※ 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.　True or False (30 points, 3 pts each)

1. The scales of observed variables need to be determined before collecting the real data of those variables.

2. When performing multiple comparisons of means, one has to set the confidence level at a higher level for each comparison than of the original experimental confidence level. This approach is adopted to prevent the hazard of making a Type I error from exceeding the pre-specified confidence level.

3. The Durbin-Waston test is employed to detect correlation between variables cross-sectionally.

4. The main objective of a statistical analysis of observed data always focuses on the average pattern presented by the observed data and therefore outliers have no places in the statistical analysis.

5. The logarithmic transformation of a variable is, in general, able to make its volatility smoother than its volatility computed using raw data.

6. The symmetric distribution of a variable means that the mean is equal to the median and the mode tends to be close to the mean.

7. When the Person's correlation coefficient between variables is significantly different from zero, it means that the variables have both linear and nonlinear relationships.

8. Researchers collected the weather data through readings recorded in the wireless weather apparatus. This data collection example belongs to an observational study instead of a designed experiment.

9. The independency between variables means that variables have no linear and nonlinear relationships among them.

10. The central limit theorem says that the average of observations obtained from the large sample size will be approximately close to the true mean of the whole population.

※ 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。

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

Use the following information to answer Questions 1 to 5.

A real estate research team investigate an issue regarding how people choose to live in either the city or the suburban area according to their income levels. They decide to adopt a 2 by 2 contingency table with income and location dimension factors in which there are only two levels for each factor. (High and Low for the income factor and City and Suburban area for the location factor)

1. If the value of Pearson's Chi-square is 2.376 and the value of Phi (φ) is 0.336, then what is the total number of subjects being enquired in this investigation?

   (A) 18
   (B) 19
   (C) 20
   (D) 21

2. If the number of low-income subjects is 11, what is the number of high-income subjects?

   (A) 8
   (B) 9
   (C) 10
   (D) 11

3. If the number of subjects choosing to live in the city is 47.62% of the total number of subjects, what is expected number of subjects who have high income and choose to live in the suburban area

   (A) 3
   (B) 5
   (C) 7
   (D) 9

4. What is Contingency Coefficient of this investigation case?

   (A) 0.319
   (B) 0.423
   (C) 0.512
   (D) 0.624

5. What is the value of Cramer's V coefficient in this investigation case?

   (A) 0.363
   (B) 0.359
   (C) 0.347
   (D) 0.336

※ 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。

Use the following information to answer Questions 6 to 10.

In addition to morale (MO), Padre, a MLB baseball team of national league, performs in each game depending on the FPCT (fielding percentage) and the BA (batting average). Statistics show some probabilities as follows: (there are only two conditions of being high and being low for each measure. HI: high; LO: low, P: probability):

6. What is the $P(BA = HI|MO = LO)$ given that the probability of a high batting average is 0.705, $P(BA = HI|MO = HI) = 0.75$, and the probability of high moral is 0.7?

   (A) 0.5
   (B) 0.6
   (C) 0.7
   (D) 0.8

7. Based on the information in the previous question, what is the probability of low morale and a low batting average?

   (A) 0.12
   (B) 0.13
   (C) 0.14
   (D) 0.15

8. Based on the information in the previous two questions, what is the $P(FPCT = HI|MO = HI, BA = HI)$ given that $P(FPCT = HI|MO = HI) = 0.4725$ and $P(FPCT = HI|MO = HI, BA = LO) = 0.6$?

   (A) 0.4
   (B) 0.5
   (C) 0.6
   (D) 0.7

9. Based on the information in the previous three questions, what is the probability of high morale, a low batting average and low fielding percentage?

   (A) 0.05
   (B) 0.06
   (C) 0.07
   (D) 0.08

10. Based on the information in the previous four questions, if the probability of high morale increases to 0.75 with the rest of information being unchanged, what is the probability of low morale and a high batting average?

    (A) 0.14
    (B) 0.15
    (C) 0.16
    (D) 0.17

※ 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。

Use the following information to answer Questions 11 to 15.

A marketing team would like to study the effects of three different media of TV, blogs and newspapers, on the purchase intention of target customers using ANOVA. Each medium has been evaluated by 28 subjects. The standard deviations of three media are 2.74 of TV, 1.686 of blogs and 2.358 of newspapers.

11. What is the sum of squares for error (SSE) for this one-way ANOVA analysis?

(A) 428.39
(B) 429.58
(C) 430.54
(D) 431.46

12. The total sum of squares (TSS) is 544.952 and then what is the mean square for treatments (MST) in this one-way ANOVA analysis? (Hint: TSS = SST (sum of squares for treatments) + SSE)

(A) 57.686
(B) 58.335
(C) 59.426
(D) 60.324

13. What is the F statistic for testing equality of purchase intention effects of media in this one-way ANOVA case?

(A) 13.144
(B) 12.233
(C) 11.344
(D) 10.877

14. In order to understand more about the purchase intention in detail, the marketing team classifies the data according to days (Monday to Sunday) in a week. The sum of squares of block (SSB for the day effect) is 151.786 and the sum of squares for error in this two-way ANOVA is 111.50. Then, what is the sum of squares for interaction between days and media?

(A) 164.575
(B) 165.132
(C) 166.294
(D) 167.712

15. What is the F statistic for testing the interaction effect between days and media in this two-way ANOVA case?

(A) 6.58
(B) 7.83
(C) 8.14
(D) 9.26

※ 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。

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

Notes:

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

    (2)　Label blanks in alphabetical order.

1.　The following table shows how paired practitioners and academicians predicted the price levels (unit: percentage) for the next year.

Practitioners	Academicians
3.10	4.20
4.20	2.30
2.50	4.20
5.10	8.25
0.50	7.23
2.45	9.83
3.20	10.25

    (1)　What is the rank sum of negative differences (using values of practitioners to deduct values of academicians) in the Wilcoxon signed rank test for related samples? ____(a)____

    (2)　What is the Spearman rank correlation between predictions of practitioners and academicians? ____(b)____

    (3)　What is the value of t statistic of the Spearman rank correlation value in (2)? ____(c)____

2.　Two stock returns have the following relationships with three different macroeconomic factors, F_{1t} F_{2t} and F_{3t}, which are uncorrelated with each other and have no correlation with error of ε_{1t} and ε_{2t}:

$$R_{1t} = 0.12 + 0.31F_{1t} + 0.35F_{2t} + \varepsilon_{1t}, t = 1,2, \dots T,$$
$$R_{2t} = 0.24 + 0.22F_{1t} + 0.42F_{3t} + \varepsilon_{2t}, t = 1,2, \dots T,$$

where R_{1t} and R_{2t} are stock returns of 1st and 2nd stocks, respectively, $\text{Var}(F_{1t}) = 0.23$, $\text{Var}(F_{2t}) = 0.15$, $\text{Var}(F_{3t}) = 0.42$, $\text{Var}(\varepsilon_{1t}) = 0.04$, $\text{Var}(\varepsilon_{2t}) = 0.09$, and the correlation between ε_{1t} and ε_{2t} is zero.

    (1)　What is the variance of R_{1t}? ____(d)____

    (2)　What is the covariance between R_{1t} and R_{2t}? ____(e)____