《統計學概要(統計)》

參考值：

$z_{0.006}=2.51$，$z_{0.01}=2.33$，$z_{0.025}=1.96$，$z_{0.05}=1.65$，$z_{0.1}=1.28$，

$t_{0.025,8}=2.31$，$t_{0.025,9}=2.26$，$t_{0.05,8}=1.86$，$t_{0.05,9}=1.83$，

$\chi^2_{3,0.025}=9.35$，$\chi^2_{4,0.025}=11.14$，$\chi^2_{5,0.025}=12.83$，$\chi^2_{3,0.05}=7.81$，$\chi^2_{4,0.05}=9.49$，$\chi^2_{5,0.05}=11.07$

$\chi^2_{3,0.1}=6.25$，$\chi^2_{4,0.1}=7.78$，$\chi^2_{5,0.1}=9.24$

一、X族人之平均身高為160公分，標準差為10公分。某製造商欲生產一款長度為200公分之床墊。假設床墊之長度必須比身長多出至少15公分方能讓使用者感覺舒適。

(一)若不知X族人身高之分配為何，試問約有多少比例族人使用此床墊感覺舒適？（10分）

(二)若X族人之身高服從常態分配，試問感覺舒適之比例為何？（10分）

試題評析	本題是考單邊柴比雪夫不等式與常態分配之計算題型，兩者都是考古題裡面的常客，要獲得高分不難。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第五章與第七章。

答：

令 X 表X族人之身高(公分) $X\sim(\mu=160,\sigma^2=10^2)$

(一) 由單邊柴比雪夫不等式 $P(X\le\mu-k)\ge\dfrac{k^2}{\sigma^2+k^2}$，$k<0$

$$P(X+15\le 200)=P(X\le 185)=P(X\le 160-(-25))\ge\frac{(-25)^2}{10^2+(-25)^2}=0.862$$

X族人至少有86.2%使用此床墊感覺舒適

(二) $X\sim N(\mu=160,\sigma^2=10^2)$

$$P(X+15\le 200)=P(X\le 185)=P(Z\le\frac{185-160}{10})$$

$$=P(Z\le 2.5)=1-0.006=0.994$$

X族人至少有99.4%使用此床墊感覺舒適

二、甲公司有4名員工，月薪分別為4萬、6萬、8萬、8萬。經由簡單計算已經求出母體平均數(μ)為6.5萬，母體變異數(σ^2)為2.75萬。今擬以抽後不放回的方式抽取2名員工，得樣本平均數 $\bar{X}_2$。將 $\bar{X}_2$ 的平均數與變異數分別記為 $\mu\bar{X}_2$ 與 $(\sigma\bar{X}_2)^2$。

(一)試寫出 $\bar{X}_2$ 之機率分配，並依此計算 $\mu\bar{X}_2$ 與 $(\sigma\bar{X}_2)^2$。（10分）

(二)假設母體個數為N，以抽後不放回的方式抽取n個值，將樣本平均數之變異數記為 $(\sigma\bar{X}_n)^2$。在考量有限母體修正係數之下，試寫出 $(\sigma\bar{X}_n)^2$ 與 σ^2 之關係式。（5分）

(三)當樣本數n遠小於母體數N時，試問(二)中之修正係數近似何值？（5分）

試題評析	本題是考樣本平均數之抽樣分配，講義都有相關例題，小心計算，獲得高分應該不難
考點命中	《高點・高上統計學講義》第二回，趙治勳撰，第九章例1。

答：

(一)

樣本	$\bar{X}_2$	機率
4，6	5	$2\times\frac{1}{4}\times\frac{1}{3}=\frac{1}{6}$
4，8	6	$2\times\frac{1}{4}\times\frac{2}{3}=\frac{1}{3}$
6，8	7	$2\times\frac{1}{4}\times\frac{2}{3}=\frac{1}{3}$
8，8	8	$\frac{2}{4}\times\frac{1}{3}=\frac{1}{6}$

$$\mu_{\bar{X}_2}=E(\bar{X}_2)=5(\frac{1}{6})+6(\frac{1}{3})+7(\frac{1}{3})+8(\frac{1}{6})=6.5$$

$$(\sigma_{\bar{X}_2})^2=V(\bar{X}_2)=E(\bar{X}_2{}^2)-[E(\bar{X}_2)]^2=43\frac{1}{6}-[6.5]^2=0.9167$$

其中 $E(\bar{X}_2{}^2)=5^2(\frac{1}{6})+6^2(\frac{1}{3})+7^2(\frac{1}{3})+8^2(\frac{1}{6})=43\frac{1}{6}$

(二)

