Esercizi di Probabilità e Statistica

Esercizio 1

In un supermercato, due macchine automatiche preparano bustine di tè: la macchina **Rossa** e la macchina **Verde**. Dopo la produzione, le bustine vengono mescolate senza etichettatura.

- La macchina Rossa produce il 3% di bustine difettose.
- La macchina Verde produce il 6% di bustine difettose.
- La macchina Rossa produce 400 bustine al giorno, la Verde 600.

Domande

1. Qual è la probabilità che una bustina sia difettosa?
2. Se una bustina è difettosa, qual è la probabilità che provenga dalla macchina Verde?

Soluzione

Definizioni:

$$P(D|R) = 0.03, \quad P(D|V) = 0.06, \quad P(R) = 0.4, \quad P(V) = 0.6$$

Teorema delle probabilità totali:

$$P(D) = P(D|R)P(R) + P(D|V)P(V) = 0.03 \cdot 0.4 + 0.06 \cdot 0.6 = 0.048$$

Teorema di Bayes:

$$P(V|D) = \frac{P(D|V)P(V)}{P(D)} = \frac{0.06 \cdot 0.6}{0.048} = 0.75$$

Risposte:

- Probabilità di bustina difettosa: 4.8%
- Se è difettosa, probabilità che provenga dalla macchina Verde: 75%

Esercizio 2 – Analisi statistica di due campioni

Dati

Consideriamo i due campioni seguenti:

$$X = \{41.2,\ 39.8,\ 40.5,\ 42.1,\ 38.7,\ 40.0,\ 39.3\}$$
$$Y = \{36.5,\ 37.2,\ 35.9,\ 38.0,\ 34.8,\ 36.1\}$$

Si assuma che i dati siano estratti da popolazioni normali con varianze ignote e non necessariamente uguali.

Domande

1. Calcolare media e deviazione standard per ciascun campione.
2. Verificare se le medie delle due popolazioni possono essere considerate uguali (livello di significatività $\alpha = 0.05$).

1. Statistiche campionarie

Campione X

- Media campionaria:

$$\bar{x}_X = \frac{1}{7}\sum x_i = \frac{281.6}{7} = \boxed{40.23}$$

- Scarti dalla media e quadrati:

x_i	$x_i - \bar{x}_X$	$(x_i - \bar{x}_X)^2$
41.2	0.971	0.943
39.8	-0.429	0.184
40.5	0.271	0.073
42.1	1.871	3.501
38.7	-1.529	2.338
40.0	-0.229	0.053
39.3	-0.929	0.864

- Deviazione standard:

$$s_X = \sqrt{\frac{\sum (x_i - \bar{x}_X)^2}{n-1}} = \sqrt{\frac{7.956}{6}} = \boxed{1.15}$$

- Varianza:

$$s_X^2 = \boxed{1.33}$$

Campione Y

- Media campionaria:

$$\bar{x}_Y = \frac{218.5}{6} = \boxed{36.42}$$

- Scarti dalla media e quadrati:

y_i	$y_i - \bar{x}_Y$	$(y_i - \bar{x}_Y)^2$
36.5	0.083	0.007
37.2	0.783	0.613
35.9	-0.517	0.267
38.0	1.583	2.506
34.8	-1.617	2.615
36.1	-0.317	0.101

- Deviazione standard:

$$s_Y = \sqrt{\frac{6.109}{5}} = \boxed{1.11}$$

- Varianza:

$$s_Y^2 = \boxed{1.22}$$

Esercizio 3 – Sondaggio elettorale (Bayes, Totali)

Un sondaggio elettorale rileva:

- 60% degli intervistati ha votato.
- Tra chi ha votato, il 55% ha votato Partito A.
- Tra chi non ha votato, il 10% simpatizza per il Partito A.

Domande

1. Qual è la probabilità che una persona simpatizzi per il Partito A?
2. Sapendo che una persona simpatizza per il Partito A, qual è la probabilità che abbia votato?
3. Il Partito A è più popolare tra i votanti o tra i non votanti?

