108年公務人員高等考試三級考試試題

類　　科：統計
科　　目：迴歸分析
考試時間：2 小時　　　　　　　　　　　　　　　　　　　座號：＿＿＿＿＿＿

※注意：㈠可以使用電子計算器。
㈡不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。
㈢本科目除專門名詞或數理公式外，應使用本國文字作答。

註：所有計算至小數點第 2 位。

一、㈠考慮下列涉及 3 條可能不同截距但相同斜率之直線的簡單線性迴歸模式：

$$y_{1i} = \beta_{01} + \beta_1 x_{1i} + \epsilon_{1i}，$$
$$y_{2i} = \beta_{02} + \beta_1 x_{2i} + \epsilon_{2i}，$$
$$y_{3i} = 1 + \beta_1 x_{3i} + \epsilon_{3i}，$$
$$i = 1, \cdots, n，$$

其中$\epsilon_{11,\cdots,}\epsilon_{1n,}\epsilon_{21,\cdots,}\epsilon_{2n,}\epsilon_{31,\cdots,}\epsilon_{3n}$為彼此獨立且期望值為0而變異數皆為$\sigma^2$的隨機誤差。請利用上述所有資料求出$\beta_{01}, \beta_{02}, \beta_1$的最小平方估計量（least squares estimator）$\hat{\beta}_{01}, \hat{\beta}_{02}, \hat{\beta}_1$及$\hat{\beta}_1$的變異數$Var(\hat{\beta}_1)$。（10 分）

㈡某國政府統計分析師利用迴歸方法分析該國經濟狀況的評估分數Y以及影響該國經濟狀況之重要指數 X，其所用之模式為

$$Y = \beta_0 + \beta_1 X + \epsilon，$$

其中隨機誤差ϵ有下列之機率密度函數表達：

$$f(x) = \frac{e^x}{(1+e^x)^2}，\ -\infty < x < \infty$$

當Y值大於0時，則該國的經濟評估為正向發展；反之即為負向發展。考慮另一變數 Q，當$Y > 0$，則 Q=1，反之當$Y \leq 0$，則 $Q = 0$，即 Q 為該國經濟是否為正向發展的指標。試求出一函數 h 使得

$$h(\mu) = \beta_0 + \beta_1 X，$$

其中$\mu = E(Q)$為 Q 的期望值。（10 分）

二、下列是關於模式選取及模式診斷的問題。

㈠下表為給定 4 種不同迴歸模式來配適 13 組資料$(y_i,\ x_{i1}, x_{i2}, x_{i3})$，$i$=1,⋯,13，所得的 AIC（Akaike's information criterion，赤池訊息標準）值。

模式	模式中的解釋變數	AIC	BIC
A	X_1，X_2	25.41	(1)
B	X_1，X_3	65.11	(2)
C	X_2，X_3	51.03	52.72
D	X_1，X_2，X_3	25.03	(3)

其中x_{ij}為解釋變數X_j的資料值，$j = 1,2,3$，隨機誤差$\epsilon_1, \cdots, \epsilon_{13}$為彼此獨立，期望值為0，變異數皆為$\sigma^2$的常態分配。請完成此表並利用 AIC 及 BIC（Bayesian information criterion，貝氏訊息標準）來選取最適合的模式。（log(a)為數字 a 的自然對數值，log(2)=0.69，log(3)=1.1，log(4)=1.39，log(9)=2.2，log(10)=2.3，log(13)=2.56）。（10 分）

㈡考慮下列複迴歸模式

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i，i = 1, \cdots, 100，$$

其中隨機誤差$\epsilon_1, \cdots, \epsilon_{100}$為彼此獨立，期望值為0，而變異數皆為$\sigma^2$的常態分配。下列的殘差圖（residual plot），請選出那些（個）不適當並請解釋為何不適當:

⑴e_i(y 軸)對$\hat{y}_i$(x 軸)的圖，即$(\hat{y}_i, e_i)$，$i = 1, \cdots, 100$;

⑵e_i(y 軸)對x_{i1}(x 軸)的圖，即(x_{i1}, e_i)，$i = 1, \cdots, 100$;

