Formulas y relaciones para cálculos básicos

1. ***Fórmulas de estadística descriptiva***

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

3. ***Inferencia estadística***

José Aurelio Pina Romero
Inmaculada Melchor
Registro de Mortalidad
Conselleria de Sanitat

1. Formulas 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** = M_0 = 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 variables con rango o posición $rq = \frac{q}{100}(n+1)$ una vez ordenadas de menor a mayor las observaciones de la variable, calculado a través del promedios 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:** $\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n}$ x_i = marca de clase **Mediana** = $Md = L_{i-1} + \left(\frac{n/2 - F_{i-1}}{F_i - F_{i-1}}\right) a_i$ **Moda = Intervalo modal** = Intervalo para el que f_i es máxima	**Rango** = R = $x_{max} - x_{min}$ **Varianza** = $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i}{n}$ **Desviación típica o estándar** = $s = \sqrt{s^2}$ **Coeficiente de Variación** = $CV = \frac{s}{\bar{x}}(x100)$ **Percentil de orden q** = $P_q = L_{i-1} + \left(\frac{\frac{qn}{100} - F_{i-1}}{F_i - F_{i-1}}\right) a_i$	**Coeficiente de asimetría:** $As = \frac{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^3 f_i}{n}}{s^3}$ **Coeficiente de curtosis:** $Cu = \frac{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^4 f_i}{n}}{s^4}$

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 KOLGOMOROV**	▪ $0 \leq p(A) \leq 1$ ▪ $p(\Omega) = 1$ ▪ $p(A \cup B) = p(A) + p(B) \quad \text{si } A \cap B = \Phi$
Dados A, B sucesos cualesquiera, Ω suceso seguro, Φ suceso imposible **PROPIEDAS BÁSICAS**	▪ $p(\overline{A}) = 1 - p(A)$ ▪ $p(\Phi) = 0$ ▪ $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ ▪ $p(\overline{A \cup B}) = p(\overline{A} \cap \overline{B})$ $p(\overline{A \cap B}) = p(\overline{A} \cup \overline{B})$
PROBABILIDAD CONDICIONAL E INDEPENDENCIA	▪ $p(A/B) = \dfrac{P(A \cap B)}{p(B)} \quad p(B/A) = \dfrac{P(B \cap A)}{p(A)}$ ▪ $p(A/B) = P(A) \Rightarrow$ A y B son independientes ▪ $p(A \cap B) = P(A) \cdot P(B) \Leftrightarrow$ A y B son independientes
TEOREMAS BÁSICOS	▪ Teorema de la probabilidad total: $p(X) = \sum_{i=1}^{n} p(X / A_i) p(A_i)$ ▪ Teorema de Bayes: $p(A_i / X) = \dfrac{p(X / A_i) p(A_i)}{p(X)}$

3. *Inferencia estadística*

INTERVALOS DE CONFIANZA PARA LOS PARÁMETROS MAS HABITUALES

PARÁMETRO (θ)	INTERVALOS DE CONFIANZA DE NIVEL 1-α $I_{1-\alpha}(\mu)=[Estimador \pm Coeficiente \cdot Error$	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 ≥ 30
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	$n\hat{p} \geq 5$ $n(1-\hat{p}) \geq 5$
Varianza σ^2	$I_{1-\alpha}(\sigma^2)=[L_{\inf};L_{\sup}]$ $L_{\inf}=\frac{(n-1)s^2}{X^2_{1-\alpha/2}}$ $L_{\inf}=\frac{(n-1)s^2}{X^2_{\alpha/2}}$	Basada en la Ji-cuadrado $X^2_{1-\alpha/2}$; $X^2_{\alpha/2}$ percentiles de una Ji-cuadrado con n-1 grados de libertad	Normalidad de la variable a estudio
Diferencias de medias $\mu_1-\mu_2$	1. **Con varianzas desconocidas pero iguales** $I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}_1-\bar{x}_2) \pm t^{n_1+n_2-2}_{1-\alpha/2}\sqrt{\frac{s^2_p}{n_1}+\frac{s^2_p}{n_2}}\right]$ **2. Con varianzas desconocidas y diferentes** $I_{1-\alpha}(\mu_1-\mu_2)=\left[(\bar{x}_1-\bar{x}_2) \pm t^{gl}_{1-\alpha/2}\sqrt{\frac{s^2_p}{n_1}+\frac{s^2_p}{n_2}}\right]$ $s^2_p=\frac{(n_1-1)s^2_1+(n_2-1)s^2_2}{n_1+n_2-2}$ $gl=\frac{\left(\frac{s^2_1}{n_1}+\frac{s^2_2}{n_2}\right)^2}{\frac{(s^2_1/n_1)^2}{n_1}+\frac{(s^2_2/n_2)^2}{n_2}}$	**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 $n_1 \geq 30$ $n_2 \geq 30$
Diferencia de	Z*	Aproximadamente	$n = n_1+n_2 \geq 20$

proporciones **p_1-p_2**		normal	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 $n_2 \hat{p}_2 \geq 5$ $n_2(1-\hat{p}_2) \geq 5$

$$Z * = I_{1-\alpha}(p_1 - p_2) = \left[(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}} \right]$$

CONTRASTES DE HIPOTESIS PARA LOS PARÁMETROS 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 ≥ 30
Proporción **p** $H_0 : p = p_0$ $H_a : p \neq p_0$	$t = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	Aproximadamente normal	$n\hat{p} \geq 5$ $n(1-\hat{p}) \geq 5$
Comparación o diferencias de media $\mu_1 - \mu_2$ $H_0 : \mu_1 = \mu_2$ $H_a : \mu_1 \neq \mu_2$	**1.** Con varianzas desconocidas pero iguales $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$ **2.** Con varianzas desconocidas y diferentes $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ $gl = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1} + \dfrac{(s_2^2/n_2)^2}{n_2}}$	**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 $n_1 \geq 30$ $n_2 \geq 30$
Comparación o Diferencia de proporcionespp $\mathbf{p_1}$**-**$\mathbf{p_2}$ $H_0 : p_1 = p_2$ $H_a : p_1 \neq p_2$	$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ $\hat{p}_1 = \dfrac{r_1}{n_1}$ $\hat{p}_2 = \dfrac{r_2}{n_2}$ $\hat{p} = \dfrac{r_1 + r_2}{n_1 + n_2}$	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 $n_2\hat{p}_2 \geq 5$ $n_2(1-\hat{p}_2) \geq 5$
Asociación entre variables cualitativas $H_0 : No asociación$ $H_a : Asociación$	$X^2 = \sum \frac{(o_i - e_i)^2}{e_i}$ o_i y e_i frecuencias observadas y esperadas $e_i = \frac{(Total fila) \times (Total Columna)}{n}$	Ji-cuadrado con (nº filas-1) x(nº de columnas -1) grados de libertad	$e_i \geq 5 \forall i$ Si la tabla tiene dimensión distinta a la 2x2, $1 \leq e_i \geq 5$ en el 20% de las celdas, como máximo