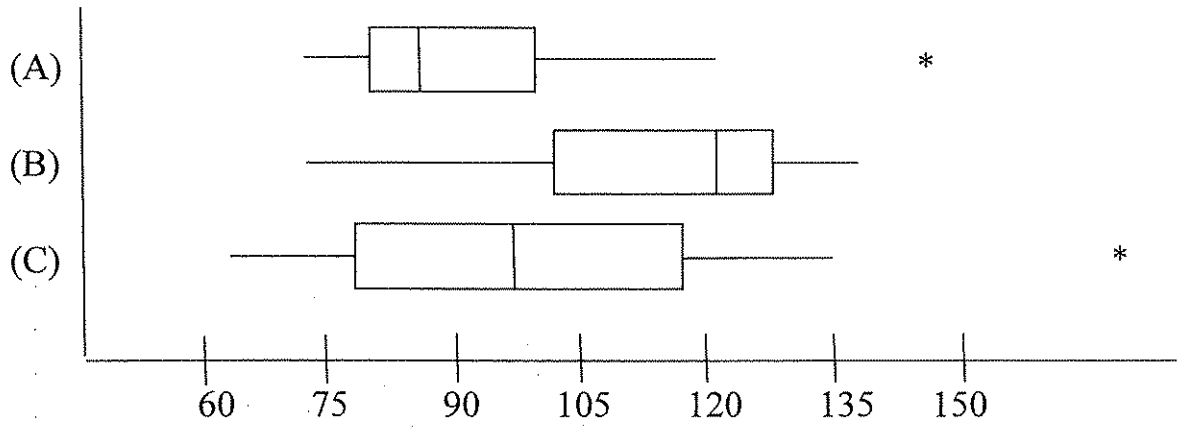
統計學系碩士班甄試考試命題紙

考試科目	統計學	所(組)別	統計學系	考試時間	104年10月31日 星期六 10：00－11：40

請注意，每一大題的作答分別寫在一份答案本上。答案紙上請清楚標示各大題及小題的題號，無法辨識者，以零分計算。每一大題皆為 **25** 分，各小題配分以 **pts** 表之。

第一大題

1. A large number fast-food restaurants with drive-through windows offer drivers and their passengers the advantages of quick service. To measure how good the service is, an organization planned a study wherein the amount of time (in seconds) taken by a sample of drive-through customers at three restaurants (A), (B), and (C) was recorded. Compare the three sets of data using the following box plots.



(1) (3pts) Which restaurant's service times display more variability?

(2) (3pts) Which restaurant's data is skewed to the left (negatively skewed)?

2. Supposed a population is divided into groups based on some characteristic so that the subjects in each group represent homogeneous behavior to the topics we wish to study.

(1) (3pts) Which of the following sample methods gives an unbiased result?

(A) simple random sampling (B) stratified random sampling

(C) both (A) & (B) (D) none of (A) & (B)

(2) (3pts) Which of the following sample methods gives a smaller variation result?

(A) simple random sampling (B) stratified random sampling

(C) cluster sampling (D) all of (A), (B), & (C) are the same

3. (3pts). Which one of the following statements is incorrect?

(A) The Chebyshev's theorem says that regardless of the shape of the distribution, at most 1-1/9 of the observations will be within 3 standard deviations of the mean.

考試科目	統計學	所(組)別	統計學系	考試時間	104 年 10 月 31 日 星期六 10：00－11：40

(B) The general rule of multiplication refers to events that are not independent is $P(A \& B)=P(B)P(A|B)$.

(C) Two events are mutually exclusive if by virtue of one event happing the other cannot happen.

(D) Events are independent if the occurrence of one event does not affect the occurrence of another event.

4. (3pts) Approximately 80% of American families have some form of life insurance coverage. Which one of the followings gives a good approximation for the probability that less than half (<50%) families are covered by life insurance among randomly selected 100 families?

(A) $\Phi((29.5)/4)$ (B) $\Phi((30.5)/4)$ (C) $1-\Phi((29.5)/4)$ (D) $1-\Phi((30.5)/4)$

5. The owner of a self-service carwash has found that customers take an average of 10 minutes to wash and dry their cars. Assuming that the self-service times tend to be exponentially distributed.

