《統計學概要》

試題評析	第一大題古典機率是要花點時間想一下的題目，剩下的二到五題分別為常態分配的應用、抽樣分配、敘述統計學的計算和迴歸分析，算是基本題型，一定要把握。

一、今從 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 等十個數字中任意選出三個不同的數字，定義下列 A，B，C 事件為 A＝三個數字中不含 0 和 9；B＝三個數字中不含 0 或 9；C＝三個數字中含 0 但不含 9。
試求下列事件的機率：
(一) A 事件的機率 P(A)＝？（5 分）
(二) B 事件的機率 P(B)＝？（5 分）
(三) C 事件的機率 P(C)＝？（5 分）

考點命中	《高點・高上統計學講義》第一回，蘇建郎編撰，2-2機率運算與性質，頁24。

答：
(一) A：三個數字不含(0和9)，即三個數字沒0或沒9

$$P(\mathrm{A})=1-\frac{C_2^2C_1^8}{C_3^{10}}=1-\frac{8}{120}=\frac{112}{120}=0.9333$$

(二) B：三個數字不含(0或9)，即三個數字沒0 且 沒9

$$P(\mathrm{B})=\frac{C_0^2C_3^8}{C_3^{10}}=\frac{56}{120}=0.4666$$

(三)C： $P(\mathrm{C})=\frac{C_1^1C_0^1C_2^8}{C_3^{10}}=\frac{28}{120}=0.2333$

二、某甲每天 8 點 10 分準時出門上班，設他由家裡到臺北車站所需之時間（單位：分鐘）呈平均數 $\mu=30$，變異數為 $\sigma^2=20$ 的常態分配，由臺北車站轉車至公司辦公室所需之時間（單位：分鐘）亦呈平均數 $\mu=14$，變異數為 $\sigma^2=16$ 的常態分配：
(一) 求某甲能趕上公司規定 9 點到達辦公室上班的機率為何？（5 分）
(二) 求某甲可能遲到 5 分鐘以上的機率為何？（5 分）

考點命中	《高點・高上統計學講義》第三回，蘇建郎編撰，4-2一維常用分配連續型，頁31。

答：
(一)設隨機變數 X 為由家裡到臺北車站之時間，即 $X \sim N(30,20)$
隨機變數 Y 為由臺北車站到公司之時間，即 $Y \sim N(14,16)$
令家裡到公司為 $W = X+Y \sim N(44,36)$
故9點前到公司的機率為 $P(W<50)=P(Z<\frac{50-44}{\sqrt{36}})=P(Z<1)=0.8413$

(二)遲到5分鐘以上的機率為 $P(W>55)=P(Z>\frac{55-44}{\sqrt{36}})=P(Z>1.83)=0.0336$

三、下列為抽自一個平均數為 μ，變異數為 σ^2 之母體的一組 $n=9$ 之隨機樣本資料：

43, 32, 54, 40, 44, 30, 41, 46, 39

請根據上列資料，回答下列問題：

(一) 試求此組資料之平均數 A 與中位數 B 分別為何？（6 分）

(二) 試求此組資料之四分位距 $IQR(=Q_3-Q_1)=$？（其中 $Q_1=$ 第一四分位數及 $Q_3=$ 第三四分位數）（5 分）

(三) 又是否可根據(二)的資訊判斷出此組資料有離群值（outlier）存在？請說明你（妳）的依據。（4 分）

(四) 若此母體為一常態母體，試以此組樣本資料之訊息，求 μ 之 90%的信賴區間為何？（5 分）

考點命中	《高點・高上統計學講義》第一回，蘇建郎編撰，1-2 統計量數，頁8。

答：

(一)資料有n=9筆 由小到大排序如下： 30,32,39,40,41,43,44,46,54

資料平均數 $A=\bar{x}=\dfrac{30+32+39+40+41+43+44+46+54}{9}=\dfrac{369}{9}=41$

n=9奇數，資料中位數 $B=X_{(\frac{9+1}{2})}=X_{(5)}=41$

(二) $9\times\dfrac{1}{4}=2.25$ 為非整數，故 $Q_1=X_{(3)}=39$

$9\times\dfrac{3}{4}=6.75$ 為非整數，故 $Q_3=X_{(7)}=44$

$IRQ=Q_3-Q_1=44-39=5$

(三)當資料在 $(39-1.5\times5\text{，}44+-1.5\times5)=(31.5\text{，}51.5)$ 範圍內則非離群值。

故離群值為30,54

(四) $n=9$，$\bar{x}=41$，$s^2=51.75$，$1-\alpha=0.9$，$t_{\alpha/2}(n-1)=t_{0.05}(8)=1.86$