Soluzione

$$P(V) = 0.6, \quad P(\bar{V}) = 0.4, \quad P(A|V) = 0.55, \quad P(A|\bar{V}) = 0.10$$

Totali:

$$P(A) = P(A|V)P(V) + P(A|\bar{V})P(\bar{V}) = 0.33 + 0.04 = \boxed{0.37}$$

Bayes:

$$P(V|A) = \frac{P(A|V)P(V)}{P(A)} = \frac{0.33}{0.37} \approx \boxed{0.892}$$

Conclusione: Il Partito A è molto più popolare tra i votanti ($P(A|V) = 55\%$) rispetto ai non votanti ($P(A|\bar{V}) = 10\%$).

Esercizio 5 - Analisi dati fatture saldate

Il proprietario di un negozio di computer vuole sapere quanto velocemente vengono saldate le fatture relative ai PC per tre diverse tipologie di clienti:

- A = Enti Pubblici,
- B = Aziende,
- C = Privati.

Per ciascuna tipologia di cliente, viene riportato il numero di fatture saldate nei seguenti intervalli temporali (*giorni trascorsi tra consegna e saldo*):

Giorni trascorsi	**A (Enti Pubblici)**	**B (Aziende)**	**C (Privati)**
0–10	26	52	40
10–20	42	60	46
20–30	12	18	14

a) Rappresentazione grafica della variabile *Tipologia del cliente*

Calcoliamo il numero totale di fatture saldate per ciascuna tipologia:

$$\text{A} = 26 + 42 + 12 = 80 \qquad \text{B} = 52 + 60 + 18 = 130 \qquad \text{C} = 40 + 46 + 14 = 100$$

Questi valori sono poi rappresentabili in un grafico a barre (omesso qui, ma può essere inserito come immagine in allegato).

b) Calcolo della mediana della variabile *Giorni trascorsi*

Assumiamo come valore rappresentativo di ciascuna classe:

$$\text{Classe } 0-10 \to 5, \qquad \text{Classe } 10-20 \to 15, \qquad \text{Classe } 20-30 \to 25$$

Costruiamo una tabella con i valori:

Valore	**Frequenza**	**Frequenza cumulata**
5	26 + 52 + 40 = 118	118
15	42 + 60 + 46 = 148	266
25	12 + 18 + 14 = 44	310

Poiché il totale è $N = 310$, la mediana è il valore corrispondente alla posizione $N/2 = 155$.

Il valore 155 si trova nella seconda classe (15), quindi:

$$\text{Mediana} = \boxed{15}$$

c) Percentuale di Enti Pubblici che hanno saldato la fattura dopo 10 giorni

Numero di fatture per Enti Pubblici:

$$\text{Totale A} = 26 + 42 + 12 = 80$$

Numero di fatture saldate dopo 10 giorni:

$$\text{A, dopo 10 giorni} = 42 + 12 = 54$$

Percentuale:

$$\frac{54}{80} \cdot 100 = \boxed{67{,}5\%}$$

d) Relazione grafica tra le variabili

Si costruisce un grafico a barre raggruppate (tipicamente con software come Python o Excel) per confrontare le frequenze di pagamento nei diversi intervalli temporali per ogni tipologia di cliente.

Dall'osservazione si nota che:

- Gli Enti Pubblici (A) hanno picco nella classe 10–20 giorni.
- Le Aziende (B) saldano in gran parte entro i primi 20 giorni.
- Anche i Privati (C) si concentrano nelle prime due fasce.

e) Indice di associazione tra le due variabili (chi-quadro)

Costruiamo la tabella di contingenza:

Giorni	A	B	C
0–10	26	52	40
10–20	42	60	46
20–30	12	18	14

Applichiamo il test chi-quadro:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{con } E_{ij} = \frac{\text{Totale riga}_i \cdot \text{Totale colonna}_j}{N}$$

