



**CONCORSO PER L'ASSUNZIONE DI 10 ESPERTI
CON ORIENTAMENTO NELLE DISCIPLINE STATISTICHE**

(Bando del 21 dicembre 2017 – lett. E)

Testo n. 2

STATISTICA E PROBABILITÀ

Due quesiti a scelta tra i tre proposti dalla Commissione

QUESITO N. 1

1. Il candidato esponga brevemente le definizioni di probabilità nei differenti approcci.
2. Sia dato un campione casuale semplice (X_1, \dots, X_n) da una popolazione X con distribuzione di probabilità esponenziale di parametro θ , cioè con funzione di densità:

$$f(x; \theta) = \theta e^{-\theta x}$$

con $x > 0$ e parametro incognito $\theta > 0$.

Ricavare:

- a. la distribuzione della variabile casuale $S = X_1 + X_2$;
- b. lo stimatore di massima verosimiglianza $\hat{\theta}$ di θ ;
- c. il quantile q_p in funzione di θ e lo stimatore di massima verosimiglianza \widehat{q}_p di q_p ;
- d. la distribuzione esatta di \widehat{q}_p .

QUESITO N. 2

1. Il candidato illustri brevemente le principali proprietà di ottimalità degli stimatori.
2. Sia dato un campione casuale semplice (X_1, \dots, X_n) da una popolazione X con distribuzione di probabilità $f(x; \theta) = \theta x^{\theta-1} I_{(0,1)}(x)$, con $\theta > 0$ parametro incognito.
 - a. Ricavare una statistica sufficiente per θ .
 - b. Identificare lo stimatore di massima verosimiglianza $\hat{\theta}$ di θ .
 - c. Definire la distribuzione approssimata di $\hat{\theta}$ per n grande.
 - d. Trovare la regione di rifiuto del test di Wald per il sistema di ipotesi $H_0: \theta = \theta_0$ vs. $H_1: \theta < \theta_0$.

QUESITO N. 3

Una banca ha classificato i propri clienti in tre classi di rating R_i , $i \in \{1,2,3\}$, a seconda della rischiosità: a tutti i clienti di ciascuna classe viene attribuita una stessa probabilità di insolvenza (ovvero, di non adempiere agli obblighi di restituzione dei prestiti). Dopo un anno la banca valuta l'efficacia del modello di classificazione. Ci si attende che, in ogni classe di rating R_i , la frequenza effettiva dei clienti insolventi e di quelli non insolventi non sia dissimile dalle corrispondenti frequenze attese.

Si sono ottenuti i dati della seguente tabella:

Classe di rating	Distribuzione Insolvenze attese	Clienti insolventi	Clienti non insolventi	Clienti totali
R ₁	\hat{n}_{11}	n_{11}	n_{12}	n_{10}
R ₂	\hat{n}_{21}	n_{21}	n_{22}	n_{20}
R ₃	\hat{n}_{31}	n_{31}	n_{32}	n_{30}
Totale	n_{01}	n_{01}	n_{02}	n_{00}

Il candidato:

1. proponga un test per valutare l'accostamento tra la distribuzione delle frequenze effettive e quella delle frequenze attese, esplicitando le ipotesi confrontate e fornendo l'espressione della statistica test per il campione descritto in tabella;
2. argomenti se il test di Kolmogorov - Smirnov possa o meno essere preso in considerazione per il problema in esame;
3. nel caso ci sia ragione di supporre che le prime due classi di rating non abbiano adeguato potere discriminante e che quindi i tassi di insolvenza \hat{f}_1 e \hat{f}_2 (stimati, rispettivamente, attraverso i rapporti $f_1 = n_{11}/n_{10}$, $f_2 = n_{21}/n_{20}$) siano uguali, indichi almeno un test statistico per verificare tale ipotesi;
4. ipotizzi che il numero di soggetti insolventi per le due classi si distribuisca come segue:
 $D_1 = X_1 + X_0$
 $D_2 = X_2 + X_0$
dove X_0 è una componente comune alle due classi e X_1 e X_2 sono specifiche di ciascuna classe. X_0 , X_1 e X_2 sono variabili casuali tra loro congiuntamente indipendenti. Si argomenti se D_1 e D_2 sono indipendenti, e se lo sono le variabili condizionate $D_1|X_0=x_0$ e $D_2|X_0=x_0$;
5. fornisca la distribuzione di probabilità di una delle variabili casuali D_1 oppure D_2 , sapendo che per ogni $i \in \{0,1,2\}$ la variabile casuale X_i ha funzione di densità $f(x; \theta_i) = \frac{\theta_i^x e^{-\theta_i}}{x!}$.

