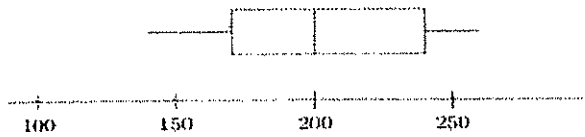


考試科目	統計學	所(組)別	統計學系	考試時間	11月7日 星期日 10:00 - 11:40
------	-----	-------	------	------	----------------------------

### 第一大題

- (4%) Which one of the following statements is not true?
  - The percentile is a measure of dispersion.
  - The coefficient of skewness based on a sample is a statistic.
  - The median of a population is a parameter.
  - The random variable is always between 0 and 1.
- (4%) The following is the box plot for the amount of time (in seconds) taken by a sample of 100 drive-through customers at McDonald's near Taipei 101.



- Which one of the following statements is most likely to be false?
- The median of the amount of time taken is 200 seconds.
  - The distribution for the amount of time taken is positively skewed.
  - The inter-quartile range for the amount of time taken is about 70 seconds.
  - The average amount of time taken is about 180 seconds.
- (4%) Let A and B be two events. Which one of the following statements is not true?
    - If A and B are independent, then  $P(A \text{ or } B) = P(A) + P(B)$ .
    - If A and B are independent, then  $P(A \text{ and } B) = P(A)P(B)$ .
    - If A and B are independent, then  $P(A|B) = P(A)$ .
    - If A and B are mutually exclusive and collectively exhaustive, then  $P(A) + P(B) = 1$ .
  - In the statistics class, 60% of the students are male and 40% are female. Also, we know 70% of the male students and 80% of the female students are Taipei residents.
    - (4%) What is the probability that a randomly selected student is either a female or a Taipei resident?
    - (4%) Suppose a randomly selected student is a Taipei resident, what is the probability that the student is male?

### 第二大題

- 請寫出二項分配(Binomial Distribution)公式, 並說明公式中符號所代表之意義。(3分)
  - 請寫出超幾何分配(Hypergeometric Distribution)公式, 並說明公式中符號所代表之意義。(3分)
  - 請說明二項分配與超幾何分配的差異點。(4分)
- 請寫出卜瓦松分配(Poisson Distribution)公式, 並說明公式中符號所代表之意義。(3分)
  - 請寫出指數分配(Exponential Distribution)公式, 並說明公式中符號所代表之意義。(3分)
  - 請說明卜瓦松分配與指數分配的差異點。(4分)

考試科目	統計學	所(組)別	統計學系	考試時間	11月7日 星期日 10:00-11:40
------	-----	-------	------	------	--------------------------

### 第三大題

1. Let  $X_1, \dots, X_n$  be a random sample from normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

If  $\mu$  is known, write down

- (1) the estimator of  $\sigma^2$ ,  
 (2) the test statistic and the critical value in order to test  $\sigma^2 = \sigma_0^2$  vs.  $\sigma^2 < \sigma_0^2$  if  $\mu$  is known.

(6%)

2. The weight of a component is uniformly distributed over the interval  $[\mu-5, \mu+5]$ .

(1) Assume a random sample of 64 components was taken. Write down the expression of the 95% confidence interval for the mean weight. (4%)

(2) Suppose a quality engineer wants to make decision about whether the mean weight is at least 20 at  $\alpha=0.05$ .

(2a) If the true mean weight is 19.5, compute the power of the test assuming the sample size is 100. (5%)

(2b) If the true mean weight is 19, and the engineer is willing to accept a 0.10 probability of making a type II error, how large a sample should be taken? (5%)

### 第四大題

For a master's thesis, Adam Psyche ran an air pollution study on automobile traffic. Air samples were taken at five different locations, and he analyzed the data (the amount of particulate matter present in the air ( $\text{mg}/\text{m}^3$ )) as a one-factor analysis of variance with five levels. However, his thesis adviser noted that the samples were taken at four different time slots. She suggested that Adam do a two-factor analysis, with one factor being location. When Adam reanalyzed the data as a two-factor design, he obtained the following ANOVA table:

Source of variation	d.f.	Sum of squares	Mean squares	F-ratio
Time	3	$C$	394.32	$G$
Location	$A$	2199.50	$E$	$H$
Error	$B$	$D$	$F$	
Total	19	3571.75		

a. What are the values for  $A, B, C, E, F, G,$  and  $H$ ? (8 pts)

b. Is there any difference in true average amount of particulate matter present in the air due to either different sampling times or different locations? Assuming that the critical values for testing the sampling times is 3.49 and is 3.26 for testing different locations. Make sure to write down your  $H_0$ s and  $H_1$ s. (6 pts)

c. Using the table above and the values obtained in (a), construct the ANOVA table that Adam got in his original one-factor analysis. (6 pts)

考試科目	統計學	所(組)別	統計學系	考試時間	11月7日 星期日 10:00-11:40
------	-----	-------	------	------	--------------------------

第五大題

1. A systems analyst wants to study whether the time to complete a certain computer programming task depends on programming experience (measured in months of experience). A sample of 15 persons were selected for the study. A simple regression model was conducted. The least squares estimates of the intercept and the slope are 19.28 and -0.445. Part of the computer output is also given below.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	10	20	15	11	11	19	11	13	17	18	16	16	17	18	10
x	24	1	10	15	17	3	20	9	3	1	7	9	7	5	20

$y$ : time to complete (minutes);  $x$ : months of experience

Source	SS	df	MS
Regression	153.41	1	
Error	16.98	13	
total	170.4	14	

- (a) (4%) Fit the data using linear model:  $y = \alpha + \beta x + e$ , where  $e \sim N(0, \sigma^2)$  is the random error. Which *ones* of the following are possible sources of the randomness in above linear model? (i) gender of the person, (ii) age of the person, (iii) average age of the 15 persons, (iv) difficulty of the task
- (b) (4%) Compute MSR and MSE.
- (c) (2%) What proportion of variation in *time to complete* can be explained by *months of experience*?
- (d) (3%) What is the estimated standard error of the regression model; that is,  $s_{y,x}$ ?
- (e) (3%) Approximately how many residuals would lie within  $\hat{y} \pm s_{y,x}$ ?
- (f) (4%) If Andy has 3 months of experience, what is the 95% prediction interval of his time to complete the task?