

Domande di autovalutazione

Capitolo 1

1) La probabilità di un evento è un numero

- ☐ reale qualsiasi
- ☐ compreso tra 0 e 1
- ☐ minore di zero
- ☐ maggiore di zero

2) La probabilità dell'unione di due eventi incompatibili, $P(A \cup B)$, è

- ☐ $P(A)+P(B)$
- ☐ $P(A)-P(B)$
- ☐ $P(A)P(B)$
- ☐ $P(A)/P(B)$

3) La probabilità condizionata $P(A|B)$ è definita come

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

4) La probabilità dell'intersezione di due eventi indipendenti, $P(A \cap B)$, è

- ☐ $P(A)+P(B)$
- ☐ $P(A)-P(B)$
- ☐ $P(A)P(B)$
- ☐ $P(A)/P(B)$

5) La funzione di ripartizione di una variabile aleatoria X è definita come

- ☐ $\Pr\{X \leq x\}$
- ☐ $\Pr\{X > x\}$
- ☐ $1 - \Pr\{X \leq x\}$
- ☐ $\Pr\{x < X \leq x\}$

6) La varianza di una variabile aleatoria è definita come

- ☐ il valore atteso della variabile aleatoria al quadrato
- ☐ il valore atteso degli scarti dalla media in valore assoluto
- ☐ il valore atteso degli scarti dalla media al quadrato
- ☐ il valore atteso degli scarti dalla media

7) Uno stimatore è

- ☐ una variabile aleatoria funzione dei dati campionari
- ☐ una variabile aleatoria indipendente dai dati campionari
- ☐ una variabile non aleatoria che fornisce informazioni sul parametro oggetto di stima
- ☐ un intervallo aleatorio che con probabilità $1-\alpha$ copre il vero valore del parametro

8) L'errore quadratico medio di uno stimatore T di θ è uguale a

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

9) Nei test l'errore di primo tipo è

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

10) Un "p-value" pari a 0.80

- ☐ ci porta ad accettare l'ipotesi nulla
- ☐ ci porta a rifiutare H_0
- ☐ non ci dà alcuna indicazione per quanto riguarda l'accettazione o il rifiuto dell'ipotesi nulla
- ☐ ci porta ad accettare l'ipotesi alternativa

11) Data la variabile aleatoria

X	-1	0	1
$P(x)$	0.3	0.4	?

completare la funzione di probabilità e calcolare il valore atteso

--

12) Dare una definizione concisa di intervallo di confidenza

--

Capitolo 2

1) Nel modello di regressione lineare omoschedastico il termine di errore

- ☐ ha media zero e varianza non costante
- ☐ ha distribuzione normale standard
- ☐ ha distribuzione normale con media zero e varianza costante
- ☐ ha distribuzione normale con media zero e varianza non costante

2) Quale relazione lega le tre quantità SSM, SSR e SST?

- ☐ SST = SSM + SSR
- ☐ SSM = SST + SSR
- ☐ SST = SSM - SSR
- ☐ SSR = SSM + SST

3) Gli stimatori dei minimi quadrati sono

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

4) Nel modello lineare, possiamo utilizzare delle variabili esplicative categoriche?

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

5) La funzione logistica è

- ☐ lineare e monotona
- ☐ non lineare e non monotona
- ☐ non lineare ma monotona
- ☐ lineare e non monotona

6) Date due variabili binarie y e x indipendenti, la stima del coefficiente di regressione logit risulta

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

7) A cosa serve il coefficiente di determinazione (R^2) ?

8) A cosa serve il test t?

Output 1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7551970	7551970		<.0001
Error	65	3302439	50807		
Corrected Total	66	10854409			

Root MSE	225.40352	R-Square	
Dependent Mean	1161.4627	Adj R-Sq	0.6911
Coeff Var	19.40687		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Intercept	1		159.89171	-4.75	<.0001	-1078.10963	-439.458
sqft	1	1.21372	0.09955			1.0149	1.41254

9) Dato l'output 1, qual è la stima della deviazione standard dell'errore?

10) Dato l'output 1, ricostruire il valore mancante della t

11) Dato l'output 1, ricostruire il valore mancante di R-squared

12) Dato l'output 1, ricostruire il valore mancante F-value

13) Dato l'output 1, ricostruire il valore mancante della stima puntuale dell'intercetta

14) Dato l'output 1, la variabile sqft è significativa?

