Prefacio

En estadística I, se presentó distribución de frecuencias como una forma útil de resumir las variaciones en los datos observados. Se explicó las tablas de distribución de frecuencias haciendo una lista de todos los resultados posibles de un experimento y, después, indicando la frecuencia observada de cada resultado posible. Fue una herramienta de cómo debemos ordenar, clasificar y evaluar los resultados de dichos experimentos, en esta guía se dará a conocer sobre aspectos relacionados con **Estadística II**, basado en las distribuciones de probabilidad quienes están relacionadas con las distribuciones de frecuencias.

De hecho, podemos pensar en la distribución de probabilidad como una distribución de frecuencias teóricas, debido a que estas distribuciones tratan sobre expectativas de que algo suceda, resultan ser modelos útiles para hacer inferencias y tomar decisiones en condiciones de incertidumbre.

La Estadística II, se basa en determinar las diferentes maneras de obtener probabilidades de acuerdo a su población a muestra que se desea estudiar.

Objetivos:

- Dotar al estudiante de los principales conceptos y teoremas de la teoría de la probabilidad que son necesarios para el estudio de la estadística.
- Proporcionar las herramientas necesarias para calcular las probabilidades, a través de variables aleatorias.
- Iniciar al estudiante en el estudio de las distribuciones probabilísticas (tanto discretas como continúas) de mayor aplicación en las distintas áreas de la carrera.

Contenido del programa de estudios:

- Distribuciones de Probabilidades, Muestreo aleatorio, distribución de muestras
- Esperanza matemática, desviación estándar y varianza
- Distribución Bernoulli
- Distribución Binomial
- Distribución Poisson
- Distribución Normal, función y propiedades
- Estimación de parámetros, estimación puntual
- Estimación por intervalos para la media y para la proporción
- Distribución t-student, inferencia respecto a la media, diferencia de dos medias
- Contraste de Hipótesis estadísticas, hipótesis nula, alterna, regla de decisión.

Distribuciones de Probabilidades, Muestreo Aleatorio

Para comenzar nuestro estudio sobre la distribución de probabilidades, primero veamos algunos conceptos importantes:

Experimento:

En la teoría de probabilidad, la actividad que origine uno de dichos eventos se conoce como experimento. Y se denota con la letra S.

Evento:

Es uno o más de los posibles resultados de hacer algo. Al lanzar una moneda al aire, si cae sello es un evento, y si cae cara es otro evento.

Espacio Muestral:

Es el conjunto de todos los resultados posibles de un experimento.

Probabilidad:

La posibilidad de que algo suceda.

Probabilidad Clásica:

Número de resultados favorables a la presentación de un evento dividido entre el número total de resultados posibles.

Probabilidad Condicional:

Probabilidad de que se presente un evento, dado que otro evento ya se ha presentado.

Ejemplo, recordemos la idea del lanzamiento de la moneda no alterada, suponiendo que se lanza dos veces la moneda cuyo resultado puede ser cara(H) o sello(T), supóngase ahora que se quiere determinar la distribución de probabilidad del **número de sellos** que podrían caer cuando lanzamos la moneda dos veces. Hay que determinar lo siguiente:

Experimento: Lanzar una moneda no alterada

Evento: Obtener el número de sellos al lanzar la moneda dos veces

Espacio Muestral:

	1er lanzamiento	2do lanzamiento	
	H	H	No salga sello la x = 0
S =	S	H	Al menos en un lanzamiento salga sello la x = 1
	H	S	
	S	S	Que salga sello en los dos lanzamientos la x =2

A continuación, se presenta la tabla de distribución de probabilidades.

Tabla °1. Distribución de probabilidad al lanzar dos veces una moneda no alterada.

Número de sellos	Casos observados	P_i	P_a	$\%P_i$	$\%P_a$
0	1	0,25	0,25	25	25
1	2	0,5	0,75	50	75
2	1	0,25	1	25	100
Total	4	1		100	

Fuente: García, I(2019)

X: número de sellos

F_i: casos observados

P_i: Probabilidad, se obtiene $\longrightarrow$ P_i= casos favorables/casos posibles

$P_i = 1/4 = 0,25$

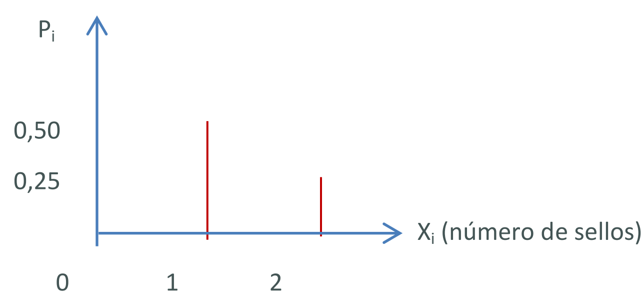
P_a: Probabilidad acumulada, se determina como la frecuencia acumulada vista en Estadística I

$\%P_i$: es el porcentaje de la probabilidad, $\longrightarrow$ $P_i * 100 = 0,25*100 = 25\%$

$\%P_a$: es el porcentaje de la probabilidad acumulada $\longrightarrow$ $P_a= 0,75 * 100 = 75\%$

Podemos representar en forma gráfica la distribución de probabilidad, según la siguiente, en el eje x se representa la variable (x_i), en el eje y la probabilidad.

Fig. °1. Distribución de probabilidad, para el número de sellos al lanzar dos veces una moneda



Esperanza matemática, varianza y desviación estándar

Variable Aleatoria: Una variable es aleatoria si toma diferentes valores como resultado de un experimento aleatorio, puede ser discreta o continua como se explicó en Estadística I, si puede tomar sólo un número limitado de valores, es una variable aleatoria discreta, ahora si toma cualquier valor dentro de un intervalo dado, es una variable aleatoria continua.

