Parámetros muestrales	
Media (*Average*): $\overline{X} = \dfrac{\sum_{i=1}^{N} x_i}{N}$	**Mediana o C_2** (*Median*): Si N es impar $\Rightarrow$ valor de la posición (N+1)/2 Si N es par $\Rightarrow$ media de los valores que ocupan las posiciones N/2 y (N/2+1)
Cuartiles:	
C_1 (*Lower quartile*) es primer cuartil si: Nº datos $\leq C_1$ es mayor o igual que N/4 Nº datos $\geq C_1$ es mayor o igual que 3N/4	**C_3** (*Upper quartile*)es tercer cuartil si: Nº datos $\leq C_3$ es mayor o igual que 3N/4 Nº datos$\geq C_3$ es mayor o igual que N/4
Varianza (*Variance*): $S^2 = \sum_{i=1}^{N} \dfrac{(x_i - \overline{X})^2}{N-1}$	**Desviación típica** (*Standar deviation*): $S = \sqrt{S^2}$

Recorrido (*Range*): $R = X_{max} - X_{min}$	**Recorrido Intercuartílico** (*Interquartile range*): $RI = C_3 - C_1$	**Coeficiente de variación** (*Coeff. of variation*): $CV = \dfrac{S}{\overline{X}}$

Coeficiente de asimetría: $CA = \dfrac{\sum_{i=1}^{N}(x_i - \overline{x})^3 / (N-1)}{S^3}$	**Coef. asimetría estandarizado o** *Stnd. Skewness* **(CAE):** Si CAE < -2 $\Rightarrow$ distribución asimétrica negativa Si CAE $\in$ [-2, 2] $\Rightarrow$ distribución simétrica Si CAE > 2 $\Rightarrow$ distribución asimétrica positiva
Coeficiente de curtosis (apuntamiento): $CC = \dfrac{\sum_{i=1}^{N}(x_i - \overline{x})^4 / (N-1)}{s^4} - 3$	**Coef. curtosis estandarizado o** *Stnd. kurtosis* **(CCE):** Si CCE < -2 $\Rightarrow$ datos planicúrticos Si CCE $\in$ [-2, 2] $\Rightarrow$ datos mesocúrticos (“normales”) Si CCE > 2 $\Rightarrow$ datos leptocúrticos
Covarianza (*Covariance*): $cov_{xy} = \dfrac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{N-1}$	**Coeficiente de correlación lineal** (*Correlation Coefficient*): $r_{xy} = \dfrac{cov_{xy}}{S_x S_y}$ $\qquad r_{xy} \in [-1, +1]$

Probabilidad		
Propiedades:		
$P(A) \geq 0$	$P(E) = 1$	Si **A** y **B** son **excluyentes** $\Rightarrow$ $P(A \cup B) = P(A) + P(B)$ y $P(A \cap B) = \varnothing$
$P(\overline{A}) = 1 - P(A)$	$P(A) \leq 1$	$P(\varnothing) = 0$
Regla de Laplace: $P(A) = \dfrac{\text{Casos favorables}}{\text{Casos posibles}}$	**Leyes de Morgan:** $\overline{A \cup B} = \overline{A} \cap \overline{B}$	$\overline{A \cap B} = \overline{A} \cup \overline{B}$

Suma de sucesos:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$

En general:

$$P(A_1\cup\ldots\cup A_n)=\sum(P(A_i))-\sum\left(P(A_i\cap A_j)\right)+\sum(P(A_i\cap A_j\cap A_k))+\ldots+(-1)^{n+1}\left(\sum(P(A_1\ldots A_n))\right)$$

Probabilidad condicional: $P(A/B)=\dfrac{P(A\cap B)}{P(B)}$	**Producto de sucesos:** $P(A\cap B)=P(A).P(B/A)$ $P(A\cap B)=P(B).P(A/B)$	Si **A** y **B** son **independientes** $\Rightarrow$ $P(A\cap B)=P(A).P(B)$

