UNIVERSIDAD DE EXTREMADURA
DEPARTAMENTO DE MATEMATICAS
Cátedra de Bioestadística
Avda. Elvas, s/n
06071-BADAJOZ (SPAIN)

Tablas y Resúmenes de Bioestadística

Agustín García Nogales y Miguel González Velasco

1er Curso de la Licenciatura en Medicina y Cirugía

Índice

RESUMEN I

Aproximación de la distribución binomial por la distribución normal

DISTRIBUCIÓN BINOMIAL: $B(n, p)$

Media=np. Varianza=npq, $(q = 1 - p)$

DISTRIBUCIÓN NORMAL: $N(\mu, \sigma^2)$

Media=μ. Varianza=σ^2

CONDICIÓN DE VALIDEZ:

$npq \geq 5$

RESUMEN II

Intervalo de confianza para la media de una población normal.

Sea $X_1, \ldots, X_n$ una muestra de una población $N(\mu, \sigma^2)$. Sean

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

σ^2 CONOCIDA	σ^2 DESCONOCIDA
$\mu \in \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$	$\mu \in \bar{x} \pm t_\alpha(n-1)\frac{s}{\sqrt{n}}$

OBSERVACIÓN: Si $n \geq 60$, se verifique o no la hipótesis de normalidad, los resultados anteriores son aproximadamente válidos utilizando z_α en lugar de $t_\alpha(n-1)$.

RESUMEN III

Contraste de hipótesis para la media de una población normal.

Sea $X_1, \ldots, X_n$ una muestra de una población $N(\mu, \sigma^2)$ y μ_0 un valor conocido. Sean

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

			Rechazamos H_0 a un nivel α si	
	H_0	H_1	σ^2 conocida	σ^2 desconocida
I.	$\mu = \mu_0$	$\mu \neq \mu_0$	$\dfrac{\lvert\bar{x} - \mu_0\rvert}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$	$\dfrac{\lvert\bar{x} - \mu_0\rvert}{\frac{s}{\sqrt{n}}} > t_\alpha(n-1)$
II.	$\mu \leq \mu_0$	$\mu > \mu_0$	$\dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{2\alpha}$	$\dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{2\alpha}(n-1)$
III.	$\mu \geq \mu_0$	$\mu < \mu_0$	$\dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{2\alpha}$	$\dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < -t_{2\alpha}(n-1)$

Observación: Si $n \geq 60$, se verifique o no la hipótesis de normalidad, los resultados anteriores son aproximadamente válidos utilizando z_α en lugar de $t_\alpha(n-1)$.

RESUMEN IV

Estimación y contraste de hipótesis para la varianza de una población normal

Sea $X_1, \ldots, X_n$ una muestra de una población $\mathrm{N}(\mu,\sigma^2)$. Sea

$$\bar{X} = \frac{1}{n}\sum X_i \quad y \quad S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$$

Intervalo de confianza para σ^2

$$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right]$$

Contraste de la hipótesis $H_0 : \sigma^2 = \sigma_0^2$ frente a $H_1 : \sigma^2 \neq \sigma_0^2$ (σ_0^2 conocida).

	H_0	H_1	Rechazamos H_0 a un nivel α si
I.	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$s^2 < \sigma_0^2 \frac{\chi^2_{1-\alpha/2}(n-1)}{n-1}$ ó $s^2 > \sigma_0^2 \frac{\chi^2_{\alpha/2}(n-1)}{n-1}$
II.	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$s^2 > \sigma_0^2 \frac{\chi^2_{\alpha}(n-1)}{n-1}$
III.	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$s^2 < \sigma_0^2 \frac{\chi^2_{1-\alpha}(n-1)}{n-1}$

RESUMEN V

Estimación y contraste de hipótesis para una proporción.

Sea $X_1, \ldots, X_n$ una muestra de una población X, donde $P(X=1)=p$ y $P(X=0)=1-p$. Sean

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \hat{q} = 1 - \hat{p}$$

Intervalo de Confianza para p.

- Condición de validez: $n\hat{p}\hat{q} \geq 5$.
- Intervalo al nivel de confianza $1-\alpha$:

$$p \in \hat{p} \pm z_\alpha \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Contraste de la hipótesis $H_0 : p = p_0$ frente a $H_1 : p \neq p_0$ (p_0 conocida).

- Condición de validez: $np_0q_0 \geq 5$ (donde $q_0 = 1 - p_0$).

