TABLAS Y FORMULAS ESTADISTICAS

Carlo Magno Araya
Profesor de Estadística
Sede de Occidente
Universidad de Costa Rica

MEDIDAS DE POSICION

Datos sin agrupar	Datos agrupados
Promedio aritmético de muestras	
$\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$	$\overline{x} = \dfrac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$
Promedio ponderado $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$	**Mediana** $M_e = L_i + \left(\dfrac{\frac{n}{2} - F_{i-1}}{f_i} \right) * c$
Mediana para n impar $M_e = X_{\left(\frac{n+1}{2}\right)}$	**Moda** $M_o = L_i + \left(\dfrac{d_1}{d_1 + d_2} \right) * c$ $d_1 = f_i - f_{i-1}$ $d_2 = f_i - f_{i+1}$
Mediana para n par $M_e = \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}$	**Percentiles** $P_m = L_i + \left(\dfrac{\frac{m.n}{100} - F_{i-1}}{f_i} \right) * c$
Percentiles $P_m = X_{\left[\frac{m}{100}(n+1)\right]}$	

Media geométrico	Media armónica
$\overline{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$	$\overline{x}_a = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

MEDIDAS DE VARIABILIDAD

Datos sin agrupar	Datos agrupados
Variancia de una muestra	
$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ $s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$	$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 . f_i$ $s_x^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}\right]$
Variancia de la población	
$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2$ $\sigma_x^2 = \frac{1}{N} \cdot \left[\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}\right]$	$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{k}\left(x_i - \mu\right)^2 . f_i$ $\sigma_x^2 = \frac{1}{N} \cdot \left[\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{N}\right]$
Coeficiente de variación de una población $CV_x = \frac{\sigma_x}{\mu} * 100$	Coeficiente de variación de una muestra $CV_x = \frac{s_x}{\overline{x}} * 100$
Desviación media	
$D.M. = \frac{\sum_{i=1}^{n} \lvert x_i - \overline{x} \rvert}{n}$	$D.M. = \frac{\sum_{i=1}^{k} \lvert x_i - \overline{x} \rvert . f_i}{\sum_{i=1}^{k} f_i}$
Medida de variabilidad para muestras pareadas $s_d^2 = \frac{1}{n-1} \cdot \left[\sum_{i=1}^{n} d_i^2\right]$ $\quad d_i = X_{1i} - X_{2i}$	Variancia para variables dicotómicas $\sigma^2 = PQ \qquad s^2 = \hat{p}\hat{q}$

INDICE DE PRECIOS	
Relativo simple de precios $$I = \frac{p_n}{p_0} \cdot 100$$	Agregado simple de precios $$I = \frac{\sum_{i=1}^{k} p_n}{\sum_{i=1}^{k} p_0} \cdot 100$$
Promedio de los relativos simples de precios $$I = \frac{\sum_{i=1}^{k} \left(\frac{p_n}{p_0} \right)}{k} \cdot 100$$	
Índices de precios ponderados	
Laspeyres $$I_{PL} = \frac{\sum p_n q_o}{\sum p_o q_o} \cdot 100$$	Paasche $$I_{PP} = \frac{\sum p_n q_n}{\sum p_o q_n} \cdot 100$$
Índices de cantidades ponderados	
Laspeyres $$I_{QL} = \frac{\sum p_o q_n}{\sum p_o q_o} \cdot 100$$	Paasche $$I_{QP} = \frac{\sum p_n q_n}{\sum p_n q_o} \cdot 100$$
Indice de precio de Fischer	
$$I_{PF} = \sqrt{\left(\frac{\sum p_n q_0}{\sum p_0 q_0} \right) \left(\frac{\sum p_n q_n}{\sum p_0 q_n} \right)} \cdot 100$$	

