ESTRATTO

CONCORSO PER L'ASSUNZIONE DI 3 COADIUTORI CON ORIENTAMENTO NELLE DISCIPLINE STATISTICHE

(Lett. H del bando del 9 novembre 2015)

Testo n.2

Statistica e probabilità

Due quesiti a scelta tra i tre proposti della Commissione

QUESITO N.1

1. Si consideri la variabile casuale X con valore atteso E(X) finito. Il candidato verifichi se le seguenti affermazioni sono vere o false motivando la risposta:
   - $E(|X|) \leq 1 + E(X^2)$
   - $Var\left[\left(X - E(X)\right)\frac{1}{E(X)}\right] = \frac{E(X^2)-[E(X)]^2}{[E(X)]^2}$
   - $Var\left[\left(X - E(X)\right)\frac{1}{X}\right] = \frac{[E(X)]^2}{E(X^2)-[E(X)]^2}$
   - Sia $E(X) < 0$ e $\theta \neq 0$ tale che $E\left(e^{\theta X}\right) = 1$ allora $\theta > 0$

2. Si ipotizzi ora per X una distribuzione Uniforme continua nell'intervallo $(0, \vartheta)$. Si calcolino media e varianza della distribuzione. Avendo estratto un campione bernoulliano $X_1, X_2, \dots X_n$ di n elementi si calcoli lo stimatore per ϑ con il metodo dei momenti, e lo si confronti con lo stimatore $T = X_{(1)} + X_{(n)}$ ricavato una volta ordinato il campione in modo non decrescente: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

QUESITO N.2

Si vuole analizzare la distribuzione del reddito di tre paesi A, B, C.

1. Si definisca il rapporto di concentrazione di Gini e si analizzi la sua relazione con la curva di Lorenz. Ipotizzando che l'indice di Gini assuma valore 0,5 nel paese A, si indichi quali considerazioni possono trarsi per il paese considerato avendo come termine di paragone la situazione italiana. Si ipotizzi che nel corso del tempo, a parità di reddito complessivo e di popolazione del paese, la quantità totale di reddito sopra una certa soglia resti invariata ma diminuisca il numero di persone che lo percepiscono, si discutano le conseguenze per l'indice di Gini.

2. Si ipotizzi che sempre nel paese A il reddito si distribuisca tra gli abitanti secondo la seguente distribuzione di Pareto:

$$f(x;\vartheta) = \begin{cases} \frac{\vartheta}{x^{\vartheta+1}} & per\ x \geq 1 \\ 0 & altrimenti \end{cases}$$

   Si calcoli per tale distribuzione il valor medio, la varianza e lo stimatore di massima

∞ [handwritten initials]

verosimiglianza per ϑ.

3. Si ipotizzi che nei paesi B e C il logaritmo del reddito si distribuisca secondo una distribuzione Normale con medie rispettivamente pari a μ_b e μ_c. Estraendo in modo indipendente in ciascun paese un campione di n elementi si indichi in che modo si può procedere per verificare l'ipotesi $\mu_b = \mu_c$.

QUESITO N.3

Sia X una variabile casuale con densità:

$$f(x;\theta,\lambda) = \lambda\theta^{\lambda}x^{-(\lambda+1)}$$

in cui $x \geq \theta$ e $\lambda > 0$ e $\theta > 0$ sono parametri incogniti. Il candidato:

1. fissato $\psi > \theta$ ricavi la distribuzione di X condizionata all'evento $\{X > \psi\}$
2. supponendo λ noto:
   a. mostri se la statistica $S = \min\{X_1,...,X_n\}$ è sufficiente per θ e valuti se risulti minimale;
   b. ricavi lo stimatore di θ con il metodo della massima verosimiglianza e stabilisca se è non distorto;
   c. ricavi la distribuzione di S e stabilisca se esiste uno stimatore non distorto per θ nella classe degli stimatori del tipo $c \cdot S$ dove c è una costante opportuna.

