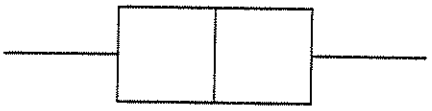
考試科目	統計學	所別		統計系	考試時間	12月4日(上)(下)午第1節 星期六

政治大學統計研究所甄試考題

共五題(每題各佔 20 分)，每題請用不同考卷繕寫

1. (a) Class A has 20 girls, class B has 15 girls and 5 boys, class C has 10 girls and 10 boys, class D has 5 girls and 15 boys, and class E has 20 boys. Which class(es) is(are) most variable (變化) in terms 'sex'? Which one(s) is/are least variable? Explain.

(b) A box and whisker plot was constructed as follows:



Which of the following statement is true? Explain.

A. The distribution is normally distributed.

B. The distribution is not skewed.

C. Mean = Median = Mode.

D. All of the above.

(c) Comment on the following statement: If a student's test score is below the class average, it means he/she did relatively poor on the exam.

考試科目	統計學		所別		統計系	考試時間	月 日 星期 上/下午 第 節

2. 兩種流行性感冒(D_1 及 D_2)的症狀不完全相同，臨床上常看到的三種症狀分別以 S_1、S_2、S_3 表示，下表為罹患這兩種流行性感冒後觀察到各症狀的比例，也就是 $P(S_j \mid D_i)$ 的機率：

$P(S_j \mid D_i)$		感冒症狀		
		S_1	S_2	S_3
感冒類型	D_1	0.15	0.6	0.15
	D_2	0.8	0.15	0.03

如果已知兩種感冒的罹患機率各為 $P(D_1)=0.6$ 及 $P(D_2)=0.4$：

(a) 計算感冒病人有症狀 S_1 的機會；

(b) 如果發現某感冒病人同時有症狀 S_1 及 S_2，試求該病人罹患 D_1 型感冒的機會（註：假設 $P(S_1 \cap S_2 \cap D_i) = P(S_1) \times P(S_2 \cap D_i)$）；

(c) 一個箱子中有 n 個球（標示為 1, 2, ..., n），隨機由箱中抽出一個球，共抽出 n 次，如果在第 i 次出現標號為 i 的球則可獲得一份獎品。試求在抽出置回(Sampling with replacement)及抽出不置回(Sampling without replacement)的假設下，至少會獲得一份獎品的機會。

考試科目	統計學	所別	統計系	考試時間	12月 4日(上/下)午第 1 節 星期六

3. A notional survey of companies included a question that asked whether the company had at least one bilingual telephone operator. The sample results of 100 companies follow (Y denotes that the company does have at least one bilingual operator; N denotes that it does not).

N N N N N **Y** N N N N N N **Y** N N N N N N N
N N **Y** N N N N N N N N N N N **Y** N N N N **Y**
N N N N N N N **Y** N N N N N N N N N N N N
Y N N N N N N N N N N N N N **Y** N N N N N
N N N N N N N **Y** N N N N N N N N N N **Y** N N

(A) Use this information to estimate with 95% confidence the proportion of the population that does have at least one bilingual operator.

(B) From the 100 observations above, count the number of companies until we find a company that does have at least one bilingual operator (i.e., count the number of "N" until we get a "Y"). If you repeat the counting process sequentially on the 100 observation above until the last "Y", you will get a sequence of numbers as follows:

6 7 10 12 5 8 13 14 12 11

Based on the data, use the method of moments to find a point estimate for the proportion of the population that does have at least one bilingual operator.

考試科目		統計學	所別		統計系		考試時間	12月4日(上/下)午第1節 星期六

4. Many female undergraduates at four-year colleges and universities switch from science, mathematics, and engineers (SME) majors into disciplines that are not science based, such as journalism, marketing and sociology. When female undergraduates switch majors, are their reasons different from those of their male counterparts? This question was investigated in ***Science Education*** (July, 1995). A sample of 335 junior/seniors undergraduates -172 females and 163 males at two large research universities were identified as "switchers," that is, they left a declared SME major for a on-SME major. Each student listed one or more factors that contributed to their switching decision.

   (a) Of the 172 females in the sample, 74 listed lack or loss of interest in SME as a major factor, compared to 72 of the 163 males. Conduct a test at .10 level of significance to determine whether the proportion of female switchers who give "lack of interest in SME" as a major reason for switching differs from the corresponding proportion of males.

   (b) Find the p-value of the test and interpret its meaning.

   (c) Experience has shown that the proportion of switchers for female and male undergraduates is in the neighborhood of .4. If the investigator wants to estimate the difference in the proportion to within .05 using a 95% confidence interval, how many female and male undergraduates must be randomly sampled?(Assume equal sample sizes)

   (d) What is the power of the test to detect a departure of .01 below the specified difference of 0.?

   (e) Draw the power curve of this test.

考試科目	統計學		所別	統計系	考試時間	12月 4日 (上)午第 1 節 星期 六 下

5.1 I ran regression of Y on variables X_1, X_2 and X_3 with an intercept (X_0), getting the following sums of squares. The error sum of squares was 1000 and the error mean square was 20. Also, I found the correlation between Y and X_3 to be 0.76.

