

考試科目	統計學	所(組)別	統計學系	考試時間	108年10月26日 星期六 10:00-11:40
------	-----	-------	------	------	-------------------------------

注意事項：本科目有四份試題，第一份試題之答案請寫在第一卷答案本上，餘類推。每份試題配分為 25 分，合計 100 分。

第一份試題(答案請寫在第一卷答案本上)

1. (17%) 下表為某年度大學指定考試，國文、英文、物理三科成績統計表：

	最小值	12%	25%	50%	平均數	75%	88%	最大值	標準差
國文	0	27	34	44	43.56	53	60	93	13.88
英文	0	8	16	34	36.68	56	69	98	23.88
物理	0	4	12	29	34.36	56	71	100	25.97

- (a) 試以上述統計數值，描述這三科成績的分佈特性。(例如：某一科成績是否左右對稱，或是成績接近常態分配等。)
- (b) 以上述表格提供的數據繪製 Boxplot，並說明圖形顯示的資訊是否和 (a) 小題的結果一致；另外，判斷這三科是否有離群值(Outlier)。
- (c) 說明什麼是「經驗法則」，再以上表結果驗證這個法則。

2. (8%) 政治大學統計系同學針對政治大學的用餐環境，執行了問卷調查，其中有兩個問項是：

● 請問您平均一週到現代連鎖餐飲店用餐次數：

__1. 0 次 __2. 1~3 次 __3. 3~6 次 __4. 6~10 次 __5. 10 次以上

● 請問您覺得現代連鎖餐飲店的價格如何？

__1. 高很多 __2. 高一些 __3. 差不多 __4. 低一些 __5. 低很多

隨機調查 400 位政大學生，整理得出以下表格，計算出兩者的相關係數 (Pearson Correlation) 為 0.08，兩者間似乎不相關。說明這種分析方法是否適合，如果不適合，又可以採用什麼方法判斷兩個問項間是否相關？

第 2 題：價格

第 1 題 ； 次 數		1	2	3	4	加總
	1	34	47	4	2	87
	2	67	199	14	0	280
	3	3	22	2	0	27
	4	1	3	0	0	4
	5	1	1	0	0	2
	加總	106	272	20	2	400

考試科目	統計學	所(組)別	統計學系	考試時間	108年10月26日 星期六 10:00-11:40
------	-----	-------	------	------	-------------------------------

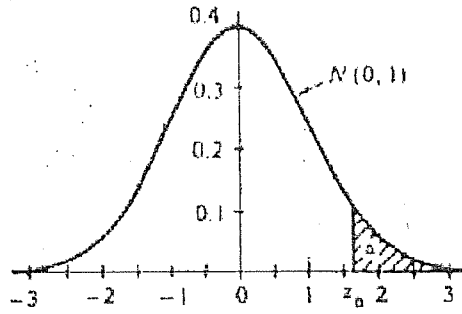
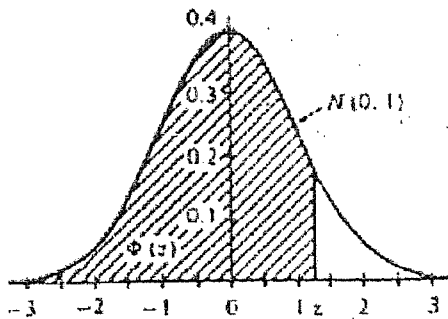
第二份試題(答案請寫在第二卷答案本上)

$(X_1, X_2, X_3, X_4, X_5, X_6)$ is a random sample, and the mean and variance of X_i are μ and σ^2 , respectively.

Let $Y_1 = (X_2 - X_1)^2 / 2, Y_2 = (X_4 - X_3)^2 / 2, Y_3 = (X_6 - X_5)^2 / 2$.

- (1) (10%) Derive the expectation of Y_i ($E(Y_i)$), $i=1, 2, 3$.
- (2) (5%) If $Y_i > E(Y_i)$ then $I_i=1$; otherwise $I_i=0$, $i=1, 2, 3$. Let $M = \sum_{i=1}^3 I_i$, derive the distribution of M .
- (3) (10%) Suppose that $X_i \sim N(\mu, \sigma^2)$, find the probability of $(Y_i > E(Y_i))$, $i=1, 2, 3$.

附表:(1)Z值表



$$P(Z \leq z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw$$

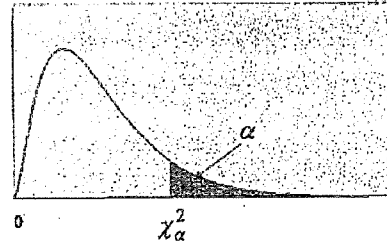
$$[\Phi(-z) = 1 - \Phi(z)]$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830

考試科目	統計學	所(組)別	統計學系	考試時間	108年10月26日 星期六 10:00-11:40
------	-----	-------	------	------	-------------------------------

(2) 卡方分配臨界值表

$$P(\chi^2 > \chi^2_\alpha) = \alpha$$



df.	$\chi^2_{0.995}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$
1	0.0000393	0.0009821	0.0039322	0.0157907	2.705541	3.841455	5.023903	6.634891
2	0.0100247	0.0506357	0.1025862	0.2107208	4.605176	5.991476	7.377779	9.210351
3	0.0717235	0.2157949	0.3518460	0.5843755	6.251394	7.814725	9.348404	11.3449
4	0.206984	0.484419	0.710724	1.063624	7.779434	9.487728	11.1433	13.2767
5	0.411751	0.831209	1.145477	1.610309	9.236349	11.0705	12.8325	15.0863
6	0.675733	1.237342	1.635380	2.204130	10.6446	12.5916	14.4494	16.8119
7	0.989251	1.689864	2.167349	2.833105	12.0170	14.0671	16.0128	18.4753
8	1.344403	2.179725	2.732633	3.489537	13.3616	15.5073	17.5345	20.0902
9	1.734911	2.700389	3.325115	4.168156	14.6837	16.9190	19.0228	21.6660
10	2.155845	3.246963	3.940295	4.865178	15.9872	18.3070	20.4832	23.2093
11	2.603202	3.815742	4.574809	5.577788	17.2750	19.6752	21.9200	24.7250

