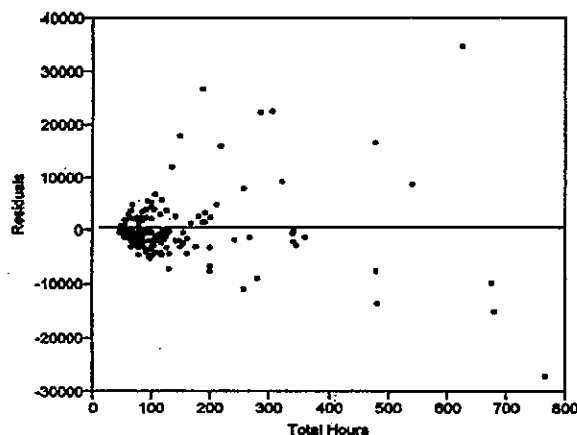


❖ 注意：請標明題號並依序作答

本試卷附有 Chi-square 分配表。

第一大題 單選題 (每題 2 分，共 30 分)

1 A construction contractor is involved in a wide variety of construction projects. The operations manager wants to investigate how the *Total Hours* of labor (design, engineering, modeling, simulation, construction, software support, etc.) required for a project is related to the *Total Cost* of completing the project. Based on data collected over many projects, the data was used to determine a predicting equation for the simple regression model: $TOTAL\ COST = F + V \times (TOTAL\ HOURS) + \epsilon$, where F and V are the fixed and variable costs respectively. After determining the predicting equation, a scatterplot of Residuals vs. *Total Hours* was determined as given below:



Which of the following statement are an appropriate interpretation of these results?

- (a) The errors are heteroscedastic.
 - (b) The similar variances condition for simple regression does not appear to be satisfied by the data.
 - (c) Prediction intervals for small values of *Total Hours* would tend to be too narrow.
 - (d) Confidence intervals for the slope of the line should still be considered reliable.
 - (e) Both (a) and (b) are appropriate interpretations.
 - (f) All of (a) – (d) are appropriate interpretations.
- 2 A regional manager for a chain of discount tire stores collected data from 30 stores over a 2-year period (24 months). The data consist of the monthly amount, A, of dollars spent on advertising by each store, and the dollar amount of sales revenue, R, for each store during the month. The manager plans to use this data to investigate the association between these two variables. The computations result in a correlation of $r = .85$. Which of the following is an appropriate interpretation of the correlation of $r=0.85$?
- (a) If A increases by \$100, we expect R to increase by \$85.
 - (b) The value for R will always be 85% of the value for A.
 - (c) An increase in A is always accompanied by an increase in R.
 - (d) The units on r are (dollars of sales revenue) per (dollar spent on advertising).
 - (e) Both (c) and (d) are appropriate interpretations.
 - (f) None of (a) - (d) is an appropriate interpretation of the correlation.
- 3 A human resources manager at a corporation has collected data for a large sample of employees. The data consist of: X=the average number of hours the employee exercised each week for the previous year, and Y=the annual dollar amount of health care claims made by the employee for the year. The results are summarized as follows:

	X	Y
Mean	10 hrs	\$5000
Standard deviation	2 hrs	\$800

見背面

Correlation $r = -0.80$

Please answer the following questions 3.1 and 3.2:

