

PROVA SCRITTA di STATISTICA_a – 10/01/2019

Canale A–D,
prof.ssa Guagnano

COGNOME

NOME

MATR.

IMPORTANTE: Bisogna sempre giustificare nell'elaborato del compito, attraverso calcoli svolti, motivazioni teoriche o altro, i risultati riportati nel seguente prospetto riepilogativo (per ciascun quesito a risposta multipla, indicare la risposta selezionata); in caso contrario, le risposte non verranno prese pienamente in considerazione ai fini della valutazione.

Prospetto riepilogativo dei risultati:

1. a) **ii (11)**; b) **iii (36)**; 2. **iii**; perché: **la varianza è una media di quadrati e non può essere negativa**;
3. a) l'informazione è coerente: sì ; no ; perché: $\text{Cov}(X,Y) = M(X \times Y) - M(X) \times M(Y)$, ma $\text{Cov}(X,Y) = 7$ e
$$M(X \times Y) - M(X) \times M(Y) = 30 - 2 \times 10 = 10 \neq 7$$

b) **ii**; perché: **l'inclinazione della retta dipende dal segno di $\text{Cov}(X,Y)$, che è positivo**;
c) **iv (8)**; perché: **vale la condizione $|\text{Cov}(X,Y)| \leq \sigma_x \times \sigma_y$** ;
4. **ii**; perché: $\text{Pr}(A \cup B) = \text{Pr}(A) + \text{Pr}(B) - \text{Pr}(A \cap B) \leq 1 \Rightarrow \text{Pr}(A \cap B) \geq 1.25 - 1 = 0.25$;
5. Dimostrazione: **Dato che $E(X_i) = \mu, \forall i$, si ha: $E(\sum_i X_i/n) = \sum_i E(X_i/n) = \sum_i \mu/n = n \times \mu/n = \mu$** ;
6. a) **i ($\mu_1' < \mu_1''$)**; perché: **la probabilità β è tanto più elevata quanto più il valore di μ_1 è prossimo a μ_0** ;
b) **0.98; 0.975**;
7. a) $M_o: 7$; $M_e: 6.82 \Rightarrow 7$; $M(X): 8$; $\text{Var}(X): 20$;
b) coordinate richieste: **($P_1=0.2$; $Q_1=0.05$)**; **($P_2=0.75$; $Q_2=0.53125$)**; **($P_3=1$; $Q_3=1$)**;
c) sistema di ipotesi da verificare: $H_0: \mu = 10$ (oppure $\mu \geq 10$); $H_1: \mu < 10$;
valore critico: $z_{0.01} = -2.235$ (oppure 8.875 per la media campionaria non standardizzata);
valore osservato per la statistica-test: **-3.975** (oppure 8, valore osservato della media campionaria, nel caso si usi
la soglia 8.875); si accetta: H_0 ; H_1 ; p-value: **0.00004**;
8. a) frequenze attese in: i) **2; 6**; ii) **0; 8**; b) percentuali richieste: **20%; 33.3%**;
c) indice utilizzato: $\eta^2_{Y|X}$; misura del grado di dipendenza in media: **0.07287 (0.07162 se per il numeratore dell'indice si usa la formula abbreviata)**;
9. a) probabilità richiesta: **0.18**; b) probabilità richiesta: **0.8626**;
c) distribuzione utilizzabile: **$N(\mu=36; \sigma^2=29.52)$** .

