《統計學概要》

一、請依序回答下列問題：

(一)直方圖（Histogram）與長條圖（Bar graph）的差異。（5分）

(二)某年普考統計學考試平均成績為76分。假設此次統計學考試成績呈鐘形分布且其標準差為8分，計算及格（亦即超過60分）人數的比例。（10分）

(三)李小明參加A、B及C三種考試，分別所獲之分數及各類考試分數之平均及標準差如下，試問李小明那一種考試所獲得之分數具有最高相對位置？（15分）

考試	李小明獲得分數	所有應考者	
		平均分數	分數標準差
A	4.3	5	0.5
B	32	46	1.5
C	63	80	20

試題評析	本題是考第三章敘述統計學的範圍內容，包含：直方圖、長條圖、經驗法則與標準化z分數，本題難度在於題目沒有註明是考什麼，考生只要能夠正確判斷出來，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》第一回，趙治勳編撰，第三章。

答：

(一)

長條圖	直方圖
針對屬質資料繪製	針對屬量資料繪製
屬質中的名目尺度資料針對累積次數繪圖無意義	所有的屬量資料皆可針對累積次數繪圖
僅呈現各分類下之次數	可呈現資料之分佈情形
長條之個數等於分類數	長條之個數(組數)非唯一
長條間要留空隙	長條間緊緊相扣

(二)根據經驗法則，約有95%之數據落在區間

$(\mu-2\sigma,\mu+2\sigma)$=(76-2×8,76+2×8)=(60,88)內，再由成績分布為對稱鐘形分配，故得到及格之人數比例約為1-0.05/2=0.975

(三)令 z_i 表考試i中李小明成績之z分數

$$z_A=\frac{4.3-5}{0.5}=-1.4,\ z_B=\frac{32-46}{1.5}=-0.9333,z_C=\frac{63-80}{20}=-0.85$$

由於考試C之z分數最大，故考試C所獲得之分數具有最高相對位置。

二、畢業學生透過人力資源公司向公司A及公司B寄出履歷找工作，若已知獲得公司A面試的機率為0.32，獲得公司B面試的機率為0.2。假設獲得公司A及公司B面試為獨立事件，請求出下列事件之機率：（每小題5分，共25分）

(一)申請者同時獲得兩家公司面試。

(二)申請者至少獲得一家公司面試。

(三)沒有任何公司提供申請者面試。

(四)公司A沒有提供申請者面試，但公司B有提供申請者面試。

(五)何謂機率加法法則（Addition Rule of Probability）？

試題評析	本題是在考事件間獨立與聯集加法公式，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，第四章。

答：
令A,B分別表獲得公司A,B之面試機會：
(一) $P(A\cap B)=P(A)P(B)=0.32\times 0.2=0.064$
(二) $P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.32+0.2-0.064=0.456$
(三) $P(A^c\cap B^c)=P(A^c)P(B^c)=0.68\times 0.8=0.544$
(四) $P(A^c\cap B)=P(A^c)P(B)=0.68\times 0.2=0.136$
(五)機率加法法則： $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

三、保險公司汽機車理賠部門相信年輕人發生汽機車意外的可能性較高，因此應該要求較高之汽機車保險費率。他們調查1,200位汽機車保險要保人過去三年對保險公司求償記錄及要保人年齡資料如下表，試問汽機車保險要保人求償與否與要保人年齡是否有關？顯著水準為0.05情況下，請使用：（每小題10分，共20分）
(一)臨界值法。
(二)p值（p-value）法。

年齡（X）	沒有求償	求償
18 < X ≤ 25	187	66
25 < X ≤ 40	228	58
40 < X ≤ 55	380	62
X > 55	186	33

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) H_0 :求償與否與要保人年齡無關 vs H_1 :求償與否與要保人年齡有關

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

R.R.: Reject H_0 at $\alpha=0.05$ if $\chi^{2*}>\chi^2_{(3)0.05}=7.815$

$\because \chi^{2*}=17.9737 \qquad \therefore rejectH_0$

結論：我們有足夠證據去推論求償與否與要保人年齡有關。

(二) $p-value=P(\chi^2_{(3)}>17.9737)<0.005$

R.R.: Reject H_0 at $\alpha=0.05$ if $\alpha>p-value$

$\because p-value<0.005 \qquad \therefore rejectH_0$

結論：我們有足夠證據去推論求償與否與要保人年齡有關。

四、某跨國企業想了解公司各部門女性及男性員工人數間的關係，隨機選取員工，各部門及不同性別員工數如下。試問：

(一)顯著水準為0.05情況下，檢定各部門女性及男性員工人數是否有關係？(15分)
(二)當企業某部門女性員工數為21位時，請預測男性員工人數及其95%預測區間。(10分)

部門	女性	男性
產品研發部	35	65
業務部	51	204
法務部	29	101
人事部	5	12
財務部	12	43
會計部	4	6
稽核部	10	32

試題評析	本題考簡迴歸，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：
令X,Y分別表女性與男性員工之人數

假設 $Y_i=\beta_0+\beta_1X_i+\varepsilon_i$，$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$，$i=1,2,\cdots,7$

$$\sum X_i=146,\ \sum X_i^2=4953,\ \sum Y_i=463,\ \sum Y_i^2=59095,\ \sum X_iY_i=16528$$

$$SS_X=\sum X_i^2-\frac{(\sum X_i)^2}{n}=4953-\frac{(146)^2}{7}=1907.8571$$

$$SS_Y=\sum Y_i^2-\frac{(\sum Y_i)^2}{n}=59095-\frac{(463)^2}{7}=28470.8571$$

$$SS_{XY}=\sum X_iY_i-\frac{(\sum X_i)(\sum Y_i)}{n}=16528-\frac{(146)(463)}{7}=6871.1429$$

(一) $\hat{\beta}_1=\frac{SS_{XY}}{SS_X}=3.6015$，$MSE=\frac{SSE}{n-2}=\frac{SS_Y-\hat{\beta}_1^2SS_X}{n-2}=744.884$

$H_0:\beta_1=0$ vs $H_1:\beta_1\neq 0$

T.S.: $T=\frac{\hat{\beta}_1-0}{\sqrt{\frac{MSE}{SS_X}}}\sim t_{(7-2=5)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $|T^*|>t_{0.025(5)}=2.571$

$\because T=\frac{3.6015-0}{\sqrt{\frac{744.884}{1907.8571}}}=5.7638$ $\therefore reject H_0$

結論：我們有足夠證據去推論女性與男性員工人數是有相關的。

(二) $\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}=-8.9741$ $\therefore \hat{y}=-8.9741+3.6015x$

當女性員工人數為 $x_0=21$ 時，男性員工之平均人數預測為

$\therefore \hat{y}=-8.9741+3.6015\times 21=66.6574$ 人

95%預測區間為

$$(\hat{\beta}_0+\hat{\beta}_1 x_0)\mp t_{\frac{\alpha}{2}(n-2)}\sqrt{MSE(1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{SS_X})}$$

$$=(66.6574)\mp 2.571\sqrt{744.884(1+\frac{1}{7}+\frac{(21-20.8571)^2}{1907.8571})}=(-8.357,141.6718)$$