3) Inferencia en poblaciones normales

3.1) Inferencia sobre μ con $\sigma 2$ conocido

Sea $\tilde{x} = N(\mu;\sigma^2)$ $\Rightarrow$ $\bar{x} = N\left(\mu;\dfrac{\sigma^2}{n}\right)$

luego $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = N(0;1)$

3.2) Inferencia sobre μ con σ^2 desconocido

Sea $\tilde{x} = N(\mu;\sigma^2)$

Puede estimarse σ^2 a partir de la muestra, mediante $s^2 = \dfrac{1}{\nu}\sum_{i=1}^{N}(x_i-\bar{x})^2$, con $\nu = n-1$ grados de libertad.

Estadístico: $t_\upsilon = \dfrac{\bar{x}-\mu`}{s/\sqrt{n}}$ (tiene distribución t-Student)

3.3) Inferencia sobre σ^2

Estadístico $\chi^2 = \dfrac{\nu.s^2}{\sigma^2}$ (tiene distribución ji-cuadrado con $\nu = n-1$ grados de libertad)

Cuando se tienen m muestras se amalgaman las varianzas:

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{x_{ij}-\bar{x}_i}{\sigma}\right)^2 = \sum_{i=1}^{n}\nu_i.\frac{s_i^2}{\sigma^2}$$

Tomando la varianza amalgamada $s^2 = \dfrac{1}{\nu}\sum_{i=1}^{n}\nu_i.s_i^2$ con $\nu = \sum \nu_i$ grados de libertad queda $\chi^2 = \nu.\dfrac{s^2}{\sigma^2}$

3.4) Comparación de poblaciones normales

3.4.1) Comparación de varianzas

Sean $\tilde{x}_1 = N(\mu;\sigma^2)$ y $\tilde{x}_2 = N(\mu;\sigma^2)$ independientes

Estadístico: $F_{\nu_1,\nu_2} = \dfrac{1}{\varphi^2}.\dfrac{s_1^2}{s_2^2}$ (tiene distribución F-Snedecor). Donde $\varphi^2 = \dfrac{\sigma_1^2}{\sigma_2^2}$ con ν_1 =g.l. del numerador y ν_2 = g.l. del denominador

3.4.2) Comparación de medias

Sean $\tilde{\ }_1 = \ (\mu;\sigma^2)$ y $\tilde{x}_2 = N(\mu;\sigma^2)$ independientes. Interesa inferir sobre el valor de $\delta = \mu_1 - \mu_2$

Estadístico: $d = \bar{x}_1 - \bar{x}_2$

3.4.2.1) Muestra no apareada con varianzas conocidas

Como d es una combinación lineal de $\bar{x}_1$ y $\bar{x}_2$

$$u = \frac{d-\delta}{\sigma_\Delta} = N(0;1) \qquad \text{siendo} \qquad \begin{array}{l} d = \bar{x}_1 - \bar{x}_2 \\ \delta = \mu_1 - \mu_2 \end{array} \qquad \text{y} \qquad \sigma_\Delta^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

3.4.2.2) Muestra no apareada con varianzas desconocidas

- **Caso Homocedástico (clásico):** $\sigma_1^2 = \sigma_2^2$

Estadístico: $t_{\nu_\Delta} = \dfrac{d-\delta}{s/\sqrt{n_e}}$ siendo $\begin{array}{l} d = \bar{x}_1 - \bar{x}_2 \\ \delta = \mu_1 - \mu_2 \end{array}$

$$\nu_\Delta = n_1 + n_2 - 2 \qquad \frac{1}{n_e} = \frac{1}{n_1} + \frac{1}{n_2} \qquad s^2 = \frac{\nu_1 . s_1^2 + \nu_2 . s_2^2}{\nu_1 + \nu_2}$$

- **Caso Heterocedástico (Welch):** $\sigma_1^2 \neq \sigma_2^2$

Estadístico $t_\nu = \dfrac{d-\delta}{s}$ siendo $\begin{array}{l} d = \bar{x}_1 - \bar{x}_2 \\ \delta = \mu_1 - \mu_2 \end{array}$

$$s^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

$$\nu = \left[\frac{1}{\nu_1} . \left(\frac{n_2 . s_1^2}{n_2 . s_1^2 + n_1 . s_2^2} \right)^2 + \frac{1}{\nu_2} . \left(\frac{n_1 . s_2^2}{n_2 . s_1^2 + n_1 . s_2^2} \right)^2 \right]^{-1}$$

3.4.2.3) Muestra apareada

Las variables son dependientes. Puede considerarse normal la diferencia entre las mismas.