在抽後不放回下，母體變異數為 $\sigma^2=\dfrac{\sum_{i=1}^{N}(X_i-\mu)^2}{N-1}$

故 $(\sigma_{\bar{X}_n})^2=(1-\frac{n}{N})\frac{\sigma^2}{n}$

(三) 當n遠小於N時，修正係數 $1-\frac{n}{N}$ 近似於1

三、甲乙丙為三名職棒左打者，下表為去年此三名打者面對左投手與右投手時之表現。

	打數（左投）	安打數（左投）	打數（右投）	安打數（右投）
甲	224	63	532	116
乙	245	49	567	238
丙	231	35	525	147

棒球打擊率之計算公式如下：打擊率＝安打數/打數

(一)試分別計算甲乙丙於左投時之打擊率的95%信賴區間，並依此兩兩比較（甲、乙）、（乙、丙）、（甲、丙）面對左投手時之打擊率是否不同？（10分）

(二)在顯著水準0.1下，試檢定甲打者面對左投與右投時，打擊率是否不同？（10分）

試題評析	本題(一)要以兩兩信賴區間比較是否存在顯著差異，雖然沒有考過這樣的比較法，但只要熟知信賴區間的意義，考生應該可以推測如何進行比較的。另外(二)是兩獨立母體成功比例差之假設檢定，屬於基本計算題，獲得滿分不難
考點命中	《高點・高上統計學講義》第三回，趙治勳撰，第十一章。

答：

(一) 令 p_1, p_2, p_3 分別表甲，乙，丙之打擊率

甲於左投時打擊率之95%信賴區間為

$$(\hat{p}_1 \mp z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}) = (\frac{63}{224} \mp 1.96\sqrt{\frac{\frac{63}{224}\frac{161}{224}}{224}}) = (0.2224, 0.3401)$$

乙於左投時打擊率之95%信賴區間為

$$(\hat{p}_2 \mp z_{0.025}\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}) = (\frac{49}{245} \mp 1.96\sqrt{\frac{\frac{49}{245}\frac{196}{245}}{245}}) = (0.1499, 0.2501)$$

丙於左投時打擊率之95%信賴區間為

$$(\hat{p}_3 \mp z_{0.025}\sqrt{\frac{\hat{p}_3(1-\hat{p}_3)}{n_3}}) = (\frac{35}{231} \mp 1.96\sqrt{\frac{\frac{35}{231}\frac{196}{231}}{231}}) = (0.1053, 0.1978)$$

題意是要利用以上各位左打者左投時打擊率之信賴區間進行兩兩比較，故判斷準則為兩者信賴區間是否包括彼此之點估計。各位左打者左投時打擊率之估計值為 $\hat{p}_1 = 0.28125, \hat{p}_2 = 0.2, \hat{p}_3 = 0.152$ ，可得（甲，乙）左投時打擊率不同，（乙，丙）左投時打擊率不同，（甲，丙）左投時打擊率不同。

(二) 令 X_1, X_2 分別表甲打者面對左投與右投時打出安打

母體： $X_1 \sim Ber(p_1) \perp X_2 \sim Ber(p_2)$ 　假設(1) $X_1 \perp X_2$ (2)隨機樣本

樣本： $X_{11}, X_{12}, \ldots, X_{1224} \overset{iid}{\sim} Ber(p_1)$ 　 $X_{21}, X_{22}, \ldots, X_{2532} \overset{iid}{\sim} Ber(p_2)$

點估計： $\hat{p}_1 - \hat{p}_2 \underset{byC.L.T.}{\sim} N(p_1 - p_2, \frac{p_1(1-p_1)}{224} + \frac{p_2(1-p_2)}{532})$

$H_0 : p_1 = p_2$ 　vs　 $H_1 : p_1 \neq p_2$

T.S.： $Z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (0)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{224} + \frac{1}{532})}} \underset{byC.L.T.}{\sim} N(0,1)$ 　其中 $\hat{p} = \dfrac{63+116}{224+532} = 0.2368$

R.R.： Reject H_0 at $\alpha = 0.1$ if $|Z^*| > z_{0.05} = 1.96$

$$\because Z^* = \frac{(\frac{63}{224} - \frac{116}{532}) - (0)}{\sqrt{0.2368(0.7632)(\frac{1}{224} + \frac{1}{532})}} = 1.867 \qquad \therefore don't reject H_0$$

結論： 我們沒有足夠證據去推論甲打者面對左投與右投時打擊不同

四、A辦事處使用抽號機供訪客抽取號碼，並使用叫號機呼叫訪客前往櫃檯接受服務。根據過去一週抽號機與叫號機之紀錄，製作訪客到訪之等候時間（分鐘）與人次如下：

等候時間	（0， 5）	[5， 10）	[10， 15）	[15， 20）	[20， 30）
人次	147	65	22	9	7

(一)試求平均等候時間。（5分）

(二)在顯著水準0.1下，試檢定等候時間是否服從指數分配。（期望值為λ的指數分配之機率密度函數為$f(x)=(1/\lambda)e^{-x/\lambda}$， $x\geq 0$；累積機率分配函數為$F(x)=1-e^{-x/\lambda}$， $x\geq 0$。）（10分）

試題評析	本題是卡方適合度檢定，相關例題於統計題庫中都有，小心計算，獲得高分應該不難
考點命中	《高點・高上統計學講義》第三回，趙治勳撰，第十三章。

答：

(一) $$\text{平均等候時間}=\frac{2.5(147)+7.5(65)+12.5(22)+17.5(9)+25(7)}{147+65+22+9+7}=5.85\ (\text{分鐘})$$

