Practice Examination

1. The weekly numbers of cars sold (y) and the daily average numbers of salespeople who work on a certain showroom floor (x) are given in the following table:

x	y
2	6
3	11
4	10
6	18
6	20

A simple linear regression model relating y to x with normal errors is fitted.

Which of the following gives the correct range for the p-value for testing the null hypothesis that the numbers of cars sold follow a common normal distribution against the alternative hypothesis that they follow the simple linear regression model?

(A) $(0, 0.005)$

(B) $(0.005, 0.01)$

(C) $(0.01, 0.02)$

(D) $(0.02, 0.05)$

(E) $(0.05, 0.1)$

2. You fit the linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ to 10 observed values (x_i, y_i).

You are given:

$$\begin{aligned} \sum (y_i - \hat{y}_i)^2 &= 2.79 \\ \sum (x_i - \bar{x})^2 &= 180 \\ \sum (y_i - \bar{y})^2 &= 152.40 \\ \bar{x} &= 6 \\ \bar{y} &= 7.78 \end{aligned}$$

Determine the width of the symmetric 95% prediction interval for y_* when $x_* = 8$.

(A) 1

(B) 2

(C) 3

(D) 4

(E) 5

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

3. Embryo (Ambrose's twin brother) is modeling monthly incurred dental claims. He has 12 monthly claims observations and three potential predictors:

   - Number of weekdays in the month
   - Number of weekend days in the month
   - Average number of insured members during the month

   Embryo obtained the following results from a linear regression:

	Coefficient	**Standard Error**
Intercept	−45,765,767.76	20,441,816.55
Number of weekdays	513,280.76	233,143.23
Number of weekend days	280,148.46	483,001.55
Average number of members	38.64	6.42

   Determine which of the following variables should be dropped, using a 5% significance level.

   I. Intercept
   II. Number of weekdays
   III. Number of weekend days
   IV. Average number of members

   (A) I and II only
   (B) I, II, and III only
   (C) I, II, and IV only
   (D) II, III, and IV only
   (E) I, II, III, and IV

Copyright © 2018 ACTEX Learning

4. For the multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$ with $i = 1, \ldots, 15$, you are given:

   (i) $(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1.00 & 0.25 & 0.25 \\ 0.25 & 0.50 & -0.25 \\ 0.25 & -0.25 & 2.00 \end{pmatrix}$

   (ii) $\hat{\beta}_0 = 10$ and $\hat{\beta}_1 = 12$

   (iii) The 98% symmetric confidence interval for β_2 is $(9.638, 20.362)$.

   Calculate the 99% symmetric prediction interval for y_* observed at $x_{*1} = 1.5$ and $x_{*2} = 4.5$.

   (A) $(79, 112)$

   (B) $(75, 116)$

   (C) $(71, 120)$

   (D) $(67, 124)$

   (E) $(63, 128)$

5. For a multiple linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, you are given:

   (i) The fitted regression function is $\hat{y} = 34.5 - 0.304 x_1 + 0.383 x_2$.

   (ii) The (incomplete) ANOVA table:

Source	Sum of Squares	*df*	Mean Square	F-ratio
Regression	270.09	?	?	?
Error	?	?	?	
Total	290.00	5		

   (iii) $(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 25.0487 & -0.8457 & 0.2864 \\ -0.8457 & 0.0294 & -0.0104 \\ 0.2864 & -0.0104 & 0.0040 \end{pmatrix}$

   Calculate the partial correlation between y and x_1.

   (A) -0.45

   (B) -0.37

   (C) -0.29

   (D) -0.21

   (E) There is not enough information to determine the answer

Copyright © 2018 ACTEX Learning

6. Interviews were conducted with 15 street vendors to study their annual incomes. Data were collected on annual income (y), age (x_1) and the number of hours worked per day (x_2). The following multiple linear regression model is suggested for the data:

$$\text{Model (1)}: \qquad y = \beta_0 + \beta_1 x_1 + \gamma_1 x_1^2 + \beta_2 x_2 + \varepsilon,$$

for some unknown parameters $\beta_0, \beta_1, \gamma_1, \beta_2$.

A number of alternative models are proposed in place of model (1) as follows:

(2) $y = \beta_0 + \beta_1 x_1 + \gamma_1 x_1^2 + \varepsilon$

(3) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

(4) $y = \beta_0 + \beta_1 x_1 + \varepsilon$

(5) $y = \beta_0 + \beta_2 x_2 + \varepsilon$

The following summarizes the residual sum of squares (RSS) obtained by fitting the above models:

Model	(1)	(2)	(3)	(4)	(5)
RSS	2,250,956	2,549,146	3,600,196	8,017,930	4,508,761

Calculate the F-statistic for testing the significance of age.

(A) 2.2

(B) 3.3

(C) 4.4

(D) 5.5

(E) 6.6

Copyright © 2018 ACTEX Learning

7. For a heteroscedastic simple linear regression model, you are given:

   (i) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, for $i = 1, 2, \ldots, 5$

   (ii) $\text{Var}(\varepsilon_i) \propto x_i^2$

   (iii)

i	x_i	y_i
1	1	1
2	2	-2
3	4	5
4	9	-10
5	16	25

   Calculate the weighted least squares estimate of β_0.

   (A) 0.15

   (B) 0.18

   (C) 0.20

   (D) 0.22

   (E) 0.25

8. You are given:

   (i) A linear regression model includes two explanatory variables: X_1 and X_2

   (ii) $R_{(j)}^2$ is the coefficient of determination obtained from regressing the j^{th} explanatory variable against the other explanatory variables

   (iii) $R_{(1)}^2 = 0.95$

   (iv) The threshold of the Variance Inflation Factor for variable j (VIF_j) for determining excessive collinearity is $\text{VIF}_j > 10$

   Determine which of the 2 variables under consideration will exceed the threshold established above.

   (A) None

   (B) X_1 only

   (C) X_2 only

   (D) X_1 and X_2

   (E) There is not enough information to determine the answer

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

9. You are given the following statements about different resampling methods:

   I. The validation set approach is a special case of k-fold cross-validation (CV).

   II. LOOCV has lower bias than k-fold CV when $k < n$.

   III. k-fold CV is less computationally expensive than LOOCV when $k < n$.

   Determine which of the above statements are correct.

   (A) I and II only

   (B) I and III only

   (C) II and III only

   (D) I, II, and III

   (E) The correct answer is not given by (A), (B), (C), or (D)

10. The following three multiple linear regression models have been fitted to the same 40 observations:

$$\begin{array}{ll} \text{Model I:} & y = \beta_0 + \beta_1 x_1 + \varepsilon \\ \text{Model II:} & y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \\ \text{Model III:} & y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \end{array}$$

    You are given:

    (i) The F-statistic for testing Model I against Model II is 30.

    (ii) The F-statistic for testing Model II against Model III is 12.

    (iii) The adjusted coefficient of determination of Model I is 0.5484.

    Calculate the adjusted coefficient of determination of Model III.

    (A) 0.75

    (B) 0.80

    (C) 0.85

    (D) 0.90

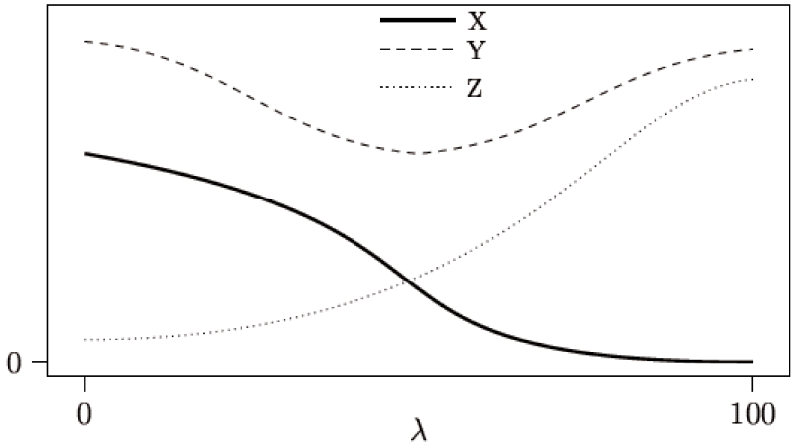
    (E) 0.95

Copyright © 2018 ACTEX Learning

11. You are estimating the coefficients of a linear regression model by minimizing the sum:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2.$$

From this model, you have produced the following plot of various statistics as a function of the parameter, λ:



Determine which of the following sets of statistics best matches the three curves.

