COLEGIO DE BACHILLERES

DIRECCION DE PLANEACION ACADEMICA

COORDINACION DEL SISTEMA DE ENSEÑANZA ABIERTA

ESTADISTICA DESCRIPTIVA E INFERENCIAL II

MODULO 3

FUNCION DE DISTRIBUCION NORMAL ESTANDAR

AUTORES: JUAN MATUS PARRA

EMIGDIO ARROLLO CERVANTES

PRESENTACION GENERAL

El Colegio de Bachilleres, dentro de su plan de trabajo 1991-1994, consideró necesario impulsar la actualización y homogeneización de los programas de su plan de estudios, en sus modalidades escolarizada y abierta.

Con este propósito, y con una amplia participación de maestros del Colegio, se desarrollaron los trabajos de actualización, orientados al fortalecimiento de la formación propedéutica universitaria de sus egresados, de tal manera que nuestra institución responda mejor, desde su ámbito de competencia, a los requerimientos del país.

Como fruto de ese esfuerzo académico de profesores del Colegio de Bachilleres en colaboración con asesores pedagógicos y de contenido, se proporcionan a nuestros estudiantes estos fascículos de apoyo al aprendizaje, los que en forma dinámica se irán mejorando en la medida que se recojan las experiencias directas y enriquecedoras que aporta el ejercicio educativo.

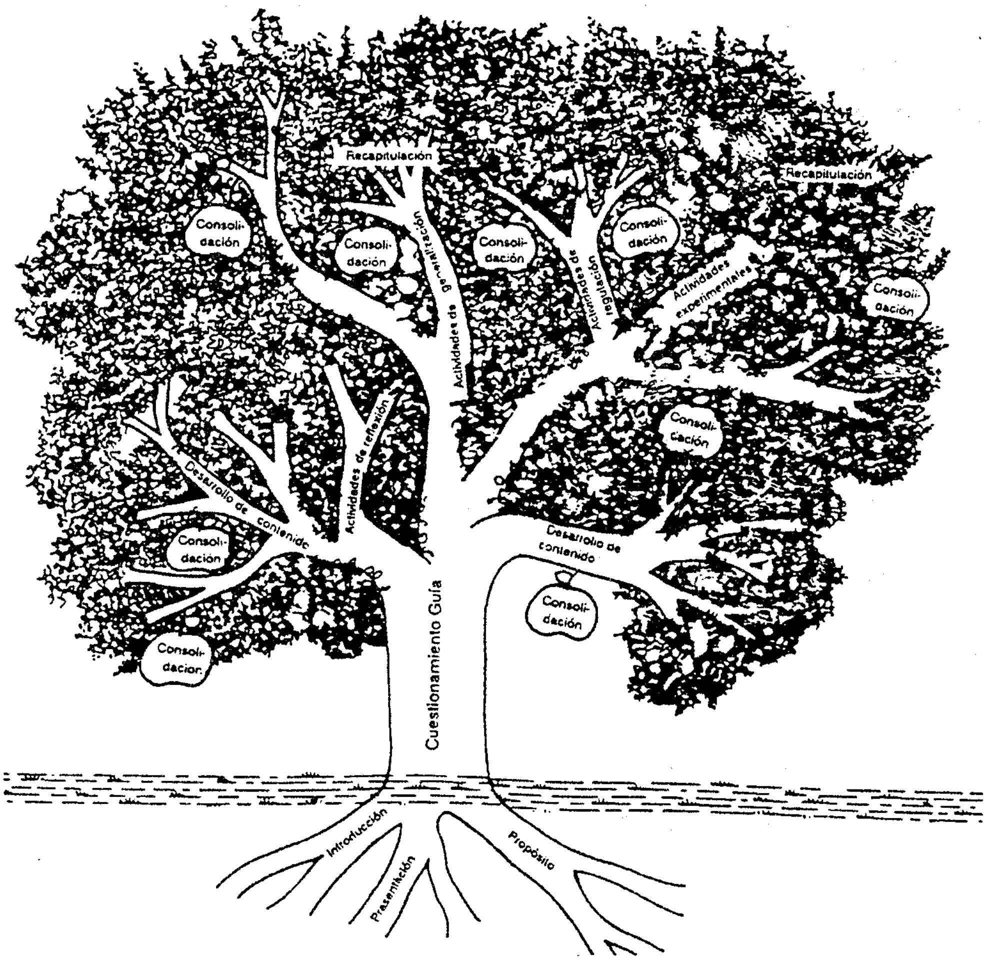
DIRECCION GENERAL

INDICE

PRESENTACION

El Colegio de Bachilleres, en apoyo a su programa "Actualización y Homogeneización de los Programas del Plan de Estudios", preparó el presente fascículo "Función de Distribución Normal Estandar", el cual constituye el tercero de una serie de cuatro que integran la asignatura Estadística Descriptiva e Inferencial II.

En su contenido se utilizan diversos elementos de manera que te facilitan el aprendizaje y la construcción del conocimiento para que, al finalizar su estudio puedas aplicar lo aprendido en las diferentes actividades de tu vida diaria. Estos elementos son:



PRESENTACION
PROPOSITO
INTRODUCCION
CUESTIONAMIENTO GUIA

Te ponen en contacto con lo que vas a aprender, cómo lo vas a lograr y la utilidad que obtendrás con su estudio; además te indican cómo se organiza el material invitándote a reflexionar sobre lo que vas a aprender.

DESARROLLO DE CONTENIDO
ACTIVIDADES
EXPLICACIONES INTEGRADORAS

Te permite analizar y aplicar los contenidos para que los construyas y reconstruyas y de esta manera los hagas tuyos, y seas un arquitecto del conocimiento.

RECAPITULACION

Te proporciona una síntesis al relacionar los conceptos relevantes de los temas tratados en el fascículo.

ACTIVIDADES DE CONSOLIDACION
LINEAMIENTOS DE AUTOEVALUACION

Te posibilita aplicar los conocimientos construidos a través de interrogantes y situaciones problemáticas para que integres lo aprendido y así confirmes tus conocimientos.

ACTIVIDADES DE GENERALIZACION
GLOSARIO
BIBLIOGRAFIA

Son elementos que te permiten aclarar conceptos técnicos y especializados propios de la asignatura, además de proyectar los contenidos del fascículo hacia otros campos científicos y tecnológicos.

PROPOSITO

Con este fascículo puedes estudiar las funciones probabilísticas contínuas su distribución normal estandar, las distribuciones muestrales y teorema central del límite, así como la distribución T de Student.

Estos temas que parecen sin sentido, toman una importancia fundamental cuando hablamos de investigación. Para cualquier ciencia o profesión la investigación juega un papel preponderante, y para que esta sea aceptada científicamente es necesario que cuente con datos fidedignos y sistematizados adecuadamente ¡y esta es la contribución de "nuestros temas sin sentido"! A través de su estudio puedes aprender a sistematizar los elementos cuantitativos de cualquier investigación, y si estás pensando ¡Yo no seré investigador, seré médico, o sociólogo, o químico, o pedagogo, o....! Nosotros tendríamos que contestarte que todos, absolutamente todos, en algún momento de su vida profesional hacen investigación y entonces requerirás estas herramientas.

Para que puedas ejercitar los contenidos que integran el fascículo, aparecen una serie de actividades, ¡no dejes de hacerlas!

INTRODUCCION

En el fascículo anterior estudiaste las distribuciones de probabilidad binomial y de Poisson. Estas son distribuciones de variable aleatoria discreta, en que cada valor de la variable se le asigna una probabilidad.

Existen otras distribuciones de probabilidad, las de variables aleatorias contínuas cuya determinación de la probabilidad difiere de las anteriores toda vez que las observaciones del experimento generan un espacio muestral infinito y cada intervalo de este tiene un número infinito no numerable de posibles resultados los cuales incluyen valores reales.

De lo anterior podemos concluir que para determinar la probabilidad de una variable aleatoria continua, se desarrolla un método distinto a los anteriores.

En este fascículo estudiarás la distribución normal como modelo de fenómenos aleatorios en los que se efectúan mediciones contínuas y te capacitarás en el cálculo de la probabilidad de fenómenos aleatorios de regularidad estadística, aplicando para ello, la distribución normal estándar.

Así mismo, estudiarás la aplicación de la distribución de medias muestrales mediante el uso del Teorema del Límite Central para muestras grandes y la distribución de Students en muestras pequeñas.

CUESTIONAMIENTO GUIA

Sabemos que las aguas negras de la Ciudad de México se utilizan para el riego de los campos de cultivo circunvecinos al Valle de México.

Estas aguas negras contienen entre otras sustancias, el cloro en cantidades perjudiciales al sembradío de cereales porque en lugar de beneficiarlo con el riego, lo quema y lo seca.

Por lo anterior, es necesario darle al agua un tratamiento con el fin de disminuir o eliminar el contenido de cloro. Para ello el Departamento del D.F. tiene establecido un laboratorio en los colectores de aguas para determinar el contenido de cloro y dar el tratamiento correspondiente antes de abrir las compuertas.

Para el análisis se toma una muestra de 5 lt. de aguas negras diariamente. Los resultados correspondientes al mes de noviembre de 1993 fueron las que se muestran en la siguiente tabla. Las cantidades de cloro se registran en partes por millón (ppm).

16.2	15.4	16.0	16.6	15.9	15.8	16.0	16.8	16.9	16.8
15.7	16.4	15.2	15.8	15.9	16.1	15.6	15.9	15.6	16.0
16.4	15.8	15.7	16.2	15.6	15.9	16.3	16.3	16.0	16.3

Usemos estos datos para realizar un recordatorio de los conceptos estudiados en tu curso de Estadística I. Esto nos servirá para abordar los nuevos conceptos que estudiarás en este fascículo y para ello realiza el siguiente ejercicio:

1. Ordena los datos en sentido creciente.
2. Determina el rango de variación de los datos.
3. Elabora una tabla de frecuencia de datos agrupados de 5 clases.
4. Determina la moda, la mediana y la media de la muestra.
5. Determina la desviación estándar.
6. Traza el histograma
7. Traza el polígono de frecuencia.
8. Analiza el polígono de frecuencias y determina:
    a) De qué tipo es (platicúrtica, mesocúrtica, etc.)
    b) Determina el sesgo
    c) Determina el orden de la media, la moda y la mediana
9. Analiza la desviación estándar y determina como es la dispersión de las puntuaciones.

FUNCIONES PROBABILISTICAS CONTINUAS

En el siglo XVIII a los jugadores profesionales les interesaba conocer a priori, las probabilidades de éxito en los distintos juegos de azar, para ello acudieron a los matemáticos de la época en busca de ayuda. Como una respuesta a una necesidad planteada a los matemáticos, en 1973 Abraham D'Moavre (1667-1754) es quien obtiene por primera vez la ecuación matemática de la curva normal.

La distribución normal nos permite el cálculo de probabilidades de variables aleatorias contínuas y discretas de cualquier problema de: Ingeniería, Medicina, Ciencias Sociales, Agricultura, Psicología, Física, Química, etc.

Otros grandes matemáticos contribuyeron dándole impulso, entre ellos podemos citar a Friedrich Gauss (1777-1855) quien la perfeccionó y la utilizó ampliamente en su teoría de errores de las mediciones físicas. Laplace la usó en el cálculo de los errores de las observaciones astronómicas. El matemático Ruso P. L. Chebyshev estableció varios teoremas relacionados con la curva de la distribución normal.

Los experimentos realizados por muchos científicos, permiten determinar que la mayor parte de las variables aleatorias se pueden estudiar considerando que tiene una función de densidad normal.

Distribución Normal Estandar

Retomemos el problema de las aguas negras. Los resultados que debiste obtener son:

R = 1.7	$\underline{M}$ = 16.01
M = 16.05	$\bar{X}$ = 16.08
S = 0.42	

El histograma y el polígono de frecuencias son los siguientes:

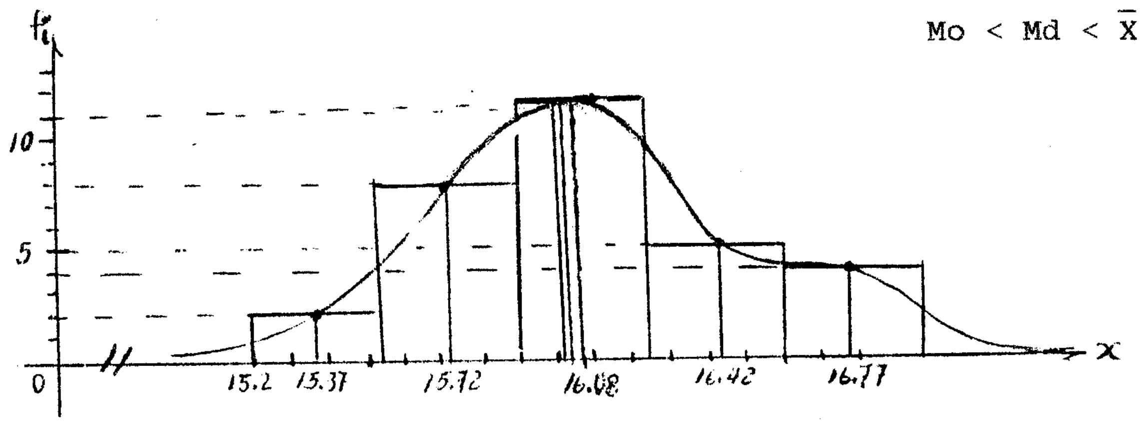


Fig.1

Del polígono de frecuencias podemos ver que la curva es asimétrica; está sesgada a la izquierda por lo tanto su asimetría es negativa. Por su puntiagudez es del tipo leptocúrtica.

Recordarás que los polígonos de frecuencias pueden ser:

1. Simétricos (Gráfica A)

2. Asimétricos (Gráficos B y C)

   a) En la asimetría positiva el sesgo es a la derecha (Gráfico B)
   b) En la asimetría negativa el sesgo es a la izquierda (Gráfico C)

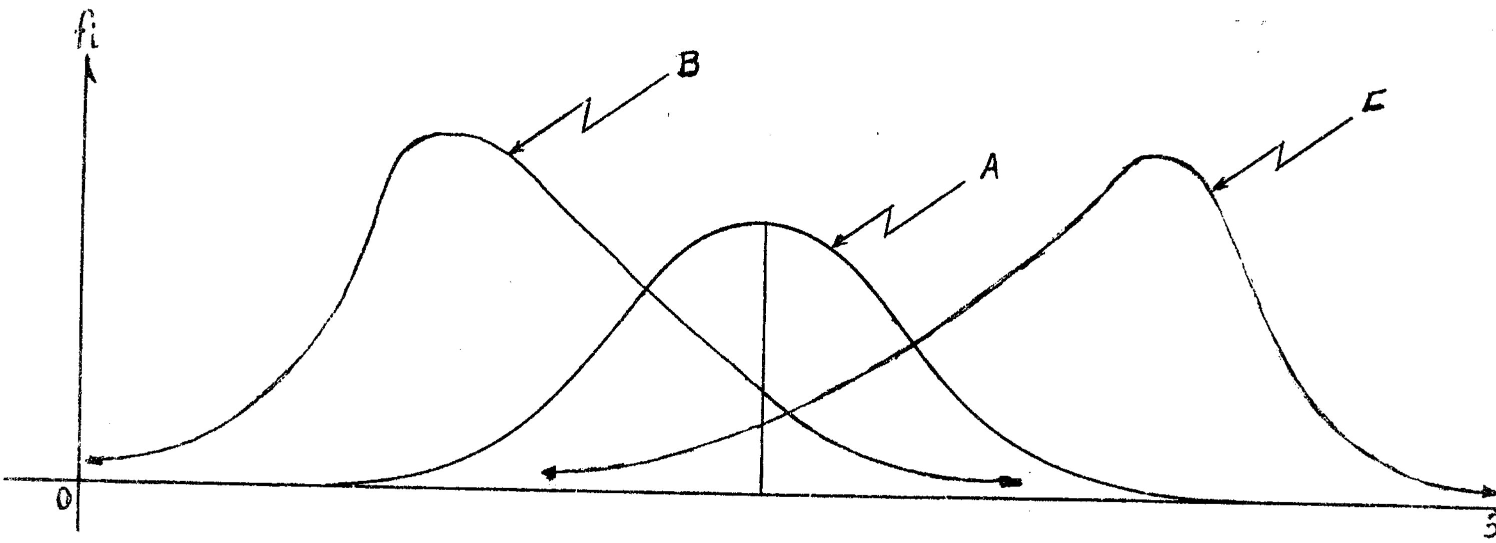


Fig. 2

1. Los polígonos simétricos se clasifican en:

   a) Platicúrtico (Gráfica A)
   b) Mesocúrtico (Gráfica B)
   c) Leptocúrtico (gráfica C)

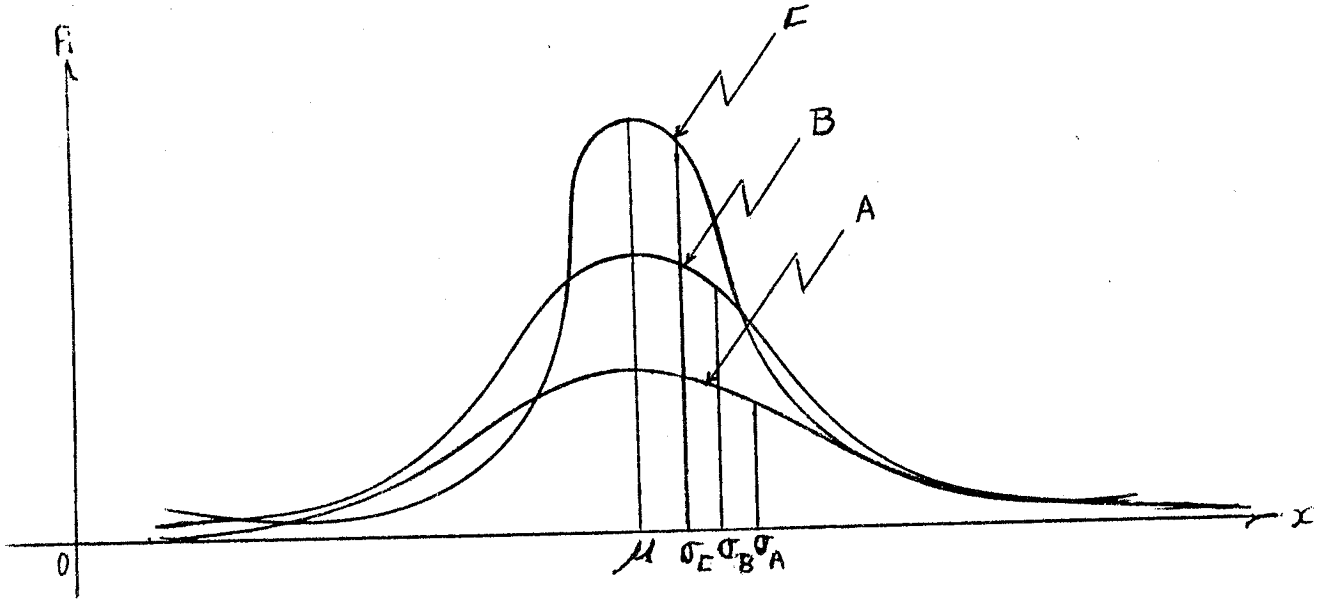


Fig. 3

De los gráficos anteriores podemos concluir que la forma de cada una, está íntimamente relacionada con las medidas de tendencia central y de dispersión.

En las simétricas, las medidas de tendencia central coinciden en el mismo punto, es decir μ= Mo= Md.

Las medidas de dispersión son diferentes, de la figura (3) obtenemos que:

$$\sigma_A = \sigma_B = \sigma_C$$

En las asimétricas las medidas de tendencia central son diferentes y lo mismo ocurre con las de dispersión.

a) Sesgo positivo μ < Md < Mo

b) Sesgo negativo μ > Md > Mo

$$\left\{ \sigma_A \neq \sigma_B \neq \sigma_C \right.$$

En cualquier problema de variable aleatoria contínua, su polígono de frecuencia es alguna de las gráficas anteriores y éstas dependen de sus parámetros de tendencia central y de dispersión.

La gráfica que tiene forma de campana, su media μ = 0 y σ= 1, se llama curva normal estandar o campana de Gauss por haber sido el primer científico que usó esta representación.

Las curvas simétricas tienen la forma de campana y las asimétricas no tienen esa forma pero pueden transformarse a simétricas.

El procedimiento para transformar las curvas asimétricas en simétricas, es mediante una normalización de los datos del problema y que etudiaremos a continuación.

NORMALIZACION

El proceso de transformación de un polígono de frecuencias a una curva normal, se llama normalización y para ello se hace un cambio de escala mediante la normalización o tipificación de las puntuaciones, es decir los valores (x) se transforman en valores Z mediante la ecuación de transformación.

$$Z = \frac{Xi - \mu}{\sigma} \quad \ldots (1)$$

Z = Puntuación normalizada o tipificada

X_i = Cada una de las puntuaciones de la población

μ = Media de las puntuaciones de la población

σ = Desviación estándar de la población

Veamos el siguiente ejemplo:

Se desea conocer el peso promedio de los alumnos del turno vespertino del plantel 2 del Colegio de Bachilleres . Para ello se toma una muestra representativa de 150 alumnos y se pesan. Los pesos ya organizados en 13 clases, se muestran en la siguiente tabla de frecuencias:

Clases en kg.	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
fi	1	5	8	8	10	18	12	36	28

75-79	80-84	85-89	90-94
12	8	3	1

Ejercicio: Con estos datos calcula

a) La media, y,
b) La desviación estándar.
c) Traza el polígono de frecuencias.

Los resultados que debiste obtener son:

N = 150
X = 63.9
S = 12.2

POLIGONO DE FRECUENCIAS

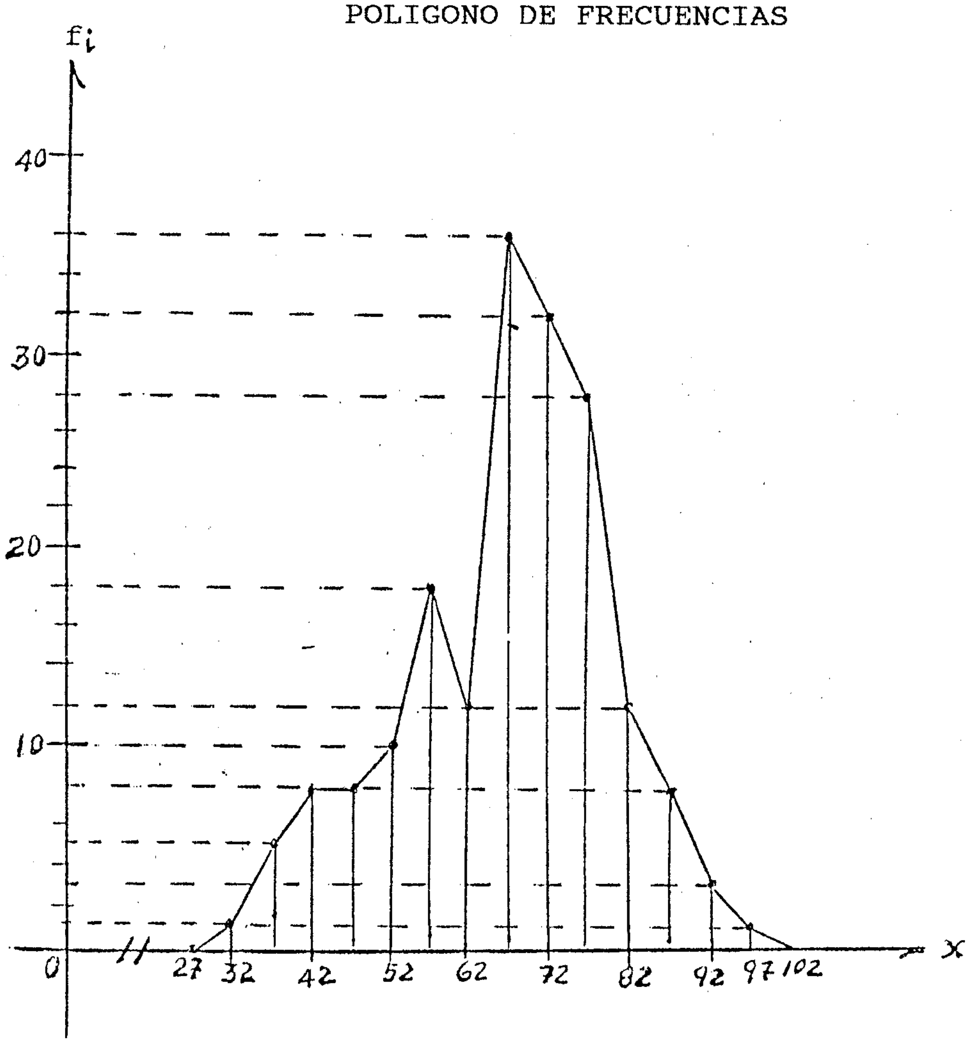


Fig. 4

De esta gráfica podemos concluir que es asimétrica con sesgo negativo y del tipo leptocúrtica.

