○ Sì, segno qua una X sul circoletto perchè sono uno studente di anni passati <u>e diverso docente</u> e desidero anche un esame orale, e consegno questo foglio piegato in 2 insieme alla bella copia.

Chi si ritira, consegna **<u>solo</u>** questo foglio: col nome e una grande R.

Gli altri, tengono per sè questo foglio, e consegnano solo i fogli di bella copia piegati in due, tutti insieme.

Questo testo deve essere costituito da un foglio stampato fronte-retro con 6 quesiti in tutto.

Se manca qualcosa chiedere un'altra copia.

La valutazione è complessiva. Tutti i quesiti valgono ugualmente.

Anche soluzioni parziali vengono valutate.

SCRIVERE I CALCOLI OVVERO PASSAGGI.

CONSEGNARE SOLO LA BELLA COPIA, non diverse versioni.

Legenda

$*$ è richiesto il valore esatto. Può anche essere $+\infty$, $-\infty$, o una frase.

$\approx$ è richiesta una ragionevole approssimazione.

$^{\%}$ è richiesto il valore in percentuale, se serve ragionevolmente approssimato.

RIQUADRARE ovvero incorniciare I RISULTATI

ES. $1_{\mu 2018}$

$*$ Calcolare quanti sono i diversi possibili 6-meri dei nucleotidi del RNA ovvero le sequenze ("parole") di 6 lettere dell'alfabeto

$$A, C, G, U$$

(di cui recentemente si è molto scritto riguardo la ricerca contro il cancro).

SVOLGIMENTO

Relativamente alla questione medica, e in futuro eventualmente farmaceutica, si può vedere sul sito governativo statunitense PubMed l'abstract dell'articolo scientifico 6mer seed toxicity in tumor suppressive microRNAs.

Per fissare le idee (ma assolutamente non sarebbe necessario) scriviamo alcune delle sequenze/parole:

AAAAAA, AAAAAC, AAAAAG, AAAAAU,

AAAACA, AAAACC, AAAACG, AAAACU,
...
UUUUUA, UUUUUC, UUUUUG, UUUUU.

Tutte (4096, come vedremo) sono elencate in https://www.6merdb.org/

Per calcolare il numero osserviamo che
la 1^ lettera può essere scelta in 4 modi
la 2^ lettera può essere scelta in 4 modi
...
la 6^ lettera può essere scelta in 4 modi
e allora la sequenza di 6 lettere può essere costituita in un numero di modi pari a

$$4^6 = 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 =$$

$$\boxed{4096}$$

Oppure, si potrebbe anche dire che l'insieme delle sequenze/parole è in corrispondenza biunivoca con il prodotto cartesiano

$$\{A, C, G, U\} \times \text{...(in tutto 6 volte)...} \times \{A, C, G, U\}$$

e allora ha cardinalità

$$\Big(\#\{A, C, G, U\}\Big)^6$$

concludendo come prima.

ES. 2$_{\mu 2019}$
$\approx$ Si supponga di avere questi dati di un ospedale in anni successivi, relativi ai consumi di un certo farmaco:
1907 4257
1908 3956
1909 3936
1910 4183
1911 4114
1912 4525
1913 4188

1914 4111
1915 4404
1916 4180
1917 0
1918 0
1919 4361
1920 4035

Dopo aver eliminato gli outlier, determinare la media interquartile dei dati sul consumo del farmaco. (Si approssimi all'intero più vicino).

SVOLGIMENTO

Gli outlier sono i due 0 (verosimilmente dovuti alla guerra).

Ordiniamo il dataset rimanente, che ha 12 elementi:

3936, 3956, 4035, 4111, 4114, 4180, 4183, 4188, 4257, 4361, 4404, 4525

Eliminiamo i primi 3 e gli ultimi 3 valori

$$\Big[3936, 3956, 4035, \Big] 4111, 4114, 4180, 4183, 4188, 4257 \Big[, 4361, 4404, 4525\Big]$$

e la media dei 6 rimanenti è

$$\approx 4172.17$$

e approssimiamo come richiesto:

$$\boxed{4172}$$

Nota. Mostriamo la straordinaria insensibilità della media interquartile agli outlier rispetto alla media aritmetica. I 12 valori non nulli provenivano da una variabile aleatoria più o meno normale, e poi erano stati aggiunti 2 zeri in corrispondenza ad anni di guerra. Ora, in questo caso è ben evidente che si tratta di outlier per una serie di motivi:

molto diversi dagli altri valori
valori esattamente 0
corrispondenza con anni di guerra.

