《統計學概要-統計》

一、下表為108年某縣政府家庭收支調查的300個樣本戶住宅面積的分組次數分配表資料：

組別	住宅面積（坪）	戶數	累積戶數
1	0～20以下	18	18
2	20～40以下	177	195
3	40～60以下	57	252
4	60～80以下	32	284
5	80～100以下	11	295
6	100～120以下	5	300

請根據此分組之次數分配表資料，試計算下列樣本統計量之值：（每小題10分，共30分）

(一)算術平均數

(二)標準差

(三)第10百分位數（P_{10}）

試題評析	本題考統計學的範圍內容，小心計算，要獲得高分應該不難。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第三章。

答：

組別(i)	住宅面積	戶數(f_i)	累積戶數	組中點(m_i)
1	0~20	18	18	10
2	20~40	177	195	30
3	40~60	57	252	50
4	60~80	32	284	70
5	80~100	11	295	90
6	100~120	5	300	110

(一) $\bar{x}=\dfrac{\sum_{i=1}^{k} f_i m_i}{n}=\dfrac{18\times10+177\times30+57\times50+32\times70+11\times90+5\times110}{300}=40.4$ (坪)

(二) $s=\sqrt{\dfrac{\sum_{i=1}^{k} f_i(m_i-\bar{x})^2}{n-1}}=\sqrt{\dfrac{18(10-40.4)^2+......+5(110-40.4)^2}{300-1}}=20.0628$ (坪)

(三) 令η_{10}為第10百分位數

第$i\times\dfrac{n}{k}=10\times\dfrac{300}{100}=30$個累積次數所處為第2組別

故第10百分位數落入該組界(20,40)內，以內插法計算

$\dfrac{30-18}{\eta_{10}-30}=\dfrac{195-30}{40-\eta_{10}}\Rightarrow\eta_{10}=30.678$ (坪)

二、設$\overline{X}_9$與$\overline{X}_{25}$為分別從常態母體$N(\mu,\sigma=3)$中隨機抽取9個與25個樣本的樣本平均數。今要對參數μ進行檢定，而我們所設定的統計假設為$H_0:\mu=0$及$H_1:\mu=1$；若有兩個檢定規則（拒絕域）如下：

規則1：若$\overline{X}_9$＞1.6則拒絕H_0；規則2：若$\overline{X}_{25}$＞1.0則拒絕H_0。

試求：

(一)請分別計算規則1與規則2之檢定力(power)為多少?(8分)
(二)請分別計算規則1與規則2之顯著水準α為多少?(8分)
(三)由(一)及(二)的結果,對所設定的統計假設,那一個檢定規則較佳?為什麼?(6分)

試題評析	本題在考統計假設檢定中有關檢定力與顯著水準,考古題也常考這一類題型,要獲得高分應該不難。
考點命中	《高點・高上統計學講義》第三回,趙治勳編撰,第十一章。

答:

母體: $X \sim N(\mu, \sigma^2 = 9)$

樣本: $X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2 = 9)$

點估計: $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n} = \frac{9}{n})$

可得 $\bar{X}_9 \sim N(\mu, \frac{\sigma^2}{n} = \frac{9}{9} = 1)$ 與 $\bar{X}_{25} \sim N(\mu, \frac{\sigma^2}{n} = \frac{9}{25} = 0.36)$

$H_0: \mu = 0$ vs $H_1: \mu = 1$

(一) 規則1: $power = P(\bar{X}_9 > 1.6 \mid \mu = 1) = P(Z > \frac{1.6-1}{\sqrt{1}}) = P(Z > 0.6) = 0.2743$

規則2: $power = P(\bar{X}_{25} > 1 \mid \mu = 1) = P(Z > \frac{1-1}{\sqrt{0.36}}) = P(Z > 0) = 0.5$

(二) 規則1: $\alpha = P(\bar{X}_9 > 1.6 \mid \mu = 0) = P(Z > \frac{1.6-0}{\sqrt{1}}) = P(Z > 1.6) = 0.0548$

規則2: $\alpha = P(\bar{X}_{25} > 1 \mid \mu = 0) = P(Z > \frac{1-0}{\sqrt{0.36}}) = P(Z > 1.67) = 0.0475$

(三) 規則2較佳。因為兩種規則策略之顯著水準都接近,但規則2之檢定力大於規則1。

三、一項有關職場菸害防制的研究,想比較臺灣北部、中部及南部等三個地區職場菸害防制狀況的差異,分別在北部、中部及南部等三個地區隨機抽訪了400位、300位及300位等總共1000位職場工作者,經整理得資料如下表:

地區 職場菸害防制狀況	北部	中部	南部
全面禁菸	262	190	188
有劃設吸菸區	26	14	20
無禁菸規定	112	96	92
合計	400	300	300

