

國立成功大學

115學年度碩士班招生考試試題

編 號：178

系 所：數據科學研究所

科 目：統計學

日 期：0203

節 次：第 1 節

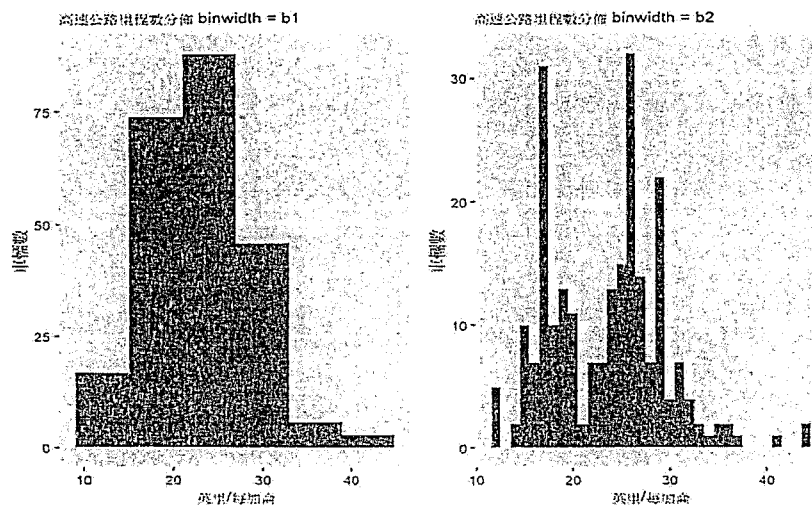
注 意：1.不可使用計算機
2.請於答案卷(卡)作答，於
試題上作答，不予計分。

1. (20%) 數據分析的基本項目是了解資料中心的位置、分散狀況和變數之間的關係。假設有兩題問卷題目 X1 和 X2，詢問受測者的意見，分數範圍由 0, 1, 2, ..., 9 分。假設在簡單隨機抽樣之下，收集到 6 位受測者，資料為 $\{(5, 3), (8, 3), (1, 0), (2, 3), (5, 3), (9, 0)\}$ 。
 - (1) (5%) 使用三種統計量計算 X1 資料中心的位置。
 - (2) (5%) 使用兩種統計量計算 X2 資料分散狀況。
 - (3) (5%) 計算 X1 和 X2 之間皮爾森相關係數。
 - (4) (5%) 敘述隨機變數 X1 與 X2 的關係。

2. (15%) 真實世界由隨機現象產生的數據分佈型態十分多元且複雜，柴比雪夫(Chebyshev)定理是一個能經由平均值和標準差掌握隨機資料分佈比例的定理。
 - (1) (5%) 請敘述柴比雪夫定理。
 - (2) (5%) 假設一個隨機變數 X，觀察到了一組隨機樣本 5, 8, 1, 2, 5, 9。請估計以下機率值
$$\Pr(5 - \sqrt{20} < X < 5 + \sqrt{20})$$
 - (3) (5%) 使用柴比雪夫定理所估計出來的機率值準確嗎？準確原因為何？不準確原因為何？

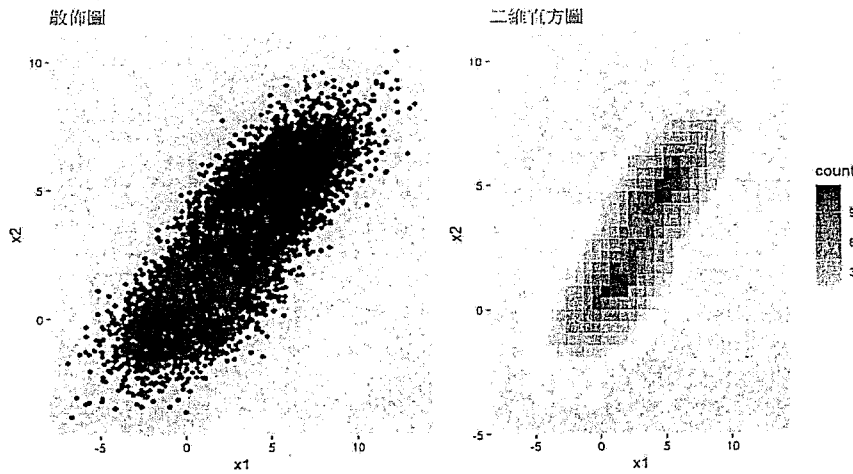
3. (20%) 變異數是衡量產品品質重要的指標，良好的產品品質需要符合既定的水準，不能有過大的產品之間的變異。假設產品品質 X 在某一個生產的過程具有常態分配 $N(\mu, \sigma^2)$ ，今有一批次 n 個同規格產品 X_1, \dots, X_n 。
 - (1) (5%) 試問 $(X_1 - \mu)/\sigma$ 和 $(X_1 - \bar{X})/\sigma$ 各具有何種分配？
 - (2) (5%) 試問 $\sum_{i=1}^n (X_i - \mu)^2 / \sigma^2$ 和 $\sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2$ 各具有何種分配？
 - (3) (5%) 今觀察 n=11 個樣本，樣本變異數 $S^2=0.6$ 。試求 σ^2 95% 信賴區間。
 - (4) (5%) 所算出來的信賴區間包含 σ^2 的機率為何？

