RESUMEN PARA CONTROL 3

Estadística

INGENIERÍA MULTIMEDIA

Violeta Migallón

RESUMEN PARA CONTROL

Intervalos de confianza para la media (μ)

Suponiendo normalidad

– Si la desviación típica poblacional (σ) es conocida, entonces

$$\left[\overline{x} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \overline{x} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right],$$

– Si σ es desconocida, entonces

$$\left[\overline{x} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2},n-1}, \overline{x} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2},n-1}\right],$$

Intervalos de confianza para la media (μ)

Muestras grandes (no normal)

– Si σ es conocida, entonces

$$\left[\overline{x} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \overline{x} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right],$$

– Si σ es desconocida, entonces

$$\left[\overline{x} - \frac{S}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \overline{x} + \frac{S}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right],$$

RESUMEN PARA CONTROL

Intervalos de confianza para la varianza σ^2

Suponiendo normalidad

– Si μ es conocida, entonces

$$\left[\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\frac{\alpha}{2},n}}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right],$$

– Si μ es desconocida, entonces

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

RESUMEN PARA CONTROL

Intervalos de confianza para proporciones (p)

- **Intervalo de confianza para p,**

$$\left[\overline{p} - \frac{1}{2\sqrt{n\alpha}}, \overline{p} + \frac{1}{2\sqrt{n\alpha}}\right],$$

donde $\overline{p}$ es la estimación de p en la muestra. El intervalo que acabamos de construir es demasiado amplio y da una estimación por intervalo pobre. En la práctica se suele utilizar el siguiente intervalo, que corresponde a muestras de tamaño grande:

- **Intervalo de confianza para p, correspondiente a muestras grandes**

$$\left[\overline{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}, \overline{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$

RESUMEN PARA CONTROL

Intervalos de confianza para μ1-μ2

Bajo normalidad y muestras independientes

- Si σ_1 y σ_2 son conocidas, el intervalo de confianza para $\mu_1 - \mu_2$ con nivel $(1-\alpha)100\%$ es

$$\left[\overline{x} - \overline{y} - \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}Z_{\frac{\alpha}{2}}, \overline{x} - \overline{y} + \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}Z_{\frac{\alpha}{2}}\right].$$

- Si σ_1 y σ_2 son desconocidas pero iguales, el intervalo de confianza para $\mu_1 - \mu_2$ con nivel $(1-\alpha)100\%$ es

$$\left[\overline{x} - \overline{y} - S_p\sqrt{\frac{1}{m} + \frac{1}{n}}t_{\frac{\alpha}{2},m+n-2}, \overline{x} - \overline{y} + S_p\sqrt{\frac{1}{m} + \frac{1}{n}}t_{\frac{\alpha}{2},m+n-2}\right],$$

siendo

$$S_p = \sqrt{\frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}}.$$

RESUMEN PARA CONTROL

Intervalos de confianza para µ1-µ2

Bajo normalidad y muestras independientes

- Si σ_1 y σ_2 son desconocidas y distintas, el intervalo de confianza para $\mu_1 - \mu_2$ con nivel $(1-\alpha)100\%$ es

$$\left[\overline{x} - \overline{y} - \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}\, t_{\frac{\alpha}{2},\nu}, \overline{x} - \overline{y} + \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}\, t_{\frac{\alpha}{2}\nu}\right],$$

con

$$\nu = \frac{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2}{\frac{\left(\frac{S_1^2}{m}\right)^2}{m-1} + \frac{\left(\frac{S_2^2}{n}\right)^2}{n-1}}.$$

RESUMEN PARA CONTROL

Intervalos de confianza para µ1-µ2

Sin normalidad y muestras independientes

- Si σ_1 y σ_2 son desconocidas y las muestras son grandes, aunque no provengan de una normal, el intervalo de confianza para $\mu_1 - \mu_2$ con nivel $(1-\alpha)100\%$ es

$$\left[\overline{x} - \overline{y} - \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} Z_{\frac{\alpha}{2}}, \overline{x} - \overline{y} + \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} Z_{\frac{\alpha}{2}}\right].$$

