Exam Statistics for Enterprise Data Analysis

Post-graduation in Enterprise Data Science & Analytics

2019

Duration: 90 m (max.)

Remark: Each question corresponds to 20/40 points. For each correct answer you get 20/40 points, for no answer or a wrong answer you get 0. No clarifications are made in a individual basis. You are allowed to use a one-sheet formulary. Choose the option that you think best fits the correct answer.

Name:_____ **Number**:_____

ANSWERS

Question	(a)	(b)	(c)	(d)	(e)	(f)	Question	(a)	(b)	(c)	(d)	(e)	(f)
1.							21.						
2.							22.						
3.							23.						
4.							24.						
5.							25.						
6.							26.						
7.							27.						
8.							28.						
9.							29.						
10.							30.						
11.							31.						
12.							32.						
13.							33.						
14.							34.						
15.							35.						
16.							36.						
17.							37.						
18.							38.						
19.							39.						
20.							40.						

1. Consider the following model multiple regression model, presented in the following table, where:

   - *wage* represents the salary of individual with 12 years of education or more;
   - *jc* represents the duration of the bachelor's degree course;
   - *rank* represents the rank of the high school where the individual studied (1 correspond to the worst position);
   - *mast* represents the duration of the master course;
   - *exper* represents the number of years of professional experience.

Parameter estimates (dependent variable: *wage*)				
Variable	Estimate	Standard error	t value	p-value
Intercept	1.47233	0.02106	69.91	< 0.0001
jc	0.06670	0.00683	9.77	< 0.0001
rank	0.00030	0.00024	1.27	0.2044
mast	0.07688	0.00231	33.30	< 0.0001
exper	0.00494	0.00015747	31.40	< 0.0001

Analysis of variance					
Source	df	Sum of Squares	Mea Square	F value	p-value
Model	3	357.75260	119.25087	644.53	< 0.0001
Error	6759	1250.54350	0.18502		
Total	6762	1608.29609			

$\hat{\sigma}^2$	0.43014	R-Square	0.2224
$\bar{y}$	2.24810	Adj R-Sq	0.2221

Combinations of bachelor's with master courses are possible. An individual with 12 years of education gets *jc*=*mast*=0.

Consider 2 individuals (A and B) with the same professional experience and that have attended the same high school. Both chose to attend a master course, of 2 and 3 years, respectively. What is the expected difference in their salaries?

(a) A salary is 0,07688% higher;

(b) B salary is 7,688% higher; (*)

(c) A salary 7,688% higher;

(d) B salary is 0,07688% higher.

(e) None of above.

2. Regarding question 1. The F value=644.533 indicates:
   (a) The model is valid from a statistical point of view. (*)
   (b) All explanatory variables are relevant.
   (c) Some explanatory variables are jointly irrelevant.
   (d) All explanatory variables are jointly irrelevant.
   (e) None of above.
3. Regarding question 1. In the hypotheses test $H_0 : \beta_{rank} = 0$ vs. $\beta_{rank} \neq 0$:
   (a) We don't reject the null hypothesis for a 5% significance because the p-value is 20,44%. (*)
   (b) We don't reject the null hypothesis for a 5% significance because the p-value is 10,22%.
   (c) We reject the null hypothesis for a 5% significance because the p-value is 20,44%.
   (d) We reject the null hypothesis for a 5% significance because the p-value is 10,22%.
   (e) None of above.
4. Regarding question 1. The value 0,2044 in column p-value indicates:
   (a) A 20,44% probability of rejecting the alternative hypothesis.
   (b) 20,44% probability of rejecting the null hypothesis, assuming it is true.
   (c) 20,44% probability of rejecting the null hypothesis, assuming it is false.
   (d) The probability of observing a value for the t-statistic that verifies $|t| > 1,27$. (*)
   (e) None of above.
5. Regarding question 1. A p-value< 0.0001, as those in column p-value, indicate:
   (a) The we never reject the null hypothesis $H_0 : \beta_k = 0$.
   (b) Strong evidence against the alternative hypothesis $H_1 : \beta_k \neq 0$.
   (c) Strong evidence supporting the null hypothesis $H_0 : \beta_k = 0$.
   (d) Strong evidence against the null hypothesis $H_0 : \beta_k = 0$. (*)
   (e) None of above.
6. Regarding question 1. The hypothesis $H_0 : \beta_{jc} = \beta_{mast}$ means that (in terms of additional salary):
   (a) Is preferable to spend one more year in a master course than in a bachelor one.
   (b) Is preferable to spend one more year in a bachelor course than in a master one.
   (c) Is preferable to attend master courses with the same duration of the bachelor one.
   (d) The marginal benefit of one more year of a master course is identical to one more year of a bachelor course. (*)
   (e) None of above.

