

《統計學概要》

一、某週刊調查臺灣地區300家餐廳，針對每家餐廳記錄其餐點價格分數（Y：1表示低價位、2表示中價位、3表示高價位），以及餐廳品質評等分數（X：1表示低品質、2表示中品質、3表示高品質），獲得以下的交叉表：（每小題10分，共30分）

品質評等 (X)	餐點價格 (Y)			合計
	1	2	3	
1	42	39	3	84
2	33	63	54	150
3	3	15	48	66
合計	78	117	105	300

(一) 試計算餐廳品質評等分數之平均值及變異數，意即 $E(X)$ 及 $Var(X)$ 。

(二) 試計算餐廳餐點價格分數之平均值及變異數，意即 $E(Y)$ 及 $Var(Y)$ 。

(三) 若 $Var(X+Y)=1.6691$ ，試計算餐廳品質評等分數 (X) 與餐點價格分數 (Y) 之相關係數，意即 $Cor(X, Y)$ ，並解釋之。

試題評析

第一大題屬於機率論的基本題型，需掌握好二元隨機變數的期望值、變異數、共變異數和相關係數的計算。

考點命中

《高點・高上統計學講義》第四回，葉哲璋編撰，頁15-32。

答：

X與Y之聯合機率分配為：

X	Y			合計
	1	2	3	
1	0.14	0.13	0.01	0.28
2	0.11	0.21	0.18	0.5
3	0.01	0.05	0.16	0.22
合計	0.26	0.39	0.35	1

故X之邊際機率分配為：

X	1	2	3
f(x)	0.28	0.5	0.22

故Y之邊際機率分配為：

Y	1	2	3
f(y)	0.26	0.39	0.35

$$(一) E(X) = \sum f(x_i)$$

$$E(X^2) = \sum x^2 f(x_i)$$

$$r(X) = \frac{E(XY) - E(X)E(Y)}{\sqrt{[E(X^2) - E(X)^2][E(Y^2) - E(Y)^2]}}$$

$$(二) E(Y) = \sum f(y_i)$$

$$E(Y^2) = \sum y^2 f(y_i) = 1^2 \times 0.26 + 2^2 \times 0.39 + 3^2 \times 0.35 = 4.97$$

$$r(Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{[E(X^2) - E(X)^2][E(Y^2) - E(Y)^2]}}$$

$$(三) \text{相關係數 } C(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

因為X和Y互相不為獨立，

$$\text{又相關係數 } C(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

其中 $Cov(X, Y)$ 為 X 和 Y 的共變異數，故

$$\begin{aligned}\text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cor}(X, Y) \times \sqrt{\text{Var}(X)} \times \sqrt{\text{Var}(Y)} \\ &\rightarrow 1.6691 = 0.4964 + 0.6019 + 2 \times \text{Cor}(X, Y) \times \sqrt{0.4964} \times \sqrt{0.6019} \\ &\rightarrow \text{Cor}(X, Y) = 0.5221\end{aligned}$$

【參考書目】

George Casella, Statistical Inference (2nd), Cengage.

二、設有一批產品，每1,000件中平均有一件為不良品。取8,000件的隨機樣本，試運用「卜瓦松 (Poisson) 機率分配近似二項機率分配」的觀念計算其中不良品少於10件的機率為何？(10分)

試題評析 第二大題需熟知當 n 很大且 p 很小時，可以利用卜瓦松機率分配近似二項機率分配的題型。

考點命中 《高點・高上統計學講義》第五回，葉哲璋編撰，頁43-44。

答：

$n=8000$ ，不良品之機率為 $\frac{8}{8000} = 0.001$ 。

設隨機變數 X 表示8000件產品中，發生不良品之次數，則

$X \sim b(8000, 0.001)$ ，因 n 很大且 p 很小，所以可以卜瓦松機率分配近似二項機率分配，

令 $\lambda = n \times p = 8000 \times 0.001 = 8$

$$f(x) \approx \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-8} \times 8^x}{x!}, x = 0, 1, 2, 3, \dots$$

$$\therefore P(X < 10) = \sum_{x=0}^9 \frac{e^{-8} \times 8^x}{x!} = 0.003 + 0.0027 + \dots + 0.1241 = 0.7166 \text{ (參考附表一)}$$

【參考書目】

George Casella, Statistical Inference (2nd), Cengage.