RESUMEN PARA CONTROL

TABLA 1: Contrastes de hipótesis bajo normalidad

Hipótesis nula	Estadístico de prueba	Hipótesis alternativa	Criterios de rechazo	P-valor
$H_0 : \mu = \mu_0$		$H_1 : \mu \neq \mu_0$	$\lvert Z_0 \rvert > Z_{\frac{\alpha}{2}}$	$2P(Z \geq \lvert Z_0 \rvert)$
σ^2 conocida	$Z_0 = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$H_1 : \mu > \mu_0$	$Z_0 > Z_\alpha$	$P(Z \geq Z_0)$
		$H_1 : \mu < \mu_0$	$Z_0 < -Z_\alpha$	$P(Z \leq Z_0)$
$H_0 : \mu = \mu_0$		$H_1 : \mu \neq \mu_0$	$\lvert t_0 \rvert > t_{\frac{\alpha}{2},n-1}$	$2P(t_{n-1} \geq \lvert t_0 \rvert)$
σ^2 desconocida	$t_0 = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}}$	$H_1 : \mu > \mu_0$	$t_0 > t_{\alpha,n-1}$	$P(t_{n-1} \geq t_0)$
		$H_1 : \mu < \mu_0$	$t_0 < -t_{\alpha,n-1}$	$P(t_{n-1} \leq t_0)$
$H_0 : \sigma^2 = \sigma_0^2$		$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi^2_{\frac{\alpha}{2},n-1}$ ó $\chi_0^2 < \chi^2_{1-\frac{\alpha}{2},n-1}$	$2P$(cola menor)
	$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$H_1 : \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{\alpha,n-1}$	$P(\chi^2_{n-1} \geq \chi_0^2)$
		$H_1 : \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi^2_{1-\alpha,n-1}$	$P(\chi^2_{n-1} \leq \chi_0^2)$

RESUMEN PARA CONTROL

TABLA 2: Contrastes de hipótesis bajo normalidad

Hipótesis nula	Estadístico de prueba	Hipótesis alternativa	Criterios de rechazo	P-valor
$H_0: \mu_1 = \mu_2$ $\sigma_1^2,\ \sigma_2^2$ conocidas	$Z_0 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$\lvert Z_0 \rvert > Z_{\frac{\alpha}{2}}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$2P(Z \geq \lvert Z_0 \rvert)$ $P(Z \geq Z_0)$ $P(Z \leq Z_0)$
$H_0: \mu_1 = \mu_2$ $\sigma_1^2 = \sigma_2^2$ desconocidas	$t_0 = \frac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$\lvert t_0 \rvert > t_{\frac{\alpha}{2}, n_1+n_2-2}$ $t_0 > t_{\alpha, n_1+n_2-2}$ $t_0 < -t_{\alpha, n_1+n_2-2}$	$2P(t_{n_1+n_2-2} \geq \lvert t_0 \rvert)$ $P(t_{n_1+n_2-2} \geq t_0)$ $P(t_{n_1+n_2-2} \leq t_0)$
$H_0: \mu_1 = \mu_2$ $\sigma_1^2 \neq \sigma_2^2$ desconocidas	$t_0 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$H_1: \mu_1 \neq \mu_2$ $H_1: \mu_1 > \mu_2$ $H_1: \mu_1 < \mu_2$	$\lvert t_0 \rvert > t_{\frac{\alpha}{2}, \nu}$ $t_0 > t_{\alpha, \nu}$ $t_0 < -t_{\alpha, \nu}$	$2P(t_\nu \geq \lvert t_0 \rvert)$ $P(t_\nu \geq t_0)$ $P(t_\nu \leq t_0)$
$H_0: \sigma_1^2 = \sigma_2^2$	$F_0 = \frac{S_1^2}{S_2^2}$	$H_1: \sigma_1^2 \neq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$F_0 > F_{\frac{\alpha}{2}, n_1-1, n_2-1}$ ó $F_0 < F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$ $F_0 > F_{\alpha, n_1-1, n_2-1}$	$2P(\text{cola menor})$ $P(F_{n_1-1, n_2-1} \geq F_0)$

RESUMEN PARA CONTROL

FÓRMULAS TABLA 2

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}},$$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

