



BANCA D'ITALIA
EUROSISTEMA

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

(Bando del 1° agosto 2019 – lettera D)

Testo n. 1



STATISTICA E PROBABILITÀ

Due quesiti a scelta tra tre proposti dalla Commissione

QUESITO N. 1

Sia X una variabile casuale discreta con supporto:

$$T \equiv \{a, a + 1, a + 2, \dots, b - 1, b\} \subset \mathbb{N}$$

e sia P la probabilità che X assuma un qualsiasi valore del supporto.

La variabile casuale X rappresenta il tempo di stacco (espresso in giorni) del dividendo delle azioni di un certo portafoglio. Secondo l'ipotesi H_0 la probabilità di stacco del dividendo P è uguale su tutte le scadenze. Dato un campione di n azioni, dopo un tempo di osservazione sufficientemente lungo, si osservano le frequenze di stacco del dividendo lungo lo scadenziario che va dal tempo a al tempo b :

$$(n_a, n_{a+1}, n_{a+2}, \dots, n_{b-1}, n_b)$$

1. Si proponga almeno una procedura di test per verificare che il tempo di stacco del dividendo si distribuisca secondo quanto ipotizzato (H_0); si scriva la statistica test e si enuncino sinteticamente i presupposti teorici della strategia proposta.
2. Supponendo che il test condotto al punto precedente abbia concluso in favore dell'ipotesi nulla (H_0), si derivi P e si calcoli il tempo medio \bar{T} dello stacco del dividendo per la distribuzione di X .
3. Assumendo che la variabile casuale X (tempo di stacco del dividendo) segua una distribuzione uniforme continua in $[a, b] \subset \mathbb{R}$, con $a = 0$ e b ignoto, si ricavi uno stimatore sufficiente per il parametro b .

I dipendenti della società XYZ percorrono in media 28 km al giorno per andare in ufficio, con una deviazione standard di 8 km.

4. Assumendo una distribuzione normale dei km percorsi, determinare la percentuale di dipendenti che percorre tra i 20 e i 36 km al giorno.
5. Il candidato esponga quali conclusioni potrebbe trarre in assenza di informazioni sulla distribuzione.

QUESITO N. 2

Una banca ha concesso prestiti a tre clienti C_1, C_2, C_3 e desidera stringere accordi di tipo assicurativo per cautelarsi contro l'eventualità di non ricevere i pagamenti dovuti (evento di insolvenza) da parte dei tre clienti.

Viene individuata un'assicurazione che propone una polizza che risarcisce la banca in caso di insolvenza anche di uno solo dei tre clienti.

Le tre variabili casuali indipendenti T_1 , T_2 e T_3 rappresentano la durata di vita dei contratti (fino all'insolvenza) dei tre clienti, con funzione di densità:

$$f_{T_i}(t; \lambda_i) = \lambda_i e^{-\lambda_i t}, i = 1, 2, 3$$

Il candidato:

1. ricavi la distribuzione di probabilità della durata di vita della polizza;
2. calcoli la durata di vita attesa della polizza;
3. assuma che la compagnia assicurativa abbia osservato sul mercato n tempi τ_i di durata di vita delle polizze: (τ_1, \dots, τ_n) ; sulla base di questo campione, determini lo stimatore di massima verosimiglianza e un intervallo di confidenza di livello α per il parametro della distribuzione.
4. Il candidato illustri le proprietà asintotiche degli stimatori.

QUESITO N. 3

Nell'azienda ABC lavorano 100 addetti, di cui 40 donne e 60 uomini. Terminato l'anno con risultati positivi, la Direzione decide di attribuire un bonus a un quinto degli addetti. Il candidato:

1. esponga sinteticamente alcuni indici di connessione e in particolare l'indice chi-quadrato di Pearson, illustrandone le modalità di costruzione;
2. supponendo che siano stati assegnati bonus a 16 uomini e a 4 donne, calcoli e commenti il valore assunto dal suddetto indice per misurare la connessione tra l'attribuzione del bonus e il genere. Sottoponga inoltre a verifica l'ipotesi nulla che non ci sia un legame fra questi due caratteri;
3. sapendo che un dipendente ha ricevuto un premio, calcoli la probabilità che sia un uomo.

Ipotizzando che la variabile casuale X si distribuisca come una binomiale negativa con parametri (k, θ) :

$$\Pr(X = x) = \begin{cases} \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k} & x = k, k+1, k+2, \dots \\ 0 & \text{altrove} \end{cases}$$

4. Il candidato calcoli la funzione generatrice dei momenti e il valore atteso.

ECONOMETRIA E STATISTICAL LEARNING

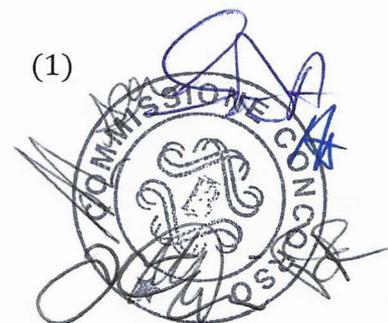
Un quesito a scelta tra due proposti dalla Commissione

QUESITO N. 4

Si consideri la seguente funzione di probabilità di tipo Poisson:

$$\Pr(Y_i = y_i | \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0, 1, 2, \dots \quad (1)$$

dove $\lambda_i > 0$.