Esperanza Matemática o valor esperado: El valor esperado es una idea fundamental en el estudio de las distribuciones de probabilidad. Durante muchos años, el concepto a ha sido puesto en práctica con bastante regularidad por las compañías aseguradoras, también ha sido utilizado ampliamente por muchas de las personas que tienen que tomar decisiones en condiciones de incertidumbre. Para obtener el valor de esperado de una variable aleatoria discreta, multiplicamos cada valor que la variable puede tomar por la probabilidad de presentación de ese valor y luego sumamos los productos.

En la siguiente tabla, se toma como referencia, una clínica de tratamiento de cáncer de pecho, para calcular el promedio de mujeres atendidas diariamente, utilizando registros sobre pacientes anteriores como base para calcular su valor esperado,

Tabla °2. Número de mujeres atendidas diariamente de cáncer de pecho.

Valores posibles de la variable aleatoria(X_i)	Probabilidad de que la variable aleatoria tome estos valores(P_i)	Multiplicando (X_i*P_i)
100	0,01	1,00
101	0,02	2,02
102	0,03	3,06
103	0,05	5,15
104	0,06	6,24
105	0,07	7,35
106	0,09	9,54
107	0,10	10,70
108	0,12	12,96
109	0,11	11,99
110	0,09	9,90
111	0,08	8,88
112	0,06	6,72
113	0,05	5,65
114	0,04	4,56
115	0,02	2,30
TOTAL	**1**	**108,02**

Fuente: García, I.(2019)

Por lo tanto el valor esperado es:

$E(x) = \sum (X_i * P_i) = 108{,}02$

Varianza y desviación estándar:

Estas medidas numéricas describen la dispersión o variabilidad de la variable aleatoria mediante el promedio o valor esperado de las desviaciones cuadráticas de los valores de x a partir de su media µ.

Sea x la variable aleatoria discreta con distribución de probabilidad f(x) y media µ que viene siendo el valor esperado, la varianza de x es σ^2 cuya fórmula es $\sigma^2 = \sum(X_i^2 * P_i) - E^2(x)$.

Y la desviación estándar se determina por $\sigma = \sqrt{\sigma^2}$.

Sea x la variable aleatoria continua con distribución de probabilidad f(x) y la media µ la varianza de x es:

$\sigma^2 = E[(x - \mu)^2] = \int (X - \mu)^2 F_x \, d_x$

Tomando los datos del ejercicio anterior y utilizando la primera fórmula, determinamos la varianza y la desviación estándar.

Tabla °3. Número de mujeres atendidas diariamente de cáncer de pecho.

Valores posibles de la variable aleatoria(X_i)	**Probabilidad (P_i)**	**X_i^2**	**$X_i^2 * P_i$**
100	0,01	10000	100
101	0,02	10201	204,02
102	0,03	10404	312,12
103	0,05	10609	530,45
104	0,06	10816	648,96
105	0,07	11025	771,75
106	0,09	11236	1011,24
107	0,10	11449	1144,9
108	0,12	11664	1399,68
109	0,11	11881	1306,91
110	0,09	12100	1089
111	0,08	12321	985,68
112	0,06	12544	752,64
113	0,05	12769	638,45
114	0,04	12996	519,84
115	0,02	13225	264,5
TOTAL	**1**		**11680,14**

Fuente: García, I.(2019)

Sustituyendo en la fórmula primera, tenemos:

Varianza: $\sigma^2 = 11680{,}14 - 108{,}02^2 = 11819{,}6$

Desviación estándar: $\sigma = \sqrt{11819{,}96} = 3{,}4376$

EJERCICIOS PROPUESTOS:

1) Determine la distribución de probabilidad si se lanzan 4 monedas al aire, y se desea obtener el N° de Sellos.

2) Al lanzar 2 dados, determine: a) tabla de distribución de probabilidad b) gráfica, de los puntos que se obtienen (suma de las caras de los dos dados)

3) Una caja contiene 30 canicas rojas, y 20 bancas, se quiere obtener el N° de canicas Rojas al extraer 4 de la caja, construya su distribución de probabilidad y gráfica

4) Se realizó una encuesta sobre el N° de hijos en familias de 3 hijos, para determinar el N° de Niñas que nacen, determine: distribución de probabilidad y represente gráficamente.

5) Los siguientes datos se refiere al número de hijos por familia, determine a) valor esperado b) varianza

N° de hijos	0	1	2	3	4	5	6
Proporción de familias	0,05	0,10	0,30	0,25	0,15	0,10	0,05

6) Para la siguiente distribución determine: a) Esperanza matemática b) Varianza

X_i	-3	-2	0	1	2	4	5	6
Casos observados	2	5	7	6	4	3	2	1

Distribución Bernoulli

Se distribuye a varias generaciones de la familia Bernoulli, quienes fueron matemáticos suizos del siglo XVIII, el nombre de Bernoulli ha quedado así asociado con esta clase de experimento, y cada recepción de experimento con solo dos resultados posibles.

Para los fines de la teoría de probabilidades, el interés se centra no en única prueba de Bernoulli, sino en una serie de pruebas de Bernoulli independientes y repetidas, es decir, estamos interesados en más de una prueba, el hecho de que estas pruebas deben ser independientes significa que el resultado de cualquiera de ellas no pueda influir en el resultado de cualquiera de las demás.

Es una distribución de probabilidad discreta, que toma valor de 1 para la probabilidad de éxito, y valor 0 para la probabilidad de fracaso. Como por ejemplo si un jugador de basquetbol está a punto de tirar hacia la parte superior del tablero, sea x = 1 si anota el tiro, sino lo hace x = 0.

