I.T. INDUSTRIAL
METODOS ESTADÍSTICOS

FORMULARIO

I. ESTADISTICA DESCRIPTIVA

X v.a	k modalidades $x_1, x_2, ..., x_k$; n datos
Media	$\bar{x} = \dfrac{\sum_{i=1}^{k} n_i x_i}{n}$
Varianza poblacional	$\sigma^2 = \dfrac{\sum_{i=1}^{k} n_i x_i^2}{n} - \bar{x}^2$
Varianza muestral	$S^2 = \dfrac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{n-1}$
Coeficiente de Asimetría	$\gamma_1 = \dfrac{\sum_{i=1}^{k} (x_i - x)^3 n_i}{n\sigma^3}$
Coeficiente de Apuntamiento	$\gamma_2 = \dfrac{\sum_{i=1}^{k} (x_i - x)^4 n_i}{n\sigma 4} - 3$
Covarianza	$Cov(X,Y) = \dfrac{\sum_{i=1}^{K} \sum_{j=1}^{p} n_{ij} x_i y_j}{N} - \bar{x}\bar{y}$
Coeficiente de correlación lineal	$\rho = \dfrac{Cov(X,Y)}{\sigma_x \sigma_y}$
Recta de regresión de Y sobre X	$\hat{y} = \left[\dfrac{Cov\,[X,Y]}{Var\,(X)}\right] x + \left[\bar{y} - \dfrac{Cov\,[X,Y]}{Var\,(X)} \bar{x}\right]$

II. PROBABILIDAD

Operaciones con sucesos

Suceso	**Ocurre siempre que**	**Probabilidad**
$\bar{A}$	**no** ocurre A	$P\left(\bar{A}\right) = 1 - P(A)$
$\mathbf{A/B}$	ocurre A **si** ya ha ocurrido B	$P(A/B) = \dfrac{P(A \cap B)}{P(B)}$
$\mathbf{A \cap B}$	ocurra A **o** B	$P(A \cap B) = P(A)P(B/A)$ $P(A \cap B) = P(B)P(A/B)$
$\mathbf{A \cup B}$	ocurren A **y** B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Si A y B son incompatibles	**Si A y B independientes**
$P(A/B) = 0$	$P(A/B) = P(A)$
$P(A \cap B) = 0$	$P(A \cap B) = P(A)P(B)$
$P(A \cup B) = P(A) + P(B)$	$P(A \cup B) = P(A) + P(B) - P(A)P(B)$

Teorema de la probabilidad total y fórmula de Bayes

Probabilidad total	$P(A) = \sum\limits_{i=1}^{k} P(A/B_i)P(B_i)$
Fórmula de Bayes	$P(B_i/A) = \dfrac{P(A \cap B_i)}{P(A)} = \dfrac{P(A/B_i)P(B_i)}{\sum\limits_{i=1}^{k} P(A/B_i)P(B_i)}$

	ALGUNAS DISTRIBUCIONES IMPORTANTES	
Distribución	**Función masa de probabilidad/función de densidad**	EX
$U(n)$	$P[X = x_i] = \frac{1}{n}$	$\frac{\sum X_i}{n}$
$B(n,p)$	$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$; $k = 0, 1, 2...n$	np
$P(\lambda)$	$P[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}; k = 0, 1, 2...$	λ
$H(N, n, p)$	$P[X = k] = \frac{\binom{N_p}{k}\binom{N_q}{n-k}}{\binom{N}{n}}$, $max\{0, n - N_q\} \leq k \leq min\{n, N_p\}$	np
$BN(K, p)$	$P(X = x) = \binom{K + x - 1}{x} p^k (1-p)^x; x = 0, 1, 2....$	$\frac{kq}{p}$
$G(p)$	$P(X = x) = pq^x; x = 0, 1, 2....$	$\frac{q}{p}$
$N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}, x \in R$	μ
$N(0,1)$	$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}z^2}, z \in R$	0
$G(\alpha, \lambda)$	$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0$ $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$	$\frac{\alpha}{\lambda}$
χ_n^2	$f(x) = \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, x > 0$	n
$B(\alpha, \beta)$	$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}; 0 < x < 1$	$\frac{\alpha}{\alpha + \beta}$
t_n	$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + \frac{t^2}{n})^{-(\frac{n+1}{2})}; t \in R$	0
$F_{n1,n2}$	$g(f) = \frac{\Gamma(n_1 + n_2)}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} f^{\frac{n_1-2}{2}} (n_1 f + n_2)^{-(\frac{n_1+n_2}{2})}; f > 0$	$\frac{n_2}{n_2 - 2}$

