



GOBIERNO DEL
ESTADO DE MÉXICO

TES OEM
TECNOLÓGICO DE ESTUDIOS SUPERIORES
ORIENTE DEL ESTADO DE MÉXICO

Secretaría de Educación, Cultura y Bienestar Social
Subsecretaría de Educación Media Superior y Superior
Tecnológico de Estudios Superiores del Oriente del Estado de México
Organismo Público Descentralizado del Gobierno del Estado de México

CUADERNO DE EJERCICIOS
ESTADISTICA ADMINISTRATIVA I
(SEGUNDO SEMESTRE)
CONTADOR PÚBLICO

Elaboró: LAE Carlos Gutiérrez Reynaga

NOVIEMBRE 2011

INDICE

INTRODUCCIÓN	6
PROPÓSITO	7
COMPETENCIAS A DESARROLLAR.....	7
METODOLOGÍA DE TRABAJO	8
UNIDAD 1 DISTRIBUCIONES DE FRECUENCIA	9
1.1 RECOPIACIÓN DE DATOS	10
1.2 DISTRIBUCIÓN DE FRECUENCIAS HISTOGRAMAS, POLÍGONOS DE FRECUENCIA, Y OJIVAS	10
1.2.1 REPRESENTACIÓN GRAFICA DE LOS DATOS.....	12
HISTOGRAMA.....	12
1.2.1 POLÍGONOS DE FRECUENCIA.	16
1.3 MEDIDAS DE TENDENCIA CENTRAL Y DE VARIABILIDAD PARA UN CONJUNTO DE DATOS NO AGRUPADOS.	17
1.4 MEDIDAS DE DISPERSIÓN.....	18
MEDIDAS DE TENDENCIA CENTRAL Y DE VARIABILIDAD EN DATOS AGRUPADOS	21
MEDIDAS DE TENDENCIA CENTRAL DATOS AGRUPADOS	22
COEFICIENTE DE VARIACION.	23
COEFICIENTE DE VARIACIÓN PEARSON.....	26
UNIDAD 2 INTRODUCCIÓN A LA PROBABILIDAD.....	28
2.1 EVENTOS MUTUAMENTE EXCLUYENTES Y NO EXCLUYENTES.....	29
2.2 REGLAS DE ADICIÓN.....	29
2.3 EVENTOS INDEPENDIENTES, DEPENDIENTES, PROBABILIDAD CONDICIONAL	30
2.3 PROBABILIDAD CONDICIONAL.....	32
2.4 REGLAS DE MULTIPLICACIÓN.....	33
2.5 DIAGRAMAS DE ÁRBOL	33
2.6 COMBINACIONES Y PERMUTACIONES.....	39
2.6 COMBINACIONES	40

UNIDAD 3. TIPOS DE DISTRIBUCIONES VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS	42
3.1 DISTRIBUCIÓN BINOMIAL.....	45
3.2 MODELO DE POISSON	47
3.3 DISTRIBUCIÓN HIPERGEOMÉTRICA DE PROBABILIDAD.	50
3.5 MODELO NORMAL.....	51
UNIDAD 4. MUESTREO Y ESTIMACIONES.....	55
4.1 DISTRIBUCIÓN MUESTRAL DE LA MEDIA.....	56
4.2 DISTRIBUCIÓN MUESTRAL DE LA DIFERENCIA ENTRE DOS MEDIAS.....	59
4.3 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA DE LA POBLACIÓN.....	61
4.4 INTERVALOS DE CONFIANZA PARA LA MEDIA, CON EL USO DE LA DISTRIBUCIÓN NORMAL Y “T” DE STUDENT.	64
4.5 INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS MEDIAS, CON EL USO DE LA DISTRIBUCIÓN NORMAL Y “t” DE STUDENT.	66
4.6 UNA SOLA MUESTRA: ESTIMACIÓN DE LA PROPORCIÓN.....	67
4.8 TAMAÑO DE LA MUESTRA COMO UNA ESTIMACIÓN DE P Y UN GRADO DE CONFIANZA $(1 - \alpha)$ 100%	70
UNIDAD 5. PRUEBA DE HIPÓTESIS.....	73
5.2 ERROR TIPO UNO I Y TIPO II EN PRUEBAS DE HIPÓTESIS.....	76
5.3 PRUEBAS UNILATERALES Y BILATERALES	79
5.4. PRUEBA DE UNA HIPÓTESIS: REFERENTE A LA MEDIA CON VARIANZA DESCONOCIDA UTILIZANDO LA DISTRIBUCIÓN NORMAL Y “t” DE STUDENT.	84
5.5. DOS MUESTRAS: PRUEBAS SOBRE MEDIAS UTILIZANDO LA DISTRIBUCIÓN NORMAL Y “t” DE STUDENT.....	86
5.6 UNA MUESTRA PRUEBA SOBRE UNA SOLA PROPORCIÓN.....	89
5.7 DOS MUESTRAS: PRUEBA SOBRE DOS PROPORCIONES	90
5.8. DOS MUESTRAS: PRUEBAS PAREADAS.....	92

TEMARIO

I. DISTRIBUCIONES DE FRECUENCIA

- 1.1 Recopilación de datos.
- 1.2 Distribución de frecuencia.
 - 1.2.1 Histogramas, polígonos de frecuencia, ojivas.
- 1.3 Medidas de tendencia central para un conjunto de datos no agrupados y datos agrupados.
 - 1.3.1 Media.
 - 1.3.2 Mediana.
 - 1.3.3 Moda.
- 1.4 Medidas de dispersión para un conjunto de datos agrupados y datos no agrupados.
 - 1.4.1 Rango.
 - 1.4.2 Varianza.
 - 1.4.3 Desviación estándar.

II. INTRODUCCIÓN A LA PROBABILIDAD

- 2.1 Eventos mutuamente excluyentes y no excluyentes
- 2.2 Reglas de adición
- 2.3 Eventos independientes, dependientes, probabilidad condicional
- 2.4 Reglas de multiplicación
- 2.5 Diagrama de árbol
- 2.6 Combinaciones y permutaciones

III. TIPOS DE DISTRIBUCIONES VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS

- 3.1 Binomial
- 3.2 Poisson
- 3.3 Hipergeométrica
- 3.4 Propiedades: media, varianza y desviación estándar
- 3.5 Normal

IV. MUESTREO Y ESTIMACIONES

- 4.1 Distribución muestral de la media
- 4.2 Distribución muestral de la diferencia entre dos medias
- 4.3 Determinación del tamaño de la muestra de una población.
- 4.4 Intervalos de confianza para la media, con el uso de la distribución Normal y "t" de student
- 4.5 Intervalos de confianza para la diferencia entre dos medias $\mu_1 - \mu_2$ con σ_1 y σ_2 , $\sigma_1 = \sigma_2$ pero conocidas, con el uso de la distribución normal y la "t" de student.
- 4.6 Una sola muestra: estimación de la proporción
- 4.7 Tamaño de la muestra como una estimación de P y un grado de confianza $(1-\alpha)$ 100%.

V. PRUEBA DE HIPÓTESIS

- 5.1 Hipótesis estadísticas.
- 5.2 Errores tipo I y II
- 5.3 Pruebas unilaterales y bilaterales
- 5.4 Prueba de una hipótesis: referente a la media con varianza desconocida utilizando la distribución normal y “t” student.
- 5.5 Dos muestras: pruebas sobre dos medias utilizando la distribución normal y “t” student.
- 5.6 Una muestra prueba sobre una sola proporción.
- 5.7 Dos muestras: prueba sobre dos proporciones.
- 5.8 Dos muestras: pruebas pareadas

Para facilitar el uso de este cuaderno de ejercicios se ha organizado su contenido empleando los siguientes símbolos de apoyo:



Identificación general del tema



Introducción del tema



Exposición del tema



Resumen del Tema



Recordar ó analizar la información para obtener sus propias conclusiones



Ejemplo del tema



Actividad, práctica o ejercicio sugerido: desarrollar la actividad indicada, realizar un procedimiento específico ó seguir detalladamente una secuencia de pasos.



Recomendación para fortalecer el aprendizaje del tema o subtema, notas importantes o tips.

INTRODUCCIÓN

En un mundo cada vez más globalizado en las áreas comerciales, financieras, tecnológicas y científicas, y donde invariablemente el flujo de información es mayor a cada momento, se hace indispensable no sólo la correcta descripción de los datos sino también su análisis e interpretación.

Es aquí donde la estadística juega un papel importantísimo, al ser esta una de las áreas del conocimiento que permite analizar la variabilidad que generalmente acompaña a los datos observados, y por ello se constituye como una herramienta que el Contador Público puede utilizar para la adecuada toma de decisiones.

Estadística Administrativa I tiene varios propósitos, pues pretende despertar en el estudiante de contaduría el interés por la investigación para la toma de decisiones, la solución de problemas y el análisis de situaciones y eventos relacionados con el entorno académico, profesional, personal y social, rigiéndose en todo momento por un código de ética profesional y personal.

Los propósitos de la asignatura en relación a la carrera de Contador Público son que el estudiante:

1. Participe en el desarrollo de investigaciones y proyectos para la solución de problemas relacionados con la administración y contaduría.
2. Adquiera la capacidad de lectura e interpretación de tablas y gráficos estadísticos para facilitar la realización de actividades administrativas.
3. Comprenda el papel que tiene de la estadística en la toma de decisiones racional y el modo en que ha contribuido al desarrollo de la sociedad.
4. Identifique, dentro del contexto empresarial, la importancia y utilidad de los análisis estadísticos para la toma de decisiones.
5. Manifieste una actitud crítica y analítica en la solución de problemas.

Esta asignatura pone especial énfasis en el enfoque práctico, tratando siempre de relacionar los conceptos, técnicas y casos de estudio con el quehacer cotidiano de la administración de una organización, esperando despertar en los estudiantes el deseo de adentrarse cada vez más a la teoría de la probabilidad y estadística, al ver lo importante que resulta su utilización en el ámbito contable y financiero.

Este cuaderno de ejercicios tratará cinco temas fundamentales para que el alumno se introduzca al estudio básico de la estadística, en el primer capítulo se abordan

ejercicios elementales de la estadística descriptiva, en el segundo; ejemplos de probabilidad y valor esperado como una medida del riesgo frente a la incertidumbre en experimentos aleatorios; en la tercera parte se realizan ejercicios de los tipos de distribuciones aleatorias discretas y continuas; el capítulo cuarto trata del muestreo y las estimaciones puntuales y por intervalo, finalmente en el capítulo quinto se abordará la prueba de hipótesis que permitirá al alumno llevar a cabo la toma de decisiones de forma racional.

PROPÓSITO

El cuaderno de ejercicios de estadística administrativa I tiene como propósito introducir al estudiante con los conceptos y técnicas básicas de la estadística aplicada a la administración y economía. El cuadernillo tiene un nivel matemático elemental, con la intención de que el estudiante comprenda la metodología y su aplicación, y no tanto la teoría matemática detrás de ella.

COMPETENCIAS A DESARROLLAR

Competencia general:

El estudiante analiza y aplica conceptos y técnicas de la probabilidad y estadística descriptiva e inferencial en la solución de problemas en el área de su competencia.

Competencias específicas:

- ✍ Aplica las fórmulas de tendencia central y de la variabilidad de datos para analizar información, relativos a datos agrupados y no agrupados y tomar decisiones.
- ✍ Aplica el concepto de valor esperado o esperanza matemática para la toma de decisiones.
- ✍ Cita ejemplos de aplicación de variables aleatorias discretas y continuas.
- ✍ Grafica una distribución de probabilidad continua y discreta.
- ✍ Aplica los tipos de distribución de variables aleatorias discretas como: binomial, Poisson, e hipergeométrica para la solución de Problemas relativos a la administración.
- ✍ Aplica los tipos de distribución de variables aleatorias continuas como: normal y aproximación de la normal a la binomial, para la toma de decisiones.
- ✍ Consulta y explica los diferentes tipos de muestreo: aleatorio, sistemático, estratificado y conglomerados.
- ✍ Aplica los métodos de muestreo para recopilación de la información que permita estimar las características poblacionales desconocidas,

examinando la información obtenida de una muestra, de una población.

- ✍ Aplica las fórmulas de tendencia central para la solución de problemas en la toma de decisiones.
- ✍ Utiliza el teorema de límite central para la solución de problemas de una muestra y la diferencia entre dos muestras cuando $\sigma^2_1 = \sigma^2_2$ es conocida.
- ✍ Utiliza la distribución z y “t” de student para hacer estimaciones de intervalo de la diferencia de dos muestras.
- ✍ Calcula intervalos de confianza para diferencia de proporciones y pruebas en aplicaciones que involucran poblaciones de datos cualitativos que deben compararse utilizando proporciones o porcentajes.
- ✍ Diferencia las variables aleatorias discretas y continuas.
- ✍ Realiza pruebas de hipótesis que conduzca a una decisión sobre una hipótesis en particular acerca de una población.

METODOLOGÍA DE TRABAJO

Para el logro de los objetivos que persigue este cuaderno de prácticas y que permitirán al alumno alcanzar la competencia, es fundamental que los procedimientos presentados se ejerciten todo el tiempo, esperamos que los contenidos no sólo se comprendan sino que se apliquen en la solución de problemas que tengan que ver con situaciones que los estudiantes pueden enfrentar en su trayectoria académica y profesional.

Por lo anterior, la estrategia metodológica de enseñanza-aprendizaje es, por un lado, el planteamiento de ejercicios y problemas, de los temas fundamentales para introducir al estudiante al estudio de la estadística y que se abordan durante el curso, esto con el objeto de que los estudiantes se ejerciten en el uso, aplicación y manejo de fórmulas y contenidos procedimentales. Por otro lado, el docente de la asignatura tendrá que orientar la aplicación de cada uno de estos ejercicios a las áreas específicas de interés de los estudiantes; es decir, el docente tendrá que ejemplificar y presentar casos y situaciones aplicables a la contaduría, que complementen los ejercicios que se están planteando.

El alumno en este esfuerzo, deberá llevar a cabo estrategias de estudio que propicien un aprendizaje verdaderamente significativo, teniendo la comprensión del contenido y relacionando éste con sus conocimientos previos, así como con sus áreas específicas de estudio, a través del estudio casos y problemas relacionados con el quehacer cotidiano donde puedan aplicar y ejercitar lo aprendido.



UNIDAD 1 DISTRIBUCIONES DE FRECUENCIA

Propósitos de la unidad

En esta unidad el alumno debe:

- ✍ Reconocer la utilidad e importancia de las medidas de tendencia central para un conjunto de datos agrupados y no agrupados.
- ✍ Identificar las operaciones que se utilizan en estadística descriptiva.
- ✍ Organizar datos en diferentes tipos de tablas y elaborar varios tipos de gráficas.
- ✍ Aplicar las fórmulas para obtener medidas de descripción de datos.

Competencia específica

Desarrolla la capacidad del razonamiento matemático utilizando las herramientas básicas de la estadística descriptiva.

Aplica los métodos de muestreo para recopilación de la información.

Aplica las fórmulas de tendencia central para la solución de problemas en la toma de decisiones.

Aplica las fórmulas de la variabilidad de datos para analizar información, relativos a datos agrupados y no agrupados para la toma de decisiones.

Aplica los parámetros de la estadística descriptiva para la representación gráfica y numérica de un conjunto de datos a través de muestras aleatorias simples.

Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.



INTRODUCCIÓN

La palabra estadística a menudo se refiere a gráficas y tablas; cifras relativas a nacimientos, muertes, impuestos, demografía, ingresos, deudas, créditos, etc. No obstante, para entender el análisis estadístico como herramienta de análisis, es necesario comprender qué representa cada concepto y la metodología mediante la cual se obtiene un dato estadístico.

Existen dos grandes divisiones de la estadística: la que se dedica a la recolección, presentación y categorización de datos, llamada estadística descriptiva, y la que se dedica a realizar inferencia en base a dichos datos, llamada estadística inferencial. Para desarrollar la capacidad del razonamiento matemático es recomendable utilizar las herramientas básicas de la estadística descriptiva para muestrear, procesar y comunicar información social y científica, para la toma de decisiones en la vida cotidiana, en un clima de colaboración y respeto



1.1 RECOPIACIÓN DE DATOS

Al recoger datos relativos a las características de un grupo de individuos u objetos, suele ser imposible o nada práctico observar todo el grupo, en especial si es muy grande. En vez de examinar el grupo entero, llamado población o universo, se examina una pequeña parte del grupo, llamada muestra. Una población puede ser finita o infinita. Por ejemplo, la población consistente en todas las tuercas producidas por una fábrica un cierto día es finita, mientras que la determinada por todos los posibles resultados (águila, sol) de sucesivas tiradas de una moneda, es infinita. Si una muestra es representativa de una población, es posible inferir importantes conclusiones sobre las poblaciones a partir del análisis de la muestra.



1.2 DISTRIBUCIÓN DE FRECUENCIAS HISTOGRAMAS, POLÍGONOS DE FRECUENCIA, Y OJIVAS



Ejemplo de distribución y construcción de tabla de frecuencias

La empresa Casa S.A presenta los siguientes datos:

35	24	26	23	50	20	25	56	30	30
38	36	35	29	28	30	40	39	38	40
27	24	30	32	35	27	29	22	28	27
48	40	48	31	39	28	36	37	52	44
49	52	41	31	31	56	58	38	26	25
24	60	55	48	37	31	30	22	20	46

Se pide distribuir y construir la tabla de frecuencias

Paso 1. Calcular el rango: Para esto, se identifica el número mayor y el número menor en los datos. El rango es el resultado de la resta del valor mayor y el menor, esto es: $R = 60 - 20 = 38$

Paso 2. Determinar el número de intervalos que se desea tener: Siguiendo con la tabla del ejercicio vamos a construir 8 intervalos. Entonces decimos que $K = 8$

Paso 3. Obtener la amplitud de intervalo: Dividir el rango entre el número de clases. $A = \frac{R}{K}$

$$A = \frac{40}{8} = 5$$

Paso 4. Se forman los intervalos: Los intervalos se forman comenzando con el valor menor se le suma la amplitud:

INTERVALOS:

20 a 25 (se cuenta 5 desde 20 hasta 25)

26 a 31

32 a 37

38 a 43

44 a 49

50 a 55

56 a 61

62 a 67

Nota: No importa que el último intervalo exceda el último dato.

Paso5. Se calcula la marca de clase (Mc)

$$Mc = \frac{(Li+Ls)}{2} \quad Mc = \frac{(20+25)}{2} = 22.5 \quad (\text{Mismo procedimiento para todas las clases})$$

Paso6. Se ubica la frecuencia absoluta (f).

Paso7. Se suman las frecuencias absolutas acumuladas hasta llegar a 60 (10 + 19 = 29), (29 + 8 = 37) etc.

Paso8. Se calcula la frecuencia relativa. Dividiendo cada frecuencia absoluta entre el total de datos, ejemplo:

$fr = \frac{10}{60} = .17$ Se repite para todas las clases hasta llegar a 1 ó 100% de los valores

Paso9. Se busca la frecuencia relativa acumulada. Se acumulan las frecuencias relativas hasta llegar a 1 (100%). La tabla de frecuencias queda de la siguiente forma:¹

Intervalos de clase		Marca de clase	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Límite inferior	Límite superior					
20	25	22.5	10	10	0.17	0.17
26	31	28.5	19	29	0.32	0.48
32	37	34.5	8	37	0.13	0.62

Media	35.6
Error típico	1.36216013
Mediana	33.5
Moda	30

¹ Resultados obtenidos en microsoft excel

Desviación estándar	10.551247
Varianza de la muestra	111.328814
Curtosis	-0.50964526
Coefficiente de asimetría	0.65175234
Rango	40
Mínimo	20
Máximo	60
Suma	2136
Cuenta	60

38	43	40.5	9	46	0.15	0.77
44	49	46.5	6	52	0.10	0.87
50	55	52.5	4	56	0.07	0.93
56	61	58.5	4	60	0.07	1.00
62	67	64.5	0		0	1.00
			60		1	

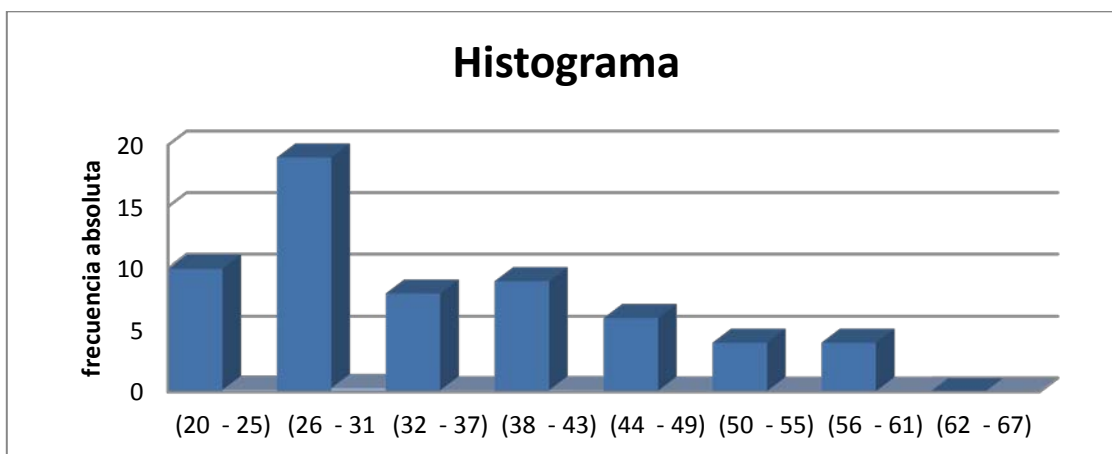


1.2.1 REPRESENTACIÓN GRÁFICA DE LOS DATOS

Histograma. Es la representación gráfica de una variable continua. Se elabora en un sistema de coordenadas rectangulares.

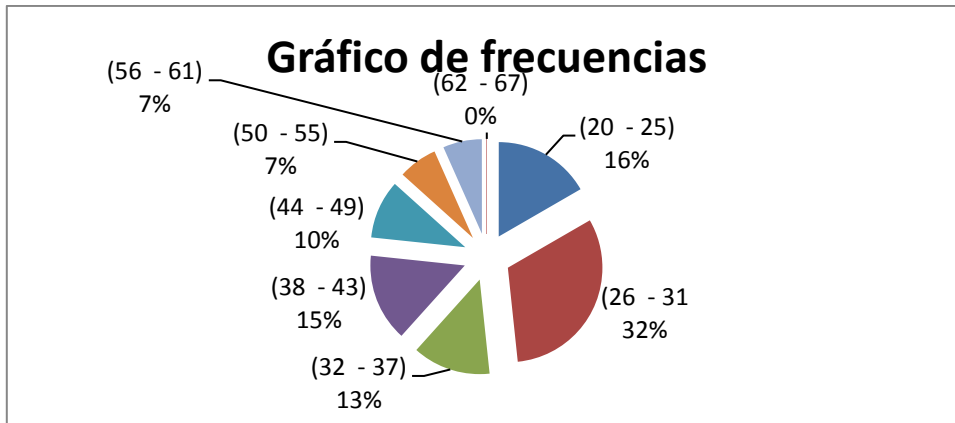
- ✍ El eje horizontal se utiliza para representar a la variable independiente, es decir, a la escala de medición o fronteras de clase.
- ✍ El eje vertical representa a la escala de frecuencias.
- ✍ Si los intervalos de clase tienen el mismo ancho, las alturas de las barras serán proporcionales a las frecuencias.
- ✍ El histograma también proporciona visualmente el aspecto de la distribución y dispersión de las mediciones.

