



INFORMÁTICA PARA ECONOMISTAS – FUNDAMENTOS BÁSICOS DE ESTADÍSTICA

MSc. Ciro Ivan Machacuay Meza

Estadística Inferencial

Definición: Rama de la estadística que usa muestras para sacar conclusiones sobre una población.

¿Para qué sirve?

- Predecir resultados (ejemplo: porcentaje de votos en elecciones).
- Tomar decisiones con incertidumbre (ejemplo: probar si un medicamento funciona).

Ejemplo:

- Población: Todos los estudiantes de una universidad (10,000).
- Muestra: Encuestas a 200 estudiantes.
- Inferencia: Si el 60% de la muestra aprueba un curso, se infiere que ~6,000 estudiantes en total lo aprobarían.

Probabilidad

Definición: Medida (entre 0 y 1) de qué tan posible es que ocurra un evento.

Tipos:

- Frecuentista: Basada en repeticiones (ej: lanzar una moneda 100 veces).
- Bayesiana: Incorpora creencias previas (ej: diagnóstico médico con historial del paciente).

Ejemplos:

- Probabilidad de sacar "cara" en una moneda justa: $P(\text{Cara})=0.5$
- Probabilidad de lluvia mañana: 30% (según datos históricos).



Variables Aleatorias

Definición: Variable que toma valores numéricos según el resultado de un fenómeno aleatorio.

Ejemplos:

- *Discreta: Número de fallas en una máquina en un mes ($X=0,1,2\dots$)*.
- *Continua: Tiempo de espera en un banco ($Y=2.3$ Y = 2.3 minutos).*

Relación con probabilidad:

- *Las variables aleatorias tienen distribuciones de probabilidad (ej: Binomial, Normal).*

Distribución de Bernoulli

¿Qué es?

Es la distribución más simple de probabilidad. Modela situaciones con **solo dos resultados posibles**:

- *Éxito (con probabilidad p). Ejemplo: "Sí", "Cara", "Gana".*
- *Fracaso (con probabilidad $1 - p$). Ejemplo: "No", "Cruz", "Pierde".*

Ejemplo Cotidiano:

Situación: Lanzar una moneda al aire (no necesariamente justa).

- *Éxito (1): "Cara" → Probabilidad $p = 0.5$ (si es justa).*
- *Fracaso (0): "Cruz" → Probabilidad $1 - p = 0.5$.*

Pregunta: Si la moneda está cargada y tiene 70% de probabilidad de salir "Cara", ¿cuál es la probabilidad de que salga "Cruz"?

Respuesta: 30% (porque $100\% - 70\% = 30\%$).

¿Para Qué Sirve?

1. *Tomar decisiones rápidas con dos opciones:*

- *¿Aprobaré el examen? (Sí/No).*
- *¿Lloverá hoy? (Sí/No).*

2. *Calcular riesgos básicos:*

- *Probabilidad de que una máquina falle (o no).*
- *Probabilidad de que un cliente compre (o no).*



3. Base para modelos más complejos:

- *La Distribución Binomial* (que cuenta éxitos en varios intentos de Bernoulli).
- *Por ejemplo: "Si lanzo la moneda 10 veces, ¿cuántas veces sale 'Cara'?"* (eso ya es Binomial).

4. En machine learning y estadística:

- *Clasificación binaria (spam/no spam, enfermo/sano).*

Distribución Binomial

¿Qué es?

Es como una extensión de la Distribución de Bernoulli, pero ahora repetimos el "juego de dos resultados" varias veces y contamos cuántos éxitos obtenemos en total.

Fórmula clave:

$$P(k \text{ éxitos en } n \text{ intentos}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Donde:

- $\binom{kn}{k}$ = combinaciones (formas de obtener "k" éxitos en "n" intentos).
- p = probabilidad de éxito en cada intento.

Ejemplo Cotidiano:

Situación: Lanzar una moneda 10 veces (¿cuántas "caras" salen?).

- **Éxito (1):** "Cara" → $p=0.5$
- **Fracaso (0):** "Cruz" → $1-p=0.5$
- **Número de intentos (n):** 10.