III. INFERENCIA ESTADISTICA

INTERVALOS DE CONFIANZA

Intervalo de confianza para la media de una normal

Varianza conocida (σ_0^2)	$\mu \in$	$\left[\bar{x} \pm \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]$
Varianza desconocida	$\mu \in$	$\left[\bar{x} \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}\right]$

Intervalo de confianza para la varianza de una normal

Media conocida (μ_0)	$\sigma^2 \in$	$\left[\frac{\sum_{i=1}^{n} (x_i - \mu_0)^2}{\chi^2_{1-\frac{\alpha}{2};n}}, \frac{\sum_{i=1}^{n} (x_i - \mu_0)^2}{\chi^2_{\frac{\alpha}{2};n}}\right]$
Media desconocida	$\sigma^2 \in$	$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right]$

Intervalo de confianza para la diferencia de medias de dos poblaciones normales e independientes

Varianza conocidas	$\mu_x - \mu_y \in$	$\left[\bar{x} - \bar{y} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x}{n_x} + \frac{\sigma_y}{n_y}}\right]$
Varianza desconocidas pero iguales (σ^2)	$\mu_x - \mu_y \in$	$\left[\bar{x} - \bar{y} \pm t_{1-\frac{\alpha}{2};n_x+n_y-2} S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right]$

con $S_p = \sqrt{\dfrac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2}}$

Intervalo de confianza para el cociente de varianzas de dos poblaciones normales e independientes

Medias conocidas	$\frac{\sigma_y^2}{\sigma_x^2} \in \left[\frac{\sum_{i=1}^{n_y} \left(y_i - \mu_y\right)^2}{\sum_{i=1}^{n_x} \left(x_i - \mu_x\right)^2} \frac{n_x}{n_y} F_{\frac{\alpha}{2};n_x,n_y}, \frac{\sum_{i=1}^{n_y} \left(y_i - \mu_y\right)^2}{\sum_{i=1}^{n_x} \left(x_i - \mu_x\right)^2} \frac{n_x}{n_y} F_{1-\frac{\alpha}{2};n_x,n_y} \right]$
Medias desconocidas	$\frac{\sigma_y^2}{\sigma_x^2} \in \left[\frac{S_y^2 F_{\frac{\alpha}{2};n_x-1,n_y-1}}{S_x^2}, \frac{S_y^2 F_{1-\frac{\alpha}{2};n_x-1,n_y-1}}{S_x^2} \right]$

Intervalo de confianza para una proporción

$$p \in \left[\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \right]$$

CONTRASTES DE HIPÓTESIS

Contraste para la media de una normal con varianza conocida

Hipótesis nula	Estadístico de contraste
$H_0 : \mu = \mu_0$	$Z = \dfrac{\bar{X} - \mu}{\sigma_0/\sqrt{n}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \mu \neq \mu_0$	$Z \leq z_{\alpha/2}$ o $Z \geq z_{1-\alpha/2}$
$H_1 : \mu > \mu_0$	$Z \geq z_{1-\alpha}$
$H_1 : \mu < \mu_0$	$Z \leq z_{\alpha}$