⑶e_i(y 軸)對y_i(x 軸)的圖，即(y_i, e_i)，$i = 1, \cdots, 100$;

⑷e_i(y 軸)對x_{i3}(x 軸)的圖，即(x_{i3}, e_i)，$i = 1, \cdots, 100$;

⑸e_i(y 軸)對i(x 軸)的圖，即(i, e_i)，$i = 1, \cdots, 100$;

其中資料$(y_i, x_{i1}, x_{i2}, x_{i3})$是代表第$i$天收集的資料，$\hat{y}_i$為第$i$個資料之配適值（fitted value），而$e_i$為第$i$個資料之殘差（residual）值。(5 分)

三、某跨國企業 A 公司其資料科學家利用下列複迴歸模式

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i，i = 1,\cdots,10，$$

來分析該公司一產品的銷售額變化量和該公司此產品價格變化量與競爭對手 B 公司其相對競爭產品價格變化量的關係，其中y_i為 A 公司在第i個地區的銷售額變化量，x_i為 A 公司的產品在第i個地區的價格變化量，z_i為 B 公司的競爭產品在第i個地區的價格變化量，而隨機誤差$\epsilon_1,\cdots,\epsilon_{10}$為彼此獨立，期望值為0，且變異數皆為$\sigma^2$的常態分配。給定銷售額變化量$y_1,\cdots,y_{10}$及下列解釋變數矩陣 X 與反應變數向量 Y 的相關資訊：

$$X^tX=\begin{bmatrix}10 & 0 & 0\\ 0 & 20 & 0\\ 0 & 0 & 20\end{bmatrix}，X^tY=\begin{bmatrix}28\\ -20\\ 12\end{bmatrix}，\bar{y}=\frac{\sum_{i=1}^{10} y_i}{10}=2.8，\sum_{i=1}^{10} y_i^2=130，$$

其中

$$X=\begin{bmatrix}1 & x_1 & z_1\\ 1 & x_2 & z_2\\ \vdots & \vdots & \vdots\\ 1 & x_{10} & z_{10}\end{bmatrix}，Y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_{10}\end{bmatrix}，$$

以及X^t為矩陣X的轉置矩陣。

㈠計算判定係數（coefficient of determination）R^2及$r_{y\hat{y}}$，其中$r_{y\hat{y}}$是觀察值 $y_1,y_2,\cdots,y_{10}$與配適值（fitted values）$\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_{10}$的相關係數（coefficient of correlation）。（8 分）

㈡在顯著水準α=0.05，利用 F 檢定法檢定

$$H_0: \beta_1 = \beta_2 = 0 及 H_1: \beta_1 \neq 0 或 \beta_2 \neq 0。$$

且完成下列關於此檢定的變異數分析表（ANOVA table）。（11 分）

來源 (source)	自由度 (degree of freedom)	平方和 (sum of squares)	均方和 (mean square)	F 統計量
迴歸	(1)	(4)	(7)	(9)
誤差	(2)	(5)	(8)	
總和	(3)	(6)		

㈢在顯著水準α＝0.05，利用 F 檢定法檢定

$$H_0: \beta_1 + 2\beta_2 = 0 及 H_1: \beta_1 + 2\beta_2 \neq 0。（8 分）$$

㈣假定因中美貿易戰的影響，若兩公司同在第 11 個地區競爭且其價格各自調高 1，即x_{11}=1及z_{11}=1。請計算在此地區 A 公司平均銷售額變化量 $E(y_{11}) = \beta_0 + \beta_1 x_{11} + \beta_2 z_{11}$的預測值及其 95%預測信賴區間，即 $E(y_{11})$的點估計及區間估計。（8 分）

四、下列是關於配適模式不正確時造成的影響以及模式適合度問題。

㈠某統計學家欲調查某一地區的當年新生人口與當年經濟成長率的關係是線性或是牽涉到更高的次方關係。此統計學家蒐集了下列在不同經濟成長率 x_i（單位為%）的新生人口資料 y_i（單位為萬人），i=1,⋯,7，

y_1=20	$y_2=22$	$y_3=26$	$y_4=30$	$y_5=37$	$y_6=39$	$y_7=42$
$x_1=-5$	$x_2=-3$	$x_3=-1$	$x_4=0$	$x_5=1$	$x_6=3$	$x_7=5$

