numero de éxitos}}{n}$

Estimación por intervalo: señala dos puntos entre los cuales se encuentran una gama de valores que pueden representar al Parámetro, están asociados a un intervalo en la linea recta. El intervalo mide, de alguna forma, el error cometido en la estimación, indicando la probabilidad con que el parámetro se encuentra en él.

4.4 INTERVALOS DE CONFIANZA

El intervalo de Confianza determina dos cantidades $\varepsilon 1$ y $\varepsilon 2$, tales que la probabilidad que el valor del parámetro ϕ desconocido se incluya en el intervalo delimitado por ellas es Γ ($\Gamma \to 1$).

$$P(\varepsilon 1 <= \phi <= \varepsilon 2) = \Gamma$$

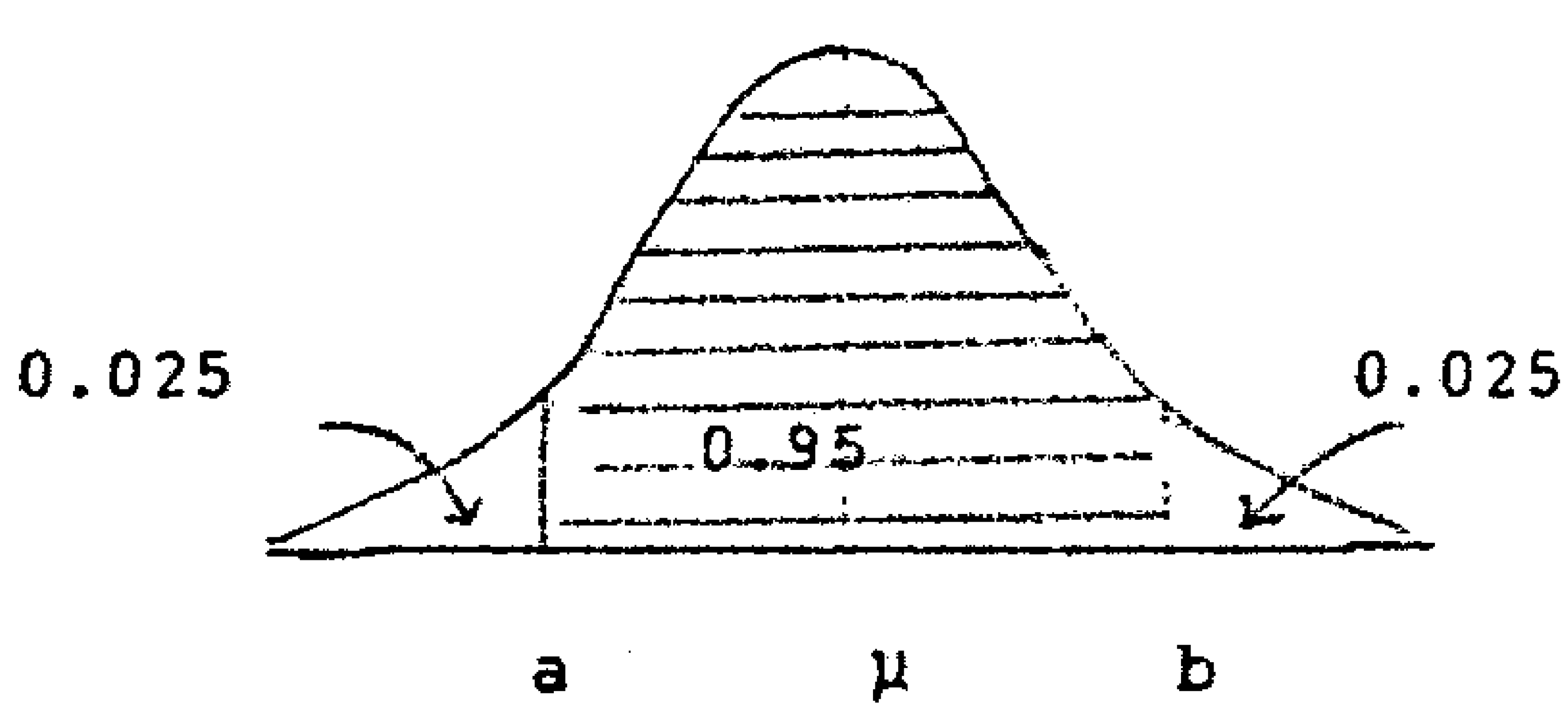
La estimación por intervalo indica dos cantidades numéricas en cuyo intervalo se incluye el valor del parámetro ϕ con una certeza conocida (con una probabilidad Γ) y es preferible a la estima puntual, ya que trata de encontrar que tanto se puede desviar, a lo más, el valor de la estima E del verdadero y desconocido parámetro ϕ.

La probabilidad Γ, de que un intervalo contenga o incluya al parámero se llama Coeficiente de Confianza, la elección de este coeficiente la hace el investigador dependiendo del grado de certeza que desee para su estimación.

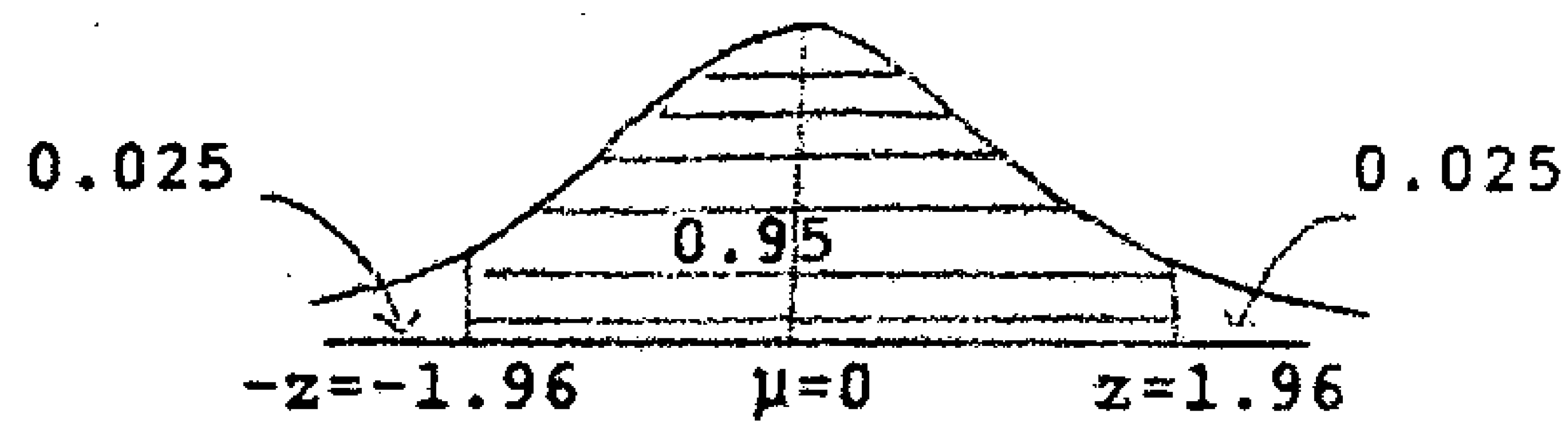
4.5 INTERVALO DE CONFIANZA PARA LA MEDIA DE UNA POBLACION NORMAL CON VARIANZA CONOCIDA

Supónga que se tiene una población infinita Normal con parámetro μ desconocido y σ^2 conocido (supóngalo igual a 4). Al seleccionar una muestra de tamaño n (n=36) de esta población el estadístico X' tiene una distribución muestral de la siguiente forma:

$\mu(X') = \mu$ $\sigma(X') = \sigma / \sqrt{n}$



Según la teoría de la distribución muestral, se pueden determinar los límites a,b de la distribución de tal forma que el intervalo que representan encierren el 95% de las medias muestrales en la forma siguiente:



En la distribución normal estandar

$$Za = \frac{a - \mu}{\sigma/\sqrt{n}} = -Z \qquad Zb = \frac{b - \mu}{\sigma/\sqrt{n}} = +Z$$

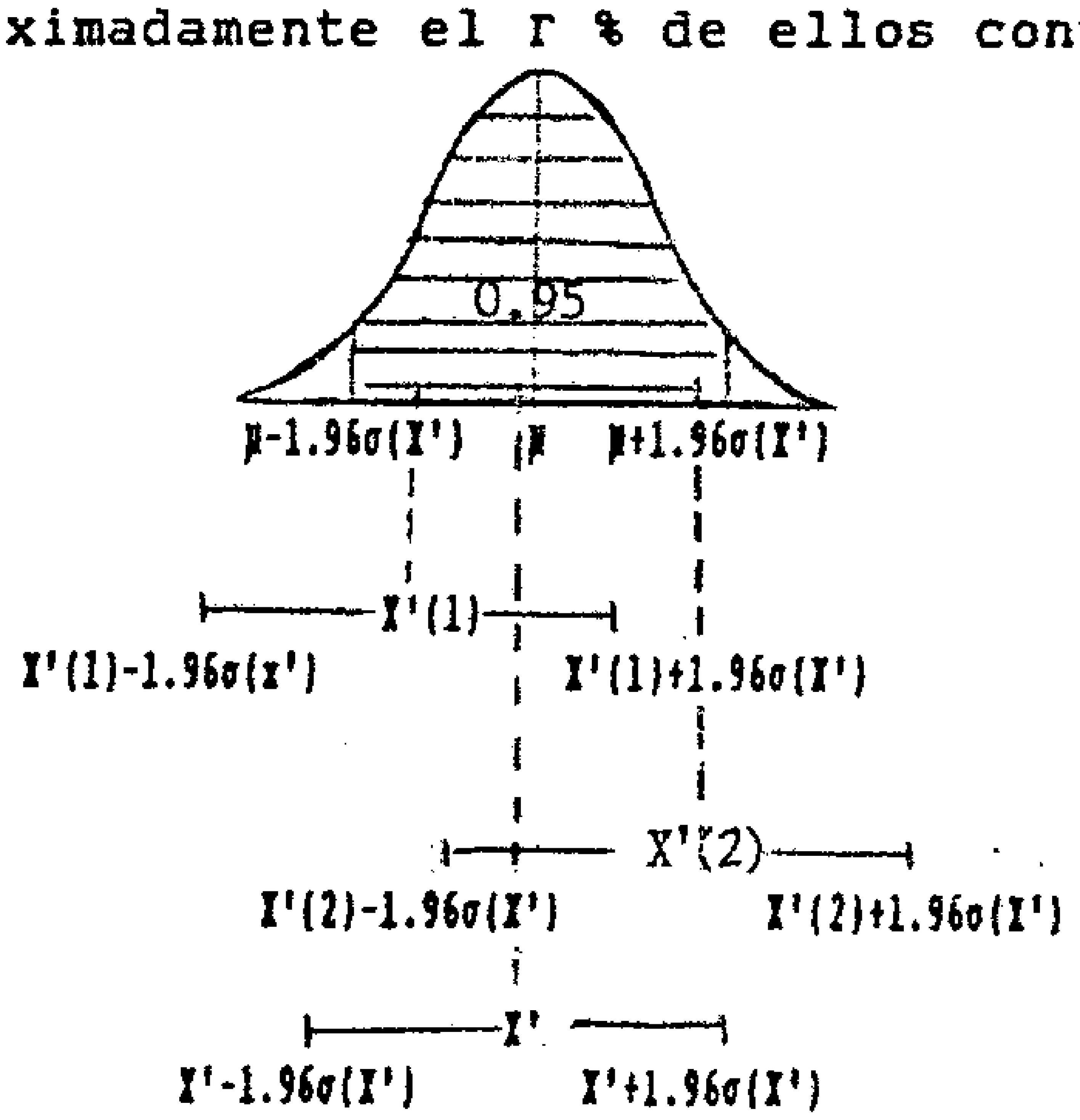
$$a = \mu - Z\ \sigma/\sqrt{n} \qquad b = \mu + Z\ \sigma/\sqrt{n}$$

$$a = \mu - 1.96 * 4/\sqrt{36} = \mu - 1.31 \qquad b = \mu + 1.96 * 4/\sqrt{36} = \mu + 1.31$$

Como μ es desconocida puede decirse que hay una probabilidad r=95% de que la media de una muestra de tamaño 36 se encuentre dentro del intervalo ± 1.96 Error Estándar a partir de la media. En otro sentido si el (r%) 95% de las muestras se halla al 1.96 error estandar (positivo y negativo), μ está a lo más a una distancia de 1.96 Error Estándar del 95% de las medias muestrales. Entonces si se construyen intervalos de r% de confianza a partir de las Xs' de las muestras, de la forma:

$$X' \pm 1.96\ \sigma/\sqrt{n}$$

aproximadamente el r % de ellos contendrán a μ.



Nota: Si se desea tener una confianza diferente al 95% entonces se debe asociar un valor Z que corresponda al r% fijado.

De esta manera, se llega a establecer la forma de cáclulo del intervalo de confianza para la Media en poblaciones Normales con Varianza conocida.

$$X' \pm Z\,\sigma/\sqrt{n}$$

donde Z representa el nivel de confianza r% seleccionado.

La cantidad ($Z\,\sigma/\sqrt{n}$) se conoce como Error de Estimación y representa la mayor distancia que puede presentarse entre la estima puntual y el vardadero valor de μ con un r% de probabilidad.

Por ejemplo:

Una muestra de 100 hombres adultos aparentemente normales de 25 años de edad, mostró una presión sistólica sanguinea media de 125. Si se sabe que la desviación estándar de la población es 15 encuentre :

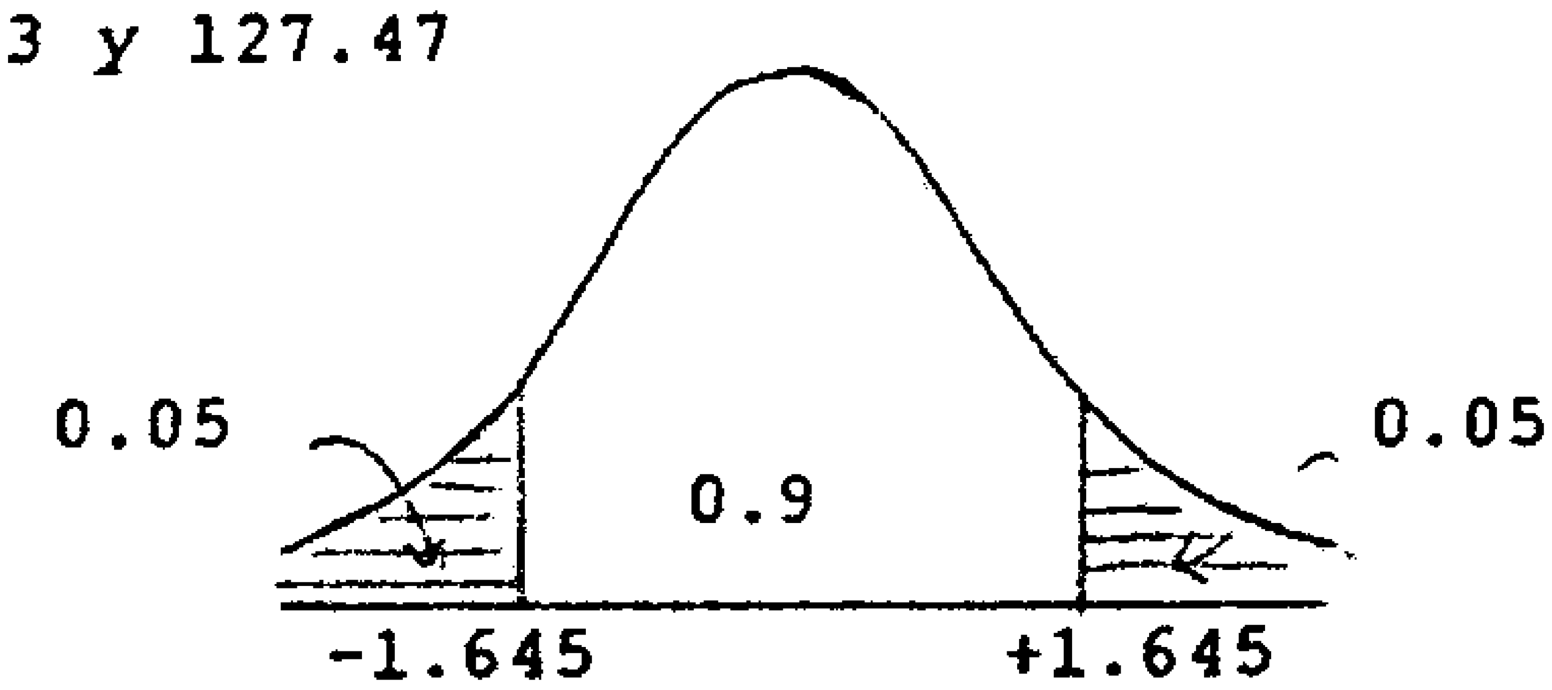
A. El intervalo del 95% de confianza para el parámetro Media de la Presión Sanguinea en hombres de 25 años.

Datos
n= 100   X' = 125 σ= 15   Z= 1.96 para r=95%

Intervalo
$X' \pm Z\,\sigma/\sqrt{n} = 125 \pm 1.96*15/\sqrt{100} = 125 \pm 2.94$
$$122.06 <= \mu <= 127.94$$
Se tiene el 95% de confianza que el intervalo (122.06,127.94) contenga al prámetro μ.

B. El intervalo del 90% de Confianza para μ
Z = 1.645 para r= 90%   (P(Z<=1.645)= 0.95)

Intervalo
$125 \pm 1.645 * 15 / \sqrt{100} = 125 \pm 2.468$
$$122.53 <= \mu <= 127.47$$
Se tiene el 90% de confianza que la media de la población esté entre 122.53 y 127.47

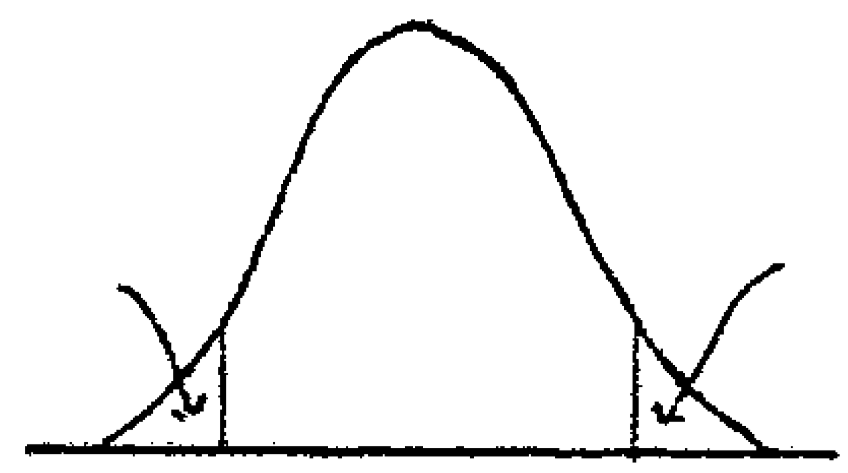


Resuelva:

En un estudio de flujo de pacientes a través del consultorio se encontró que en promedio 35 pacientes llegaban 17.2 minutos tarde a las citas. Una investigación previa habia demostrado que la desviación estándar de la población era de 8 minutos.
Si se supone que la distribución de la población es normal, clacule el intervalo del 99% de confianza para la cantidad de tiempo promedio de llegada tarde a las citas para la poblaciòn de pacientes en el consultorio.

Datos
$n=$ $\quad X'=$ $\quad S^2=$ $\quad S=$ $\quad Z=$

Intervalo



En algunas oportunidades no se tienen las condiciones de normalidad y varianzas conocidas para la estimación de la media, por lo que se han planteado procedimientos alternativos.

El cuadro siguiente reúne las fórmulas para el cálculo de Intervalos de Confianza para la media y sus aplicaciones.

INTERVALOS DE CONFIANZA PARA LA MEDIA

Poblaciones Normales con varianza conocida y aplicable a:	$X' \pm Z\ \sigma/\sqrt{n}$
Muestras grandes cuando la población no puede suponerse Normal y σ^2 es conocida o desconocida.	$X' \pm Z\ \sigma/\sqrt{n} * \sqrt{((N-n)/(N-1))}$ Poblaciones finitas
Muestras grandes y población normal con σ^2 desconocida y es estimada por S^2.	$X' \pm Z\ S\ /\sqrt{n}$ $X' \pm Z\ S/\sqrt{n} * \sqrt{((N-n)/(N-1))}$ Poblaciones finitas
Poblaciones Normales con Varianza desconocida y muestras pequeñas (n<30) Donde t es un estadístico con distribución T de Student, n-1 grados de libertad, y representa el porcentaje Γ de confianza en el intervalo.	$X' \pm t\ S\ \sigma/\sqrt{n}$

Por ejemplo:

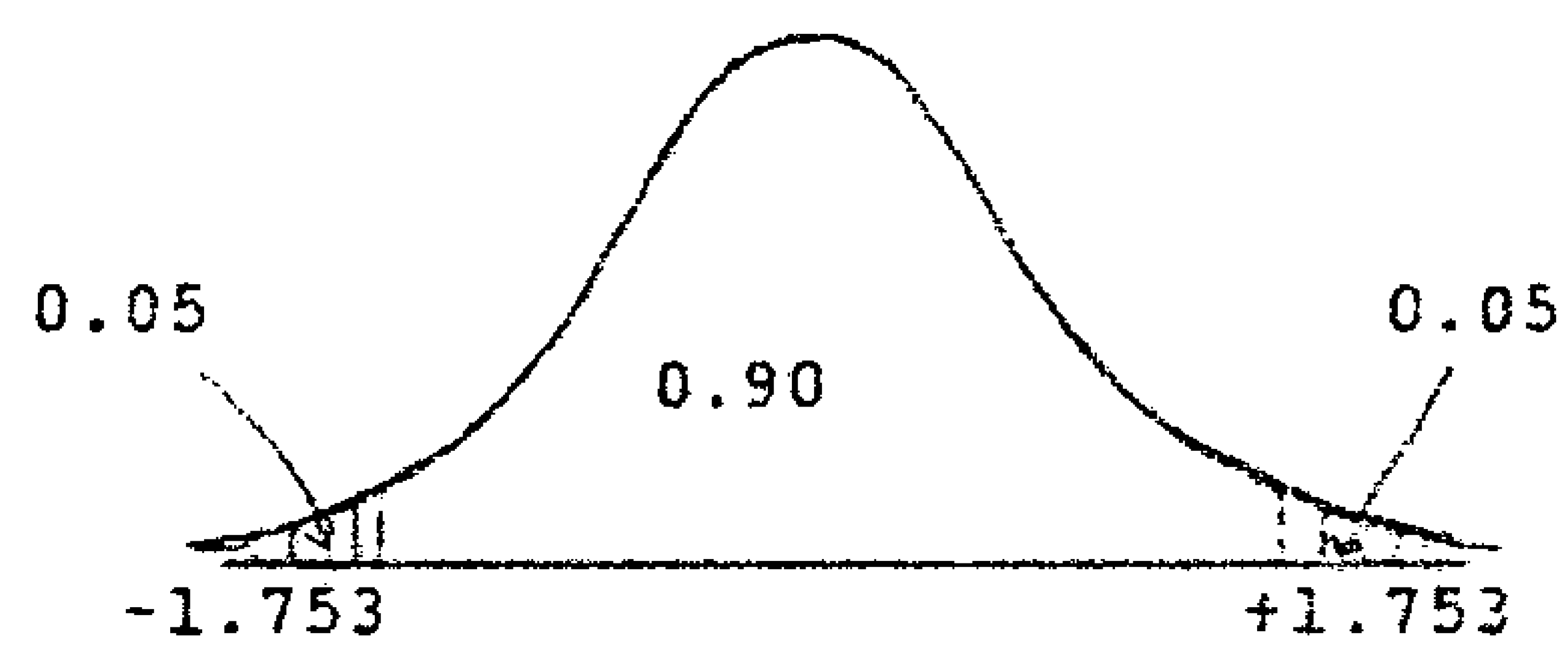
Se encontró que el nivel indirecto medio de bilirrubina en el suero de 16 niños de 4 días de nacidos era de 5.98 mg/100 cc con una desviación estándar de 3.5 mg/ 100 cc. Suponiendo que los niveles de bilirrubina presentan una distribución aproximadamente normal; el intervalo del 90% de confianza para el nivel medio de bilirrubina en el suero de la población de niños de 4 días de nacidos es:

Datos
X'= 5.98 S= 3.5
n= 16 muestra pequeña
Poblacion con Distribución Normal
Γ= 90% t= ± 1.753

De la tabla T con 15 grados de libertad

Intervalo:
$X \pm t\ S/\sqrt{n} = 5.98 \pm 1.753 * 3.5 /\sqrt{16} = 5.98 \pm 1.534$
$4.446 <= \mu <= 5.514$

Se tiene un 90 % de confianza que la media de la población está etre 4.446 y 7.514 mg/ 100 cc.

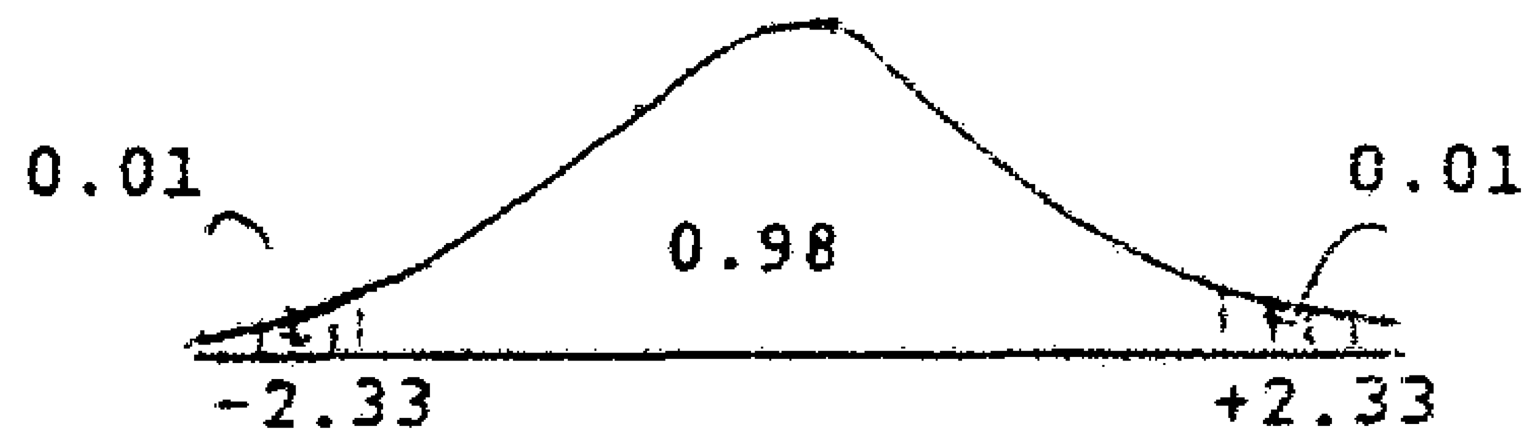


El administrador de un hospital tomó una muestra de 40 cuentas vencidas, a partir de las cuales se calculó la media, Q250.00 y la desviación estándar, Q75.
El intervalo del 98% de confianza para la media de la población de cuentas vencidas es:

Datos
X'= 250 S = 75 n = 40
Distribución de la población no identificada
Muestras grandes
Γ= 98%

Intervalo
$X' \pm Z\ S /\sqrt{n}$
$Z = 2.33$
$250 \pm 2.33 * 75 / \sqrt{40} = 250 \pm 27.63$
$222.36 <= \mu <= 277.63$

Se tiene el 98 % de confianza que la media de la población está entre Q 22.36 y Q 277.63.



Resuelva:
Se estudió la actividad total del complemento serológico (CH50) en 20 personas aparentemente sanas y se obtuvo los siguientes resultados: Media 47.2, Desviación Estándar 10.1. Los Investigadores tenían razón al pensar que la población muestreada está distribuida en forma aproximadamente Normal. Encuentre el intervalo del 95 % de confianza para la media de la actividad serológica de la población estudiada.

Datos:
n= *X=* *S=*
Distribución:
γ=

Intervalo:

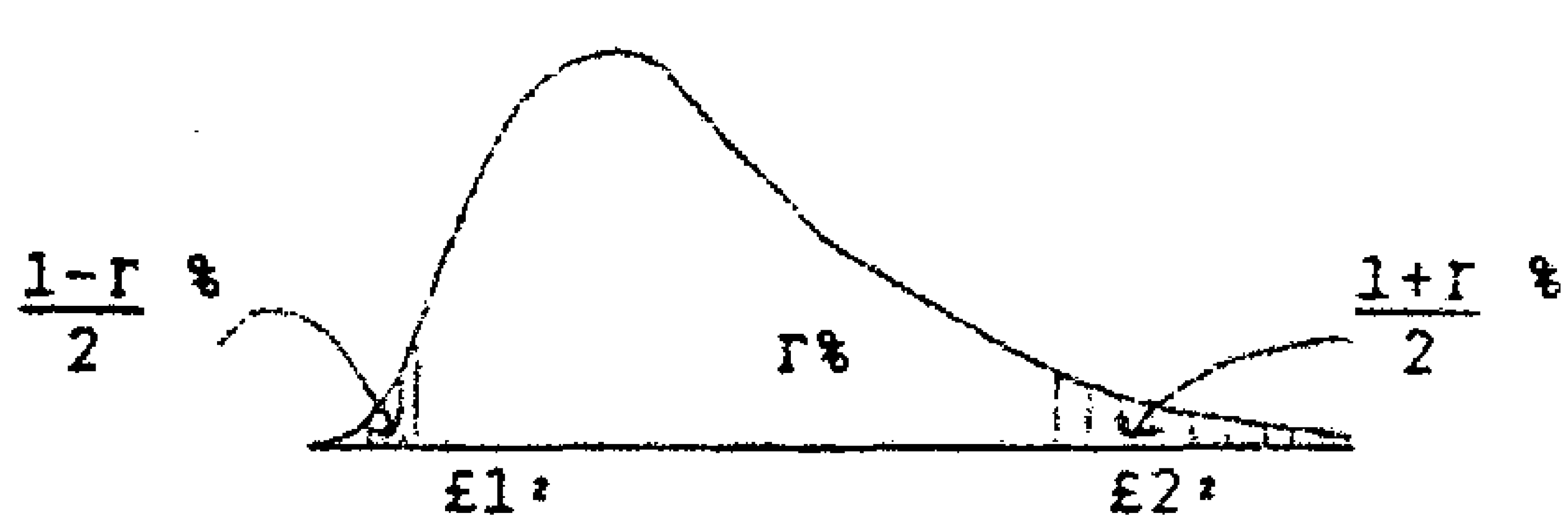
4.6 INTERVALO DE CONFIANZA PARA LA VARIANZA DE UNA POBLACION NORMAL.

Para estimar la varianza , el estadístico utilizado es la varianza muestral S^2 y el intervalo de confianza se basa en la distribución muestral del estadístico $\chi^2 = (n-1)*S^2/\sigma^2$, el cual tiene una distribución Chicuadrado con (n-1) grados de libertad.

INTERVALO DE CONFIANZA PARA LA VARILANZA POBLACIONAL

$$\frac{(n-1) * S^2}{\chi 1^2} <= \sigma^2 <= \frac{(n-1) * S^2}{\chi 2^2}$$

donde $\chi 1^2$ y $\chi 2^2$ son los valores de la variable Chicuadrado con (n-1) grados de libertad y $(1-\gamma)/2$ $(1+\gamma)/2$ de area acumulada respectivamente.



Por ejemplo:

Se hicieron determinaciones de hemoglobina en 16 animales expuestos a un compuesto químico nocivo, se registraron los siguientes valores:

15.6, 14.8, 14.4, 16.6, 13.8, 14.0, 17.3, 17.4, 18.6, 16.2, 14.7, 15.7, 16.4, 13.9, 14.8, 17.5.