Esta distribución está basada en dos resultados ´´cierto´´ y ´´falso´´, donde **p** representa el éxito, y **q** representa el fracaso. Los ensayos repetidos cumplen un papel muy importante en probabilidad y estadística, en especial cuando el número de ensayos es fijo, el parámetro es el mismo para cada ensayo y cuando todos los ensayos son independientes, existen algunas variables aleatorias que aparecen en conexión con ensayos repetidos.

Podríamos por tanto definir este experimento mediante una variable discreta x que toma los valores:

X = 0 Si el suceso no ocurre, X = 1 en caso contrario y se denota de la siguiente manera

$$X \sim Ver(p) \longleftrightarrow X = \begin{cases} 0 \Rightarrow q = 1 - p = P[X = 0] \\ 1 \Rightarrow p = P[X = 1] \end{cases}$$

Un ejemplo típico de este tipo de variable aleatoria consiste en lanzar una moneda al aire y considerar la variable:

$$X = \text{Número de caras obtenidas} = \begin{cases} 0 \Rightarrow q = ½ \\ 1 \Rightarrow p = ½ \end{cases}$$

Para una variable aleatoria de Bernoulli, tenemos que su función de probabilidad es:

$$F(x) = \begin{cases} Q & \text{si } x = 0 \\ P & \text{si } x = 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Distribución Binomial

Este tipo de distribución fue desarrollada por Jakob Bernoulli (Suiza, 1654 – 1705), es la principal distribución de probabilidad discreta para variables dicotómicas, es decir, que sólo pueden tomar dos posibles resultados. Es una distribución de probabilidad ampliamente utilizada de una variable aleatoria, esta describe varios procesos de interés para los administradores. Describe datos discretos, resultantes de un experimento denominado proceso de Bernoulli en honor al matemático suizo Jacob Bernoulli, quien vivió en el siglo XVII.

En estadística, la distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de *n* ensayos de Bernoulli independientes entre si, con una probabilidad fija *p* de ocurrencia del éxito entre los ensayos. Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, solo dos resultados son posibles. A uno de estos se denomina *éxito* y tiene una probabilidad de ocurrencia *p* y al otro, *fracaso*, con una probabilidad q = 1 – p.

La distribución de probabilidad binomial es uno de los modelos matemáticos (expresión matemática para representar una variable) que se utiliza cuando la variable aleatoria discreta es el número de éxitos en una muestra compuesta por n observaciones.

En la distribución Binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para n = 1, la binomial se convierte, de hecho, en una distribución Bernoulli.

Algunos ejemplos típicos de una distribución binomial son: al nacer un bebe puede ser hembra o varón, un equipo de baloncesto puede ganar o perder, en un examen con preguntas cerradas dicotómicas puede ser verdadero o falso, un tratamiento médico, como por ejemplo la vacuna para la gripe A, puede ser efectivo o inefectivo, etc,.

Para representar una variable aleatoria x, que sigue una distribución binomial de parámetros n y p, se escribe:

$$\boxed{X \sim B(n,p)}$$

Propiedades:

1. Cada observación se clasifica en una de dos categorías, mutuamente excluyentes (los eventos no pueden ocurrir de manera simultánea. Ejemplo: una persona no puede ser de ambos sexos) y colectivamente

exhaustivos (uno de los eventos debe ocurrir, ejemplo: al lanzar una moneda, sino ocurre cruz, entonces ocurre cara). A estas categorías se les denomina éxito y fracaso.

2. La probabilidad de que una observación se clasifique como éxito (p), es constante de una observación a otra, de la misma forma, la probabilidad de que una observación se clasifique como fracaso (q), es constante en todas las observaciones.
3. La variable aleatoria binomial tiene un rango de 0 a n.

Ecuaciones:

1) $P(x=k) = \binom{n}{k} p^k . q^{n-k}$ $\qquad \binom{n}{k} = {}_nC_k$ "se determina en la calculadora científica"

2) $$P(x=k) = \frac{n!}{k! * (n-k)!} * p^k * (1-p)^{n-k}$$

Depende del ejercicio se puede utilizar cualquier ecuación.
X: variable aleatoria, representa el número de éxitos obtenidos en cada prueba del experimento
K: número de aciertos o éxitos deseados
n: número de experimentos o ensayos efectuados
p: probabilidad de éxito, como por ejemplo que salga "*cara*" al lanzar una moneda
q: probabilidad de fracaso, también conocida como 1 – p,

Parámetros de la distribución:

Valor esperado, esperanza matemática o media es: $\mu = n.p$
Varianza: $\sigma^2 = n.p.q$
Desviación típica: $\sigma = \sqrt{n.p.q}$

Con el siguiente **ejemplo** veremos que la distribución binomial es fácil de entender y es importante en nuestra vida cotidiana imaginemos una escuela primaria donde los alumnos llegan tarde a menudo, 5 alumnos están en el jardín de niños. La directora lleva tiempo estudiando el problema, habiendo llegado a la conclusión de que hay una probabilidad de 0,4 de que un alumno llegue tarde y de que los alumnos lleguen independientemente uno de otro ¿Cómo trazamos una distribución binomial de probabilidad que ilustre las probabilidades de que 0,1,2,3,4 ó 5 estudiantes lleguen tarde simultáneamente?

Solución:

Datos:
n = 5
p = 0,4
q = 1 – p = 1 – 0,4 = 0,6
P(x = 0,1,2,3,4,5) = ?

