1.- Dada la siguiente distribución de frecuencias. Calcular:

a) Mediana *b) Moda* *c) Media* *d) Varianza y desviación típica o estándar*

VARIABLE	FRECUENCIA
37	2
38	1
39	3
40	4
41	6
42	2
43	3

2.- Con los datos que se muestran en la siguiente tabla, construir una ojiva y contestar las siguientes preguntas:

a) ¿cuál es el valor de la mediana?

b) ¿hasta que valor abarca el tercer cuartil?

clase	frecuencia
2 – 5	12
6 – 10	25
11 – 15	10
16 – 20	15
21 – 25	8

EJEMPLO 3.- ¿Qué puede decir y cuál sería su conclusión al comparar estadísticamente los siguientes conjuntos de datos?:

a) 1, 2 y 12

b) 4, 5 y 6

II.- *LA TEORÍA DE LA PROBABILIDAD*

II.1.- Conceptos básicos de probabilidad

A continuación se presentan algunos conceptos que son básicos para el estudio de la probabilidad.

Probabilidad: es la posibilidad de que un hecho o fenómeno ocurra.

Experimento: es el proceso mediante el cual se obtiene una observación o una medición de un fenómeno o hecho.

Evento: es el resultado de la realización de un experimento.

El valor de una probabilidad se puede mostrar a través de fracciones, cuyos valores se encuentran de o a 1, esto significa que los valores pueden ser expresados de la siguiente manera: 0.3, 0.5, 0.23, 0.034, 0.456, etc., o bien, como: ½, 2/3, 1/4, 3/5, 1/10, etc.

Si la probabilidad es cero significa que el fenómeno nunca aparecerá, en cambio, una probabilidad de uno, significa que el fenómeno siempre ocurrirá.

La probabilidad desempeña un papel muy importante en la estadística inferencial. Para demostrar esto último, podemos utilizar el siguiente ejemplo: Supongamos que la empresa "TIRSA" fabrica cable de acero de una pulgada de diámetro y de 10 metros de largo, la empresa realiza esfuerzos para que la resistencia al rompimiento de este producto se mantenga en 5850 kilogramos por centímetros cuadrado en promedio. para mantener un control apropiado se estableció un esquema de muestreo, que consiste en seleccionar 20 piezas al azar en cada turno de fabricación y realizar las pruebas de resistencia pertinentes y si el promedio resultante se encuentra entre 5830 y 5870 kg/cm, inclusive, el proceso de producción se mantiene "bajo control", de otra manera, se considera "fuera de control". Se sabe que esta empresa fabrica 500 piezas diariamente de este producto, en sus tres turnos de labores. Como podemos observar, las 500 piezas representan la población y la muestra está representada por las 60 piezas examinadas al día. El resultado del análisis de la muestra lo podemos utilizar para obtener conclusiones de la población.

II.2.- Tipos de probabilidades

Iniciaremos el estudio de la teoría de la probabilidad, con el señalamiento de la probabilidad “clásica”, recibe este nombre debido a que se basa en la repetición de experimentos realizados bajo las mismas condiciones. Por ejemplo, el lanzamiento de una moneda varias veces.

La fórmula representativa de este tipo de probabilidad:

$$P = h / n$$

Ejemplo 1: Se lanza una moneda al aire una sola vez, calcular la probabilidad de que caiga “águila”.

La suma de la probabilidad de cada uno de los eventos del espacio muestral es igual a la unidad. O sea, p(obtener “águila”) + p(obtener “sol”) = 1. Esto también se puede interpretar como: p(obtener “águila”) + p(no obtener “águila”) = 1.

Ejemplo 2: Se lanza un dado al aire una sola vez, calcular la probabilidad de obtener un punto.

Ejemplo 3: El señor Juan Pérez, recibe un boleto para participar en la rifa de tres premios, si el número de boletos que se reparten para dicho sorteo son 200, ¿cuál es la probabilidad de que el señor Juan Pérez obtenga uno de los tres premios?.

Ejemplo 4: En un partido de futbol entre los equipos "A" y "B", ¿cuál es la probabilidad de que gane el equipo "A"?.

Ejemplo 5: A continuación se transcribe una distribución de frecuencias de las comisiones por ventas anuales, resultado de una encuesta a 124 vendedores de medios publicitarios. basándonos en esta información, ¿cuál es la probabilidad de que un vendedor seleccionado al azar logre una comisión: a) entre $5,000 y $10,000. b) menos de $15,000. c) más de 20,000.

COMISIÓN ANUAL	FRECUENCIA
$ 0 – 4,999	15
5,000 – 9,999	20
10,000 – 14,999	12
15,000 – 19,999	35
20,000 – 24,999	30
25,000 – 29,999	22

Eventos mutuamente excluyentes y no excluyentes.

Eventos mutuamente excluyentes:

Se dice que dos eventos “a“ y “b” son mutuamente excluyentes, cuando ocurre “a”, pero no puede ocurrir “b”. Por lo tanto, la probabilidad de que ocurra “a” o “b” es:

$$p(\,a \text{ o } b\,) = p(a) + p(b)$$

Ejemplo 1: ¿Cuál es la probabilidad de que en un día cualquiera, la paridad del dólar con respecto al peso mexicano, aumente o disminuya?.

- **Ejemplo 2: En una encuesta de mercadeo para un gran almacén, se clasificó a los clientes de la tienda según el sexo y según su residencia, en el centro de la ciudad o en las colonias. La proporción (probabilidad) de los clientes que caen en las cuatro categorías se muestra en la siguiente tabla.**

Residencia	sexo	
	Masculino	femenino
Colonias	0.17	0.67
Centro de la ciudad	0.04	0.12

si se selecciona al azar un cliente de este grupo de consumidores, calcular la probabilidad de que. a) el consumidor resida en las colonias, b) el consumidor sea mujer y viva en el centro de la ciudad, c) el consumidor sea hombre.

Eventos que no son mutuamente excluyentes.

Existen eventos que pueden aparecer al mismo tiempo, éstos reciben el nombre de eventos que no son mutuamente excluyentes. su fórmula es:

p(a o b) = p(a) + p(b) - p(a y b)

Ejemplo 3: según la información mostrada en el ejemplo 2, calcular la probabilidad de que el consumidor seleccionado, sea mujer o viva en la ciudad.

