《統計學》

一、美國職棒大聯盟的世界大賽系列賽是採7戰4勝制，即兩對戰球隊先取得4勝者為世界冠軍。因此，此系列賽最少要打4場而最多要打到7場才能決定世界冠軍。如果某年世界大賽系列賽是由球隊A對上球隊B，且給定每場比賽球隊A贏球機率及球隊B贏球機率皆為0.5，即5-5波，且每場比賽結果皆彼此獨立。如果隨機變數X代表此系列賽總售票數，且總售票數與此系列賽的總比賽場數關係為：總售票數＝6400×（總比賽場數）^{2}，例如，如果此系列賽總比賽場數為4場，則總售票數為6400×4^2=102400。請問本系列賽總比賽場數最可能為幾場以及算出X的期望值E(X)，即本系列賽預計的總售票數。（10分）

試題評析	本題是排列組合的範圍內容，再加上隨機變數期望值之計算。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第四章。

答：

令 Y 表總比賽場數, $y=4,5,6,7$

$$P(Y=4)=\binom{2}{1}0.5^4=0.125$$

$$P(Y=5)=\binom{2}{1}\binom{4}{3}0.5^5=0.25$$

(說明：$\binom{2}{1}$表 A 隊或 B 隊獲勝，$\binom{4}{3}$表前四場裡面某一隊有勝 3 場)

$$P(Y=6)=\binom{2}{1}\binom{5}{3}0.5^6=0.3125$$

$$P(Y=7)=\binom{2}{1}\binom{6}{3}0.5^7=0.3125$$

故本系列賽總比賽場數最有可能是 6 與 7 場

$$E(X)=E(6400Y^2)=6400E(Y^2)=6400\times 34.8125=222800$$

其中 $E(Y^2)=4^2\times 0.125+\cdots+7^2\times 0.3125=34.8125$

二、若一新生產機台其故障時間服從一平均值為λ（天）的指數分配（exponential distribution）。假定 p_1 為此機台連續運轉不故障超過3天的機率，p_2 為給定此機台已連續運轉2天不故障情形下再連續運轉超過1天不故障的條件機率，且 $\log(p_2)-\log(p_1)=4$，其中 $\log(a)$ 為數字a的自然對數值。試算出此生產機台在12小時內故障的機率。（指數e＝2.718）（10分）

試題評析	本題是指數分配的範圍內容，需要先以題目給定資訊計算出參數值，再計算某事件機率，相關題型在題庫中都有練習過。
考點命中	《高點・高上統計學講義》第二回，趙治勳編撰，第七章。

答：

令 X 表故障時間(單位時間：天)

$X\sim Exp(\lambda)$, $E(X)=\lambda$

已知 $p_1 = P(X>3) = e^{-\frac{3}{\lambda}} \Rightarrow \ln p_1 = -\frac{3}{\lambda}$

$p_2 = P(X>3 \mid X>2) \overset{\text{無記憶性質}}{=} P(X>1) = e^{-\frac{1}{\lambda}} \Rightarrow \ln p_2 = -\frac{1}{\lambda}$

又 $\ln p_2 - \ln p_1 = (-\frac{1}{\lambda}) - (-\frac{3}{\lambda}) = \frac{2}{\lambda} = 4 \quad \Rightarrow \lambda = \frac{1}{2}$

$P(X<\frac{1}{2}) = 1 - e^{-\frac{1/2}{1/2}} = 1 - e^{-1} = 0.632$

三、下列是關於最大概似估計量（Maximum likelihood estimator）以及最小變異不偏估計量（minimum variance unbiased estimator）的問題。

（一）考慮下列隨機變數X其機率密度函數（probability density function）：

$$f(x) = \begin{cases} f_1(x), & x \le 1 \\ f_2(x), & x > 1 \end{cases} = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-1)^2}{2\sigma^2}}, & x \le 1 \\ \frac{1}{2\sqrt{2\pi\sigma^2}} e^{\frac{-(x-1)^2}{8\sigma^2}}, & x > 1 \end{cases}$$