Aplicando la ecuación **1**, se tiene:

$P(X=0) = {}_5C_0 . 0{,}4^0 . 0{,}6^{5-0} = 0{,}07776x100 = 7{,}776\%$

$P(X=1) = {}_5C_1 . 0{,}4^1 . 0{,}6^{5-1} = 0{,}2592x100 = 25{,}92\%$

$P(X=2) = {}_5C_2 . 0,4^2 . 0,6^{5-2}$ = 0,3456x100 = 34,56%

$P(X=3) = {}_5C_3 . 0,4^3 . 0,6^{5-3}$ = 0,2304x100 = 23,04%

$P(X=4) = {}_5C_4 . 0,4^4 . 0,6^{5-4}$ = 0,0768x100 = 7,68%

$P(X=5) = {}_5C_5 . 0,4^5 . 0,6^{5-5}$ = 0,01024x100 = 1,024%

Otro ejemplo para demostrar el cálculo de la distribución binomial, sería el siguiente, con el propósito de verificar si se aceptan los lotes de piezas que se reciben en una determinada fábrica, se lleva a cabo un plan de control consistente en seleccionar 10 artículos al azar de cada lote y determinar el número de piezas defectuosas. Un lote rechaza si se encuentran dos o más piezas defectuosas. ¿Cuál es la probabilidad de aceptar lotes con un 5% de piezas defectuosas?

Solución:

Sea el suceso A = ser pieza defectuosa

n = 10 artículos

P = 5%/100 = 0,05 es la proporción de piezas defectuosas

q = 1 – 0,05 = 0,95

x = número de piezas defectuosas en el lote

P(aceptar lotes de piezas defectuosas) = $P(X < 2)$

$P(X < 2) = P(X = 0) + P(X = 1)$ Aplicando la ecuación 1

$= {}_{10}C_0 . 0,05^0 . 0,95^{10-0} + {}_{10}C_1 . 0,05^1 . 0,95^{10-1}$

= 0,5987 + 0,3151

= 0,9138*100

$P(X < 2)$ = 91,38%

Interpretación: el 91,38% es la probabilidad de aceptar lotes de piezas defectuosas.

EJERCICIOS PROPUESTOS DE DISTRIBUCIÓN BINOMIAL

1) La última novela de un autor ha tenido un gran éxito, hasta el punto de que el 80% de los lectores ya la han leído. Un grupo de 4 amigos son aficionados a la lectura, Determine:

1 ¿Cuál es la probabilidad de que en el grupo hayan leído la novela 2 personas?

2 ¿Y cómo máximo 2?

2) Un laboratorio afirma que una droga causa efectos secundarios en una proporción de 3 de cada 100 pacientes. Para contrastar esta afirmación, otro laboratorio elige al azar a 5 pacientes a los que aplica la droga. ¿Cuál es la probabilidad de los siguientes sucesos?

1 Ningún paciente tenga efectos secundarios
2 Al menos dos tengan efectos secundarios

3) Se sabe que el 30% de las piezas que fabrica una máquina son defectuosas, determine:
1 La probabilidad de que una muestra de 4 piezas tomadas al azar todas son defectuosas
2 La probabilidad de que en un grupo de 5 piezas elegidas al azar haya 2 buenas
3 Cuál es la probabilidad de elegir una muestra de 6 piezas, al menos la mitad sean buenas
4 La probabilidad de que una muestra aleatoria de 4 piezas haya más de 2 defectuosas

Distribución Poisson

Distribución Poisson, es una distribución muy usada en medicina y biología, se deriva del proceso de Poisson en honor al matemático francés Simeón Dennis Poisson (1781 – 1840). Se aplica para variables aleatorias discretas.

Los experimentos que resultan en valores numéricos de una variable aleatoria X, quien representa el número de resultados durante el intervalo de tiempo dado o una región específica, frecuentemente se llaman experimentos de Poisson. El intervalo de tiempo dado puede ser de cualquier duración de tiempo, por ejemplo, un minuto, un día, una semana, un mes o inclusive un año.

De aquí que un experimento de Poisson puede generar observaciones para la variable aleatoria X, que representa el número de algún evento en un lapso de tiempo dado, por ejemplo: número de llamadas telefónicas, número de vehículos que pasan por un peaje, número de defectos de una pieza, etc.

Un experimento de Poisson tiene las siguientes propiedades:

- ✓ El número de resultados que ocurren en un intervalo de tiempo o región específico es independiente del número que ocurre en cualquier otro intervalo de tiempo.
- ✓ La probabilidad de que un resultado muy sencillo ocurra en un intervalo de tiempo muy corto o en una región muy pequeña es proporcional a la longitud del intervalo de tiempo o al tamaño de la región.
- ✓ La probabilidad de que más de un resultado ocurra en un intervalo de tiempo tan corto o en esa región tan pequeña es despreciable.
- ✓ El número promedio de veces que ocurre un éxito por cada unidad de tiempo o de espacio es constante.

Hay que hacer notar que en esta distribución el número que éxitos que ocurren por unidad de tiempo, área o producto es totalmente al azar y que cada intervalo de tiempo es independiente de otro intervalo dado, así como cada área es independiente de otra área dada y cada producto es independiente de otro producto dado. La fórmula a emplear es la siguiente:

$$\boxed{P(X) = \frac{e^{-\lambda} \cdot \lambda^{K}}{K!}}$$

P(X): Es la probabilidad K: Número medio de ocurrencia/unidad

X: Variable aleatoria λ: Promedio

!: Factorial de un número

Ejemplos:

