Marco Minozzo e Annamaria Guolo
Laurea in Economia del Commercio Internazionale
Laurea in Economia e Amministrazione delle Imprese
Università degli Studi di Verona (sede di Vicenza)

Statistica Inferenziale
Soluzioni 3. Verifica di ipotesi

Introduzione.
Siano $X_1, \ldots, X_n$ v.a. indipendenti e con identica distribuzione. Sia $f(x_1, \ldots, x_n; \theta)$ la densità o la distribuzione di probabilità congiunta di $X_1, \ldots, X_n$, con parametro incognito θ, appartenente allo spazio parametrico $\Theta \subseteq \mathbb{R}^p$. Sia $\{\Theta_0, \Theta_1\}$ una partizione di Θ e sia dato in corrispondenza il sistema di verifica d'ipotesi, $H_0 : \theta \in \Theta_0$ (*ipotesi nulla*) contro $H_1 : \theta \in \Theta_1$ (*ipotesi alternativa*).
Data una statistica T, condurre la verifica d'ipotesi per valutare l'appartenenza o meno di θ a Θ_0 consiste nella divisione dello spazio campionario in una zona di rifiuto R (nella quale H_0 viene rifiutata) ed in una zona di accettazione A (nella quale H_0 viene accettata).
Nel condurre la procedura si possono compiere due tipi di errore:

- *errore di primo tipo*: si rifiuta H_0 quando in realtà è vera. La probabilità di compiere l'errore di primo tipo si indica solitamente con α, livello di significatività;

- *errore di secondo tipo*: si accetta H_0 quando in realtà è falsa. La probabilità di compiere questo errore si indica solitamente con β. La quantità $(1 - \beta)$ è nota come *potenza* del test.

Non è possibile annullare le probabilità di commettere entrambi gli errori. Solitamente, si procede cercando la regione di rifiuto (o la soglia critica che la definisce) in modo che l'errore di primo tipo (il livello di significatività del test) sia pari ad un valore prefissato, spesso 0,05.

Esemplificazioni.

- *Verifica d'ipotesi sulla media di una popolazione normale con varianza nota - ipotesi semplici*. Siano $X_1, \ldots, X_n$ v.a. indipendenti ed identicamente distribuite come $N(\mu, \sigma^2)$. Sia dato il sistema di verifica d'ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu = \mu_1$ (ipotesi semplici), con $\mu_0 < \mu_1$. Consideriamo la statistica $\overline{X}$ che, sotto H_0 , vale a dire quando $\mu = \mu_0$, ha distribuzione
$$\overline{X} \sim N(\mu_0, \sigma^2/n).$$
Il test rifiuta H_0 in favore di H_1 per valori grandi della statistica $\overline{X}$, vale a dire che la regione di rifiuto è formata dai valori di X tali che $\overline{x} > k$, $R = \{x : \overline{x} > k\}$. Il valore k si determina in modo che il livello del test sia pari ad α, vale a dire in modo che
$$P\left(\overline{X} > k; \mu = \mu_0\right) = \alpha.$$
Considerando la distribuzione di $\overline{X}$ sotto H_0 si ha che $k = \mu_0 + z_\alpha\sqrt{\sigma^2/n}$, dove z_α è il quantile di una normale standard che lascia alla propria destra una probabilità pari a α.

- *Verifica d'ipotesi sulla media di una popolazione normale con varianza nota - ipotesi alternativa composta*. Si considerino le v.a. distribuite come $N(\mu, \sigma^2)$ del punto precedente. Sia ora $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ (ipotesi alternativa composta). Il test in questione è un test bilaterale o a due code. Consideriamo la statistica test
$$Z = \frac{\overline{X} - \mu_0}{\sqrt{\sigma^2/n}},$$

che sotto H_0 ha distribuzione normale standard. Allora la regione di rifiuto è $R = \{z : |z| > z_{\alpha/2}\}$, ovvero

$$R = \{x : \overline{x} < \mu_0 - z_{\alpha/2}\sqrt{\sigma^2/n}\} \cup \{x : \overline{x} > \mu_0 + z_{\alpha/2}\sqrt{\sigma^2/n}\}.$$

Nel caso di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu < \mu_0$ (test unilaterale o ad una coda), si ha $R = \{z : z < -z_\alpha\}$. Nel caso di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$ si ha $R = \{z : z > z_{1-\alpha}\}$.