	X	Y	Z
(A)	Squared bias	Test MSE	Training MSE
(B)	Squared bias	Variance	Training MSE
(C)	Variance	Squared bias	Training MSE
(D)	Squared bias	Test MSE	Variance
(E)	Variance	Test MSE	Squared bias

Copyright © 2018 ACTEX Learning

12. A multiple linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ is fitted, leading to the following table of parameter estimates:

Variable	Estimate	Standard Error
Intercept	0.6	0.15
x_1	-0.2	0.30
x_2	0.5	0.65
x_3	-1.4	0.45

Which of the following variables will be eliminated in the first step of the backward selection procedure?

(A) Intercept

(B) x_1

(C) x_2

(D) x_3

(E) None should be dropped from the model

13. You are given the following information about a GLM:

- The model uses four categorical explanatory variables:
  (a) x_1 is a categorical variables with three levels.
  (b) x_2, x_3 are categorical variables with two levels.
  (c) x_4 is a categorical variable with six levels.
- The model also uses a continuous explanatory variable x_5 modeled with a first order polynomial.
- There is only one interaction in the model, which is between x_1 and x_5.

Determine the maximum number of parameters in this model.

(A) Less than 13

(B) 13

(C) 14

(D) 15

(E) At least 16

Copyright © 2018 ACTEX Learning

Use the following information for Questions 14 and 15.

You are given the following GLM output:

Response variable	Pure Premium	
Response distribution	Gamma	
Link	log	
Scale parameter	1	
Parameter	**df**	$\hat{\beta}$
Intercept	1	4.78
Risk Group	2	
Group 1	0	0.00
Group 2	1	−0.20
Group 3	1	−0.35
Vehicle Symbol	1	
Symbol 1	0	0.00
Symbol 2	1	0.42

14. Calculate the predicted pure premium for an insured in Risk Group 3 with Vehicle Symbol 1.

(A) 0.23

(B) 84

(C) 127

(D) 7044

(E) 591253

15. Calculate the estimated variance of the pure premium for an insured in Risk Group 3 with Vehicle Symbol 1.

(A) 0.23

(B) 84

(C) 127

(D) 7044

(E) 591253

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

16. You are given the following table for model selection:

Model	Negative Loglikelihood	Number of Parameters	AIC
Intercept + Age	A	5	435
Intercept + Vehicle Body	196	11	414
Intercept + Age + Vehicle Value	196	X	446
Intercept + Age + Vehicle Body + Vehicle Value	B	Y	500

Calculate B.

(A) 211

(B) 212

(C) 213

(D) 214

(E) 215

17. Determine which of the following statements about GLMs with normal responses and identity link function is/are true.

I. A large deviance indicates a poor fit for a model.

II. Deviance cannot be used directly as a goodness of fit statistic.

III. The deviance residual is the same as the Pearson residual.

(A) I only

(B) II only

(C) III only

(D) All but III

(E) All

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

Use the following information for Questions 18 and 19.

18. You are given the following information for a logistic regression model to estimate the probability of a claim for a portfolio of independent policies:

- The model uses two explanatory variables:
  (a) Age group, which is treated as a continuous explanatory variable taking values of 1, 2 and 3, modeled with a second order polynomial
  (b) Sex, which is a categorical explanatory variable with two levels
- Observations:

	Sex					
	Male			**Female**		
	Age Group			**Age Group**		
Response	1	2	3	1	2	3
No Claim	20	28	30	24	28	22
Claim	8	7	3	16	13	1

- Parameter estimates:

Parameter	$\hat{\beta}$
Intercept	-1.1155
Sex	
Female	0.0000
Male	-0.4192
Age group	1.2167
$(\text{Age group})^2$	-0.5412

Calculate the estimated variance of the number of claims from the male policyholders belonging to age group 2 in the portfolio.

(A) 0.2

(B) 1.2

(C) 2.4

(D) 4.8

(E) 6.0

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

19. A policy is predicted to have a claim if the fitted probability of a claim is greater than 0.25.

Which of the following policies is/are predicted to have claims?

Policy	Sex	Age Group
I	Male	1
II	Male	2
III	Female	3

(A) I only
(B) II only
(C) I and II only
(D) I and III only
(E) II and III only

20. You are given:

- $y_1, y_2, \ldots, y_n$ are independent Poisson random variables with respective means μ_i for $i = 1, 2, \ldots, n$.
- A Poisson GLM was fitted to the data with a log link function:

$$\ln \mu_i = \beta_0 + \beta_1 x_i$$

where x_i refers to the value of the explanatory variable of the ith observation.

- Analysis of the data produced the following output:

x_i	y_i	$\hat{\mu}_i$	$y_i \log(y_i/\hat{\mu}_i)$
-1	2	?	??
-1	3	?	??
0	6	7.45163	-1.30004
0	7	7.45163	-0.43766
0	8	7.45163	0.56807
0	9	7.45163	1.69913
1	10	12.38693	-2.14057
1	12	12.38693	-0.38082
1	15	12.38693	2.87112

Calculate the deviance of the model.

(A) 0.4
(B) 0.9
(C) 1.4
(D) 1.9
(E) There is not enough information to determine the answer.

Copyright © 2018 ACTEX Learning

21. You are given the following sample of size 6 from a time series:

$$1 \qquad 1.5 \qquad 1.6 \qquad 1.4 \qquad 1.5 \qquad 1.7$$

Calculate the sample lag-3 autocorrelation.

(A) -0.25
(B) -0.04
(C) -0.03
(D) 0.21
(E) 0.25

22. You are given:

(i) The random walk model

$$y_t = y_0 + c_1 + c_2 + \cdots + c_t$$

where c_t, $t = 1, 2, \ldots, 8$ denote observations from a Gaussian white noise process.

(ii) The following eight observed values of y_t:

t	1	2	3	4	5	6	7	8
y_t	2	-1	4	7	11	13	17	16

(iii) $y_0 = 0$

(iv) The 7-step ahead forecast of y_{15}, $\hat{y}_{15}$, is determined based on the observed value of y_8.

Determine the 95% symmetric prediction interval for y_{15}.

(A) $(20, 40)$
(B) $(18, 42)$
(C) $(16, 44)$
(D) $(14, 46)$
(E) $(12, 48)$

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

23. You are performing out-of-sample validation for exponential smoothed forecasts with $w = 0.8$ and $\hat{s}_0 = 25$. The validation sample is:

t	y_t
1	20
2	30
3	60
4	40
5	15

Calculate the mean absolute percentage error.

(A) 0.30

(B) 0.35

(C) 0.40

(D) 0.45

(E) 0.50

24. The following table gives the prices of a stock over a 5-year period:

Year	Stock Price
1	11.0
2	10.0
3	9.0
4	10.5
5	9.5

You have fitted a first-order autoregressive model of the form

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \qquad t = 1, 2, \ldots, 5.$$

The parameters of the model are estimated by the method of conditional least squares.

Calculate the mean square error.

(A) 0.18

(B) 0.36

(C) 0.54

(D) 0.72

(E) 0.90

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

25. For a stationary first-order autoregressive process $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$, you are given:

(i) Based on the observed series $\{y_{2001}, y_{2002}, \ldots, y_{2018}\}$, the estimated parameters are

$$\hat{\beta}_0 = 0.75, \qquad \hat{\beta}_1 = -0.6, \qquad s^2 = 0.5.$$

(iii) $y_{2018} = 10$

Determine the symmetric 95% forecast interval for y_{2020}.

(A) $(2.5, 5.3)$

(B) $(2.4, 5.4)$

(C) $(2.3, 5.5)$

(D) $(2.2, 5.6)$

(E) $(2.1, 5.7)$

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

26. Consider a test and a training data set, denoted by $\mathcal{D}_{test}$ and $\mathcal{D}_{training}$. A decision tree is fitted on the training data set, we denote this decision tree by $\hat{f}$. We also construct a bagged tree $\hat{f}^{bag}$ and a random forest $\hat{f}^{randForest}$, with $B = 500$. The test error of the single decision tree, the bagged decision tree and the random forest are denoted by Error $\left(\hat{f}\right)$, Error $\left(\hat{f}^{bag}\right)$ and Error $\left(\hat{f}^{randForest}\right)$, respectively.