1) Si un banco recibe en promedio 6 cheques sin fondo por día, ¿cuáles son las probabilidades de que reciban, a) 4 cheques sin fondo en un día dado. b) 10 cheques sin fondo en cualquiera de 2 días consecutivos?
Solución:
a) Datos
X= 4, es la variable que nos define el número de cheques sin fondo que llegan al banco en un día cualquiera, 1,2,3,4……..
λ= 6 cheques sin fondo/día
e= 2,718 (constante)
Utilizando la fórmula anterior tenemos:

$$P(X=4, \lambda=6) = \frac{(6^4)x(2{,}718^{-6})}{4!} = \frac{1296x0{,}00248}{24} = \boxed{0{,}13392}$$

b) X= 10, número de cheques sin fondo que llegan al banco en 2días consecutivos
λ= 6x2 = 12, cheques sin fondo en promedio que llegan al banco en 2 días consecutivos
e= 2,718

$$P(X= 10, \lambda= 12) = \frac{(12^{10})x(2{,}718^{-12})}{10!} = \boxed{0{,}104953}$$

EJERCICIOS PROPUESTOS

1) Un 10 % de las herramientas producidas por una fábrica son defectuosos, hallar la probabilidad de que en una muestra de 10 herramientas tomadas al azar, exactamente sean defectuosas.
2) De los estudiantes que asisten a un liceo, el 5% posee gran aptitud para las artes plásticas, se elige una muestra de 100 alumnos al azar, ¿Cuál es la probabilidad 1,2 y 3 con referida aptitud?
3) En una localidad son albinos 2 de cada mil personas, se elige al azar una muestra de 100 y se desea saber la probabilidad de que sean albinos a) 3 b) entre 1 y 4 c) ninguno
4) Si el 2% de los libros encuadernados en cierto taller tiene encuadernación defectuosa, para obtener la probabilidad de que 5 de 400 libros encuadernados en este taller tengan encuadernaciones defectuosas usamos la distribución de Poisson.
5) La probabilidad de tener un accidente de tráfico es de 0,02 cada vez que se viaja, si se realizan 300 viajes, ¿cuál es la probabilidad de tener 3 accidentes?
6) La probabilidad de que un niño nazca pelirrojo es de 0,012. ¿Cuál es la probabilidad de que entre 800 recién nacidos haya 5 pelirrojos?
7) El número de pulsos que llegan a un contador GEIGER se presentan en promedio de 6 pulsos por minuto. Hallar la probabilidad de que en 15 minutos se reciban exactamente 20 pulsos.
8) En la inspección de hojalata producida por un proceso electrolítico continuo, se identifican 0,2 imperfecciones en promedio por minuto. Determine la probabilidad de: a) una imperfección en 3 minutos. b) 2 imperfecciones en 5 minutos c) una imperfección en 15 minutos.

9) Supongamos que el número de imperfecciones en un alambre delgado de cobre sigue una distribución de Poisson, con una media de 2,3 imperfecciones por milímetro.
   a) Determine la probabilidad de 2 imperfecciones en un milímetro de alambre.
   b) Determine la probabilidad de 10 imperfecciones en 5 milímetros de alambre.
   c) Determine la probabilidad de una imperfección en 2 milímetros de alambre.

Distribución Normal

En estadística y probabilidad se llama distribución Normal, distribución de Gauss o, distribución gaussiana, a una de las distribuciones de probabilidad de variable continua, que con más frecuencia aparece en fenómenos reales. La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro. Esta curva se conoce como campana de Gauss.

La importancia de esta distribución radica en que permite modernizar numerosos, fenómenos naturales, sociales y psicológicos, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

La distribución normal también aparece en muchas áreas de la propia estadística, ya que muchos test estadísticos están basados en una supuesta normalidad. En probabilidad la distribución normal aparece como el límite de varias distribuciones de probabilidades continuas y discretas.

La distribución normal estándar, o tipificada o reducida, es aquella que tiene por media el valor cero, es decir $\mu = 0$, y por desviación típica la unidad, es decir $\sigma = 1$.

En resumen, la importancia de la distribución se debe principalmente a que hay muchas variables asociadas a fenómenos naturales que sigue el modelo de la normal, como por ejemplo:

- ✓ Caracteres morfológicos de individuos (personas, animales, plantas, etc.,) de una especie, como tallas, pesos, envergaduras, diámetros, perímetros.
- ✓ Caracteres fisiológicos, por ejemplo: efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.
- ✓ Caracteres sociológicos, por ejemplo: consumo de cierto producto por un mismo grupo de individuo, puntuaciones de un examen.
- ✓ Caracteres psicológicos, por ejemplo: cociente intelectual, grado de adaptación a un medio.
- ✓ Errores cometidos al medir ciertas magnitudes.
- ✓ Valores estadísticos muéstrales, por ejemplo: la media.
- ✓ Otras distribuciones como la de Poisson son aproximaciones normales. En general cualquier característica que se obtenga como suma de muchos factores.

Propiedades de la Distribución Normal:

1. La curva tiene forma de campana.
2. Es simétrica con respecto al eje y, ya que tiene una única moda, que coincide con su media y mediana.
3. La curva normal es asintótica, con respecto al eje x. Por ello, cualquier valor entre menos infinito y más infinito es teóricamente posible.
4. Para éste tipo de variable existe una probabilidad de un 50% de observar un dato mayor que la media, y un 50% de observar un dato menor.

5. La distancia entre la línea trazada en la media y el punto de inflexión de la curva es igual a una desviación típica representado por el signo sigma (Ϭ), cuando mayo sea sigma, más aplanada será la curva de la densidad.

Se deduce que no existe una única distribución normal, sino una familia de distribuciones con una forma común diferenciadas por los valores de su media y varianza, cuya fórmula es la siguiente:

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

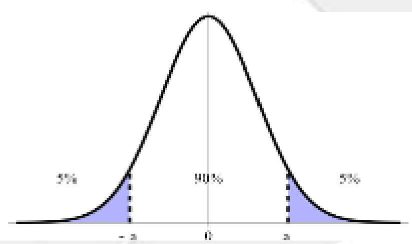
Z: Es la variable estándar o Tipificada

X: Variable aleatoria Continua

µ: Media, media de la población, promedio

Ϭ: Desviación de la población

Representación gráfica:



Ejemplo:

Supongamos que se sabe la cantidad promedio de cervezas, que 20 alumnos del 3er semestre de Administración Industrial del IUTEPI se toman los días viernes, es de 45 latas de cerveza con una desviación estándar de 8 latas de cervezas. Determine, a) ¿La probabilidad de que tomen entre 35 a 47 latas de cerveza? b) ¿Podemos saber cuántos alumnos han tomado más de51 latas de cervezas?