Ma in altri casi non è così evidente quali valori considerare outlier, per eliminarli "a mano". Allora potremmo fare la media aritmetica di tutti i valori, che però verrebbe fortemente influenzata dagli outlier. Invece la media interquartile dei 12 valori degli ultimi 12 anni, dal 1909 al 1919, senza escludere gli zeri,

1909 3936
1910 4183
1911 4114
1912 4525
1913 4188
1914 4111
1915 4404
1916 4180
1917 0
1918 0
1919 4361
1920 4035

è 4135, alquanto simile a quella di prima. Gli zeri sono finiti nelle porzioni scartate.

Insoma la media interquartile ci mostra di cogliere il *vero valore "medio"* della variabile aleatoria retrostante al fenomeno quando esso avviene *normalmente*, e lo fa in un modo *automatico*, con una formula, che non richiede l'esclusione a mano degli outlier, cosa impossibile se i dati sono milioni.

La media aritmetica 3503 invece è molto minore (essenzialmente a causa dei 2 zeri) della media 4188 dei 12 valori non nulli iniziali (e qua si vede la sensibilità della media aritmetica agli outlier).

Sarebbe bello mostrare che similmente avverrebbe considerando tutti i 14 valori iniziali, ma sfortunatamente la definizione di media interquartile per campioni con un numero non quadruplo di elementi non è facilissima per un calcolo a mano.

ES. 3$_{\mu 2019}$

$*$ In un certo procedimento chimico per la produzione di un farmaco, la temperatura nel reattore è, in $°C$,

$$T(z) := \frac{100}{z} + \frac{100}{1-z}$$

essendo z un parametro che possiamo modificare come vogliamo fra 0.1 e 0.9. Qual è il valore minimo della temperatura ottenibile?

SVOLGIMENTO

Da un punto di vista matematico dobbiamo trovare il minimo assoluto du

$$T(z) := \frac{100}{z} + \frac{100}{1-z} \qquad z \in [0.1, 0.9] \quad \text{ovvero } 0.1 \leq z \leq 0.9$$

e poi da un punto di vista fisico dobbiamo aggiungere $^\circ C$ al risultato.

Deriviamo la funzione

$$T'(z) := -\frac{100}{z^2} + \frac{100}{(1-z)^2}$$

e risolviamo la disequazione *derivata* > 0 ovvero $T'(z) > 0$

$$-\frac{100}{z^2} + \frac{100}{(1-z)^2} > 0$$

denominatore comune

$$\frac{-100\,(1-z)^2 + 100\,z^2}{(z-1)^2 z^2} > 0$$

semplificazioni algebriche

$$\frac{100\,(2\,z-1)}{(z-1)^2\,z^2} > 0$$

equivalente a

$$2\,z - 1 > 0$$
$$z > \frac{1}{2}$$

cioè la funzione decresce a 0.1 a $\frac{1}{2} = 0.5$ e poi cresce fino a 0.9, avendo un punto di minimo relativo e assoluto in $\frac{1}{2}$, e quel minimo è

$$T\left(\frac{1}{2}\right) = \frac{100}{\frac{1}{2}} + \frac{100}{1-\frac{1}{2}}$$

che fa 400, e allora con l'unità di misura

$$\boxed{400^\circ C}$$

Grafico online su WolframAlpha: link->

ES. 4$_{\mu 2019}$
% Consideriamo 4 neonati che vengono sottoposti ad una terapia con proba-

bilità di esito fatale (morte) del 7%. Considerando indipendenti gli eventi, che probabilità c'è che sopravvivano tutti?

