題號： 321
科目：統計學(K)
節次： 7

國立臺灣大學 112 學年度碩士班招生考試試題

題號： 321
共 6 頁之第 1 頁

第一部份：單選題 (50%)，請用 2B 鉛筆作答於答案卡；每題 5 分。

Multiple Choice: Choose the best possible answer out of the choices from the list. (50%, 5 points for each question)

1. A marketing research company conducted a survey and found that 60% of the respondents subscribed to Netflix, 30% subscribed to Disney+, and 25% subscribed to both. What is the proportion that a respondent does not subscribe to either?
   A) 0.10
   B) 0.65
   C) 0.50
   D) 0.45
   E) 0.35

2. A researcher was interested in determining whether the prices of fiction books in the bookstore were higher than if purchased online. Therefore, he randomly selected 6 fiction books and found the price for each book at the bookstore and online. He ran an analysis and the results were as follows:

```
Paired T for Bookstore - Online
            N     Mean   StDev  SE Mean
Bookstore   6   115.00   22.36    9.13
Online      6   105.83   13.20    5.39
Difference  6     9.17   13.20    5.39
95% lower bound for mean difference: -1.69
T-Test of mean difference = 0 (vs > 0): T-Value = 1.70  P-Value = 0.075
```

Based on his analysis, what can we conclude at the 0.05 level of significance?
   A) The prices are higher in the bookstore.
   B) The prices are higher online.
   C) The prices are the same in the bookstore and online.
   D) The prices are not equal in the bookstore and online.
   E) The sample size is too small to make a conclusion.

3. A professor was investigating whether gender had an effect on job category in large investment firms. She sampled 1050 employees and summarized the number of respondents in each job category by gender, as shown in the table below. Which of the following statements is true about gender and job category?

Job Category	Male	Female
Clerical / Technical	40	100
Professional Staff	360	300
Executive / Managerial	200	50

   A) Gender and job category are independent.
   B) Gender and job category are not independent.
   C) A chi-square test should be used to determine whether the association between gender and job category existed.
   D) Both A and C are correct.
   E) There was not sufficient information to determine whether gender and job category were independent.

見背面

題號： 321
科目：統計學(K)
節次： 7

國立臺灣大學 112 學年度碩士班招生考試試題

題號： 321
共 6 頁之第 2 頁

4. Which of the following yields a simple random sample?
   A) All students in a class are grouped according to their gender. A random sample of 5 is selected from the males and a separate random sample of 5 is drown from the females.
   B) The best 10 students, according to the exam scores, in a class are selected.
   C) The names of 50 students in a class are written on 50 different pieces of paper and put in a hat. The first 10 pieces of paper are selected blindly one at a time without replacing them back in the hat after shuffling the papers thoroughly.
   D) 50 students in a class are divided into 10 groups. One student is randomly chosen from each group.
   E) 50 students in a class are divided into 5 groups according to the rows that they are seated. One of the groups is randomly selected.

5. A researcher is running a regression analysis where the dependent variable is the number of weeks a worker is unemployed (Unemployed) due to a layoff and the independent variables are the age of the worker (Age) and a dummy variable for management position (Manager: $1 = $ yes, $0 = $ no). Results of the analysis are as follows.

DV: Unemployed

	Coefficient	Standard Error	t-statistic	p-Value
Intercept	-0.21	11.58	-0.02	0.985
Age	1.44	0.32	4.57	0.000
Manager	-22.58	11.35	-1.99	0.054

Which of the following is a correct statement?
   A) On average, a worker who is a year older is estimated to stay jobless shorter by approximately 0.21 weeks while holding constant the effects of the manager dummy variable.
   B) On average, a worker who is a year older is estimated to stay jobless shorter by approximately 1.44 weeks while holding constant the effects of the manager dummy variable.
   C) On average, those who are in a management position are estimated to stay jobless longer by approximately 22.58 weeks while holding constant the effects of age.
   D) On average, those who are in a management position are estimated to stay jobless shorter by approximately 22.58 weeks while holding constant the effects of age.
   E) On average, those who are in a management position are estimated to stay jobless longer by approximately 11.35 weeks while holding constant the effects of age.