Teorema de la probabilidad total: $P(B)=\sum_{j=1}^{n}P(A_j)P(B/A_j)=P(A_1)P(B/A_1)+\ldots+P(A_n)P(B/A_n)$	**Teorema de Bayes** $P(A_i/B)=\dfrac{P(A_i\cap B)}{P(B)}=\dfrac{P(A_i)P(B/A_i)}{\sum_{j=1}^{n}P(A_j)P(B/A_j)}$

Distribuciones de probabilidad

Función de distribución: $F(x)=P(X\le x)$	**Propiedad:** $P(a<X\le b)=F(b)-F(a)$
Variables aleatorias discretas	**Variables aleatorias continuas**
Función de probabilidad: $P(X=x_i)$	**Función de densidad:** $f(x)=\dfrac{dF(x)}{dx}$

Esperanza matemática y parámetros poblacionales

Media: $m=E(X)$	**Varianza:** $\sigma^2=E(X-m)^2$	Desviación típica: $\sigma=\sqrt{\sigma^2}$

Propiedades de la media:

Si $Y=a_0\pm a_1\cdot X_1\pm a_2\cdot X_2\pm\ldots\pm a_n\cdot X_n \Rightarrow m_Y=a_0\pm a_1\cdot m_{X1}\pm a_2\cdot m_{X2}\pm\ldots\pm a_n\cdot m_n$

Casos particulares:

Si $Y=a+b\cdot X \Rightarrow m_Y=a+b\cdot m_X$	Si $Y=a-b\cdot X \Rightarrow m_Y=a-b\cdot m_X$
Si $Y=X_1+X_2 \Rightarrow m_Y=m_{X1}+m_{X2}$	Si $Y=X_1-X_2 \Rightarrow m_Y=m_{X1}-m_{X2}$

Propiedades de la varianza:

Si $Y=a_0\pm a_1\cdot X_1\pm a_2\cdot X_2 \Rightarrow \sigma_Y^2=a_1^2.\sigma_{X_1}^2\pm a_2^2.\sigma_{X_2}^2\pm 2.a_1.a_2.Cov_{X_1X_2}$

Casos particulares:

Si $Y=a+b\cdot X \Rightarrow \sigma_Y^2=b^2.\sigma_X^2$	Si $Y=a-b\cdot X \Rightarrow \sigma_Y^2=b^2.\sigma_X^2$

Si X_1 y X_2 son **independientes**: $Y=X_1\pm X_2 \Rightarrow \sigma_Y^2=\sigma_{X_1}^2+\sigma_{X_2}^2$

Coeficiente de variación: $CV_X=\dfrac{\sigma_X}{m_X}$	**Recorrido Intercuartílico:** C_3-C_1	**Recorrido:** $X_{max}-X_{min}$

Covarianza: $\sigma_{X_1X_2}^2=E\left(\left(X_1-m_{X_1}\right)\left(X_2-m_{X_2}\right)\right)$	**Coeficiente de correlación:** $\rho_{X_1X_2}=\dfrac{Cov_{X_1X_2}}{\sigma_{X_1}.\sigma_{X_2}}$

Distribuciones más importantes

Binomial: X~B(n, p) (X = 0, 1, ..., n)

Función de probabilidad: $P(X=x)=\binom{n}{x}.p^x.(1-p)^{n-x}$	$P(X \le x)=\sum_{x_i=0}^{x} P(X=x_i)$	**Media:** $m_X = E(X) = n.p$	**Varianza:** $\sigma_X^2 = n.p.(1-p)$

Propiedades:

$$\left.\begin{array}{l} X_1 \approx B(n_1, p) \\ \ldots\ldots\ldots\ldots\ldots \\ X_N \approx B(n_N, p) \end{array}\right. \Rightarrow Y = X_1 + \ldots + X_N \approx B(n_1 + \ldots + n_N, p)$$

Poisson: X ~ Ps (λ) (X = 0, 1, ..., ∞)

