ESTADISTICA 1 (REPASO)

DEFINICIONES

- **Experimento** = Procedimiento que genera datos
- **Experimento Aleatorio** = Genera distintos resultados así se repita con las mismas condiciones
- **Espacio Muestral** = (S), conjunto de todos los posibles resultados
- **Evento** = $(A_1, A_2, A_3 \ldots)$ Subconjunto del espacio muestral (S)
- **Evento nulo** = Subconjunto sin ningún elemento (Φ)

Definiendo (S) y los (A_1, A_2, A_3) se aplica toda la teoría de conjuntos

TIPOS DE EVENTOS (An)

1. Mutuamente excluyentes
   $A \cap B = \Phi$
2. Exhaustivos
   $A \cup B = S$

Unión = "O"
Intersección = "Y"

3. Unión de eventos
   $A \cup B = \{x \in S \,/\, x \in A \vee x \in B\}$
4. Intersección de eventos
   $A \cap B = \{x \in S \,/\, x \in A \wedge x \in B\}$
5. Complemento
   $\bar{A}, A', A^c$

PROPIEDADES DE LOS EVENTOS

1. $A \cup \Phi = A$
2. $A \cap \Phi = \Phi$
3. $A \cup A' = S$
4. $A \cap A' = \Phi$
5. $S' = \Phi$
6. $\Phi' = S$
7. $A'' = A$
8. $(A \cap B)' = A' \cup B'$
9. $(A \cup B)' = A' \cap B'$

PROBABILIDAD

$$P(A) = \frac{\text{\# de casos favorables}}{\text{\# de casos posibles}}$$

$$P(A) = \frac{\text{\# de elementos de A}}{\text{\# elementos de S}}$$

FUNCION DE PROBABILIDAD P(A) (AXIOMAS)

Una función $P: S \to \mathbb{R}$, será llamada $P(A)$ si cumple:

I) $P(A) \geq 0 \quad \forall A \subseteq S$

II) $P(S) = 1$

III) a) Eventos mutuamente excluyentes finitos

$$P\left(\bigcup_{i=1}^{K} A_i\right)$$

$$\sum_{i=1}^{K} P(A_i) = P(A_1) + P(A_2) + \ldots + P(A_K)$$

b) Eventos mutuamente excluyentes infinitos

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

$$\sum_{i=1}^{\infty} P(A_i)$$

PROPIEDADES P(A) (TEORAMA)

I) $P(\Phi) = 0$

II) $P(A') = 1 - P(A)$

III) $0 \leq P(A) \leq 1$

IV) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

V) Si $A \subseteq B$ entonces $P(A) \leq P(B)$

VI) $P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$

VII) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

PROBABILIDAD CONDICIONAL

Probabilidad condicional de A dado B se denota $P(A|B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \; ; \; P(B) > 0$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \; ; \; P(A) > 0$$

De lo anterior podemos deducir

$$P(A|B)(PB) = P(A \cap B)$$

$$P(B|A)\,P(A) = P(A \cap B)$$

TEOREMA DE PROBABILIDAD TOTAL

$A_1, A_2, A_3 \ldots A_N$

- Eventos disyuntos $A_1, A_2, \ldots, A_N$ (Mutuamente excluyentes)
- Exhaustivos $(A_1 \cup A_2 \cup A_3, \ldots = S)$

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i)$$

TEOREMA DE BAYES

$A_1, A_2, A_3, \ldots A_N$

- Eventos disyuntos
- Exhaustivos

$$P(A_i|B) = \frac{P(A_i)\,P(B|A_i)}{\sum_{i=1}^{K} P(A_i)P(B|A_i)}$$

INDEPENDENCIA ENTRE EVENTOS

Son eventos independientes si se cumple cualquiera de las siguientes proposiciones:

I) $P(A|B) = P(A)$
II) $P(B|A) = P(B)$
III) $P(A \cap B) = P(A) \cdot P(B)$

VARIABLES ALEATORIAS (V.A)

Es una funcion que asigna a cada posible resultado w en el espacio muestral S un unico numero real $X(w)$

V.A. DISCRETAS (V.A.D)

Conjunto de valores que toma la variable es finito o numerable (conteos)

1. Distribucion de Probabilidad o funcion de masa de probabilidad (f.m.p)

$$\sum P(x_i) = 1$$

$$P(x_1) + P(x_2) + P(x_3) + \ldots P(x_n) = 1$$

2. Funcion de distribucion acumulada (f.d.a)

$$F(x) = P(X \leq x) = \sum_{x' \leq x} P(x') \quad \forall x \in \mathbb{R}$$

- fda siempre posee saltos en los puntos correspondientes a valores de la V.A. y la magnitud del salto es igual a la probabilidad F.M.P en dicho punto

Proposicion

$$P(a \leq X \leq b) = F(b) - F(a^-)$$

$$F(a^-) = \lim_{x \to a^-} F(x)$$

V.A. CONTINUAS (V.A.C)

Es un intervalo o union de intervalos reales (mediciones)

1 Funcion de densidad de probabilidad (f.d.p)

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

- $\int_{-\infty}^{\infty} f(x)dx$ = Area total bajo la curva,
= 1
= Probabilidad total

2. Funcion de distribucion Acumulada

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

Proposiciones

- $P(X > a) = 1 - F(a)$
- $P(a \leq X \leq b) = F(b) - F(a)$
- $F'(x) = dF(x)/dx = f(x)$

■ Valor esperado de un V.A.

$$E[X] = \begin{cases} \sum_x x P(x) & \text{Discreta} \\ \int_{-\infty}^{\infty} x f(x) & \text{Continua} \end{cases}$$

Proposiciones

- $E[a] = a$
- $E[aX+b] = aE[X]+b$
- Si $g(x)$ es una funcion de X, entonces

$$E[g(x)] = \begin{cases} \sum_x g(x)\,P(x) & \text{V.A.D} \\ \int_{-\infty}^{\infty} g(x)\,f(x)\,dx & \text{V.A.C} \end{cases}$$

VARIANZA σ_x^2

$$Var[X] = E[X^2] - E[X]^2$$

- $Var[a] = 0$
- $Var[aX+b] = a^2\,Var[X]$

DESVIACION ESTANDAR σ

Es la raiz cuadrada de la varianza

$$\sigma_x = \sqrt{Var[X]}$$

TECNICAS DE CONTEO

↳ conteo de resultados

1. Regla multiplicativa

Secuencia de pasos de un mismo proceso

En una manufactura se produce una pieza con 3 operaciones cada operacion tiene ciertos herramientas

Maquinado = 3 herramientas
Pulido = 4 herramientas
Pintado = 3 herramientas

¿Cuantas rutas distintas son posibles para hacer una pieza?