6. Two samples each of size 25 are taken from independent populations assumed to be normally distributed with equal variances. The first sample has a mean of 23.9 and standard deviation of 3.6 while the second sample has a mean of 20.1 and standard deviation of 3.2. A researcher would like to test if there is a difference between the population means at the 0.05 significance level. Which of the following statement is correct?
   A) The null hypothesis is that the population means are not equal.
   B) The degrees of freedom in the analysis is 50.
   C) There is not sufficient evidence to reject the null hypothesis; the population means are equal.
   D) There is sufficient evidence to reject the null hypothesis and conclude that the two population means are different.
   E) The information provided is not enough to achieve a conclusion.

接次頁

題號： 321
科目：統計學(K)
節次： 7

國立臺灣大學 112 學年度碩士班招生考試試題

題號： 321
共 6 頁之第 3 頁

7. Random samples of two freshmen, two sophomores, two juniors, and two seniors each from four dormitories were asked to rate on a scale from 1 (poor) to 10 (excellent) the quality of the dormitory environment for studying. The results are shown in the following table:

Ratings of the two students for each dormitory

	Dormitory A		Dormitory B		Dormitory C		Dormitory D	
Freshman	7	5	8	6	9	8	9	9
Sophomore	6	8	5	5	7	8	8	9
Junior	5	4	7	6	6	7	7	8
Senior	7	4	6	8	7	5	6	7

ANOVA Table

Source of Variation	SS	df	MS	F	Sig.
Dormitory	20.3438	3	6.7813	5.8649	0.0067
Student year	10.5938	3	3.5313	3.0541	0.0587
Interaction	16.0313	9	1.7813	1.5405	0.2160
Error	18.5000	16	1.1563		
Total	65.4689	31			

Which of the following statement is correct?
A) The population mean ratings for at least two of the four dormitories are not the same.
B) There is insufficient evidence that the population mean ratings of the four student years are not the same.
C) There is insufficient evidence that interaction exists between year and dormitory ratings.
D) Both A and C are correct.
E) A, B, and C are correct.

8. For a random sample of 263 professionals, the correlation between their age and their income was found to be 0.38. You are interested in testing the null hypothesis that there is no linear relationship between these two variables against the alternative that there is a positive relationship. Which of the following statement is correct?
A) Any professional's income increases as he gets older.
B) Income would be considered as the dependent variable and age would be considered as the independent variable.
C) We can reject the null hypothesis and conclude that the population correlation coefficient is greater than zero.
D) We fail to reject the null hypothesis and conclude that the population correlation coefficient is not greater than zero.
E) The information provided is not enough to determine the statistical significance.

9. A professor would like to test the hypothesis that the average number of minutes that a student needs to complete a statistics exam is equal to 30 minutes. A Type II error would occur if the professor concludes that the average exam time is _____.
A) not equal to 30 minutes when, in reality, the average time is equal to 30 minutes
B) not equal to 30 minutes when, in reality, the average time is greater than 30 minutes
C) equal to 30 minutes when, in reality, the average time is greater than 30 minutes
D) equal to 30 minutes when, in reality, the average time is less than 30 minutes
E) equal to 30 minutes when, in reality, the average time is not equal to 30 minutes

見背面

題號： 321
科目：統計學(K)
節次： 7

國立臺灣大學 112 學年度碩士班招生考試試題

題號： 321
共 6 頁之第 4 頁

10. There are five rows of students seated in a statistics class. The following table shows the number of students in each row and the average exam score for that row. What is the average exam score for this class?