Consider the following statements:

I Error $\left(\hat{f}^{bag}\right) \leq$ Error $\left(\hat{f}\right)$.

II Error $\left(\hat{f}^{randForest}\right) \leq$ Error $\left(\hat{f}\right)$.

III Error $\left(\hat{f}^{bag}\right) \leq$ Error $\left(\hat{f}^{randForest}\right)$.

Which inequalities would you expect to hold true?

(A) I and II.

(B) II and III.

(C) I and III.

(D) I,II and III.

(E) None of the inequalities.

27. Consider a classification tree where the response variable can only belong to two possible classes. Consider the following statements.

I. A pure node has a misclassification error of zero.

II. A pure node has a misclassification error of one.

III. The misclassification error is always bounded from above by the Gini Index.

IV. The misclassification error is always bounded from below by the Gini Index.

Which of these statements is correct?

(A) I and III are correct.

(B) I and IV are correct.

(C) II and III are correct.

(D) II and IV are correct.

(E) None of these statements is correct.

Copyright © 2018 ACTEX Learning

28. John has been asked to develop a predictive model for claims from automobile insurance. He starts with a toy model which utilizes observations of two descriptive features, namely, driver's age (age) and vehicle market value (veh). He decides to use the principal components analysis to identify two key variables (principal components). John's calculation shows that the first principal component is given by the formula

$$Z_1 = 0.489 \times (\texttt{age} - \overline{\texttt{age}}) + 0.872 \times (\texttt{veh} - \overline{\texttt{veh}}).$$

For the 5-th driver, the age variable $\texttt{age}_5 = 34.16$, the vehicle market value variable $\texttt{veh}_5 = 13.51$ and means of the two variables are given by $\overline{\texttt{age}} = 35.04$ and $\overline{\texttt{veh}} = 13.18$. Denote the first and the second principal component scores by z_{51} and z_{52} respectively, and their loading vectors by $(\phi_{11}, \phi_{21})^\top$ and $(\phi_{12}, \phi_{22})^\top$ respectively.

Determine which of the following statements is incorrect.

(A) The correlation coefficient of the first principal component and the second principal component is zero.

(B) The mean of the second principal component is zero.

(C) The first principal component score is given by $z_{51} = -0.147$.

(D) The second principal component score is given by $z_{52} = -0.937$.

(E) The two principal component loading vectors provide an approximation of the fifth observation of $\texttt{age}_5$, i.e.

$$\texttt{age}_5 \approx z_{51}\phi_{11} + z_{52}\phi_{12}.$$

29. Consider a cluster C with 4 observations. The observations are given by

$$\begin{aligned} x_1 &= (2,3,4,5), \\ x_2 &= (1,2,3,4), \\ x_3 &= (1,2,1,2), \\ x_4 &= (3,2,3,2). \end{aligned}$$

Determine the average squared Euclidean distance $W(C)$ defined by

$$W(C) = \frac{1}{|C|} \sum_{i,i' \in C} \sum_{j=1}^{4} (x_{ij} - x_{i'j})^2 .$$

(A) 40.

(B) 30.

(C) 20.

(D) 10.

(E) 5.

Copyright © 2018 ACTEX Learning

30. Assume K-means clustering is applied to group a set of observations in five different clusters. The clusters are denoted by C_1, C_2, C_3, C_4 and C_5. The data points are given by:

$$\begin{aligned} x_1 &= (1,1,1) \\ x_2 &= (1,2,1) \\ x_3 &= (2,1,2) \\ x_4 &= (3,3,3) \\ x_5 &= (2,3,2) \\ x_6 &= (3,2,3) \\ x_7 &= (4,4,4) \\ x_8 &= (6,6,6) \\ x_9 &= (7,7,7) \\ x_{10} &= (2,2,2) \\ x_{11} &= (2,4,2). \end{aligned}$$

Assume at a certain step in the K-means clustering algorithm, the clusters contain the following observations:

$$\begin{aligned} C_1 &= \{1,2,3\} \\ C_2 &= \{4,5,6\} \\ C_3 &= \{7,8\} \\ C_4 &= \{9\} \\ C_5 &= \{10,11\}. \end{aligned}$$

Determine to which cluster observation x_{11} will be assigned in the next step of the algorithm.

(A) C_1.

(B) C_2.

(C) C_3.

(D) C_4.

(E) C_5.

Copyright © 2018 ACTEX Learning

31. Consider the following statements about $K-$means clustering and hierarchical clustering.

    I. If you know how many clusters there are, $K-$means clustering is always preferred over hierarchical clustering.

    II. Consider applying K-means clustering with $K = 5$ and 5 observations. Then each cluster contains only one observation.

    III. In one step of the $K-$means clustering algorithm, only one observation can change to another cluster.

    Which of the following is correct?

    (A) Only I is correct.

    (B) Only II is correct.

    (C) Only III is correct.

    (D) Only I and II are correct.

    (E) Only II and III are correct.

32. Consider a data set with n observations, where $n > 10$. Hierarchical clustering is applied. One then considers the solution with $n - 3$ clusters. How many elements does each cluster have at most?

    (A) 1

    (B) 2

    (C) 3

    (D) 4

    (E) more than 4

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

33. Consider a life insurer having data about policyholders purchasing life insurance. There are five explanatory variables, which are all assumed to be categorical with two categories. A classification tree is used to predict if someone will buy Life Insurance using the five categorical variables Salary, Sex, Age, Location, Children. Determine the variable which will be used in the first split, when the Gini index is used to determine the optimal split.

Salary	Sex	Age	Location	Children	Purchase Life Insurace
Low	M	Young	Urban	No	Yes
Low	M	Old	Rural	No	No
Low	F	Young	Urban	Yes	No
Low	M	Old	Rural	Yes	No
Low	F	Old	Rural	No	No
High	F	Young	Rural	Yes	Yes

(A) Income.

(B) Sex.

(C) Age.

(D) Location.

(E) Children.

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

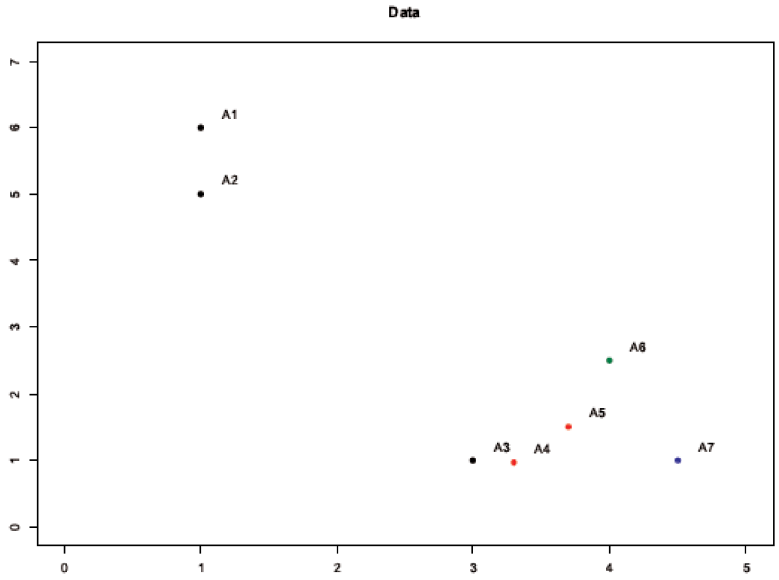


Figure 11.5.2: Data set

34. Consider seven data points, denoted by $A_1, A_2, \ldots, A_7$. Hierarchical clustering with single linkage is used to determine the clusters. If we choose four clusters, denoted by C_1, C_2, C_3, C_4, the clusters are given by:

$$\begin{aligned} C_1 &= \{A_1, A_2, A_3\}, \\ C_2 &= \{A_4, A_5\}, \\ C_3 &= \{A_6\}, \\ C_4 &= \{A_7\}. \end{aligned}$$

The data is shown in Figure 11.5.2, where the clusters are represented by different colors. In order to obtain the solution for three clusters, the hierarchical clustering algorithm then merges two clusters into one cluster. If we use single linkage to measure dissimilarity between clusters, which two clusters will be merged?