SVOLGIMENTO

Probabilità dell'evento complementare, sopravvivenza: 93%, ovvero 0.93.

Sopravvive il primo con probabilità 0.93

sopravvive <u>anche</u> il secondo con probabilità $0.93 \cdot 0.93$

sopravvive <u>anche</u> il terzo con probabilità $0.93 \cdot 0.93 \cdot 0.93$

e infine considerando analogamente anche il quarto $0.93^4 \approx 0.748$, cioè in forma percentuale

$$\boxed{\approx 74.8\%}$$

ovvero molto più ragionevolmente (considerando la più che verosimile imprecisione del dato iniziale del 7%)

$$\boxed{\approx 75\%}$$

Oppure, si può vedere la questione così, individuando uno schema successo-insuccesso e allora una distribuzione binomiale:

$$p := P(\text{successo=sopravvivenza}) = 1 - 0.07 = 0.93$$

$$P(\text{sopravvivono tutti}) =$$

$$= P(\text{4 successi su 4 prove indipendenti}) = \binom{4}{4} \cdot p^4 \cdot (1-p)^{4-4} =$$

$$= \frac{4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot 4} \cdot 0.93^4\, 0.07^0 = 0.93^4$$

concludendo come prima.

Oppure, chiamando successo la morte (un po' sgradevolmente)

$$p := P(\text{successo=morte}) = 0.07$$

la sopravvivenza diviene un insuccesso:

$$P(\text{sopravvivono tutti}) =$$

$$= P(0 \text{ successi su 4 prove indipendenti}) = \binom{4}{0} \cdot p^0 \cdot (1-p)^{4-0} =$$

$$= \frac{4!}{0! \cdot (4-0)!} \cdot 0.07^0 \cdot 0.93^4 = 0.93^4$$

concludendo come prima.

ES. 5$_{\mu 2019}$
% Per un campione gaussiano $N(20.19, 13.5)$ calcolare

$$P(X_1 + ... + X_{19} \leq 415)$$

SVOLGIMENTO
Applichiamo l'Approssimazione Normale

$$P\Big(X_1 + ... + X_n \leq x\Big) \approx \Phi\Big(\frac{x - n\mu}{\sigma\sqrt{n}}\Big)$$

con
$n = 19$
$x = 415$
$\mu = 20.19$
$\sigma^2 = 13.5$ ovvero $\sigma = \sqrt{13.5}$
trovando

$$P\Big(X_1 + ... + X_{19} \leq 415\Big) \approx \Phi\Big(\frac{415 - 19 \cdot 20.19}{\sqrt{13.5}\sqrt{19}}\Big) \approx$$

$$\approx \Phi(1.96) \approx$$

valore (approssimatamente) ben noto, anche nella forma $\phi_{0.975} \approx 1.96$,

$$\approx 0.975$$

concludendo

$$\boxed{\approx 97.5\%}$$

(E certo, se l'argomento di Φ non veniva proprio 1.96 – oppure 0, altro valore di cui classicamente si conosce a memoria $\Phi(x)$ – avremmo dovuto usare le tavole dei quantili normali, o un software adatto).

ES. 6$_{\mu 2019}$

$*$ Supponiamo che per un test statistico, con ipotesi (nulla) H_0 vera e alternativa A, ad un certo livello $\alpha > 5\%$, la regione critica sia $\{T > 12.65\}$ e il calcolo dello stimatore del test dia $T = g(x_1, ..., x_n) = 20.19$. Quale delle seguenti affermazioni è vera?

- Era il caso in generale sperato
- Si commette un errore di prima specie
- Si commette un errore di seconda specie
- Il risultato è sostanzialmente inutile perchè $\alpha > 5\%$
- Non è possibile rispondere perchè non è specificato il livello α.

Svolgimento

Lo stimatore T vale 20.19 che $\in$ alla regione critica, e l'ipotesi (nulla) è vera. Allora "male respingo ipotesi vera", cioè

$$\boxed{\text{si commette un errore di prima specie}}$$

(Che è il caso peggiore).