

Metodi statistici per il Management

Nome Cognome Matricola Firma

1) La funzione di densità di probabilità di una variabile aleatoria X può assumere valori

- ☐ anche maggiori di 1
- ☐ solo compresi tra 0 e 1
- ☐ anche negativi
- ☐ minori di 100

2) Dati due eventi A e B , la probabilità di A condizionata a B , $P(A|B)$, è

- ☐ $P(A)$
- ☐ $P(A \cup B)/P(A)$
- ☐ $P(A \cap B)/P(B)$
- ☐ $P(A)/P(B)$

3) La distorsione (bias) di uno stimatore T di θ è

- ☐ $V(T) + [E(T) - \theta]^2$
- ☐ $[E(T) - \theta]^2$
- ☐ $[E(T) - \theta]$
- ☐ $V(T)$

4) Un intervallo di confidenza è

- ☐ un intervallo aleatorio che con una certa probabilità “copre” il vero valore del parametro
- ☐ un intervallo non aleatorio dove il parametro assume valori con una certa probabilità
- ☐ un intervallo che contiene con certezza il vero valore del parametro
- ☐ un intervallo dove con probabilità $1 - \alpha$ dovremmo trovare il vero valore del parametro

5) Un “p-value” pari a 0.001

- ☐ ci porta ad accettare l’ipotesi nulla
- ☐ ci porta a rifiutare l’ipotesi nulla
- ☐ non da alcuna indicazione per quanto riguarda l’accettazione o il rifiuto dell’ipotesi nulla
- ☐ ci porta a rifiutare l’ipotesi alternativa

6) Quali sono le ipotesi sulla distribuzione della variabile risposta nel modello logit?

7) Nel modello di regressione lineare, gli stimatori dei minimi quadrati sono

- ☐ distorti
- ☐ non corretti
- ☐ hanno un bias che dipende dal vero valore dei parametri
- ☐ non distorti

8) Nel modello logit, possiamo utilizzare delle variabili esplicative categoriche?

- ☐ no
- ☐ dipende dalla natura della variabile risposta y
- ☐ sì, codificandole opportunamente con variabili dummy
- ☐ sì, è sufficiente numerare le categorie da 1 a K

9) In un modello logit come interpretiamo l'esponentiale del coefficiente di regressione?

10) Nell'output 1, ricostruire il numero di osservazioni ed il valore della t mancante (motivare la risposta)

11) Nell'output 1, possiamo accettare l'ipotesi nulla che tutti i coefficienti di regressione sono simultaneamente uguali a zero? (motivare la risposta)

12) Qual è lo scopo della backward elimination?

13) Nell'output 2, quali variabili risultano non significative? (motivare la risposta)

14) Nell'output 4, possiamo accettare l'ipotesi nulla $\rho = 0$? (motivare la risposta)

15) Quale delle seguenti affermazioni è vera ? Le componenti principali sono

- ☐ tra loro incorrelate
- ☐ tutte della stessa varianza
- ☐ incorrelate con le variabili osservate
- ☐ a media 1

16) In base all'output 3, quale correlazione ci aspettiamo tra occ e sch?

- ☐ 0
- ☐ alta positiva
- ☐ alta negativa
- ☐ moderata negativa

17) In base all'output 3, quale variabile ha la comunalità minore?

- ☐ occ
- ☐ pop
- ☐ serv
- ☐ sch

18) Come sceglie i gruppi da "fondere" il metodo del legame completo?

19) Descrivere la relazione approssimata che intercorre tra DW e la correlazione di lag 1

20) In rif. all'output 4, possiamo affermare che

- ☐ se aumenta di una unità il reddito il consumo medio aumenta di .53
- ☐ nulla perchè la variabile è non significativa
- ☐ se aumenta di una unità il reddito il consumo medio diminuisce di .53
- ☐ se aumenta di una unità il reddito il consumo medio aumenta di .69

21) In rif. all'output 2, possiamo accettare l'ipotesi nulla che tutti i coef. di regressione siano zero?

- ☐ sì
- ☐ no, rifiutiamo
- ☐ non abbiamo sufficienti informazioni per rispondere
- ☐ no perché almeno un p-value è praticamente 0

22) In riferimento all'output 2, calcolare il valore di BIC.

23) In rif. all'output 4, possiamo accettare l'ipotesi nulla che l'intercetta sia 0? (motivare la risposta)

24) Quali sono le differenze tra il metodo del legame medio e singolo?

25) La curva ROC è costruita riportando i valori di

- ☐ sensitivity e specificity
- ☐ 1-sensitivity e specificity
- ☐ sensitivity e 1-specificity
- ☐ 1-sensitivity e 1-specificity

Output 1

Source	SS	df	MS			
Model	1018.65323	2	509.326616	Number of obs =		
Residual	982.314275	997	.985270085	F(2, 997) = 516.94		
Total	2000.96751	999	2.00297048	Prob > F = 0.0000		
				R-squared = 0.5091		
				Adj R-squared = 0.5081		
				Root MSE = .99261		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	-2.026722	3.193331		0.526	-8.293144	4.239699
x2	3.032726	3.191807	0.95	0.342	-3.230704	9.296157
_cons	1.015805	.031413	32.34	0.000	.9541615	1.077448

Output 2 (kwit=1 non è assunto dopo il periodo di prova)						
Logistic regression		Number of obs		=	683	
		LR chi2(3)		=	56.60	
		Prob > chi2		=	0.0000	
Log likelihood = -368.65709		Pseudo R2		=	0.0713	

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

sex		.4760307	.1860024	2.56	0.010	.1114728 .8405887
dex		-.094538	.0138168	-6.84	0.000	-.1216185 -.0674575
lex		-.00757	.078257	-0.10	0.923	-.160951 .145811
_cons		2.960625	1.103127	2.68	0.007	.7985371 5.122714

Output 3			
Variable	Fac(a)	Fac(b)	
pop	0.7557	-0.0100	
sch	0.9874	0.0100	
occ	0.0100	-0.9974	
serv	0.0100	0.8631	

Output 4						
ARIMA regression						
Sample: 1 to 100		Number of obs		=	100	
		Wald chi2(2)		=	156.01	
Log likelihood = 19.63483		Prob > chi2		=	0.0000	

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

consumption						
income	.5266249				.4202224	.6330274
_cons	.1521163				-.1003656	.4045982

ARMA						
ar						
L1.	.6925841	.0807578	8.58	0.000	.5343017	.8508664

/sigma	.1981852	.0178379	11.11	0.000	.1632235	.2331469