Ahora vamos a normalizar estos datos y trazar la curva normal estándar sobre este polígono de frecuencias para poder constatar el cambio de escala.

Para explicar el procedimiento vamos a construir la siguiente tabla:

Normalización de una Distribución Asimétrica

1 CLASE	2 fi	3 Lr SUP	4 $\Delta x = x_i - x$	5 $Z=\Delta x/\sigma$	6 PARTE MAYOR	7 PARTE MENOR	8 fe	9 fe redondeada
90-94	1	94.5	30.6	2.51	0.9940	0.0119	1.785	1.8
85-89	3	89.5	25.6	2.10	0.9821	0.0276	4.140	4.1
80-84	8	84.5	20.6	1.69	0.9545	0.0548	8.220	8.2
75-79	12	79.5	15.6	1.28	0.8997	0.0919	13.875	13.9
70-74	28	74.5	10.6	0.87	0.8078	0.1306	19.590	19.6
65-69	36	69.5	5.6	0.46	0.6772	0.1573	23.595	23.6
60-64	12	64.5	0.6	0.05	0.5199	0.1605	24.075	24.1
55-59	18	59.5	-4.4	-0.36	0.3594	0.1388	20.820	20.8
50-54	10	5405	-9.4	-0.77	0.2206	0.1016	15.240	15.2
45-49	8	49.5	-14.4	-1.18	0.1190	0.0631	9.465	9.5
40-44	8	44.5	-19.4	-1.59	0.0599	0.8331	4.965	5.0
35-39	5	39.5	-24.4	-2.00	0.0228	0.0148	2.220	2.2
30-34	1	34.5	-29.4	-2.41	0.0080	0.0080	1.200	1.2

Las columnas 1 y 2 corresponden a la clase y la frecuencia establecidas en la primera tabla.

La columna 3 corresponde al límite real superior de cada clase el cual se determina aumentando medio punto a cada valor del límite superior.

La columna 4 es igual a la desviación de cada puntuación con respecto a la media y se obtiene mediante la ecuación:

$$\Delta x = x_i - \bar{x} \qquad \ldots (2)$$

Se toma como x_i al límite real superior de cada clase.

La columna 5 es el valor de Z correspondiente a cada puntuación y se obtiene mediante la ecuación de normalización o tipificación, esto es:

$$Z = \frac{\Delta x}{\sigma} = \frac{x_i - \bar{x}}{\sigma} \qquad . . . (1)$$

La columna 6 se determina de los valores de la tabla del apéndice (A) "Areas y ordenadas de la curva de distribución normal en función de $\frac{\Delta x}{\sigma}$ " que posteriormente ejemplificaremos.

En la primera columna de esta se localiza el valor de Z, en la tercera columna se lee el valor del área bajo la curva normal de la parte mayor.

En la cuarta columna se lee el área bajo la curva normal de la parte menor y se registra en la columna siete de nuestra tabla.

Ejemplo:

Para Z=2.51 el área de la parte mayor que se lee en la tercera columna es 0.9940.

Para Z=-2.41 en la cuarta columna se lee el área de la parte menor correspondiente a 0.0080

La columna ocho de nuestra tabla corresponde a la frecuencia esperada (fe) y se calcula multiplicando el total de casos N=150 por el área de la parte menor (columna 7) de cada puntuación.

Ejemplo:

$$(0.0119)\ (150) = 1.785$$

La columna nueve es la frecuencia esperada (fe) redondeada a una cifra decimal.

Con estos valores de la tabla trazamos el polígono de frecuencias y la curva normalizada para ver el cambio que sufre el polígono de frecuencias de cualquier problema que se normalizan los datos:

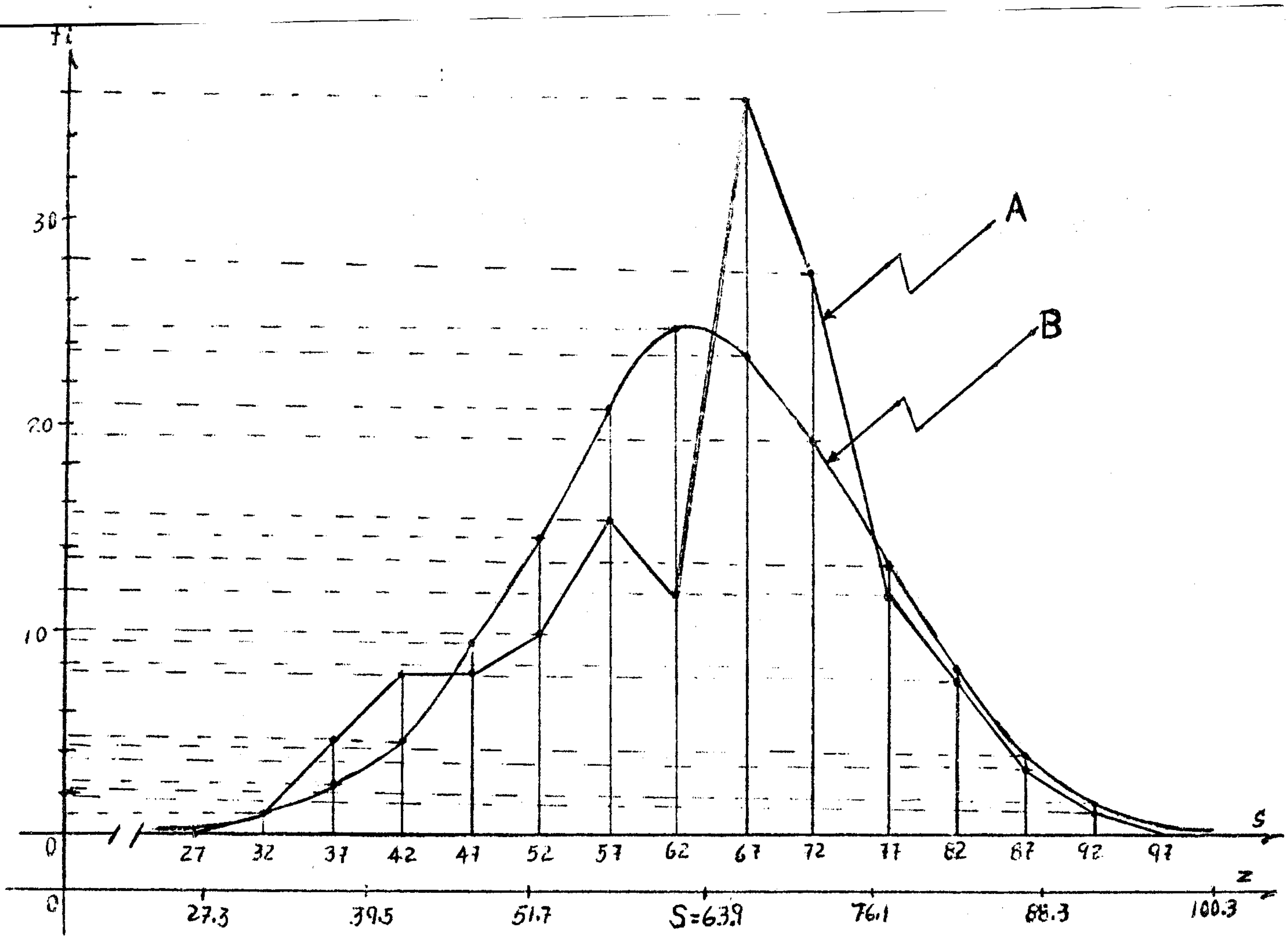


Fig. 5

A = Polígono de frecuencia de los pesos de 150 alumnos del plantel 2 del turno vespertino del Colegio de Bachilleres.

B = Curva normal del mismo problema

En la escala Z de la figura 5, se determinan los valores de la desviación típica (s), a uno y otro lado de la media ($\overline{X}$).

Del ejemplo anterior habrás notado que normalizar los datos de un problema es equivalente a cambiar la escala "x" por la "z" y calcular las nuevas frecuencias que son las ordenadas de cada punto. Para ello usamos los valores de la tabla. Estos valores corresponden a las áreas bajo la curva normal y se han calculado mediante la ecuación que define a la función normal y ésta es:

$$y = f(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-\mu)^2}{2\sigma}} \quad . . . (3)$$

Ɖ = 3.1416.... μ = media poblacional

e = 2.718281... σ = Desviación estandar de la población

x = cada uno de los datos u observaciones.

Con la ecuación (3) podemos trazar la curva normal que tiene la forma de campana. Primero obtenemos μ y σ de los datos del problema y sustituimos en la fórmula (3). Para obtener un par ordenado, usamos un valor arbitrario de x y obtenemos un valor de y. Esta sucesión de puntos nos da la curva normal.

La Curva Normal Estandar

La curva normal estándar o campana de Gauss es la misma curva normalizada solamente que mediante una traslación se lleva la curva hasta el origen. En este caso usamos $\mu = 0$ y $\sigma = 1$.

La siguiente gráfica muestra la curva normalizada con $\mu = 30$ y $\sigma = 10$

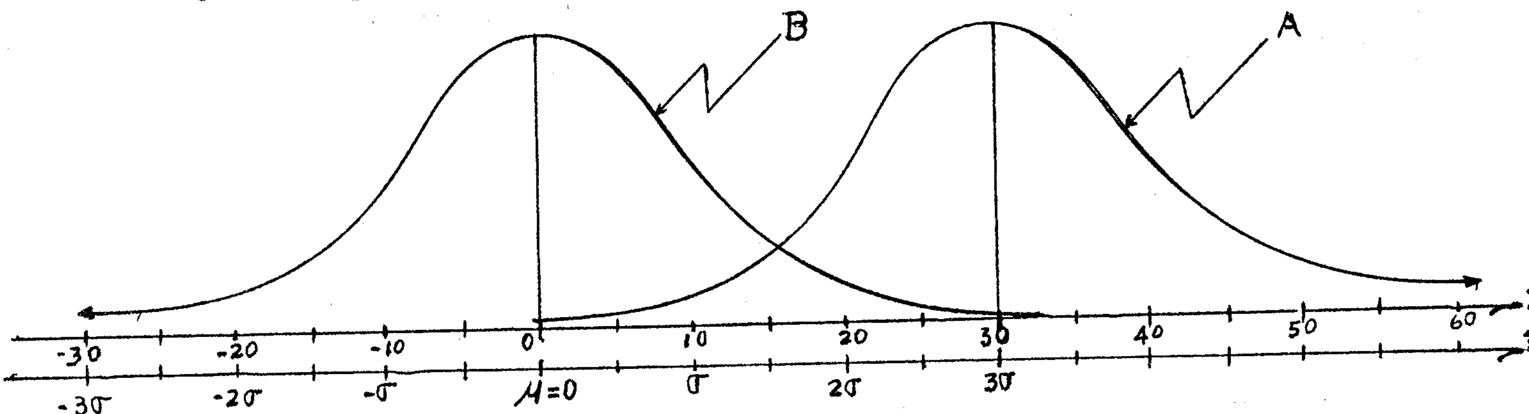


Fig. 6

A= Curva normalizada
B= Curva normalizada y estandarizada

Al trasladar la gráfica anterior al origen, hemos transformado los parámetros μ y σ en:

$$\mu = 0 \quad \text{y} \quad \sigma = 1$$

Con estos valores reducidos, la curva normal estándar se obtiene mediante la gráfica de la función:

$$y = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \ldots (4)$$

y la ecuación de tipificación es la ya conocida:

$$x = \mu + \sigma Z \quad \therefore \quad Z = \frac{x-\mu}{\sigma} \quad \ldots (1)$$

Ejercicio:

Con los datos del problema de las aguas negras, elabora la tabla con los datos normalizados y traza la curva normal sobre el polígono de frecuencias que ya obtuviste antes.

VALORES NORMALIZADOS "Z" Y AREA BAJO LA CURVA NORMAL

Ya quedó establecido que para normalizar el polígono de frecuencias y transformarlo en una campana de Gauss, se tipifican las observaciones (Xi) del problema cambiándolos a una escala (Z) mediante la ecuación (1).

Esta curva normal es necesaria estandarizarla para poder calcular la probabilidad mediante una misma tabla ya elaborada para toda curva normal estandarizada, que se obtiene trasladando la media al origen como ya se indicó.

La curva normal estandarizada tiene las siguientes características:

a) La altura alcanza su valor máximo con $\mu = 0$ y su valor es 0.4, es decir; el punto máximo es P_m (0,0.4)

b) La curva normal estándar es simétrica con respecto a la media por lo tanto los parámetros de tendencia central son iguales, es decir:

$$\mu = Mo = Md = 0 \qquad . . . \quad (5)$$

c) La desviación estándar es $\sigma = 1$
d) El área bajo la curva es $A = 1$

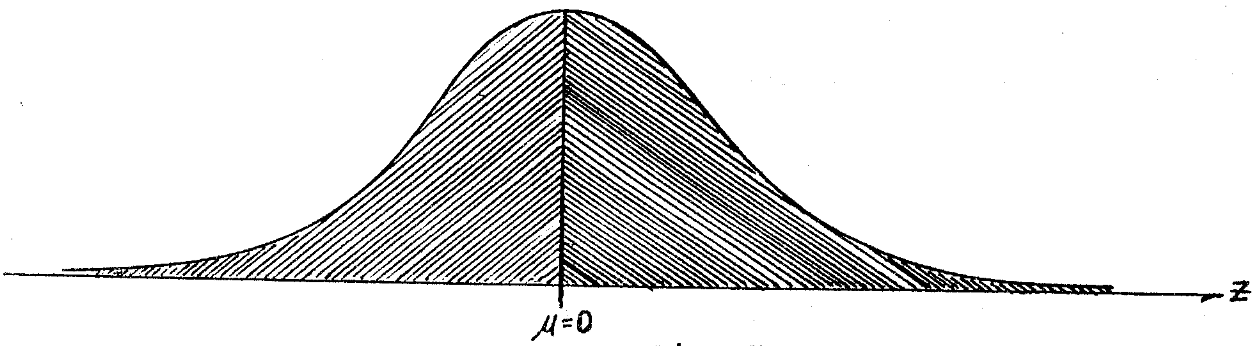


Fig. 7

El área sombreada vale 1 y como la curva es simétrica cada región a los lados del eje y vale 0.5

e) El eje Z es una asíntota horizontal de la curva ya que $\lim_{z \to \infty}(z)=0$

f) El área más importante donde se distribuye la probabilidad de un suceso, se encuentra comprendida entre $\pm 3\sigma$ y esto lo puedes constatar en la siguiente gráfica de la fig. 8.

g) De acuerdo con el teorema de Chebishev relacionado con la desviación estándar y el área bajo la curva, podemos establecer los siguientes porcentajes de la misma:

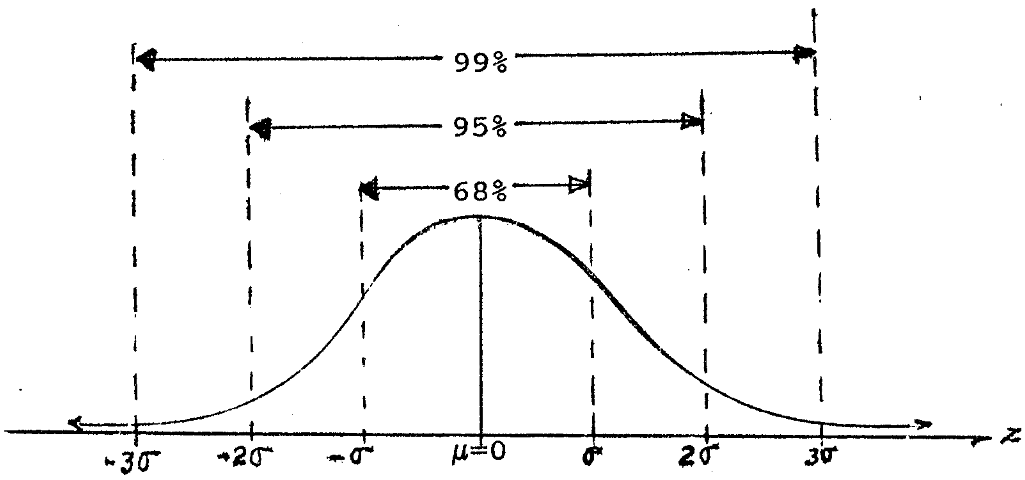


Fig. 8

De esta gráfica podemos ver que el área antes y después de $\pm 3\sigma$ corresponde al 1%, es decir el 0.5% para cada lado de la gráfica.

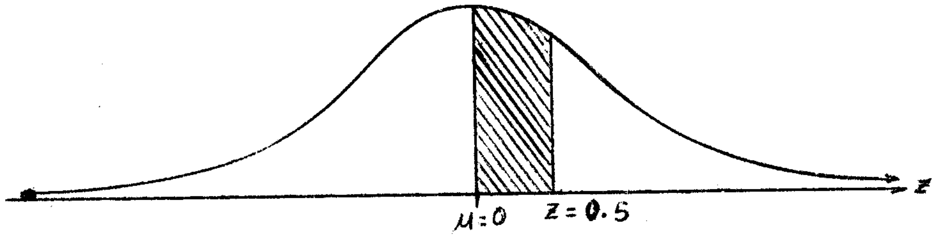
Por la simetría que tiene la curva normal estándar, existen tablas correspondientes al área bajo la curva que únicamente contemplan la parte positiva de la gráfica y estos mismos valores se usan para el lado negativo.

Ejemplo: Con los siguientes valores de "Z" determinaremos el valor del área bajo la curva y trazaremos un esquema del área correspondiente:

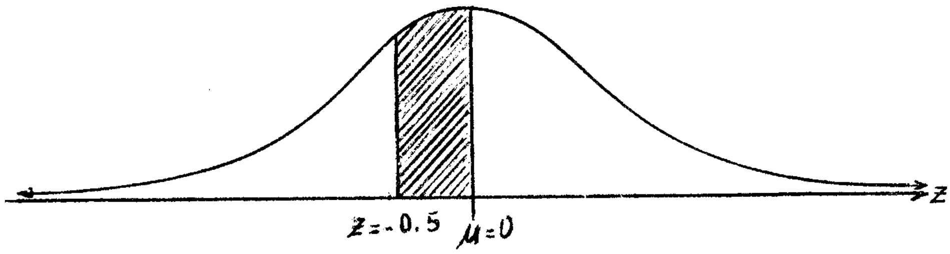
$$Z = \pm 0.5, \pm 0.7, \pm 1.5$$

En la primera columna de la tabla localizamos el valor de Z=0.5 y en la segunda columna leemos el valor del área.

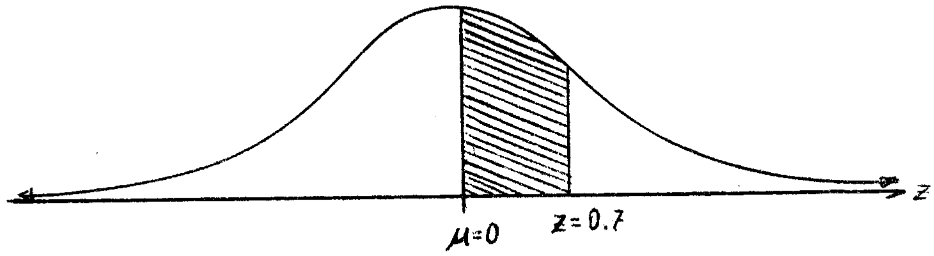
Z=0.5 ; A=0.1915



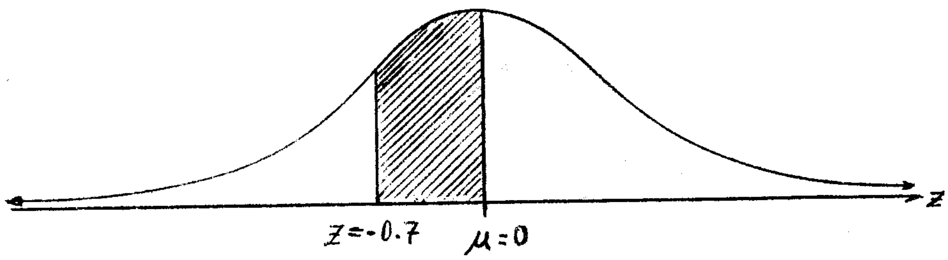
Z=-0.5 ; A=0.1915



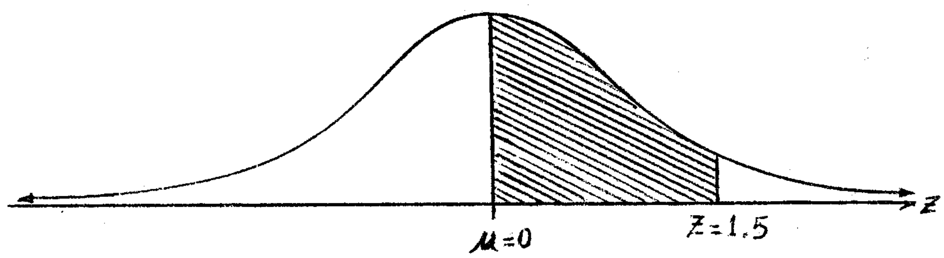
Z=0.7 ; A = 0.2580



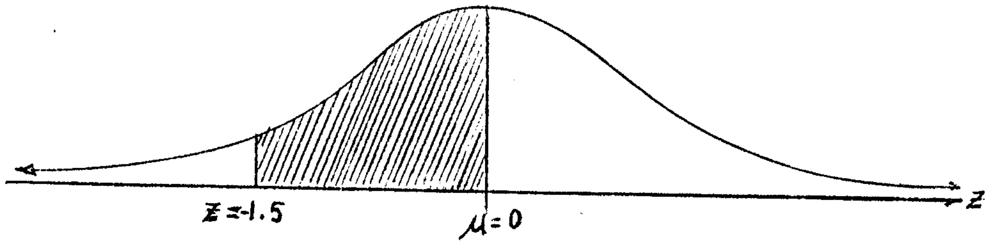
Z=-0.7 ; A=0.2580



Z=1.5 ; A=0.4322



$Z=-1.5$; $A=0.4322$



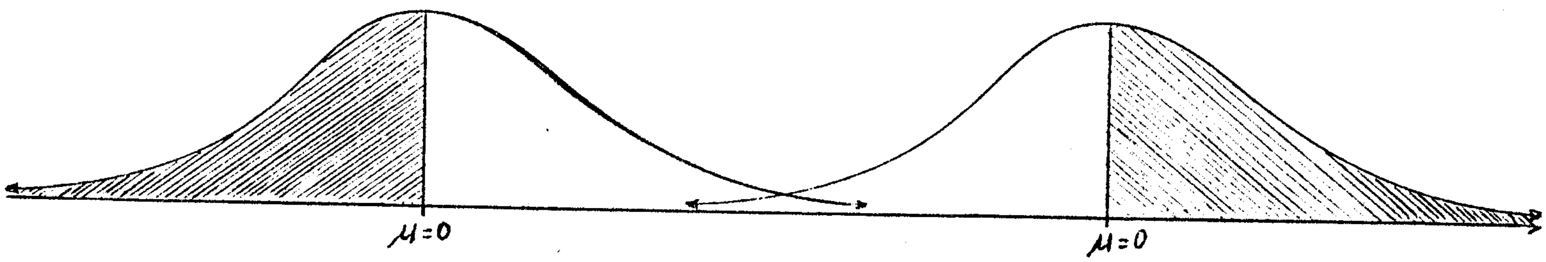
De las gráficas anteriores podemos ver, que el valor del área es el mismo para valores positivos y negativos de Z solamente que para el valor negativo, el área se representa a la izquierda de la media.

Ejercicio:

Normaliza los valores de X=4,6,9,12,18,20, usando μ =10 y σ=5. Traza una gráfica para cada valor de x, compárala con la gráfica de los valores normalizados y traza una para cada "Z" sombreando en ambas gráficas la región correspondiente.

El área bajo la curva normal estándar representa la probabilidad de un evento; toda el área bajo la curva vale uno y representa la probabilidad del evento seguro.

El área de cada mitad de la gráfica, es 0.5



$P(Z)=0.5$ $P(Z)=0.5$

Si queremos la probabilidad de un evento cuyo valor está limitado por dos puntuaciones, por ejemplo:

$$P(x_1 \leq X \leq x_2)$$

Probabilidad de "X" comprendida entre X_1 y X_2.

