《統計學》

一、某人為瞭解稻米產量、耕作面積及氮肥用量間之關係，蒐集了50戶農家的資料，得到敘述統計量如下：

	產量（千公斤）	耕作面積（甲）	氮肥用量（百公斤）
樣本平均	10.68	6.78	6.64
樣本變異數	10.30	4.18	6.24
相關係數	（產量、耕作面積）；0.71 （產量、氮肥用量）：0.52 （耕作面積、氮肥用量）：0.68		

(一) 今若以簡單線性迴歸模型

產量= a + b × 耕作面積

研析產量與面積之關係，並以耕作面積為解釋變數，請問該模型中a與b之估計值為何？本模式下之判定係數為何？（10分）

(二) 請寫出本模型下之變異數分析（ANOVA）表，並檢定耕作面積在95%之信心水準下是否與產量有顯著之關係。請詳列虛無以及對立假設、檢定統計量、檢定原則及檢定結果。（10分）

(三) 若將氮肥用量加入以多元線性迴歸模型

產量= a + b_1 ×耕作面積+ b_2 ×氮肥用量

分析結果得到b_1及b_2之估計值分別為$\hat{b}_1 = 1.04$及$\hat{b}_2 = 0.1$，請問該b_1之估計值的解釋與(一)之簡單線性迴歸中b的估計值之解釋有何異同？（5分）

(四) 在(三)的多元線性迴歸模型所得之變異數分析（ANOVA）表如下：

變異來源	平方和	自由度	均方	F值
迴歸	256.61	2	128.305	24.291
殘差	248.27	47	5.282	
總變異	504.88	49		

請檢定在耕作面積已經列為線性迴歸模型的解釋變數之情形下，當信心水準為95%時，氮肥用量是否為顯著之解釋變數。請敘述恰當之虛無及對立假設、檢定統計量、檢定原則及檢定結果。（10分）

試題評析	本題屬於簡迴歸與複迴歸之計算題型，包含迴歸係數之意義解釋，屬於常考問題。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第八章。

答：

(一)

令Y表產量(千公斤)

X_1表耕作面積(甲)，X_2表氮肥用量(百公斤)

已知 $\bar{y} = 10.68$，$\bar{x}_1 = 6.78$，$\bar{x}_2 = 6.64$，$s_Y^2 = 10.30$，$s_1^2 = 4.18$，$s_2^2 = 6.24$，$r_{Y1} = 0.71$，

$r_{Y2} = 0.52$，$r_{12} = 0.68$

$\hat{b} = r_{Y1}\dfrac{s_Y}{s_1} = 1.1145$ ，$\hat{a} = \overline{Y} - \hat{b}\overline{X}_1 = 3.12369$

$R^2 = r_{Y1}^2 = 0.5041$

(二)

ANOVA TABLE

Source	SS	d.f.	MS	F
Reg	250.409	1	250.409	$F^{*}=47.267$
Error	254.291	48	5.2977	
Total	504.7	49		

其中 $SST=SS_Y=(n-1)S_Y^2=504.7$

$SSR=SS_1=\hat{b}^2(n-1)S_1^2=254.409$

假設迴歸模型： $Y_i=a+bX_{1i}+\varepsilon_i,\varepsilon_i\overset{iid}{\sim}N(0,\sigma^2),i=1,2,\cdots,50$

$H_0:b=0$ vs $H_1:b\neq 0$

T.S.: $F=\dfrac{\text{MSR}}{\text{MSE}}\sim F_{(1,48)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $F^{*}>F_{(1,48)0.05}=4.043$

$\because F^{*}=47.267$ $\quad\therefore$ Reject H_0

結論：我們有足夠證據去推論耕作面積與產量有顯著之關係。

(三)

在(一)的模型中,$\hat{b}=1.1145$表當耕作面積每增加1甲,平均稻米產量會增加1.1145千公斤

$\hat{b}_1=1.04$表固定氮肥用量(X_2)下,當耕作面積每增加1甲,平均稻米產量會增加1.04千公斤

可見$\hat{b}_1$是一種邊際影響

(四)

假設迴歸模型: $Y_i=a+b_1X_{1i}+b_2X_{2i}+\varepsilon_i,\varepsilon_i\overset{iid}{\sim}N(0,\sigma^2),i=1,2,\cdots,50$