Il risultato del test è:

$$\chi^2 \approx 1{,}44, \quad p\text{-value} \approx 0{,}837$$

Poiché $p > 0{,}05$, non vi è evidenza significativa di dipendenza tra le due variabili:

$$\boxed{\text{Le due variabili sono statisticamente indipendenti.}}$$

Esercizio: Intervalli di confidenza per la media

Sono date le seguenti misurazioni della velocità del suono (espresse in centinaia di metri al secondo):

3,25; 3,27; 3,30; 3,33; 3,32; 3,29

1. Statistiche descrittive

Numero di osservazioni: $n = 6$

Media campionaria

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{3{,}25 + 3{,}27 + 3{,}30 + 3{,}33 + 3{,}32 + 3{,}29}{6} = \boxed{3{,}293}$$

Deviazione standard campionaria

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \boxed{0{,}0301}$$

2. Intervallo di confidenza al 95% per la media

Utilizziamo la distribuzione t di Student:

$$IC_{95\%} = \bar{x} \pm t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}}$$

Gradi di libertà: $df = n - 1 = 5$
Quantile $t_{0{,}025{,}5} \approx 2{,}571$

$$\text{Errore standard} = \frac{s}{\sqrt{n}} = \frac{0{,}0301}{\sqrt{6}} \approx 0{,}0123$$

$$\text{Margine} = 2{,}571 \cdot 0{,}0123 \approx 0{,}0312$$

$$IC_{95\%} = 3{,}293 \pm 0{,}0312 \Rightarrow \boxed{(3{,}262;\ 3{,}325)}$$

3. Intervallo di confidenza al 99% per la media

Quantile $t_{0{,}005{,}5} \approx 4{,}032$

$$\text{Margine} = 4{,}032 \cdot 0{,}0123 \approx 0{,}0496$$

$$IC_{99\%} = 3{,}293 \pm 0{,}0496 \Rightarrow \boxed{(3{,}244;\ 3{,}343)}$$

Esercizio 6

Test sulla media nota.

Dati

- $n = 150$
- $\bar{x} = 38000$
- $\sigma = 5000$
- $\mu_0 = 40000$
- Test unilatero a sinistra

Ipotesi

$$H_0 : \mu = 40000$$
$$H_1 : \mu < 40000$$

Statistica test

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{38000 - 40000}{5000/\sqrt{150}} = \frac{-2000}{408.25} \approx -4.90$$

Decisione

- Livello di significatività: $\alpha = 0.01$
- Valore critico: $z_{0.01} = -2.33$
- Poiché $z = -4.90 < -2.33$, rifiutiamo H_0

Conclusione

Il reddito medio è significativamente inferiore a 40000.

Intervallo di confidenza al 99%

$$IC = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 38000 \pm 2.576 \cdot 408.25 = 38000 \pm 1052.77$$
$$IC = (36947.2;\ 39052.8)$$

Esercizio 7

Test t su media con deviazione standard stimata.

Dati

- $n = 50$
- $\bar{x} = 75.2$
- $s = 14.22$
- $\mu_0 = 70$
- Test unilatero a destra

Ipotesi

$$H_0 : \mu = 70$$
$$H_1 : \mu > 70$$

Statistica test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{75.2 - 70}{14.22/\sqrt{50}} = \frac{5.2}{2.0105} \approx 2.59$$

Gradi di libertà

$$df = n - 1 = 49$$

Decisione

- $t_{0.05,49} \approx 1.677 \Rightarrow$2.59 ¿ 1.677$\Rightarrow rifiutiamo$H$_0$
- $t_{0.01,49} \approx 2.405 \Rightarrow$2.59 ¿ 2.405$\Rightarrow rifiutiamo$H$_0$ anche al 1%

Conclusione

Evidenza che la media di grassi è maggiore di 70g. L'ipotesi dei ricercatori è supportata a entrambi i livelli di significatività.