Función de probabilidad: $P(X=x)=e^{-\lambda}.\frac{\lambda^x}{x!}$	$P(X \le x)=\sum_{x_i=0}^{x} P(X=x_i) \Rightarrow$ Ábaco de Poisson	**Media:** $m_X = E(X) = \lambda$	**Varianza:** $\sigma_X^2 = \lambda$

Propiedades:

$$\left.\begin{array}{l} X_1 \approx Ps(\lambda_1) \\ \ldots\ldots\ldots\ldots\ldots \\ X_N \approx Ps(\lambda_N) \end{array}\right. \Rightarrow Y = X_1 + \ldots + X_N \approx Ps(\lambda_1 + \ldots + \lambda_N)$$

Uniforme: X ~U (a,b) (0 ≤ X ≤ ∞)

$P(X \le x) = \frac{x-a}{b-a} \quad a < x \le b$	**Media:** $m_X = E(X) = \frac{a+b}{2}$	**Varianza:** $\sigma_X^2 = \frac{(b-a)^2}{12}$

Exponencial: X ~ Exp (α) (0 ≤ X ≤ ∞)

$P(X \le x) = 1 - e^{-\alpha x} \quad x \ge 0 \Rightarrow$	**Media:** $m_X = E(X) = \frac{1}{\alpha}$	**Varianza:** $\sigma_X^2 = \frac{1}{\alpha^2}$

Normal: X ~ N(m, σ) (-∞ ≤ X ≤ ∞)

Z ~ Normal tipificada $P(Z \le z) \Rightarrow$ Tabla	**Media:** $m_X = E(X) = m$	**Varianza:** $\sigma_X^2 = \sigma^2$

Normal tipificada: $Z \sim N(m_Z = 0, \sigma_Z = 1)$ Si $X \sim N(m_X, \sigma_X) \Rightarrow Z = \frac{X - m_X}{\sigma_X}$

Propiedades:

$$\left.\begin{array}{l} X_1 \approx N(m_{X_1}, \sigma_{X_1}) \\ \ldots\ldots\ldots\ldots\ldots \\ X_N \approx N(m_{X_N}, \sigma_{X_N}) \end{array}\right. \Rightarrow Y = X_1 + \ldots + X_N \approx N\left(m_Y = m_{X_1} + \ldots + m_{X_N}; \sigma_Y = \sqrt{\sigma_{X_1}^2 + \ldots + \sigma_{X_N}^2}\right)$$

$$\sigma_Y^2 = \sigma_{X_1}^2 + \ldots + \sigma_{X_N}^2$$

Casos particulares

Si $Y = a + b \cdot X \Rightarrow Y \sim N(m_Y = a + b.m_X; \sigma_Y = \sqrt{b^2 \sigma_X^2})$	Si **X** ~ N($\mathbf{m_x}$, $\boldsymbol{\sigma_x}$) e **Y** ~ N($\mathbf{m_y}$, $\boldsymbol{\sigma_y}$) **independientes** $Z = X \mp Y \sim N(m_z = m_x \mp m_y, \sigma_z^2 = \sigma_x^2 + \sigma_y^2)$

Si $\mathbf{X} \sim N(\mathbf{m_x}, \boldsymbol{\sigma_x}) \Rightarrow \begin{cases} 68{,}26\% \text{ de los valores de } X \in [m-\sigma, m+\sigma] \\ 95{,}44\% \text{ de los valores de } X \in [m-2\cdot\sigma, m+2\cdot\sigma] \\ 99{,}73\% \text{ de los valores de } X \in [m-3\cdot\sigma, m+3\cdot\sigma] \end{cases}$

Aproximaciones normales

Teorema Central del Límite:

$$\left.\begin{array}{l} X_1 \sim g_1(m_{X_1}, \sigma^2_{X_1}) \\ \cdots\cdots\cdots \\ X_N \sim g_N(m_{X_N}, \sigma^2_{X_N}) \end{array}\right. \Rightarrow \quad Y = X_1 + \cdots + X_N \sim N(m_Y = m_{X_1} + \cdots + m_{X_N};\ \sigma^2_Y = \sigma^2_{X_1} + \cdots + \sigma^2_{X_N})$$

Siendo:

g $\rightarrow$ cualquier distribución (Binomial, Poisson, etc.)