Histograma correspondiente al ejemplo de la empresa Casa S.A



Graficas de área (pastel)

Para trazar la gráfica, se hace una distribución proporcional de las frecuencias del problema anterior con respecto a la circunferencia determinando sectores circulares para cada categoría. Siguiendo con el ejemplo de la empresa Casa S.A



Gráfica de pastel empresa Casa SA 1



Ejemplo para la elaboración de un histograma.

Paso 1. En una serie de números, se cuenta el número de datos que contiene la muestra.

9.9	9.3	10.2	9.4	10.1	9.6	9.9	10.1	9.8
9.8	9.8	10.1	9.9	9.7	9.8	9.9	10.0	9.6
9.7	9.4	9.6	10.0	9.9	9.8	10.1	10.4	10.0
10.2	10.1	9.8	10.1	10.3	10.0	10.2	9.8	10.7
9.9	10.7	9.3	10.3	9.9	9.8	10.3	9.5	9.9
9.3	10.2	9.2	9.9	9.7	9.9	9.8	9.5	9.4
9.0	9.5	9.7	9.7	9.8	9.8	9.3	9.6	9.7
10.0	9.7	9.4	9.8	9.4	9.6	10.0	10.3	9.8
9.5	9.7	10.6	9.5	10.2	10.0	9.8	10.1	9.6
9.6	9.4	10.1	9.5	10.1	10.2	9.8	9.5	9.3
10.3	9.6	9.7	9.7	10.1	9.8	9.7	10.0	10.0
9.5	9.5	9.8	9.9	9.2	10.0	10.0	9.7	9.7
9.9	10.4	9.3	9.6	10.2	9.7	9.7	9.7	10.7
9.9	10.2	9.8	9.3	9.6	9.5	9.6	10.7	

Esta muestra contiene 125 datos.

Paso 2 Se determina el rango (**R**) En este caso, el número mayor es 10.7 y el menor es 9.0 por tanto, el rango es 1.7

Paso 3 Se determina el número de clase (k) a formar. Este número se selecciona de acuerdo con una tabla ya establecida que sirve de guía para determinar el número recomendado de clases.

La tabla es la siguiente:

Número de datos	Números de clases (k)
Menos de 50	5-7
50-99	6-10
100-250	7-12
Más de 250	10-20

En este ejercicio, como los datos son 125 se establece considerar **10 clases**.

CLASE	LIMITE DE CLASE	FRECUENCIA	TOTAL
1	9.00-9.19	I	1
2	9.20-9.39	IIII III	9
3	9.40-9.59	IIII IIII IIII I	16
4	9.60-9.79	IIII IIII IIII IIII IIII II	27
5	9.80-9.99	IIII IIII IIII IIII IIII IIII I	31
6	10.0-10.19	IIII IIII IIII IIII III	23
7	10.20-10.39	IIII IIII II	12
8	10.40-10.59	II	2

9	10.60-10.79	IIII	4
10	10.88-10.99		0

Paso 4 Se determina la amplitud de la clase. La fórmula para hacer esto es la siguiente: $A = \frac{R}{K}$ Aplicando esta fórmula a nuestro ejemplo, se tiene:

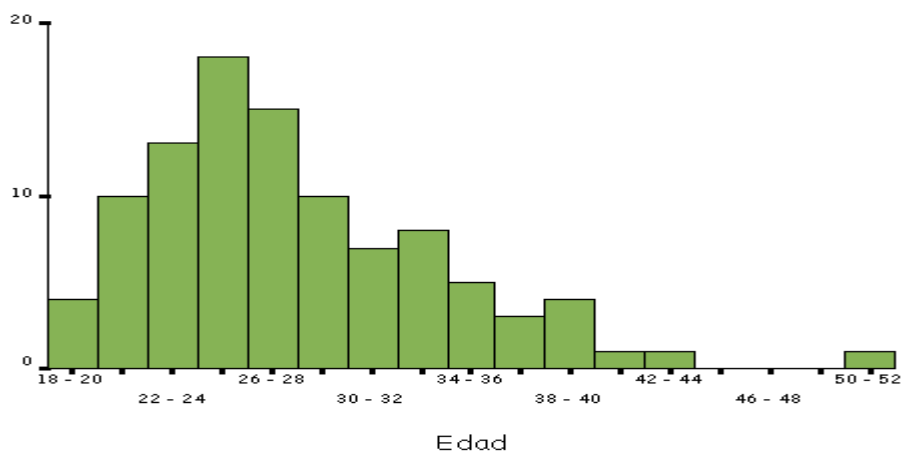
$A = \frac{1.7}{10} = .17$ En la mayoría de los casos es conveniente redondear a un número adecuado. En nuestro caso, 0.17 se redondea a 0.20

Paso 5 Se determina los límites de clase. Para esto se toma la medición individual menor del conjunto de datos. Este es el punto inferior del límite de la primera clase. Se suma a este el número la amplitud de clase. El número que resulta para a ser el límite inferior de la segunda clase y así sucesivamente.

Paso 6. Se Construye la tabla de frecuencias con base en los valores obtenidos (número de clases, intervalo de clases y límite de clases). La tabla de frecuencias que resulta es ya un histograma en forma tabular.

Paso 7 se construye el histograma con base en la tabla de frecuencias. Estas se presentan en forma de barras.

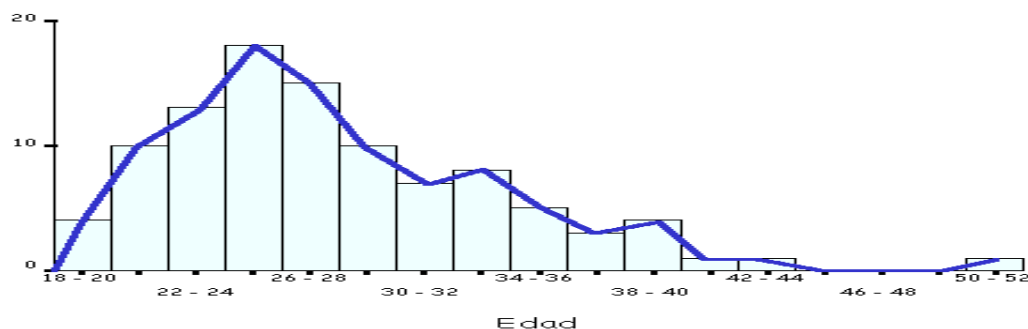
Las barras se elevan a partir de la línea horizontal, en la que se indica los límites de clase. Su altura se determina tomando en cuenta la frecuencia de datos incluidos dentro del límite de clase. La línea vertical del eje de coordenadas se gradúa para indicar precisamente dicha frecuencia. El histograma es una herramienta de diagnóstico muy importante, ya que proporciona una vista panorámica de la variación en la distribución de los datos. El histograma tiene que observarse semejante a este:





1.2.1 POLÍGONOS DE FRECUENCIA.

Se puede obtener uniendo cada punto medio (marca de clase) de los rectángulos del histograma con líneas rectas, teniendo cuidado de agregar al inicio y al final marcas de clase adicionales, con el objeto de asegurar la igualdad del áreas.



Ejercicios: grafique el histograma y polígono de frecuencia a partir de los siguientes datos.

7.9	7.3	8.2	7.4	8.1	7.6	8.1	7.8	7.8
7.8	8.1	7.9	7.7	7.8	7.9	7.6	7.7	7.4
7.6	8.0	7.8	7.9	8.1	8.4	8.2	8.1	7.8
8.1	8.3	8.0	8.2	7.8	8.7	8.7	7.3	8.3
7.9	7.8	8.3	7.5	7.9	7.3	7.2	7.9	7.7
7.9	7.8	7.5	7.4	7.0	7.5	7.7	7.8	7.8
7.3	7.6	7.7	8.0	7.7	7.4	7.4	7.6	8.0
8.3	7.8	7.5	7.7	8.6	7.5	8.0	7.8	8.1
7.6	7.6	7.4	8.1	7.5	8.1	7.8	7.5	7.3
8.3	7.6	7.7	7.7	8.1	7.8	8.0	8.0	7.5
7.5	7.8	7.9	7.2	8.0	8.0	7.7	7.9	8.4
7.3	7.6	8.2	7.7	7.7	7.7	7.9	8.2	7.8
7.3	7.6	7.5	7.6	8.7				



1.3 MEDIDAS DE TENDENCIA CENTRAL Y DE VARIABILIDAD PARA UN CONJUNTO DE DATOS NO AGRUPADOS.



Ejemplo: Supongamos que tenemos los siguientes valores no agrupados: 2, 4, 0, 8, 6, 4, 7, 1, 1, 0, 8, 6, 9. Se pide obtener:

a) Media, Mediana, Moda, Varianza, Desviación Estándar

Solución Media aritmética:

$$\bar{x} = \sum_{i=1}^n \frac{X_i}{n} = \frac{(2 + 4 + 0 + 8 + 6 + 4 + 7 + 1 + 1 + 0 + 8 + 6 + 9)}{13} = 4.31$$



Mediana. Para cuando la cantidad de valores de la distribución es impar:

1. Ordenamos los valores de menor a mayor.
2. Buscamos el valor del centro.

Ordenamos: 0, 0, 1, 1, 2, 4, 4, 6, 6, 7, 8, 8, 9

El dato que divide a la mitad es: 4, por lo tanto la Mediana = 4



Para cuando la cantidad de valores es par:

1. Ordenamos los valores de menor a mayor.
2. Buscamos los valores del centro.
3. Promediamos los valores del centro.

Agregamos un valor a los datos anteriores para ejemplificar

0, 0, 1, 1, 2, 4, 4, 4, 6, 6, 7, 8, 8, 9

1. **Ordenamos:** 0, 0, 1, 1, 2, 4, 4, 4, 6, 6, 7, 8, 8, 9
2. **Buscamos los datos del centro:** 4, 4

3. **Promediamos:** $4 + 4 = 8/2 = 4$, por lo tanto Me: 4



Moda. Es el valor que más se repite. Ejemplo:

0, 0, 1, 1, 2, 4, 4, 4, 6, 6, 7, 8, 8, 9

La moda es el 4



1.4 MEDIDAS DE DISPERSIÓN

Varianza:

Siguiendo con el mismo ejemplo:

$$\sigma^2 = \sum_{i=1}^n \frac{(x - \bar{x})^2}{n - 1}$$

$$S^2 = \sum_{i=1}^n \frac{(2 - 4.31)^2 + (4 - 4.31)^2 + (0 - 4.31)^2 + (8 - 4.31)^2 + (6 - 4.31)^2 + (4 - 4.31)^2 + (7 - 4.31)^2 + (1 - 4.31)^2 + (1 - 4.31)^2 + (0 - 4.31)^2 + (8 - 4.31)^2 + (6 - 4.31)^2 + (2 - 4.31)^2}{13 - 1}$$

$$= 10.56$$

$$S^2 = 10.56$$

Desviación típica o estándar

La desviación típica muestra qué tan alejado está un dato del valor de la media aritmética, es decir, la diferencia que hay entre un dato y la media aritmética. Se denota como s ó σ , según se calcule en una muestra o en toda la población, respectivamente. Se define como la raíz cuadrada positiva de la varianza. Para el ejemplo anterior:

$$S = \sqrt{S^2} \quad S = \sqrt{10.56} \quad S = 3.25$$



Ejercicios. Calcule las medidas de tendencia central, así como las medidas de dispersión (media, moda, mediana, rango, varianza y desviación estándar) de cada conjunto de datos. Analice resultados e indique observaciones.

1. La oficina de correos envió durante julio a diferentes estados de la república, el siguiente número de paquetes:

78, 38, 47, 84, 49, 55, 42, 32, 66, 60, 94, 67, 66, 68, 70.

2. Las tallas más comunes de los vestidos que vendió una boutique durante julio son:

7, 10, 14, 9, 14, 9, 18, 9, 16, 12, 14, 11, 14.

3. En el departamento de control de calidad se tomó una muestra al azar de 10 focos para determinar el número de horas de vida de cada uno obteniéndose los siguientes datos.

Número de muestra.	1	2	3	4	5	6	7	8	9	10
Número de horas	865	850	841	850	820	843	830	848	840	838

4. La producción de tornillos elaborados por un empleado durante la semana que se toma de muestra es :

Día de la semana	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
Número de tornillos	240	225	215	208	295	230

5. La edad de las 10 finalistas de un concurso de belleza es:

18 años, 19, 25, 19, 20, 21, 20, 22, 18, y 18

6. De acuerdo con el informe sobre los pacientes atendidos en un hospital durante la primera semana de julio, se obtuvieron los siguientes datos: lunes 25, martes 24, miércoles 20, jueves 30, viernes 26, sábado 35 y domingo 29

7. Un gerente de personal entrevistó a 15 personas para su contratación, el tiempo(en minutos) que duró la entrevista de cada aspirante fue:

37, 30, 23, 46, 18, 40, 58, 43, 39, 55, 64, 42, 28, 20, 35

8. Al estibar varias cajas de jeringas en un almacén se detectó que algunas de éstas se habían roto, por lo que se tomaron 10 cajas al azar para su revisión habiéndose obtenido la siguiente información: De las primeras cajas dos jeringas rotas, de las siguientes:

3, 1, 0, 4, 2, 1, 3, 0, 2, 3

9. Se tomaron 11 mediciones de diámetro de los anillos para los pistones del motor de un automóvil. Los resultados en milímetros fueron:

74.001, 74.003, 74.025, 74.005, 74.000, 74. 015, 74.005, 74.002, 74.005, 74.002 , 74.004.

RESULTADO

1.	Media	Desv.Est.	Varianza	Mediana	Moda
	61.07	17.38	302.21	66.00	66
2.	Media	Desv.Est.	Varianza	Mediana	Moda
	12.077	3.226	10.410	12.000	14
3.	Media	Desv.Est.	Varianza	Mediana	Moda
	842.50	12.20	148.94	842.00	850
4.	Media	Desv.Est.	Varianza	Mediana	Moda
	235.5	31.2	975.5	227.5	
5.	Media	Desv.Est.	Varianza	Mediana	Moda
	20.000	2.211	4.889	19.500	18
6.	Media	Desv.Est.	Varianza	Mediana	Moda
	27.00	4.83	23.33	26.00	
7.	Media	Desv.Est.	Varianza	Mediana	Moda
	38.53	13.61	185.27	39.00	
8.	Media	Desv.Est.	Varianza	Mediana	Moda
	1.900	1.370	1.878	2.00	3
9.	Media	Desv.Est.	Varianza	Mediana	Moda
	74.006	0.00742	0.00006	74.004	74.005



MEDIDAS DE TENDENCIA CENTRAL Y DE VARIABILIDAD EN DATOS AGRUPADOS

Las fórmulas para calcular la media con los datos agrupados son:

EN UNA MUESTRA	EN UNA POBLACIÓN
$\bar{x} = \sum_{i=1}^n \frac{M_{ci} f_i}{n}$	$\mu = \sum_{i=1}^n \frac{M_{ci} f_i}{N}$

Donde:

M_c = Marca de clase en la i ésima clase

f_i = frecuencia absoluta en la i ésima clase

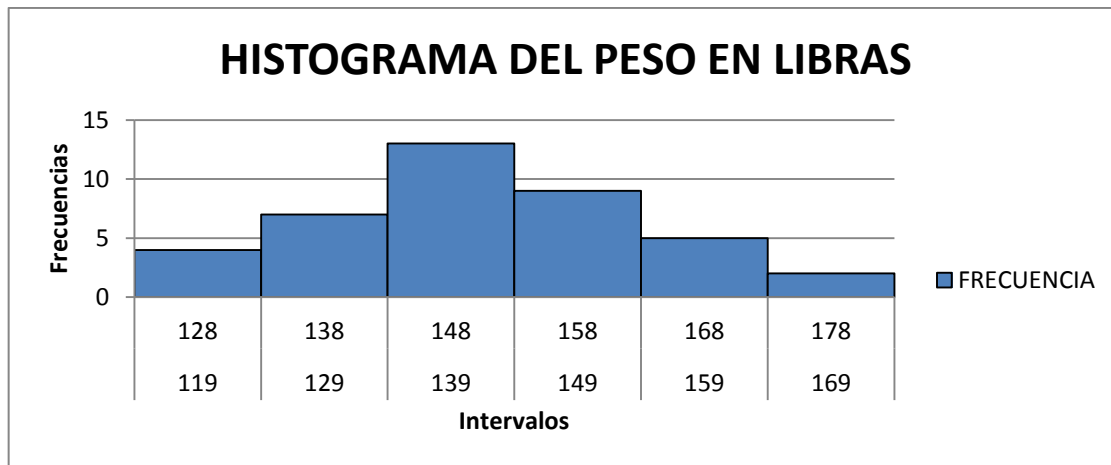
n = Número total de frecuencias



Ejemplo. A partir de la siguiente lista de datos obtener la tabla de distribución de frecuencias agrupadas, medidas de tendencia central (Media, Moda, Mediana), así como las medidas de dispersión (Desviación estándar, varianza y rango). Los datos que se enlistan corresponden a los pesos en libras de los estudiantes de la secundaria.

**138 164 150 132 144 125 149 157 146 158 152 144 168 126 138 176
163 119 154 165 135 153 140 135 161 145 135 142 150 156 147 173
128 136 142 148 147 140 146 145.**

INTERVALOS DE CLASE		MARCA DE CLASE	FRECUE NCIA	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA	FR. REL. %
LI	LS					
119	128	123.5	4	4	0.1	10
129	138	133.5	7	11	0.175	17.5
139	148	143.5	13	24	0.325	32.5
149	158	153.5	9	33	0.225	22.5
159	168	163.5	5	38	0.125	12.5
169	178	173.5	2	40	0.05	5
			40		1	100



MEDIDAS DE TENDENCIA CENTRAL DATOS AGRUPADOS

Media de datos agrupados = $\bar{x} = \sum_{i=1}^n \frac{M_{c i} f_i}{n}$

$$\bar{x} = \frac{4 * 123.5 + 7 * 133.5 + 13 * 143.5 + 9 * 153.5 + 5 * 163.5 + 2 * 173.5}{40}$$

$$= \frac{5840}{40} = 146$$



Mediana de datos agrupados= $ME = L.i.e \left(\frac{N}{2} - \sum f \right) \div f * A$

$$Me = 138.5 + (20 - 11) \div 13 * 10 = 145.42$$

$$N = \frac{40}{2} = 20$$

$$L_{ie} = 138.5$$

$$\sum f = 11$$

$$A = 10$$



$$\text{Moda para datos agrupados} = mo = Lie + \frac{d1}{d1+d2} * A$$

$$Moda = 138.5 + \frac{6}{(6+4)} * 10 = 144.5$$

$$Lie = 138.5$$

$$d1 = 13 - 7 = 6$$

$$d2 = 13 - 9 = 4$$

$$A = 10$$

$$\text{Varianza} = S^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{(123.5 - 146)^2 + (133.5 - 146)^2 + (143.5 - 146)^2 + (153.5 - 146)^2 + (163.5 - 146)^2 + (173.5 - 146)^2}{39}$$

$$= \frac{1787.5}{39} = 45.83$$

$$s^2 = 45.83$$

$$\text{Desviación estándar} = S = \sqrt{S^2}$$

$$S = \sqrt{45.83} = s = 6.77$$



COEFICIENTE DE VARIACION.

$$\left(\frac{\text{Desviación estándar}}{\text{Promedio}} \right) \times 100$$

$$\left(\frac{6.77}{146} \right) \times 100 = .005 \times 100 = 4.63$$



Ejemplo 2. Los datos que a continuación se enlistan corresponden a los diámetros interiores de inyectores.

424 430 433 435 436 437 426 431 433 435 436 438 428 431 434 435
437 438 429 432 434 436 437 438 430 432 434 436 437 438 430 432
434 436 437 439 442 439 444 440 443 440 444 441 446

$$\text{MEDIA} = \bar{X} = \frac{\sum_{i=1}^n f * MC}{n}$$

$$= \bar{X} = \frac{3*426+9*431+17*436+9*441+2*446}{40} = 435.75 = 436$$

$$\text{MEDIANA} = Me = L.med + \left(\frac{\frac{n}{2} - f}{+med} \right)$$

$$ME = 434 + \left(\frac{\frac{40}{2} - 429}{17} \right) * 4 = 434 + \left(\frac{-409}{17} \right) * 4 = 434 + (-24.05) * 4 = 434 - 96.2 = 337.8$$

$$\text{MODA} = Mo = Li + \left(\frac{d_1}{d_1 + d_2} \right) * c$$

$$MO = 429 + \left[\frac{(9-3)}{(9-3)(9-17)} \right] * 4 = 429 + \left[\frac{(6)}{(6)(-8)} \right] * 4 = 429 + \left(\frac{6}{-48} \right) * 4 = 429 - 0.125 * 4 = 429 - 0.5 = 428.5$$

$$\text{VARIANZA } S^2 = \frac{\sum_{i=1}^n f_i (M_i - \bar{x})^2}{n-1}$$

$$S^2 = \frac{\sum_{i=1}^n (426 - 436)^2 + (431 - 436)^2 + (436 - 436)^2 + (441 - 436)^2 + (446 - 436)^2}{40 - 1}$$

$$S^2 = \frac{\sum_{i=1}^n 250}{39} = 6.41$$

DESVIACION ESTANDAR MUESTRAL PARA DATOS AGRUPADOS

$$s = \sqrt{S^2} \quad s = \sqrt{6.41} = 2.53$$

COEFICIENTE DE VARIACION.

$$\left(\frac{\text{Desviación estándar}}{\text{Promedio}} \right) \times 100 \quad \left(\frac{2.53}{436} \right) \times 100 = .005 \times 100 = .5$$



EJERCICIOS

1. El gerente de producción de la imprenta "x" desea determinar el tiempo promedio que se necesita para fotografiar una placa de impresión; utilizando un cronometro y observando a los operadores registran los siguientes tiempos:

20.4, 22, 20, 24.07, 22.2, 25.7, 23.8, 24.9, 22.7, 25.1, 24.4, 21.2, 24.3, 22.4, 23.6, 22.8, 23.2, 24.3, 21

- ✍ Construye una tabla de datos
- ✍ Construye una tabla de frecuencias
- ✍ Construye el histograma, polígonos de frecuencia u ojivas una gráfica de línea y una gráfica de barras.
- ✍ Calcular media, moda, mediana, varianza y desviación estándar para datos agrupados
- ✍ Encuentra en cada ejemplo el coeficiente de variación

2. En un grupo de 30 estudiantes se preguntó cuánto dinero llevaban en ese momento. Los resultados obtenidos, en pesos, fueron los siguientes:

45.00, 11.55, 25.00, 30.00, 17.50, 8.00, 2.50, 268.00, 60.50, 78.50, 159.50, 230.00, 500.00, 120.00, 10.00, 5.00, 18.00, 20.00, 67.50, 50.00, 37.50, 150.00, 20.50, 98.50, 18.50, 12.50, 31.50, 42.50, 56.00 y 110.00.

