

本份統計考題總分為 100 分。

一、單選題(共 20 題，每題 3 分)：※ 注意：本大題請於「選擇題作答區」依序作答。

- An estimator is said to be consistent if:
 - it is an unbiased estimator
 - the variance of the estimator is close to one
 - the expected value of the estimator is known and positive
 - it is an unbiased estimator and the difference between the estimator and the population parameter grows smaller as the sample size grows larger
 - None of the above
- Which of the following statements is **NOT** correct?
 - The sample mean is an unbiased estimator of the population mean
 - The sample proportion is an unbiased estimator of the population proportion
 - The sample standard deviation is an unbiased estimator of the population standard deviation
 - The difference between two sample means is an unbiased estimator of the difference between two population means
 - None of the above
- The sample size needed to estimate a population mean within 2 units with a 95% confidence when the population standard deviation equals 8 is?
 - 62
 - 72
 - 82
 - 92
 - None of the above
- The power of a test is denoted by: (for $\theta \in H_1$)
 - β
 - $1-\beta$
 - α
 - $1-\alpha$
 - χ^2
- A robust estimator is one that:
 - is unbiased and symmetrical about zero
 - is consistent and is also mound-shaped
 - is efficient and less spread out
 - is not sensitive to moderate departure from the assumption of normality in the population
 - None of the above
- A sample of size 125 selected from one population has 55 successes, and a sample of size 140 selected from a second population has 70 successes. The test statistic for testing the equality of the population proportions equal to:
 - 0.060
 - 0.940
 - 0.977
 - 0.472
 - None of the above
- To determine whether a single coin is fair, the coin was tossed 100 times, and head was observed 60 times. The value of the test statistic is
 - 2
 - 6
 - 20
 - 60
 - None of the above
- Consider the following two independent samples:

Sample A:	16	17	19	22	47
Sample B:	27	31	34	37	40

 The value of the test statistic for a left-tail Wilcoxon rank sum test is:
 - 15
 - 20
 - 30
 - 35
 - None of the above
- In the two-way ANOVA where a is the number of factor A levels, b is the number of factor B levels, and r in the number of replicates, the number of degrees of freedom for interaction is:
 - $ab-r$
 - $abr-1$
 - $(a-1)(b-1)(r-1)$
 - $ab(r-1)$
 - None of the above
- A nonparametric method to compare two populations, when the samples are matched pairs and the data are ordinal, is the:
 - Sign test
 - Run test
 - Wilcoxon rank sum test
 - Mann-Whitney U test
 - None of the above
- The time-series multiplicative model is used for forecasting, where T_t, C_t, S_t , and R_t are respectively the trend, cyclical, seasonal, and random variation components of the time series, and y_t is the value of the time series at time t . The following estimates are obtained: $\hat{T}_t = 120$, $\hat{C}_t = 1.02$, $\hat{S}_t = .95$, $\hat{R}_t = .90$. The model will produce a forecast of:
 - 122.870
 - 102.300
 - 116.280
 - 104.652
 - None of the above