	H_0	H_1	Rechazamos H_0 a un nivel α si
I.	$p = p_0$	$p \neq p_0$	$\dfrac{\lvert\hat{p} - p_0\rvert}{\sqrt{\dfrac{p_0q_0}{n}}} > z_\alpha$
II.	$p \leq p_0$	$p > p_0$	$\dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0q_0}{n}}} > z_{2\alpha}$
III.	$p \geq p_0$	$p < p_0$	$\dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0q_0}{n}}} < -z_{2\alpha}$

RESUMEN VI

Test de normalidad de D'Agostino

MUESTRA: Sean $x_1, \ldots, x_n$ los valores observados y

$$x_{(1)} \leq \cdots \leq x_{(n)}$$

esos mismos valores escritos en orden creciente.

ESTADÍSTICO DE CONTRASTE:

$$D = \frac{\sum_{i=1}^{n} i x_{(i)} - \frac{n+1}{2} \sum_{i=1}^{n} x_i}{n\sqrt{n\left(\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n\right)}}$$

ACEPTAMOS LA HIPÓTESIS DE NORMALIDAD SI

$$a_1(n, \alpha) < D < a_2(n, \alpha)$$

donde los valores $a_1(n, \alpha) < a_2(n, \alpha)$ aparecen en la tabla publicada por D'Agostino (Tabla VII) en función del tamaño de muestra n y el nivel de significación α.

RESUMEN VII

Comparación de varianzas para poblaciones normales (muestras independientes).

Sea $X_1,\dots,X_m$ e $Y_1,\dots,Y_n$ dos muestras independientes de $N(\mu_1,\sigma_1^2)$ y $N(\mu_2,\sigma_2^2)$, respectivamente. Sean

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{m}X_i,\quad \bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i,\quad S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\bar{X})^2,\quad S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

	H_0	H_1	Rechazamos H_0 a un nivel α si
I.	$\sigma_1^2=\sigma_2^2$	$\sigma_1^2\neq\sigma_2^2$	$\frac{s_1^2}{s_2^2}>\mathrm{F}_{\alpha/2}(m-1,n-1)$ si $s_1^2\geq s_2^2$ ó $\frac{s_2^2}{s_1^2}>\mathrm{F}_{\alpha/2}(n-1,m-1)$ si $s_1^2\leq s_2^2$
II.	$\sigma_1^2\leq\sigma_2^2$	$\sigma_1^2>\sigma_2^2$	$\frac{s_1^2}{s_2^2}>\mathrm{F}_{\alpha}(m-1,n-1)$
III.	$\sigma_1^2\geq\sigma_2^2$	$\sigma_1^2<\sigma_2^2$	$\frac{s_2^2}{s_1^2}>\mathrm{F}_{\alpha}(n-1,m-1)$

Intervalo de confianza para el cociente de varianzas para poblaciones normales (muestras independientes).

$$\frac{\sigma_1^2}{\sigma_2^2}\in\left[\frac{s_1^2}{s_2^2}\mathrm{F}^{-1}_{\alpha/2}(m-1,n-1)\,,\ \frac{s_1^2}{s_2^2}\mathrm{F}_{\alpha/2}(n-1,m-1)\right]$$

RESUMEN VIII

Comparación de medias para poblaciones normales (muestras independientes).

Sea $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ dos muestras independientes de $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, respectivamente. Sean

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad S^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

	H_0	H_1	RECHAZAMOS H_0 A UN NIVEL α SI: σ_1^2, σ_2^2 CONOCIDAS	σ_1^2, σ_2^2 DESCONOCIDAS, $\sigma_1^2 = \sigma_2^2$	σ_1^2, σ_2^2 DESCONOCIDAS, $\sigma_1^2 \neq \sigma_2^2$ (*)
I.	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\dfrac{\lvert\bar{x}-\bar{y}\rvert}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} > z_\alpha$	$\dfrac{\lvert\bar{x}-\bar{y}\rvert}{s\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} > t_\alpha(m+n-2)$	$\dfrac{\lvert\bar{x}-\bar{y}\rvert}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}} > t_\alpha(f)$
II.	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} > z_{2\alpha}$	$\dfrac{\bar{x}-\bar{y}}{s\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} > t_{2\alpha}(m+n-2)$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}} > t_{2\alpha}(f)$
III.	$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} < -z_{2\alpha}$	$\dfrac{\bar{x}-\bar{y}}{s\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} < -t_{2\alpha}(m+n-2)$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}} < -t_{2\alpha}(f)$

(*) $f = \dfrac{(s_1^2/m + s_2^2/n)^2}{\dfrac{(s_1^2/m)^2}{m-1} + \dfrac{(s_2^2/n)^2}{n-1}}$ (TEST DE WELCH).