- 3.1 If an employee exercises an average of $X=13$ hours each week, complete the following: "The number of dollars in health care claims for this employee would be predicted to be _____ dollars."
- (a) \$4040 (b) \$5960 (c) \$3800 (d) \$6200 (e) \$3080
- 3.2 Using the information given, calculate the slope of the line of best fit.
- (a) -0.002 (b) 0.002 (c) -320 (d) 320
- 4 Events R and S are defined on a sample space. If $P(R)=0.2$ and $P(S)=0.5$, which of the following statements is true?
- (a) If R and S are mutually exclusive, then $P(R \cup S)=0.10$.
 (b) If R and S are independent, then $P(R \cup S)=0.6$.
 (c) If R and S are mutually exclusive, then $P(R \cap S)=0.7$.
 (d) If R and S are mutually exclusive, then $P(R \cup S)=0.6$.
 (e) Both (a) and (c) are true statements.
 (f) None of (a)-(d) is true statement.
- 5 X, Y, Z are independent and $E(X)=5$, $E(Y)=7$, $E(Z)=4$, $\text{Var}(X)=2$, $\text{Var}(Y)=3$, $\text{Var}(Z)=1$, the mean and variance of $X+2Y+3Z$ is _____ and _____, respectively.
- (a) 31, 11 (b) 31, 23 (c) 31; 14 (d) 10.3; 11 (e) 10.3; 23 (f) 10.3; 14
- 6 A machining process produces screws for which the population proportion of screws with defective threads is .10 (10%). A new laser-enhanced process has become available and the developer claims that the population proportion of screws with defective threads will be less than .10. To test the validity of this claim, 900 screws produced by the new process are selected at random and the sample proportion of defectives is computed. This result will be used to make a decision as to whether or not the manufacturer should invest in the new process. Please answer the following questions 6.1 and 6.2.
- 6.1 Which of the following should be used as the alternative hypothesis?
- (a) $p > 0.10$ (b) $p < 0.10$ (c) $p = 0.10$ (d) $p \geq 0.10$ (e) $p \leq 0.10$
- 6.2 $\hat{p} = 0.067$ is determined. If a Type II error is made in carrying out this test, what are the consequences to the manufacturer?
- (a) They will purchase the new process when it is no better than the current process being used.
 (b) They will not purchase the new process when it is no better than the current process being used.
 (c) They will not purchase the new process when it is no better than the current process being used.
 (d) They will not purchase the new process when it is an improvement over the process currently being used.
 (e) A Type II error cannot be made if $\hat{p} = .076$ as given in the question statement.
- 7 An insurance agent has selected a sample of drivers that she insures whose ages are in the range from 16-42 years old. For each driver, she records the age of the driver and the dollar amount of claims that the driver filed in the previous 12 months. A scatterplot showing the dollar amount of claims as the dependent variable and the age as the predictor shows a linear trend. The OLS regression line is determined to be: $\hat{y} = 3715 - 75.4x$. A plot of the residuals versus age of the drivers showed no pattern. r^2 is reported to be 0.822 and the standard deviation of the residuals (also known as the *root mean squared error*, RMSE) to be 312.1. Please answer the following questions 7.1-7.4.
- 7.1 Using the fitted line given above to estimate the dollar amount of claims for a driver whose age is 55 years would provide a prediction that is unreliable because it is an _____.
- (a) Unsolvable problem
 (b) Extended result
 (c) Extrapolation
 (d) Extorted point
 (e) None of (a)-(d) is correct.

- 7.2 What percentage of the variation in the dollar amount of claims is due to factors other than age?
- (a) 82.2%
 - (b) 75.4%
 - (c) 17.8%
 - (d) 31.21%
 - (e) Unable to determine with the given information.
- 7.3 A driver in the data set whose age is 25 years had a residual of -\$150 using the fitted line given above. Did the regression line overestimate or underestimate the driver's dollar amount of claims?
- (a) Overestimate, because the residual is negative and therefore the data value fell below the regression line.
 - (b) Overestimate, because the residual is negative and therefore the data value fell above the regression line.
 - (c) Underestimate, because the residual is negative and therefore the data value fell below the regression line.
 - (d) Underestimate, because the residual is negative and therefore the data value fell above the regression line.
 - (e) Unable to determine with the given information.
- 7.4 The histogram summarizing the residuals is reasonably symmetric around zero and bell-shaped. Using the Empirical Rule, complete the sentence: Approximately 68% of the dollar amounts of claims are within _____ dollars of the regression line.
- (a) $\pm 1(312.1)$
 - (b) $\pm 2\left(\frac{312.1}{\sqrt{12}}\right)$
 - (c) $\pm 1\left(\frac{312.1}{\sqrt{12}}\right)$
 - (d) $\pm 2(312.1)$
 - (e) Unable to determine with the given information
- 8 The inevitable difference between the mean of a sample and the mean of a population based on chance alone is a _____.
- (a) Type I error
 - (b) Probability
 - (c) Sampling error
 - (d) Random sample
 - (e) None of (a)-(d) is correct.
- 9 Researchers deal with the uncertainty in estimating the standard error of mean with the sample standard deviation by using which of the following distribution:
- (a) Normal
 - (b) Z
 - (c) Unsymmetrical
 - (d) t
- 10 When increasing the level of confidence, we:
- (a) Lose confidence but gain precision.
 - (b) Lose precision but gain confidence that our answer interval is accurate.
 - (c) Lose precision and lose confidence that our answer interval is accurate
 - (d) None of (a)-(d) is correct.

第二大題 計算與簡答題(共 20 分)

1. (共 12 分) All new drugs must go through a drug study before being approved by the U.S. Food and Drug Administration. A drug study typically includes clinical trials whereby participants are randomized to receive different dosages as well as a placebo but are unaware of which group they are in. To control as many factors as possible, it is best to assign participants randomly but homogeneously across the treatments. Consider the following arrangement for homogeneity with respect to gender (# of participants in the cells):

Treatment	10-mg Drug	20-mg Drug	Placebo
Female	54	56	60
Male	32	27	26

Please test if at the 0.01 level of significance, is the distribution of drug the same for both genders?