Realiza lo siguiente:

- ✍ Organiza los datos en orden ascendente (del menor al mayor)
- ✍ Obtén el rango de los datos
- ✍ Realiza una tabla con 10 intervalos con las siguientes columnas:
- ✍ Intervalo
- ✍ Límite inferior
- ✍ Límite superior
- ✍ Marca de clase
- ✍ Frecuencia
- ✍ Frecuencia acumulada
- ✍ Frecuencia relativa
- ✍ Frecuencia relativa acumulada
- ✍ Obtén las medidas de tendencia central para datos agrupados por intervalos
- ✍ Obtén las medidas de dispersión para datos agrupados por intervalos

3. En una escuela se midió el peso de 21 alumnos en kilogramos y se obtuvieron los siguientes resultados:

58, 42, 51, 54, 40, 39, 49, 56, 58, 57, 59, 63, 58, 63, 70, 72, 71, 69, 70, 68, 64

Realiza lo siguiente:

- ✍ Organiza los datos en una tabla de datos
- ✍ Organiza los datos en una tabla de frecuencias
- ✍ Organiza los datos en una tabla que tenga 7 intervalos
- ✍ Calcula las medidas de tendencia central para cada una de las tablas
- ✍ Calcula las medidas de dispersión para cada una de las tablas

4. Una compañía que fabrica llantas investiga la duración promedio de un nuevo compuesto de caucho. Para ello se probaron 30 llantas en una carretera hasta alcanzar la vida útil de éstas. Los resultados obtenidos, en kilómetros, fueron:

60, 613	59, 836	60, 135	60, 222 5	9, 554	60, 252
60, 613	59, 784	60, 221	59, 997	60, 311	50, 040
60, 222	60, 220	60, 545	60, 222	60, 257	60, 000
59, 997	59, 997	69, 947	60, 135	60, 220	60, 311
59, 784	60, 222	60, 554	60, 225	59, 838	60, 523

Realiza lo siguiente:

- ✍ Organiza los datos en una tabla de datos
- ✍ Organiza los datos en una tabla de frecuencias
- ✍ Organiza los datos en una de intervalos que tenga 10 intervalos
- ✍ Saca la media, la mediana y la moda para cada una de las tablas
- ✍ Saca el rango, la varianza y la desviación estándar para cada una de las tablas



COEFICIENTE DE VARIACIÓN PEARSON

Formula $V_P = \frac{s}{\bar{x}}$

Ejemplo. Tenemos dos grupos de mujeres de 11 y 25 años con medias y desviaciones típicas dadas por la tabla siguiente:

	Peso Medio ($\bar{x}$)	Desviación Típica (s)
11 años	40 Kg.	2 kg
25 años	50 Kg	2 kg

Puede parecernos, al observar en ambos grupos una desviación típica igual, que ambos grupos de datos tienen la misma dispersión. No obstante, como parece lógico, no es lo mismo una variación de dos kilos en un grupo de elefantes que en uno de conejos. El *coeficiente de Variación de Pearson* elimina esa posible confusión al ser una medida de la variación de los datos pero en relación con su media. En el ejemplo anterior, al grupo de mujeres de 11 años le corresponde un coeficiente de variación de Pearson igual a

$$V_P = \frac{2}{40} \cdot 100 = 5$$

Y al grupo de las mujeres de 25 años

$$V_P = \frac{2}{50} \cdot 100 = 4$$

Lo que indica una mayor dispersión en el grupo de mujeres de 11 años.

Ejercicio 1. Se va a comparar la dispersión en los precios anuales de las acciones que se venden a menos de \$10 (dólares) y la dispersión en los precios de aquellas que se venden por arriba de \$60. El precio medio de las acciones que se venden a menos de \$10 es \$5.25 y la desviación estándar es \$1.52. El precio medio de las acciones que se negocian a más de \$60 es \$92.50 y su desviación estándar es \$5.28.



- ¿Porque debe utilizarse el coeficiente de variación para comparar la dispersión de los precios?
- Calcule los coeficientes de variación.Cuál es su conclusión

2. Suponga que Usted trabaja en una compañía de ventas, que ofrece como premio de incentivo al mejor vendedor del trimestre anterior las entradas al palco empresarial en la serie final de béisbol de las grandes ligas en los Estados Unidos. De los registros de ventas se tienen los siguientes datos de ventas, expresados en porcentajes de cumplimiento de las metas fijadas mensualmente:

Vendedor A 95 105 100

Vendedor B 100 90 110

El promedio trimestral de cumplimiento de las metas de ventas de ambos vendedores es igual y equivale al 100%, pero Ud. Sólo le puede dar el premio de incentivo a uno de ellos. ¿Cuál usted escogería? En base a que criterio. Explique su respuesta.



REFERENCIAS:

- Montgomery, Douglas C. y George C. Runger (1996). Probabilidad y Estadística aplicadas a la ingeniería. McGraw-Hill, México, cuarta edición.
 - Walpole, Ronald E., Raymond H. Myers et al. (2007). Probabilidad y Estadística para Ingeniería y ciencias. México: Pearson Educación, octava edición.
 - Intervalos de clase, consultado en:
http://www.virtual.unal.edu.co/cursos/odontologia/2002890/lecciones/estadistica_descriptiva_2/estadistica_descriptiva_2.htm
 - Censo y entrevista, en:
- <http://www.indec.gov.ar/proyectos/censo2001/maestros/quees/masinfo.doc>.
 - http://www.tec.url.edu.gt/boletin/URL_03_BAS01.pdf
 - Medidas de tendencia central y dispersión, consultado en:
- <http://bibliotecavirtual.lasalleurubamba.edu.pe/Estadistica/res/pdf/estadisticadescriptivavariabales2.pdf>
 - <http://www.vitutor.com/estadistica.html>



UNIDAD 2 INTRODUCCIÓN A LA PROBABILIDAD

Propósitos de la unidad

En esta unidad el alumno:

- ✍ Identifica los conceptos básicos de la teoría de probabilidad.
- ✍ Utiliza las reglas y postulados de la probabilidad para resolver problemas en eventos aleatorios.
- ✍ Obtiene las variables aleatorias y las distribuciones de probabilidad de experimentos aleatorios simples.
- ✍ Aplica los modelos de probabilidad para solucionar problemas.

Competencia específica

Aplica la teoría de la probabilidad en la toma de decisiones en problemas del área económica administrativa.

Aplica el concepto de valor esperado o esperanza matemática para la toma de decisiones.

Utiliza los modelos de probabilidad para el análisis de eventos y situaciones en diferentes contextos a través de experimentos aleatorios.

Identifica los conceptos básicos de probabilidad para la solución de problemas mediante experimentos aleatorios.



INTRODUCCIÓN

La utilidad de la teoría de la probabilidad en cualquier disciplina que se aplique, es que puede proporcionar un modelo matemático adecuado para la descripción de los fenómenos aleatorios con los que nos encontremos. Muy frecuentemente, estos fenómenos tienen un comportamiento similar al de modelos como Binomial, de Poisson y Normal.

En esta unidad se abordarán algunos ejercicios básicos de probabilidad. Ésta es una de las mejores herramientas que existen para el manejo del riesgo en las sociedades modernas, pues día a día se presentan múltiples situaciones en las que la toma de decisiones se debe realizar sin contar con que todas las variables estén bajo un perfecto control. De hecho esta situación de control total rara vez (o nunca) se da. En estadística la probabilidad nos ayudará a hacer inferencias con los resultados obtenidos a través del manejo de los datos.



2.1 EVENTOS MUTUAMENTE EXCLUYENTES Y NO EXCLUYENTES

Definición. Dos eventos A y B se dicen ser mutuamente excluyentes si el evento $A \cap B$ no contiene ningún punto muestral.



2.2 REGLAS DE ADICIÓN

La probabilidad de la unión ($A \cup B$) es $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Si A y B son mutuamente excluyentes, entonces

$$P(A \cup B) = P(A) + P(B) \text{ ya que } P(A \cap B) = 0$$



Un taller sabe que por término medio acuden: por la mañana tres automóviles con problemas eléctricos, ocho con problemas mecánicos y tres con problemas de chapa, y por la tarde dos con problemas eléctricos, tres con problemas mecánicos y uno con problemas de chapa.

	Electricidad	Mecánica	Chapa	
Mañanas	3	8	3	14
Tardes	2	3	1	6
Total	5	11	4	20

- a) Calcular la probabilidad de que un automovilista acuda por la tarde (T)

$$P(T) = \frac{6}{20} = 0.30 = 30\%$$

- b) Calcular la probabilidad de que un automovilista acuda por la mañana ($\tilde{N}$)

$$P(\tilde{N}) = \frac{14}{20} = 0.70 = 70\%$$

- c) Calcular la probabilidad de que un automovilista acuda por problemas mecánicos (M).

$$P(M) = \frac{11}{20} = 0.55 = 55\%$$

- d) Calcular la probabilidad de que un automovilista acuda por problemas eléctricos (E).

$$P(E) = \frac{5}{20} = 0.25 = 25\%$$

- e) Calcular la probabilidad de que un automóvil con problemas eléctricos acuda por la mañana.

$$P(\tilde{N} \cap E) = \frac{3}{5} = 0.60 = 60\%$$



2.3 EVENTOS INDEPENDIENTES, DEPENDIENTES, PROBABILIDAD CONDICIONAL

Definición. Dos eventos A y B se dicen ser independientes si

$$P(A|B) = P(A) \quad \text{ó bien} \quad P(B|A) = P(B)$$

En caso contrario, los eventos se dirán ser dependientes



Ejemplo de eventos independientes. La experiencia indica que un determinado tipo de negociación obrero patronal ha resultado en la firma de un convenio dentro de dos semanas de pláticas el 50% de las veces. También la experiencia indica que el fondo de soporte monetario para la huelga ha sido adecuado para soportar la huelga el 60% de las veces y que ambas de estas condiciones se han satisfecho el 30% de las veces. ¿Cuál es la probabilidad de que en una negociación determinada se logre una firma de convenio dentro de dos semanas de pláticas dado que se tiene un fondo adecuado para la huelga? ¿Es la firma de convenio dentro de dos semanas dependiente de si se tiene o no un fondo adecuado para la huelga?

Solución Se definen primero dos eventos:

Evento A: se firma convenio dentro de dos semanas de pláticas

Evento B: el fondo de soporte para huelga es adecuado

Se desea encontrar $P(B|A)$, con base en $P(A) = .50$, $P(B) = .60$ $P(A \cap B) = .30$

$$\text{Se tiene: } P(A|B) = \left(\frac{P(A \cap B)}{P(B)} \right) \left(\frac{.30}{.60} \right) = .50$$

Para determinar si los eventos son o no independientes, observa $P(A|B) = .50$
Que por definición indica que si son independientes



EJEMPLO DE EVENTOS DEPENDIENTES.

Cuando se recibe una entrega de un proveedor, el comprador usualmente inspecciona la calidad del envío. Un almacén de descuento ha recibido 100 aparatos de televisión del proveedor, de los cuales les es desconocido, que 10 están defectuosos. Si se seleccionan al azar 2 aparatos para ser sometidos a una inspección muy minuciosa, ¿cuál es la probabilidad de que ambos estén defectuosos?

Solución Se definen primero dos eventos:

Evento A: el primer aparato de TV está defectuoso

Evento B: el segundo aparato de TV está defectuoso

El evento de interés es el evento $(A \cap B)$, que ambos estén defectuosos, y

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

$P(A) = .10$ ya que hay 10 defectuosos en el lote de 100. Sin embargo $P(B|A) = \frac{9}{99}$ ya que tras haber seleccionado el primero que resultó defectuoso, habrá 9 defectuosos restantes en el lote, ahora de 99 solamente.

$$\text{Por tanto } P(A \cap B) = P(A) P(B|A) = \left(\frac{10}{100} \right) \left(\frac{9}{99} \right) = \left(\frac{1}{110} \right)$$



2.3 PROBABILIDAD CONDICIONAL

La probabilidad condicional de B dado que A ha ocurrido, es

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

La probabilidad condicional de A dado que B ha ocurrido, es

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



EJEMPLOS DE PROBABILIDAD CONDICIONAL. Sean A y B dos sucesos aleatorios con:

$$P(A) = \frac{1}{3}, \quad P(B) = \frac{1}{4} \quad P(A \cap B) = \frac{1}{5} \quad \text{Determinar:}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{5}}{\frac{1}{4}} = \frac{4}{5}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{5}}{\frac{1}{3}} = \frac{3}{5}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} = \frac{23}{60}$$

$$P(\bar{B}|\bar{A}) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{A})} = \frac{P(\bar{A} \cup \bar{B})}{1 - P(A)} = \frac{1 - P(A \cup B)}{1 - P(A)} = \frac{1 - \frac{23}{60}}{1 - \frac{1}{3}} = \frac{37}{40}$$

$$P(\bar{A}|B) = \frac{P(\bar{A} \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = \frac{\frac{1}{4} - \frac{1}{5}}{\frac{1}{4}} = \frac{1}{5}$$

$$P(\bar{B}|A) = \frac{P(\bar{A} \cap A)}{P(A)} = \frac{P(A) - P(A \cap B)}{P(A)} = \frac{\frac{1}{3} - \frac{1}{5}}{\frac{1}{3}} = \frac{2}{5}$$

EJERCICIOS

1. Sean A y B dos sucesos aleatorios con $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$,

$$P(A \cap B) = \frac{1}{4} \quad \text{Determinar:}$$

a) $P(A|B) =$

b) $P(B|A) =$

c) $P(A \cup B) =$

d) $P(\bar{A}|\bar{B}) =$

e) $P(\bar{B}|\bar{A}) =$

Respuestas:

a) $\frac{3}{4}$ b) $\frac{1}{2}$ c) $\frac{7}{12}$ d) $\frac{5}{8}$ e) $\frac{5}{6}$



2.4 REGLAS DE MULTIPLICACIÓN

Dados dos eventos A y B la probabilidad de la intersección ($A \cap B$) es

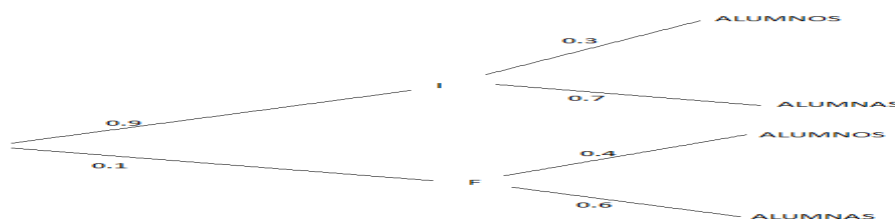
$P(A \cap B) = P(A) P(B|A)$ Si A y B son independientes $P(A \cap B) = P(A) P(B)$



2.5 DIAGRAMAS DE ÁRBOL



Ejemplo. En el tecnológico los alumnos pueden optar por cursar como lengua extranjera inglés o francés. En un determinado curso, el 90% de los alumnos estudia inglés y el resto francés. El 30% de los que estudian inglés son hombres y de los que estudian francés son hombres el 40%. Se ha elegido un alumno al azar, ¿cuál es la probabilidad de que sea mujer?

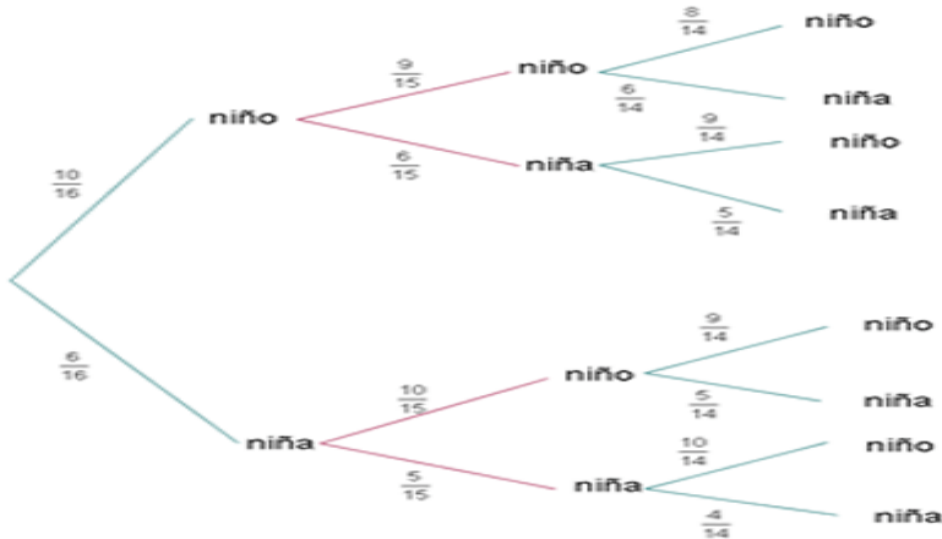


$$P(\text{Mujer}) = (0.9)(0.7) + (0.1)(0.6) = \mathbf{0.69}$$



Una clase consta de seis niñas y 10 niños. Si se escoge un comité de tres al azar, hallar la probabilidad de:

a) Seleccionar tres niños.



$$P(3 \text{ niños}) = \left(\frac{10}{16}\right)\left(\frac{9}{15}\right)\left(\frac{8}{14}\right) = 0.214 = 21.4\%$$

b) Seleccionar exactamente dos niños y una niña.

$$P(2 \text{ niños y } 1 \text{ niña}) = \left(\frac{10}{16}\right)\left(\frac{9}{15}\right)\left(\frac{6}{14}\right) + \left(\frac{10}{16}\right)\left(\frac{6}{15}\right)\left(\frac{9}{14}\right) + \left(\frac{6}{16}\right)\left(\frac{10}{15}\right)\left(\frac{9}{14}\right) = 0.482$$

$$= 48.2\%$$

c) Seleccionar por lo menos un niño.

$$P(\text{al menos } 1 \text{ niño}) = 1 - (\text{todas niñas}) = 1 - \left(\frac{6}{16}\right)\left(\frac{5}{15}\right)\left(\frac{4}{14}\right) = 0.964$$

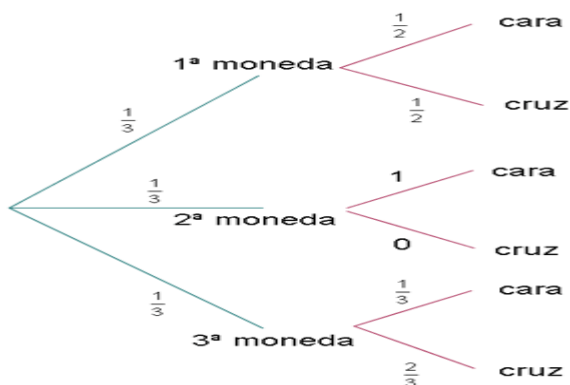
d) Seleccionar exactamente dos niñas y un niño.

$$P(2 \text{ niñas y } 1 \text{ niño}) = \left(\frac{10}{16}\right)\left(\frac{6}{15}\right)\left(\frac{5}{14}\right) + \left(\frac{6}{16}\right)\left(\frac{10}{15}\right)\left(\frac{5}{14}\right) + \left(\frac{6}{16}\right)\left(\frac{5}{15}\right)\left(\frac{10}{14}\right)$$

$$= 0.268 = 26.8\%$$



Una caja contiene tres monedas. Una moneda es corriente, otra tiene dos caras y la otra está cargada de modo que la probabilidad de obtener cara es de $\frac{1}{3}$. Se selecciona una moneda lanzar y se lanza al aire. Hallar la probabilidad de que salga cara.



$$P(\text{cara}) = \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)(1) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = 0.611 = 61.1\%$$



EJERCICIOS

- En un aula hay 100 alumnos, de los cuales: 40 son hombres, 30 alumnos usan lentes, y de este grupo 15 son varones y usan lentes. Si seleccionamos al azar un alumno de dicho curso:

	Con lentes	Sin Lentes	
HOMBRES	15	25	40
MUJERES	15	45	60
	30	70	100

- a) ¿Cuál es la probabilidad de que sea mujer y no use lentes?
 - b) Si sabemos que el alumno seleccionado no usa gafas, ¿qué probabilidad hay de que sea hombre?
2. Disponemos de dos urnas: la urna A contiene 6 bolas rojas y 4 bolas blancas, la urna B contiene 4 bolas rojas y 8 bolas blancas. Se lanza un dado, si aparece un número menor que 3; nos vamos a la urna A; si el resultado es 3 ó más, nos vamos a la urna B. A continuación extraemos una bola. Se pide:
- a) Probabilidad de que la bola sea roja y de la urna B.
 - b) Probabilidad de que la bola sea blanca.
3. Un estudiante cuenta, para un examen con la ayuda de un despertador, el cual consigue despertarlo en un 80% de los casos. Si oye el despertador, la probabilidad de que realiza el examen es 0.9 y, en caso contrario, de 0.5.
- a) Si va a realizar el examen, ¿cuál es la probabilidad de que haya oído el despertador?
 - b) Si no realiza el examen, ¿cuál es la probabilidad de que no haya oído el despertador?
4. En una estantería hay 60 novelas y 20 libros de poesía. Una persona A elige un libro al azar de la estantería y se lo lleva. A continuación otra persona B elige otro libro al azar.
- a) ¿Cuál es la probabilidad de que el libro seleccionado por B sea una novela?
 - b) Si se sabe que B eligió una novela, ¿cuál es la probabilidad de que el libro seleccionado por A sea de poesía?
5. Se supone que 25 de cada 100 hombres y 600 de cada 1000 mujeres usan gafas. Si el número de mujeres es cuatro veces superior al de hombres, se pide la probabilidad de encontrarnos:
- a) Con una persona sin gafas.
 - b) Con una mujer con gafas.
6. En una casa hay tres llaveros A, B y C; el primero con cinco llaves, el segundo con siete y el tercero con ocho, de las que sólo una de cada llavero abre la puerta del trastero. Se escoge al azar un llavero y, de él una llave para abrir el trastero. Se pide:
- a) ¿Cuál será la probabilidad de que se acierte con la llave?
 - b) ¿Cuál será la probabilidad de que el llavero escogido sea el tercero y la llave no abra?

- c) Y si la llave escogida es la correcta, ¿cuál será la probabilidad de que pertenezca al primer llavero A?