$3 \times 4 \times 3 = 36$

2. Permutaciones

↳ Variacion del orden de los elementos

$${}_nP_r$$

Proposiciones

- Para todos los elementos

$${}_nP_n = n!$$

- Con condiciones

$${}_nP_r = \frac{n!}{(n-r)!}$$

- Cuando se permiten repeticiones

$${}_nP_r = n^r$$

- Si cada elemento (evento) es de un tipo distinto

$$P^n_{n_1 \ldots} = \frac{n!}{n_1!\,n_2!\,n_3!\ldots}$$

Juan, Carlos, Ana y Milena esperan el la parada del bus. ¿de cuantas maneras se pueden filar? Si solo hay dos puestos ¿de cuantas maneras se pueden organizar en los dos puestos?

- $4! = 24$
- ${}_4P_2 = 4!/(4-2)! = 12$

3. Combinacion

↳ No se tiene en cuenta el orden

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

ENSAYO BERNOULLI VAD

$$X \sim Ber(P)$$

P: Probabilidad del evento

El ensayo de bernoulli se caracteriza por tener dos resultados

Exito = P
Fracaso = $1-P$

DISTRIBUCION BINOMIAL VAD

$X \sim bin(n, p)$

n = # ensayos
p = Propabilidad de cada ensayo

$$\binom{n}{x} p^{(x)} (1-p)^{n-x}$$

Recordemos que para que un experimento sea llamado binomial debe cumplir

1. Consta de n pruebas identicas e independientes
2. Cada prueba tiene dos posibles resultados Exito o fracaso
3. La probabilidad de Exito es constante en las n pruebas y se denota p
4. La VA de interes es X = # exitos en los n ensayos

Media

$$E[X] = n \cdot p$$

Varianza

$$Var[X] = n \cdot p \cdot (1-p)$$

DISTRIBUCION HIPERGEOMETRICA VAD

[redacted]

$X \sim hiper(n, M, N)$

n = Tamaño de la muestra
M = Exitos dentro de la muestra
N = Poblacion total

$$P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\text{casos favorables}}{\text{casos posibles}}$$

Recordemos las condiciones de una distribucion hipergeometrica:

1. Poblacion de N elementos
2. Hay M exitos y N-M fallos
3. Cualquier muestra de n elementos es seleccionada al azar, sin remplazo y con p constante

Media

$$E[X] = n \cdot \frac{M}{N}$$

$$\boxed{p = \frac{M}{N} = \text{Proporcion de exito}}$$

$$E[X] = n \cdot p$$

Varianza

$$Var[X] = \left(\frac{N-n}{N-1}\right) n \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

$$Var[X] = \left(\frac{N-n}{N-1}\right) \cdot E[X] \cdot \left(1 - \frac{M}{N}\right)$$

$$\boxed{q = 1 - p = \text{Proporcion de fracaso}}$$

$$Var[X] = \left(\frac{N-n}{N-1}\right) \cdot E[X] \cdot q$$

APROXIMACION BINOMIAL DE LA HIPERGEOMETRICA

Si $\left(\frac{N-n}{N-1}\right) \approx 1$ entonces

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^{(x)} (1-p)^{n-x}$$

Donde $p = \frac{M}{N}$ = Proporcion de exito

De la misma manera se puede hacer aproximacion binomial cuando

$$n < 5\% \cdot N$$

DISTRIBUCION DE POISSON VAD

$X \sim P_{ois}(\lambda)$ $\quad \frac{e^{-\lambda}\lambda^x}{x!}$

λ = # promedio de ocurrencia por unidad de tiempo o espacio

$$E[X] = \lambda = Var[X]$$

La f.m.p. = $f(a) = P(X=a)$
La f.d.a. = $F(a) = P(X \leq a)$

Recordemos que para que un experimento tenga distribucion de poisson debe cumplir

- X = # Ocurrencias en el intervalo real
- Resultados en intervalos reales: tiempo, espacio
- En cada unidad establecida, el número de eventos que ocurren es independiente de los que ocurren en otras unidades
- La probabilidad en cada subintervalo es despreciable
- La probabilidad es proporcional al intervalo.

APROXIMACION POISSON DE LA BINOMIAL

Sea $X \sim bin(n, p)$ y

$n \geq 100$
$p < 0{,}10$
$np \leq 10$

Entonces

$$\binom{n}{x} p^{(x)} (1-p)^{n-x} \approx \frac{e^{-\lambda}\lambda^x}{x!}$$

Donde $\lambda = n \cdot p$

DISTRIBUCION UNIFORME VAC

$$X \sim U(a,b)$$

(a,b): Intervalo

Funcion de densidad de probabilidad (f.d.p)

$$f(x) = \begin{cases} \frac{1}{b-a} & ; \ a < x < b \\ 0 & ; \ \text{en otros casos} \end{cases}$$

$f(a)b) = P(a < x \, b)$
cuando X pertenezca a cualquier subintervalo de la longitud total

Funcion de distribucion acumulada F.d.a

$$F(x) = \begin{cases} 0 & ; \ \text{si } x < a \\ \frac{x-a}{b-a} & ; \ a \leq x \leq b \\ 1 & ; \ x > b \end{cases}$$

$$\int_{-\infty}^{x} f(u)\,du$$

$E[X] = \frac{a+b}{2}$ $V[X] = \frac{(b-a)^2}{12}$

Recordemos que

$V[X] = E[X^2] - E[X]^2$

DISTRIBUCION NORMAL VAC

$X \sim n(\mu, \sigma^2)$

μ: Media, Mediana, Moda = Localizacion
σ^2: Varianza = Escala
σ = Desviacion estandar

F.d.P.

$$f(x) = \frac{e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}}{\sqrt{2\pi}\,\sigma} \; ; \quad \begin{matrix} -\infty < x < \infty \\ \mu \in R \\ \sigma > 0 \end{matrix}$$

Recordemos que una distribucion normal cumple:

- La distribucion normal tiene forma de campana.
- El area de la curva entre $(-\infty, \infty)$ es igual a 1.
- La normal es simetrica respecto a μ, es decir que el area bajo la curva en $[\mu, \infty)$ es igual a 1/2

- Distribucion Normal Estandar (Estandarizacion)

$$\begin{matrix}\mu = 0 \\ \sigma = 1\end{matrix} \Rightarrow \begin{matrix} Z \sim n(0,1) \\ X \sim N(0,1)\end{matrix}$$

F.d.a

(Dada en general para la distribucion Normal a partir de la estandar)