Ejemplo 4: De los 250 empleados de una empresa, 130 fuman cigarros. Hay 150 hombres que trabajan en esta empresa, de los cuales 85 fuman cigarros. ¿Cuál es la probabilidad de que un empleado seleccionado aleatoriamente, a) no fume cigarros?, b) sea mujer y fume cigarros?, c) sea hombre o fume cigarros?.

Ejemplo 5: Se efectuó un estudio de mercado en escala nacional para determinar la preferencia de los hombres de 10 a 59 de edad por los diferentes deportes. Se seleccionó una muestra aleatoria de 1,000 hombres y a cada uno se le pidió indicar su deporte favorito. Los resultados fueron los siguientes.

Si se hace una selección aleatoria de un entrevistado, ¿cuál es la probabilidad de que: a) prefiera el béisbol?, b) tenga entre 20 y 29 años?, c) tenga entre 20 y 29 años y prefiera el baloncesto?, d) tenga cuando menos 50 años o prefiera el béisbol?.

Edad	Deporte			
	Beisbol	Futbol	Baloncesto	Golf
10 – 19	26	47	41	36
20 – 29	38	84	80	48
30 – 39	72	68	38	22
40 – 49	96	48	30	26
50 – 59	134	44	18	4

Eventos independientes, eventos dependientes y probabilidad condicional.

Eventos independientes

Estos eventos ocurren, cuando existen dos o mas experimentos, en Donde la aparición de uno de ellos no afecta la ocurrencia del otro. Los eventos son independientes, principalmente cuando se realizan con reemplazo.

La fórmula que se aplica para calcular la probabilidad de la ocurrencia del evento "a" y del evento "b" en estas condiciones, es:

$$P(a \text{ Y } b) = P(a) * P(b)$$

Ejemplo 1: Se lanza una moneda dos veces en forma sucesiva. calcular la probabilidad de que en el primer lanzamiento se obtenga un "águila" y un "sol" en el segundo lanzamiento.

Ejemplo 2: Se encuentran en una bolsa 3 canicas verde, dos canicas azules 2 y dos canicas blancas. Si se extraen dos canicas al azar y con reemplazo, calcular la probabilidad de

obtener dos canicas blancas.

Eventos dependientes

Los eventos "a" y "b" son dependientes cuando el resultado del primero afecta la aparición del segundo.

La fórmula que se utiliza para analizar este tipo de eventos es:

$$p(a \text{ y } b) = p(a) * p(b/a)$$

Ejemplo 3: Se introducen 5 fichas iguales pero de diferente color en una caja. Dos fichas son rojas, dos son verdes y una es blanca. Si se efectúan dos extracciones al azar y con reemplazo, calcular la probabilidad de obtener una ficha roja y una blanca.

Ejemplo 4: Se lanza un dado dos veces, hallar la probabilidad de los siguientes eventos: a) la suma total de puntos en ambos lanzamientos sea 7, b) la suma total de puntos en ambos lanzamientos sea mayor que 5.

PROBABILIDAD CONDICIONAL

En algunas situaciones debemos dar respuesta a la probabilidad de un evento cuando éste se halla sujeto a ciertas condiciones. El siguiente problema nos muestra este caso.

El centro de salud de San Agustín Etla, lleva a cabo un censo de todas las personas que viven en esa comunidad. Los encuestadores anotan el número de visitas al año que una persona hace al centro de salud y las condiciones sanitarias de la vivienda que habita. Los resultados fueron como se indica en la siguiente tabla.

Con base en esta tabla, si preguntamos, ¿cuál es la probabilidad de que una persona escogida al azar viva en malas condiciones sanitarias? La respuesta es 500/2000 = 0.25.

Número de visitas	Condiciones sanitarias		Total
	Buenas	Malas	
2 o menos	700	100	800
Más de 2	800	400	1200
Total	1500	500	2000

Si la pregunta ahora es, ¿cuál es la probabilidad de que una persona seleccionada al azar viva en malas condiciones sanitarias, *dado* que visita dos veces o menos al año el centro de salud?. En este caso la respuesta hay que buscarla en el primer renglón y sería 100/800 = 0.125. Esta última probabilidad se le conoce con el nombre de "probabilidad condicional", puesto que se da al tener en cuenta la condición de que la persona visitó dos veces o menos el centro de salud.

Cuando se calcula una probabilidad condicional siempre hay dos eventos involucrados. EL EVENTO CONDICIONANTE: la persona visitó dos veces o menos en un año el centro de salud y EL EVENTO CONDICIONADO: la persona vive en malas condiciones sanitarias

En general, si denotamos como A el evento condicionado y como B el evento condicionante, la probabilidad condicional se escribe P (A/B), que se lee: la probabilidad de A dado B. En nuestro caso, A es: la persona que vive en malas condiciones sanitarias, B es: la persona visita dos veces o menos el centro de salud, entonces, P(A/B), se traduce como: probabilidad de que la persona viva en malas condiciones sanitarias *dado* que visitó dos veces o menos el centro de salud. Que por lo ya explicado P(A/B) = 0.125.

Sean A y B dos eventos, la probabilidad condicional de A dado B se indica de la siguiente manera:

P(A/B) = P(A y B) / P(B), donde P(B) > 0

Igualmente se tiene P(B/A) = P(A y B) / P(A), donde P(A) > 0 como la probabilidad condicional de B dado A.

Por lo tanto, P(A y B) se traducirá como: la persona vive en malas condiciones sanitarias y visitó dos veces o menos el centro de salud. En total hay 100 personas y así P(A y B) = 100/2000 = 1/20. Por otra parte, P(B) = 800/2000 = 8/20. Como P(A y B) / P(B) = 1/8. Que coincide con P(A/B). Esto no es una coincidencia, sino todo lo contrario, esto se basa en la definición de probabilidad condicional.