OBSERVACIONES:

1. Si $m+n-2 \geq 60$ (resp. $f \geq 60$) utilizamos z_α en lugar de $t_\alpha(m+n-2)$ (resp. $t_\alpha(f)$).
2. Si no se verifica la hipótesis de normalidad pero $m, n \geq 60$, entonces:

	H_0	H_1	RECHAZAMOS H_0 A UN NIVEL α SI: σ_1^2, σ_2^2 CONOCIDAS	σ_1^2, σ_2^2 DESCONOCIDAS
I.	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\dfrac{\lvert\bar{x}-\bar{y}\rvert}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} > z_\alpha$	$\dfrac{\lvert\bar{x}-\bar{y}\rvert}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}} > z_\alpha$
II.	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} > z_{2\alpha}$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}} > z_{2\alpha}$
III.	$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} < -z_{2\alpha}$	$\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}} < -z_{2\alpha}$

RESUMEN IX

Intervalo de confianza para la diferencia de medias en poblaciones normales (muestras independientes)

Sea $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ dos muestras independientes de $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, respectivamente. Sean

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad S^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

σ_1^2, σ_2^2 CONOCIDAS	σ_1^2, σ_2^2 DESCONOCIDAS, $\sigma_1^2 = \sigma_2^2$	σ_1^2, σ_2^2 DESCONOCIDAS, $\sigma_1^2 \neq \sigma_2^2$
$\mu_1 - \mu_2 \in (\bar{x} - \bar{y}) \pm z_\alpha \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\mu_1 - \mu_2 \in (\bar{x} - \bar{y}) \pm t_\alpha(m+n-2) s \sqrt{\frac{1}{m} + \frac{1}{n}}$	$\mu_1 - \mu_2 \in (\bar{x} - \bar{y}) \pm t_\alpha(f) \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$ (*)

(*) $f = \frac{(s_1^2/m + s_2^2/n)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$

OBSERVACIONES:

1. Si $m + n - 2 \geq 60$ (resp. $f \geq 60$) utilizamos z_α en lugar de $t_\alpha(m+n-2)$ (resp. $t_\alpha(f)$).
2. Si no se verifica la hipótesis de normalidad pero $m, n \geq 60$, entonces:

σ_1^2, σ_2^2 CONOCIDAS	σ_1^2, σ_2^2 DESCONOCIDAS
$\mu_1 - \mu_2 \in (\bar{x} - \bar{y}) \pm z_\alpha \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\mu_1 - \mu_2 \in (\bar{x} - \bar{y}) \pm z_\alpha \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$

RESUMEN X

Test de de suma de rangos de Mann-Whitney-Wilcoxon (muestras independientes)

Sea $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ dos muestras independientes de dos poblaciones X e Y, respectivamente. Supondremos que $m \leq n$.

- Se mezclan las dos muestras en una sola.
- Se ordenan los datos de dicha muestra de menor a mayor y se asignan rangos (rangos medios en caso de empate).
- Se calculan R_X =suma de los rangos correspondientes a los datos de la muestra de menor tamaño, $X_1, \ldots, X_m$ y R_Y =suma de los rangos correspondientes a los datos de la muestra de mayor tamaño, $Y_1, \ldots, Y_n$. Se comprueba que $R_X + R_Y = (m+n)(m+n+1)/2$
- Se denota $R_{exp} = R_X$.
- Si $m + n > 30$, se calculan:
  1. $E(R) = m(m+n+1)/2$
  2. $V(R)$
     - Si no hay empates: $V(R) = mn(m+n+1)/12$
     - Si hay empates: $V(R) = \dfrac{mn}{m+n-1} \dfrac{(m+n)((m+n)^2-1) - \sum_{j=1}^{r} T_j}{12(m+n)}$,
       siendo $T_j = t_j^3 - t_j$ y t_j el número de observaciones en el j-ésimo grupo de empates, $1 \leq j \leq r$.