三、假設臺灣有15%的成年人有抽菸的習慣。今隨機抽出200位成年的臺灣人，試問其中至少有35位成年人有抽菸習慣的機率為何？(10分)

試題評析 第三大題需熟知當 n 趨近於無限大時，可以利用常態機率分配近似二項機率分配的題型。但須注意在求算機率時，以連續型的常態機率分配近似離散型的二項機率分配，必須考慮「連續性校正」。

考點命中 《高點・高上統計學講義》第六回，葉哲璋編撰，頁17-19。

答：

設隨機變數 X 表示200位成年的臺灣人中抽菸之人數，

則 $X \sim b(200, 0.15)$ ，

又 $n \times p = 200 \times 0.15 = 30 > 5$ ，且 $n \times (1 - p) = 200 \times (1 - 0.15) = 170 > 5$

則可利用常態機率分配近似二項機率分配的機率，

$$\mu = n \times p = 200 \times 0.15 = 30 \quad \text{【版權所有，重製必究！】}$$

$$\sigma^2 = n \times p \times 1 - p = 200 \times 0.15 \times 0.85 = 25.5$$

$$P(X \geq 35) = P\left(\frac{X - \mu}{\sigma} \geq \frac{(35 - 0.5) - 30}{\sqrt{25.5}}\right) = P(Z \geq 0.89) = 1 - 0.8133 = 0.1867$$

【參考書目】

George Casella, Statistical Inference (2nd), Cengage.

四、根據某樣本求得母體平均數 μ 的99%信賴區間落在244.051與255.949之間。假設母體標準差為 $\sigma=12$ ，試問樣本平均數 $\bar{x}$ 及樣本個數 n 分別是多少？（10分）

試題評析	第四大題屬於信賴區間的觀念題，需掌握好信賴區間的公式及樣本數的計算。
考點命中	《高點·高上統計學講義》第九回，葉哲瑋編撰，頁5-11。

答：

(一)母體平均數 μ 之100(1- α)%的區間估計量為：

$$\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = (\bar{X} - d, \bar{X} + d) = (244.051, 255.949)$$

其中 d 為誤差界限，故

$$\bar{X} = \frac{(\bar{X} - d) + (\bar{X} + d)}{2} = \frac{244.051 + 255.949}{2} = 250$$

(二) $\sigma = 12$ ， $\alpha = 0.01$ ， $Z_{\frac{\alpha}{2}} = Z_{\frac{0.01}{2}} = Z_{0.005} = 2.575$

$$\text{因為 } \bar{X} + d = 255.949 \rightarrow 250 + d = 255.949 \rightarrow d = 5.949$$

$$\text{所以 } n = \left(\frac{Z_{\frac{\alpha}{2}} \times \sigma}{d}\right)^2 = \left(\frac{2.575 \times 12}{5.949}\right)^2 = 26.98$$

取 $n=27$

【參考書目】

George Casella, Statistical Inference (2nd), Cengage.

五、某大學調查近年來經濟學系和統計學系的大學部畢業生之平均起薪（月薪）議題，獲得以下數據訊息：

	經濟學系	統計學系
樣本數	$n_1=10$	$n_2=12$
樣本平均數（元）	$\bar{x}_1 = 30,000$	$\bar{x}_2 = 30,500$
樣本標準差（元）	$s_1=230$	$s_2=250$

(一)試計算此兩學系畢業生之起薪差異的95%信賴區間，並解釋之。（10分）

(二)在 $\alpha=0.05$ 顯著水準之下，試檢定此兩學系畢業生之起薪的變異數是否存在顯著差異。（15分）

(三)在 $\alpha=0.05$ 顯著水準之下，試檢定此兩學系畢業生之平均起薪是否存在顯著差異。（15分）

（以上(二)、(三)皆需正確寫出虛無假設、對立假設、檢定統計量、拒絕域、P-值、檢定結果與結論。）

試題評析	第五大題屬於信賴區間和假設檢定的基本題型，需掌握好信賴區間的公式及正確寫出假設檢定的流程(依題意)。
考點命中	1.《高點·高上統計學講義》第九回，葉哲瑋編撰，頁25-27。 2.《高點·高上統計學講義》第十回，葉哲瑋編撰，頁49-54、90-93。

答：

(一)假設兩母體資料分配都是常態分配，
由第二題的檢定結果可知，兩母體變異數相等(但未知)。

$$\text{故 } \sigma_1^2 = \sigma_2^2 = s_p^2$$

設 X 為經濟學系的起薪

Y 為統計學系的起薪

$$\text{則 } \alpha = 0.05, t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0.025}(20) = 2.086$$

$$n_1 = 10, n_2 = 12$$

$$\bar{x}_1 = 30000, \bar{x}_2 = 30500$$

$$s_1^2 = 230^2, s_2^2 = 250^2$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(10 - 1) \times 230^2 + (12 - 1) \times 250^2}{(10 - 1) + (12 - 1)} = 58180$$

