



CONCORSO PUBBLICO, PER TITOLI ED ESAMI, PER LA COPERTURA DI N. 2 POSTI DI COLLABORATORE TECNICO PROFESSIONALE – STATISTICO (CATEGORIA D – POSIZIONE ECONOMICA D), A TEMPO PIENO E INDETERMINATO, PER LE ESIGENZE DELL’INMP, (CODICE PROCEDURA COLL.STAT./2018)

PROVA SCRITTA

QUESTIONARIO N.1 (Non sorteggiato)

Nei seguenti test a risposta multipla il candidato individui la risposta corretta (una e solo una) barrando con una x in corrispondenza della lettera e descriva sinteticamente il motivo della scelta, utilizzando al massimo le 5 righe riportate sotto le risposte

1. Un ricercatore ha bisogno di 90 soggetti e una potenza dell’80% al fine di rilevare una reale differenza di almeno il 20% tra due bracci di studio. Seleziona la frase corretta:
- a. Il ricercatore dovrebbe ridurre il numero dei soggetti a 80 per ridurre la probabilità di errore di 2° tipo.
 - b. Il ricercatore avrebbe bisogno di meno di 90 soggetti per rilevare una reale differenza di almeno 15% tra due bracci di studio, mantenendo la potenza all’80%.
 - c. Il ricercatore avrebbe bisogno di meno di 90 soggetti e una potenza del 90% per rilevare una reale differenza di almeno 20% tra due bracci di studio.
 - d. Il ricercatore avrebbe bisogno di meno di 90 soggetti e una potenza del 80% per rilevare una reale differenza superiore al 20% tra due bracci di studio.

2. In uno studio comparativo, un film è stato mostrato a un gruppo di bambini alle 8 di mattina, un altro film alle 11 di mattina. È stato scoperto che durante il film delle 11 sono stati mangiati più crackers rispetto al film delle 8. I ricercatori concludono che differenti film hanno un effetto diverso sull’appetito. I risultati non possono essere accettati perché:
- a. Lo studio non era in doppio cieco. Né gli investigatori né i bambini avrebbero dovuto sapere che film era loro mostrato
 - b. I ricercatori erano parziali. Sapevano in anticipo quali risultati speravano di ottenere dallo studio
 - c. I ricercatori avrebbero dovuto usare molte ciotole con crackers contenuti in modo random in ciascuna ciotola
 - d. L’orario del film mostrato è una variabile di confondimento



-
-
-
-
3. In un articolo scientifico gli autori studiano un test diagnostico in una popolazione di pazienti in cui la prevalenza di disfunzione ventricolare era il 5%. Le proprietà del test sono studiate usando un campione casuale di 155 pazienti estratto dalla popolazione:
- a. Scegliendo un campione casuale, gli autori possono ottenere approssimativamente uguale precisione per specificità e sensibilità
 - b. Il valore predittivo positivo può essere stimato con maggiore precisione rispetto alla specificità nel campione in esame
 - c. La specificità può essere stimata con maggiore precisione rispetto alla sensibilità nel campione in esame
 - d. Un campione stratificato avrebbe dato migliore precisione per la sensibilità rispetto alla specificità

-
-
-
-
-
4. All'inizio di uno studio di coorte l'esposizione è misurata attraverso un questionario. Durante lo studio non ci sono persi al follow-up. Alla fine del follow-up il numero dei casi è conosciuto e diviso tra esposti e non esposti. L'Odds Ratio (OR) è usato come misura di associazione. Quale commento è più appropriato?
- a. L'OR ha sempre una interpretazione utile
 - b. I ricercatori avrebbero dovuto calcolare il rischio relativo
 - c. L'OR approssima l'incidence rate ratio se la malattia è rara
 - d. L'OR sottostima l'incidence rate ratio

5. Se l'incidenza nei non esposti è 0.1 e quella negli esposti è 0.4:

- a. Il rischio attribuibile è 4
- b. Il rischio relativo è 0,3
- c. Il rischio attribuibile è 0,3
- d. Il rischio attribuibile è 3

6. I voti dell'esame di storia sono stati usati per creare il seguente grafico:

```
Stem-and-leaf of visitors  N = 40
Leaf Unit = 1.0

 2  5
 3
 4
 5  26
 6  1359
 7  00338
 8  01122335789
 9  00111233445677899
```

Il ricercatore decide di aggiungere 1 punto ai voti di ognuno. Qual è l'effetto?

