

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

DEPARTAMENTO DE ESTADÍSTICA

ACADEMIA DE MÉTODOS ESTADÍSTICOS AVANZADOS A2

Formulario para los cursos:
Estadística II
Inferencia estadística

Author:

Dr. José Antonio GUERRERO DÍAZ DE LEÓN

2018

VALOR ESPERADO $E[X]$

Sea X una v.a. con distribución $f(x)$,

$$E[X] = \begin{cases} \sum x_i f(x_i) & \text{c. discreto} \\ \int x f(x) dx & \text{c. continuo} \end{cases}$$

Propiedades

Sean X y Y v.a. y $c \in R$,

- $E[c] = c$
- $E[cX] = cE[X]$
- $E[X + c] = E[X] + c$
- $E[X + Y] = E[X] + E[Y]$

VARIANZA $Var[X]$

Sea X una v.a. con distribución $f(x)$,

$$Var[X] = \begin{cases} \sum (x_i - E[X])^2 f(x_i) & \text{c. discreto} \\ \int (x - E[X])^2 f(x) dx & \text{c. continuo} \end{cases}$$

Propiedades

Sean X y Y v.a. independientes y $c \in R$,

- $Var[c] = 0$
- $Var[cX] = c^2 Var[X]$
- $Var[X + c] = Var[X]$
- $Var[X + Y] = Var[X] + Var[Y]$
- $Var[X] = E[X^2] - E[X]^2$

MOMENTOS

Momentos muestrales

$$\hat{m}_k = \sum_{i=1}^n \frac{(x_i)^k}{n}$$

Momentos poblacionales

$$m_k = E[X^k] = \begin{cases} \sum (x_i)^k f(x_i) & \text{c. discreto} \\ \int x^k f(x) dx & \text{c. continuo} \end{cases}$$

PROPIEDADES DE \prod (PRODUCTO)

Sean $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n, c \in R$

- $\prod_{i=1}^n c = c^n$
- $\prod_{i=1}^n (a_i b_i) = \left(\prod_{i=1}^n a_i \right) \left(\prod_{i=1}^n b_i \right)$
- $\prod_{i=1}^n (a_i)^c = \left(\prod_{i=1}^n a_i \right)^c$
- $\prod_{i=1}^n c^{a_i} = c^{\left(\sum_{i=1}^n a_i \right)}$
- $\prod_{i=1}^n e^{a_i} = e^{\left(\sum_{i=1}^n a_i \right)}$
- $\ln \prod_{i=1}^n a_i = \sum_{i=1}^n \ln(a_i)$

DISTRIBUCIONES MUESTRALES

- $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$
- $\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$
- $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$

$$F_{\alpha, \nu_1, \nu_2} = \frac{1}{F_{1-\alpha, \nu_2, \nu_1}}$$

Parámetro	Estimador	Error estándar	Intervalo del $(1 - \alpha)100\%$	Fórmulas adicionales
Media μ (σ conocida)	\bar{x}	$\frac{\sigma}{\sqrt{n}}$	$\left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} , \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$	Tamaño de muestra $n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{E} \right)^2$
Media μ (σ desconocida)	\bar{x}		$\left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} , \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \right)$	
Varianza σ^2	s^2		$\left(\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}, n-1}^2} , \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2} \right)$	
Proporción p	\hat{p}	$\sqrt{\frac{p(1-p)}{n}}$	$\left(\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} , \hat{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$	Tamaño de muestra $n = \left(\frac{Z_{\frac{\alpha}{2}}}{E} \right)^2 p^*(1-p^*)$ p^* =estimación previa
Dif. de medias $\mu_1 - \mu_2$ (σ_1 y σ_2 conocidas)	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\left((\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$	
Dif. de medias $\mu_1 - \mu_2$ ($\sigma_1 = \sigma_2$ desconocidas)	$\bar{x}_1 - \bar{x}_2$		$\left((\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}, \nu} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} , (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}, \nu} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$	$\nu = n_1 + n_2 - 2$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
Dif. de medias $\mu_1 - \mu_2$ ($\sigma_1 \neq \sigma_2$ desconocidas)	$\bar{x}_1 - \bar{x}_2$		$\left((\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$	$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 - 1}}$
Dif. de medias μ_d (muestras pareadas)	\bar{d}		$\left(\bar{d} - t_{\frac{\alpha}{2}, n-1} \frac{s_d}{\sqrt{n}} , \bar{d} + t_{\frac{\alpha}{2}, n-1} \frac{s_d}{\sqrt{n}} \right)$	d =diferencia de las observaciones
Diferencia de proporciones $p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$	$\left((\hat{p}_1 - \hat{p}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} , (\hat{p}_1 - \hat{p}_2) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right)$	$q_i = 1 - p_i$ $\hat{q}_i = 1 - \hat{p}_i$
Cociente de varianzas $\frac{\sigma_1^2}{\sigma_2^2}$	$\frac{s_1^2}{s_2^2}$		$\left(\frac{s_1^2}{s_2^2} \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}} , \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1} \right)$	$F_{\omega, \nu_1, \nu_2} = \frac{1}{F_{1-\omega, \nu_2, \nu_1}}$