Pregunta: ¿Cuál es la probabilidad de obtener exactamente 7 caras en 10 lanzamientos?

Respuesta:

$$P(7) = \binom{10}{7} (0.5)^7 (0.5)^3 = 120 \times 0.0078 \times 0.125 \approx 11.7\%.$$



¿Para Qué Sirve?

1. Predecir resultados repetitivos:

- Ejemplo: Si un vendedor tiene un 30% de probabilidad de cerrar una venta, ¿cuántas ventas logrará en 20 intentos?

2. Control de calidad:

- En una fábrica, si 1% de productos son defectuosos, ¿cuántos defectos se esperan en 1000 unidades?

3. Experimentos médicos:

- Si una medicina cura al 60% de los pacientes, ¿cuál es la probabilidad de que 5 de 10 pacientes se curen?

4. Juegos y apuestas:

- Probabilidad de ganar 3 de 5 partidas de póker si tienes 40% de chances por partida.

Distribución de Poisson

¿Qué es?

La distribución de Poisson se usa para calcular la probabilidad de que ocurra **un número específico de eventos raros en un intervalo de tiempo o espacio**, cuando estos eventos pasan con una **frecuencia promedio conocida (λ)**.

Fórmula clave:

$$P(k \text{ eventos}) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Donde:

- λ = promedio de eventos en el intervalo (ejemplo: 3 fallas por mes).
- k = número de eventos que queremos predecir (ejemplo: ¿cuál es la probabilidad de que haya exactamente 2 fallas?).
- $e \approx 2.71828$ (constante matemática).

Ejemplo Cotidiano:

Situación: Un call center recibe **en promedio 5 llamadas por hora ($\lambda = 5$)**.

Pregunta: ¿Cuál es la probabilidad de que reciba **exactamente 3 llamadas en una hora**?

**Cálculo:**

$$P(3) = \frac{e^{-5} \cdot 5^3}{3!} = \frac{0.0067 \cdot 125}{6} \approx 13.9\%.$$

Interpretación: Hay un 13.9% de probabilidad de que lleguen exactamente 3 llamadas en una hora.

¿Para Qué Sirve?**1. Eventos de baja frecuencia:**

- Accidentes en una carretera por semana.
- Errores de impresión en un libro.

2. Gestión de recursos:

- ¿Cuántos cajeros se necesitan en un banco si llegan 10 clientes por hora?

3. Control de calidad:

- Número de defectos en un metro cuadrado de tela.

4. Ciencias naturales:

- Cantidad de estrellas en un sector del cielo.

Distribución Normal

¿Qué es?

La distribución normal (o campana de Gauss) es la **reina de las distribuciones** porque describe fenómenos naturales, sociales y físicos donde los datos se agrupan **alrededor de un valor promedio** y son menos frecuentes a medida que se alejan de él.

Fórmula clave:

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

- μ = media (centro de la campana).
- σ = desviación estándar (qué tan "ancha" es la campana).

**Ejemplo Cotidiano:**

Situación: Las estaturas de los adultos en una ciudad.

- **Promedio (μ):** 1.70 m.
- **Desviación estándar (σ):** 0.1 m.

¿Para Qué Sirve?**1. Ciencias sociales:**

- Puntajes de CI (coeficiente intelectual) (media = 100, $\sigma=15\sigma$).
- Calificaciones de exámenes.

2. Control de calidad:

- Peso de productos en una fábrica (ej: paquetes de arroz).

3. Finanzas:

- Rendimientos de acciones (a largo plazo).

4. Medicina:

- Presión arterial en adultos sanos.

Distribución Chi-Cuadrado (χ^2)

¿Qué es?

La distribución chi-cuadrado es una **herramienta estadística** que mide cómo se distribuyen las **diferencias al cuadrado** entre los valores observados y los esperados. Se usa principalmente en:

- **Pruebas de independencia** (tablas de contingencia).
- **Pruebas de bondad de ajuste** (¿mis datos siguen cierta distribución?).
- **Estimación de varianzas** (intervalos de confianza).

Fórmula clave (para curiosos):

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- O_i : Valor observado.
- E_i : Valor esperado (bajo una hipótesis).



