



BANCA D'ITALIA
EUROSISTEMA

CONCORSO PER L'ASSUNZIONE DI 5 ESPERTI LAUREATI
CON ORIENTAMENTO NELLE DISCIPLINE STATISTICHE
(Bando 15/09/2020 – Lett. B)

Testo n. 4

STATISTICA E PROBABILITÀ

Due quesiti a scelta tra tre proposti dalla Commissione

QUESITO N. 1

La candidata/il candidato:

- 1 applichi il metodo dei momenti per la stima dei parametri θ_1 e θ_2 di una popolazione con distribuzione uniforme:

$$f(X; \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1}$$

- 2 stimi, applicando il metodo della massima verosimiglianza, il parametro θ nella funzione di densità:

$$f(X; \theta) = \frac{1}{2} e^{-|X - \theta|} \quad (-\infty \leq X, \theta \leq \infty)$$

- 3 illustri i seguenti metodi di stima puntuale dei parametri, evidenziandone le principali proprietà e le condizioni che devono essere rispettate affinché possano essere applicati: metodo dei momenti, metodi dei minimi quadrati e metodo della massima verosimiglianza.

QUESITO N. 2

1. Siano $X_1, \dots, X_i, \dots, X_n$ n variabili aleatorie indipendenti e identicamente distribuite secondo una distribuzione uniforme definita fra $(0, 1)$, X_i i.i.d. $\sim Unif(0, 1)$.
Si consideri la seguente successione di v.a.:

$$Y_n = (\prod_{i=1}^n X_i)^{1/n}$$

Individuare il limite in probabilità di Y_n applicando la legge dei grandi numeri.



2. Illustrare le differenze fra la teoria della probabilità classica, quella frequentista e l'impostazione soggettiva, soffermandosi sugli eventuali svantaggi degli approcci richiamati.
3. Immaginiamo di avere tre urne numerate da 1 a 3. Ciascuna urna contiene 100 palline. Ogni urna contiene un determinato numero di palline difettose: la numero 1 contiene una pallina difettosa, la numero 2 contiene due palline difettose e la numero 3 contiene tre palline difettose. Si selezionano casualmente una delle tre urne e si estraggono una pallina dall'urna selezionata. Qual è la probabilità di selezionare una pallina difettosa? Se avessimo già selezionato una pallina e notato che è difettosa, qual è la probabilità che questa sia stata estratta dall'urna numero 1?

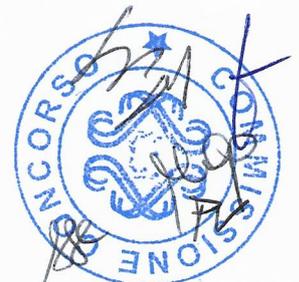
QUESITO N. 3

Su un campione di 1.000 intervistati sono state rilevate le seguenti variabili: titolo di studio dell'intervistato (variabile X), titolo di studio della madre (variabile Y) e reddito familiare annuo. Le tabelle 1 e 2 riportate sotto mostrano le frequenze di risposta, rispettivamente, degli 800 intervistati con reddito familiare più basso e dei restanti 200 intervistati con reddito familiare più elevato.

Tabella 1 (reddito familiare più basso)						Tabella 2 (reddito familiare più elevato)					
	Titolo di studio della madre						Titolo di studio della madre				
	Licenza elementare	Licenza media	Diploma	Laurea	Totale		Licenza elementare	Licenza media	Diploma	Laurea	Totale
Titolo di studio dell'intervistato						Titolo di studio dell'intervistato					
Licenza Elementare	53	30	18	8	109	Licenza elementare	0	0	0	0	0
Licenza Media	50	43	80	20	193	Licenza media	0	40	2	4	46
Diploma	81	80	131	46	338	Diploma	0	4	62	10	76
Laurea	42	35	66	17	160	Laurea	0	0	0	78	78
Totale	226	188	295	91	800	Totale	0	44	64	92	200

La candidata/il candidato:

1. illustri i concetti di connessione nulla e di dipendenza perfetta tra le due variabili X e Y. Sulla base di tali concetti, descriva quali conclusioni si possono trarre in merito alla relazione tra le tre variabili rilevate nell'indagine in base all'osservazione delle due tabelle riportate;
2. riporti, per le due tabelle illustrate sopra, le corrispondenti tabelle di connessione nulla e illustri i principi alla base degli indici di connessione (o contingenza);
3. descriva quali caratteristiche devono avere le distribuzioni doppie per essere definite di "massima omogeneità" e di "massima eterogeneità".



ECONOMETRIA E STATISTICAL LEARNING

Un quesito a scelta tra due proposti dalla Commissione

QUESITO N. 4

1. Si supponga di voler prevedere una risposta binaria sulla base di un insieme di predittori multipli attraverso il seguente modello di regressione lineare generalizzato:

$$\Pr(Y) = G(\beta_0 + \beta_1 X_1 + \dots + \beta_K X_K)$$

dove $X = (X_1, X_2, \dots, X_K)$ sono K predittori e Y la variabile risposta.