Pruebas de hipótesis

Prueba	Estadístico de prueba	Distribución	Fórmulas adicionales
Para una media (σ conocida)	$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	$N(0, 1)$	
Para una media (σ desconocida)	$T = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$	t con $n - 1$ gl	
Para la varianza	$X^2 = \frac{(n - 1)S^2}{\sigma_0^2}$	χ^2 con $n - 1$ gl	
Para una proporción	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$N(0, 1)$	
Para diferencia de medias (σ_1 y σ_2 conocidas)	$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$N(0, 1)$	
Para diferencia de medias ($\sigma_1 = \sigma_2$ desconocidas)	$T = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t con $n_1 + n_2 - 2$ gl	$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
Para diferencia de medias ($\sigma_1 \neq \sigma_2$ desconocidas)	$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	t con ν gl	$\nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{S_2^2}{n_2})^2}{n_2 - 1}}$
Para diferencia de medias (muestras pareadas)	$T = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$	t con $n - 1$ gl	$d_i = X_{1,i} - X_{2,i}$
Para diferencia de proporciones	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{1,2}(1 - \hat{p}_{1,2}) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$	$N(0, 1)$	$\hat{p}_{1,2} = \frac{x_1 + x_2}{n_1 + n_2}$, donde $\hat{p}_i = \frac{x_i}{n_i}$
Para cociente de varianzas	$f = \frac{s_1^2}{s_2^2}$	F con $n_1 - 1, n_2 - 1$ gl	$F_{\alpha; \nu_1, \nu_2} = \frac{1}{F_{1-\alpha; \nu_2, \nu_1}}$

PRUEBAS DE HOMOGENEIDAD E INDEPENDENCIA

$$X = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{\nu}^2$$

donde

- La tabla es de m renglones por k columnas
- n_{ij} : total en la celda (i, j)
- $n_{i\bullet}$: total de reglón i
- $n_{\bullet j}$: total de columna j
- N : total de observaciones en la tabla
- O_{ij} : Frecuencia observada para la celda (i, j)
- $E_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{N}$: Frecuencia esperada para la celda (i, j)
- $\nu = (m - 1)(k - 1)$

PRUEBA DE BONDAD DE AJUSTE

$$X = \sum_{i=1}^m \frac{(O_i - E_i)^2}{E_i} \sim \chi_{\nu}^2$$

donde

- La tabla tiene en total m celdas.
- p_i : probabilidad teórica de pertenecer a la celda i .
- N : total de observaciones en la tabla
- $E_i = Np_i$
- r : número de parámetros a estimar
- $\nu = m - r - 1$

Sumas de cuadrados

$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$	$S_{yy} = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2$	$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$
$SSR = \hat{\beta}_1 S_{xy}$	$SSE = S_{yy} - SSR$	$SST = S_{yy}$
$MSR = \frac{SSR}{1}$	$MSE = \frac{SSE}{n-2}$	$MST = \frac{SST}{n-1}$

Parámetro	Estimador	Error estándar	Intervalo de C.	Estadístico de prueba
σ	$S = \sqrt{MSE}$			
ρ	$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$	$S_r = \sqrt{\frac{1-r^2}{n-2}}$		$T = \frac{r - \rho_0}{S_r} \sim t \text{ con } n-2 \text{ gl}$
β_1	$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$	$S_{\hat{\beta}_1} = \frac{S}{\sqrt{S_{xx}}}$	$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-2} S_{\hat{\beta}_1}$	$T = \frac{\hat{\beta}_1 - (\beta_1)_0}{S_{\hat{\beta}_1}} \sim t \text{ con } n-2 \text{ gl}$
β_0	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$	$S_{\hat{\beta}_0} = S \sqrt{\frac{\sum_{i=1}^n x_i^2}{n S_{xx}}}$	$\hat{\beta}_0 \pm t_{\frac{\alpha}{2}, n-2} S_{\hat{\beta}_0}$	$T = \frac{\hat{\beta}_0 - (\beta_0)_0}{S_{\hat{\beta}_0}} \sim t \text{ con } n-2 \text{ gl}$
$\mu_{Y x_0}$	$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$	$S_{\hat{Y}_0} = S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$	$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2} S_{\hat{Y}_0}$	
Y_0	$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$	$S_{\hat{Y}_0 - Y_0} = S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$	$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2} S_{\hat{Y}_0 - Y_0}$	

Coefficientes

de determinación $r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$	de determinación ajustado $r_{adj}^2 = 1 - (1 - r^2) \frac{n-1}{n-2}$
---	--

Residuales

sin estandarizar $e_i = y_i - \hat{y}_i$	estandarizados $\frac{e_i}{\sqrt{MSE}}$
---	--

