

Esercitazione di Statistica

Inferenza, Regressione e Teorema del Limite Centrale

Parte 1: Teorema del Limite Centrale (TLC)

Esercizio 1 (Distribuzione della Media Campionaria)

Il peso delle confezioni di pasta prodotte da un macchinario ha una media $\mu = 500$ g e una deviazione standard $\sigma = 15$ g. Viene estratto un campione casuale di $n = 100$ confezioni.

- a) Qual è la distribuzione approssimata della media campionaria $\bar{X}$?
- b) Calcolare la probabilità che il peso medio del campione sia inferiore a 497 g.

Esercizio 2 (TLC sulla Somma - Caso Ascensore)

Il peso delle persone che utilizzano un ascensore si distribuisce con media $\mu = 75$ kg e deviazione standard $\sigma = 10$ kg. L'ascensore ha una portata massima di 2000 kg. Se salgono 25 persone ($n = 25$):

- a) Qual è la probabilità che la somma dei pesi delle 25 persone superi la portata massima? (Suggerimento: $S_{25} > 2000 \iff \bar{X} > 80$).

Esercizio 3 (TLC per le Proporzioni)

In una grande popolazione, il 40% degli individui è favorevole a una nuova legge ($p = 0.40$). Si estrae un campione di $n = 200$ individui.

- a) Calcolare la probabilità che la proporzione campionaria $\hat{P}$ di favorevoli sia compresa tra il 35% e il 45%.

Parte 2: Inferenza (Stima e Test d'Ipotesi)

Esercizio 4 (Intervallo di Confidenza - Varianza Nota)

Si vuole stimare il tempo medio di reazione di un pilota. Si assume che la popolazione sia Normale con $\sigma = 0.05$ secondi. Su un campione di $n = 25$ piloti, la media campionaria è $\bar{x} = 0.40$ secondi.

- a) Costruire un intervallo di confidenza al 99% per la media della popolazione μ . ($z_{0.005} = 2.576$).

Esercizio 5 (Test d'Ipotesi - Varianza Ignota, t-Student)

Un produttore dichiara che le batterie durano in media 60 ore. Un'associazione consumatori testa $n = 16$ batterie trovando una durata media $\bar{x} = 58$ ore con deviazione standard corretta $s = 4$ ore.

- a) Verificare l'ipotesi $H_0 : \mu = 60$ contro $H_1 : \mu < 60$ con un livello di significatività $\alpha = 0.05$. ($t_{15,0.05} = 1.753$).

Esercizio 6 (Inferenza sulla Proporzione)

Un sito web afferma che il 20% dei visitatori finalizza un acquisto. Su un campione di $n = 400$ visitatori analizzati oggi, 96 hanno acquistato.

- a) Verificare al livello $\alpha = 0.05$ se la proporzione di acquisti è significativamente aumentata rispetto al 20% storico. ($z_{0.05} = 1.645$).

Esercizio 7 (Confronto tra due Medie)

Vogliamo confrontare il rendimento medio di due gruppi di studenti sottoposti a metodi didattici diversi (A e B).

- Gruppo A ($n_A = 50$): $\bar{x}_A = 26$, $s_A^2 = 9$.
- Gruppo B ($n_B = 50$): $\bar{x}_B = 24$, $s_B^2 = 11$.

- a) Testare l'ipotesi che le medie siano diverse ($H_1 : \mu_A \neq \mu_B$) con $\alpha = 0.05$. (Considerare grandi campioni, usare Normale standard).

Parte 3: Regressione Lineare

Esercizio 8 (Calcolo Parametri da Statistiche)

Dati i seguenti valori riassuntivi su $n = 10$ osservazioni tra Spesa pubblicitaria (X) e Ricavi (Y):

$$\bar{x} = 4, \quad \bar{y} = 15, \quad S_{xx} = 20, \quad S_{xy} = 60$$

- a) Calcolare la pendenza $\hat{\beta}_1$ e l'intercetta $\hat{\beta}_0$.
- b) Scrivere l'equazione della retta stimata.

Esercizio 9 (Bontà di Adattamento e Devianze)

In un modello di regressione lineare semplice, la Devianza Totale (SST) è 500 e la Devianza Residua (SSE) è 120.

- a) Calcolare la Devianza Spiegata (SSR).
- b) Calcolare il coefficiente di determinazione R^2 e interpretarlo.

Esercizio 10 (Relazione Correlazione-Regressione)

Si consideri un modello di regressione $Y = \beta_0 + \beta_1 X + \epsilon$.

- a) Se il coefficiente di determinazione è $R^2 = 0.64$ e si sa che la relazione tra X e Y è inversa (all'aumentare di X , Y diminuisce), qual è il valore del coefficiente di correlazione lineare r ?
- b) Quale sarà il segno di $\hat{\beta}_1$?