A partir de un cambio de Variable donde

$Z = \frac{X - \mu}{\sigma^2}$; Donde σ^2 es igual a σ ya que son = 1

$dz = \frac{1}{\sigma}$

Tenemos que:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}}{\sqrt{2\pi}\,\sigma}\, dx$$

$$P(z_1 < Z < z_2) = \int_{z_1}^{z_2} \frac{e^{-(z^2/2)}}{\sqrt{2\pi}}\, dz$$

Donde:

$z_1 = \frac{x_1 - \mu}{\sigma}$ $z_2 = \frac{x_2 - \mu}{\sigma}$

Concluimos que:

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2)$$

Donde:

$$\Phi(z) = P(Z < z) = \int_{-\infty}^{z} \frac{e^{-(z^2/2)}}{\sqrt{2\pi}}\, dz$$

Entonces

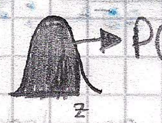
Teorema fundamental del calculo

$$P(z_1 < Z < z_2) = \Phi(z_2) - \Phi(z_1)$$

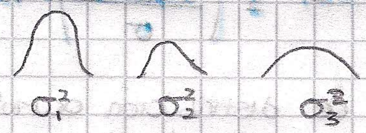
Sabiendo que

$$\Phi(z) = \begin{cases} \frac{1}{2} + \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{\sqrt{2\pi}\,(2n+1)\,2^n\,n!} \; ; & z \geq 0 \\ \frac{1}{2} - \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{\sqrt{2\pi}\,(2n+1)\,2^n\,n!} \; ; & z < 0 \end{cases}$$

Para la fda encontramos tablas que resumen todo el procedimiento que nos permite hallar las areas que necesitamos



Nota$_1$: Cuando σ^2 se hace mas grande la curva se achata



$\sigma_3^2 > \sigma_2^2 > \sigma_1^2$

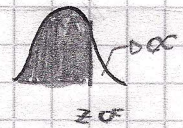
Nota$_2$ = $P(Z > z_\alpha) = \alpha$

Para uso practico de la tabla de probabilidades es bueno tener en cuenta

- Si $a<0$, $\Phi(a) = 1 - \Phi(-a)$ Por simetria
- $Z = \dfrac{X-\mu}{\sigma}$ (Estandarizacion)
- Cuando tenemos una distribucion Normal no estandarizada

* $P(a \leq X \leq b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)$

* $P(X \leq b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right)$



$z_\alpha = 100(1-\alpha)\%$
= percentil superior

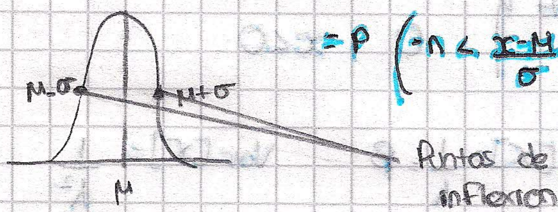
Area = $\Phi(n) = P$

* [redacted] $\sigma \cdot n_{est} + \mu = n_{\mu,\sigma}$

n desviaciones estandar

$$P(|x-\mu| \leq n\sigma) = P\left(\left|\frac{x-\mu}{\sigma}\right| \leq n\right)$$

$$= P\left(-n < \frac{x-\mu}{\sigma} < n\right)$$



APROXIMACION NORMAL DE LA BINOMIAL

Sea $X \sim bin(n,p)$

$np \geq 10$
$n(1-p) \geq 10$

Entonces

$$P(X < x) = P(X < x - 1/2)$$

$$\approx P\left(Z \leq \frac{x - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

Donde

$1/2$ = Factor de correccion
$n \cdot p = E[X]$
$n \cdot p(1-p) = Var[X]$

DISTRIBUCION EXPONENCIAL

$X \sim Exp(\lambda)$

Es un caso especial de la distribucion gamma con $K=1$

$\lambda > 0$ $\quad B = 1/\lambda$

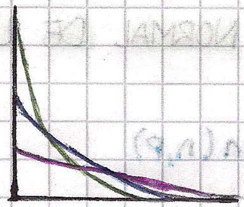
Funcion de densidad de probabilidad (f.d.p)

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x>0 \\ 0, & \text{otro caso} \end{cases}$$

Funcion de distribucion acumulada (f.d.a)

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$E[X] = \frac{1}{\lambda} = \beta$ Var $[X] = \frac{1}{\lambda^2}$


$\lambda_1 > \lambda_2 > \lambda_3$

RELACION DISTRIBUCION EXP Y PROCESO DE POISSON

Sea $X \sim Pois(\lambda^{\circ})$ y

T = Variable aleatoria que representa el tiempo (t) entre ocurrencias de proceso Poisson

Entonces

$T \sim Exp(\lambda)$

Donde

λ° = # Promedio de ocurrencias por unidad de tiempo ($t\lambda$)

- Propiedad de carencia de memoria

$P(X < t_0 + t_1 \mid X \geq t_0) = P(X < t_1)$

\# como nuevo

DISTRIBUCION LOGNORMAL VAC

Si la variable aleatoria $Y = Ln(X)$ tiene una distribucion normal con media μ y desviacion estandar σ entonces X tiene una distribucion Lognormal

Funcion de densidad de probabilidad (F.d.P)

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[Ln(x)-\mu]^2}{2\sigma^2}} ; x > 0$$

Funcion de distribucion acumulada (F.d.a)

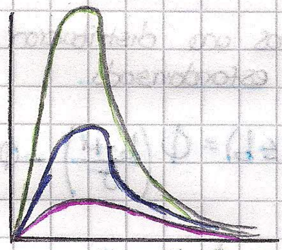
$$F(x) = P(X \leq x) = \Phi\left(\frac{Ln(x) - \mu}{\sigma}\right) ; \forall x > 0$$

La funcion de distribucion acumulada (Φ) Proviene de la distribucion normal estandar (Tabla)

$E[X] = e^{\mu + \frac{\sigma^2}{2}}$; Var $[X] = e^{2\mu + \sigma^2} * (e^{\sigma^2} - 1)$

Donde μ y σ^2 son la media y la Varianza de $Ln(x)$

La mediana es el percentil 50 y se relaciona con la media con $\Phi(0) = 0,5$; Por lo que $Ln\,\tilde{\mu} = 0$ y Mediana $(\tilde{\mu}) = e^{\mu}$


$\sigma_1^2 > \sigma_2^2 > \sigma_3^2$

Exp: tiempo

Poison: Eventos

Binomial = Exito - Fracaso

DISTRIBUCIONES BIVARIADAS

Involucra dos variables aleatorias X y Y

VARIABLES ALEATORIAS DISCRETAS (VAD)

