



NB: (A) ai fini della valutazione verranno considerate solo le risposte riportate dallo studente negli appositi riquadri bianchi del testo d'esame. (B) nello svolgimento del compito, si utilizzino tre cifre decimali dopo la virgola. (C) Allo studente/studentessa che consulti foglietti, appunti, libri, ecc o che parli con altri sarà annullata la prova d'esame.

SOLUZIONI

ESERCIZIO 1 (8 punti)

Si consideri un campione casuale semplice (X_1, X_2, X_3) estratto da una popolazione X con valore atteso μ e varianza $\sigma^2 = 9$. Allo scopo di stimare μ , viene proposto il seguente stimatore:

$$T_1 = \frac{4}{3}X_1 + \frac{1}{3}X_2 - \frac{1}{3}X_3$$

a) Si fornisca la definizione, in generale, di correttezza di uno stimatore.

Uno stimatore T si dice **corretto** (o **non distorto**) per un parametro θ se:

$$E(T) = \theta, \quad \forall \theta \in \Theta$$

b) Si verifichi se lo stimatore proposto T_1 è corretto e, in caso negativo, se ne calcoli la distorsione.

$$E(T_1) = E\left(\frac{4}{3}X_1 + \frac{1}{3}X_2 - \frac{1}{3}X_3\right) = \frac{4}{3}\mu + \frac{1}{3}\mu - \frac{1}{3}\mu = \frac{4}{3}\mu \neq \mu$$

quindi lo stimatore non è corretto e la sua distorsione è pari a:

$$D_T(\mu) = E(T) - \mu = \frac{4}{3}\mu - \mu = \frac{1}{3}\mu$$

c) Si fornisca la definizione generale di errore quadratico medio di uno stimatore.

Dato uno stimatore T per il parametro $\theta \in \Theta$, si definisce **errore quadratico medio** (EQM) la quantità:

$$EQM_T(\theta) = E[(T - \theta)^2]$$

L'EQM può essere equivalentemente espresso come:

$$EQM_T(\theta) = VAR(T) + D_T^2(\theta)$$

dove $D_T(\theta)$ indica la distorsione dello stimatore.

d) Si calcoli l'errore quadratico medio di T_1 .

$$EQM_{T_1}(\mu) = VAR(T_1) + D_{T_1}^2(\mu) = \left[\left(\frac{4}{3}\right)^2 \sigma^2 + \left(\frac{1}{3}\right)^2 \sigma^2 + \left(\frac{1}{3}\right)^2 \sigma^2 \right] + \left(\frac{1}{3}\mu\right)^2 = \frac{18}{9}\sigma^2 + \frac{1}{9}\mu^2 = 18 + \frac{\mu^2}{9}$$

ESERCIZIO 2 (4 punti)

Sia X una variabile aleatoria con valore atteso $E(X)=250$ e varianza $\text{VAR}(X) = 1280$. Supponete di estrarre da X un campione casuale semplice di ampiezza $n=80$.

a) Calcolate valore atteso e varianza della media campionaria \bar{X}

$$E(\bar{X})=E(X)=250$$

$$\text{Var}(\bar{X})=\frac{\text{Var}(X)}{n}=\frac{1280}{80}=16$$

b) Calcolate approssimativamente la probabilità che \bar{X} sia minore di 242.

Per il teorema centrale del limite, $\bar{X} \sim N(250,16)$, quindi:

$$\Pr(\bar{X} < 242) = \Pr\left(\frac{\bar{X} - 250}{\sqrt{16}} < \frac{242 - 250}{\sqrt{16}}\right) = \Pr(Z < -2) = 1 - 0.9772 = 0.0228$$

ESERCIZIO 3 (6 punti)

Si supponga che la variabile casuale $X \sim N(\mu, \sigma^2)$ rappresenti il peso di un certo prodotto industriale. Disponete di un campione di ampiezza $n=10$, grazie al quale osservate una media campionaria pari a 45 e una varianza campionaria corretta pari a 250. Considerate infine le ipotesi statistiche $H_0 : \mu = 40$ e $H_1 : \mu \neq 40$.

a) Costruite la regione di rifiuto per il test di cui sopra al livello $\alpha = 0.1$ e decidete se rifiutare o meno l'ipotesi nulla.

La regione di rifiuto è:

$$R = \left\{ (x_1, \dots, x_n) : |\bar{x} - \mu_0| > t_{1-\frac{\alpha}{2}}^{(n-1)} \sqrt{\frac{s_C^2}{n}} \right\} = \left\{ (x_1, \dots, x_n) : |\bar{x} - 40| > t_{0.95}^{(9)} \sqrt{\frac{250}{10}} \right\} =$$

$$= \left\{ (x_1, \dots, x_n) : |\bar{x} - 40| > 1.833 \cdot 5 \right\} = \left\{ (x_1, \dots, x_n) : |\bar{x} - 40| > 9.165 \right\}$$

Siccome $|\bar{x} - 40| = 5 < 9.165$, si **accetta** l'ipotesi nulla.

b) Costruite l'intervallo di confidenza per μ con coefficiente di confidenza 0.9.