<u>**Econometria**</u>

Un quesito a scelta tra i due proposti della Commissione

QUESITO N.4

Si supponga di volere stimare la probabilità che un generico individuo sia proprietario di una abitazione attraverso il seguente modello:

$$P(Y = 1|X) = G(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)$$

dove Y è una variabile dicotomica che assume valore 0 (non possiede) o 1 (possiede), X è un vettore di variabili socio-demografiche e G è una funzione da definire. Il candidato:

1. discuta le problematiche che si avrebbero utilizzando un *linear probability model*;
2. illustri i vantaggi del modello *logit* come alternativa al precedente;
3. espliciti l'equazione di verosimiglianza per il modello *logit* (con condizioni di 1° e 2° ordine);
4. illustri uno stimatore per la varianza degli stimatori dei parametri;
5. supponga di disporre di un campione di n osservazioni e di avere stimato un modello *logit* con 5 regressori. Presenti un test per verificare l'ipotesi (β_2=β_3=0);
6. mostri che cosa accadrebbe alle stime dei parametri di un modello *logit* nell'ipotesi che il modello riesca a discriminare perfettamente fra possessori di una abitazione e non possessori per un determinato insieme di regressori.

QUESITO N.5

Si assuma che il processo stocastico $\{X_t\}$ segua un modello $AR(1)$ (autoregressivo di ordine 1) a media nulla, cioè che per ogni t si abbia:

$$X_t = \Phi X_{t-1} + e_t$$

con e_t *white noise* (con media nulla e varianza unitaria). Il candidato:

1. fornisca la rappresentazione MA(∞) del processo, cioè esprima, per ogni t, X_t come combinazione lineare (infinita) di variabili del processo *white noise*;
2. indichi sotto quali vincoli sui parametri il processo $\{X_t\}$ soddisfa le condizioni del teorema di Wald e fornisca la rappresentazione corrispondente;
3. assumendo che $\{X_t\}$ sia stazionario (in senso debole):
    - ricavi $\rho_k = Corr(X_t, X_{t-k})$ per $k = 1,2, \dots$;
    - supponga di trovarsi nell'istante t e indichi con $X_t(h) = \varphi(X_t, X_{t-1}, \dots)$ una opportuna funzione delle informazioni disponibili da utilizzare per prevedere come si manifesterà la variabile X_{t+h} riferita all'istante futuro $t + h$. Individui la funzione φ e quindi il previsore $X_t(h)$ tale per cui l'errore di previsione $e_t(h) = X_{t+h} - X_t(h)$ abbia media nulla e media quadratica più piccola possibile;
    - commenti il comportamento di $X_t(h)$ e di $Var(e_t(h))$ al crescere di h;
    - ricavi il limite per $h \to \infty$ di $X_t(h)$ e di $Var(e_t(h))$.
4. Assumendo ancora che $\{X_t\}$ sia stazionario in senso debole e che il processo *white noise* sia Gaussiano, cioè che ciascuna variabile e_t sia Normale:
    - discuta se il processo $\{X_t\}$ sia stazionario in senso forte;
    - mostri se il processo $\{X_t\}$ è un processo di Markov.

Metodi di campionamento

Un quesito a scelta tra i due proposti della Commissione

QUESITO N.6

Si vuole determinare la proporzione (p) di una popolazione finita che possiede un carattere qualitativo attraverso un'indagine campionaria. Il candidato:

1. illustri gli stimatori della proporzione e della relativa varianza nel caso del campionamento casuale semplice e costruisca un intervallo di confidenza per la proporzione stimata;
2. illustri, sempre nel caso del campionamento casuale semplice, la relazione tra l'errore di stima e la numerosità campionaria distinguendo tra il caso in cui non si disponga di alcuna informazione sulla proporzione da quello in cui sia noto che essa è pari almeno a 0,8;
3. illustri in che modo si modifica lo stimatore della proporzione nel caso si adotti un disegno campionario stratificato e si discutano i possibili guadagni di efficienza rispetto al campionamento casuale semplice;
4. supponga che per l'indagine sia stato selezionato un campione casuale semplice e che il tasso di risposta sia stato pari al 50 per cento: si indichino quali problemi si possono presentare per la stima di p;
5. ipotizzi che alcuni individui, pur partecipando all'indagine, non rispondano alla domanda sul carattere oggetto di rilevazione. Esponga sinteticamente una tecnica per imputare i dati mancanti,

discutendone le ipotesi alla base e motivandone la scelta. Illustri inoltre le accortezze da tenere in fase di stima della varianza della proporzione in presenza di dati imputati;

6. ipotizzi che il carattere oggetto di indagine riguardi un argomento sensibile e che gli intervistati potrebbero essere dunque reticenti a rispondere in modo sincero; si indichino possibili tecniche statistiche per condurre la rilevazione.

QUESITO N.7

Si immagini di dover realizzare una indagine per valutare il grado di conoscenza finanziaria degli studenti al termine del loro percorso di istruzione obbligatoria. La valutazione viene fatta sottoponendo un questionario ad un campione selezionato di studenti e calcolando per ciascuno studente la media dei punteggi ottenuti. Si supponga di disporre di una lista esaustiva delle scuole (con indicazione del numero totale di studenti per ogni scuola). Il candidato:

1. descriva un piano di campionamento in base al quale selezionerebbe il campione motivando le ragioni della propria scelta e calcoli la probabilità π_{ij} che lo studente i della scuola j ha di essere incluso nel campione, sulla base del piano di campionamento scelto;
2. ipotizzi di selezionare un campione casuale semplice di scuole e di includere nel campione tutti gli studenti di tali scuole (campionamento a grappoli ad uno stadio). Definisca il coefficiente di omogeneità nei grappoli (δ) e ne indichi il campo di variazione (per semplicità ipotizzi qui che le scuole abbiano lo stesso numero di studenti M);
3. supponga di aver selezionato un campione di 500 studenti attraverso il piano di campionamento descritto nel punto precedente e che l'effetto del disegno (*deff*) sia pari a 2. Dopo aver definito il *deff* indichi quale numerosità sarebbe stata necessaria per ottenere la stessa precisione della media campionaria nel caso in cui le osservazioni fossero state selezionate tramite un campionamento casuale semplice degli studenti;
4. supponga di voler fare un test di ipotesi sul punteggio medio conseguito dagli studenti; cosa accadrebbe se ignorasse la correlazione che esiste fra studenti della stessa scuola (ossia se si ipotizzasse che $\delta=0$)?
5. ipotizzi di avere a disposizione la lista degli studenti e di effettuare il seguente campionamento: prima si estrae un campione casuale semplice di n studenti, poi si identificano le scuole alle quali questi studenti appartengono ed infine si includono nel campione tutti gli studenti di quelle scuole. Discuta brevemente le implicazioni di questo tipo di campionamento sulle proprietà dello stimatore della media;
6. indichi con N_1 e N_2 il totale degli studenti maschi e femmine nella popolazione e con n_1 e n_2 i rispettivi totali nel campione selezionato. Sia inoltre Y_{ij} la variabile casuale che indica il punteggio ottenuto dallo studente i-esimo della scuola j-esima al test e $\bar{y}_U$ il valore medio nella popolazione. Supponendo di aver selezionato un campione casuale semplice senza ripetizione di n studenti e di conoscere i totali N_1 e N_2, indichi quale stimatore di $\bar{y}_U$ utilizzerebbe e discuta la relativa varianza rispetto al tradizionale stimatore di *Horvitz Thompson*.

<u>Prova in lingua inglese</u>

Homelessness is a serious issue in present-day society. Are we doing enough to fight it? In your opinion, what can we do to help homeless people?