

## Esercizi in preparazione dell'esame

Cosa significa che la mediana per il tempo di attesa per un esame è pari a 10 giorni?

E cosa significa se sappiamo anche che il 95° percentile è 14 giorni?

In particolare: quante persone in percentuale hanno aspettato tra 10 e 14 gg? E quante hanno aspettato più di 14 gg?

Disegnare un boxplot, avendo a disposizione i seguenti dati: mediana=30, I° quartile=26, III° quartile=32, max=37, min=10, e sapendo che ci sono solo due persone (di 10 e di 12 anni) che hanno meno di 25 anni. In particolare, dove 'termina' il baffo inferiore, e dove il superiore?

Utilizzando un boxplot, dire se ci sono degli outlier avendo a disposizione le seguenti informazioni: min=18, q1=20, q3=30, max=44.

Quanto è lungo il baffo superiore e quanto quello inferiore?

Si sa che una malattia è indipendente dal 'genere' e si sa pure che la popolazione oggetto dello studio è costituita per il 55% da maschi e per il 45% da femmine. E' noto che la prevalenza della malattia è pari al 3%. Completare la tavola di contingenza (tavola 2x2) e dire in particolare quanti maschi malati e quante femmine sane ci aspettiamo.

Un test diagnostico ha una sensibilità  $Se=80\%$  e una specificità  $Sp=90\%$  e viene utilizzato per una malattia che ha una prevalenza del 5%. Calcolare i valori predittivi del test. Cosa succede se la prevalenza passa dal 5% al 40%?

In un test diagnostico  $Se=85\%$ ,  $Sp=95\%$ . Quanto vale la probabilità che una persona sana risulti positiva al test?

I dati di un certo parametro si distribuiscono secondo una gaussiana di media 105 e deviazione standard 12 (scriviamo quindi  $N(105,12)$ ). Ricordando le proprietà della gaussiana, che percentuale di dati mi aspetto superino il valore 129?

Chi è (circa) il 16° percentile?

La sensibilità e la specificità di un test diagnostico sono pari a 60% e a 80%, mentre il valore predittivo positivo è pari a 47%. Calcolare la probabilità che una persona sia sana, sapendo che è risultata positiva al test.

Una popolazione è costituita per il 10% da mancini. In tale popolazione gli ultrasessantenni sono il 30%. Quanti sono i mancini con meno di 70 anni?

In un gruppo composto da 80 persone, le età sono tali che: min=25°perc.=19, Me=35, 75°perc.=41, max=54. Quante persone hanno 19 anni?

In un outlet si effettua lo sconto del 40% su dei capi che erano già scontati del 20%. Se un capo inizialmente costava 80 Euro, quale sarà il suo prezzo finale all'outlet?

Un test diagnostico ha sensibilità  $Se=80\%$ , specificità  $Sp=90\%$ , e la malattia ha una prevalenza  $Pr=5\%$ ? Quanti veri positivi mi aspetto?

Calcolare l'intervallo di fiducia al 95% per la stima della media di una popolazione di dati gaussiani, sapendo che la media campionaria è pari a 30, la deviazione standard campionaria è pari a 3 e sapendo che il campione ha numerosità pari a 50 (prendere  $q=2$ ).

Cosa succede all'intervallo se la numerosità aumenta?

Quanto dovrebbe essere grande il campione se vogliamo arrivare ad un intervallo di fiducia con misura minore o uguale a 1?

Abbiamo eseguito il test del chi quadro 'sulle variabili malattia e genere', e abbiamo ottenuto  $p=36\%$ . Cosa significa?

E' stato sperimentato un nuovo farmaco (rispetto ad uno normalmente utilizzato) ... Il test t di Student ha dato un output  $p=0,0001$ . Come leggiamo questo risultato?

E se il p-value fosse venuto pari a  $5,1\%$ ?

In un articolo leggiamo che il RR (rischio relativo) per lo sviluppo di una certa malattia rispetto ad un certo fattore di esposizione ha un intervallo di fiducia al  $95\%$  pari a  $(2,3; 3,8)$ . Cosa significa?

Cosa succede a un confidence interval se si chiede una 'fiducia' del  $99\%$  anziché del  $95\%$  (cioè: l'intervallo rimane uguale, diventa più grande, più piccolo oppure viene traslato)?

Cosa succede a un confidence interval se cambia la media (cioè: l'intervallo rimane uguale, diventa più grande, più piccolo oppure viene traslato)?

Su un articolo leggiamo che è stato effettuato il test esatto di Fisher per le variabili genere e malattia e che  $p=0,37\%$ .

Chi è l'ipotesi  $H_0$  di questo test?

A che risultato siamo pervenuti?

Un certo parametro si distribuisce secondo una gaussiana  $N(42;4)$ .

Che valore ha la mediana?

Quali sono i punti di flesso?

Che percentuale di dati è compresa tra 42 e 50?

Chi è l' $84^\circ$  percentile?

Le età di un gruppo di persone sono tali che  $\text{minimo}=75^\circ\text{percentile}=22$ .

Quante persone (in percentuale) hanno 22 anni?

Un gruppo di 15 persone ha età media pari a 24 anni. Entrano 2 persone di 40 anni. Quanto vale la nuova età media (scrivere solo la formula – non occorre fare i calcoli)?