N $\rightarrow \infty$ (N muy grande)

X~B(n, p) Si $\sigma^2 \geq 9 \Rightarrow X \sim N\left(m = np, \sigma^2 = np(1-p)\right)$	**X ~ Ps (λ)** Si $\sigma^2 \geq 9 \Rightarrow X \sim N\left(m = \lambda, \sigma^2 = \lambda\right)$

Distribuciones en el muestreo de poblaciones normales

$X \sim N(m,\sigma^2)$ y $\bar{X}$ es la media de una muestra de tamaño N	$\dfrac{\bar{x} - m}{\sigma/\sqrt{N}} \sim N(0,1)$
$X \sim N(m,\sigma^2)$ y S^2 es la varianza de una muestra de tamaño N	$(N-1)\dfrac{s^2}{\sigma^2} \sim \chi^2_{N-1}$
$X \sim N(m,\sigma^2)$ y $\bar{X}$ y S^2 son la media y la varianza de una muestra de tamaño N	$t = \dfrac{\bar{x} - m}{s/\sqrt{N}} \sim t_{N-1}$
$X_1 \sim N(m_1,\sigma^2_1)$, $X_2 \sim N(m_2,\sigma^2_2)$ independientes. S^2_1 y S^2_2 son las varianzas muestrales de X_1 y X_2 (tamaños N_1 y N_2)	$\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{N_1-1, N_2-1}$

Inferencia en poblaciones normales

Test de hipótesis para la media (Test t)

$H_0: m = m_0$ $H_1: m \neq m_0$	Si $\left\lvert \dfrac{\bar{X} - m_0}{S/\sqrt{N}} \right\rvert \leq t^{\alpha/2}_{N-1} \Rightarrow$ Aceptar H_0 Si $\left\lvert \dfrac{\bar{X} - m_0}{S/\sqrt{N}} \right\rvert > t^{\alpha/2}_{N-1} \Rightarrow$ Rechazar H_0	Si ***p-value*** $\geq \alpha \Rightarrow$ Aceptar H_0 Si ***p-value*** $< \alpha \Rightarrow$ Rechazar H_0
	$t^{\alpha/2}_{N-1} \Rightarrow$ valor en tabla	α = Riesgo de 1ª especie

Intervalo de confianza para la media (IC_m)

$IC_m \Rightarrow \left[\bar{x} - t^{\alpha/2}_{N-1}\dfrac{S}{\sqrt{N}},\ \bar{x} + t^{\alpha/2}_{N-1}\dfrac{S}{\sqrt{N}}\right]$	$t^{\alpha/2}_{N-1} \Rightarrow$ valor en tabla

Test de hipótesis para la media mediante IC

$$\left.\begin{array}{l} H_0: m = m_0 \\ H_1: m \neq m_0 \end{array}\right\}$$

$$\text{Si} \quad m_0 \in IC_m \Rightarrow \text{Aceptar} \quad H_0$$
$$\text{Si} \quad m_0 \notin IC_m \Rightarrow \text{Rechazar} \quad H_0$$

Intervalo de confianza para la varianza

$$IC_{\sigma^2} \Rightarrow \left[\frac{(N-1)S^2}{g_2}, \frac{(N-1)S^2}{g_1}\right]$$

g_1 y g_2 $\Rightarrow$ valores en tabla

$$g_1 \,/\, P(\chi^2_{N-1} > g_1) = 1 - \frac{\alpha}{2} \quad \text{y} \quad g_2 \,/\, P(\chi^2_{N-1} > g_2) = \frac{\alpha}{2}$$

Test de hipótesis para la varianza mediante IC

$$\left.\begin{array}{l} H_0: \sigma^2 = \sigma^2{}_0 \\ H_1: \sigma^2 \neq \sigma^2{}_0 \end{array}\right\}$$

$$\text{Si} \quad \sigma_0^2 \in IC_{\sigma^2} \Rightarrow \text{Aceptar} \quad H_0$$
$$\text{Si} \quad \sigma_0^2 \notin IC_{\sigma^2} \Rightarrow \text{Rechazar} \quad H_0$$