並利用下列兩種迴歸模式來配適資料

模式 A：$y_i=\beta_0+\beta_1 x_i+\epsilon_i$

模式 B：$y_i=\beta_0+\beta_1 x_i+\beta_2 x_i^2+\beta_3 x_i^3+\epsilon_i$

其中$\epsilon_{1,\ldots,}\epsilon_7$為彼此獨立且期望值為0，變異數皆為$\sigma^2$的隨機誤差。但是真正的迴歸模式是

$$y_i=\beta_0+\beta_1 x_i+\beta_2 x_i^2+\epsilon_i。$$

如果$\hat{\beta}_{0A}$及$\hat{\beta}_{1A}$為使用模式 A 所得之β_0及β_1的最小平方估計量（least squares estimator），而$\hat{\beta}_{0B}$, $\hat{\beta}_{1B}$及$\hat{\beta}_{2B}$為使用模式 B 所得之β_0,β_1及β_2的最小平方估計量，請得到這些估計量的期望值向量，即$\begin{bmatrix}E(\hat{\beta}_{0A})\\E(\hat{\beta}_{1A})\end{bmatrix}$及$\begin{bmatrix}E(\hat{\beta}_{0B})\\E(\hat{\beta}_{1B})\\E(\hat{\beta}_{2B})\end{bmatrix}$。（7 分）

㈡針對模式 A，請問是否可利用此統計學家所蒐集的資料作模式缺適檢定（lack of fit test）？如果可，請算出檢定統計量的值；如果不可，請解釋原因。（3 分）

五、某工業研究所欲研究某反應過程所散發之熱能Y與某化合物含量X間的關係。利用簡單線性迴歸模式

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i，i = 1,\cdots,5，$$

其中y_i為第i次反應過程所散發熱量之測量值，x_i為第i次反應過程此化合物含量，且隨機誤差$\epsilon_1,\cdots,\epsilon_5$為彼此獨立，期望值為0，變異數皆為$\sigma^2$的常態分配。根據 5 次反應過程所得之資料可得估計迴歸關係式為

$$\hat{y} = 0.2+2.6x$$

且判定係數（coefficient of determination）R^2為 0.845。

㈠計算調整判定係數（adjusted coefficient of determination）及$x_1,\cdots,x_5$與$y_1,\cdots,y_5$ 的相關係數（coefficient of correlation）。（5 分）

㈡在顯著水準$\alpha = 0.05$，利用 F 檢定法檢定

$$H_0: \beta_1 = 0及H_1: \beta_1 \neq 0。（5 分）$$

㈢計算β_1的 95%信賴區間估計。（5 分）

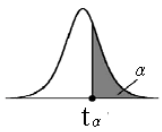
㈣$\hat{\beta}_0$為β_0的最小平方估計量（least squares estimator），且給定$\hat{\beta}_0$的標準誤為 2.13。在顯著水準$\alpha = 0.05$，利用 t 檢定法檢定

$$H_0: \beta_0 \geq 4及H_1: \beta_0 < 4。（5 分）$$

代號：21650
頁次：6-6

Percentage Points of the t Distribution; $t_{\nu,\alpha}$

$P(T > t_{\nu,\alpha}) = \alpha$



	α													
ν	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.02**	**0.015**	**0.01**	**0.0075**	**0.005**	**0.0025**	**0.0005**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	15.895	21.205	31.821	42.434	63.657	127.322	636.590
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303	4.849	5.643	6.965	8.073	9.925	14.089	31.598
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052	2.158	2.291	2.473	2.598	2.771	3.057	3.690
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045	2.150	2.282	2.462	2.586	2.756	3.038	3.659
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960	2.054	2.170	2.326	2.432	2.576	2.807	3.291

F - Distribution ($\alpha = 0.05$ in the Right Tail)

$df_2 \backslash df_1$	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9
Denominator Degrees of Freedom 1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799