

# MODELOS DE PROBABILIDAD MÁS COMUNES

- $X$ : Variable aleatoria.
- $E(X)$ : Esperanza (media) de  $X$
- $V(X)$ : Varianza de  $X$
- $f(x)$ : Función de densidad
- $F(x) = P(X \leq x)$ : Función de Distribución.

1. **Binomial:**  $X \sim B(n; p)$  con  $n \in \mathbb{N}$ ,  $0 < p < 1$  y  $q = 1 - p$ .

- $P(X = k) = \binom{n}{k} p^k q^{n-k}$ , para  $k = 0, 1, 2, \dots, n$ .
- $E(X) = np$ ,  $V(X) = npq$ .

2. **Poisson:**  $X \sim P(\lambda)$  con  $\lambda > 0$ .

- $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ , para  $k = 0, 1, 2, \dots$
- $E(X) = \lambda$ ,  $V(X) = \lambda$ .

3. **Geométrica:**  $X \sim \text{Geom}(p)$ ;  $0 < p < 1$ .

- $P(X = k) = (1 - p)^{k-1} p$ , para  $k = 1, 2, 3, \dots$
- $E(X) = \frac{1}{p}$ ,  $V(X) = \frac{1 - p}{p^2}$ .

4. **Uniforme:**  $X \sim U(a, b)$  con  $a < b$ ,  $a, b \in \mathbb{R}$ .

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

5. **Exponencial:**  $X \sim \exp(\lambda)$ , con  $\lambda > 0$ .

- $f(x) = \lambda e^{-\lambda x}$ ,  $x > 0$ .  $F(x) = 1 - e^{-\lambda x}$ ,  $x > 0$ .
- $E(X) = 1/\lambda$ ,  $V(X) = 1/\lambda^2$ .

6. **Normal:**  $X \sim N(\mu, \sigma)$  con  $\mu \in \mathbb{R}$  y  $\sigma > 0$ .

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ,  $x \in \mathbb{R}$ .
- $E(X) = \mu$ ,  $V(X) = \sigma^2$ .

7.  $\chi^2$  de Pearson con  $n$  grados de libertad:  $X \sim \chi_n^2$ .

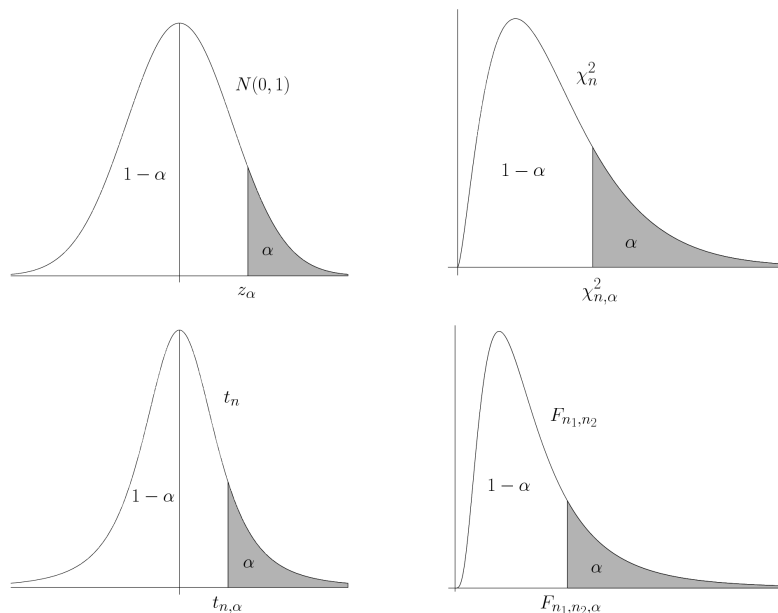
- $f(x) = \frac{x^{n/2-1} e^{-x/2}}{\text{constante} \cdot 2^{n/2}}$ ,  $x > 0$ .
- $E(X) = n$ ,  $V(X) = 2n$ .