L'intervallo di confidenza è:

$$\left(\bar{x} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \sqrt{\frac{s_C^2}{n}} \right) = (35.835; 54.165)$$

c) E' possibile arrivare alla stessa conclusione del test utilizzando l'intervallo di confidenza costruito al punto b)? Se sì, come?

Sì, è possibile: si decide infatti di **accettare** l'ipotesi nulla poiché $\mu_0 = 40$ appartiene all'intervallo di confidenza.

ESERCIZIO 4 (4 punti)

Supponete che p rappresenti la proporzione di utenti di internet che utilizzano una linea ADSL. Su 80 utenti intervistati, 62 dichiarano di avere la linea ADSL.

a) Proponete uno stimatore corretto per p e fornite la stima puntuale per p .

Uno stimatore corretto per la proporzione p è la media campionaria; la corrispondente stima puntuale è:

$$\bar{x} = \frac{1}{80} \sum_{i=1}^{80} x_i = \frac{62}{80} = 0.775$$

b) Secondo una precedente indagine, oltre il 75% degli utenti utilizza una linea ADSL. Sulla base del vostro campione, ritenete attendibile questa indagine? Rispondete utilizzando un test di livello $\alpha = 0.05$ ($H_0: p < 75\%$, $H_1: p > 75\%$).

La regione di rifiuto è:

$$R = \left\{ (x_1, \dots, x_n) : \bar{x} > p_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}} \right\} = \left\{ (x_1, \dots, x_n) : \bar{x} > 0.75 + z_{0.975} \sqrt{\frac{0.75(1-0.75)}{80}} \right\} = \\ = \left\{ (x_1, \dots, x_n) : \bar{x} > 0.75 + 1.96 \cdot 0.0484 \right\} = \left\{ (x_1, \dots, x_n) : \bar{x} > 0.8449 \right\}$$

Siccome la media campionaria è pari a 0.775, si ACCETTA l'ipotesi nulla. Dunque, sulla base del campione e del livello α dato, l'indagine non è attendibile.

ESERCIZIO 5 (2 punti)

Enunciate le ipotesi deboli di un modello lineare semplice.

Definendo il **modello di regressione** lineare semplice come: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, per $i=1, \dots, n$, le **ipotesi deboli** sono:

1. $E(\varepsilon_i) = 0$, per ogni $i=1, \dots, n$.
2. $\text{VAR}(\varepsilon_i) = \sigma^2$, per ogni $i=1, \dots, n$.
3. $\text{COV}(\varepsilon_i, \varepsilon_j) = 0$, per ogni $i \neq j$.

ESERCIZIO 6 (6 punti)

E' stato rilevato il rendimento giornaliero dell'indice azionario americano NASDAQ negli ultimi sei mesi (121 osservazioni). Allo scopo di stabilire la dipendenza dei rendimenti dal tempo, è stata condotta un'analisi basata sul modello lineare mediante il software SPSS. Il risultato di tale analisi è di seguito riportato:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.030 ^a	.001	-.007	.7871

a. Predictors: (Constant), TEMPO

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.646E-02	1	6.646E-02	.107	.744 ^a
	Residual	73.715	119	.619		
	Total	73.782	120			

a. Predictors: (Constant), TEMPO

b. Dependent Variable: NASDAQ

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.136	.144		.943	.348	-.149	.421
	TEMPO	-6.71E-04	.002	-.030	-.328	.744	-.005	.003

a. Dependent Variable: NASDAQ

a) Ritenete che il modello lineare sia adeguato per stabilire la dipendenza tra le due variabili? Rispondete sulla base di un opportuno indice.

Dalla tabella "Model Summary", il coefficiente di determinazione (R^2) risulta essere pari soltanto a 0.001, dunque il modello lineare NON è adeguato ai dati.

b) Qual è la stima dei coefficienti β_0 e β_1 del modello lineare?

Dalla tabella "Coefficients", le stime dei due parametri β_0 e β_1 sono rispettivamente 0.136 e $-6.71 \cdot 10^{-4}$.

c) Considerate le ipotesi $H_0: \beta_1 = 0$ e $H_1: \beta_1 \neq 0$. Quale ipotesi va accettata? Motivate la risposta.

Dalla tabella "Coefficients" si evince che l'intervallo di confidenza per il parametro β_1 è (-0.005 , 0.003): siccome lo zero appartiene a tale intervallo, l'ipotesi nulla va **accettata**.