RESUMEN PARA CONTROL

TABLA 3: Contrastes de hipótesis para muestras grandes

Hipótesis nula	Estadístico de prueba	Hipótesis alternativa	Criterios de rechazo	P-valor
$H_0 : \mu = \mu_0$	$Z_0 = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$	$H_1 : \mu \neq \mu_0$ $H_1 : \mu > \mu_0$ $H_1 : \mu < \mu_0$	$\|Z_0\| > Z_{\frac{\alpha}{2}}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$2P(Z \geq \|Z_0\|)$ $P(Z \geq Z_0)$ $P(Z \leq Z_0)$
$H_0 : \mu_1 = \mu_2$	$Z_0 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$H_1 : \mu_1 \neq \mu_2$ $H_1 : \mu_1 > \mu_2$ $H_1 : \mu_1 < \mu_2$	$\|Z_0\| > Z_{\frac{\alpha}{2}}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$2P(Z \geq \|Z_0\|)$ $P(Z \geq Z_0)$ $P(Z \leq Z_0)$
$H_0 : p = p_0$	$Z_0 = \frac{p^* - p_0}{\sqrt{\frac{(1-p_0)p_0}{n}}}$	$H_1 : p \neq p_0$ $H_1 : p > p_0$ $H_1 : p < p_0$	$\|Z_0\| > Z_{\frac{\alpha}{2}}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$2P(Z \geq \|Z_0\|)$ $P(Z \geq Z_0)$ $P(Z \leq Z_0)$
$H_0 : p_1 = p_2$	$Z_0 = \frac{p_1^* - p_2^*}{\sqrt{p^*(1-p^*)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$H_1 : p_1 \neq p_2$ $H_1 : p_1 > p_2$ $H_1 : p_1 < p_2$	$\|Z_0\| > Z_{\frac{\alpha}{2}}$ $Z_0 > Z_\alpha$ $Z_0 < -Z_\alpha$	$2P(Z \geq \|Z_0\|)$ $P(Z \geq Z_0)$ $P(Z \leq Z_0)$

RESUMEN PARA CONTROL

- Contrastes de hipótesis: contingencia y homogeneidad

El test χ^2 se basa en la siguiente regla: Rechazar H_0 si

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(\theta_{ij} - E_{ij})^2}{E_{ij}} \geq \chi_{\alpha,\nu}^2,$$

donde $\nu = (r-1)(c-1)$. El valor r es el número de filas de la tabla y c el número de columnas. El valor θ_{ij} es la frecuencia absoluta observada en la celda (i,j) y E_{ij} es la frecuencia esperada bajo la hipótesis nula. Por último, α es el nivel de significación.

Otra forma de estudiar si se rechaza el test es usando el P-valor. Así, se rechaza la hipótesis nula para todo α tal que

$$\alpha > P(\chi_\nu^2 \geq \chi_0^2).$$

RESUMEN PARA CONTROL

Estudio de las celdas de interés:

En caso de rechazar la hipótesis nula parece interesante estudiar qué casillas han sido las que han contribuido en mayor medida a esto.

Para ello, calculamos para cada casilla

$$Z_{ij} = \frac{(\theta_{ij} - E_{ij})}{\sqrt{E_{ij}}},$$

y asignamos símbolos a las casillas de la siguiente forma:

- Si $|Z_{ij}| \leq 1.645$ asignamos el símbolo $\cdot$.
- Si $1.645 < |Z_{ij}| \leq 1.96$ asignamos el símbolo o.
- Si $1.96 < |Z_{ij}| \leq 2.576$ asignamos el símbolo O.
- Si $|Z_{ij}| > 2.576$ asignamos el símbolo @.

RESUMEN PARA CONTROL

- H_0: la muestra procede de la población especificada
- H_1: no H_0

El test de bondad de ajuste se basa en la regla: Rechazar H_0 si $\chi_0^2 = \sum_{i=1}^{k} \frac{(\theta_i - E_i)^2}{E_i} \geq \chi^2_{\alpha,k-1}$, donde θ_i es la frecuencia observada en cada categoría, E_i la frecuencia esperada de cada categoría bajo la hipótesis nula y k el número de categorías.

Otra forma de estudiar si se rechaza el test es con el P-valor. Así se rechaza la hipótesis nula para todo α tal que

$$\alpha > P(\chi^2_{k-1} \geq \chi^2_0).$$