第三份試題(答案請寫在第三卷答案本上)

The World Series ends when one of the two teams playing has won four out of seven games. (七場比賽先贏四場者為勝隊)

- (a) (12%) If it is assumed that each World Series game is an independent event and that the probability of Team A's winning any particular game is p , find the probability distribution of X , the number of games in the Series.
- (b) (13%) Listed below is the length distribution of World Series competition for the 50 years from 1926 to 1975.

Number of games, X	Number of years
4	9
5	11
6	8
7	22
	50.

Test whether these data are compatible with the model that each World Series game is an independent Bernoulli trial with $p=0.5$.

(Note: Make sure your probabilities sum to 1. If they do not, your answer to (a) must be wrong. If you cannot solve (a) correctly, use for part (b) $P(X=4)=0.1$, $P(X=5)=0.2$, $P(X=6)=0.3$, $P(X=7)=0.4$, although these are not the correct probabilities.)

考試科目	統計學	所(組)別	統計學系	考試時間	108年10月26日 星期六 10:00 - 11:40
------	-----	-------	------	------	---------------------------------

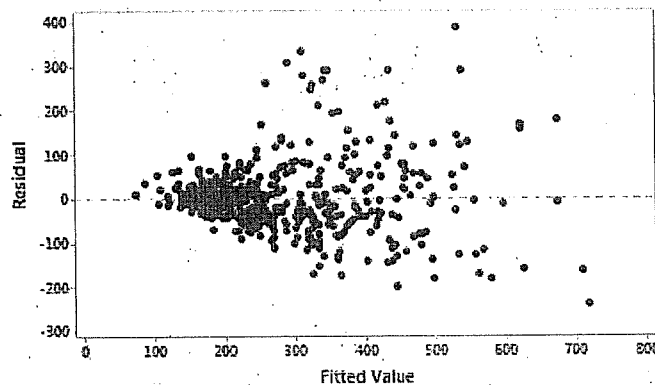
第四份試題(答案請寫在第四卷答案本上)

單選題 (每題 3 分, 共 15 分)

Q1) Which of the following statement is correct when the ordinary least squares method is used for linear regression based on sampled data?

- _____ (A) The coefficients can be estimated even when there exists collinearity among independent variables (B) The coefficients can be estimated for any sample sizes (C) The least squares regression model provides the best prediction for unobserved data (D) The least squares estimates for the coefficients are the same as the maximum likelihood estimates under the Gaussian-noise assumption (E) The coefficients cannot be estimated when the relationship between dependent and independent variables is nonlinear

Q2) The following residual plot obtained by fitting a multiple linear regression model apparently violates the regression assumption.



Which of the following is the best strategy to deal with the problem?

- _____ (A) Exclude some independent variables and fit the model again (B) Fit a higher order polynomial model (C) Transform the response variable Y (D) Remove outliers or high-leverage data (E) None of the above is appropriate.

Q3) Which of the following assumptions is not necessary when performing the one-way ANOVA?

- _____ (A) All populations are normally distributed (B) The sample sizes of all populations must be equal (C) All population variances must be equal (D) All populations are independent (E) None of the above

The following ANOVA table is the output of regression analysis with a dependent variable Y and an independent variable X .

Source	df	SS	MS	F
Regression	1	300	---	---
Error	9	600	---	
Total	10	900		

(The upper critical values of the F distribution are: $F_{0.05}(1, 9) = 5.117$, $F_{0.01}(1, 9) = 10.561$)

考試科目	統計學	所(組)別	統計學系	考試時間	108年10月26日 星期六 10:00-11:40
------	-----	-------	------	------	-------------------------------

Answer Q4) and Q5).

Q4) If $\hat{Y} = a + bX$ is the least squares regression equation associated with the ANOVA output and $b < 0$, what is the value of (sample) correlation coefficient between X and Y ?

- _____ (A) -0.577 (B) 0.577 (C) 0.33 (D) -0.33 (E) 0.50

Q5) Continuing with Q4), which of the following statements is not correct?

- _____ (A) The standard error of estimate for the residual is 8.165 (B) The proportion of variation in Y that can be explained by X is 0.33 (C) The linear relationship between X and Y is highly significant (D) Adding a quadratic term X^2 in the model will never reduce the R^2 (E) None of the above

填充題 (每題 5 分, 共 10 分)

Consider a dependent variable Y and an independent variable X , each has 15 sampled observations.

Suppose we know:

The sample means $\bar{X} = 7.6$, $\bar{Y} = 12.3$, sample standard deviations $S_X = 1.58$ and $S_Y = 1.85$, and the sample covariance $S_{XY} = 2.55$.

Answer Q6) and Q7).

Q6) Compute the least squares regression equation, $\hat{Y} = a + bX =$ _____.

Q7) Continuing with Q6), what is the unbiased estimate for the "standard error of residuals" under the ordinary regression assumptions? Answer = _____.