8.  $t$  de Student con  $n$  grados de libertad:  $X \sim t_n$ .

- $f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ .
- $E(X) = 0$ ;  $V(X) = \frac{n}{n-2}$ , ( $n > 2$ ).

9.  $F$  de Snedecor con  $m$  y  $n$  grados de libertad:  $X \sim F_{m,n}$

- $f(x) = \frac{m}{n} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}x\right)^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}$ .
- $E(X) = \frac{n}{n-2}$ , ( $n > 2$ );  $V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ , ( $n > 4$ ).



#### APROXIMACIONES DE UNA DISTRIBUCIÓN BINOMIAL

- Aproximación de una Binomial por una Normal

Para  $n$  grande en relación a  $p$  y  $q = 1 - p$  ( $np \geq 10q$  y  $nq \geq 10p$ ):

$$B(n, p) \approx N\left(\mu = np, \sigma = \sqrt{np(1-p)}\right).$$

- Aproximación de una Binomial por una Poisson

Para  $n$  grande ( $n \geq 30$ ) y  $0 < p < 0.1$ :

$$B(n, p) \approx P(\lambda = np).$$

#### ESTIMADORES

$(X_1, \dots, X_n)$  muestra aleatoria simple (m.a.s.) de  $X$ .

Media muestral:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

Cuasi-varianza:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Distribución de  $\bar{X}$  cuando  $X \sim N(\mu, \sigma)$

Si  $X \sim N(\mu, \sigma)$  y  $(X_1, X_2, \dots, X_n)$  es una m.a.s. de  $X$ , entonces  $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ .

Distribución de  $S^2$  cuando  $X \sim N(\mu, \sigma)$

Si  $X \sim N(\mu, \sigma)$  y  $(X_1, X_2, \dots, X_n)$  es una m.a.s. de  $X$ , entonces

$$\frac{S^2}{\sigma^2/(n-1)} \sim \chi_{n-1}^2.$$

**Intervalos de confianza más usuales**

1.  $X \sim N(\mu, \sigma)$

$$\text{Intervalo de confianza } 1 - \alpha \text{ para } \mu: \begin{cases} I = \left[ \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] & (\sigma \text{ conocida}) \\ I = \left[ \bar{x} \pm t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} \right] & (\sigma \text{ desconocida}) \end{cases}$$

$$\text{Intervalo de confianza } 1 - \alpha \text{ para } \sigma^2: I = \left[ \frac{(n-1)s^2}{\chi_{n-1; \alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1; 1-\alpha/2}^2} \right]$$

2.  $X \sim B(1, p)$  (muestras grandes).

$$\text{Intervalo de confianza } 1 - \alpha \text{ para } p: I = \left[ \bar{x} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \right]$$

3.  $X \sim P(\lambda)$  (muestras grandes).

$$\text{Intervalo de confianza } 1 - \alpha \text{ para } \lambda: I = \left[ \bar{x} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}}{n}} \right]$$

4. Dos poblaciones Normales independientes

$X \sim N(\mu_x, \sigma_x)$ ,  $Y \sim N(\mu_y, \sigma_y)$  independientes

$(X_1, \dots, X_{n_x})$  m.a.s. de  $X$ ; se calcula  $\bar{x}$  y  $s_x^2$ .