(二)

等候時間	(0，5)	[5，10)	[10，15)	[15，20)	20以上
O_i	147	65	22	9	7
機率 p_i	$1-e^{-\frac{5}{5.85}}=0.575$	$e^{-\frac{5}{5.85}}-e^{-\frac{10}{5.85}}=0.244$	0.104	0.044	0.033
$E_i=250p_i$	143.75	61	26	11	8.25

H_0:等候時間服從指數分配 vs H_1:等候時間不服從指數分配

T.S.： $\chi^2=\sum_{i=1}^{5}\frac{(O_i-E_i)^2}{E_i}\overset{d}{\sim}\chi^2_{(5-1-1=3)}$

R.R.： Reject H_0 at $\alpha=0.1$ if $\chi^{2*}>\chi^2_{(3)0.1}=6.25$

$\because \chi^{2*}=1.504$ $\quad\therefore don't\,reject H_0$

結論： 我們沒有足夠證據去推論等候時間不服從指數分配

五、賈先生欲購買一輛油電混合二手車。他蒐集一組隨機樣本，其中車齡（年）與價格（萬元）之資料如下：

X：車齡	8	3	6	5	5	2	8	10	9	8
Y：價格	55	145	82	68	100	140	35	40	65	70

已經算出$S_{xx}=\Sigma(x-\bar{x})^2=62.4$，$S_{yy}=\Sigma(y-\bar{y})^2=12948$，$\Sigma(x-\bar{x})(y-\bar{y})=-808$。

(一)在顯著水準0.05下，試檢定車齡與價格之相關係數是否為負值。（5分）

(二)考慮以價格為依變數、車齡為因變數之線性回歸模型。試求出回歸方程式，並於顯著水準0.05下，試檢定斜率係數是否為負值。（10分）

(三)試比較(一)(二)之檢定統計量之異同。（5分）

(四)試求出判定係數，並說明其意義。（5分）

試題評析	本題是簡迴歸之計算題型，屬於常考題型，小心計算，獲得高分不難
考點命中	《迴歸分析申論題完全制霸》高點文化出版，趙治勳編著，第八章第六節。

答：

假設模型 $Y_i=\beta_0+\beta_1 X_i+\varepsilon_i, \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2), i=1,2,\cdots,10$

(一)

$$r_{XY}=\frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}}=\frac{-808}{\sqrt{62.4}\sqrt{12948}}=-0.8989$$

$H_0:\rho_{XY}\ge 0$ vs $H_1:\rho_{XY}<0$

T.S.： $T=\dfrac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}\sim t_{(10-2=8)}$

R.R.： Reject H_0 at $\alpha=0.05$ if $T^*<-t_{(8)0.05}=-1.86$

$\because T^*=\dfrac{-0.8989\sqrt{10-2}}{\sqrt{1-0.8989^2}}=-5.803$ $\quad\therefore reject H_0$

結論： 我們有足夠證據去推論車齡與價格之相關係數為負值

(二)

$\hat{\beta}_1=\dfrac{SS_{XY}}{SS_X}=\dfrac{-808}{62.4}=-12.9487$ $\qquad \hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}=80-(-12.9487)6.4=162.87168$

$\therefore \hat{y}=162.87168-12.9487x$

$H_0:\beta_1\ge 0$ vs $H_1:\beta_1<0$

T.S.： $T=\dfrac{\hat{\beta}_1-0}{\sqrt{\dfrac{MSE}{SS_X}}}\sim t_{(10-2=8)}$

R.R.： Reject H_0 at $\alpha=0.05$ if $T^*<-t_{(8)0.05}=-1.86$

$\because T^*=\dfrac{(-12.9487)-0}{\sqrt{\dfrac{310.6831}{62.4}}}=-5.803$ $\quad\therefore reject H_0$

其中 $MSE=\dfrac{SSE}{n-2}=\dfrac{SS_Y-\hat{\beta}_1^2 SS_X}{n-2}=\dfrac{12948-(-12.9487)^2 62.4}{10-2}=310.6831$

結論： 我們有足夠證據去推論斜率係數為負值

(三)

兩個檢定統計量為相同的，因為

$$T=\frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}=\hat{\beta}_1\sqrt{\frac{SS_X}{SS_Y}}\frac{\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}=\frac{\hat{\beta}_1\sqrt{SS_X}}{\sqrt{\dfrac{SS_Y(1-r_{XY}^2)}{n-2}}}=\frac{\hat{\beta}_1-0}{\sqrt{\dfrac{SS_Y(1-r_{XY}^2)/(n-2)}{SS_X}}}=\frac{\hat{\beta}_1-0}{\sqrt{\dfrac{MSE}{SS_X}}}$$

(四)

$R^2=r_{XY}^2=(-0.8989)^2=0.80802$

表示考慮之自變數車齡與模型可以解釋價格(Y)之變異達到80.802%