			Rechazamos H_0 a un nivel α si	
	H_0	H_1	$m+n \leq 30$	$m+n > 30$
I.	X e Y igual centralización	X e Y distinta centralización	$R_{exp} \leq V_I(\alpha)$ ó $R_{exp} \geq V_S(\alpha)$	$\dfrac{\lvert R_{exp} - E(R)\rvert}{\sqrt{V(R)}} > z_\alpha$
II.	X e Y igual centralización	X mayor centralización que Y	$R_{exp} \geq V_S(2\alpha)$	$\dfrac{R_{exp} - E(R)}{\sqrt{V(R)}} > z_{2\alpha}$
III.	X e Y igual centralización	X menor centralización que Y	$R_{exp} \leq V_I(2\alpha)$	$\dfrac{R_{exp} - E(R)}{\sqrt{V(R)}} < -z_{2\alpha}$

(*) Valores $V_I(\alpha)$ y $V_S(\alpha)$ en la tabla de Mann-Whitney-Wilcoxon (Tabla VIII) para muestras independientes.

RESUMEN XI

Comparación de medias para poblaciones normales (muestras relacionadas)

Sea $(X_1, Y_1), \ldots, (X_n, Y_n)$ una muestra de una población bidimensional, de vector media (μ_1, μ_2), tal que $D_i = X_i - Y_i$, $i = 1, \ldots, n$ es una muestra de una población $N(\mu_D, \sigma_D^2)$, con $\mu_D = \mu_1 - \mu_2$. Sean

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$$

	H_0	H_1	Rechazamos H_0 a un nivel α si
I.	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\dfrac{\lvert\bar{d}\rvert}{\frac{s_D}{\sqrt{n}}} > t_\alpha(n-1)$
II.	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\bar{d}}{\frac{s_D}{\sqrt{n}}} > t_{2\alpha}(n-1)$
III.	$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$\dfrac{\bar{d}}{\frac{s_D}{\sqrt{n}}} < -t_{2\alpha}(n-1)$

Intervalo de confianza para $\mu_1 - \mu_2$

$$\mu_1 - \mu_2 \in \bar{d} \pm t_\alpha(n-1)\frac{s_D}{\sqrt{n}}$$

Observación: Si $n \geq 60$, se verifique o no la hipótesis de normalidad, los resultados anteriores son aproximadamente válidos utilizando z_α en lugar de $t_\alpha(n-1)$.

RESUMEN XII

Test de rangos con signo de Wilcoxon (muestras relacionadas)

Sea $(X_1, Y_1), \ldots, (X_n, Y_n)$ una muestra de una población bidimensional (X, Y).

- Se calculan las diferencias $d_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Se eliminan las diferencias nulas y se actualiza el tamaño muestral, n.
- Se ordena de menor a mayor la muestra de los valores absolutos de las diferencias no nulas y se asignan rangos (rangos medios en caso de empate).
- Se calculan $R(+)$ = suma de los rangos correspondientes a las diferencias positivas y $R(-)$ = suma de los rangos correspondientes a las diferencias negativas.
- Si $n > 100$, se calculan:
  1. $E(R) = n(n+1)/4$
  2. $V(R)$
     - Si no hay empates: $V(R) = n(n+1)(2n+1)/24$
     - Si hay empates: $V(R) = \dfrac{2n(n+1)(2n+1) - \sum_{j=1}^{r} T_j}{48}$,
       siendo $T_j = t_j^3 - t_j$ y t_j el número de observaciones en el j-ésimo grupo de empates, $1 \leq j \leq r$.

			Rechazamos H_0 a un nivel α si	
	H_0	H_1	$n \leq 100$	$n > 100$
I.	X e Y igual centralización	X e Y distinta centralización	$R(+) \leq T_\alpha(n)$ ó $R(-) \leq T_\alpha(n)$ (*)	$\dfrac{\lvert R(+) - E(R)\rvert}{\sqrt{V(R)}} > z_\alpha$
II.	X e Y igual centralización	X mayor centralización que Y	$R(-) \leq T_{2\alpha}(n)$ (*)	$\dfrac{R(+) - E(R)}{\sqrt{V(R)}} > z_{2\alpha}$
III.	X e Y igual centralización	X menor centralización que Y	$R(+) \leq T_{2\alpha}(n)$ (*)	$\dfrac{R(+) - E(R)}{\sqrt{V(R)}} < -z_{2\alpha}$

(*) Valores $T_\alpha(n)$ en la tabla de Wilcoxon (Tabla IX) para muestras relacionadas.