Ejemplo 1: El 50% de los estudiantes de una Universidad tiene clases por la mañana, el 30% tiene clases por la tarde y el 20% tiene clases por la mañana y por la tarde. Se escoge un estudiante al azar de esta Universidad. Calcule la probabilidad de que a) tenga clases por la tarde, dado que tiene clases por la mañana, b) tenga clases por la mañana, dado que tiene clases por la tarde.

Ejemplo 2: En cierto grupo de estudiantes de secundaria formado por 60 mujeres y 40 hombres, se observa que 24 de éstos usan lentes, lo mismo que 16 mujeres. Se escoge un estudiante al azar; halle la probabilidad de que a) sea mujer, dado que usa lentes. b) use lentes, dado que sea hombre.

Ejemplo 3: En un estudio clínico se ha llegado a la conclusión de que la probabilidad de que una persona sufra problemas coronarios es de 0.1. Además, la probabilidad de que una persona sufra problemas de obesidad es el 0.25 y la probabilidad de que una persona sufra a la vez problemas de obesidad y coronarios es del 0.05. Calcular la probabilidad de que una persona seleccionada al azar tenga problemas coronarios si está obesa.

Respuesta: 0.20.

II.3.- ***El teorema de Bayes***

Tema de investigación para el próximo viernes.

II.4.- Distribuciones de probabilidad

La distribución binomial

La distribución de probabilidad para una variable aleatoria discreta es una fórmula, una tabla o una gráfica que proporciona la probabilidad asociada a cada valor de la variable aleatoria. Por ejemplo, considérese los experimentos que consisten en tirar al aire una moneda dos veces en forma consecutiva, y sea x el número de caras observadas en las tiradas, hallar la distribución de probabilidad para x.

x	Primer lanzamiento	Segundo lanzamiento	Probabilidad
0	Cruz	Cruz	¼
1	Cruz	Cara	
			½
1	Cara	Cruz	
2	Cara	Cara	¼

LA REPRESENTACIÓN MAS ÚTIL DE UNA DISTRIBUCIÓN DE ESTE TIPO ES A TRAVÉS DE UNA FÓRMULA.

LA DISTRIBUCIÓN BINOMIAL ES UNA DE LAS DISTRIBUCIONES DISCRETAS DE PROBABILIDAD MÁS ÚTILES. SUS ÁREAS DE APLICACIÓN INCLUYEN INSPECCIÓN DE CALIDAD, VENTAS, MERCADOTECNIA, MEDICINA, INVESTIGACIÓN DE OPINIÓN Y OTRAS.

LAS DOS SUPOSICIONES CLAVES PARA LA DISTRIBUCIÓN BINOMIAL SON:

1.- LA PROBABILIDAD DE ÉXITO PERMANECE CONSTANTE PARA CADA EXPERIMENTO.

2.- LOS n EXPERIMENTOS SON INDEPENDIENTES ENTRE SI.

si x es una variable aleatoria que representa el número de éxitos en n experimentos y "p" la probabilidad de éxito con cualquiera de éstos. se dice entonces que x tiene una distribución binomial con función de probabilidad.

La fórmula general para calcular la probabilidad binomial es:

$$P(X;\ n,\ p) = \Sigma \frac{[\ n!\ p^{x}\ (1-p)^{(n-x)}\ ]}{(n-X)!\ X!}$$

.

EJEMPLO 1: TODOS LOS DIAS SE SELECCIONAN, DE MANERA ALEATORIA, 15 UNIDADES DE UN PROCESO DE PRODUCCIÓN CON EL PROPÓSITO DE VERIFICAR EL PORCENTAJE DE UNIDADES DEFECTUOSAS EN LA PRODUCCIÓN. CON BASE EN INFORMACIÓN PASADA, LA PROBABILIDAD DE TENER UNA UNIDAD DEFECTUOSA ES DE 0.05. LA GERENCIA A DECIDIDO DETENER LA PRODUCCIÓN CADA VEZ QUE UNA MUESTRA DE 15 UNIDADES TENGA DOS O MAS DEFECTUOSAS. ¿CUÁL ES LA PROBABILIDAD DE QUE, EN CUALQUIER DÍA, LA PRODUCCIÓN SE DETENGA?

tablas

$P(\, M \geq X \,) \;=\; 1 - P(\, X - 1;\, n,\, p \,)$

$P(\, M \geq 2 \,) \;=\; 1 - P(\, 1;\, 15,\, 0.05 \,) \;=\; 1 - 0.8290 \;=\; 0.1710$

Comprobar el resultado obtenido aplicando la fórmula general.

La Distribución Poisson

Esta distribución es otro ejemplo de las distribuciones discretas de probabilidad y es el principal modelo de probabilidad empleado para analizar problemas de línea de espera. Además, ofrece una aproximación excelente a la función de probabilidad binomial, cuando p es pequeño y n es grande.

También es muy fácil identificar este tipo de distribución de probabilidad, ya que si se menciona que una cantidad se representa como un promedio, esto significa que es una distribución de poisson

Su fórmula general es:

$$P(X;\ \lambda) = \Sigma \left(\frac{e^{-\lambda}\ \lambda^{x}}{X\,!} \right)$$

Ejemplo 1: si el número promedio de clientes que llega a un banco es de 120 por hora, ¿cuál es la probabilidad de que en un minuto dado lleguen por lo menos 3 clientes?

Este problema representa el cuarto de los casos que se presentan en el formulario respectivo.

