

UNIVERSIDAD SERGIO ARBOLEDA

Análisis de Datos
Nivel Integrador

PREPARACIÓN SEMANA 3

DISTRIBUCIONES DE PROBABILIDAD

Discreta

- Las variables discretas son aquellas que toman valores aislados y contables. Estos valores suelen ser números enteros.
- Ejemplos de variables discretas incluyen el número de clientes que ingresan a una tienda en un día o la cantidad de errores en un documento.

Continua:

- Las variables continuas pueden tomar cualquier valor dentro de un rango específico y generalmente se miden con precisión.
- Ejemplos de variables continuas incluyen la altura de las personas, la temperatura en grados Celsius, o el tiempo que tarda en completarse una tarea.

DISTRIBUCIONES DE PROBABILIDAD

Distribuciones Comunes: Binomial, Poisson y Exponencial:

1. Distribución Binomial:

- **Definición:** Modela el número de éxitos en una secuencia de ensayos independientes, donde cada ensayo tiene solo dos resultados posibles: éxito o fracaso.
- **Parámetros clave:** Número de ensayos (n) y probabilidad de éxito en un solo ensayo (p).
- **Ejemplo:** El número de caras obtenidas al lanzar un dado equilibrado varias veces.

2. Distribución de Poisson:

- **Definición:** Modela el número de eventos que ocurren en un intervalo de tiempo o espacio específico.
- **Parámetros clave:** Tasa de ocurrencia (λ).
- **Ejemplo:** El número de llamadas recibidas por una central telefónica en una hora.

DISTRIBUCIONES DE PROBABILIDAD

3. Distribución Exponencial:

- **Definición:** Modela el tiempo hasta que ocurra un evento, dado que los eventos ocurren de manera continua e independiente.
- **Parámetros clave:** Tasa de ocurrencia (λ).
- **Ejemplo:** El tiempo entre llegadas de clientes a un servicio de atención al cliente.

MUESTREO E INFERENCIA ESTADÍSTICA

Muestreo:

- **Definición:** El muestreo es el proceso de seleccionar un subconjunto representativo de una población más grande para realizar observaciones y análisis.
- **Objetivo:** Obtener información precisa y significativa sobre la población completa a través de la observación de una muestra más pequeña y manejable.
- **Métodos de Muestreo:** Pueden incluir muestreo aleatorio simple, estratificado, por conglomerados, entre otros.

Inferencia Estadística:

- **Definición:** La inferencia estadística es el proceso de hacer afirmaciones o tomar decisiones sobre una población basándose en los resultados observados en una muestra.
- **Objetivo:** Extender los resultados de la muestra a la población, permitiendo generalizaciones y toma de decisiones basadas en evidencia estadística.
- **Conceptos Clave:** Puntos estimados y intervalos de confianza.

MUESTREO E INFERENCIA ESTADÍSTICA

Intervalos de Confianza:

- **Definición:** Un intervalo de confianza es un rango estimado de valores que es probable que contenga el parámetro de la población de interés.
- **Confianza:** Se expresa como un porcentaje de certeza (por ejemplo, 95% de confianza) y cuantifica la certeza con la que se puede afirmar que el intervalo captura el parámetro.

MUESTREO E INFERENCIA ESTADÍSTICA

Proceso para Construir un Intervalo de Confianza:

1. **Recopilación de Datos:** Obtén una muestra representativa de la población.
2. **Estimación Puntual:** Calcula un punto estimado del parámetro de interés (puede ser la media, proporción, etc.).
3. **Cálculo del Error Estándar:** Determina la variabilidad de la estimación en función del tamaño de la muestra.
4. **Selección del Nivel de Confianza:** Decide el nivel de confianza deseado para el intervalo (por ejemplo, 90%, 95%, 99%).
5. **Cálculo del Intervalo:** Utiliza el error estándar y la distribución de probabilidad adecuada para calcular el rango del intervalo.
6. **Interpretación:** Afirma con cierto nivel de confianza que el parámetro de la población está contenido en el intervalo.

PRUEBAS DE HIPÓTESIS

- Una prueba de hipótesis es un procedimiento estadístico utilizado para tomar decisiones sobre una afirmación o hipótesis acerca de un parámetro poblacional.
- La hipótesis nula (H_0) es una afirmación que se asume como verdadera hasta que se presente evidencia en su contra. La hipótesis alternativa (H_1 o H_a) es la afirmación que se busca respaldar con evidencia.

Pasos Básicos:

1. Formulación de Hipótesis:

- H_0 : Hipótesis nula (afirmación a probar o refutar).
- H_1 : Hipótesis alternativa (afirmación opuesta a la nula).

2. Selección del Nivel de Significancia (α):

- Representa la probabilidad de cometer un error tipo I (rechazar H_0 cuando es verdadera).
- Comúnmente se utiliza $\alpha=0.05$ o $\alpha=0.01$, pero puede variar según el contexto.

PRUEBAS DE HIPÓTESIS

1. Elección de la Prueba Estadística:

- Selecciona la prueba estadística adecuada según el tipo de datos y la naturaleza de la hipótesis.

2. Recolección de Datos y Cálculo de la Estadística de Prueba:

- Recolecta datos de la muestra y calcula la estadística de prueba.
- La estadística de prueba es una medida que cuantifica la evidencia en contra de la hipótesis nula.

3. Decisión y Comparación con el Valor Crítico o el P-valor:

- Compara la estadística de prueba con un valor crítico o utiliza el p-valor.
- Si la estadística de prueba cae en la región crítica (o el p-valor es menor que α), se rechaza H_0 .

4. Interpretación y Conclusión:

- Concluye sobre la evidencia a favor o en contra de la hipótesis nula.
- Informa sobre cualquier conclusión significativa y relevante.

PRUEBAS DE HIPÓTESIS

Tipos de Errores: Tipo I y Tipo II:

- **Error Tipo I (α):**
 - Ocurre cuando se rechaza incorrectamente la hipótesis nula que es verdadera.
 - La probabilidad de cometer un error tipo I es igual al nivel de significancia (α).
- **Error Tipo II (β):**
 - Ocurre cuando no se rechaza la hipótesis nula que es falsa.
 - Relacionado con el poder de la prueba, que es la probabilidad de rechazar la hipótesis nula cuando es falsa.

APLICACIÓN EN PYTHON

Veamos los siguientes ejemplos aquí <https://acortar.link/dGOLx0>

¿Preguntas?