Bioestadística

Fórmulas y relaciones para cálculos básicos

1. *Fórmulas de estadística descriptiva*
2. *Propiedades y teoremas básicos de probabilidad. Modelos de probabilidad*
3. *Inferencia estadística*

©Andreu Nolasco Bonmatí
Departamento de Enfermería Comunitaria,
Medicina preventiva y Salud Pública e Historia de la Ciencia
Universidad de Alicante

1. Fórmulas de estadística descriptiva

Tipo de datos	Medidas de tendencia central	Medidas de dispersión	Medidas de forma
No agrupados	**Media** = $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ **Mediana** = Md = Valor de la observación que ocupa la posición o rango $r_{Md} = \frac{n+1}{2}$ En caso de que r_{Md} no sea entero, Md se calcula como la semisuma de los valores anterior y posterior **Moda** = Mo = Valor de la variable con mayor frecuencia	**Rango** = R = $x_{max} - x_{min}$ **Varianza** = $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$ **Desviación típica o estándar** = $s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$ **Coeficiente de Variación** = $CV = \frac{s}{\bar{x}}(x100)$ **Percentil de orden q** = Valor de la variable con rango o posición $r_q = \frac{q}{100}(n+1)$, una vez ordenadas de menor a mayor las observaciones de la variable, calculado a través del promedio ponderado entre los valores que ocupen los rangos anterior y posterior: $p_q = (1-f)\, x_i + f\, x_{i+1}$ con f parte fraccionaria de r_q	**Coeficiente de asimetría:** $As = \frac{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n}}{s^3}$ **Coeficiente de curtosis** $Cu = \frac{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{n}}{s^4}$
Agrupados	**Media**= $\bar{x} = \frac{\sum_{i=1}^{k} xm_i f_i}{n}$ **Mediana =** $Md = x_i + \left(\frac{n/2 - F_{i-1}}{F_i - F_{i-1}}\right) a_i$ **Moda** = **Intervalo modal** = Intervalo para el que f_i es máxima	**Rango** = $R = xm_{max} - xm_{min}$ **Varianza** = $s^2 = \frac{\sum_{i=1}^{k} (xm_i - \bar{x})^2 f_i}{n}$ **Desviación típica** = $s = \sqrt{s^2}$ **Coeficiente de variación** = $CV = \frac{s}{\bar{x}}(x100)$ **Percentil de orden q** = $P_q = x_i + \left(\frac{\frac{qn}{100} - F_{i-1}}{F_i - F_{i-1}}\right) a_i$	**Coeficiente de asimetría** = $As = \frac{\frac{\sum_{i=1}^{k} (xm_i - \bar{x})^3 f_i}{n}}{s^3}$ **Coeficiente de curtosis** = $Cu = \frac{\frac{\sum_{i=1}^{k} (xm_i - \bar{x})^4 f_i}{n}}{s^4}$

	Regresión y correlación lineal
X, Y variables cuantitativas continuas $\{ \mathbf{x_i}, \mathbf{y_i} \}_{\mathbf{i=1}}^{\mathbf{n}}$ **muestra de n observaciones**	**Coeficiente de correlación lineal de Pearson** $r = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \dfrac{n\sum_{i=1}^{n} x_i\, y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}}$ **Modelo de regresión lineal simple** $\hat{y} = \hat{\alpha} + \hat{\beta}\, x$ $\hat{\beta} = \dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$ $\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$

2. Propiedades y teoremas básicos de probabilidad. Modelos de probabilidad

Tipo	Relación
Dados A, B sucesos cualesquiera, Ω suceso seguro, Φ suceso imposible **AXIOMAS DE KOLMOGOROV**	✒ $0 \le p(A) \le 1$ ✒ $p(\Omega) = 1$ ✒ $p(AUB) = p(A)+p(B)$ si $A \cap B = \Phi$
Dados A, B sucesos cualesquiera, Ω suceso seguro, Φ suceso imposible, $\overline{A}$ = suceso complementario de A **PROPIEDADES BASICAS**	✒ $p(\overline{A}) = 1 - p(A)$ ✒ $p(\Phi) = 0$ ✒ $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ ✒ $p(\overline{A \cup B}) = p(\overline{A} \cap \overline{B})$ $p(\overline{A \cap B}) = p(\overline{A} \cup \overline{B})$
Dados A, B sucesos cualesquiera **PROBABILIDAD CONDICIONAL E INDEPENDENCIA**	✒ $p(A / B) = \dfrac{p(A \cap B)}{p(B)}$ $p(B / A) = \dfrac{p(A \cap B)}{p(A)}$ ✒ $p(A / B) = p(A)$ $\Rightarrow$ A y B son independientes ✒ $p(A \cap B) = p(A) \cdot p(B)$ $\Leftrightarrow$ A y B son independientes
PARTICIÓN	✒ Familia de sucesos $\{A_i\}_{i=1}^{n}$ tal que: $A_i \cap A_j = \Phi \quad \forall i, j$ $\bigcup_{i=1}^{n} A_i = \Omega$
Dada una partición $\{A_i\}_{i=1}^{n}$, y X otro suceso cualquiera **TEOREMAS BÁSICOS**	✒ Teorema de la probabilidad total: $p(X) = \sum_{i=1}^{n} p(X / A_i)\, p(A_i)$ ✒ Teorema de Bayes: $p(A_i / X) = \dfrac{p(X / A_i)\, p(A_i)}{p(X)}$