La candidata/il candidato:

- 1.a discuta l'importanza di scegliere una funzione *link* G nei modelli per dati binari;
1.b illustri le differenze tra il modello di probabilità lineare, il modello logistico e il modello probit.
2. A un campione di individui è stato chiesto se fossero favorevoli o non favorevoli all'introduzione di una certa politica sociale. Al fine di valutare la propensione dei cittadini verso la politica proposta, è stato stimato il seguente modello logistico:

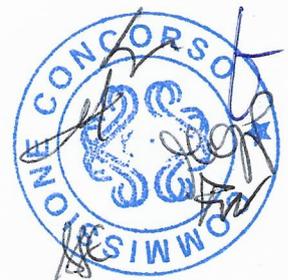
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.68413	0.03698	18.502	< 2e-16	***
rescale(income)	0.12453	0.05330	2.336	0.019615	*
male	-0.09739	0.02472	-3.940	8.58e-05	***
employ	0.10098	0.02643	3.820	0.000139	***
married	0.06564	0.02998	2.190	0.028724	*
college	-0.08486	0.02896	-2.930	0.003446	**
rescale(income):married	-0.13658	0.05928	-2.304	0.021384	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

dove:

- la variabile risposta rappresenta l'essere o meno favorevole all'introduzione della politica proposta e assume valore 1 se l'individuo si dichiara favorevole e 0 altrimenti;
- "rescale(income)" rappresenta il reddito disponibile familiare standardizzato rispetto alla media e a due volte la deviazione standard;
- "male" è una variabile binaria che assume valore 1 se l'individuo intervistato è uomo e 0 se è donna;



- “employ” è una variabile binaria che assume valore 1 se l’individuo intervistato ha una occupazione a tempo pieno e 0 altrimenti;
- “married” è una variabile binaria che assume valore 1 se l’individuo intervistato è sposato/convivente e 0 altrimenti;
- “college” è una variabile binaria che assume valore 1 se l’individuo dichiara di avere almeno una laurea triennale e 0 altrimenti;
- “rescale(income):married” rappresenta l’interazione tra reddito (standardizzato) e la condizione di sposato/convivente.

La candidata/il candidato:

- 2.a spieghi come stimare la probabilità che una donna occupata a tempo pieno, con reddito medio, sposata e laureata si dichiara favorevole alla politica proposta;
 - 2.b commenti l’effetto interazione, confrontando la propensione verso la politica degli individui sposati/conviventi rispetto agli individui non sposati/non conviventi al crescere del reddito.
3. La devianza residua del modello stimato è pari a 250, mentre la devianza del modello nullo (ossia senza predittori) è pari a 260. La candidata/il candidato discuta se l’introduzione dei predittori abbia effettivamente migliorato la stima del modello. Si illustrino i principali strumenti di diagnostica per valutare un modello logistico.

QUESITO N. 5

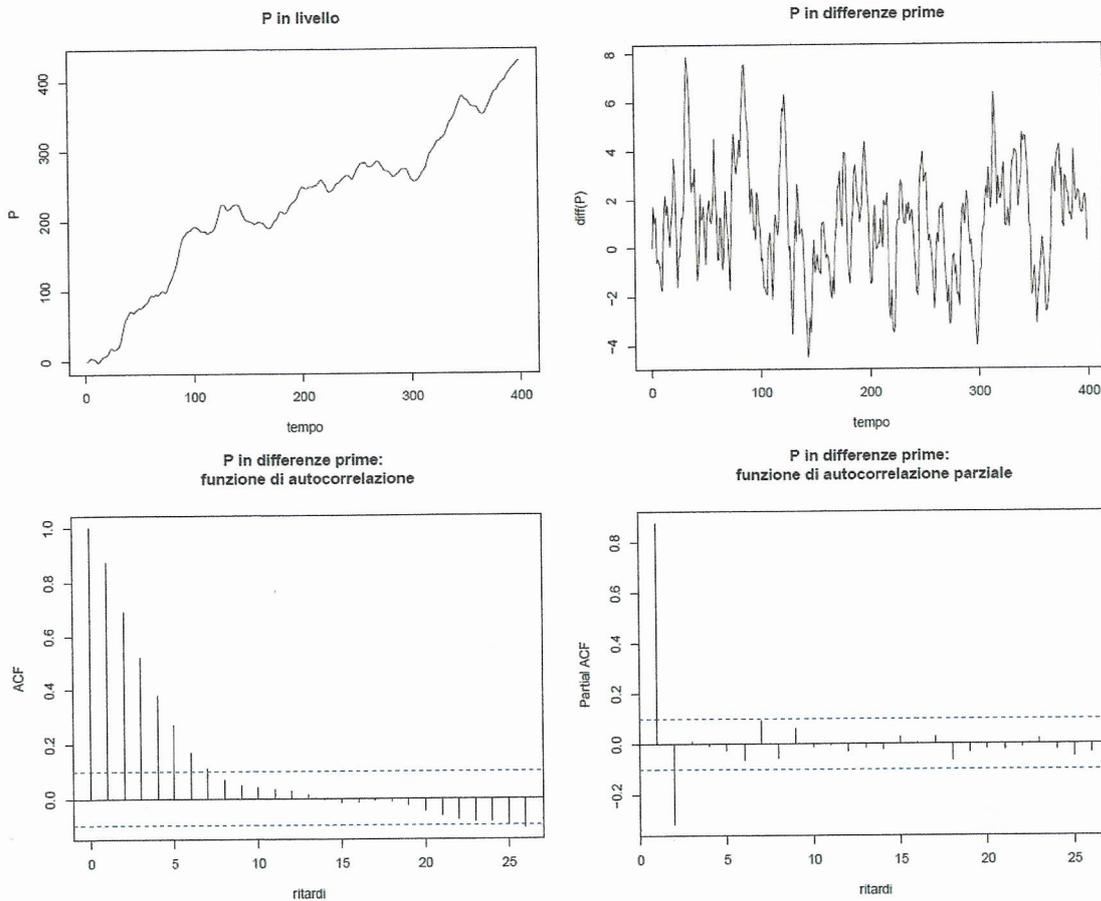
1. Si supponga che il processo stocastico Y_t segua il seguente modello:

$$Y_t = 0.2 + Y_{t-1} + a_t, \quad \text{dove } a_t \sim WN(0, \sigma^2 = 4) \quad \text{e} \quad Y_0 = 0$$

- 1.a Si scriva il processo nella sua rappresentazione MA e si calcolino la media e la varianza del processo.
- 1.b Che cosa rappresenta la costante 0.2? In un modello MA, invece, che significato assume la costante?



2. Si considerino i quattro grafici sottostanti. In senso orario si ha: l'andamento nel tempo della variabile P , l'andamento nel tempo della sua differenza prima ($\Delta(P)$), la funzione di autocorrelazione di $\Delta(P)$ e la funzione di autocorrelazione parziale di $\Delta(P)$.



2.a Osservando i quattro grafici, si discuta cosa si può intuire sul processo sottostante alla variabile P .

2.b Effettuando un test di Dickey-Fuller aumentato sulla variabile P , si ottiene un p -value pari a 0.7236. Cosa ne si evince? Si illustri sinteticamente la logica del test.

3. La candidata/il candidato scriva un esempio di processo stazionario attorno a un trend (*trend-stationary*) e indichi in cosa si differenzia principalmente dal modello del punto 1.



METODI DI CAMPIONAMENTO

Un quesito a scelta tra due proposti dalla Commissione

QUESITO N. 6

Si intende stimare il fatturato medio di 100 imprese ubicate in una zona industriale suddivisa in due aree. A tale scopo si esegue un campionamento stratificato con allocazione proporzionale di numerosità campionaria complessiva pari a 10. I risultati del campionamento sono riportati nella seguente tavola dove, per ciascuno strato h , N_h è l'ampiezza della popolazione, \bar{y}_h è la media campionaria del fatturato, e s_h^2 è la varianza campionaria del fatturato.

Strato	N_h	\bar{y}_h	s_h^2
Area A	60	5	4
Area B	40	4	3

1. Si calcoli l'allocazione proporzionale del campione e si stimino il fatturato medio e la sua varianza. Si calcoli, inoltre, l'effetto del disegno e si commentino i risultati.
2. Si discuta brevemente in cosa consiste il campionamento stratificato con allocazione di Neyman. Con riferimento ai risultati del punto precedente, si confrontino i due metodi di allocazione.
3. Si supponga che le imprese con fatturati più elevati riportino nell'indagine valori di fatturato mediamente più bassi di quelli effettivi. Si discutano le conseguenze dell'errore di misura sulla stima del fatturato medio.



QUESITO N. 7

Si supponga di voler stimare la proporzione di studenti che intendono iscriversi all'università al termine della scuola secondaria di secondo grado. La popolazione di studenti, di numerosità pari a 300, è suddivisa in 10 classi e ogni classe è composta dallo stesso numero di studenti. Si decide di selezionare tre classi attraverso un campionamento casuale semplice senza reintroduzione e di intervistare tutti gli studenti che compongono ciascuna classe. I risultati dell'indagine sono riassunti nella tavola che segue.

Classe	Totale campionario di studenti che intendono iscriversi all'università
1	15
2	8
3	22

1. Si stimi la proporzione di studenti che intendono iscriversi all'università e la sua varianza. Si calcoli, inoltre, l'effetto del disegno e si commentino i risultati.
- 2.a Si ricavi il coefficiente di correlazione intra-classe a partire dall'effetto del disegno campionario.
- 2.b Per l'indagine in corso si dispone di un budget totale pari a C . A parità di budget totale, il campionamento illustrato consente di ottenere una numerosità campionaria pari a 7 volte quella di un campionamento causale semplice. Volendo utilizzare il budget C interamente, quale disegno di campionamento è preferibile tra i due?
3. Si supponga di voler stimare la proporzione degli studenti con cittadinanza non italiana che intendono iscriversi all'università. Si illustrino brevemente le strategie che si possono adottare e gli eventuali problemi.

PROVA IN LINGUA INGLESE

The Covid-19 pandemic has caused an exponential increase in online shopping. What are the advantages and disadvantages of this development?