$H_0:b_2=0$ vs $H_1:b_2\neq 0$

T.S.: $F=\dfrac{SSR(X_2\mid X_1)/1}{SSE(X_1,X_2)/(50-2-1)}\sim F_{(1,47)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $F^{*}>F_{(1,47)0.05}=4.047$

$\because F^{*}=\dfrac{[SSR(X_1,X_2)-SSR(X_1)]/1}{SSE(X_1,X_2)/(50-2-1)}=\dfrac{[256.61-250.409]/1}{248.27/(50-2-1)}=1.1739$

$\therefore$ don't Reject H_0

結論：我們沒有足夠證據去推論氮肥用量為顯著之解釋變數。

二、對便利商店所進行之經營狀況調查中，隨機選擇100家便利商店後記錄其週營業額，但因原始資料遺失，僅存次數分配資料如下：

週營業額（萬元）	便利商店數
$20\leq$週營業額<25	35
$25\leq$週營業額<30	30
$30\leq$週營業額<35	20
$35\leq$週營業額<40	10
$40\leq$週營業額<45	5

請根據本表推算便利商店之週營業額樣本平均、樣本變異數以及樣本中位數。（15分）

試題評析	本題屬於敘述統計學之計算題型，考生沒有計算錯誤，應該可輕易獲滿分。
考點命中	《高點‧高上統計學講義》第一回，趙治勳編撰，第二章。

答：

週營業額X	次數 f_i	組中點 m_i	累積次數
$20\le X<25$	35	22.5	35
$25\le X<30$	30	27.5	65
$30\le X<35$	20	32.5	85
$35\le X<40$	10	37.5	95
$40\le X<45$	5	42.5	100

$$\overline{X}=\frac{\sum f_i m_i}{\sum f_i}=\frac{35\times 22.5+\cdots+5\times 42.5}{100}=28.5$$

$$S^2=\frac{\sum f_i m_i^2-n\overline{X}^2}{n-1}=\frac{84625-100\times 28.5^2}{100-1}=34.343$$

第 $x=\frac{100}{2}=50$ 個累積次數落在第二組，故樣本中位數落在週營業額為[25,30)(萬元)之間

$$\frac{50-35}{Md-25}=\frac{65-50}{30-Md}$$

$\Rightarrow Md=27.5$(萬元)

三、為鼓勵提升生育率，評估若要對甲鎮內家戶中有12歲以下之家戶成員數目為2人或以上之家戶實施補助所需之資源，於鎮內隨機選取200家戶，各家戶中12歲以下家戶成員數目如下表：

12歲以下家戶成員數	0	1	2	3	4	5
家戶數	70	82	33	10	3	2

(一) 請問此一未成年家戶成員數資料在95%之信心水準下是否服從卜瓦松分配（Poisson distribution）？（15分）

(二) 請說明在95%之信心水準下是否有顯著之證據指出甲鎮內家戶12歲以下未成年家戶成員數目大於或等於2人之比例大於15%。請敘述恰當之虛無及對立假設、檢定統計量及檢定結果。（10分）

試題評析	本題屬於卡方適合度檢定與母體成功比例之假設檢定，考古題常有，且跟課本例題幾近相同，考生獲得滿分不難。
考點命中	1.母體成功比例之假設檢定：《高點‧高上統計學講義》第一回，趙治勳編撰，第六章。 2.卡方適合度檢定：《高點‧高上統計學講義》第一回，趙治勳編撰，第九章例6。

答：

(一) 在 H_0 為真下, $X\sim Poisson(\lambda=\overline{x}=1)$

$$E_1=n\times P(X=0)=200\times\frac{e^{-1}\times 1^0}{0!}=73.58$$

$$E_2=n\times P(X=1)=200\times\frac{e^{-1}\times 1^1}{1!}=73.58$$ 以此類推

X	0	1	2	3	4	5	
O_i	70	82	33	10	3	2	200
E_i	73.58	73.58	36.79	12.26	3.07	0.72	200

合併X=3,4,5

X	0	1	2	3	
O_i	70	82	33	15	200
E_i	73.58	73.58	36.79	16.05	200

H_0 : 資料服從Poisson分配　vs　H_1 : not H_0

T.S.: $\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}\overset{d}{\sim}\chi^2_{(4-1-1=2)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $\chi^{2*}>\chi^2_{(2)0.05}=5.991$