Solución A:

Datos:

N = 20 alumnos
µ = 45 latas de cervezas
Ϭ = 8 latas de cervezas
Z = desconocida
X_1= 35 y X_2 = 45

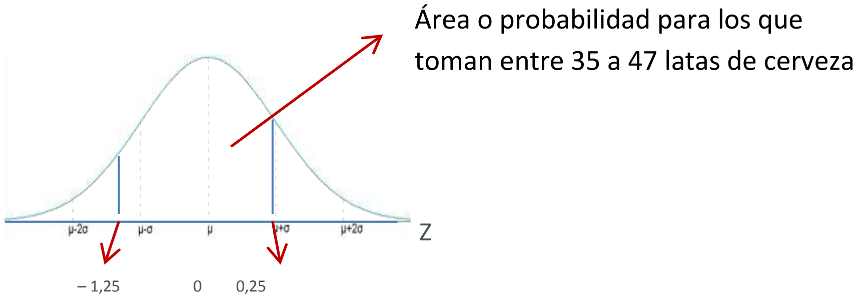
Sustituyendo los datos, calculamos Z:

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

$$Z_1 = \frac{35 - 45}{8} \Rightarrow Z_1 = -1,25$$

$$Z_2 = \frac{47 - 45}{8} \Rightarrow Z_2 = 0,25$$

Representación gráfica:



Se ubican los dos valores de Z en la tabla de distribución normal, para éste ejemplo se tomó la siguiente tabla como referencia: ESTAD II\TABLA DISTRIBUCIÓN NORMAL TIPIFICADA.docx

Donde se obtienen los valores de las Probabilidades, es decir,

Para Z = – 1,25 su P = 0,8944*100 = 89,44%

Para Z = 0,25 su P = 0,5987*100 = 59,87%

Por lo tanto,

P(– 1,25 ≤ Z ≤ 0,25) = P(Z ≤ 0,25) – P(Z ≤ – 1,25) = 89,44% – 59,87% = 29,57%

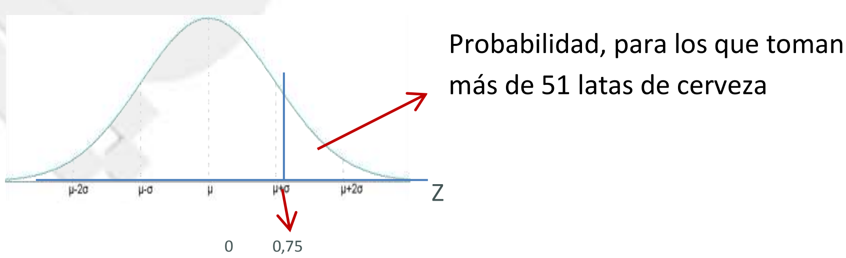
Solución B:

N = 20 alumnos
μ = 45 latas de cervezas
σ = 8 latas de cervezas
Z = desconocida
X_1= 35 y X_2 = 45

Sustituyendo los datos, calculamos Z

$$Z = \frac{51 - 45}{8} = 0{,}75$$

Representación gráfica:



P(Z > 0,75) = 1 – P(Z ≤ 0,75) = 1 – 0,7734 = 0,2266*100
P(Z > 0,75) = 22,66%

Ahora calculamos la cantidad de alumnos que se toman más de 51 latas de cervezas,

$n = N*P / 100 = 20*22{,}66\%/100$
$n = 4{,}532$
$n \approx 5$ Alumnos

Podemos concluir que 5 alumnos se toman más de 51 latas de cervezas.

Estimación de parámetros, estimación puntual

En este objetivo se trata el problema de la estimación de parámetros. Para ello, comenzamos recordando algunos conceptos básicos de la inferencia estadística que ya fueron introducidos en el tema anterior, y que serán necesarios para la construcción y el estudio de los estimadores:

Población: conjunto homogéneo de individuos sobre los que se estudian características observables con el objetivo de extraer alguna conclusión. Por abuso de notación, en ocasiones nos referimos a la distribución que sigue la variable de interés en vez de al conjunto de individuos. Así, se dice que estamos ante una población Normal indicando que la variable que nos interesa sigue una distribución normal.

Parámetro: característica de la población, como la media y la varianza (o desviación típica) en la distribución Normal o la probabilidad de éxito en la Binomial son parámetros. Si conocemos su valor (o si somos capaces de aproximarlo con suficiente precisión) podremos responder a cualquier pregunta sobre la distribución.

Estadístico: cualquier función de la muestra. Por ejemplo, la media o la varianza muéstrales son estadísticos.

Estimadores: son estadísticos independientes de los parámetros de la población, y que se utilizan para aproximarlos. Si θ es el parámetro de interés, el estimador se denotará por ˆθ. En el caso de una población Normal, podemos considerar la media muestral como estimador de la media poblacional (es decir, X = μ) y la varianza muestral como estimador de la varianza poblacional ($s^2 = \sigma^2$). Para una distribución Bi(m, p), donde m denota el número de pruebas de Bernoulli, la proporción p se puede estimar a partir de la proporción poblacional (que denotaremos por p). Por tanto, X, s^2 y p son estimadores puntuales de μ, σ^2 (en distribución Normal) y p (en distribución Binomial), respectivamente.