$(Y_1, \dots, Y_{n_y})$  m.a.s. de  $Y$ ; se calcula  $\bar{y}$  y  $s_y^2$ .

$$s_p^2 = \frac{(n_x - 1) s_x^2 + (n_y - 1) s_y^2}{n_x + n_y - 2}$$

Intervalo de confianza  $1 - \alpha$  para  $\mu_x - \mu_y$ :

$$I = \left[ \bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right] \quad \sigma_x, \sigma_y \text{ conocidas}$$

$$I = \left[ \bar{x} - \bar{y} \pm t_{n_x+n_y-2; \alpha/2} s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \right] \quad \sigma_x, \sigma_y \text{ desconocidas, } \sigma_x = \sigma_y$$

El caso  $\sigma_x, \sigma_y$  desconocidas,  $\sigma_x \neq \sigma_y$  es complicado

$$\text{Intervalo de confianza } 1 - \alpha \text{ para } \sigma_x^2/\sigma_y^2: I = \left[ \frac{s_x^2/s_y^2}{F_{n_x-1; n_y-1; \alpha/2}}, (s_x^2/s_y^2) F_{n_y-1; n_x-1; \alpha/2} \right]$$

5. Comparación de proporciones (muestras grandes e independientes)

$X \sim B(1, p_x)$ ,  $Y \sim B(1, p_y)$ , independientes.

$(X_1, \dots, X_{n_x})$  m.a.s. de  $X$ ; se calcula  $\bar{x}$  y  $s_x^2$ .

$(Y_1, \dots, Y_{n_y})$  m.a.s. de  $Y$ ; se calcula  $\bar{y}$  y  $s_y^2$ .

$$\text{Intervalo de confianza } 1 - \alpha \text{ para } p_x - p_y: I = \left[ \bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n_x} + \frac{\bar{y}(1-\bar{y})}{n_y}} \right]$$

## 6. Datos emparejados

$$X \sim N(\mu_x, \sigma_x), Y \sim N(\mu_y, \sigma_y).$$

$$D = X - Y \sim N(\mu = \mu_x - \mu_y, \sigma),$$

donde el cálculo de  $\sigma$  supera el nivel de este curso.

### Contrastes de hipótesis más usuales

- $\alpha$  = nivel de significación del contraste.
- $n$  = tamaño de la muestra.
- $H_0$  = hipótesis nula.
- $R$  = región crítica o de rechazo de  $H_0$ .

#### 1.- $X \sim N(\mu, \sigma)$

$$H_0 : \mu = \mu_0 \text{ (}\sigma \text{ conocida)} \quad R = \left\{ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

$$H_0 : \mu = \mu_0 \text{ (}\sigma \text{ desconocida)} \quad R = \left\{ |\bar{x} - \mu_0| > t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} \right\}$$

$$H_0 : \mu \leq \mu_0 \text{ (}\sigma \text{ conocida)} \quad R = \left\{ \bar{x} - \mu_0 > z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

$$H_0 : \mu \leq \mu_0 \text{ (}\sigma \text{ desconocida)} \quad R = \left\{ \bar{x} - \mu_0 > t_{n-1; \alpha} \frac{s}{\sqrt{n}} \right\}$$

$$H_0 : \mu \geq \mu_0 \text{ (}\sigma \text{ conocida)} \quad R = \left\{ \bar{x} - \mu_0 < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

$$H_0 : \mu \geq \mu_0 \text{ (}\sigma \text{ desconocida)} \quad R = \left\{ \bar{x} - \mu_0 < t_{n-1; 1-\alpha} \frac{s}{\sqrt{n}} \right\}$$

$$H_0 : \sigma = \sigma_0 \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 \notin \left[ \chi_{n-1; 1-\alpha/2}^2, \chi_{n-1; \alpha/2}^2 \right] \right\}$$

$$H_0 : \sigma \leq \sigma_0 \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 > \chi_{n-1; \alpha}^2 \right\}$$

$$H_0 : \sigma \geq \sigma_0 \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 < \chi_{n-1; 1-\alpha}^2 \right\}$$

#### 2.- $X \sim B(1, p)$ (muestras grandes)

$$H_0 : p = p_0 \quad R = \left\{ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

$$H_0 : p \leq p_0 \quad R = \left\{ \bar{x} - p_0 > z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

$$H_0 : p \geq p_0 \quad R = \left\{ \bar{x} - p_0 < z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