其中 $f_1(x)$ 為一平均值為1及標準差為σ的常態機率密度函數（normal probability density function），而 $f_2(x)$ 為一平均值亦為1及標準差為2σ的常態機率密度函數。

下列為一組服從上述機率分配所得的隨機樣本：

2	5	3	0	−2	−3	7

請算出 σ^2 之最大概似估計量的值。（10分）

（二）若某一次國家考試其某考試科目共有25題單選題，隨機變數X代表考生答對題數，且X之分配是每題答對機率為p的二項式分配（binomial distribution），下列是隨機取得6個考生答對題數資料：

12	15	20	10	5	13

根據上述資料，請算出X的標準差之最大概似估計量的值及p的最小變異不偏估計量的值。（10分）

試題評析	本題是 MLE 與 UMVUE 的範圍內容，對於非統計類組的考生來說確實不容易拿到高分。
考點命中	《迴歸分析經典題型》，高點文化出版，趙治勳編著，第一篇第四章。

答：

(一) $L(\sigma^2) = f_1(0) f_1(-2) f_1(-3) f_2(2) f_2(5) f_2(3) f_2(7)$

$$= (\frac{1}{\sqrt{2\pi\sigma^2}})^3 e^{-\frac{1}{2\sigma^2}} e^{-\frac{9}{2\sigma^2}} e^{-\frac{16}{2\sigma^2}} (\frac{1}{2\sqrt{2\pi\sigma^2}})^4 e^{-\frac{1}{8\sigma^2}} e^{-\frac{16}{8\sigma^2}} e^{-\frac{4}{8\sigma^2}} e^{-\frac{36}{8\sigma^2}}$$

$$= 2^{-4} (2\pi\sigma^2)^{-\frac{7}{2}} e^{-\frac{161}{8\sigma^2}}$$

$$\ln L(\sigma^2) = -4\ln 2 - \frac{7}{2}\ln(2\pi\sigma^2) - \frac{161}{8\sigma^2}$$

令 $\dfrac{d\ln L(\sigma^2)}{d\sigma^2} = -\dfrac{7}{2}\dfrac{1}{2\pi\sigma^2}(2\pi) + \dfrac{161}{8\sigma^4} = 0$ 且 $\dfrac{d^2\ln L(\sigma^2)}{d(\sigma^2)^2} < 0$

$\Rightarrow \hat{\sigma}^2 = 5.75$

因此，σ^2 之最大概似估計量值為 5.75

(二)母體：$X \sim Bin(n=25, p)$

樣本：$X_1, X_2, \ldots, X_6 \overset{iid}{\sim} Bin(n=25, p)$

$$L(p) = f(x_1, x_2, \ldots, x_n; p) = \prod_{i=1}^{6}\binom{25}{x_i} p^{\sum x_i}(1-p)^{6\times 25 - \sum x_i}$$

$$\ln L(p) = \ln(\prod_{i=1}^{6}\binom{25}{x_i}) + \sum x_i \ln p + (150 - \sum x_i)\ln(1-p)$$

$$\frac{d\ln L(p)}{dp} = \frac{\sum x_i}{p} - \frac{150 - \sum x_i}{1-p} = 0 \quad 且 \quad \frac{d^2\ln L(p)}{dp^2} < 0$$

$$\Rightarrow \hat{p}_{MLE} = \frac{\sum x_i}{150} = \frac{75}{150} = 0.5$$

由 MLE 不變性

$\sqrt{25\hat{p}_{MLE}(1-\hat{p}_{MLE})} = \sqrt{25\times 0.5\times 0.5} = 2.5$ 為 $\sqrt{V(X)} = \sqrt{25p(1-p)}$ 之最大概似估計量之值

又因為

$f_X(x;p) = \binom{25}{x} p^x (1-p)^{25-x} = \binom{25}{x}(1-p)^{25} e^{\ln(\frac{p}{1-p})\times x}$ 為指數族