Método de muestreo: procedimiento para seleccionar una muestra. Si en una población queremos obtener una muestra de un cierto tamaño n (siendo n menor que el tamaño de la población), la manera de obtener esta muestra no es única.

Los estimadores siempre suministran dispersión aleatoria. Como sabemos del monográfico sobre muestreo, el conjunto de todas las muestras de un mismo diseño que provienen de una misma población suministran valores diferentes. Esta circunstancia indica que existe una variación aleatoria con la que hay que vivir porque es inevitable.

Es posible que el estimador escogido tenga sesgo, es decir, que no solo esté alrededor de un punto, sino que el punto sobre el que varía no es el valor poblacional, verdadero u objetivo de nuestro interés. Esto si es evitable. Así que los estimadores que utilizamos intentamos que sean insesgados, es decir, que carezcan de sesgo.

El recurso que utilizamos para ello es el valor esperado, es decir la media aritmética de la distribución muestral del estimador. El valor esperado es, como dice la expresión, el valor que esperamos. Cabe elegir un estimador tal que el valor esperado coincida con el parámetro. Esto ocurre si utilizamos la media aritmética de la muestra como estimador de la media aritmética de la población, pues E(X) = μ.

También ocurre con las proporciones, pues E(p) = π. Pero no ocurre así con la varianza (y, por tanto, tampoco con la desviación tipo). En el muestreo aleatorio simple donde las poblaciones son de gran tamaño, es la cuasivarianza el estadístico escogido como estimador de la varianza poblacional, es decir, la cuasivarianza es un estimador insesgado de la varianza poblacional.

Por ejemplo, podemos tener interés en conocer cuántas personas votarán al partido HH en las próximas elecciones o cuántos cigarrillos o cuantos cigarrillos van a consumirse en el mes de abril. Para responder, utilizamos un recurso indirecto que parte de una estimación previa, bien sea de una media aritmética o de una proporción.

Ejemplo:

Supongamos que la población que nos interesa cuenta con un millón de habitantes, tomando una muestra de 200 personas, de los que 38 dicen que votarán por el partido HH, esto significa que: 38/200*100 = 19%.

Una estimación puntual establece que el 19% de la población votará a HH, como hay un millón de habitantes, entonces, podemos determinar: 1.000.000*19/100 = 190.000 personas.

Supongamos también que se fuman 50 cigarrillos por término medio cada mes. Si ese es el valor de la media aritmética de la muestra, la estimación puntual afirmará que en la población se fumarán 50 cigarrillos por persona durante el mes de abril, por término medio.

Como hay un millón de habitantes, para el mes de abril habrán consumido 50 millones de cigarrillos, así pues, en la estimación de totales no realizamos un camino alternativo específico, sino que ampliamos la estimación realizada previamente, sea de una proporción o de una media.

Estimación por intervalos, para la media y la proporción

Intervalo de confianza: Es aquel intervalo que con cierto nivel de confianza, contiene al parámetro que se está estimando. Se llama intervalo de confianza a un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto. Formalmente, estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional.

Nivel de confianza: Es la probabilidad de éxito, es el intervalo calculado que contiene el verdadero valor del parámetro (1 – α). En estas circunstancias, α es el llamado error aleatorio o nivel de significación, esto es, una medida de las posibilidades de fallar en la estimación mediante tal intervalo.

El nivel de confianza y la amplitud del intervalo varían conjuntamente, de forma que un intervalo más amplio tendrá más probabilidad de acierto (mayor nivel de confianza), mientras que para un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error.

Intervalo de confianza para la media de una población: Dado una población con media μ y desviación típica σ, se pueden demostrar que la media de todas las medias muéstrales coincide con la media poblacional.

Pero además, si el tamaño de las muestras es lo suficientemente grande, o la distribución poblacional es normal, la distribución de medias muéstrales es, prácticamente, una distribución normal (o gaussiana) con media μ y una desviación típica dad por la siguiente expresión:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

En la distribución normal de medias se puede calcular el intervalo de confianza donde se encontrará la media poblacional si solo se conoce una media muestral ($\overline{X}$), con una confianza determinada, habitualmente se manejan valores de confianza del 95% y del 99%.

Fórmula para determinar el Intervalo de confianza para la media:

$$\overline{X} \pm z_{\alpha/2} * \sigma/\sqrt{n}$$

Ejemplo:

Para una muestra de 81 habitantes de cierta población se obtuvo una estatura media de 167 cm, por estudios anteriores se sabe que la desviación típica de la altura de la población es de 8 cm, construya un intervalo de confianza para la estatura media de la población al 95%.

Solución:

Datos:

$n = 81$
$\overline{X} = 167$ cm.
$\sigma = 8$ cm.
$1 - \alpha = 95\%$ ➡ Utilizando la tabla de distribución normal, Z = 1,96

Empleando la fórmula anterior se tiene lo siguiente;

167 ± 1,96*8/√81 → 167 ± 1,74 → 167 + 1,74 = 168,74
167 ± 1,74 → 167 – 1,74 = 165,26

El intervalo de confianza (IC), queda de la siguiente manera, IC: $165,26 \leq \mu \leq 168,74$

Estimación de parámetros para la proporción:

Construiremos intervalos de confianza para la proporción p en la distribución Binomial y para la media μ en la distribución Normal. Los estimadores que hemos introducido para la proporción y la media (pˆ y X, respectivamente) son simétricos y podemos calcular o aproximar su error típico.

Consideremos pˆ, proporción muestral, como estimador de p, podemos construir un intervalo de confianza de nivel (cobertura) (1 − α) para p. En este caso, el error típico se aproxima por qpˆ(1−pˆ) n y, dado que pˆ tiene una distribución Normal, consideramos los cuartiles de una N(0, 1). En concreto, para los intervalos de confianza usuales.