- a. La deviazione standard aumenta di 1
- b. La mediana varia, ma la media non varia perché la distribuzione non è simmetrica
- c. La deviazione standard non varia, ma la media e la mediana aumentano
- d. La deviazione standard, la media e la mediana aumentano



7. Supponiamo di lanciare due dadi. Si consideri la somma dei due dadi: sia A l'evento "si osserva un numero pari" e B l'evento "si osserva un numero maggiore di 7". Quale delle seguenti affermazioni è vera?

- a. Gli eventi A e B sono mutuamente esclusivi.
- b. Gli eventi A e B sono indipendenti.
- c. L'intersezione tra A e B è l'insieme [6, 8, 10, 12].
- d. Nessuna delle precedenti.

8. Un test a risposta multipla ha 5 domande, ognuna con 5 possibili risposte, di cui una sola corretta. Se rispondi sempre a caso, qual è la probabilità di rispondere correttamente a 3 domande?

- a. 0,00032
- b. 0,0082
- c. 0,0512
- d. 0,0216

9. Nel costruire un intervallo di confidenza per la media della popolazione è stato utilizzato un campione di 100 osservazioni. La stima intervallare risultante è stata $28,76 \pm 1,48$. Se la numerosità campionaria fosse stata 400 invece che 100, la stima intervallare sarebbe stata:

- a. $28,76 \pm 0,74$
- b. $28,76 \pm 0,37$
- c. $28,76 \pm 0,49$
- d. $28,76 \pm 1,74$



10. Si considerino i seguenti importi in euro riguardanti un campione casuale di richieste di copertura assicurativa per incidenti stradali derivato da un noto istituto assicurativo. L'azienda intende analizzare i dati per studiare la distribuzione degli importi:

6.751	9.908	3.461	2.336	21.147	2.332	189	1.185	370
1.414	4.668	1.953	10.034	735	802	618	180	1.657

Indicare quali delle seguenti affermazioni è corretta:

- a. La distribuzione dei dati è simmetrica.
- b. La distribuzione dei dati presenta una asimmetria positiva
- c. La distribuzione dei dati presenta una asimmetria negativa
- d. La distribuzione dei dati è bimodale



QUESTIONARIO N.2 (Sorteggiato)

Nei seguenti test a risposta multipla il candidato individui la risposta corretta (una e solo una) barrando con una x in corrispondenza della lettera e descriva sinteticamente il motivo della scelta, utilizzando al massimo le 5 righe riportate sotto le risposte

1. In uno studio cross-sectional basato su dati censuari il 13% dei 20enni era costituito da mancini, solo il 5% dei 50enni e nessuno degli ultra 80enni. Questo trend con l'età mostra:
- a. Un bias nella selezione della popolazione studiata
 - b. Un possibile incremento nella proporzione di mancini negli ultimi anni
 - c. Un reale svantaggio di sopravvivenza dei mancini
 - d. La necessità di standardizzare per età

2. La distribuzione dei livelli sierici di alfa tocoferolo (siero vitamina E) in un campione è approssimativamente normale con media 800 $\mu\text{g/dL}$. La deviazione standard della popolazione è 200 $\mu\text{g/dL}$. Avendo a disposizione un campione di 400 individui, l'intervallo approssimativo entro il quale la popolazione presenta il livello medio con il 99% della probabilità, varia:
- a. tra 450 e 1150 $\mu\text{g/dL}$
 - b. tra 600 e 1000 $\mu\text{g/dL}$
 - c. tra 774 e 826 $\mu\text{g/dL}$
 - d. tra 200 e 1400 $\mu\text{g/dL}$



