數三統計學 (二) 小考參考題 日期:98.06.15 系級:____________學號:____________姓名:____________

⨂每小題10分 (可以使用計算器)6/18請記得帶計算機及分配表來考試, 勿過度依賴參考題

1. 箭牌口香糖每片的厚度$X \sim n(\mu, \sigma^2)$ 今有10片的厚度如下: 7.65, 7.60, 7.65, 7.70, 7.55, 7.55, 7.40, 7.40, 7.50, 7.50, 試檢定$H_0 : \mu = 7.5$ vs $\mu \neq 7.5$

   (a) 求$p - value$
   解:$\bar{x} = 7.55, s_x = 0.1027 \Rightarrow t = \frac{\bar{x}-7.5}{s_x/\sqrt{n}} = 1.539$介於$t_{.05,9} = 1.833$及 $t_{0.1,9} = 1.383$之間,$\Rightarrow$ $p - value$介於0.05×2及0.1×2之間, 也就是$p - value$介於0.1及0.2之間。

   (b) 求μ的95%CI
   解:$\bar{x} \pm t_{.025,9} s_x/\sqrt{n} = 7.55 \pm 2.262 \times 0.1027/\sqrt{10} = (7.4765, 7.6234)$.

   (c) 結論若$\alpha = 0.05$
   解: 由 (a)p-value$>$ 0.05或由 (b)7.5$\in$ 95% CI of $\mu \Rightarrow$ Fail to reject H_0

2. 擲一骰子120次得

點數	1	2	3	4	5	6
次數	40	10	10	10	10	40

   試檢測H_0 :骰子是公正的 vs H_a : not H_0 , $\alpha = 0.05$
   解:Under $H_0, E(X_i) = 120 \times \frac{1}{6} = 20, i = 1, 2, \cdots, 6 \Rightarrow Q_5 = \frac{(40-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(40-20)^2}{20} = 20 + 5 + 5 + 5 + 5 + 20 = 60 > \chi^2_{.05,5} = 11.07 \Rightarrow$ Reject H_0

3. 設$X_1, X_2, \ldots, X_n \overset{r.s.}{\sim} Poisson(\theta)$且$\theta$滿足 prior distribution $f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad x > 0$

   (a) 求θ的 posterior distribution.
   解:The likelihood function $L(\mathbf{x}|\theta) = e^{-n\theta}\theta^{\sum x_i}/\prod x_i! \propto e^{-n\theta}\theta^{\sum x_i}$
   The prior pdf of θ, $\xi(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta}$
   The posterior pdf is proportional to $L(\mathbf{x})\xi(\theta)$, it follows that
   $\xi(\theta|\mathbf{x}) \propto \theta^{\alpha+\sum x_i - 1} e^{-(\beta+n)\theta}$
   The right side of this relation can be recognized as the pdf of a gamma distribution with parameters $\alpha + \sum x_i$ and $\beta + n$(有人寫成倒數也是對的).

   (b) Suppose the squared error loss function is used, Find the Bayes estimate of θ
   解:The Bayes estimate $\delta^*(\mathbf{x}) = \frac{\alpha+\sum x_i}{\beta+n}$, which is the mean of the posterior distribution.

4. 隨機調查100位同學最喜歡的休閒運動如下表

	喜歡的休閒運動					
性別	籃球	棒球	游泳	慢跑	網球	合計
男生	21	5	9	12	13	60
女生	9	3	1	15	12	40
合計	30	8	10	27	25	100

   試檢驗喜好和性別是否無關

   (a) 求檢定統計量.
   解:$Q_4 = \frac{\left(21 - \frac{(60)(30)}{100}\right)^2}{\frac{(60)(30)}{100}} + \frac{\left(5 - \frac{(60)(8)}{100}\right)^2}{\frac{(60)(8)}{100}} + \cdots + \frac{\left(12 - \frac{(40)(25)}{100}\right)^2}{\frac{(40)(25)}{100}} = 0.5 + 0.008 + 1.5 + 1.089 + 0.267 + 0.750 + 0.013 + 2.250 + 1.633 + 0.400 = 8.410$