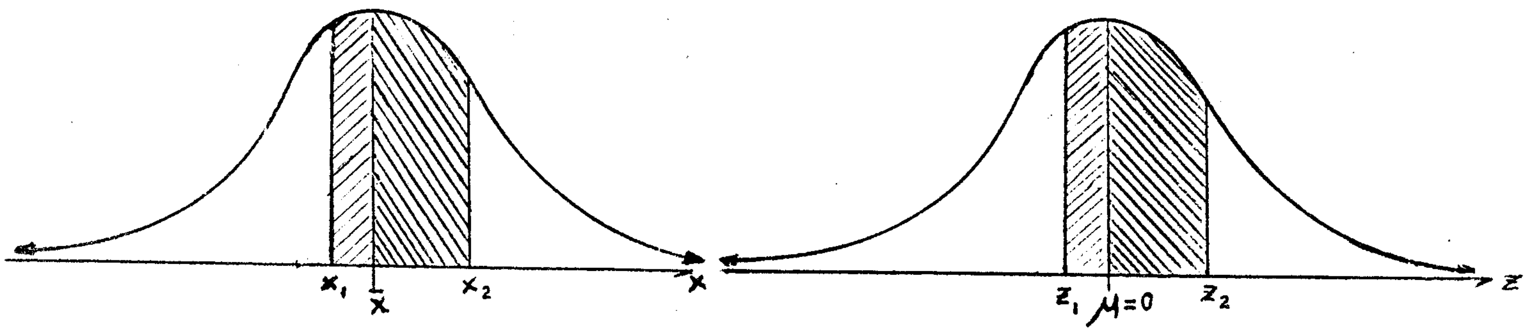
Para determinar esta probabilidad, tipificamos los valores "X".

Sabemos que la curva normalizada de la escala "X" es equivalente a la curva normal estándar en la escala "Z"

$$\therefore\ P(x_1 \leq x \leq x_2) = P(Z_1 \leq Z \leq Z_2)$$

Determinamos $\quad Z_1 = \dfrac{X_1 - \mu}{\sigma} \Rightarrow \quad P(Z)=P(Z_1)+P(Z_2)$

$$Z_2 = \frac{X_2 - \mu}{\sigma}$$

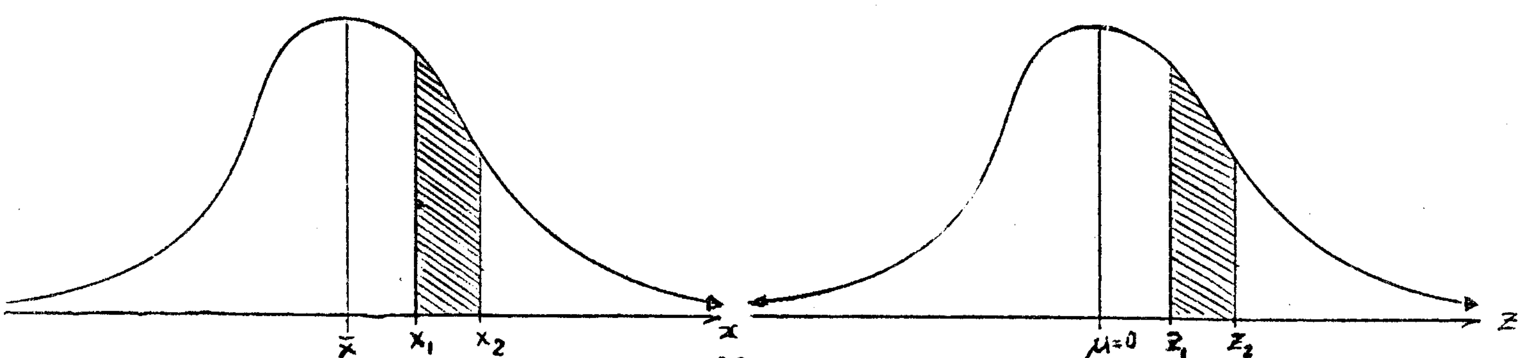


Las gráficas correspondientes en ambas escalas son:

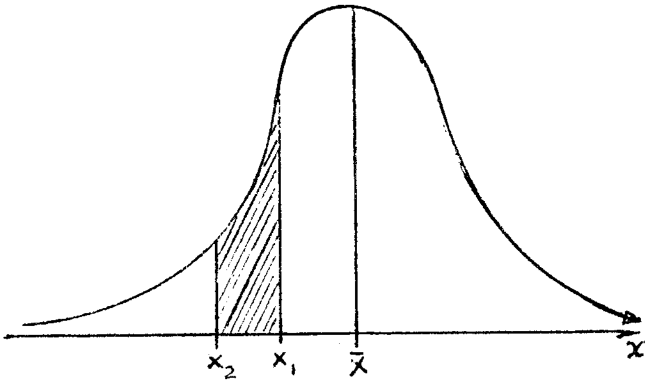
Si las "X" están en el lado positivo entonces, debemos recordar que los valores que se leen en la tabla normalizada son a partir de la media hasta el valor de Z.

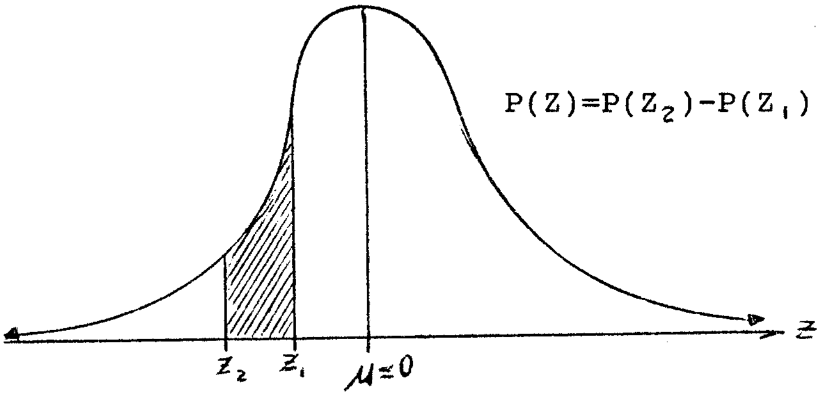
Las gráficas de las variables x y z son las que se muestran a continuación:

$$P(Z)=P(Z_2)-P(Z_1)$$



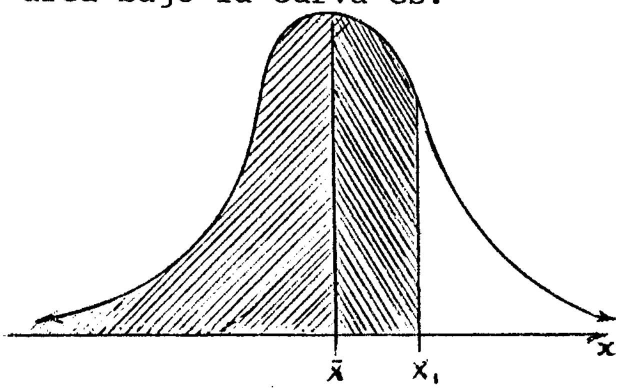
Si los valores de "X" están en la parte negativa de "Z" es decir a la izquierda de la media, entonces los gráficos son:

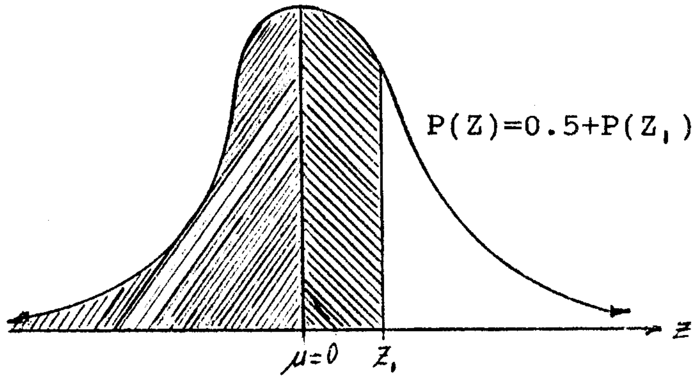




$P(Z)=P(Z_2)-P(Z_1)$

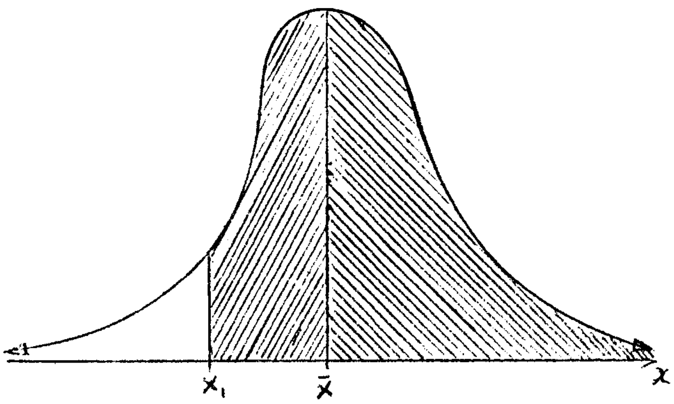
Si solamente tenemos una "X" a la derecha de la media entonces el área bajo la curva es:

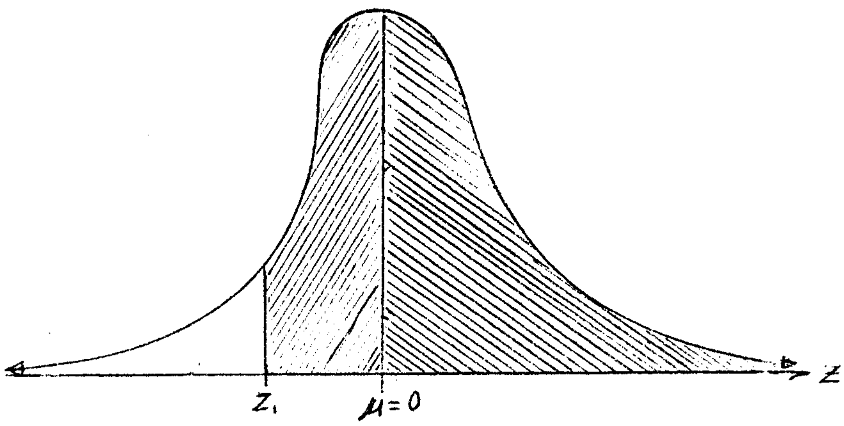




$P(Z)=0.5+P(Z_1)$

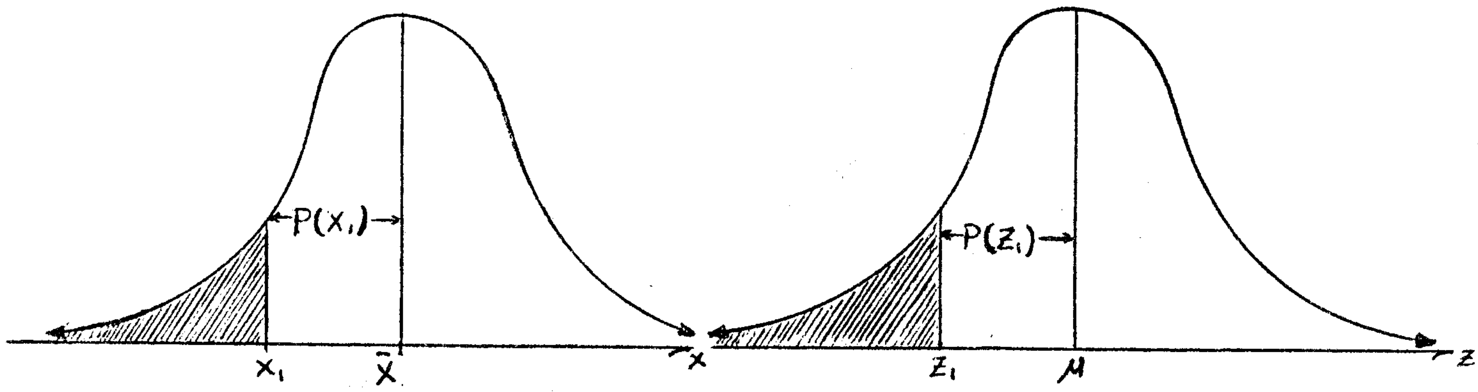
Recuerda que la primera mitad del área bajo la curva vale 0.5, es por eso que a la probabilidad de Z, le sumamos 0.5.





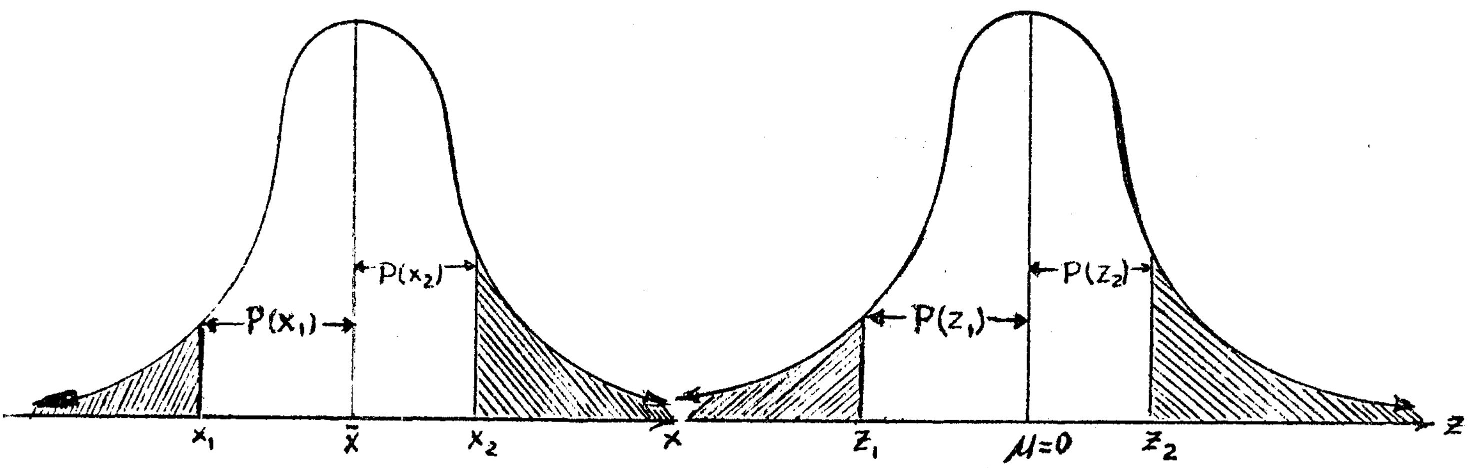
Si nos interesa la probabilidad de $x<\mu$, entonces la gráfica es:

$$P(Z)=0.5-P(Z_1)$$



Si queremos la probabilidad de las partes sombreadas de la siguientes gráficas:

$$P(Z)=[0.5-P(Z_1)]+[0.5-P(Z_2)]$$



Recuerda que la probabilidad de "Z", es la parte sin sombrear de la media a la izquierda hasta "Z_1", y la probabilidad de Z_2, es la parte sin sombrear de la media a la derecha hasta Z_2.

Ejercicio: Con los siguientes valores determina la probabilidad y traza la gráfica correspondiente para cada inciso.

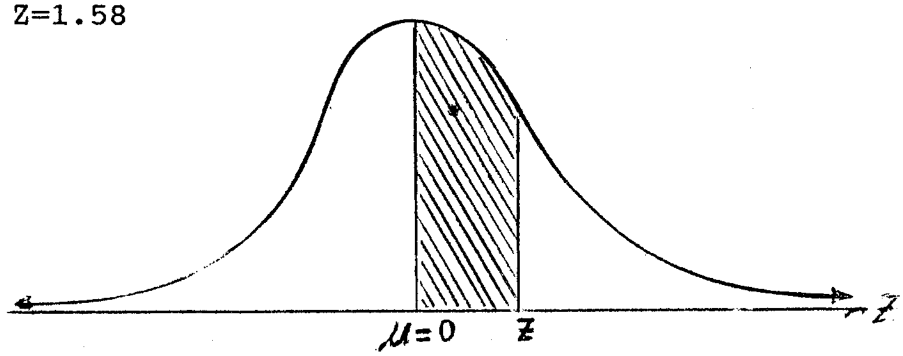
1. a) Z =0 y Z = 0.94
   b) Z =0 y Z =-2.15
   c) A la derecha de Z= 0.62
   d) A la derecha de Z=-0.93
   e) A la izquierda de Z= 0.84
   f) A la izquierda de Z=-0.35

2.-a) Z =-0.59 y Z = 0.59
   b) Z =-0.71 y Z = 1.99
   c) Z = 0.32 y Z = 0.92
   d) Z =-0.81 y Z =-0.42
   e) Z =-1.65 y Z =-0.25

Si se conoce la probabilidad de un evento y queremos determinar el valor de Z, entonces nos situamos en la segunda columna de la tabla (Area desde la media), localizamos el valor de la probabilidad y en el mismo renglón y en la columna 1 determinamos el valor de Z.

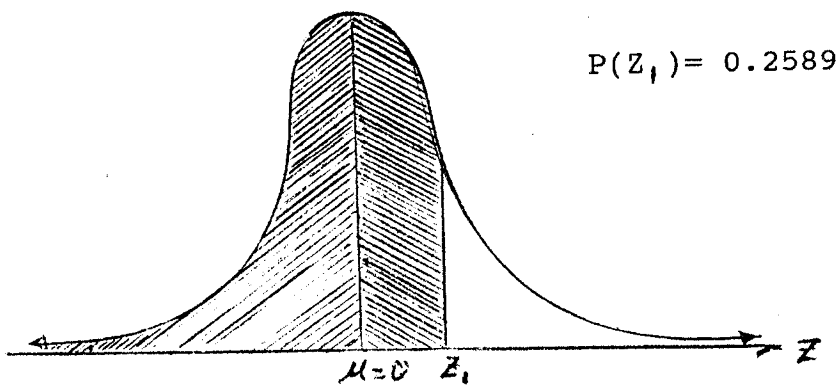
Ejemplo: Si P(Z)=0.4429 entonces¿el valor de Z es?

De la tabla obtenemos que Z=1.58



Determina Z si P(Z)=.7580. Este valor es mayor que 0.5 correspondiente a la mitad de la gráfica por lo tanto hacemos la siguiente transformación:

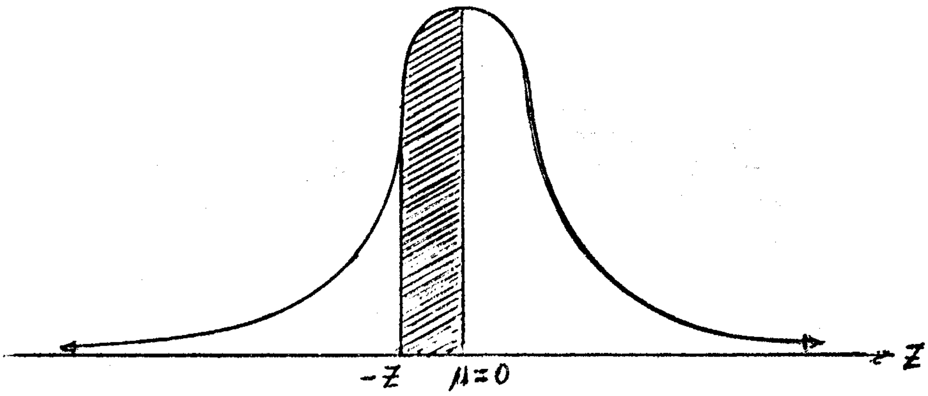
$P(Z)=0.5+P(Z_1)$ $\therefore$ $P(Z_1)=P(Z)-0.5=0.7589-0.5=0.2589$



En la tabla nos situamos en este valor y en la columna 1 está el valor de Z =0.70

Si la probabilidad de Z a la izquierda de la media es P(Z)=0.1331, entonces¿Z es?

En la segunda columna de la tabla localizamos el valor de la probabilidad y en la misma línea en la primera columna determinamos el valor de Z =-0.34. El valor del signo es por estar a la izquierda de la media.



Ejercicio.

1. Determina el valor de Z y traza la gráfica de cada inciso, si la probabilidad de z es:

   a) Entre 0 y Z, P(Z)=0.4864
   b) A la izquierda de Z, P(Z)=0.9983
   c) A la derecha de Z, P(Z)=0.7324
   d) A la derecha de Z, P(Z)=0.2981
   e) A la izquierda de Z, P(Z)=0.1314
   f) Entre -Z y Z, P(Z)=0.7286

2. Una variable aleatoria tiene una distribución normal con media μ=60 y desviación estandard σ =5.2 ¿Cuales son las probabilidades de que la variable aleatoria tome un valor como el que se indica? Traza la gráfica de cada inciso.

   a) Menor que 62.5
   b) Mayor que 70.5
   c) Entre 60.0 y 66.2
   d) Entre 48 y 72

Si tienes alguna duda consulta a tu profesor o a tu consultor.

Ejemplo:

Con los siguientes valores, calcula la probabilidad de la distribución binomial y traza el polígono de frecuencia de cada una.

