國立成功大學 108 學年度碩士班招生考試試題

系　　所：國際企業研究所

考試科目：統計學

考試日期：0224，節次：3

※ 考生請注意：本試題可使用計算機。　請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

I. Choose the Best Answer - Concept Questions (30 points, 3 pts each)

1. Two data analysts had the following conversation:
Data Analyst A: We should try to enlarge our sample size as possible since large sample size could increase our estimation precision for sure.
Data Analyst B: The sample data collection should include both discrete and continuous data for categorical and numerical data. Otherwise, the statistical analysis will not be complete.

Which data analyst is correct in a statistical sense?

(A) Data analyst A
(B) Data Analyst B
(C) Both data analysts are incorrect

2. The nonlinear relationship between two data sets can be measured by

(A) Pearson's correlation coefficient
(B) Spearman's rank correlation coefficient
(C) Both methods above can do the job

3. Which of the following graphs will specifically label the quartiles?

(A) Box and Whisker chart
(B) Bar chart
(C) Cumulative frequency graph

4. Two data analysts had the following conversation of sampling methods:
Data Analyst A: If the population is grouped first, and then the sample is composed of every member of some groups. This is the clustered sampling method.
Data Analyst B: If the population is grouped first, and then the sample is composed of some members of each group. This is the clustered sampling method.

(A) Data Analyst A
(B) Data Analyst B
(C) Both data analysts are incorrect

5. Which of following survey methods might have higher response rate?

(A) Self-Administered survey
(B) Telephone interview
(C) Personal interview

6. Which of following measure could be used for numerical and categorical data regarding the central tendency of a distribution?

(A) Mean
(B) Mode
(C) Median

7. Two data analysts had the following conversation:
Data Analyst A: If the value of a random sample variable is tripled, then the standard deviation of this random variable will be tripled as well.
Data Analyst B: If the value of a random sample variable is tripled, then the mean of this random variable will be tripled as well.

(A) Data Analyst A
(B) Data Analyst B
(C) Both data analysts are incorrect

8. Two data analysts had the following conversation:
Data Analyst A: If we would like to measure the dispersion of a population, the mean value of absolute deviations from the mean could do the job.
Data Analyst B: If we would like to measure the dispersion of a population, the interquartile could do the job.

(A) Data Analyst A
(B) Data Analyst B
(C) Both data analysts are correct

9. Two data analysts had the following conversation:
Data Analyst A: If there is a possibility of having extremely higher positive value on the right-hand side of a distribution, then this distribution should have high kurtosis.
Data Analyst B: If there are higher possibilities of tail values in a distribution than a normal distribution, then this distribution should be highly skewed.

(A) Data Analyst A
(B) Data Analyst B
(C) Both data analysts are incorrect

10. Two data analysts had the following conversation:
Data Analyst A: The efficient estimator is the one which has the lowest estimation error.
Data Analyst B: The unbiased estimator is the one which has the lowest estimation error.

(A) Data Analyst A
(B) Data Analyst B
(C) Both data analysts are correct

國立成功大學 108 學年度碩士班招生考試試題

系　　所：國際企業研究所

考試科目：統計學

考試日期：0224，節次：3

II. Choose the BEST answer -Computational Questions (45 points, 3 pts each)

Use the following information to answer Questions 1 to 5.

The following table contains two variables, X and Y:

X	Y
0.16	0.11
0.22	0.14
0.11	0.08
0.27	0.20
0.10	0.09
0.15	0.16
0.12	0.10
0.21	0.15
0.17	0.12
0.18	0.17
0.14	0.07
0.25	0.19

1. What is the Pearson's Correlation Coefficient between X+2 and 3Y?

(A) 0.873
(B) 0.862
(C) 0.851
(D) 0.839

2. What is the Spearman's rank correlation coefficient between X and Y+2?

(A) 0.842
(B) 0.839
(C) 0.828
(D) 0.817

3. What is the number of concordant pairs between X and Y? (Hint: the concordant pair is the pair with the following condition: if two pairs (x_1, y_1) and (x_2, y_2), either $x_1 > x_2$ and $y_1 > y_2$ or $x_1 < x_2$ and $y_1 < y_2$)

(A) 55
(B) 54
(C) 53
(D) 52

4. What is the number of discordant pairs between X and Y? (Hint: the discordant pair is the pair with the following condition: if two pairs (x_1, y_1) and (x_2, y_2), either $x_1 < x_2$ and $y_1 > y_2$ or $x_1 > x_2$ and $y_1 < y_2$)

(A) 10
(B) 11
(C) 12
(D) 13

5. What is the Kendaull's tau-b correlation coefficient between X+2 and Y+2?

(A) 0.689
(B) 0.678
(C) 0.667
(D) 0.656

Use the following information to answer Questions 6 to 10.