(A) C_1 and C_2.

(B) C_1 and C_3.

(C) C_1 and C_4.

(D) C_2 and C_3.

(E) C_1 and C_4.

Copyright © 2018 ACTEX Learning

35. Consider the following statements:

I. Scaling the variables has no effect on the results of the Principal Component Analysis.

II. Each principal component loading vector is unique.

III. If the number of principal components is one less than the number of observations, then the representation of observed data in terms of principal components is exact.

IV. The number of principal components can be used as a tuning parameter to be selected via cross-validation in an unsupervised analysis.

Determine which of the statements are correct.

(A) Statements II and III only

(B) Statement III only

(C) Statements II and IV only

(D) Statements I, II and III only

(E) Statements I, II, III and IV are all correct

****END OF EXAMINATION****

Copyright © 2018 ACTEX Learning

Solutions to Practice Examination

Answer Key

Question #	Answer
1	C
2	C
3	B
4	D
5	B
6	D
7	E
8	D
9	C
10	B
11	E
12	B
13	B
14	B
15	D
16	C
17	E
18	E
19	A
20	D

Question #	Answer
21	D
22	C
23	E
24	C
25	E
26	A
27	C
28	E
29	C
30	E
31	D
32	D
33	A
34	A
35	B

1. **(Performing a two-sided t-test given raw data in an SLR setting)**

Solution. Using a financial calculator, we can find the LSEs $\hat{\beta}_0 = -0.125$ and $\hat{\beta}_1 = 3.125$. Moreover,

$$\begin{aligned} \text{TSS} &= \sum_{i=1}^{5} y_i^2 - 5\bar{y}^2 = 981 - 5(13)^2 = 136, \\ \text{Reg SS} &= \hat{\beta}_1^2 S_{xx} = 3.125^2(101 - 5 \times 4.2^2) = 125, \\ \text{RSS} &= 136 - 125 = 11. \end{aligned}$$

The fact that the y_i's follow a common normal distribution is identical to the fact that $\text{H}_0 : \beta_1 = 0$ (i.e., i.i.d. model). The F-statistic for testing this hypothesis is

$$F = \frac{\text{Reg SS}/1}{\text{RSS}/(n-2)} = \frac{125/1}{11/(5-2)} = 34.0909.$$

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

Because $\hat{\beta}_1 > 0$, the t-statistic for testing H_0 against $H_a : \beta_1 \neq 0$ is the positive square root of the F-statistic, or $t(\hat{\beta}_1) = \sqrt{34.0909} = 5.8387$. Under H_0, the t-statistic has a $t_{5-2} \equiv t_3$ distribution. As $t_{3,0.01} = 4.5407$ and $t_{3,0.005} = 5.8409$, the p-value of the test is between $2(0.005) = 0.01$ and $2(0.01) = 0.02$. **(Answer: (C))** □

Remark. Because we are not provided with the F-quantiles in the SRM exam, we have no choice but to use the t-statistic as the test statistic and look up the t-table.

2. **(Construction of a prediction interval, SLR setting)**

Solution. The MSE is

$$s^2 = \frac{\text{RSS}}{n-2} = \frac{2.79}{10-2} = 0.34875.$$

The width of the 95% prediction interval for y when $x = 8$ is

$$\begin{aligned} 2t_{8,0.025}\sqrt{s^2\left[1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}\right]} &= 2(2.3060)\sqrt{0.34875\left[1+\frac{1}{10}+\frac{(8-6)^2}{180}\right]} \\ &= \boxed{2.8853}. \quad \textbf{(Answer: (C))} \end{aligned}$$

□

3. **(Dropping insignificant variables using t-tests)**

Comments: This is a modification of SRM Sample Question 27.

Solution. The t-statistics for testing the significance of the four variables are:

	Coefficient	**Standard Error**	***t* Stat**
Intercept	−45,765,767.76	20,441,816.55	$\boxed{-2.2388}$
Number of weekdays	513,280.76	233,143.23	$\boxed{2.2016}$
Number of weekend days	280,148.46	483,001.55	$\boxed{0.5800}$
Average number of members	38.64	6.42	6.0187

At the 1% significance level, a variable should be dropped if its t-statistic is smaller than $t_{12-4,0.025} = t_{8,0.025} = 2.3060$ in absolute value. All except the average number of members satisfies this criterion, so the intercept, number of weekdays, and number of weekend days should all be dropped. **(Answer: (B))** □

4. **(Given a CI, find a PI)**

Copyright © 2018 ACTEX Learning

Solution. From the confidence interval for β_2, we deduce that $\hat{\beta}_2 = (9.638+20.362)/2 = 15$ and that

$$t_{12,0.01}\sqrt{2s^2} = 2.6810\sqrt{2s^2} = 20.362 - 15,$$

giving $s^2 = 2$. Then the 99% symmetric prediction interval for y_* observed at $x_{*1} = 1.5$ and $x_{*2} = 4.5$ is

$$\begin{aligned}
& (\hat{\beta}_0 + 1.5\hat{\beta}_1 + 4.5\hat{\beta}_2) \pm t_{12,0.005}\sqrt{s^2[1 + \mathbf{x}_*'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_*]} \\
= \; & [10 + 1.5(12) + 4.5(15)] \\
& \pm 3.0545\sqrt{2\left[1 + \underbrace{\begin{pmatrix} 1 & 1.5 & 4.5 \end{pmatrix}\begin{pmatrix} 1.00 & 0.25 & 0.25 \\ 0.25 & 0.50 & -0.25 \\ 0.25 & -0.25 & 2.00 \end{pmatrix}\begin{pmatrix} 1 \\ 1.5 \\ 4.5 \end{pmatrix}}_{42.25}\right]} \\
= \; & 95.5 \pm 28.4085 \\
= \; & \boxed{(67.09, 123.91)}. \qquad \textbf{(Answer: (D))}
\end{aligned}$$

□

5. **(Calculation of partial correlation given summarized model output)**

Solution. The MSE is $s^2 = (290 - 270.09)/(5-2) = 6.636667$. The t-statistic for testing $\mathrm{H}_0 : \beta_1 = 0$ is

$$t(\hat{\beta}_1) = \frac{-0.304 - 0}{\sqrt{6.636667(0.0294)}} = -0.6882.$$

By (2.2.1), the partial correlation between y and x_1 is

$$\begin{aligned}
r(y, x_1|x_2) &= \frac{t(\hat{\beta}_1)}{\sqrt{t(\hat{\beta}_1)^2 + \text{df of RSS}}} \\
&= \frac{-0.6882}{\sqrt{(-0.6882)^2 + 3}} = \boxed{-0.3693}. \quad \textbf{(Answer: (B))}
\end{aligned}$$

□

6. **(Generalized F-test)**

Solution. We should use Model (5), which involves no x_1, to test $\mathrm{H}_0 : \beta_1 = \gamma_1 = 0$ (Model (5)) against H_a : β_1 and γ_1 not both zero (Model 1) by the generalized F-test. The value of the F-statistic is

$$F = \frac{(\mathrm{RSS}_5 - \mathrm{RSS}_1)/2}{\mathrm{RSS}_1/(15-4)} = \frac{(4,508,761 - 2,250,956)/2}{2,250,956/11} = \boxed{5.5167}. \quad \textbf{(Answer: (D))}$$

□

Copyright © 2018 ACTEX Learning

Remark. Erroneously taking Model (3) as the reduced model would produce an F-statistic of 6.5935, leading to Answer (E).