12. The trend equation for annual sales data (in millions of dollars) is given by $\hat{y}_t = 65 + 2.5t$, where $t = 1$ for 1990. The monthly seasonal index for December is 0.97. What is the forecasted sales for December of 1999?
- A. 90 B. 7.28 C. 7.5 D. 7.69 E. None of the above
13. Suppose that the estimated regression equation of 200 College of Business graduates is: $\hat{y} = 20,000 + 2000x + 1500I$ where y is the starting salary, x is the grade point average and I is an indicator variable which takes the value of 1 if the student is a computer information systems major and 0 if not. What is the average starting salary of a business administration major graduate with 3.5 grade point average?
- A. \$22,000 B. \$27,890 C. \$28,500 D. \$20,000 E. None of the above
14. Which of the following is not an advantage of multiple regression as compared with analysis of variance?
- A. Multiple regression can be used to estimate the relationship between the dependent variable and independent variables.
 B. Multiple regression handles problems with more than two independent variables easier than analysis of variance.
 C. Multiple regression handles qualitative variables better than analysis of variance.
 D. All of the above are advantages of multiple regression as compared with analysis of variance.
 E. None of the above
15. Suppose that the sample regression line of the first order model is $\hat{y} = 8 + 2x_1 + 3x_2$. What do we observe, if we examine the relationship between y and x_1 for four different values of x_2 ?
- A. Effect of x_1 on y remains the same no matter what the value of x_2
 B. Effect of x_1 on y remains the same no matter what the value of x_2
 C. Only difference in the four equations produced is the coefficient of x_2
 D. Not enough information is given to answer this question
 E. None of the above
16. Suppose that we want to model the randomized block design of the analysis of variance with, say, three treatments and four blocks. How many indicator variable should we create?
- A. 7 indicator variables B. 5 indicator variables
 C. 6 indicator variables D. 4 indicator variables E. None of the above
17. For the following regression equation $\hat{y} = 20 + 8x_1 + 5x_2 + 3x_1x_2$ which combination of x_1 and x_2 , respectively, results in the largest average value of y ?
- A. 3 and 5 B. 5 and 3 C. 7 and 4 D. 4 and 7 E. None of the above
18. If X and Y are random variables with $V(X) = 7.5$, $V(Y) = 6$ and $COV(X, Y) = 4$, then what is the value of $V(2X + 3Y)$?
- A. 133 B. 37 C. 88 D. 132 E. None of the above
19. Which of the following models might be appropriate to describe a new product that has experienced a rapid early growth rate followed by the inevitable leveling off?
- A. Autoregressive model B. Linear model for long-term trend
 C. Quadratic model for long-term trend D. Lognormal model for long-term trend
 E. None of the above
20. An investment firm has classified its clients according to their gender and the composition of their investment portfolio (primarily bonds, primarily stocks, or a balanced mix of bonds and stocks). The proportions of clients falling into the various categories are shown in the following table. Please find the probability that the employee selected is male, given that the employee has an unbalanced portfolio.

Gender	Portfolio Composition		
	Bonds	Stocks	Balanced
Male	0.18	0.20	0.25
Female	0.12	0.10	0.15

- A. 0.633 B. 0.625 C. 0.3968 D. 0.375 E. None of the above

二、計算題(共4題，每題10分)：請於試卷上依序作答，並應註明作答之大題及其題號。

1. The following data give the number of pets owned for a population of 4 families

Family	A	B	C	D
Number of Pets Owned	2	1	4	3

Samples of size 2 will be drawn at random from the population (without replacement).

- (1) Find the mean of the sampling distribution of the sample means.
- (2) Find the standard deviation of the sampling distribution of the sample means.
- (3) Find the sampling distribution of \bar{X} , and use it directly to recalculate the mean and the standard deviation of \bar{X} .

2. Partial results of one-way ANOVA for airline data are summarized in the following table:

ANOVA table

Source of Variation	SS	df	MS	F	P-value
Between Groups	212.4	3		8.305	0.00145
Within Groups	136.4		8.525		
Total	348.8				

- (1) What is the within groups degrees of freedom?
 - (2) What is the total number of degrees of freedom?
 - (3) What is the between group means square?
 - (4) What is the null hypothesis?
 - (5) At a significance level of 1%, what is the proper conclusion?
3. An avid football fan was in the process of examining the factors that determine the success or failure of football teams. He noticed that teams with many rookies and teams with many veterans seem to do quite poorly. To further analyze his beliefs he took a random sample of 20 teams and proposed a second-order model with one independent variable. The selected model is $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ where y = winning team's percentage and x = average years of professional experience. The computer output is shown below.

Predictor	Coef	StDev	T	p-value
Constant	32.6	19.3	1.689	0.1095
x	5.96	2.41	2.473	0.0243
x^2	-0.48	0.22	-2.182	0.0434

S = 16.1 R-Sq = 43.9%

ANALYSIS OF VARIANCE

Source of Variation	df	SS	MS	F	p-value
Regression	2	3452	1726	6.663	0.0073
Error	17	4404	259.059		
Total	19	7856			

- (1) What is the coefficient of determination? Explain what this statistic tells you about the model.
 - (2) Do these results allow us to conclude at the 5% significance level that the model is useful in predicting the team's winning percentage?
 - (3) Test to determine at the 10% significance level if the linear term and the x^2 term should be retained.
 - (4) Based on the result of (2) and (3), please predict the value of winning team's percentage with 12 years of professional experience.
4. Gasohol (a mixture of gasoline and methyl alcohol) sales in Iowa City have been recorded over the past 10 months as shown below.

Month	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.
Sales	75	72	81	92	90	105	112	107	110	93

- (1) Please compute the four-month moving average.
- (2) Please compute the exponentially smoothed sales with $\alpha = .8$.
- (3) Compare the results of (1) and (2) and determined which one provide the smaller MAPE.
- (4) Please forecast November's sales using the method selected in (3).