Prueba F para el modelo de regresión

$$F = \frac{MSR}{MSE} \sim F_{1, n-2 \text{ gl}}$$

Estadístico de Durbin-Watson

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

Nombre de la distribución	Función de distribución de probabilidad	Dominio	Parámetros	Media $E[X]$	Varianza $V[X]$	Generadora de momentos $M_X(t)$
Uniforme	$P(X = x) = \frac{1}{N}$	$x \in \{x_1, x_2, \dots, x_N\}$		\bar{x}	S_N^2	$\frac{1}{N} \sum_{i=1}^n e^{tx_i}$
	$P(X = x) = \frac{1}{N}$ $N = b - a + 1$	$x = a, \dots, b$	a : Límite inferior b : Límite superior	$\frac{a + b}{2}$	$\frac{N^2 - 1}{12}$	$\frac{e^{at} - e^{(b+1)t}}{N(1 - e^t)}$
Bernoulli	$P(X = x) = p^x(1 - p)^{1-x}$	$x = 0, 1$	p : Probabilidad de éxito	p	$p(1 - p)$	$pe^t + (1 - p)$
Binomial	$P(X = x) = \binom{N}{x} p^x(1 - p)^{N-x}$	$x = 0, \dots, N$	p : Probabilidad de éxito N : Número de ensayos	Np	$Np(1 - p)$	$(pe^t + (1 - p))^N$
Geométrica	$P(X = x) = p(1 - p)^{x-1}$	$x = 1, 2, \dots$	p : Probabilidad de éxito	$\frac{1}{p}$	$\frac{1 - p}{p^2}$	$\frac{pe^t}{1 - (1 - p)e^t}$
	$P(Y = y) = p(1 - p)^y$ $Y = X - 1$	$y = 0, 1, \dots$	p : Probabilidad de éxito	$\frac{1 - p}{p}$	$\frac{1 - p}{p^2}$	$\frac{p}{1 - (1 - p)e^t}$
Binomial negativa	$P(X = x) = \binom{x-1}{r-1} p^r(1 - p)^{x-r}$	$x = r, r + 1, \dots$	p : Probabilidad de éxito r : Número de éxitos deseados	$\frac{r}{p}$	$\frac{r(1 - p)}{p^2}$	$\left(\frac{pe^t}{1 - (1 - p)e^t}\right)^r$
Poisson	$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$	$x = 0, 1, \dots$	λ : tasa	λ	λ	$e^{\lambda(e^t - 1)}$
Hipergeométrica	$P(X = x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$	$x = A, \dots, B$ $A = \max(0, n + K - N)$ $B = \min(n, K)$	N : Tamaño de la población n : Tamaño de la muestra K : Tamaño de subconjunto de interés	$n \frac{K}{N}$	$\frac{nK(N - K)(N - n)}{N^2(N - 1)}$	No es práctica

Nombre	Función de densidad	Parámetros	Media $E[X]$	Varianza $V[X]$	Función generadora de momentos $M_X(t)$
Uniforme	$f_X(x) = \frac{1}{b-a}$ $a \leq x \leq b$	a : localización (límite inferior) b : localización (límite superior) $-\infty < a, b < \infty$ $a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt} - e^{at}}{(b-a)t}$
Exponencial	$f_X(x) = \lambda e^{-\lambda x}$ $x > 0$	λ : tasa $\lambda > 0$ $\lambda = 1/\theta$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t} \quad t < \lambda$
	$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$ $x > 0$	θ : escala $\theta > 0$	θ	θ^2	$\frac{1}{1-t\theta} \quad t < \frac{1}{\theta}$
Normal	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty$	μ : localización (central) $-\infty < \mu < \infty$ σ : escala $\sigma > 0$	μ	σ^2	$e^{\mu t + \frac{t^2 \sigma^2}{2}}$
Gamma	$f_X(x) = \frac{\lambda^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\lambda x}$ $x > 0$	λ : tasa β : forma $\beta, \lambda > 0$	$\frac{\beta}{\lambda}$	$\frac{\beta}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^\beta \quad t < \lambda$
	$f_X(x) = \frac{1}{\theta^\beta \Gamma(\beta)} x^{\beta-1} e^{-\frac{x}{\theta}}$ $x > 0$	θ : escala β : forma $\beta, \theta > 0$	$\beta\theta$	$\beta\theta^2$	$\left(\frac{1}{1-t\theta}\right)^\beta \quad t < \frac{1}{\theta}$
Weibull	$f_X(x) = \beta \alpha^\beta x^{\beta-1} e^{-(\alpha x)^\beta}$ $x > 0$	α : tasa β : forma $\alpha, \beta > 0$ $\alpha = \frac{1}{\theta}$	$\frac{1}{\alpha} \Gamma\left(1 + \frac{1}{\beta}\right)$	$\frac{1}{\alpha^2} \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2 \right]$	$\sum_{k=0}^{\infty} \frac{t^k}{k! \alpha^k} \Gamma\left(1 + \frac{1}{\beta}\right)$ $\beta \geq 1$
	$f_X(x) = \frac{\beta}{\theta} \left(\frac{x}{\theta}\right)^{\beta-1} e^{-\left(\frac{x}{\theta}\right)^\beta}$ $x > 0$	θ : escala β : forma $\theta, \beta > 0$	$\theta \Gamma\left(1 + \frac{1}{\beta}\right)$	$\theta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2 \right]$	$\sum_{k=0}^{\infty} \frac{t^k \theta^k}{k!} \Gamma\left(1 + \frac{1}{\beta}\right)$ $\beta \geq 1$