RESUMEN XIII

Comparación de proporciones (muestras independientes)

$$H_0 : p_1 = p_2 \quad \text{frente} \quad \text{a} \quad H_1 : p_1 \neq p_2$$

Muestras	Características Si	No	
A	x_1	$n_1 - x_1$	n_1
B	x_2	$n_2 - x_2$	n_2

$\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{q}_1 = 1 - \hat{p}_1$.

$\hat{p}_2 = \frac{x_2}{n_2}$, $\hat{q}_2 = 1 - \hat{p}_2$.

$\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2}$, $\hat{q} = 1 - \hat{p}$.

- Condiciones de validez:

  $n_i\hat{p}\hat{q} \geq 5$, $i = 1, 2$.

- Estadístico de contraste y valor teórico:

$$t_{exp} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

  z_α en N(0,1)

- Regla de decisión:

  Si $|t_{exp}| > z_\alpha$, entonces rechazamos H_0 al nivel α.

- Test unilaterales:

	H_0	H_1	Rechazamos H_0 al nivel α, si
I.	$p_1 \leq p_2$	$p_1 > p_2$	$t_{exp} > z_{2\alpha}$
II.	$p_1 \geq p_2$	$p_1 < p_2$	$t_{exp} < -z_{2\alpha}$

- Intervalo de confianza para $p_1 - p_2$ (si se verifican las condiciones de validez: $n_i\hat{p}_i\hat{q}_i \geq 5$, $i = 1, 2$):

$$p_1 - p_2 \in (\hat{p}_1 - \hat{p}_2) \pm z_\alpha\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

RESUMEN XIV

Comparación de proporciones (muestras relacionadas): Test de McNemar.

$$H_0 : p_1 = p_2 \quad \text{frente} \quad \text{a} \quad H_1 : p_1 \neq p_2$$

V1 \ V2	SI	NO	
SI	n_{11}	n_{12}	
NO	n_{21}	n_{22}	
			n

- Condición de validez:

  $n_{12} + n_{21} \geq 20.$

- Estadístico de contraste y valor teórico:

  $$t_{exp} = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$$

  z_α en N(0,1)

- Regla de decisión:

  Si $|t_{exp}| > z_\alpha$, entonces rechazamos H_0 al nivel α.

- Test unilaterales:

	H_0	H_1	Rechazamos H_0 al nivel α, si
I.	$p_1 \leq p_2$	$p_1 > p_2$	$t_{exp} > z_{2\alpha}$
II.	$p_1 \geq p_2$	$p_1 < p_2$	$t_{exp} < -z_{2\alpha}$

- Intervalo de confianza para $p_1 - p_2$ (si $n_{12}, n_{21} > 5$):

$$p_1 - p_2 \in \frac{1}{n}\left((n_{12} - n_{21}) \pm z_\alpha \sqrt{(n_{12} + n_{21}) - \frac{(n_{12} - n_{21})^2}{n}} \right)$$

RESUMEN XV

Análisis de la varianza de una vía (ANOVA de una vía)

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \qquad H_1 : \mu_i \neq \mu_j, \text{ para algún } i \text{ y } j$$

- Muestra:

Nivel del factor			
1	X_{11}	$\ldots$	X_{1n_1}
$\vdots$	$\ldots$	$\ldots$	$\ldots$
k	X_{k1}	$\ldots$	X_{kn_k}

Supondremos que $X_{ij}, 1 \leq j \leq n_i, 1 \leq i \leq k$, son variables independientes con distribución $N(\mu_i, \sigma^2)$. Sean:

$$T_{i\cdot} = \sum_{j=1}^{n_i} X_{ij} \qquad \bar{X}_{i\cdot} = T_{i\cdot}/n_i, \qquad 1 \leq i \leq k.$$

$$T_{\cdot\cdot} = \sum_{i=1}^{k} T_{i\cdot} \qquad \bar{X}_{\cdot\cdot} = T_{\cdot\cdot}/N.$$

- Tabla de Análisis de la Varianza

Fuentes Variación	Suma Cuadrados	Grados Libertad	Media Cuadrática	F
Entre	$V_e = \sum_{i=1}^{k} n_i(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$	$k-1$	$mc_e = \frac{V_e}{k-1}$	$F = \frac{mc_e}{mc_d}$
Dentro	$V_d = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2$	$N-k$	$S_d^2 = mc_d = \frac{V_d}{N-k}$	
Total	$V_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{\cdot\cdot})^2$	$N-1$		

- Decisión: Se rechaza H_0 a un nivel de significación α si $F > F_\alpha(k-1, N-k)$

RESUMEN XVI

Test de Kruskal-Wallis

- Muestra:

Nivel del factor			
1	X_{11}	$\dots$	X_{1n_1}
$\vdots$	$\dots$	$\dots$	$\dots$
k	X_{k1}	$\dots$	X_{kn_k}

- Se mezclan las k muestras en una sola.
- Se ordenan los datos de dicha muestra de menor a mayor y se asignan rangos.
- Se calculan R_i =suma de los rangos correspondientes a los datos de la muestra i-ésima, $1 \leq i \leq k$.