Row	Number of Students	Row Average
1	5	82
2	5	91
3	4	85
4	5	78
5	6	90

A) 85.4
B) 83.1
C) 84.7
D) 86.7
E) 82.8

接 次 頁

題號： 321
科目：統計學(K)
節次： 7

國立臺灣大學 112 學年度碩士班招生考試試題

題號： 321
共 6 頁之第 4 頁

題號： 321
國立臺灣大學 112 學年度碩士班招生考試試題
科目：統計學(K)
題號： 321
節次： 7
共 6 頁之第 5 頁

第二部份：簡答題 (50%)，請依題號順序作答，每小題作答字數不得超過 5 行，可以用中文作答。

Question I (14 points; 7 points each)

Consider the regression model $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i$. One can show that the ordinary least squares (OLS) estimator for β_1 can be written as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$

where $\hat{r}_{i1}$ is the OLS residual from the regression of x_1 on x_2. The variance of the OLS estimator $\hat{\beta}_1$ can be written as

$$\text{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n} \hat{u}_i^2 / (n-3)}{TSS_1 (1 - R_1^2)},$$

where $\hat{u}_i$ is the OLS residual from the regression of y on x_1 and x_2, $TSS_1 = \sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)^2$ is the total variation in x_1, and R_1^2 is the R-square from the regression of x_1 on x_2.

11. Will the estimator for β_1 be estimated more precisely when the sample size n grows? Explain.

12. Will the estimator for β_1 be estimated more precisely when each explanatory variable (i.e., x_1 and x_2) has a large amount of variation independent of other explanatory variables? Explain.

Question II (12 points; 6 points each)

By 2017, the workplace wellness business had ballooned into an eight-billion-dollar industry in the U.S. alone. One report suggests that half of firms with more than fifty employees offer wellness programs of some sort. The exact structures of corporate wellness programs vary, but the approach is grounded in preventative medicine. Wellness programs often involve disease screening, health education, fitness activities, nutritional advice, weight loss, and stress management. As wellness programs raise ethical questions about employers having this level of control and ownership over employees' bodies, employers cannot force employees to participate. In other words, if you work for a large company, you may have the option to participate in such a program.

While employers say that they offer these programs because they care about their employees and want to improve their quality of life, the primary rationale for implementing a wellness program is that by improving the health of its employees, a company can lower insurance costs and decrease absenteeism. But there is a fundamental question: Do they work? Meta-analyses—studies that aggregate the results of previous studies—seem encouraging. They compare employees within the same company who did take part in wellness activities with those who did not, controlling for age, gender, weight, and other characteristics. Such studies typically found significant effects that wellness programs reduce medical costs and absenteeism, generating considerable savings for employers.

13. Can we conclude that offering a wellness program had such beneficial effects? Was there something particularly ineffective in the previous studies?

14. Find a way to deal with the problem above and investigate whether offering a wellness program has such beneficial effects.

見背面

題號： 321
科目：統計學(K)
節次： 7

國立臺灣大學 112 學年度碩士班招生考試試題

題號： 321
共 6 頁之第 6 頁

Question III (24 points; 6 points each)

The following table contains the ordinary least squares (OLS) coefficient estimates of the income model, where *Income* is a dependent variable. Estimated standard errors are reported in parentheses below the coefficient estimates.

Income   = Annual income of an individual, measured in thousands of dollars
Age   = Age of an individual, measured in years
Experience = Work experience of an individual, measured in years
Female   = 1 if an individual is female, 0 otherwise

	(1)	(2)	(3)
Intercept	9.684	45.667	58.425
	(3.826)	(16.220)	(14.311)
Age	0.407	-1.458	-0.921
	(0.103)	(0.938)	(0.361)
Age Squared		0.012	0.008
		(0.004)	(0.002)
Experience			1.199
			(0.454)
Female			-57.737
			(30.090)
*Female * Experience*			-0.967
			(0.803)
R-squared	0.146	0.168	0.592
USS	5,344	5,207	2,549
Observations	80	80	80

15. According to regression (1), explain how age is related to income. Is age statistically significant in regression (1)? Test the hypothesis at a 5 percent significance level.
16. Do the results of regression (2) conflict with these results of regression (1)? Explain.
17. In regression (3), what is the marginal effect of being female on income?
18. Test whether regression (3) or regression (2) is preferable. State your hypotheses. Calculate the required test statistic, state the decision rule you use, and the inference you draw from the test.

試題隨卷繳回