Si se supone que la población muestreada es normal, los intervalos del 95% de confianza para la varianza y para la desviación estándar de la población son:

Datos:

$n=16$ $X' = \Sigma xi/n = 15.73$

$S^2=\Sigma(xi-X')^2/(n-1) = 2.18$

$S= \sqrt{S^2}= 1.48$

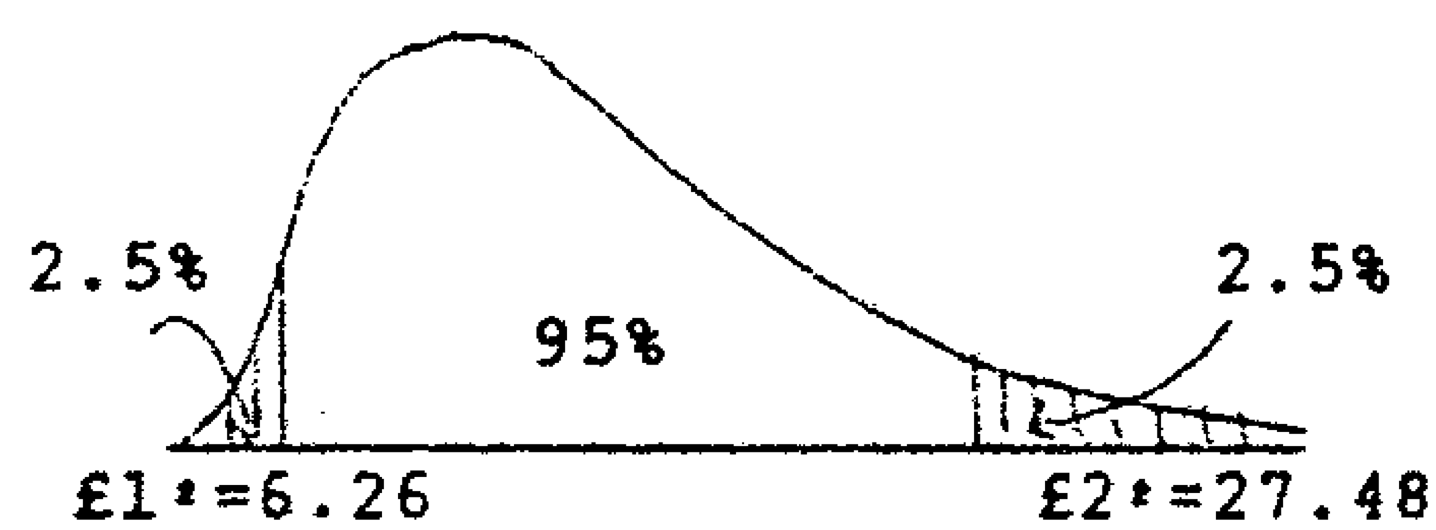
r= 95%

$£1^2 = 6.262$

valor de Chi-cuadrado con area de 2.5% y 15 grados de libertad

$£2^2 = 27.488$

Valor de Chi-cuadrado con area de 97.5% y 15 grados de libertad.



Intervalo

$(n-1)S^2/£1 <= \sigma^2 <= (n-1)\ S^2/£2 =$

$15*2.18/27.488 <= \sigma^2 <= 15*2.18/6.262$

Los intervalos del 95% de confianza para la varianza y la desviación estándar poblacional son respectivamente:

$1.19 <= \sigma^2 <= 5.22$

$1.09 <= \sigma <= 2.285$

Resuelva:

A cada uno de los miembros de una muestra de 51 estudiantes de odontología se les hizo una prueba estandarizada para medir su nivel de responsabilidad. Se obtuvo un valor de la varianza de 12.

Construya un intervalo del 99% de confianza para la desviación estándar de la población de estudiantes.

Datos:
n= *S^2=* *S=*
£1^2=
£2^2=
Intervalo:

4.7 INTERVALO DE CONFIANZA PARA PROPORCIONES

Para estimar la proporción de éxitos p en una población, se extrae una muestra de la misma y se calcula el estadístico P (número de exitos en la muestra/n). De acuerdo a la distribución muestral de proporciones el intervalo de confianza para muestras grandes (n>30) y np >5 está dado por:

INTERVALO DE CONFIANZA PARA PROPORCIONES

$$P \pm Z \sqrt{(PQ/n)}$$

donde Z tiene una distribución normal estándar y representa la confianza de la estimación.

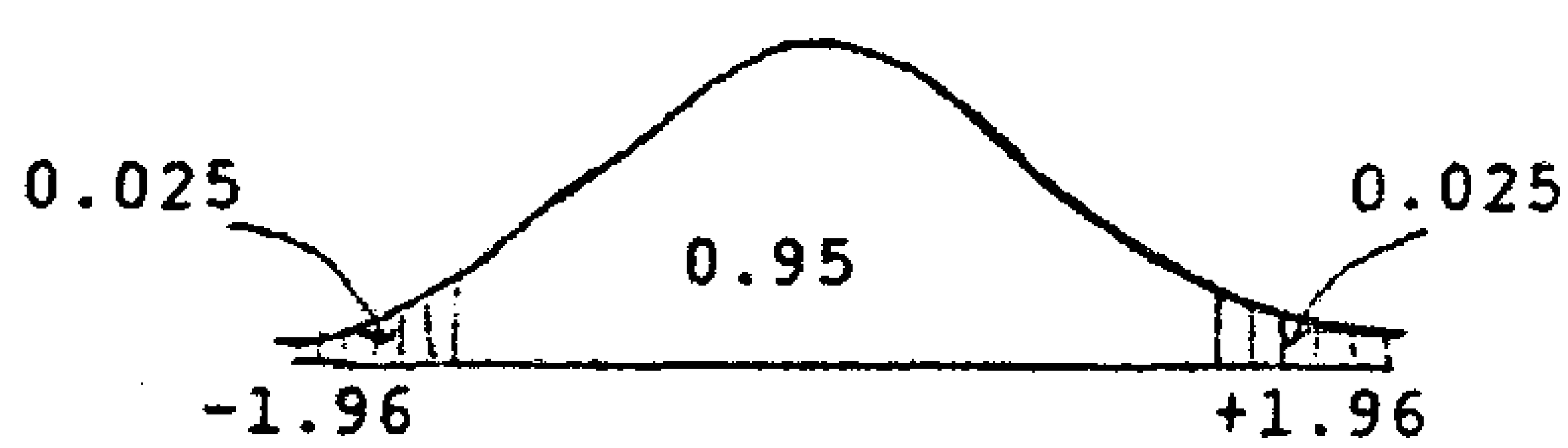
Por ejemplo:

Se llevó a cabo una encuesta para estudiar las prácticas de salud bucal de cierta población urbana de adultos. De los 300 adultos entrevistados 123 de ellos manifestaron que se someten regularmente a una revisión dental dos veces al año.

La estimación puntual de la proporción de adultos, que se someten a una revisión dental dos veces al año, es:
P= 123/300 = 0.41

El intervalo del 95% de confianza para la proporción de adultos que se someten a una revisión dental dos veces al año es:

Z = 1.96 valor de la variable Z con distribución Normal Estándar y 97.5% de area acumulada.



$P \pm Z \sqrt{(PQ/n)} = 0.41 \pm 1.96 \sqrt{(0.41*0.59/300)} = 0.41 \pm 0.05$

$$0.36 <= p <= 0.46$$

Se dice que se tiene el 95 % de confianza de que la proporción verdadera "p" de adultos que se someten a una revisión dental dos veces al año está entre 36% y 46%

Resuelva:

En una encuesta dental se pidió a 220 adultos que dieran la razón de su última visita al dentista, cuarenta y cuatro de ellos señalaron que lo habían hecho por razones de prevención .
Contruya un intervalo del 99% de confianza para la proporción poblacional de adultos que asisten al dentista por razones preventivas.

Datos.
n=
P=
Z=

Intervalo:

4.8 ELECCION DEL TAMAÑO DE LA MUESTRA

Supónga que se desea estimar el nivel promedio de PH en la saliva de los niños de una escuela primaria, además que el error de estimación sea menor de 0.5 unidades con una probabilidad de 0.95.

Puesto que aproximadamente el 95% de las medias muestrales se encuentran a una distancia de μ menor que $1.96\sigma(X')$ al repetirse el muestreo, lo que se desea es que $1.96\ \sigma/\sqrt{n}$ sea a lo más 0.5 unidades, que puede escribirse
como:

$$Z\ \sigma/\sqrt{n} = \text{error de estima deseado } D.$$

$$1{,}96\ \sigma/\sqrt{n} = 0.5$$

Para cumplir con este requisito el tamaño de muestra adecuado es:

$$n = z^2\sigma^2/D^2$$

$$n = 1.96^2 * \sigma^2/0.5^2$$

Inmediatamente se evidencia que no se puede obtener valor numérico de n, a menos que se conozca la desviación estándar de la población σ. A falta de un valor exacto de σ se puede usar la mejor aproximación disponible, tal como una estimación S obtenida de una muestra previa.

Si para el ejemplo se considera que $\sigma = 1.2$

$n = 1.96^2 * 1.2^2/\ 0.5^2 = 22.13 \approx 23$

El tamaño de muestra adecuado para obtener una estimación con un error a lo más de 0.5 unidades es de 23 niños.

El método de selección del tamaño de la muestra para los otros procedimientos de estimación son similares al que se acaba

de describir, esto es, el investigador debe especificar la cota del error de estimación D que desea y el nivel de confianza Γ% asociado, e igualando D al error de estimación en la fórmula del intervalo despejar el tamaño de la muestra.
El cuadro siguiente resume los procedimientos de cálculo del tamaño de muestra.

CALCULO DEL TAMAÑO DE MUESTRA

Para estimar la Media de la Población

$n = Z^2 \sigma^2 / D^2$ Para poblaciones infinitas

$n = N Z^2 \sigma^2 / (D^2(N-1) + Z^2\sigma^2)$ Para poblaciones finitas

---

Para estimar la Proporción de Exitos de la Población

$n = Z^2 p'q'/D^2$ Para poblaciones infinitas

$n = N Z^2 p'q'/(D^2(N-1)+Z^2p'q')$ para poblaciones finitas

donde p' y q' es una estimación preliminar de p obtenida de muestras anteriores. Si no existe esta estimación, se acostumbra maximizar el valor de n con la igualdad:

$$p = q = 0.5$$

Resuelva:

A un médico le gustaría estimar el valor medio de glucosa en la sangre, en mg/100ml, de 10000 pacientes en ayunas tratados en una clínica en los últilmos 4 años. Determine el número de registros que el médico tendrá que examinar para obtener un intervalo del 90% de confianza para μ, si el error de estimación (amplitud del intervalo) deseado es de 6 mg/100ml y una muestra piloto dió una varianza de 60.

Datos:
N= *D=* *Γ=* *Z=* $\sigma^2=$

Tamaño de la muestra n=

Se está planeando una encuesta con el fin de determinar que proporción de familias en cierta área son medicamente indigentes.
Se cree que la proporción no puede ser mayor que 0.35, y se desea un intervalo

del 95% de confianza con un error típico de estimación de 0.05. Cuántas familias deben seleccionarse para el estudio?.

Datos:
N= *D=* *r=* *Z=* $\sigma^2=$

Tamaño de muestra n=

4.9 EVALUACION

I. Complete las siguientes oraciones:

1. *Un solo número usado para estimar el parámetro de una población es una estimación* ____________

2. *Cuando ofrecemos una estimación por intervalo del parámetro de una población mostramos el grado de seguridad de que el intervalo contenga al parámetro real de la población estableciendo un nivel* ____________

3. *El estadístico muestral que se utiliza para estimar el parámetro de una población desconocida se denomina* ____________

II. Seleccione la respuesta correcta:

1. *Después de seleccionar una muestra y calcular X' un estadístico dice "Tengo confianza de 88% que la media de la población fluctúa entre 106 y 122". Qué es lo que realmente está diciendo?*

   a. *Que hay una probabilidad de 0.88 de que μ fluctúe entre 106 y 122.*
   b. *Hay una probabilidad de 0.88 de que μ=114 o sea el punto medio del intervalo.*
   c. *88% de los intervalos calculados de las muestras de éste tamaño contendrán a la media de la población.*
   d. *Todos los anteriores*
   e. *a y c pero no b.*

2. *Suponga que a 200 miembros de un grupo se les preguntó si visitaban al Odontólogo al menos una vez al año, 50 contestaron afirmativamente y 150 contestaron negativamente. Supondiendo que la respuesta afirmativa significa éxito, cuál de las siguientes expresiones es correcta.*

   a. $P = 0.75$
   b. $P = 0.25$
   c. $p = 0.75$
   d. $p = 0.25$
   e. *b y d solamente.*

III. Resuelva los siguientes problemas:

1 *Se determinan los niveles de pH de la saliva en una muestra de niños de la escuela primaria. Los resultados fueron los siguientes: 7.14, 7.11, 7.61, 7.98, 7.21, 7.16, 7.89, 7.24, 7.86, 7.47, 7.82, 7.37, 7.66, 7.62, 7.65*
Construya un intervalo del 90% de confianza para la media y la desviación estándar de la población.

2. *Un epidemiólogo desea saber que proporción de adultos que vivien en un área metropolitana grande tienen el tipo ay del virus B de Hepatitis. Determine el tamaño de muestra que se requerirá para estimar la proporción verdadera cercana en 0.03 con un intervalo de confianza del 95%. En un area similar se reportó que la proporción de adultos con la caracteristica de interés fue de 0.20.*

MODULO 5

HIPOTESIS

INTRODUCCION

Una hipótesis es una suposición que se hace respecto a alguna característica de una población.

Este módulo presenta los conceptos fundamentales y los procedimientos de Prueba de Hipótesis cuyo objetivo es determinar cuando es razonable concluir a partir del análisis de una muestra, que la población entera posee la determinada propiedad o parámetro supuesto y cuando no es razonable llegar a esa conclusión. Presenta especificamente las pruebas relacionadas con: la media de una o más poblaciones;la proporción de éxitos de una o más poblaciones; la varianza de una o dos poblaciones; y las pruebas de Chicuadrado de la independencia.

OBJETIVOS

Al finalizar el Módulo el lector estará en capacidad de:

1. Formular la hipótesis Nula y Alternativa en un problema dado.

2. Identificar la aplicación de cada uno de los procedimientos de Prueba de Hipótesis expuestos.

3. Interpretar el término Significancia.

4. Definir el Error Tipo I y el Error Tipo II.

5. En situaciones particulares evaluar la importancia de los riesgos de hacer falsas decisiones.

6. Comprobar con el procedimiento adecuado la hipótesis Nula a determinado nivel de significancia.

7. Considerando los Errores Tipo I y Tipo II calcular el tamaño de muestra para efectuar una prueba de hipótesis.

5.1 HIPOTESIS

Es una teoría tentativa o suposición adoptada previamente para explicar ciertos hechos y guiar una investigación.

Es una aseveración o conjetura relacionada con el comportamiento de una o más poblaciones.

Es un enunciado que se hace acerca de una característica cuantitativa de la población.

Por ejemplo:

Supónga que un investigador lleva a cabo un estudio para identificar diferencias entre dos tipos de sellantes de fosas y fisuras, un enunciado preliminar lo plantearía de la siguiente forma: Hay diferencia entre un sellante autopolimerizable y un sellante fotopolimerizable.

De otra forma: Un sellante autopolimerizable ofrece mayor tiempo de vida útil que un sellante fotopolimerizable.
En este caso las poblaciónes serán todas las restauraciones que se hacen usando cada uno de los sellantes y se hace un supuesto sobre una característica: La vida útil de la Población.

5.2 HIPOTESIS NULA Y ALTERNATIVA

HIPOTESIS NULA (Ho): Está fundamentada en la teoría o principio propuesto y establece que la diferencia entre el resultado de la muestra y la teoría no es significativa sino debida al azar.

En el ejemplo, a pesar que el investigador plantea el supuesto que el sellante autopolimerizable ofrece mayor tiempo de vida, la Ho la plantearía como: No existe diferencia entre los dos sellantes, la diferencia que pueda aparecer en las muestras estudiadas es debida al azar, ya que el tiempo de vida de cada una de las restauraciones es una variable aleatoria.

HIPOTESIS ALTERNATIVA (H1): Es una suposición que contradice la hipótesis nula e implica que las variaciones son significativas, es decir no se deben al azar sino a algún o algunos factores determinados.

En el ejemplo, H1 sería planteada como: Existe diferencia entre los sellantes, la vida útil del sellante autopolimerizable es superior a la del sellante fotopolimerizable; la diferencia en el tiempo de vida se debe al tipo de sellantes y no al azar (esta hipótesis es congruente con la teoría del investigador).

Resuelva:
Supónga que se tiene interés en investigar si la aspirina disminuye en los pacientes el dolor posterior a la extracción tal como un compuesto de aspirina con codeina.
*¿ Cuál será la Hipótesis Nula apropiada?*____*¿ Cuál será la Hipótesis Alternativa apropiada?*____

Cualquiera que sea el planteamiento conceptual de las hipótesis estas deben traducirse a supuestos sobre los parámetros de una variable.

En los planteamientos anteriores las hipótesis están dadas en función a la Vida Util de una clase de sellante, sin embargo para comprobar que una de ellas es cierta, se deben representar en función de un parámetro característico, este puede ser: El Promedio de la Vida Util de los sellantes, que es una medida para evaluar el comportamiento de la población.

La hipótesis nula Ho se plantea:
La Vida Util promedio es igual para ambos sellantes

$$Ho : \mu(1) = \mu(2)$$

La hipótesis alternativa H1 es:
La Vida Util promedio del sellador autopolimerizable es mayor que la del sellador fotopolimerizable.

$$H1: \mu(1) > \mu(2)$$

Representar las hipótesis en función de parámetros es un requisito para contrastarlas estadisticamente.

Las hipótesis que se refieren a parámetros pueden clasificarse en:

Hipótesis Simple: cuando asigna un valor al Parámetro.

Por ejemplo:

Ho: Las Vida Util promedio del sellante fotopolimerizable es de 5 años.

$$Ho: \mu=5$$

H1: La Vida Util promedio del sellante fotopolimerizable es de 4 años.

$$H1: \mu=4$$

Ho: La diferencia entre la vida útil promedio de un sellante fotopolimerizable y un sellante autopolimerizable es cero.

$$Ho: \mu(2)-\mu(1)= 0$$

H1: La diferencia entre la vida útil promedio de un sellante fotopolimerizable y un sellador autopolimerizable es de un año.

$$H1.\ \mu(2)-\mu(1)=1$$

Hipótesis Compuesta: cuando asigna un conjunto de valores posibles al parámetro.

Por ejemplo:

Ho: La Vida Util promedio del Sellante Fotopolimerizable es mayor o igual a 4 años.

$$Ho: \mu>=4$$

H1: La Vida útil promedio del Sellante Fotopolimerizable es menor de 4 años.

$$H1: \mu<4$$

Ho: La diferencia entre los promedios de Vida Util de ambos sellantes es mayor o igual a un año

$$Ho.\ \mu(1)-\mu(2)>=1$$

H1: La diferencia entre los promedios de Vida Util de ambos sellantes es menor a un año.

$$H1: \mu(1)-\mu(2)<1$$

Generalmente la hipóteis nula se plantea como un hipótesis simple y la hipótesis alternativa como un hipótesis compuesta.

5.3 COMPROBACION DE HIPOTESIS

Es un procedimiento formal que utilizan los investigadores para probar las teorías propuestas; en éste se supone que los resultados del experimento están de acuerdo a cierto modelo o teoría y que las variaciones son debidas al azar. Para llevar a cabo la comprobación de hipótesis, se toma una muestra de los elementos de la población, calculando un estadístico, se determina a partir de él si los resultados son consistentes o imposibles con la teoría planteada, lo que implica la aceptación o rechazo de la hipótesis.

Tomando en cuenta que el resultado de la muestra, estadístico, puede variar por causas al azar o significativas, la prueba consiste en establecer, si Ho es cierta, el conjunto de posibles valores del estadístico resultante y hacer en éste conjunto una partición de dos regiones, Región Crítica y Región de Aceptación, es decir, se identifica la partición de forma tal que marque los límites entre lo probable (variaciones al azar) y lo imposible (variaciones significativas) del comportamiento muestral (comportamiento del estadístico). Estos límites quedan determinados por el Nivel de Significancia α, que es la probabilidad suficientemente pequeña, que si la hipótesis nula es verdadera, el estadístico tenga un valor dentro de cierto intervalo de valores marcado como región de rechazo.

Por ejemplo:

En el análisis sobre efectos secundarios que cierto medicamento tiene en los pacientes, se estableció el 5 % de afección.
Para determinar por un muestreo de 30 pacientes que el porcentaje se mantiene y no es mayor del 5%, se establece el nivel de significancia y las regiones de aceptación y rechazo de la siguiente forma.

I. Las hipótesis del estudio se plantean en función del parámetro p, proporción de pacientes afectatos.

Ho. La proporción de pacientes con efectos secundarios es 5%.

H1. La proporción de pacientes con efectos secundarios es mayor del 5%.

II. En una muestra de 30 pacientes es posible encontrar desde cero a treinta pacientes con efectos secundarios. La probabilidad de éstos sucesos son:

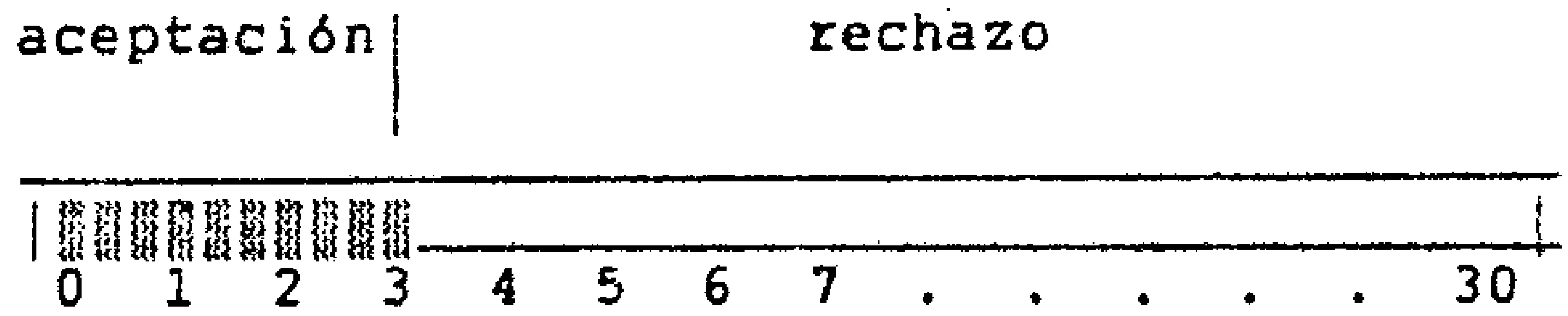
P(cero pacientes con efectos secundarios)	= 0.223
P(un paciente)	= 0.335
P(dos pacientes)	= 0.251
P(tres pacientes)	= 0.125
P(cuatro pacientes)	= 0.047
P(cinco pacientes)	= 0.015
P(seis pacientes)	= 0.003
P(siete pacientes)	= → 0
..	..
..	..

Cada una de estas probabildades fueron calculadas por la distribución binomial.

P(X=Xo)= (30!/(X!*(30-X)!)* 0.05^X*0.95^(30-X)

Esto implica que si la hipótesis es verdadera la muestra presenta a los más 3 pacientes afectados con probabilidad 0.934, es decir es muy probable que los resultados no sean mayores que tres si la hipótesis es verdadera.
Los resultados muestran al menos cuatro pacientes afectados con probabilidad de 0.066, es poco probable, si la hipótesis es verdadera, que el número de pacientes con efectos secundarios sean mayores que tres.

El nivel de significancia α puede fijarse igual a 6.6% que representa una probabilidad pequeña que si Ho es verdadera el resultado de la muestra tenga valores mayores que tres, región critica.



Número de pacientes en la muestra con efectos secundarios

Si la muestra presenta a lo más tres pacientes con efectos secundarios se acepta Ho, pués los resultados son consistentes con la hipótesis (altamente probables). Si la muestra presenta más de tres pacientes con efectos secundarios se rechaza la hipótesis porque los resultados son poco probables si Ho fuera cierta.

5.4 PRUEBA DE HIPOTESIS PARA LA MEDIA DE UNA POBLACION NORMAL CON VARIANZA CONOCIDA.

Las pruebas de hipótesis para muestras de poblaciones normales, se basan en el análisis del comportamienteo del estadístico X' y en su distribución muestral. A continuación se presenta el razonamiento de la Prueba:

Se lleva a cabo una investigación para identificar el precio medio cargado a cierto tratamiento, se considera que el precio justo es $10.00

I Planteamiento de las hipótesis.
Ya que el problema se refiere al precio promedio de un tratamiento, las hipótesis deben relacionarse con el parámetro μ, precio medio de todos los tratamientos similares.

La hipótesis nula debe plantearse congruente a cierta teoría. Ho: $\mu = \$10$

La hipótesis alternativa debe contradecir a al hipótesis nula y puede seleccionarse entre:

H1: $\mu \neq \$10$ H1: $\mu < \$10$ H1 > \$10

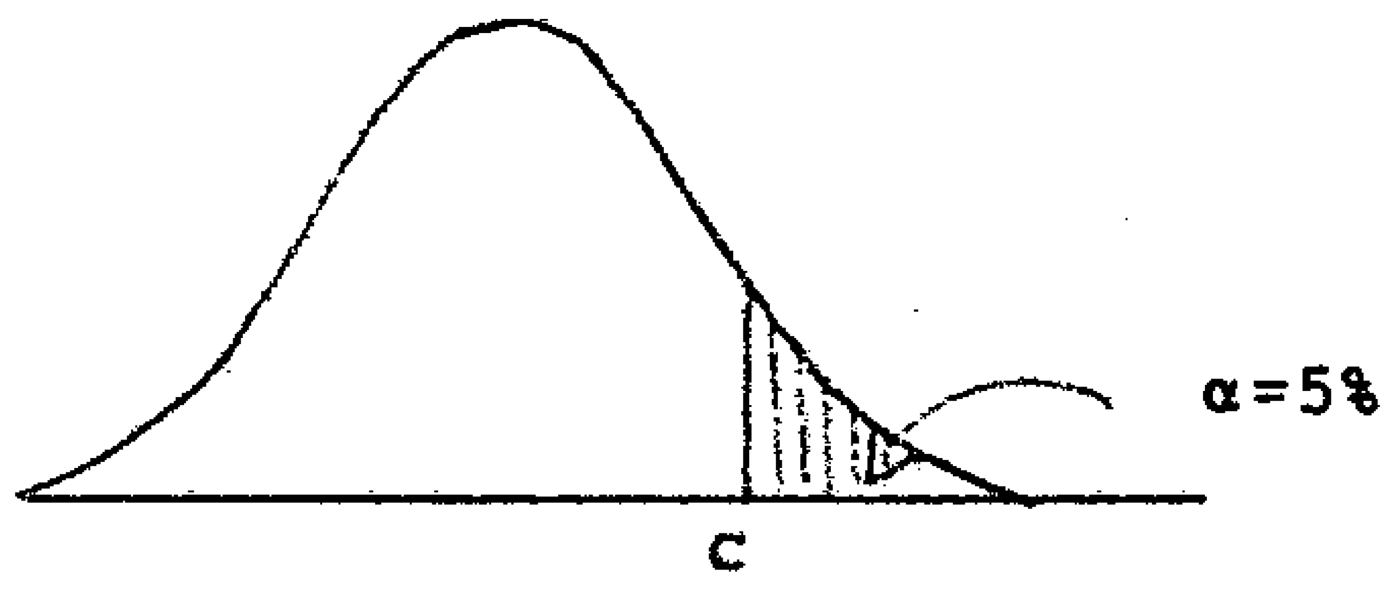
Las dos últimas conducen a una prueba unilateral y la primera a una prueba bilateral.

II Selección del estadístico de Prueba:
Al ser una prueba de hipótesis relacionada con el parámetro μ, el estadístico será el promedio artimético X'.

III Delimitación de areas de aceptación y rechazo.
Supónga que se selecciona, Ho: $\mu = \$10$ y H1: $\mu > \$10$ con un

nivel de significancia de 5% (el investigador fija a su discreción el valor de α).
Se delimitan las areas de aceptación y rechazo de acuerdo a un ensayo unilaterial a la derecha.

Como para el estadístico X' de una muestra obtenida de poblaciones normales con varianza conocida, la distribución muestral es normal con $\mu(X) = \mu$ y $\sigma(X) = \sigma/\sqrt{n}$ las áreas pueden graficarse de la siguiente forma.



Aceptación | Rechazo

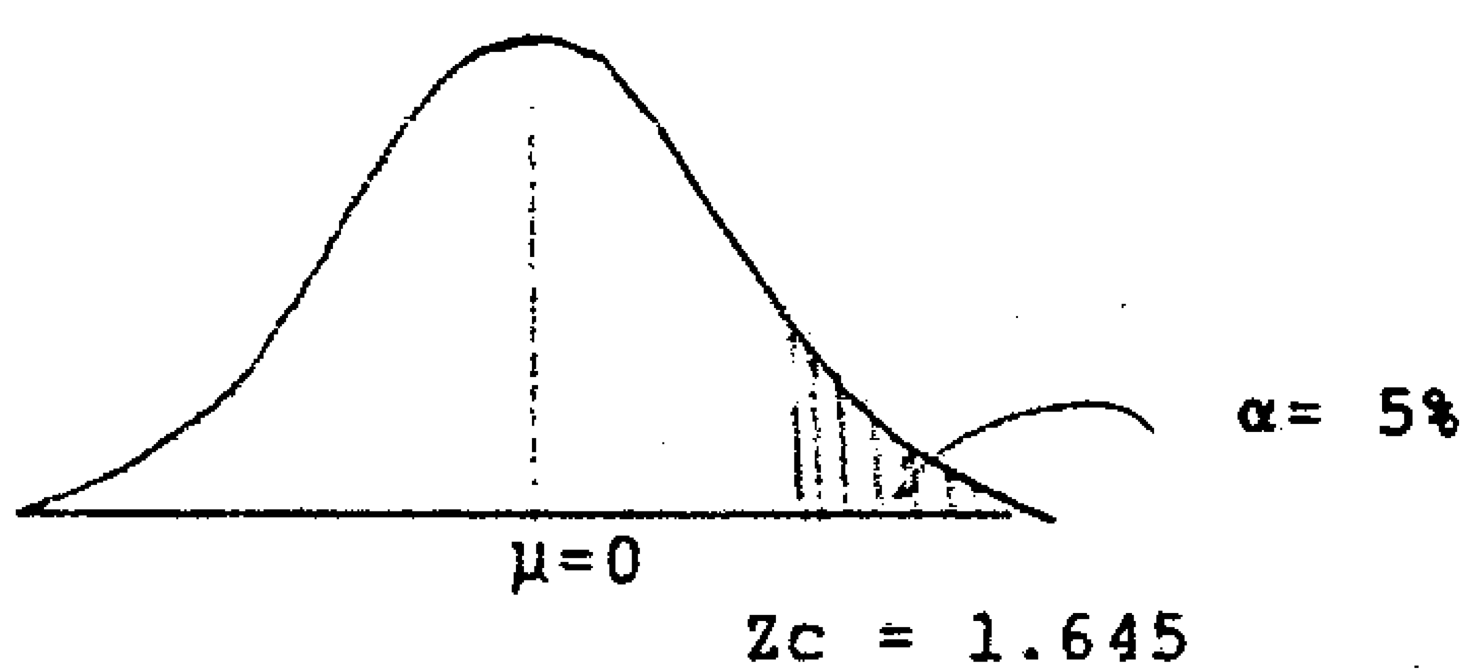
Observe que si la hipótesis nula es cierta, existe una probabilidad de 5% que la X' de la muestra sea superior a c y existe una probabilidad de 95% que el valor de X sea menor que c.

El valor c de X' que separa las regiones se llama Valor Crítico del estadístico.

¿Cuál es el valor de c?...
Hay que recordar que al trabajar con la distribución normal es práctico usar la transformación:

$$Z = \frac{X' - \mu(X')}{\sigma(X')}$$

que es una variable normal estándar, por lo que las áreas en esta distribución se representan:



El valor de Zc se denomina valor de z crítico y se obtiene de la tabla normal y limita también las áreas de aceptación y rechazo.

El criterio de decisión puede establecerse en esta distribución así: Si la transformación Z del estadístico, es menor que Zc se acepta Ho. porque los resultados son consistentes con la hipótesis nula. Si el valor de Z es mayor que Zc se rechaza Ho. porque existe diferencia significativa entre el resultado del estadístico y la teoría presentada, ya que la muestra presenta resultados poco probables si las variaciones fueran debidas al azar.

IV Si en el problema se sabe que σ es \$6 y que el precio está distribuido en forma normal; al tomar una muestra de 64 clínicas que revelan un precio promedio X' de \$12 la prueba se efectúa:

$$Z = \frac{X' - \mu}{\sigma/\sqrt{n}} = \frac{12-10}{6/\sqrt{64}} = 2.66$$

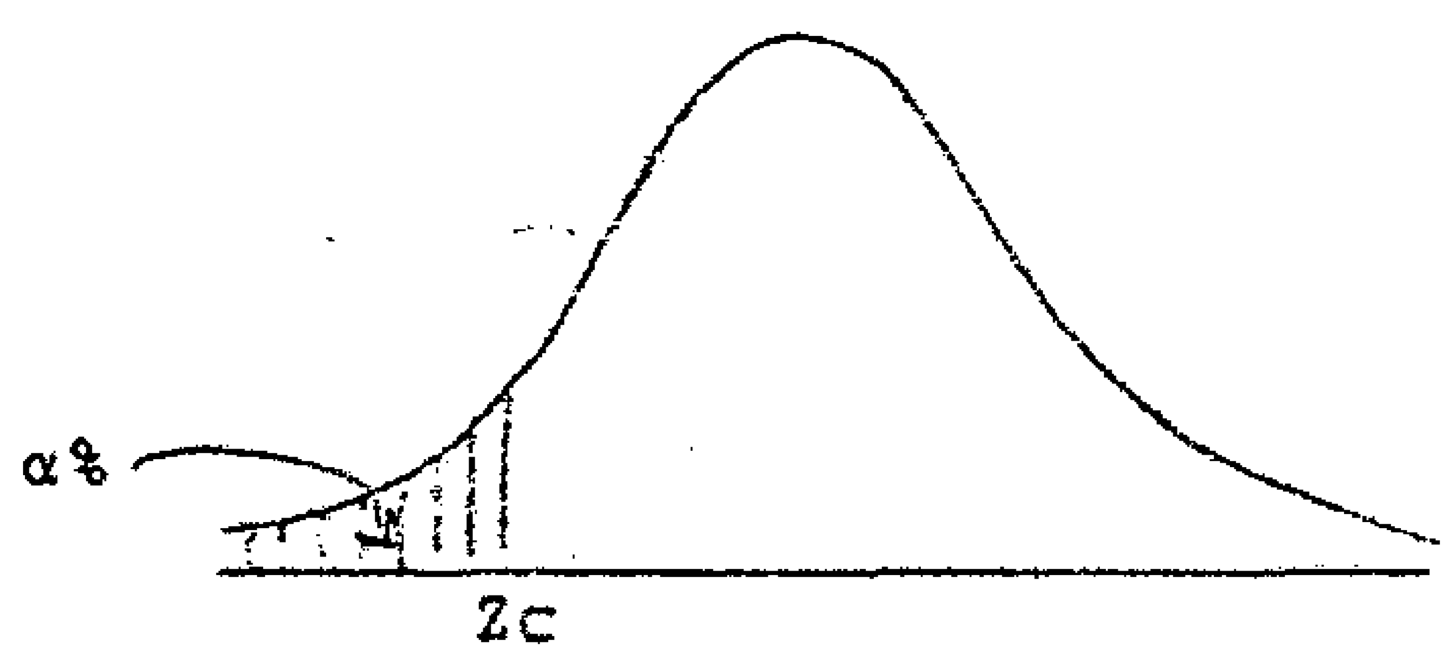
Z= 2.66 es mayor que Zc, Z está en el área de rechazo por lo tanto no hay evidencia para aceptar Ho; si ésta fuera cierta la probabilidad que el estadístico tenga un valor mayor o igual a \$12 es 0.003 (resultado obtenido de la tabla normal) que es menor del nivel de significancia fijado.