試以顯著水準α=0.05下,檢定北部、中部及南部等三個地區,職場菸害防制狀況有無差異:
(一)可以用那一種統計方法進行檢定?請寫出統計方法之名稱。(5分)
(二)承(一)所選的統計方法,完成該項檢定與結論。(需寫出必須的檢定步驟)(15分)

試題評析	本題考卡方獨立性檢定,相關題型考古題甚多,獲得高分應該不難。
考點命中	《高點・高上統計學講義》第三回,趙治勳編撰,第十三章。

答:

(一) 卡方獨立性檢定

(二)

O_{ij}/E_{ij}	北部中部	中部	南部	
全面禁菸	$262/256$	$190/192$	$180/192$	640
有劃設吸菸區	$26/24$	$14/18$	$20/18$	60
無禁菸規定	$112/120$	$96/90$	$92/90$	300
	400	300	300	1000

H_0 : 三個地區之職場菸害防制狀況無差異

vs

H_1 : 三個地區之職場菸害防制狀況有差異

T.S.: $\chi^2=\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\overset{d}{\sim}\chi^2_{((3-1)(3-1)=4)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $\chi^{2*}>\chi^2_{(4)0.05}=9.49$

Q $\chi^{2*}=3.167$ $\quad\therefore don'trejectH_0$

結論: 我們沒有足夠證據去推論三個地區之職場菸害防制狀況有差異

四、某一速食店經理為瞭解其廣告促銷的效果，乃抽取20家分店做實驗，比較廣告促銷前後的銷售量效果。設 X_i 表示分店廣告促銷前的銷售量，而 Y_i 表示分店廣告促銷後的銷售量。實驗結束後，經統計得相關資料如下：廣告促銷前的平均銷售量 $\overline{x}=1000$ ，變異數 $s_X^2=2500$ ；廣告促銷後的平均銷售量 $\overline{y}=1200$ ，變異數 $s_Y^2=3600$ ；廣告促銷前後銷售量的相關係數 $r_{XY}=0.8$ 。又假設 X_i 與 Y_i 呈現線性相關，其模式為 $Y_i=\alpha+\beta X_i+\varepsilon_i$ ，其中 ε_i 為隨機誤差項，且 $\varepsilon_i\sim N(0,\sigma^2)$ ，i=1,2,...,20。則：

(一)試根據上述資料，以最小平方法估計迴歸直線 $\hat{Y}=\hat{\alpha}+\hat{\beta}X$ 。（10分）

(二)試求迴歸誤差之母體變異數 σ^2 的估計值為何？（8分）

(三)試以顯著水準 $\alpha=0.05$ ，檢定 $H_0:\beta=1$ ， $H_1:\beta\neq1$ 。（10分）

註1：本試題可能使用的統計表之參考值如下：

$t_{0.025}(18)=2.101$ ， $t_{0.025}(19)=2.093$ ， $t_{0.025}(20)=2.086$ ， $t_{0.05}(18)=1.734$ ， $t_{0.05}(19)=1.729$ ， $t_{0.05}(20)=1.725$ ， $\chi^2_{0.025}(4)=11.14$ ， $\chi^2_{0.025}(9)=19.02$ ， $\chi^2_{0.05}(4)=9.49$ ， $\chi^2_{0.05}(9)=16.92$

註2：標準常態機率分配表，Z～N(0,1)

試題評析	本題考簡單迴歸，講義中也有收錄相關題型，考生只要小心計算，獲得高分應該不難。
考點命中	《高點・高上迴歸分析講義》第一回，趙治勳編撰，第二個部份。

答：

(一) $\hat{\beta}=r_{XY}\frac{S_Y}{S_X}=0.8\times\frac{\sqrt{3600}}{\sqrt{2500}}=0.96$

$\hat{\alpha}=\overline{Y}-\hat{\beta}\overline{X}=1200-0.96\times1000=240$

$\therefore\hat{y}=240+0.96x$

(二) $\hat{\sigma}^2=MSE=\frac{SSE}{n-k-1}=\frac{SST-SSR}{n-2}=\frac{(n-1)S_Y^2-\hat{\beta}^2(n-1)S_X^2}{n-2}$

$$=\frac{(20-1)3600-0.96^2(20-1)2500}{20-2}=1368$$

(三) $H_0:\beta=1 \text{ vs } H_1:\beta\neq 1$

T.S.: $T=\dfrac{\hat{\beta}-1}{\sqrt{\dfrac{MSE}{(n-1)S_X^2}}}\sim t_{(20-2=18)}$

R.R.: Reject $\mathrm{H_0}$ at $\alpha=0.05$ if $|T^*|>t_{(18)0.025}=2.101$

$\because |T^*|=0.2357 \quad \therefore don't\ reject H_0$

結論: 我們沒有足夠證據去推論 $H_1:\beta\neq 1$