- *Verifica d'ipotesi sulla media di una popolazione normale con varianza incognita - ipotesi alternativa composta*. Nelle ipotesi dei casi precedenti, sia dato il sistema di verifica d'ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$. Consideriamo la statistica test

$$t = \frac{\overline{X} - \mu_0}{\sqrt{S^2/n}},$$

che sotto H_0 ha distribuzione t_{n-1}. Allora la regione di rifiuto è $R = \{t : |t| > t_{n-1;\alpha/2}\}$, dove $t_{n-1;\alpha/2}$ è il quantile di una t_{n-1} che lascia alla propria destra una probabilità pari a $\alpha/2$. Nel caso di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu < \mu_0$, si ha $R = \{t : t < -t_\alpha\}$. Nel caso di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$ si ha $R = \{t : t > t_\alpha\}$.

- *Verifica d'ipotesi sulla varianza di una popolazione normale con media incognita - ipotesi alternativa composta*. Nelle ipotesi dei casi precedenti, sia dato il sistema di verifica d'ipotesi $H_0 : \sigma^2 = \sigma_0^2$ contro $H_1 : \sigma^2 \neq \sigma_0^2$. Consideriamo la statistica test

$$\chi_c^2 = \frac{S^2(n-1)}{\sigma_0^2},$$

che sotto H_0 ha distribuzione χ^2_{n-1}. Allora la regione di rifiuto risulta pari a

$$R = \{\chi_c^2 : \chi_c^2 < \chi^2_{n-1;1-\alpha/2}\} \cup \{\chi_c^2 : \chi_c^2 > \chi^2_{n-1;\alpha/2}\}.$$

- *Verifica d'ipotesi sull'uguaglianza delle medie di due popolazioni normali con uguale varianza incognita*. Siano $X_1, \ldots, X_n$ v.a. indipendenti ed identicamente distribuite come $N(\mu_X, \sigma_X^2)$ e siano $Y_1, \ldots, Y_m$ v.a. indipendenti ed identicamente distribuite come $N(\mu_Y, \sigma_Y^2)$. Si considerino un campione di dimensione n da $X_1, \ldots, X_n$ e un campione di dimensione m da $Y_1, \ldots, Y_m$. Sia dato il sistema di verifica d'ipotesi $H_0 : \mu_X = \mu_Y$ contro $H_1 : \mu_X \neq \mu_Y$. Si supponga che le due varianze siano uguali, $\sigma_X^2 = \sigma_Y^2$. Siano S_X^2 e S_Y^2 le varianze campionarie corrette di X e Y. Consideriamo la statistica test

$$D = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_x^2(n-1)+S_Y^2(m-1)}{n+m-2}(1/n + 1/m)}},$$

che sotto H_0 ha distribuzione t_{n+m-2}. La regione di rifiuto è $R = \{d : |d| > t_{n+m-2;\alpha/2}\}$.

- *Verifica d'ipotesi sull'uguaglianza delle varianze di due popolazioni normali con medie ignote*. Nelle ipotesi del punto precedente, si consideri il sistema di verifica d'ipotesi $H_0 : \sigma_X^2 = \sigma_Y^2$ contro $H_1 : \sigma_X^2 \neq \sigma_Y^2$. Consideriamo la statistica test F data dal rapporto tra le varianze campionarie corrette

$$F = \frac{S_X^2}{S_Y^2}.$$

Allora, sotto H_0, $F \sim F_{n-1,m-1}$. La regione di rifiuto del test è $R = \{F : |F| > F_{n-1,m-1;\alpha/2}\}$, dove $F_{n-1,m-1;\alpha/2}$ è il quantile di una $F_{n-1,m-1}$ che lascia alla propria destra una probabilità pari a $\alpha/2$. Nel caso di ipotesi $H_0 : \sigma_X^2 = \sigma_Y^2$ contro $H_1 : \sigma_X^2 < \sigma_Y^2$, si ha $R = \{F : F < -F_\alpha\}$. Nel caso di ipotesi $H_0 : \sigma_X^2 = \sigma_Y^2$ contro $H_1 : \sigma_X^2 > \sigma_Y^2$ si ha $R = \{F : F > F_\alpha\}$.