V Después de efectuada la prueba se puede concluir que la muestra evidencia que el precio promedio del tratamiento es superior al precio justo \$10.

OBSERVACIONES:

Si en el planteamiento se hubiera seleccionado cualquiera de las otras hipótesis alternativas la delimitación de las areas sería:

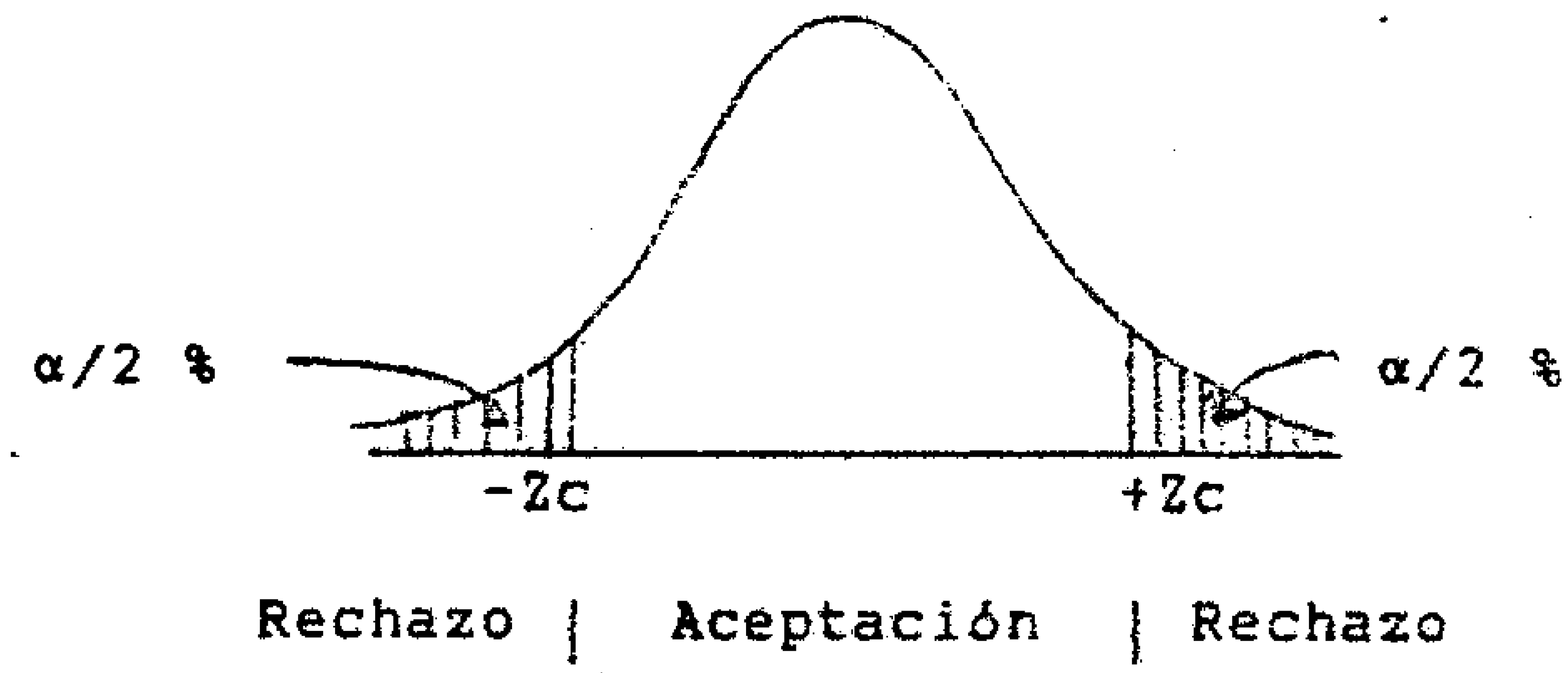
Ho: μ=\$10
H1: μ<\$10



Rechazo | Aceptación

Ensayo unilateral a la izquierda

Ho: $\mu=\$10$
H1: $\mu\$10$



Ensayo bilateral

5.5 PLANTEAMIENTO GENERAL DE LAS PRUEBAS DE HIPOTESIS

El razonamiento expuesto anteriormente para probar una hipótesis de medias se puede resumir así:

1. Definir que es lo que se pretende probar con la investigación o experimento.

2. Aceptando que se ha hecho un enunciado claro del problema expresar Ho y H1 en términos cuantitativos, en función de algún (algunos) parámetro (s) de la (s) población (es) involucrada (s) en el estudio.

3. Elegir el nivel de significancia α.

4. Seleccionar el método de prueba de acuerdo a las condiciones del experimento y tomando en cuenta las suposiciones y limitaciones que se aplican a cada una de ellas (ver sección siguiente).

5. Elegir el tamaño de muestra adecuado para realizar el experimento.

6. Realizar el experimento.

7. Efectuar la prueba estadística.

8. Elaborar conclusiones de acuerdo al problema planteado.

Ejemplo:

Como parte de un estudio de tiempos y movimientos conducido en cierto Centro de Salud, una muestra de 100 pacientes demoró en promedio 17 minutos en la sala de espera entre su registro y su atención por un médico. La desviación estándar estimada para la población en investigaciones anteriores es de 10 minutos. Si se supone que la distribución del tiempo de espera es normal. ¿Proporciona el resultado de esta muestra suficiente evidencia para indicar que el tiempo medio de permanencia en la sala de espera es menor de 20 minutos?

1. Con la investigación se desea comprobar que los pacientes esperan menos de 20 minutos para se atendidos por un médico.

2. Las hipótesis pueden enunciarse:
   Ho. La medida del tiempo de espera de la población de pacientes es 20 minutos

   $$Ho: \mu=20$$

   H1 La medida del tiempo de espera es menor de 20 minutos (De acuerdo con el planteamiento del problema).

   $$H1: \mu<20$$

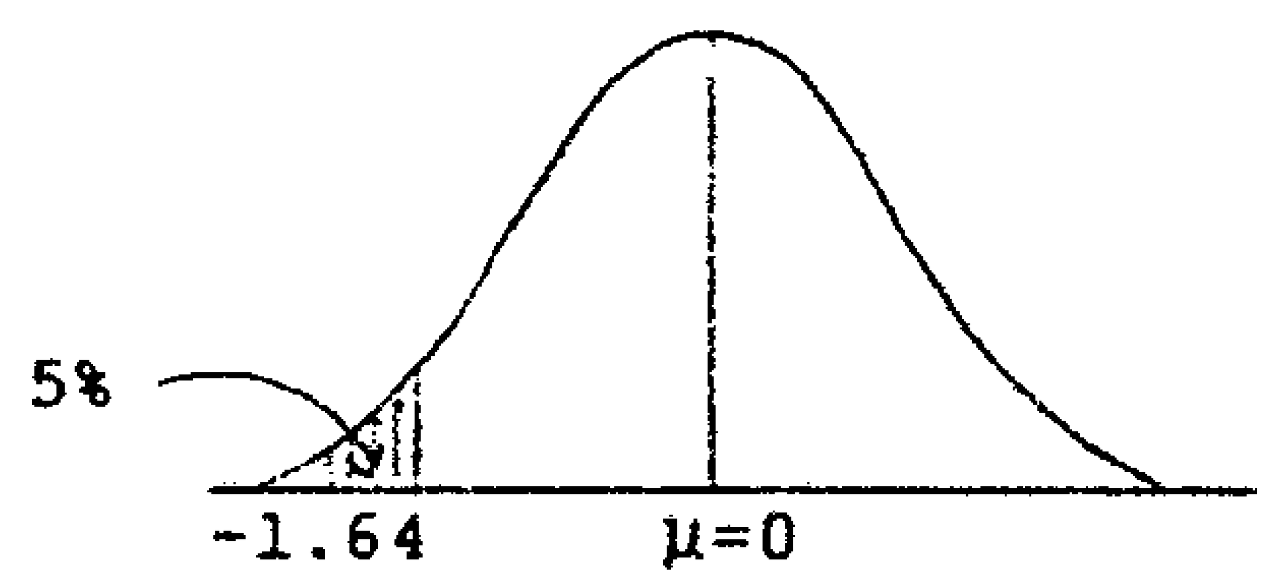
3. El nivel de significancia lo elige el investigador de acuerdo a las necesidades del estudio, $\alpha=0.05$.

4. Se supone que la población es normal con σ^2 conocida e igual a 10.

   El estadístico de prueba bajo estas condiciones es:

   $$Z= \frac{X' - \mu}{\sigma/\sqrt{n}}$$

   Zc= -1.64 un ensayo unilateral a la izquierda.



5. El tamaño de muestra ya está explícito en el planteamiento del problema n = 100

6. El experiemento se realizó y el resultado de la muestra es X'= 17 minutos

7. Prueba:

   $$Z= \frac{17 - 20}{10/\sqrt{100}} = -3 \qquad Z<Zc, -3 < -1.64$$

8. Los resultados de la muestra evidencian que el tiempo medio de espera es menor de 20 minutos.

Resuelva:

Una muestra de 25 estudiantes de primer año tuvo una calificación media de 77 puntos en una prueba para medir su actitud hacia el paciente.
Si se considera que la distribución de las calificaciones de los estudiantes es normal con varianza 100 puntos². ¿Proporiconan los datos suficiente evidencia que la media de la población es menor de 80 puntos.

1. *¿Qué se pretende con las investigación?*

2. *Ho:*

   H1:

3. $\alpha=$

4. *El estadístico de prueba es:*

5. $n=$

6. *Prueba:*

7. *Conclusiones:*

____

____

____

5.6 OTRAS PRUEBAS DE HIPOTESIS

Existen situaciones donde las condiciones de la población expuestas en la prueba anterior no son pertinentes, así mismo hay problemas que relacionan a más de una población o se refieren a parámetros diferentes, por lo que es necesario plantear nuevos Estadísticos de prueba.

El cuadro que se presenta a continuación muestra diferentes situaciones y los estadisticos que corresponden a cada una de ellas.

PRUEBAS APLICABLES A HIPOTESIS SOBRE EL COMPORTAMIENTO DE UNA POBLACION

Ho	Estadístico	Condición de la prueba	Est. de prueba	Distribución	Hl	Area de rechazo
$\mu = \mu_o$	X'	Poblaciones normales o casi normales con σ conocida, aplicable a: $n>30$ al conocer σ o estimarla por S. En poblaciones no normales si $n>30$.	$Z = \frac{X'-\mu}{\sigma/\sqrt{n}}$	Normal estandar	$\mu > \mu_o$ $\mu < \mu_o$ $\mu \neq \mu_o$	$Z>Zc$ $Z<Zc$ $Z=Zc$
$\mu = \mu_o$	X'	Poblaciones Normales con σ desconocida y $n<30$	$t = \frac{X'-\mu}{S/\sqrt{n}}$	de Student con n-1 grados de libertad	$\mu > \mu_o$ $\mu < \mu_o$ $\mu \neq \mu_o$	$t>tc$ $t<tc$ $t=tc$
$\sigma^2=\sigma_o^2$	S^2	Poblaciones Normales	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	Chicuadrado con n-1 grados de libertad	$\sigma^2>\sigma_o^2$ $\sigma^2<\sigma_o^2$ $\sigma^2\neq\sigma_o^2$	$\chi^2>\chi c^2$ $\chi^2<\chi c^2$ $\chi^2=\chi c^2$
$p=p_o$	P	Poblaciones binomiales $n>30$	$Z = \frac{P - p_o}{\sqrt{(pq/n)}}$	Normal Estandar	$p>p_o$ $p<p_o$ $p\neq p_o$	$Z>Zc$ $Z<Zc$ $Z=Zc$

Ejemplo:

Una muestra aleatoria de 20 profesores aparentemente sanos, presentó los siguientes valores de capacidad respiratoria máxima:

132,33,91,108,67,169,54,203,190,133,96,30,187,21,63,166,84, 110,157,138.

Si se supone que la población es normal, ¿Proporcionan éstos datos suficiente evidencia para concluir que la media poblacional es distinta a 110, nivel de significancia α 5%?

Con el estudio se pretende probar que la media de la capacidad respiratoria máxima es distinta a 110

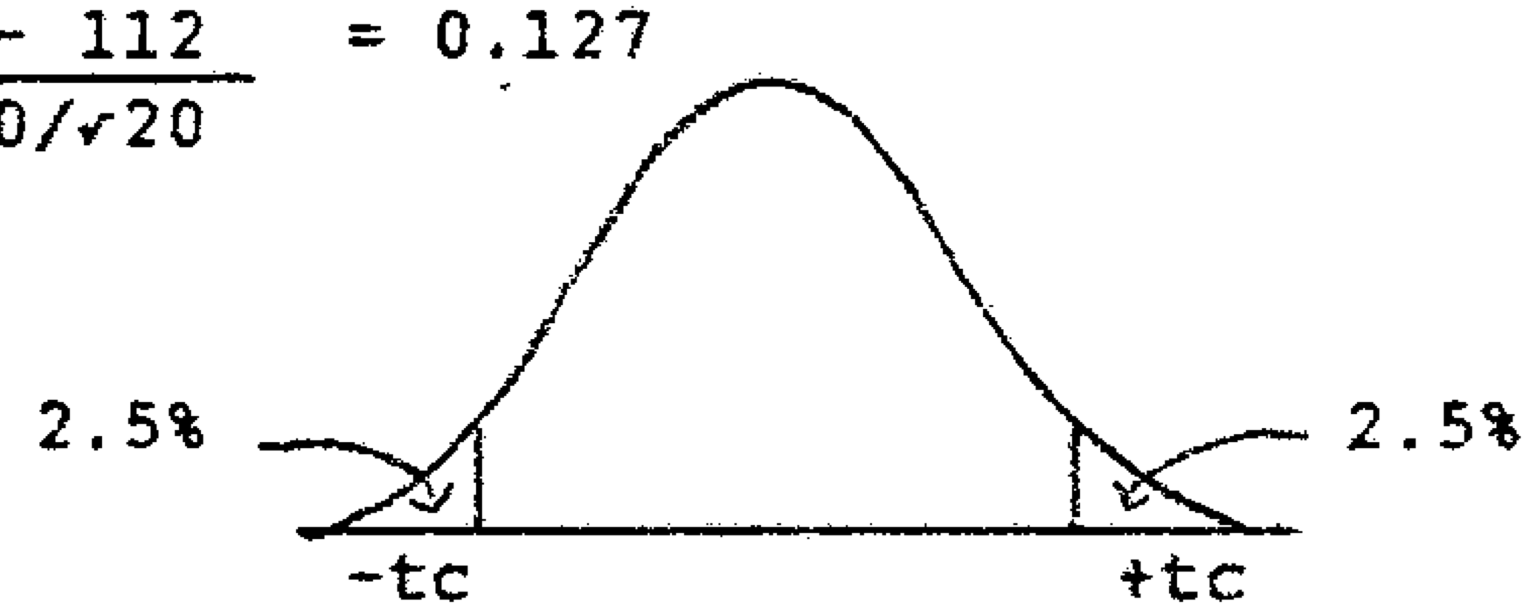
Ho: $\mu=110$ H1: $\mu \neq 110$

Población normal con σ desconocida, muestra pequeña.

Datos:
X'= 111.6 S= 56.30 n=20

El estadístico de prueba es:

$$t = \frac{X'-\mu}{S/\sqrt{n}} = \frac{111.6 - 112}{56.30/\sqrt{20}} = 0.127$$



tc=±2.86 Valor de t con 19 grados de libertad y α=5%

t< tc y t> -tc

Se acepta la Hipótesis Nula, la muestra no evidencia que la media de la población sea diferente a 110.

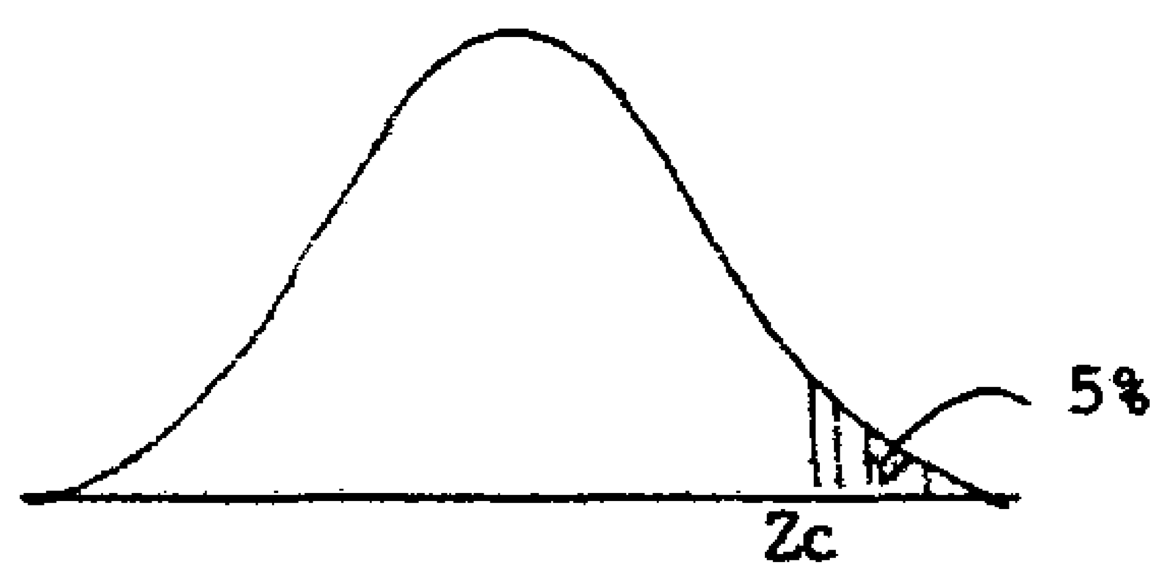
Una muestra de 100 empleados de un hospital, los cuales habían estado en contacto con Sangre, fué examinada por presentar evidencia serológica de hepatitis B. Se encontró que 23 de ellos tuvieron reacción positiva.
¿Puede concluirse a partir de esto que la proporción de positivos en la población muestreada es mayor de 15%? α=5%

Con el análisis de la muestra se pretende establecer si la proporcion de empleados con reacción positiva (éxitos) es

mayor de 15%.
Ho. $p=0.15$ H1: $p>0.15$

Datos;
$P= 23/100 =0.23$ $n=100$
Población binomial muestra grande

$Zc= 1.645$ valor de la Normal estándar con $\alpha=5\%$



Estadístico de Prueba

$$Z = \frac{P - p}{\sqrt{(pq/n)}} = \frac{0.23 - 0.15}{\sqrt{(0.15*0.85/100)}} = 2.24$$

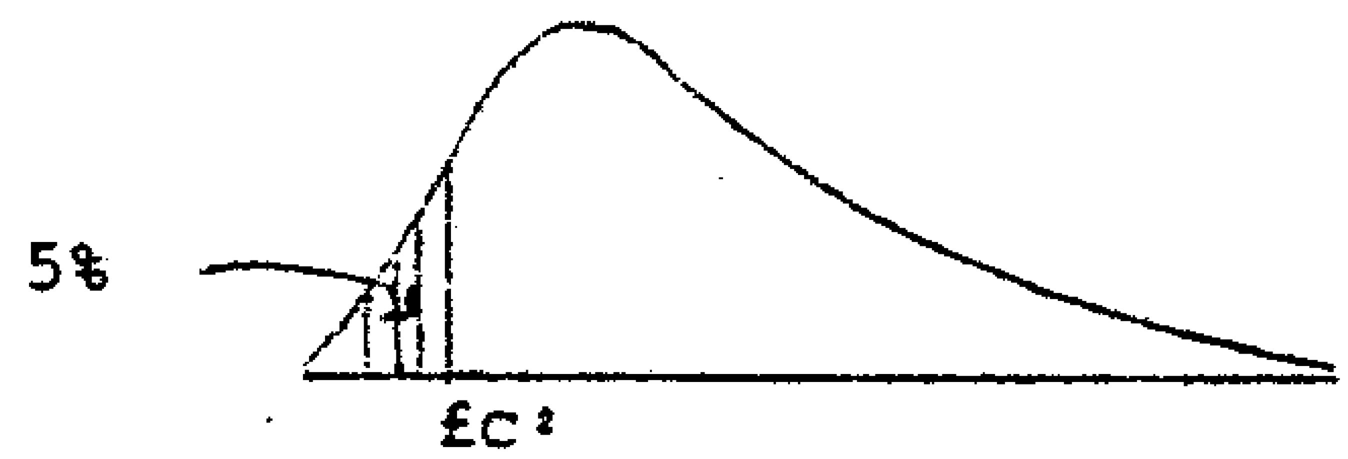
Z>Zc Se rechaza la hipótesis nula, la muestra evidencia que la proporción de empleados con reacción positiva es mayor de 15%

Se registraron los valores de la capacidad vital de una muestra de 10 pacientes con obstrucción crónica en las vías respiratorias. La varianza de las 10 observaciones fué de 0.75. Pruebe la hipótesis nula de que la varianza de la población no es menor que 1.00. $\alpha=5\%$

Se debe probar la hipótesis que la capacidad vital de los pacientes tiene una $\sigma^2= 1$
Ho: $\sigma^2=1$ H1: $\sigma^2< 1$

Datos:
$S= 0.75$ $n=10$
Se supone la población Normal

Estadístico de Prueba
$\chi^2= (n-1)S^2/\sigma^2 = 9*(0.75)^2/1^2 = 5.06$
$\chi c^2=$ 3.32 valor de Chi-cuadrado con $\alpha=5\%$ y 9 grados de libertad.



$£^{2} > £c^{2}$, 5.06 >3.32

Se acepta la hipótesis nula, la muestra no evidencia que la varianza de la población sea menor que uno.

Resuelva:

Un Director de hospital piensa que el 18 % de los empleados trabajan horas extras todas las semanas. Si la proporción observada en esta semana es de 0.13, en una muestra de 1250 empleados, de 2500 que tiene el hospital. ¿Se puede aceptar como razonable la opinión del Director?

Nueve animales de laboratorio fueron infectados con cierta bacteria y luego inmunosuprimidos. El número de organismos en promedio aislados posteriormente en los tejidos de dichos animales fue de 6.5 (datos codificados) con una desviación estándar de 0.6.
¿Puede concluirse a partir de éstos datos que la media de organismos aislados en la población de todos los animales que han sido o serán infectados con esa bacteria es mayor a 6?.
Use un nivel de significancia del 5%

Se supone que la desviación estándar de la distribución de una variable aleatoria es 300. Si una muestra de 20 de esa población muestra una desviación estándar de 250. ¿Se debe rechazar la hipótesis de que la desviación estandar es 300?
Utilice un nivel de significancia de 10%.

PRUEBAS APLICABLES A HIPOTESIS SOBRE EL COMPORTAMIENTO DE DOS POBLACIONES

Ho	Estadístico	Condiciones de la prueba	Est. de prueba	Distribución	H1	Área de rechazo
$\mu1=\mu2$ $\mu1-\mu2=0$ $\mu1-\mu2=d$	$X1'-X2'$	Poblaciones normales con $\sigma1$ y $\sigma2$ conocidas, Muestras extraídas en forma independiente de cada población, no necesariamente del mismo tamaño. Aplicable cuando n1 y n2 son mayores de 30	$Z=\dfrac{(X1'-X2')-d}{\sqrt{(\frac{\sigma1^2}{n1}+\frac{\sigma2^2}{n2})}}$	Normal Estándar	$\mu1-\mu2>d$ $\mu1-\mu2<d$ $\mu1-\mu2\neq d$	$Z>Zc$ $Z<Zc$ $Z\ Zc$
$\mu1-\mu2=d$	$X1'-X2'$	Poblaciones Normales con varianzas iguales pero desconocidas, estimadas por Sp. Muestras independientes. $Sp=\dfrac{(n1-1)S1^2+(n2-1)S2^2}{n1+n2-2}$	$t=\dfrac{(X1'-X2')-d}{Sp\sqrt{(\frac{1}{n1}+\frac{1}{n2})}}$	Student con n1+n2-2 grados de libertad	$\mu1-\mu2>d$ $\mu1-\mu2<d$ $\mu1-\mu2\neq d$	$t>tc$ $t<tc$ $t=tc$
$\mu1-\mu2=d$	$X1'-X2'$	Poblaciones Normales con varianzas desconocidas y no iguales, estimadas por S1 y S2 Muestras Independientes. $v=\dfrac{(S1^2/n1+S2^2/n2)^2}{\frac{(S1^2/n1)^2}{n1-1}+\frac{(S2^2/n2)^2}{n2-1}}$	$t=\dfrac{(X1'-X2')-d}{\sqrt{(S1^2/n1+S2^2/n2)}}$	Student con v grados de libertad.	$\mu1-\mu2>d$ $\mu1-\mu2<d$ $\mu1-\mu2\neq d$	$t>tc$ $t<tc$ $t=tc$
Comparación por parejas $\mu(D)=d$	$D=X1i-X2i$ $X'(D)$ $S(D)$	Muestras no independientes aplicable cuando se quiere comparar muestras del mismo tamaño y a cada valor de le corresponde un valor de la otra.	$t=\dfrac{X'(D)-d}{S(D)/\sqrt{n}}$ n= parejas	Student con n-1 grados de libertad	$\mu(D)>d$ $\mu(D)<d$ $\mu(D)\neq d$	$t>tc$ $t<tc$ $t=tc$
$\sigma1^2=\sigma1^2$ $\sigma1^2/\sigma2^2=1$	S1,S2	Poblaciones Normales Muestras independientes	$F=S1^2/S2^2$ $S1^2$ Varianza mayor $S2^2$ Varianza menor.	F con (n1-1),(n2-1) grados de libertad.	$\sigma1^2/\sigma2^2>1$	$F>Fc$
$p1-p2=d$ $p1=p2$ $p1-p2=d=$	P1,P2	Poblaciones Binomiales Con proporciones de éxitos p1,p2 y muestras grandes $p=(n1p1+n2p2)/(n1+n2)$ $q=1-p$	$Z=\dfrac{(P1-P2)-d}{\sqrt{(\frac{P1Q1}{n1}+\frac{P2Q2}{n2})}}$ $Z=\dfrac{P1-P2}{\sqrt{(pq(\frac{1}{n1}+\frac{1}{n2})}}$	Normal	$p1-p2>d$ $p1-p2<d$ $p1-p2\neq d$	$Z>Zc$ $Z<Zc$ $Z=Zc$

Ejemplos:

Un grupo de 10 sujetos se someten a un estímulo suave, la tabla siguiente muestra las mediciones de presión sanguínea en milímetros antes y después del estímulo.

Antes	Después	Diferencia D
118	127	9
120	128	8
128	136	8
124	131	7
136	138	2
132	132	0
130	131	1
140	141	1
140	132	-8
128	120	-8

¿Soportan los datos en forma aceptable la hipótesis que el estímulo aumenta la presión sanguínea?

Se debe probar la hipótesis que el estímulo aumenta la presión.

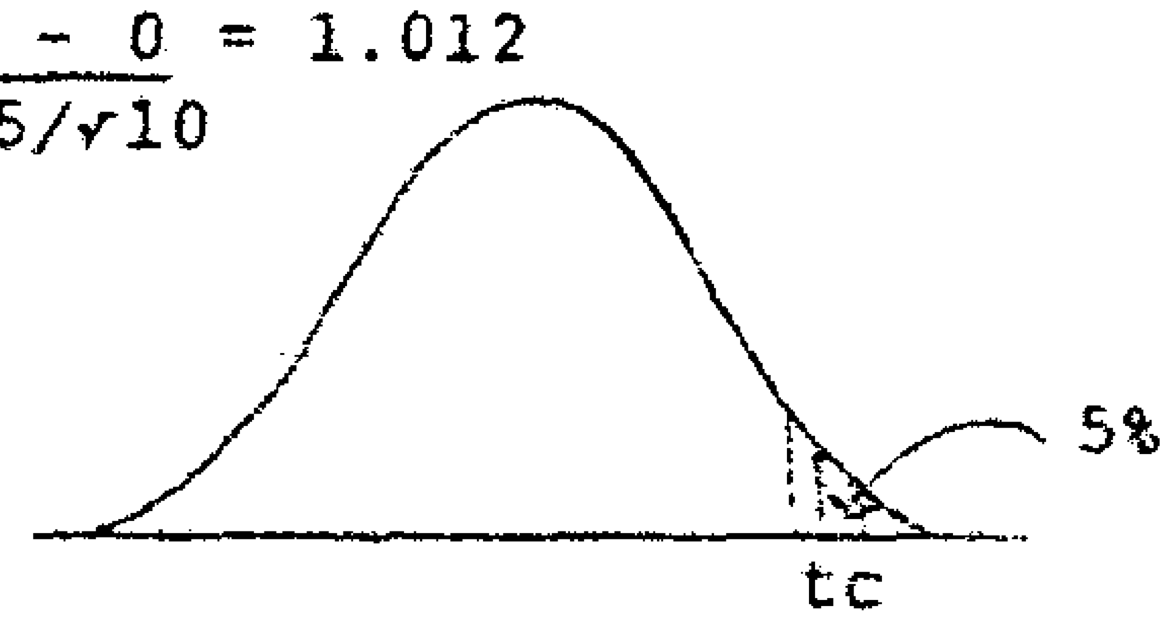
Ho: $\mu(1)-\mu(2)=0$ Ho: $\mu(D)=0$
H1: $\mu(1)-\mu(2)>0$ H1: $\mu(D)>0$

Se supone que las poblaciones son normales.
Observaciones por parejas (antes y después del estímulo a las mismas personas).

Datos:
n=10   X'(D)= 2   S(D)=6.25

Estadístico de Prueba

$$t= \frac{X'(D) - \mu(D)}{S(D)/\sqrt{n}} = \frac{2 - 0}{6.25/\sqrt{10}} = 1.012$$



tc= 1.645, valor de T con 9 grados de libertad y un nivel de significancia de 5%.
t<tc 1.012< 1.645, se acepta la hipótesis Nula.

Los datos no evidencian que el estímulo aumente la presión sanguínea.

Una preparación normalizada de Protombina, que se usa como control, se prueba en dos días diferentes, midiendo el tiempo en segundos que tarda en formarse un cuágulo. Los resultados de 5 mediciones hechas cada día son los siguientes:

Primer día: 11.3, 11.7, 12.8, 13.0, 12.1
Segundo día: 13.0, 13.6, 13.7, 13.9, 11.9

Se considera que un aumento de un segundo en el tiempo de coagulación es evidencia suficiente de que se ha deteriorado la preparación y hay que componer una nueva.
¿Los datos evidencian que efectivamente ha habido un aumento superior a un segundo? Se utiliza un nivel de significancia de 5%.