$p(m \geq x) = 1 - p(x - 1;\ \lambda)$

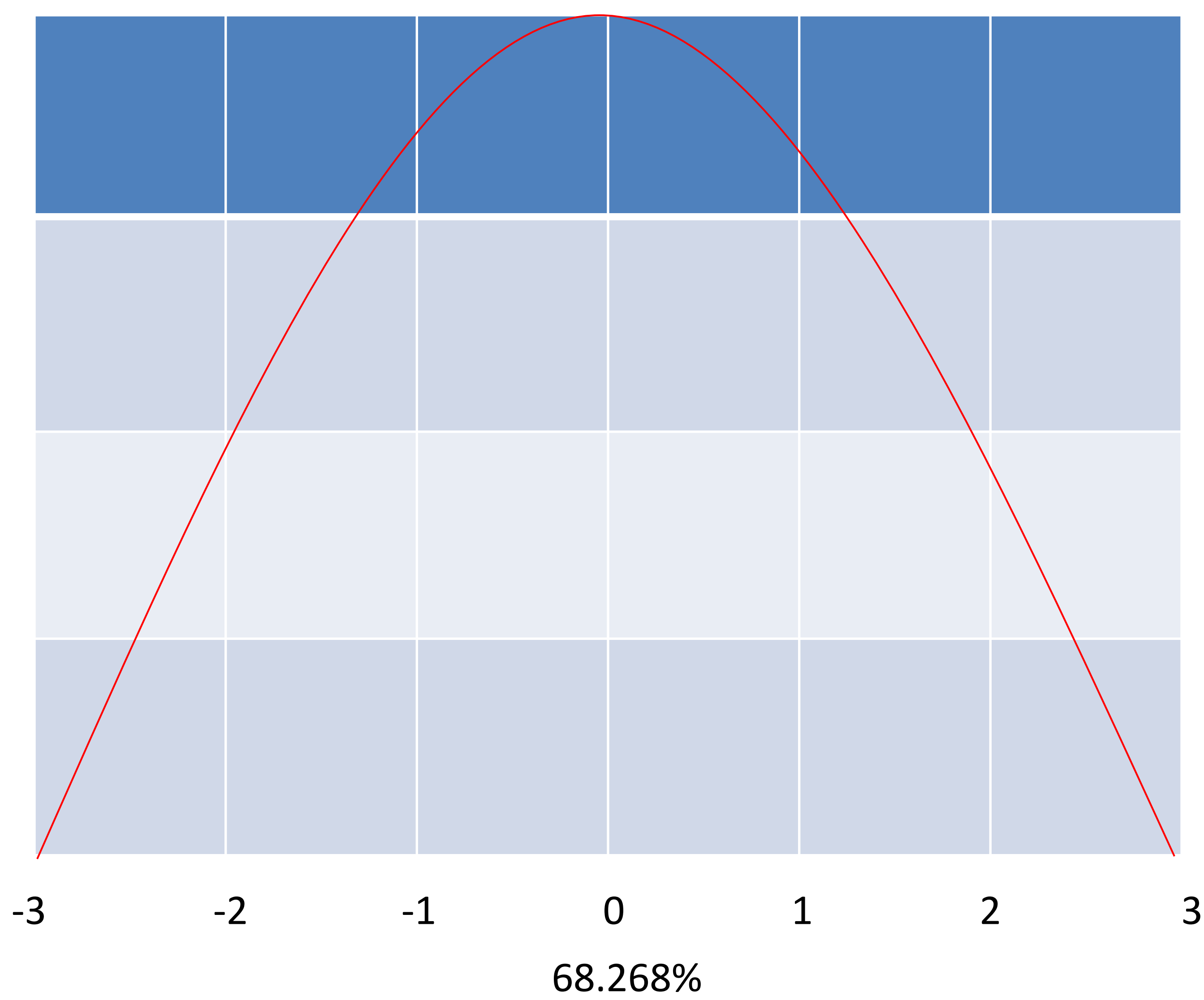
$p(m \geq 3) = 1 - p(2; 2) = 1 - 0.6767 = 0.3233$

La Distribución Normal.

Una de las distribuciones continuas, y tal vez la más importante, es la distribución normal, la cual ocupa un lugar destacado en la estadística inferencial. Su curva que recibe el nombre de curva normal.

Los problemas que se presentan en la administración empresarial, representan una distribución normal o por lo menos en la mayoría de estos problemas.

A continuación se muestra la gráfica de una distribución normal, donde se indican los diferentes porcentajes en que se divide el área bajo la curva. ésta área representa todos los datos bajo análisis.


-3
-2
-1
0
1
2
3
68.268%

La fórmula que se debe aplicar para calcular el área bajo la curva normal es:

$$Z = \Sigma \ \frac{(X - \bar{X})}{S}$$

Ejemplo 1: si la media poblacional es 400 y su desviación estándar es 100, ¿cuál es la probabilidad (área) de valores: a) entre 250 y 500? y b) menores de 250?.

Ejemplo 2: supongamos que el ingreso mensual promedio de los 10,000 trabajadores de la empresa "el sol" es de \$5,000 y la desviación estándar es de \$1,000. Si la distribución es normal, encontrar el número de trabajadores que tienen un ingreso mensual: a) inferior a \$5,500, b) superior a \$5,000 pero inferior a \$6,000, c) superior a \$6,000.

Ejemplo 3: Con referencia a la información del problema anterior, ¿cúal es la probabilidad de ingreso arriba del cual están los salarios del 60% de los trabajadores?.

Ejemplo 4: Se sabe que el tiempo útil de un componente eléctrico tiene una distribución normal con media de 2,000 horas y desviación estándar de 200 horas. ¿Calcular la probabilidad de que un componente eléctrico elegido al azar dure: a) entre 2,000 y 2,400 horas, b) más de 2,200 horas.

Ejemplo 5: Con referencia a los componentes eléctricos que se mencionan en el problema anterior. Determinar su vida útil, de manera que solo el 10% de ellos fallen antes de ese tiempo

LA ESTADÍSTICA INFERENCIAL

III.- LAS DISTRIBUCIONES MUESTRALES

III.1.- El muestreo de una población

Para analizar una población "grande", debemos extraer en forma aleatoria, una muestra o un conjunto de muestras.

Si se extraen todas las muestras posibles de igual tamaño, en una población, la suma de los promedios de cada muestra es igual a la media de la población.

Esto lo podemos demostrar a través de la utilización de la distribución en el muestreo de la media :

La distribución en el muestreo de la media obtenida de todas las muestras posibles del mismo tamaño extraídas de la población, proporciona información valiosa para el análisis estadístico.

Distribución en el muestreo.

Un cierto estadístico puede ser calculado para cada una de las muestras posibles extraídas de la población. Una distribución del estadístico obtenida de las muestras es llamada distribución en el muestreo del estadístico. Por ejemplo, si la muestra es de tamaño 2 y la población de tamaño tres(elementos a, b, c), es posible extraer 3 muestras(ab, bc, ac) de la población. Podemos calcular la media de cada muestra. Por lo tanto, tenemos 3 medias muestrales para las 3 muestras. las 3 medias forman una distribución, que recibe el nombre de distribución de las medias muestrales, o la distribución en el muestreo de la media. De la misma manera, la distribución de las proporciones obtenida de todas las muestras posibles del mismo tamaño, extraídas de una población, es llamada la distribución en el muestreo de la proporción.