There is the following frequency table with the sample size of 20

Group	Cumulative frequency
11.22-14.22	0.2
15.35-18.35	0.5
21.28-24.28	0.9
31.84-34.84	1.0

6. What is the group mode in this frequency table if we shift the values with an increase of 0.25 in this distribution?

(A) 23.03
(B) 24.14
(C) 25.25
(D) 26.36

7. What is the group mode of this frequency table if we shift the values with an increase of 0.25 and scale the values with 0.75?

(A) 27.48
(B) 28.59
(C) 29.60
(D) 30.71

8. What is the group mean of this frequency table if we shift the values with an increase of 0.25 and scale the values with 0.75?

(A) 26.95
(B) 27.06
(C) 28.17
(D) 29.28

9. What is the group median of this frequency table if we shift the values with an increase of 0.25 and scale the values with 0.75?

(A) 20.68
(B) 21.79
(C) 22.80
(D) 23.81

10. What is the group standard deviation of this frequency table if we shift the values with an increase of 0.25 and scale the values with 0.75?

(A) 69.54%
(B) 67.43%
(C) 65.32%
(D) 63.21%

Use the following information to answer Questions 11 to 15.

The hotel rating table (rating ranging from 1 to 10) is listed as follows:

Hotel Rating		
Hotel A	Hotel B	Hotel C
2.03	3.89	6.65
3.08	7.58	5.74
3.53	9.48	5.86
4.68	8.52	4.71
5.46	8.43	6.72

11. What are the median ratings of these three hotels (A, B, C), respectively?

(A) 2.03, 3.89, 6.65
(B) 5.46, 7.58, 4.71
(C) 3.08, 8.52, 5.74
(D) 3.53, 8.43, 5.86

12. What are the sums of the absolute deviations from the medians for these three hotels (A, B, C), respectively?

(A) 5.03, 6.53, 2.92
(B) 6.14, 7.64, 3.03
(C) 7.25, 8.75, 4.14
(D) 8.35, 9.86, 5.25

13. What is the F statistic value of modified Levene test for the homogeneity test among three hotels?

(A) 0.233
(B) 0.344
(C) 0.455
(D) 0.566

14. If we would like to perform the post hoc tests using the Tukey-Kramer procedure in hotel ratings for these three hotels, what is the number of pairs needed to be compared?

(A) 2
(B) 3
(C) 4
(D) 5

15. If the Q_α in the Tukey-Kramer procedure in this case is 3.77, and what is the critical range of the Tuckey-Kramer procedure in this case and is the rating difference between Hotel A and Hotel B significant?

(A) 2.58, significant
(B) 2.61, significant
(C) 2.72, insignificant
(D) 2.83, insignificant

III. Partial Credit Questions and Fill in the Blanks (25points, 5 pts each)

Notes:

(1) Label blanks in alphabetical order.

(2) Write down your answers along with associated blanks.

(3) Write down your problem-solving steps in order to possibly obtain partial credits

1. An economic analyst tries to forecast the economic performance in a country. The current information he holds is as follows:

Year (t: time period)	Income (unit = billion, y_t)
2014 (t = 1)	$102.32
2015 (t = 2)	$105.24
2016 (t = 3)	$109.25
2017 (t = 4)	$114.35

(1) If this analyst finds that a quadratic trend model (i.e., $y_t = \beta_0 + \beta_1 \times t^2$) is a perfect fit to the data above, what is the estimated coefficient of a quadratic trend term, β_1, using the difference method without using the least squares estimation method? ___(a)___

(2) What is estimated income for the year of 2018 if β_0 is equal to $103.7025 and the estimated β_1 in (a) is used? ___(b)___

(3) What is the estimated income for the year of 2018 if β_0 is equal to $103.7025 and the estimated β_1 in (a) is used plus that annual estimation errors are estimated using the exponential smoothing method with an exponential smoothing constant of 0.97? ___(c)___

2. A company employs the randomized block design to understand which promotion campaign is much more effective to obtain the higher click-through rate (CTR). The marketing team collects the following information regarding the preference feedback (on a scale from one with being less preferred to 5 with being most preferred):

	Promotion Campaign		
Block	**Campaign A**	**Campaign B**	**Campaign C**
1	3	4	5
2	2	4	3
3	2	4	3
4	2	2	1
5	4	5	5

The marketing team employs the Friedman F_r-statistic, a nonparametric test, to test whether these three promotion campaigns differ.

(1) What is the rank sum of Campaign B under the Friedman F_r-statistic? ___(d)___

(2) What is the value of the Friedman F_r-statistic? ___(e)___