ECONOMETRIA

Un quesito a scelta tra i due proposti dalla Commissione

QUESITO N. 4

1. Il candidato enunci le ipotesi alla base di un modello di regressione lineare multipla:

$$Y = X\beta + \varepsilon$$

dove Y è la variabile dipendente, X è una matrice contenente k variabili esplicative (inclusa l'intercetta) ed ε è il termine di errore. Si mostri che lo stimatore OLS di β è non distorto e si ricavi la varianza dello stimatore (si assuma per semplicità il caso con le X non stocastiche).

2. Si consideri il seguente modello di regressione lineare:

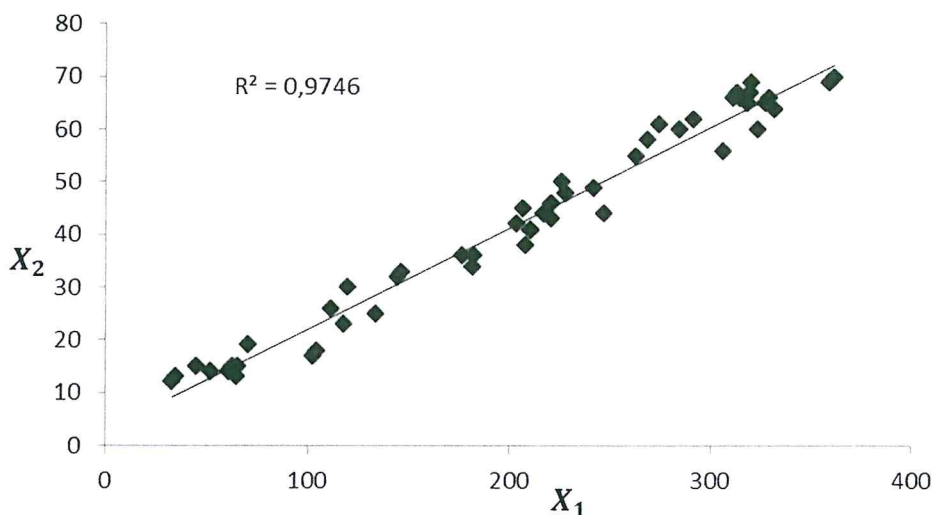


$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (1)$$

in cui Y è la variabile dipendente, X_1 ed X_2 sono due variabili quantitative ed ε è un termine di errore omoschedastico.

- Si supponga che $X_2 = c_1 + c_2 X_1$; il candidato indichi quali conseguenze si avrebbero sulla stima dei parametri.
- Si supponga, invece, che la figura sottostante rappresenti i valori di X_1 ed X_2 rilevati sulle n unità del campione. Il candidato spieghi se possono sorgere problemi di qualche tipo nella stima dei parametri del modello (1) ed eventualmente come potrebbe gestire il problema.

Figura 1



- Si consideri il seguente modello (2), analogo al modello (1), in cui si aggiunge un effetto interattivo fra le variabili X_1 ed X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon \quad (2)$$

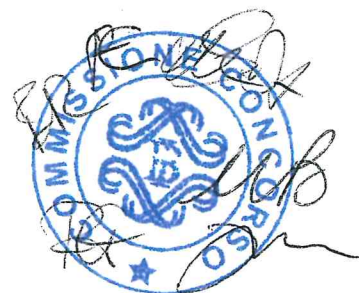
Il candidato spieghi il significato di effetto marginale in un modello di regressione lineare ed indichi come calcolerebbe l'effetto marginale di X_2 sulla variabile dipendente nel modello in questione.