- *Verifica d'ipotesi per il confronto di proporzioni.* Si considerino due campioni casuali da $X_1, \ldots, X_n$ v.a. indipendenti distribuite come $Ber(p_1)$ e $Y_1, \ldots, Y_m$ distribuite come $Ber(p_2)$. Siano $\hat{p}_1$ e $\hat{p}_2$ le frequenze relative derivate dai due campioni. Supponendo che n e m siano sufficientemente grandi si ha che

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \hat{p}(1-\hat{p})(1/n + 1/m)\right),$$

dove $\hat{p} = (n\hat{p}_1 + m\hat{p}_2)/(n+m)$. Allora, per la verifica d'ipotesi $H_0 : p_1 = p_2$ contro $H_1 : p_1 \neq p_2$, facciamo riferimento alla statistica test

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n + 1/m)}},$$

che sotto H_0 tende ad una normale standard. Allora la regione di rifiuto è $R = \{z : |z| > z_{\alpha/2}\}$. Nel caso di ipotesi $H_0 : p_1 = p_2$ contro $H_1 : p_1 < p_2$, si ha $R = \{z : z < -z_{\alpha/2}\}$. Nel caso di ipotesi $H_0 : p_1 = p_2$ contro $H_1 : p_1 > p_2$, si ha $R = \{z : z > z_{\alpha/2}\}$.

Esercizio A.
a) Dai 10 valori campionari osservati si ottiene una media campionaria pari a $\bar{x} = 14,2$ e una varianza campionaria corretta pari a $S^2 = 112,63$. Si rifiuta H_0 per valori di $\bar{X}$ maggiori di $\bar{x}_\alpha$ dove $\bar{x}_\alpha$ è tale che $P(\bar{X} \geq \bar{x}_\alpha; H_0) = \alpha = P(T \geq \frac{\bar{x}_\alpha - 8}{S/\sqrt{n}})$ dove T ha distribuzione t di Student con 9 g.d.l. Dalle tavole si trova che $\frac{\bar{x}_\alpha - 8}{S/\sqrt{n}} = 2,821$ da cui $\bar{x}_\alpha = 2,821 \cdot 3,356 + 8 = 17,467$. Essendo $\bar{x} = 14,2$ minore del valore soglia, si accetta H_0.
b) La probabilità richiesta è data da

$$P(\bar{X} > 15,59; \mu = 8) = P\left(T > \frac{15,59 - 8}{\sqrt{112,63/10}}; \mu = 8\right) = P(T > 2,262) = 0,025,$$

dove $T = (\bar{X} - 8)/\sqrt{112,63/10}$ ha distribuzione t di Student con 9 g.d.l.
c) Si rifiuta l'ipotesi nulla per valori osservati di S^2 maggiori di s^2_α, dove s^2_α è tale che $P(S^2 \geq s^2_\alpha; H_0) = \alpha = P\left[\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)s^2_\alpha}{\sigma_0^2}; H_0\right]$, dove σ_0^2 è il valore di σ^2 sotto l'ipotesi nulla e $\frac{(n-1)S^2}{\sigma_0^2}$ si distribuisce come una χ^2 con 9 g.d.l. Dalle tavole si trova che $\frac{(n-1)s^2_\alpha}{\sigma_0^2} = 16,919$, da cui $s^2_\alpha = 16,919 \cdot 10/9 = 18,8$, ed essendo S^2 pari a $112,63$, si rifiuta H_0.

Esercizio B.
La v.a. media campionaria $\overline{X}$ risulta distribuita come una normale di media μ e varianza σ^2/n. Sulla base del test di verifica d'ipotesi indicato, si ha che

$$\alpha = 0,05 = P\left(\overline{X} > 147\right),$$

da cui si ricava che

$$\Phi\left(\frac{\overline{x} - 145}{\sigma/\sqrt{30}}\right) = 0,05.$$

Da qui risulta che $\sigma^2 = 44,6$.
b) La potenza del test è data

$$P\left(\frac{\overline{X} - 150}{\sigma/\sqrt{20}} > \frac{147 - 150}{\sigma/\sqrt{20}}\right) = 0,993.$$