Con el análisis se pretende determinar si el tiempo de coagulación a aumentado más de un segundo.
Ho: $\mu 2-\mu 1 = 1$
H1: $\mu 2-\mu 1 > 1$

Datos:
X1'= 12.8 S1= 0.72
X2'= 13.42 S2= 0.443

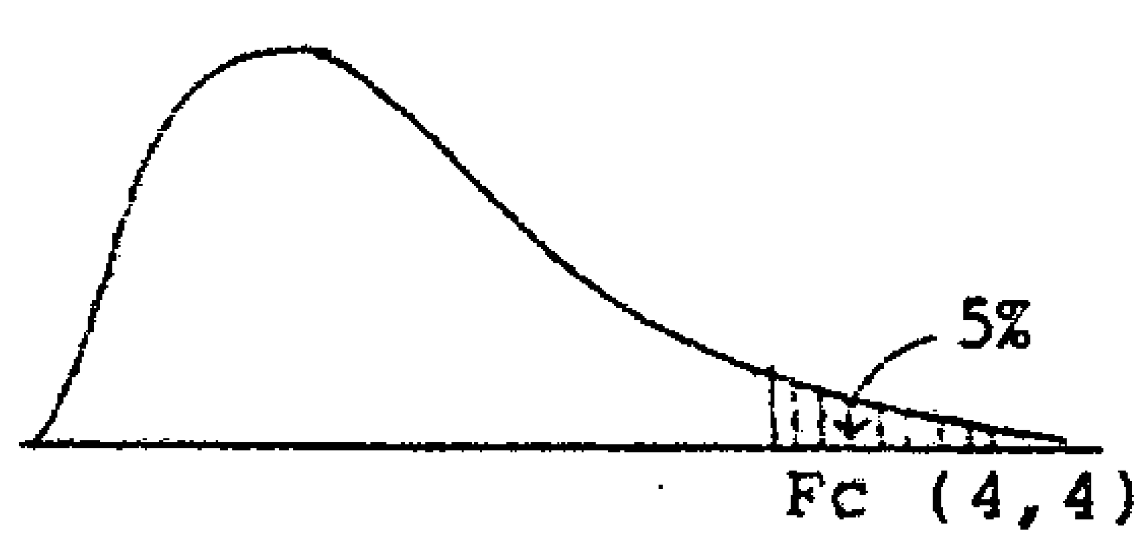
Se supone que las poblaciones son normales e independientes.

Previo a seleccionar el estadístico de prueba, se debe evaluar el comportamiento de las varianzas.

Ho: $\sigma 1^2=\sigma 2^2$
H1: $\sigma 1^2>\sigma 2^2$

Estadístico de Prueba:

$F= S1^2/S2^2 = 0.72^2/0.443^2 = 2.642$



Fc= 6.35 , de la tabla de la distribución F con (4,4) grados de libertad.

F<Fc se acepta la hipótesis nula.
Los datos no evidencian que las varianzas sean diferentes.

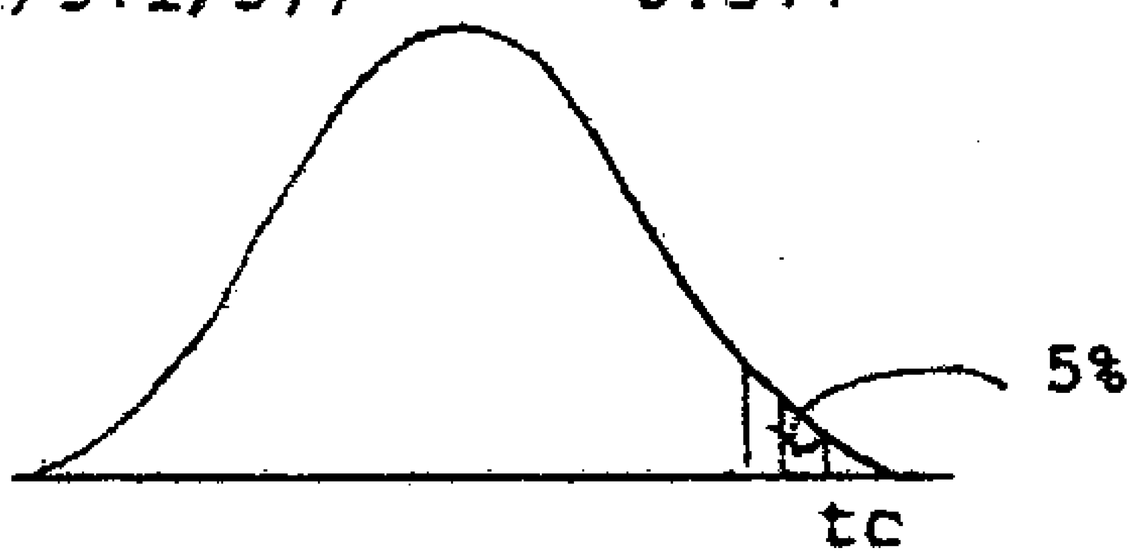
Para las hipótesis iniciales se supone que las varianzas son desconocidas pero iguales, estimadas por Sp^2.

$Sp^2 = (4(0.72^2) + 4(0.443^2)) / 8 = 0.357$
$Sp = 0.597$

Estadístico de Prueba

$$t = \frac{[X1'-X2'] - [\mu1-\mu2]}{Sp\ (\sqrt{(1/n1 + 1/n2)}}$$

$$t = \frac{(13.42 - 12.18) - 1}{0.597(\sqrt{(1/5+1/5)})} = \frac{0.24}{0.377} = 0.636$$



tc= 1.833 de la tabla de la distribución T con 8 grados de libertad.

$t<tc \quad 0.636<1.833$

Se acepta la hipótesis nula, la muestra no evidencia que haya aumentado el tiempo de coagulación.

Se están evaluando dos métodos para determinar la concentración de Oxígeno en la Sangre, una muestra de 10 observaciones con el método A produjo una desviación de 10 respecto a la media. Con el método B una muestra de 8 valores porodujo una desviación de 8 respecto a la media.
¿Apoyan esos resultados la hipótesis que no hay diferencia en la precisión de los métodos?

Se desea probar si la variabilidad es igual en ambos métodos.

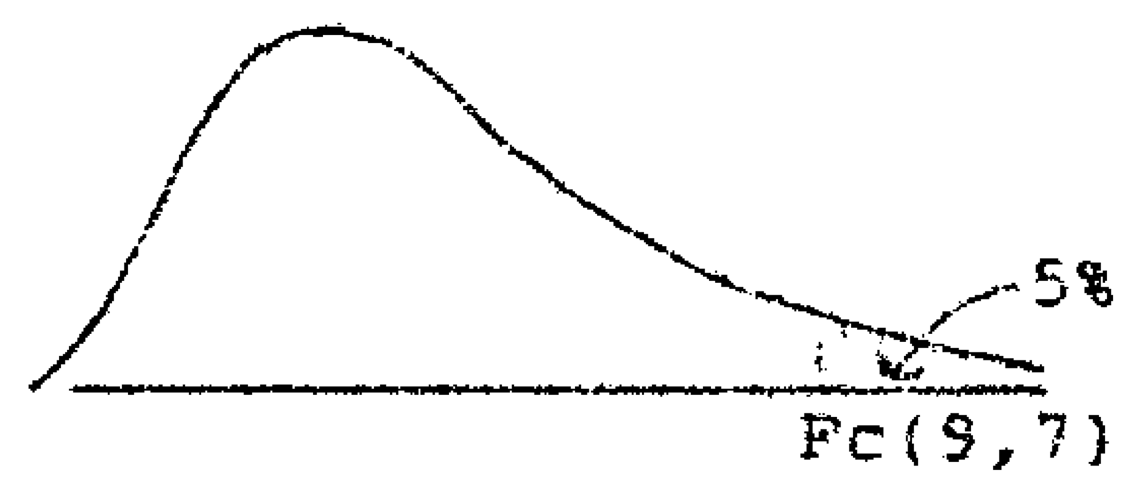
Ho: $\sigma1^2=\sigma2^2$

H1: $\sigma1^2>\sigma2^2$

Se asume que las dos poblaciones tienen una distribución normal.

Datos:
S1= 10 n1=10
S2= 8 n2=8

$F= S1^2/S2^2= 10^2/8^2= 1.56$



Fc= 3.68 de la tabla de la distribución F con (9,7) grados de libertad y un nivel de significancia del 5%.

$$F<Fc, \ 1.56 < 3.68$$

Se acepta la hipótesis nula, de varianzas iguales.

Un grupo de médicos clínicos, está efectuando pruebas con algunos pacientes para determinar la eficiencia de un nuevo antihipertensivo, un grupo de enfermos con hipertensión fueron elegidos al azar y luego asignados aleatoriamente, al grupo de control, que recibía un antihipertensivo bien probado, o al grupo experimental, que recibía el nuevo fármaco. Los médicos anotaron el porcentaje de pacientes cuya presión sanguínea se redujo a un nivel normal en el lapso de un año. Con un nivel de significancia de 0.01 pruebe la hipótesis apropiada para determinar si el nuevo medicamento es significativamente más eficaz que el anterior para reducir la hipertensión.

Grupo	Proporción	Número de pacientes
Experimental	0.45	120
Control	0.36	150

Si el medicamento es más eficaz, la proporción de pacientes cuya presión sanguínea se reduce a nivel normal es mayor en el grupo experimental que en el de control.

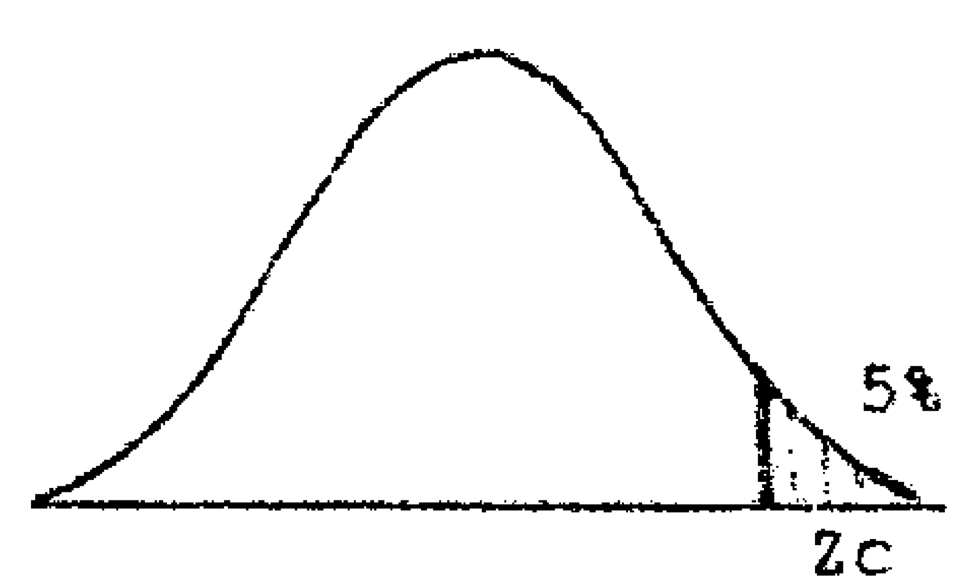
Ho: p2=p1

H1: p2>p1

Datos:
P1=0.36 n1= 120
P2=0.45 n2= 150

Estadístico de Prueba:

$$Z = \frac{(P1 - P2) - 0}{\sqrt{(pq(1/n1+1/n2))}}$$



$$p = \frac{54 + 54}{270} = 0.40$$

$$Z = \frac{0.45 - 0.36}{\sqrt{(0.4*0.6(1/120+1/150))}} = 1.5$$

Zc= 2.33 con un nivel de significancia de 1%, distribución normal estándar.

$Z<Zc \quad 1.5 < 2.33$

Se acepta la hipótesis Nula, los datos no evidencian que el nuevo medicamento sea más efectivo.

Resuelva:

Se usan 10 pares de ratas para comparar dos dietas, sus pesos en gramos se muestra en la tabla siguiente.

Dieta A	*Dieta B*
122	*110*
120	*108*
112	*94*
126	*138*
112	*102*
126	*122*
118	*114*
112	*108*
88	*124*
122	*116*

¿Puede concluirse que existe diferencia significativa en el peso promedio de las ratas con ambas dietas?

Se quiere establecer si el nivel de fluorocal salivarus encontrado en la saliva, difiere entre dos vacunas. Pruebe la hipótesis que las dos vacunas anticaries producen el mismo nivel de componente fluorocal salivarus, si dos muestras de ocho individuos cada una presentan los siguientes resultados:

Flurocal Salivarus presenta después de tres días en la saliva

Vacuna A	*Vacuna B*
22	30
17	18
19	27
30	42
37	35
18	37
24	31
20	29

Se efectuó un estudio para comparar la resistencia compresiva de dos tipos de aleaciones de amalgama dental tres minutos después de la trituración y condensación.
Se ensayaron 31 mezclas de la aleación A, una amalgama de endurecimiento rápido, y 31 mezclas de B, una amalgama de endurecimiento lento. Los resultados fueron los siguientes:

$n_A = 31$ $\quad X_A' = 9.530$ psi $\quad S_A^2 = 3.410$ psi^2
$n_B = 31$ $\quad X_B' = 9.420$ psi $\quad S_B^2 = 2.88$ psi^2

Determinar si hay una diferencia significativa a un nivel de 0.01 entre las resistencias compresivas de las dos aleaciones.

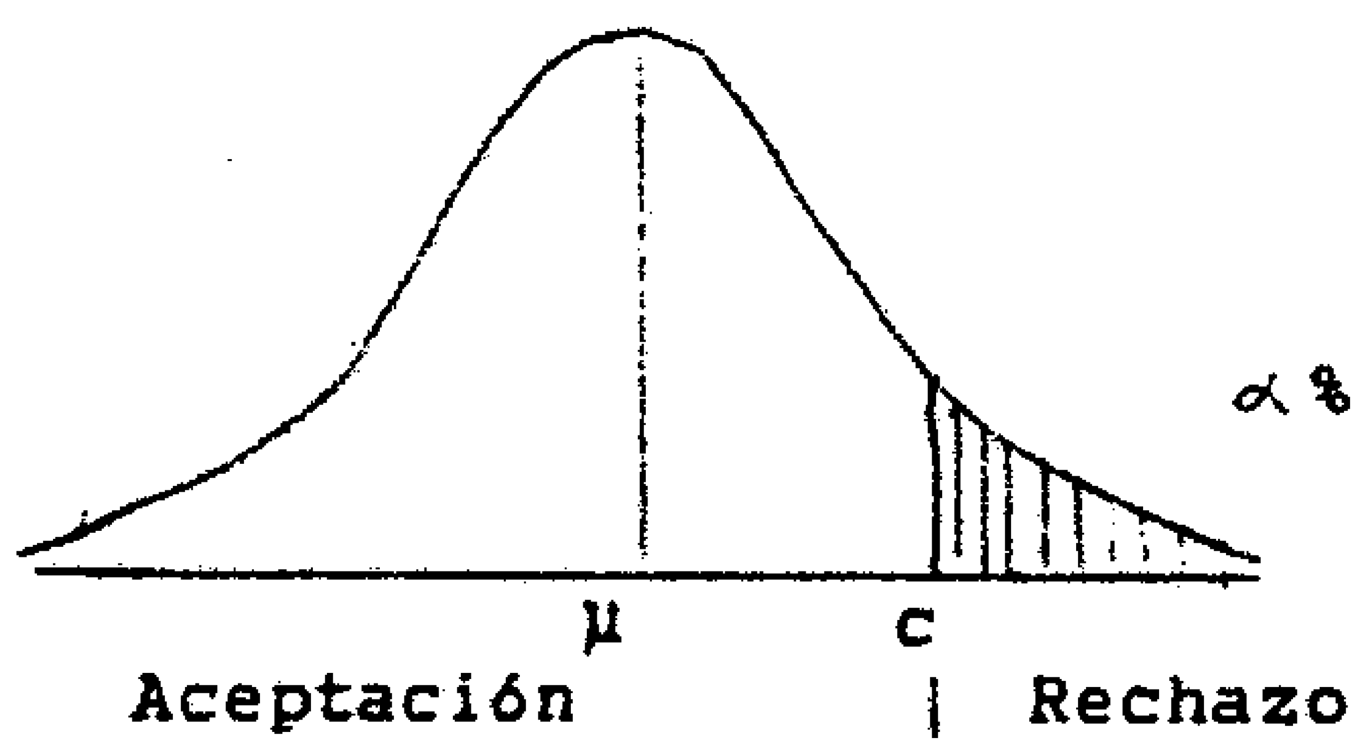
5.7 RIESGOS DE HACER FALSAS DECISIONES

Cuando se toma la decisión de rechazar o no rechazar una hipótesis se corre el riesgo de cometer uno de los siguientes errores:

ERROR TIPO I Rechazar la hipótesis nula cuando esta es verdadera.

El error tipo I está ligado al planteamiento de la hipótesis nula y al nivel de significancia α. Anteriormente se dijo que α representa una probabilidad lo suficientemente pequeña que si la Ho es verdadera, la muestra presente un valor del estadístico significativamente diferente a ella, por ocurrir este hecho , poco probable pero posible, la Ho debe ser rechazada, sin embargo la decisión está equivocada porque Ho es verdadera.

Por ejemplo en la hipótesis para la media de una población Normal con Varianza conocida si Ho: $\mu=\mu o$ y H1: $\mu > \mu o$



Ho. se rechaza si el estadístico X' es mayor que c, esto es, se rechaza porque un valor de X' mayor que c es poco probable que ocurra si Ho es vardadera, pero Ho puede ser cierta con una probabilidad α%. Al rechazarla se está cometiendo un error, el Error Tipo I.

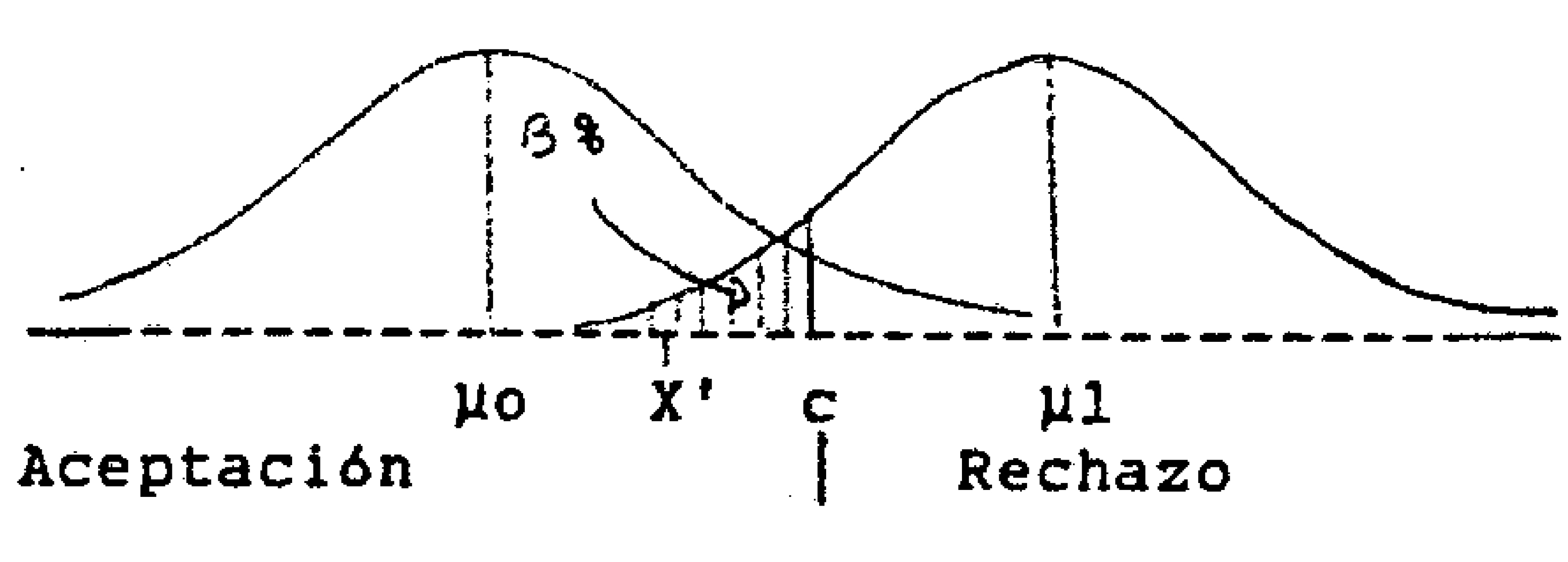
α es la probabilidad de rechazar una hipótesis nula cuando es verdadera, es la probabilidad de comenter el error tipo I.

ERROR TIPO II Aceptar la hipótesis nula cuando es falsa.

El error tilpo II está ligado con el planteamiento de la hipótesis alternativa, pués cuando Ho es falsa, el verdadero valor del parámetro se encuentra contenido en el intervalo que H1 representa.

La magnitud del Error tipo II se representa por β y es la probabilidad que el estadístico muestre un valor consistente con Ho a pesar que es falsa.

Supónga como ejemplo que Ho es falsa que $\mu= \mu 1$ ($\mu 1 > \mu o$), entonces si el estadístico X' ofrece un valor menor que c, Ho se acepta por norma, sin embargo X' es una observación aleatoria de la distribución de la población centrada en $\mu 1$ y la probabilidad que X' sea menor que c es β %. Por presentar X' un valor menor que c se acepta Ho pero es falsa $\mu=\mu 1$, se está cometiendo un error, el Error Tipo II.



β es la probabilidad de aceptar la hipótesis nula cuando es falsa, es la probabilidad de cometer el error Tipo II.

β varía dependiendo del verdadero valor del parámetro, si H1 es una hipótesis compuesta, β va a existir para cada posible valor que cumpla con ella.

	Ho verdadera	Ho falsa
Se acepta Ho	Decisión correcta	Error Tipo II
Se rechaza Ho	Error Tipo I	Decisión correcta

Ejemplo:

Un Odontólogo supone que la proporción de pacientes que regresan al chequeo semestral después de haber terminado el tratamiento es del 50%, quiere probar esa hipótesis; para ello decide observar el comportamiento de 10 de sus pacientes, si al menos dos regresan , acepta su hipótesis. Los errores Tipo I y Tipo II asociados a este experimento son:

Ho: $p=0.50$

H1: $p<0.50$ Una de las alternativas que se incluyen en ese planteamiento es $p=0.40$

Ahora, si Ho es cierta, $p=0.50$, la muestra de 10 pacientes puede presentar los siguientes resultados con sus respectivas probabilidades:

Número de Pacientes que regresan X	probabilidad*
0	0.001
1	0.01
2	0.044
3	0.117
4	0.205
5	0.246
6	0.205
7	0.117
8	0.044
9	0.01
10	0.001

* Calculadas a partir de la distribución binomial $p=q=0.5$

La probabilidad de aceptar Ho si es cierta es $P(X>=2)=0.988$
La probabilidad de rechazar Ho cuando es cierta, α, es $P(X<2)=0.011$

El criterio de decisión implica un riesgo de 1.1% de rechazar la hipótesis cuando es cierta, 1.1% la probabilidad de cometer el error tipo I.

Si la hipótesis verdadera fuera H1, p=0.40, la muestra presentaría la siguiente distribución de probabilidades.

Número de Pacientes que regresan X	Probabilidad*
0	0.006
1	0.04
2	0.121
3	0.215
4	0.251
5	0.201
6	0.111
7	0.042
8	0.008
9	$\rightarrow 0$
10	$\rightarrow 0$

* Calculada a partir de la distribución binomial con p=0.4 q=0.6

Sin embargo si X>=2 la decisión será aceptar la hipótesis nula Ho.
$P(x>=2) = 0.954 = \beta$

β es la probabilidad que no obstante p=0.4 se acepta la hipótesis nula p=0.5 por la evidencia de la muestra.
La probabilidad de rechazar Ho cuando es falsa es:

$1-\beta=1-0.954=0.046$ y se conoce como potencia de la prueba.

La Potencia de la prueba $(1-\beta)$ es la probabilidad de detectar que la Ho es falsa y depende de la magnitud de la diferencia entre el valor planteado en ésta y el verdadero valor del parámetro, que puede encontrarse incuído en el intervalo de H1.

ALGUNAS OBSERVACIONES SOBRE LOS ERRORES:

- Los errores tipo I y tipo II están relacionados, un decremento en la probabilidad de ocurrencia de uno de ellos generalmente conduce a un incremento en la probabilidad de ocurrencia del otro.

- Un incremento en el tamaño de la muestra del experimento reduce simultaneamente a α y β.

- Si la hipótesis nula es falsa, entonces β se hacees máxima a medida que el valor verdadero del parámetro se aproxima al valor hipotético. Mientras más grande es la distancia entre los valores: hipotético, planteado en Ho, y el verdadero, ß es más pequeña.

- El investigador es el responsable de fijar cada uno de los errores, dándoles la importancia que merecen de acuerdo a la naturaleza del experimento.

- Es necesario estar concientes que siempre se comete un error tipo II.

5.8 ELECCION DEL TAMAÑO DE LA MUESTRA

Para controlar la magnitud de los errores Tipo I y Tipo II, es necesario utilizar en el experimento el tamaño de muestra apropiado. A continuación se presentan algunas fórmulas aplicables en el cálculo del tamaño de la muestra cuando la o las varianzas de las poblaciones son conocidas.

Hipotesis relacionadas con medias

$$Ho: \mu=\mu. \qquad H1: \mu>\mu.$$

Con nivel de significancia α, una potencia de la prueba $1-\beta$, y se desean correr esos riesgos máximos cuando el verdadero valor de μ difiera del hipotético $\mu.$ en δ.

$$n= \frac{(Z(\alpha)+Z(\beta))^2 * \sigma^2}{\delta^2}$$

$Z(\alpha)$ y $Z(\beta)$ son los valores de la variable Normal Estándar con α y β áreas acumuladas respectivamente.

Hipótesis relacionadas con diferencias se medias

$$Ho: \mu1-\mu2 = d. \qquad H1: \mu1-\mu2 \neq d.$$

Con nivel de significancia α, potencia de la prueba $1-\beta$ y se desean correr esos riesgos máximos cuando el verdadero valor de d difiera del hipotetico d. en δ.

$$n= \frac{(Z(\alpha)+Z(\beta))^2 * (\sigma1^2+\sigma2^2)}{\delta^2}$$

$Z(\alpha)$ y $Z(\beta)$ son los valores de la variable Normal Estándar con áreas acumuladas de $\alpha/2$ y β respectivamente.

Por ejemplo:

Suponga que se desea probar la hipótesis:

Ho: $\mu=68$ kg. H1: $\mu>68$ kg.

Utilizando un nivel de significancia de 5% y conociendo σ igual a 5 kg.

El tamaño requerido de muestra para tener una potencia del 90% cuando la verdadera media es 69 kg., se obtiene de la forma siguiente:

$\delta^2=(68-69)^2= 1$

$Z(\alpha)= Z(5\%)= 1.645$

$Z(\beta)= Z(10\%)= 1.28$

$$n= (1.645+1.28)^2 * 5^2/1 = 213.16$$

$$n= 214$$

Resuelva:

Un investigador afirma que los ratones cuyo promedio de vida es de 32 meses podrían vivir 40 meses cuando el 40% de las calorías de su comida sea reemplazada por proteínas y vitaminas. Si la distribución de los tiempos de vida es normal con una desviación estándar de 5.8 meses. ¿Qué tamaño de muestra debe seleccionarse para que la probabilidad de cometer un error tipo II sea 0.10 cuando la verdadera media es 35.9 meses, y en la prueba se usará un nivel de significancia de 2.5%?

Ejemplo:

Supónga que se desea probar la hipótesis:

Ho: $\mu1 = \mu2$, $\mu1-\mu2=0$

H1: $\mu1 = \mu2$, $\mu1-\mu2$ 0

Con un nivel de significancia de 5%.

Se considera que las desviaciones de ambas poblaciones son iguales. Además es esencial detectar una diferencia (d) de 0.8σ entre las medias de las poblaciones y se desea tener una probabilidad de cometer el error tipo II del 10%. El tamaño de muestra a seleccionar es:

$\delta^2= (d.-d)^2 = (0.8\sigma)^2$

$\sigma1=\sigma2=\sigma$

$Z(\alpha/2)= Z(2.5\%)= 1.96$

$Z(\beta)= Z\ (10\%) = 1.28$

$$n=(1.96 + 1.28)^2*(2\sigma^2)/(0.8\sigma)^2= 32.8$$

$$n= 33$$

Resuelva:

Supónga dos poblaciones normales con $\sigma 1= 6.8$ y $\sigma 2= 5.6$.
Se desea probar la hipótesis $Ho: \mu 1=\mu 2$ $H1: \mu 1 \neq \mu 2$
Con un nivel de significancia de 5%.
¿Qué tamaño de muestra debe seleccionarse a fin de tener una potencia de la prueba del 5% cuando la verdadera diferencia entre las medias de la población sea 8?

5.9 EL ANALISIS DE VARIANZA

Es el procedimiento empleado para probar la hipótesis de que cierto número de muestras proceden todas de poblaciones con igual media, basándose en las suposiciones que las poblaciones están distribuidas en forma normal y con varianzas iguales.

El análisis de Varianza se utiliza cuando se quiere inferir sobre las diferencias significativas entre las media de varias muestras.

$$Ho: \mu 1 = \mu 2 = \ldots\ldots = \mu k$$

H1: No todas las medias poblacionales son iguales.

Mediante el análisis de varianza la variación total presente en un conjunto de datos se distribuye en varios componentes, asociando a cada uno de ellos una fuente especifica de varianción, de modo que en el análisis es posible averiguar la magnitud de las contribuciones de cada una de las fuentes de variación.

5.10 ANALISIS DE VARIANZA DE UN FACTOR.

Se utiliza un arreglo experimental sencillo donde las diferentes k poblaciones se clasifican con base a un sólo criterio, tratamientos,y de cada una de ellas se toman Ni observaciones (i=12...k).

	TRATAMIENTOS					
MUESTRA	1	2	3	i	**	k
1	X11	X21	X31	Xi1	**	Xk1
2	X12	X22	X32	Xi2	**	Xk2
3	X13	X23	X33	Xi3		Xk3
.	...	...	...	...	..	...
j	X1j	X2j	X3j	Xij		Xkj
.	...	...	...	...	..	...
.	...	...	...	...	..	...
N	X1N	X2N	X3N	XiN		XkN

Por Ejemplo:

En un estudio sobre el efecto de la glucosa sobre la liberación de insulina, se trataron muestras de animales de laboratorio con cinco estimulantes distintos, el interés recae en probar si los estimulantes influyen sobre la liberación de insulina.

En este caso las muestras representan 5 poblaciones diferentes, cada una influenciada por un factor de varianción que son los estimulantes, 5 tratamientos.

Liberación de insulina

estimulantes

1	2	3	4	5
x11	x21	x31	x41	x51
x12	x22	x32	x42	x52
.	.	.	.	.
.	.	.	.	.

Otra situación:

Se desea comparar a tres odontólogos respecto a la duración de una consulta específica de sus pacientes, para ello se seleccionaron 8 pacientes de cada odontólogo ($Ni=8$ para todo $i=1, 2, 3..$), en este caso las muestras son representativas de tres poblaciones diferentes, el factor de variación es el odontólogo, tres tratamientos.

Tiempo de consulta

Odontólogo

1	2	3
X11	X21	X31
X12	X22	X32
.	.	.
.	.	.
X18	X28	X38

Resuelva

Otro planteamiento en el cual se puede utilizar el análisis de varianza de un factor es:

PRUEBA DE HIPOTESIS

El análisis se lleva a cabo descomponiendo la Variación Total de las muestras en dos componentes: una que corresponde al factor cuyo efecto deseamos investigar, tratamientos; la otra corresponde al factor aleatorio. Con cada una de las componentes se obtiene una estimación de la varianza poblacional, si el efecto de los tratamientos no tienen influencia significativa, es decir si $\mu 1=\mu 2=....=\mu k$, las dos estimaciones de las varianzas deben ser iguales y su razón debe ser aproximadamente igual a uno. En caso que la razón no se aproxime a uno puede ser evidencia que los tratamientos si tienen influencia significativa y las μi no son todas iguales.

El análisis de varianza se basa en una comparación de dos distintas estimaciones de la varianza σ^2 de la población.

La estimación basada en la varianza entre las medias muestrales de los tratamientos y la estimación determinada por la variancián dentro de los grupos de tratamientos.

Variación entre tratamientos: Cuadrado Medio entre tratamientos (CMa^2)

$$CMa^2 = \frac{\Sigma\ Ni\ (Xi'-X')^2}{k-1} = \frac{SCa}{k-1} = \frac{\text{suma de cuadrados del factor}}{\text{grados de libertad del factor}}$$

Variación dentro los grupos: Cuadrado Medio dentro de grupos (CMe^2)

$$CMe^2 = \frac{\Sigma\ \Sigma\ (xij - Xi')^2}{(\ \Sigma\ Ni) - k} = \frac{SCe}{(\Sigma Ni)-k}$$

$$CMe^2 = \frac{\text{Suma de cuadrados dentro de los grupos}}{\text{Grados de libertad del factor aleatorio}}$$

Variación Total: Cuadrado Medio (CM)

Fórmulas para calcular la suma de cuadrados

$$SCt = SCa + SCe$$

$$SCe = SCt - SCa$$

$$SCt = \Sigma\ \Sigma\ Xij^2 - \frac{(\Sigma\ \Sigma\ Xij)^2}{\Sigma\ Ni}$$

$$SCa = \Sigma\ \left(\frac{(\Sigma\ Xij)^2}{Ni} \right) - \frac{(\ \Sigma\Sigma\ Xij)^2}{\Sigma Ni}$$

PRUEBA F

Si Ho: $\mu 1=\mu 2= \ldots = \mu k$, es verdadera

$$F= \frac{\text{Variación entre taratamientos}}{\text{Variación detro de tratamientos}}$$

Debe tender a uno, ya que las estimaciones de la varianza son iguales.

Cuando las poblaciones no son iguales las variación entre tratamientos tenderá a ser mayor que la variación dentro de tratamientos, y el valor de F tendrá que aumentar.

Ho': $CMa^2/CMe^2=1$
H1': $CMa^2/CMe^2 >1$

Estadístico de prueba

$$F= \frac{SCa^2/\ (k-1)}{SCe^2/\ (N-k)} \qquad N= \Sigma Ni$$

F tiene una distribución F con (k-1),(N-k) grados de libertad.