1) $n = 10,\ p = 0.2,\ q = 0.8,\ x = 0,1,2,...10$

2) $n = 10,\ p = 0.8,\ q = 0.2,\ x = 0,1,2,...10$

3) $n = 10,\ p = 0.5,\ q = 0.5,\ x = 0,1,2,...10$

Para el problema 1 sustituimos valores en (6) y obtenemos:

$$f(0)=\binom{10}{0}(0.2)^{0}(0.8)^{10}=(1)(1)(0.1073)=0.1073$$

$$f(1)=\binom{10}{1}(0.2)^{1}(0.8)^{9}=(10)(0.2)(0.1342)=0.2684$$

$$f(2)=\binom{10}{2}(0.2)^{2}(0.8)^{8}=(45)(0.04)(0.1677)=0.3019$$

$$f(3)=\binom{10}{3}(0.2)^{3}(0.8)^{7}=(120)(0.0008)(0.2097)=0.2013$$

$$f(4)=\binom{10}{4}(0.2)^{4}(0.8)^{6}=(210)(0.0016)(0.2621)=0.0881$$

$$f(5)=\binom{10}{5}(0.2)^{5}(0.8)^{5}=(252)(0.00032)(0.3276)=0.0264$$

Calcula los siguientes valores:

$f(6)=$

$f(7)=$

$f(8)=$

$f(9)=$

$f(10)=$

2. Cálculo del segundo problema:

$$f(0)=\binom{10}{0}(0.8)^{0}(0.2)^{10}=(1)(1)(0.000000102)=0.0000001$$

$$f(1)=\binom{10}{1}(0.8)^{1}(0.2)^{9}=(10)(0.8)(0.000000512)=0.0000009$$

$$f(2)=\binom{10}{2}(0.8)^{2}(0.2)^{8}=(45)(0.64)(0.000002)=0.0000737$$

$$f(3)=\binom{10}{3}(0.8)^{3}(0.2)^{7}=(120)(0.512)(0.000012)=0.0008$$

$$f(4)=\binom{10}{4}(0.8)^{4}(0.2)^{6}=(210)(0.4096)(0.000064)=0.0055$$

$$f(5)=\binom{10}{5}(0.8)^{5}(0.2)^{5}=(252)(0.3276)(0.00032)=0.02642$$

Calcula los siguientes valores:

$f(6)=$

$f(7)=$

$f(8)=$

$f(9)=$

$f(10)=$

3. Cálculo para el tercer problema:

$$f(0)=\binom{10}{0}(0.5)^{0}(0.5)^{10}=(1)(1)(0.00097)=0.0009$$

$$f(1)=\binom{10}{1}(0.5)^{1}(0.5)^{9}=(10)(0.5)(0.00195)=0.009$$

$$f(2)=\binom{10}{2}(0.5)^{2}(0.5)^{8}=(45)(0.25)(0.0039)=0.0439$$

$$f(3)=\binom{10}{3}(0.5)^{3}(0.5)^{7}=(120)(0.125)(0.00781)=0.1172$$

$$f(4)=\binom{10}{4}(0.5)^{4}(0.5)^{6}=(210)(0.0625)(0.0156)=0.2051$$

$$f(5)=\binom{10}{5}(0.5)^{5}(0.5)^{5}=(252)(0.03125)(0.03125)=0.2461$$

f(6)=

f(7)=

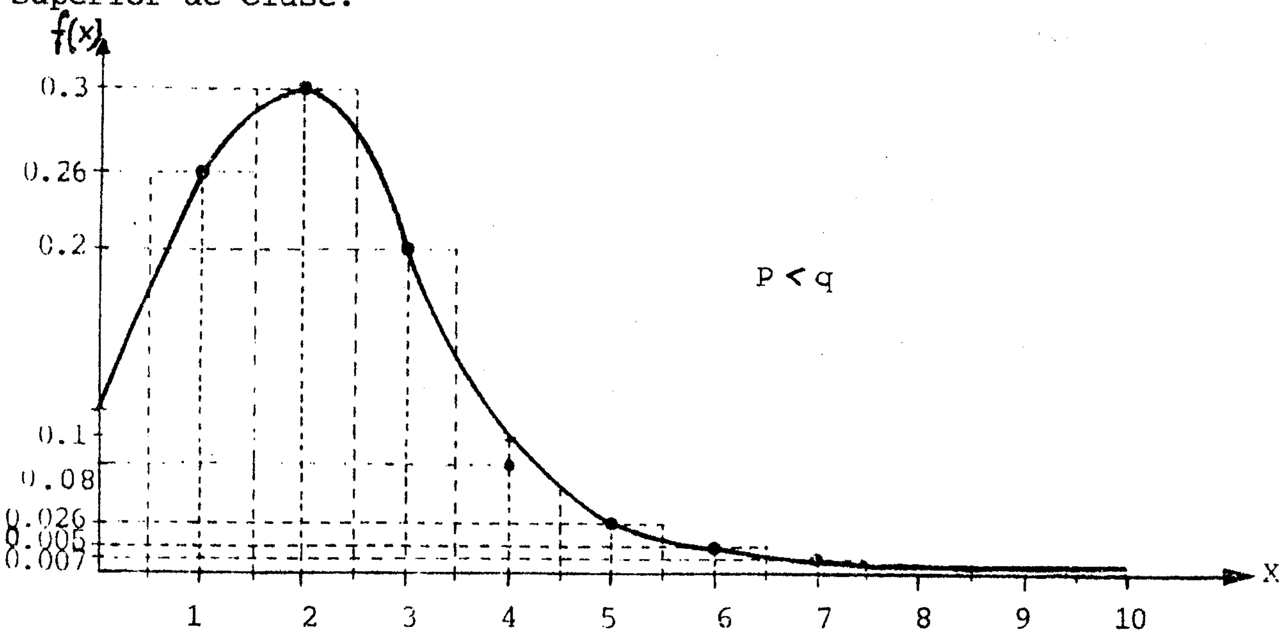
f(8)=

f(9)=

f(10)=

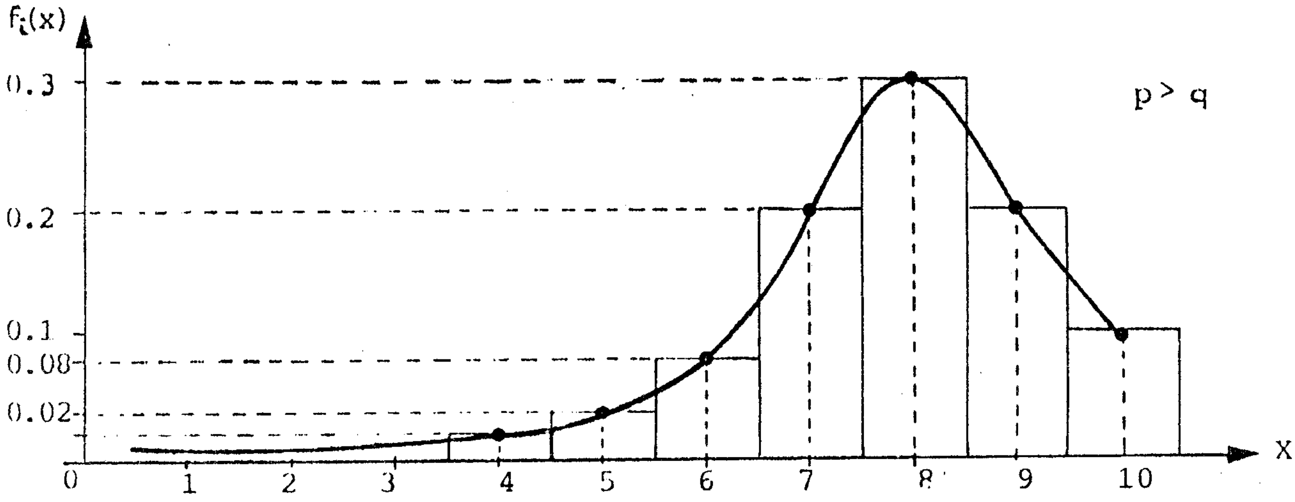
Representación gráfica de las probabilidades de cada uno de los problemas:

Para poder trazar la gráfica como si fuese una variable contínua, cerramos los espacios entre cada barra del histograma, para ello tomemos medio punto después de cada valor para obtener el límite real superior de clase.



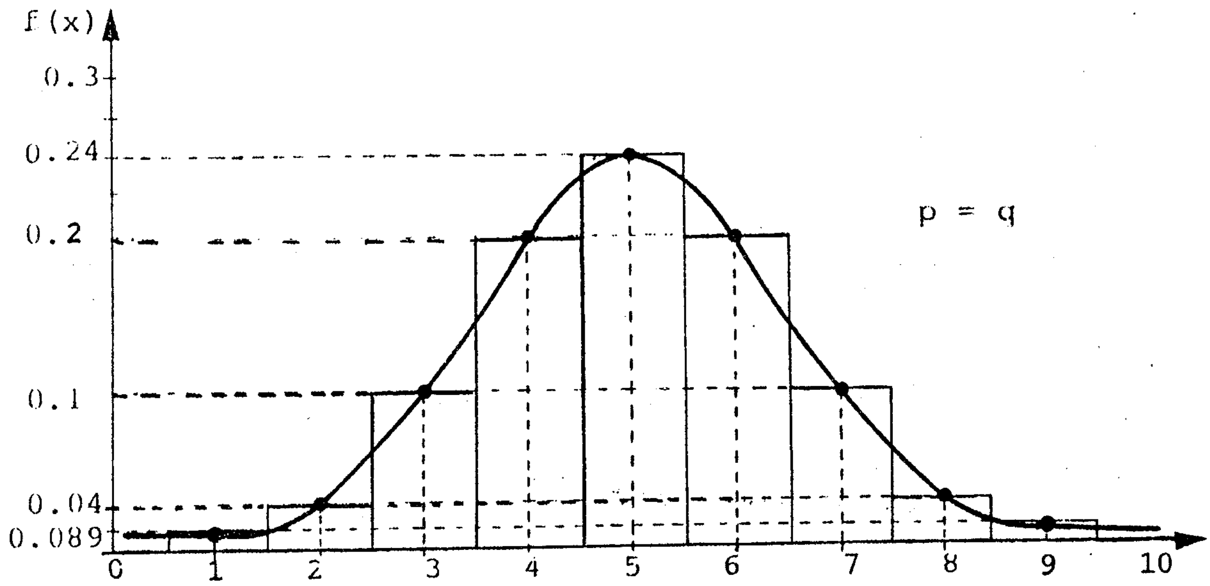
Gráfica del problema 1

La gráfica 1 es antisimétrica y sesgada a la derecha.



Gráfica del problema 2

La gráfica 2 es antisimétrica y sesgada a la izquierda.

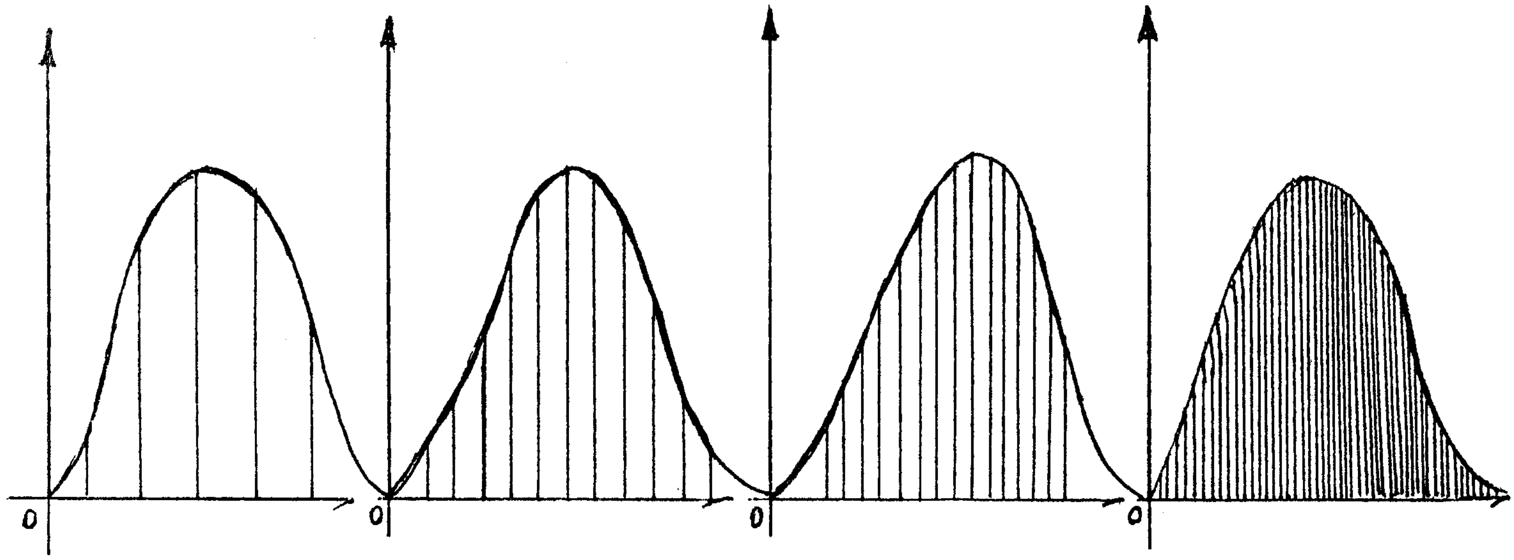


Gráfica del problema 3

La gráfica 3 es simétrica muy parecida a la campana de Gauss.

En el fascículo anterior estudiastes el cálculo de probabilidades de variables discretas cuya distribución es binomial.

Veamos la representación gráfica de una variable de distribución binomial cuando n (número de elementos de la población) aumenta.



n=5 n=11 n=15 n= 50

En las gráficas anteriores podemos ver que si "n" aumenta, los espacios entre las barras se van cerrando y la gráfica se aproxima a la campana de Gauss que es la gráfica de una variable aleatoria contínua.

Veamos el cálculo de los siguientes problemas correspondientes a una distribución binomial definida por la ecuación:

$$f(x)=\binom{n}{x}P^{x}(1-P)^{n-x} \qquad \ldots (6)$$

n = Número de observaciones de la variable aleatoria
x = Número de éxitos esperados
p = Probabilidad de éxito
q = 1-p = Probabilidad de fracaso

Con estos ejemplos podrás notar que el cálculo en la distribución binomial, es muy laborioso, aunque existen tablas para algunos valores; pero no son suficientes cuando "n" crece.

Por ejemplo si en un problema de distribución binomial se han realizado 100 observaciones y se desea saber la probabilidad de obtener al menos 45 éxitos.

Para determinar esta probabilidad tenemos que calcular

$$f(45)+f(46)+...f(100)=P(x) \qquad ...(7)$$

Otra forma de calcular esta probabilidad es restándole a la unidad las probabilidades de la siguiente forma:

$$P(x)=1-[f(0)+f(1)+f(2)+...+f(44)] \qquad ...(8)$$

Una forma de ahorrar este trabajo laborioso es haciendo el cálculo de probabilidades por medio de la distribución normal.

Ya vimos en las gráficas anteriores cómo el polígono de frecuencias de un problema de distribución binomial se aproxima a la campana de Gauss, por lo tanto podemos usar la distribución normal para calcular una probabilidad binomial con una aproximación aceptable.

Se recomienda usar la distribución normal cuando "n" es grande y P se aproxima al valor de 0.5. Se considera que n es grande si n>30

Para usar la distribución normal se calculan los parámetros aplicando las siguientes ecuaciones:

$$\mu = np \qquad ...(9)$$

$$\sigma = \sqrt{np(1-p)} \qquad ...(10)$$

Veamos el siguiente ejemplo:

Determinar la probabilidad de obtener 6 águilas en 15 lanzamientos de una moneda equilibrada y comparar el resultado mediante la distribución normal.

Solución:

$$f(x)=f(6)= \binom{15}{6}\left(\frac{1}{2}\right)^{6}\left(1-\frac{1}{2}\right)^{15-6} = \binom{15}{6}\left(\frac{1}{2}\right)^{15} =5005(0.0000305)=0.1527$$

$f(x)=0.1527$

Solución usando la distribución normal.

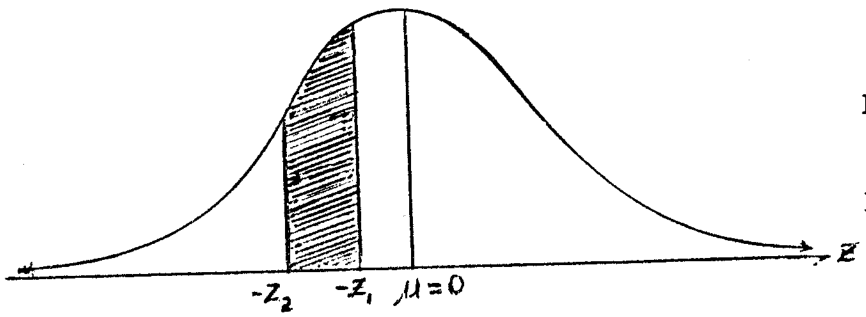
Para aplicar esta distribución corregimos los espacios para considerar a la variable como si fuese contínua o sea para 6 águilas tomamos medio punto antes y medio punto después, es decir:

$x=5.5$ $\qquad\qquad \mu=np=15\left(\frac{1}{2}\right)=7.5$

$x=6.5$ $\qquad\qquad \sigma=\sqrt{np(1-p)}=\sqrt{(15)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}=\frac{1}{2}\sqrt{15}=1.9365$

$$Z_1=\frac{5.5-7.5}{1.9365}=\frac{-2}{1.9365}=-1.033$$

$$Z_2=\frac{6.5-7.5}{1.9365}=\frac{-1}{1.9365}=-0.5164$$



$P(Z)=P(Z_2)-P(Z_1)=$
$=P(-1.033)-P(-0.5164$
$=0.3485-0.1950=0.1535$
$P(Z)-f(x)=0.1535-0.1527$
$=\underline{0.0008}$

De estos cálculos concluimos que la diferencia de la probabilidad normal y binomial es de 8 diezmilésimos. Este ejemplo nos ilustra que podemos usar la distribución normal para calcular la probabilidad de una distribución binomial, con una aproximación tal que no afecta la toma de decisiones.

Veamos otro ejemplo:

Una editorial de libros técnicos obsequia un porcentaje de libros para dar a conocer una nueva edición. Con el libro de obsequio se envia un cuestionario que deben contestar los lectores y devolver a la editorial. En el cuestionario se incluyen preguntas con respecto al contenido del libro. A la editorial le interesa conocer la opinión de las personas para mejorar su contenido y preparar nuevos tirajes; pero la experiencia de ésta es que la probabilidad de que devuelvan el cuestionario es de $P(x)=0.18$

Se envían por correo 100 ejemplares a profesionistas que pudiesen interesarles el contenido, del libro "Aplicación industrial de las probabilidad". A la editorial le interesa saber la probabilidad que al menos reciban 15 cuestionarios de regreso.

Solución: El problema es de recibir o no el cuestionario, por lo tanto es una distribución binomial con n=100 y P=0.18 por lo que para hallar el resultado debemos calcular:

$$P(x)=f(15)+f(16)+...+f(100) \qquad ... \quad (11)$$

o bien

$$P(x)=1-[f(0)+f(1)+...+f(14) \qquad ... \quad (12)$$

El segundo cálculo es menos laborioso, sin embargo no deja de serlo. Sabemos que una buena aproximación es mediante la distribución normal cuyo cálculo es más sencillo.

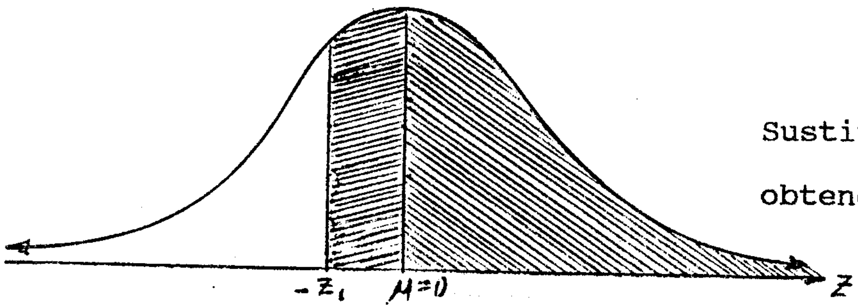
Veamos el desarrollo:

$$\mu=np=100(.18)=18$$

$$\sigma=\sqrt{np(1-p)} = \sqrt{100(.18)(.82)} = 3.84$$

Para transformar la variable binomial a contínua tomamos el límite real inferior de clase X =14.5 o sea medio punto antes. Con este valor calculamos $Z_1 = \frac{14.5-18}{3.84} = -0.9114$

$$P(Z)=P(Z_1)+0.5 \qquad ... \quad (13)$$



De las tablas obtenemos:

$P(Z_1)=P(-0.9114)=0.3186$

Sustituyendo este valor en (13)

obtenemos: $f(x)=P(Z)=0.3186+0.5$

$f(x)=0.8186$

De acuerdo con este resultado la editorial recibirá el 82% de los cuestionarios enviados.

Ejercicio:

1. Aplica la distribución binomial y determina la probabilidad de recibir al menos 15 cuestionarios. Compara los resultados e indica el error de aproximación, si es positivo o negativo.

Ejercicio:

1. Realiza los siguientes problemas aplicando la distribución binomial y compara el resultado con el resultado usando la distribución normal.

2. La policía tiene conocimiento que la probabilidad del robo de automóviles en la ciudad de México es de $P(x)=0.4$ y tiene reportados 10 automóviles robados en el mes de diciembre. Calcular la probabilidad de recuperar:

   a) A lo más 3 de los 10 robados
   b) Al menos 6 de los 10 robados

3. En el plantel 11 del Colegio de Bachilleres se tiene el conocimiento de que la probabilidad de mujeres en el primer ingreso es de 0.45. Si seleccionamos una muestra al azar de 10 alumnos de primer ingreso, ¿Cuál es la probabilidad de que:

   a) Seis de los 10 sean mujeres
   b) A lo más 6 de los 10 sean mujeres
   c) Al menos 5 de los 10 sean mujeres

Distribuciones Muestrales y Teorema Central del Limite

En el fascículo (1) del curso de estadística descriptiva se definieron los conceptos de:

1) Población
   a) finita e
   b) infinita
2) Muestra aleatoria
3) Estadística
4) Parámetros

También se estableció porqué es conveniente estudiar una muestra aleatoria en lugar de la población.

Se recomienda que repases estos conceptos que usaremos en esta unidad.

DISTRIBUCIONES MUESTRALES

Cuando vamos al mercado nos dan una prueba de barbacoa y del sabor de esta muestra se infiere el sabor de toda y si nos gusta entonces la compramos. Lo mismo ocurre si queremos comprar queso, pedimos una prueba y de esta deducimos si todo el queso está bueno o no.

Si el industrial quiere determinar el número de horas de vida que tiene un foco, toma una muestra de todo el lote y los mantiene encendidos hasta que se funden.

De estos casos podemos deducir que no es posible analizar todo el queso o la barbacoa porque no quedaría para vender. El industrial no puede fundir todos los focos porque no tendría qué vender.

En toda investigación estadística el objetivo general de ésta, es hacer generalizaciones de inferencias válidas obtenidas de la muestra. En otras palabras, se trata de conocer las características de la población a partir de los datos de una o más muestras obtenidas de la población.

Las muestras pueden ser:

a) No probabilísticas y
b) Probabilísticas

a) Las muestras no probabilísticas no nos permiten hacer generalizaciones.

b) Las muestras probabilísticas son la base de la inferencia estadística y a este tipo corresponde el muestreo aleatorio.

DEFINICION:

Se llama muestreo aleatorio de una población finita de n elementos, si cada muestra tiene la misma probabilidad de ser seleccionada y cada elemento de la población tiene la misma probabilidad de ser incluido en la muestra.

Los tipos de muestras aleatorias son:

1. Muestreo sistemático

2. Muestreo estratificado

3. Muestreo por conglomerados

4. Muestreo aleatorio simple

En lo que sigue nos ocuparemos de cada uno de ellos.

Muestreo Sistemático

En este muestreo los elementos de la población se seleccionan con un intervalo uniforme que se mide en el tiempo, en el espacio o en el orden.

Ejemplo:

Se desea entrevistar a cada décimo estudiante del S.E.A. del Plantel 2 del Colegio de Bachilleres, para ello se toma una lista de todos los estudiantes. Supongamos que escogimos el 5o., entonces el siguiente será de los 10 primeros seleccionados al azar y a partir de este vamos tomando los números décimos de toda la lista.

Este muestreo tiene ventajas y desventajas.

a) Ventajas:

1. Cada elemento de la población tiene la misma probabilidad de ser seleccionado.

2. El muestreo requiere de poco tiempo.

3. El costo es reducido.

b) Desventajas:

1. No todas las muestras tienen la misma probabilidad de ser seleccionadas.

2. Debido a lo anterior se puede cometer el grave error de tomar una muestra que no sea representativa, por ejemplo:

   Se muestrea un determinado número de familias para saber si el miércoles está incluido un platillo de carne de res en su alimentación. La respuesta es negativa porqué solamente el domingo la consume ya que es el día en que van al pueblo a comprarla. Esta forma de tomar la muestra no es representativa.

Muestreo Estratificado

Para este muestreo, dividimos la población en grupos homogéneos llamados estratos. Determinamos la proporción correspondiente de cada estrato en base a la población y esta misma proporción se toma de cada estrato para formar la muestra.

Este método es útil cuando la población ya esta dividida en grupos.

Por ejemplo:

Los estudiantes del S.E.A. del plantel 2 del Colegio de Bachilleres están divididos por edades con intervalos de 5 años y los porcentajes son los siguientes:

de 18 a 23	30%
de 24 a 29	25%
de 30 a 35	20%
de 36 a 41	10%
de 42 a 47	7%
de 48 a 53	5%
de 54 y mas	3%
	100

Se desea saber cuantas horas estudian diariamente; para ello de cada grupo se toma un porcentaje igual al del grupo, es decir del primer grupo tomamos el 30% del grupo. De la misma forma se toma el porcentaje de los siguientes grupos para formar la muestra representativa para su estudio.

Muestreo por Conglomerados

Para este tipo de muestreo, dividimos a la población en grupos conglomerados y de estos seleccionamos una muestra aleatoria, para su estudio.

Por ejemplo:

En una investigación de mercados se desea saber el número de coches por familia de la ciudad de México. Para ello dividimos las colonias en manzanas y de este número seleccionamos aleatoriamente un número de manzanas para entrevistar a cada familia.

Muestreo Aleatorio Simple

El muestreo aleatorio simple tiene las características establecidas en la definición dada en la página 30. Es el muestreo más recomendable para el estudio estadístico, solamente que tiene sus inconvenientes.

Para poder hacer el muestreo aleatorio simple debemos contestarnos las siguientes preguntas:

1. Dada una población finita de N elementos, ¿Cuantas muestras de "n" elementos podemos formar?

2. Conociendo las "n" muestras ¿Cómo podemos tomar una de ellas que sea representativa de la población?

Para dar respuesta a la primera pregunta, nos trasladamos al fascículo donde estudiaste el análisis combinatorio y aplicamos la ecuación:

$$C\binom{N}{n} = \frac{N!}{n!(N-n)!} \quad . . . . (14)$$

Ejemplo:

Determina ¿Cuantas muestras de tamaño n se pueden formar de una población finita N para los siguientes datos?

a) n = 2 y N = 20

b) n = 3 y N = 100

Solución:

$$a) \quad C\binom{20}{2} = \frac{20!}{2!(20-2)!} = \frac{20\ 19\ 18!}{2!(18)!} = 190$$

Este resultado nos dice que con una población de 20 elementos podemos tomar 190 muestras de dos elementos cada una.

$$b) \quad C\binom{100}{3} = \frac{100!}{3!(100-3)!} = \frac{100\ 99\ 98\ 97!}{1\ 2\ 3(97)!} = 50\ 33\ 98 = 161,700$$

Este resultado nos índica que de una población de 100 elementos podemos formar 161,700 muestras de 3 elementos.

Para contestar la segunda pregunta observamos lo siguiente:

Para que estas muestras sean representativas en el primer caso cada muestra debe tener $\frac{1}{190}$ de probabilidad de ser seleccionada.