- Il candidato supponga ora di voler stimare la probabilità che un'impresa sia stata in grado di ripagare il proprio debito utilizzando il modello:

$$P(Y = 1|X) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

dove Y è una variabile dicotomica che assume il valore 1 se l'impresa ha ripagato il proprio debito e 0 in caso contrario ed X è un vettore di variabili relative allo stato patrimoniale ed al fatturato dell'impresa. Il candidato espliciti che forma assume F nei casi dei modelli di scelta binari *logit* e *probit*.

- Il candidato fornisca la definizione di effetto marginale in un modello *logit* o *probit* di cui al punto precedente (senza esplicitare la forma della funzione di densità). Inoltre, sempre considerando il modello *logit* o *probit*, si supponga che fra le k variabili esplicative vi sia una variabile di tipo dicotomico X_d (che assume i valori 0 ed 1). Il candidato spieghi come definirebbe gli effetti marginali di X_d sulla variabile dipendente in questo caso.



QUESITO N. 5

1. Dato un modello di regressione lineare, il candidato illustri quali sono le conseguenze sulle stime dei parametri se viene meno l'ipotesi di esogeneità dei residui.
2. Si indichino quali sono le caratteristiche desiderabili di una variabile strumentale da utilizzare per affrontare la problematica descritta al punto precedente.
3. Si supponga di voler studiare l'effetto della partecipazione ad un corso di grammatica inglese sulla capacità degli studenti di comprendere testi in tale lingua. Per misurare la comprensione del testo, si sottopongono gli studenti ad un test, al termine del quale viene assegnato loro un voto. Si supponga che il voto al test sia un variabile quantitativa Y continua.

Per valutare l'efficacia del corso di grammatica inglese, si contattano i presidi di due scuole (A e B) che decidono di adottare i seguenti approcci:

- Il preside della scuola A seleziona casualmente un gruppo di studenti e li sottopone due volte al test, misurando la loro performance prima del corso di grammatica inglese ($t = 0$) e dopo aver fatto partecipare gli studenti al corso ($t = 1$).
- Il preside della scuola B forma due gruppi di alunni, scelti a caso, e sottopone entrambi i gruppi al test una prima volta al tempo $t = 0$. Successivamente, solo il primo dei due gruppi partecipa al corso di grammatica e, terminato il corso, entrambi i gruppi sono sottoposti nuovamente al test di comprensione.

Il candidato discuta punti di forza e di debolezza dei due approcci adottati, illustrando le eventuali ipotesi necessarie e indichi, a suo avviso, il metodo preferibile. Il candidato spieghi sinteticamente cosa potrebbe accadere se gli studenti partecipanti al corso fossero selezionati su base volontaria.

4. Il candidato indichi quale dei due approcci può essere utilizzato per stimare un modello "difference-in-difference". Ne illustri inoltre i principali punti di forza e fornisca una possibile specificazione per il caso in esame.
5. Si utilizzino i dati della tabella (tra le due seguenti) che consente la stima del modello *difference-in-difference* e si fornisca la stima dell'effetto della partecipazione al corso.