Fc es el valor de la variable F con nivel de significancia α y (k-1),(N-k) grados de libertad.

Si F <Fc aceptar la hipótesis nula de igualdad de medias. Si F >Fc rechazar la hipótesis nula, al menos una de las poblaciones tiene media diferente.

TABLA RESUMEN DEL ANALISIS DE VARIANZA

Fuente de variación	SC	grados de Libertad	CM	F
Tratamientos	SCa	k-1	SCa/(k-1)	$\frac{SCa/(K-1)}{SCe/(N-k)}$
Aleatroria	SCe	N-k	SCe/(n-k)	
Total	SCt	N-1	SCt/(N-1)	

Ejemplo

Un Odontólogo desea estudiar el efecto de la hora del día en la duración de las visitas que realizan sus pacientes, para ello reunió los siguientes datos:

PROPIEDAD DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
Biblioteca Central

Duración de las visitas

Paciente	Temprano por la mañana	Tarde por la mañana	Temprano por la tarde	Tarde por la tarde
1	27	28	30	23
2	31	30	27	20
3	35	38	34	30
4	20	18	20	14
ΣXij	113	114	111	87
ΣXij^2	3315	3452	3185	2025

$\Sigma\Sigma Xij = 425$
$\Sigma\Sigma Xij^2 = 11977$

Ho: La hora de la visita es un factor que no tiene efecto en el tiempo de misma. Ho: $\mu 1=\mu 2=\mu 3=\mu 4$

H1: El tiempo de visita se ve afectado por la hora en que se efectúa.
H1: al menos una de las μi es diferente $i=1,2,3,4$

$N1=N2=N3=N4= 4$
$N=16$
$k=4$

$SCt= 11977 - (425)^2/16) = 687.937$

$$SCa= \frac{(113)^2}{4} + \frac{(114)^2}{4} + \frac{(111)^2}{4} + \frac{(87)^2}{4} - \frac{(425)^2}{16} = 124.697$$

$SCe= 687.937 - 124.697 = 563.2495$

Fuente de variación	SC	Gl.	CM	F
Hora de la visita	124.697	3	41.566	0.8855
Aleatorio	563.2495	12	46.937	
Total	687.937	15	-	

Si el nivel de significancia usado es 5 % Fc = 3.49 valor de la variable F con (3,12) grados de libertad.
$F<Fc \quad 0.8855<3.49$

Se acepta Ho. No hay evidencia que el tiempo de la visita se vea afectado por la hora de la misma.

Resuelva:

Se llevó a cabo un experimento para probar el efecto de 4 medicamentos distintos sobre el tiempo de coagulación de la sangre (en minutos). Se extrajeron muestras de sangre y se asignaron al azar a los 4 medicamentos, los resultados fueron los siguientes:

Medicamento			
W	*X*	*Y*	*Z*
1.5	1.8	1.7	1.9
1.4	1.4	1.3	1.5
1.8	1.6	1.5	1.9
1.3	1.2	1.2	1.4
2.0	2.1	2.2	2.3
1.7	1.0	1.0	1.2
1.5	1.6	1.5	1.7
1.5	1.5	1.5	1.7
1.2	1.0	1.3	1.5
1.5	1.6	1.6	1.9

Con base en esos datos puede concluirse que los medicamentos tienen efectos distintos. Sea α= 5%

5.11 ANALISIS DE FRECUENCIAS Y LAS PRUEBAS DE CHICUADRADO

En el análisis estadístico suele suceder que los datos se clasifican en diferentes categorías y es de interés determinar las relaciones que existen entre unas y otras, en estos casos las pruebas de Chicuadrado tienen aplicación

PRUEBA DE LA INDEPENDENCIA: Util para probar la hipótesis Ho que dos criterios de clasificanción cuando se aplican al mismo conjunto de entidades son independientes.

Por ejemplo:

Probar la hipótesis que el estado socioeconómico y el área de residencia son independientes, o probar la hipótesis que los hábitos de higiene bucal son independientes del sexo.

Para la ejecución de la prueba es necesario clasificar los datos observados (O) en una tabla de Contingencia; esta tabla se construye con r renglones que representan los niveles de uno de los criterios o categorías de clasificación y c columnas que representan los niveles del segundo criterio.

Nivel de clasificación 1	Nivel de clasificación 2				
	1	2	3	j	c
1	x11	x12	x13	x1j	x1c
2	x21	x22	x23	x2j	x2c
3	x31	x32	x33	x2j	x3c
.	.	.	.	.	.
.	.	.	.	.	.
i	xi1	xi2	xi3	xij	xic
.	.	.	.	.	.
r	xr1	xr2	xr3	xrj	xrc

donde xij es el número de elementos en el i-ésimo nivel 1 y el j-ésimo nivel 2.

Si la hipótesis planteada en cierta, se espera que la distribución relativa del nivel 1 sea igual en cada una de las clasificaciones del nivel 2 y viceversa. La prueba consiste en clacular las frecuencias esperadas (E) en cada una de las celdas bajo este criterio. Si Ho es verdadera la diferecia entre las frecuencias observadas y esperadas debe ser mínimo (tendiendo a cero).

Para comprobarlo se utiliza el estadístico:

$$£^2 = \Sigma \frac{(O-E)^2}{E}$$

que tiene una distribución chicuadrado con (c-1)*(r-1) grados de libertad.

Si $£^2 > £c^2$ con un nivel de significancia α determinado, se considera que la diferencia entre los valores observados y esperados es lo suficientemente grande para rechazar Ho.

Por ejemplo:

Supónga que se está analizando los resultados de una encuesta llevada a cabo para determinar la relación entre la actitud que se manifiesta respecto a la fluoración y el conocimiento que de ella se tiene. Los encuestados se seleccionaron de modo que fuesen representativos del público en general. Los resultados se resumen en la tabla siguiente.

Frecuencias Observadas
Conocimiento sobre la fluoración

Actitud	Correcto	Incierto	Incorrecto	Total
Favorable	209	80	151	440
Indeciso	61	67	31	159
Desfavorable.	32	25	20	77
Total	302	172	202	676

Lo que interesa es determinar si hay independencia entre el conocimiesnto que las personas tienen sobre la fluoración (clasilficación 2) y la actitud que hacia ella adopta (clasificación 1).
La hipótesis nula puede enuciarse como:
Ho. No existe diferencia entre quienes demostraron un conocimiento correcto, incorrecto o incierto de lo que es la fluoración, con respecto a su actitud, los dos criterios son independientes.
La muestra puede clasificasrse según el nivel 1, Actitud, en 65% favorable, 23.5% indiciso y 11.4% desfavorable. Si Ho. es verdadera se espera igual proporción se encuentre en cada uno de los niveles de la clasificación 2

Frecuencia Esperada
Calsificación 2

Calsificación 1	Correcto	Incierto	Incorrecto	total
Favorable	197	112	132	440
Indeciso	71	40	47	159
Desfavorable	34	20	23	77
Total	302	172	202	676

El valor esperado puede calcularse facilmente por la fórmula

$$Eij = \frac{\text{Total de fila * total de columna}}{\text{total total}}$$

$E11 = 440*302 / 676 = 196.9$

El valor del estadístico $\pounds^2$ es

$$\pounds^2 = \frac{(209-196.6)^2}{196.6} + \frac{(61-71.3)^2}{71.3} + \frac{(32-30.2)^2}{30.2} + \ldots + \frac{(20-23)^2}{23}$$

$\pounds^2 = 39.38$

Si el Nivel de significancia seleccionado es 1 %
$\pounds c^2 = 13.277$ de la distribución chicuadrado con 2*2=4 grados de libertad.

Como $\chi^2>\chi_c^2$, 39.8>13.277 se rechaza la hipótesis nula y se debe concluir que los dos criterios de clasificación, actitudes y conocimientos, no son independiente.

Resuelva:

Se clasificó en una tabla de contingencia una muestra de 25 odontólogos en base a su especialidad y la zona de la comunidad en que trabajan, los resultados son los siguientes:

	Especialidad				
Zona	*A*	*B*	*C*	*D*	*Total*
Norte	*20*	*18*	*12*	*17*	*67*
Sur	*6*	*22*	*15*	*13*	*56*
Oriente	*4*	*6*	*14*	*11*	*35*
Occidente	*10*	*19*	*23*	*40*	*92*
Total	*40*	*65*	*64*	*81*	*250*

¿Proporcionan los datos evidencia que indique una falta de independencia entre los dos criterios de clasificación, zona y especialidad, use α*= 5%*

PRUEBAS DE HOMOGENEIDAD

En ocasiones el investigador puede seleccionar muestras independientes de varias poblaciones, cada una de tamaño N1, N2, N3,..,Nr y trata de encontrar si las r poblaciones son homogéneas de acuerdo a un criterio de clasificación.

En la prueba de homogeneidad se trata de probar la Ho. que las proporciones en cada nivel son iguales para cada una de las poblaciones.

Por ejemplo:

En el estudio de caries dental en niños de seis comunidades, con niveles variables de fluoruro en el suministro de agua, se seleccionó una muestra de 125 niños de cada una de ellas y se les practicó un examen dental. La tabla siguiente señala el número de niños con caries en cada una de las muestras.

Frecuencia Observada

Comunidad	No. de niños	Sin caries	Con caries
A	125	38	87
B	125	8	117
C	125	30	95
D	125	40	85
E	125	64	61
F	125	32	93
total	750	212	538

Para probar si estos dato son compatibles con la hipótesis de que las poblaciones son homogéneas con respecto a la proporción de niños con caries se efectúa la prueba de Chicuadrado de la siguiente forma.

Ho La proporción de niños sin cares es igual en las seis comunidades.
P= 212/750= 0.3533

Frecuencia Esperada

Comunidad	Sin caries	Con caries
A	35.33	89.67
B	35.33	89.67
C	35.33	89.67
D	35.33	89.67
E	35.33	89.67
F	35.33	89.67

$£^2=\Sigma((O-E)^2/E)=(38-35.33)^2+(87-89.67)^2+(8-35.33)^2+....=$
$£^2= 65.85$

$£c^2=$ 11.07 Valor de la variable chicuadrado con $\alpha=5\%$ y $1*5=5$ grados de libertad.

$£^2>£c^2$, $65.85 >11.07$

La hipótesis nula se rechaza, las poblaciones no son homogéneas con relación a la proporción de niños sin caries.

La enfermería de una Universidad realizó un experimento para determinar el grado de alivio prporcionado por tres remedios para la tos. Se probó cada remedio con 50 estudiantes y se registraron los siguientes datos

Frecuencia Observada

	Remedio			
	Nyquil	Robitussin	Triminic	total
Sin alivio	11	13	9	33
Algún alivio	32	28	27	87
Alivio total	7	9	14	30
Total	50	50	50	150

¿Con un nivel de significancia de 5% se puede decir que los tres remedios son igualmente efectivos?

Ho: Las proporciones de estudiantes, sin alivio, algún alivio, y alivio total son iguales para cada remedio,respectivamente: 33/150,87/150, 30/150

Frecuencia Esperada

Nyquil	Robitussin	Triminic
11	11	11
29	29	29
10	10	10

$\pounds^2=\Sigma((O-E)^2/E)=(11-11)^2/11+(32-29)^2/29....+(14-10)^2/10$

$\pounds^2= 3.81$

$\pounds c^2=$ 9.488 Valor de Chicuadrado con $\alpha= 5\%$ y 4 grados de libertad.

$\pounds^2<\pounds c^2$, 3.81<9.488 se acepta la hipótesis nula, los remedios son igualmente efectivos.

Resuelva:

En un estudio sobre la contaminación atmosférica realizado en tres comunidades se seleccionó una muestra aleatoria de 200 familias, en cada una de ellas se interrogó a uno de los miembros acerca de a si alguno de la familia se sentía afectado por la contaminación. Las respuestas son las siguientes.

	Respuestas	
Comunidad	*Si*	*No*
I	*43*	*157*
II	*81*	*119*
III	*70*	*130*

¿Pueden concluir los investigadores que las tres comunidades difieren respecto a la variable Proporción de familias afectadas por la contaminación?

5.12 EVALUACION

I. Para cada una de las aseveraciones siguientes contestar si es cierta o falsa:

1. Si cometemos el error Tipo I rechazamos una hipótesis Nula que es realmente verdadera.

2. Una prueba de diferencia pareada es adecuada cuando las dos muestras que van a probarse son dependientes.

3. El valor $1-\beta$ es conocido como Potencia de la Prueba.

II. Seleccione la alternativa correcta para cada uno de los siguientes planteamientos:

1. Si decidimos que α es 0.10 para una prueba entonces estamos diciendo que:
- *a. 10% es el estándar mínimo de la probabilidad de aceptación.*
- *b. 10% es el riesgo que corremos de rechazar la Hipótesis Nula si es verdadera.*
- *c. 10% es el riesgo que corremos de aceptar la Hipótesis Nula si es falsa.*
- *d. a y b solamente.*
- *e. a y c solamente.*

2. Supónga que se está efectuando una prueba de hipótesis para un proceso en el cual un error tipo I sería muy costoso pero un error tipo II sería realmente barato y tendría poca importancia. Cuál de los siguientes valores será la mejor opción para α?
- *a. 0.01*
- *b. 0.10*
- *c. 0.25*
- *d. 0.50*

3. *Si queremos probar si las proporciones de más de dos poblaciones son iguales usamos:*
   a. *Análisis de Varianza*
   b. *Estimación*
   c. *La Varianza*
   d. *Ninguno de los anteriores*

III. *Resuelva los siguientes problemas:*

1. *En el desarrollo de nuevos medicamentos para el tratamiento de la ansiedad es importante verifcar los efectos que tienen en las funciones motoras del organismo, uno de los cuales es la conducción de vehiculos. Una compañía está probando cuatro tranquilizantes para investigar sus efectos en la pericia de conducir. Los sujetos se someten a una prueba simulada de manejo y sus puntuaciones reflejan los errores que cometen, los errores más graves dan origen a puntuaciones más altas. El resultado de la prueba produjo la siguiente tabla.*

   Medicamento:

1	245	258	239	241	
2	277	276	263	274	
3	215	232	225	247	226
4	241	253	237	246	240

   En un nivel de significancia de 0.05, afectan los medicamentos en forma diferente la habilidad de conducir?

2. *Un Químico está desarrollando repelente de insectos y quiere probar si una fórmula recién inventada brinda mayor protección contra las mordeduras de insectos que la proporcionada por el producto líder en el mercado.*
   En el experimento de 14 voluntarios, se les roció un brazo con el viejo producto y el otro con la nueva fórmula. Después cada uno metió los brazos en dos cámaras llenas con igual número de insectos, en número de picaduras recibidas en cada brazo se transcriben aquí.
   Cuando α es 0.01, debe el Químico concluir que la nueva fórmula realmente es más eficaz que el actual producto líder?

Sujeto	1	2	3	4	5	6	7	8
F. Antigua	5	2	5	4	3	6	2	4
F. Nueva	3	1	5	1	1	4	4	2

Sujeto	9	10	11	12	13	14
F. Antigua	2	6	5	3	1	3
F. Nueva	5	2	3	3	1	2

3. *Para probar Ho: $\mu=14$ H1: $\mu > 14$ si se está considerando una Prueba t de nivel $\alpha=0.05$*
   Qué tamaño de muestra se necesita con el objeto de que la probabilidad de aceptar erróneamente Ho sea 0.1 cuando la media poblacional real difiere de 14 por 0.5? De una muestra preliminar se estima que la desviación poblacional es 1.25.

MODULO 6

REGRESION Y CORRELACION

INTRODUCCION

En algunas oportunidades se tiene interés en estudiar la relación que existe entre dos variables, el análisis de Regresión y Correlación facilita los modelos teóricos para hacerlo.

La Regresión permite identificar la forma más adecuada de la relación de variables, se utiliza para predecir o estimar el valor de una en función de un valor determinado de la otra. La Correlación mide la intensidad de la relación de variables.

Los conceptos de Regresión Lineal y Correlación simples, serán expuestos en éste módulo, orientando al lector para el cálculo de la recta de Minimos Cuadrados y el Coeficiente de Correlación de Pearson. Además se presenta el procedimiento de prueba de hipótesis del coeficiente de correlación y la estimación por intervalo del mismo.

OBJETIVOS

Al finalizar el Módulo, el lector estará en capacidad de:

1. Interpretar la aplicabilidad de la Recta de Regresión.

2. Construír un diagrama de dispersión que muestre la relación entre dos variables X,Y.

3. Calcular la recta de Regresión por el Método de Mínimos Cuadrados.

4. Predecir los valores que corresponde a al variable dependiente en función de un valor dado de la variable independiente.

5. Interpretar el significado del coeficiente de Correlación ρ.

6. Calcular el coeficiente de Correlación Muestral r de Pearson.

7. Estimar por Intervalo de Confianza el Coeficiente de Correlación Poblacional ρ.

8. Comprobar Hipótesis relacionadas con el coeficiente de Correlación Poblacional ρ.

6.1 REGRESION

Con el Análisis de Regresión se desea encontrar una ecuación que exprese la relación entre dos variables X,Y, una de las variables es considerada como variable Ordinaria (X) que puede medirse sin error apreciable y la otra (Y) como una variable aleatoria. X es denominada la variable Independiente y el interés es estudiar la dependencia de Y en funión de X, llegando así a determinar la mejor ecuación que estime la relación de las variables. Es importante decir que las relaciones descritas son de asociación pero no necesariamente causales.

La regresión Lineal busca establecer una ecuación de la forma: $\mu(Y)=\alpha+\beta X$, llamada recta de regresión que expresa la relación entre las variables, e indica que la media de Y varía proporcionalmente en función de los valores de X. En la ecuación, el coeficiente α es el valor de $\mu(y)$ cuando X es igual a cero y β es el Coeficiente de Regresión, pendiente de la Recta.

La utilidad de la recta de regresión es el poder predecir los valores de $\mu(Y)$ que corresponden a valores predeterminados de X.

Por ejemplo:

Se puede estudiar la relación que existe entre la presión sanguínea y el peso de una persona, al establecer una ecuación $\mu(Y)=\alpha+\beta X$ que describa esa relación, la Variable ordinaria X, independiente, es el peso de la población de individuos y $\mu(Y)$ es la media de cada subpoblación de Y que depende de X. Observe que Y es una variable aleatoria ya que para individuos del mismo peso pueden encontrarse variaciones en los niveles de presión por causas al azar.

6.2 SUPOSICIONES QUE FUNDAMENTAN LA REGRESION LINEAL SIMPLE.

1. Los valores de X son preseleccionados por el investigador y se mide sin error aparente, esto significa que se desprecia la magnitud de los errores de medición en X.

2. Para cada valor de X existe una subpoblación de valores de Y que tiene un distribución Normal y la media $\mu(Y)$ es función de la variable X

$$\mu(Y) = \alpha+\beta X$$

3. Todas la varianzas de las subpoblaciones de Y son iguales.

4. Los valores de Y son estadísticamente independientes.

6.3 ESTIMACION DE LA RECTA POR EL METODO DE MINIMOS CUADRADOS

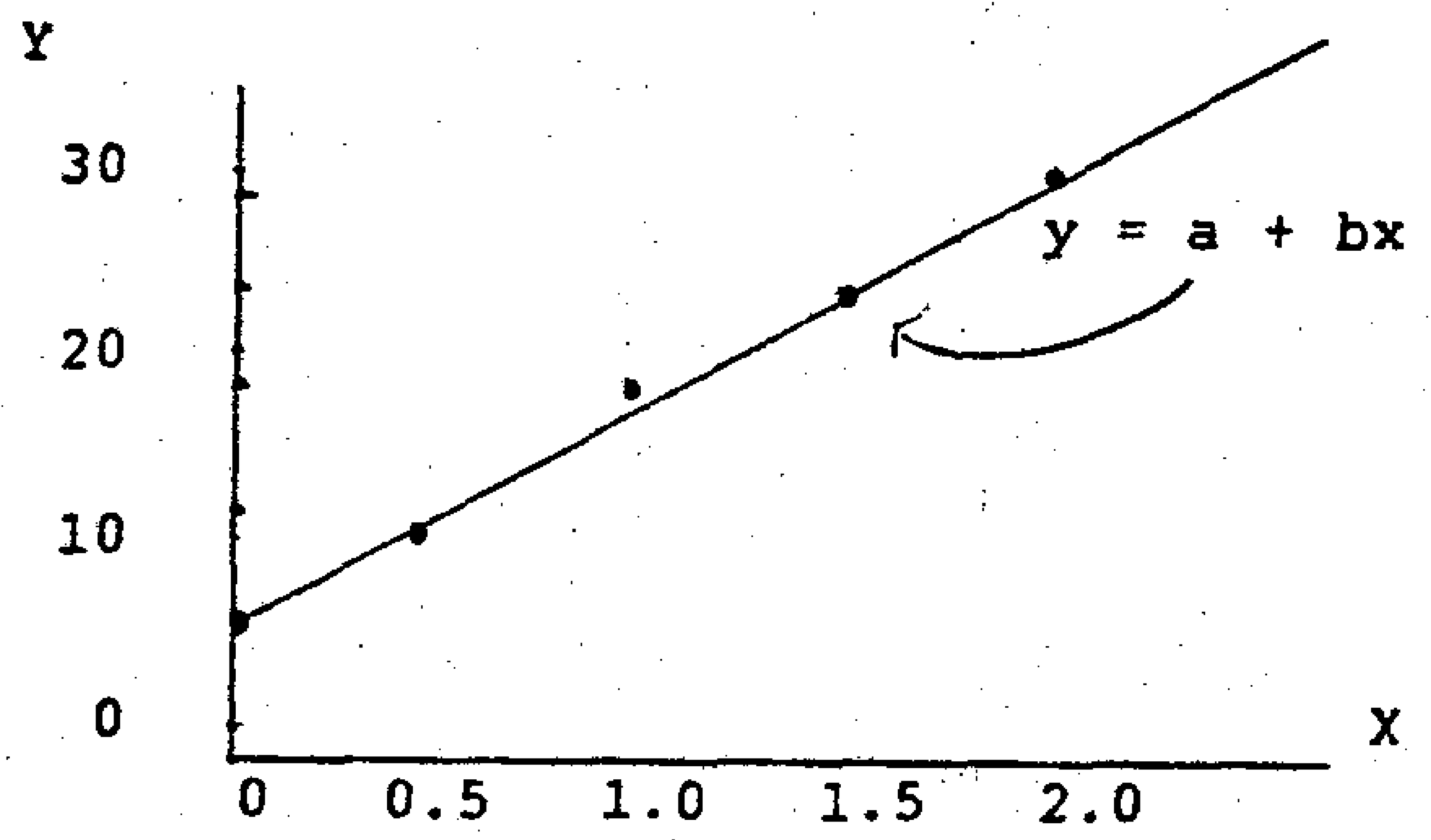
La ecuación $\mu(Y) = \alpha + \beta X$ describe el comportamiento de las subpoblaciones de Y en función de la variable X, sin embargo los parámetros α y β son a menudo desconocidos, por lo que deben ser estimados a partir del análisis de una muestra, encontrando la recta de regresión muestral $\tilde{Y} = a + b X$, en la que a y b son las estimaciones de α y β, además $\tilde{y}$ se considera como una estimación de $\mu(Y)$.

El primer paso para ésta estimación es obtener una colección de datos que muestren los valores de las variables consideradas y su correspondencia, esta colección de datos se presenta en forma de tabla o por un Diagrama de Dispersión.

Por ejemplo:

Para medir la relación que existe entre la resistencia compresiva de una amalgama y el tiempo, se pueden coleccionar los valores de una muestra de 5 amalgamas así:

Tiempo en horas que Transcurre desde que fue hecha. X	Resistencia Compresiva Y lb./pulg^2
0.0	5.21
0.5	11.21
1.0	19.47
1.5	24.19
2.0	28.32



En el diagrama se puede ubicar una Recta que describa la relación entre las variables: tiempo, resistencia. El problema es buscar la recta de ajuste que represente mejor a los datos.

Resuelva:

Trace el diagrama de dispersión y una recta de ajuste que relacione las variables X: ingreso bruto del Odontólogo, Y porcentaje correspondiente a prótesis fijas entre el total de prestaciones que realiza en su clínica.

X (Q)	*Y (%)*
70000	*0.10*
160000	*0.60*
110000	*0.45*
95000	*0.30*
150000	*0.55*
80000	*0.15*

Para evitar el juicio personal en la identificación de la recta, es necesario investigar el mejor ajuste, para ello se considsera que yi˜= a +bxi es también una estimación de los valores de yi observados, sin embargo esta estimación difiere de ellos en una cantidad denominada residuo (yi-yi˜), éstos residuos describen el error de ajuste de la recta en el dato xi. Si la recta se ajustara completamente los residuos serían iguales a cero y el problema es seleccionar la recta de ajuste que haga cero los residuos o en su defecto las minimice, esto se consigue estimando los coeficientes a y b por el METODO DE MINIMOS CUADRADOS.

El Método de los Mínimos Cuadrados permite encontrar la mejor recta de estimación de la relación lineal entre varibles modo que:

$$\Sigma (yi-yi^{\sim})^2 = \Sigma(yi - (a+bxi))^2 \text{ sea mímima.}$$

y lo hace por medlio de la estimación de α y β calculadas a partir de las fórmulas.

$$b = \frac{\Sigma\, xi\, yi - n\, X'\, Y'}{\Sigma\, xi^2 - n\, (X')^2}$$

$$a = Y' - b\, X'$$

En el ejemplo anterior la Recta de Mínimos Cuadrados está dada por:

$$yi^{\sim} = 5.84 + 11.84\ yi$$

Cálculos:

n= 5	X'=1	Y'=17.68
Σxi= 5	Σyi= 88.4	Σxiyi=118
Σxi^2= 7.5	Σyi^2=1919.07	

$$b = \frac{118 - 5 * 1 * 17.68}{7.5 - 5 * 1^2} = 11.84$$

$$a = 17.68 - 11.84 * 1 = 5.84$$

Tiempo en horas que Transcurre desde que fue hecha. X	Resistencia Compresiva Y lb./pulg²	Estimacion Y˜	Residuo
0.0	5.21	5.84	-0.63
0.5	11.21	11.76	-0.55
1.0	19.47	17.68	1.79
1.5	24.19	23.60	0.59
2.0	28.32	29.52	-1.20

Cuando la recta de estimación es calculada por el Método de Mínimos Cuadrados la suma de los residuos es cero.

Resuelva:

Se reunieron los siguientes datos en un estudio de la relación entre la inteligencia y el tamaño de la familia.

Número de niños en la familia	*Calificación promedio de la Inteligencia para todos los niños de la familia*
1	105
2	102
3	104
4	100
5	97
6	101
7	95
8	93
9	97
10	88

Encuentre la recta de mímimos cuadrados y los valores estimados de y para cuando el número de niños en la familia es 4,6,8, y 10.

6.4 OBSERVACIONES ADICIONALES A LA REGRESION

Lo que se presentó hasta el momento acerca de la regresión involucra una sola variable independiente, y supone que el modelo elegido es el correcto, que Y está relacionada linealmente con X. No debe esperarse que la predicicón sea buena si hay muchas variables independientes asociadas o la estructura que relaciona a Y con X es extremadamente no lineal.

6.5 CORRELACION

El Análisis de Correlación es una herramienta estadística que se utiliza para medir el grado de la relación entre dos variables X,Y, es decir, la asociación que existe entre ellas, por medio de un sólo número ρ denominado Coeficiente de Correlación.

A menudo se usa junto con el Análisis de Regresión para medir la eficacia con que la línea de regresión explica la variación de la variable dependiente, sin embargo en el modelo de Correlación se considera que ambas variables son aleatorias, X es una variable distribuida Normalmente con media $\mu(X)$ y varianza $\sigma^2(X)$ y la distribución de Y condicionada a X es normal con $\mu(Y)= \alpha + \beta X$ y varianza $\sigma^2(Y)$ independiente de X.

Dos medidas son usadas para describir la correlación entre dos variables, estas son: El Coeficiente de Determinación y el Coeficiente de Correlación.

6.6 COEFICIENTE DE DETERMINACION

El Coeficiente de Determinación Muestral (r^2), es una manera de medir el grado de relación entre dos variables y es calculado a partir de los datos de una muestra aleatoria.

El coeficiente se obtiene de la relación entre tres tipos de variación:

La variación de los valores de Y en el conjunto de datos al rededor de su media , variación total $(\Sigma(y_i-Y')^2)$.

La de los valores de $\tilde{Y}$ en el conjunto de datos al rededor de la media, variación explicada ($\Sigma(\tilde{y}_i-Y')^2$).

La de los valores de Y al rededor de la recta de regresión ajustada, variación no explicada $(\Sigma(y_i-\tilde{y}_i)^2)$.

Para calcular la fracción de la variación total que no es explicada por la recta de regresión se encuentra la razón

$$\frac{\Sigma\,(\,yi - yi^{\sim})^2}{\Sigma\,(\,yi - Y'\,)^2}$$

Entonces el Coeficiente de determinación

$$r^2 = 1 - \frac{\Sigma\,(yi-yi^{\sim})^2}{\Sigma\,(yi-Y')^2}$$

Se define como la fracción de la variación total que es explicada por la recta de regresión.

El valor de r^2 es +1 siempre que la línea de regresión tenga un ajuste perfecto, esto es $\Sigma(yi-y^{\sim})=0$, toda la variación es explicada. El valor de r^2 es cero cuando no hay relación entre las variables, la variación es toda no explicada.

r^2 se encontrará entre los extremos 1 y 0, r^2 cercano a 1 indica una fuerte correlación entre X,Y y un r^2 cercano a 0 significa que las variables tienen poca correlación.

Una fórmula abreviada para calcular el coeficiente de determinación es:

$$r^2 = \frac{a\,\Sigma yi + b\,\Sigma xiyi + n\,Y'^2}{\Sigma yi^2 - n\,Y'^2}$$

Ejemplo:

Los valores siguientes son 15 lecturas sobre la congestión del tráfico y la concentración de monóxido de carbono efectuadas en un sitio de muestreo, para determinar la calidad del aire en cierta ciudad.

Congestionamiento del tráfico (autos/hora) X	CO (ppm) Y
100	8.8
110	9.0
125	9.5
150	10.0
175	10.5
190	10.5
200	10.5
225	10.6
250	11.0
275	12.1
300	12.1
325	12.5
350	13.0
375	13.2
400	14.5

Datos:

$\Sigma xi = 3550$ $\Sigma xi^2 = 974450$ $X' = 236.67$

$\Sigma yi = 167.8$ $\Sigma yi^2 = 1915.36$ $Y' = 11.19$

$\Sigma xiyi = 41945$

La recta de regresión para describir la concentración de CO en función del Número de autos por hora es:

$$b= \frac{\Sigma\, x_iy_i - n\, X'\, Y'}{\Sigma x_i^2 - n\, X'^2} = \frac{41945 - 15*236.67 * 11.19}{974450 - 15 * (236.67)^2} = 0.0165$$

$$a= Y' - b\, X' = 11.19 - 0.0165\ (236.67) = 7.2849$$

$$\tilde{Y}= 7.2849 - 0.0165\ X$$

El coeficiente de determinación se calcula,

$$r^2= \frac{a\ \Sigma y_i + b\Sigma x_iy_i - n\ Y'^2}{\Sigma\ y_i^2 - n\ (Y')^2} =$$

$$r^2= \frac{7.2849 * 167.8 + 0.0165 * 41945 - 15 * (11.19)^2}{1915.36 - 15\ (11.19)^2}$$

$$r^2= 0.9767$$

El 97.67% de la variación total de la concentración de monóxido de carbono es explicada por la regresión respecto al número de autos por hora, existe una fuerte relación entre las dos variables.

