FORMULARIO DE ESTADÍSTICA

- Recta de regresión de Y sobre X y coeficiente de correlación lineal:

$$y - \bar{y} = \frac{\mathrm{Cov}(X,Y)}{V(X)}(x - \bar{x}) \qquad\qquad r = \frac{\mathrm{Cov}(X,Y)}{dt(X)dt(Y)}$$

- Distribución binomial, $B(n, p)$: nº de "éxitos" en n experimentos (p=probabilidad de "éxito")

$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}, \quad r = 0, 1, \ldots, n; \qquad E(X) = np, \quad V(X) = np(1 - p)$$

- Distribución geométrica, $G(p)$:

  X: nº de experimentos **antes** del primer "éxito" (p=probabilidad de "éxito")

$$P(X = r) = p\,(1 - p)^r, \quad r = 0, 1, 2, \ldots; \qquad E(X) = \frac{1 - p}{p}, \qquad V(X) = \frac{1 - p}{p^2}$$

  X: nº de experimento **en el que aparece** el primer "éxito" (p=probabilidad de "éxito")

$$P(X = r) = p\,(1 - p)^{r-1}, \quad r = 1, 2, 3, \ldots; \qquad E(X) = \frac{1}{p}, \quad V(X) = \frac{1 - p}{p^2}$$

- Distribución de Poisson, $P(\lambda)$: nº de "éxitos" en un intervalo

$$P(X = r) = \frac{e^{-\lambda} \lambda^r}{r!}, \quad r = 0, 1, 2, \ldots; \qquad E(X) = V(X) = \lambda$$

- Distribución Uniforme, $U(a,b)$:

$$f(x) = \frac{1}{b - a}, \quad \text{si } x \in (a,b); \qquad E(X) = \frac{a + b}{2}, \qquad V(X) = \frac{(b - a)^2}{12}$$

- Distribución normal, $N(\mu,\sigma)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \forall x \in \Re$$

- Distribución gamma, $\gamma(\alpha,\beta)$:

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}, \quad \text{si } x > 0; \qquad E(X) = \frac{\alpha}{\beta}, \quad V(X) = \frac{\alpha}{\beta^2}$$

- Distribución exponencial, $\mathrm{Exp}(\beta)$:

$$f(x) = \beta\, e^{-\beta x}, \quad \text{si } x > 0; \qquad F(x) = 1 - e^{-\beta x}, \quad \text{si } x \geq 0; \qquad E(X) = \frac{1}{\beta}, \quad V(X) = \frac{1}{\beta^2}$$

- Distribución de Pareto, $P(\alpha,k)$:

$$f(x) = \alpha k^{\alpha} x^{-(\alpha+1)}, \quad \text{si } x > k; \qquad E(X) = \frac{\alpha k}{\alpha - 1} \text{ si } \alpha > 1, \qquad V(X) = \frac{\alpha k^2}{(\alpha - 2)(\alpha - 1)^2} \text{ si } \alpha > 2$$

- Definiciones de χ^2, t de Student y F de Snedecor:

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2, \text{ con } X_i \sim N(0,1) \text{ independientes}; \qquad t_n = \frac{N(0,1)}{\sqrt{\frac{\chi_n^2}{n}}}; \qquad F_{m,n} = \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}}$$

Distribución de estadísticos para intervalos y contrastes

- Para la media, μ: $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ (población normal) ; $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \approx t_{n-1} \approx N(0,1)$ (poblaciones cualesquiera con n suficientemente grande)

- Para la varianza, σ^2: $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ (**SÓLO** población normal)

- Para el cociente de varianzas, $\dfrac{\sigma^2_X}{\sigma^2_Y}$: $\dfrac{S^2_X / S^2_Y}{\sigma^2_X / \sigma^2_Y} \sim F_{n_X-1, n_Y-1}$ (**SÓLO** poblaciones normales)

- Para la diferencia de medias, $\mu_X - \mu_Y$, con **varianzas iguales**: (**SÓLO** poblaciones normales)

$$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_p\sqrt{\dfrac{1}{n_X}+\dfrac{1}{n_Y}}} \sim t_{n_X+n_Y-2}, \text{ siendo } S_p = \sqrt{\frac{(n_X-1)S^2_X+(n_Y-1)S^2_Y}{n_X+n_Y-2}}$$

- Para la diferencia de medias, $\mu_X - \mu_Y$, con varianzas distintas (aproximación de Welch) (poblaciones normales):

$$X^* = \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\dfrac{S^2_X}{n_X}+\dfrac{S^2_Y}{n_Y}}} \approx t_f, \text{ donde } f = \frac{\left(\dfrac{S^2_X}{n_X}+\dfrac{S^2_Y}{n_Y}\right)^2}{\dfrac{\left(\dfrac{S^2_X}{n_X}\right)^2}{n_X+1}+\dfrac{\left(\dfrac{S^2_Y}{n_Y}\right)^2}{n_Y+1}} - 2$$

Para poblaciones cualesquiera, con n_X y n_Y **suficientemente grandes**: $X^* \approx N(0,1)$

- Intervalo para la proporción, P: $\dfrac{\hat{P}-P}{\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n-1}}} \approx t_{n-1}$ (n suficientemente grande)

- Contraste para la proporción, P: $\dfrac{\hat{P}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \approx N(0,1)$ (n suficientemente grande)

- Intervalo para la diferencia de proporciones, $P_1 - P_2$: $\dfrac{(\hat{P}_1-\hat{P}_2)-(P_1-P_2)}{\sqrt{\dfrac{\hat{P}_1(1-\hat{P}_1)}{n_1-1}+\dfrac{\hat{P}_2(1-\hat{P}_2)}{n_2-1}}} \approx N(0,1)$

- Contraste para la diferencia de proporciones, $P_1 - P_2$:

$$\frac{\hat{P}_1-\hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \approx N(0,1), \text{ donde } \hat{P} = \frac{r_1+r_2}{n_1+n_2}$$

- Estadístico del test χ^2 de Pearson: $D^* = \displaystyle\sum_{i=1}^{K} \frac{(n_i - n\,p_i)^2}{n\,p_i} \approx \chi^2_{K-r-1}$ (discretas, $n \geq 30$, $np_i \geq 5$ para todo i ; r: nº de parámetros estimados)

- Estadístico del test de Kolmogorov-Smirnov: $D_n = \max\left|F_n(x) - F(x)\right|$ (continuas)

$$D_n(x_i) = \max\left(\left|F_n(x_{i-1}) - F(x_i)\right|, \left|F_n(x_i) - F(x_i)\right|\right)$$