國立政治大學
111 學年度

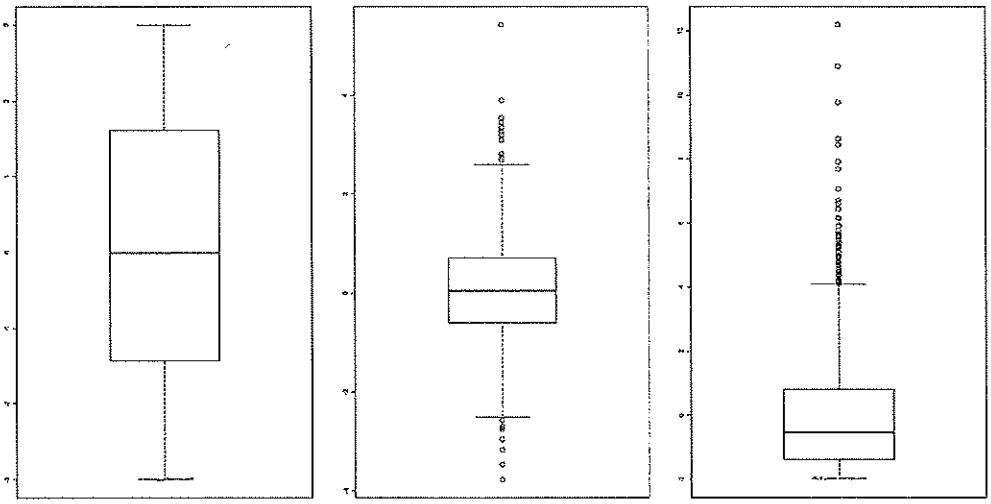
統計學系碩士班甄試考試試題

考試科目	統計學	所(組)別	統計學系	考試時間	110 年 11 月 13 日 星期六 10：00－11：40

注意: 本科目有四份試題，第一份試題答案請寫在第一卷答案本上，依此類推。每份試題配分為 25 分，合計共 100 分。

第一份試題（共 25 分，答案請寫在第一卷答案本上）

(1) 請參考下列三組資料的 Boxplot：



第一組　　第二組　　第三組

(a)（9 分）請根據圖形分佈，猜測這三組資料分別來自於什麼分配？

(b)（6 分）請說明 Boxplot 圖形的繪製方法，以及如何根據判斷離群值、這些離群值出現機率大約等於多少？

(2) 以下是大樂透「特別號」歷年出現次數：

號碼	1	2	3	4	5	6	7	8	9	10
次數	29	41	41	41	36	48	41	42	50	47
號碼	11	12	13	14	15	16	17	18	19	20
次數	38	35	42	43	39	28	25	45	40	30
號碼	21	22	23	24	25	26	27	28	29	30
次數	46	32	29	49	32	44	25	37	43	44
號碼	31	32	33	34	35	36	37	38	39	40
次數	46	32	46	27	33	38	46	38	42	43
號碼	41	42	43	44	45	46	47	48	49	50
次數	47	45	42	30	41	33	28	34	47	50

備考：試題隨卷繳交

統計學系碩士班甄試考試試題

考試科目	統計學	所(組)別	統計學系	考試時間	110 年 11 月 13 日 星期六 10：00－11：40

根據號碼所屬範圍，分別統計各組出現總次數，如下表：

號碼範圍	1～50	1～10	11～20	21～30	31～40	41～50	3 的倍數
出現次數	共 1950	416	365	381	391	397	668

(a)（5 分）有人認為特別號是 3 的倍數機會比較高，請提出自己的看法。

(b)（5 分）如果將號碼分成五個區間：1～10、11～20、21～30、31～40、41～50，以此作為特別號出現是否均勻的依據，請說明分析結果。

（**註**：上述請以顯著水準$\alpha = 0.05$ 為標準做檢測。）

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of X^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81

第二份試題（共 25 分，答案請寫在第二卷答案本上）

(1)（6 分）About 20% of the population has a non-fatal disease, but many of them do not know. A scan test had been designed to detect that disease. For a person actually has the disease, the probability that a positive test result, indicating the presence of the disease, is 0.6; while for a person actually does not have the disease, the probability of a positive result is 0.3. A person got a negative test result, should he/she be confident about not having the disease? Why?

(2)（7 分）An electric board contains 4 chips. The board functions if no more than two chips fails. Suppose the lifetime of each chip is independently and uniformly over 1 to 4 years. Find the probability that the board's lifetime is at least 3 years.

(3) The monthly water usage for each family is normally distributed with $\mu = 22\ m^3$ and $\sigma = 4\ m^3$.

(a)（5 分）To save the water usage, it is decided to make families with the top 20% usage pay more. What is the maximum water usage in order not to pay more?

(b)（7 分）Mr. A and Mr. B are next door neighbors. If one family uses more water, the other family will have fewer water to use, and the correlation between their monthly water usages is −0.2.
What is the probability that their monthly total water usage is at least 55 m^3?

統計學系碩士班甄試考試試題

考試科目	統計學	所(組)別	統計學系	考試時間	110 年 11 月 13 日 星期六 10：00－11：40

第三份試題（共 25 分，答案請寫在第三卷答案本上）

多選題：選出所有正確的敘述 (Select all of the following statements which you believe to be true)

(1)（5分）When doing a significance test, a student gets a p-value of 0.003. This means that:

(A) Assuming H_0 were true, this sample's results were an unlikely event.
(B) 99.97% of samples should give results which fall in this interval.
(C) We reject H_0 at any reasonable alpha level.
(D) The p-value is the probability that the null hypothesis is true.
(E) The probability of making a Type I error is 0.003.