Contraste para la media de una normal con varianza desconocida

Hipótesis nula	Estadístico de contraste
$H_0 : \mu = \mu_0$	$T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \mu \neq \mu_0$	$T \leq t_{\alpha/2,n-1}$ o $T \geq t_{1-\alpha/2,n-1}$
$H_1 : \mu > \mu_0$	$T \geq t_{1-\alpha,n-1}$
$H_1 : \mu < \mu_0$	$T \leq t_{\alpha,n-1}$

Contraste para la varianza de una normal con media conocida

Hipótesis nula	Estadístico de contraste
$H_0 : \sigma^2 = \sigma_0^2$	$\chi^2 = \dfrac{\sum_{i=1}^{n} (x_i - \mu_0)^2}{\sigma_0^2}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi^2 \leq \chi^2_{\alpha/2,n}$ o $\chi^2 \geq \chi^2_{1-\alpha/2,n}$
$H_1 : \sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi^2_{1-\alpha,n}$
$H_1 : \sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi^2_{\alpha,n}$

Contraste para la varianza de una normal con media desconocida

Hipótesis nula	Estadístico de contraste
$H_0 : \sigma^2 = \sigma_0^2$	$\chi^2 = \dfrac{(n-1)\,S^2}{\sigma_0^2}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi^2 \leq \chi^2_{\alpha/2,n-1}$ o $\chi^2 \geq \chi^2_{1-\alpha/2,n-1}$
$H_1 : \sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi^2_{1-\alpha,n-1}$
$H_1 : \sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi^2_{\alpha,n-1}$

Contraste para el cociente de varianzas de dos normales independientes con medias conocidas

Hipótesis nula	Estadístico de contraste
$H_0 : \sigma_X^2 = \sigma_Y^2$	$F = \dfrac{\sum_{i=1}^{n_X} (x_i - \mu_X)^2 / n_X}{\sum_{i=1}^{n_Y} (y_i - \mu_Y)^2 / n_Y}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \sigma_X^2 \neq \sigma_Y^2$	$F \leq 1/f_{1-\alpha/2,n_Y,n_X}$ o $F \geq f_{1-\alpha/2,n_X,n_Y}$
$H_1 : \sigma_X^2 > \sigma_Y^2$	$F \geq f_{1-\alpha,n_X,n_Y}$
$H_1 : \sigma_X^2 < \sigma_Y^2$	$F \leq 1/f^2_{1-\alpha,n_Y,n_X}$

Contraste para el cociente de varianzas de dos normales independientes con medias desconocidas

Hipótesis nula	Estadístico de contraste
$H_0 : \sigma_X^2 = \sigma_Y^2$	$F = \dfrac{S_X^2}{S_Y^2}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \sigma_X^2 \neq \sigma_Y^2$	$F \leq 1/f_{1-\alpha/2,n_Y-1,n_X-1}$ o $F \geq f_{1-\alpha/2,n_X-1,n_Y-1}$
$H_1 : \sigma_X^2 > \sigma_Y^2$	$F \geq f_{1-\alpha,n_X-1,n_Y-1}$
$H_1 : \sigma_X^2 < \sigma_Y^2$	$F \leq 1/f_{1-\alpha,n_Y-1,n_X-1}$

Contraste para la diferencia de medias de dos normales independientes con varianzas conocidas

Hipótesis nula	Estadístico de contraste
$H_0 : \mu_X - \mu_Y = \delta_0$	$Z = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{\sigma^2_{n,X}}{n_X} + \frac{\sigma^2_{n,Y}}{n_Y}}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \mu_X - \mu_Y \neq \delta_0$	$Z \leq z_{\alpha/2}$ o $Z \geq z_{1-\alpha/2}$
$H_1 : \mu_X - \mu_Y > \delta_0$	$Z \geq z_{1-\alpha}$
$H_1 : \mu_X - \mu_Y < \delta_0$	$Z \leq z_\alpha$