3. Durante uno studio di 20 anni 6 persone sono seguite per misurare l'insorgenza di infezioni del tratto respiratorio superiore. Questa infezione può verificarsi più di una volta, e in questo studio sono considerati come esito tutti gli eventi patologici:

- 1 persona è persa al follow-up dopo 10 anni;
- 2 persone muoiono rispettivamente dopo 15 e 18 anni per causa diversa;
- 1 persona ha avuto la prima infezione del tratto respiratorio dopo sette anni e la seconda infezione dopo 12 anni di follow-up. Entrambe le infezioni richiedono un anno e mezzo prima della completa guarigione. Questa persona è seguita fino alla fine dello studio;
- 2 persone sono seguite per tutto il periodo senza che la malattia insorga.

Qual è il tasso di incidenza nello studio?

- a. 0,06 anni persona
- b. 0,02 anni persona
- c. 0,03 anni persona
- d. 0,15 anni persona

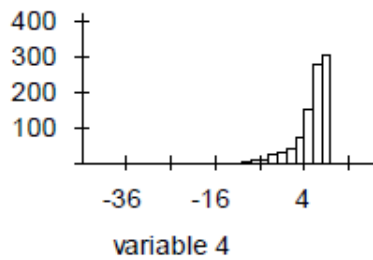
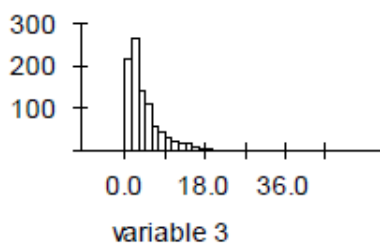
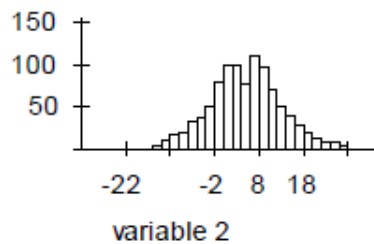
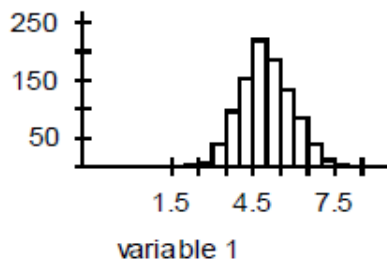
4. L'information bias in uno studio di coorte può essere evitato assicurando:

- a. Che si verifichino altre malattie concomitanti
- b. Che il follow-up sia completo
- c. Che le modalità di misura della malattia siano altamente standardizzate
- d. Che le persone che misurano la malattia non siano a conoscenza dell'esposizione dei partecipanti



5. Una malattia cronica con 80 malati presenti su 1000 persone e 2 nuovi casi ogni anno su 100 persone a rischio:
- Ha un'incidenza annua dell'8%
 - Ha una prevalenza del 2%
 - Ha una stima della durata media di 4 anni
 - Ha una stima della durata media di 0.4 anni

6. Quali dei seguenti indici di tendenza centrale e dispersione sintetizzano meglio la distribuzione della variabile 4?



- Media= 4.99, mediana = 3.13, deviazione standard= 5.49
- Media= 4.89, mediana = 4.83, deviazione standard= 7.99
- Media= 5.01, mediana = 6.87, deviazione standard= 5.49
- Media= 4.96, mediana = 4.93, deviazione standard= 0.96



7. Se in un esperimento casuale si ha che $P(A) = 0,7$ può accadere che $P(A \cup B) = 0,5$?

- a. NO mai
- b. Solo se A e B sono eventi incompatibili
- c. Solo se $B = \emptyset$
- d. Solo se $A \cap B$ è un insieme vuoto

8. Il numero di televisori che escono ogni giorno da una certa linea di produzione si distribuisce come una variabile casuale con deviazione standard di 15. La media giornaliera della linea di produzione determinata su un campione di 25 giorni è 452,3. Quale dei seguenti intervalli rappresenta un intervallo di confidenza al 95% per la media della produzione in un giorno?

