**Conceptos Básicos en Estadística**

**Media (Promedio)**

**Definición**: La media aritmética, o simplemente la media, es una medida de tendencia central que se calcula sumando todos los valores de un conjunto de datos y dividiendo la suma por el número total de valores.

**Fórmula:**

= ​

donde es la media, es el número de observaciones, y ​ son los valores individuales.

**Ejemplo:** Consideremos el siguiente conjunto de datos que representa las calificaciones de 5 estudiantes en un examen:

**Para calcular la media:**

Interpretación: La media de 86 indica que, en promedio, los estudiantes obtuvieron una calificación de 86 en el examen.

**Mediana**

Definición: La mediana es el valor que separa la mitad superior de la mitad inferior de un conjunto de datos ordenado. Si el número de observaciones es impar, la mediana es el valor central. Si el número de observaciones es par, la mediana es el promedio de los dos valores centrales.

Cálculo:

1. Ordenar los datos.
2. Identificar el valor central si el número de observaciones es impar, o calcular el promedio de los dos valores centrales si el número de observaciones es par.

**Ejemplo:**

Usando el mismo conjunto de datos ordenado: 75, 85, 88, 90, 92.

Dado que hay 5 valores (número impar), la mediana es el tercer valor:

Mediana = 88

**Interpretación:** La mediana de 88 indica que la mitad de los estudiantes obtuvieron calificaciones menores o iguales a 88, y la otra mitad obtuvo calificaciones mayores o iguales a 88.

**Desviación Estándar**

Definición: La desviación estándar es una medida de dispersión que indica cuánto se desvían los valores de un conjunto de datos respecto a la media. Una desviación estándar baja indica que los valores tienden a estar cerca de la media, mientras que un alta indica mayor dispersión.

**Fórmula:**

donde es la desviación estándar, es la media, y ​ son los valores individuales.

**Ejemplo:**

Consideremos el conjunto de datos:

85,90,75,88,92

La media ya calculada es 86.

Las desviaciones de cada valor respecto a la media son:

85 – 86 = −1

90 − 86 = 4

75 – 86 = −11

88 – 86 = 2

92 − 86 = 6

Cuadrados de las desviaciones:

Sumamos los cuadrados de las desviaciones:

1 + 16 + 121 + 4 + 36 = 178

Calculamos la varianza:

== 35.6

Finalmente, la desviación estándar:

σ =≈ 5.97

**Interpretación**: Una desviación estándar de aproximadamente 5.97 indica que, en promedio, las calificaciones de los estudiantes se desvían de la media en 5.97 puntos.

**Varianza**

Definición: La varianza es una medida de dispersión que calcula el promedio de los cuadrados de las desviaciones respecto a la media. Es esencialmente el cuadrado de la desviación estándar y proporciona una idea de la variabilidad de los datos.

**Fórmula:**

donde es la varianza, es la media, y son los valores individuales.

**Ejemplo:**

Usando el conjunto de datos y cálculos anteriores:

= 35.6

Interpretación: Una varianza de 35.6 indica el grado de dispersión de las calificaciones respecto a la media. Al ser el cuadrado de la desviación estándar, proporciona una medida de variabilidad en las mismas unidades cuadradas que los datos originales.

**Conclusión**

Entender y calcular la media, mediana, desviación estándar y varianza es fundamental en estadística, ya que estas medidas proporcionan información valiosa sobre la tendencia central y la dispersión de los datos. Estas estadísticas son esenciales para interpretar y analizar datos en diversos campos, incluyendo la investigación académica, la ingeniería, las ciencias sociales y la economía.

**Distribución Binomial**

**Definición**: La distribución binomial es una distribución de probabilidad discreta que describe el número de éxitos en una secuencia de ensayos de Bernoulli independientes. Cada ensayo de Bernoulli tiene solo dos posibles resultados: éxito (con probabilidad ) y fracaso (con probabilidad 1−). Los ensayos son idénticos e independientes.

**Función de Masa de Probabilidad**: La probabilidad de obtener exactamente éxitos en ensayos está dada por:

* ()= es el coeficiente binomial.
* es el número de ensayos.
* es el número de éxitos.
* es la probabilidad de éxito en un ensayo.
* es la probabilidad de fracaso.

**Ejemplo Práctico**: Supongamos que estamos interesados en encontrar la probabilidad de obtener exactamente 3 éxitos en 10 lanzamientos de una moneda justa. Aquí, un "éxito" puede ser definido como obtener cara, con p=0.5.