2. (共 8 分) The director of an advertising agency is concerned with the effectiveness of a television commercial.
- 1) What null hypothesis is she testing if she commits a type I error when she erroneously says that the commercial is effective? (4 分)
 - 2) What null hypothesis is she testing if she commits a type II error when she erroneously says that the commercial is effective? (4 分)

第三大題 申論題 (共 50 分)

1. 請先詳讀研究問題的背景說明，再針對研究問題進行研究設計。(共 35 分)

台灣目前大學的招生方式主要是採用「繁星推薦」、「個人申請」和「考試入學」三種模式。「個人申請」的考生均須參加「學科能力測驗」(學測)，學科能力測驗包括國文、英文、數學、社會、自然等五科，成績均採級分制。採用「個人申請」招生方式的大學校系，在選取學生時，除了以學測成績進行篩選，可能還會採用考試成績以外的項目，如書面審查、小論文筆試、面試等方式進行甄試。各校系自訂申請條件和錄取方式，甄試項目及各項目佔分比重，各不相同，即使同一學校不同學系(如台大 54 個設有學士班的學系)之間也有差異。

相較之下，採用「考試入學」的校系，則以學測、指定科目考試，或其他考試(如術科)的測驗成績來選取學生。指定科目考試包括國文、英文、數學甲、數學乙、物理、化學、生物、地理、歷史、公民與社會等十科，成績均採百分制。各校系可指定考生選考 3-5 科。

近年來，「個人申請」佔大學入學錄取人數的比率，不斷擴充，「考試入學」的比率則持續下降。2004 年「考試分發」佔大學錄取人數比率為 80.5%，2014 年則降為 37.6%。「個人申請」和「考試入學」佔錄取人數比率的消長，引起很多辯論。支持「個人申請」擴大比率的人認為，這種招生管道不以考試成績為選才的唯一標準，可以從多元資料看到學生的學習歷程與動機、特殊能力、社團活動和其他課外表現，因此有助於大學校系適性選才，也能鼓勵學生多元表現、均衡學習，發展學科知識以外的能力。反對者則指出，「個人申請」還是以學測成績來選取學生，並無法鼓勵學生多元能力的表現。即使「個人申請」會以考試成績以外的甄試項目來挑選學生，但是，家庭社經地位比較高的學生，比較有資源可以參與課外活動，發展學科測驗以外的能力，在面試時表現也比較好，因此，「個人申請」管道有利於家庭社經地位高的學生，這種招生管道的擴大，無異是增加社會不平等。

這些爭辯，各自言之成理，但都沒有堅實的證據來佐證。「個人申請」是否只以學測成績來選取學

生？與「考試入學」比較，「個人申請」這種管道是否確實能夠以考試成績以外的表現和學習經驗來挑選學生？是否以「個人申請」管道入學的學生，比較可能來自社經地位高的家庭？這是有待探究的問題。

假設你想用問卷調查和量化分析的方法，以台大大學部學生（共 11 個學院、54 學系，約 16500 人）為對象，抽取 1000 個樣本，來探討以上的問題。請詳述研究設計，包括研究假設、主要變項（自變項、依變項和控制變項）及測量方式、抽樣、統計方法，並討論以本研究設計的發現進行推論的限制。請注意：以何種統計分析方法，使用哪些變項來檢證假設，這些變項如何測量，務必詳細說明。

2. 請針對與上題相同的研究問題，亦即，「個人申請」是否只以學測成績來選取學生？與「考試入學」比較，「個人申請」這種管道是否確實能夠以考試成績以外的表現和學習經驗來挑選學生？是否以「個人申請」管道入學的學生，比較可能來自社經地位高的家庭？設計一個質化的研究。

你可以選擇任何研究對象（不一定要以台大學生為對象），你也可以選擇任何資料來源或方法（如深度訪談、焦點團體座談、生命史敘事、檔案文本分析、歷史分析等等）。請說明你要收集何種資料，使用何種方法來分析，請詳細說明你使用這類資料和方法的理由，並指出，相較於前面的量化研究，你的質化研究設計有何優點？（共 15 分）

Upper Critical Values of the Chi-Square Distribution

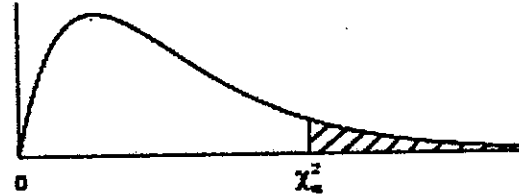
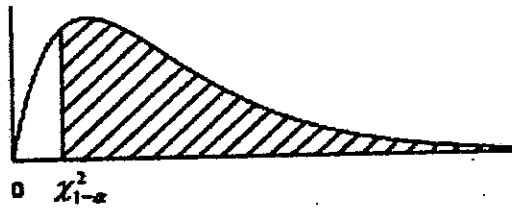