且 $w(p) = \ln(\dfrac{p}{1-p})$ 為線性獨立且至少包含一個一維矩形

$\therefore \sum x_i$ 為 p 之完備充分統計量

$$E(\sum X_i) = 6\times 25p \Rightarrow E(\frac{\sum X_i}{150}) = p$$

由 Lehman-Scheffe 得知，$\dfrac{\sum x_i}{150} = \dfrac{75}{150} = 0.5$ 為 p 之 UMVUE 估計值。

四、一保險公司精算人員利用下列簡單線性迴歸模型來建立某美食外送平台，其外送員車險理賠金額（Y）與理賠請求機車騎行里程數（X）之間的關係：

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{，} i = 1, \ldots, 20$$

其中 y_i（單位為千元）為第i件理賠請求的理賠金額，而 x_i（單位為千公里）為第i件理賠請求機車之騎行里程數，共計20件理賠請求；隨機誤差 $\varepsilon_1, \ldots, \varepsilon_{20}$ 為彼此獨立，期望值為0變異數皆為 σ^2 的常態分配。根據上述20件理賠請求，且利用最小平方法（least squares method）來估計 β_0 及 β_1，得到下列資訊：$x_1, \ldots, x_{20}$ 與 $y_1, \ldots, y_{20}$ 的相關係數為0.693，$y_1, \ldots, y_{20}$ 之變異數為 $x_1, \ldots, x_{20}$ 之變異數的0.52倍，估計迴歸關係式為：

$$\hat{Y} = 0.5 + 0.5X$$

(一)請算出判定係數(coefficient of determination) R^2。(2分)

(二)在顯著水準 $\alpha=0.05$,利用F檢定法檢定 $H_0:\beta_1=0$ 對 $H_1:\beta_1\neq 0$。(4分)

(三)在顯著水準 $\alpha=0.05$,利用t檢定法檢定 $H_0:\beta_1\geq 1$ 對 $H_1:\beta_1<1$。(4分)

試題評析	本題是迴歸分析的範圍內容,獲得高分應該不難。
考點命中	《迴歸分析經典題型》,高點文化出版,趙治勳編著,第二篇第二章。

答:

由 $S_Y^2=0.52S_X^2 \Rightarrow \dfrac{S_Y^2}{S_X^2}=\dfrac{(n-1)SS_Y}{(n-1)SS_X}=\dfrac{SS_Y}{SS_X}=0.52$

(一) $R^2=r^2=0.693^2=0.48$

(二) $H_0:\beta_1=0$ vs $H_1:\beta_1\neq 0$

T.S.: $F=\dfrac{n-2}{1}\dfrac{R^2}{1-R^2}\sim F_{(1,18)}$

R.R.:Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(1,18)}=4.41$

$\because F=\dfrac{20-2}{1}\dfrac{0.48}{1-0.48}=16.615 \qquad \therefore reject H_0$

結論:我們有足夠的統計證據去推論斜率不為 0($\beta_1\neq 0$)。

(三) $H_0:\beta_1\geq 1$ vs $H_1:\beta_1<1$

T.S.: $T=\dfrac{\hat{\beta}_1-1}{\sqrt{\dfrac{MSE}{SS_X}}}\sim t_{(18)}$

R.R.:Reject H_0 at $\alpha=0.05$ if $T^*<-t_{0.05(18)}=-1.734$

$$\because T=\frac{\hat{\beta}_1-1}{\sqrt{\dfrac{MSE}{SS_X}}}=\frac{\hat{\beta}_1-1}{\sqrt{\dfrac{\frac{SS_Y(1-r^2)}{n-2}}{SS_X}}}=\frac{\hat{\beta}_1-1}{\sqrt{\dfrac{SS_Y}{SS_X}\dfrac{1-r^2}{n-2}}}=\frac{0.5-1}{\sqrt{0.52\dfrac{1-0.48}{20-2}}}=-4.079$$

$\therefore reject H_0$

結論:我們有足夠的統計證據去推論斜率小於 1($\beta_1<1$)。

五、下列是關於母體平均以及母體比率之估計與檢定的問題:

(一)假定某地區的每日最高溫服從一平均值為μ及標準差為4的常態分配。針對下列假設, $H_0:\mu=28$ 對 $H_1:\mu=30$,隨機取得16筆此地區當日最高溫資料。估計量 $\bar{X}$ 為樣本平均值。給定下列3種檢定法:

檢定法A:若 $\bar{X}>c$ 則拒絕 H_0,反之則不拒絕 H_0;

檢定法B:若 $\bar{X}>30.17$ 則拒絕 H_0,反之則不拒絕 H_0;

檢定法C:若 $\bar{X}<30.5$ 則拒絕 H_0,反之則不拒絕 H_0。

已知檢定法A之檢定力(power)為0.1587,且設定顯著水準為 $\alpha=0.015$,請計算c的值並決定上述3個檢定法那一個或那一些符合設定並有最大檢定力。(10分)

(二)一軟體公司欲比較一新版軟體是否較舊版軟體更能有效率執行程式;假定 μ_1 為舊版軟體執

行測試程式的平均執行時間，而μ_2為新版軟體執行測試程式的平均執行時間，且$x_1,\ldots,x_6$為舊版軟體執行測試程式之執行時間的隨機資料，而$y_1,\ldots,y_6$為新版軟體執行測試程式之執行時間的隨機資料。給定下列資訊：$x_1,\ldots,x_6$之平均值為7且變異數為5.6，而$y_1,\ldots,y_6$之平均值為6且變異數為2.4，$x_1,\ldots,x_6$與$y_1,\ldots,y_6$之相關係數為0。請分別用獨立樣本（independent samples）t檢定法，即將$x_1,\ldots,x_6$與$y_1,\ldots,y_6$視為獨立樣本，及成對樣本（Paired samples or matched samples）t檢定法，即將(x_1,y_1)，⋯，(x_6,y_6)視為成對樣本，在顯著水準$\alpha=0.01$，檢定H_0：$\mu_1 \le \mu_2 - 2.5$對H_1：$\mu_1 > \mu_2 - 2.5$。（10分）

（三）某民調機構欲比較候選人A的支持率p_1與候選人B的支持率p_2的差異。下列是此民調機構關於兩候選人支持度的兩份問卷：

問卷	調查人數	支持此候選人人數
關於候選人A的問卷	4900	2450
關於候選人B的問卷	4900	n

已知根據上述問卷所得p_1-p_2之95%信賴區間其長度為0.0392，以及候選人A的支持人數大於候選人B的支持人數，即$n<2450$。請算出n的值，並利用上述問卷得到p_2的95%信賴區間。（10分）

試題評析	本題是假設檢定的範圍內容，雖然難度不高但題量多，要在短時間內寫出完整且清晰的解答，對考生來說是另一種考驗。
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：

(一)令 X 表每日最高溫

母體：$X \sim N(\mu,4^2)$

樣本：$X_1,X_2,\ldots,X_{16} \overset{iid}{\sim} N(\mu,4^2)$

點估計：$\bar{X} \sim N(\mu,\frac{4^2}{16}=1)$

$H_0:\mu=28$ vs $H_1:\mu=30$

$$power_A = P(\bar{X}>c\mid\mu=30)=P(Z>\frac{c-30}{\sqrt{1}})=P(Z>c-30)$$

$\Rightarrow c-30=z_{0.1587}=1$

$\Rightarrow c=31$

$$\alpha_A=P(\bar{X}>31\mid\mu=28)=P(Z>\frac{31-28}{\sqrt{1}})=P(Z>3)=0.0013$$

$$\alpha_B=P(\bar{X}>30.17\mid\mu=28)=P(Z>\frac{30.17-28}{\sqrt{1}})=P(Z>2.17)=0.015$$

$$\alpha_C=P(\bar{X}<30.5\mid\mu=28)=P(Z<\frac{30.5-28}{\sqrt{1}})=P(Z<2.5)=0.9938$$

只有檢定法 A 與 B 符合設定之顯著水準 0.015

$power_A=0.1587$ (題目已給定)