15) Dato l'output 1, come interpreto il coefficiente di sqft?

16) In un modello logit come interpretiamo l'esponenziale del coefficiente di regressione?

17) In un modello logit ha senso assumere la variabile risposta come omoschedastica?

Capitolo 3

1) La distanza euclidea tra due unità può essere interpretata come

- ☐ una similarità
- ☐ una correlazione
- ☐ una dissimilarità
- ☐ una covarianza

2) Le componenti principali sono

- ☐ tra loro correlate
- ☐ tutte della stessa varianza
- ☐ a media zero
- ☐ a media non necessariamente nulla

3) Uno dei criteri per la scelta del numero di componenti seleziona quelle con varianza

- ☐ maggiore di 0.80
- ☐ pari a 1
- ☐ maggiore di 1
- ☐ minore di 1

4) La somma delle varianze delle componenti principali è

- ☐ maggiore di 0.80
- ☐ pari a K (numero di variabili originali)
- ☐ maggiore di 1
- ☐ minore di 1

5) La somma delle correlazioni al quadrato tra le variabili originali e la k -ma componente è

- ☐ maggiore di 0.80
- ☐ pari a K (numero di variabili originali)
- ☐ pari alla varianza della componente
- ☐ minore di 1

6) Il metodo delle K -medie è

- ☐ non gerarchico
- ☐ gerarchico
- ☐ sceglie automaticamente il numero di gruppi
- ☐ gerarchico o non gerarchico a seconda del tipo di applicazione

7) Nel modello di regressione lineare cosa distingue il test t da quello F?

--

8) Qual è lo scopo della backward elimination?

Output 1

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.87331359	1.0766535		
2	1.79666009	1.58182321	0.3593	
3	0.21483689	0.11490283	0.043	0.977
4	0.09993405	0.08467868	0.02	0.9969
5	0.01525537		0.0031	1

9) Dato l'output 1, ricostruire i valori mancanti

10) Dato l'output 1, scegliere il numero di componenti

11) Con quale modello/i possiamo analizzare la dipendenza di una variabile continua Y da una variabile categorica X ?

12) Qual è la differenza tra il metodo di Ward e quello del legame singolo?

13) Quali sono gli indici utilizzati per la scelta del numero di gruppi?

Capitolo 4

1) Nel modello AR(1) le osservazioni sono

- ☐ indipendenti
- ☐ incorrelate
- ☐ correlate
- ☐ dipendenti ma incorrelate

2) Se le osservazioni sono indipendenti, ci aspettiamo un valore del DW pari a

- ☐ zero
- ☐ uno
- ☐ due
- ☐ tre

3) Il criterio di informazione AIC serve per

- ☐ selezionare il modello
- ☐ stimare i parametri del modello
- ☐ costruire intervalli di confidenza
- ☐ verificare la bontà di adattamento del modello

4) Un valore del Durbin-Watson pari a 0.63 indica

- ☐ autocorrelazione di lag 1 dei residui negativa
- ☐ autocorrelazione di lag 1 dei residui positiva
- ☐ possiamo accettare l'ipotesi nulla di assenza di autocorrelazione degli errori
- ☐ il modello si adatta abbastanza bene ai dati

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

Output 1

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t	Variable Label
Intercept	1	0.152	0.123	1.23	0.2206	
income	1	0.527		9.78	<.0001	income
AR1	1	-0.693	0.077	-9.05	<.0001	

6) In riferimento all'output 1, calcolare lo standard error mancante

7) In riferimento all'output 1, qual è la stima del coefficiente di correlazione di lag 1?

- ☐ .527
- ☐ .152
- ☐ .693
- ☐ .198

Output 2

Maximum Likelihood Estimates			
SSE	3.928	DFE	97
MSE	0.040	Root MSE	0.201
SBC		AIC	
MAE	0.164	AICC	-33.020
MAPE	15.430	HQC	-30.107
Log Likelihood	19.635	Regress R-Square	0.496
Durbin-Watson	2.053	Total R-Square	0.599
		Observations	100

8) In riferimento all'output 2, calcolare i valori di AIC e BIC

9) La varianza dello stimatore media campionaria nel campionamento con ripetizione

- ☐ è maggiore di quella ottenuta con il campionamento senza ripetizione
- ☐ è minore di quella ottenuta con il campionamento senza ripetizione
- ☐ è $(1 - f)$ volte quella ottenuta con il campionamento senza ripetizione
- ☐ è la stessa di quella ottenuta con il campionamento senza ripetizione