1. Tra due caratteri statistici vale la seguente relazione: $Y=3+2X$. Sapendo che $\mu_X=4$ e $\sigma_X^2=9$,
 - a) indicare quale valore è ammissibile per μ_Y : i) 8; ii) 11; iii) 14; iv) 20;
 - b) indicare quale valore è ammissibile per σ_Y^2 : i) 18; ii) 21; iii) 36; iv) 39.
2. Ad uno studente viene chiesto di determinare la varianza di una certa distribuzione di frequenza e il risultato che egli ottiene è pari a -9 . Commentare il risultato scegliendo una tra le seguenti spiegazioni alternative e indicandola nel prospetto riepilogativo: i) le modalità presentate dal carattere sono prevalentemente inferiori alla media aritmetica; ii) le modalità presentate dal carattere sono prevalentemente negative; iii) è stato commesso un errore di calcolo; iv) la distribuzione presenta una asimmetria negativa.
3. In un campione di 60 unità, relativamente alle variabili X e Y sono state calcolate le seguenti sintesi: $M(X)=2$; $M(Y)=10$; $\text{Var}(X)=4$; $\text{Var}(Y)=16$; $\text{COV}(X,Y)=7$.
 - a) Stabilire se l'informazione $M(X*Y)=30$ risulti coerente con le precedenti informazioni, motivando la risposta.
 - b) Con riferimento alla retta di regressione che esprime Y in funzione di X, indicare il tipo di inclinazione che dovremmo attenderci, motivando la risposta: i) parallela all'asse delle X; ii) positiva (retta crescente); iii) negativa (retta decrescente); iv) parallela all'asse delle Y.
 - c) Indicare il massimo valore che può assumere la covarianza, motivando la risposta: i) informazioni insufficienti per rispondere; ii) 64; iii) ∞ ; iv) 8.
4. Con riferimento a due eventi A e B, per i quali risulta $P(A)=0.6$ e $P(B)=0.65$, indicare cosa possa dirsi in merito alla probabilità $P(A \cap B)$, motivando la risposta: i) è nulla; ii) è almeno pari a 0.25; iii) è al massimo pari a 0.25; iv) è esattamente pari a 0.3; v) informazioni insufficienti per rispondere.
5. Dimostrare che la media campionaria è uno stimatore non distorto della media di popolazione.
6. In un test unilaterale destro sulla media di una popolazione normale, condotto al livello $\alpha=0.025$, si considerano due diversi valori del parametro, μ_1' e μ_1'' , coerenti con l'ipotesi alternativa H_1 . Condizionatamente a tali valori, la probabilità β vale rispettivamente 0.02 e 0.015.
 - a) Stabilire in che relazione sono tra loro i due valori μ_1' e μ_1'' , motivando la risposta: i) $\mu_1' < \mu_1''$; ii) $\mu_1' > \mu_1''$; iii) tra i due valori non c'è alcun legame; iv) informazioni insufficienti per rispondere.
 - b) Indicare la probabilità di rifiutare correttamente l'ipotesi nulla, supponendo che $\mu = \mu_1'$, e quella di accettare correttamente H_0 .

PARTE B (23 punti)

7. In un campione casuale di 80 individui estratti da una popolazione Normale viene rilevato il numero di libri letti in un anno, pervenendo alla seguente sintesi tabellare:

| Numero di libri | Numero di individui |
|-----------------|---------------------|
| 1 – 3 | 16 |
| 4 – 10 | 44 |
| 11 – 19 | 20 |
| Tot. | 80 |

- a) Determinare la moda, la mediana, la media e la varianza del numero di libri letti.
 - b) Per ciascuna classe di frequenza, indicare le coordinate della spezzata di concentrazione.
 - c) Verificare se, al livello di significatività dell'1%, il numero di libri mediamente letto in un anno nella popopossa ritenersi inferiore a 10, oppure no. Determinare, inoltre, il p-value del test.
8. Si consideri la seguente distribuzione di 20 aziende secondo il livello di know-how tecnologico utilizzato (variabile X) e la quota di mercato detenuta (variabile Y):
- | X | Basso | Medio/Alto | Totale |
|-------------|-------|------------|--------|
| Y | | | |
| 0.01 – 0.05 | 3 | 3 | 6 |
| 0.05 – 0.15 | 1 | 7 | 8 |
| 0.15 – 0.25 | 1 | 5 | 6 |
| Totale | 5 | 15 | 20 |
- a) Indicare le frequenze che dovremmo attenderci nella seconda riga della tabella, sotto l'ipotesi di:
 - i) indipendenza assoluta; ii) massima associazione tra i due caratteri.
 - b) Indicare la percentuale di aziende con quota di mercato più elevata, tra quelle con basso livello di know-how e tra quelle con medio/alto livello di know-how.
 - c) Valutare il grado di dipendenza in media di Y da X, sapendo che la varianza di Y vale 0.004389, mentre la quota di mercato mediamente detenuta vale: 0.109 per il complesso delle 20 aziende, 0.078 per le aziende con basso livello di know-how e 0.1193 per quelle con medio/alto livello di know-how.

9. I pezzi prodotti da un certo macchinario possono presentare due tipi di difetti, D1 e D2. La probabilità che un pezzo prodotto presenti il difetto D1 è 0.08, la probabilità che presenti il difetto D2 è 0.12, mentre la probabilità che presenti entrambi i difetti è 0.02.
- Calcolare la probabilità che un pezzo preso a caso presenti almeno un difetto.
 - Calcolare la probabilità che, prendendo a caso 10 pezzi, ve ne sia almeno uno difettoso (con difetto qualsiasi).
 - Indicare la distribuzione di probabilità che si potrebbe utilizzare per determinare la probabilità in b) qualora si estraessero 200 pezzi, specificandone il/i parametro/i.