Table Area 1- α

df	.995	.99	.975	.95	.90
1	0.000	0.000	0.001	0.004	0.016
2	0.010	0.020	0.051	0.103	0.211
3	0.072	0.115	0.216	0.352	0.584
4	0.207	0.297	0.484	0.711	1.064
5	0.412	0.554	0.831	1.145	1.610
6	0.676	0.872	1.237	1.635	2.204
7	0.989	1.239	1.690	2.167	2.833
8	1.344	1.646	2.180	2.733	3.490
9	1.735	2.088	2.700	3.325	4.168
10	2.156	2.558	3.247	3.940	4.865
11	2.603	3.053	3.816	4.575	5.578
12	3.074	3.571	4.404	5.226	6.304
13	3.565	4.107	5.009	5.892	7.042
14	4.075	4.660	5.629	6.571	7.790
15	4.601	5.229	6.262	7.261	8.547
16	5.142	5.812	6.906	7.962	9.312
17	5.697	6.408	7.564	8.672	10.085
18	6.265	7.015	8.231	9.390	10.865
19	6.844	7.633	8.907	10.117	11.651
20	7.434	8.260	9.591	10.851	12.443
21	8.034	8.897	10.283	11.591	13.240
22	8.643	9.542	10.982	12.338	14.041
23	9.260	10.196	11.689	13.091	14.848
24	9.896	10.856	12.401	13.848	15.659
25	10.520	11.524	13.120	14.611	16.473
26	11.160	12.198	13.844	15.379	17.292
27	11.808	12.879	14.573	16.151	18.114
28	12.461	13.565	15.308	16.928	18.939
29	13.121	14.256	16.047	17.708	19.768
30	13.787	14.953	16.791	18.493	20.599
31	14.458	15.655	17.539	19.281	21.434
32	15.134	16.362	18.291	20.072	22.271
33	15.815	17.073	19.047	20.867	23.110
34	16.501	17.789	19.806	21.664	23.952
35	17.192	18.509	20.569	22.465	24.797
36	17.887	19.233	21.336	23.269	25.643
37	18.586	19.960	22.106	24.075	26.492
38	19.289	20.691	22.878	24.884	27.343
39	19.996	21.426	23.654	25.695	28.196
40	20.707	22.164	24.433	26.509	29.051
41	21.421	22.906	25.215	27.326	29.907
42	22.138	23.650	25.999	28.144	30.765
43	22.859	24.398	26.785	28.965	31.625
44	23.584	25.148	27.575	29.787	32.487
45	24.311	25.901	28.366	30.612	33.350
46	25.041	26.657	29.160	31.439	34.215
47	25.775	27.416	29.956	32.268	35.081
48	26.511	28.177	30.755	33.098	35.949
49	27.249	28.941	31.555	33.930	36.818
50	27.991	29.707	32.357	34.764	37.689

Table Area α

df	.10	.05	.025	.01	.005
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156
19	27.204	30.144	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997
21	29.615	32.671	35.479	38.932	41.401
22	30.813	33.924	36.781	40.289	42.796
23	32.007	35.172	38.076	41.638	44.181
24	33.196	36.415	39.364	42.980	45.559
25	34.382	37.652	40.646	44.314	46.928
26	35.563	38.885	41.923	45.642	48.290
27	36.741	40.113	43.195	46.963	49.645
28	37.916	41.337	44.461	48.278	50.993
29	39.087	42.557	45.722	49.588	52.336
30	40.256	43.773	46.979	50.892	53.672
31	41.422	44.985	48.232	52.191	55.003
32	42.585	46.194	49.480	53.486	56.328
33	43.745	47.400	50.725	54.776	57.648
34	44.903	48.602	51.966	56.061	58.964
35	46.059	49.802	53.203	57.342	60.275
36	47.212	50.998	54.437	58.619	61.581
37	48.363	52.192	55.668	59.892	62.883
38	49.513	53.384	56.895	61.162	64.181
39	50.660	54.572	58.120	62.428	65.476
40	51.805	55.758	59.342	63.691	66.766
41	52.949	56.942	60.561	64.950	68.053
42	54.090	58.124	61.777	66.206	69.336
43	55.230	59.304	62.990	67.459	70.616
44	56.369	60.481	64.201	68.709	71.893
45	57.505	61.656	65.410	69.957	73.166
46	58.641	62.830	66.617	71.201	74.437
47	59.774	64.001	67.821	72.443	75.704
48	60.907	65.171	69.023	73.683	76.969
49	62.038	66.339	70.222	74.919	78.231
50	63.167	67.505	71.420	76.154	79.490