10) La varianza dello stimatore media campionaria nel campionamento stratificato

- ☐ dipende in maniera diretta dalle varianze negli strati
- ☐ dipende in maniera inversa dalle varianze negli strati
- ☐ non dipende dalle varianze negli strati
- ☐ è minima se le medie negli strati sono tutte uguali

11) La varianza dello stimatore media campionaria nel campionamento stratificato

- ☐ aumenta se passiamo da una estrazione con ripetizione ad una senza
- ☐ diminuisce se passiamo da una estrazione con ripetizione ad una senza
- ☐ rimane invariata se passiamo da una estrazione con ripetizione ad una senza
- ☐ raddoppia se passiamo da una estrazione con ripetizione ad una senza

12) Descrivere lo stimatore della media nel caso di un campionamento casuale stratificato

Capitolo 5

1) Quali sono le due assunzioni distribuzionali generalmente adottate per il termine di errore del modello lineare di utilità?

2) In cosa differisce il modello logit dal probit?

Output 1: Logit model

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	204.382	8	<.0001
Score	190.987	8	<.0001
Wald	152.622	8	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	95% Confidence Limits	
Intercept	1	0.67	0.3199	4.387	0.0362	0.043	1.297
cuenta1	1	-0.6346	0.1764	12.943	0.0003	-0.9804	-0.2889
cuenta2	1	-1.9517	0.2061	89.710	<.0001	-2.3555	-1.5478
dm	1	0.000032	0.000033	0.945	0.3309	-0.00003	0.000098
estc1	1	-0.3854	0.2194	3.087	0.0789	-0.8154	0.0445
mes	1	0.035	0.00785	19.914	<.0001	0.0196	0.0504
ppag1	1	-0.9884	0.253	15.268	<.0001	-1.4841	-0.4926
sexo1	1	-0.2235	0.2208	1.025	0.3115	-0.6563	0.2093
uso1	1	-0.4744	0.1605	8.740	0.0031	-0.7889	-0.1599

3) In rif. all'output 1, possiamo accettare l'ipotesi nulla che tutti i coef. di regressione siano zero?

- ☐ sì
- ☐ no perchè il p-value corrispondente è praticamente zero
- ☐ non abbiamo sufficienti informazioni per rispondere
- ☐ no perché almeno un p-value è praticamente 0

4) In rif. all'output 1 e sapendo che `sexo1` indica la dummy uomo, possiamo concludere

- ☐ gli uomini hanno maggior probabilità di andare in default rispetto alle donne
- ☐ nulla perché la variabile è non significativa
- ☐ gli uomini hanno minor probabilità di andare in default rispetto alle donne
- ☐ gli uomini hanno la stessa probabilità di andare in default delle donne

5) In rif. all'output 1 e sapendo che `uso1` indica la dummy uso privato, possiamo concludere

- ☐ i prestiti per uso privato hanno maggior probabilità di andare in default rispetto a quelli per altro uso
- ☐ nulla perché la variabile è non significativa
- ☐ i prestiti per uso privato hanno minor probabilità di andare in default rispetto a quelli per altro uso
- ☐ i prestiti per uso privato hanno la stessa probabilità di andare in default di quelli per altro uso

Output2

Classification Table									
Prob Level	Correct		Incorrect		Percentages				
	Event	Non- Event	Event	Non- Event	Correct	Sensi- tivity	Speci- ficity	False POS	False NEG
0.5	113	632	68	187					

6) In riferimento all'output 2, calcolare i valori di sensitivity e specificity

7) In riferimento all'output 2, calcolare i valori di false positive e negative rate

8) In riferimento all'output 2, la percentuale di corretta classificazione è

- ☐ 34.43
- ☐ 74.50
- ☐ 22.03
- ☐ 91.00

9) La sensitivity

- ☐ dipende in maniera diretta dal valore di cut-off
- ☐ dipende in maniera inversa dal valore di cut-off
- ☐ non dipende dal valore di cut-off
- ☐ è massima se il cut-off è pari a 0.5

10) 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

11) La curva ROC è costruita utilizzando come cut off

- ☐ 0.5
- ☐ 1
- ☐ tutti i valori tra 0.5 e 1
- ☐ tutti i valori tra 0 e 1

12) Cosa indica l'area sotto la curva ROC?