En el 2ª caso cada muestra debe tener $\frac{1}{161700}$ de probabilidad de ser seleccionada.

¿Como debemos tomar cada muestra para que sea representativa?

Hay varias formas de tomar la muestra. Estas formas son las siguientes: en el primer caso cuando el número de muestras no es muy grande, se pueden numerar recortes de papel, doblarlos y meterlos en un recipiente donde se puedan mezclar ampliamente. Una vez mezclados, se saca la muestra.

Por ejemplo:

En una empresa se premiará con un viaje a Europa a <u>solo</u> 2 de los 5 empleados de mayor eficiencia. ¿Cómo seleccionamos a los dos que deben ir?

Solución:

A cada empleado lo representamos con la primera letra de su nombre.

1. Abraham (A)
2. Dionisio (D)
3. Efraín (E)
4. Fausto (F)
5. Iván (I)

Determinamos el número de muestras

$$C\binom{N}{n} = C\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5\ 4\ 3!}{2(3)!} = \frac{20}{2} = 10$$

$$P(n) = \frac{1}{10}$$

Cada muestra la escribimos en un recorte de papel y éstas son:

1	2	3	4	5	6	7	8	9	10
A-D	A-E	A-F	A-I	D-E	D-F	D-I	E-F	E-I	F-I

Doblamos bien el corte de papel de cada muestra, la introducimos en una vasija; la agitamos ampliamente y extraemos a la pareja afortunada.

Quizá haya visto este procedimiento en el sorteo de los equipos para el campeonato mundial de fútbol. En el sorteo se usaron esferas huecas bisectadas y en su interior se colocó el nombre de cada equipo, se revolvían ampliamente, se sacaba una esfera de la cual se tomaba el nombre del equipo y se colocaba en el grupo correspondiente.

Ejercicio:

1) Si para el campeonato mundial de fútbol hay 24 equipos de los cuales se deben formar 6 grupos de 4. ¿Como organizarías los equipos para que cada muestra sea aleatoria? ¿Como tomarías cada muestra y qué equipos la compondrían? Determina los dos posibles finalistas.

2) Calcula el número de muestras de tamaño 3 para una población de:
   a) 7 elementos
   b) 15 elementos
   c) 50 elementos

3) Calcula el número de muestras de tamaño 5 para una población de:
   a) 10 elementos
   b) 25 elementos
   c) 75 elementos

Si tienes alguna duda consulta a tu profesor o a tu consultor académico.

b) Si el número de muestras es muy grande como en el último ejercicio, 15^{5}, que son 17,259,390; la forma explicada con recortes de papel no es la adecuada. Para estos casos se usa otro procedimiento que consiste en usar una tabla de números aleatorios como la que se incluye en el apéndice "B".

Esta tabla de números aleatorios se puede construir fácilmente con un programa de computación.

Uso de la tabla de números aleatorios.

Para explicar su uso, veamos el siguiente ejemplo:

El Banco Nacional de México tiene una promoción para tarjetahabientes que consiste en condonarles la cuenta a 10 personas de cada sucursal, en la primera quincena del mes de enero de 1994. La lista de cuentahabientes es de 550 y para determinar la muestra aleatoria numeramos cada cliente con tres cifras en orden ascendente esto es: 001,002, 003, ..., 550 y nos situamos al azar en una columna de números aleatorios y nos desplazamos en ella en la dirección que queramos analizando las tres primeras cifras de cada número hasta completar los 10 números de la muestra.

Para nuestro ejemplo nos situamos en la última página de números aleatorios del apéndice "B", en la columna 27 renglón 31 y nos desplazamos hacia abajo, los números obtenidos de 3 cifras son:

187, 155, 388, 320, 281, 088, 520, 275, 480 y 273

Como la tabla es de números aleatorios, podemos asegurar que esta muestra es aleatoria.

Como habrás notado mediante el uso de números aleatorios, es muy fácil tomar una muestra aleatoria.

Ejercicio:

Mediante el uso de las tablas del apéndice "B", realiza el siguiente ejercicio.

En una empresa de 120 empleados se desea obtener una muestra aleatoria de 10 empleados para darles un curso de actualización ¿Qué empleados formarían la muestra?

Distribución de Media Muestrales

Ya sabemos cómo determinar el número de muestras de una población y cómo seleccionar una muestra aleatoria, ahora estudiaremos cómo se organiza una distribución de medias muestrales.

La distribución de medias muestrales son las probabilidades de todas las medias posibles de las muestras de una población finita.

Toda distribución de probabilidad puede describirse mediante su media y su desviación estándar.

Al tomar las muestras aleatorias se cometen ciertos errores que se reflejan en que la media y la distribución de cada muestra no son iguales, y por lo tanto la media y la desviación estándar de la población tampoco coinciden con los de la muestra. Por esta razón, la desviación estándar de la distribución de un estadístico muestral recibe el nombre de error estándar del estadístico.

El error estándar no solamente indica el tamaño del error accidental, sino también la exactitud que alcanzaremos si usamos un estadístico muestral para estimar un parámetro de la población.

Veamos el siguiente ejemplo:

De una población cuyos elementos son {1,3,5,7,9}, formar el número de muestras aleatorias de 2 elementos, construir la distribución de medias muestrales, determinar la media de la distribución de medias ($\mu_{\bar{x}}$); determinar la desviación estándar de la distribución, de medias y comparar estos resultados con los parámetros de la población.

Solución:

$n = 2$
$N = 5$

Media de la población:

$$\mu = \frac{1+3+5+7+9}{5} = \frac{25}{5} = 5 \qquad \therefore \ \mu = 5$$

σ^2 = Varianza de la población:

$$\sigma^2 = \frac{1}{5}[(1-5) + (3-5) + (5-5) + (7-5) + (9-5)]$$

$$\sigma^2 = 8 \qquad \therefore \qquad \sigma = \sqrt{8} = 2.83 \qquad \therefore \quad \sigma = 2.83$$

Número de muestras $C\binom{5}{2} = \dfrac{5!}{2!(5-2)!} = 10$

Conjunto de muestras

{(1,3), (1,5), (1,7), (1,9), (3,5), (3,7), (3,9), (5,7), (5,9),(7,9)}

Conjunto de medias muestrales

{2, 3, 4, 5, 4, 5, 6, 6, 7, 8}

Probabilidad de las medias muestrales

$\bar{X}$	Probabilidad
2	$\frac{1}{10}$
3	$\frac{1}{10}$
4	$\frac{2}{10}$
5	$\frac{2}{10}$
6	$\frac{2}{10}$
7	$\frac{1}{10}$
8	$\frac{1}{10}$

Media de la distribución de medias muestrales

$$\mu_{\bar{X}} = 2(\frac{1}{10})+3(\frac{1}{10})+4(\frac{2}{10})+5(\frac{2}{10})+6(\frac{2}{10})+7(\frac{1}{10})+8(\frac{1}{10})$$

$$\mu_{\bar{X}} = 5$$

Varianza de la distribución de medias muestrales:

$$\sigma^2 =(2-5)(\frac{1}{10})+(3-5)(\frac{1}{10})+(4-5)(\frac{1}{10})+(5-5)(\frac{1}{10})+(6-5)(\frac{1}{10})+(7-5)(\frac{1}{10})$$

$$+(8-5)(\frac{1}{10})$$

$$\sigma^2 = 3 \qquad \therefore \qquad \sigma = \sqrt{3} = 1.73$$

De estos resultados concluimos que:

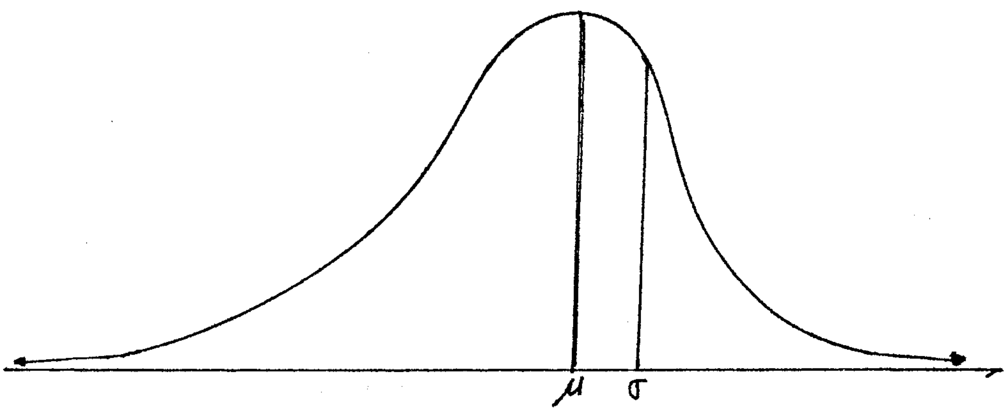
1. La media de la distribución de medias $\mu_{\bar{X}}$ es igual a la media poblacional (μ)

2. La desviación estándar de la distribución de medias $\sigma_{\bar{X}}$ es menor que la desviación estándar poblacional (σ)

De este ejemplo podemos ver el error estándar de la media en que $\sigma_{\bar{X}} < \sigma$, el cual ya habíamos mencionado.

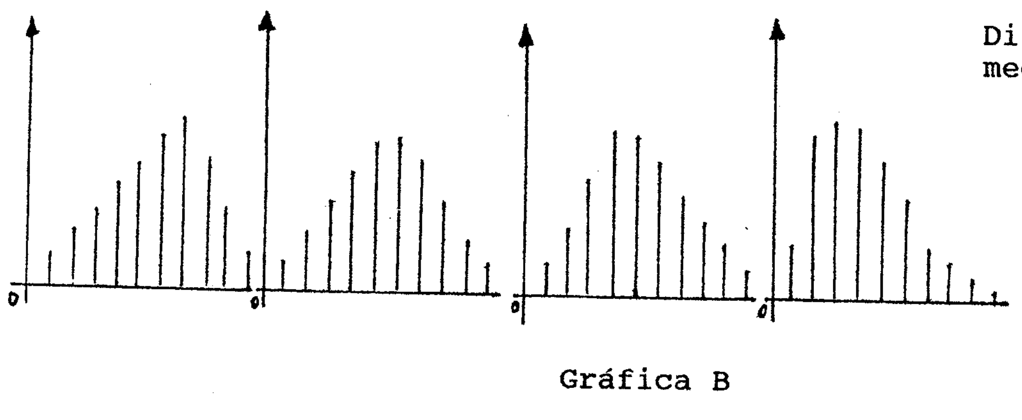
Ilustramos el proceso de la distribución de media muestrales mediante las siguientes gráficas.

Dada una población de N elementos, ésta tiene una media μ y una desviación estándar σ cuya relación entre ellos se muestra en la gráfica siguiente:



Distribución de la población

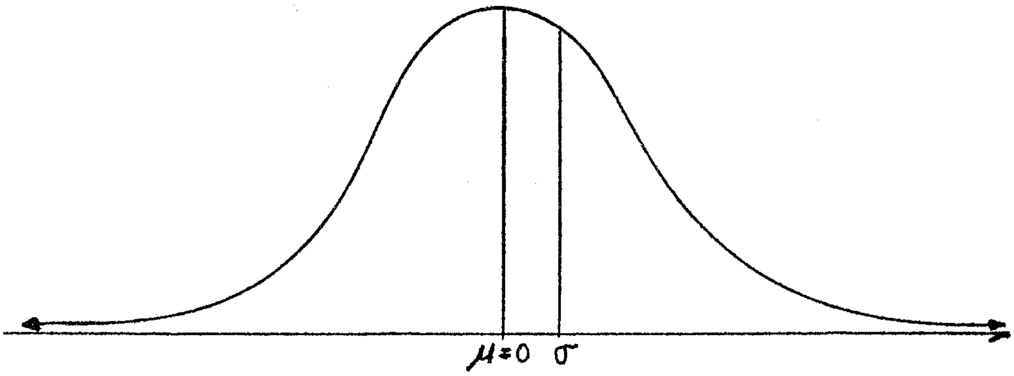
Gráfica A



Distribución de medias muestrales

Gráfica B

De esta población se pueden formar un gran numero de muestras pero solamente mostramos 4 de ellas para ilustrar el procedimiento.



Gráfica C

Distribución de medias muestrales con $\mu_{\bar{X}} = \mu$

$\sigma_{\bar{X}} < \sigma$

Con estas gráficas podemos darnos mejor idea de la secuencia de operaciones que realizamos para obtener la distribución de medias muestrales representada por la gráfica C. Esta gráfica es simétrica y tiene la forma de la curva normal o campana de Gauss. De esta misma gráfica podemos constatar que la media poblacional es igual a la media de la distribución de medias, lo cual no ocurre con la desviación estándar en la que hay un error.

La desviación estándar de la distribución muestral de medias para poblaciones finitas de tamaño N, se puede calcular por la ecuación

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \ldots (15)$$

Esta ecuación se llama error estandard de las medias.

A la raíz $\sqrt{\frac{N-n}{N-1}}$ le llamamos factor de corrección por población finita, toda vez que para poblaciones infinitas se aplica la ecuación

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \ldots (16)$$

Si la muestra es al menos el 5% de la población entonces el factor de corrección no afecta porque tiende a la unidad.

Veamos el ejemplo que usamos para la distribución de medias en que:

N=5 n=2 $\sigma_{\bar{X}} = \sqrt{3}$ $\sigma = \sqrt{8}$

Con estos valores sustituimos en la fórmula y obtenemos:

$$\sigma_{\bar{x}} = \frac{\sqrt{8}}{\sqrt{2}} \cdot \sqrt{\frac{5-2}{5-1}} = \sqrt{\frac{8}{2}}\sqrt{\frac{3}{4}} = \sqrt{\frac{3(8)}{2(4)}}$$

$$\sigma_{\bar{x}} = \sqrt{3}$$

De este resultado concluimos que por ser la muestra al menos el 5% de la población, el factor de corrección no afecta a la distribución estándar de medias.

Ejercicio:

1. De una población finita N ={3, 4, 5, 6, 7, 8 y 9}, se toman muestras aleatorias de 2 elementos.

   a) Calcula la media μ y la desviación estándar de la población, σ.
   b) Calcula el núumero de muestras aleatorias que se pueden formar, establece el conjunto de muestras y determina la probabilidad de cada una.
   c) Construye la distribución de medias muestrales de la población.
   d) Calcula la media, la varianza de la distribución de medias; compara los resultados con los de la población y comprueba el valor de la desviación estándar de las medias, aplicando la ecuación del error estándar.
   e) Realiza las gráficas de la secuencia de operaciones.

2. Determina el factor de corrección para una población N=10,000 con muestras de n=100 e indica si afecta o no a la desviación estándar de la distribución de medias muestrales $\sigma_{\bar{x}}$

EL TEOREMA CENTRAL DEL LIMITE

En los ejemplos anteriores quedó establecido que las muestras aleatorias tomadas de una población tienen diferentes medias y comparadas con la media muestral, hay un determinado error.

Con respecto a este error, el teorema de Chebyshev dice:

> Podemos afirmar con una probabilidad de cuando menos $1-\frac{1}{k^2}$ que la media de una muestra aleatoria de tamaño (n) difiere de la media de la población en un valor igual a $\sigma_{\bar{x}}$ k.

Este teorema de Chebyschev afirma que para estimar la media poblacional, cuando utilizamos la media de una muestra aleatoria podemos afirmar con una probabilidad de cuando menos $1-\frac{1}{k^2}$ que nuestro error será menor que: $\sigma_{\bar{X}}(K)$.

Ejemplo:

Dada una población de N elementos ¿Cuál es el error para K=2, si tomamos una muestra n=64 con una desviación estándar Í = 20?

Solución:

Calculamos $\sigma_{\bar{X}} = \frac{20}{\sqrt{64}} = 2.5$

Se afirma con una probabilidad de $1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 0.75$ que la media de la muestra difiere de la media de la población, y que el error que se comete es menor que:

$$\sigma_{\bar{X}}\, k = 2.5\ (2) = (5)$$

Con este teorema podemos conocer el error que cometemos sin tener que hacer el desarrollo de la distribución de medias muestrales.

Existe otro teorema aun mas preciso que el de Chebyshev, éste teorema se llama:

Teorema del limite central y dice:

Si el tamaño de la muestra (n) es grande, entonces la distribución muestral teórica de las medias, se puede aproximar con una distribución normal.

Este teorema es fundamental en la estadística, ya que justifica el uso de los métodos de la curva normal en la solución de una amplia gama de problemas. Se aplica a poblaciones infinitas y a poblaciones donde n es una parte de la población. Es difícil especificar con exactitud cuan grande debe ser (n) para poder aplicar el teorema central del limite. Sin embargo para n=20 ya se puede obtener un polígono de frecuencias simétricas y en forma de campana para n=30 ya podemos considerar a (n) suficientemente grande.

Si la población que muestreamos tiene un polígono de frecuencias simétrico y en forma de campanas, entonces podemos aplicar el teorema del límite central sin importar el tamaño de (n).

Ejemplo:

Apliquemos el teorema del limite central en el mismo problema donde aplicamos el teorema de Chebyshev o sea $n=64$ $\sigma=20$

Chebyshev dice: ¿Cual es la probabilidad de que el error que se comete al tomar la media de la muestra como parámetro de la población sea menor que 5? y con su teorema se obtiene cuando menos de 0.75. Este resultado nos indica que puede ser más pero no se precisa.

Veamos el cálculo con el teorema central del límite.

El área bajo la curva es para:

$$Z_1 = \frac{-5}{20\ /\ \sqrt{64}} = -2 \quad \text{y} \quad Z_2 = \frac{5}{20\ /\ \sqrt{64}} = 2$$

Con los valores de Z nos vamos a las tablas del apéndice A, que se encuentra al final del fascículo, obtenemos

$$P(Z\ ) = P(Z_1) = 0.4772$$

$$\therefore \quad P(Z) = P(Z_1) + P(Z_2) = 2(0.4772) = 0.9544$$

Con este ejemplo podemos ver como el teorema central del limite es mas preciso que el de Chebyshev, toda vez que Chebyshev da un rango de aproximación y el del limite central nos fija el valor de la probabilidad.

La gráfica de la curva normal de este problema se muestra en la siguiente figura cuya área está sombreada.

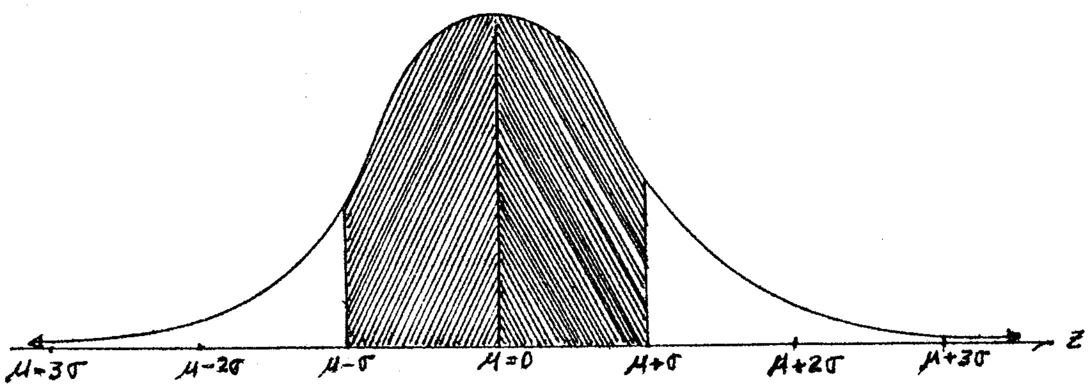


Fig. 9

El tamaño de (n) cobra importancia cuando el polígono de frecuencias no es simétrica, en estos casos en la medida en que (n) aumenta, el error estándar disminuye.

Veamos el siguiente ejemplo:

Si tomamos a $\bar{X}$ como estimación de μ ¿Cómo es el error estándar de la media si n=50 se incrementa a n=200

Solución:

$$\frac{\frac{\sigma}{\sqrt{200}}}{\frac{\sigma}{\sqrt{50}}} = \frac{\sigma\sqrt{50}}{\sigma\sqrt{200}} = \sqrt{\frac{50}{200}} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

Con este ejemplo podemos ver que al aumenta el valor de (n), el error de la media disminuye; en nuestro ejemplo disminuyó a la mitad.

Si la naturaleza del problema que se está resolviendo tiene distribución normal entonces el teorema del limite central cobra mayor importancia en el cálculo del error estándar de la media. Veamos el siguiente ejemplo.

Ejemplo:

Dada una población normal de μ=100 y σ=25, formamos muestras de 5 elementos y determinamos la media de cada muestra X. Sin duda le media de cada muestra es mayor que la media poblacional y la desviación estándar de la distribución muestral es menor que la de la población, porque la dispersión de la muestra es menor que la de la población. Gráficamente lo podemos ver de la siguiente forma:

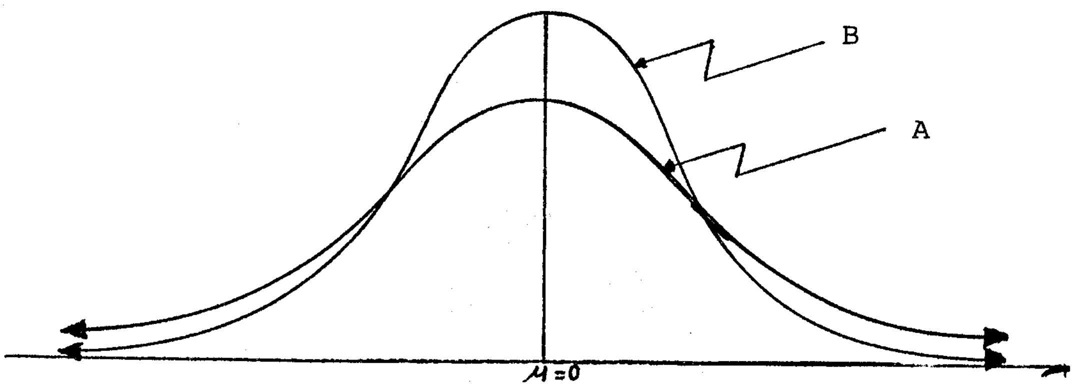


Fig.

La gráfica A es la distribución muestral de la población μ=100 y σ=25

La gráfica B es la distribución de las medias muestrales con n=5 y $\sigma_{\bar{X}}<25$

Ahora formemos muestras con n=20 y la gráfica de esta nueva distribución de medias muestrales es la C

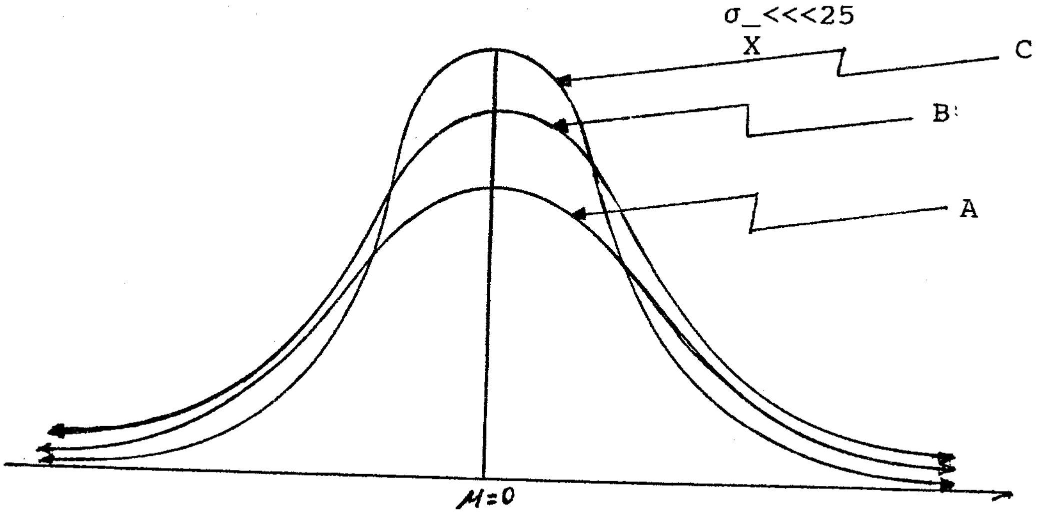


Fig. 11

De la gráfica C concluímos que al aumentar el valor de (n) estamos intensificando el efecto de promediar la muestra y por ello la dispersión disminuye aún mas, es decir en la gráfica C $\sigma_{\bar{X}} << 25$

De lo anterior concluimos que si (n) crece el error estándar que se comete al tomar a la media muestral como estimador de la población (μ) es cada vez mas pequeño.

Ya dijimos que para n>30 podemos considerar que (n) es grande y aunque el teorema central del limite se puede aplicar a una muestra cuya n<30, el error estandard es mayor. En estos casos se recomienda aplicar otra distribución que nos permite cálculos mas precisos en muestras pequeñas y que veremos a continuación.