$power_B = P(\bar{X} > 30.17 \mid \mu = 30) = P(Z > \frac{30.17-30}{\sqrt{1}})$

$= P(Z > 0.17) = 0.4325$

$\because power_A < power_B$

故檢定法 B 具有最大檢定力

(二)令 X_1, X_2 分別表舊版與新版軟體之執行時間

題目已給 $\bar{x}_1 = 7, \bar{x}_2 = 6, S_1^2 = 5.6, S_2^2 = 2.4, r_{12} = 0$

<u>獨立樣本 t 檢定法</u>

假設(1) $X_1 \sim N, X_2 \sim N$ (2) $X_1 \perp X_2$

<註>理應進行 $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$，但考卷未提供 $F_{0.005(5,5)}$，故以下解答以 $\sigma_1^2 \neq \sigma_2^2$ 回答

母體： $X_1 \sim N(\mu_1, \sigma_1^2) \perp X_2 \sim N(\mu_2, \sigma_2^2)$

樣本： $X_{11}, X_{12}, \cdots, X_{16} \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ $X_{21}, X_{22}, \cdots, X_{26} \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$

點估計： $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{6} + \frac{\sigma_2^2}{6})$

$H_0 : \mu_1 - \mu_2 \leq -2.5$ vs $H_1 : \mu_1 - \mu_2 > -2.5$

T.S.： $T = \frac{(\bar{X}_1 - \bar{X}_2) - (-2.5)}{\sqrt{\frac{S_1^2}{6} + \frac{S_2^2}{6}}} \sim t_{(v=9)}$ 其中 $v = \frac{(\frac{S_1^2}{6} + \frac{S_2^2}{6})^2}{\frac{(\frac{S_1^2}{6})^2}{6\text{-}1} + \frac{(\frac{S_2^2}{6})^2}{6\text{-}1}}$ =8.631 ≈ 9

R.R.：Reject H_0 at $\alpha = 0.01$ if $T^* > t_{0.01(9)} = 2.821$

$\because T^* = \frac{(7-6)-(-2.5)}{\sqrt{\frac{5.6}{6} + \frac{2.4}{6}}}$=3.031 $\therefore reject H_0$

結論：我們有足夠的統計證據去推論 $(\mu_1 - \mu_2 > -2.5)$ 。

<u>成對樣本 t 檢定法</u>

假設 $X_1 \sim N, X_2 \sim N$

令 $D = X_1 - X_2$

母體： $D \sim N(\mu_D, \sigma_D^2)$

樣本： $D_1, D_2, \cdots, D_6 \overset{iid}{\sim} N(\mu_D, \sigma_D^2)$

點估計： $\bar{D} \sim N(\mu_D, \frac{\sigma_D^2}{6})$

$H_0 : \mu_1 - \mu_2 \leq -2.5$ vs $H_1 : \mu_1 - \mu_2 > -2.5$

$\Rightarrow H_0 : \mu_D \leq -2.5$ vs $H_1 : \mu_D > -2.5$

T.S.：$T=\dfrac{\bar{D}-(-2.5)}{\dfrac{S_D}{\sqrt{6}}}\sim t_{(6-1=5)}$

R.R.：Reject H_0 at $\alpha=0.01$ if $T^*>t_{0.01(5)}=3.365$

$\because T^*=\dfrac{1-(-2.5)}{\dfrac{\sqrt{8}}{\sqrt{6}}}=3.031$ 其中 $\bar{D}=\bar{X}_1-\bar{X}_2$

$$S_D^2=S_1^2+S_2^2-2r_{12}S_1S_2\overset{r_{12}=0}{=}S_1^2+S_2^2=5.6+2.4=8$$

$\therefore don'trejectH_0$

結論：我們沒有足夠的統計證據去推論$(\mu_1-\mu_2>-2.5)$。

(三) $2z_{0.025}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{4900}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{4900}}=0.0392$

$$\Rightarrow 2(1.96)\sqrt{\frac{\frac{2450}{4900}(1-\frac{2450}{4900})}{4900}+\frac{\frac{n}{4900}(1-\frac{n}{4900})}{4900}}=0.0392$$