Soluzioni e Svolgimenti

Parte 1: Teorema del Limite Centrale (TLC)

Esercizio 1 (Distribuzione della Media Campionaria)

Dati: $\mu = 500, \sigma = 15, n = 100$.

- a) **Distribuzione approssimata $\bar{X}$:** Per il TLC, poiché n è grande ($n = 100 > 30$), la distribuzione della media campionaria $\bar{X}$ si approssima a una **distribuzione Normale**.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\bar{X} \sim N\left(500, \frac{15^2}{100}\right) = N\left(500, \frac{225}{100}\right) = N(500, 2.25)$$

L'errore standard è $SE = \sqrt{2.25} = 1.5$.

- b) **Probabilità $P(\bar{X} < 497)$:** Si standardizza $\bar{X}$ usando la statistica Z :

$$Z = \frac{\bar{X} - \mu}{SE} = \frac{\bar{X} - 500}{1.5}$$

$$Z_{oss} = \frac{497 - 500}{1.5} = \frac{-3}{1.5} = -2$$

$$P(\bar{X} < 497) = P(Z < -2)$$

Dalle tavole della Normale Standard $\Phi(2) \approx 0.9772$.

$$P(Z < -2) = 1 - \Phi(2) = 1 - 0.9772 = \mathbf{0.0228} \quad (\text{ovvero } 2.28\%)$$

Esercizio 2 (TLC sulla Somma - Caso Ascensore)

Dati: $\mu = 75, \sigma = 10, n = 25$. Limite totale 2000 kg.

- a) **Probabilità $P(S_{25} > 2000)$:** Il problema è equivalente a calcolare $P(\bar{X} > 80)$ (poiché $80 = 2000/25$). La distribuzione approssimata di $\bar{X}$ è $N(\mu, \sigma^2/n)$:

$$\bar{X} \sim N\left(75, \frac{100}{25}\right) = N(75, 4). \quad SE = 2$$

$$Z_{oss} = \frac{80 - 75}{2} = \frac{5}{2} = \mathbf{2.5}$$

$$P(Z > 2.5) = 1 - \Phi(2.5)$$

Dalle tavole, $\Phi(2.5) \approx 0.9938$.

$$P = 1 - 0.9938 = \mathbf{0.0062} \quad (\text{ovvero } 0.62\%)$$

Esercizio 3 (TLC per le Proporzioni)

Dati: $p = 0.40, n = 200$.

- a) **Probabilità** $P(0.35 < \hat{P} < 0.45)$: Usiamo l'approssimazione Normale per la porzione campionaria $\hat{P}$.

$$\hat{P} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Calcolo dell'errore standard: $SE = \sqrt{\frac{0.40 \cdot 0.60}{200}} = \sqrt{0.0012} \approx 0.0346$.

Standardizzazione dei limiti:

$$Z_1 = \frac{0.35 - 0.40}{0.0346} \approx \mathbf{-1.45}$$

$$Z_2 = \frac{0.45 - 0.40}{0.0346} \approx \mathbf{1.45}$$

$$P = P(-1.45 < Z < 1.45) = \Phi(1.45) - \Phi(-1.45) = 2\Phi(1.45) - 1$$

Assumendo $\Phi(1.45) \approx 0.9265$:

$$P \approx 2(0.9265) - 1 = 1.853 - 1 = \mathbf{0.853} \quad (\text{ovvero } 85.3\%)$$

—

Parte 2: Inferenza (Stima e Test d'Ipotesi)**Esercizio 4 (Intervallo di Confidenza - Varianza Nota)**

Dati: $\bar{x} = 0.40, \sigma = 0.05, n = 25$. Fiducia 99% $\rightarrow z_{\alpha/2} = 2.576$.

- a) **Intervallo di Confidenza al 99% per μ :** Formula per σ nota: $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$IC = 0.40 \pm 2.576 \cdot \frac{0.05}{\sqrt{25}}$$

$$IC = 0.40 \pm 2.576 \cdot (0.01)$$

$$IC = 0.40 \pm 0.02576$$

$$IC_{99\%} = \mathbf{[0.37424, 0.42576]} \quad \text{secondi}$$

Esercizio 5 (Test d'Ipotesi - Varianza Ignota, t-Student)

Dati: $H_0 : \mu = 60, H_1 : \mu < 60$. $n = 16, \bar{x} = 58, s = 4$. $\alpha = 0.05$. $gdl = 15$.

- a) **Verifica d'ipotesi (μ):** Poiché σ è ignota e $n < 30$, usiamo la distribuzione **t-Student**.