7. **(Calculation of WLS estimate in an SLR setting)**

Solution. To restore homoscedasticity, we divide both sides of the model equation by x_i, leading to

$$\frac{y_i}{x_i} = \frac{\beta_0}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i}, \quad \text{where } \frac{\varepsilon_i}{x_i} \overset{\text{i.i.d.}}{\sim} \text{N}(0, \sigma^2), \quad i = 1, 2, \ldots, 5.$$

This is a homoscedastic SLR model whose response variable is y/x, and the explanatory variable is $1/x$. The corresponding intercept and slope parameters are β_1 and β_0 (not β_0 and β_1!), respectively. The dataset for this model is:

i	y_i/x_i	$1/x_i$
1	1	1
2	-1	1/2
3	5/4	1/4
4	$-10/9$	1/9
5	25/16	1/16

Inputting $\{(1/x_i, y_i/x_i)\}_{i=1}^5$ into a financial calculator, we get $\hat{\beta}_1^{\text{WLS}} = \boxed{0.2539}$ (and $\hat{\beta}_0^{\text{WLS}} = 0.2245$). **(Answer: (E))** □

8. **(Detecting collinearity via VIF)**

Solution. The VIF of x_1 is

$$\text{VIF}_1 = \frac{1}{1 - R^2_{(1)}} = \frac{1}{1 - 0.95} = 20,$$

which is higher than 10. For x_2, note that $R^2_{(2)} = R^2_{(1)}$. This is because for simple linear regression, R^2 is the square of the correlation between the response and explanatory variables, and the correlation between x_1 and x_2 is the same as the correlation between x_2 and x_1. It follows that $\text{VIF}_2 = \text{VIF}_1 = 20 > 10$. Both X_1 and X_2 display excessive collinearity. **(Answer: (D))** □

9. **(True-or-false statements about resampling methods)**

Solution. I. Incorrect. Neither the validation set approach nor k-fold CV is a special case of another.

II. Correct. LOOCV has lower bias than k-fold cross-validation because of the larger training set.

Copyright © 2018 ACTEX Learning

III. Correct. k-fold CV is less computationally expensive than LOOCV because of the need for fitting k instead of n models.

(Answer: (C)) □

10. *Solution.* With a slight abuse of notation, we write "I", "II," and "III" to represent the residual sums of squares RSS_I, RSS_II, and RSS_III, respectively. From (i), we are given that

$$\frac{(\text{I}-\text{II})/1}{\text{II}/(40-3)} = 37\left(\frac{\text{I}}{\text{II}}-1\right) = 30 \quad \Rightarrow \quad \frac{\text{I}}{\text{II}} = \frac{67}{37}.$$

From (ii), we also have

$$\frac{(\text{II}-\text{III})/1}{\text{III}/(40-4)} = 36\left(\frac{\text{II}}{\text{III}}-1\right) = 12 \quad \Rightarrow \quad \frac{\text{II}}{\text{III}} = \frac{4}{3}.$$

It follows that

$$\frac{\text{I}}{\text{III}} = \frac{\text{I}}{\text{II}} \times \frac{\text{II}}{\text{III}} = \frac{67}{37} \times \frac{4}{3} = \frac{268}{111}.$$

Moreover, we deduce from R_a^2 for Model I that

$$1 - \frac{\text{I}/38}{\text{TSS}/39} = 0.5484 \quad \Rightarrow \quad \frac{\text{I}}{\text{TSS}} = 0.440021$$
$$\Rightarrow \quad \frac{\text{III}}{\text{TSS}} = \frac{\text{I}}{\text{TSS}} \times \frac{\text{III}}{\text{I}} = 0.182247.$$

Finally, the adjusted R^2 for Model III is

$$R^2_{a,\text{III}} = 1 - \frac{\text{III}/(40-4)}{\text{TSS}/(40-1)} = 1 - 0.182247 \times \frac{39}{36} = \boxed{0.8026}. \quad \textbf{(Answer: (B))}$$

□

11. **(Effects of shrinkage parameter on various statistics)**

Comments: This problem is motivated from and should be compared with Example 4.4.3.

Solution. As the shrinkage parameter λ increases, the model becomes less flexible and the amount of shrinkage increases. The squared bias should increase, corresponding to Curve Z, and the variance should decrease, corresponding to Curve X. As a function of the squared bias and variance, the test MSE typically exhibits a U-shape, corresponding to Curve Y. **(Answer: (E))** □

Remark. The training MSE should also increase as λ increases.

12. **(Performing the first step of backward selection)**

Solution. Using backward selection, we should first drop the variable with the smallest t-statistic *in absolute value*:

Variable	Estimate	Standard Error	t-statistic
Intercept	0.6	0.15	4
x_1	-0.2	0.30	$\boxed{-0.6667}$
x_2	0.5	0.65	0.7692
x_3	-1.4	0.45	-3.1111

The variable to be eliminated first is x_1. **(Answer: (B))** □

13. **(Counting the number of parameters in a GLM)**

Solution. The number of parameters corresponding to each explanatory variable is:

- x_1: 2
- x_2 and x_3: 1 each, for a total of 2
- x_4: 5
- x_5: 1

There are also two $(= 2 \times 1)$ interaction terms between x_1 and x_5. Together with the intercept, in total there are $2 + 2 + 5 + 1 + 2 + 1 = \boxed{13}$ parameters. **(Answer: B)** □

14. **(Point prediction for a GLM)**

Solution. The estimated linked mean is $4.78 - 0.35 = 4.43$, so the predicted pure premium is $\hat{\mu} = \mathrm{e}^{4.43} = \boxed{83.9314}$. **(Answer: (B))** □

15. **(Variance estimation for a GLM)**

Solution. It has been calculated in the preceding question that $\hat{\mu} = \mathrm{e}^{4.43}$. The estimated variance is $\hat{\mu}^2 = \mathrm{e}^{8.86} = \boxed{7,044.48}$. **(Answer: (D))** □

16. **(Manipulating loglikelihood and AIC)**

Comments: This problem is an extension of Problem 5.4.21 on page 342.

Solution.

- From the first model, we deduce that Age has $5 - 1 = 4$ parameters.
- The second model shows that Vehicle Body has $11 - 1 = 10$ parameters.

Copyright © 2018 ACTEX Learning

- The AIC of the third model is 446. Thus

$$2(196) + \underbrace{2p}_{X} = 446,$$

resulting in $p = 27$ parameters. Subtracting the $4+1 = 5$ parameters from the intercept and age, the number of parameters associated with Vehicle Value is $27 - 5 = 22$.

Therefore, the number of parameters in the last model is

$$Y = 1 + 4 + 10 + 22 = 37$$

and

$$B = \frac{500 - 2Y}{2} = \boxed{213}. \qquad \textbf{(Answer: (C))}$$

□

17. **(Miscellaneous facts about the normal linear model as a GLM)**

Solution. All of the three statements are correct.

I. This follows directly from the definition of deviance.

II. Note that the deviance D involves σ^2, which is often an unknown parameter. As a result, D cannot be directly calculated.

III. Recall from Example 5.1.11 on page 287 (or you can derive) that

$$D = \frac{\text{RSS}}{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{\sigma^2},$$

so that each deviance residual is

$$d_i = \text{sign}(y_i - \hat{\mu}_i) \times \sqrt{\frac{(y_i - \hat{\mu}_i)^2}{\sigma^2}} = \frac{y_i - \hat{\mu}_i}{\sigma},$$

which agrees with the Pearson residual (equal to the observation less the fitted value, divided by the standard deviation). **(Answer: (E))** □

Remark. As with deviance, both the deviance residual and Pearson residual cannot be used directly as a goodness of fit measure.