Sean $\begin{array}{l} \tilde{x}_1 = N(\mu_1;\sigma_1^2) \\ \tilde{x}_2 = N(\mu_2;\sigma_2^2) \end{array}$ donde μ_1 y μ_2 varían dato a dato

manteniendo $\mu_1 - \mu_2 = c$ y $\tilde{d} = \tilde{x}_1 - \tilde{x}_2 = N(\delta;\sigma_d^2)$

Estadístico: $t_\nu = \dfrac{\bar{d} - \delta}{s_d / \sqrt{n}}$ con $\nu = n - 1$

3.5) Inferencia sobre ρ

Sea $(\tilde{x}; \tilde{y}) = N(\mu_x; \mu_y; \sigma_x^2; \sigma_y^2)$ Normal Bidimensional

Si ρ=0

Estadístico: $t_\nu = r.\sqrt{\dfrac{n-2}{1-r^2}}$ con $\nu = n - 2$

donde $r = \dfrac{s_{xy}}{s_x . s_y}$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y}) = \frac{1}{n-1}\sum_{i=1}^{n} x_i . y_i - n.\bar{x}.\bar{y}$$

Para inferir acerca de $\rho = \rho_0$

Estadístico (aproximación de Fisher) $z = \dfrac{1}{2}.\ln\left(\dfrac{1+r}{1-r}\right) \cong N(\mu_z; \sigma_z^2)$ con

$$\mu_z = \frac{1}{2}.\ln\left(\frac{1+\rho}{1-\rho}\right) \qquad \sigma_z^2 = \frac{1}{n-3}$$

4) *Regresión*

4.1) Regresión Lineal Simple

Modelo: $\tilde{y}_i = \beta_0 + \beta_1 . x_i + \tilde{\varepsilon}_i$ $\qquad i = 1,2,\ldots,n$

$\tilde{y}$: variable que se desea estudiar

x: variable que se utilizará para explicar a $\tilde{y}$

β_0 y β_1: parámetros de la función lineal que vincula a $\tilde{y}$ con x

$\tilde{\varepsilon}$: error (no observable)

4.1.1) Estimación de los parámetros

Función Predictora: $\hat{y} = b_0 + b_1 . x$

$$b_1 = \frac{\sum (x_i - \bar{x}).(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad b_0 = \bar{y} - b_1 . \bar{x}$$

Otra fórmula más sencilla (aunque aumenta los errores de redondeo):

$$_1 = \frac{\cdot\sum \ _i \cdot \ _i - \sum \ _i \sum \ _i}{\cdot\sum \ _i^{\,2} - (\sum \ _i)^2}$$

4.1.2) Coeficientes de Determinación y de Correlación

$$SC_{TOTAL} = SC_{RESIDUAL} + SC_{REGRESION}$$

Suma de Cuadrados Residual $Q = \sum \left(y_i - \hat{y}_i\right)^2 = \sum \left(y_i - b_0 - b_1.x_i\right)^2$

$$Q = \sum y_i^{\,2} - b_0 \sum y_i - b_1 \sum x_i.y_i$$

Coeficiente de Determinación: $R^2 = 1 - \dfrac{Q}{\sum (y_i - \bar{y})^2}$

Coeficiente de Correlación: $R = \dfrac{\sum x_i.y_i - n.\bar{x}.\bar{y}}{\sqrt{(\sum x_i^{\,2} - n.\bar{x}^2)(\sum y_i^{\,2} - n.\bar{y}^2)}}$

4.1.3) Estimación de la varianza residual

$$s^2 = \frac{Q}{n-2}$$

4.1.4) Inferencia sobre β_1

Estadístico $t_\nu = \dfrac{b_1 - \beta_1}{\hat{D}(b_1)}$ con $\nu = n - 2$

$$\hat{D}^2(b_1) = \frac{s^2}{\sum x_i^{\,2} - n.\bar{x}^2}$$

4.1.5) Inferencia sobre E $\tilde{y}/x_0 = \beta_0 + \beta_1.x_0$ (para Intervalos de confianza)

Estimador: $\tilde{y}/x_0 = b_0 + b_1.x_0$ (el punto de la recta para x_0)

Estadístico: $t_\nu = \dfrac{(b_0 + b_1 x_0) - (\beta_0 + \beta_1.x_0)}{s_{\tilde{y}/x_0}}$

donde $s^2_{\tilde{y}/x_0} = \left(\dfrac{1}{n} + \dfrac{\left(x_0 - \bar{x}\right)^2}{\sum x_i^{\,2} - n\bar{x}}\right).s^2$

$$\nu = n - 2$$

La inferencia sobre β_0 (ordenada al origen) es un caso particular, basta usar las fórmulas anteriores con $x_0=0$