El error estándar

La desviación estándar de una distribución, en el muestreo de un estadístico, es frecuentemente llamada el error estándar del estadístico. Por ejemplo, la desviación estándar de las medias de todas las muestras posibles del mismo tamaño, extraídas de una población, es llamada el error estándar de la media. La diferencia que existe entre los términos "desviación estándar" y "error estándar" es que la primera se refiere a los valores originales, mientras que la última está relacionada con valores calculados. Un estadístico es un valor calculado, obtenido con los elementos incluidos en la muestra.

Error muestral o error de muestreo

La diferencia entre el resultado obtenido de una muestra(un estadístico) y el resultado que se debería obtener de una población (parámetro correspondiente), recibe el nombre de error muestral o error de muestreo. Por ejemplo, si el valor de la media poblacional es 6 y el valor obtenido en la muestra es 4, el error muestral es de 2. Esto usualmente ocurre cuando no se lleva a cabo la encuesta completa de la población, sino que se toma una muestra para estimar las características de la población.

Metodos de selección de muestras

Una muestra debe ser representativa si va a ser usada para estimar las características de una población. Los métodos utilizados para seleccionar una muestra representativa son numerosos, dependiendo del tiempo, dinero y habilidad disponibles para tomar una muestra.

Los métodos de selección de muestras se clasifican de acuerdo a:

1.- El número de muestras tomadas de una población dada para el estudio.

2.- La manera usada en seleccionar los elementos incluidos en la muestra.

Tarea: Investigar cuales son estos métodos y explicar la forma de aplicarlos.

para calcular el número de muestras de orden "n" que se pueden formar con los "N" elementos de una población.

$$ {}_{N}C_{n} = \frac{N!}{(N-n)!\ n!} $$

Para calcular el error estándar se utiliza la fórmula:

$$ \sigma = (\sigma / n) \sqrt{\frac{N - n}{N - 1}} $$

Por ejemplo: el salario por hora de los 6 trabajadores de un pequeño taller se indican en la siguiente tabla, encontrar: a) la media de las muestras de tamaño 2(salarios de 2 trabajadores), b) la media de las medias muestrales, c) el error estándar de la media, d) las probabilidades de las medias muestrales.

III.2.- La distribución en el muestreo de la media y de la proporción.

Estimaciones puntuales y por intervalo.

Una estimación es el proceso mediante el cual se utiliza el valor de un estadístico (valor de una característica de una muestra) para calcular el valor de un parámetro (valor de una característica de una población) desconocido.

La estimación de un parámetro puede ser expresada de dos maneras: una estimación de punto y una estimación de intervalo. Una estimación de punto es un número único que es usado para representar la estimación del parámetro. Una estimación de intervalo es un recorrido establecido dentro del cual podemos esperar que esté el parámetro. Por ejemplo, si la estimación del salario promedio por hora en una ciudad es expresada como $4, es una estimación de punto, en cambio si es expresada como entre $3.5 y $4.5, es una estimación de intervalo.

INTERVALOS, LÍMITES Y COEFICIENTES DE CONFIANZA.

Un intervalo es el recorrido que existe entre dos desviaciones estándar, donde se incluye el valor de la media. Su representación simbólica o numérica recibe el nombre de intervalo de confianza.

$$X \pm Z\,(S / \sqrt{n})$$

Si la media poblacional es $6 y su error de la media $0.50, el intervalo de confianza será:

$6 = $0.50 o bien como: $5.50 a $6.50

Los dos valores, $5.50 y $6.50 reciben el nombre de límites de confianza. Y la probabilidad, 68.26%, es llamada el coeficiente de confianza o también nivel de confianza.

Ejemplo: Se selecciona una muestra aleatoria de 500 compradores de un centro comercial de la Ciudad de Oaxaca, para determinar la distancia promedio que recorren los clientes para llegar al centro comercial. Un análisis de los resultados de la muestra revelan que el promedio es de 23.5 km y la desviación estándar es de 10.4 km. ¿cuál es el valor de la estimación puntual de la media poblacional?, ¿cuál es el valor por intervalo de la media poblacional si se desea un nivel de confianza del 95%?

Distribución muestral de proporciones de las muestras.

Como anteriormente se indicó la distribución binomial implica determinar las probabilidades de diferentes números de “éxitos” en un experimento binomial. Con frecuencia es conveniente investigar la proporción de “exitos”más que el número absoluto.

Por ejemplo: Una empresa cuenta con cinco empleados, de los cuales dos son profesionistas. Se pide formar todas las muestras posibles de dos empleados, mostrando la proporción de profesionistas en cada muestra. En la siguiente tabla se indica quienes son los profesionistas.

Empleado	profesionista
A	si
B	no
C	no
D	si
E	no

La proporción de la población debe ser igual a la media de la distribución muestral de las proporciones.

Igualmente, el error estándar de la proporción debe ser igual a la desviación estándar de la distribución muestral de las proporciones.

***EL ERROR ESTANDAR DE LA PROPORCIÓN*.**

El término "error estándar" se usa para identificar la desviación estándar de una distribución muestral. El error estándar de la proporción es la desviación estándar de la distribución muestral de las proporciones de la muestra.

El error estándar de la proporción se calcula utilizando la siguiente fórmula:

$$\sigma_p = \sqrt{p(1-p) / n}$$

Ejercicio para resolver: Una población consiste de 6 unidades producidas el día de hoy. Esta población tiene una unidad defectuosa. Si se toman muestras de 2 elementos cada una de ellas. A) calcule la proporción de la población para las unidades defectuosas. B) Elabore una tabla que contenga todas las muestras posibles de tamaño 2. C) Calcule la proporción de unidades defectuosas para cada muestra con 2 elementos que se obtuvo en el punto B). D) Calcule la media de esta distribución muestral de proporciones. ¿equivale esta proporción media a la proporción de la población?. E) Calcule el error estándar de la proporción para esta distribución muestral. F) Calcule la desviación estándar de la distribución muestral de proporciones. ¿Equivale esta desviación estándar al error estándar de la proporción?.