Il confidence interval all' $80\%$  per la stima di un certo parametro è  $(26; 28)$ . Cosa succede all'intervallo se passo da una fiducia (confidence) dell' $80\%$  a una del  $90\%$

Sempre in riferimento all'esercizio di sopra, cosa succede se la media campionaria  $m$  passa dal valore 25 al valore 29?

In un test diagnostico sono stati trovati 425 test negativi, 38 test positivi. Si sa pure che i falsi negativi erano 12, mentre erano 17 i falsi positivi. Riempire la tavola di contingenza e calcolare la sensibilità, la specificità e i valori predittivi del test.

	<b>T-</b>	<b>T+</b>	
<b>M-</b>			
<b>M+</b>			

Di un test diagnostico sono note la Sensibilità  $Se = 75\%$  e la Specificità  $Sp = 85\%$  e viene somministrato ad una popolazione in cui la prevalenza della malattia  $M$  è dell'  $8\%$ .

a: Con che probabilità una persona, risultata positiva al test, è malata?

b: E se il test venisse somministrato ad una popolazione con prevalenza della malattia del  $26\%$ ?

Ricordando la formula per la stima a priori della numerosità campionaria  $n \geq (2qs/k)^2$ , e sapendo che abbiamo ottenuto  $n \geq 100$ , cosa succede se dimezziamo  $k$

In un articolo troviamo scritto: t-test per dati accoppiati:  $p < 0,1\%$ .

Che tipo di studio è stato effettuato?

A che risultato siamo pervenuti?

In un articolo troviamo scritto: test del chi-quadro:  $p=0,001\%$ .

Chi è l'ipotesi  $H_0$  di questo test?

A che risultato siamo pervenuti?

In uno studio longitudinale (prima – dopo – follow up) è stato somministrato un farmaco. I risultati, prima, dopo, follow up, sono descritti nel boxplot qui a fianco (le diagonali sono un errore di visualizzazione...).

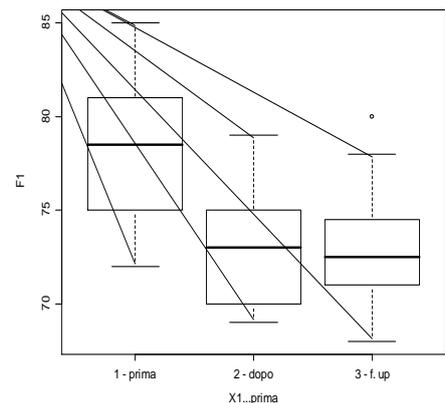
Commentare i risultati. In particolare:

il medicinale ha avuto effetto?

In caso affermativo, quando ha avuto effetto?

L'effetto del farmaco si è mantenuto anche al follow up?

Pensando di fare un t-test, che valori  $p$  ci aspettiamo?



Cosa significa se il RR (esposizione ad un certo fattore vs malattia) ha una stima con c.i. pari a  $(0,8 ; 1,2)$ ?

Cosa significa se il RR (esposizione ad un certo fattore vs malattia) ha una stima con c.i. pari a  $(8 ; 12)$ ?

Uno studente ha una media (pesata sui crediti) pari a 27 nei primi 4 esami per un totale di 22 crediti.

Quanto vale la nuova media dopo aver preso nel 5° esame (con 5 crediti) un voto pari a 30?

Quanto varrebbe la media se non venisse pesata sui crediti ma solo sul numero di esami sostenuti?

Le età di un gruppo di persone vanno dai 10 ai 29 anni. Il 25° percentile è pari a 18 anni, mentre il 75° percentile è pari a 24 anni.

Ci sono outlier, pensando di disegnare un boxplot?

Quanto è lungo il baffo inferiore e quanto quello superiore?

Le età di un gruppo di persone si distribuiscono secondo una gaussiana  $N(27;3)$ .

Calcolare l'età mediana

Quante persone in percentuale sono circa presenti tra 27 e 30 anni?

In un articolo troviamo scritto che è stato eseguito il test del chi quadro per le variabili malattia e fascia d'età, ottenendo un valore p pari a 36%.

Che test è stato effettuato (cioè, chi è  $H_0$ )?

A che risultato siamo pervenuti?

Stimare la numerosità a priori di un campione se si vuole ottenere un 'confidence interval' al 95% di lunghezza non superiore a 1, sapendo che la media campionaria è uguale a 34, la deviazione standard del campione è uguale a 5 (prendere  $q=2$ ).

Cosa significa se, in un test di ammissione, i risultati sono stati tali che il 76° percentile è stato pari a 4 (su 10)?

Per un certo parametro è noto che:  $\min=16$ ,  $q_1=18$ ,  $Me=20$ ,  $q_3=24$ ,  $\max=32$ .

Utilizzando il grafico boxplot, quanti outlier mi aspetto?

Qui a fianco la tavola dei valori osservati Malattia/Genere.

Quanti uomini malati mi aspetto in percentuale in caso di malattia indipendente dal genere?

E' *ragionevole* pensare che la malattia sia indipendente dal genere?

	M+	M-	
U	20	120	
D	5	95	