Resuelva:

La siguiente tabla ilustra los valores del consumo de metilmerecurio y la cantidad total de mercurio en la sangre de 12 individuos expuestos a la primera sustancia por haber consumido pescado contaminado:

Consumo de Metil mercurio MgHg/dia	*Mercurio en la sangre Mg/g*
180	*90*
200	*120*
230	*125*
410	*290*
600	*310*
550	*290*
275	*170*
580	*375*
105	*70*
250	*105*
460	*205*
650	*480*

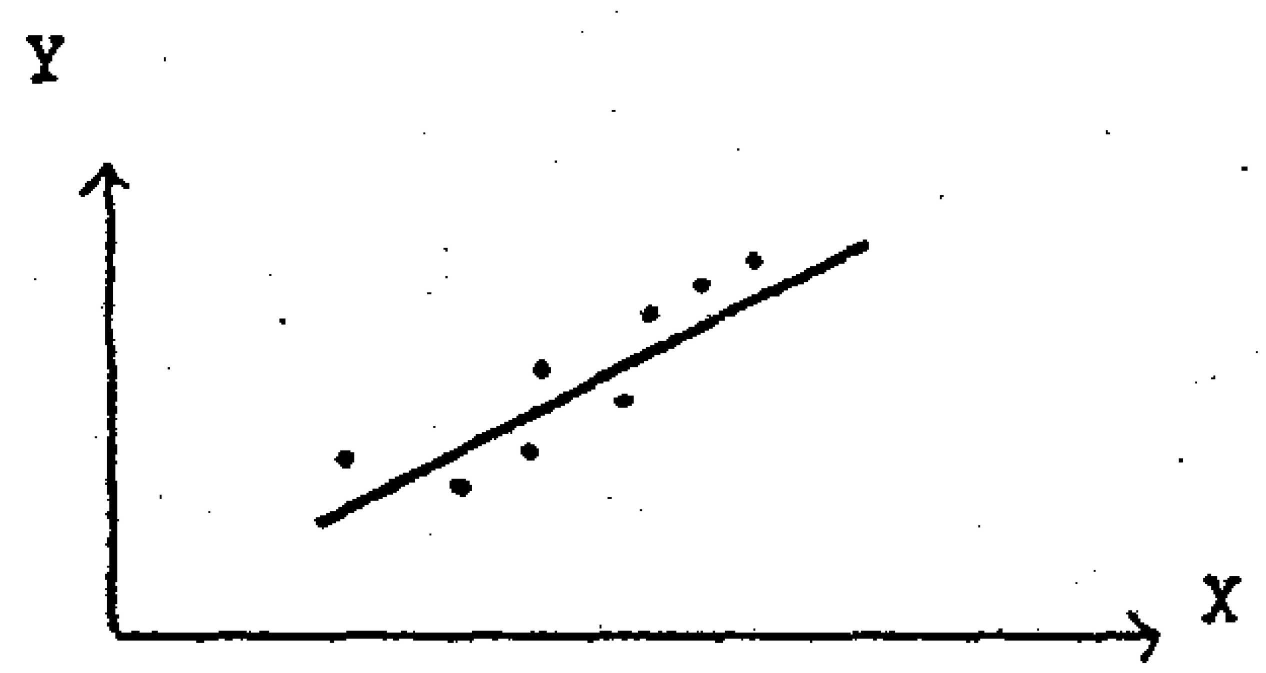
Encuentre la ecuación de regresión que describa la relación lineal entre las dos variables, calcule e interprete el coeficiente de determinación r^2.

6.7 COEFICIENTE DE CORRELACION PRODUCTO MOMENTO DE PEARSON (r)

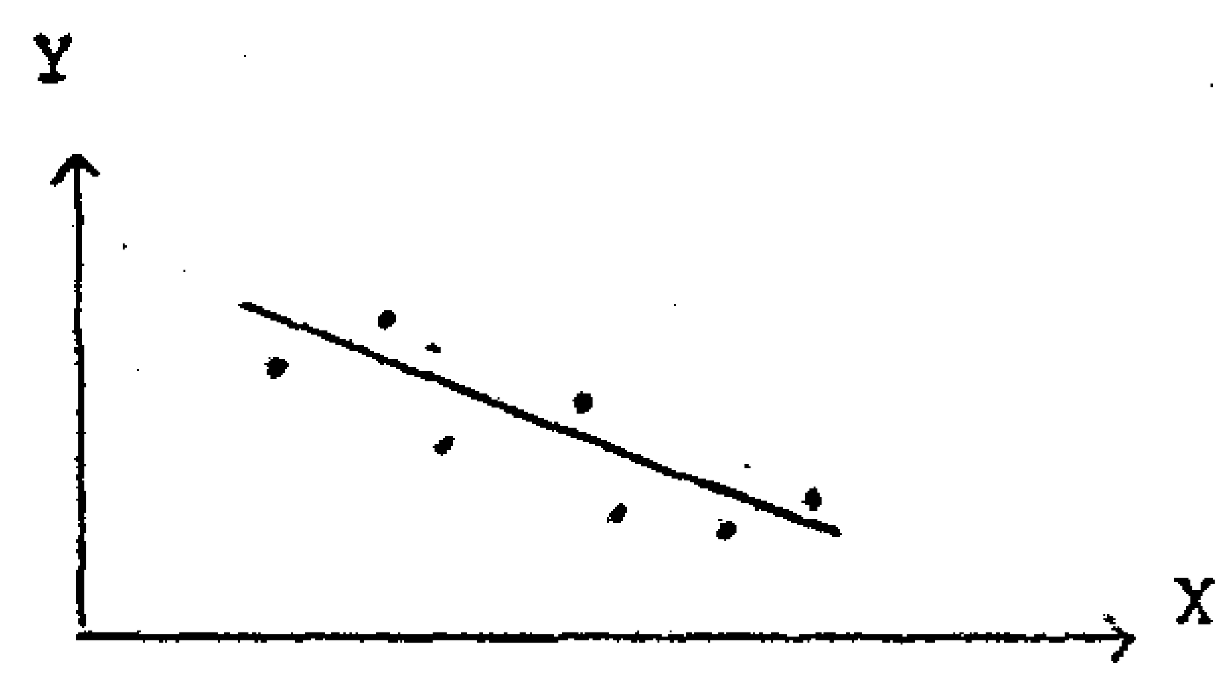
El Coeficiente de Correlación Muestral se denota por r y es la raíz cuadrada del Coeficiente de Determinación.

$$r = \sqrt{r^2}$$

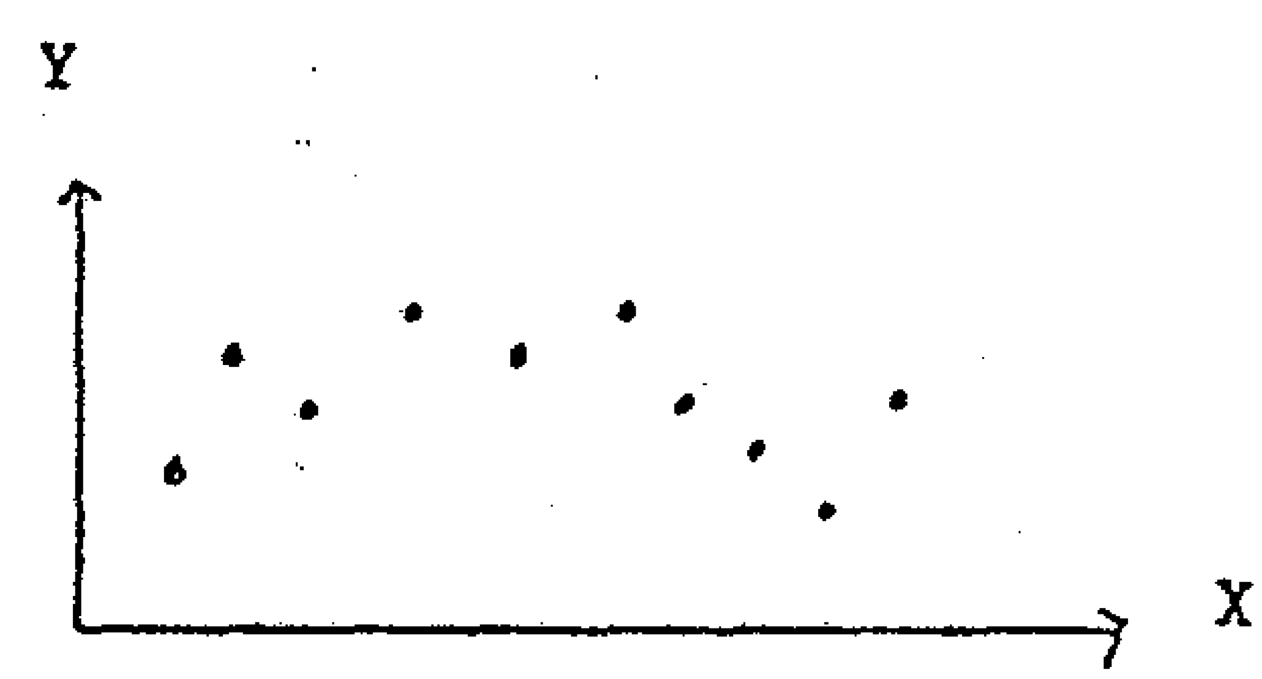
Cuando la Pendiente de la recta de estimación (b) es positiva, la raíz cuadrada es positiva, pero si es negativa, la raíz cuadrada es negativa, así, r varía entre -1 y +1, el signo indica la dirección de la relación entre las variables X,Y. Si hay una relación directa Y aumenta al hacerlo X y r está entre 0 y 1, si existe una relación inversa r está entre -1 y 0, Y disminuye al aumentar X.



Correlación Positiva



Correlación Negativa



No existe Correlación r=0

En el ejemplo anterior:

$$r = \sqrt{r^2} = \sqrt{0.9768} = 0.988$$

La pendiente b es positiva, por lo que r también tiene el signo positivo, e implica una relación directa entre el congestionamiento y la concentración de CO.

Resuelva:
Encuentre el Coeficiente de Correlación del ejercicio precedente.

6.8 INFERENCIAS RELATIVAS AL COEFICIENTE DE CORRELACION ρ

Como se dijo anteriormente ρ, Coeficiente Poblacional, mide el grado de relación entre variables, y es estimado por r el Coeficiente de Correlación Muestral, por lo que es útil hacer inferencias sobre el verdadero valor del parámetro a partir de r ya sea por medio de la estimación por interevalo o por las pruebas de hipótesis relacionadas con ρ.

Las fórmulas para el cálculo del Intervalo de Confianza y los Estadísticos de Prueba se presentan a continuación.

INTERVALOS DE CONFIANZA PARA ρ

Limite inferior: $\tanh\left(\tfrac{1}{2}\ln((1+r)/(1-r)) - Z/(\sqrt{(n-3)})\right)$

Limite superior: $\tanh\left(\tfrac{1}{2}\ln((1+r)/(1-r)) - Z/(\sqrt{(n-3)})\right)$

ln= logaritmo Natural
n= tamaño de la muestra
Z= valor de la variable normal estándar con r% de confianza.
Tanh= tangente hiperbólica.

PRUEBAS DE HIPOTESIS:

$$H_o.\ \rho = 0 \qquad H_1:\ \rho \neq 0$$

Estadístico de Prueba:

$$t = r\left(\sqrt{(n-2)/(1-r^2)}\right)$$

tc valor de la distribución t de Student con n-2 grados de libertad y un nivel de significancia α.

Región de aceptación: si $-tc < t < tc$ aceptar Ho

Región de rechazo: $t > tc$ ó $t < -tc$ rechazar Ho.

$$H_o:\ \rho = \rho_. \qquad H_1:\ \rho \neq \rho_.$$

Estadístico de Prueba

$$Z = \frac{\sqrt{(n-3)}}{2} * \ln\left(\frac{(1+r)\ (1-\rho_.)}{(1-r)\ (1+\rho_.)}\right)$$

Zc es el valor de la variable normal estándar con un nivel de significancia α

Región de Aceptación $-Zc < Z < Zc$ Aceptar Ho.

Región de rechazo $Z < -Zc$ ó $Z > Zc$ Rechazar Ho.

Ejemplo:
Del ejercicio anterior, el intervalo de confianza para el coeficiente de correlación entre las variables congestionamiento de tráfico y concentración de CO es:

Datos
r= 0.988 n= 15
Z = 1.96 con 95% de confianza.

Limite Inferior:

$\text{Tanh}(\frac{1}{2}\ln((1+0.988)/(1-0.988)) - 1.96/\sqrt{12}) =$
$\text{Tanh}(\frac{1}{2}\ln(165.67) - 0.5658) = \text{Tanh}(3.121) = 0.9632$

Limite superior

$\text{Tanh}(\frac{1}{2}\ln(165.67) + 0.5658) =$
$\text{Tanh}(3.121) = 0.9961$

$$0.9632 <= \rho <= 0.996$$

La prueba de hipótesis correspondiente para:

Ho: $\rho = 0.96$ H1: $\rho = 0.96$

con 5 % de nivel de significancia, Zc = 1.96, ensayo bilateral.

Estadistico de Prueba

$Z = \sqrt{12}/2 * \ln((1.98*0.4)/(0.012*1.96)) = 2.1099$

$$Z > Zc, \quad 2.1099 > 1.96$$

Se rechaza la hipótesis nula, la muestra evidencia que el coeficiente de correlación es diferente de 0.96

En un estudio sobre el efecto de un compuesto de dieta sobre la composición de los lipidos del plasma, se obtuvo en una muestra de 15 animales un r=-0.74, pruebe la hipótesis que el coeficiente de correlación es igual a cero.

Ho: $\rho = 0$ H1: $\rho \neq 0$

Datos:

n= 15 r= - 0.74
tc= ± 2.16 valor de la varliable t con 13 grados de libertad.

$t = -0.74 \sqrt{(13/(1-(-0.74)^2))} = -3.9668$

$$t < tc, -3.9668 <- 2.16 \text{ se rechaza Ho.}$$

El coeficiente de correlación entre las variables es significativamente menor que cero.

Resuelva:

En un estudio acerca de la relación entre el peso y el tamaño del tórax de los niños al nacer, se obtuvieron los siguientes datos.

Peso Kg.	Tamaño del Tórax cm.
2.75	29.5
2.15	26.3
4.41	32.2
5.52	36.5
3.21	27.2
4.32	27.7
2.31	28.3
4.30	30.3
3.71	28.7

Calcule el interevalo del 95% de confianza para ρ.
Pruebe la hipótesis nula que $\rho=0$ contra la alternativa $\rho>0$ con un nivel de signficancia de 1%.

6.9 EVALUACION

I. *Responda si cada una de las siguientes aseveraciones son verdaderas o falsas:*

1. *El Análisis de Regresión se usa para determinar la exactitud con que una ecuación de estimación describe la relación que está siendo estudiada entre dos variables.*

2. *Un valor de r^2 cercano a 0 indica una estrecha correlación entre X y Y.*

3. *Si $r = 0.8$ entonces la ecuación de regresión explica el 80% de la variación total de la variable dependiente.*

II. *Seleccione la alternativa que complemente correctamente los siguientes planteamientos:*

1. *Supongase que conocemos la altura de un estudiante pero no su peso. Aplicamos una ecuación para obtener una estimación del peso basandonos en su talla, por tanto suponemos que:*
   a. *El peso es la variable independiente.*
   b. *La altura es la variable dependiente.*
   c. *Ninguna de las anteriores.*

2. *Si la variable dependiente aumenta en una ecuación de estimación y también lo hace la variable independiente, el coefieciente de correlación se encontrará en el intervalo:*
   a. *De 0 a -1*
   b. *De 0 a 2*
   c. *De 0 a 1*

3. *Suponga que según cálculos efectudos a es 4 y b es 2 para determinar la línea de regresión. Si la variable independiente tiene un valor de dos que valor cabe esperar que tenga la variable dependiente?*
   - a. 8
   - b. -1
   - c. 10
   - d. 0

III. *Resuelva los siguientes problemas:*

1. *El Centro Estadístico analizó datos acerca de marmotas normales. Las variables de interés fueron el peso del cuerpo, en gramos, y el peso del corazón también en gramos. Desarrolle una ecuación de regresión lineal, con el objeto de determinar si existe relación significativa entre el peso del corazón y el del cuerpo. Utilice el peso del corazón como variable independiente. Pruebe la hipótesis que $\rho = 0$.*

Peso del Cuerpo	Peso del Corazón
4050	11.2
2465	12.4
3120	10.5
5700	13.2
2595	9.8
3640	11.0
2050	10.8
4235	10.4
2935	12.2
4975	11.2
3690	10.8
2800	14.2
2775	21.2
2170	10.0
2370	12.3
2055	12.5
2025	11.8
2645	16.0
2675	13.8

2. *Las cantidades de un compuesto químico que se disuelven en 100 gramoa de agua a diferentes stermperaturas X se registraron como siguie:*

X (°C)	Y (gramos)		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	42
75	48	51	44

a. *Determine la ecuación de la línea de regresión por el Método de Mínimos Cuadrados.*
b. *Grafique la recta en un Diagrama de Dispersión.*
c. *Estime la cantidad de compuesto que se disuelve en 100 gramos de agua a 50 °C.*

MODULO 7

METODOS NO PARAMETRICOS

INTRODUCCION

La mayor parte de las pruebas estadísticas estudiadas se refieren a inferencias sobre los parámetros de las poblaciones, y para formularla se plantean suposiciones restrictivas en cuanto a la población, en especial al empleo de la distribución normal. Sin embargo esas suposiciones no siempre pueden ser válidas, para resolver casos en que se ignora la clase de población que se está muestreando, es necesario buscar alternativas de análisis algunas las ofrecen los Métodos no Paramétricos.

El presente módulo es una introducción a los Métodos no Paramétricos y muestra las aplicaciones de las pruebas del Rango con Signo de Wilcoxon, de Kruskal-Wallis, U de Man-Witney, de la Mediana y la Correlación de Rangos de Spearman.

OBJETIVOS

Al finalizar el estudio del módulo el lector estará en capacidad de:

1. Identificar la aplicación de los Métodos No Paramétricos.

2. Señalar la utilidad de cada una de las pruebas presentadas.

3 Efectuar correctamente el contraste de hipótesis para cada una de las pruebas estudiadas.

4 Calcular el Coeficiente de Correlación de Spearman.

5 Interpretar el significado del Coeficiente de Correlación de Spearman.

7.1 ESTADISTICOS NO PARAMETRICOS

Son técnicas útiles que no hacen suposiciones restrictivas sobre la forma de la distribución y se conocen con el nombre de Pruebas No Paramétricas o Sin Distribución.

Ventajas de los Métodos No paramétricos

1. No exigen hacer la suposición de que una población está distribuida en forma normal.
2. Son fáciles de llevar a cabo
3. A veces ni se requiere que de la ordenación formal, lo único que se hace es describir un resultado como mejor que otro, por lo que cuando las mediciones no son tan exactas como lo exigen las pruebas paramétricas es posible aplicarlos.

Desventajas

1. Ignoran cierta cantidad de información.
2. No son tan eficientes como la prueba paramétrica.
3. La prueba No Paramétrica requiere un tamaño de muestra más grande para la misma probabilidad de cometer el Error Tipo II.

7.2 PRUEBA DE RANGOS CON SIGNO DE WILCOXON

Utilizada para probar hipótesis nula que se está muestreando una pobación simétrica continua con media μ.

Ho: $\mu = \mu.$

H1: $\mu < \mu.$

$\mu > \mu.$

$\mu \neq \mu.$

Procedimiento:

Se sustrae primero $\mu.$ de cada valor muestral y se descartan todas las las diferencias iguales a cero, las diferencias restantes son clasificadas sin importar el signo, asignando a la diferencia absoluta más pequeña el valor uno, a la siguiente más pequeña el valor 2, y asi sucesivamente. Cuando el valor absoluto de dos o más diferencias es el mismo se asigna a cada uno el promedio de los valores que les hubiere asignado si fueran diferentes.

Si Ho. es verdadera el total de los valores correspondientes a las diferencias positivas W+ debe ser casi igual al total de los valores correspondientes a las diferencias negativas W-, la más pequeña de las dos W+ y W- se designa por W.

Aceptación y Rechazo:

La Ho puede rechazarse a favor de $\mu<\mu.$ si W+ es pequeña y W- es grande. Ho puede rechazarse a favor de $\mu>\mu.$ si W+ es grande y W- es pequeña. A Favor de μ $\mu.$ si W es lo suficientemente pequeña.

La tabla de Apéndice F muestra los valores críticos de W para niveles de significancia de 1,2.5,5 % unilaterales y 2,5,10 % bilaterales. La hipótesis nula se rechaza si W+, W- o W son menores que el Wc.

Ejemplo:

Las cifras que siguen son cantidades de tiempo en minutos que tardó una muestra de 20 técnicos en realizar un exámen de laboratorio. Para probar la Ho que esta muestra tomada al azar es de una población continua con $\mu=19.4$, contra H1: $\mu \neq 19.4$ con $\alpha=5\%$ se procede de la siguiente forma.

Valor Observado	X- $\mu.$	Rangos asignados
18.1	-1.3	11.0
20.3	0.9	7.5
18.3	-1.1	9.0
15.6	-3.8	20.0
22.5	3.1	19.0
16.8	-2.6	17.0
17.6	-1.8	13.0
16.9	-2.5	16.0
18.2	-1.2	10.0
17.0	-2.4	15.0
19.3	-0.1	1.5
16.5	-2.9	18.0
19.5	0.1	1.5
18.6	-0.8	6.0
20.0	0.6	4.5
18.8	-0.6	4.5
19.1	-0.3	3.0
17.5	-1.9	14.0
18.5	-0.9	7.5
18.0	-1.4	12.0

W-= 177.5 W+= 32.5 W=32.5

Con α= 5% ensayo bilateral n=20 el valor crítico Wc= 52. Se rechaza Ho. la media de la distribución es diferente a 19.4

Resuelva:

Los siguientes datos representan el tiempo en minutos que tuvo que esperar un paciente en 12 visitas al consultorio de una doctora antes de ser atendido por ésta: 17,15,20,20,32,28,12,26,25,25,35,24.

Con un nivel de significancia de 5% pruebe la afirmación de la doctora que en promedio sus pacientes no esperan más de 20 minutos antes de ser atendidos.

7.3 PRUEBA U DE WILCOXON O DE MAN Y WHITNEY

Es una alternativa de la prueba t, útil para probar la hipótesis nula que se muestrean poblaciones continuas idénticas, contra la alternativa que las poblaciones tienen medias desiguales, o para probar que muestras independientes fueron extraídas de la misma población.

Ho: $\mu 1=\mu 2$ H1: $\mu 1 < \mu 2$

$\mu 1 > \mu 3$

$\mu 1 \neq \mu 2$

Procedimiento:

Para probar la hipótesis nula, primero se selecciona una muestra aleatoria de cada una de las poblaciones, N1 es el número de las observaciones de la muestra más pequeña y N2 es el número de observaciones de la muestra más grande. Se ordenan las observaciones N1+N2 de las muestras combinadas en orden creciente y se sustituye un valor de 1,2,.....N1+N2 para cada observación, en caso de empates se reemplazan las observaciones por la media de los valores que tendrían si fueran distinguibles.
La suma de los valores correspondientes a las N1 observaciones se denota W1 y la suma de las N2 observaciones se denota W2.

Aceptación y Rechazo:

Ho. será rechazada en favor de la aternativa $\mu 1<\mu 2$ so W1 es pequeño y W2 es grande. A favor de $\mu 1>\mu 2$ sio W1 es grande y W2 es pequeño. La alternativa $\mu 1$ $\mu 2$ es aceptada si el mínimo de W1 y W2 es lo suficientemente pequeño.

Los estadísticos para hacer el contraste de Hipótesis son:

$$U1 = W1 - N1\,(N1+1)/2 \qquad U2 = W2 - N2\,(N2+1)/2$$

La hipótesis nula se rechaza si U1, U2 o el mínimo U toma un valor menor o igual al valor crítico Uc dado en la tabla del Apéndice G.

Ejemplo:

El contenido de nicotina de dos marcas de cigarrillos medidos en miligramos es el siguiente:

Marca A	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3
Marca B	4.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2
	1.9	5.4						

Para probar la Hipótesis, con un nivel de significancia de 5%, que el contenido medio de nicotina de las dos marcas son iguales se procede: Ho: $\mu 1=\mu 2$ H1: $\mu 1$ $\mu 2$

$\alpha=5\%$

Uc= 17 de la tabla del apéndice G

	Observaciones ordenadas	Rango
	0.6	1
	1.6	2
	1.9	3
*	2.1	4
	2.2	5
	2.5	6
	3.1	7
*	3.3	8
*	3.7	9
*	4.0	10
	4.0	11
	4.1	12
*	4.8	13
	5.4	14
*	5.4	15
*	6.1	16
	6.2	17
	6.3	18

*valores que corresponden a N1

W1= 93 W2= 78

U1= 93 - (8*9)/2 = 57 U2= 78 - (10*11)/2 = 23

U =23 mínimo

Como U>Uc se acepta Ho, no hay diferencia en el contenido medio de nicotina en las dos muestras.

Resuelva:

Los siguientes valores son las respuestas de los anticuerpos en personas que han recibido una dósis de refuerzo de uno de dos tipos de vacuna contra la rabia.

Tipo 1: 1.25 5.30 1.70 1.00 8.5 3.75 8.10 2.25 5.6 7.85
Tipo 2: 0.57 3.90 8.20 1.20 1.70 1.00 4.55 5.20 2.16
1.90 4.60

Puede concluirse con base a esos datos que los dos tipos de vacunas difieren en su efectos? use $\alpha=0.05$

7.4 PRUEBA H DE KRUSKAL Y WALLIS

Se utiliza para probar la hipótesis nula de que k muestras independientes provienen de poblaciones idénticas, es una altenativa para la prueba F en el Análisis de Varianza de un factor.

Ho: $\mu 1 = \mu 2 = \mu 2 \ldots\ldots = \mu k$

H1: no tadas las medias poblacionales son iguales.

Procedimiento:

Sea N1, N2,....Nk el número de observaciones de cada una de las muestras. Se procede a combinar todas las k muestras y se ordenan las N= N1+N2+...Nk observaciones en forma creciente, sustituyendo el valor de 1,2,3....N para cada observación, en caso de empates sustituir el valor por la media de los valores que le corresponderian a éstos si hubieran sido distinguibles. La suma de las N1 observaciones de la población 1 se denota por R1, la suma de las N2 observaciones de la población 2 se denota R2, etc.

Aceptación y Rechazo:

El estadístico de prueba es: $H = \frac{12}{N(N+1)} \Sigma \frac{Ri^2}{Ni} - 3(N+1)$

El estadístico H tiene una distribución Chicuadrado con k-1 grados de libertad, si cada muestra cuenta con 5 observaciones por lo menos.

$\pounds c^2$ es el valor de la distribución Chicuadrado con k-1 grados de libertad y un nivel de significancia $\alpha\%$. Si $H > \pounds c^2$ se rechaza Ho. de otra forma se acepta.

Ejemplo:

Al estudiar los efectos de dos medicamentos en el tiempo de reacción a cierto estímulo, en animales, se asignaron tres grupos el I y II fueron tratados con el medicamento A y B respectivamente el II sirvió como control.

Para decir si las tres poblaciones de animales difieren en el tiempo de reacción se procede de la siguiente forma.

Ho: $\mu 1 = \mu 2 = \mu 3$

H1: al menos un grupo tiene una media diferente.

Tiempo de reacción en segundos

Grupo 1	Grupo 2	Grupo3
24.0	23.2	18.4
16.7	19.8	19.1
22.8	18.1	17.3
19.8	17.6	17.3
18.9	20.2	19.7
	17.8	18.9
		8.8
		19.3

N1= 5 N2= 6 N3= 8 N= 19

Rangos asignados

Grupo 1	Grupo 2	Grupo 3
19.0	18.0	7.0
1.0	14.5	11.0
17.0	6.0	2.5
14.5	4.0	2.5
9.5	16.0	13.0
	5.0	8.0
		12.0
R1= 61	R2=63.5	R3= 65.5

$$H = \frac{12}{19*20}\left(61^2/5 + 63.5^2/6 + 65.5^2/8\right) - 3*20$$

$$H = 61.659 - 60 = 1.66$$

$\chi_c^2 = 5.991$ valor de la variable chicuadrado con 2 grados de libertad y 5% de significancia.

$H < \chi_c^2$, $1.66 < 5.991$ se acepta la hipótesis nula, las muestras no evidencian que las poblaciones tengan diferentes medias.

Resuelva.

Los siguientes valores son los honorarios diarios promedio cobrados a los pacientes por cierta intervención quirúrgica Puede concluirse al nivel de significancia de 5% que los tres grupos estudiados difieren con respecto a los honorarios diarios promedio?

	Grupo	
I	*II*	*III*
$80.75	*$58.63*	*$84.21*
78.15	*72.70*	*101.76*
85.40	*64.20*	*107.74*
71.94	*62.50*	*115.30*
82.05	*63.20*	*126.15*

7.5 LA PRUEBA DE LA MEDIANA

Es un procedimiento para probar si k grupos independientes difieren en sus tendencias centrales, especificamente, que k muestras independientes se han extraído de poblaciones con medianas iguales.

Procedimiento:

El primer paso en el procedimiento es calcular la mediana común de las muestras combinadas (la mediana Me, es el valor central de la distribución, tiene el 50 % de los datos muestrales superiores a él y el restante 50% inferiores). Luego para cada uno de los grupos se determina el número de observaciones que son mayores que la mediana común y el número de observaciones que son menores o iguales a ella, los resultados se disponen en una tabla de contingencia 2*k

	Poblaciones			
Observaciones	1	2	3	.... k
Mayores Me.				
Menores o Iguales Me				
Total				

Si en efecto las muestras provienen de poblaciones con la misma mediana es de esperar que la mitad de las observaciones de cada muestra estén por arriba de la mediana y la mitad por debajo, éste es el supuesto para obtener las frecuencias observadas y utilizar la prueba de Chicuadrado para comprobar la hipótesis nula que las medianas de las poblaciones son iguales.

Aceptación y Rechazo:

Estadístico de Prueba es $\chi^2 = \Sigma((O-E)^2/E)$

χc^2 es el valor de la distribucicón chicuadrado con (k-1) grados de libertad y un nivel de significancia α.

Si $\chi^2 < \chi c^2$ se acepta la hipótesis nula concluyendo que las medianas de las poblaciones son iguales. Si $\chi^2 > \chi c^2$ la diferencia entre los valores observados y esperados es los suficientemente grande para concluir que las medianas son diferentes.

Ejemplo:

Se obtienen los siguientes valores de albúmina en el suero de 17 personas normales y 13 hospitalizadas, se considera con un nivel de significancia de 0.05 que las medianas de las dos poblaciones muestreadas son distintas, para comprobarlo se procede de la siguiente forma:

Albúmina en el suero (g/100ml)

Personas Normales			Personas Hospitalizadas		
2.4	5.0	4.0	1.5	3.1	2.0
3.5	2.9	3.5	1.3	3.4	1.5
3.1	3.0	3.6	1.7	1.8	2.0
4.0	3.2	4.2	2.0	3.8	1.5
3.5	3.4	3.8	3.5		
4.5	3.9				

Número de observacione n=30

Ordenando las observaciones de menor a mayor

1.3, 1.5, 1.5, 1.5, 1.7, 1.8, 2.0, 2.0, 2.0, 2.4, 2.9, 3.0, 3.1, 3.1, 3.2, 3.4, 3.4, 3.5, 3.5, 3.5, 3.5, 3.6, 3.8, 3.8, 3.9, 4.0, 4.0, 4.2, 4.5, 5.0

Me= 3.3 (15 observaciones son mayores que éste valor y 15 observaciones son menores que él.

Tabla de Contingencia:

VALORES OBSERVADOS

	Personas Normales	Hospitalizadas
Mayores que Me	12	3
Menores que Me	5	10
Total	17	13

VALORES ESPERADOS

	Personas Normales	Personas Hospitalizadas
Mayores que Me	8.5	6.5
Menores que Me	8.5	6.5

$$£^2 = \frac{(12-8.5)^2}{8.5} + \frac{(3-6.5)^2}{6.5} + \frac{(5-8.5)^2}{8.5} + \frac{(10-6.5)^2}{6.5} = 6.65$$

$£c^2 = 3.84$ con 5% de nivel de significancia y 1 grado de libertad. $£^2 > £c^2$ se rechaza Ho, las medianas de las poblaciones son diferentes.

Resuelva:

Se estudiaron los sueros de dos grupos de personas, después de haber sufrido una infección por estreptococo, para observar la acción neutralizante de los anticuerpos ante la esptolisena O. Los resultados fueron los siguientes.

Anticuerpos (unidades IDD)

Grupo A	*Grupo B*
324	558
375	108
349	291
604	863
566	303
810	640
340	358
295	503
357	646
580	689
344	

Proporcionan estos datos evidencia suficiente que indique una diferencia en las medianas de las poblaciones? Sea $\alpha = 0.01$

7.6 COEFICIENTE DE CORRELACION DE RANGOS DE SPEARMAN (Rs)

Al igual que el Coeficiente de Correlación de Muestras r de Pearson, el Coeficiente de Correlación de Rangos es una medida de la relacion entre variables X,Y generalmente numéricas, sin embargo, a menudo en el análisis de correlación no se dispone de valores numéricos pero se puede asignar un ordenamiento por rangos a los elementos de las variables que se están estudiando, en estos casos se aplica también el coeficiente Rs.

El coeficiente de correlación de Rangos Rs, es una medida de la correlación que existe entre dos conjuntos de rangos.

Una de las ventajas de utilizar Rs en lugar que r, es que ya no se asume que la relación desconocida entre X,Y es lineal, por esa razón cuando los datos poseen una relación distinta el Coeficiente de Rangos será probablemente más confiable que r, además ya no se hacen las suposiciones de normalidad respecto a las distribuciones de X,Y.

Procedimiento:

El procedimiento para calcular Rs es asignar rangos de 1 a n a las observaciones de X en orden de magnitud o importancia y de modo análogo a las observaciones de Y, cuando hay empates se procede asignando el promedio de los valores que corresponderian a las observaciones si fueran diferentes.
Posteriormente se calcula Di que es la diferencia entre los valores asignados a Xi y Yi, calculando Rs por la fórmula:

$$Rs = 1 - \frac{6 \Sigma Di^2}{n(n^2-1)}$$

n es el número de parejas muestreadas.

También se puede calcular Rs con la fórmula del Coeficiente de Correlación de Pearson r, sustituyendo los valores de X, Y por los rangos que corresponden a las observaciones.