Esercizio C.
a) Si rifiuta l'ipotesi nulla per valori osservati della differenza $D = \bar{X} - \bar{Y}$ superiori a $D_{\alpha/2}$ oppure inferiori a $-D_{\alpha/2}$, dove $D_{\alpha/2}$ è tale che $P(|D| \geq D_{\alpha/2}; H_0) = \alpha = P\left(|Z| \geq \frac{D_{\alpha/2}}{S_p\sqrt{1/n_1 + 1/n_2}}\right)$, e dove $S^2_p = (S^2_X(n_1 - 1) + S^2_Y(n_2 - 1))/(n_1 + n_2 - 2)$ e Z ha approssimativamente distribuzione normale standardizzata. Dalle tavole si trova che $\frac{D_{\alpha/2}}{S_p\sqrt{1/n_1 + 1/n_2}} = 2,576$, da cui $D_{\alpha/2} = 2,576 \cdot S_p\sqrt{1/n_1 + 1/n_2}$. Ed essendo $S_p = 1,12$, si ha che $D_{\alpha/2} = 0,178$. Per cui avendo ossevato un valore di D pari a $-0,2$, si rifiuta H_0. Si noti che mancando l'assunzione di normalità dei dati di partenza, ed essendo n_1 e n_2 sufficientemente grandi, abbiamo assunto che Z fosse approssimativamente distribuita come una normale standardizzata in base all'estensione del teorema del limite centrale.
b) Per questa verifica di ipotesi, si rifiuta l'ipotesi nulla $H_0 : \mu_X = \mu_Y$ per valori di D maggiori di $D_\alpha = z_\alpha\ \sigma\ \sqrt{1/n_1 + 1/n_2} = 0,16027$, dove $z_\alpha = 2,326$. Quindi, la potenza del test è data da $P(D > D_\alpha; H_1) = P\left(Z > \frac{0,16027 - 0,5}{\sqrt{1,25(1/n_1 + 1/n_2)}}\right) = P(Z > -4,93) = 1$, dove Z ha approssimativamente distribuzione normale standardizzata.

c) Si rifiuta l'ipotesi nulla per valori osservati della differenza $D = \hat{p}_{95} - \hat{p}_{96}$ inferiori a D_α, con D_α tale che $P(D \leq D_\alpha; H_0) = \alpha = P\left(Z \leq \frac{D_\alpha}{S_{\hat{p}_{95}-\hat{p}_{96}}}\right)$, dove $S_{\hat{p}_{95}-\hat{p}_{96}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $\hat{p} = (\hat{p}_{95}n_1 + \hat{p}_{96}n_2)/(n_1+n_2)$, $n_1 = 80$, $n_2 = 90$ e Z ha approssimativamente distribuzione normale standardizzata. Dalle tavole si trova che il percentile che lascia alla sua destra una probabilità del 95% è pari a $-1,645$, per cui $\frac{D_\alpha}{S_{\hat{p}_{95}-\hat{p}_{96}}} = -1,645$, ed essendo $\hat{p}_{95} = 0,624$, $\hat{p}_{96} = 0,645$, $\hat{p} = 0,635$, $S_{\hat{p}_{95}-\hat{p}_{96}} = 0,0740$, $D_\alpha = -1,645 \cdot 0,0740 = -0,12$. Perciò avendo osservato un valore di D pari a $-0,021$, si accetta H_0.

Esercizio D.
Sia $X \sim N(\mu_X, \sigma_X^2)$ la v.a. che descrive il peso dei prodotti del macchinario A e sia $Y \sim N(\mu_Y, \sigma_Y^2)$ la v.a. che descrive il peso dei prodotti del macchinario B. L'interesse è su $H_0 : \sigma_X^2 = \sigma_Y^2$ contro $H_1 : \sigma_X^2 \neq \sigma_Y^2$. Sia $n = 8$ la numerosità del campione da X e $m = 6$ la numerosità del campione da Y. Consideriamo la statistica F data dal rapporto tra le varianza campionarie corrette

$$F = \frac{S_X^2}{S_Y^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)}{\sum_{i=1}^{m}(Y_i - \overline{Y})^2/(m-1)},$$

che sotto H_0 ha distribuzione $F_{n-1,m-1}$.
Sulla base dei dati campionari, si ha che $F = 0.8260$. Tale valore risulta essere minore del quantile $F_{n-1,m-1;\alpha}$, pari a 4,876 e ciò conduce all'accettazione dell'ipotesi nulla di uguaglianza tra le varianze σ_X^2 e σ_Y^2.