Compito scritto dell'esame di Statistica e analisi dei dati 18 luglio 2024	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri, appunti, cellulari.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.
Si misurano i diametri medi e le varianze di due campioni di cavi di rame, di taglia rispettivamente 9 e 12, prodotti il primo con un processo classico e il secondo con una tecnica innovativa che dovrebbe ridurre la variabilitá del diametro Effettuata la misurazione, le varianze del primo e del secondo campione risultano essere rispettivamente 20 e 8.

(a) Sembrerebbe che la tecnica introdotta abbia significativamente migliorato la produzione. Si verifichi che tale conclusione sia statisticamente plausibile al livello di significativitá 0.05, ragionando mediante intervalli di confidenza e assumendo che la distribuzione dei diametri segua una legge normale.

Soluzione:
I dati del problema ci dicono che: $n_1 = 9$, $n_2 = 12$, $s_1^2 = 20$, $s_2^2 = 8$.
L'IC per il rapporto delle due varianze, con $\alpha = 0.05$ é:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{0.025}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{0.025}(\nu_2,\nu_1)$$

con $\nu_1 = 8, \nu_2 = 11$, $\frac{s_1^2}{s_2^2} = 2.5$ $f_{0.025}(8,11) = 3.66$, $f_{0.025}(11,8) \simeq 4.25$

$$2.5 \cdot 0.27 < \frac{\sigma_1^2}{\sigma_2^2} < 2.5 \cdot 4.25$$

$$0.68 < \frac{\sigma_1^2}{\sigma_2^2} < 10.6$$

L'intervallo ammette la possibilitá che $\frac{\sigma_1^2}{\sigma_2^2} = 1$, dunque non possiamo concludere che la tecnica abbia significativamente migliorato la produzione (ridotta la variabilitá) .

Esercizio 2. Giorgio pensa che se beve due pinte di birra, la probabilitá di essere coinvolto in un incidente guidando casa dal pub sia una su cinquecento. Assumendo che ogni sera guidi dal medesimo pub a casa dopo aver bevuto due pinte:

(a) Calcolare la probabilitá che Giorgio sia coinvolto in almeno un incidente nel corso di un anno.

Soluzione: Indichiamo con
- I_i l'evento "nessun incidente nel giorno i" , con $i = 1, \cdots, 365$
- A l'evento "almeno 1 incidente nell'anno"

La credenza di Giorgio é

$$P(\sim I_i) = \frac{1}{500} = 0.002 \implies P(I_i) = 1 - 0.002 = 0.998$$

Allora, usando l'indipendenza statistica tra eventuali incidenti:

$$P(A) = 1 - P(\sim A) = 1 - P(\cup_{i=1}^{365} I_i) = 1 - \prod_{i=1}^{365} P(I_i) = 1 - (0.998)^{365} = 0.5184$$

Esercizio 3. Una macchina produce astucci di cui l'1% é difettoso.

(a) Quanti astucci posso imballare in una scatola di modo che l'eventualitá di trovare uno o piú astucci difettosi abbia probabilitá inferiore a 0.5?

Soluzione: (Distribuzione binomiale) Sia X_n il numero di astucci difettosi quando n astucci vengano messi in una scatola. X_n segue legge binomiale, $X_n \sim Bin(X_n; n, p)$.
Qui $p = 0.01$ dunque $q = 0.99$. Pertanto

$$P(X_n = 0) = Bin(X_n; n, p = 0.01) = (0.99)^n$$

Da cui si richiede che

$$P(X_n \geq 1) = 1 - P(X_n = 0) = 1 - (0.99)^n < 0.5 \implies (0.99)^n > 0.5$$

Pertanto, eliminando l'esponenziale con il logaritmo ed invertendo la disuguaglianza poichè la base $0.99 < 1$, deve valere

$$n < \frac{\log 0.5}{\log 0.99} = 68.97$$

Essendo n intero possiamo porre

$$n = 68$$

(b) Quale sarebbe il valore atteso e la deviazione standard del numero di pezzi difettosi in una scatola di tale dimensione?

Soluzione: Per la binomiale

$$E[X_{68}] = np = 68 \times 0.01 = 0.68$$
$$stdev(X_{68}) = \sqrt{68 \times 0.01 \times 0.99} = 0.8205$$

Esercizio 4. Si definisca una variabile aleatoria $Y = -2X + 3$ dove X ha densitá

$$f_X(x) = \begin{cases} kx, & per \quad 0 \leq x \leq 1 \\ \\ 0 & \qquad altrove \end{cases}$$

(a) Trovare valore atteso e varianza della v.a. Y.

Soluzione:
Per le proprietá di valore atteso e varianza

$$E[Y] = E[-2X + 3] = -2E[X] + 3$$
$$var(Y) = var(-2X + 3) = 4var(X)$$

Calcoliamo allora $E[X]$ e $var(X) = E[X^2] - E[X]^2$
Per procedere, trovo la costante k imponendo la normalizzazione

$$\int_0^1 f_X(x)dx = \int_0^1 kxdx = 1 \implies k\left[\frac{1}{2}x^2\right]_0^1 = 1 \implies k = 2$$

Allora

$$E[X] = \int_0^1 x \cdot 2xdx = \frac{2}{3}$$
$$E[X^2] = \int_0^1 x^2 \cdot 2xdx = \frac{1}{2}$$

da cui

$$var(X) = E[X^2] - E[X]^2 = \frac{1}{2}(\frac{2}{3})^2 = \frac{1}{18}$$

Infine

$$E[Y] = -2E[X] + 3 = \frac{5}{3}$$
$$var(Y) = 4var(X) = \frac{2}{9}$$

Esercizio 5. L' erogatore di acqua della mensa di via Golgi versa nominalmente in ciascun bicchiere 200 ml in media con deviazione standard di 15 ml (si assuma che la quantitá erogata segua una legge normale) Viene effettuato un controllo periodico su un campione di nove bicchieri e se il contenuto medio misurato é compreso nell'intervallo tra i 191ml e i 209 ml si conclude che l'erogatore funziona correttamente.