Ejercicio:

1. La media de una muestra aleatoria de tamaño n=400 se utiliza para estimar la media de una población infinita que tiene desviación estándar σ=5. ¿Que podemos decir acerca de la probabilidad de que el error será menor que 0.4 mediante el uso de:

   a) El teorema de Chebyshev
   b) El teorema central del limite.

2. En los equipos de detección de la contaminación por humo, se usan pequeñas baterías cuya duración tiene una desviación σ=77 horas. Se utiliza la media de una muestra de tamaño n=49 para estimar la media poblacional. Mediante la aplicación del limite central ¿Que podemos decir acerca de la probabilidad de que la estimación tenga un error

   a) Menor de 10 hrs?
   b) Menor de 20 hrs?

DISTRIBUCION T - DE STUDENT

En la inferencia estadística se hacen generalizacines con base en muestras, mediante estimaciones y pruebas de hipótesis.

La estimación consiste en asignar un valor numérico a un parámetro de una población sobre la base de datos de muestras; y la prueba de hipótesis está basada en la aceptación o rechazo de suposiciones concernientes a los parámetros de una población.

En el subtema 3.2.2 se hicieron estimaciones de medias poblacionales a través de medias muestrales cuando el tamaño de la muestra es grande (Teorema del limite central).

Sin embargo, cuando la muestra involucrada es pequeña es muy probable que la desviación típica muestral S sea bastante distinta de la desviación típica de la población σ; en consecuencia en estos casos no se puede utilizar el teorema central del limite para estimar la media de una población a través de la media de una muestra. En estos casos se utiliza otra distribución llamada t de Student.

La teoría de las muestras pequeñas sacadas de una población normal de desviación típica Í desconocida, fue descubierta por el inglés William Gosset en 1908 con el seudónimo de Student.

La distribución t de Student se representa mediante la expresión

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} , \qquad \ldots\ldots \text{(12)}$$

donde:

$\bar{X}$ = media de la muestra
μ = media de la población
s = desviación típica de la muestra
n = tamaño de la muestra

La distribución t de Student se basa en la consideración de que la población a partir de la cual se obtiene la muestra tiene una distribución normal, ó al menos aproximadamente normal.

Con la distribución de Student es posible estimar parámetros de una población a partir de los estadísticos calculados para una muestra cuando ésta es pequeña.

Dicha estimación puede ser puntual o por intervalos.

La estimación es puntual cuando se estiman parámetros empleando valores de una muestra única; y por intervalos cuando se establece un rango de valores dentro de los cuales se espera que el parámetro caiga.

Como ejemplo para ilustrar un problema en la estimación de medias, considérese un estudio en el cual un médico desea determinar el incremento promedio real del pulso cardiaco de una persona que realiza cierta tarea ardua. Los siguientes datos representan los incrementos de pulso cardiaco en pulsaciones por minuto que el médico obtuvo en relación con 32 personas:

27, 25, 19, 28, 35, 23, 24, 22,
14, 30, 32, 34, 23, 26, 29, 27,
27, 24, 31, 22, 23, 38, 25, 16,
32, 29, 26, 25, 28, 26, 21, 28.

Calculando la media de la muestra se tiene que $\bar{X} = 26.2$ pulsaciones por minuto y en ausencia de otra información este número sirve como estimador de la media de la población μ.

Una estimación de este tipo ese una estimación puntual ya que consta de un solo número. Pero esta manera de estimar un parámetro no es la mas confiable ya que no nos dice en cuanta información se basa la estimación y tampoco nos dice nada acerca del posible tamaño del error. Una estimación por intervalos es mucho más útil, que una estimación puntual debido a que posee más información; no solo da el valor estimado, sino también la precisión y el nivel de confianza.

Propiedades de la Distribución T - Student

Comparando la variable normal estandarizada $Z = \frac{X-\mu}{\sigma/\sqrt{n}}$ y la variable "t de student $t = \frac{X-\mu}{s/\sqrt{n}}$ se observa que son muy similares y que el único cambio está en el denominador donde se sustituye S en lugar de σ.

Como la distribución normal estándar Z, la distribución t también es contínua, en forma de campana y perfectamente simétrica. La única diferencia entre las dos distribuciones, es que la distribución t tiene mayor variabilidad; la curva t esta más extendida en la parte de las clases y es más achatada en la zona del centro.

En la siguiente figura se comparan los dos tipos de curvas.

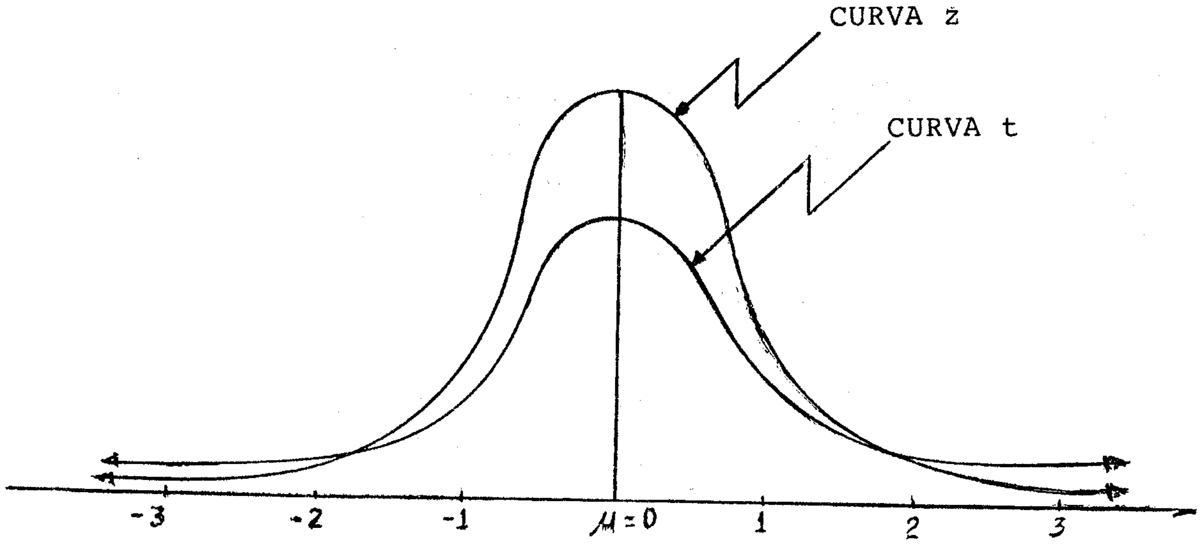


Fig. 12

La siguiente figura muestra el comportamiento de la distribución t comparada con la distribución Z

g.l.= grados de libertad. Los grados de libertad están definidos en la página 49.

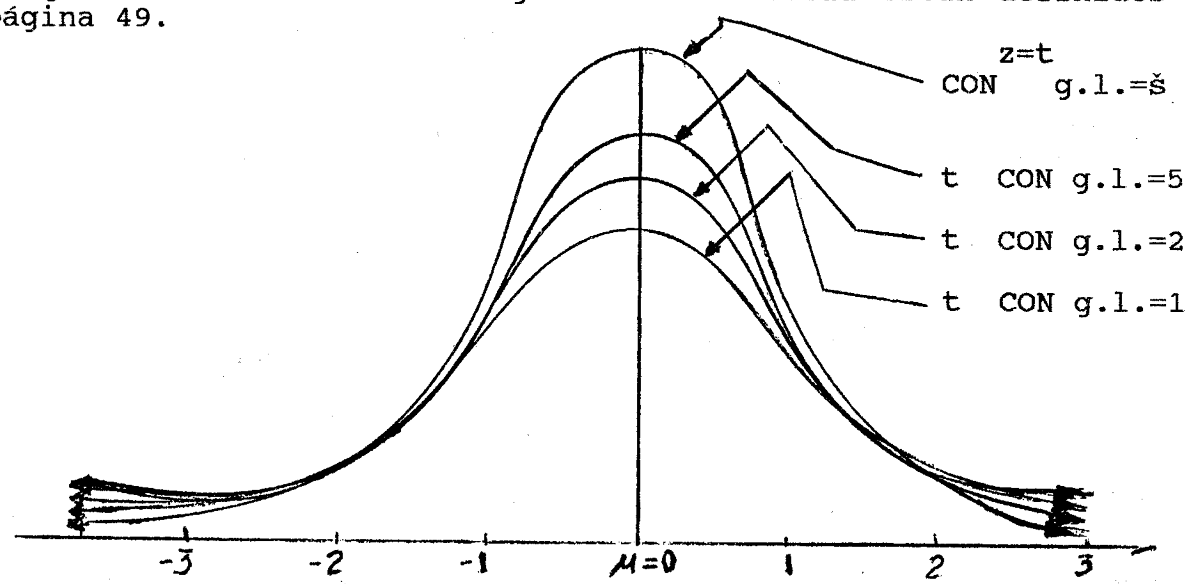


Fig. 13

De la figura se puede observar que conforme aumenta el tamaño de la muestra, la curva t se aproxima a la curva normal; cuando el tamaño de la muestra n tiende a infinito, la curva t es idéntica a la curva normal. También de la figura se puede afirmar que no hay una sola distribución para la distribución t de Student, sino una familia de distribuciones; esto es debido al efecto del tamaño de la muestra. Si n es pequeña, la t de Student correspondiente es muy ancha, pero si n • 30 la distribución t y la normal Z son casi indistinguibles. De todo lo anterior se pueden establecer propiedades de la distribución de t de Student.

Características de la distribución t de Student.

1. Es simétrica con respecto a la media.
2. Tiene media $\mu=0$ y $\sigma>1$
3. La desviación típica $\sigma \approx 1$, cuando el tamaño de la muestra tiende a infinito.
4. La distribución Z tiene solamente una distribución con media $\mu=0$ y desviación típica $\sigma=1$; mientras que la distribución t tiene una familia de distribuciones.
5. La distribución t no se tabula según el tamaño de la muestra, sino en términos del número de grados de libertad.
6. La distribución t es contínua, en forma de campana.
7. La distribución t se basa en la consideración de que la población a partir de la cual se obtienen la muestra tiene una distribución normal o aproximadamente normal.
8. La variabilidad de la distribución t, depende de dos variables aleatorias (S y $\bar{X}$)
9. La distribución t de student se utiliza para estimar parámetros poblacionales a través de los valores de las muestras, para muestras pequeñas (n<30) y cuando la desviación típica S es conocida.
10. El número de grados de libertad es el único parámetro de la distribución t. Esto es, la forma de la curva t está totalmente definida cuándo se conoce el número de grados de libertad (g.l.=n-1)

El término "grados de libertad" abreviado g.l., se refiere al número de datos que pueden variar libremente, después de haber impuesto ciertos restricciones a nuestros datos.

El número de grados de libertad es el tamaño de la muestra menos uno; es decir g.l.=n-1.

Cuando se quiere calcular la media de una población a través de la media muestral, debido a la variabilidad de la media muestral X; ésta no será exactamente igual a la media poblacional μ, por lo tanto siempre habrá un margen de error, llamado error muestral; es decir,

$$\mu = \bar{X} \pm \text{error muestral.}$$

El máximo error que se comete cuando se utiliza $\bar{X}$ como estimación de μ, cuando n·30 está dado por

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$, donde $Z_{\frac{\alpha}{2}}$ denota el valor

de Z para el cual el área situada debajo de la curva normal estándar a su derecha es igual a $\alpha/2$.

La selección del valor de α es arbitraria, depende de qué tanto error se esté dispuesto a tolerar.

El error que se está dispuesto a tolerar se llama <u>nivel de confianza</u>.

Ejemplo:

$Z_{0.05}$, significa que estamos dispuestos a tolerar un 5% de error.

Hallando este nivel de confianza a una gráfica se tiene:

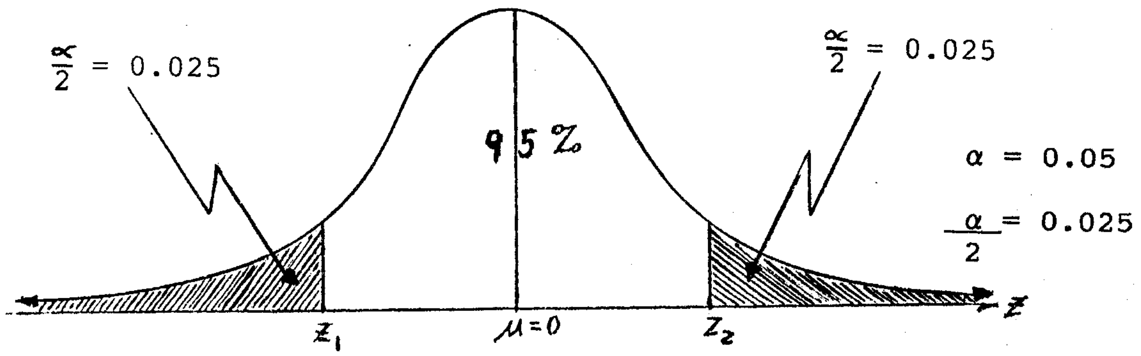


Fig. 14

En las tablas del área bajo la curva normal se obtiene $Z_2 = 1.96$; y como la curva es simétrica $Z_1 = -1.96$.

Lo anterior significa que el 95% de las diferencias muestrales cae entre -1.96 y 1.96 desviaciones estándares.

En base al ejemplo anterior, obtener $Z_{0.01}$ y representarlo en una gráfica.

Ejemplo:

Un experto en mecánica utiliza la media de una muestra aleatoria de tamaño n=30 para estimar el tiempo promedio que le toma a un mecánico realizar cierta tarea. Si con base en la experiencia, el experto puede suponer Í=2.5 minutos para estos datos ¿Qué se puede decir con un nivel de confianza del 1% acerca del tamaño máximo de su error?

Solución:

$n=30 \qquad \sigma=2.5 \qquad \alpha=1\%=0.01 \Rightarrow \alpha/2=0.005$

Utilizando las tablas del área bajo la curva normal se tiene que

$$Z_{0.005} = 2.57$$

Sustituyendo estos datos en la formula $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{x}}$ se tiene

$$E = (2.57)\left(\frac{2.5}{\sqrt{30}}\right) = \frac{6.425}{5.477} = 1.17$$

El resultado obtenido significa que el experto en mecánica puede afirmar con una certeza del 99% que su error será cuando mucho de 1.17 minutos.

Ejercicio:

Con referencia al problema de los pulsos cardíacos de las 32 personas, ¿Qué se puede decir con un nivel de confianza del 5% acerca del error máximo si se utiliza X=26.5 como estimación del incremento promedio real del pulso de una persona que realiza la tarea dada?

<u>Formato de una muestra para estimar la media μ cuando $n \geq 30$.</u>

La fórmula $E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, también se puede utilizar para determinar el tamaño de la muestra que se necesita para lograr un grado de exactitud deseada. Despejando n de la expresión anterior se tiene:

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E}\right]^2 \quad . . . . . (16)$$

Fórmula para Determinar el Tamaño de la Muestra

Ejemplo:

Un profesor de estadística de la Universidad desea emplear la media de una muestra tomada al azar para estimar la cantidad promedio de tiempo que los estudiantes requieren para pasar de una clase a la siguiente. Además desea que esta estimación tenga un error de cuando mucho 0.30 minutos con probabilidad 0.95. Si se sabe de estudios similares anteriores que es razonable tomar σ=1.50 minutos ¿De qué tamaño tendrá que tomar una muestra?

Solución:

La probabilidad 0.95 de que al hacer la estimación se tenga un error de cuando mucho 0.30 significa que se está tomando un nivel de confianza del 5% $\therefore\ \alpha=5\% \Rightarrow \alpha/2 = 2.5\% = 0.025$

De tablas se tiene $Z_{0.025} = 1.96$; además σ=1.50, E=0.30

Sustituyendo los datos en la formula

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E} \right]^2 \quad \text{se tiene}$$

$$n = \left[\frac{(1.96)(1.50)}{0.30} \right]^2 = 96.04$$

$\therefore$ Se requiere una muestra aleatoria de tamaño n=96 para la estimación.

Ejercicio:

1. En un estudio de los hábitos de ver televisión, se busca estimar el número de horas en promedio que los alumnos de bachillerato ven televisión por semana. Si es razonable suponer σ=3 horas, ¿De qué tamaño deberá ser la muestra de manera que se pueda afirmar con la probabilidad de 0.99 que la media de la muestra fallará cuando mucho en 35 minutos?

Intervalos de Confianza

Anteriormente ya se dijo que para estimar parámetros, lo más adecuado es formar un intervalo de confianza, el cual generalmente incluirá al parámetro por estimar.

Como ya vimos al estimar μ en base a la media de la muestra $\overline{X}$, la estimación no será perfecta; es decir, siempre habrá un margen de error, tal que:

$$\mu = \overline{X} \pm \text{error muestral};$$

pero ya vimos que el máximo error muestral que se puede cometer es $E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, por lo tanto podemos escribir

$$\mu = \overline{X} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \qquad \ldots\ldots\ (17)$$

donde:

$\overline{X}$ = media muestral
$Z_{\frac{\alpha}{2}}$ = Es el valor de Z para el cual el área bajo la curva normal a la derecha de Z es $\alpha/2$
α = nivel de confianza
$\frac{\sigma}{\sqrt{n}}$ = desviación t_pica de la media.

Puesto que los niveles de confianza más utilizados son 0.05 y 0.01; entonces podemos establecer los siguientes intervalos de confianza.

$$\mu = \overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}} \text{ ; intervalo de confianza de 95\%}$$

$$\mu = \overline{X} \pm 2.58 \frac{\sigma}{\sqrt{n}} \text{ ; intervalo de confianza de 99\%}$$

Recuerda que para α=0.01 => $\alpha/2$=0.025 y que $Z_{0.025} = 1.96$

Ejercicio:

1. Para α=0.01; obtener: $Z_{\alpha}/2$=__________, $Z_{\frac{\alpha}{2}}$=__________.

Ejemplo:

Obtener el intervalo de confianza del 95% del conjunto de datos:

{ 1, 5, 2, 3, 4, 1, 2, 2, 4, 3.}

x	x^2
1	1
5	25
2	4
3	9
4	16
1	1
2	4
2	4
4	16
3	9
27	89

1er. Paso: Se determina la media

$$\bar{X} = \frac{\Sigma X}{N} \qquad \therefore \quad \bar{X} = \frac{27}{10} = 2.7$$

$$\Sigma X = 27 \qquad \therefore \quad \bar{X} = 2.7$$

$$N = 10$$

2do. Paso: Se obtiene la desviación estándar de la muestra

$$S = \sqrt{\frac{\Sigma X}{N} - \bar{X}}$$

$$S = \sqrt{\frac{89}{10} - (2.7)} = \sqrt{8.9-7.29}$$

$$S = \sqrt{1.61} = 1.27 \qquad \therefore \quad S=1.27$$

3er. Paso: se obtiene el error estándar de la media.

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{N-1}}$$

nótese que en el denominador, en la fórmula se escribió N-1 en vez de N; la razón es que N-1 corrige el sesgo del error estándar.

$$\sigma_{\bar{X}} = \frac{1.27}{\sqrt{10-1}} = \frac{1.27}{\sqrt{9}} = \frac{1.27}{3} = 0.42$$

$$\therefore \quad \sigma_{\bar{X}} = 0.42$$

4o. Paso: Se multiplica el error estándar de la media por 1.96 que es el valor de z al nivel de confianza de 0.05

$$\mu = \bar{X} \pm 1.96(\sigma_{\bar{X}})$$

$$\mu = 2.7 \pm 1.96(0.42)$$

$$\mu = 2.7 \pm 0.82$$

$$\therefore \quad 1.88 \leq \mu \leq 3.52$$

Lo anterior significa que se puede asegurar con un 95% de confianza que la verdadera muestra poblacional está entre 1.88 y 3.52.

Ejercicio:

Del conjunto de datos del problema anterior, encontrar el intervalo de confianza del 99%.

Ejemplo:

Un fabricante de productos especiales de acero, necesita estimar la dureza media de un lote grande de piezas de acero que acaba de recibir. Es muy importante la determinación de la dureza ya que si ésta sale de un cierto rango, es necesario aplicar un tratamiento costoso para llevarla al grado de dureza deseado. Imagínese que usted trabaja en el departamento de pruebas de cierta compañía y le han enviado el resultado de una prueba de dureza efectuada en una muestra aleatoria de 40 piezas, siendo la media de la muestra X = 70 y desviación estandard S = 2 ¿Qué haría usted?

Solución:

Se tiene que estimar la dureza media μ en base a una muestra con n=40, $\bar{X}$=70 y S=2 y un nivel de confianza del 99%, ya que la situación es bastante delicada.

$$\therefore \quad \mu = \bar{X} \pm 2.58\frac{\sigma}{\sqrt{n}}$$

$$\mu = 70 \pm (2.58)\left(\frac{2}{\sqrt{40}}\right)$$

$$\mu = 70 \pm (2.58)(0.82)$$

o sea:

$$69.18 \leq \mu \leq 70.82$$

El gerente, al recibir el informe, observa que el resultado cae muy cerca del extremo del rango aceptable (de 68.3 a 70.6), pide que se aumente la precisión del intervalo de confianza del 0.82 a 0.50, preservando el nivel de confianza en 99% ¿Que haría usted?

Solución:

Hay que determinar el tamaño de la muestra necesaria para alcanzar la precisión de E = 0.50

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E} \right]^2 \quad \text{podemos tomar } \sigma=5$$

$$n = \left[\frac{(2.58)(2)}{0.50} \right]^2 = \left(\frac{5.16}{0.50} \right)^2 = (10.32)^2 = 106.50$$

Entonces nos bastaría una muestra de 107 piezas. Como ya teníamos 40 piezas, se manda completar la muestra probando la dureza de 67 piezas adicionales. Se calculan las nuevas $\bar{X}$ y S en base a la muestra total y se obtiene el nuevo intervalo de confianza a 99% con precisión de 0.50.

Ejercicio:

La actividad de ciertas vacunas puede mediarse únicamente a través de pruebas en organismos vivos (conejos, por ejemplo). Este procedimiento es costoso y tardado pro esencial para asegurar el funcionamiento correcto de estas vacunas.

a) Si la muestra de 30 pruebas dió un índice medio de actividad de $\bar{X}$=880 unidades con S=110, forme un intervalo de confianza de 95% para la actividad media de la vacuna.

b) Calcule el tamaño de muestra total necesaria para tener un error de estimación $E \leq 25$ unidades con 99% de confianza.

Confiabilidad de Promedios en Muestras Pequeñas.

Anteriormente ya se comparó la distribución Z con la distribución t de student.

Si en la distribución $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, se reemplaza t por Z y σ por S se

tiene la distribución:

$$t = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \qquad \text{distribución de Student}$$

Esta distribución se utiliza para estimar parámetros para muestras pequeñas.

Los intervalos de confianza se forman de la misma manera que en la distribución Z.

La forma de la curva de la distribución t de Student está basada en el número de grados de libertad (g.l.=n-1) en lugar del tamaño n de la muestra. A medida que aumenta el número de grados de libertad, la curva de la distribución t es menos variable.

Una muestra la vamos a considerar pequeña cuando sea $n<30$.

La tabla de valores t es diferente de la de valores Z. En la tabla de valores de la distribución t de Student cada fila corresponde a una distribución t distinta. La última columna da el número de grados de libertad.

Ejemplo:

Para 10 g.l., el 10% del área de la curva está a la derecha del valor t=1.383, y como la curva es simétrica, el 10% del área de la curva está a la izquierda del valor t = -1.383.