   (b) 問檢定統計量的自由度
   解:$(r-1)(c-1) = (2-1)(5-1) = 4$

   (c) p-值大約是多少?
   解:8.41介於$\chi^2_{0.1,4} = 7.779$及$\chi^2_{0.05,4} = 9.488$之間, 因此 p-值介於0.05及0.1之間.(統計軟體給值爲0.078)

   (d) 若$\alpha = 0.05$, 是否棄卻虛無假設?
   解: 因爲 p-值大於0.05因此無法棄卻虛無假設

5. 吾人欲知補習對 IQ 測驗是否有效, 隨機取10位學童, 記錄補習前與補習後 IQ 的測驗成績如下:

補習前 (x)	97	106	106	95	102	111	115	104	90	96
補習後 (y)	113	113	101	119	111	122	121	106	110	126

若資料滿足常態分配, 試檢驗補習是否對提高 IQ 有效?($\alpha = 0.05$)

解: 欲檢定$H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x < \mu_y$, 令$d_i = x_i - y_i, i = 1, 2, \ldots, 10 \Rightarrow \bar{d} = -12, s_d = 10.5830$ 所以$t = \frac{\bar{d}}{s_d/\sqrt{10}} = \frac{-12}{10.5830/\sqrt{10}} = -3.5857 < -t_{9,0.05} = -1.8331$, 所以棄卻虛無假設, 因此補習對提高 IQ 測試成績有顯著效果。

6. 肺塵病的醫療研究以天竺鼠做實驗, 利用甲、乙兩種藥做治療, 從85隻感染肺塵病天竺鼠中隨機選取45隻以甲種藥做治療一個月後有20隻死亡, 其餘40隻以乙種藥治療一個月後有13隻死亡, 試問在$\alpha = 0.05$下, 這兩種藥的效果是否有顯著差異?

解: 欲檢定$H_0 : p_1 = p_2$ vs $p_1 \neq p_2$, 其中p_1, p_2分別表示甲, 乙兩種藥治療成功的比例,$\hat{p_1} = \frac{25}{45} = 0.5556, \hat{p_2} = \frac{27}{40} = 0.675, \bar{p} = \frac{52}{85} = 0.6118, z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.5556 - 0.675}{0.6118 \times 0.3882 \times (\frac{1}{45} + \frac{1}{40})} = -1.1275$, 因爲$|z| < z_{0.025} = 1.96$ 所以無法棄卻H_0, 即無證據說兩種藥效有顯著差異。

7. 有甲、乙兩位侯選人參加某次鎮長選舉, 從選民中隨機問50位 (設選民的支持意向清楚且沒有模稜兩可者), 試問回答支持甲侯選人的支持者要幾位以上能有足夠證據說甲的得票率比乙的得票率多10%以上? ($\alpha = 0.05$)

解: 設p爲甲的得票率, 欲檢定$H_0 : p \leq 0.55$ vs $H_1 : p > 0.55$, $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\hat{p} - 0.55}{\sqrt{\frac{0.55 \times 0.45}{50}}} \geq z_{0.05} = 1.645 \Rightarrow \hat{p} = \frac{x}{50} \geq 0.55 + 1.645 \times \sqrt{\frac{0.55 \times 0.45}{50}} = 0.6657 \Rightarrow x \geq 50 \times 0.6657 = 33.285 \Rightarrow x \geq 34$ 所以 至少要有34位支持甲侯選人才有足夠證據說甲的得票率比乙的得票率多10%以上

8. 有四種品種的玉蜀黍, 隨機指定在面積相同的土地上栽種, 其產量整理如下表:(資料假設是常態分配)

品種	樣本數n_i	平均數$\bar{x}_i$	標準差s_i
1	9	90.5556	3.6439
2	10	86.4000	3.7771
3	7	95.7143	3.6384
4	8	79.8750	1.7269