18. **(Logistic regression with grouped data)**

Solution. The estimated claim probability of male policyholders belonging to age group 2 is

$$\hat{\pi} = \frac{e^{-1.1155-0.4192+1.2167(2)-0.5412(2)^2}}{1 + e^{-1.1155-0.4192+1.2167(2)-0.5412(2)^2}} = \frac{e^{-1.2661}}{1 + e^{-1.2661}} = 0.219926.$$

Copyright © 2018 ACTEX Learning

As the 35 policyholders act independently (notice the word "independent" in the question!), the number of claims from these 35 policyholders is a binomial random variable with parameters $n = 35$ and π, and with an estimated variance of

$$\widehat{\text{Var}}(y) = 35\hat{\pi}(1-\hat{\pi}) = \boxed{6.0045}. \qquad \textbf{(Answer: (E))}$$

□

19. **(Classification using logistic regression)**

Solution. The fitted probabilities for the three policies are

$$\begin{aligned}
\hat{\pi}_I &= \frac{1}{1+\mathrm{e}^{1.1155+0.4192-1.2167(1)+0.5412(1)^2}} = \frac{1}{1+\mathrm{e}^{0.8592}} = 0.2975, \\
\hat{\pi}_{II} &= \frac{1}{1+\mathrm{e}^{1.1155+0.4192-1.2167(2)+0.5412(2)^2}} = \frac{1}{1+\mathrm{e}^{1.2661}} = 0.2199, \\
\hat{\pi}_{III} &= \frac{1}{1+\mathrm{e}^{1.1155-1.2167(3)+0.5412(3)^2}} = \frac{1}{1+\mathrm{e}^{2.3362}} = 0.0882.
\end{aligned}$$

As only $\hat{\pi}_I > 0.25$, $\boxed{\text{only Policy I}}$ is predicted to have a claim. **(Answer: (A))** □

20. **(Calculation of deviance given raw data, Poisson case)**

Solution. From Example 5.1.9 (or you can derive it from scratch), the deviance formula is

$$D = 2\sum_{i=1}^{9}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right].$$

In the presence of the intercept, we have $\sum_{i=1}^{9}\hat{\mu}_i = \sum_{i=1}^{9} y_i$, so the estimated mean $\hat{\mu}$ when $x = -1$ can be deduced via solving

$$2\hat{\mu} + 4(7.45163) + 3(12.38693) = 2 + 3 + 6 + \cdots + 15,$$

so $\hat{\mu} = 2.516345$. The deviance is

$$\begin{aligned}
D &= 2\left[2\ln\frac{2}{2.516345} + 3\ln\frac{3}{2.516345}\underbrace{-1.3004 - 0.43766 + \cdots + 2.87112}_{0.87923}\right] \\
&= \boxed{1.89}. \qquad \textbf{(Answer: (D))}
\end{aligned}$$

□

Remark. Do not forget to multiply the loglikelihood difference by 2.

21. **(Calculation of sample autocorrelation)**

Copyright © 2018 ACTEX Learning

Solution. The sample mean is

$$\bar{y} = \frac{1 + 1.5 + 1.6 + 1.4 + 1.5 + 1.7}{6} = 1.45.$$

The sum of squares of the whole series is

$$\sum_{t=1}^{6}(y_t - \bar{y})^2 = (-0.45)^2 + 0.05^2 + 0.15^2 + (-0.05)^2 + 0.05^2 + 0.25^2 = 0.295,$$

and the sum of lag-3 cross products is

$$\sum_{t=4}^{6}(y_{t-3} - \bar{y})(y_t - \bar{y}) = (-0.45)(-0.05) + (0.05)(0.05) + 0.15(0.25) = 0.0625.$$

By (6.1.1), the sample ACF at lag 3 is

$$r_3 = \frac{0.0625}{0.295} = \boxed{0.2119}. \qquad \textbf{(Answer: (D))}$$

□

22. **(Prediction interval for random walk)**

Comments: This is a modification of SRM Sample Questions 3 and 4.

Solution. Given $y_0, y_1, \ldots, y_8$, we can deduce the white noise values by differencing:

t	1	2	3	4	5	6	7	8
$c_t = y_t - y_{t-1}$	2	-3	5	3	4	2	4	-1

Then we can estimate the white noise mean and variance, respectively, as

$$\begin{aligned} \bar{c} &= \frac{2 + (-3) + \cdots + (-1)}{8} = 2 \qquad \left(\text{or more simply } \frac{y_8 - y_0}{10} = \frac{16 - 0}{8} = 2\right) \\ s_c^2 &= \frac{(2-2)^2 + (-3-2)^2 + \cdots + (-1-2)^2}{\underbrace{7}_{\text{not } 8}} = \frac{52}{7}. \end{aligned}$$

Since $y_{15} = y_8 + c_9 + \cdots + c_{15}$, the approximate 95% symmetric prediction interval for y_{15} is

$$\begin{aligned} (y_8 + 7\bar{c}) \pm 2\sqrt{7s_c^2} &= [16 + 7(2)] \pm 2\sqrt{7(52/7)} \\ &= \boxed{(15.58, 44.42)}. \qquad \textbf{(Answer: (C))} \end{aligned}$$

□

Copyright © 2018 ACTEX Learning

23. **(Evaluating simple exponential smoothed forecasts)**

Solution. We determine $\hat{s}_t$ for $t = 1, 2, 3, 4, 5$ recursively in accordance with (7.1.3):

t	y_t	$\hat{s}_t = 0.2y_t + 0.8\hat{s}_{t-1}$
1	20	$0.2(20) + 0.8(25) = 24$
2	30	$0.2(30) + 0.8(24) = 25.2$
3	60	$0.2(60) + 0.8(25.2) = 32.16$
4	40	$0.2(40) + 0.8(32.16) = 33.728$
5	15	(not needed)

The mean absolute percentage error is

$$\begin{aligned}
\text{MAPE} &= \frac{100}{5}\sum_{t=1}^{5}\left|\frac{e_t}{y_t}\right| \\
&= \frac{100}{5}\sum_{t=1}^{5}\left|\frac{y_t - \hat{s}_{t-1}}{y_t}\right| \\
&= \frac{100}{5}\left(\left|\frac{20-25}{20}\right| + \left|\frac{30-24}{30}\right| + \left|\frac{60-25.2}{60}\right|\right. \\
&\qquad \left. + \left|\frac{40-32.16}{40}\right| + \left|\frac{15-33.728}{15}\right|\right) \\
&= \boxed{49.49}. \qquad \textbf{(Answer: (E))}
\end{aligned}$$

□

24. **(Parameter estimation by conditional least squares)**

Solution. The sample mean is $\bar{y} = 10$. The centered data points are:

t	1	2	3	4	5
$y_t - \bar{y}$	1	0	-1	0.5	-0.5

The conditional LSEs of β_1 and β_0 are respectively

$$\hat{\beta}_1 = \frac{1(0) + 0(-1) + (-1)(0.5) + 0.5(-0.5)}{1^2 + 0^2 + (-1)^2 + 0.5^2} = -1/3$$

and $\hat{\beta}_0 = \bar{y}(1 - \hat{\beta}_1) = 40/3$. Then the fitted values and residuals are computed as follows:

Copyright © 2018 ACTEX Learning

Year	Stock Price	Fitted Stock Price	Residual
1	11.0	N.A.	N.A.
2	10.0	29/3	1/3
3	9.0	10	−1
4	10.5	31/3	1/6
5	9.5	59/6	−1/3

Then $\bar{e} = -5/24$ and the MSE is

$$\begin{aligned} &\frac{s^2}{} \\ = &\frac{[1/3-(-5/24)]^2+[-1-(-5/24)]^2+[1/6-(-5/24)]^2+[-1/3-(-5/24)]^2}{4-2} \\ = &\ 155/288 = \boxed{0.5382}. \qquad \textbf{(Answer: (C))} \end{aligned}$$

□

25. **(AR(1) prediction interval)**

Solution. Note that y_{2020} is two steps ahead of 2018. By (7.2.5),

$$\begin{aligned} \hat{y}_{2019} &= \hat{\beta}_0 + \hat{\beta}_1 y_{2018} = 0.75 + (-0.6)(10) = -5.25, \\ \hat{y}_{2020} &= \hat{\beta}_0 + \hat{\beta}_1 \hat{y}_{2019} = 0.75 + (-0.6)(-5.25) = 3.9. \end{aligned}$$

Then the symmetric 95% prediction interval for y_{2020} is

$$\begin{aligned} \hat{y}_{2020} \pm t_{18-3,0.025}\sqrt{s^2(1+\beta_1^2)} &= 3.9 \pm 2.1314\sqrt{0.5[1+(-0.6)^2]} \\ &= 3.9 \pm 1.7576 \\ &= \boxed{(2.14, 5.66)}. \qquad \textbf{(Answer: (E))} \end{aligned}$$

□

Remark. Taking $z_{0.025} = 1.96$ in place of $t_{15,0.025} = 2.1314$ leads to Answer (C).

26. *Solution.* I. This inequality is expected to be true. The bagged tree aggregates single trees, which avoids overfitting and thus we expect that the test error will decrease.