4. (15%) 統計圖是數據分析重要的視化工具，適當的應用，對資料分佈的了解可以有很大的幫助。圖一的左圖和右圖都是 234 車輛高速公路的里程數的直方圖，具有不相同的 binwidth。里程數資料的最小值和最大值各為 12 和 44 英里/每加侖。圖二是 5000 筆 (x1, x2) 資料圖。左圖為散佈圖，右圖為二維直方圖。X1 的資料範圍(-6.90, 9.64)，X2 的資料範圍(-3.53, 13.42)。



圖一：234 車輛高速公路的里程數的直方圖

- (1) (5%)求取圖一 binwidth b1 的數值。敘述直方圖繪製的流程。
- (2) (5%)高速公路的里程數的資料分布應視為單峰還是多峰？如何判斷？

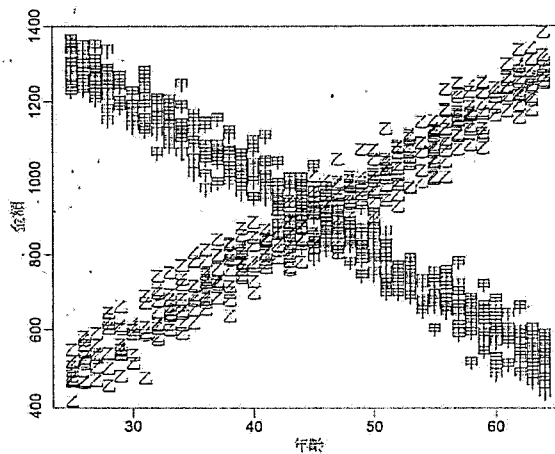


圖二 5000 筆 (x1, x2)資料。左圖為散佈圖，右圖為二為直方圖。

- (3) (5%)估計圖二右圖 binwidth 的數值。敘述二維直方圖繪製的流程。此 5000 筆 (x1, x2)資料分布應視為單峰還是多峰？如何判斷？

5. (30%)消費行為研究常常需要了解顧客特徵和消費現象之間的關係。有一個項消費研究收集 925 消費者的消費金額資料以及年齡和族群特徵，其中 466 來自族群甲，459 人來自族群乙。圖三為年齡與金額的散佈圖，同時標註了族群。今以回歸分析來進行研究，定義族群指標變數，當消費者屬於族群甲，族群指標變數數值設為 0，當消費者屬於族群乙，族群指標變數數值設為 1。回歸模型為

$$\text{金額} = \alpha_0 + \alpha_1 \text{年齡} + \alpha_2 \text{族群指標} + \alpha_3 \text{年齡: 族群指標} + \epsilon$$



圖三: 年齡與金額的散佈圖，同時標註了族群

回歸分析結果

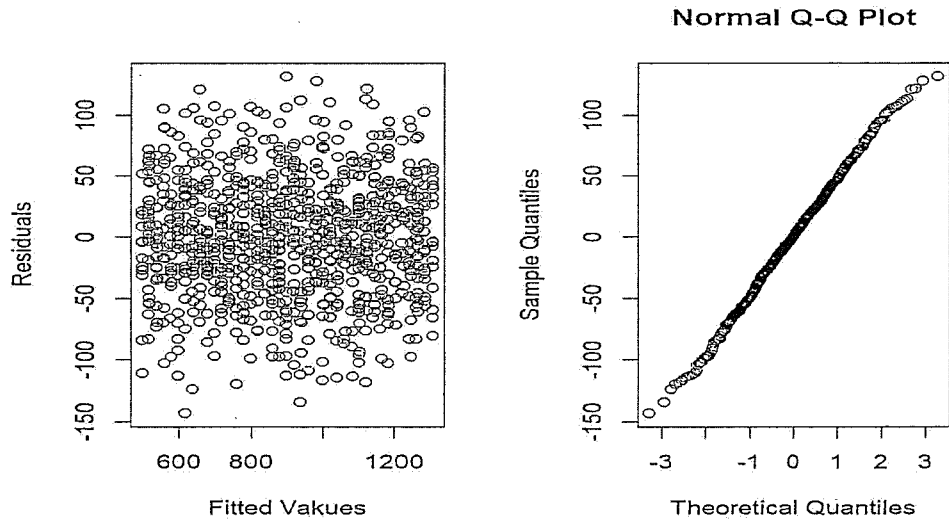
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1814.5445	8.6011	211.0	<2e-16 ***
年齡	-20.3175	0.1878	-108.2	<2e-16 ***
族群指標	-1821.2337	12.5736	-144.8	<2e-16 ***
年齡: 族群指標	40.4461	0.2724		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47.63 on 921 degrees of freedom