$\Rightarrow (4900-n)n=5762400$

$\Rightarrow n^2-4900n+5762400=0$

$$\Rightarrow n=\frac{-(-4900)\pm\sqrt{(-4900)^2-4(1)(5762400)}}{2(1)}$$

$\Rightarrow n=1960$ 及 2940 (不合)

令 X_1,X_2 表支持候選人A與B

母體: $X_2\sim Ber(p_2)$

樣本: $X_1,X_2,\ldots,X_{4900}\overset{iid}{\sim}Ber(p_2)$

點估計: $\hat{p}_2\underset{byC.L.T.}{\sim}N(p_2,\dfrac{p_2(1-p_2)}{4900})$

樞紐量: $\dfrac{\hat{p}_2-p_2}{\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{4900}}}\underset{byC.L.T.}{\sim}N(0,1)$

機率區間: $P(-z_{0.025}\le\dfrac{\hat{p}_2-p_2}{\sqrt{\dfrac{\hat{p}_2(1-\hat{p}_2)}{4900}}}\le z_{0.025})=0.05$

結論: p_2 之95%C.I.為

$$(\hat{p}_2\mp z_{0.025}\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{4900}})=(\frac{1960}{4900}\mp1.96\sqrt{\frac{\frac{1960}{4900}(1-\frac{1960}{4900})}{4900}})$$
$$=(0.3863,0.4137)$$

六、欲比較3種品牌汽車其耗油程度，給定A品牌汽車其每公升行駛平均里程數為μ_1，B品牌汽車其每公升行駛平均里程數為μ_2，而C品牌汽車其每公升行駛平均里程數為μ_3。每種品牌各隨機抽測8台汽車並得到其每公升行駛里程數。下列是關於此3組樣本其相關資訊以及利用這3組樣本所做的統計分析：
此3組樣本其平均值為A品牌之平均里程數最大，而B品牌之平均里程數最小，此3組樣本其標準差皆相同，檢定 $H_0：\mu_1=\mu_2$ 對 $H_1：\mu_1\neq\mu_2$ 的獨立樣本t檢定其t統計量絕對值為2，檢定 $H_0：\mu_2=\mu_3$ 對 $H_1：\mu_2\neq\mu_3$ 的獨立樣本t檢定其t統計量絕對值為1。在顯著水準$\alpha=0.05$，請利用單因子變異數分析法（one-way ANOVA）來檢定此3種品牌汽車其耗油程度是否一致，即檢定 $H_0：\mu_1=\mu_2=\mu_3$ 對 $H_1：\mu_1$、μ_2 以及 μ_3 並不完全相等。（10分）

試題評析	本題是一因子變異數分析的範圍內容，題目所給的資訊並非考古題出現過的，非常考驗考生的臨場反應，且也需要對於變異數分析理論非常熟悉。
考點命中	《高點・高上統計學講義》第三回，趙治勳編撰，第十二章。

答：

題目已給 $n_1=n_2=n_3=8$ ，$\bar{Y}_1>\bar{Y}_3>\bar{Y}_2$ ，令 $S_1=S_2=S_3=S$

$H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$

$$\text{T.S.}：|T|=|\frac{(\bar{Y}_1-\bar{Y}_2)-(0)}{\sqrt{\frac{S_1^2}{8}+\frac{S_2^2}{8}}}|=|\frac{\bar{Y}_1-\bar{Y}_2}{\sqrt{\frac{S^2}{4}}}|=2\Rightarrow\bar{Y}_1-\bar{Y}_2=2\sqrt{\frac{S^2}{4}}\text{ -------(1)}$$

