S115_LE4_001

Statistics 115 April 2, 2013
Fourth Long Examination TGCapistrano

I. FILL IN THE BLANKS.

1. If Y_i is the value of the response variable of the i^{th} element and X_i is the value of the explanatory variable of the i^{th} element then the simple linear regression model is defined as (a) $Y_i =$ _____ that satisfies the following conditions about the error terms, ε_is: (b) _____.
2. The population linear correlation coefficient of X and Y, denoted by ρ, is defined as $\rho =$ _____.
3. In a 6x3 table, the number of cells with expected frequencies less than 5 but greater than 1 must not be more than _____ for the chi-square test to be valid.
4. In the test for homogeneity comparing 3 populations in terms of the distribution of the categorical variable Y that has 5 categories, the null distribution of the test statistic is (a) _____ and the value of its parameter is (b) _____.
5. The critical region of the test Ho: $\rho=0$ vs Ha: $\rho\neq0$ at 0.01 level of significance based on a sample of size 40 is the collection of real numbers t that satisfies the condition that _____.
6. The individual deviations of the observations y_i from the predicted values using the estimated regression equation are called _____.
7. In simple linear regression analysis based on a sample of size 50, if the SSE=157.2 then the value of the estimator for the variance of ε_i is equal to _____.
8. If n=12, $\sum_{i=1}^{12} X_i = 31, \sum_{i=1}^{12} X_i^2 = 97, \sum_{i=1}^{12} Y_i = 729, \sum_{i=1}^{12} Y_i^2 = 59531, \sum_{i=1}^{12} X_iY_i = 2383, MSE = 48,$
   a) the value of b_1 is _____
   b) the value of b_0 is _____
   c) the value of the test statistic to test Ho:$\beta_1=0$ vs Ha: $\beta_1\neq0$ is _____
   (Round-off final answers only to 3 decimal places).
9. The three basic principles of experimental design are randomization, blocking and _____.
10. The component of a time series that describes the long-term movement in the time series and usually modeled by a smooth curve is called _____.
11. In computing S_5 using single exponential smoothing where $\alpha=0.2$, the weight of Y_3 is equal to _____.
12. A study was conducted to compare the immune response of mice to the venom of three different species of poisonous snakes, labeled as Treatments A, B, and C. The increase in antibody activity as measured from a blood sample of the mouse is the

response variable. All the mice in the experiment were each assigned a random number as follows:

	1	2	3	2	1	3	3	2	1	3	2	1	1	3	2
Mouse	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Random No.	25	37	89	42	10	64	83	52	8	72	68	15	2	94	22
	6	7	14	8	3	10	13	9	2	12	11	4	1	15	5

Each mouse in the experiment is randomly assigned a species of snake using these random numbers and receives minute amounts of the venom of the corresponding snake by injection.

a) Using CRD with 5 replicates per treatment, Mouse #1 is assigned Treatment ___.
b) Using CRD with 5 replicates per treatment, Mouse #2 is assigned Treatment ___.

Suppose Mouse nos. 1, 2, 3 all belong in the same litter and will form 1 block. Mouse nos. 4, 5, 6 form another block; and so on.

c) Using RCBD with 5 blocks, Mouse #1 is assigned Treatment ____.
d) Using RCBD with 5 blocks, Mouse #2 is assigned Treatment ____.