MUESTREO ALEATORIO SIMPLE

Población finita	Población infinita
Variancia del promedio	
$s_{\bar{x}}^2 = \frac{N - n}{N - 1} . \frac{s_x^2}{n}$	$s_{\bar{x}}^2 = \frac{s_x^2}{n}$
$\sigma_{\bar{x}}^2 = \frac{N - n}{N - 1} . \frac{\sigma_x^2}{n}$	$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$
Variancia de una proporción	
$s_{\hat{p}}^2 = \frac{N - n}{N - 1} . \frac{\hat{p}\hat{q}}{n}$	$s_{\hat{p}}^2 = \frac{\hat{p}\hat{q}}{n}$
$\sigma_{\hat{p}}^2 = \frac{N - n}{N - 1} . \frac{PQ}{n}$	$\sigma_{\hat{p}}^2 = \frac{PQ}{n}$
Tamaño de muestra para la estimación De un promedio y una proporción poblacional	
$n = \frac{n_1}{1 + \frac{n_1}{N}}$ donde $n_1 = \left(\frac{Z_{\alpha/2}\,\sigma}{d}\right)^2$	$n = \left(\frac{Z_{\alpha/2}\,\sigma}{d}\right)^2$
$n = \frac{n_1}{1 + \frac{n_1}{N}}$ donde $n_1 = \left(\frac{Z_{\alpha/2}\sqrt{PQ}}{d}\right)^2$	$n = \left(\frac{Z_{\alpha/2}\sqrt{PQ}}{d}\right)^2$
Intervalos de confianza para el promedio cuando la variancia de la población es conocida	
$L_i = \bar{x} \pm Z_{\alpha/2} * \sqrt{\frac{N - n}{N - 1}} * \frac{\sigma_x}{\sqrt{n}}$	$L_i = \bar{x} \pm Z_{\alpha/2} * \frac{\sigma_x}{\sqrt{n}}$
Intervalos de confianza para el promedio cuando la variancia de la población es desconocida y n≤30	
$L_i = \bar{x} \pm t_{\alpha/2(n-1)gl} * \sqrt{\frac{N - n}{N - 1}} * \frac{s_x}{\sqrt{n}}$	$L_i = \bar{x} \pm t_{\alpha/2(n-1)gl} * \frac{s_x}{\sqrt{n}}$
Intervalos de confianza para una proporción si np>5 y nq>5	
$L_i = \hat{p} \pm Z_{\alpha/2} * \sqrt{\frac{N - n}{N - 1}} * \sqrt{\frac{\hat{p}\hat{q}}{n}}$	$L_i = \hat{p} \pm Z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}}$

ESTADISTICO PARA PRUEBA DE HIPOTESIS

Promedios	Proporciones
Para un promedio: variancia conocida $$Z_c = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$	Para una proporción $$Z_c = \frac{\hat{p} - P}{\sqrt{\frac{PQ}{n}}}$$
Para un promedio: variancia desconocida $$t_c = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$	Diferencia de proporciones $$Z_c = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$
Diferencia de dos promedios: variancia conocida $$Z_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$	Otra alternativa de cálculo: $$Z_c = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}}$$ $$p = \frac{x_1 + x_2}{n_1 + n_2}$$
Diferencia de dos promedios: variancia desconocida $$t_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}} * k$$ donde $$k = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$	
Estadístico de prueba de independencia y de homogeneidad Ji-Cuadrada $$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$ $$E_{ij} = \frac{N_i N_j}{N}$$	Estadístico de prueba para muestras pareadas $$t_c = \frac{\bar{d}}{S_d / \sqrt{n}}$$

ANALISIS DE REGRESION LINEAL SIMPLE

Constante de regresión $$a = \bar{y} - b\bar{x}$$	Coeficiente regresión lineal $$b = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$
Intervalos de confianza para el promedio de y dado un x_0 $$L_i = \hat{y} \pm t_{\alpha/2(n-2)gl} * S_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SC_x}}$$	Intervalos de confianza para una observación de y dado un x_0 $$L_i = \hat{y} \pm t_{\alpha/2(n-2)gl} * S_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SC_x}}$$
Error estándar de estimación $$S_e = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a \sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i y_i}{n-2}}$$	Suma de cuadrados de x $$SC_x = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$
Inferencia sobre la constante y coeficiente de regresión	
Intervalos de confianza $$a \pm t_{(n-2)gl} S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SC_x}}$$ $$b \pm t_{(n-2)gl} \frac{S_e}{\sqrt{SC_x}}$$	Estadístico de prueba de hipótesis $$t_c = \frac{a}{S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SC_x}}}$$ $$t_c = \frac{b}{S_e / \sqrt{SC_x}}$$

ANALISIS DE CORRELACION LINEAL SIMPLE

Coeficiente de correlación lineal $$r = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] * \left[n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$	
Estadístico para prueba de hipótesis sobre el coeficiente de correlación $$t_c = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n-2}}}$$	Coeficiente de correlación parcial $$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

MUESTREO ALEATORIO ESTRATIFICADO

Afijación de la muestra proporcional	Afijación de la muestra óptima o Neyman
$n_h = n \cdot \dfrac{N_h}{\sum_{h=1}^{L} N_h}$	$n_h = n \cdot \left(\dfrac{N_h \sigma_h}{\sum_{h=1}^{L} N_h \sigma_h} \right)$
Promedio aritmético estratificado	**Proporción estratificada**
$\overline{y}_{st} = \dfrac{1}{N} \sum_{h=1}^{L} N_h \overline{y}_h = \sum_{h=1}^{L} W_h \overline{x}_h$ $W_h = \dfrac{N_h}{N}$	$\hat{p}_{st} = \dfrac{1}{N} \sum_{h=1}^{L} N_h \hat{p}_h = \sum_{h=1}^{L} W_h \hat{p}_h$
Variancia del promedio estratificada	**Variancia de la proporción estratificada**
$Var(\overline{y}_{st}) = \sum_{y=1}^{l} W_h^2 \cdot Var\left(\overline{y}_h\right)$	$Var(\hat{p}_{st}) = \sum_{y=1}^{l} W_h^2 \cdot Var\left(\hat{p}_h\right)$
Tamaño de la muestra para la estimación de la media de la población	**Tamaño de la muestra para proporciones**
$n = \dfrac{\sum \dfrac{W_h^2 s_h^2}{w_h}}{V + \dfrac{1}{N} \sum \dfrac{W_h^2 s_h^2}{w_h}}$ $V = \dfrac{1}{n} \sum \dfrac{W_h^2 s_h^2}{w_h} - \dfrac{1}{N} \sum W_h s_h^2$	**Proporcional:** $n = \dfrac{n_0}{1 + \dfrac{n_0}{N}}$ donde $n_0 = \dfrac{\sum W_h p_h q_h}{V}$ **Optimo supuesto**: $n = \dfrac{n_0}{1 + \dfrac{1}{NV} \sum W_h p_h q_h}$ $n_0 = \dfrac{\left(\sum W_h \sqrt{p_h q_h} \right)^2}{V}$