- a. $452,3 \pm 5,4$
- b. $452,3 \pm 5,88$
- c. $452,3 \pm 11,3$
- d. $452,3 \pm 7,63$



9. Un idraulico afferma di poter completare l'installazione di un box doccia in meno di un'ora. Per un campione di 16 interventi, l'idraulico impiega una media di 65 minuti e una deviazione standard di 8 minuti. Qual è la statistica test osservata?

- a. $t = 1,79$
- b. $t = 2,50$
- c. $Z = 1,79$
- d. $Z = 2,07$

10. Una prova del carico di rottura di 6 cavi d'acciaio costruiti da una ditta ha mostrato un carico di rottura medio campionario pari a 7750 kg e uno scarto quadratico medio campionario di 145 kg, mentre il costruttore afferma che il carico di rottura medio è di 8000 kg. Si vuole stabilire se l'affermazione del costruttore non sia corretta e che il carico di rottura è inferiore con un livello di significatività del 5%. Indicare quali delle seguenti affermazioni è corretta.

- a. Ipotizzando che la popolazione abbia distribuzione normale, utilizzo un test Z per la media unilaterale.
- b. Ipotizzando che la popolazione abbia distribuzione normale, utilizzo un test Z per la media bilaterale.
- c. Nessuna delle precedenti.
- d. Ipotizzano che la popolazione abbia distribuzione di Poisson



QUESTIONARIO N.3 (Non sorteggiato)

Nei seguenti test a risposta multipla il candidato individui la risposta corretta (una e solo una) barrando con una x in corrispondenza della lettera e descriva sinteticamente il motivo della scelta, utilizzando al massimo le 5 righe riportate sotto le risposte

1. In uno studio sulla relazione tra utilizzo della pillola anticoncezionale e incremento del peso corporeo sono state seguite per un anno 500 donne di età compresa tra i 20 e i 40 anni, di cui 250 non assumevano la pillola e 250 la assumevano. Dopo un anno di osservazione la loro distribuzione rispetto all'incremento del peso è descritta nella tabella.

	Pillola sì	Pillola no	Totale
Incremento di peso sì	50	40	90
Totale donne	250	250	500

Quale dei seguenti valori corrisponde al rischio attribuibile di incremento di peso tra le donne che prendono la pillola, cioè descrive la quantità di donne che hanno avuto un incremento di peso attribuibile al fatto di essere state esposte agli ormoni estroprogestinici?

- a. RA = 10 donne su 500
- b. RA = 10 donne su 250
- c. RA = 50 donne su 250
- d. RA= 50 donne su 500

2. Quale dei seguenti termini significa che l'effetto osservato è reale?

- a. α level
- b. β level
- c. $1 - \alpha$
- d. $1 - \beta$



3. Qual è la frazione dei malati tra gli esposti attribuibile all'esposizione?

	Non esposti	Esposti
Malati	9	17
Sani	7	5

- a. 0,27
- b. 0,60
- c. 0,54
- d. 0,77

4. In uno studio caso-controllo è stata esaminata l'associazione tra fumo e rischio di malattia di Parkinson. La tabella seguente fornisce i risultati. Quale dei seguenti odds ratio è corretto?

	Controlli	Casi
Fumatori	55	30
Non fumatori	45	70

- a. 0,78
- b. 0,35
- c. 2,85
- d. 1,83



5. Quale delle seguenti affermazioni è FALSA

- a. L'analisi della varianza è un'estensione del t-test
- b. Quando uno studio ha solo una variabile indipendente con 2 gruppi, F e t sono identici.
- c. La direzione delle differenze tra gruppi deve essere specificata nel valore critico F
- d. I gradi di libertà sono il numero di punteggi liberi di variare una volta che le medie sono note.
- e. Il t-test è un test parametrico