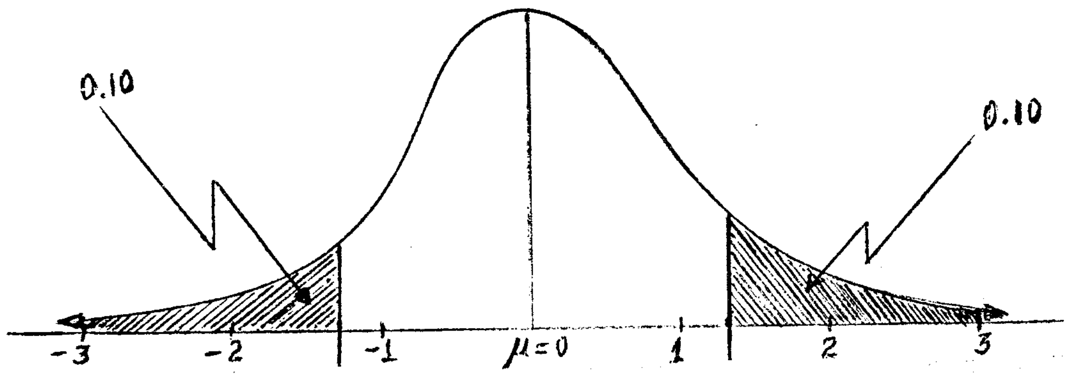


Fig. 14

Distribución t para 10 g.l. y $\alpha/2=0.10$

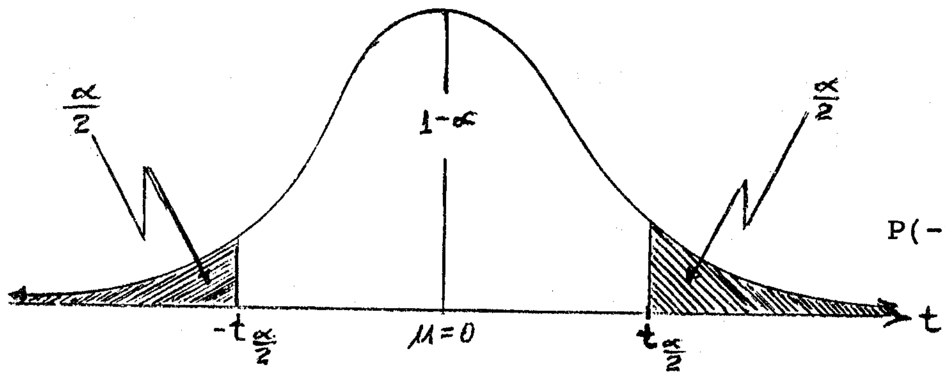
De la figura se tiene que el 80% de los casos están comprendidos entre -1.383 y 1.383.

Ejercicio:

Para 10 g.l. y un nivel de confianza de 5%, determinar el intervalo de confianza y dibujar su gráfica.

Para la distribución t de Student se define $t_{\frac{\alpha}{2}}$, de la misma forma como se difinió $z_{\frac{\alpha}{2}}$, de manera que el área situada debajo de la curva que está a la derecha de $t_{\frac{\alpha}{2}}$ es igual a $-t_{\frac{\alpha}{2}}$. Sin embargo $t_{\frac{\alpha}{2}}$ depende del número de grados de libertad.

Utilizando el hecho de que la distribución t es simétrica con respecto a t=0 (media de la distribución t); entonces la probabilidad de que la variable aleatoria que tiene una distribución t tome un valor entre $-t\alpha/2$ y $t\alpha/2$; es decir, $-t\alpha/2 \leq t \leq t\alpha/2$ es $1-\alpha$.



De la figura podemos afirmar que:

$$P(-t\alpha/2 \leq t \leq t\alpha/2)=1-\alpha$$

Distribución t

Igual que en la distribución Z, el intervalo de confianza en muestras pequeñas se puede escribir:

$$\mu = \bar{X} \pm t\ \alpha/2\ \frac{S}{\sqrt{n}}$$

Fórmula para determinar intervalos de confianza para muestras pequeñas.

Ejemplo:

La curva de la distribución t con 10 g.l. se muestra en la siguiente figura. Hallar el valor de t para que:

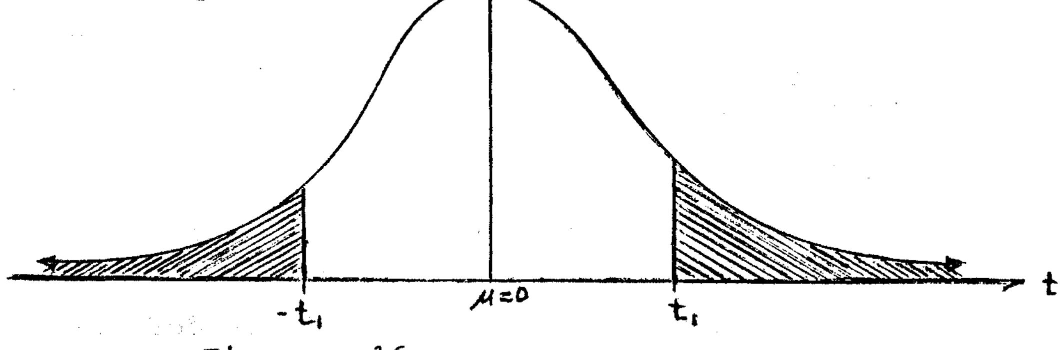


Fig. 16

a) El área rayada de la derecha sea 0.05
b) El área total rayada sea 0.10
c) El área no rayada sea 0.99
d) El área rayada de la izquierda sea 0.01

Solución:

a) $\alpha/2=0.05$

En las tablas de la distribución t se busca el nivel de confianza 0.05 con 10 g.l.(n=10+1=11) es decir, el tamaño de la muestra n=11

$$\therefore \quad t_{0.05} = 1.812$$

Ejercicio:

Resolver los demás incisos del problema anterior.

Ejemplo:

Los contenidos de ácido sulfúrico en siete recipientes similares son: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 y 9.6 litros. Encuentre un intervalo de confianza al 95% para la media de todos los recipientes, suponiendo una distribución aproximadamente normal.

Solución:

La media y la desviación estándar de la muestra para los datos proporcionados, son:

$X=10.0$ y $S=0.283$ (Checar estos resultados)

Empleando la tabla de la distribución t, se encuentra que $t_{0.025}=2.447$, para 6 g.l. Recuerde que $\alpha=0.05 \quad == \alpha/2=0.025$

.. el intervalo de confianza al 95% para μ es:

$$\mu = 10.0 \pm (2.447)\left(\frac{0.283}{\sqrt{7}}\right)$$

$$\mu = 10.0 \pm \frac{0.6925}{2.64575}$$

$$\mu = 10.0 \pm 0.26174$$

$$\therefore \quad 9.738 < \mu < 10.26174 \qquad \text{redondeando}$$

$$\therefore \quad 9.74 < \mu < 10.26$$

Ejercicio:

1. Una muestra aleatoria de 25 automóviles del mismo modelo se conducen de la misma forma y usando la misma calidad de gasolina. Los automóviles recorren un promedio de 9 km. por litro de gasolina, con una desviación tipo de 1.2 km. Estimar el recorrido medio por litro, y dar su intervalo de confianza a un nivel de confianza del 95%.

Verifica la respuesta: $8.505 \leqslant \mu \leqslant 9.495$

2. En un estudio de la contaminación del aire, una estación de experimentos obtuvo una media de 2.36 miligramos de materia orgánica suspendida soluble de benceno por metro cúbico con una desviación estándar de 0.48 de una muestra tomada al azar de tamaño n=10.

a) Construye un intervalo de confianza del 99% de la media de la población muestreada.

b) ¿Qué se puede afirmar con el 95% de confianza acerca del error máximo, si $\bar{X}=2.36$ miligramos se utiliza como estimación de la media de la población muestreada?

Verifica las respuestas.

a) $1.87 \leqslant \mu \leqslant 2.85$

b) $E = 0.34$

Pruebas de Hipótesis

Al hacer inferencias de características de poblaciones a través de muestras se utilizan los métodos de Estimación y Pruebas de Hipótesis.

Cuando se analizan características e poblaciones por el método de pruebas de hipótesis es necesario tener en cuenta los siguientes conceptos:

NIVEL DE CONFIANZA. Es el nivel de error que se esté dispuesto a tolerar.

ESTADISTICO DE PRUEBA. Es una variable aleatoria cuyo valor se utiliza para llegar a la decisión de rechazar o no la hipótesis nula.

REGION CRITICA. Es el conjunto de valores para el estadístico de prueba que llevará a rechazar la hipótesis nula.

REGION DE ACEPTACION. Es el conjunto de valores para el estadístico de prueba que provocará la aceptación de la hipótesis nula.

VALOR CRITICO. Es el valor que separa a la región de rechazo y la región de aceptación.

HIPOTESIS ESTADISTICA. Es una afirmación o conjetura acerca del parámetro o parámetros de una población.

La siguiente gráfica muestra el valor critico, la región de rechazo y la región de aceptación.

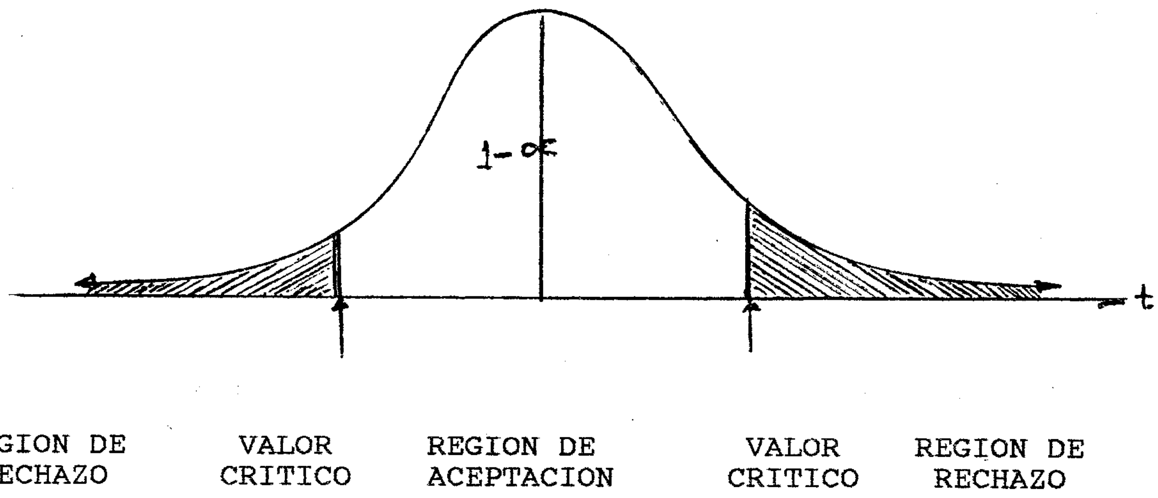
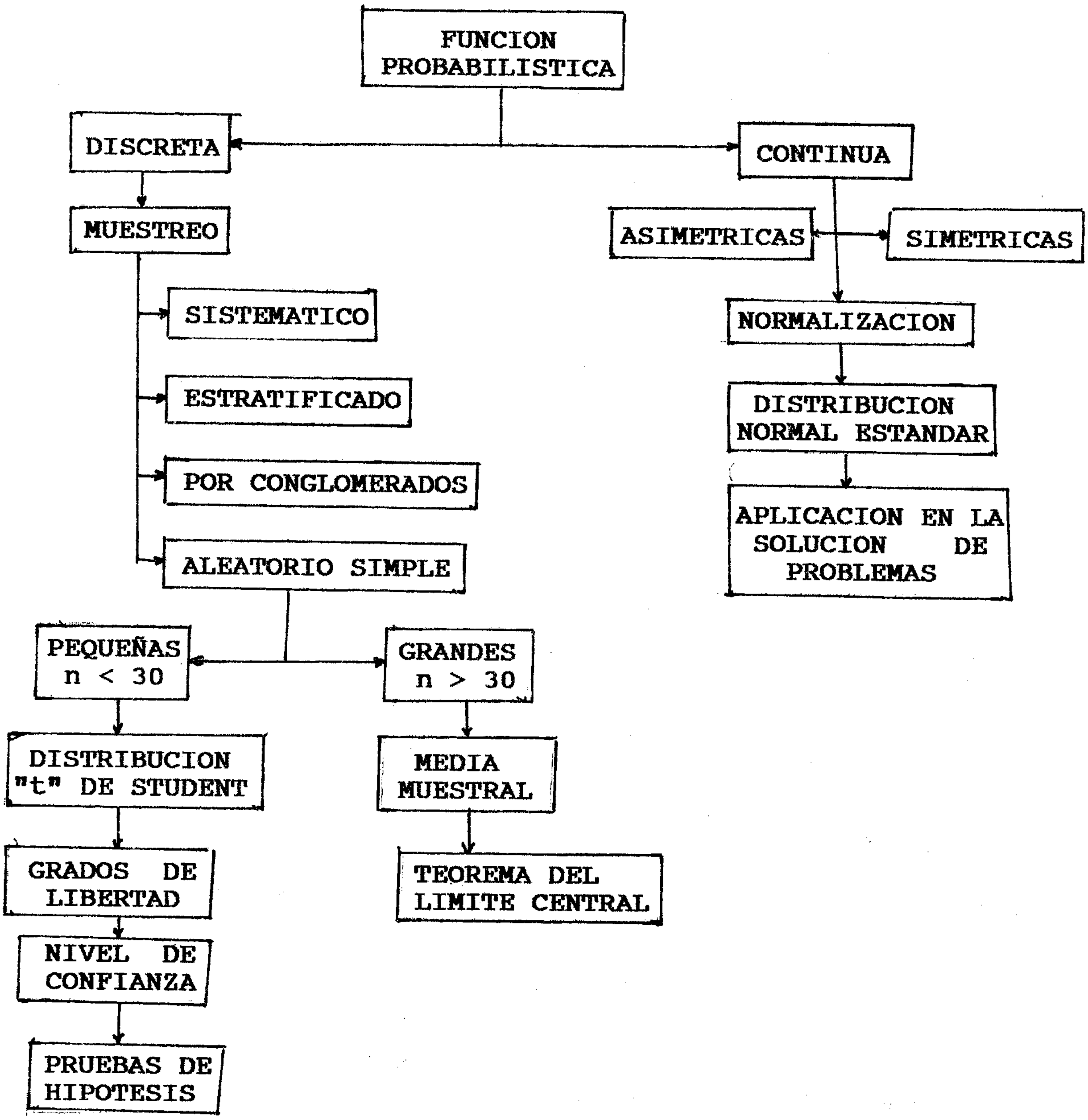


Fig. 17

Se ha aprendido a estimar la media de una población μ, dando un intervalo de confianza o acompañando la estimación de punto X con una evaluación del error posible. Ahora aprenderá còmo demostrar una hipótesis referente a la media de una población μ; es decir, se presentarán métodos para decidir si se acepta o se rechaza una afirmación acerca de un valor específico de μ.

Estos conceptos serán abordados en la siguiente unidad, esté preparado para acceder a ellos.

RECAPITULACION



ACTIVIDADES DE CONSOLIDACION

Para reafirmar los conceptos aprendidos resuelve el siguiente ejercicio. Si tienes alguna duda consulta con tu maestro o asesor.

I. La Cía. General Motor Company tiene la intención de promover a sus trabajadores a un tabulador salarial mejor que el actual y para ello aplica un examen de conocimientos culturales, habiendo obtenido las siguientes puntuaciones:

27, 28, 28, 28, 29, 30, 30, 30, 30, 31, 31, 31, 32, 32, 32,
33, 33, 33, 33, 34, 34, 34, 34, 35, 35, 35, 35, 35, 35, 35,
36, 36, 36, 36, 36, 37, 37, 37, 37, 37, 37, 38, 38, 38, 38,
38, 38, 38, 39, 39, 39, 39, 40, 40, 40, 40, 40, 40, 40, 40,
41, 41, 41, 41, 41, 41, 41, 41, 42, 42, 42, 42, 42, 42, 43,
43, 43, 43, 44, 44, 44, 44, 44, 44, 44, 45, 45, 45, 45, 45,
46, 46, 46, 47, 47, 47, 47, 47, 47, 47, 48, 48, 48, 48, 49,
49, 49, 49, 50, 50, 51, 51, 51, 52, 52, 53, 53, 53, 54, 54,
55, 56, 56, 57, 57, 58, 59, 61, 62, 62.

Determina:

1) La media.

2) La moda.

3) La mediana.

4) la varianza.

5) La desviación estándar

6) Traza el polígono de frecuencias.

7) Normaliza los datos y traza la curva de mejor ajuste sobre la gráfica anterior para contrastar el cambio.

8) Determina el tanto por ciento de casos que se espera hallar entre la media y las puntuaciones 28,38 y 60.

9) Calcula el tanto por ciento y el número de casos esperados entre los siguientes pares de puntuaciones:

a) 35 y 45
b) 50 y 55
c) 56 y 60

10) ¿Cuántos casos se espera hallar por encima de una puntuación igual a 50?. ¿Cuántos por debajo de 35?.

Problema:

II. Los datos dados a continuación corresponden a incrementos de pulso cardíaco en pulsaciones por minuto que un médico determina en relación con diez personas que realizan una tarea ardua:

27, 14, 27, 32, 25, 30, 24, 29, 19, 32.

a) Estimar el incremento promedio real del pulso cardíaco de una persona que realiza una tarea ardua, mediante el estimador puntual X.

b) Estimar el alejamiento de las pulsaciones por minuto con respecto al promedio, utilizando un estimador puntual.

c) Determinar el número de grados de libertad para la muestra dada.

d) Determinar el error máximo que se comete al estimar el incremento promedio del pulso cardíaco de una persona mediante el estimador puntual X, con un nivel de confianza del 95%.

e) Obtener el tamaño que deberá tener la muestra, de tal manera que al emplear la media X, de una muestra para estimar el incremento promedio del pulso cardíaco de una persona, se tenga un error máximo de 25 pulsaciones por minuto con un nivel de confianza de 95%.

f) Construir un intervalo con un nivel del 99% en relación con el incremento promedio real del pulso de personas que realizan la tarea dada.

LINEAMIENTOS DE AUTOEVALUACION

Los resultados que debiste obtener son los siguientes, si alguno no coincide, entonces revisa tus cálculos, localiza el error y corrígelo:

Solución del problema I.

1) = 47.1 | 2) Mo = 48.7 | 3) Md = 50.6

4) = 64.2 | 5) = 9.04

6) TABLA DE FRECUENCIAS

CLASE	X_m	Fi	Fa	XmFi	$Xm-\bar{X}$	$(Xm-\bar{X})^2$	$Fi(Xm-\bar{X})^2$
60-62	61	5	5	305	13.9	193.21	966.5
57-59	58	10	15	580	10.9	118.81	1188.1
54-56	55	15	30	825	7.9	62.1	936.2
51-53	52	18	48	936	4.9	24.01	432.2
48-50	49	20	68	980	1.9	3.61	72.2
45-47	46	17	85	782	-1.1	1.21	20.6
42-44	43	14	99	602	-4.1	16.81	235.3
39-41	40	10	109	400	-7.1	50.41	504.1
36-38	37	8	117	296	-10.1	102.01	816.1
33-35	34	6	123	204	-13.1	171.61	1029.7
30-32	31	4	127	124	-16.1	259.21	1036.8
27-29	28	3	130	84	-19.1	364.81	1094.4
				6118			8332.2

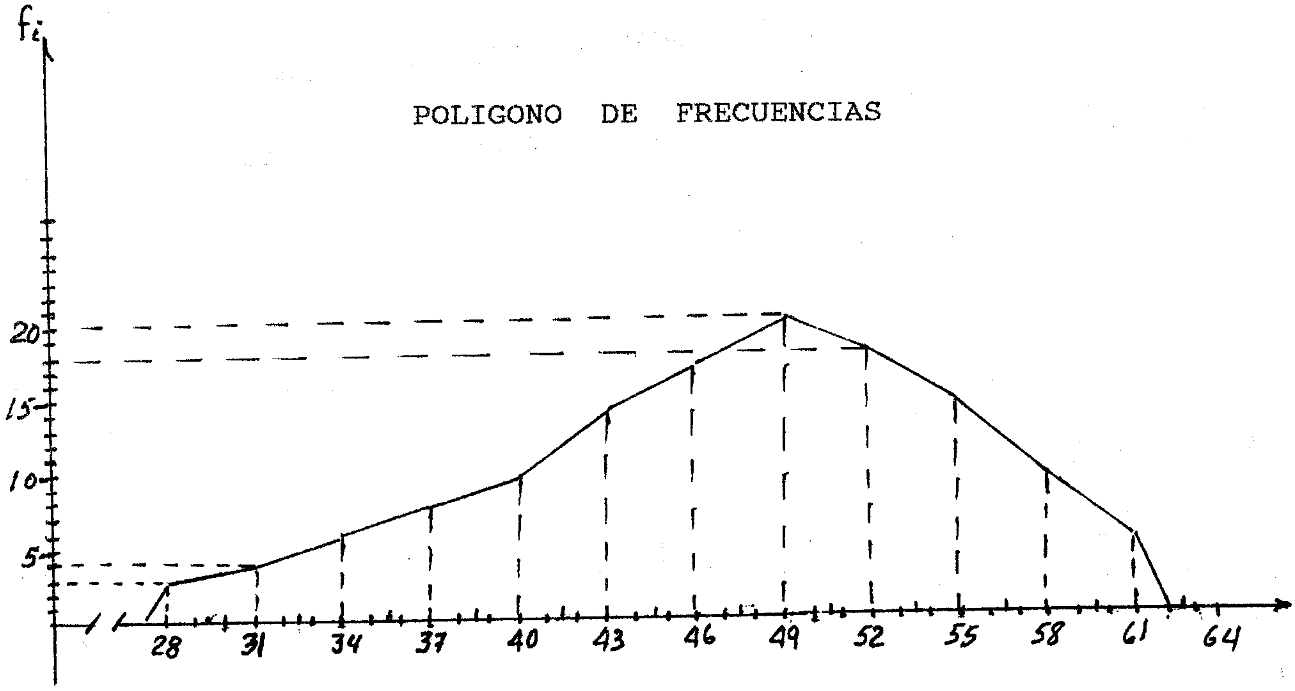
$$\bar{X} = \frac{\Sigma Xmfi}{n} = \frac{6118}{130} = 47.1$$

$$\sigma^2 = \frac{\Sigma fi(Xm-\bar{X})}{n-1} = \frac{8332.2}{130-1} = \frac{833.2}{129} = 64.6$$

$$\sigma = \sqrt{64.590698} = 8.04$$

$$Mo = 47.5 + 3\left(\frac{2}{2+3}\right) = 47.5 + \frac{6}{5} = \underline{48.7}$$

$$Md = 48 + 3\left(\frac{\frac{130}{2} - 48}{20}\right) = 48 + \frac{51}{20} = \underline{50.6}$$



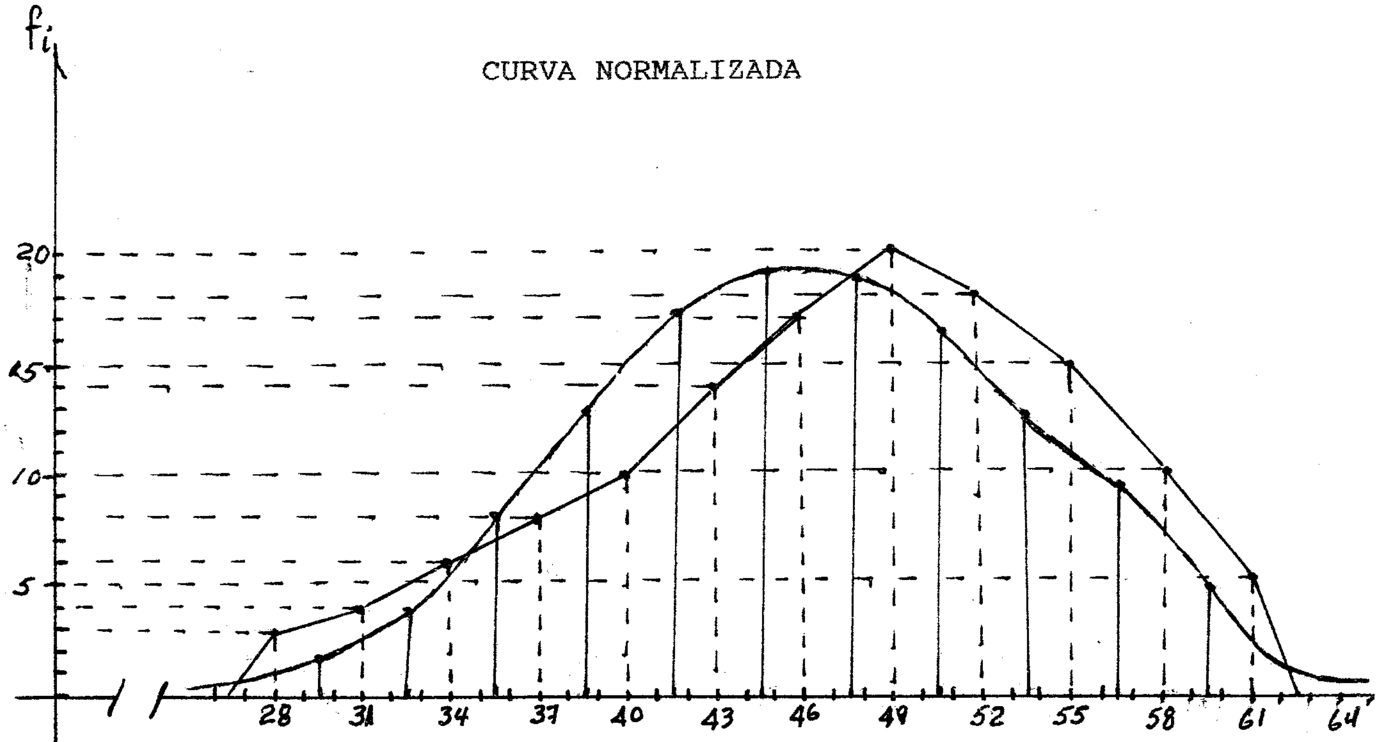
7. TABLA DE NORMALIZACION PARA EL AJUSTE DE CURVAS

CLASE	fi	Ls.	$\Delta X=X_i-\bar{X}$	$Z=\frac{\Delta X_i}{\sigma}$	DEBAJO	ENCIMA	fe	fe.red.
60-62	5	62.5	15.4	1.92	0.9726	0.0344	4.47	4.5
57-59	10	59.5	12.4	1.54	0.9382	0.0696	9.05	9.1
64-56	15	56.5	9.4	1.12	0.8686	0.0805	10.47	10.5
51-53	18	53.5	6.4	0.80	0.7881	0.1253	16.29	16.3
48-50	20	50.5	3.4	0.42	0.6628	0.1429	18.58	18.6
45-47	17	47.5	0.4	0.05	0.5199	0.1454	18.90	18.9
42-44	14	44.5	-2.6	-0.32	0.3745	0.1325	17.23	17.2
39-41	10	41.5	-5.6	-0.70	0.2420	0.0997	12.96	13.0
36-38	8	38.5	-8.6	-1.07	0.1423	0.0674	8.76	8.8
33-35	6	35.5	-11.6	-1.44	0.0749	0.0405	5.27	5.3
30-32	4	32.5	-14.6	-1.82	0.0344	0.0201	2.61	2.6
27-29	3	29.5	-17.6	-2.19	0.0143	0.0143	1.86	1.9

$\bar{X}=47.1$

$\sigma=8.04$

N= 130



A) POLIGONO DE FRECUENCIAS (A)
B) CURVA NORMALIZADA (B)

8)	1) 49%	2) 37%	3) 45%
9)	1) 33% 43 casos	2) 19.6% 26 casos	
10)	1) 47 casos	2) 9 casos	

Solución del problena II.

a) Para estimar la media μ de una población, hay varios estimadores puntuales, los más conocidos son: media, mediana y moda.