Ejemplo: La empresa KARLA ha estado recibiendo reclamaciones respecto a un tostador que comercializa en el centro del país. Para resolver este problema, la empresa quiere conocer primero los porcentajes de tostadores vendidos que tienen problema. Durante los últimos 6 meses ha estado vigente un programa de descuentos, de manera que la empresa decide usar los nombres de la lista de descuentos como la población de la que se obtendrá la muestra. Se escoge una muestra aleatoria de 1800 nombres a partir de esta lista, y se les manda por correo un breve cuestionario junto con el vale de descuento por valor de $80. Contestaron un total de 1675 encuestados, lo que representa un buen porcentaje (93%). La empresa piensa que obtuvo una muestra representativa. El 12% de los que contestaron dijo haber experimentado problemas con el tostador. Estimar el porcentaje de compradores que han tenido problemas con el tostador, con un nivel de confianza del 99%.

EJERCICIOS PARA RESOLVER

1.- El 70% de una población de empleados es masculina. En una muestra de 10 empleados, ¿ cuál es la media y la desviación estándar de la distribución muestral para la proporción de hombres seleccionados de dos en dos?.

2.- Una máquina automática que llena latas de sopa tiene una media de llenado de 16 onzas y una desviación estándar de 0.5. a) ¿cuál es la probabilidad de obtener una muestra de 49 latas con una media más grande que 16.1 onzas?. b) Determine la probabilidad de que la media muestral esté a 0.05 onzas o menos de la media poblacional.

3.- Una población se compone de las siguientes unidades producidas el día de hoy por cuatro trabajadores.

Trabajador	unidades producidas
A	5
B	3
C	7
D	8

A) calcule la media y la desviación estándar de la población. B) elabore una tabla que presente todas las muestras posibles de dos elementos. C) calcule la media de cada una de las muestras de tamaño dos. D) calcule la media de esta distribución muestral. ¿Es igual esta media a la media de la población?. E) calcule el error estándar de la media para esta distribución muestral. F) calcule la desviación estándar de la distribución muestral de las medias. ¿es esta desviación estándar igual al error estándar de la media?.

4.- La media de una población de 500 elementos con distribución normal es 175 y la desviación estándar es 19. Se selecciona una muestra de 81 elementos de esta población. Determinar la probabilidad de que:

a) El primer elemento seleccionado al azar de la población sea mayor que 184.

b) La media de la muestra sea mayor que 180.

c) La media de la muestra esté entre 173 y 177

6.- Una compañía hipotecaria sabe que el 8% de los receptores de préstamos de ahorro-vivienda dejan de cumplir los pagos durante los primeros cinco años. ¿Cuál es la probabilidad de que de 350 receptores de préstamos, menos de 25 no cumplan con los pagos durante los próximos cinco años?.

ESTIMACION CON MUESTRAS PEQUEÑAS

Como mencionamos anteriormente, una muestra pequeña es aquella que tiene menos de 30 datos, y para analizarla se utiliza la distribución "t" de student.

La estimación por intervalo con una muestra pequeña es:

$$\bar{X} \pm t\, S/\sqrt{n}$$

El uso de la tabla "t" requiere que se conozca el área en las colas de la curva. Suponga que se desea un nivel de confianza del 95% para la estimación por intervalo. Esto significa que debe existir un 5% en ambas colas de la curva, o sea el 2.5% en cada cola. Se debe tomar el 5% para buscar el valor en la tabla.

El segundo valor requerido para usar la tabla es el número de grados de libertad(gl). Esto es: gl = n – 1. Este valor solo se presenta cuando se usa un estadístico para estimar un parámetro. En este caso, el estadístico que se usa es la desviación estándar.

Cuando se usan dos estadísticos se emplea n – 2 y así sucesivamente.

Ejemplo 1.- debe formarse una estimación por intervalo alrededor de una media muestral. Se ha seleccionado una muestra de 8 elementos y se elige un nivel de confianza del 98%. ¿cuál es el valor correcto de "t"?

RESPUESTA: un nivel de confianza del 98% implica un área del 1% en cada cola de la distribución y el número de grados de libertad es 7 (8 – 1). Por lo tanto, la tabla se lee usando la fila 7 y la columna 0.02 para el área de las dos colas. El valor correcto de "t" es 2.998.

Ejemplo 2.- Josefina López, dueña de una tienda de productos naturales en la Ciudad de Oaxaca, acaba de leer un artículo titulado "EL MITO DEL COLESTEROL". La lectura hizo que josefina estimara el promedio del contenido de grasa por libra de carne para hamburguesas que se venden en la ciudad. Para su muestra, compró una libra de carne en cada una de las nueve tiendas elegidas al azar. Cocinó la carne, escurrió la grasa y la pesó.

Los resultados en onzas fueron: 3.3, 4.8, 5.1, 4.5, 4.0, 3.9, 4.7, 5.0 y 3.6. Josefina quiere ahora una estimación por intervalo con el 90% de confianza para el contenido promedio de onzas de grasa por libra de carne para hamburguesas.

Ejemplo 3.- En cada uno de los siguientes casos, ¿debe usarse la distribución normal o la distribución "t" para formar la estimación por intervalos?. Determine esas estimaciones.

	Media muestral	s	tamaño de la muestra	nivel de confianza
a)	10	2	25	95%
b)	30	5.5	55	99%
c)	50	9.4	10	90%
d)	1	0.2	100	98%

TAMAÑO DE LA MUESTRA

Tres factores afectan a la determinación del tamaño de la muestra para estimar la media de la población. 1) el nivel de confianza, 2) el error tolerable máximo y 3) la variación de la población.

ERROR TOLERABLE MÁXIMO

Es la cantidad máxima en la que el estadístico de la muestra difiere del parámetro de la población, para un nivel de confianza específico.

TAMAÑO DE LA MUESTRA PARA MEDIAS.