7. Sean A y B dos sucesos aleatorios con:

$$P(A) = \frac{3}{8}$$

$$P(B) = \frac{1}{2}$$

$$P(A \cap B) = \frac{1}{4}$$

Hallar:

a) $P(A \cup B) =$

b) $P(\bar{A}) =$

c) $P(\bar{B}) =$

d) $P(\bar{A} \cap \bar{B}) =$

e) $P(\bar{A} \cup \bar{B}) =$

f) $P(A \cap \bar{B}) =$

8. Se sacan dos bolas de una urna que se compone de una bola blanca, otra roja, otra verde y otra negra. Escribir el espacio muestral cuando:

a) La primera bola se devuelve a la urna antes de sacar la segunda.

b) La primera bola no se devuelve.

9. Una urna tiene ocho bolas rojas, 5 amarilla y siete verdes. Si se extrae una bola al azar calcular la probabilidad de:

a) Sea roja.

b) Sea verde.

c) Sea amarilla.

d) No sea roja.

e) No sea amarilla.

10. Una urna contiene tres bolas rojas y siete blancas. Se extraen dos bolas al azar. Escribir el espacio muestral y hallar la probabilidad de los sucesos:
- a) Con reemplazamiento.
 - b) Sin reemplazamiento.
11. Se extrae una bola de una urna que contiene 4 bolas rojas, 5 blancas y 6 negras, ¿cuál es la probabilidad de que la bola sea roja o blanca? ¿Cuál es la probabilidad de que no sea blanca?
12. En una clase hay 10 alumnas rubias, 20 morenas, cinco alumnos rubios y 10 morenos. Un día asisten 45 alumnos, encontrar la probabilidad de que un alumno:
- a) Sea hombre.
 - b) Sea mujer morena.
 - c) Sea hombre o mujer.
13. Un dado está trucado, de forma que las probabilidades de obtener las distintas caras son proporcionales a los números de estas. Hallar:
- a) La probabilidad de obtener el 6 en un lanzamiento.
 - b) La probabilidad de conseguir un número impar en un lanzamiento.
14. Se lanzan dos dados al aire y se anota la suma de los puntos obtenidos. Se pide:
- a) La probabilidad de que salga el 7.
 - b) La probabilidad de que el número obtenido sea par.
 - c) La probabilidad de que el número obtenido sea múltiplo de tres.



2.6 COMBINACIONES Y PERMUTACIONES

PERMUTACIONES



EJEMPLO: 1.- ¿De cuantas maneras posibles se pueden sentar 10 personas en una banca si solamente hay 4 puestos disponibles?

SOLUCIÓN

El primer puesto puede ocuparse de cualquiera de 10 maneras, luego el segundo puede ocuparse de 9 maneras, el tercero de 8 maneras diferentes y el cuarto de 7, por lo tanto: El numero de ordenaciones de 10 personas tomadas de 4 a la vez $= 10 \cdot 9 \cdot 8 \cdot 7 = 5040$

2.- calcule

a) ${}_8P_3$

b) ${}_6P_4$

c) ${}_{15}P_1$

d) ${}_3P_3$

SOLUCIÓN:

(a) ${}_8P_3 = 8 \cdot 7 \cdot 6 = 336$

(b) ${}_6P_4 = 6 \cdot 5 \cdot 4 \cdot 3 = 360$

(c) ${}_{15}P_1 = 15$

(d) ${}_3P_3 = 3 \cdot 2 \cdot 1 = 6$



EJERCICIOS. Se necesita sentar 5 hombres y 4 mujeres en fila, de manera que las mujeres ocupen los lugares pares, ¿de cuantas maneras pueden sentarse?

Calcule:

a) ${}_8P_4$

b) ${}_5P_2$

c) ${}_{10}P_{13}$

d) ${}_{13}P_5$



2.6 COMBINACIONES



EJEMPLO ¿de cuantas maneras se pueden dividir 10 objetos en dos grupos que contengan 4 y 6 objetos respectivamente?

SOLUCIÓN:

En general, el número de selecciones de r de n objetos, llamados *el número de combinaciones de n objetos tomados a la vez*, se describe por ${}_nC_r$ ó $\binom{n}{r}$ y esta dado por:

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1) \cdots (n-r+1)}{r!} = \frac{nPr}{r!}$$

Esto es lo mismo que el número de ordenaciones de 10 objetos, de los cuales 4 son semejantes entre si y los otros 6 también lo cual podemos determinar que:

$$\frac{10!}{4!6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = 210$$

2.- calcule

a) ${}_7C_4$

b) ${}_6C_5$

c) ${}_4C_4$

SOLUCIÓN:

$$(a) \quad {}_7C_4 = \frac{7!}{4!3!} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{4!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$$

$$(b) \quad {}_6C_5 = \frac{6!}{5!1!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{5!} = 6$$

$$(c) \quad {}_4C_4 = \frac{4!}{4!0!} = 1 \quad \text{definimos } 0! = 1$$

3.- ¿de cuantas maneras se puede formar un comité de 5 personas a partir de un grupo de 9?

SOLUCIÓN:

$$\binom{9}{5} = \frac{9!}{5!4!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{5!} = 126$$



Análisis combinatorio

Estudia los diversos arreglos o selecciones que podemos formar con los elementos de un conjunto dado los cuales nos permite resolver muchos problemas prácticos.

Principios fundamentales del análisis combinatorio

En la mayoría de problemas de análisis combinatorios se observa que una operación o actividad aparece en forma repetitiva y es necesario conocer las formas o maneras de realizar dicha operación



EJEMPLO 1- Para calcular el número de combinaciones con repetición se aplica:

$$C_n^m = \binom{n}{m} = \frac{n!}{m! (n - m)!}$$

SOLUCION: son las combinaciones de 10 elementos agrupándolos en subgrupos de 4 elementos,

$$C_4^{10} = \frac{10!}{4! (10 - 4)} = 210$$



EJERCICIOS: 1.-Con 3 personas: Antonio, Beto y Carlos ¿cuántos grupos diferentes de dos se podrán formar?

2.- se tienen cinco personas A, B, C, D, y E y queremos formar grupos diferentes de tres personas lo cual podríamos combinarlos de la siguiente manera:

3.-¿Cuántas comisiones de tres alumnos se pueden formar con 4 varones y 5 mujeres.

Fuentes de consulta

1. Douglas C. Montgomery, George C. Runger. Probabilidad y Estadística aplicadas a la ingeniería. Primera Edición, McGraw-Hill, México, 1999.
2. Walpole Ronald E., Myers Raymond H. Probabilidad y Estadística. Cuarta Edición, Thomson, México, 1999.

- <http://www.vitutor.com/estadistica.html>
- <http://www.uaq.mx/matematicas/estadisticas/xu4.html>
- <http://www.eumed.net/cursecon/libreria/drm/ped-drm-est.htm>



UNIDAD 3. TIPOS DE DISTRIBUCIONES VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS

Propósitos de la unidad

En esta unidad el alumno debe:

- ✍ Identificar los principios básicos de probabilidad discreta y continua para la toma de decisiones.
- ✍ Graficar una distribución de probabilidad.
- ✍ Diferenciar las variables aleatorias continuas y discretas.
- ✍ Aplicar las técnicas de distribución de probabilidad continua como: normal y aproximación de la normal a la binomial, para la toma de decisiones

Competencia específica

- ✍ Diferencia las variables aleatorias discretas y continuas.
- ✍ Aplica las técnicas de distribución de probabilidad discreta y continua para la toma de decisiones



Introducción

La utilidad de la teoría de la probabilidad en cualquier disciplina que se aplique, es que puede proporcionar un modelo matemático adecuado para la descripción de los fenómenos aleatorios con los que nos encontremos. Y muy frecuentemente, estos fenómenos tienen un comportamiento similar al de modelos ya conocidos como binomial, de Poisson y Normal, que es lo que corresponde tratar en esta unidad.



Una variable aleatoria continua es aquella que puede tomar valores infinitos. Una forma útil de diferenciar este tipo de variables es que típicamente las variables continuas representan datos medidos, tales como alturas, distancias, pesos, temperaturas, tiempo de vida, etc.,

Mientras que las variables discretas representan conteo de datos, tales como el número de productos defectuosos, el número de contagios de una enfermedad, etc.

1. El número de canicas escogidas aleatoriamente de un lote de producción para la inspección de calidad DISCRETA
2. Cantidad de bebes nacidos en el hospital general de zona numero 197 en un día DISCRETA.

3. Estaturas de los alumnos del TESOEM comprendidas en 1.50m. al 1.90m.
CONTINUA.
4. Número de tarjetas de debito dadas por un banco local en un cuatrimestre.
DISCRETA.



Ejemplo de distribución, valor esperado, varianza y desviación estándar en variables aleatorias discretas

Ejemplo: obtener el valor esperado, varianza y desviación estándar de los siguientes problemas.

1. En el siguiente cuadro se muestran la probabilidad de artículos de un producto que se esperan vender en un día normal.

N° De productos (x_i)	Probabilidad $P(x_i)$	E(X) $(x_i) P(x_i)$
0	0.10	$(0)(0.10) = 0$
10	0.15	$(10)(0.15) = 1.5$
20	0.15	$(20)(0.15) = 3$
30	0.40	$(30)(0.40) = 12$
40	0.20	$(40)(0.20) = 8$
	1.00	$\mu = E(x) = 24.5$

Solución:

$$\text{Media} = \mu = E(x) = (x_i) P(x_i)$$

Varianza:

$$\sigma^2 = \sum_{i=1}^n [X_i - E(X)]^2 P(X_i)$$

$$= (0 - 24)^2(0.10) + (10 - 24.5)^2(0.15) + (20 - 24.5)^2(0.15) + (30 - 24.5)^2(0.40) + (40 - 24.5)^2(0.20)$$

$$= 60.025 + 31.5375 + 3.0375 + 3.0375 + 12.1 + 48.05 = 154.75$$

Desviación estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^n [X_i - E(X)]^2 P(X_i)} \quad \sigma = \sqrt{154.75} = 12.4399$$



En el siguiente cuadro se muestran la probabilidad de bebés que se esperan que nazcan en una semana. Encuentre la media, varianza y desviación estándar en los datos discretos.

N° De bebés(x_i)	probabilidad $P(x_i)$	$(x_i) P(x_i)$
0	0.05	$(0)(0.05) = 0$
2	0.20	$(2)(0.20) = 0.4$
4	0.25	$(4)(0.25) = 1$
6	0.20	$(6)(0.20) = 1.2$
8	0.30	$(8)(0.30) = 2.4$
	1.00	$\mu = E(x) = 5$

Varianza:

$$\sigma^2 = \sum_{i=1}^n [Xi - E(X)]^2 P(Xi)$$

$$= (0 - 5)^2(0.05) + (2 - 5)^2(0.20) + (4 - 5)^2(0.25) + (6 - 5)^2(0.20) + (8 - 5)^2(0.30)$$

$$= 1.25 + 1.8 + 0.25 + 0.20 + 2.7 = 6.2$$

Desviación estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^n [Xi - E(X)]^2 P(Xi)} \quad \sigma = \sqrt{6.2} = 2.489$$



Ejercicio. En el siguiente cuadro se muestran la probabilidad de pares de botas que se esperan vender en un mes. Encuentre la media, varianza y desviación estándar en los datos discretos

No. De pares de botas(x_i)	probabilidad $P(x_i)$	$(x_i) P(x_i)$
4	0.19	
8	0.40	
14	0.30	
20	0.11	



En la siguiente distribución de probabilidad nos muestra la cantidad de bolsas que se esperan vender en un día de una fábrica. Encuentre la media, varianza y desviación estándar en los datos discretos

No. De bolsas(x_i)	probabilidad $P(x_i)$	$(x_i) P(x_i)$
0	0.01	
50	0.02	
125	0.14	
150	0.35	
200	0.48	
	1.00	



3.1 DISTRIBUCIÓN BINOMIAL

La distribución binomial es una distribución discreta de probabilidad que tiene muchas aplicaciones. Se relaciona con un experimento de etapas múltiples que llamamos binomial. La variable aleatoria X que denota el número de éxitos en n ensayos de Bernoulli tiene una distribución binomial dada por $p(x)$, donde:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n$$



Propiedades de un experimento binomial

1. El experimento consiste en una sucesión de n intentos o ensayos idénticos.
2. En cada intento o ensayo son posibles dos resultados. A uno lo llamaremos éxito y a otro fracaso.
3. La probabilidad de un éxito, representada por p , no cambia de un intento o ensayo a otro. En consecuencia, la probabilidad de un fracaso, representada por $1 - p$, no cambia de un intento a otro.
4. Los intentos o ensayos son independientes.



Media, varianza y desviación estándar de la distribución binomial

La media de la distribución binomial puede determinarse como

$$E(X) = \sum_{x=0}^n x * \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= np \sum_{x=1}^n \frac{(n-1)!}{y!(n-1-y)!} p^y q^{n-1-y}$$

Y dejando $y = x - 1$ $E(X) = np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{n-1-y}$

Por lo que $E(X) = np$

Al emplear un enfoque similar encontramos la varianza como

$$V(X) = \sum_{x=0}^n \frac{n!}{x!(n-x)!} p^x q^{n-x} - (np)^2$$

$$= n(n-1)p^2 \sum_{x=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} p^y q^{n-2-y} + np - (np)^2$$

De manera que

$$V(X) = npq$$

La desviación estándar se obtiene:

$$\sigma = \sqrt{npq}$$

Refirámonos al caso de arrojar 3 monedas, $n = 3$ y $p = 1/2$ obtenemos:

$$\sigma = \sqrt{npq} = \sqrt{(3)(1/2)(1/2)} = \sqrt{3/4} = \sqrt{0.75} = 0.87$$



Ejemplo 1: Si la probabilidad de que cualquier elector registrado (seleccionado al azar de las listas oficiales) vote en una elección determinada es 0.70 ¿Cuál es la probabilidad de que 2 de 5 electores registrados voten en la elección?

Datos:

$$\frac{n!}{r!(n-r)!} p^r q^{n-r}$$

$$r = 2 \quad n = 5 \quad \binom{5}{2} = 10$$

$$P(r = 2) = \binom{5}{2} (0.70)^2 (1 - 0.70)^{5-2} \\ = 10(0.70)^2 (0.30)^3 = 0.132$$



Ejemplo 2. Una máquina fabrica una determinada pieza y se sabe que produce 7 defectuosas de cada 1000 piezas. Hallar la probabilidad de que al examinar 50 piezas sólo haya una defectuosa.

Solución: Se trata de una distribución binomial de parámetros B (50, 0.007) y debemos calcular la probabilidad $P(r = 1)$.

$$P(r = 1) = \binom{50}{1} 0.007^1 * 0.993^{49} = 0.248$$



3.2 MODELO DE POISSON



Existen otros experimentos en los que lo que se busca es determinar el número de eventos que suceden en tiempo o espacio finito y no si el resultado es éxito o fracaso. Por ejemplo, conocer el número de autos que pasan por una cierta ruta en un intervalo de tiempo, determinar el número de llamadas simultáneas que está procesando una antena de telefonía celular, saber el número de accesos que tiene un servidor web por segundo, etc. Para llevar a cabo el análisis de este tipo de experimentos, se utiliza el modelo de Poisson.



PROPIEDADES DEL MODELO DE POISSON

La distribución de Poisson se calcula con la fórmula: $\frac{\lambda^x e^{-\lambda}}{x!}$ donde:

$p(x, \lambda)$ = probabilidad de que ocurran x éxitos, cuando el número promedio de ocurrencia de ellos es λ

λ = media o promedio de éxitos por unidad de tiempo, área o producto

$e = 2.718$

x = variable que nos denota el número de éxitos que se desea que ocurra



Ejemplo Si un banco recibe en promedio 6 cheques sin fondo por día, cuáles son las probabilidades de que reciba:

- a) cuatro cheques sin fondo en un día dado,
- b) 10 cheques sin fondos en cualquiera de dos días consecutivos

Solución:

- a) X = variable que nos define el número de cheques sin fondo que llegan al banco en un día cualquiera = 0, 1, 2, 3,....., etc.

$\lambda = 6$ cheques sin fondo por día

$e = 2.718$

$$p(x = 4, \lambda = 6) = \frac{(6)^4(2.718)^{-6}}{4!} = \frac{(1226)(0.00248)}{24} = 0.13392$$

- b) X = variable que nos define el número de cheques sin fondo que llegan al banco en dos días consecutivos = 0, 1, 2, 3,....., etc., etc.

$\lambda = (6 \times 2) = 12$ cheques sin fondo en promedio que llegan al banco en dos días consecutivos. Nota: λ siempre debe de estar en función de x siempre o dicho de otra forma, debe “hablar” de lo mismo que x .

$$p(x = 10, \lambda = 12) = \frac{(12)^{10}(2.718)^{-12}}{10!} = \frac{(6.191736)(0.000006151)}{3628800} = 0.104953$$



Ejemplo. En la inspección de hojalata producida por un proceso continuo, se identifican 0.2 imperfecciones en promedio por minuto. Determine las probabilidades de identificar:

- a. una imperfección en 3 minutos,
- b. al menos dos imperfecciones en 5 minutos,
- c. cuando más una imperfección en 15 minutos.

Solución:

- a) x = variable que nos define el número de imperfecciones en la hojalata por cada 3 minutos = 0, 1, 2, 3,....., etc.

$\lambda = 0.2 \times 3 = 0.6$ imperfecciones en promedio por cada 3 minutos en la hojalata

$$p(x = 1, \lambda = 0.6) = \frac{(0.6)^1(2.718)^{-0.6}}{1!} = \frac{(0.6)(0.548845)}{1} = 0.329307$$

- b) x = variable que nos define el número de imperfecciones en la hojalata por cada 5 minutos = 0, 1, 2, 3,....., etc.

$\lambda = 0.2 \times 5 = 1$ imperfección en promedio por cada 5 minutos en la hojalata

$$\begin{aligned} p(x = 2, 3, 4, \text{etc} \dots \lambda = 1) &= 1 - p(x = 0, 1, \lambda = 1) \\ &= 1 - \left(\frac{(1)^0 (2.718)^{-1}}{0!} + \frac{(1)(2.718)^{-1}}{1!} \right) \\ &= 1 - (0.367918 + 0.367918) = 0.26419 \end{aligned}$$

c) x = variable que nos define el número de imperfecciones en la hojalata por cada 15 minutos = 0, 1, 2, 3, ..., etc.

$\lambda = 0.2 \times 15 = 3$ imperfecciones en promedio por cada 15 minutos en la hojalata

$$\begin{aligned} p(x = 0, 1, \lambda = 3) &= p(x = 0, \lambda = 3) + p(x = 1, \lambda = 3) \\ &= \left(\frac{(3)^0 (2.718)^{-3}}{0!} + \frac{(3)^1 (2.718)^{-3}}{1!} \right) \\ &= 0.049800226 + 0.149408 = 0.1992106 \end{aligned}$$



EJERCICIO 1: Se sabe que el 2% de los libros que se encuadernan en un taller tienen una encuadernación defectuosa. Use la aproximación de Poisson para la distribución binomial para encontrar la probabilidad de que 5 de 400 libros encuadernados en este taller tengan una encuadernación defectuosa.



La distribución de Poisson tiene muchas aplicaciones importantes y no se relacionan en forma directa con la distribución binomial. En este caso, np se sustituye por λ y calculamos la probabilidad de tener x triunfos por medio de la fórmula.

$$f(x) = \frac{\lambda^x \cdot e^{-1}}{x!}$$

Para $x = 0, 1, 2, 3 \dots$



EJERCICIO 2: Si un banco recibe en promedio $\lambda = 6$ cheques sin fondos por día. ¿Cuál es la probabilidad de que reciba cuatro cheques sin fondos en un día determinado?



3.3 DISTRIBUCIÓN HIPERGEOMÉTRICA DE PROBABILIDAD.

Con la distribución hipergeométrica los intentos no son independientes. La notación que se acostumbra al aplicar la distribución hipergeométrica de probabilidad es que r representa la cantidad de elementos en la población de tamaño N , que se identifican como éxitos, y que $N - r$ representa la cantidad de elementos en la población que se identifican como fracasos.



La distribución hipergeométrica de probabilidad se usa para calcular la probabilidad de que, en una muestra aleatoria de n artículos, seleccionados sin remplazo, obtengamos x elementos identificados como éxitos y $n - x$ identificados como fracasos. Para que suceda esto debemos obtener x éxitos de los r en la población, y $n - x$ fracasos de los $N - r$ de la población. La siguiente función hipergeométrica de probabilidad determinada $f(x)$, la probabilidad de obtener x éxito en una muestra de tamaño n .

Función de probabilidad hipergeométrica:

$$f(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \quad \text{para } 0 \leq x \leq r$$

En donde:

$f(x)$ = probabilidad de x éxitos en n intentos

n = cantidad de intentos

N = la cantidad de elementos en la población

r = la cantidad de elementos identificados con éxito en la población

Obsérvese que $\binom{N}{n}$ representa la cantidad de formas en la que se puede seleccionar una muestra de tamaño n de una población de tamaño N ; que $\binom{r}{x}$ representa la cantidad de maneras que se pueden seleccionar x éxitos de un total r éxitos de la población; y que $\binom{N-r}{n-x}$ representa la cantidad de maneras en que se pueden seleccionar $n - x$ fracasos de un total de $N - r$ fracasos en la población.



EJEMPLO: Seleccionar dos miembros de comité, entre cinco, que asistan a una convención en Las Vegas. Suponga que el comité de cinco miembros está formado por tres mujeres y dos hombres .para determinar la probabilidad de seleccionar dos mujeres al azar.

Aplicando la ecuación:

$$f(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

$$n = 2$$

$$N = 5$$

$$r = 3$$

$$x = 2$$

$$f(x) = \frac{\binom{3}{2} \binom{5-3}{2-2}}{\binom{5}{2}} = \frac{\binom{3}{2} \binom{2}{0}}{\binom{5}{2}} = \frac{\left(\frac{3!}{2!1!}\right) \left(\frac{2!}{2!0!}\right)}{\left(\frac{5!}{2!3!}\right)} = \frac{3}{10} = .30$$



EJERCICIO: Una población consiste en 10 artículos, cuatro de los cuales son defectuosos y los seis restantes son no defectuosos . ¿Cuál es la probabilidad de que una muestra aleatoria de tamaño tres contenga dos artículos defectuosos? (En este caso podemos imaginar que un éxito consiste en obtener un artículo defectuoso)



3.5 MODELO NORMAL

1. El máximo ocurre para $\bar{x} = \mu$
2. La curva es simétrica alrededor de μ
3. La curva tiene sus puntos de inflexión (puntos en que la curva cambia de cóncava a convexa) en $\bar{x} = \mu \pm \sigma$
4. La curva se aproxima al eje horizontal de forma asintótica.
5. El área total de la curva normal es igual a 1 (toda posible gama de posibilidades está contemplada $p = [0,1]$)



Fórmula para calcular distribución normal

La distribución normal depende de 2 parámetros, la media μ y la desviación estándar σ .

La fórmula para la distribución normal de una variable discreta es la siguiente:

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde:

μ es la media

σ es la desviación estándar

$\pi=3.14159...$



Ejemplo sobre cómo convertir una distribución normal a una normal tipificada.