故 $\mu_1 - \mu_2$ 之100(1- α)%的區間估計量為：

$$\begin{aligned} & \left(\bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \bar{x}_1 - \bar{x}_2 + t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right) \\ &= \left((30000 - 30500) - 2.086 \times \sqrt{58180 \times \left(\frac{1}{10} + \frac{1}{12} \right)}, (30000 - 30500) + 2.086 \right. \\ & \quad \left. \times \sqrt{58180 \times \left(\frac{1}{10} + \frac{1}{12} \right)} \right) \\ &= (-715.4378, -284.5622) \end{aligned}$$

我們有95%的信心，此區間能夠包含兩畢業生之真正平均起薪差異。

(二) 假設兩母體資料分配都是常態分配，

$$\alpha = 0.05, n_1 = 10, n_2 = 12$$

$$s_1^2 = 230^2, s_2^2 = 250^2$$

$$F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) = F_{0.975}(9, 11) = \frac{1}{F_{0.025}(11, 9)} = \frac{1}{3.912} = 0.2556$$

$$F_{\frac{\alpha}{2}}(n_1-1, n_2-1) = F_{0.025}(9, 11) = 3.588$$

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$\text{拒絕域: } C = \{F | F < F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) \text{ 或 } F > F_{\frac{\alpha}{2}}(n_1-1, n_2-1)\}$$

$$= \{F | F < 0.2556 \text{ 或 } F > 3.588\}$$

$$\text{檢定統計量: } F = \frac{s_1^2}{s_2^2} = \frac{230^2}{250^2} = 0.8464 \notin C, \text{ 不拒絕 } H_0。$$

結論：在 $\alpha = 0.05$ 下，無充分證據顯示兩學系畢業生之起薪的變異數有顯著差異。

$$p\text{-value} = 2 \times \min\{P(F > \frac{s_1^2}{s_2^2}), P(F < \frac{s_1^2}{s_2^2})\} = 2 \times \min\{P(F > \frac{s_1^2}{s_2^2} = 0.8464), P(F < \frac{s_1^2}{s_2^2} = 0.8464)\}$$

(三) 假設兩母體資料分配都是常態分配，

由第二題的檢定結果可知，兩母體變異數相等(但未知)。

$$\text{故 } \sigma_1^2 = \sigma_2^2 = s_p^2$$

則由第一題可知，

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

$$\text{拒絕域 } C = \{t | t < -t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) \text{ 或 } t > t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)\}$$

$$= \{t | t < -2.086 \text{ 或 } t > 2.086\}$$

$$\text{檢定統計量 } t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(30000 - 30500) - 0}{\sqrt{58180 \times \left(\frac{1}{10} + \frac{1}{12} \right)}} = -4.8413 \in C, \text{ 拒絕 } H_0$$

結論：在 $\alpha = 0.05$ 下，有充分證據顯示兩學系畢業生之平均起薪有顯著差異。

$$p\text{-value} = 2 \times P(T > |t| | T \sim t(n_1 + n_2 - 2)) = 2 \times P(T > 4.8413 | T \sim t(20))$$

【參考書目】

George Casella, Statistical Inference (2nd), Cengage.

附表一：

卜瓦松機率表 $X \sim \text{Poisson}(\mu)$

x	μ									
	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993

附表二：

標準常態機率分布表（左尾機率），例如： $P(Z \leq 1.96) = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

【版權所有，重製必究！】

附表三：

t 機率分布表（右尾機率），例如： $P(t_{25} > 0.856) = 0.20$

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
20	.860	1.325	1.725	2.086	2.528	2.845
21	.859	1.323	1.721	2.080	2.518	2.831
22	.858	1.321	1.717	2.074	2.508	2.819
23	.858	1.319	1.714	2.069	2.500	2.807
24	.857	1.318	1.711	2.064	2.492	2.797

附表四：

F 分布臨界值與對應之機率值， $P(F(v_1, v_2) > F_\alpha(v_1, v_2)) = \alpha$

$F_{0.1}(9, 11) = 2.274$, $F_{0.05}(9, 11) = 2.896$, $F_{0.025}(9, 11) = 3.588$, $F_{0.01}(9, 11) = 4.632$, $F_{0.005}(9, 11) = 5.537$
 $F_{0.1}(10, 12) = 2.188$, $F_{0.05}(10, 12) = 2.753$, $F_{0.025}(10, 12) = 3.374$, $F_{0.01}(10, 12) = 4.296$, $F_{0.005}(10, 12) = 5.085$
 $F_{0.1}(11, 9) = 2.396$, $F_{0.05}(11, 9) = 3.103$, $F_{0.025}(11, 9) = 3.912$, $F_{0.01}(11, 9) = 5.178$, $F_{0.005}(11, 9) = 6.314$
 $F_{0.1}(12, 10) = 2.284$, $F_{0.05}(12, 10) = 2.913$, $F_{0.025}(12, 10) = 3.621$, $F_{0.01}(12, 10) = 4.706$, $F_{0.005}(12, 10) = 5.661$

【版權所有，重製必究！】