#### 3.- $X \sim P(\lambda)$ (muestras grandes)

$$H_0 : \lambda = \lambda_0 \quad R = \left\{ |\bar{x} - \lambda_0| > z_{\alpha/2} \sqrt{\lambda_0/n} \right\}$$

$$H_0 : \lambda \leq \lambda_0 \quad R = \left\{ \bar{x} - \lambda_0 > z_{\alpha} \sqrt{\lambda_0/n} \right\}$$

$$H_0 : \lambda \geq \lambda_0 \quad R = \left\{ \bar{x} - \lambda_0 < z_{1-\alpha} \sqrt{\lambda_0/n} \right\}$$

#### 4.- Dos poblaciones Normales independientes

( $s_p^2$  calculado como en los intervalos de confianza)

$$\begin{array}{ll}
 H_0 : \mu_x = \mu_y \ (\sigma_x, \sigma_y \text{ conocidas}) & R = \left\{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right\} \\
 H_0 : \mu_x = \mu_y \ (\sigma_x = \sigma_y) & R = \left\{ |\bar{x} - \bar{y}| > t_{n_x+n_y-2; \alpha/2} s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \right\} \\
 H_0 : \mu_x \leq \mu_y \ (\sigma_x, \sigma_y \text{ conocidas}) & R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right\} \\
 H_0 : \mu_x \leq \mu_y \ (\sigma_x = \sigma_y) & R = \left\{ \bar{x} - \bar{y} > t_{n_x+n_y-2; \alpha} s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \right\} \\
 H_0 : \mu_x \geq \mu_y \ (\sigma_x, \sigma_y \text{ conocidas}) & R = \left\{ \bar{x} - \bar{y} < z_{1-\alpha} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right\} \\
 H_0 : \mu_x \geq \mu_y \ (\sigma_x = \sigma_y) & R = \left\{ \bar{x} - \bar{y} < t_{n_x+n_y-2; 1-\alpha} s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \right\} \\
 H_0 : \sigma_x = \sigma_y & R = \left\{ s_x^2/s_y^2 \notin [F_{n_x-1; n_y-1; 1-\alpha/2}, F_{n_x-1; n_y-1; \alpha/2}] \right\} \\
 H_0 : \sigma_x \leq \sigma_y & R = \left\{ s_x^2/s_y^2 > F_{n_x-1; n_y-1; \alpha} \right\} \\
 H_0 : \sigma_x \geq \sigma_y & R = \left\{ s_x^2/s_y^2 < F_{n_x-1; n_y-1; 1-\alpha} \right\}
 \end{array}$$

#### 5.- Comparación de proporciones (muestras grandes e independientes)

$$\left. \begin{array}{l}
 X \sim B(1, p_x), \quad (X_1, \dots, X_{n_x}) \text{ m.a.s. de } X \\
 Y \sim B(1, p_y), \quad (Y_1, \dots, Y_{n_y}) \text{ m.a.s. de } Y
 \end{array} \right\} \rightsquigarrow \bar{p} = \frac{\sum_i x_i + \sum_i y_i}{n_x + n_y} = \frac{n_x \bar{x} + n_y \bar{y}}{n_x + n_y}$$

$$\begin{array}{ll}
 H_0 : p_x = p_y & R = \left\{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\bar{p}(1-\bar{p}) \left( \frac{1}{n_x} + \frac{1}{n_y} \right)} \right\} \\
 H_0 : p_x \leq p_y & R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\bar{p}(1-\bar{p}) \left( \frac{1}{n_x} + \frac{1}{n_y} \right)} \right\} \\
 H_0 : p_x \geq p_y & R = \left\{ \bar{x} - \bar{y} < z_{1-\alpha} \sqrt{\bar{p}(1-\bar{p}) \left( \frac{1}{n_x} + \frac{1}{n_y} \right)} \right\}
 \end{array}$$

#### 6.- Datos emparejados

$$X \sim N(\mu_x, \sigma_x), Y \sim N(\mu_y, \sigma_y).$$

$$D = X - Y \sim N(\mu = \mu_x - \mu_y, \sigma),$$

se utilizan los métodos para una muestra con  $\sigma$  desconocida.