El salario medio de los empleados de una empresa se distribuye según una distribución normal, con media 5 mil pesos y desviación típica 1 mil pesos. Calcular el porcentaje de empleados de la empresa con un sueldo inferior a 7 mil pesos.

1. Transformamos esa distribución en una normal tipificada, para ello se crea una nueva variable (Z):

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

1. Sustituimos la fórmula y la nueva variable sería:

$$Z = \frac{x - 5}{1}$$

2. Esta nueva variable se distribuye como una normal tipificada. La variable Z que corresponde a una variable X de valor 7 es:

$$Z = \frac{7 - 5}{1} = 2$$

Ya podemos consultar en la tabla Z la probabilidad acumulada para el valor 2 (equivalente a la probabilidad de sueldos inferiores a 7 mil pesos). Esta probabilidad es 0.97725. Por lo tanto, el porcentaje de empleados con salarios inferiores a 7 mil pesos es del 97.725%.



Cómo se usa la tabla de valores para la distribución normal estándar

La tabla de probabilidad normal estándar se utiliza de la siguiente manera.

La columna de la izquierda indica el valor cuya probabilidad acumulada queremos conocer. La primera fila nos indica el segundo decimal del valor que estamos consultando.

1. Se localiza en una Tabla de la distribución normal estándar acumulada el valor de z buscado en la primera columna, aproximando la unidad y una décima.
2. Una vez localizado, se recorre el renglón de la tabla hasta encontrar la z que corresponda a la centésima más próxima.
3. En la intersección de la columna y renglón aparece la probabilidad buscada.



Ejemplo: Suponga que Z es una variable normal estándar. Encuentre la $P(Z \leq 1.34)$.

TABLA Probabilidades de una Normal Estándar										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706

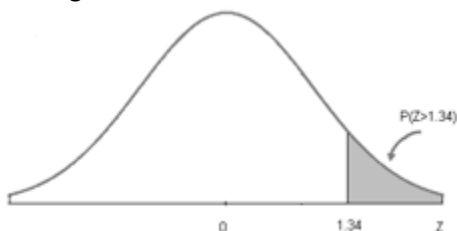
Buscando en la tabla nos da un valor de $P(Z \leq 1.34) = 0.9099$, es decir, tiene el 90.1% del área total de la curva de probabilidad hasta $Z = 1.34$, como se muestra a continuación.



Continuando con el ejemplo anterior, si quisiéramos calcular la $P(Z > 1.34)$ entonces, sería más conveniente calcularlo así:

$$P(Z > 1.34) = 1 - P(Z \leq 1.34) = 1 - 0.9099 = 0.0901$$

Y su gráfica se muestra a continuación,



Si quisiéramos la probabilidad entre 2 valores, tendríamos que realizar la resta de áreas, por ejemplo:

$$P(1.21 < Z \leq 1.34) = P(Z \leq 1.34) - P(Z \leq 1.21) = 0.9099 - 0.8869 = 0.023$$

Y su gráfica se muestra a continuación,



Ejercicios. Los resultados en el examen de admisión al TESOEM tienen una distribución normal con media 75 y desviación estándar 10.

- a. ¿Qué fracción de los resultados quedó entre 80 y 90?
 - b. Obtén la variable aleatoria normal estándar.
1. En una compañía refresquera se ajusta una máquina de refrescos de tal manera que llena las latas de refresco con un promedio de 300 mililitros. El número de mililitros por lata tiene una distribución normal con una desviación estándar de 10 mililitros.
 - a) ¿Cuál debe ser la capacidad mínima de las latas para que se derrame cuando mucho el 1% de ellas?
 - b) Obtén la variable aleatoria normal estándar.
 2. El diámetro del agujero de las tuercas de una fábrica tienen una distribución normal con una media de 15.0 milímetros y una desviación estándar de 0.1 milímetros. Los tornillos diseñados aceptan tuercas de entre 14.888 y 5.112
 - a) ¿Cuál es la probabilidad de que una tuerca escogida al azar no sirva?
 - b) Obtén la variable aleatoria normal estándar.



UNIDAD 4. MUESTREO Y ESTIMACIONES

Propósitos de la unidad

En esta unidad el alumno debe:

- ✍ Identificar los conceptos básicos de muestreo.
- ✍ Reconocer la utilidad e importancia de las medidas de tendencia central.
- ✍ Identificar operaciones que se utilizan en distribución de muestreo de la media.
- ✍ Organizar datos en diferentes tipos de Intervalos de confianza para la media, con el uso de la distribución Normal y “t” de student
- ✍ Aplicar las fórmulas para obtener Intervalo de confianza para la diferencia entre dos medias $\mu_1 - \mu_2$ con $\sigma_1 = \sigma_2$ pero conocidas, con el uso de la distribución normal y la “t” de student cuando no se conoce la varianza de la población.

Competencia específica

- ✍ Utiliza los tipos de muestreo para asegurar que las muestras que se tomen sea una representación real de la población.
- ✍ Conoce y comprende las características de la distribución normal.
- ✍ Conoce y comprende las características de la distribución t de student
- ✍ Determina el tamaño de la muestra óptimo para un análisis poblacional, utilizando grado de confianza y estimación de μ .
- ✍ Aplica los métodos de estimación por intervalos para la solución de problemas relativos a la Administración.



Introducción

Los estudios estadísticos normalmente se hacen con una parte de la población, ya que realizarlos sobre la totalidad resultaría demasiado complicado. Para que la información obtenida tenga validez es necesario que la muestra cumpla con ciertas condiciones específicas, relacionadas con el método para determinar el tamaño y características de la muestra y los individuos que la componen.



Los métodos de muestreo se pueden clasificar en:

- **Muestreo probabilístico:** en él, todos los elementos de una población y, por lo tanto, todas las muestras posibles tienen la misma posibilidad de ser elegidas. Las muestras obtenidas a través de este tipo de muestreo son contables porque aseguran la condición de representatividad que es muy importante para hacer generalizaciones.
- **Muestreo no probabilístico:** en este tipo de muestreo los elementos de la población no comparten las mismas posibilidades de ser seleccionados. Las muestras obtenidas no cumplen con la condición de representatividad, por lo que no es probable hacer generalizaciones a toda la población.



Metodología del muestreo aleatorio simple

Definir la población de estudio y el parámetro a estudiar. Recordemos que la población es el grupo formado por el conjunto total de individuos, objetos o medidas que poseen algunas características comunes observables en un lugar y en un momento determinado. Por lo tanto:

1. Es determinar el que se va a estudiar.
2. Enumerar a todas las unidades de análisis que integran la población, asignándoles un número de identidad o identificación.
3. Determinar el tamaño de la población, determinar el porcentaje de error y el porcentaje de confianza y obtener una muestra preliminar.



4.1 DISTRIBUCIÓN MUESTRAL DE LA MEDIA



EJEMPLO

1. La media de la población normal, es $\mu = 60$ y la desviación estándar poblacional es $\sigma = 12$. Se toma una muestra aleatoria de $n = 9$. Calcule la probabilidad de que la media muestral sea;
 - a) Mayor que 63
 - b) Menor que 56
 - c) Entre 56 y 63.

Solución:

$$\text{a) } P(\bar{x} > 63) \quad \mu = 60 \quad \sigma = 12$$

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} \quad Z = \frac{63 - 60}{\frac{12}{\sqrt{9}}} = \frac{3}{4} = .75$$



El valor estandarizado se busca en tabla Z y se tiene que la probabilidad es .2734 ó 27.34%, como se busca que sea mayor se resta de .5 la cantidad que no interesa para el estudio quedando:

$$.5 - .2734 = .2266 \quad 1 - .2734 = 0.2266 = \mathbf{22.66\%}$$

b) P ($\bar{x} < 56$)

$$Z = \frac{56 - 60}{\frac{12}{\sqrt{9}}} = \frac{-4}{4} = -1 \quad 1 - .8298 = .1702 = \mathbf{17.02\%} \quad .5 - .3298 = .1702$$

c) Este entre 56 y 63

$$\mathbf{P (56 < \bar{x} < 63)} \quad .3298 + .2734 = 0.6032 \times 100 = \mathbf{60.32\%}$$



EJERCICIOS

- Se sabe que la resistencia a la ruptura de cierto tipo de cuerda se distribuye normalmente con media de 2000 libras y una varianza de 25000 libras. Si se selecciona una muestra aleatoria de 100 cuerdas; determine la probabilidad de que en esa muestra:
 - La resistencia media encontrada sea de por lo menos 1958 libras.
 - La resistencia media sea mayor de 2080 libras.
- Como parte de un proyecto general de mejoramiento de la calidad, un fabricante textil decide controlar el número de imperfecciones encontradas en cada pieza de tela. Se estima que el número promedio de imperfecciones por cada pieza de tela es de 12, determine la probabilidad de que en la próxima pieza de tela fabricada se encuentren:
 - Entre 10 y 12 imperfecciones.
 - Menos de 9 y más de 15 imperfecciones.
- En una prueba de aptitud la puntuación media de los estudiantes es de 72 puntos y la desviación estándar es de 8 puntos. ¿Cuál es la probabilidad de que dos grupos de estudiantes, formados de 28 y 36 estudiantes, respectivamente, difieran en su puntuación media en:
 - 3 ó más puntos.
 - 6 ó más puntos.
 - Entre 2 y 5 puntos
- Un especialista en genética ha detectado que el 26% de los hombres y

- el 24% de las mujeres de cierta región del país tiene un leve desorden sanguíneo; si se toman muestras de 150 hombres y 150 mujeres, determine la probabilidad de que la diferencia muestral de proporciones que tienen ese leve desorden sanguíneo sea de:
- Menos de 0.035 a favor de los hombres.
 - Entre 0.01 y 0.04 a favor de los hombres.
5. Una urna contiene 80 bolas de las que 60% son rojas y 40% blancas. De un total de 50 muestras de 20 bolas cada una, sacadas de la urna con reemplazamiento, ¿en cuántas cabe esperar
- Igual número de bolas rojas y blancas?
 - 12 bolas rojas y 8 blancas?
 - 8 bolas rojas y 12 blancas?
 - 10 ó mas bolas blancas?
6. Los pesos de 1500 cojinetes de bolas se distribuyen normalmente con media de 2.40 onzas y desviación estándar de 0.048 onzas. Si se extraen 300 muestras de tamaño 36 de esta población, determinar la media esperada y la desviación estándar de la distribución muestral de medias si el muestreo se hace:
- Con reemplazamiento
 - Sin reemplazamiento
7. La vida media de una máquina para hacer pasta es de siete años, con una desviación estándar de un año. Suponga que las vidas de estas máquinas siguen aproximadamente una distribución normal, encuentre:
- La probabilidad de que la vida media de una muestra aleatoria de 9 de estas máquinas caiga entre 6.4 y 7.2 años.
 - El valor de la X a la derecha del cual caería el 15% de las medias calculadas de muestras aleatorias de tamaño nueve.
8. Se llevan a cabo dos experimentos independientes en lo que se comparan dos tipos diferentes de pintura. Se pintan 18 especímenes con el tipo A y en cada uno se registra el tiempo de secado en horas. Lo mismo se hace con el tipo B. Se sabe que las desviaciones estándar de la población son ambas 1.0. Suponga que el tiempo medio de secado es igual para los dos tipos de pintura. Encuentre la probabilidad de que la diferencia de medias en el tiempo de secado sea mayor a uno a favor de la pintura A.



4.2 DISTRIBUCIÓN MUESTRAL DE LA DIFERENCIA ENTRE DOS MEDIAS



Inicialmente estaremos interesados en verificar si ambas distribuciones tienen la misma media poblacional, es decir si $\mu_1 = \mu_2$ ó equivalentemente $\mu_1 - \mu_2 = 0$, por lo que debemos hacer las siguientes consideraciones:

- Distribución de la diferencia entre dos medias cuando las varianzas son conocidas.
- Distribución de la diferencia entre dos medias cuando las varianzas son conocidas y diferentes
- Distribución de la diferencia entre dos medias cuando las varianzas son desconocidas pero iguales.
- Distribución de la diferencia entre dos medias cuando las varianzas son desconocidas y diferentes



Ejemplo de cuando las varianzas son conocidas: En un estudio para comparar los pesos promedio de niños y niñas de sexto grado en una escuela primaria se usará una muestra aleatoria de $n_1 = 20$ niños y otra de $n_2 = 25$ niñas. Se sabe que tanto para niños como para niñas los pesos siguen una distribución normal. El promedio de los pesos de todos los niños de sexto grado de esa escuela es de $\mu_1 = 100$ libras y su desviación estándar es de $\sigma_1 = 14.142$, mientras que el promedio de los pesos de todas las niñas del sexto grado de esa escuela es de $\mu_2 = 85$ libras y su desviación estándar es de $\sigma_2 = 12.247$ libras. Si $\bar{x}_1$ representa el promedio de los pesos de 20 niños y $\bar{x}_2$ es el promedio de los pesos de una muestra de 25 niñas, encuentre la probabilidad de que el promedio de los pesos de los 20 niños sea al menos 20 libras más grande que el de las 25 niñas.

Solución:

Datos:

$$\mu_1 = 100 \text{ libras}$$

$$\sigma_1 = 14.142 \text{ libras}$$

$$\mu_2 = 85 \text{ libras}$$

$$\sigma_2 = 12.247 \text{ libras}$$

$$n_1 = 20 \text{ niños}$$

$$n_2 = 25 \text{ niñas}$$

$$\bar{x}_1 - \bar{x}_2 = 20$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{20 - (100 - 85)}{\sqrt{\frac{(14.142)^2}{20} + \frac{(12.247)^2}{25}}} = 1.25$$

Por lo tanto, la probabilidad de que el promedio de los pesos de la muestra de niños sea al menos 20 libras más grande que el de la muestra de las niñas es 0.1056.



EJEMPLO de cuando las varianzas poblacionales son conocidas e iguales. De una población se toma una muestra de $n_1 = 40$ observaciones. La media muestral es de $\bar{x}_1 = 102$ y la desviación estándar de $\sigma_1 = 5$. De otra población se toma una muestra de $n_2 = 50$ observaciones y la media muestral es ahora $\bar{x}_2 = 99$ y la desviación estándar es 6. Calcule el valor estadístico de la prueba. Se debe suponer que las medias poblacionales son iguales.

$$\bar{x}_1 = 102$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\bar{x}_2 = 99$$

$$\sigma_1 = 5$$

$$Z = \frac{(102 - 99) - (0)}{1.18} = \frac{3}{1.18} = 2.54$$

$$\sigma_2 = 6$$

$$\sigma^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2} = \frac{(40 - 1)5^2 + (50 - 1)6^2}{40 + 50 - 2} = \frac{975 + 1764}{88} = \frac{2739}{88} = 31.13$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{31.13}{40} + \frac{31.13}{50}} = \sqrt{0.77 + 0.62} = \sqrt{1.3926} = 1.18$$

$$.5 + .3810 = 0.119$$



EJERCICIOS:

1. Uno de los principales fabricantes de televisores compra los tubos de rayos catódicos a dos compañías. Los tubos de la compañía A tienen una vida media de 7.2 años con una desviación estándar de 0.8 años, mientras que los de la B tienen una vida media de 6.7 años con una desviación estándar de 0.7. Determine la probabilidad de que una muestra aleatoria de 34 tubos de la compañía A tenga una vida promedio de al menos un año más que la de una muestra aleatoria de 40 tubos de la compañía B.
2. Se prueba el rendimiento en km/L de 2 tipos de gasolina, encontrándose una desviación estándar de 1.23km/L para la primera gasolina y una desviación estándar de 1.37km/L para la segunda gasolina; se prueba la primera gasolina en 35 autos y la segunda en 42 autos.
 - a. ¿Cuál es la probabilidad de que la primera gasolina de un rendimiento promedio mayor de 0.45km/L que la segunda gasolina?
 - b. ¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre 0.65 y 0.83km/L a favor de la gasolina 1?



4.3 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA DE LA POBLACIÓN.



Con el muestreo aleatorio simple estratificado se puede considerar que la elección del tamaño de la muestra es un proceso en dos etapas.

Primero, se debe elegir un tamaño total de muestra n . En segundo lugar, decidir cuando asignar las unidades muestrales a los diversos estratos.

En forma alterna, se podría decidir primero el tamaño de la muestra que se tomará de cada estrato, y después sumar los tamaños de muestra para obtener el tamaño total.

La distribución consiste en decidir que fracción de la muestra total se debe asignar a cada estrato. Esta fracción determina el tamaño de la muestra aleatoria simple en cada estrato. Los factores que se consideran más importantes en la asignación son:

1. La cantidad de elementos en cada estrato
2. La varianza de los elementos dentro de cada estrato
3. El costo de selección de elementos dentro de cada estrato

Las muestras más grandes se deben asignar a los principales estratos y a los estratos con varianzas mayores. Al revés para obtenerla máxima información a

determinado costo, las muestras mas pequeñas se deben asignar a los estratos en los que es máximo el costo por unidad muestreada.

El costo de selección puede ser muy importante cuando se requiere de desplazamientos significativos del encuestador entre las unidades muestreadas en determinados estratos, pero no en otros, este caso se presenta más cuando algunos de los estratos implican áreas rurales y otras ciudades.



Las siguientes fórmulas presentan el costo total de muestreo para determinado nivel de precisión. El método se conoce como asignación de Neyman, y asigna total n para los diversos estratos en la forma siguiente:

Ecuación 1:
$$n_h = n \frac{N_h S_h}{\sum_{i=1}^n N_h S_h}$$

Dado un nivel B de precisión, podemos usar las siguientes fórmulas para elegir el tamaño total de la muestra y así estimar la media de la población y el total de la población.

Ecuación 2:

Tamaño de la muestra para estimar la media de la población

$$n = \frac{(\sum_{h=1}^H N_h S_h)^2}{\frac{N^2 B^2}{4} + \sum_{h=1}^H N_h S_h^2}$$

Ecuación 3:

Tamaño de la muestra para estimar el total de la población

$$n = \frac{\sum_{h=1}^H N_h S_h^2}{\frac{B^2}{4} + \sum_{h=1}^H N_h S_h^2}$$

Donde:

N_h = La cantidad de elementos en cada estrato

S_h^2 = La varianza de los elementos dentro de cada estrato

B^2 = El costo de selección de elementos dentro de cada estrato



Ejemplo: Imaginemos el caso de un distribuidor Chevrolet, que desea encuestar a los clientes que le compraron un Corvette, un Corsa o un Cavalier, para obtener información que cree le será útil para elaborar sus promociones en el futuro. En especial supongamos que la agencia desea estimar la media del ingreso mensual para estos clientes con una cuota de 100 dólares en el error del muestreo. Los 600 clientes del distribuidor se han dividido en tres estratos: 100 dueños de Corvette, 200 de Corsa y 300 de Cavalier. Se hizo una encuesta de piloto para estimar la desviación estándar en cada estrato, cuyos resultados fueron $s_1 = \$1,300$, $s_2 = \$900$, y $s_3 = \$500$, respectivamente, para los dueños de Corvette, Corsa y Cavalier.

El primer paso para elegir un tamaño de la muestra para esta encuesta es usar la ecuación 2 y determinar el tamaño de la muestra necesario para obtener una cuota de $B = \$100$ en el estimado de la media de la población. Primero se calcula:

$$\sum_{h=1}^3 N_h S_h = 100(1300) + 200(900) + 300(500) = 460,000$$

$$\sum_{h=1}^3 N_h S_h^2 = 100(1300)^2 + 200(900)^2 + 300(500)^2 = 406,000,000$$

Sustituimos esos valores en la ecuación 2, a fin de poder determinar el tamaño total de la muestra necesario para obtener una cuota de $B = \$100$ en el error del muestreo.

$$n = \frac{(460,000)^2}{\frac{(600)^2(100)^2}{4} + 406,000,000} = 162$$

Con un tamaño total de muestra igual a 162 se obtendrá la precisión deseada. Para asignar la muestra total a los tres estratos usamos la ecuación 1.

$$n = 162 \frac{100(1300)}{460,000} = 46$$

$$n_2 = 162 \frac{200(900)}{460,000} = 63$$

$$n_3 = 162 \frac{300(500)}{460,000} = 53$$



4.4 INTERVALOS DE CONFIANZA PARA LA MEDIA, CON EL USO DE LA DISTRIBUCIÓN NORMAL Y “T” DE STUDENT.

Tamaño de muestra pequeña y varianza poblacional σ^2 desconocida



Supóngase que la varianza de la población es desconocida. ¿Qué sucede con la distribución de esta estadística si se reemplaza σ por s ? La distribución t proporciona la respuesta a esta pregunta.

Fórmula para muestras <30

$$Ic = \bar{x} \pm t S_{\bar{x}}$$

Fórmula para muestras >30

$$Ic = \bar{x} \pm Z \sigma_{\bar{x}}$$



Ejemplo: El señor Juan Pérez se dedica a hacer tarjetas postales y los vende en 50 papelerías; como el negocio no marcha como él espera, desea saber cómo está el ausentismo entre sus trabajadores, y ver si esa es la causa de la baja en las ventas. A continuación se da el número de días de ausencia durante una quincena en una muestra de 10 trabajadores 4,1, 2, 2, 1, 2, 2, 1, 0, 3

- ✍ Determine la media y desviación estándar de la muestra
- ✍ ¿Cuál la mejor estimación de ese valor?
- ✍ Proporcione un intervalo de confianza de 95 % para la media poblacional
- ✍ Explique porque se usa la distribución t como parte del intervalo de confianza
- ✍ ¿Es razonable concluir que el trabajador promedio no faltó ningún día durante una quincena?

Media	Desviación Estándar	Varianza
1.8	1.135	1.289

Se obtiene el coeficiente y grados libertad

$$\alpha = 1 - .95 = .05/2 = 0.025$$

$$gl = n - 1 = 9$$

Buscando en la tabla “t”

$n = 9$ y $\alpha = 0.025$ se encuentra el valor **2.262**

$$\bar{x} = 1.8$$

$$S = \sqrt{S^2} \quad S = 1.13$$

$$n = 10$$

$$Ic = \bar{x} \pm t S_{\bar{x}}$$

$$Ic = 1.8 + (2.262) (.35) = 2.612$$

$$Ic = 1.8 - (2.262) (.35) = 0.988$$

Respuesta, la verdadera media poblacional de ausencia en una quincena va de los 0.988 a los 2.612 días.

¿Es razonable concluir que el trabajador promedio no faltó ningún día durante una quincena? No, porque según el resultado el intervalo está entre los valores (0.988, 2.612) y el "0" se encuentra fuera del intervalo, por lo tanto no es razonable pensar que hubo cero ausencias en la quincena.