- Condiciones de validez:

$n_i \geq 5$, $1 \leq i \leq k$.

- Estadístico de contraste y valor teórico:

- Valor experimental:

1. Si no hay empates: $H = \dfrac{12}{N(N+1)} \left(\sum_{i=1}^{k} \dfrac{R_i^2}{n_i} \right) - 3(N+1)$.
2. Si hay empates: $H = \dfrac{12}{CN(N+1)} \left(\sum_{i=1}^{k} \dfrac{R_i^2}{n_i} \right) - \dfrac{3(N+1)}{C}$,

siendo $C = 1 - \dfrac{1}{N^3 - N} \sum_{j=1}^{r} T_j$, donde $T_j = t_j^3 - t_j$ y t_j es el número de observaciones en el j-ésimo grupo de empates, $1 \leq j \leq r$.

- Valor teórico: $\chi^2_\alpha(k-1)$

- Regla de decisión:

Se rechaza H_0 al nivel de significación α si $H > \chi^2_\alpha(k-1)$.

RESUMEN XVII

Comparaciones múltiples

• Comparaciones múltiples tras ANOVA de una vía:

$$H_0 : \mu_i = \mu_j \text{ para todo } i, j \qquad H_1 : \mu_i \neq \mu_j \text{ para algún par } (i, j)$$

Método de Tukey-Kramer:

1. Sean $m_1 \leq \cdots \leq m_k$ las k medias muestrales ordenadas de menor a mayor, $\mu'_1, \ldots, \mu'_k$ las medias poblacionales correspondientes y $n'_1, \ldots, n'_k$ los tamaños muestrales. Sea $N = n'_1 + \cdots + n'_k$. Hagamos $i = 1$ y $j = k$.
2. Si $n'_1 = \cdots = n'_k = n$, se comparan las medias μ'_i y μ'_j utilizando el estadístico de rango estudentizado
$$q_{ij} := \frac{m_j - m_i}{\sqrt{\frac{S_d^2}{2}\left(\frac{1}{n'_j} + \frac{1}{n'_i}\right)}}$$
3. Las medias μ'_i y μ'_j se declaran homogéneas si $q_{ij} \leq q_\alpha(N-k, k)$ (véase la Tabla VIII de la distribución q de Tukey). Se declaran igualmente homogéneas todas las medias comprendidas entre ellas, y no es necesario comparar dos medias dentro de un grupo homogéneo.
4. Las medias μ'_i y μ'_j se declaran no homogéneas si $q_{ij} > q_\alpha(N-k, k)$. En ese caso, se vuelve al paso 2 reemplazando i por $i+1$ para comparar las medias μ'_{i+1} con μ'_j.
5. Si μ'_j se declara no homogénea a μ'_i para todo $i < j$, concluiremos que existen diferencias significativas al nivel α entre μ'_j y todas las medias menores que ella. Se reemplaza j por $j-1$ y se vuelve al paso 2.
6. El proceso continúa hasta que no quede nada que comparar.

• Comparaciones múltiples tras el test de Kruskal-Wallis: Se procede como en el método de Tukey-Kramer ordenando los rangos medios R_i/n_i de menor a mayor, y las muestras i y j se declaran homogéneas (junto con cualquier otra muestra con rango medio contenido entre ambos) si

$$\frac{\left|\dfrac{R_i}{n_i} - \dfrac{R_j}{n_j}\right|}{\sqrt{\dfrac{CN(N+1)}{12}\left(\dfrac{1}{n_i} + \dfrac{1}{n_j}\right)}} \leq Q_\alpha(k)$$

(C es como en Resumen XVI y $Q_\alpha(k)$ en Tabla XI).