Statistica Test:

$$T_{oss} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{58 - 60}{4/\sqrt{16}} = \frac{-2}{4/4} = \mathbf{-2}$$

Valore Critico: Il test è **unilaterale sinistro**. Dal testo, $t_{15,0.05} = 1.753$. Il valore critico è -1.753 .

Conclusione: Poiché $T_{oss} = -2$ cade nella regione di rifiuto (dato che $-2 < -1.753$), si **Rifiuta** H_0 . C'è evidenza significativa che la durata media delle batterie sia inferiore a 60 ore.

Esercizio 6 (Inferenza sulla Proporzione)

Dati: $H_0 : p = 0.20, H_1 : p > 0.20$. $n = 400, \hat{p} = 96/400 = 0.24$.

- a) **Verifica d'ipotesi (p):** Usiamo l'approssimazione Normale per la proporzione.

Statistica Test (Z):

$$Z_{oss} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.24 - 0.20}{\sqrt{\frac{0.20(1-0.20)}{400}}} = \frac{0.04}{\sqrt{0.0004}} = \frac{0.04}{0.02} = \mathbf{2}$$

Valore Critico: Il test è **unilaterale destro** ($\alpha = 0.05$). Dal testo, $z_{0.05} = 1.645$.

Conclusione: Poiché $Z_{oss} = 2$ è maggiore del valore critico 1.645, si **Rifiuta** H_0 . La proporzione di acquisti è significativamente aumentata oltre il 20%.

Esercizio 7 (Confronto tra due Medie)

Dati: $n_A = 50, \bar{x}_A = 26, s_A^2 = 9$. $n_B = 50, \bar{x}_B = 24, s_B^2 = 11$. $H_0 : \mu_A = \mu_B, H_1 : \mu_A \neq \mu_B$. $\alpha = 0.05$.

- a) **Test d'Ipotesi (Differenza tra Medie):** Poiché i campioni sono grandi ($n \geq 30$) e indipendenti, usiamo la statistica Z .

Statistica Test (Z):

$$Z_{oss} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_{A0} - \mu_{B0})}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{(26 - 24) - 0}{\sqrt{\frac{9}{50} + \frac{11}{50}}}$$

$$Z_{oss} = \frac{2}{\sqrt{\frac{20}{50}}} = \frac{2}{\sqrt{0.4}} \approx \frac{2}{0.6324} \approx \mathbf{3.16}$$

Valore Critico: Il test è **bilaterale** ($\alpha = 0.05$). I valori critici sono $\pm z_{\alpha/2} =$

Parte 3: Regressione Lineare Semplice

Esercizio 8 (Calcolo Parametri da Statistiche)

Dati: $\bar{x} = 4, \bar{y} = 15, S_{xx} = 20, S_{xy} = 60$.

- a) **Calcolo Pendenza ($\hat{\beta}_1$) e Intercetta ($\hat{\beta}_0$):** * La pendenza OLS è data da:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{60}{20} = \mathbf{3}$$

* L'intercetta OLS è data da:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 15 - 3(4) = 15 - 12 = \mathbf{3}$$

- b) **Equazione della retta stimata:**

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad \Longrightarrow \quad = \mathbf{3 + 3X}$$

Esercizio 9 (Bontà di Adattamento e Devianze)

Dati: Devianza Totale (SST) = 500, Devianza Residua (SSE) = 120.

- a) **Calcolo Devianza Spiegata (SSR):** La scomposizione della varianza totale è $SST = SSR + SSE$.

$$SSR = SST - SSE = 500 - 120 = \mathbf{380}$$

- b) **Calcolo R^2 e interpretazione:** Il coefficiente di determinazione è dato da $R^2 = SSR/SST$.

$$R^2 = \frac{380}{500} = \mathbf{0.76}$$

Interpretazione: Il modello di regressione lineare spiega il ****76%**** della variabilità totale della variabile dipendente (Y). L'adattamento è forte.

Esercizio 10 (Relazione Correlazione-Regressione)

Dati: $R^2 = 0.64$, relazione inversa.

- a) **Valore del Coefficiente di Correlazione r :** La correlazione r è la radice quadrata di R^2 , e ha lo stesso segno della pendenza $\hat{\beta}_1$. Poiché la relazione è inversa (pendenza negativa), anche r è negativo.

$$r = -\sqrt{R^2} = -\sqrt{0.64} = \mathbf{-0.8}$$

- b) **Segno di $\hat{\beta}_1$:** Essendo la relazione inversa, all'aumentare di X , Y diminuisce, quindi $\hat{\beta}_1$ deve essere **Negativo** ($\hat{\beta}_1 < 0$).