7. Which of the following conditions are necessary for a series to be classifiable as a weakly stationary process?

   (i) It must have a constant mean

   (ii) It must have a constant variance

   (iii) It must have constant autocovariances for given lags

   (iv) It must have a constant probability distribution

   (a) (ii) and (iv) only

   (b) (i) and (iii) only

   (c) (i), (ii), and (iii) only (*)

   (d) (i), (ii), (iii), and (iv)

   (e) None of above

8. Consider the following sample autocorrelation estimates obtained using 250 data points:

Lag	1	2	3
Coefficient	0.2	-0.15	-0.1

   Assuming that the coefficients are approximately normally distributed, which of the coefficients are statistically significant at the 5% level?

   (a) 1 only.

   (b) 1 and 2 only. (*)

   (c) 1, 2 and 3 only.

   (d) It is not possible to determine the statistical significance since no standard errors have been given.

   (e) None of above

9. Which of the following statements are true concerning the autocorrelation function (acf) and partial autocorrelation function (pacf)?

   i) The acf and pacf will always be identical at lag one whatever the model

   ii) The pacf for an MA(q) model will in general be non-zero beyond lag q

   iii) The pacf for an AR(p) model will be zero beyond lag p

   iv) The acf and pacf will be the same at lag two for an MA(1) model

   (a) (ii) and (iv) only

   (b) (i) and (iii) only (*)

   (c) (i), (ii), and (iii) only

   (d) (i), (ii), (iii), and (iv)

10. Consider a monthly time series data exhibiting an increasing linear trend and a seasonal component. There are no problems with nonconstant variance. You should consider using a stationary process to model:

    (a) $(1-B)^{12}(1-B)x_t$.
    (b) $(1-B^{12})(1-B)x_t$. (*)
    (c) $(1-B)^{24}x_t$.
    (d) $(1-B^{12})(1-B^2)x_t$.
    (e) None of above.

11. For a 100 observations time series, the 95% confidence limits for the autocorrelation coefficients are

    (a) ± 0.196. (*)
    (b) ± 0.186.
    (c) ± 0.2.
    (d) None of the above.

12. Which of the following sentences represent a response bias?

    (a) Comments about the quality of the food in the homepage of a restaurant.
    (b) Answers to an anonimous questionnaire regarding leisure activities.
    (c) Answers to a questionniare by mothers of new borns regarding diapers price.
    (d) All of the above.
    (e) None of the above.

    (a) (i) and (ii) only
    (b) (i) only (*)
    (c) (i), and (iii) only
    (d) (v) only

13. Which of the following sentences are true regarding Pearson's ρ?

    (a) The sign of the correlation provides direction only.
    (b) Correlation is unit free.
    (c) Correlation between x and y is always different from correlarion between y and x.
    (d) It can be use for all types of relationships between variables.
    (e) Is resistant to outliers.

    (a) (i) and (ii) only (*)
    (b) (i), (ii), and (iv) only
    (c) (i) and (iii) only
    (d) (i) and (v) only

14. Which of the following sentences is true?

    (a) Type I errors happen when we reject the null hypothesis and the null hyphotesis is true.(*)
    (b) Type II errors happen when we reject the null hypothesis and the null hyphotesis is true.

15. Assume we use the t.test function in R to compare the means of two independent groups A and B. We are trying to measure the effect of a new promotional campaingh on the sales of our product, on both groups. Participants of both groups were randomly selected. Both sample sizes were 45. The p-value is 0.1167.You have to provide a recommendation. Which recommendation should we propose. (Assume the usual null hyphotesis)

    (a) Stop the promotional efforts.(*)
    (b) Mantain the campaingh, or expand it to all groups.