

Metodi statistici per il Management

Nome Cognome Matricola Firma

Importante: nelle domande aperte motivare la risposta.

1) Nei test l'errore di primo tipo è

- ☐ non rifiutare l'ipotesi nulla quando è vera
- ☐ non rifiutare l'ipotesi nulla quando è falsa
- ☐ rifiutare l'ipotesi nulla quando è falsa
- ☐ rifiutare l'ipotesi nulla quando è vera

2) Un p -value minore di 0.0001

- ☐ ci porta a non rifiutare 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

3) Descrivere il modello AR(1) per il termine di errore w_i nel modello di regressione lineare per dati temporali.

4) Dato un campione dove le due variabili y e x sono incorrelate, il coefficiente di determinazione (R^2) della regressione di y su x risulta

- ☐ uguale a zero
- ☐ maggiore di zero
- ☐ minore di zero
- ☐ dipende dal campione

5) Come si distribuisce la variabile risposta in un modello logit?

6) output 2: quali variabili sono non significative al 9% ($\alpha = 0.09$)?

7) output 1: possiamo accettare l'ipotesi nulla che tutti i coefficienti di regressione sono simultaneamente uguali a zero?

8) output 2: ricostruire il valore mancante della statistica Wald Chi-square.

9) output 1: come interpreto il coef. di sqft (ampiezza appartamento in feet²)?

10) output 2: sapendo che dm indica l'ammontare del prestito in marchi, cosa possiamo concludere?

- ☐ all'aumentare dell'ammontare del prestito aumenta la probabilità di default
- ☐ la probabilità di default non dipende dall'ammontare del prestito
- ☐ all'aumentare dell'ammontare del prestito diminuisce la probabilità di default
- ☐ nulla perché la variabile è non significativa

11) output 2: definire e calcolare il valore di sensitivity

12) La specificity

- ☐ cresce al crescere del valore di cut-off
- ☐ decresce al crescere del valore di cut-off
- ☐ non dipende dal valore di cut-off
- ☐ è massima se il cut-off è pari a 0.5

13) Date 5 variabili estraiamo la prima componente principale ed osserviamo che ha correlazione pari a 0.5 con ognuna delle 5 variabili osservate. Cosa possiamo dedurre?

- ☐ la sua media è 5.00
- ☐ la sua varianza è 1.25
- ☐ la sua varianza è 2.50
- ☐ la sua media è 2.50

14) Nel grafico detto "cerchio delle correlazioni", rappresentiamo le

- ☐ variabili come punti con coordinate date dalle correlazioni con le componenti
- ☐ unità come punti con coordinate date dalle correlazioni con le componenti
- ☐ variabili come punti con coordinate date dalle correlazioni al quadrato con le componenti
- ☐ unità come punti con coordinate date dai punteggi sulle componenti

15) output 0: qual è la correlazione tra pop e sch?

- ☐ non si può dedurre dall'output
- ☐ quasi nulla
- ☐ negativa
- ☐ positiva

16) Come si decompone la devianza totale nella cluster analysis? Spiegare il significato dei termini.

17) Qual è la differenza tra il metodo di Ward ed il k -medie?

18) Nella cluster analysis come è definita e a cosa serve la pseudo T^2 ?

19) Il metodo di cluster analysis Ward è un metodo

- ☐ gerarchico dove ad ogni passo si minimizza l'incremento di devianza within
- ☐ non gerarchico dove si massimizza la devianza within
- ☐ non gerarchico dove si minimizza la devianza within
- ☐ gerarchico dove ad ogni passo si minimizza l'incremento di devianza between

20) Il criterio di informazione AIC è dato da

- ☐ $-2\ln(\max L)$
- ☐ $-2\ln(\max L) + (\text{numero parametri})$
- ☐ $-2\ln(\max L) + 2(\text{numero parametri})$
- ☐ $-2\ln(\max L) + \ln(n)(\text{numero parametri})$

21) Un valore del Durbin-Watson pari a 3.5 indica

- ☐ autocorrelazione di lag 1 degli errori negativa
- ☐ autocorrelazione di lag 1 degli errori positiva
- ☐ assenza di autocorrelazione degli errori
- ☐ il modello si adatta abbastanza bene ai dati

22) Come è costruita la curva ROC? A cosa serve?

Output 0

	Factor Pattern	
	Factor1	Factor2
pop	-0.00032	0.85257
sch	0.93475	0.00027
occ	0.39361	-0.75056
serv	0.95330	0.10856
house	0.95043	0.01692

Output 1 (price = prezzo di vendita di un appartamento in centinaia di dollari)

The REG Procedure
Dependent Variable: price

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5944174	1188835		<.0001
Error	23	1097071	47699		
Corrected Total	28				

Root MSE	218.40038	R-Square	
Dependent Mean	1142.58621	Adj R-Sq	0.8103
Coeff Var	19.11456		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-632.89944	252.94500	-2.50	0.0199
age	1	-6.87044	3.20660	-2.14	0.0430
feats	1	9.36038	30.08579	0.31	0.7585
ne	1	136.96799	103.34023	1.33	0.1981
cor	1	184.16942	108.01161	1.71	0.1016
sqft	1	1.10195	0.16124	6.83	<.0001

Output 2 (y = 1 se va in default)

The LOGISTIC Procedure
Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	639.114	523.024
SC	643.378	561.395
-2 Log L	637.114	505.024

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	132.0900	8	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.6450	0.4461	2.0905	0.1482
cuenta bad running	1	-0.4486	0.2524	3.1586	0.0755
cuenta good running	1	-2.0084	0.2946		<.0001
mes	1	0.0419	0.0109	14.7824	0.0001
ppag pre buen pagador	1	-1.1134	0.3643	9.3391	0.0022
uso privado	1	-0.7110	0.2340	9.2308	0.0024
dm	1	0.000043	0.000048	0.8085	0.3686
sexo hombre	1	0.00740	0.3277	0.0005	0.9820
estc no vive solo	1	-0.5599	0.3212	3.0394	0.0813

Classification Table

		Correct		Incorrect		Percentages			
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.700	24	361	9	131					