Ejemplo 2: Una cámara de comercio quiere determinar cuánto tiempo necesitan los empleados para llegar a su trabajo. Los siguientes datos en minutos corresponden a una muestra de 15 empleados: 29, 39, 38, 33, 38, 21, 45, 34, 40, 37, 37, 42, 30, 29, 35. Determine un intervalo de confianza de 98% para la media poblacional, interprete el resultado.

$$\alpha = 1 - .98 = .02/2 = 0.01 \text{ con } 14 \text{ grados libertad} = \text{En tabla } t = 2.262$$

$$S = 6.06$$

$$\bar{x} = 35.13$$

$$n = 15$$

$$S_{\bar{x}} = \frac{6.06}{\sqrt{15}} = 1.56$$

$$Ls = IC_{\alpha=98\%} = 35.13 + (2.262) (1.56) = 39.24$$

$$Li = IC_{\alpha=98\%} = 35 - (2.262) (1.56) = 31.03 \quad (31.03, 39.24)$$

Lo que significa que un empleado tarda en promedio de 31 a 39 minutos aproximadamente para llegar a su trabajo.



4.5 INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS MEDIAS, CON EL USO DE LA DISTRIBUCIÓN NORMAL Y “t” DE STUDENT.



Una empresa comercial que procesa muchos de sus pedidos por teléfono tiene 2 tipos de clientes: generales y comerciales. Se recogen los pedidos de tiempo telefónico por artículo requerido, por una muestra aleatoria de 12 llamadas de clientes generales y 10 llamadas de clientes comerciales. Se supone que las cantidades de tiempos para cada tipo de llamadas tiene una distribución aproximadamente normal. Obtenga el Intervalo de Confianza de 95% para la diferencia de la cantidad media de tiempo por artículo requerida para cada llamada

Clientes generales

48
66
106
84
146
139
154
150
177
156
122
121
1469

$$\bar{x}_1 = 122.42$$

$$s_1^2 = 1560.44$$

$$s_1 = 39.50$$

$$n_1 = 12$$

Clientes Comerciales

81
137
107
110
107
40
154
142
34
165
1077

$$\bar{x}_2 = 107.7$$

$$s^2 = 2021.78$$

$$s_2 = 44.96$$

$$n_2 = 10$$

$$\alpha = \frac{.05}{2} = .025 = t = 2.086$$

$$s^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} = \frac{(12-1)39.50^2 + (10-1)44.96^2}{10+12-2} = \frac{+437500}{20} = \frac{797,500}{16}$$

$$s^2 = \frac{(12-1)39.50^2 + (10-1)44.96^2}{12+10-2}$$

$$s^2 = \frac{17,162.75 + 18,192.61}{20} = \frac{35,355.36}{20} \quad s^2 = 1,767.76$$

$$s_{\bar{x}_1 - \bar{x}_2} = \frac{\sqrt{1767.76}}{20} + \frac{\sqrt{1767.76}}{10} \quad s_{\bar{x}_1 - \bar{x}_2} = \sqrt{147.31 + 176.77} \quad s_{\bar{x}_1 - \bar{x}_2} = \sqrt{324.08} = 18.00$$

$$(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t s_{\bar{x}_1 - \bar{x}_2} = 14.72 \pm 2.086(18) = 14.72 \pm 37.55$$

(-22.83, 52.27)

Interpretación: como el cero se encuentra incluido en el intervalo, se puede decir con un 95% de confianza que no hay diferencia en el tiempo medio de cada llamada requerida para cada artículo.



4.6 UNA SOLA MUESTRA: ESTIMACIÓN DE LA PROPORCIÓN



EJEMPLO. Se elige una muestra de 2000 electores potenciales en el Estado de México; se encontró que 1550 planean votar por el gobernador actual para presidente de la república. En una encuesta previa se determinó que el 80% de la población total del padrón votante elegiría a dicho candidato. ¿Cuál será la probabilidad de que más del 77.5% de la población lo elija presidente?

$$p = \frac{x}{n} = P(A) = \frac{1550}{2000} = .775$$

$$P = .80$$

$$Z = \frac{.775 - .80}{\sqrt{\frac{(.775)(.225)}{2000}}} = -2.67$$

$$\bar{p} = .775$$

$$Z = \frac{\bar{p} - P}{\sqrt{\bar{p}}}$$

$$n = 2000$$

$$q = .225$$

$$P(\bar{p} > .775) = .5 + .4962 = 0.9962$$

Hay un 99.62% de probabilidad de ganar la presidencia de la república



INTERVALO DE CONFIANZA PARA ESTIMAR UNA PROPORCIÓN



EJEMPLO. Una compañía textil produce pantalones para hombre, los pantalones se confeccionan y venden con corte regular o con corte de bota. En un esfuerzo por estimar la proporción del mercado de sus pantalones para hombre en el centro de la ciudad que prefiere pantalones con corte de bota, el analista toma una muestra aleatoria de 212 ventas de pantalones de las 2 tiendas de venta al público de la ciudad, solo 34 de las ventas fueron de pantalones de corte de bota. Construya un intervalo de confianza de 90% para estimar la proporción de la población en toda la ciudad que prefieren pantalones con corte de bota.

$$P(A) = \frac{34}{212} = 0.16 \quad n = 212 \quad \bar{p} = 0.16$$

$$\sigma_{\bar{p}} = \frac{\sqrt{pq}}{n}$$

$$\sigma_{\bar{p}} = \frac{\sqrt{(0.16)(0.84)}}{212}$$

$$\sigma_{\bar{p}} = 0.025$$

$$P = IC_{\alpha} = \bar{p} \pm z\sigma_{\bar{p}}$$

$$IC = 0.16 \pm (1.65)(0.025) = 0.16 \pm 0.041$$

$$IC = 0.16 + 0.041 = 0.2015$$

$$IC = 0.16 - 0.041 = 0.1190$$

Conclusión. La proporción de la población que prefiere los pantalones corte bota va del 11% al 20% de la población.



PROBLEMAS

Use la información sobre cada una de las siguientes muestras para calcular el intervalo de confianza para estimar la proporción de la población.

- a) $n = 44$ $\bar{p} = .51$; calcule un intervalo de confianza del 99%
- b) $n = 300$ $\bar{p} = .82$; calcule un intervalo de confianza del 95%
- c) $n = 1,150$ $\bar{p} = .48$; calcule un intervalo de confianza del 90%
- d) $n = 95$ $\bar{p} = .32$; calcule un intervalo de confianza del 88%



Muchas aplicaciones involucran poblaciones de datos cualitativos que deben compararse utilizando proporciones o porcentajes. A continuación se citan algunos ejemplos:

- **Educación.**- ¿Es mayor la proporción de los estudiantes que aprueban matemáticas que las de los que aprueban inglés?
- **Medicina.**- ¿Es menor el porcentaje de los usuarios del medicamento A que presentan una reacción adversa que el de los usuarios del fármaco B que también presentan una reacción de ese tipo?
- **Administración.**- ¿Hay diferencia entre los porcentajes de hombres y mujeres en posiciones gerenciales?
- **Ingeniería.**- ¿Existe diferencia entre la proporción de artículos defectuosos que genera la máquina A los que genera la máquina B?



Quando el muestreo procede de dos poblaciones binomiales y se trabaja con dos proporciones muestrales, la distribución muestral de diferencia de proporciones es aproximadamente normal para tamaños de muestra grande ($n_1p_1 \geq 5$, $n_1q_1 \geq 5$, $n_2p_2 \geq 5$ y $n_2q_2 \geq 5$). Entonces p_1 y p_2 tienen distribuciones muestrales aproximadamente normales, así que su diferencia $p_1 - p_2$ también tiene una distribución muestral aproximadamente normal.

Formula:

$$Z = \frac{(p_1 - p_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$$



Ejemplo: Los hombres y mujeres adultos radicados en una ciudad grande del norte difieren en sus opiniones sobre la promulgación de la pena de muerte para personas culpables de asesinato. Se cree que el 12% de los hombres adultos están a favor de la pena de muerte, mientras que sólo 10% de las mujeres adultas lo están. Si se pregunta a dos muestras aleatorias de 100 hombres y 100 mujeres su opinión sobre la promulgación de la pena de muerte, determine la probabilidad de que el porcentaje de hombres a favor sea al menos 3% mayor que el de las mujeres.

Solución:

Datos:

$$PH = 0.12$$

$$PM = 0.10$$

$$NH = 100$$

$$P(P_H - P_M) \geq 0.03$$

Se recuerda que se está incluyendo el factor de corrección de 0.5 por ser una distribución binomial y se está utilizando la distribución normal.

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} = \frac{0.025 - (0.12 - 0.10)}{\sqrt{\frac{(0.12)(0.88)}{100} + \frac{(0.10)(0.90)}{100}}} = 0.11$$

$$p_H - p_M = 0.02 \quad p_H - p_M = 0.03 \quad 0.03 - \left(\frac{0.5}{100}\right) = 0.025$$

Se concluye que la probabilidad de que el porcentaje de hombres a favor de la pena de muerte, al menos 3% mayor que el de mujeres es de 0.4562.



4.8 TAMAÑO DE LA MUESTRA COMO UNA ESTIMACIÓN DE P Y UN GRADO DE CONFIANZA $(1 - \alpha)$ 100%.



Determinación del tamaño de la muestra que se requiere para estimar la proporción.

Antes de tomar una muestra se puede determinar el tamaño de la muestra mínimo requerido especificando el nivel de confianza que desea, el error de muestreo aceptable y haciendo una estimación inicial (subjetiva) de π la proporción poblacional desconocida:

$$n = \frac{z^2 \pi (1 - \pi)}{E^2}$$

En esta ecuación z es el valor para el intervalo de confianza especificado, π es una estimación inicial de la proporción poblacional y E es el error del muestreo es mas y en menos tolerado por el intervalo (siempre un medio de todo intervalo de confianza)

Si no es posible hacer una estimación inicial de π , entonces se debe estimar que es .50. Esta estimación es conservadora ya que es el valor para el que se requiere mayor tamaño para la muestra. Bajo esta suposición la formula general para el tamaño de la muestra se simplifica como sigue:

$$n = \left(\frac{z}{2E}\right)^2$$

Cuando se calcula el tamaño de la muestra cualquier resultado fraccionario se redondea siempre hacia arriba.

Además cualquier tamaño de muestra menor que 100 que se obtenga con los cálculos debe incrementarse a 100 debido a que las formulas se basan en el uso de la distribución normal.



Ejemplo: Suponga que se especifica que la estimación mediante un intervalo de 95% debe ser ± 0.05 y que no se hace ninguna suposición previa acerca del posible valor de π . El tamaño mínimo de la muestra que debe tomarse es:

$$n = \left(\frac{z}{2E} \right)^2 = n = \left(\frac{1.96}{.10} \right)^2 = (19.6)^2 = 384.16 = 385$$

Además de estimar la proporción poblacional, también se puede estimar el número total en una categoría de la población.



EJERCICIOS

1. Se probó una muestra aleatoria de 400 pantallas planas de computadora y se encontraron 40 defectuosas. Estime el intervalo que contiene, con un coeficiente de confianza de 90%, a la verdadera fracción de elementos defectuosos.
2. Se planea realizar un estudio de tiempos para estimar el tiempo medio de un trabajo, exacto dentro de 4 segundos y con una probabilidad de 0.90, para terminar un trabajo de montaje. Si la experiencia previa sugiere que $\sigma=16$ segundos mide la variación en el tiempo de montaje entre un trabajador y otro al realizar una sola operación de montaje, ¿cuántos operarios habrá que incluir en la muestra?
3. El decano registró debidamente el porcentaje de calificaciones 6 y 7 otorgadas a los estudiantes por dos profesores universitarios de estadística. El profesor I alcanzó un 32%, contra un 21% para el profesor II, con 200 y 180 estudiantes, respectivamente. Estime la diferencia entre los porcentajes de calificaciones 6 y 7 otorgadas por los dos profesores. Utilice un nivel de confianza del 95% e interprete los resultados.
4. Suponga que se quiere estimar la producción media por hora, en un proceso que produce antibiótico. Se observa el proceso durante 100 períodos de una hora, seleccionados al azar y se obtiene una media de 34 onzas por hora con una desviación estándar de 3 onzas por hora. Estime la producción media por hora para el proceso, utilizando un nivel de confianza del 95%.
5. Un ingeniero de control de calidad quiere estimar la fracción de elementos defectuosos en un gran lote de lámparas. Por la experiencia, cree que la fracción real de defectuosos tendría que andar alrededor de 0.2. ¿Qué tan

- grande tendría que seleccionar la muestra si se quiere estimar la fracción real, exacta dentro de 0.01, utilizando un nivel de confianza de 95%?
6. Se seleccionaron dos muestras de 400 tubos electrónicos, de cada una de dos líneas de producción, A y B. De la línea A se obtuvieron 40 tubos defectuosos y de la B 80. Estime la diferencia real en las fracciones de defectuosos para las dos líneas, con un coeficiente de confianza de 0.90 e interprete los resultados.
 7. Se tienen que seleccionar muestras aleatorias independientes de $n_1=n_2=n$ observaciones de cada una de dos poblaciones binomiales, 1 y 2. Si se desea estimar la diferencia entre los dos parámetros binomiales, exacta dentro de 0.05, con una probabilidad de 0.98. ¿qué tan grande tendría que ser n ? No se tiene información anterior acerca de los valores P_1 y P_2 , pero se quiere estar seguro de tener un número adecuado de observaciones en la muestra.
 8. Se llevan a cabo pruebas de resistencia a la tensión sobre dos diferentes clases de largueros de aluminio utilizados en la fabricación de alas de aeroplanos comerciales. De la experiencia pasada con el proceso de fabricación se supone que las desviaciones estándar de las resistencias a la tensión son conocidas. La desviación estándar del larguero 1 es de 1.0 Kg/mm^2 y la del larguero 2 es de 1.5 Kg/mm^2 . Se sabe que el comportamiento de las resistencias a la tensión de las dos clases de largueros son aproximadamente normal. Se toma una muestra de 10 largueros del tipo 1 obteniéndose una media de 87.6 Kg/mm^2 , y otra de tamaño 12 para el larguero 2 obteniéndose una media de 74.5 Kg/mm^2 . Estime un intervalo de confianza del 90% para la diferencia en la resistencia a la tensión promedio.
 9. Se quiere estudiar la tasa de combustión de dos propelentes sólidos utilizados en los sistemas de escape de emergencia de aeroplanos. Se sabe que la tasa de combustión de los dos propelentes tiene aproximadamente la misma desviación estándar; esto es $\sigma_1 = \sigma_2 = 3 \text{ cm/s}$. ¿Qué tamaño de muestra debe utilizarse en cada población si se desea que el error en la estimación de la diferencia entre las medias de las tasas de combustión sea menor que 4 cm/s con una confianza del 99%?

Respuesta a los Problemas propuestos

1. $0.07532 \leq P \leq 0.1246$
2. $n = 44$
3. $0.0222 \leq P_1 - P_2 \leq 0.1978$
4. $33.412 \leq \mu \leq 34.588$
5. $n = 6147$
6. $0.059 \leq P_B - P_A \leq 0.141$
7. $n = 1086$
8. $12.22 \leq \mu_1 - \mu_2 \leq 13.98$
9. $n = 8$



UNIDAD V. PRUEBA DE HIPÓTESIS.

Propósitos de la unidad

En esta unidad el alumno debe:

- ✍ Comprender la teoría de las hipótesis estadísticas nula y alternativa.
- ✍ Aplicar los conceptos de error tipo I y II para el planteamiento del problema.
- ✍ Establecer y probar pruebas de hipótesis relativas a medias y proporciones.
- ✍ Diferenciar y aplicar las pruebas de hipótesis sobre dos medias de muestras independientes utilizando la distribución normal y "t" student.
- ✍ Aplicar las pruebas de hipótesis sobre la diferencia de dos proporciones.
- ✍ Aplicar la prueba de hipótesis, para pruebas dependientes. (pareadas)

Competencia específica

Aplica el uso de las pruebas de hipótesis y reconoce la potencia de dichas pruebas para inferir características poblacionales

Aplica pruebas de hipótesis con dos o más poblaciones para inferir características de las mismas



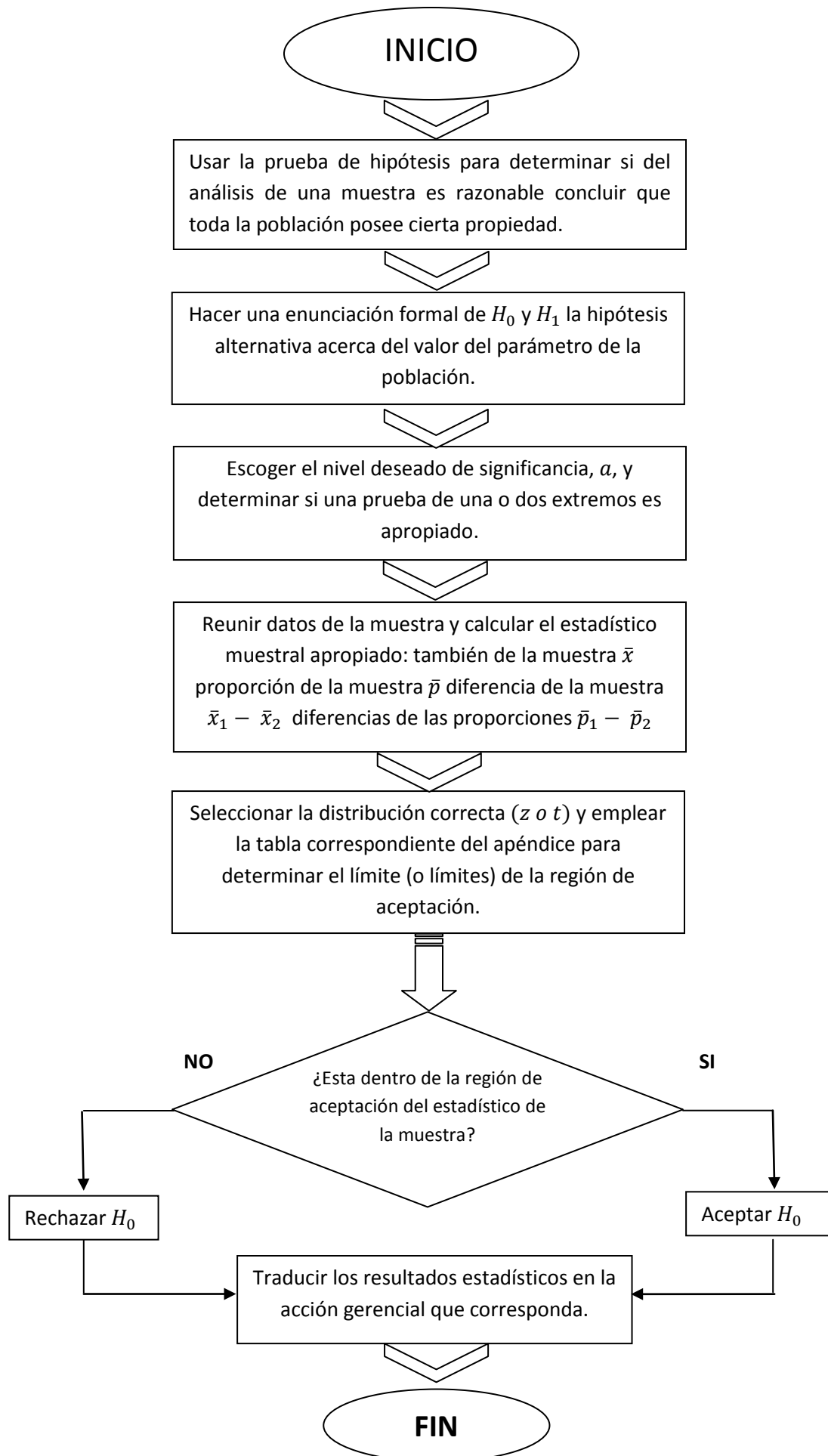
Introducción

Al intentar alcanzar una decisión, es útil hacer hipótesis (o conjeturas) sobre la población aplicada. Tales hipótesis, que pueden ser o no ciertas, se llaman hipótesis estadísticas. Son, en general, enunciados acerca de las distribuciones de probabilidad de las poblaciones.

En muchos casos formulamos una hipótesis estadística con el único propósito de rechazarla o invalidarla. Análogamente, si deseamos decidir si un procedimiento es mejor que otro, formulamos la hipótesis de que no hay diferencia entre ellos (o sea. Que cualquier diferencia observada se debe simplemente a fluctuaciones en el muestreo de la misma población). Tales hipótesis se suelen llamar hipótesis nula y se denotan por H_0 .



Al responder a un problema, es muy conveniente proponer otras hipótesis en que aparezcan variables independientes distintas de las primeras que formulamos. Por tanto, para no perder tiempo en búsquedas inútiles, es necesario hallar diferentes hipótesis alternativas como respuesta a un mismo problema y elegir entre ellas cuáles y en qué orden vamos a tratar su comprobación





HIPÓTESIS ESTADÍSTICAS

Para todo tipo de investigación en la que tenemos dos ó más grupos, se establecerá una hipótesis nula. La hipótesis nula es aquella que nos dice que no existen diferencias significativas entre los grupos.

Por ejemplo, supongamos que un investigador cree que si un grupo de jóvenes se somete a un entrenamiento intensivo de natación, éstos serán mejores nadadores que aquellos que no recibieron entrenamiento. Para demostrar su hipótesis toma al azar una muestra de jóvenes, y también al azar los distribuye en dos grupos: uno que llamaremos experimental, el cual recibirá entrenamiento, y otro que no recibirá entrenamiento alguno, al que llamaremos control. La hipótesis nula señalará que no hay diferencia en el desempeño de la natación entre el grupo de jóvenes que recibió el entrenamiento y el que no lo recibió.



Una hipótesis nula es importante por varias razones:

Es una hipótesis que se acepta o se rechaza según el resultado de la investigación. El hecho de contar con una hipótesis nula ayuda a determinar si existe una diferencia entre los grupos, si esta diferencia es significativa, y si no se debió al azar.

No toda investigación precisa de formular hipótesis nula. Recordemos que la hipótesis nula es aquella por la cual indicamos que la información a obtener es contraria a la hipótesis de trabajo.

Al formular esta hipótesis, se pretende negar la variable independiente. Es decir, se enuncia que la causa determinada como origen del problema fluctúa, por tanto, debe rechazarse como tal.

HIPÓTESIS ALTERNATIVA.

Toda hipótesis que difiere de una dada se llamará una hipótesis alternativa. Por ejemplo: Si una hipótesis es $p = 0.5$, la hipótesis alternativa podrían ser $p = 0.7$, $p < 0.5$ ó $p > 0.5$. Una hipótesis alternativa a la hipótesis nula se denotará por H_1 .



Es importante recordar que las hipótesis siempre son proposiciones sobre la población o distribución bajo estudio, proposiciones sobre la muestra. Por lo general, el valor del parámetro de la población especificado en la hipótesis nula se determina en una de tres maneras diferentes:

1. Puede ser resultado de la experiencia pasada o del conocimiento del proceso, entonces el objetivo de la prueba de hipótesis usualmente es determinar si ha cambiado el valor del parámetro.