## CONTRASTES $\chi^2$

- $\alpha$  = nivel de significación del contraste.
- $H_0$  = hipótesis nula.
- $n$  = tamaño de la muestra.
- $R$  = región crítica o de rechazo de  $H_0$ .

### 1. Contraste de la bondad del ajuste: Primer caso

- $H_0$  : La población  $X$  sigue el modelo  $P$  indicado.
- $A_1, A_2, \dots, A_k$ :  $k$  clases de los posibles valores de  $X$ .
- $O_i$  = frecuencia observada en la clase  $A_i$ .
- $e_i = n P(A_i)$  = frecuencia esperada en la clase  $A_i$ , suponiendo que  $H_0$  es cierta.

$$R = \left\{ \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^k \frac{O_i^2}{e_i} - n > \chi_{k-1; \alpha}^2 \right\}$$

### 2. Contraste de la bondad del ajuste: Segundo caso.

- $H_0$  : La población  $X$  sigue algún modelo  $P_\theta$  de una cierta familia de distribuciones
- $r$  = número de los parámetros desconocidos:  $\theta = (\theta_1, \theta_2, \dots, \theta_r)$ .
- $A_1, A_2, \dots, A_k$ :  $k$  clases de los posibles valores de  $X$ .
- $O_i$  = frecuencia observada en la clase  $A_i$ .
- $e_i = n P_{\hat{\theta}}(A_i)$  = frecuencia esperada en la clase  $A_i$ , suponiendo que  $H_0$  es cierta (y usando el estimador de máxima verosimilitud  $\hat{\theta}$  del parámetro  $\theta$ ).

$$R = \left\{ \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^k \frac{O_i^2}{e_i} - n > \chi_{k-1-r; \alpha}^2 \right\}$$

### 3. Contraste de homogeneidad de poblaciones

- $H_0$  : Las  $p$  poblaciones  $X_1, X_2, \dots, X_p$  son homogéneas
- $A_1, A_2, \dots, A_k$ :  $k$  clases de los posibles valores de  $X$ .
- $O_{ij}$  = frecuencia observada en la clase  $A_i$  con la muestra  $j$ -ésima.
- $e_{ij} = n_j \hat{P}(A_i) = \frac{(\sum \text{columna } i\text{-ésima}) \cdot (\sum \text{fila } j\text{-ésima})}{n} =$  frecuencia esperada en la clase  $A_i$  con la muestra  $j$ -ésima, si  $H_0$  es cierta.

$$R = \left\{ \sum_{i=1}^k \sum_{j=1}^p \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^k \sum_{j=1}^p \frac{O_{ij}^2}{e_{ij}} - n > \chi_{(k-1)(p-1); \alpha}^2 \right\}$$

### 4. Contraste de independencia

- $H_0$  : Las características  $X$  e  $Y$  de la población son independientes.
- $A_1 \times B_1, \dots, A_i \times B_j, \dots, A_k \times B_p$ :  $k p$  clases de los posibles valores de  $X \times Y$ .
- $O_{ij}$  = frecuencia observada en la clase  $A_i \times B_j$ .
- $e_{ij} = n \hat{P}(A_i) \hat{P}(B_j) = \frac{(\sum \text{columna } i\text{-ésima}) \cdot (\sum \text{fila } j\text{-ésima})}{n} =$  frecuencia esperada en la clase  $A_i \times B_j$  suponiendo que  $H_0$  es cierta.

$$R = \left\{ \sum_{i=1}^k \sum_{j=1}^p \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^k \sum_{j=1}^p \frac{O_{ij}^2}{e_{ij}} - n > \chi_{(k-1)(p-1); \alpha}^2 \right\}$$