Ejemplo

La tabla siguiente muestra los consumos de calorías (cal/día/kg.)
y de Oxigeno VO2 (ml/min/kg) de 10 niños. El coeficiente Rs es:

Consumo de calorías (X)	VO2(Y)	Rango X	Rango Y	Di
50	7.0	2	1	1
70	8.0	3	2	1
90	10.5	5	6	-1
120	11.0	8	8	0
40	9.0	1	3	-2
100	10.8	6	7	-1
150	12.0	9	10	-1
110	10.0	7	5	2
75	9.5	4	4	0
160	11.9	10	9	1

$\Sigma Di^2 = 14$ $\qquad Rs = 1 - \frac{6 * 14}{10*(10^2-1)} = 0.915$

Sugiere una asociación creciente fuerte entre el consumo de calorías y de oxigeno.

La tabla siguiente contienen 5 personas y los rangos tanto académico que alcanzaron en la universidad como del nivel que obtuvieron en cierto hospital 10 años después de graduados, el 5 es el nivel más alto y el 1 el más bajo.
Se puede calcular el coeficiente de correlación de rangos entre el éxito en la universidad X y el nivel alcanzado al cabo de 10 años de graduados en la forma siguiente:

Estudiante	Rango Universitario	Rango Hospitalario	Di
Juan	4	5	-1
Margarita	3	1	2
Delia	1	3	-2
Esteban	2	2	0
Raúl	5	4	1

$\Sigma Di^2 = 10$ $\qquad Rs = 1 - \dfrac{6 * 10}{5*(5^2-1)} = 0.5$

Existe una leve correlación positiva entre el éxito en la universidad y el nivel alcanzado 10 años después de graduado.

Resuelva:

La tabla siguiente muestra como un grupo de expertos en nutrición y un grupo de amas de casa clasificaron por rangos 9 alimentos para el desayuno.

Alimento	*Experto*	*Ama de Casa*
A	*3*	*5*
B	*7*	*4*
C	*11*	*8*
D	*9*	*14*
E	*1*	*2*
F	*4*	*6*
G	*10*	*12*
H	*8*	*7*
I	*5*	*1*

Determine Rs como una medida de consistencia de las dos clasificaciones.

7 .7 EVALUACION:

I. *Asocie las definiciones de la columna A a las definiciones presentadas en la columan B, colocando en el paréntesis el número que las identifica.*

1. Rango con Signo de Wilcoxon.	*( ) Método no Paramétrico de probar si tres o más muestras independientes han sido extraídas de poblaciones que tienen la misma distribución.*
2. Coeficiente de Correlación de Spearman	*( ) Técnicas estadísticas que no hacen suposiciones sobre la forma de la distribución de una población cuando se realiza una prueba de hipótesis.*
3. Kruskal-Wallis	*( ) Medida del grado de asociación entre dos variables que se basa en los rangos de las Observaciones.*
4. Métodos No Paramétricos	*( ) Método no paramétrico que se emplea para determinar si dos muestras independientes han sido extraída de poblaciones con la misma media.*
5. U de Man Wintney	
6. Mediana	

II *Resuelva los siguientes problemas:*

1. *Para determinar si un nuevo suero eliminará la Leucemia se seleccionaron 9 pacientes quienes habían alcanzado etapas avenzadas de la enfermedad, cinco recibieron tratamiento y cuatro no. Los períodos de sobrevivencia en años desde el momento en que se inició el experimento son:*

Tratamiento	2.1	5.3	1.4	4.6	0.9
Sin Tratamiento	1.9	0.5	2.8	3.1	

A un nivel de significancia de 0.05 determine si el suero es eficaz.

2. *Se revisaron quince expedientes de pacientes de dos hospitales y se asignó una calificación diseñada para estimar el nivel de cuidados recibidos. Las calificaciones fueron las siguientes:*

Hospital A: 99, 85, 73, 98, 83, 88, 99, 80, 74, 91, 80, 94, 94, 98, 80.
Hospital B: 78, 74, 69, 79, 57, 78, 79, 68, 59, 91, 89, 55, 60, 55, 79.

Concluiría al nivel de significancia de 0.05 que las medianas de las dos poblaciones son distintas.

PRUEBA FINAL

I. Cada uno de los términos listados en la columna B asócielos con los conceptos de la columna A, anotando en el peréntesis el número que los identifica.

Columna A	Columna B
1. Número de resultados favorables a la ocurrencia de un evento dividido entre el número total de resultados.	( ) Probabilidad
	( ) Valor Esperado
	( ) Parámetros
2. Valores que describen las características de una población.	( ) Estima
3. Valor específico calculado de un estimador.	( ) Nivel de Significancia
4. Distribución de probabilidades de todos los posibles valores que un estadístico puede asumir para cierto tamaño de muestra.	( ) Distribución Muestral.
5. Posibilidad de que algo suceda.	
7. Medidas que describen el comportamiento de una muestra.	
8. Limite superior o inferior de un intervalo de confianza.	
9. Probabilidad de rechazar una Hipótesis Nula cuando es verdadera.	
10. Rechazo de una Hipótesis cuando es Verdadera.	

II. Efectúe los cálculos que corresponden y seleccione la alternativa correcta para cada una de las afirmaciones siguientes.

1. Un especialista en alergias alega que el 50% de los pacientes que examina son alérgicos a algún tipo de hierba. La probabilidad que exactamente tres de sus siguientes cuatro pacientes sean alérgicos es:
   A. 0.25
   B. 0.50
   C. 0.125

2. Si A y B son mutuamente excluyentes y P(A)=0.3 y P(B)=0.5 la probabilidad de A^ (A Complemento) es:
   A. P(A^)=0.5
   B. P(A^)=0.7
   C. P(A^)=0.8

3. Cierta empresa tiene 2000 empleados, durante el año el monto promedio por empleado debido a costos médicos fue de $31.50 y desviación estandar $6.00. La probabilidad de que una muestra aleatoria de 36 empleados proporcione una media superior a $30.00 es:
   A. 0.9345
   B. 0.0655
   C. 0.9332

4. Siete amas de casa fueron muestreadas aleatoriamente y se averiguó que caminan en promedio 39.2 millas por semana durante sus tareas domésticas con una desviación de 3.2 millas. El intervalo del 95% de confianza para la media de la población es:
   A. 39.2 ± 2.9594
   B. 39.2 ± 2.3502
   C. 39.2 ± 2.3705

5 Al comparar el comportamiento de dos catalizadores sobre un cultivo se lleva a cabo una prueba t de dos muestras con $\alpha=0.05$. Las varianzas en el cultivo se consideran iguales para los dos catalizadores. El tamaño de muestra adecuado para probar Ho: $\mu 1=\mu 2$ y H1: $\mu 1$ $\mu 2$, si es esencial detectar una diferencia de 0.8σ entre los dos catalizadores con probabilidad 0.9 es:
   A. 34
   B. 27
   C. 41

6 Si el 75% de la variación de la variable dependiente es explicada por la línea de regresión el valor de r es:
   A. 0.75
   B. 0.866
   C. 0.25

III. Argumente brevemente cada una de las siguientes cuestiones.

1. Un fabricante de baterías para Marca Pasos especifica que la vida de cada una es de 28 meses o más. Si la programación de la cirugía de sustitución de las baterías debe basarse en esa declaración, explique al fabricante las consecuencias de los errores Tipo I y Tipo II.

2. Explique de que se ocupa el Análisis de Regresión y su utilidad en la Investigación.

3. Exponga las ventajas que tiene la aplicación de los Métodos No Paramétricos en el Análisis de Datos.

4. Explique cual es la ventaja de utilizar una estimación por Intervalo y no la estimación Puntual.

IV. Resuelva los siguientes problemas.

1. Efectúe una prueba de hipótesis para determinar si el número medio de días para cobrar una cuenta después de que un Odontólogo terminó el tratamiento es de 37 días, si una muestra de 15 pacientes presentó los siguientes resultados. (Se desconoce la distribución de la Población)

32, 35, 33, 36, 44, 41, 36, 32, 39, 31, 47, 30, 34, 29, 41.

2. Los datos de la tabla siguiente representan el número de horas de descanso proporcionadas por cinco diferentes marcas de tabletas para el dolor de cabeza administradas a 25 sujetos que experimentaban fiebres de 39 °C o más. Formúle y pruebe la hipótesis correspondiente si se desea comparar la efectividad de las 5 marcas con un nivel de significancia de 5%.

Tableta

A	B	C	D	E
5	9	3	2	4
4	7	5	3	6
8	8	2	4	9
6	6	3	1	4
3	9	7	4	7

BIBLIOGRAFIA

1. CANAVOS, GEORGE C. Probabiliad y Estadística, aplicaciones y métodos. Mc Graw Hill, México 1988.

2. ESCOTET, MIGUEL A. Estadística Psicoeducativa. Trillas, México 1973.

3. FREUND, JOHN, MAMNING SMITH. Estadística. Prentice Hall Hispanoamericana, México 1989.

4. FREUND, JOHN, RONALD WALPOLE. Estadística Matemática con Aplicaciones. Prentice Hall Hispanoamericana, México 1990.

5. GANE GLASE, JULIAN STANLEY. Métodos Estadísticos Aplicados a las Ciencias Sociales. Prentice Hall, México 1988.

6. LEVIN RICHARD. Estadística para Administradores. Prentice Hall Hispanoamericana, México 1988.

7. MENDENHALL, SCHEAFFER, WACKELY. Estadística Matemática con Aplicaciones. Grupo Editorial Iberoamericana, México 1986.

8. MENDENHALL, WADSWORTH. Introducción a la Probabilidad y Estadística. Internacional Iberoamericana 1979.

9. SIDNEY SIEGUEL. Estadística No Paramétrica. Trillas, México 1989.

10. WALPOLE, MYERS. Probabilidad y Estadística. Mac Graw Hill Interamericana, México 1992.

11. WAYNE W. DANIEL. Bioestadística. Limusa, México 1990.

12. WEINTRAUB, DOUGLASS Y GILLINGS. Bioestadística en Salud Bucodental Organización Panamericana de la Salud, USA 1989.

APENDICE

SIMBOLOGIA

a	Valor de la constante en la recta de Regresión
A^	Complemento de un Suceso
A,B..	Eventos de un experimento aleatorio
b	Coeficiente de Regresión Muestral, pendiente de la Recta de Regresión.
E	Estadístico
E(X)	Esperanza, valor esperado de una Variable Aleatoria X
f(x)	Función de Densidad de Probabilidad
F(X)	Función acumulada de una Variable Aleatoria X
F	Valor de la Variable con distribución F
fr	Frecuencia Relativa de un suceso
H	Estadístico definido para la prueba de Kruskal y Wallis
Ho	Hipótesis Nula
H1	Hipótesis Alternativa
(m,n)	Grados de Libertad de una Variable con distribución F
Me	Mediana
n	Número de elementos en la muestra
N	Número de elementos en la población
p	Probabilidad de éxito en una prueba de Bernoulli
p.	Proporción de Exitos en una población con distribución Binomial
P	Proporción de Exitos en una Muestra Aleatoria
P(A)	Probabilidad de un suceso
P(E)	Probabilidad de ocurrencia de un Estadístico
P(A/B)	Probabilidad Condicionada
q	Probabilidad de Fracaso en una prueba de Bernoulli
Q	Proporción de fracasos en una muestra aleatoria
r	Coeficiente de Correlación de Pearson
r^2	Coeficiente de Determinación
Rs	Coeficiente de Correlación de Spearman
Rx	Recorrido de una variable Aleatoria
S	Espacio Muestral
S	Desviación Estándar de Una Muestra
S^2	Varianza de una Muestra
t	Valor de la variable con distribución de Student
U1	Estadístico usado en la Prueba de Man y Whintney
U2	Estadístico usado en la Prueba de Man y Whintney
v	Grados de Libertad para la variable con distribución Chicuadrado, t de Student y F
V(X)	Varianza de la Variable Aleatoria X
W+	Suma de las diferencias positivas en la prueba del Rango con Signo.
W-	Suma de las diferencias Negativas en la prueba del Rango con Signo.
x,y..	Observaciones de la variable X, Y
X,Y..	Variables Aleatorias
X',Y'	Promedio Aritmético de los valores de una muestra, media muestral

$\tilde{Y}$	Variable Y estimada por la Recta de Regresión
Z	Valor de la Variable con distribución Normal Estándar
α	Nivel de significancia, probabilidad de cometer el error Tipo I
α	Valor de la Constante en la Recta de Regresión Poblacional
β	Probabilidad de cometer el error Tipo II
β	Coeficiente de Regresión Poblacional
$(1-\beta)$	Potencia de la Prueba
Γ	Porcentaje de Confianza, Nivel de Confianza
Σ	Sumatoria
σ^2	Varianza de una Variable Aleatoria
σ	Desviación Estándar de una Variable Aleatoria
$\sigma^2(E)$	Varianza de Un Estadístico
$\sigma(E)$	Error Estándar de un Estadístico
μ	Media de una Variable Aleatoria
$\mu(E)$	Media de un Estadístico
Φ	Parámetro
£²	Valor de la Variable con Distribución Chicuadrado
ρ	Coeficiente de Correlación Poblacional
∞	Infinito
$\cap$	Intersección
^k	Potencia k

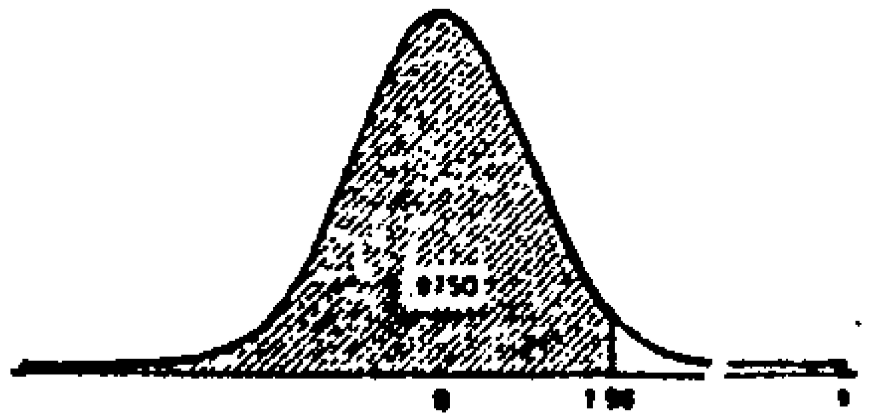


TABLA A

AREAS DE LA CURVA NORMAL ESTANDAR

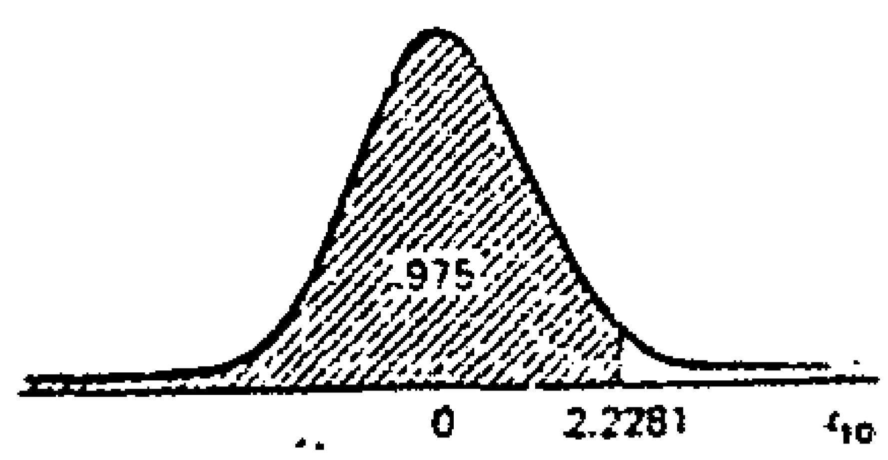
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60
2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	2.70
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	2.80
2.90	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986	2.90
3.00	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990	3.00
3.10	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993	3.10
3.20	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	3.20
3.30	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997	3.30
3.40	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	3.40
3.50	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	3.50
3.60	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.60
3.70	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.70
3.80	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.80

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	0.00	z
−3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.80
−3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.70
−3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	−3.60
−3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	−3.50
−3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	−3.40
−3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	−3.30
−3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	−3.20
−3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	−3.10
−3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	−3.00
−2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	−2.90
−2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	−2.80
−2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	−2.70
−2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	−2.60
−2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	−2.50
−2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	−2.40
−2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	−2.30
−2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	−2.20
−2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	−2.10
−2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	−2.00
−1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	−1.90
−1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	−1.80
−1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	−1.70
−1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	−1.60
−1.50	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	−1.50
−1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	−1.40
−1.30	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968	−1.30
−1.20	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	−1.20
−1.10	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	−1.10
−1.00	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	−1.00
−0.90	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	−0.90
−0.80	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	−0.80
−0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420	−0.70
−0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	−0.60
−0.50	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085	−0.50
−0.40	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	−0.40
−0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	−0.30
−0.20	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	−0.20
−0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	−0.10
0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	0.00

Tomado de: Bioestadística, Wayne Daniel, Limusa 1990

TABLA B

PERCENTILES DE LA DISTRIBUCION T

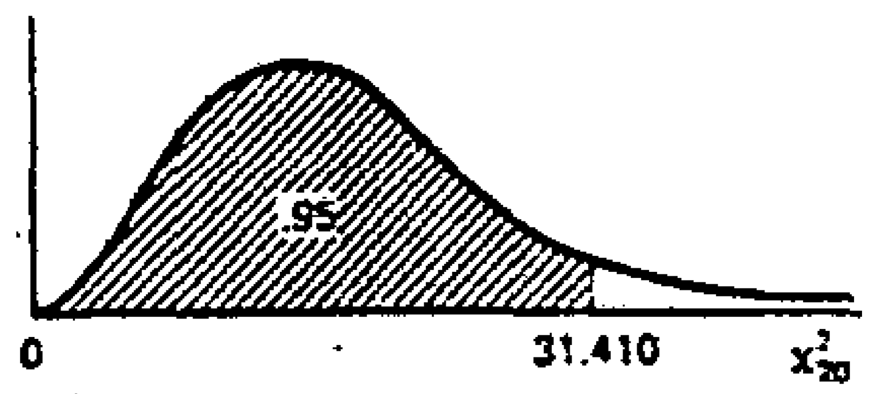


g.l.	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	3.078	6.3138	12.706	31.821	63.657
2	1.886	2.9200	4.3027	6.965	9.9248
3	1.638	2.3534	3.1825	4.541	5.8409
4	1.533	2.1318	2.7764	3.747	4.6041
5	1.476	2.0150	2.5706	3.365	4.0321
6	1.440	1.9432	2.4469	3.143	3.7074
7	1.415	1.8946	2.3646	2.998	3.4995
8	1.397	1.8595	2.3060	2.896	3.3554
9	1.383	1.8331	2.2622	2.821	3.2498
10	1.372	1.8125	2.2281	2.764	3.1693
11	1.363	1.7959	2.2010	2.718	3.1058
12	1.356	1.7823	2.1788	2.681	3.0545
13	1.350	1.7709	2.1604	2.650	3.0123
14	1.345	1.7613	2.1448	2.624	2.9768
15	1.341	1.7530	2.1315	2.602	2.9467
16	1.337	1.7459	2.1199	2.583	2.9208
17	1.333	1.7396	2.1098	2.567	2.8982
18	1.330	1.7341	2.1009	2.552	2.8784
19	1.328	1.7291	2.0930	2.539	2.8609
20	1.325	1.7247	2.0860	2.528	2.8453
21	1.323	1.7207	2.0796	2.518	2.8314
22	1.321	1.7171	2.0739	2.508	2.8188
23	1.319	1.7139	2.0687	2.500	2.8073
24	1.318	1.7109	2.0639	2.492	2.7969
25	1.316	1.7081	2.0595	2.485	2.7874
26	1.315	1.7056	2.0555	2.479	2.7787
27	1.314	1.7033	2.0518	2.473	2.7707
28	1.313	1.7011	2.0484	2.467	2.7633
29	1.311	1.6991	2.0452	2.462	2.7564
30	1.310	1.6973	2.0423	2.457	2.7500
35	1.3062	1.6896	2.0301	2.438	2.7239
40	1.3031	1.6839	2.0211	2.423	2.7045
45	1.3007	1.6794	2.0141	2.412	2.6896
50	1.2987	1.6759	2.0086	2.403	2.6778
60	1.2959	1.6707	2.0003	2.390	2.6603
70	1.2938	1.6669	1.9945	2.381	2.6480
80	1.2922	1.6641	1.9901	2.374	2.6388
90	1.2910	1.6620	1.9867	2.368	2.6316
100	1.2901	1.6602	1.9840	2.364	2.6260
120	1.2887	1.6577	1.9799	2.358	2.6175
140	1.2876	1.6558	1.9771	2.353	2.6114
160	1.2869	1.6545	1.9749	2.350	2.6070
180	1.2863	1.6534	1.9733	2.347	2.6035
200	1.2858	1.6525	1.9719	2.345	2.6006
∞	1.282	1.645	1.96	2.326	2.576

Tomado de: Bioestadística, Wayne Daniel, Limusa 1990

TABLA C

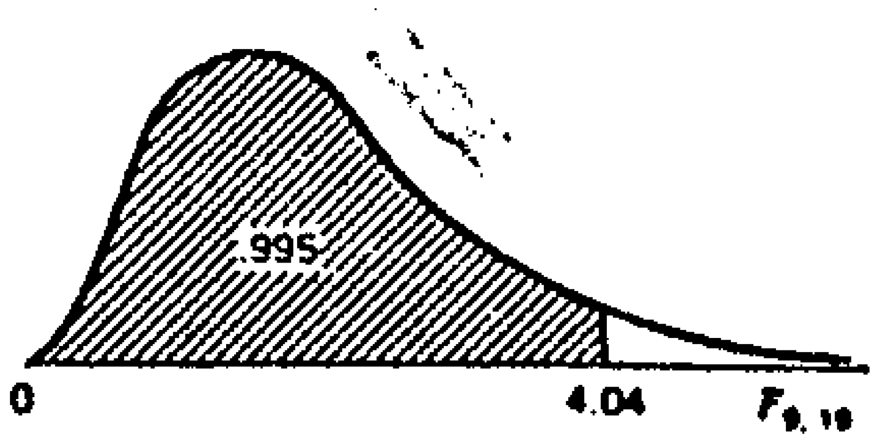
PERCENTILES DE LA DISTRIBUCION CHICUADRADO



$P(\chi^2_{20} < 31.410) = .95$

g.l.	$\chi^2_{.005}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.0000393	.000982	.00393	2.706	3.841	5.024	6.635	7.879
2	.0100	.0506	.103	4.605	5.991	7.378	9.210	10.597
3	.0717	.216	.352	6.251	7.815	9.348	11.345	12.838
4	.207	.484	.711	7.779	9.488	11.143	13.277	14.860
5	.412	.831	1.145	9.236	11.070	12.832	15.086	16.750
6	.676	1.237	1.635	10.645	12.592	14.449	16.812	18.548
7	.989	1.690	2.167	12.017	14.067	16.013	18.475	20.278
8	1.344	2.180	2.733	13.362	15.507	17.535	20.090	21.955
9	1.735	2.700	3.325	14.684	16.919	19.023	21.666	23.589
10	2.156	3.247	3.940	15.987	18.307	20.483	23.209	25.188
11	2.603	3.816	4.575	17.275	19.675	21.920	24.725	26.757
12	3.074	4.404	5.226	18.549	21.026	23.336	26.217	28.300
13	3.565	5.009	5.892	19.812	22.362	24.736	27.688	29.819
14	4.075	5.629	6.571	21.064	23.685	26.119	29.141	31.319
15	4.601	6.262	7.261	22.307	24.996	27.488	30.578	32.801
16	5.142	6.908	7.962	23.542	26.296	28.845	32.000	34.267
17	5.697	7.564	8.672	24.769	27.587	30.191	33.409	35.718
18	6.265	8.231	9.390	25.989	28.869	31.526	34.805	37.156
19	6.844	8.907	10.117	27.204	30.144	32.852	36.191	38.582
20	7.434	9.591	10.851	28.412	31.410	34.170	37.566	39.997
21	8.034	10.283	11.591	29.615	32.671	35.479	38.932	41.401
22	8.643	10.982	12.338	30.813	33.924	36.781	40.289	42.796
23	9.260	11.688	13.091	32.007	35.172	38.076	41.638	44.181
24	9.886	12.401	13.848	33.196	36.415	39.364	42.980	45.558
25	10.520	13.120	14.611	34.382	37.652	40.646	44.314	46.928
26	11.160	13.844	15.379	35.563	38.885	41.923	45.642	48.290
27	11.808	14.573	16.151	36.741	40.113	43.194	46.963	49.645
28	12.461	15.308	16.928	37.916	41.337	44.461	48.278	50.993
29	13.121	16.047	17.708	39.087	42.557	45.722	49.588	52.336
30	13.787	16.791	18.493	40.256	43.773	46.979	50.892	53.672
35	17.192	20.569	22.465	46.059	49.802	53.203	57.342	60.275
40	20.707	24.433	26.509	51.805	55.758	59.342	63.691	66.766
45	24.311	28.366	30.612	57.505	61.656	65.410	69.957	73.166
50	27.991	32.357	34.764	63.167	67.505	71.420	76.154	79.490
60	35.535	40.482	43.188	74.397	79.082	83.298	88.379	91.952
70	43.275	48.758	51.739	85.527	90.531	95.023	100.425	104.215
80	51.172	57.153	60.391	96.578	101.879	106.629	112.329	116.321
90	59.196	65.647	69.126	107.565	113.145	118.136	124.116	128.299
100	67.328	74.222	77.929	118.498	124.342	129.561	135.807	140.169

Tomado de: Bioestadística, Wayne Daniel, Limusa 1990



$P(F_{9,19} < 4.04) = .995$

TABLA D

PERCENTILES DE LA DISTRIBUCION F

$F_{.995}$

Grados de libertad del denominador	Grados de libertad del numerador 1	2	3	4	5	6	7	8	9
1	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

Grados de libertad del denominador	Grados de libertad del numerador 10	12	15	20	24	30	40	60	120	∞
1	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29
28	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25
29	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21
30	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

PERCENTILES DE LA DISTRIBUCION F $(F_{.99})$

continuación

Grados de libertad del denominador	Grados de libertad del numerador 1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Grados de libertad del denominador	Grados de libertad del numerador 10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

PERCENTILES DE LA DISTRIBUCION F ($F_{.975}$)

Continuación

Grados de libertad del denominador	1	2	3	4	5	6	7	8	9
	Grados de libertad del numerador								
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

Grados de libertad del denominador	10	12	15	20	24	30	40	60	120	∞
	Grados de libertad del numerador									
1	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

PERCENTILES DE LA DISTRIBUCIØN F ($F_{.95}$)

Continuación

Grados de libertad del denominador	Grados de libertad del numerador 1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Grados de libertad del denominador	Grados de libertad del numerador 10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

PERCENTILES DE LA DISTRIBUCION F ($F_{.90}$)

Continuación

Grados de libertad del denominador	Grados de libertad del numerador 1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.34	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Grados de libertad del denominador	Grados de libertad del numerador 10	12	15	20	24	30	40	60	120	∞
1	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Tomado de: Bioestadística, Wayne Daniel, Limusa 1990

TABLA E

NUMEROS ALEATORIOS

85967	73152	14511	85285	36009	95892	36962	67835	63314	50162
07483	51453	11649	86348	76431	81594	95848	36738	25014	15460
96283	01898	61414	83525	04231	13604	75339	11730	85423	60698
49174	12074	98551	37895	93547	24769	09404	76548	05393	96770
97366	39941	21225	93629	19574	71565	33413	56087	40875	13351
90474	41469	16812	81542	81652	45554	27931	93994	22375	00953
28599	64109	09497	76235	41383	31555	12639	00619	22909	29563
25254	16210	89717	65997	82667	74624	36348	44018	64732	93589
28785	02760	24359	99410	77319	73408	58993	61098	04393	48245
84725	86576	86944	93296	10081	82454	76810	52975	10324	15457
41059	66456	47679	66810	15941	84602	14493	65515	19251	41642
67434	41045	82830	47617	36932	46728	71183	36345	41404	81110
72766	68816	37643	19959	57550	49620	98480	25640	67257	18671
92079	46784	66125	94932	64451	29275	57669	66658	30818	58353
29187	40350	62533	73603	34075	16451	42885	03448	37390	96328
74220	17612	65522	80607	19184	64164	66962	82310	18163	63495
03786	02407	06098	92917	40434	60602	82175	04470	78754	90775
75085	55558	15520	27038	25471	76107	90832	10819	56797	33751
09161	33015	19155	11715	00551	24909	31894	37774	37953	78837
75707	48992	64998	87080	39333	00767	45637	12538	67439	94914
21333	48660	31288	00086	79889	75532	28704	62844	92337	99695
65626	50061	42539	14812	48895	11196	34335	60492	70650	51108
84380	07389	87891	76255	89604	41372	10837	66992	93183	56920
46479	32072	80083	63868	70930	89654	05359	47196	12452	38234
59847	97197	55147	76639	76971	55928	36441	95141	42333	67483
31416	11231	27904	57383	31852	69137	96667	14315	01007	31929
82066	83436	67914	21465	99605	83114	97885	74440	99622	87912
01850	42782	39202	18582	46214	99228	79541	78298	75404	63648
32315	89276	89582	87138	16165	15984	21466	63830	30475	74729
59388	42703	55198	80380	67067	97155	34160	85019	03527	78140
58089	27632	50987	91373	07736	20436	96130	73483	85332	24384
61705	57285	30392	23660	75841	21931	04295	00875	09114	32101
18914	98982	60199	99275	41967	35208	30357	76772	92656	62318
11965	94089	34803	48941	69709	16784	44642	89761	66864	62803
85251	48111	80936	81781	93248	67877	16498	31924	51315	79921
66121	96986	84844	93873	46352	92183	51152	85878	30490	15974
53972	96642	24199	58080	35450	03482	66953	49521	63719	57615
14509	16594	78883	43222	23093	58645	00257	89250	63266	90856
37780	07688	65533	72126	23611	[illegible]	01848	03920	38552	17472
85466	56302	72722	15473	73205	40759	56157	60477	83254	56367
52969	55363	42312	67842	05673	[illegible]	82734	36563	79540	51935
42744	68315	17514	02878	97201	[illegible]	47728	[illegible]	[illegible]	62941
26640	13536	67726	61876	20071	[illegible]	[illegible]	52817	96039	85211
95589	56349	14563	24071	00910	96555	18195	32291	79257	04224
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Tomado de : Wayne Daniel, Bioestadística, limusa 1990

TABLA F

VALORES CRITICOS PARA LA PRUEBA DEL RANGO CON SIGNO

n	Unilateral $\alpha = 0.01$ Bilateral $\alpha = 0.02$	Unilateral $\alpha = 0.01$ Bilateral $\alpha = 0.05$	Unilateral $\alpha = 0.01$ Bilateral $\alpha = 0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

Tomado de: Probabilidad y Estadística, Ronald E. Walpole, McGraw Hill Interamericana , 1991

TABLA G

VALORES CRITICOS DE U

U .05

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2							0	0	0	0	1	1	1	1
3				0	1	1	2	2	3	3	4	4	5	5
4			0	1	2	3	4	4	5	6	7	8	9	10
5		0	1	2	3	5	6	7	8	9	11	12	13	14
6		1	2	3	5	6	8	10	11	13	14	16	17	19
7		1	3	5	6	8	10	12	14	16	18	20	22	24
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34
10	0	3	5	8	11	14	17	20	23	26	29	30	36	39
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49
13	1	4	8	12	16	20	24	28	30	37	41	45	50	54
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64

U .02

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2												0	0	0
3						0	0	1	1	1	2	2	2	3
4				0	1	1	2	3	3	4	5	5	6	-
5			0	1	2	3	4	5	6	7	8	9	10	11
6			1	2	3	4	6	7	8	9	11	12	13	15
7		0	1	3	4	6	7	9	11	12	14	16	17	19
8		0	2	4	6	7	9	11	13	15	17	20	22	24
9		1	3	5	7	9	11	14	16	18	21	23	26	28
10		1	3	6	8	11	13	16	19	22	24	27	30	33
11		1	4	7	9	12	15	18	22	25	28	31	34	37
12		2	5	8	11	14	17	21	24	28	31	35	38	42
13	0	2	5	9	12	16	20	23	27	31	35	39	43	47
14	0	2	6	10	13	17	22	26	30	34	38	43	47	51
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56

U .10

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2				0	0	0	1	1	1	1	2	2	3	3
3		0	0	1	2	2	3	4	4	5	5	6	7	7
4		0	1	2	3	4	5	6	7	8	9	10	11	12
5	0	1	2	4	5	6	8	9	11	12	13	15	16	18
6	0	2	3	5	7	8	10	12	14	16	17	19	21	23
7	0	2	4	6	8	11	13	15	17	19	21	24	26	28
8	1	3	5	8	10	13	15	18	20	23	26	28	31	33
9	1	4	6	9	12	15	18	21	24	27	30	33	36	39
10	1	4	7	11	14	17	20	24	27	31	34	37	41	44
11	1	5	8	12	16	19	23	27	31	34	38	42	46	50
12	2	5	9	13	17	21	26	30	34	38	42	47	51	55
13	2	6	10	15	19	24	28	33	37	42	47	51	56	61
14	3	7	11	16	21	26	31	36	41	46	51	56	61	66
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72

U .01

n_1 \ n_2	3	4	5	6	7	8	9	10	11	12	13	14	15
3							0	0	0	1	1	1	2
4				0	0	1	1	2	2	3	3	4	5
5			0	1	1	2	3	4	5	6	7	7	8
6		0	1	2	3	4	5	6	7	9	10	11	12
7		0	1	3	4	6	7	9	10	12	13	15	16
8		1	2	4	6	7	9	11	13	15	17	18	20
9	0	1	3	5	7	9	11	13	16	18	20	22	24
10	0	2	4	6	9	11	13	16	18	21	24	26	29
11	0	2	5	7	10	13	16	18	21	24	27	30	33
12	1	3	6	9	12	15	18	21	24	27	31	34	37
13	1	3	7	10	13	17	20	24	27	31	34	38	42
14	1	4	7	11	15	18	22	26	30	34	38	42	46
15	2	5	8	12	16	20	24	29	33	37	42	46	51

Tomado de: Estadística Matemática con Aplicaciones, John Freund/Ronald Walpole
Prentice Hall 1990

PROBLEMAS RESUELTOS

MODULO 1 "PROBABILIDAD"

1. Un fabricante tiene 5 terminales de Computadora, aparentemente idénticas, él sabe que dos de las cinco son defectuosas. Recibe un pedido de dos terminales y lo surte seleccionando al azar dos de las cinco.