II. TRUE OR FALSE. Write 'True' if the statement is always true; otherwise, write 'False'.

1. The alternative hypothesis in a test for homogeneity is, "The relative frequency distributions of the c populations are all different from each other."
2. In the test for independence of X and Y, the expected frequencies are computed under the assumption that X and Y are related.
3. The population linear correlation coefficient of X and Y, denoted by ρ, can only take on values from -1 to 1.
4. If ρ is equal to 0 then there is no relationship between X and Y.
5. If ρ is equal to 1 then either X causes Y or Y causes X.
6. If r is equal to -1 then all of the sample points (X_i, Y_i) will fall on a line whose slope is negative.
7. The criterion used to derive the estimates for the regression coefficients using the method of least squares is to minimize the sum of the squares of the error terms ε_is.
8. If the estimated regression equation based on X-values ranging from -5 to 5 is $\hat{Y} = 5 + 10X$ then when X=0 the mean value of Y is estimated to be equal to 5.
9. The length of the $(1-\alpha)100\%$ confidence interval estimate for the regression coefficient β_0 is shorter when the MSE is smaller .
10. In the completely randomized design (CRD), the experimental error will be large when the experimental units are heterogeneous.
11. In the randomized complete block design (RCBD), the experimental units belonging in the same block must be homogeneous.
12. The cyclical component of a time series recurs in varying length and magnitude.
13. The smaller the value of T in the single moving average then the greater the smoothing effect.
14. If the weight α used in single exponential smoothing is closer to 0 then the greater the smoothing effect.
15. The exponential decrease in the weights of the past observations is faster if the weight α used in single exponential smoothing is closer to 0.

III. DATA ANALYSIS. Always show important steps in your solution. No immediate rounding-off. Whenever necessary, round-off final answer only to 4 decimal places.

1. The following data are the amounts y of an unconverted substance from six similar chemical reactions after x minutes:

x_i (minutes)	5	7	8	8	10	10
y_i (milligrams)	65.0	41.4	32.2	30.9	8.6	7.0

a) Find the estimated regression equation with X as the explanatory variable and Y as the response variable.
b) Find the predicted value of Y when X=8.5 using the estimated regression equation in (a).
c) What is the critical region when we test Ho: $\beta_1 = 0$ vs Ha: $\beta_1 \neq 0$ at α=0.01?
d) What percentage of total variation in the values of Y in the sample can be accounted for or explained by its linear relationship with the values of X?

2. In a research to study the relationship of hypertension and smoking habits, the data were taken from a random sample of 208 individuals and the collected data were summarized in the following contingency table:

	Nonsmokers	Moderate smokers	Heavy smokers
With hypertension	18	30	56
No hypertension	54	25	25

a) State Ho and Ha.
b) Write the formula of the test statistic to be used.
c) State the decision rule at 0.1 level of significance.
d) Compute for the value of the test statistic. Present all the computed expected frequencies.
e) Is there sufficient evidence at 0.1 level of significance to conclude that hypertension and smoking habits are related?
f) Compute for Cramer's V.

3. The following data show the gross value added (in millions of pesos) in water supply, 1st quarter 2010 to 4th quarter 2012:

2010				2011				2012			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
4.49	4.84	4.33	4.84	4.41	4.78	4.29	5.02	4.94	5.09	4.56	5.23

a) Compute for the smoothed value for 4th quarter of 2011 using single moving average
(i) T=4
(ii) T=2

b) Compute for the smoothed value for 4th quarter 2010 using single exponential smoothing and initial value=Y_1
   (i) with α=0.3
   (ii) with α=0.9

4. Five different baking temperatures (A, B, C, D, E) are being tested to bake a ready-mix cake. The experiment will be performed on five days (1st day of the months Dec, Jan, Feb, March, and April) and five timeslots for each day (8:00, 10:00, 12:00, 2:00, 4:00.) Copy the following table in your bluebook and indicate the baking temperature to be used for each time slot of each month using the Latin square design:

Month	Time Slot				
	8:00	10:00	12:00	2:00	4:00
Dec					
Jan					
Feb					
March					
April					

Use the following random numbers and the 5x5 latin square below:

Row 1: 43	**Row 2: 27**	**Row 3: 95**	**Row 4: 12**	**Row 5: 58**
Col 1: 20	**Col 2: 24**	**Col 3: 12**	**Col 4: 83**	**Col 5: 62**

Formulas:

$$S_{b_1} = \sqrt{\frac{MSE}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}}$$

Latin square:

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A