

※ 注意：請於試卷內之「非選擇題作答區」作答，並應註明作答之題號。

第一大題 改錯題：下面各題目皆有一個選項不正確，請找出錯誤的選項並更正其錯誤(請寫清楚題號，每題三分，選出答案得一分，更正得兩分)。

1. Which of the following statements regarding variance and mean is incorrect?

- (A) The mean of the sum of two random variables is the sum of the means.
- (B) The mean of the difference of two random variables is the difference of the means.
- (C) The variance of the sum of two random variables is the sum of their individual variances.
- (D) If the random variables are independent, the variance of their difference is the sum of the variances.

2. Let $\text{Cov}(X, Y)$: the covariance of X and Y , $E(X)$: expected value of X , $\text{Var}(X)$: variance of X , $E(X) = u$ and $E(Y) = v$, which of the following is incorrect?

- (A) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- (B) $\text{Cov}(X, X) = \text{Var}(X)$
- (C) $\text{Cov}(cX, dY) = \text{Cov}(X, Y)$ for any constant c and d
- (D) $\text{Cov}(X, Y) = E(XY) - uv$

3. Which of the following formula regarding the correlation coefficient is incorrect?

- (A) $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$
- (B) $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y}$
- (C) $r = \frac{\sum z_x z_y}{n-1}$, where $z_x = \frac{x-\bar{x}}{s_x}$ $z_y = \frac{y-\bar{y}}{s_y}$
- (D) $M_i = aX_i + b$, $N_i = cY_i + d$ 且 $a > c > 0$, then $r_{m,n} > r_{x,y}$

4. Which of the following statements is not correct?

- (A) In a Simple Random Sample (SRS), every possible group of n individuals has an equal chance of being our sample.
- (B) Response bias arises when sampled individuals will not or cannot respond.
- (C) Stratified samples can reduce sampling variability by identifying homogeneous subgroups and then randomly sampling within each.
- (D) Cluster samples randomly select among heterogeneous subgroups that each resemble the population at large, making our sampling tasks more manageable.

5. Which of the following statements is incorrect?

- (A) The Central Limit Theorem depends crucially on the assumption of independence.
- (B) Sampling distribution is the distribution of the sample.
- (C) The standard deviation of the sample means is the population's standard deviation divided by the square root of the sample size
- (D) Regardless of the shape of the original population, the shape of the distribution of the means of all possible samples can be described by a Normal model, provided the samples are large enough.

見背面

6. Which of the following statements is incorrect?
- (A) The P-value is the probability that the null hypothesis is true.
- (B) The power of a test is the probability that it correctly rejects a false null hypothesis.
- (C) When we choose significant level α , we are setting the probability of a Type I error to α .
- (D) The only way to reduce both types of error is to collect more evidence or, in statistical terms, to collect more data.

第二大題 簡答題：

7. 實驗設計中有四個重要的原則（The Four Principles of Experimental Design）：control, randomize, replicate, and block，請舉例說明這四個原則的意義。（八分）

8. The TVBS poll which estimated that 82% of all voters believed global warming exists had a margin of error of $\pm 3\%$ with a confidence level of 95%. Suppose an environmental group planning a follow-up survey of voters' opinions on global warming wants to determine a 95% confidence interval with a margin of error of no more than $\pm 2\%$. How large a sample do they need? (六分)

9. 以下為利用「台灣社會變遷基本調查」的資料來檢定男女每月平均薪資（單位：萬元）是否有差異的t-test。請根據此電腦分析的結果回答下列問題（每小題兩分，共十分）：

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0_Male	712	4.80618	.1524203	4.06708	4.506932	5.105427
1_Female	583	3.236707	.1222607	2.952032	2.996581	3.476833
combined	1295	4.099614	.1025481	3.690305	3.898435	4.300793
diff		1.569473	.2015286		1.174114	1.964832

diff = mean(0_Male) - mean(1_Female) $t = 7.7878$
 Ho: diff = 0 degrees of freedom = 1293

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

- (1) 請寫出"男女薪資差異"的抽樣分配。
- (2) 請寫出t值(=7.7878)的計算公式
- (3) 請寫出std.Err (standard error)那一欄中0.2015286的計算公式。
- (4) 請解釋std.Err那一欄中，0.1222607, 0.1025481, 及0.2015286三個統計量的差別。
- (5) 請解釋男性(0_Male)那一系列中的0.1524203(Std Err)與4.06708(Stander Deviation)的意義，兩者有何差別？

10. 將上題中的每月平均薪資當成依變項，以虛擬變數(dummy variable)female (1=女性, 0=男性)為自變項進行簡單迴歸分析，請根據以下電腦分析結果回答下列問題(每小題兩分，共八分)：

接次頁

regress income0 female

Source	SS	df	MS			
Model	789.562521	1	789.562521	Number of obs =	1295	
Residual	16832.5873	1293	13.0182423	F(1, 1293) =	60.65	
Total	17622.1498	1294	13.6183538	Prob > F =	0.0000	
				R-squared =	0.0448	
				Adj R-squared =	0.0441	
				Root MSE =	3.6081	

income0	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	-1.569473	.2015286	-7.79	0.000	-1.964832	-1.174114
_cons	4.80618	.1352185	35.54	0.000	4.540908	5.071452

- (1). 請解釋迴歸模型中_cons與female的系數(coef.)的數值-1.569473與4.80618實際上所代表的意義。
- (2). 請寫出t=-7.79的計算公式。
- (3). 請問t=-7.79是在檢定甚麼?請寫出所欲檢定的虛擬假設(null hypothesis)。
- (4). 請問SS那一欄中，16832.5873這個數值所代表的意義為何?

第三大題 申論題：

11. 一位社會學研究生想探討近年來台灣年輕世代的社會處境。請你一起腦力激盪，根據這個主題，提出兩組具體的研究問題。請針對這兩組研究問題，提出最適合回答這些研究問題的研究方法(不限質性或量性)，並說明採用這類研究方法的理由。(共三十分)

12. 上述這位研究生設想過以Facebook或twitter上年輕世代的圖文紀錄與活動，作為研究資料。請問，這可能會遇到哪些研究倫理的問題?你覺得可以採用什麼樣的方法，來處理這些倫理爭議?(共二十分)

試題隨卷繳回