MUESTREO POR CONGLOMERADOS

Estimación del promedio	Estimación de una proporción
$\overline{y} = \dfrac{\sum_{i=1}^{A} y_i}{\sum_{i=1}^{A} m_i}$	$\hat{p} = \dfrac{\sum_{i=1}^{A} a_i}{\sum_{i=1}^{A} m_i}$

MODELOS DE CRECIMIENTO

Modelo aritmético	**Modelo geométrico**
$N_t = N_0(1+rt)$ $r = \dfrac{1}{t} \cdot \dfrac{N_t - N_0}{N_0}$	$N_t = N_0(1+r)^t$ $r = \left(\dfrac{N_t}{N_o}\right)^{1/t} - 1$

Modelo exponencial

$$N_t = N_0 e^{rt} \qquad r = \frac{1}{t} \ln\left(\frac{N_t}{N_0}\right)$$

DISTRIBUCIONES DE PROBABILIDADES

Distribución binomial	Distribución de Poisson
$f(x) = \binom{n}{x} p^x q^{n-x}$ x=0, 1,..., n	$f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ x=0, 1,...
Distribución hipergeométrica	Distribución geométrica
$f(x) = \dfrac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$ x=0, 1, 2,..., min(n,D)	$f(x) = q^{x-1} p$ x=1, 2,...

TEOREMA DE BAYES

$$P(A/D) = \frac{P(A)P(D/A)}{P(A)P(D/A) + P(B)P(D/A)}$$

TECNICAS DE CONTEO

Permutaciones	Combinaciones
$n\Pr = \dfrac{n!}{(n-x)!}$	$nCr = \dfrac{n!}{r!(n-x)!}$

ANALISIS DE VARIANCIA A UNA VIA: DISEÑO COMPLETAMENTE ALEATORIZADO

Media total $$\overset{=}{x} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}}{n}$$	Suma de cuadrados total $$SCT = \sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij} - \overset{=}{x})^2$$
Suma de cuadrados de los tratamientos $$SCTR = \sum_{j=1}^{c} r_j(\bar{x}_j - \overset{=}{x})^2$$	Suma del cuadrado de error $$SCE = \sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij} - \bar{x}_j)^2$$

Prueba para diferencias entre pares de medias

Diseños balanceados	Diseños no balanceados
Criterio de Tukey $$T = q_{\alpha,c,n-c}\sqrt{\frac{CME}{R}}$$ Diferencia mínima significativa $$DMS = \sqrt{\frac{2(CME)F_{\alpha,1,n-c}}{r}}$$	Diferencia mínima significativa $$DMS_{J,K} = \sqrt{\left[\frac{1}{r_j} + \frac{1}{r_k}\right](CME)F_{\alpha,c-1,n-c}}$$

ANALISIS DE VARIANCIA A DOS VÍAS: DISEÑO ALEATORIZADO EN BLOQUES

Suma de cuadrados de bloques $$SCBL = \sum_{i=1}^{r} c_i(\bar{x}_i - \overset{=}{x})^2$$	Suma de cuadrados del error $$SCE = SCT - SCTR - SCBL$$

PRUEBAS NO PARAMETRICAS

Prueba U de Mann-Whitney	Media y desviación estándar de la distribución muestral para la prueba U de Mann-Whitney
$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1$ $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$	$\mu_u = \frac{n_1 n_2}{2}$ $\sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$
Valor Z para normalizar la prueba U de Mann-Whitney	Prueba de independencia Chi-Cuadrada
$Z = \frac{U_i - \mu_u}{\sigma_u}$	$\chi^2_{obs} = \sum_{i=1}^{rc} \frac{(O_i - E_i)^2}{E_i}$
Coeficiente de correlación de Spearman	Desviación normal para la prueba de rangos de Spearman
$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$	$Z = r_s\sqrt{n-1}$

Prueba de Kruskal-Wallis

$$K = \frac{12}{n(n+1)}\left[\sum \frac{R_i^2}{n_i}\right] - 3(n+1)$$

Valor crítico para la prueba de Kruskal-Wallis

$$C_k = \sqrt{\chi^2_{\alpha,k-1}\left[\frac{n(n+1)}{12}\right]\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}$$