6. Un epidemiologo in Tanzania vuole studiare l'efficacia dell'integrazione di ferro per la prevenzione di infezione da HIV. Vuole assicurarsi che solo i soggetti ancora liberi da HIV siano arruolati nel suo trial. Quindi esamina un grande gruppo di persone usando un test diagnostico. In base ai risultati del test, decide chi può partecipare al suo trial con integrazione di ferro. Per questo scopo, è importante che il suo test diagnostico abbia un'alta

- a. sensibilità
- b. valore predittivo positivo
- c. specificità
- d. valore predittivo negativo



7. Sia X una variabile casuale binomiale con $n = 10$ e $p = 0,5$. Allora

- a. La X ha una moda in $X = 1$
- b. La X ha due mode in $X = 4$ e $X = 6$
- c. La X ha una moda in $X = 5$
- d. La X ha una moda in $X = 8$

8. Quale delle seguenti frasi NON è vera?

- a. La regione di rifiuto è l'insieme di tutti i valori della statistica test per cui l'ipotesi alternativa viene rifiutata.
- b. Una statistica test è una funzione dei dati campionari sulla base della quale si decide se rifiutare o meno l'ipotesi nulla.
- c. La regione di rifiuto è l'insieme di tutti i valori della statistica test per cui l'ipotesi nulla viene rifiutata.
- d. Una buona procedura di test delle ipotesi deve comportare una probabilità dell'errore del I tipo e del II tipo piccola.



9. In un sondaggio precedente ad un referendum è stato estratto un campione casuale di 1600 persone dalla popolazione dei potenziali votanti. Al fine di valutare se verrà raggiunto il quorum necessario, a ciascuna persona del campione stato chiesto se andrà a votare per il referendum oppure no, e 800 persone hanno detto che andranno a votare. Si vuole calcolare un intervallo di confidenza al 95% per la proporzione di persone che andranno a votare nella popolazione. Tale intervallo è pari a:

- a. [0,47; 0,53]
- b. [0,13; 0,97]
- c. [0,34; 0,66]
- d. [0,69; 0,76]

10. In uno studio di coorte il Rischio Relativo per COPD per fumatori moderati rispetto ai non fumatori era 4. Per i forti fumatori rispetto ai non fumatori il Rischio Relativo era 10. Cosa sarebbe accaduto in questo studio se per il Rischio Relativo per COPD fossero stati usati i forti fumatori come categoria di riferimento?

- a. Per i non fumatori 0.1 e per i fumatori moderati 0.4
- b. Per i non fumatori 0.2 e per i fumatori moderati 0.6
- c. Per i non fumatori 4 e per i fumatori pesanti 10
- d. Non può essere calcolato con i dati disponibili

PROVA PRATICA

PROBLEMA N.1 (Non sorteggiato)

E' stato effettuato uno studio randomizzato su un piccolo campione (25 pazienti) per valutare l'efficacia di un farmaco calcio-antagonista nel trattamento di una grave patologia coronarica correlata all'ipertensione. L'end-point principale, ovvero l'ESITO (morto=1, vivo=0), è stato osservato dopo un periodo di follow-up fisso di 2 anni. Il tempo all'evento non è stato considerato in quanto tutti gli eventi avvenivano entro il primo anno dall'arruolamento. Oltre all'esito è stato valutato come end-point surrogato la riduzione pressoria dopo 1 mese dall'inizio del trattamento (PRESS_DELTA), mentre PRESS_BASE rappresenta il livello di pressione alla randomizzazione. Alcune variabili potenzialmente influenzanti l'esito ed il delta pressorio sono state osservate, come il sesso (SEX=0 femmina, SEX=1 maschio), l'età (AGE in anni), il tempo intercorso tra l'inizio dei sintomi e la diagnosi (T_DIA, in mesi), la PO2.