Intervalo de confianza para el cociente de varianzas

$$IC_{\sigma_1^2/\sigma_2^2} \Rightarrow \left(\frac{S_1^2}{S_2^2\, f_2}, \frac{S_1^2}{S_2^2\, f_1}\right)$$

f_1 y f_2 $\Rightarrow$ valores en tabla

$$f_1 \,/\, P(F_{(N_1-1),(N_2-1)} > f_1) = 1 - \frac{\alpha}{2} \quad \text{y} \quad f_2 \,/\, P(F_{(N_1-1),(N_2-1)} > f_2) = \frac{\alpha}{2}$$

Test de hipótesis para la comparación de varianzas mediante IC

$$\left.\begin{array}{l} H_0: \sigma^2{}_1 = \sigma^2{}_2 \\ H_1: \sigma^2{}_1 \neq \sigma^2{}_2 \end{array}\right\}$$

$$\text{Si} \quad 1 \in IC_{\sigma_1^2/\sigma_2^2} \Rightarrow \text{Aceptar} \quad H_0$$
$$\text{Si} \quad 1 \notin IC_{\sigma_1^2/\sigma_2^2} \Rightarrow \text{Rechazar} \quad H_0$$

Análisis de la Varianza (ANOVA)

Nomenclatura

N = nº total de observaciones	$\mathbf{F_i}$ = factor i $\mathbf{F_j}$ = factor j $\mathbf{F_i}$x $\mathbf{F_j}$ = interacción entre $\mathbf{F_j}$ y $\mathbf{F_j}$	**I** = nº niveles/variantes factor $\mathbf{F_i}$ **J** = nº niveles/variantes factor $\mathbf{F_j}$	**SC** = Suma de Cuadrados **gl** = grados de libertad

SCT = SC total $\mathbf{SC_{Fi}}$ = SC factor i $\mathbf{SC_{Fj}}$ = SC factor j $\mathbf{SC_{FixFj}}$ = SC de la interacción $\mathbf{F_i}$ x $\mathbf{F_j}$ **SCR** = SC Residual	$\mathbf{gl_{Tot}}$ = gl totales $\mathbf{gl_{Fi}}$ = gl asociados a la SC del $\mathbf{F_i}$ $\mathbf{gl_{Fj}}$ = gl asociados a la SC del $\mathbf{F_j}$ $\mathbf{gl_{Fi}}$ = gl asociados a la SC de la interacción $\mathbf{F_ixF_j}$ $\mathbf{gl_{Res}}$ = gl residuales

$gl_{Tot} \rightarrow (N-1)$	$\mathbf{gl_{Fi}} \rightarrow (I-1)$	$\mathbf{gl_{Fj}} \rightarrow (J-1)$	$\mathbf{gl_{FjxFj}} \rightarrow (I-1)x(J-1)$	$gl_{Res} = gl_T - \left(\sum_{\forall Fact} gl_{Factores} + \sum_{\forall Int} gl_{Interacciones}\right)$

Ecuación fundamental del ANOVA

$SCT = \sum_{\forall Fact} SC_{Factores} + \sum_{\forall Int} SC_{Interacciones} + SCR$	CM =SC/gl

Test de hipótesis para el ANOVA (Test F)

$\left.\begin{array}{l} H_0 : m_1 = m_2 = \cdots m_k \\ H_1 : \exists i,j \quad i \neq j \ / \ m_i \neq m_j \end{array}\right\}$	$F_{ratio} = \dfrac{CMF}{CMR} \sim F_{gl_F, gl_{Res}}$	**CMF** = Cuadrado medio asociado el efecto de un factor o interacción **CMR** = Cuadrado medio residual