Contraste para la diferencia de medias de dos normales independientes con varianzas desconocidas pero iguales

Hipótesis nula	Estadístico de contraste
$H_0 : \mu_X - \mu_Y = \delta_0$	$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{S_p^2 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \mu_X - \mu_Y \neq \delta_0$	$T \leq t_{\alpha/2,n}$ o $T \geq t_{1-\alpha/2,n}$
$H_1 : \mu_X - \mu_Y > \delta_0$	$T \geq t_{1-\alpha,n}$
$H_1 : \mu_X - \mu_Y < \delta_0$	$T \leq t_{\alpha,n}$

donde

$$\begin{aligned} n &= n_X + n_Y - 2 \\ S_p^2 &= \frac{(n_X - 1)\, S_X^2 + (n_Y - 1)\, S_Y^2}{n} \end{aligned}$$

Contraste para la diferencia de medias de dos normales relacionadas (muestras apareadas) con varianzas desconocidas pero iguales

Hipótesis nula	Estadístico de contraste
$H_0 : \mu_X - \mu_Y = \delta_0$	$T = \dfrac{\bar{D} - \delta_0}{\sqrt{\frac{S_D^2}{n}}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : \mu_X - \mu_Y \neq \delta_0$	$T \leq t_{\alpha/2,n-1}$ o $T \geq t_{1-\alpha/2,n-1}$
$H_1 : \mu_X - \mu_Y > \delta_0$	$T \geq t_{1-\alpha,n-1}$
$H_1 : \mu_X - \mu_Y < \delta_0$	$T \leq t_{\alpha,n-1}$

donde $D = X - Y$

Contraste para una proporción

Hipótesis nula	Estadístico de contraste
$H_0 : p = p_0$	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 (1 - p_0)}{n}}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : p \neq p_0$	$Z \leq z_{\alpha/2}$ o $Z \geq z_{1-\alpha/2}$
$H_1 : p > p_0$	$Z \geq z_{1-\alpha}$
$H_1 : p < p_0$	$Z \leq z_{\alpha}$

Contraste para la comparación de dos proporciones

Hipótesis nula	Estadístico de contraste
$H_0 : p_1 = p_2$	$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{p_T (1 - p_T)/n_1 + p_T (1 - p_T)/n_2}}$
Hipótesis alternativa	**Criterios de rechazo**
$H_1 : p_1 \neq p_2$	$Z \leq z_{\alpha/2}$ o $Z \geq z_{1-\alpha/2}$
$H_1 : p_1 > p_2$	$Z \geq z_{1-\alpha}$
$H_1 : p_1 < p_2$	$Z \leq z_{\alpha}$

donde

$$\hat{p}_T = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n1 + n_2}$$

TABLA ANOVA Y CONTRASTE DE LA F

Fuentes de Variación	Suma de Cuadrados	Grados de libertad	Varianzas
Entre grupos (VE)	$\sum_{i=1}^{I} n_i(\bar{x}_i - \bar{x})^2$	$I-1$	$S_e^2 = \frac{VE}{I-1}$
Interna, no explicada o residual (VNE)	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2 = \sum_{i=1}^{I} n_i\sigma_i^2$	$n-I$	$S_R^2 = \frac{VNE}{n-I}$
TOTAL (VT)	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2$	$n-1$	$S_y^2 = \frac{VT}{n-1}$

donde n_i denota el número de observaciones en el grupo i, $\bar{x}_i$ la media , σ_i^2 la varianza (en calculadora σ_n al cuadrado), y $\bar{x}$ la media del conjunto total de observaciones. I es el n^{o}de grupos y n el n^{o}total de observaciones.

Hipótesis nula	Estadístico de contraste
$H_0 : \mu_1 = ... = \mu_I$	$F = \frac{S_e^2}{S_R^2}$
Hipótesis alternativa	Rechazar H_0 si:
No todas las medias son iguales	$F > F_{1-\alpha,I-1,n-I}$