Fórmula para calcular intervalo de confianza para la proporción:

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

Ejemplo:

Una maquina fabrica piezas de precisión y en una caja de 200 piezas, resultan 7 piezas defectuosas, a un nivel de confianza del 99%, ¿Entre que valores se puede esperar que esté la verdadera proporción de piezas defectuosas por la máquina?

Solución:

Datos:

n = 200 piezas $\rightarrow$ 7 piezas defectuosas

$\hat{p}$ = probabilidad de piezas defectuosas

$\hat{q}$ = probabilidad de piezas buenas

$1 - \alpha = 99\%$ $\rightarrow$ Por la tabla de la distribución normal, $Z_{99\%} = 2,58$

$\hat{p} = 7/200 = 0,035$

$\hat{q} = 1 - \hat{p} = 1 - 0,035 = 0,965$

Aplicando la fórmula se tiene:

$$0,035 \pm 2,58 * \sqrt{(0,035*0,965/200)} = 0,035 \pm 0,034$$

0,035 + 0,034 = 0,069

0,035 – 0, 034 = 0,001

IC: $0,001 \leq \hat{p} \leq 0,069$

Distribución T- studen

En estadística y probabilidad, la distribución t de student es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño, es la base para la determinación de las diferencias entre dos medias muéstrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones.

La distribución surge, en la mayoría de los estudios estadísticos prácticos, cuando la desviación típica de una población se desconoce y debe ser estimada a partir de los datos de una muestra, aparece de manera natural al realizar la prueba de student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce la desviación típica de una población y ésta debe ser estimada a partir de dos datos de una muestra.

Por lo que sigue una distribución Normal de media 0 y varianza 1, sin embargo, dado que la desviación estándar no siempre es conocida de antemano, Gosset estudió un cociente relacionado:

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}},$$

Características:

1. Al igual que la distribución normal, es una distribución continua.
2. La distribución t tiene una media de cero, es simétrica respecto a la media, y se extiende hasta el infinito en ambas direcciones.
3. Tiene forma acampanada y simétrica.
4. No hay una distribución t, sino una familia de distribuciones t, todas con la misma media cero, pero con su respectiva desviación estándar diferente de acuerdo con el tamaño de la muestra n.
5. La distribución t – student, es más ancha y más plana en el centro que la distribución normal estándar como resultado de ello se tiene una mayor variabilidad en las medias de muestra calculadas a partir de muestras más pequeñas. Sin embargo a medida que aumenta el tamaño de la muestra, la distribución t se aproxima a la distribución normal estándar.
   En el siguiente video se le da un ejercicio de la distribución:
   https://www.youtube.com/watch?v=vNnakZ5oJTc

EJERCICIOS PROPUESTOS:

1. Los puntajes de un grupo de estudiantes se comportan normal, con promedio de 50, sin embargo, Se tomó una muestra de 9 estudiantes encontrando una desviación de 6 y un promedio de 52. ¿Cuál es la probabilidad de que el promedio Sea de 50?

2. Las puntuaciones en un test que mide la variable de creatividad de los adolescentes, siguen una distribución normal de media 11,5, en un centro escolar se ha implantado un programa de estimulación de creatividad tomando una muestra de 30 alumnos, la media muestral es de 12,47 y la desviación típica de 5,22, determine: a) t_α , si el nivel de significación es del 95%   b) t_c

3. Un fabricante de focos afirma que su producto dura en promedio 500 horas de trabajo, para conservar este promedio, el fabricante toma una muestra de 25 focos cada mes, y arrojo un promedio de 505,36 horas, una desviación de 12,07 y un nivel de confianza del 90% determine el t_α y el t_c

Contraste de Hipótesis, estadísticas, hipótesis nula, alterna

Los métodos de contraste de hipótesis son modelos utilizados en inferencia estadística cuyo objetivo es comprobar si una estimación se adapta a los valores poblacionales. El objetivo de los métodos de contraste de hipótesis es comprobar si una estimación se adapta a la realidad de forma fiable.

Los supuestos se denominan hipótesis paramétricas. Es decir, se estable un criterio de decisión. Si con esa condición se acepta la hipótesis de referencia, entonces podemos afirmar con cierta probabilidad que la estimación puede ser muy cercana al supuesto valor real.

Prueba de Hipótesis: Es un procedimiento estadístico por medio del cual se toma una muestra de una población y con los datos de la muestra se trata de estimar si un parámetro es igual o no a uno pre-establecido.

Es decir se formulan hipótesis y se verifica si se puede rechazar o aceptar dicha hipótesis. En todo contraste de hipótesis existen dos supuestos. La Hipótesis nula (Ho), que recoge la idea de que un valor tiene un valor predeterminado. Si se rechaza la hipótesis nula (Ho), entonces se acepta la Hipótesis alternativa (H_1).

Para profundizar los conocimientos sobre el contraste de Hipótesis revisar el siguiente link: https://es.wikipedia.org/wiki/Contraste_de_hip%C3%B3tesis

Referencia Bibliográfica:

- ✓ Valera R. *Manual de Estadística Básica.* 4ta
- ✓ Dura Peiró, J. M. y López Cuñat, J.M.(1992) *Fundamentos de Estadística. Estadística Descriptiva y Modelos Probabilísticos para la Inferencia*. Madrid: Ariel Editorial.
- ✓ Martin Pliego, F. y Ruiz – Maya, L. (1995) *Estadística I: Probabilidad.* Madrid: AC.
- ✓ Montiel, A.M., Rius, F. y Baron, F.J. (1997) *Elementos Básicos de Estadística Económica y Empresarial.* Madrid: Prentice Hall.