Tipo	Relación
Modelos discretos de probabilidad	**Modelo binomial** $p(k) = p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ media o esperanza = E(X) = np varianza = V(X) = np(1-p) **Modelo Poisson** $p(X = k) = \frac{\lambda^k e^{-\lambda}}{k!};\quad k = 0, 1, 2, 3, ...$ media o esperanza = E(X) = λ = V(X) = varianza
Modelos continuos de probabilidad	Modelo normal (ver tablas de probabilidad) Modelo t de student (ver tablas de probabilidad) Modelo F de Snedecor (ver tablas de probabilidad) Modelo Ji.cuadrado (ver tablas de probabilidad)

3. Inferencia estadística

INTERVALOS DE CONFIANZA PARA LOS PARAMETROS MAS HABITUALES

PARÁMETRO (θ)	**INTERVALO DE CONFIANZA DE NIVEL 1 - α** **$I_{1-\alpha}(\theta)$=[Estimador$\pm$Coeficiente$\times$Error Estándar]**	**DISTRIBUCIÓN MUESTRAL**	**REQUERIMIENTOS**
Media μ	$I_{1-\alpha}(\mu)=\left[\bar{x}\pm t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}}\right]$	t de Student con n-1 grados de libertad	Normalidad de la variable a estudio o $n\geq30$
Proporción **p**	$I_{1-\alpha}(p)=\left[\hat{p}\pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$	Aproximadamente normal	$n\hat{p}\geq5$ $n(1-\hat{p})\geq5$
Varianza σ^2	$I_{1-\alpha}(\sigma^2)=[L_{inf};L_{sup}]$ $L_{inf}=\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$ $L_{sup}=\frac{(n-1)s^2}{\chi^2_{\alpha/2}}$	Basada en la Ji-cuadrado, $\chi^2_{\alpha/2}$; $\chi^2_{1-\alpha/2}$ percentiles de una Ji-cuadrado con n–1 grados de libertad	Normalidad de la variable a estudio
Diferencia de Medias $\mu_1 - \mu_2$	1. Con varianzas desconocidas pero iguales $I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}_1-\bar{x}_2)\pm t^{n_1+n_2-2}_{1-\alpha/2}\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}\right]$ 2. Con varianzas desconocidas y diferentes $I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}_1-\bar{x}_2)\pm t^{gl}_{1-\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right]$ $s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ $gl=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1}+\frac{(s_2^2/n_2)^2}{n_2}}$	1. t de Student con n_1+n_2-2 grados de libertad 2. t de Student con gl grados de libertad	Normalidad de la variable a estudio o $n_1\geq30$ y $n_2\geq30$
Diferencia de Proporciones p_1-p_2	$I_{1-\alpha}(p_1-p_2)=\left[(\hat{p}_1-\hat{p}_2)\pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_1}}\right]$	Aproximadamente normal	$n=n_1+n_2\geq20$ Si $n=n_1+n_2$, está entre 20 y 40, $n_1\hat{p}_1\geq5$ $n_1(1-\hat{p}_1)\geq5$ y $n_2\hat{p}_2\geq5$ $n_2(1-\hat{p}_2)\geq5$

CONTRASTES DE HIPOTESIS PARA LOS PARAMETROS MAS HABITUALES

PARÁMETRO (θ)	ESTADÍSTICO DE CONTRASTE	DISTRIBUCIÓN MUESTRAL	REQUERIMIENTOS
Media μ $H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	t de Student con n-1 grados de libertad	Normalidad de la variable a estudio o $n \geq 30$
Proporción p $H_0: p = p_0$ $H_a: p \neq p_0$	$z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	Aproximadamente normal	$n\hat{p} \geq 5$ $n(1-\hat{p}) \geq 5$
Comparación o dferencia de Medias $\mu_1 - \mu_2$ $H_0: \mu_1 = \mu_2$ $H_a: \mu_1 \neq \mu_2$	1. <u>Con varianzas desconocidas pero iguales</u> $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$ 2. <u>Con varianzas desconocidas y diferentes</u> $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ $\quad gl = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1} + \dfrac{(s_2^2/n_2)^2}{n_2}}$	1. t de Student con $n_1 + n_2 - 2$ grados de libertad 2. t de Student con gl grados de libertad	Normalidad de la variable a estudio o $n_1 \geq 30$ y $n_2 \geq 30$
Comparación o Diferencia de Proporciones $p_1 - p_2$ $H_0: p_1 = p_2$ $H_a: p_1 \neq p_2$	$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ $\hat{p}_1 = \dfrac{r_1}{n_1}$ $\quad \hat{p}_2 = \dfrac{r_2}{n_2}$ $\quad \hat{p} = \dfrac{r_1 + r_2}{n_1 + n_2}$	Aproximadamente normal	$n = n_1 + n_2 \geq 20$ Si $n = n_1 + n_2$, está entre 20 y 40, $n_1\hat{p}_1 \geq 5$ $\; n_1(1-\hat{p}_1) \geq 5$ y $n_2\hat{p}_2 \geq 5$ $\; n_2(1-\hat{p}_2) \geq 5$
Comparación de varianzas $H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \neq \sigma_2^2$	$F = \dfrac{s_1^2}{s_2^2}$ Se pone en el numerador la mayor de las varianzas	F de Snedecor con n_1-1 y n_2-1 grados de libertad	Normalidad de la variable a estudio
Asociación entre variables cualitativas H_0: No asociación H_a: Asociación	$\chi^2 = \sum \dfrac{(o_i - e_i)^2}{e_i}$ o_i y e_i frecuencias observadas y esperadas $e_i = \dfrac{(\text{Total fila}) \times (\text{Total columna})}{n}$	Ji-cuadrado con (n° filas-1)x (n° columnas-1) grados de libertad	$e_i \geq 5 \quad \forall i$ Si la tabla tiene dimensión distinta a la 2x2, $1 \leq e_i < 5$ en el 20% de las celdas, como máximo