2. Puede obtenerse a partir de alguna teoría o modelo que se relaciona con el proceso bajo estudio. En este caso, el objetivo de la prueba de hipótesis es verificar la teoría o modelo

3. Cuando el valor del parámetro proviene de consideraciones externas tales como las especificaciones de diseño o ingeniería, o de obligaciones contractuales. En esta situación, el objetivo usual de la prueba de hipótesis es probar el cumplimiento de las especificaciones.

Los procedimientos de prueba de hipótesis dependen del empleo de la información contenida en la muestra aleatoria de la población de interés.



5.2 ERROR TIPO UNO I Y TIPO II EN PRUEBAS DE HIPÓTESIS

La probabilidad máxima de error tipo I se designa con la letra griega α alfa. Esta probabilidad es siempre igual al nivel de significancia que se usa para probar la hipótesis nula. Esto se debe a que por definición la proporción de área en la región de rechazo es igual a la proporción de resultados muestrales que se darían en esa región dado que la hipótesis nula fuera verdadera.



Ejemplo. La hipótesis nula es que la media de todas las cuentas por cobrar es de \$ 260 y la hipótesis alternativa es que la media sea menor que esta cantidad; la prueba se realiza con 5% como nivel de significancia. El auditor indica, además, que una media verdadera de \$ 240 ó menos, sería considerada como diferencia importante en relación con el valor hipotético \$260. Como antes $\sigma = \$43$ y el tamaño de la muestra es $n = 36$ cuentas.

La determinación de la probabilidad del error tipo II requiere:

- Formular la hipótesis nula y alternativa de esta prueba.
- Determinar el valor crítico de la media muestral necesario para probar la hipótesis nula con 5% de nivel de significancia
- Determinar la probabilidad del error tipo I correspondiente al uso del valor crítico arriba calculado como base para la regla de decisión.
- Determinar la probabilidad del error tipo II correspondiente a la regla de decisión dado el valor alternativo para la media \$240

Solución:

$$1.- H_0: \mu = \$260.00$$

$$H_1: \mu < \$260.00$$

$$2. \quad \bar{X}_{CR} = \mu_0 \pm Z\sigma_{\bar{X}} = 260 + (-1.641)(7.17) = 248.21$$

$$\text{Donde: } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{43}{\sqrt{36}} = \frac{43}{6} = 7.17$$

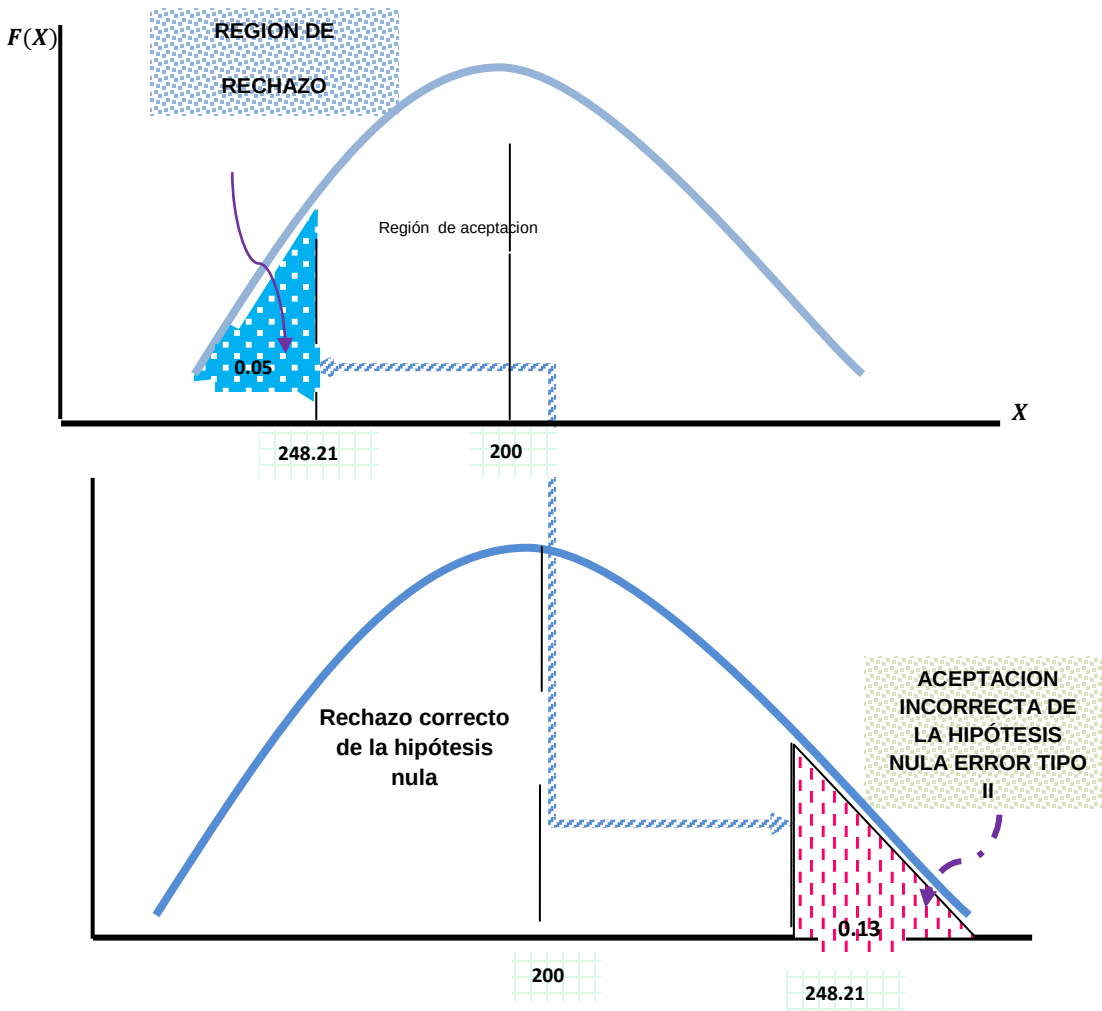
3.- La probabilidad máxima de error tipo I es igual a 0.05 (el nivel de significancia que se usa para probar la hipótesis nula)

4.- La probabilidad del error tipo II es la probabilidad de que la media de la muestra aleatoria sea mayor o igual que \$284.21, dado que la media de todas las cuentas en realidad es \$240.

$$Z = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{X}}} = \frac{248.21 - 240}{7.17} = \frac{8.21}{7.17} = 1.15$$

$$P(\text{error tipo II}) = P(z \geq 1.15) = 0.5000 - 0.3749 = 0.1251 = 0.13$$

Manteniendo constantes el nivel de significancia y el tamaño de la muestra, la probabilidad de error tipo II disminuye a medida que el valor alternativo para la media se elige más alejado de la hipótesis nula y aumenta a medida que este valor alternativo se elige más cerca del valor de la hipótesis nula.





Ejemplo. Suponga que el desarrollador consideraría discrepancia grave el hecho de que el ingreso doméstico promedio fuera de inferior a \$43,500, en lugar del nivel de ingreso propuesto, que es \$45,000.

Determine:

- a) la probabilidad del error tipo I,
- b) la probabilidad del error tipo II.
- c) La potencia asociada con esta prueba de la cola inferior

Solución:

a) P (error tipo I) = 0.05 (nivel α , ó nivel de significancia)

b) P (error tipo II) = P (el valor critico $\bar{X}$ sea sobrepasado dado que $\mu = \$43,500$)

El valor critico inferior $\bar{X} = \mu_0 + Z\sigma_{\bar{X}} = 45000 + (-1.645)(516.80) = \$44,149.86$

Donde $\mu_0 = \$45000$

$Z = -1.645$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2000}{\sqrt{15}} = \frac{43}{3.87} = \$516.80$$

$P(\text{error tipo II}) = P(\bar{X} \geq \$44,149.86)$

$\mu_1 = 43000$

$\sigma_{\bar{X}} = 516.80$

$$Z = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{X}}} = \frac{44,149.86 - 43,500}{516.80} = \frac{649.86}{516.80} = 1.26$$

$P(\text{error tipo II}) = P(z \geq +1.26) = 0.500 - 0.3962 = 0.1038 = 0.10$

c) Potencia = 1 - P (error tipo II) = 1 - .10 = .90



5.3 PRUEBAS UNILATERALES Y BILATERALES

Se pueden presentar dos tipos de pruebas de hipótesis que son:

1. De dos colas, o bilateral.
2. De una cola, o unilateral.

Este último puede ser de cola derecha o izquierda. La hipótesis es una afirmación sobre un parámetro de la población, como la media, la varianza o la desviación estándar. La hipótesis inicial que se define sobre la población se llama hipótesis nula; pero si rechazamos esa hipótesis nula debemos tener una hipótesis alternativa, la cual tomaremos si la hipótesis inicial o nula es falsa.

El proceso de revisión de la hipótesis para determinar si se considera Verdadera o falsa se llama Prueba de Hipótesis. Una prueba de hipótesis es una regla que especifica

1. Para que valores de la muestra se toma la decisión de que H_0 es verdadera.
2. Para que valores de la muestra se rechaza H_0 y se acepta H_1 como verdadera.



PRUEBAS UNILATERALES



Ejemplo. Suponga que el auditor parte de la hipótesis alternativa de que el valor medio de todas las cuentas por cobrar es menor que \$260. Dado que la media muestral es \$240, a continuación se prueba esta hipótesis con un 5% como nivel de significancia mediante los procedimientos siguientes.

Determinando el valor crítico para la media muestral, cuando

$$H_0: \mu = \$260.00$$

$$H_1: \mu < \$260.00)$$

$$\bar{X}_{CR} = \mu_0 \pm Z\sigma_{\bar{X}} = 260 + (-1.641)(7.17) = 248.21$$

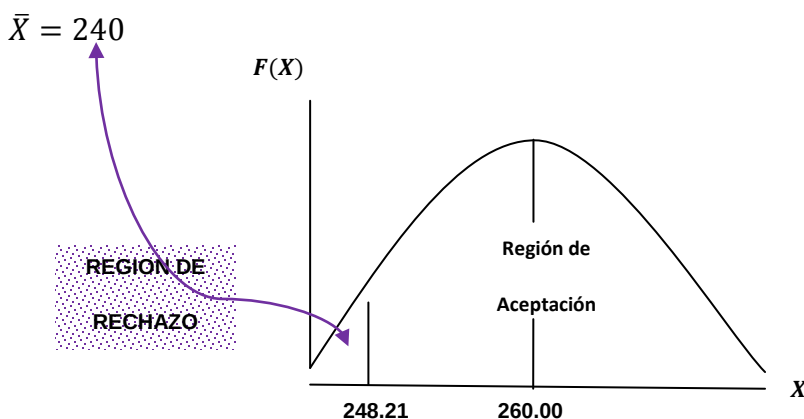
Como $\bar{X} = 240$ este valor se encuentra en la región de rechazo. Por tanto se rechaza la hipótesis nula y se acepta la hipótesis alternativa $\mu < \$260$.

Determinando el valor crítico en términos de Z, donde z crítico

$(\alpha = 0.05) = -1.645$:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{240 - 260}{7.17} = \frac{-20}{7.17} = -2.79$$

Como $Z = -2.79$, esta región de rechazo a la izquierda del valor crítico -1.64 , la hipótesis nula se rechaza. Y esto se representa en la grafica siguiente.



PRUEBAS BILATERALES PASOS BÁSICOS EN LAS PRUEBAS DE HIPÓTESIS USANDO EL MÉTODO DE VALOR CRÍTICO

 **Ejemplo 1:** Un auditor toma una muestra de $n = 36$ y calcula la media muestral, desea probar la suposición de que el valor medio de todas las cuentas por cobrar en una determinada empresa sea \$260.00. El auditor desea rechazar este valor supuesto de \$260.00 solo si la media muestral lo contradice claramente, y así, en este procedimiento de prueba, al valor hipotético deberá otorgársele el beneficio de la duda.

 **Paso 1.** Formular la hipótesis nula y la hipótesis alternativa. La hipótesis nula H_0 es valor paramétrico hipotético que se compara con el resultado muestral. La hipótesis nula se rechaza solo si es poco probable que el resultado muestral se dé siendo la hipótesis correcta. La hipótesis alternativa H_1 se acepta solo si la hipótesis nula se rechaza. Las hipótesis nulas y alternativa en esta prueba son:

$$H_0: \mu = \$260.00$$

$$H_1: \mu \neq \$260.00.$$

Paso 2. Especificar el nivel de significancia que habrá de usarse. El nivel de significancia es el criterio estadístico que se establece para rechazar la hipótesis nula. Si se establece $\alpha = 5\%$ como nivel de significancia, entonces la hipótesis nula se rechaza solo si el resultado muestral es tan diferente del valor hipotético que la probabilidad de que una diferencia de esa magnitud o mayor se dé por casualidad es de por casualidad es de 0.05 o menos.

Observe que si se usa como nivel de significancia 5%, existe una probabilidad de 0.05 de rechazar la hipótesis nula aun cuando sea verdadera. A esto se le conoce como error tipo I. La probabilidad de un error de tipo I es siempre igual al nivel de significancia que se utiliza como criterio para rechazar la hipótesis nula; al error tipo I se le designa mediante la letra griega minúscula α alfa y entonces α también designa el nivel de significancia. Un error de tipo II ocurre cuando no se rechaza la hipótesis nula, y por lo tanto se acepta, siendo falsa.

		Situaciones posibles.	
		Hipótesis nula verdadera	Hipótesis nula falsa
Aceptar la hipótesis nula	la	Aceptación correcta	Error tipo II
Rechazar la hipótesis nula	la	Error tipo I	Rechazo correcto

Paso 3. Elegir el estadístico de prueba. El estadístico de prueba es el estadístico muestral o una versión estandarizada del estadístico muestral. Por ejemplo, con objeto de probar un valor hipotético de la media poblacional, como estadístico de prueba puede emplearse la media de una muestra aleatoria tomada de esa población. Sin embargo, si la distribución de muestreo para la media tiene distribución normal, entonces es común que el valor de la media muestral se convierta a un valor Z el cual sirve entonces como estadístico de prueba.

Paso 4. Establecer el valor o los valores críticos del estadístico de prueba. Una vez especificados la hipótesis nula, el nivel de significancia y el estadístico de prueba que se usaran, se establecen los valores críticos del estadístico de prueba. Puede haber uno o dos de estos valores, dependiendo de si se trata de una prueba unilateral o bilateral. En cualquiera de los dos casos un valor crítico establece el valor del estadístico de prueba que se requiere para rechazar la hipótesis nula.

Paso 5. Determinar el valor del estadístico de prueba. Por ejemplo, al probar un valor hipotético para la media poblacional se toma una muestra aleatoria y se determina el valor de la media muestral. Si el valor crítico se fija como un valor Z , entonces la media muestral se convierte a un valor Z .

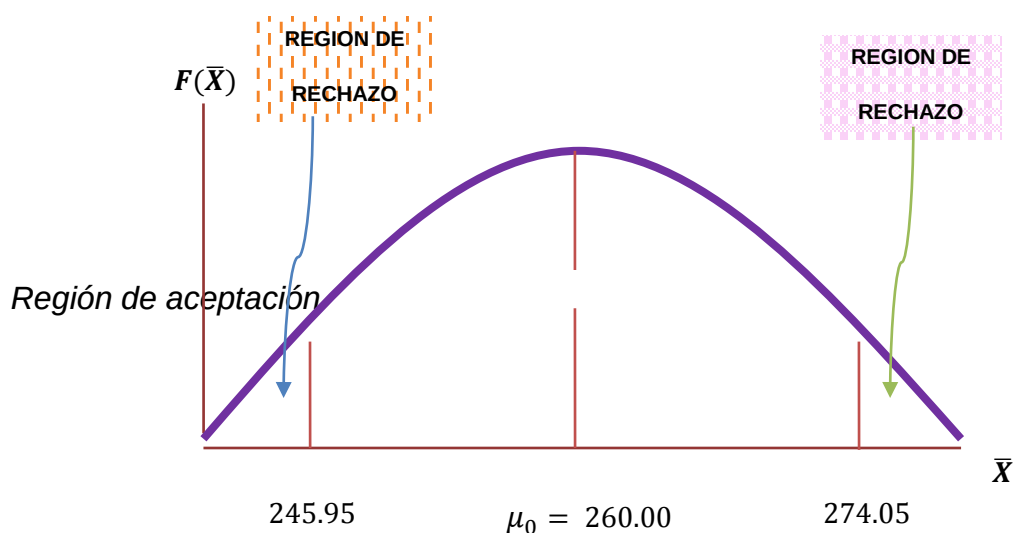
Paso 6. Tomar la decisión. El valor del estadístico muestral obtenido se compara con los valores críticos del estadístico de prueba. A continuación la hipótesis nula se acepta o se rechaza. Si se rechaza la hipótesis nula, se acepta la alternativa. La distribución de probabilidad normal se puede usar para probar un valor hipotético para la media poblacional siempre que $n \geq 30$, debido al teorema del límite central, ó cuando $n < 30$ pero la población tiene distribución y se conoce σ .

Fórmula para calcular valores críticos $\bar{X}_{CR} = \mu_0 \pm Z\sigma_{\bar{X}}$

Dada la hipótesis nula formulada anteriormente, determine los valores críticos para la media muestral si se quiere probar la hipótesis con un nivel de significancia $\alpha = 5\%$. Dado que se sabe que la desviación estándar de los montos de las cuentas por cobrar es $\sigma = \$43.00$ los valores críticos son:

$$\begin{aligned}\bar{X}_{CR} &= \mu_0 \pm Z\sigma_{\bar{X}} = 260 \pm 1.96 \frac{\sigma}{\sqrt{n}} = 260 \pm 1.96 \frac{43}{\sqrt{36}} \\ &= 260 \pm 1.96 (7.17) = 260 \pm 14.05 = \$245.95 \text{ y } \$274.05\end{aligned}$$

Por tanto, para rechazar la hipótesis nula la media muestral debe tener un valor menor que \$245.95 o mayor que \$274.05. Así, en el caso de una prueba bilateral hay dos regiones de rechazo. Los valores $Z \pm 1.96$ se usan para establecer los valores críticos, debido a que en la distribución normal estándar en las dos colas queda una proporción de 0.05 del área, lo que corresponde al valor $\alpha = 0.05$ que se fijó.



En las pruebas de hipótesis los valores críticos suelen especificarse en términos de valores de Z en lugar de establecer en términos de la media muestral. Por ejemplo, los valores críticos Z para el nivel de significancia de 5% en la prueba bilateral son -1.96 y +1.96. Cuando se determina el valor de la media muestral, este se convierte a un valor Z de modo que este valor pueda compararse con los valores críticos Z. La formula de conversión, de acuerdo con si se conoce o no σ , es:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} \quad \text{ó si se desconoce } \sigma^2 \text{ se utilizará } S^2 \quad Z = \frac{\bar{X} - \mu_0}{S_{\bar{X}}}$$

En el mismo problema de la prueba de hipótesis, suponga que la media muestral es $\bar{X} = \$240$. Para determinar si se debe rechazar la hipótesis nula, esta media se convierte a un valor Z y se compara con los valores críticos ± 1.96 como sigue:

$$\sigma_{\bar{X}} = 7.17$$

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{240 - 260}{7.17} = \frac{-20}{7.17} = -2.79$$

En el modelo para las pruebas de hipótesis, este valor de Z se encuentra en la región de rechazo de la cola izquierda. Así la hipótesis nula se rechaza y se acepta la hipótesis alternativa

$$H_1: \mu \neq \$260.00 .$$



Ejercicio. El representante de un grupo comunitario le informa al posible desarrollador de un centro comercial al sur de la ciudad, el ingreso promedio por hogar en la zona es de \$45,000. Supongamos que puede asumirse que, para el tipo de zona del que se trata, el ingreso hogar tiene una distribución aproximadamente normal y que puede aceptarse que la desviación estándar es igual a \$2,000, con base a un estudio anterior. A partir de una muestra aleatoria de 15 hogares se determina que el ingreso domestico medio es $\bar{x} = \$44,000$. Pruebe la hipótesis nula $\mu = \$45,000$ estableciendo los limites críticos de la media muestral en términos de pesos y con un nivel de significancia del 5%. Pruebe la hipótesis del problema con la variable normal estándar Z como estadístico de prueba



5.4. PRUEBA DE UNA HIPÓTESIS: REFERENTE A LA MEDIA CON VARIANZA DESCONOCIDA UTILIZANDO LA DISTRIBUCIÓN NORMAL Y “t” DE STUDENT.



Ejemplo: La Comisión Federal de Electricidad publica cifras del número anual de Kilowatt-hora que gastan varios aparatos electrodomésticos.

Se afirma que una aspiradora gasta un promedio de 46 kilowatt-hora al año. Si una muestra aleatoria de 12 hogares que se incluye en un estudio planeado indica que las aspiradoras gastan un promedio de 42 kilowatt-hora al año con una desviación estándar de 11.9 kilowatt-hora, ¿esto sugiere con un nivel de significancia de 0.05 que las aspiradoras gastan, en promedio, menos de 46 kilowatt-hora anualmente? Suponga que la población de kilowatt-hora es normal.

Datos:

$$\bar{x} = 42, \quad s = 11.9 \quad n = 12 \quad \mu = 46$$

$$H_0: \mu = 46$$

$$H_1: \mu < 46$$

$$gl = n - 1$$

$$gl = 12 - 1 = 11$$

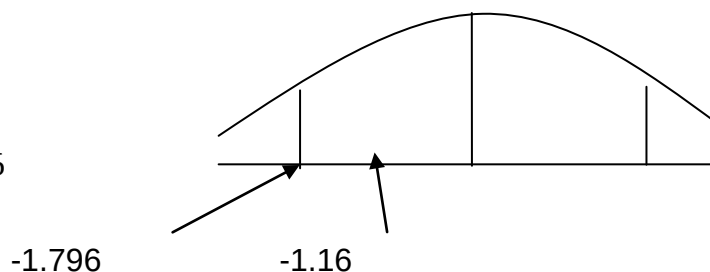
$$\alpha = \frac{.05}{2} = .025 \text{ valor } t = -1.796$$

Formula:

$$t = \frac{\bar{X} - \mu_0}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$t = \frac{42 - 46}{\frac{11.9}{\sqrt{12}}} = \frac{-4}{3.43} = -1.16$$



$\therefore$ aceptamos la H_0 : que establece que el promedio de kilowatt hora que gastan las aspiradoras al año no es significativamente menor a 46



Ejemplo 2: Una revista de negocios desea clasificar los aeropuertos internacionales de acuerdo con una evaluación hecha por la población de viajeros de negocios. Se usa una escala de valuación que va desde un mínimo de 0 hasta un máximo de 10, y aquellos aeropuertos que obtengan una media mayor que 7 serán considerados como aeropuertos de servicio superior. Para obtener datos de evaluación, el personal de la revista entrevista una muestra de 60 viajeros de negocios de cada aeropuerto. En la muestra tomada en el aeropuerto Heathrow de Londres la media muestral es $\bar{x} = 7.25$ y la desviación estándar es $s=1.052$. De acuerdo con estos datos muestrales. ¿Deberá ser designado el aeropuerto de Londres como un aeropuerto de servicio superior?

$$H_0: \mu = 7$$

$$H_1: \mu > 7$$

En esta prueba se usa como nivel de significancia $\alpha = .05$

$$gl = n - 1$$

$$gl = 60 - 1 = 59$$

$$\alpha = 59, .05 = 1.671$$

$$\bar{x} = 7.25, \quad S = 1.052, \quad n = 60, \quad \mu = 7$$

$$t = \frac{\bar{X} - \mu_0}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

$$t = \frac{7.25 - 7}{\frac{1.052}{\sqrt{60}}} = \frac{.25}{.135} = 1.84$$

$\therefore$ se rechaza H_0 : y se concluye que Heathrow se debe considerar como aeropuerto de servicio superior.