Para calcular dicho tamaño de muestra se utilizan los 4 pasos siguientes:

1.- determinar el nivel de confianza, para calcular z.

2.- determinar cuál es el error máximo (E), posible que se admite.

3.- determinar la desviación estándar de la población, si no se conoce se debe estimar. Para determinar el valor de la desviación estándar, se debe llevar a cabo un estudio "piloto" con una pequeña muestra y usar la desviación estándar calculada.

4.- usar la siguiente fórmula:

$$N = \frac{Z^2 \sigma^2}{E^2}$$

Ejemplo: El Sr. Jiménez desea estimar el peso medio de las cajas de cartón que le envían de una empresa. La media muestral no debe diferir en más de 2 libras de la media poblacional con un 95% de confianza. Se estima que el valor de la desviación estándar de la población es 10 libras. ¿qué tamaño deberá tener la muestra seleccionada?

$n = 1.96^2 \ (10)^2 / 2^2 = 97$ cajas aprox.

Para determinar el tamaño de la muestra para estimar una proporción se siguen los mismos pasos indicados anteriormente. Su fórmula es

$$n = Z^2 P(1 - P) / E^2$$

Cuando no se conoce la proporción de la población, ésta debe tomarse como 0.5, ya que cuando se toma esta cantidad la amplitud es la más grande.

Ejemplo: La agencia de publicidad CAMERON quiere medir la proporción de la población que responde de manera favorable a un nuevo comercial de chocolates. La agencia desea estimar la proporción de respuestas favorables con una diferencia máxima de 0.04 a un nivel de confianza del 90%. ¿ qué tamaño de muestra deberá tomar si no tiene idea de la proporción de la población?.

Ejemplo: El gerente de una gran empresa quiere comprobar los registros de inventario contra los inventarios físicos mediante un estudio muestral. El gerente indica que el máximo error muestral no deberá ser más del 5% hacia arriba o hacia debajo de la verdadera proporción de los registros inexactos. El nivel de confianza deberá ser del 99.73% y la proporción de los registros inexactos se estima en el 35% de acuerdo con la experiencia pasada. Encontrar el tamaño de la muestra a analizar.

Si se conoce el número de elementos de la población se utiliza la fórmula:

$$N = \frac{N\ Z^2 p(1 - P)}{(n - 1)E^2 + Z^2 P(1 - P)}$$

$$N = \frac{n\ z^2\ \sigma^2}{E^2(N - 1) + z^2\sigma^2}$$

PRUEBAS DE HIPOTESIS

Una hipótesis estadística o simplemente hipótesis, es una suposición con referencia a una población.

Antes de aceptar o rechazar una hipótesis, se deberá probar la validez de la misma, puesto que pueden ser verdaderas o falsas. Desde luego que un método seguro de probar la hipótesis sería un examen de toda la población. Sin embargo, el examen puede ser impráctico o imposible. Una forma práctica es usar una muestra de acuerdo con la teoría de la probabilidad y según el resultado obtenido se toma la decisión de aceptar o rechazar la hipótesis. Los métodos estadísticos, los cuales son usados para decidir si se acepta o se rechaza una hipótesis, son llamados; Pruebas de hipótesis, Pruebas de significación o métodos para formular reglas de decisión.

En las pruebas de hipótesis siempre es posible cometer uno de los siguientes tipos de errores:

ERROR TIPO I: RECHAZAR UNA HIPOTESIS NULA VERDADERA (α)

ERROR TIPO II: ACEPTAR UNA HIPOTESIS NULA FALSA (β)

La probabilidad de rechazar una hipótesis nula que es cierta se llama NIVEL DE SIGNIFICANCIA (α) de una prueba de hipótesis. Es común designar un nivel de significancia de : 1%, 2%, 5% o 10%. Debe entenderse que una probabilidad pequeña de cometer un error tipo I genera una mayor probabilidad de cometer un error tipo II. O viceversa.

En la prueba de hipótesis, hemos de formular el supuesto valor del parámetro de la población antes de iniciar el muestreo. La suposición que deseamos probar recibe el nombre de HIPOTESIS NULA O HIPOTESIS INICIAL. Y se representa por el símbolo: Ho.

Cada vez que rechazamos la hipótesis nula, la conclusión que aceptamos se llama HIPOTESIS ALTERNATIVA y se representa por el símbolo: H1.

Cuando la hipótesis nula nos representa una igualdad, esto significa que es una prueba de 2 extremos. Siempre y cuando no se indique otra cosa en el problema.

Cuando en la hipótesis nula se indique "mayor o igual que" o "menor o igual que", significa que es una prueba de un extremo, izquierdo o derecho respectivamente.

PASOS PARA PRUEBAS DE HIPOTESIS.

1.- Establecer las hipótesis nula y alternativa.

2.- Suponer que la hipótesis nula es verdadera y determinar la distribución muestral apropiada para esta suposición.

3.- Obtener una muestra aleatoria a partir de la población bajo estudio y con los datos seleccionados, calcular el estadístico apropiado para la prueba.

4.- Con el resultado obtenido en el paso anterior, determinar si se acepta o se rechaza la hipótesis.

Ejemplo: Un ingeniero de seguridad en carreteras decide probar la capacidad de carga de un puente que tiene 20 años de construido. Dispone de abundantes datos recabados de pruebas semejantes efectuadas en la misma clase de puentes. Considerando que la capacidad mínima de carga del puente ha de ser de 10 toneladas, ¿conviene aplicar una prueba de hipótesis de uno o de dos extremos?. ¿cuál es la hipótesis nula y cuál es la alternativa?

RESPUESTA: $H_o: \mu \geq 10$

$H_1: \mu < 10$

Por lo anterior, la prueba de hipótesis que se debe aplicar es de un extremo.

Ejemplo: Un fabricante vende ejes traseros de camiones. Los ejes han de soportar 80,000 kg. por centímetro cuadrado en las pruebas de esfuerzo, pero los ejes demasiado fuertes elevan de manera considerable los costos de fabricación. La experiencia indica que la desviación estándar de la resistencia de los ejes es de 4000 kg. por centímetro cuadrado. El fabricante selecciona una muestra al azar de 100 ejes de la última serie de producción, los somete a pruebas y averigua que la capacidad media de resistencia de la muestra es de 79,600 kg./ cm cuadrado. Si el fabricante usa un nivel de significancia del 5% en las pruebas. ¿cumplirán los ejes con los requisitos de esfuerzo?