Funcion de masa de probabilidad (f.m.p) conjunta

$$P(x,y) = P(X=x, Y=y), \forall (x,y) \in A \subseteq \mathbb{R}^2$$

- Propiedades

1. $P(x,y) \geq 0 \ \forall (x,y) \in A$
2. $\sum_{x} \sum_{y} P(x,y) = \sum \sum_{(x,y) \in A} P(x,y) = 1$
3. Si $A \subseteq \mathbb{R}^2 \Rightarrow P((X,Y) \in A) = \sum \sum_{(x,y) \in A} P(x,y)$

Recordar

$$P(X \cap Y) = P(X|Y)\,P(Y)$$

$$P(Y \cap X) = P(Y|X)\,P(X)$$

$Y \backslash X$	x_1	x_2	x_3	x_4	..
y_1					
y_2					
y_3					
y_4					
$\vdots$					

Dependiendo de las probabilidades se puede presentar simetria

VARIABLES ALEATORIAS CONTINUAS (VAC)

Funcion de distribucion acumulada (F.d.a) Conjunta

$$F(x,y) = P(X \leq x, Y \leq y);\ \forall (x,y) \in \mathbb{R}^2$$

F es continua en $\mathbb{R}^2$, si existe una función $f: \mathbb{R}^2 \to \mathbb{R}^+ \cup \{0\}$ tal que:

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$$

$\forall (x,y)$ donde exista $\frac{\partial^2 F}{\partial x \partial y}$ entonces

La funcion de densidad de probabilidad (f.d.p) conjunta es f

$$f(x,y)$$

- Propiedades

1. $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t,s)\, ds\, dt$
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\, dx\, dy = 1$
3. Si $B \subseteq \mathbb{R}^2 \Rightarrow$

$$P((x,y) \in B) = \iint_{B} f(x,y)\, dx\, dy$$

DISTRIBUCIONES MARGINALES Y CONDICIONALES

VARIABLES ALEATORIAS DISCRETAS (VAD)

- *Distribucion Marginal*

$$P_X(x) = \sum_{y} P(x,y)$$

$$P_Y(y) = \sum_{x} P(x,y)$$

$Y \backslash X$	x_1	x_2 .. x_i	$\sum$
y_1			$* = P_Y(y)$
y_2			
$\vdots$			
y_j			
$\sum$	$* = P_X(x_1)$		

• Distribucion Condicional "Y dado $X=x$"

$$P_{Y|X}(y) = \frac{P(x,y)}{P_X(x)} \text{, si } P_X(x)>0$$

Sea $g(X,Y)$ una funcion de X e Y $= x \cdot y$. $g: A \subseteq \mathbb{R}^2 \to \mathbb{R}$ Entonces

$$E[g(x,y)] = \sum_x \sum_y g(x,y)\, p(x,y)$$

Si $g(X,Y)=X$ se obtiene $E[X]$

Si $g(X,Y)=Y$ se obtiene $E[Y]$

Si $g(X,Y)=(X-\mu_x)^2$ se obtiene $Var = \sigma_x^2$

Sea X e Y variables aleatorias diremos que son Estadisticamente independientes (EI), si:

$$P(x,y) = P_X(x) \cdot P_Y(y)$$

VARIABLES ALEATORIAS CONTINUAS (VAC)

• Distribucion Marginal

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\, dx$$

• Distribucion Condicional de "Y dado $X=x$"

$$f_{Y|X}(y) = \frac{f(x,y)}{f_X(x)} \text{, si } f_X(x)>0$$

Por deduccion

$$f(x,y) = f_X(x) \cdot f_{Y|X}(y) = f_Y(y)\, f_{X|Y}(x)$$

Sean X e Y variables aleatorias diremos que son estadisticamente independientes (EI) si:

$$f(x,y) = f_X(x) \cdot f_Y(y), \ \forall (x,y) \in A$$

Si existe un par (x,y), para el cual la igualdad no es cierta, diremos que X e Y son estadisticamente Dependientes (ED)

Sea $g(X,Y)$ una funcion de X e Y $= X \cdot Y$. $g: A \subseteq \mathbb{R}^2 \to \mathbb{R}$ entonces

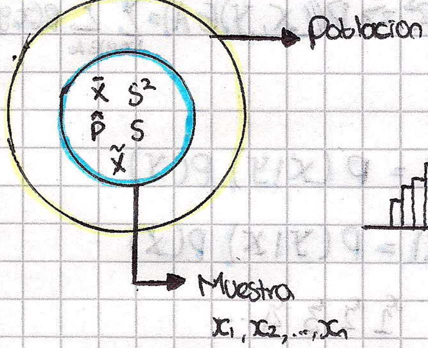
$$E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f(x,y)\, dx\, dy$$

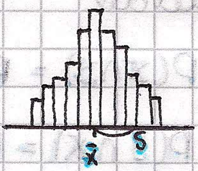
Si $g(X,Y)=X$ se obtiene $E[X]$

Si $g(X,Y)=Y$ se obtiene $E[Y]$

Si $g(X,Y)=(X-\mu_x)^2$ se obtiene $Var = \sigma_x^2$

ESTADISTICOS Y SUS DISTRIBUCIONES





Muestra — Poblacion

estimacion

$\bar{x} \longrightarrow \mu$

$S^2 \longrightarrow \sigma^2$

$S \longrightarrow \sigma$

$\hat{p} \longrightarrow P$

Estadistico — Parametro

En general estimamos 3 parametros dependiendo del caso

- Una Proporcion (P)
- Un Promedio (μ)
- Una Varianza (σ^2)

Un estadistico es la estimacion de un parametro a partir de una muestra

- *Muestra Aleatoria (M.a)*

Una M.a. de tamaño n, es un conjunto de VA EI e identicamente distribuidas. Si $X_1, \dots, X_n$ es una M.a. entonces

I. $f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

II. $f_{X_i}(x_i) = f(x_i) \; ; \; \forall i = 1, \dots, n$

- *Estimaciones* (Estimadores insesgados)

$\mu \Rightarrow \bar{x} = \frac{1}{n} \sum x_i$

$\sigma^2 \Rightarrow S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

$p \Rightarrow \frac{X}{n}$, $X \sim \text{bin}(n, p)$

DISTRIBUCION DE LA MEDIA MUESTRAL

Sea $X_1, \dots, X_n$ una muestra aleatoria (m.a.) de una distribucion con media μ y varianza σ^2 y

Sea $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Entonces

$$E[\bar{X}] = E \; \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} E[X_i]$$

$$E[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$$

$$E[\bar{X}] = \mu$$

$$Var[\bar{X}] = Var\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \sum_{i=1}^{n} Var[X_i]$$

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Siendo $T_0 = \sum_{i=1}^{n} X_i$. Entonces