RESUMEN XVIII

Test de homogeneidad de varias muestras (caso discreto)

- Muestra:

O_{ij}	1	...	j	...	s	Totales
1	O_{11}	...	O_{1j}	...	O_{1s}	F_1
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	O_{i1}	...	O_{ij}	...	O_{is}	F_i
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
r	O_{r1}	...	O_{rj}	...	O_{rs}	F_r
Totales	C_1	...	C_j	...	C_s	T

-Valores esperados: $E_{ij} = F_i C_j / T$, $i = 1, \ldots, r$; $j = 1, \ldots, s$.

- Condiciones de validez:

  - Ningún E_{ij} es < 1.

  - A lo sumo un 20% de los E_{ij} son < 5.

- Estadístico de contraste y valor teórico:

  $\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_i \sum_j \frac{O_{ij}^2}{E_{ij}} - T.$

  $\chi^2_\alpha((r-1)(s-1)).$

- Regla de decisión:

  Se rechaza H_0 al nivel de significación α si $\chi^2 > \chi^2_\alpha((r-1)(s-1))$.

- Observaciones:

1. Para una **tabla** 2×2 las condiciones de validez son: a) los marginales F_i y C_j son todos mayores que $T/10$; b) todas las E_{ij} son > 5. El valor experimental es:

$$\chi^2 = \frac{(O_{11}O_{22} - O_{12}O_{21})^2}{F_1 F_2 C_1 C_2} T.$$

2. La **comparación de varias proporciones para muestras independientes** es un caso particular de este test en el que $s = 2$.

RESUMEN XIX

Relación entre variables cuantitativas

Regresión lineal

- Modelo: $Y = \alpha + \beta X + \varepsilon$, ε sigue distribución N(0,σ^2)
- Muestra: Sea (X_i,Y_i) $i = 1, \ldots, n$ una muestra de la población bidimensional (X, Y).
- Estimación puntual de α, β y σ^2:

  1. **Estimador puntual de β:** $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \dfrac{\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)/n}{\sum x_i^2 - \left(\sum x_i\right)^2/n}$.
  2. **Estimador puntual de α:** $a = \bar{y} - b\bar{x}$.
  3. **Recta de regresión muestral:** $y = a + bx$.
  4. **Estimador puntual de σ^2:** $s^2 = \dfrac{1}{n-2}[S_{yy} - bS_{xy}]$, siendo $S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \left(\sum y_i\right)^2/n$.

- Contraste de hipótesis e intervalos de confianza:

  1. **Contraste de la independencia lineal:** $H_0 : \beta = 0$ frente a $H_1 : \beta \neq 0$

     Se rechaza H_0 al nivel de significación α si $t_{exp} = \dfrac{|b|}{s/\sqrt{S_{xx}}} > t_\alpha(n-2)$.
  2. **Intervalo de confianza para β:** $\beta \in b \pm t_\alpha(n-2)\dfrac{s}{\sqrt{S_{xx}}}$.
  3. **Intervalo de confianza para α:** $\alpha \in a \pm t_\alpha(n-2)s\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}}$.
  4. **Intervalo de confianza para la media de Y dado un valor de X:** $\alpha + \beta x \in a + bx \pm t_\alpha(n-2)s\sqrt{\dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{S_{xx}}}$.
  5. **Intervalo de confianza para una predicción:** $y \in a + bx \pm t_\alpha(n-2)s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{S_{xx}}}$

Correlación Lineal

- Estimación puntual del coeficiente de correlación poblacional ρ:

  Coeficiente de correlación lineal muestral: $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$.

- Contraste de la independencia lineal: $H_0 : \rho = 0$ frente a $H_1 : \rho \neq 0$

  Se rechaza H_0 a un nivel de significación α si $t_{exp} = \sqrt{\dfrac{(n-2)r^2}{1-r^2}} > t_\alpha(n-2)$.

Correlación No Paramétrica

- Coeficiente de correlación de Spearman: $r_s = 1 - \dfrac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - R_i')^2$,
  siendo (R_i, R_i'), $i = 1, \ldots, n$ los pares de rangos asociados a las observaciones (x_i, y_i), $i = 1, \ldots, n$
- Contraste de independencia: $H_0 : X$ e Y son independientes

  Si $n \leq 50$, se rechaza H_0 al nivel de significación α si $|r_s| > r_\alpha(n)$, $r_\alpha(n)$ en la tabla de Spearman.

  Si $n > 50$, se rechaza H_0 al nivel de significación α si $|r_s|\sqrt{n-1} > z_\alpha$.