II. This statement is true. A Random forest is essentially a bagged tree and it is expected to reduce the test error, just like a bagged tree.

III. This statement is not true. A random forest decorrelates the individual trees. The goal is to decrease the test error w.r.t. the bagged tree.

(Answer: (A)) □

Copyright © 2018 ACTEX Learning

27. *Solution.* All observations belonging to the same pure node belong to the same class. Therefore, if R_j is a pure node with all observations belonging to class k, we have that $\hat{p}_{j,k} = 1$ and $\hat{p}_{j,l} = 0$ for all $l \neq k$. Therefore, the misclassification error E is given by

$$E = 1 - \max_l \hat{p}_{j,l} = 0.$$

So I is correct and II is False.

Since there are only two classes, we can introduce the following notation: $\hat{p}_j = \hat{p}_{j,1}$ and then

$$1 - \hat{p}_j = \hat{p}_{j,2}.$$

The misclassification error E_j of leaf R_j is then given by:

$$E_j\left(\hat{p}_j\right) = 1 - \max\left\{\hat{p}_j, 1 - \hat{p}_j\right\}.$$

Note that this is a piecewise linear function in $\hat{p}_j$, with:

$$\begin{aligned} E_j(0) &= 0, \\ E_j(1) &= 1, \\ E_j\left(\frac{1}{2}\right) &= \frac{1}{2}. \end{aligned}$$

The Gini Index G_j of this node is given by:

$$G_j\left(\hat{p}_j\right) = 2\hat{p}_j\left(1 - \hat{p}_j\right) = -2\hat{p}_j^2 + 2\hat{p}_j.$$

This function satisfies the following equalities:

$$\begin{aligned} G_j(0) &= 0 \\ G_j(1) &= 1 \\ G_j\left(\frac{1}{2}\right) &= \frac{1}{2}. \end{aligned}$$

Moreover, this function is twice differentiable and concave. Indeed, one can easily verify that its second derivative is negative. As a result, one should have that:

$$E_j\left(\hat{p}_j\right) \leq G_j\left(\hat{p}_j\right).$$

So statement III is true and statement IV is False. **(Answer: (C))** □

28. *Solution.* Similar to Example 10.1.2, this question tests basic properties of the PCA in the case of two dimensional data. Let us consider each of the answers.

(A) The correlation coefficient of the two principal components is given by

$$\rho(Z_1, Z_2) = \frac{\mathrm{Cov}(Z_1, Z_2)}{\sqrt{\mathrm{Var}(Z_1)}\sqrt{\mathrm{Var}(Z_2)}} = 0,$$

because it is known that Z_1 and Z_2 are uncorrelated.

Copyright © 2018 ACTEX Learning

(B) Recall that all principal components are linear combinations of the feature values. For example, the first principal component score for the i-th policyholder is given by

$$z_{i1} = \phi_{11}(\mathsf{age}_i - \overline{\mathsf{age}}) + \phi_{21}(\mathsf{veh}_i - \overline{\mathsf{veh}}).$$

Suppose that there are n policyholders in total. Then it follows that

$$\sum_{i=1}^{n} z_{i1} = \phi_{11}\sum_{i=1}^{n}(\mathsf{age}_i - \overline{\mathsf{age}}) + \phi_{21}\sum_{i=1}^{n}(\mathsf{veh}_i - \overline{\mathsf{veh}}) = 0,$$

since

$$\sum_{i=1}^{n}(\mathsf{age}_i - \overline{\mathsf{age}}) = \sum_{i=1}^{n}\mathsf{age}_i - n\overline{\mathsf{age}} = 0, \qquad \sum_{i=1}^{n}(\mathsf{veh}_i - \overline{\mathsf{veh}}) = \sum_{i=1}^{n}\mathsf{veh}_i - n\overline{\mathsf{veh}} = 0.$$

Therefore, the average of principal component scores must be zero.

(C) The first principal component score is given by

$$\begin{aligned} z_{51} &= \phi_{11}(\mathsf{age}_5 - \overline{\mathsf{age}}) + \phi_{21}(\mathsf{veh}_5 - \overline{\mathsf{veh}}) \\ &= 0.489 \times (34.15 - 35.04) + 0.872 \times (13.51 - 13.18) \\ &= -0.147. \end{aligned}$$

(D) Although the loading vector of the second principal component is not shown explicitly, we can infer its value from the property that the two loading vectors must be orthogonal to each other, i.e.,

$$\phi_{11}\phi_{12} + \phi_{21}\phi_{22} = 0, \qquad \phi_{12}^2 + \phi_{22}^2 = 1.$$

The second principal component score is given by

$$\begin{aligned} z_{52} &= \phi_{12}(\mathsf{age}_5 - \overline{\mathsf{age}}) + \phi_{22}(\mathsf{veh}_5 - \overline{\mathsf{veh}}) \\ &= 0.872 \times (34.15 - 35.04) - 0.489 \times (13.51 - 13.18) \\ &= -0.937 \end{aligned}$$

(E) There are two problems with the statement. First, when the number of principal components is equal to the number of descriptive features, then the representation is exact, not an approximation. Second, all descriptive features must be centralized. Therefore, the linear combination representation gives the difference between age and its mean, rather than the age value. In other words,

$$\mathsf{age}_5 - \overline{\mathsf{age}} = z_{51}\phi_{11} + z_{52}\phi_{12}.$$

Since we know exact values of both principal component scores, loading vectors as well as the age values, it is in fact possible to verify this identity numerically. Observe that on the left-hand side

$$\mathsf{age}_5 - \overline{\mathsf{age}} = -0.88,$$

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

on the right-hand side

$$z_{51}\phi_{11} + z_{52}\phi_{12} = -0.147 \times 0.489 - 0.937 \times 0.872 = -0.889.$$

The minor difference is due to the rounding error from using only three decimal places. One can observe a similar result for the vehicle market value variable

$$\texttt{veh}_5 - \overline{\texttt{veh}} = z_{51}\phi_{21} + z_{52}\phi_{22}.$$

In conclusion, the only incorrect statement is **E**. **(Answer: (E))** □

29. *Solution.* The squared Euclidean distance between x_i and x_j is denoted by $d_{i,j}^2$ and:

$$\begin{aligned}
d_{1,2}^2 &= (2-1)^2 + (3-2)^2 + (4-3)^2 + (5-4)^2 \\
&= 4 \\
d_{1,3}^2 &= (2-1)^2 + (3-2)^2 + (4-1)^2 + (5-2)^2 \\
&= 20 \\
d_{1,4}^2 &= (2-3)^2 + (3-2)^2 + (4-3)^2 + (5-2)^2 \\
&= 12 \\
d_{2,3}^2 &= (1-1)^2 + (2-2)^2 + (3-1)^2 + (4-2)^2 \\
&= 8 \\
d_{2,4}^2 &= (1-3)^2 + (2-2)^2 + (3-3)^2 + (4-2)^2 \\
&= 8 \\
d_{3,4}^2 &= (1-3)^2 + (2-2)^2 + (1-3)^2 + (2-2)^2 \\
&= 8.
\end{aligned}$$

We then find:

$$W(C) = 2 \times \frac{4 + 20 + 12 + 8 + 8 + 8}{4} = 20.$$

(Answer: (C)) □

30. *Solution.* In order to determine to which cluster the observation x_{11} will be assigned, we need to determine the distance between the observation x_{11} and each of the centroids of the clusters. The centroid of cluster C_i is denoted by $\bar{x}_i$. We can

Copyright © 2018 ACTEX Learning

determine the centroids as follows:

$$\begin{aligned}
\bar{x}_1 &= \left(\frac{1+1+2}{3}, \frac{1+2+1}{3}, \frac{1+1+2}{3}\right) \\
&= \left(\frac{4}{3}, \frac{4}{3}, \frac{4}{3}\right), \\
\bar{x}_2 &= \left(\frac{3+2+3}{3}, \frac{3+3+2}{3}, \frac{3+2+3}{3}\right) \\
&= \left(\frac{8}{3}, \frac{8}{3}, \frac{8}{3}\right) \\
\bar{x}_3 &= \left(\frac{6+4}{2}, \frac{6+4}{2}, \frac{6+4}{2}\right) \\
&= (5,5,5) \\
\bar{x}_4 &= (7,7,7) \\
\bar{x}_5 &= \left(\frac{2+2}{2}, \frac{2+4}{2}, \frac{2+2}{2}\right) \\
&= (2,3,2)\,.
\end{aligned}$$