$E[T_0] = n\mu$; $Var[T_0] = n\sigma^2$

T_0: Total de la muestra

Proposicion 1

Sean X y Y VA EI con $\sim N(\mu, \sigma^2)$ cada una con sus parametros

$$X \sim N(\mu_X, \sigma_X^2)$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

y sea Z VA

$$Z = aX \pm bY$$

Entonces

$$\mu_Z = a\mu_X \pm b\mu_Y$$

$$\sigma_Z^2 = a^2 \sigma_X^2 + b^2 \sigma_Y^2$$

Proposicion 2

$X_1, X_2, \dots, X_n$ una muestra aleatoria de una $\sim n(\mu, \sigma^2)$ Entonces

$$\bar{X} \sim n\left(\mu, \frac{\sigma^2}{n}\right) \quad y$$

$$T_0 \sim n(n\mu, n\sigma^2)$$

Entonces estandarizamos

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim n(0,1) \quad y \quad \frac{T_0 - n\mu}{\sigma\sqrt{n}} \sim n(0,1)$$

TEOREMA DEL LIMITE CENTRAL (TLC)

Suponga que $X_1, X_2, \dots, X_n$ es una M.a. de una poblacion con cualquier distribucion, media μ y varianza σ^2. Sea $\bar{X}$ la media muestral (que depende de n) entonces cuando $n \to +\infty$, la distribucion muestral de

$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ es aproximadamente normal estandar. Escribimos

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow[n\to+\infty]{\text{Aprox}} n(0,1)$$

De igual manera con T_0

$$\frac{T_0-n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\text{Aprox}} n(0,1)$$

- Entre mayor sea n, mejor es la aproximacion
- Regla empirica: Si $n>30$ se puede utilizar (T.L.C)
- Si la distribucion de la muestra es simetrica y continua, los tamaños muestrales relativamente pequeños permiten obtener buenas aproximaciones.
- Si la distribucion es discreta, se requiere de tamaños muestrales grandes

Asi

$$P(\bar{X}\le a)=P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{a-\mu}{\sigma/\sqrt{n}}\right)$$

$$\approx P\left(Z \le \frac{a-\mu}{\sigma/\sqrt{n}}\right)$$

Si σ^2 es desconocida y n grande se puede remplazar σ^2 por S^2. Así:

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow[n\to+\infty]{\text{Aprox}} n(0,1)$$

$$\frac{T_0-n\mu}{S\sqrt{n}} \xrightarrow[n\to+\infty]{\text{Aprox}} n(0,1)$$

ESTIMACIÓN PUNTUAL

Parametro	Estimador
θ	$\hat{\theta}$
- Media	- Media recortada
- Mediana	- Media Min-Max

Hay dos tipos de estimadores en las m.a.

1. Puntuales (unico valor)
2. Por intervalos

• Propiedades deseables de un estimador

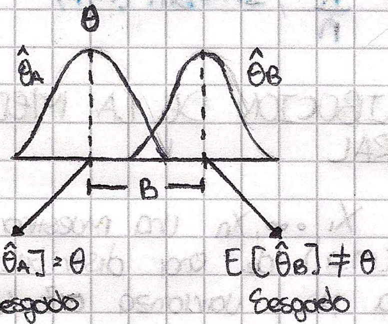
Un estimador puntual $\hat{\theta}$ de θ es INSESGADO si

$$E[\hat{\theta}]=\theta \quad ; \quad B=0$$

Si $\hat{\theta}$ es sesgado, se define como:

$$B=E[\hat{\theta}]-\theta$$

B= Sesgo

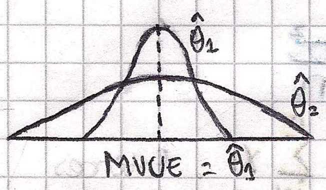


- Entre dos estimadores insesgados para θ, se prefiere aquel con menor varianza

$$Var[\hat{\theta}_1] < Var[\hat{\theta}_2]$$

Entonces $\hat{\theta}_1$ es mejor estimador de θ que $\hat{\theta}_2$.

El mejor estimador ($\hat{\theta}_1$) insesgado que tiene menor variaza es llamado: Estimador Insesgado de Minima Varianza (MVUE)



$\underset{A}{\int}$ = Integral Doble

- Error Cuadrado Medio (ECM)

Medida para comprobar el mejor estimador, se elige aquel con menor ECM

$$ECM[\hat{\theta}] = E[(\hat{\theta}-\theta)^2] = Var[\hat{\theta}] + B^2$$

↳ cuadrado de los errores

Para los estimadores insesgados $B=0$, por lo cual el mejor estimador sera el de menor varianza

• Teorema

Sea $X_1, \dots, X_n$ una muestra aleatoria de una $n \sim (\mu, \sigma^2)$. Entonces, $\bar{X}$ es el MVUE para μ.

• Error Estandar

Si $\hat{\theta}$ es un estimador de θ, el error estandar sera su desviacion estandar

$$\sqrt{Var[\hat{\theta}]}$$

Para una M.A de una poblacion con media μ e varianza σ^2, se tiene que un estimador insesgado para μ es $\bar{X}$, cuya varianza es σ^2/n. Asi el error estandar de $\bar{X}$ esta dado por

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

Si σ se desconoce se puede usar la desviacion estandar muestral

$$S_{\bar{x}} = S/\sqrt{n}$$

Recordemos

- $E[a] = a$
- $E[ax+b] = aE[X]+b$
- $Var[ax+b] = a^2 var[X]$
- $Var[X] = E[X^2] - (E[X])^2$

Para distribuciones bivariadas en VAD usamos sumatorias y en VAC integrales.

$$E[X] = \underset{A}{\int} x \, f(x,y) \, dA$$

$$E[X^2] = \underset{A}{\int} x^2 \, f(x,y) \, dA$$

$$E[XY] = \underset{A}{\int} xy \, f(x,y) \, dA$$

- Covarianza

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

$* \; Cov[aX+b, cY+d] = ac \; cov[X,Y]$

- Correlacion

$$Corr[X,Y] = \frac{Cov[X,Y]}{\sigma_x \cdot \sigma_y}$$

$* \; Corr[aX+b, cY+d] = Corr[X,Y]$

Si dos muestras estan relacionadas $Cov[X,Y] \neq 0$ entonces

$$Var[aX+bY] = a^2 var[X] + b^2 var[X] + 2ab \; cov[X,Y]$$

Si $cov[X,Y]=0$ entonces

$$Var[aX+bY] = a^2 var[X] + b^2 var[Y]$$

* Las muestras aleatorias nunca estan relacionadas, $Cov[X,Y]=0$

MAXIMA VEROSIMILITUD

Es un metodo para estimar parametros desconocidos, sean un # finito de parametros.

Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribucion f que depende de un(os) parametro(s) θ, Δ. La funcion de verosimilitud se define como

$$L(\theta,\Delta) = \prod_{i=1}^{n} f(x_i;\theta,\Delta)$$

EMV = Estimador de maximo verosimilitud de θ o Δ

EMV, es aquel que maximiza a $L(\theta,\Delta)$; en general se maximiza con mayor facilidad el logaritmo natural de $L(\theta)$.

• ¿Como se calculan los EMV?

1. Hallamos la funcion de verosimilitud $L(\theta,\Delta)$
2. Hallamos el $Ln[L(\theta,\Delta)] = \ell(\theta,\Delta)$
3. Hallamos las derivadas parciales respecto a los parametros (ℓ_θ) y (ℓ_Δ) e igualamos a cero

$\ell_\theta =$ EMV de θ

$\ell_\Delta =$ EMV de Δ.

Nota = * Los EMV no son siempre insesgados

* Los EMV no tiene siempre varianza minima

Para verificar que efectivamente son EMV utilizamos la segunda derivada

$$\begin{bmatrix} \ell_{\theta\theta} & \ell_{\Delta\theta} \\ \ell_{\theta\Delta} & \ell_{\Delta\Delta} \end{bmatrix}$$

Si la matriz es negativa se cumple que ℓ_θ y ℓ_Δ son EMV

[redacted]

Para un solo parametro verificamos que la segunda derivada tenga signo negativo (-)

En los EMV si n es grande es aproximadamente insesgado y MVUE.

- Principio de Invarianza

Si $\hat{\theta}$ es EMV de θ, entonces $h(\hat{\theta})$ es el estimador máximo verosimil de $h(\theta)$

Sea $f(\theta) = \frac{x}{\theta} e^{-x^2/2\theta}$

y sea $\hat{\theta}_{EMV} = \frac{\sum x_i^2}{2n}$

Entonces si tenemos que la mediana de $f(\theta) \Rightarrow \sqrt{2Ln(2)\cdot\theta}$

Entonces un EMU de la mediana sera:

$$\sqrt{2Ln(2)\cdot\hat{\theta}} = \sqrt{2Ln(2)\cdot\frac{\sum x_i^2}{n}}$$

- EMV para la varianza de una distribucion normal

$$S = \sqrt{\frac{\sum (x_i-\bar{x})^2}{n}}$$

Por principio de invarianza

$$S^2 = \frac{\sum (x_i-\bar{x})^2}{n}$$

Nota = Recordando que $\hat{\sigma}^2$ es un estimador insesgado de σ^2

$$\hat{\sigma}^2 = \frac{\sum (x_i-\bar{x})^2}{n-1}$$

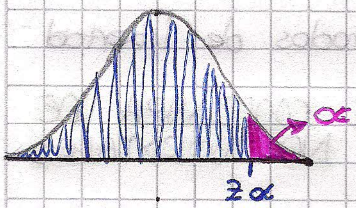
Entonces

$$\frac{\hat{\sigma}^2}{S^2} = \frac{n}{n-1}$$ cuando $n \to \infty$ este coeficiente se hace 1

↳ tiene menor varianza que $\hat{\sigma}^2$

Intervalo de confianza
"Donde esta el parametro"
Intervalo de prediccion
"Valor de un nuevo evento"
Limite de tolerancia
"Donde estan la mayoria de datos"

INTERVALOS DE CONFIANZA



$\Phi(z_\alpha) = 1-\alpha$

α: Nivel de significancia

Niveles de confianza
$(1-\alpha)100\%$
95%, 98%, 99%

- Intervalos Unilaterales Aleatorios

- $P(L < z) = 1-\alpha$
- $P(z < U) = 1-\alpha$

Para una muestra particular tenemos

- $(l, +\infty)$
- $(-\infty, u)$

En los intervalos unilaterales no tenemos precisión

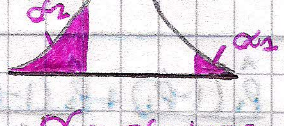
- Intervalos Bilaterales Aleatorios

$P(L < \theta < U) = 1-\alpha$

↳ intervalo (L, U)

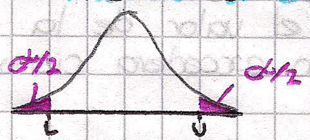
↳ Probabilidad de cobertura

* Intervalo Asimetrico



$\alpha = \alpha_1 + \alpha_2$

* Intervalo Simetrico



$\alpha = \frac{\alpha}{2} + \frac{\alpha}{2}$

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \leq z_{\alpha/2}\right)$$

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right)$$

$$P\left(-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$P\left(\bar{X}-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Intervalo

$$\left(\bar{X}-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X}+z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

- Interpretacion

* Con un nivel de confianza $100(1-\alpha)\%$ el verdadero valor del parametro estara entre L y U

* De 100 intervalos de confianza, el 95% de ellos tendra el valor del parametro

- Precision

Longitud = Limite superior - Limite inferior
$= U - L$

Teorema

$$n \geq \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2$$

A mayor confiabilidad menor precision

* Tamaño de muestra que permite estimar μ

* Maximo error que que se comete al estimar a μ con un nivel de confianza $100(1-\alpha)\%$

* Con un nivel de confianza de $100(1-\alpha)\%$ el valor de la precision no excederá cierta cota E

$$S^2 = \sum \frac{x_i^2 - n\bar{x}^2}{n-1}$$

Para calculadora

Boostrap

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{\text{Aprox}}{\sim} N(0,1)$$

$$(\hat{\theta} - P_{97.5},\ \hat{\theta} - P_{2.5})$$

- Se generan M muestras de tamaño n
- Con cada M se calcula $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_M$
- $$\hat{\hat{\theta}} = \frac{1}{M}\sum^{M} \hat{\theta}_i$$
- Calcule $\hat{\theta}_1 - \hat{\hat{\theta}}, \ldots, \hat{\theta}_M - \hat{\hat{\theta}}$
- Halle el percentil 97,5 y 2,5 de los M valores

INTERVALO DE CONFIANZA PARA UNA PROPORCION P DE UNA POBLACION

Si $X \sim Bin(n, p)$
$n \geq 30$

tenemos que

$$\hat{P} \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

$$\hat{P} = \frac{X_{obs}}{n}$$

Ademas para la precision tenemos

$$n \geq \frac{1}{4} \cdot \left(\frac{Z_{\alpha/2}}{E}\right)^2$$

INTERVALOS DE CONFIANZA DE POBLACIONES NORMALES

- Si la varianza σ^2 es conocida

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Si la varianza σ^2 es desconocida

$$\bar{X} \pm t_{\alpha/2, v} \cdot \frac{S}{\sqrt{n}}$$

Donde $v = n-1$
→ Grados de libertad

INTERVALOS DE CONFIANZA DE POBLACIONES NO NORMALES

Se debe cumplir la regla empírica. $n \geq 30$

- Si la varianza σ^2 es conocida

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$ Aproximado por TLC

- Si la varianza σ^2 es desconocida

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$ Aproximado por TLC

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE 2 PROPORCIONES $\hat{P}_1 - \hat{P}_2$ DE DOS POBLACIONES INDEPENDIENTES