Scuola A

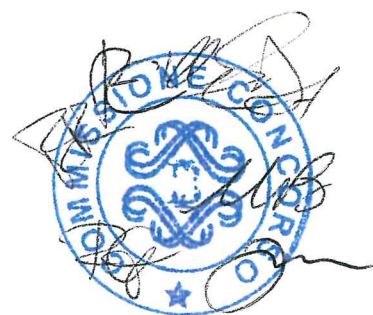
Punteggi medi degli studenti al test per la scuola A

Prima del corso ($t=0$)	Dopo il corso ($t=1$)
29	33

Scuola B

Punteggi medi degli studenti al test per la scuola B

	Prima del corso ($t=0$)	Dopo il corso ($t=1$)
Frequentanti il corso	28	34
Non frequentanti il corso	27	30



METODI DI CAMPIONAMENTO

Un quesito a scelta tra i due proposti dalla Commissione

QUESITO N. 6

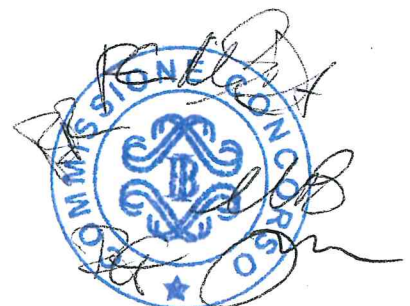
Si immagini di voler stimare la quota di individui dotati di titolo universitario da una popolazione finita tramite una indagine telefonica, selezionando il campione dagli elenchi dei numeri di telefono fissi.

1. Adottando un campionamento casuale semplice di n individui, corrispondenti ad altrettanti numeri di telefono, si derivi lo stimatore corretto della proporzione e il suo intervallo di confidenza.
2. Se si venisse a conoscenza, da fonti aggiuntive, che vi sia una relazione tra il titolo universitario e la zona di residenza (a cui corrisponde un prefisso telefonico), si illustri in che modo questa informazione potrebbe modificare il piano di campionamento scelto.
3. Si supponga che per ogni numero di telefono estratto vengano intervistati a grappolo tutti i componenti con oltre 25 anni all'interno della famiglia contattata. Si derivi lo stimatore della quota media di individui laureati e la sua varianza.
4. Rispetto alle strategie di campionamento casuale semplice, quali sarebbero i vantaggi e gli svantaggi del campionamento a grappolo? Si indichi e si discuta la formula dell'effetto del disegno.
5. Si discutano eventuali problemi di copertura derivanti dall'utilizzo delle liste dei telefoni fissi.

QUESITO N. 7

Si supponga di aver condotto una indagine campionaria per stimare il totale della spesa sostenuta dalle banche per i servizi sul territorio italiano utilizzando un piano di campionamento casuale semplice senza ripetizione. Il questionario rileva la spesa sostenuta per i servizi (Y), il numero di clienti (X) e altre caratteristiche di ciascuna banca.

1. Si derivi lo stimatore corretto del totale di Y e la sua varianza.
2. Ipotizzando una relazione di proporzionalità approssimata tra X e Y , dove X è una variabile ausiliaria nota sulla popolazione, come si potrebbe impiegare questa informazione per costruire uno stimatore per quoziente del totale di Y e quali sarebbero le proprietà di questo stimatore rispetto a quelle dello stimatore del punto precedente?
3. Si supponga invece che si sappia che la spesa in servizi e il numero di clienti siano legati da una relazione lineare $Y = \alpha + \beta X + \epsilon$ e che si conoscano i valori dei coefficienti della retta di regressione. Come si utilizzerebbe questa informazione per costruire lo stimatore per regressione del totale della spesa in Y ? Si derivi la formula della varianza di questo stimatore e si discuta in quali circostanze tale stimatore sarebbe più efficiente.
4. Come cambierebbero le proprietà di quest'ultimo stimatore se i coefficienti della retta di regressione non fossero noti, ma fosse necessario stimarli utilizzando i dati rilevati con l'indagine campionaria stessa, conoscendo la variabile ausiliaria X sulla popolazione?
5. Nel caso in cui alcune banche non riportino alcune informazioni richieste (mancate risposte parziali), si illustri brevemente quali strategie possono essere applicate e le conseguenze in termini di proprietà degli stimatori.



PROVA IN LINGUA INGLESE

Some governments today seek to monitor the general public's electronic communications (phone calls, texts and emails) saying that this reduces crime. Many people oppose this, saying it erodes individual freedom. Discuss both views and give your own conclusion to the discussion.