We have to determine the centroid which is closest to the observation x_{11}. The squared Euclidean distance between x_{11} and centroid $\bar{x}_i$ is denoted by $d^2(x_{11}, \bar{x}_i)$. We can calculate the distances as follows:

$$\begin{aligned}
d^2(x_{11}, \bar{x}_1) &= \left(2-\frac{4}{3}\right)^2 + \left(4-\frac{4}{3}\right)^2 + \left(2-\frac{4}{3}\right)^2 \\
&= \frac{4}{9} + \frac{64}{9} + \frac{4}{9} \\
&= 8. \\
d^2(x_{11}, \bar{x}_2) &= \left(2-\frac{8}{3}\right)^2 + \left(4-\frac{8}{3}\right)^2 + \left(2-\frac{8}{3}\right)^2 \\
&= \frac{4}{9} + \frac{16}{9} + \frac{4}{9} \\
&= 2.66. \\
d^2(x_{11}, \bar{x}_3) &= (2-5)^2 + (4-5)^2 + (2-5)^2 \\
&= 19. \\
d^2(x_{11}, \bar{x}_4) &= (2-7)^2 + (4-7)^2 + (2-7)^2 \\
&= 59 \\
d^2(x_{11}, \bar{x}_5) &= (2-2)^2 + (4-3)^2 + (2-2)^2 \\
&= 1.
\end{aligned}$$

The distance to cluster 5 is the smallest and hence in the next step of the algorithm, x_{11} will again be assigned to cluster 5. **(Answer: (E))** □

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

31. *Solution.*
- Statement I is false. K-means clustering requires specifying the number of clusters upfront. However, hierarchical clustering can also be applied if the number of clusters is known at the start and can still result in accurate results.
- Statement II is correct. In this particular case, in the first step of the algorithm, each observation is assigned to a cluster. Since each cluster contains one observation, the centroid and the observation are the same. Then it is clear that each observation is closest to its own centroid and as a result, the algorithm stops after one step and each cluster contains one observation.
- Statement III is false. In a single step of the algorithm of K-means clustering, observations are assigned to the cluster which centroid is the closest. Therefore, more than one observation can move to another cluster.

(Answer: (D)) □

32. *Solution.* The algorithm starts with the solution where there are n clusters. In this case, each cluster exactly has 1 element. In the next step, 2 clusters are merged, yielding $n-1$ clusters. Since two clusters are merged, there is one cluster with 2 elements, all other clusters have one element. In the next step, again two clusters are merged, giving $n-2$ clusters. There are two situations, or two clusters with a single element are merged or the cluster with 2 elements is merged with a single-element cluster. Hence, the maximum amount of element in each cluster is now 3. In the next step, there will be $n-3$ clusters. One creates the largest cluster if one merges a cluster with 3 elements with a cluster with 1 element. As a result, the maximum amount of elements in a cluster when there are $n-3$ clusters is 4.
(Answer: (D)) □

33. *Solution.* The best split is the one that gives the highest reduction in Gini index. Since there are 2 observations who purchased life insurance and 4 who didn't, the Gini index before the split is given by:

$$G = 2 \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{9} = 0.44.$$

Consider now the variable Income. If we use this variable to split the predictor space, we split the data in two groups: The group called High and the group Low. The response variables in each group are given by:

$$\begin{aligned} High &= \{Yes\} \\ Low &= \{No, No, No, No, Yes\}. \end{aligned}$$

Then, the Gini index for each group is given by

$$\begin{aligned} G_{High}^{Income} &= 2 \times 1 \times 0 = 0, \\ G_{Low}^{Income} &= 2 \times \frac{4}{5} \times \frac{1}{5} = \frac{8}{25}. \end{aligned}$$

Copyright © 2018 ACTEX Learning

Then, the Gini index G^{Income} of the tree with two leaves when Income is used to split, is given by

$$G^{Income} = \frac{1}{6} \times 0 + \frac{5}{6} \times \frac{8}{25} = 0.27.$$

Similarly, using the variable Sex to split yields:

$$\begin{aligned} M &= \{Yes, No, No\} \\ F &= \{No, No, Yes\}. \end{aligned}$$

The Gini index can then be determined as follows:

$$\begin{aligned} G^{Sex} &= \frac{1}{2} \times G_M^{Sex} + \frac{1}{2} \times G_F^{Sex} \\ &= \frac{1}{2} \times 2 \times \frac{1}{3} \times \frac{2}{3} + \frac{1}{2} \times 2 \times \frac{1}{3} \times \frac{2}{3} \\ &= 0.44. \end{aligned}$$

If the variable Age is used, we end up with the following two categories:

$$\begin{aligned} Young &= \{Yes, No, Yes\} \\ Old &= \{No, No, No\}. \end{aligned}$$

The Gini index for the variable Age can be determined as follows:

$$\begin{aligned} G^{Age} &= \frac{1}{2} \times G_{Young}^{Age} + \frac{1}{2} \times G_{Old}^{Age} \\ &= \frac{1}{2} \times 2 \times \frac{1}{3} \times \frac{2}{3} + \frac{1}{2} \times 2 \times 1 \times 0 \\ &= 0.22. \end{aligned}$$

If Location is used as a first split, the data is grouped as follows:

$$\begin{aligned} Urban &= \{Yes, No\} \\ Rural &= \{No, No, No, Yes\} \end{aligned}$$

The Gini index for the variable Location can be determined as follows:

$$\begin{aligned} G^{Location} &= \frac{1}{2} \times_{Urban}^{Location} + \frac{1}{2} \times G_{Rural}^{Location} \\ &= \frac{2}{6} \times 2 \times \frac{1}{2} \times \frac{1}{2} + \frac{4}{6} \times 2 \times \frac{3}{4} \times \frac{1}{4} \\ &= 0.41. \end{aligned}$$

The Gini index for the variable Children can be determined as follows:

$$\begin{aligned} G^{Children} &= \frac{1}{2} \times_{No}^{Children} + \frac{1}{2} \times G_{Yes}^{Children} \\ &= \frac{1}{2} \times 2 \times \frac{1}{3} \times \frac{2}{3} + \frac{1}{2} \times 22 \times \frac{1}{3} \times \frac{2}{3} \\ &= 0.44. \end{aligned}$$

We conclude that the lowest Gini index is achieved when Age is used. **(Answer: (A))** □

Copyright © 2018 ACTEX Learning

精算师学院 www.iswant.com | 精算师论坛 www.lwwhy.com

34. *Solution.* Single linkage to measure dissimilarity between clusters means that we measure the minimal distance. As a result, it is clear from the plot that the points A_3 and A_4 are the closest together. Since A_3 belongs to cluster C_1 and A_4 belongs to cluster C_2, the clusters C_1 and C_2 are the closest to each other and will be merged when we have to go from four to three clusters. **(Answer: (A))** □

35. *Solution.* Let us analyze why each of the statements I, II and IV is wrong.

    I. The PCA requires that all variables are individually scaled to have standard deviation one. Keep in mind that in order to find principal component loading vectors each step is used to maximize the sample variance of the scores of principal components. If variables are unscaled, then principal components are always unduely influenced by variables with largest variances. Therefore, scaling does have a substantial effect on the results obtained. For this reason, it is undesirable for the principal components to depend on an arbitrary choice of scaling. Thus the constraint of standard deviation one is imposed in each step of finding principal component loading vectors.

    II. Each principal component loading vector is unique, up to a sign flip. The signs may differ because each principal component loading vector specifies a direction in a lower dimension space for approximation: flipping the sign has no effect as the direction does not change.

    IV. The number of principal components can be used as a tuning parameter to be selected via cross-validation in a supervised analysis, such as linear regression. Since it is impossible to cross validate in an unsupervised learning, the original statement is self-contradictory. The number of principal components to use in an unsupervised PCA is typically determined by examining a scree plot, which visualizes the proportion of variance explained by each principal component.

    (Answer: (B)) □

Copyright © 2018 ACTEX Learning