(a) 試完成 ANOVA 表

變異來源	自由度	平方和	均方和	F 值	P 值
品種	A	D	F	H	I
誤差	B	E	G		
總和	C	1363.5294			

解:A=k-1=4-1=3

B=$\sum(n_i - 1) = (9-1) + (10-1) + (7-1) + (8-1) = 30$

C=A+B=33

E=$\sum(x_{ij} - \bar{x}_i)^2 = \sum(n_i - 1)s_i^2 = 334.9258$

D=1363.5294-E=1028.6036

F=1028.6036/3=342.8679

G=334.9258/30=11.1642

H=342.8679/11.1642=30.71

F 值$> F_{0.01,3,30} = 4.51$所以 P 值$< 0.01, I < 0.01$

(b) 檢定四種品種產量平均數是否相等?($\alpha = 0.05$)

解:$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs H_1 :不是所有μ_i都相等

由於 P 值$< 0.01 < 0.05$所以棄卻盧無假設也就是四種品種產量平均數顯著的不相等

(c) 求$\mu_1 - \mu_2$的95%信賴區間

解:$s_p = \sqrt{MSE} = \sqrt{11.1642} = 3.3413$, $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2,n-k} s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 90.5556 -$

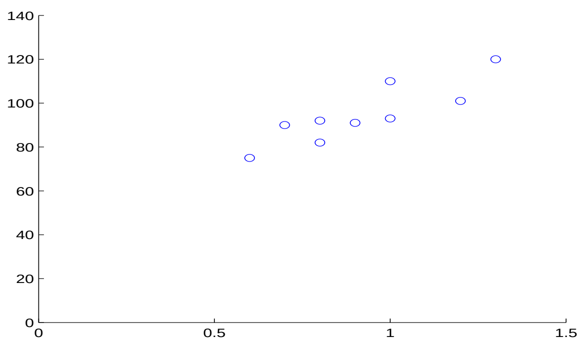
$$86.4000 \pm 2.0423 \times 3.3413 \times \sqrt{\tfrac{1}{9} + \tfrac{1}{10}} = (1.0202, 7.2910)$$

9. 收集某行業其每月銷售量與廣告費用, 資料如下:

廣告量 (x)	1.2	0.8	1.0	1.3	0.7	0.8	1.0	0.6	0.9
銷售量 (y)	101	92	110	120	90	82	93	75	91

(a) 求x, y的相關係數。
解: 相關係數$r = \frac{\sum(x_i-\bar{x})(y-i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y-i-\bar{y})^2}} = \frac{21.72222}{0.41556 \times 1508.88889} = 0.8675$

(b) 畫y對x的散佈圖
解:



(c) 試求迴歸線$E(Y) = b_0 + b_1 x$的估計
解:$\hat{b}_1 = \frac{\sum(x_i-\bar{x})(y-i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{21.72222}{0.41556} = 52.2727$
$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 94.88889 - 52.2727 \times 0.92222 = 46.6818$
所以迴歸線$y = 46.6818 + 52.2727x$

(d) 若$Y|x \sim n(b_0 + b_1 x, \sigma^2)$, 求$\sigma^2$的估計值
解: $\hat{\sigma^2} = \frac{\sum(y_i - \hat{b}_0 - \hat{b}_1 \times x_i)^2}{n-2} = 53.3440$

(e) 預測某日廣告費用投入1.0時, 其銷售量爲何?
解: $\hat{y} = 46.6818 + 52.2727 \times 1.0 = 98.9545$

(f) 預測某日廣告費用投入1.0時, 其銷售量95%的預測區間爲何?
解: $\hat{y} \pm t_{\alpha/2, n-2} \times \hat{\sigma} \times \sqrt{1 + \frac{1}{n} + \frac{(1-\bar{x})^2}{\sum(x_i-\bar{x})^2}} = 98.9545 \pm 2.3646 \times 7.3037 \times \sqrt{1 + \frac{1}{9} + \frac{(1-0.92222)^2}{0.41556}} = (80.631, 117.278)$

(g) 在$\alpha = 0.05$下, 檢定廣告費與銷售量是否相關
解: 檢定$H_0 : b_1 = 0$ vs $H_1 : b_1 \neq 0$
$t = \frac{\hat{b}_1}{\hat{\sigma}/\sqrt{\sum(x_i-\bar{x})^2}} = \frac{52.2727}{7.3037/\sqrt{0.41556}} = 4.614 > t_{7,0.025} = 2.3646$
所以棄卻H_0, 即廣告費與銷售量有顯著相關。