(1) (4pts) What is the probability that a customer will require less than 10 minutes to complete the job given that he/she has already started the wash for 5 minutes? Show your work.

(2) (3pts) What is the distribution for the number of cars that have been completed the wash between 1pm-2pm? Please give the name of the distribution and the value(s) of the parameter(s).

第二大題

1. (10pts) 圓周率(π=3.14159265358...)大概是我們最早接觸的無理數，歷來是數學家有興趣的研究議題，近年也受到電腦專家的青睞，開發新且快的計算方式，已知的記錄是日本科學家在 2002 年創下，可計算至小數點後面 1.2 兆位以上的有效數字。這些圓周率的有效數字，出現並無規律可遵循，請以統計的角度思考，定義「無規律」的數字應該具有哪些特性？

2. (15pts) 上述提到的「無規律」的統計特性，可由哪些統計方法檢定？（如果給定圓周率小數點之後的最前面一千位有效數字。）

第三大題

1. Multiple choice.

(1) (1pt) The sample proportion from a survey is 0.5. Which of the following intervals for the population proportion has the highest confidence?

(A) (0.1, 0.5) (B) (0.1, 0.9) (C) (0.2, 0.8) (D) (0.3, 0.7) (E) (0.5, 0.9)

(2) (1pt) In hypothesis testing,

(A) type II error will not be affected by Type I error

(B) the sum of Type I and Type II errors must equal to 1

考試科目	統計學	所(組)別	統計學系	考試時間	104年10月31日 星期六 10：00－11：40

(C) the smaller the Type I error, the smaller the Type II error will be.

(D) the smaller the Type I error, the larger the Type II error will be.

2. Monthly salary information for samples of employees of a big company is shown below.

	Bachelor Degree	Master Degree
Sample size	31	21
Sample mean (in $1,000)	38	42
Sample variance	60	80

Assume distributions of monthly salaries are about normal.

(1) (6pts) At α=0.05, test whether the variances of the two groups are equal.

(2) (8pts) At α=0.05, test whether the mean salary of employees with bachelor degree is less than that of employees with master degree.

(3) (4pts) Based on the conclusion of (b), do you agree that 學歷 determines salary? Briefly state your reason.

(4) (5pts) Assume $\sigma^2_{Bachelor} = \sigma^2_{Master} = 70$, find the 90% confidence interval for $\mu_{Bachelor} - \mu_{Master}$.

第四大題

1. With the ongoing energy crisis, researchers for the major oil companies are attempting to find alternative sources of oil. It is known that some types of shale (頁岩) contain small amounts of oil that feasibly could be extracted. 4 methods have been developed for extracting oil from shale, and the government has decided that some experimentation should be done to determine whether the methods differ significantly in the average amount of oil that each can extract from the shale. Method 4 is known to be the most expensive methods to implement, and method 1 is the least expensive, so inferences about the differences in performance of these two methods are of particular interest. 16 bits of shale were randomly subjected to the 4 methods, with the results shown in the accompanying table (the units are in liters per cubic meter).

Method 1	Method 2	Method 3	Method 4
3	2	5	5
2	2	2	2
1	4	5	4
2	4	1	5

(1) (10 pts) Assuming that the 16 experimental units were as alike as possible, write down the model and implement the appropriate ANOVA to determine whether there is any significant difference among the mean amounts extracted by the 4 methods. Use α = 0.05.

(2) (10 pts) Set up a 95% confidence interval for the difference in the mean amounts extracted by the two methods of particular interest. Interpret the result.

2. (5 pts) For a high-performance tissue machine used in processing of paper by paper mills, the following data

考試科目	統計學	所(組)別	統計學系	考試時間	104 年 10 月 31 日 星期六 10：00－11：40

was collected on machine speed x (m/min) and temperature in the drying hood y (°C). Suppose there is a linear probabilistic relationship between speed and hood temperature.

x:	1000	1100	1200	1250	1300	1400	1450
y:	220	280	350	375	450	470	500

Now $\sum x_i = 8,700$, $\sum x_i^2 = 10,965,000$, $\sum y_i = 2,645$, $\sum y_i^2 = 1,063,325$, $\sum x_i y_i = 3,384,750$, $r = \sum (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2} = 0.988$, and $r^2 = 0.976$. What does r measure? What does r^2 describe?