RESUMEN XX

Relación entre variables cualitativas

- Test de Hipótesis:

  Sean A y B dos caracteres cualitativos con r y s niveles, respectivamente.

  H_0 : A y B son independientes $\qquad$ H_1 : A y B son dependientes

- Muestra:

O_{ij}	B_1	$\ldots$	B_j	$\ldots$	B_s	Totales
A_1	O_{11}	$\ldots$	O_{1j}	$\ldots$	O_{1s}	F_1
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	O_{i1}	$\ldots$	O_{ij}	$\ldots$	O_{is}	F_i
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_r	O_{r1}	$\ldots$	O_{rj}	$\ldots$	O_{rs}	F_r
Totales	C_1	$\ldots$	C_j	$\ldots$	C_s	T

  -Valores esperados: $E_{ij} = F_i C_j / T$, $i = 1, \ldots, r$; $j = 1, \ldots, s$.

- Condiciones de validez:

  - Ningún E_{ij} es < 1.

  - A lo sumo un 20% de los E_{ij} son < 5.

- Estadístico de contraste y valor teórico:

  $\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_i \sum_j \frac{O_{ij}^2}{E_{ij}} - T.$

  $\chi^2_\alpha((r-1)(s-1)).$

- Regla de decisión:

  Se rechaza H_0 al nivel de significación α si $\chi^2 > \chi^2_\alpha((r-1)(s-1))$.

- Observaciones: Para una **tabla** 2×2 las condiciones de validez son: a) los marginales F_i y C_j son todos mayores que $T/10$; b) todas las E_{ij} son > 5. El valor experimental es:

$$\chi^2 = \frac{(O_{11}O_{22} - O_{12}O_{21})^2}{F_1 F_2 C_1 C_2} T.$$

RESUMEN XXI

Medidas de asociación en tablas de contingencias

- TABLAS $r \times s$: COEFICIENTE DE CONTINGENCIA DE PEARSON

$$C = \sqrt{\frac{\chi^2}{T + \chi^2}}$$

C toma valores entre 0 (asociación nula) y $\sqrt{\frac{q-1}{q}}$, siendo $q = \min\{r, s\}$.

- TABLAS 2×2: COEFICIENTE PHI:

$$\phi = \sqrt{\frac{\chi^2}{T}}$$

ϕ toma valores entre 0 (asociación nula) y 1 (asociación máxima).

- TABLAS 2×2: DIFERENCIA DE RIESGOS:

Estimación:

$$\hat{D} = \frac{O_{11}}{F_1} - \frac{O_{21}}{F_2}.$$

Toma valores entre -1 y 1, correspondiendo el valor cero a la independencia entre la enfermedad y el factor de riesgo.

- TABLAS 2×2: RIESGO RELATIVO:

Estimación:

$$\hat{R} = \frac{\frac{O_{11}}{F_1}}{\frac{O_{21}}{F_2}} = \frac{O_{11}F_2}{O_{21}F_1}.$$

Test de hipótesis: La región crítica contrastar la hipótesis nula $H_0 : R = 1$ de no asociación entre factor de riesgo y enfermedad contra la hipótesis alternativa $H_1 : R \neq 1$ es

$$\frac{(\log \hat{R})^2}{S^2_{\log \hat{R}}} > \chi^2_\alpha(1).$$

Intervalo de confianza: un intervalo al $(1-\alpha)100\%$ de confianza para el riesgo relativo es

$$e^{\log \hat{R} \pm z_\alpha S_{\log \hat{R}}}.$$

- TABLAS 2×2: RAZÓN DEL PRODUCTO CRUZADO MUESTRAL:

Estimación:

$$\widehat{OR} = \frac{\frac{O_{11}/F_1}{O_{12}/F_1}}{\frac{O_{21}/F_2}{O_{22}/F_2}} = \frac{O_{11}O_{22}}{O_{21}O_{12}}.$$

Test de hipótesis: La región crítica contrastar la hipótesis nula $H_0 : OR = 1$ de no asociación entre factor de riesgo y enfermedad contra la hipótesis alternativa $H_1 : OR \neq 1$ es

$$\frac{(\log \widehat{OR})^2}{S^2_{\log \widehat{OR}}} > \chi^2_\alpha(1),$$

Intervalo de confianza: un intervalo al $(1-\alpha)100\%$ de confianza para la razón del producto cruzado es

$$e^{\log \widehat{OR} \pm z_\alpha S_{\log \widehat{OR}}}.$$