5.5. DOS MUESTRAS: PRUEBAS SOBRE MEDIAS UTILIZANDO LA DISTRIBUCIÓN NORMAL Y “t” DE STUDENT.



EJEMPLO DE DIFERENCIA DE DOS MUESTRAS UTILIZANDO LA DISTRIBUCIÓN NORMAL. El salario anual para una muestra de $n_1=50$ empleados de una empresa comercial del estado de México es de $\bar{x}_1 = \$190\,000$, con desviación estándar muestral de $\sigma_1 = \$10\,000$. En otra empresa grande del estado de Colima, una muestra aleatoria de $n_2 = 30$ empleados tiene un salario anual promedio de $\bar{x}_2 = \$170\,000$, con una desviación estándar muestral de $\sigma_2 = \$14\,000$. Se prueba la hipótesis nula de que no existe diferencia entre los salarios promedio anuales de las dos empresas, utilizando un nivel de significancia del 5% de la siguiente manera:

$$H_0: (\mu_1 - \mu_2) = 0$$

$$H_1: (\mu_1 - \mu_2) \neq 0$$

$$n_1 = 50 \quad n_2 = 30$$

$$Z \text{ Crítica } (\alpha = 0.05) = \pm 1.96$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{\$190\,000 - \$170\,000}{2917.1} = \frac{20\,000}{2917.1} = 6.85$$

Donde

$$\sigma_{\bar{x}_1} = \frac{\sigma_1}{\sqrt{n_1}} = \frac{10\,000}{\sqrt{50}} = \frac{10\,000}{7.10} = \$1,408.45 \quad \sigma_{\bar{x}_2} = \frac{\sigma_2}{\sqrt{n_2}} = \frac{14\,000}{\sqrt{30}} = \frac{14\,000}{5.477} = \$2,556.14$$

$$\begin{aligned} \sigma_{\bar{x}_1 - \bar{x}_2} &= \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} = \sqrt{(1,408.45)^2 + (2,556.14)^2} = \sqrt{1,975,289.7 + 6,533,851.7} \\ &= \sqrt{8,509,141.4} = 2917.1 \end{aligned}$$

+6.85 que salió de la distribución normal z, se encuentra en la región de rechazo de la hipótesis, que se encuentra en la gráfica presentada en la parte superior. Por ello se rechaza la hipótesis nula y se acepta la hipótesis alternativa de que el salario promedio anual de las dos empresas es diferente con un nivel de significancia del 5%.



EJEMPLO DE DIFERENCIA DE DOS MUESTRAS UTILIZANDO LA DISTRIBUCIÓN “t” DE STUDENT.

En una muestra aleatoria de $n_1=10$ focos el promedio de vida de los focos es $\bar{X}_1 = 4000$ horas, con una desviación de $S_1=200$ horas. Para otra marca de focos de cuya vida útil también se presume que sigue una distribución normal, una muestra aleatoria de $n_2= 8$ focos tiene una media muestral de $\bar{X}_2 = 4300$ horas y una desviación estándar muestral de $S_2 = 250$, pruebe la hipótesis de que no existe ninguna diferencia entre el ciclo medio de vida útil de las 2 marcas de focos con un nivel de significancia del 1%

$$n_1 = 10$$

$$\bar{X}_1 = 4000$$

$$S_1 = 200$$

$$n_2 = 8$$

$$\bar{X}_2 = 4300$$

$$S_2 = 250 \quad t = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{\sigma_{x_1 - x_2}}$$

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(10 - 1)200^2 + (8 - 1)250^2}{16} = \frac{360000 + 437500}{16} = \frac{797,500}{16}$$

$$= 49,843.75$$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$= \sqrt{\frac{49,843.75}{10} + \frac{49,843.75}{8}} = 105.90$$

$$t = \frac{(4000 - 4300) - (0)}{105.90} = \frac{-300}{105.90} = -2.83$$

$$\alpha = .01/2 = 0.005$$

$$gl = 10 + 8 - 2 = 16 = 2.921$$

∴ Como cae en la zona de aceptación se acepta con un nivel de confianza

del 99% la hipótesis nula, de que no existe diferencia entre las dos marcas de focos



EJERCICIOS

1. Un desarrollador considera dos ubicaciones alternativas para un centro comercial regional dado que el ingreso domestico de la comunidad es una consideración importante en la selección del sitio, él desea probar la hipótesis nula de que no existe ninguna diferencia entre los montos de ingreso domestico medio de las dos comunidades. Se supone que la desviación estándar del ingreso domestico también es igual en las dos comunidades. En una muestra de $n_1 = 30$ hogares de la primera comunidad el ingreso anual promedio es de $\bar{x}_1 = 45,500$ con una desviación estándar $S_1 = 1,800$. En una muestra de $n_2 = 40$ hogares de la segunda comunidad $\bar{x}_2 = 44,600$ y $S_2 = 2,400$. Pruebe la hipótesis nula al nivel de significancia de 5%.
2. Una muestra aleatoria de $n_1 = 12$ estudiantes de Contaduría tiene un promedio de calificación media de 2.70 (donde $A=4$) con una desviación estándar de .40 en el caso de los estudiantes de ingeniería en sistemas una muestra aleatoria de $n_2 = 10$ estudiantes tiene un promedio de calificación media de 2.90 con una desviación estándar de .30 se supone que los valores de calificación sigue una distribución normal ,pruebe la hipótesis nula de que el promedio de calificación de las 2 categorías de estimación no es diferente con un nivel de significancia de 5%
3. El salario medio diario de una muestra de $n_1=30$ empleados de una gran empresa manufacturera es $\bar{X}_1=280$, por una distribución estándar de 14 pesos. En otra gran empresa una muestra aleatoria $n_2=40$ empleados tiene un salario medio de $\bar{X}_2=270$ pesos, con una desviación estándar de 10 pesos. Pruebe la hipótesis de que no existe diferencia entre los montos salariales semanales medio de las dos empresas con un nivel de significancia del 5%.
4. La altura promedio de 50 palmas que tomaron parte de un ensayo es de 78 cm. con una desviación estándar de 2.5 cm.; mientras que otras 50 palmas que no forman parte tienen media y desviación estándar igual a 77.3 y desviación estándar poblacional de 2.8 cm. Se desea probar la hipótesis de que las palmas que participan en el ensayo son más altas que las otras.
5. Para una muestra aleatoria de $n_1 = 10$ lámparas de gas, se encuentra que la vida promedio es $\bar{x}_1 = 6000$ horas con $s_1 = 200$. Para otra marca de lámparas, para los cuales se supone también que tiene una vida útil con distribución normal, una muestra aleatoria de $n_2 = 15$ lámparas de gas tiene una media muestral de $\bar{x}_2 = 5600$ horas y una desviación estándar muestral de $s_2 = 250$. Pruebe la hipótesis de que no existe diferencia entre la vida útil promedio de las dos marcas de lámparas de gas, utilizando un nivel de significancia del 1%.



5.6 UNA MUESTRA PRUEBA SOBRE UNA SOLA PROPORCIÓN



Ejemplo: Se plantea la hipótesis de que no más del 5% de las refacciones que se fabrican en una empresa manufactura tienen defectos. Para una muestra aleatoria de $n = 200$ refacciones, se encuentran que 30 están defectuosas. Prueba la hipótesis nula al 5% del nivel de significancia.

$$H_0: \pi \leq 0.05$$

$$H_1: \pi > 0.05 \quad Z \text{ critica } (\alpha=0.05)=+1.645$$

$$\sigma_{\bar{p}} = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} = \sqrt{\frac{(0.05)(0.95)}{200}} = \sqrt{\frac{0.0475}{200}} = \sqrt{0.0002375} = 0.015$$

$$z = \frac{\hat{p} - \pi_0}{\sigma_{\bar{p}}} = \frac{0.10 - 0.05}{0.015} = \frac{0.05}{0.015} = 3.33$$

El valor calculado de z de 3.33 es mayor que el valor crítico de 1.645 para esta prueba del extremo superior. Por lo tanto, como se encuentran 30 refacciones defectuosas en el lote de 200, se rechaza la hipótesis de que la proporción de artículos defectuosos en la población es de 5% o menor, utilizando el nivel de significancia al 5% en la prueba.



Ejemplo 2: Se plantea la hipótesis de que no más del 5% de las refacciones que se fabrican en proceso de manufactura tienen defectos. Para una muestra aleatoria de $n = 100$ refacciones, se encuentran que 10 están defectuosas. Prueba la hipótesis nula al 5% del nivel de significancia.

$$H_0: \pi \leq 0.05$$

$$H_1: \pi > 0.05$$

$$z \text{ critica } (\alpha = 0.05) = +1.645$$

$$\sigma_{\bar{p}} = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} = \sqrt{\frac{(0.05)(0.95)}{100}} = \sqrt{\frac{0.0475}{100}} = \sqrt{0.000475} = 0.022$$

$$z = \frac{\hat{p} - \pi_0}{\sigma_{\bar{p}}} = \frac{0.10 - 0.05}{0.022} = \frac{0.05}{0.022} = +2.27$$

El valor calculado de z de + 2.27 es mayor que el valor crítico de + 1.645 para esta prueba del extremo superior. Por lo tanto, como se encuentran 10 refacciones defectuosas en el lote de 100, se rechaza la hipótesis de que la proporción de artículos defectuosos en la población es de 0.05 o menor, utilizando el nivel de significancia el 5% en la prueba.



El administrador estipula que la probabilidad de tener el proceso para ajustarlo, cuando de hecho no es necesario, debe ser a un nivel de solo el 1%, mientras la probabilidad de no detener el proceso cuando la proporción verdadera de defectuosos es de $\pi = 0.10$ puede fijarse en el 5%. ¿Qué tamaño de muestra debe obtenerse, como mínimo para satisfacer esos objetivos de prueba?

$$\begin{aligned}
 n &= \left[\frac{z_0 \sqrt{\pi_0(1-\pi_0)} - z_1 \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2 \\
 &= \left[\frac{2.33 \sqrt{(0.05)(0.95)} - (-1.645) \sqrt{(0.10)(0.90)}}{0.10 - 0.05} \right]^2 \\
 &= \left[\frac{2.33(0.218) + 1.645(0.300)}{0.05} \right]^2 = \left(\frac{1.0014}{0.05} \right)^2 = (20.03)^2 = 401.2
 \end{aligned}$$

= 402 refacciones

Se trata de una muestra un tanto grande para efectos de muestreo industrial, por lo que el administrador podrá reconsiderar los objetivos de la prueba con respecto a la P (error típico 1) de 0.01 y la P (error tipo 2) de 0.05



5.7 DOS MUESTRAS: PRUEBA SOBRE DOS PROPORCIONES

Prueba para la diferencia entre dos proporciones poblacionales



Ejemplo: Un fabricante está evaluando dos tipos de equipo para fabricar un artículo. Se obtiene una muestra aleatoria de $n_1 = 50$ para la primera marca de equipo y se encuentra que 5 de ellos tiene defectos. Se obtiene una muestra aleatoria de $n_2 = 80$ para la segunda marca y se encuentra que 6 de ellos tienen defectos. La tasa de fabricación es la misma para las dos marcas. Sin embargo, como la primera cuesta bastante menos, el fabricante le otorga a esa marca el beneficio de la duda y plantea la hipótesis $H_0: \pi_1 \leq \pi_2$. Pruebe la hipótesis en el nivel de significancia del 5%.

Datos

$$n_1 = 50 \quad n_2 = 80$$

$$\bar{p}_1 = .10 \quad \bar{p}_2 = .075$$

$$H_0: (\pi_1 - \pi_2) \leq 0$$

$$H_1: (\pi_1 - \pi_2) > 0$$

$$z_{critica} = (\alpha = 0.05) = 1.645$$

Operaciones

$$\hat{\pi} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{50(0.10) + 80(0.075)}{50 + 80} = \frac{5 + 6}{130} = 0.085$$

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{\pi}(1 + \hat{\pi})}{n_1} + \frac{\hat{\pi}(1 + \hat{\pi})}{n_2}} = \sqrt{\frac{(0.085)(0.915)}{50} + \frac{(0.085)(0.915)}{80}}$$

$$= \sqrt{\frac{0.0778}{50} + \frac{0.0778}{80}} = \sqrt{0.0016 + 0.0010} = 0.051$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}} = \frac{0.10 - 0.075}{0.051} = \frac{0.025}{0.051} = 0.49$$

El valor calculado de z de 0.49 no es mayor que 1.645 para esta prueba del extremo superior. Por ello, no puede rechazarse la hipótesis nula en el nivel de significancia del 5%.



Ejemplo 2: Se desea saber si existe una diferencia de proporciones entre los alumnos que reprobaron la materia de física de las escuelas Ignacio Ramírez Y Venustiano Carranza la encuesta se realiza a 70 alumnos de la primera escuela de los cuales el 58% dijo haber reprobado y a 60 alumnos de la segunda escuela y de estos el 70% reprobó.

- Establecer la hipótesis nula y alternativa.
- Establecer se rechaza o se acepta la hipótesis con un nivel de significancia del 5%.

Datos

$$n_1 = 70 \quad n_2 = 60$$

$$\bar{p}_1 = .58 \quad \bar{p}_2 = .70$$

$$H_0: (\pi_1 - \pi_2) = 0$$

$$H_1: (\pi_1 - \pi_2) \neq 0$$

$$z_{critica} = (\alpha = 5\%) = 1 - .95 = .4750 = 1.96$$

Operaciones

$$\hat{\pi} = \frac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2} = \frac{70(0.58) + 60(0.70)}{70 + 60} = \frac{82.6}{130} = 0.63$$

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{\pi}(1 + \hat{\pi})}{n_1} + \frac{\hat{\pi}(1 + \hat{\pi})}{n_2}} = \sqrt{\frac{(.63)(.37)}{70} + \frac{(.63)(.37)}{60}}$$

$$= \sqrt{\frac{0.2331}{70} + \frac{0.2331}{60}} = \sqrt{0.0033 + 0.0038} = 0.084$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}} = \frac{0.58 - 0.70}{0.084} = \frac{-0.12}{0.084} = -1.42$$

Se acepta la hipótesis nula de que no hay diferencia en el nivel de reprobados de las dos escuelas.



5.8. DOS MUESTRAS: PRUEBAS PAREADAS.



En muchas situaciones las muestras se recolectan como pares de valores, como cuando se determina el nivel de productividad de cada trabajador después de un curso de capacitación. Estos valores se llaman observaciones apareadas o pares asociados mismos y a diferencia de las muestras independientes, dos muestras que contienen observaciones apareadas se llaman **muestras dependientes**



En el caso de observaciones apareadas, el método apropiado para probar la diferencia entre las medias de dos muestra consiste en determinar primero la diferencia **d** entre cada par de valores, para después probar la hipótesis nula de que la **diferencia** poblacional media es **cero**. Así, desde el punto de vista de los cálculos de la prueba se aplica a **una** muestra de valores **d**, con $H_0: \mu_d = 0$

La media y desviación estándar es la muestra de valores d se obtiene por medio de la aplicación de las fórmulas básicas, excepto que d es sustituida por X . La diferencia media de un conjunto de diferencias entre observaciones apareadas es:

$$\bar{d} = \frac{\sum d}{n}$$



La fórmula de desviaciones y la fórmula de cálculo para la desviación estándar de las diferencias entre observaciones apareadas son, respectivamente:

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

$$s_d = \sqrt{\frac{\sum d^2 - n\bar{d}^2}{n - 1}}$$

El error estándar de la diferencia media entre observaciones apareadas se obtiene por medio de la formula. Para el error estándar de la media, excepto que d es sustituida de nueva cuenta por X :

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

Dado que el error estándar de la diferencia media calcula con base en la desviación estándar de la muestra de diferencias (**el valor σ_d es desconocido**) y por lo general puede suponerse que los valores de d siguen una distribución normal.



La estadística de prueba empleada para probar la hipótesis de que no existe diferencia entre las medias de un conjunto de las medias de un conjunto de observaciones apareadas es:

$$t = \frac{\bar{d}}{s_d}$$



Ejemplo: un fabricante de automóviles recolecta datos sobre millaje de $n = 10$ autos de diversas categorías de peso usando gasolina de calidad estándar con y sin cierto aditivo. Por supuesto, los motores

fueron ajustados a las mismas especificaciones antes de cada corrida, y los mismos conductores sirvieron para los dos casos de gasolina (aunque no se les hizo saber que gasolina se usaba en una corrida en particular). Dados los datos de millaje en la tabla, probamos la hipótesis de que no existe diferencia entre el millaje medio obtenido con y sin el aditivo, empleando el nivel de significancia del 5% y se resuelve de la siguiente manera:

$$\text{promedio con aditivo} = \frac{276.8}{10} = 27.68 \text{ mpg}$$

$$\text{peomedio sin aditivo} = \frac{275.1}{10} = 27.51 \text{ mpg}$$

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

$$t \text{ critica } (gl = 9, \alpha = 0.05) = \pm 2.262$$

$$\bar{d} = \frac{\sum d}{n} = \frac{1.7}{10} = 0.17$$

$$s_d = \sqrt{\frac{\sum d^2 - n\bar{d}^2}{n-1}} = \sqrt{\frac{1.31 - 10(0.17)^2}{10-1}} = \sqrt{\frac{1.31 - 10(0.0289)}{9}}$$

$$= \sqrt{0.1134} = 0.337$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{0.337}{\sqrt{10}} = \frac{0.337}{3.16} = 0.107$$

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{0.17}{0.107} = +1.59$$

Automóvil	Millaje con aditivo	Millaje sin aditivo	d	d^2
1	36.7	36.2	0.5	0.25
2	35.8	35.7	0.1	0.01
3	31.9	32.3	-0.4	0.16
4	29.3	29.6	-0.3	0.09
5	28.4	28.1	0.3	0.09
6	25.7	25.8	-0.1	0.01
7	24.2	23.9	0.3	0.09
8	22.6	22.0	0.6	0.36
9	21.9	21.5	0.4	0.16
10	20.3	20.0	0.3	0.09
<i>total</i>	276.8	275.1	+1.7	1.31



Ejercicio. El director de la capacitación de una compañía desea comparar un nuevo método de capacitación técnica, que supone la combinación de diskettes instructivos de cómputo y resolución de problemas en el laboratorio con el método tradicional de impartición de clases. Se asocian así doce pares de aprendices de acuerdo con sus antecedentes y desempeño académico, en tanto que uno de los miembros de cada par asignado al curso tradicional y el otro al nuevo método. Al final del curso se determina el nivel de aprendizaje por medio de un examen sobre información básica y la capacidad de aplicarla. Dado que el director de capacitación desea conceder el beneficio de la duda al sistema de instrucción establecido, se formula la hipótesis nula de que el desempeño medio del sistema establecido es igual o mayor que el nivel medio de desempeño del nuevo sistema. Pruebe esta hipótesis al nivel de significancia de 5%. Los datos muestrales de desempeño se presentan en las tres primeras columnas de la siguiente tabla:

Par de aprendices	Método tradicional	Nuevo método	d ($x_1 - x_2$)	d^2
1	89	94		
2	87	91		
3	70	68		
4	83	88		
5	67	75		
6	71	66		
7	92	94		
8	81	88		
9	97	96		
10	78	88		
11	94	95		
12	79	87		
total	988	1030		

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d < 0$$

REFERENCIAS:

- Borrego, Silvia (2008). "Estadística descriptiva e inferencial". Revista digital innovación y experiencias educativas 13. Recuperado el 10 de marzo de 2010 desde: <http://www.csi-csif>.
- Castillo Manrique, Isabel (2006). *Estadística descriptiva y cálculo de probabilidades*. México: Pearson Educación.
- Galbiati Riesco, Jorge M. Conceptos Básicos de Estadística (Versión electrónica). Pontificia Universidad Católica de Valparaíso, Instituto de Estadística. Recuperado el 1 de marzo de 2010 desde: http://www.jorgegalbiati.cl/ejercicios_4/ConceptosBasicos.pdf

- Jordi Casal, Enric Mateu. (2003). Tipos de muestreo (versión electrónica). Rev. Epidem.Med.Prev. (2003), 1: 3-7. Recuperado el 1 de marzo de 2010 en <http://minnie.uab.es/~veteri/21216/TiposMuestreo1.pdf>
- Larios Osorio, Víctor (1999). "Unidad 5. Teoría de muestreo". Recuperado el 12 de marzo de 2010 desde: <http://www.uaq.mx/matematicas/estadisticas/xu5.html>
- Lind, Douglas, William Marchal y Samuel Wathen (2008). *Estadística aplicada a los negocios y la economía* decimotercera edición. México: McGraw-Hill.
- Montgomery, Douglas C. y George C. Runger (1996). *Probabilidad y Estadística aplicadas a la ingeniería*. Cuarta edición. McGraw-Hill, México.
- Ritchey, Ferris (2008). Estadística para las ciencias sociales. Segunda edición. México: McGraw-Hill.
- Ruiz Muñoz, David (2004). Manual de estadística (versión electrónica). Recuperado el 9 de marzo de 2010 desde: <http://www.eumed.net/cursecon/libreria/drm/ped-drm-est.htm>
- Wackerly, Dennis D., William Mendenhall III y Richard L. Scheaffer (2010). *Estadística Matemática con Aplicaciones*. Séptima edición. México: Cengage Learning.
- Walpole Ronald E., Raymond H. Myers et al. (2007). *Probabilidad y Estadística para Ingeniería y ciencias*. Octava Edición. México: Pearson Educación.

Bibliografía complementaria:

- Wackerly Dennis D., Mendenhall William III, Scheaffer, Richard L. Estadística Matemática con Aplicaciones. Séptima Edición, Cengage Learning, México, 2010.
- Ferris Ritchey. Estadística aplicada a las ciencias sociales. Segunda Edición. Mc Graw Hill, 2008.
- Douglas L., William M., Samuel W. Decimotercera Edición, Estadística aplicada a los negocios y la economía, Mc Graw Hill, 2008.
- Isabel Castillo Manrique, Estadística descriptiva y cálculo de probabilidades, Primera Edición, Pearson México, 2006.