	Parameter Estimates	(10%)
Variable	Type I SS (Sequential Sum of Squares)	Type II SS (Partial Sum of Squares)
Intercept	16328	1
X_1	5277	326
X_2	1269	1382
X_3	211	211

(a) What will be the corrected regression sum of squares _____ if I do a simple linear regression of Y on just variable X_3 (with an intercept).

(b) Compute the F test of the hypothesis that the last two regression coefficients are both 0, that is, the F test for $H_0 : \beta_2 = \beta_3 = 0$. F=_____.

(c) If possible, compute the sequential (Type I) sums of squares and error mean square (MSE) .

	Df	Type I SS
Intercept	1	16328
X_1	1	_____
X_3	1	_____
X_2	1	_____

5.2 An absent-minded professor analyzed four separate data sets using one-way (single-factor) analysis of variance (ANOVA) to compare treatment means. He calculated the four ANOVA tables, and constructed box plots for each of the four data sets. The following page is the four ANOVA tables and the four box plots. Examine the box plots and the ANOVA tables. The units of measurements are the same in all four data sets. Also, the vertical axis scales of the four box plots are equal. (10%)

(a) Determine which box plot best matches each ANOVA.

i. ANOVA A matches Box Plot _____. Explanation.

ii. ANOVA B matches Box Plot _____. Explanation.

iii. ANOVA C matches Box Plot _____. Explanation.

iv. ANOVA D matches Box Plot _____. Explanation.

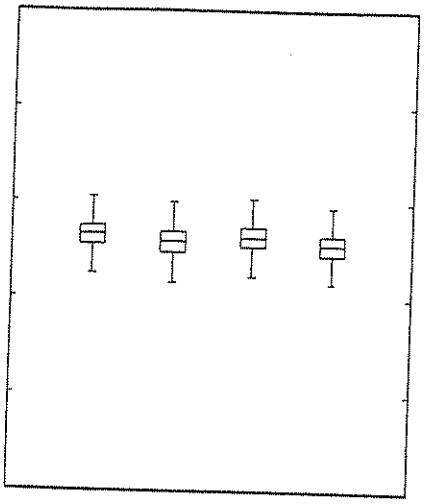
(b) Treatment sample sizes (n) in ANOVA B are all equal. What is n?_____. Explanation

考試科目	統計學	所別	統計系	考試時間	12 月 4 日 上/下 午第 1 節 星期 六

ANOVA Table A

Source of Variation	df	Sums of Squares	Mean Squares	F
Treatments	3	89.43680	29.81227	0.2934
Error	52	5284.42905	101.62364	
Total	55	5373.86585		

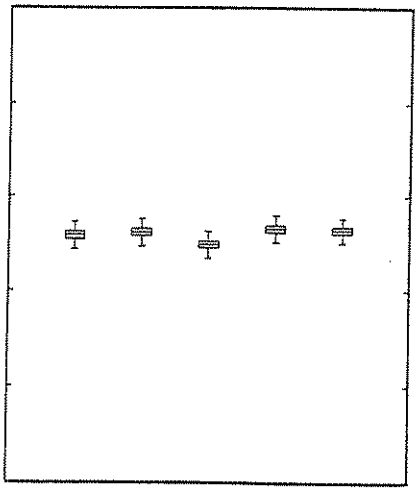
Box Plot 1



ANOVA Table B

Source of Variation	df	Sums of Squares	Mean Squares	F
Treatments	3	4975.47354	1658.49118	318.8538
Error	52	270.47359	5.20142	
Total	55	5245.94713		

Box Plot 2



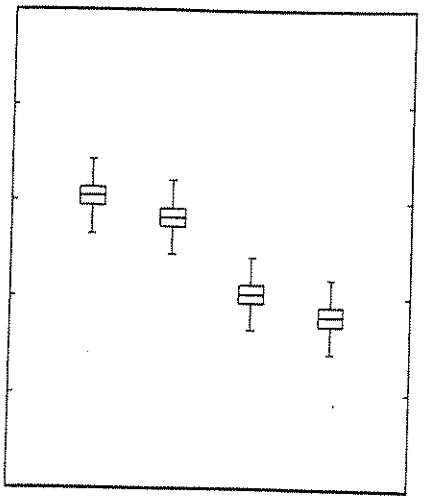
ANOVA Table C

Source of Variation	df	Sums of Squares	Mean Squares	F
Treatments	4	76.79411	19.19853	26.5560
Error	65	46.99135	0.72294	
Total	69	123.78546		

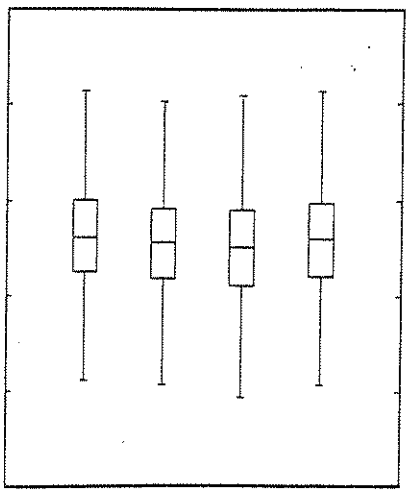
ANOVA Table D

Source of Variation	df	Sums of Squares	Mean Squares	F
Treatments	3	26.03820	8.67940	1.2087
Error	52	373.41472	7.18105	
Total	55	399.45292		

Box Plot 3



Box Plot 4



備考	試題隨卷繳交