Ejemplo Cotidiano:

Situación: Un profesor quiere saber si sus estudiantes **prefieren ciertos días para exámenes** (lunes, miércoles, viernes).

- **Resultados observados (O):** lunes: 10, miércoles: 25, viernes: 15.
- **Hipótesis nula (E):** No hay preferencia → esperaríamos 16.67 estudiantes por día (50/3).

Cálculo χ^2 :

$$\chi^2 = \frac{(10 - 16.67)^2}{16.67} + \frac{(25 - 16.67)^2}{16.67} + \frac{(15 - 16.67)^2}{16.67} = 6.8$$

Interpretación:

- Si $\chi^2 >$ valor crítico (de tablas χ^2 con 2 grados de libertad), rechazamos la hipótesis nula (hay preferencia).

¿Para Qué Sirve?**1. Pruebas de independencia:**

- ¿El color de pelo afecta la preferencia por un producto? (Tabla 2x2).

2. Bondad de ajuste:

- ¿Mis datos siguen una distribución normal? (Comparar frecuencias observadas vs. esperadas).

3. Intervalos de confianza para varianza:

- En control de calidad, estimar la variabilidad de un proceso.

Distribución t-Student

¿Qué es?

La distribución **t-Student** es una herramienta estadística para analizar datos cuando:

- La muestra es **pequeña** (típicamente < 30).
- La **desviación estándar poblacional (σ)** es desconocida.

¿Por qué existe?

Porque con muestras pequeñas, la distribución normal **no es precisa**, y la distribución t corrige este problema.



Ejemplo Cotidiano:

Situación: Un nutricionista quiere saber si su nueva dieta reduce el peso promedio.

- **Muestra:** 10 pacientes.
- **Datos:** Pérdida de peso promedio = 3 kg.
- **Desviación estándar muestral (s)** = 1.5 kg.

Pregunta: ¿La pérdida de peso es significativa (diferente de 0)?

Solución:

1. Calcula el estadístico t:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3 - 0}{1.5/\sqrt{10}} \approx 6.32$$

2. Compara con el valor crítico de la tabla t (para 9 grados de libertad).
3. **Conclusión:** Si $|t| >$ valor crítico, la dieta sí funciona.

¿Para Qué Sirve?

1. **Pruebas de hipótesis para medias** (muestras pequeñas).
 - Ejemplo: ¿El rendimiento de un nuevo fertilizante difiere del estándar?
2. **Intervalos de confianza para la media poblacional** (con σ desconocida).
 - Ejemplo: "El tiempo promedio de entrega está entre 20 y 30 mins (95% de confianza)".
3. **Comparación de dos medias** (muestras independientes o pareadas).

Teorema del Límite Central

¿Qué es?

El **Teorema del Límite Central (TLC)** es uno de los conceptos más importantes en estadística. Nos dice que:

"Si tomas muchas muestras grandes de una población (con cualquier distribución), las medias de esas muestras se distribuirán de forma aproximadamente **normal** (en forma de campana), incluso si la población original no es normal."



Ejemplo Cotidiano:

Situación: Quieres saber el **peso promedio de manzanas** en un huerto (la población tiene formas raras: algunas muy pesadas, otras livianas).

1. Toma 30 muestras de 50 manzanas cada una.
2. Calcula la media de cada muestra (ejemplo: 102g, 98g, 101g...).
3. Grafica esas medias: ¡Formarán una **campana de Gauss!** (aunque las manzanas individuales no lo hagan).

Conclusión:

- Aunque las manzanas individuales tengan pesos dispares, el **promedio de grupos** es predecible y normal.

¿Para Qué Sirve?

1. Inferencia estadística:

- Permite hacer **pruebas de hipótesis e intervalos de confianza** incluso si los datos originales no son normales.
- Ejemplo: "El salario promedio en la ciudad está entre \$1,000 y \$1,200 (95% de confianza)".

2. Control de calidad:

- Analizar promedios de producción (ej: grosor de láminas de metal).

3. Machine Learning:

- Muchos algoritmos asumen normalidad (gracias al TLC, esto a menudo es válido para promedios).