4.1.6)) Inferencia sobre $\tilde{y}/x_0 = \beta_0 + \beta_1.x_0 + \tilde{\varepsilon}$ (para Intervalos de predicción)

Por ser una variable aleatoria se la predice o pronostica

Predictor: $\hat{y}/x_0 = b_0 + b_1.x_0$

Estadístico: $t_\nu = \dfrac{(b_0 + b_1 x_0) - (\beta_0 + \beta_1.x_0 + \tilde{\varepsilon})}{s_{\hat{y}/x_0 - \tilde{y}/x_0}}$

con $s^2_{\hat{y}/x_0 - \tilde{y}/x_0} = \left(1 + \dfrac{1}{n} + \dfrac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}}\right).s^2$

$\nu = n - 2$

4.1.7) Regresión con intercepto conocido

Modelo: $\tilde{z}_i = \tilde{y}_i - \beta_0 = \beta_1 x_i + \tilde{\varepsilon}$

La ecuación predictora es $\hat{z} = b_1 x$

Coeficiente de regresión: $b_1 = \dfrac{\sum x_i.z_i}{\sum x_i^2}$

Estimación de la varianza residual: $s^2 = \dfrac{Q}{n-1}$

grados de libertad $\nu = n - 1$

La suma de cuadrados de los residuos $Q = \sum (z_i - b_1.x_i) = \sum z_i^2 - \dfrac{(\sum x_i.z_i)^2}{\sum x_i^2}$

Coeficiente de correlación: $R = \sqrt{1 - \dfrac{Q}{\sum z_i^2}} = \dfrac{\sum x_i.z_i}{\sqrt{\sum x_i^2.\sum z_i^2}}$

Varianza del estimador b_1: $\hat{D}^2(b_1) = \dfrac{s^2}{\sum x_i^2}$

Dado un valor x_0 de $\tilde{x}$, la varianza de $\hat{z}_0$ es: $\hat{D}^2(\hat{z}_0) = s^2.\dfrac{x_0}{\sum x_i^2}$

4.2) Regresión Múltiple

Siendo:

$$\tilde{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & x_{21} & & x_{k1} \\ 1 & x_{12} & x_{22} & & x_{k2} \\ \\ 1 & x_{1n} & x_{2n} & & x_{kn} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \\ \beta_k \end{bmatrix} \qquad \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_n \end{bmatrix}$$

n x 1   n x p   p x 1   n x 1

p = k + 1

Modelo: $\tilde{Y} = X.\beta + \tilde{\varepsilon}$
Cantidad de variables k
Esta función toma en el dato X el valor $\tilde{y}_i = \beta_0 + \beta_1.x_{1i} + \beta_2.x_{2i} + \ldots + \beta_n.x_{ni} + \tilde{\varepsilon}_i$

La ecuación predictora es $\hat{Y}_0 = x_0.b$

4.2.2) Coeficientes de Determinación y de Correlación

$$SC_{TOTAl} = SC_{RESIDUAL} + SC_{REGRESION} = Q + SC_{REGRESION}$$

Coeficiente de Determinación: $R^2 = \dfrac{SC_{REGRESION}}{SC_{TOTAL}} = 1 - \dfrac{Q}{\sum (y_i - \bar{y})^2}$

Coeficiente de Determinación Ajustado :

$$R^2 = 1 - \left(\frac{n-1}{n-p}\right).\frac{Q}{\sum y_i^{\,2}} = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-p}\right)$$

4.2.3) Desvío Std. residual

$$s_{RES} = \sqrt{\frac{SC_{RESIDUAL}}{n-p}}$$

Estadístico: $\chi^2{}_\nu = \dfrac{n-p}{\sigma^2}.s^2$ con $\nu = n - p$

4.1.4) Inferencia sobre β

Matriz de covarianza de los estimadores $\left\|\text{cov}(b_i, b_j)\right\| = \sigma^2.(X'.X)^{-1}$, se designa con c_{ij} a; elemento ubicado en la fila I y la columna j de la matriz $(X'.X)^{-1}$

Estadístico $t_\nu = \dfrac{b_i - \beta_i}{\hat{D}(b_i)}$ con

4.1.5) Intervalo de confianza

Estadístico: $t_\nu = \dfrac{.x'_o.b - .x'_0.\beta}{s..\sqrt{x_0'.(X'.X)^{-1}.x_0}}$ con $\nu = n - p$

4.1.6)) Intervalo de Predicción

Estadístico: $t_\nu = \dfrac{x'_o.b - (x_0.\beta + \tilde{\varepsilon})}{s..\sqrt{1 + x'_0\,'.(X'.X)^{-1}.x_0}}$ con $\nu = n - p$