ID	T_DIA	SEX	AGE	PO2	PRESS_BASE	PRESS_DELTA	ESITO
1	225	0	34	64	124	-24	0
2	36	0	44	73	130	-31	1
3	152	1	12	77	126	-37	0
4	82	1	35	71	102	-22	1
5	147	1	45	74	141	-27	1
6	12	1	26	66	210	-17	1
7	9	0	13	63	140	-7	0
8	24	0	27	55	134	-15	0
9	67	0	57	59	108	-14	1
10	9	0	47	70	120	-20	1
11	14	1	49	76	129	-17	1
12	20	0	8	65	152	-12	1
13	1	0	49	77	119	-10	0
14	5	1	45	68	147	-10	0
15	36	1	62	85	117	-26	1
16	49	1	33	80	128	-16	1
17	13	0	58	56	145	-25	1
18	10	0	42	63	149	-44	1
19	43	0	48	70	98	-18	1
20	38	0	36	81	130	-17	1
21	29	1	18	82	144	-10	0
22	1	1	43	62	134	-17	0
23	2	1	60	81	119	-10	1
24	9	0	32	62	150	-10	1
25	2	0	53	52	103	-15	0

1. Il farmaco può essere considerato efficace se è in grado di produrre una riduzione della pressione arteriosa superiore a 10 mmHg. Sapendo che la riduzione media della pressione osservata nel campione



del disco e a parità di età. Di seguito sono elencati gli elementi costitutivi dell'obiettivo conoscitivo sottostante e tre formulazioni (F1, F2, F3) sul ruolo che ognuno di essi gioca nell'obiettivo conoscitivo, descritto con una lettera (E per esito di salute, D per determinante, C per confondente, M per modificatore d'effetto, O per misura di occorrenza, A per misura di associazione); scegli la formulazione di ruoli che descrive esattamente l'obiettivo conoscitivo.

	F1	F2	F3
Insorgenza di ernia del disco	E	O	E
Rischio di insorgenza di ernia	O	A	A
Posizioni di lavoro scorrette	D	D	C
Predisposizione all'ernia	C	E	D
Età	C	M	C

PROBLEMA N.2 (Sorteggiato)



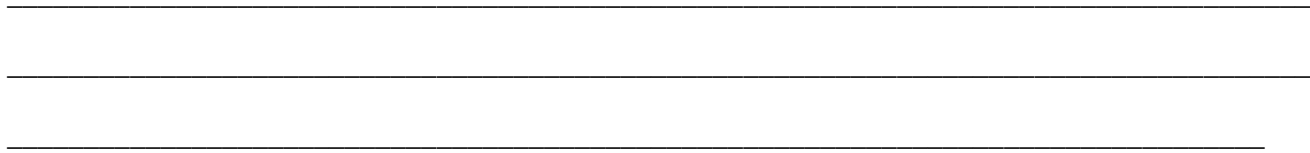
1. In uno studio condotto su un campione di 15 studenti di statistica è stato valutato l'impatto del parlare in pubblico sulla pressione arteriosa. All'inizio di una lezione agli studenti è stata misurata la pressione sanguigna sistolica. Durante la lezione l'insegnante ha invitato ciascuno studente a rispondere alle domande sugli argomenti della lezione. Dopo aver parlato, gli studenti hanno nuovamente registrato la pressione sanguigna. I valori risultanti sono indicati di seguito insieme alle statistiche riassuntive.

Prima	102	123	128	125	136	108	130	113	105	136	146	138	108	149	132
Dopo	141	142	146	136	144	111	132	114	103	133	140	130	100	133	113
Dopo- prima	39	19	18	11	8	3	2	1	-2	-3	-6	-8	-8	-16	-19

Prima	Dopo	Differenza (Dopo- prima)
Media = 125.3	Media = 127.9	Media = 2.6
DevSt = 15.03	DevSt = 15.5	DevSt = 14.87

Utilizzando le informazioni disponibili al candidato viene richiesto: 1) di formalizzare ed eseguire un test statistico adeguato a valutare la significatività dell'incremento della pressione sanguigna; 2) di descrivere il risultato ottenuto

2. Si consideri la seguente tabella di contingenza, che sintetizza la distribuzione di un campione di 40 soggetti secondo una patologia e un fattore di rischio ad essa potenzialmente associato.



PROBLEMA N.3 (Non sorteggiato)