Si $X_1 \sim Bin(n, p)$ con $n_1 \geq 30$
y $X_2 \sim Bin(n, p)$ con $n_2 \geq 30$

$$\hat{P} = \frac{X_{obs}}{n}$$

Entonces:

$$\hat{P}_1 - \hat{P}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$$

Si el intervalo contiene al cero se dice que P_1 y P_2 no son significativamente distintos. Es decir $P_1 \approx P_2$

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE DOS MEDIAS $\mu_1 - \mu_2$ DE DOS POBLACIONES NORMALES

- Si σ_1^2 y σ_2^2 Conocidas

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Si σ_1^2 y σ_2^2 Desconocidas
  - Si $\sigma_1^2 = \sigma_2^2$

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2, v} \cdot SP \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$v = (n_1 + n_2) - 2$$

$$SP = \sqrt{\frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1 + n_2 - 2}}$$

  - Si $\sigma_1^2 \neq \sigma_2^2$

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2, v} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$ Aproximado

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS $\mu_1 - \mu_2$ DE DOS POBLACIONES NO NORMALES

Se debe cumplir la regla empírica $n_1 \geq 30$ y $n_2 \geq 30$

- Si σ_1^2 y σ_2^2 Conocidas

Aproximado

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Si σ_1^2 y σ_2^2 Desconocidas

Aproximado

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Si el intervalo contiene al cero los promedios son muy similares

Si el intervalo está a la derecha de cero $\mu_1 > \mu_2$

Si el intervalo está a la izquierdo de cero $\mu_2 > \mu_1$

TEST DE HIPOTESIS

"Afirmación sobre un valor de un parámetro de una distribución"

En cualquier problema Ph (Prueba de hipotesis) hay dos hipotesis contradictorias que se contrastan entre sí

- Hipotesis nula (Ho) Desigualdad no estricta o igualdad ($\leq, \geq, =$)

- Hipotesis alternativa (Ha) Desigualdad estricta o diferencia ($<, >, \neq$)

Objetivo: Decidir rechazar Ho o no rechazar Ho

*** PROCEDIMIENTO Ph**
- Definir estadístico de prueba (Z_c, T_c, F_c, X_c)
- Definir región de rechazo

- Hallar P-value

TIPOS DE ERRORES

- Error tipo (I)

P(Rechazar Ho | Ho cierta)

α = P(Error tipo I)
↳ Nivel de significancia
Maximo probabilidad de error

- Error tipo (II)

P(Aceptar Ho | Ho Falsa)

β = P(Error tipo II)

$1-\beta$ = Potencia de la prueba

- Interpretaciones

  - α (Significancia)
  Cada vez que rechazamos Ho y Ho es verdadera, nos podemos estar equivocando el P(Error tipo I) de las veces

  - β (Potencia)
  Cada vez que no rechazamos Ho y Ho es falsa, nos podemos estar equivocando aproximadamente el P(Error tipo II) de las veces

DECISIÓN CON P-VALUE

El valor-P de una Ph es la probabilidad calculada, suponiendo que Ho es verdadera, de que el estadístico de prueba tome un valor obtenido

- Se rechaza Ho si P-Value < α
- Aumentar tamaño de muestra si P-value = α
- No se rechaza Ho si P-Value > α

[redacted]

Ph Para μ

Ho: $\mu = \mu_0$

Ha:
$\mu > \mu_0$ (1)
$\mu < \mu_0$ (2)
$\mu \neq \mu_0$ (3)

- *Poblaciones Normales*

- Si la varianza σ^2 es conocida

$$Z_c = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim n(0,1)$$

- Si la varianza σ^2 es desconocida

$$T_c = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

- *Poblaciones No Normales*

- Se debe cumplir la regla empírica $n \geq 30$

- Si la varianza σ^2 es conocida

$$Z_c = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim n(0,1)$$

- Si la varianza σ^2 es desconocida

$$Z_c = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim n(0,1)$$

Región de rechazo y P-Value

(1)
RR: $Z_c > Z_\alpha$
RR: $T_c > t_{\alpha, n-1}$
$P(Z > Z_c)$
$P(T > T_c)$

(2)	RR: $Z_c < -Z_\alpha$	$P(Z<Z_c)$
	RR: $T_c < -t_{\alpha,n-1}$	$P(T<T_c)$
(3)	RR: $\lvert Z_c\rvert > Z_{\alpha/2}$	$2P(Z>\lvert Z_c\rvert)$
	RR: $\lvert T_c\rvert > t_{\alpha/2,n-1}$	$2P(T>\lvert T_c\rvert)$

Ph para $\mu_1 - \mu_2$

$H_o: \mu_1 - \mu_2 = \delta_o$

$$H_a: \begin{cases} \mu_1 - \mu_2 > \delta_o & (1) \\ \mu_1 - \mu_2 < \delta_o & (2) \\ \mu_1 - \mu_2 \neq \delta_o & (3) \end{cases}$$

- *Poblaciones Normales*

- Si σ_1^2 y σ_2^2 Conocidas

$$Z_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_o}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim n(0,1)$$

- Si σ_1^2 y σ_2^2 desconocidas e iguales (Prueba F)

$$T_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_o}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$$S_P = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

- Si σ_1^2 y σ_2^2 desconocidas y diferentes (Prueba F)

$$T_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_o}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$$

$$v \Rightarrow \frac{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$

- *Poblaciones No Normales*

■ Se debe cumplir la regla empirica $n_1, n_2 \geq 30$

- Si σ_1^2 y σ_2^2 conocidas

$$Z_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_o}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim n(0;1)$$

Aproximadamente

- Si σ_1^2 y σ_2^2 desconocidas

$$Z_c = \frac{\bar{X}_1 - \bar{X}_2 - \delta_o}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Aproximadamente