Si $F_{ratio} \leq f \Rightarrow$ Aceptar H_0 Si $F_{ratio} > f \Rightarrow$ Rechazar H_0	Si ***p-value*** $\geq \alpha \Rightarrow$ Aceptar H_0 Si ***p-value*** $< \alpha \Rightarrow$ Rechazar H_0	$f \ / \ P(F_{gl_F, gl_{Res}} > f) = \alpha$ $P - value \ / \ P(F_{gl_F, gl_{Res}} > F_{ratio}) = p - value$

Introducción a los Modelos de Regresión Lineal

Nomenclatura

Modelo $E(Y/X_1=x_{1t},....,X_l=x_{lt}) = \beta_0 + \beta_1 x_{1t} +.... + \beta_l x_{lt}$	$\mathbf{E(Y/X_1=x_{1t},....,X_l=x_{lt})}$ es la media de la distribución condicional de Y cuando $X_1=x_{1t}$....y $X_l=x_{lt}$

N = nº total de observaciones	$\boldsymbol{\beta_i}$ = parámetros del modelo $\mathbf{b_i}$ = estimadores de los parámetros **l** = nº variables explicativas	**SCT = SCE + SCR**

SCT = SC total **SCE** = SC Explicada **SCR** = SC Residual	**gl**$_{Tot}$ = gl totales → (N-1) **gl**$_{Exp}$ = gl asociados a la SC Explicada → I **gl**$_{Res}$ = gl residuales → (N-1)-I	CME = SCE / gl$_{Exp}$ CMR = SCR / gl$_{Res}$

Test de significación global del ajuste (ANOVA)

$\left.\begin{array}{l} H_0 : \beta_1 = \beta_2 = \ldots \beta_I = 0 \\ H_1 : \exists \beta_i \,/\, \beta_i \neq 0 \end{array}\right\}$	$F\text{-ratio} = \frac{CME}{CMR}$	Si **F-ratio** $\leq F_{I,N-1-I}^{\alpha}$ Aceptar H_0 Si **F-ratio** $> F_{I,N-1-I}^{\alpha}$ Rechazar H_0	Si ***p-value*** $\geq \alpha \Rightarrow$ Aceptar H_0 Si ***p-value*** $< \alpha \Rightarrow$ Rechazar H_0

Coef. de determinación (*R-squared*): $R^2 = SCE/SCT * 100$	**Varianza residual:** S^2_R = CMR	**Desv. típica residual** (Standard Error of Est.): $S_R = \sqrt{CMR}$

Test de significación del efecto de una variable X_i (Test t)

$\left.\begin{array}{l} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{array}\right\}$	$t-calc = \left\lvert \frac{b_i}{s_{b_i}} \right\rvert$	Si **t-calc** $\leq t_{N-1-I}^{\alpha}$ Aceptar H_0 Si **t-calc** $> t_{N-1-I}^{\alpha}$ Rechazar H_0	Si ***p-value*** $\geq \alpha \Rightarrow$ Aceptar H_0 Si ***p-value*** $< \alpha \Rightarrow$ Rechazar H_0

Predicciones

$(Y/X_1=x_{1t},\ldots,X_I=x_{It}) \sim \text{Normal}(m = E(Y/X_1=x_{1t},\ldots,X_I=x_{It})\ ;\ \sigma^2 = \sigma^2_R)$

Modelos de Regresión Lineal Simple

Modelo $E(Y/X=x_t) = \beta_0 + \beta_1 x_t$	**$E(Y/X=x_t)$** es la media de la distribución condicional de Y cuando $X=x_t$

Estimación Recta de regresión: $Y = a + bX$	(*Slope*) $b = r_{xy} \frac{S_y}{S_x}$	(*Intercept*) $a = \overline{Y} - b.\overline{X}$
Coeficiente de determinación: $R^2 = (r_{xy})^2 \ .\ 100$	**Varianza residual:** $S^2_{residual} = S^2_y(1-r^2_{xy})$	**Residuo:** $e_i = y_i - (a + bx_i)$

Material docente previo de R. Alcover (DEIOAC - UPV)