De estos estimadores el más adecuado es la media, ya que es la más confiable por considerar todos los datos de la muestra, cosa que no ocurre con la mediana y la moda.

$$\bar{X} = \frac{27+14+24+32+25+30+24+29+19+32}{10} = \frac{259}{10} = 25.9$$

$$\therefore \quad \bar{X} = 25.69$$

Esto significa que el incremento promedio real del pulso cardiaco es 25.9 pulsaciones por minuto.

Este error que se comete al estimar a través de $\bar{X}$ se determina mediante la fórmula

$$E = t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad ,$$

donde:

E= error
$t_{\alpha/2}$= área bajo la curva a la derecha de $\alpha/2$.
σ= desviación estándar de la población
n= N° de datos

NOTA: Recuerda que en ausencia de σ se puede utilizar s.

$\alpha = 0.05$ $\qquad\qquad$ $\alpha/2 = 0.025$

De las tablas de la distribución t de Student y tomando g.l.=9 se tiene:

$$t(0.025) = 2.262$$

$$\therefore \quad E = 2.262\left(\frac{5.44}{\sqrt{10}}\right) = \frac{12.305}{3.162} = 3.89$$

$$\therefore \quad E = 3.89$$

Esto significa que podemos asegurar con un grado de confianza del 95% que el error que se comete al estimar a través de x es menor de 3.89 pulsaciones por minuto.

e) Para determinar el tamaño que deberá tener la muestra con un nivel de confianza del 95% para tener un error máximo de 2.5 pulsaciones por minuto se utiliza la fórmula

$$E = t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad \text{despejando n se tiene}$$

$$n = \frac{t_{\alpha/2} \cdot \sigma}{E}$$

$$n = \frac{(2.262)(5.44)^2}{2.5} = (4.922112)^2 = 24.22$$

redondeando se tiene n=24

La confiabilidad de $\bar{X}$ como estimador de la media de la población depende del tamaño de la muestra y el tamaño de la desviación estándar de la población.

b) Para estimar el alejamiento promedio de las pulsaciones por minuto con respecto al incremento promedio real existen varios estimadores. Los más usuales son: la desviación media, varianza y desviación estándar.

De ellos utilizaremos el estimador s (desviación estándar de la muestra).

Dado que generalmente no se conoce el parámetro, que es la desviación estándar de la muestra; el estadístico s (desviación estándar de la muestra) puede servir como estimador de σ.

x	f_i	x^2	$f_i x^2$
14	1	196	196
19	1	361	361
24	1	576	576
25	1	625	625
27	2	729	1458
29	1	841	841
30	1	900	900
32	2	1024	2048
		5252	7005

Para determinar s se utiliza la fórmula

$$s = \sqrt{\frac{\Sigma fx^2}{N} - \bar{x}^2}$$

donde: N = Nº de datos
f = frecuencia de cada dato
X = media de la muestra

$$\therefore\ s = \sqrt{\frac{7005}{10} - (25\ 9)^2} = \sqrt{700.5 - 670.81} = \sqrt{29.69} = 5.44$$

$\therefore\ s = 5.44$

Esto significa que en promedio el incremento promedio del pulso cardíaco se aleja 5.44 pulsaciones por minuto de la media.

c) g.l. = n-1 $\qquad \therefore$ g.l. = 10-1 = 9

$\therefore$ g.l.= 9

d) Sabemos que al estimar la media poblacional a través de la media muestral X existe un error, es decir:

$$\mu = \bar{X} \pm \text{error muestral}$$

Esto significa que el tamaño de la muestra debe ser 24 para cometer un error menor de 2.5 pulsaciones por minuto al estimar a través de $\bar{X}$.

f) La estimación de parámetros puede ser puntual o por intervalos.

La estimación de la media poblacional por intervalos tiene la ventaja sobre la estimación puntual de que en la estimación por intervalos es posible conocer el tamaño del error, así como la precisión y el nivel de confianza, cosa que no se puede tener con la estimación puntual.

Como ya vimos al estimar en base a la media muestral X, la estimación no es perfecta, es decir, siempre hay un margen de error.

$\therefore \mu = \bar{X} \pm E$ pero $E = t\,\alpha/2 . \frac{\sigma}{\sqrt{n}}$

$\therefore \mu = \bar{X} \pm t\,\alpha/2 . \frac{\sigma}{\sqrt{n}}$ intervalo de confianza para estimar μ

donde:

α = nivel de confianza

Para un nivel de confianza del 99% se tiene que:

$\alpha = 0.01 \quad \alpha/2 = 0.005$

$\therefore$ t(0.005) = 4.032 con 9 grados de libertad

$$\therefore \quad \mu = 25.9 \pm 4.032 \left(\frac{5.44}{\sqrt{10}}\right)$$

$$\mu = 25.9 \pm \frac{21.934}{3.162}$$

$$\mu = 25.9 \pm 6.936$$

$$\therefore \quad 18.96 \leq \mu \leq 32.83$$

Esto significa que se puede asegurar con un 99% de confianza que la verdadera muestra poblacional esta entre 18.96 y 32.83 pulsaciones por minuto.

A P E N D I C E S

APENDICE A

AREAS Y ORDENADAS DE LA CURVA DE DISTRIBUCION NORMAL EN FUNCION DE $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
0.00	.0000	.5000	.5000	.3989
0.01	.0040	.5040	.4960	.3989
0.02	.0080	.5080	.4920	.3989
0.03	.0120	.5120	.4880	.3988
0.04	.0160	.5160	.4840	.3986
0.05	.0199	.5199	.4801	.3984
0.06	.0239	.5239	.4761	.3982
0.07	.0279	.5279	.4721	.3980
0.08	.0319	.5319	.4681	.3977
0.09	.0359	.5359	.4641	.3973
0.10	.0398	.5398	.4602	.3970
0.11	.0438	.5438	.4562	.3965
0.12	.0478	.5478	.4522	.3961
0.13	.0517	.5517	.4483	.3956
0.14	.0557	.5557	.4443	.3951
0.15	.0596	.5596	.4404	.3945
0.16	.0636	.5636	.4364	.3939
0.17	.0675	.5675	.4325	.3932
0.18	.0714	.5714	.4286	.3925
0.19	.0753	.5753	.4247	.3918
0.20	.0793	.5793	.4207	.3910
0.21	.0832	.5832	.4168	.3902
0.22	.0871	5871	.4129	.3894
0.23	.0910	.5910	.4090	.3885
0.24	.0948	.5948	.4052	.3876
0.25	.0987	.5987	.4013	.3867
0.26	.1026	.6026	.3974	.3857
0.27	.1064	.6064	.3936	.3847
0.28	.1103	.6103	.3897	.3836
0.29	.1141	.6141	.3859	.3825
0.30	.1179	.6179	.3821	.3814
0.31	.1217	.6217	.3783	.3802
0.32	.1255	.6255	.3745	.3790
0.33	.1293	.6293	.3707	.3778
0.34	.1331	.6331	.3669	.3765
0.35	.1368	.6368	.3632	.3752
0.36	.1406	.6406	.3594	.3739
0.37	.1443	.6443	.3557	.3725
0.38	.1480	.6480	.3520	.3712
0.39	.1517	.6517	.3483	.3697
0.40	.1554	.6554	.3446	.3683
0.41	.1591	.6591	.3409	.3668
0.42	.1628	.6628	.3372	.3653
0.43	.1664	.6664	.3336	.3637
0.44	.1700	.6700	.3300	.3621

APENDICE A

AREAS Y ORDENADAS DE LA CURVA DE DISTRIBUCION NORMAL EN FUNCION DE $\Delta x/\sigma$

(1) Z. Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
0.45	.1736	.6736	.3264	.3605
0.46	.1772	.6772	.3228	.3589
0.47	.1808	.6808	.3192	.3572
0.48	.1844	.6844	.3156	.3555
0.49	.1879	.6879	.3121	.3538
0.50	.1915	.6915	.3085	.3521
0.51	.1950	.6950	.3050	.3503
0.52	.1985	.6985	.3015	.3485
0.53	.2019	.7019	.2981	.3467
0.54	.2054	.7054	.2946	.3448
0.55	.2088	.7088	.2912	.3429
0.56	.2123	.7123	.2877	.3410
0.57	.2157	.7157	.2843	.3391
0.58	.2190	.7190	.2810	.3372
0.59	.2224	.7224	.2776	.3352
0.60	.2257	.7257	.2743	.3332
0.61	.2291	.7291	.2709	.3312
0.62	.2324	.7324	.2676	.3292
0.63	.2357	.7357	.2643	.3271
0.64	.2389	.7389	.2611	.3251
0.65	.2422	.7422	.2578	.3230
0.66	.2454	.7454	.2546	.3209
0.67	.2486	.7486	.2514	.3187
0.68	.2517	.7517	.2483	.3166
0.69	.2549	.7549	.2451	.3144
0.70	.2580	.7580	.2420	.3123
0.71	.2611	.7611	.2389	.3101
0.72	.2642	.7642	.2358	.3079
0.73	.2673	.7673	.2327	.3056
0.74	.2704	.7704	.2296	.3034
0.75	.2734	.7734	.2266	.3011
0.76	.2764	.7764	.2236	.2989
0.77	.2794	.7794	.2206	.2966
0.78	.2823	.7823	.2177	.2943
0.79	.2852	.7852	.2148	.2920
0.80	.2881	.7881	.2119	.2897
0.81	.2910	.7910	.2090	.2874
0.82	.2939	.7939	.2061	.2850
0.83	.2967	.7967	.2033	.2827
0.84	.2995	.7995	.2005	.2803
0.85	.3023	.8023	.1977	.2780
0.86	.3051	.8051	.1949	.2756
0.87	.3078	.8078	.1922	.2732
0.88	.3106	.8106	.1894	.2709
0.89	.3133	.8133	.1867	2685

APENDICE A

AREAS Y ORDENADAS DE LA CURVA DE DISTRIBUCION NORMAL EN FUNCION DE $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0.93	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
0.98	.3365	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444
1.00	.3413	.8413	.1587	.2420
1.01	.3438	.8438	.1562	.2396
1.02	.3461	.8461	.1539	.2371
1.03	.3485	.8485	.1515	.2347
1.04	.3508	.8508	.1492	.2323
1.05	.3531	.8531	.1469	.2299
1.06	.3554	.8554	.1446	.2275
1.07	.3577	.8577	.1423	.2251
1.08	.3599	.8599	.1401	.2227
1.09	.3621	.8621	.1379	.2203
1.10	.3643	.8643	.1357	.2179
1.11	.3665	.8665	.1335	.2155
1.12	.3686	.8686	.1314	.2131
1.13	.3708	.8708	.1292	.2107
1.14	.3729	.8729	.1271	.2083
1.15	.3749	.8749	.1251	.2059
1.16	.3770	.8770	.1230	.2036
1.17	.3790	.8790	.1210	.2012
1.18	.3810	.8810	.1190	.1989
1.19	.3830	.8830	.1170	.1965
1.20	.3849	.8849	.1151	.1942
1.21	.3869	.8869	.1131	.1919
1.22	.3888	.8888	.1112	.1895
1.23	.3907	.8907	.1093	.1872
1.24	.3925	.8925	.1075	.1849
1.25	.3944	.8944	.1056	.1826
1.26	.3962	.8962	.1038	.1804
1.27	.3980	.8980	.1020	.1781
1.28	.3997	.8997	.1003	.1758
1.29	.4015	.9015	.0985	.1736
1.30	.4032	.9032	.0968	.1714
1.31	.4049	.9049	.0951	.1691
1.32	.4066	.9066	.0934	.1669
1.33	.4082	.9082	.0918	.1647
1.34	.4099	.9099	.0901	.1626

APENDICE A

AREAS Y ORDENADAS DE LA CURVA DE DISTRIBUCION NORMAL EN FUNCION DE $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
1.35	.4115	.9115	.0885	.1604
1.36	.4131	.9131	.0869	.1582
1.37	.4147	.9147	.0853	.1561
1.38	.4162	.9162	.0838	.1539
1.39	.4177	.9177	.0823	.1518
1.40	.4192	.9192	.0808	.1497
1.41	.4207	.9207	.0793	.1476
1.42	.4222	.9222	.0778	.1456
1.43	.4236	.9236	.0764	.1435
1.44	.4251	.9251	.0749	.1415
1.45	.4265	.9265	.0735	.1394
1.46	.4279	.9279	.0721	.1374
1.47	.4292	.9292	.0708	.1354
1.48	.4306	.9306	.0694	.1334
1.49	.4319	.9319	.0681	.1315
1.50	.4332	.9332	.0668	.1295
1.51	.4345	.9345	.0655	.1276
1.52	.4357	.9357	.0643	.1257
1.53	.4370	9370	.0630	.1238
1.54	.4382	.9382	.0618	.1219
1.55	.4394	.9394	.0606	.1200
1.56	.4406	.9406	.0594	.1182
1.57	.4418	.9418	.0582	.1163
1.58	.4429	.9429	.0571	.1145
1.59	.4441	.9441	.0559	.1127
1.60	.4452	.9452	.0548	.1109
1.61	.4463	.9463	.0537	.1092
1.62	.4474	.9474	.0526	.1074
1.63	.4484	.9484	.0516	.1057
1.64	.4495	.9495	.0505	.1040
1.65	.4505	.9505	.0495	.1023
1.66	.4515	.9515	.0485	.1006
1.67	.4525	.9525	.0475	.0989
1.68	.4535	.9535	.0465	.0973
1.69	.4545	.9545	.0455	.0957
1.70	.4554	.9554	.0446	.0940
1.71	.4564	.9564	.0436	.0925
1.72	.4573	.9573	.0427	.0909
1.73	.4582	9582	.0418	.0893
1.74	.4591	.9591	.0409	.0878
1.75	.4599	.9599	.0401	.0863
1.76	.4608	.9608	.0392	.0848
1.77	.4616	.9616	.0384	.0833
1.78	.4625	.9625	.0375	.0818
1.79	.4633	.9633	.0367	.0804

APENDICE A

AREAS Y ORDENADAS DE LA CURVA DE DISTRIBUCION NORMAL EN FUNCION DE $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
1.80	.4641	.9641	.0359	.0790
1.81	.4649	.9649	.0351	.0775
1.82	.4656	.9656	.0344	.0761
1.83	.4664	.9664	.0336	.0748
1.84	.4671	.9671	.0329	.0734
1.85	.4648	.9678	.0322	.0721
1.86	.4686	.9686	.0314	.0707
1.87	.4693	.9693	.0307	.0694
1.88	.4699	.9699	.0301	.0681
1.89	.4706	.9706	.0294	.0669
1.90	.4713	.9713	.0287	.0656
1.91	.4719	.9719	.0281	.0644
1.92	.4726	.9726	.0274	.0632
1.93	.4732	.9732	.0268	.0620
1.94	.4738	.9738	.0262	.0608
1.95	.4744	.9744	.0256	.0596
1.96	.4750	.9750	.0250	.0584
1.97	.4756	.9756	.0244	.0573
1.98	.4761	.9761	.0239	.0562
1.99	.4767	.9767	.0233	.0551
2.00	.4772	.9772	.0228	.0540
2.01	.4778	.9778	.0222	.0529
2.02	.4783	.9783	.0217	.0519
2.03	.4788	.9788	.0212	.0508
2.04	.4793	.9793	.0207	.0498
2.05	.4798	.9798	.0202	.0488
2.06	.4803	.9803	.0197	.0478
2.07	.4808	.9808	.0192	.0468
2.08	.4812	.9812	.0188	.0459
2.09	.4817	.9817	.0183	.0449
2.10	.4821	.9821	.0179	.0440
2.11	.4826	.9826	.0174	.0431
2.12	.4830	.9830	.0170	.0422
2.13	.4834	.9834	.0166	.0413
2.14	.4838	.9838	.0162	.0404
2.15	.4842	.9842	.0158	.0396
2.16	.4846	.9846	.0154	.0387
2.17	.4850	.9850	.0150	.0379
2.18	.4854	.9854	.0146	.0371
2.19	.4857	.9857	.0143	.0363
2.20	.4861	.9861	.0139	.0355
2.21	.4864	.9864	.0136	.0347
2.22	.4868	.9868	.0132	.0339
2.23	.4871	.9871	.0129	.0332
2.24	.4875	.9875	.0125	.0325

APENDICE A

AREAS Y ORDENADAS DE LA CURVA DE DISTRIBUCION NORMAL EN FUNCION DE $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
2.32	.4898	.9898	.0102	.0270
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.0055	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
2.57	.4949	.9949	.0051	.0147
2.58	.4951	.9951	.0049	.0143
2.59	.4952	.9952	.0048	.0139
2.60	.4953	.9953	.0047	.0136
2.61	.4955	.9955	.0045	.0132
2.62	.4956	.9956	.0044	.0129
2.63	.4957	.9957	.0043	.0126
2.64	.4959	.9959	.0041	.0122
2.65	.4960	.9960	.0040	.0119
2.66	.4961	.9961	.0039	.0116
2.67	.4962	.9962	.0038	.0113
2.68	.4963	.9963	.0037	.0110
2.69	.4964	.9964	.0036	.0107

APENDICE A

Areas y Ordenadas de La Curva de Distribucion Normal en Funcion de $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
2.70	.4965	.9965	.0035	.0104
2.71	.4966	.9966	.0034	.0101
2.72	.4967	.9967	.0033	.0099
2.73	.4968	.9968	.0032	.0096
2.74	.4969	.9969	.0031	.0093
2.75	.4970	.9970	.0030	.0091
2.76	.4971	.9971	.0029	.0088
2.77	.4972	.9972	.0028	.0086
2.78	.4973	.9973	.0027	.0084
2.79	.4974	.9974	.0026	.0081
2.80	.4974	.9974	.0026	.0079
2.81	.4975	.9975	.0025	.0077
2.82	.4976	.9976	.0024	.0075
2.83	.4977	.9977	.0023	.0073
2.84	.4977	.9977	.0023	.0071
2.85	.4978	.9978	.0022	.0069
2.86	.4979	.9979	.0021	.0067
2.87	.4979	.9979	.0021	.0065
2.88	.4980	.9980	.0020	.0063
2.89	.4981	.9981	.0019	.0061
2.90	.4981	.9981	.0019	.0060
2.91	.4982	.9982	.0018	.0058
2.92	.4982	.9982	.0018	.0056
2.93	.4983	.9983	.0017	.0055
2.94	.4984	.9984	.0016	.0053
2.95	.4984	.9984	.0016	.0051
2.96	.4985	.9985	.0015	.0050
2.97	.4985	.9985	.0015	.0048
2.98	.4986	.9986	.0014	.0047
2.99	.4986	.9986	.0014	.0046
3.00	.4987	.9987	.0013	.0044
3.01	.4987	.9987	.0013	.0043
3.02	.4987	.9987	.0013	.0042
3.03	.4988	.9988	.0012	.0040
3.04	.4988	.9988	.0012	.0039
3.05	.4989	.9989	.0011	.0038
3.06	.4989	.9989	.0011	.0037
3.07	.4989	.9989	.0011	.0036
3.08	.4990	.9990	.0010	.0035
3.09	.4990	.9990	.0010	.0034
3.10	.4990	.9990	.0010	.0033
3.11	.4991	.9991	.0009	.0032
3.12	.4991	.9991	.0009	.0031
3.13	.4991	.9991	.0009	.0030
3.14	.4992	.9992	.0008	.0029

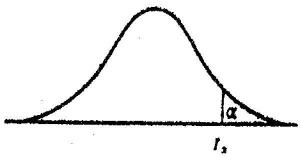
APENDICE A

Areas y Ordenadas de La Curva de Distribucion Normal en Funcion de $\Delta x/\sigma$

(1) Z Puntuacion tipificada $\frac{\Delta x}{\sigma}$	(2) A Area desde la media a $\frac{\Delta x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) Y Ordenada en $\frac{\Delta x}{\sigma}$
3.15	.4992	.9992	.0008	.0028
3.16	.4992	.9992	.0008	.0027
3.17	.4992	.9992	.0008	.0026
3.18	.4993	.9993	.0007	.0025
3.19	.4993	.9993	.0007	.0025
3.20	.4993	.9993	.0007	.0024
3.21	.4993	.9993	.0007	.0023
3.22	.4994	.9994	.0006	.0022
3.23	.4994	.9994	.0006	.0022
3.24	.4994	.9994	.0006	.0021
3.30	.4995	.9995	.0005	.0017
3.40	.4997	.9997	.0003	.0012
3.50	.4998	.9998	.0002	.0009
3.60	.4998	.9998	.0002	.0006
3.70	.4999	.9999	.0001	.0004

APENDICE B

Tabla . Valores críticos de t



n	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	*d.f.*
2	3.078	6.314	12.706	31.821	63.657	1
3	1.886	2.920	4.303	6.965	9.925	2
4	1.638	2.353	3.182	4.541	5.841	3
5	1.533	2.132	2.776	3.747	4.604	4
6	1.476	2.015	2.571	3.365	4.032	5
7	1.440	1.943	2.447	3.143	3.707	6
8	1.415	1.895	2.365	2.998	3.499	7
9	1.397	1.860	2.306	2.896	3.355	8
10	1.383	1.833	2.262	2.821	3.250	9
11	1.372	1.812	2.228	2.764	3.169	10
12	1.363	1.796	2.201	2.718	3.106	11
13	1.356	1.782	2.179	2.681	3.055	12
14	1.350	1.771	2.160	2.650	3.012	13
15	1.345	1.761	2.145	2.624	2.977	14
16	1.341	1.753	2.131	2.602	2.947	15
17	1.337	1.746	2.120	2.583	2.921	16
18	1.333	1.740	2.110	2.567	2.898	17
19	1.330	1.734	2.101	2.552	2.878	18
20	1.328	1.729	2.093	2.539	2.861	19
21	1.325	1.725	2.086	2.528	2.845	20
22	1.323	1.721	2.080	2.518	2.831	21
23	1.321	1.717	2.074	2.508	2.819	22
24	1.319	1.714	2.069	2.500	2.807	23
25	1.318	1.711	2.064	2.492	2.797	24
26	1.316	1.708	2.060	2.485	2.787	25
27	1.315	1.706	2.056	2.479	2.779	26
28	1.314	1.703	2.052	2.473	2.771	27
29	1.313	1.701	2.048	2.467	2.763	28
30	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.