$\sigma = 4000, \quad \bar{X} = 79{,}600, \quad n = 100, \quad \alpha = 5\%.$

$H_o: \mu = 80{,}000$

$H_1: \mu \neq 80{,}000$

LA PRUEBA "t" STUDENT

Cuando no se conoce la desviación estándar de la población y el error estándar es estimado a partir de una muestra, usaremos la distribución "t". Esta distribución tiene n – 1 grados de libertad. Solo se utiliza cuando tenemos una muestra pequeña(menos de 25 datos). Se calcula el estadístico de prueba mediante la siguiente fórmula.

$$t = (X - \bar{X}) / (S/\sqrt{n})$$

Ejemplo: La empresa TAPICO desea saber cuál es la máxima tensión de ruptura que soportan los cables de acero que fabrica. Un cliente importante está interesado en la compra de un gran número de cables y ha establecido que el punto promedio de ruptura no debe ser menor que una tonelada. Tapico piensa que una tonelada es aproximadamente el punto de ruptura de los cables, pero decide probar la hipótesis de que la tensión de ruptura promedio es una tonelada, frente a la alternativa de que es menor. La empresa Tapico a establecido que una muestra pequeña es apropiada para realizar su prueba, ya que cada cable seleccionado se destruye en el proceso y una muestra grande sería demasiado costosa. Por lo anterior la muestra es de 10 cables y se decide un nivel de significancia del 5% para la prueba. ¿El cliente de la empresa se encontrará satisfecho con el producto que va a comprar?. Los resultados de la prueba fueron:

$$\bar{X} = 0.96, \quad S = 0.15, \quad n = 10, \quad gl. = 10 - 1 = 9, \quad \alpha = 5\%$$

Ejemplo: Un especialista en personal que labora en una gran empresa, está reclutando un gran número de empleados para un trabajo en el extranjero. Durante la realización de las pruebas, la gerencia pregunta cómo marchan las cosas y el especialista contesta: “Bien, creo que la puntuación promedio en el test de aptitudes será 90 “. Cuando la gerencia revisa 20 de los resultados de la prueba, averigua que la puntuación media es 84 y la desviación estándar de esta puntuación es 11. La gerencia quiere probar la hipótesis del especialista en personal con un nivel de significancia del 10?

Ejemplo: La dueña de una tienda de productos naturales en la Ciudad de Oaxaca, decide verificar el contenido promedio de grasa por libra de carne para hamburguesas que se vende en las carnicerías de esta ciudad. Para hacerlo compró una libra de carne en cada una de las 9 carnicerías elegidas al azar. Cocinó la carne, escurrió la grasa y la pesó. Los resultados para cada carnicería, en onzas de grasa, fueron: 3.3, 4.8, 5.1, 4.5, 4.0, 3.9, 4.7, 5.0, 3.6. La dueña de la tienda de productos naturales desea probar la hipótesis de que el nivel medio de grasa es 4 onzas usando un nivel de significancia del 5%.

PRUEBAS DE HIPOTESIS SOBRE LA PROPORCION DE UNA POBLACIÓN

Los analistas también están interesados en el porcentaje de elementos de una población que cumple con ciertas características. Los políticos están preocupados por el porcentaje de votantes que favorecerán un asunto específico, los directores de empresa se interesan en el porcentaje de artículos producidos o comprados que salen defectuosos, etc.

La metodología de prueba depende de si se usa una muestra pequeña o grande. Para muestras pequeñas(20 o menos datos), se usa una distribución binomial. Para muestras grandes(más de 20 datos) es aceptable la distribución normal si $np > 5$ y $n(1 - p) > 5$. Como la mayor parte del análisis de proporciones maneja muestras grandes, sólo se estudiarán las pruebas de hipótesis para este tipo de muestras.

La fórmula que se utiliza para esta clase de pruebas es la siguiente:

Ejemplo: La empresa LA ROCA dirige su atención a la calidad de uno de sus productos que fabrica, un compuesto ácido usado para hacer un analgésico que se vende sin receta. Según sus propios estándares de control de calidad, no más del 3% de sus lotes de 200 libras producidos en la fábrica pueden fallar una serie de pruebas exhaustivas y meticulosas de análisis químicos. Un fallo significa que el producto no es aceptable para los clientes de la empresa. Construya la prueba de hipótesis.

Ejemplo: Productos Arquitectónicos ROSAS está verificando la producción de cables conductores de acero en su planta principal. Como estos componentes por lo general se encuentran encerradas en paredes o estructuras, el nivel de calidad debe ser bastante alto. Se ha puesto en práctica un programa de mejora de la calidad en la línea de conductores durante los últimos 6 meses, y el supervisor asegura que se ha reducido la tasa anterior de defectos del 2%. La dirección de la empresa duda de esta aseveración, por lo tanto, decide probar lo que piensa contra la aseveración del supervisor. Para esto selecciona una muestra aleatoria de 358 cables y de estos, se encontraron 4 unidades defectuosas. Se decide emplear un nivel de significancia del 10%. ¿qué se puede decir de la aseveración del supervisor?.

Ejemplo: Un artículo en el periódico "EL GRITÓN" habla sobre la severa competencia que presentan los nuevos bancos, forzando a las instituciones bancarias existentes a mejorar sus productos y servicios. Uno de estos bancos está intentando mejorar el nivel de sus cuentas por cobrar en nuestro país. Durante varios años, el porcentaje de cuentas vencidas cada mes ha sido de alrededor del 6%. El departamento de contabilidad considera que tiene evidencias de que las condiciones económicas han sido la causa del aumento en este porcentaje. Durante el mes anterior, este banco mandó facturar a 478 proveedores. De estas cuentas por cobrar 42 están vencidas. Probar de que el porcentaje de cuentas vencidas por cobrar se ha mantenido igual(alrededor del 6%) contra la aseveración de su empleado del departamento de contabilidad. Con un nivel de significancia del 5%.