1. **Datos del problema**:
   * Número de ensayos (): 10
   * Número de éxitos (): 3
   * Probabilidad de éxito (): 0.5
2. **Cálculo**:

P (X = 3) = ()

P (X = 3) =

P (X = 3) =

P (X = 3) = 120 \*

P (X = 3) = 120 \* 0.0009765625

**Interpretación**: La probabilidad de obtener exactamente 3 caras en 10 lanzamientos de una moneda justa es aproximadamente 0.117, o 11.7%.

**Distribución de Poisson**

**Definición**: La distribución de Poisson es una distribución de probabilidad discreta que expresa la probabilidad de un número dado de eventos que ocurren en un intervalo de tiempo o espacio fijo, si estos eventos ocurren con una tasa promedio constante λ y son independientes del tiempo transcurrido desde el último evento.

**Función de Masa de Probabilidad**: La probabilidad de observar eventos en un intervalo está dada por: donde:

* λ es la tasa promedio de ocurrencia de los eventos (media).
* es el número de eventos observados.
* es la base del logaritmo natural (aproximadamente 2.71828).

**Ejemplo Práctico**: Supongamos que estamos interesados en encontrar la probabilidad de que ocurran exactamente 4 eventos en un intervalo de tiempo dado, si el promedio de eventos por intervalo es 2.

1. **Datos del problema**:
   * Tasa promedio (λ): 2
   * Número de eventos (): 4
2. **Cálculo**:

P( X = 4) =

P( X = 4) =

P( X = 4) =

P( X = 4) =

P( X = 4) = 0.0122

**Interpretación**: La probabilidad de que ocurran exactamente 4 eventos en un intervalo dado, cuando la tasa promedio es de 2 eventos por intervalo, es aproximadamente 0.0902, o 9.02%.

**Conclusión**

La comprensión y aplicación de las distribuciones binomial y de Poisson son fundamentales en estadística para modelar eventos discretos. La distribución binomial es adecuada para situaciones donde se tienen ensayos repetidos e independientes con dos resultados posibles, mientras que la distribución de Poisson es ideal para modelar la ocurrencia de eventos en un intervalo continuo de tiempo o espacio. Estos conceptos y técnicas son vitales en diversos campos como la ingeniería, las ciencias sociales y las ciencias de la vida, permitiendo una mejor interpretación y toma de decisiones basada en datos estadísticos.

**Distribución Normal Estándar y Uso de la Tabla Normal**

**Distribución Normal Estándar**

**Definición**: La distribución normal estándar es una distribución normal con una media (μ) de 0 y una desviación estándar (σ) de 1. Esta distribución se denota como

**Función de Densidad de Probabilidad**:

donde:

* es el valor de la variable aleatoria normal estándar.
* es la media.
* es la desviación estándar.

La distribución normal estándar se utiliza para simplificar los cálculos de probabilidades de cualquier distribución normal mediante la transformación de cualquier variable aleatoria con media y desviación estándar a la variable mediante la fórmula:

**Uso de la Tabla Normal**

La Tabla Normal, también conocida como Tabla , proporciona las áreas bajo la curva normal estándar para valores específicos de. Estas áreas corresponden a las probabilidades acumuladas desde la cola izquierda de la distribución hasta el valor .

**Pasos para Encontrar una Probabilidad Usando la Tabla Normal**:

1. **Transformar la Variable**: Si la variable XXX no es estándar, transformarla a la variable utilizando:

Z =

1. **Buscar el Valor en la Tabla**:
   * Localiza el valor de en la Tabla Normal. La tabla está organizada de modo que las filas representan el primer decimal de y las columnas representan el segundo decimal de .
2. **Interpretar el Valor**:
   * El valor encontrado en la tabla representa la probabilidad acumulada es decir, la probabilidad de que la variable aleatoria estándar tome un valor menor o igual a .

**Ejemplo Práctico**:

Supongamos que queremos encontrar la probabilidad de que sea menor que 1.5

1. **Buscar el Valor de** :
   * Divide 1.5 en sus componentes: 1.5 se busca como 1.5 (donde 1.5 se considera directamente sin separar).
2. **Consultar la Tabla Normal**:
   * Busca en la fila correspondiente a 1.5. Si utilizamos una tabla que incluye directamente 1.5, encontramos que el valor es aproximadamente 0.9332.

**Interpretación**:

La probabilidad de que sea menor que 1.5 es 0.9332, o 93.32%. Esto significa que hay un 93.32% de probabilidad de que un valor aleatorio tomado de una distribución normal estándar sea menor que 1.5.

**Conclusión**

La distribución normal estándar es una herramienta fundamental en estadística para simplificar los cálculos de probabilidad y es utilizada ampliamente en diversos campos. La Tabla Normal permite encontrar rápidamente las probabilidades acumuladas asociadas a valores específicos de , facilitando el análisis e interpretación de datos distribuidos normalmente.