Multiple R-squared: 0.9601, Adjusted R-squared: 0.9599

F-statistic: 7379 on 3 and 921 DF



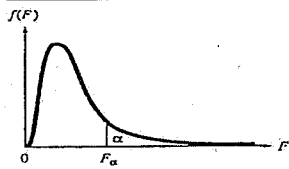
圖四：回歸模型殘差圖

- (1) (5%)列出迴歸係數估計方法的公式。
- (2) (5%)列出 R-square 和 Adjusted R-square 的公式。並說明本回歸模型配適資料的優劣。
- (3) (5%)計算年齡為 40 歲時，本回歸模型估計族群甲和族群乙的顧客平均消費金額各為多少。
- (4) (5%)圖四為回歸模型兩張殘差圖。請利用圖四兩張殘差圖檢查隨機誤差項 ε 的四項分佈假設是
否符合。
- (5) (5%)列出虛無假設 $\alpha_3 = 0$ 時 t 檢定的公式。顯著水準設為 0.05，討論檢定結果。
- (6) (5%)列出回歸模型 F 檢定的虛無假設和檢定公式。顯著水準設為 0.05，討論檢定結果。

Chi-Square Table

df	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.878	14.573	16.151	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.994
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.335
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	44.985	48.232	52.191	55.002
32	15.134	16.362	18.291	20.072	46.194	49.480	53.486	56.328
33	15.815	17.073	19.047	20.867	47.400	50.725	54.775	57.648
34	16.501	17.789	19.806	21.664	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	52.192	55.668	59.893	62.883
38	19.289	20.691	22.878	24.884	53.384	56.895	61.162	64.181
39	19.996	21.426	23.654	25.695	54.572	58.120	62.428	65.475
40	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	124.342	129.561	135.807	140.170

TABLE IX Percentage Points of the F-distribution, $\alpha = .05$



		NUMERATOR DEGREES OF FREEDOM								
		1	2	3	4	5	6	7	8	9
DENOMINATOR DEGREES OF FREEDOM	v_1									
	v_2									
1	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	1	18.51	19.00	19.16	19.25	19.33	19.35	19.35	19.37	19.38
3	1	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	1	7.71	6.94	6.59	6.26	6.16	6.16	6.09	6.04	6.00
5	1	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	1	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	1	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	1	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	1	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	1	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	1	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	1	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	1	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	1	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	1	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	1	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	1	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	1	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	1	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	1	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	1	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	1	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	1	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	1	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	1	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	1	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	1	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	1	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	1	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	1	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	1	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.13
60	1	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	1	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	1	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Source: From M. Merrington and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta (F)-Distribution," *Biometrika*, 1943, 33, 73-88. Reproduced by permission of the *Biometrika* Trustees.

(continued)

Student T Table

Level of Significance of a One-Tailed or Directional Test						
$H_0: \mu - \mu \geq 0$ or $H_0: \mu - \mu \leq 0$						
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = .025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0005$
	$1 - \alpha = 0.90$	$1 - \alpha = 0.95$	$1 - \alpha = 0.975$	$1 - \alpha = 0.99$	$1 - \alpha = 0.995$	$1 - \alpha = 0.9995$
Level of Significance of a Two-Tailed or Nondirectional Test						
$H_0: \mu - \mu = 0$						
	$\alpha = 0.20$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	$\alpha = 0.001$
df	$1 - \alpha = 0.80$	$1 - \alpha = 0.90$	$1 - \alpha = 0.95$	$1 - \alpha = 0.98$	$1 - \alpha = 0.99$	$1 - \alpha = 0.999$
1	3.078	6.314	12.706	31.821	63.656	636.578
2	1.886	2.920	4.303	6.965	9.925	31.600
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.689
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.660
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
50	1.299	1.676	2.009	2.403	2.678	3.496
60	1.296	1.671	2.000	2.390	2.660	3.460
70	1.294	1.667	1.994	2.381	2.648	3.435
80	1.292	1.664	1.990	2.374	2.639	3.416
90	1.291	1.662	1.987	2.368	2.632	3.402
100	1.290	1.660	1.984	2.364	2.626	3.390
150	1.287	1.655	1.976	2.351	2.609	3.357
200	1.286	1.653	1.972	2.345	2.601	3.340
500	1.283	1.648	1.965	2.334	2.586	3.310
1000	1.282	1.646	1.962	2.330	2.581	3.300
∞	1.282	1.645	1.960	2.326	2.576	3.290