(2)（5分）A political poll of Americans was conducted to investigate their opinions on gun control. Each person was asked if they were in favor of gun control or not in favor of gun control - no respondents were removed from the results. The survey found that 25% of people contacted were not in favor of gun control laws. These results were accurate to within 3 percentage points, with 95% confidence.

(A) The 95% confidence interval is approximately from (22% to 28%).
(B) We are 95% confidence that the true proportion of people not in favor of gun control is within 3 percentage points of 25%.
(C) In approximately 95% of polls on this issue, the confidence interval will include the sample proportion.
(D) If another poll of similar size were taken, the percentage of people IN FAVOR of gun control would likely range from 72% to 78%.
(E) To narrower the confidence interval, we should decrease the sample size.

(3)（5分）A 95% confidence interval for the mean number of televisions per household in Taiwan is (1.15, 4.20).

(A) We are 95% confident that the true mean number of televisions per household in Taiwan is between 1.15 and 4.20.
(B) 95% of all samples should have x-bars between 1.15 and 4.20.
(C) 95% of all households in Taiwan have between 1.15 and 4.20 televisions.
(D) Of 100 intervals calculated the same way (95%), we expect 95 of them to capture the population mean.
(E) Of 100 intervals calculated the same way (95%), we expect 100 of them to capture the sample mean.

(4)（5分）In a study of workplace bullying in a company, staff were asked whether workplace bullying had affected their working environment. Staff who had been bullied had lower levels of job satisfaction (mean 10.5 [SD 2.7] vs. 12.2 [2.3], $P < 0.001$) and higher levels of job-induced stress (mean 22.5 [SD 6.1] vs. 16.9 [5.8], $P < 0.001$) than those who had not been bullied.

考試科目	統計學	所(組)別	統計學系	考試時間	110 年 11 月 13 日 星期六 10：00 – 11：40

(A) A suitable null hypothesis for this would be that staff in the population who have been bullied at work have similar levels of job satisfaction to staff in the population who have not been bullied at work.

(B) There is a significant difference in the levels of job satisfaction between those who had and those who had not been bullied

(C) The p-value for the comparison of job-induced stress was less than 0.001. This means that we cannot reject the null hypothesis at the 5% level of significance.

(D) P-value < 0.001 means that there is a less than 1 in 1000 chance that we would have obtained these results, or more extreme results, if the alternative hypothesis was true.

(E) We can conclude that workplace bullying causes lower levels of job satisfaction and higher levels of job-induced stress.

(5)（5分）The following statements are about Type I error, Type II error, and hypothesis testing.

(A) When a new drug is created, the pharmaceutical company must subject it to testing before receiving the necessary permission from the Food and Drug Administration (FDA) to market the drug. Suppose the null hypothesis is "the drug is unsafe." The Type II Error is "not to conclude the drug is safe when, in fact, it is safe."

(B) A COVID-19 test gave a negative PCR result for a woman who in fact has an infection with COVID-19. This is an example of making a Type II error.

(C) A null hypothesis H_0 was rejected at level $\alpha = 0.10$, but we will fail to reject H_0 at level $\alpha = 0.05$.

(D) Suppose the p-value for a right-tailed test of population proportion π was 0.037. The p-value will be 0.074 if we change the alternative hypothesis to $Ha: \pi \neq 0.8$

(E) We can construct a confidence interval for a mean μ to test the hypothesis H_0: $\mu \geq 2$ vs $Ha: \mu < 2$.

第四份試題（共 25 分，答案請寫在第四卷答案本上）

When we have repeated measurements for different predictor variables X_j, it is possible to use the F test for determining whether a specific type of regression function (here is the simple linear regression) adequately fits the data (the so-called lack of fit test). Suppose that data can be expressed in the form:

$$\{(X_j, Y_{ij}): i = 1, \ldots, n_j; j = 1, \ldots, c\}, \text{ where } c > 2.$$

(1)（4 分）What are the assumptions of this lack of fit test?

(2)（2 分）What is the full model used for this lack of fit test?

(3)（2 分）What is the reduced model used for this lack of fit test?

(4)（2 分）What is the null hypothesis for this lack of fit test?

統計學系碩士班甄試考試試題

考試科目	統計學	所(組)別	統計學系	考試時間	110 年 11 月 13 日 星期六 10：00 – 11：40

(5)（10 分）The Growth rate data are available on the effect of dietary supplement on the growth rates of rats. Here X = dose of dietary supplement and Y = growth rate. The following table presents the data in a form suitable for the analysis ($c = 6$, $n = 12$). Construct a general ANOVA Table (including Source of Variation, Sum of Square, Degree of Freedom, Mean Square and F statistics) for testing lack of fit of a simple linear regression function.

		$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
Replicate		X_1 =10	X_2 = 15	X_3 = 20	X_4 = 25	X_5 = 30	X_6 = 35
Y_{ij}	$i=1$	73	85	90	87	75	65
	$i=2$	78	88	91	86		63
	$i=3$				91		
n_j		2	2	2	3	1	2
Mean $\bar{Y}_j$		75.5	86.5	90.5	88	75	64

(6)（5 分）State the test statistics, decision rule and your conclusion. (The level of significance is 5% for all j)

[Hint: Suppose we have the error sum of squares for the reduced model (SSE(R)) = 891.73, regression sum of squares (SSR) = 204.27, total sum of squares (SSTO) = 1096.00, the critical values of F distributions are $F(0.95; 5, 5) = 5.050$, $F(0.95; 6, 4) = 6.163$, $F(0.95; 4, 6) = 4.534$, $F(0.95; 1, 10)$, $F(0.95; 10, 1) = 241.881$, $F(0.95; 2, 10) = 4.103$, $F(0.95; 2, 9) = 4.256$, $F(0.95; 2, 8) = 4.459$; and $\hat{Y}_{ij} = 92.003 - 0.498\, X_j$]