$H_0:\mu_2=\mu_3$ vs $H_1:\mu_2\neq\mu_3$

$$\text{T.S}：|T|=|\frac{(\bar{Y}_2-\bar{Y}_3)-(0)}{\sqrt{\frac{S_2^2}{8}+\frac{S_3^2}{8}}}|=|\frac{\bar{Y}_2-\bar{Y}_3}{\sqrt{\frac{S^2}{4}}}|=1\Rightarrow\bar{Y}_2-\bar{Y}_3=-\sqrt{\frac{S^2}{4}}\text{ -------(2)}$$

(1)+(2)可得 $\bar{Y}_1-\bar{Y}_3=\sqrt{\frac{S^2}{4}}$

(1)/(2)可得 $\bar{Y}_1+\bar{Y}_2=2\bar{Y}_3$

ANOVA TABLE

source	SS	d.f.	MS	F
品牌	$4S^2$	2	$2S^2$	F^*=2
Error	$21S^2$	21	S^2	
Total	$25S^2$	23		

$$SSE=\sum(n_i-1)S_i^2=(8-1)3S^2=21S^2$$

$$SSR=\sum n_i(\bar{Y}_i-\bar{\bar{Y}})^2=8(\bar{Y}_1-\bar{Y}_3)^2+8(\bar{Y}_2-\bar{Y}_3)^2+8(\bar{Y}_3-\bar{Y}_3)^2$$

$$=8(\sqrt{\frac{S^2}{4}})^2+8(-\sqrt{\frac{S^2}{4}})^2=4S^2$$

其中 $\bar{\bar{Y}}=\frac{\bar{Y}_1+\bar{Y}_2+\bar{Y}_3}{3}=\frac{2\bar{Y}_3+\bar{Y}_3}{3}=\bar{Y}_3$

假設模型：$y_{ij}=\mu+\alpha_i+\varepsilon_{ij}, \varepsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2)$ ，$i=1,2,3$，$j=1,2,\cdots,8$

$H_0:\mu_1=\mu_2=\mu_3$ vs H_1：至少一個 $\mu_i\neq\mu_j$

T.S.：$F=\dfrac{MSR}{MSE}\sim F_{(2,21)}$

R.R.：Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(2,21)}=3.47$

$\because F^*=2$　　$\therefore don't\ reject H_0$

結論：我們沒有足夠的統計證據去推論 3 種品牌汽車之耗油程度有顯著差異

七、根據一份國人讀報偏好之報告指出，最常看A報的比率為 $p_1=0.25$，B報的比率為 $p_2=0.2$，C報的比率為 $p_3=0.15$，D報的比率為 $p_4=0.15$，而其他報的比率為 $p_5=0.25$。下列是一份關於民眾看報的問卷資料：

最常看報紙	A報	B報	C報	D報	其他報
人數	225	215	160	125	275

在顯著水準α＝0.1，利用上述問卷資料以及卡方檢定（chi-squared test）來驗證上述報告是否可靠，即檢定：

H_0：$p_1=0.25$、$p_2=0.2$、$p_3=0.15$、$p_4=0.15$、$p_5=0.25$ 對

H_1： 並非 $p_1=0.25$、$p_2=0.2$、$p_3=0.15$、$p_4=0.15$、$p_5=0.25$。（10分）

試題評析	本題是卡方適合度檢定的範圍內容，這一類題型是考古題的常客之一，考生獲得高分不難。
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：

	A	B	C	D	其他
O_i	225	215	160	125	275
E_i	250	200	150	150	250

$H_0:p_1=0.25, p_2=0.2, p_3=0.15, p_4=0.15, p_5=0.25$ vs H_1：not H_0

T.S.：$\chi^2=\sum_{i=1}^{5}\dfrac{(O_i-E_i)^2}{E_i}\overset{d}{\sim}\chi^2_{(5-1-0=4)}$

R.R.：Reject H_0 at $\alpha=0.1$ if $\chi^{2*}>\chi^2_{(4)0.1}=7.779$

$\because \chi^{2*}=9.958$　　$\therefore reject H_0$

結論：我們有足夠的統計證據去推論各報之最常看比率並非

$p_1=0.25, p_2=0.2, p_3=0.15, p_4=0.15, p_5=0.25$