$\because \chi^{2*}=1.5968$　　$\therefore$ don't reject H_0

結論：我們沒有足夠證據去推論資料不是服從Poisson分配。

(二) 令 X 表 12 歲以下成員數大於或等於 2 人

母體: $X\sim Ber(p)$

樣本: $X_1,X_2,\cdots,X_{200}\overset{iid}{\sim}Ber(p)$

點估計: $\hat{p}=\frac{\sum X_i}{n}\underset{byC.L.T.}{\sim}\mathrm{N}(p,\frac{p(1-p)}{200})$

$H_0: p\le 0.15$　vs　$H_1: p>0.15$

T.S.: $Z=\frac{\hat{p}-0.15}{\sqrt{\frac{0.15(1-0.15)}{200}}}\overset{d}{\sim}N(0,1)$

R.R.: Reject H_0 at $\alpha=0.05$ if $Z^*>z_{0.05}=1.645$

$\because Z^*=\frac{\frac{48}{200}-0.15}{\sqrt{\frac{0.15(1-0.15)}{200}}}=3.56$　　$\therefore$ reject H_0

結論：我們有足夠證據去推論12歲以下成員數大於或等於2人之比例大於15%。

四、某人想要研析三種不同的灌溉方式對產量的影響，規劃了15塊面積相同的農地，每種灌溉方式各隨機選擇5塊農地加以應用，三種灌溉方式所得產量資料整理後如下表：

灌溉方式	A	B	C
樣本平均	15.4	18.4	28.4
樣本變異數	9.3	20.8	18.3

經一因子變異數分析後，發現三種灌溉方式之產量確有顯著之差距，請回答下列問題：

(一) 請寫出一因子變異數分析之變異數分析表（含檢定三種灌溉方式之產量是否相同之檢定統計量值）。（10分）

(二) 因為一因子變異數分析顯示三種灌溉方式之產量顯著不同，請用費雪之最小顯著差異法（Fisher's Least Significant Difference），以單一成對比較信心水準為95%之基準下，判斷那幾組成對之兩種灌溉方式產量比較有顯著之不同。（10分）

(三) 承(二)中之整體型一誤差機率為何（亦即產生至少一次型一誤差之機率）？若用邦佛洛尼校正法（Bonferroni correction）調整單一成對檢定之顯著水準，調整後之單一成對檢定之顯著水準以及整體型一誤差機率為何？（5分）

試題評析	本題屬於ANOVA與多重比較之計算題，考生若沒有計算錯誤，應該可輕易獲得滿分。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第七章。

答：

(一)

ANOVA TABLE

Source	SS	d.f.	MS	F
灌溉	463.333	2	231.667	$F^{*}=14.36$
Error	193.6	12	16.133	
Total	656.933	14		

其中 $SSE=\sum_{i=1}^{k}(n_i-1)S_i^2=193.6$

$SSR=\sum_{i=1}^{k}n_i(\overline{Y}_i-\overline{\overline{Y}})^2=463.333$ 其中 $\overline{\overline{Y}}=\dfrac{15.4+18.4+28.4}{3}=20.733$

(二)

$t_{(12)0.025}\sqrt{MSE(\frac{1}{5}+\frac{1}{5})}=2.179\sqrt{16.133(\frac{1}{5}+\frac{1}{5})}=5.535$

$\mu_A-\mu_B:(15.4-18.4)\mp5.535=(-,+)$

$\mu_A-\mu_C:(15.4-28.4)\mp5.535=(-,-)$

$\mu_B-\mu_C:(18.4-28.4)\mp5.535=(-,-)$

可見A vs C 與 B vs C有顯著不同

(三)

整體: P(至少一次發生型I誤差)=1-P(沒有發生型I誤差)=$1-(1-\alpha)^{\binom{k}{2}}$

$=1-(1-0.05)^{\binom{3}{2}}=0.143$

Bonferroni校正法

單一: 顯著水準=$\dfrac{\alpha}{\binom{k}{2}}=\dfrac{0.05}{\binom{3}{2}}=0.0167$

整體: P(至少一次發生型I誤差)=$1-(1-0.0167)^{\binom{3}{2}}=0.0493$