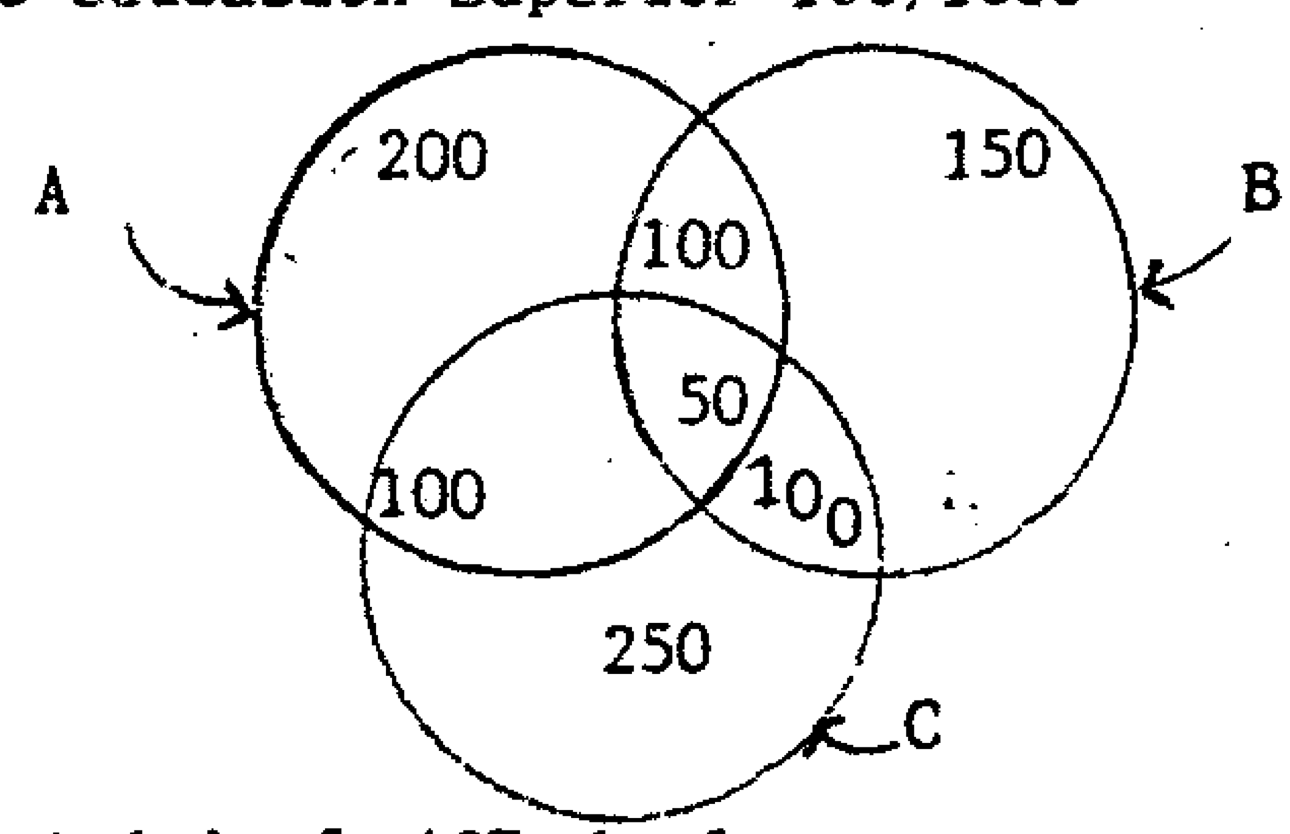
A. El espacio muestral del experimento es:
$S= D_1D_2,\ D_1B_2, D_1B_3, D_2B_1, D_2B_2, D_2B_3, B_1B_2, B_1B_3, B_2B_3, D_1B_1$

B. El evento A, surte con dos terminales no defectuosas
$A= B_1B_2,\ B_1B_3,\ B_2B_3$

C. La probabilidad de evento A es
P(A)= 3/10

2. Cierta población de 1000 hombres adultos presentan tres características: A. ser casado, B. Tener grado de educación superior, C. Ser originario de un Estado específico.

Según las cifras indicadas en el diagrama, la probabilidad de que:
Un individuo sea casado es 500/1000
Tenga un grado de educación superior y esté casado 150/1000
Que no sea del Estado específico pero si casado y que tenga un grado de educación superior 100/1000



3. En cierta ciudad el 40% de los votantes son Republicanos y el 60% Demócratas, el 70% de los Republicanos y el 80% de los Demócratas están a favor de una inversión particular de bonos, se selecciona al azar un votante de la ciudad. La probabilidad que esté a favor de la inversión de bonos es:

P(R)= 0.4 P(D)= 0.6 P(F/R)= 0.7 P(F/D)= 0.8
P(F)= 0.7*0.4 + 0.6*0.8 = 0.76

4. Una compañía dedicada al montaje de ventiladores usa motores de dos proveedores, la compañía A proporciona el 90% de los motores y la compañía B el otro 10%. Se sabe que el 5% de los motores de A son defectuosos mientras que el 3% de los motores de B. Se encuentra que un ventilador que ha sido probado tiene motor defectuoso. La probabilidad que haya sido proporcionado por la compañía B es:

P(A)= 0.9 P(B)= 0.1 P(D/A)= 0.05 P(D/B)= 0.03
P(D)= 0.9*0.05 + 0.1*0.03 = 0.048 $P(B/D)= \dfrac{0.1*0.3}{0.048} = 0.0625$

5. Un sistema para detectar humo utiliza dos dispositivos A y B. Si hay humo la probabilidad de detectarlo por A es 0.95 y por B es 0.98, operan en forma independiente. Si hay humo la probabilidad que sea detectado por almenos uno de los dos dispositivos (A o B o ambos) es

$P(A \cup B) = 0.95+0.98-0.95*0.98=0.999$

La probabilidad que no sea detectado el humo es:

$1- P(A \cup B) = 1-0.999 = 0.001$

MODULO 2 "DISTRIBUCIONES DE PROBABILIDAD"

1. Se sabe que un grupo de cuatro componentes tienen dos defectuosos, un inspector prueba uno a uno hasta encontrar los dos defectuosos, una vez encontrados se concluye la prueba pero se prueba el segundo defectuoso como comprobación. Sea Y el número de pruebas necesarias para encontrar el segundo defectuoso.

El recorrido de la variable Y es $R_y = (2,3,4)$

La distribución de probabilidades de Y es

y	P(Y)
2	1/261/3=1/6
3	1/2*2/3*1/2 +1/2*2/3*1/2= 1/3
4	1/2*1/3*1 +1/2*2/3*1/2 + 1/2*2/3*1/2 = 1/2

El valor esperado de la Varieble es :

Y	P(Y)
2	1/6
3	2/6
4	3/6

$E(Y) = 2/6 + 6/6 + 12/6 = 20/6 = 3.33$

La varianza de la Variable es:

$E(Y^2) = 4/6 + 18/6 + 48/6 = 11.67$

$V(Y) = 11.667 - 3.33^2 = 0.577$

La desviación estándar de la variable es

$\sqrt{V(Y)} = 0.76$

2. Un fruticultor afirma que 1/3 de su consecha de duraznos está contaminada. La probabilidad que al selecciónar 4 duraznos al azar ninguno está contaminado.

X= Número de frutos contaminados $p = 1/3$ $q = 2/3$

$$P(x=0) = {}_4C_0 * (2/3)^4 * (1/3)^0 = 0.1976$$

3. Se su pone que el 10% de los vasos fabricados por determinada máquina tienen algún tipo de defecto, si se selecciónan al azar 10 de éstos vasos, la probabilidad de encontrar menos de 3 defectuosos es:

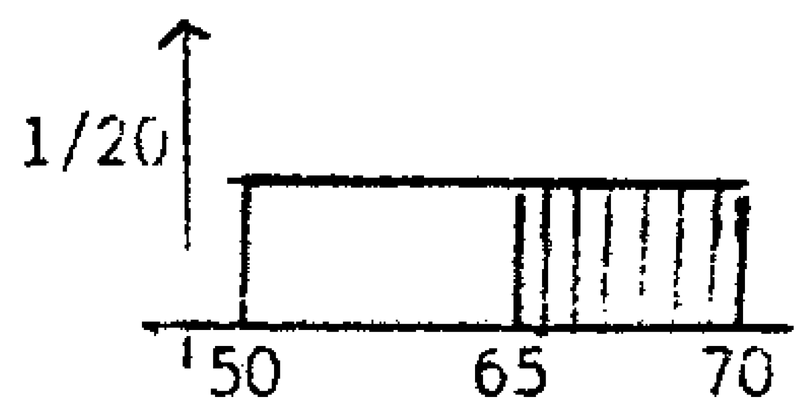
$n = 10$ $p=0.1$ $q=0.9$

$$P(X < 3) = P(X=0,1,2) = {}_{10}C_0*0.9^{10} + {}_{10}C_1*0.9^9*0.1^1 + {}_{10}C_2*0.9^8*0.1^2$$

$$= 0.9278$$

El número de vasos defectuosos esperados es: $n*p = 10 * 0.1 = 1$

4. El tiempo de viaje, ida y vuelta, de los caminoes que transportan concreto hacia una obra de construción en una carretera está distribuído uniformemente en un intervalo de 50 a 70 minutos. La probabilidad que la duración del viaje sea mayor de 65 minutos es



$f(x)=1/(70-50) = 1/20$

$P(X > 65) = (70-65) * 1/20 = 5/20$

5. Una variable tiene una distribución Normal con una desviación estándar de 21.5 unidades. La media de la distribución, si la probabilidad de que la variable tome un valor menor que 120.5 es de 0.8849, se calcula:

$P(X < 120.5) = 0.8849$

$P(Z < 1.20) = 0.8849$ $\qquad Z = \dfrac{X - \mu}{\sigma} = \dfrac{120.5 - \mu}{21.5} = 1.2$

$\mu = 94.7$

6. El tiempo requerido para ensamblar una pieza mecánica es un variable aleatoria cuya distribución de probabilidad es aproximadamente Normal, con media 12.9 y desviación estándar 2. La probabilidad que el ensamblado de tal pieza tarde entre 11.0 y 14,8 minutos es:

$\mu = 12.9$ minutos $\qquad \sigma = 2.0$ minutos

$P(11 < X < 14.8) = P\left(\dfrac{11 - 12.9}{2} < Z < \dfrac{14.8 - 12.9}{2} \right)$

$P(-0.95 < Z < 0.95) = 0.8298 - 0.1711 = 0.6578$

MODULO 3 "DISTRIBUCIONES MUESTRALES"

1. Una compañía de alimentos que tiene 122 supermecados ha sido adquirida por un inversionista que antes de cerrar el trato quiere estar seguro que hará una buena inversión, para ello decide analizar el registro de 40 tiendas. La gerencia de esta afirma que las utilidades de cada estableciemiento tiene una distribución normal con media \$14,000 y desviación estándar \$1000. Los límites entre los cuales espera el inversionista encontrar el 90% de las medias muestrales son:

$\mu = 14000 \qquad \sigma = 1000 \qquad \sigma_{\bar{x}} = 1000/\sqrt{40} = 158.1$

$Z_{90\%} = 1.645$ $\qquad$ Límites: $14000 \pm 1.645 * 158.11$

$13739.90 \ , \ 14260.10$

La probabilidad de que la media de las 40 tiendas difiera de la media real en ± $150 es:

$$Z = \pm \frac{150}{158.11} = \pm 0.948$$

Area a - 0.948 es 0.32894
Area a + 0.948 es 0.32894
area total 0.66

2. Según las publicaciones de la revista Epoca los censos muestran que el 53% de las viviendas familiares de cierta ciudad están ocupadas por 3 o 4 personas, suponiendo que éste porcentaje es válido, la probabilidad de que en 1000 viviendas seleccionadas al azar por lo menos el 52% estén ocupadas por 3 o 4 personas.

$p=0.53 \qquad q=0.47 \qquad n=1000$

$$P(P > 0.52) = P\left(Z > \frac{0.52-(1/2000) - 0.53}{\sqrt{\frac{0.53 * 0.47}{1000}}} \right)$$

$$P\left(Z > \frac{-0.0105}{0.0157}\right) \doteq P(Z > -0.665) = 1-0.2514$$

3. Un fabricante de acumuladores para automóvil garantiza que sus productos durarán en promedio 3 años, con una desviación estándar de 1 año. Si cinco de estos acumuladores tiene duracion de 1.9, 2.4, 3.0, 3.5, 4.2 años ¿ Estará aún convencido el fabricante que su producto tiene una desviación estándar de un año?

La varianza da la muestra es $S^2 = 0.815$

$X^2 = 4 * 0.815 / 1 = 3.26$

Que es un valor de la distribución Chi-cuadrado con 4 grados de libertad. Puesto que el 95% de los valores de X^2 con 4 grados de libertad se ubican entre 0.484 y 11.14 es razonable el valor calculado con $\sigma^2=1$ Por consiguiente, el fabricante no tiene razones para esperar que la desvición estándar no sea de un año.

MODULO 4 " ESTIMACION "

1. Las medidas de los diámetros de una muestra al azar de 200 cojinetes hechos por una máquina durante una semana dieron una media de 0.824 pulgadas y una desviación típica de 0.042. Los límites de confianza del 99% para el diámetro medio de la población de cojinetes es:

$X' = 0.824 \qquad S = 0.042 \qquad Z_{99\%} = \pm 2.58$

Intervalo : $0.824 \pm 2.58 * (0.042/\sqrt{200}) = 0.824 \pm 0.0077$

2. Una muestra de 100 votantes elegidos al azar entre todos los de un distrito indican que el 55 % están a favor de un determinado candidato. El Intervalo del 99.73% de confianza para la proporción de votantes que están a favor de ese candidato es:

$P = 0.55 \quad Q = 0.45 \quad Z_{99.73\%} = 3$

Límites de p : $0.55 \pm 3 * \sqrt{\frac{0.55*0.45}{100}}$

0.55 ± 0.15

3. Al medir el tiempo de reacción un psicólogo estima que la desviación típica del mismo es 0.05 segundos. El número mínimo de medidas que deberán hacerse para que se tenga el 95% de confianza de que el error de su estima para el promedio no exceda de 0.01 segundos es:

$Z_{95} = 1.96$ Error de estima 0.01

$n = (Z*\sigma / E)^2 = (1.96*0.05/0.01)^2 = 96.04$

El tamaño de muestra no debe ser menor de 97 mediciones.

4. Los contenidos de ácido sulfúrico de 7 recipientes similares son: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6 litros. Obtenga un intervalo del 95% de confianza para la media de todos los recipientes, suponiendo que la distribución es aproximadamente normal.

La media y la desviación estándar de la muestra son: X'=10.0 S=0.283

El valor de t para 6 grados de libertad es: $t_{0.95\%} = 2.447$

Intervalo: $10.0 \pm 2.477 \; (0.283/\sqrt{7})$ lo que se reduce a

$9.74 < \mu < 10.26$

5. Los siguintes son los pesos, en kilogramos. de 10 paquetes de semillas de pasto, distribuidos por cierta compañía: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, 46.0. El intervalo de confianza de 95% para la varianza de la población de todos los paquetes de semillas distribuidos por la empresa, suponiendo una población normal, es:

La varianza de la muestra es $S^2 = 0.286$

Utilizando la distribución Chi-cuadrado con 95% de confianza y 9 grados de libertad se obtiene:

$X^2_{0.025} = 19.023$ y $X^2_{0.975} = 2.70$

El intervalo está dado por: $\frac{9*0.286}{19.023} < \sigma^2 < \frac{9*0.286}{2.70}$

$0.135 < \sigma^2 < 0.953$

MODULO 5 "HIPOTESIS"

1. Cierta compañía ha usado en sus empleados una prueba de inteligencia, en los empleados presentes el resultado medio es de 50 con una desviación estándar de 10. Un investigador desea comprobar la teoría que los principales supervisores obtienen resultados más altos que los empleados promedio, para ratificar ésta teoría se selecciona una muestra al azar de 100 supervisores y se calcula la media de los resultados, siendo de 53. Son significativos los resultados para la teoría del Investigador?

Problema: La teoría del investigador supone que los supervisores tienen resultados más altos.
Ho = El promedio de los supervisores es igal al de los otros empleados.
H1 = El promedio de los supervisores es mayor que el de los otros empleados.

Ho: $\mu = 50$ H1: $\mu > 50$
Nivel de significancia 0.01 $Z_c = 2.33$

$\mu = 50$ $\sigma = 10$ $X' = 53$ $Z = \dfrac{53 - 50}{10/\sqrt{100}} = 3$

Son significativos los resultados $Z > Z_c$ $3 > 2.33$

2. De experimentos anteriores se ha determinado que un operador calificado para cierta máquina trabajando 400 artrículos produce 20 o menos defectuosos. Un operador nuevo es empleado para éste trabajo considerando que es un trabajador calificado. Si en un día de trabajo hizo 32 defectuosos de los 400, compruebe la hipótesis "es un operador Calificado" con un nivel de significancia de 5%.

Ho La proporción de defectuosos es 0.05 o menos para un operador calificado $p = 20/400 = 0.05$

H1 La proporción es mayor $p > 0.05$, no es un operador calificado

Nivel de significancia 0.05 $Z_c = 1.64$

Distribución Binomial $p = 0.05$ $q = 0.95$ $n=400$ $P = 32/400$

$$Z = \frac{P - p}{\sqrt{\frac{p\,q}{n}}} = \frac{0.08 - 0.05}{\sqrt{\frac{0.05*0.95}{400}}} = 2.75$$

Son significativos los resultados $Z > Z_c$ $2.75 > 1.64$
El operador no es calificado para la máquina.

3. Se realizó un experimento para comparar la resistencia abrasiva de dos diferentes materiales laminados, se probaron 12 piezas del material 1 exponiendo cada pieza a una máquina medidora de dicha resistencia. Análogamente se probaron 10 piezas del material 2. En cada caso se observó la capacidad de resistencia. La muestras del material 1 dieron un promedio de 85 unidades con una desviación estándar de 4, en tanto que

las muestras del material 2 dieron un promedio de 81 con una desviación estándar de 5. ¿ Puede concluirse que con un nivel de significancia de 0.05 la resistencia abrasiva del material 1 supera a la del material 2 en más de 2 unidades?. Suponga que las poblaciones son aproximadamente normales con varianzas iguales.

$H_o \mu_1 - \mu_2 = 2 \qquad H_1 \mu_1 - \mu_2 > 2$

Nivel de significancia 5 % $t_c = 1.725$ con 20 grados de libertad

$X'_1 = 85 \quad X'_2 = 81 \quad S_1 = 4 \quad S_2 = 5 \quad N_1 = 12 \quad N_2 = 10$

$$S_p = \sqrt{\frac{11*16 + 9*25}{12+10-2}} = 4.478$$

$$t = \frac{(X'_1 - X'_2) - D_o}{S_p \sqrt{(1/n_1 + 1/n_2)}}$$

$$t = \frac{85 - 81 - 2}{4.478 \sqrt{(1/12)+(1/10)}} = 1.04$$

Se acepta H_o y se concluye que la resistencia abrasiva del material 1 no supera en más de 2 unidades a la del material 2

4. Se va a realizar una votación entre los residentes de un pueblo y sus alrededores para determinar si una planta química propuesta debería construirse. El lugar de contrucción está dentro de los límites del pueblo y por esta razón, muchos de los votantes de la zona creen que la propuesta será aprobada debido a la gran proporción de votantes del pueblo que favorecen la construcción de dicha planta. Para determinar si existe una diferencia significativa en la proporción de votantes del pueblo y los votantes de la zona que favorecen la propuesta, se llevó a cabo un escrutinio. Si 120 de los 200 votantes del pueblo favorecen la propuesta y 240 de 500 residentes de la zona la favorecen también. ¿ Se estaría de acuerdo en que la proporción de votantes del pueblo a favor de la propuesta es mayor que la proporción de votantes de la zona? Utilizar un nivel de significancia de 0.025.

$H_o: p_1 = p_2 \qquad H_1: p_1 > p_2$

Nivel de significancia 2.5% $Z_c = 1.96$

$P_1 = 120/200 = 0.6 \qquad P_2 = 240/500 = 0.48$

$P' = (120 + 240) / (200 + 500) = 0.51$

$$Z = \frac{0.60 - 0.48}{\sqrt{(0.51*0.49) ((1/200) + (1/500))}} = 2.9$$

Rechazar H_o y aceptar que la proporción de votantes del pueblo a favor de la propuesta es mayor que la proporción de votantes de la zona.

5. Un fabricante de baterías para automóviles afirma que la vida de sus baterías está aproximadamente distribuída en forma normal con una desviación estándar de 0.9 años. Si una muestra aleatoria de 10 baterías tiene una desviación estándar de 1.2 años. ¿ Se puede pensar que σ es 0.9 años?

H_o $\sigma^2 = 0.81$ H_1 $\sigma^2 > 0.81$

Nivel de significancia 0.05 $X^2_c = 16.919$ con 9 grados de libertad

$X^2 = \dfrac{(n-1)\ S^2}{\sigma^2_o}$ donde $S^2 = 1,44$ n=10 $X^2 = 9 * 1.44/0.81 = 16$

Aceptar H_o y concluir que no hay razón suficiente para dudar de que la desviación estándar es 0.9 años

MODULO 6 "REGRESION Y CORRELACION"

1. Un coeficiente de correlación basado en una muestra de tamaño 18 resultó ser 0.32 ¿ Se puede deducir al nivel de significancia de 0.05 que el coeficiente de la población correspondiente difiere de cero?

H_o $\rho = 0$ H_1 $\rho > 0$

$t = r \sqrt{(n-2)/\ (1-r^2)} = 0.32 \sqrt{((18-2)/\ (1-0.32^2))} = 1.35$

Según el ensayo unilateral de una distribución de Student al nivel de significancia de 0.05 se rechazaría H_o si $t > t_{95} = 1.75$ para 16 grados de libertad. Así que no se puede rechazar H_o al nivel de 0.05

2. Cuál es el tamaño mínimo de muestra necesario para que se pueda deducir que un coeficiente de correlación de 0.32 difiere significativamente de cero al nivel 0.05?

Al nivel de 0.05 y mediante un ensayo unilateral de la distribución de Student el valor mínimo de n debe ser tal que:

$0.32 \sqrt{((n-2)\ /\ (1-o.32^2))} = T_{95}$ para n-2 grados de libertad

Para un número infinito de grados de libertad $t_{0.95}$ es 1.64 de aquí n= 25.6

para n= 26 v=24 $t_{0.95} = 1.71$ $t = 0.32 * \sqrt{(24/\ (1-0.32^2))} = 1.65$

para n= 27 v=25 $t_{0.95} = 1.71$ $t=0.32 * \sqrt{(25/\ (1-0.32^2))} = 1.69$

para n = 28 v=26 $t_{0.95} = 1.71$ $t=0.32 * \sqrt{(26\ /(1-0.32^2))} = 1.72$

El tamaño mínimo de muestra es n= 28

3. Un coeficiente de correlación basado e una muestra de tamaño 24 resultó ser 0.75. ¿ Puede rechazarse la hipótesis de que el coeficiente de correlación poblacional sea 0.60 a un nivel de significancia de 0.05?

$z = \sqrt{(24-3)/2} * \ln\ (\ (1+0.75)(1-0.6)/(1-0.75)(1+0.6) = 1.38$

Al nivel de significaicón de 0.05 mediante un ensayo unilateral de la distribución normal, se rechazará la hipótesis solamente si Z es mayor de 1.64. Así no se puede rechazar la hipótesis de que el coeficiente de correlación Poblacional sea tan pequeño como 0.6

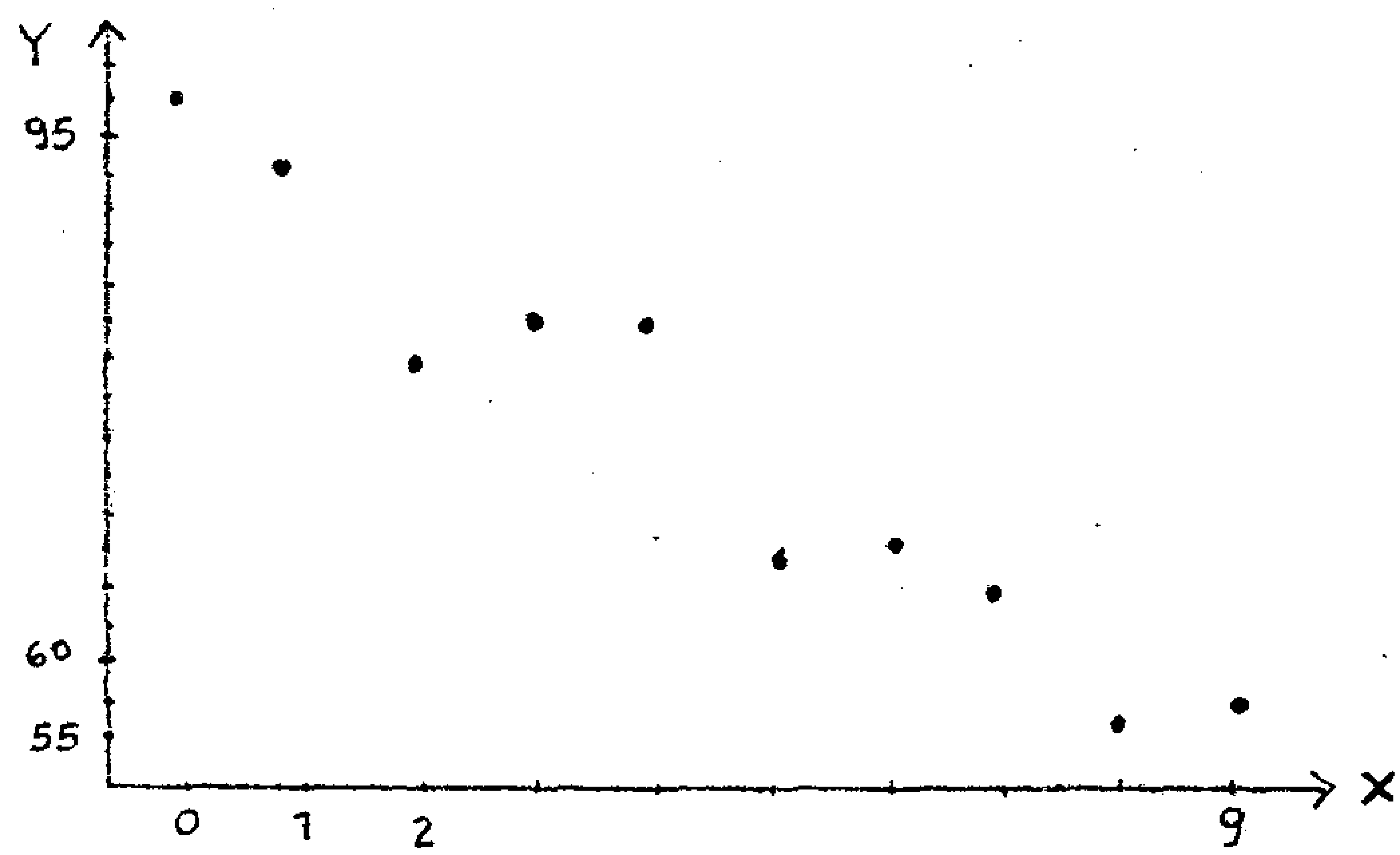
4. La producción en Estados Unidos de Cigarros puros durante los años 1974 y 1983 aparece en la siguiente tabla.

A. Representar los datos gráficamente

B. Hallar la ecuación de la recta de mínimos cuadrados

C. Estimar la producción de cigarros puros durante el año 84

Año	X	Y
74	0	98.2
75	1	92.3
76	2	80.0
77	3	83.5
78	4	83.5
79	5	68.9
80	6	69.2
81	7	67.1
82	8	58.3
83	9	61.2



Linea de Regrasión: $a=96.2$ $b=-4.30$

$$Y= 96.2 - 4.30\ X$$

Donde el origen es cero y las unidades de X son un año

Mediante la ecuación con X= 10 correspondiente al año 84 Y resulta igual a : $Y = 96.2 - 4.3\ (10)= 53.2$

MODULO 7 "METODOS NO PARAMETRICOS"

1. Un experimento diseñado para comprobar tres métodos preventivos contra la corrosión produjo las siguientes profundidades máximas de las cavidades (en milésimas de pulgada) en piezas de alambre que fueron sumetidas a los tratamientos respectivos:

Método A:	77	54	67	74	71	66	
Médodo B:	60	41	59	65	62	64	52
Método C:	49	52	69	47	56		

Con un nivel de significancia de 0.05 pruébese la hipótesis nula de que las muestras provienen de poblaciones idénticas.

Hipótesis Nula: las poblaciones son idénticas
Hipótesis Alternativa: las poblaciones no son inguales
Nivel de significancia 0.05
Criterio : se rechaza la H_o si $H > 5.991$ o sea el valor de $X^2_{0.05}$ para dos grados de libertad.

Cálculos: Clasifincando esas observaciones conjuntamente de la más pequeña a la mayor se tiene que las de la primera ocupan los rangos 6,13, 14, 16,17 y 18. Las mediciones de la segunda muestra ocupan los rangos, 1, 4.5,8,9,10,11,12 y las de la tercera muestra ocupan los rangos 2,3,4.5, 5,7,y 15. En consecuancia $R_1 = 84$ $R_2 = 55.5$ $R_3 = 31.5$

$$H = \frac{12}{18*19} \left(84^2/6 + 55.5^2/7 + 31.5^2/5 \right) - 3.19 = 6.7$$

Dado que H = 6.7 exceda a 5.991 la hipótesis nula debe rechazarse Se concluye que los tres métodos preventivos contra la corrosión no tienen igual eficacia.

2. Los siguientes datos representan el número de horas que un temporizador opera entes de que deba recargarse: 1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8,1.5 2.0, 1.2, y 1.7 . Pruebe la hipótesis al nivel de significancia de 0.05 que éste temporizador en particular opera con una media de 1.8 horas antes de requerir recarga.

H_o $\mu = 1.8$ H_1 $\mu \neq 1.8$ Nivel de significancia 0.05

Dado que n = 10 después de descartar la única medición que es igual a 1.8 se encuentra que la región crítica es $w < 8$

Al restar 1.8 a cada medición y después dándole rangos a las diferencias sin considerar el signo se tiene:

d_i	-0.3	0.4	-0.9	-.05	0.2	-0.2	-0.3	0.2	-0.6	-0.1
Rangos	5.5	7	10	8	3	3	5.5	3	9	1

Ahora W+ = 13 y W- = 42 de tal forma que W= 13 el más pequeño.

No se rehcaza H_o y se concluye que el tiempo de operación promedio no es significativamente diferente de 1.8 horas.

PROPIEDAD DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
Biblioteca Central

3. Las cifras que aparecen en la tabla obtenidas por la Federal Trade Commisison, indican los miligramos de alquitrán y nicotina encontrados en diez marcas de cigarros. Calcule el coeficiente de correlación de rangos para medir el grado de relación entre los contenidos de alquitrán y nicotina en los cigarros.

Sea X y Y los contenidos de alguitrán y nicotína, respectivamente. Primero se asignan rangos a cada conjunto de mediciones, con el rango 1 asignado al número más pequeño en cada conjunto, el rango 2 al segundo número más pequeño y así consecutivamente. Posteriormente se calculan las diferencias en rangos para los diez pares de observaciones.

Al sustituir en la fórmula para r_s se encuentra que :

$$r_s = 1 - \frac{6 * 5.5}{10 * (100-1)} = 0.97$$

Indica una correlación positiva alta entre las cantidades de alquitrán y nicotina encontradas en los cigarros.

Marca de Cigarros	Alquitrán	Nicotina	X_i	Y_i	d_i
Viceroy	14	0.9	2	2	0
Malboro	17	1.1	4.5	4	0.5
Chesterfiel	28	1.6	9	9	0
Kool	17	1.3	4.5	6	-1.5
Kent	16	1.0	3	3	0
Raleigh	13	0.8	1	1	0
Old Gold	24	1.5	7	8	-1
Philip Morris	25	1.4	8	7	1
Oasis	18	1.2	6	5	1
Players	31	2.0	10	10	0