**Uso de las Tablas de Distribuciones Chi Cuadrado y T-Student**

**Distribución Chi Cuadrado**

**Definición**: La distribución chi cuadrado es una distribución de probabilidad continua que se utiliza principalmente en pruebas de hipótesis sobre la varianza de una población y en análisis de independencia en tablas de contingencia. Se define como la suma de los cuadrados de variables aleatorias independientes, cada una con una distribución normal estándar.

**Ejemplo Práctico**: Supongamos que queremos encontrar el valor crítico de la distribución chi cuadrado para un nivel de significancia (α) de 0.05 con 10 grados de libertad.

1. **Datos del problema**:
   * Nivel de significancia (α): 0.05
   * Grados de libertad (): 10
2. **Consulta en la Tabla Chi Cuadrado**:
   * Busca en la tabla de valores críticos de la distribución chi cuadrado el valor correspondiente a 0.05 (columna) y 10 grados de libertad (fila).
   * El valor crítico encontrado es aproximadamente 18.307.

**Interpretación**: El valor crítico chi cuadrado para un nivel de significancia de 0.05 y 10 grados de libertad es 18.307. Este valor se utiliza en pruebas de hipótesis para determinar si la varianza observada es significativamente diferente de la varianza esperada.

**Distribución T-Student**

**Definición**: La distribución T-Student es una distribución de probabilidad continua que se utiliza cuando se estima la media de una población pequeña y se desconoce la desviación estándar de la población. Es especialmente útil para muestras pequeñas

**Ejemplo Práctico**: Supongamos que queremos encontrar el valor crítico de la distribución T-Student para un nivel de significancia (α) de 0.05 con 15 grados de libertad.

1. **Datos del problema**:
   * Nivel de significancia (α): 0.05 (dos colas, por lo tanto para cada cola)
   * Grados de libertad (): 15
2. **Consulta en la Tabla T-Student**:
   * Busca en la tabla de valores críticos de la distribución T-Student el valor correspondiente a 0.025 (columna) y 15 grados de libertad (fila).
   * El valor crítico encontrado es aproximadamente 2.131.

**Interpretación**: El valor crítico T-Student para un nivel de significancia de 0.05 (dos colas) y 15 grados de libertad es 2.131. Este valor se utiliza en pruebas de hipótesis para determinar si la media de una muestra es significativamente diferente de una media hipotética.

**Ejemplos Detallados**

**Ejemplo con Distribución Chi Cuadrado**

**Situación**: Un investigador quiere determinar si la variabilidad en la duración de vida de un producto es diferente de la especificada por el fabricante. La varianza especificada es , y se toma una muestra de 25 productos con una varianza muestral = 5.

1. **Hipótesis**:
   * ​: = 4 (varianza poblacional especificada)
   * : ≠ 4 (varianza poblacional diferente)
2. **Estadístico de Prueba**:
3. **Valor Crítico**:
   * Grados de libertad (): 24
   * Nivel de significancia (α): 0.05 (dos colas, por lo tanto para cada cola)
   * Consulta en la tabla chi cuadrado: Los valores críticos para 24 grados de libertad y son aproximadamente 12.401 (izquierda) y 39.364 (derecha).
4. **Decisión**:
   * Como el valor calculado de está entre los valores críticos, no rechazamos ​. La variabilidad no es significativamente diferente de la especificada.

**Ejemplo con Distribución T-Student**

**Situación**: Un científico quiere determinar si el tiempo promedio de resolución de un problema es diferente de 30 minutos. Se toma una muestra de 16 estudiantes con una media muestral minutos y una desviación estándar muestral .

1. **Hipótesis**:
   * ​: (tiempo promedio especificado)
   * ​: (tiempo promedio diferente)
2. **Estadístico de Prueba**:
3. **Valor Crítico**:
   * Grados de libertad 15
   * Nivel de significancia 0.05 (dos colas, por lo tanto para cada cola)
   * Consulta en la tabla T-Student: El valor crítico para 15 grados de libertad y es aproximadamente 2.131.
4. **Decisión**:
   * Como el valor calculado de es menor que el valor crítico 2.131, no rechazamos ​:. No hay suficiente evidencia para afirmar que el tiempo promedio de resolución es diferente de 30 minutos.

**Conclusión**

La utilización de las tablas de distribución chi cuadrado y T-Student es esencial para llevar a cabo pruebas de hipótesis y análisis estadísticos en situaciones donde se desconoce la varianza poblacional o cuando se trabaja con muestras pequeñas. Estos ejemplos ilustran cómo aplicar estos conceptos de manera práctica y precisa, permitiendo una toma de decisiones informada en contextos de investigación y análisis de datos.