Region de rechazo y P-Value

(1)	RR: $Z_c > Z_\alpha$	$P(Z>Z_c)$
	RR: $T_c > t_{\alpha,gl}$	$P(T>T_c)$
(2)	RR: $Z_c < -Z_\alpha$	$P(Z<Z_c)$
	RR: $T_c < -t_{\alpha,gl}$	$P(T<T_c)$
(3)	RR: $\lvert Z_c\rvert > Z_{\alpha/2}$	$2P(Z>\lvert Z_c\rvert)$
	RR: $\lvert T_c\rvert > t_{\alpha/2,gl}$	$2P(T>\lvert T_c\rvert)$

gl = grados de libertad

$T(gl) = T$

Ph para P

$H_o: P = P_o$

$$H_a: \begin{cases} P > P_o & (1) \\ P < P_o & (2) \\ P \neq P_o & (3) \end{cases} \qquad \hat{P} = \frac{x}{n}$$

$$Z_c = \frac{\hat{P} - P_o}{\sqrt{\frac{P_o(1-P_o)}{n}}} \sim n(0,1)$$

Aproximadamente

Region de rechazo y P-Value

(1)	RR: $Z_c > Z_\alpha$	$P(Z>Z_c)$
(2)	RR: $Z_c < -Z_\alpha$	$P(Z<Z_c)$
(3)	RR: $\lvert Z_c\rvert > Z_{\alpha/2}$	$2P(Z>\lvert Z_c\rvert)$

DD MM AA

Ph para $P_1 - P_2$

$Ho: P_1 - P_2 = \delta_0$

$Ha:$
- $P_1 - P_2 > \delta_0$ (1)
- $P_1 - P_2 < \delta_0$ (2)
- $P_1 - P_2 \neq \delta_0$ (3)

$\hat{P} = \frac{x}{n}$

Para Z_c hay dos opciones:

$$Z_c = \frac{\hat{P}_1 - \hat{P}_2 - \delta_0}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}} \quad ó$$

$$Z_c = \frac{\hat{P}_1 - \hat{P}_2 - \delta_0}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow \boxed{\text{Cuando } \delta_0 = 0}$$

Donde

$$P = \frac{x_1 + x_2}{n_1 + n_2}$$

Region de rechazo y P-value

(1) RR: $Z_c > Z_\alpha$	$P(Z > Z_c)$
(2) RR: $Z_c < -Z_\alpha$	$P(Z < Z_c)$
(3) RR: $\lvert Z_c \rvert > Z_{\alpha/2}$	$2P(Z > \lvert Z_c \rvert)$

Ph Para σ_1^2 / σ_2^2

Razon de varianzas o homocedasticidad

$Ho: \sigma_1^2 = \sigma_2^2$

$Ha:$
- $\sigma_1^2 > \sigma_2^2$ (1)
- $\sigma_1^2 < \sigma_2^2$ (2)
- $\sigma_1^2 \neq \sigma_2^2$ (3)

$$F_c = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,\, n_2-1}$$

Region de rechazo y P-Value

(1) RR: $F_c > F_{1-\alpha,\,(n_1-1,\, n_2-1)}$

$P(F_{v_1,v_2} > F_c)$

(2) RR: $F_c < \frac{1}{F_{1-\alpha,\,(n_2-1,\, n_1-1)}}$

$P\left(F_{v_1,v_2} < \frac{1}{F_{1-\alpha,\,(n_2-1),\, n_1-1}}\right)$

(3) RR:

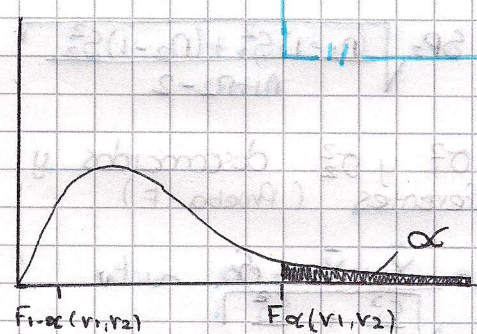
$F_c < \frac{1}{F_{1-\alpha/2}\,(n_2-1)\, n_2-1}$ ó

$F_c > F_{1-\alpha/2}\,(n_1-1,\, n_2-1)$

$\min \{ P(F_{(v_1,v_2)} > F_c),\ P(F_{(v_1,v_2)} < F_c) \}$

$F_c < 1,\ P(F_{(v_1,v_2)}) + P\left(F_{(v_2,v_1)} > \frac{1}{F_c}\right)$

Nota: $F_{\alpha/2}(n_1-1,\, n_2-1) = \frac{1}{F_{1-\alpha/2}\,(n_2-1,\, n_1-1)}$



$\alpha = P(F > F_\alpha(v_1,v_2))$

Ph Bondad de ajuste

Se utiliza la prueba Chi Cuadrado en VAD

Generalmente se utiliza en un experimento Multinomial, con K posibles categorias

Se utiliza para comprobar ciertas distribuciones o/y las proporciones que corresponden a cierto experimento

Cuando queremos verificar proporciones utilizamos

Ho: Todas las probabilidades P_i son iguales a las propuestas P_i^0

$\forall P_i = P_i^0$

Ha: Al menos una de las probabilidades P_i es distinta a la propuesta.

$\exists_i$ tal que $P_i \neq P_i^0$

Cuando queremos verificar cierta distribucion con su parametro

Ho: Los datos corresponden a la distribucion con ese parametro

$X \sim Poiss(\lambda)$

. . . . . $X \sim$ cierta $dist^n$

. . . .

Ha: Los datos no corresponden a la distribucion con ese parametro

$X \nsim Poiss(\lambda)$

. . . . . $X \nsim$ cierta $dist^n$

. . . . .

$$X_c = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}$$

O_i = Frecuencias Observadas

e_i = Frecuencias esperadas $= nP_i$

Para aplicar el test Chi-Cuadrado se debe cumplir la regla empirica

$nP_i \geq 5$

$e_i \geq 5$

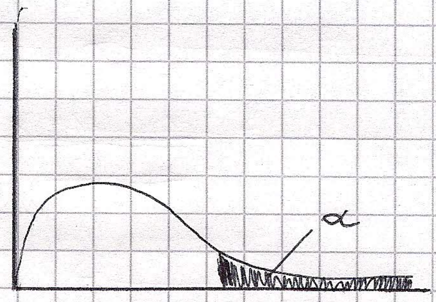
De no cumplirse juntamos categorias

P-Value

$$\boxed{P(X^2_{(k-r-1)} > X_c)}$$

K = Clases

r = Parametros estimados



Para concluir las pruebas de hipotesis utilizamos la hipotesis (Ha) de tal manera que al:

Rechazar Ho

Basados en el criterio de ...(1)... se rechaza Ho, es decir que hay evidencia muestral suficiente para sugerir que...(Ha cierta)(2)... con una significancia de $\alpha\%$

No Rechazar Ho

Basados en el criterio de ...(1)... no se rechaza Ho, es decir que no hay evidencia muestral suficiente para sugerir que ...(Ha cierta)(2)... con una significancia de $\alpha\%$

(1) decision del valor P
la region de rechazo

(2) $>, <, \neq$ Parametro(s)