(a) Si calcoli la probabilità di rischio associato al criterio di decisione sopra riportato

Soluzione:
Il contenuto medio erogato $\bar{X}$ é distribuito con legge normale $\bar{X} \sim \mathcal{N}(\mu_{\bar{X}} = 11, \sigma^2_{\bar{X}} = \frac{\sigma}{n})$.
Si ha che $n = 9$, $\mu = 200$ $\sigma = 15$ dunque $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 5$ml
Le ipotesi sono

$$\begin{cases} H_0 & : \mu = 200 = \mu_0 \\ H_1 & : \mu \neq 200. \end{cases} \tag{1}$$

e il criterio adottato corrisponde ad accettare H_0 se $(191 < \bar{x} < 209)$ml.
Dalla teoria dei test di ipotesi sappiamo che il rischio (del consumatore) é definito come la probabilitá α di errore di tipo I associato al criterio di decisione :

$$P(\texttt{rifiuto } H_0 \mid H_0 \texttt{ vera}) = \alpha.$$

Considerato il criterio di decisione nell'ipotesi che H_0 sia vera, $\mu = \mu_0 = 200$ml; dunque:

$$\alpha = P(\texttt{rifiuto } H_0 \mid H_0 \texttt{ vera}) = P(\bar{X} < 191 \mid \mu = \mu_0) + P(\bar{X} > 209 \mid \mu = \mu_0)$$

Standardizzando e usando le tavole della normale standard si ha che

$$P(\bar{X} < 191 \mid \mu = 200) = P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{191 - \mu}{\sigma/\sqrt{n}})$$

$$P(\bar{X} > 209 \mid \mu = 200) = P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{209 - \mu}{\sigma/\sqrt{n}})$$

dunque il rischio é

$$\alpha = P(Z < -1.8) + P(Z > 1.8) = 2 \cdot (1 - \Phi(1.8)) = 0.0718$$

(b) Si valuti la potenza del test associato al criterio di decisione sopra riportato qualora la media fosse di 215 ml

Soluzione:
La potenza del test é definita nella teoria dei test come complemento della probabilitá β dell'errore di tipo II

$$P_{test} = 1 - \beta = 1 - P(\texttt{non rifiuto } H_0 \mid H_0 \texttt{ falsa}) = P(\texttt{rifiuto } H_0 \mid H_1 \texttt{vera}).$$

Nell'ipotesi che H_1 sia vera, $\mu = \mu_1 = 215$. Pertanto

$$\beta = P(191 < \bar{X} < 209 \mid \mu = 215) = P(\frac{191 - 215}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{209 - 215}{\sigma/\sqrt{n}}) =$$
$$P(-4.8 < Z < -1.2) = (1 - \Phi(1.2)) - (1 - \Phi(4.8)) \approx 0.115$$

Dunque la potenza é pari a $P_{test} = 1 - \beta \simeq 88{,}5\%$

Esercizio 6. Un certo titolo azionario nel 2022 veniva accreditato sul mercato ad un valor medio di 6 euro con una volatilitá/rischio (deviazione standard) di 4,74 euro. Nel 2023 vengono peró rilevati mensilmente i seguenti valori del titolo

$$16, 10, 12, 8, 0, 12, 10, 6, 10, 8, 4, 2$$

Assumiamo che i valori seguano una legge normale

(a) Si verifichi con un test di ipotesi se possiamo ritenere plausibile che la volatilitá del titolo sia rimasta invariata, con un livello di significativitá del 5%

Soluzione: Usiamo un test bilaterale sulla varianza $\sigma^2 = (4.74)^2 \approx 22.5$
Le ipotesi sono

$$\begin{cases} H_0 & : \sigma^2 = 22.5 = \sigma_0^2 \\ H_1 & : \sigma^2 \neq 22.5. \end{cases} \tag{2}$$

Si utilizza la statistica:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

che ha una distribuzione chi-quadro con $\nu = n - 1 = 11$ gradi di libertà.
Per $\alpha = 0.05$, la regione di rifiuto si ha per

$$\chi^2 < \chi^2_{0.975}(\nu = 11) = 3.816$$

$$\chi^2 > \chi^2_{0.025}(\nu = 11) = 21.92$$

La media del campione casuale di dimensione $n = 12$ calcolata dai dati a disposizione vale $\bar{x} = 8.166$
La varianza é:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{12} (x_i - \bar{x})^2 = 20.697$$

Quindi,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{11 \cdot 20.697}{22.5} = 10.12$$

Poiché $3.82 < \chi^2 < 21.92$ e' statisticamente accettabile che la volatilitá nominale sia ancora quella dichiarata nel 2023