- a) Si fornisca almeno un esempio di un evento che occorra in modo casuale nello spazio o nel tempo e che possa essere descritto da un modello di tipo Poisson come nella equazione (1).
- b) Si supponga di voler stimare il parametro λ_i attraverso un modello log-lineare del tipo $\log(\lambda_i) = x_i'\beta$, dove x_i indica un vettore di caratteristiche osservate. Il candidato spieghi se tale modello è stimabile con il metodo dei minimi quadrati ordinari (OLS).
- c) Si derivi la funzione di log-verosimiglianza per il modello dell'equazione (1) supponendo un modello log-lineare, ricavando anche le condizioni del primo e del secondo ordine.
- d) Si supponga di avere raccolto i dati sulle informazioni accademiche di 316 studenti. La variabile dipendente indica i giorni di assenza durante l'anno scolastico (**daysabs**); le variabili indipendenti sono: il punteggio standardizzato in matematica (**mathnce**), il punteggio standardizzato dei test linguistici (**langnce**) e il genere (**female**). Si commenti la tavola seguente, con particolare attenzione all'interpretazione dei coefficienti nello spiegare l'andamento della variabile **daysabs**.

Poisson regression		Number of obs =	316			
		LR chi2(3) =	175.27			
		Prob > chi2 =	0.0000			
Log likelihood = -1547.9709		Pseudo R2 =	0.0536			

daysabs	Coef.	Std. Err.	z	P > z	[95% Conf. Interval]	

mathnce	-.0035232	.0018213	-1.93	0.053	-.007093	.0000466
langnce	-.0121521	.0018348	-6.62	0.000	-.0157483	-.008559
female	.4009209	.0484122	8.28	0.000	.3060348	.495807
_cons	2.286745	.0699539	32.69	0.000	2.149638	2.423852

- e) Il candidato spieghi qual è il limite principale del modello di Poisson. Se i dati a disposizione contenessero più zeri di quanti il modello di Poisson riesca a predire, si proponga un modello alternativo per la stima di λ_i .

QUESITO N. 5

- a) Sia dato un modello di regressione lineare multipla:

$$Y = X\beta + \varepsilon \quad (1)$$

dove la Y è la variabile dipendente, X è una matrice contenente k variabili esplicative (inclusa l'intercetta) ed ε è il termine di errore.

Spiegare sinteticamente in cosa consiste il problema della multicollinearità. Si discutano, inoltre, gli effetti sulle stime dei parametri e sul modello e si descrivano i metodi che possono essere utilizzati per individuare tale problema.

- b) In presenza di eteroschedasticità, si assuma che per tutte le n unità la varianza dell'errore sia proporzionale al quadrato della variabile esplicativa x_k , ovvero:

$$\sigma_i^2 = \sigma^2 x_{ik}^2, \quad i = 1, 2, \dots, n$$

Il candidato proponga uno stimatore da utilizzare in questa circostanza e illustri come si modificherebbe il modello di regressione.



- c) Il candidato enunci le condizioni di stazionarietà in senso debole per un generico processo autoregressivo di ordine p $AR(p)$ e verifichi se il seguente processo è stazionario in senso debole, assumendo che ϵ_t sia un *white noise* a media nulla e varianza unitaria:

$$Y_t = Y_{t-1} - 0,5Y_{t-2} + \epsilon_t \quad (2)$$

- d) Il candidato definisca la funzione di risposta ad impulso per uno shock di grandezza unitaria al tempo t . Si utilizzi tale strumento per definire le caratteristiche del processo (2) analizzandolo per un numero opportuno di periodi.

METODI DI CAMPIONAMENTO

Un quesito a scelta tra due proposti dalla Commissione

QUESITO N. 6

Un'indagine prevede di intervistare sul telefono fisso un campione probabilistico di 1.000 individui. Le domande del questionario prevedono la rilevazione di comportamenti legati alla religione.

1. Si discutano sinteticamente i vantaggi dell'utilizzo di un'indagine campionaria rispetto a una rilevazione totale e gli eventuali problemi di stima derivanti dalla tecnica di rilevazione utilizzata.
2. Il disegno campionario prevede che, al momento del contatto telefonico, si intervisti solo un membro della famiglia scelto a caso. Si definisca il piano di campionamento adottato e si discutano le differenze nel caso in cui si intervistino tutti i membri della famiglia. In particolare si fornisca lo stimatore della media di una generica variabile quantitativa Y nei due casi e si discutano le differenze in termini di efficienza.
3. Si ipotizzi che l'indagine raccolga anche informazioni su reddito Y e consumi C e che si voglia individuare uno stimatore per la propensione al consumo $c = C/Y$. Proporre un metodo di stima della variabile c .

QUESITO N. 7

Un'azienda di spedizione di semi per corrispondenza vuole conoscere l'opinione dei propri clienti su una nuova varietà di prodotto. Decide quindi di svolgere un'indagine campionaria a partire dalla lista completa dei clienti, in cui sono elencati i nominativi, gli indirizzi e il totale dei semi venduti della nuova varietà.

1. Si descriva come potrebbe essere divisa la popolazione per applicare un campionamento stratificato. Quali sono le condizioni per applicare questo piano di campionamento e quali i suoi vantaggi?
2. Si definisca inoltre lo stimatore della percentuale di clienti soddisfatti in caso di allocazione proporzionale tra gli strati e la formula della sua varianza.
3. Si discuta in quali situazioni può essere opportuno allontanarsi dal criterio dell'allocazione proporzionale tra gli strati, sia in termini di efficienza sia in termini di costi.
4. Nel caso in cui l'indagine venga ripetuta nel tempo, si discuta quali vantaggi potrebbe comportare la reintervista di tutto il campione o di una sola parte di esso.

PROVA IN LINGUA INGLESE

“The world's most valuable resource is no longer oil, but data” (*The Economist*, May 2017). Discuss.