$\bar{x}\pm t_{\alpha/2}(n-1)\sqrt{\dfrac{s^2}{n}}\Rightarrow 41\pm1.86\sqrt{\dfrac{51.75}{9}}\Rightarrow(36.54,45.46)$

此為母體平均數 μ 的90%信賴區間。

四、有一隨機變數 X 之機率分配如下：

x	1	5	9
$f(x)$	0.2	0.5	0.3

令 X_1 與 X_2 分別表由此母體以隨機所抽取的 $n=2$ 個樣本，令 $\bar{X}=\dfrac{X_1+X_2}{2}$，則：

(一) 試列出此 $n=2$ 之所有可能的樣本組合與其對應的被抽出機率。（10 分）

(二) 試求出樣本平均數（$\bar{X}$）的抽樣分配，並求出 $\bar{X}$ 之期望值 $E(\bar{X})$ 與 $\bar{X}$ 之變異數 $V(\bar{X})$ 之值。（10 分）

(三) 試求 $P(\bar{X}>5)=$？（5 分）

考點命中	《高點・高上統計學講義》第四回，蘇建郎編撰，5-2抽樣分配，頁10。

答：

(一)

可能組合 (x_1,x_2)	$\overline{x}=\frac{x_1+x_2}{2}$	發生機率
(1,1)	1	$0.2\times0.2=0.04$
(1,5)	3	$0.2\times0.5=0.1$
(1,9)	5	$0.2\times0.3=0.06$
(5,1)	3	$0.5\times0.2=0.1$
(5,5)	5	$0.5\times0.5=0.25$
(5,9)	7	$0.5\times0.3=0.15$
(9,1)	5	$0.3\times0.2=0.06$
(9,5)	7	$0.3\times0.5=0.15$
(9,9)	9	$0.3\times0.3=0.09$

(二)

$\overline{x}$	1	3	5	7	9
$f_{\overline{X}}(\overline{x})$	0.04	0.2	0.37	0.3	0.09

$E(\overline{X})=\sum\overline{x}f_{\overline{X}}(\overline{x})=1\times0.04+3\times0.2+5\times0.37+7\times0.3+9\times0.09=5.4$

$E(\overline{X}^2)=\sum\overline{x}^2f_{\overline{X}}(\overline{x})=1^2\times0.04+3^2\times0.2+5^2\times0.37+7^2\times0.3+9^2\times0.09=33.08$

$V(\overline{X})=E(\overline{X}^2)-E[(\overline{X})]^2=33.08-5.4^2=3.92$

(三) $P(\overline{X}>5)=0.3+0.09=0.39$

五、一位分析師為探討 A 地區勞工的所得（X）如何影響其休閒娛樂支出（Y）而建立如下之迴歸模型：

$Y=\alpha+\beta X+\varepsilon$，其中 ε 為隨機誤差項，且 $\varepsilon\sim N(0,\sigma^2)$。該分析師在 A 地區隨機抽樣了 101 位勞工，經蒐集此 101 位勞工之相關資料，並得統計資料如下：（單位：千元）

相關係數 $r_{XY}=0.5$，X 的樣本標準差 $s_X=10$，Y 的樣本標準差 $s_Y=2$

註：$s_X=\sqrt{\frac{1}{n-1}\sum(x_i-\overline{x})^2}$，$s_Y=\sqrt{\frac{1}{n-1}\sum(y_i-\overline{y})^2}$，$r_{XY}=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2}\sqrt{\sum(y_i-\overline{y})^2}}$

(一) 試根據上述資料，以最小平方法估計勞工的所得（X）增加 1 千元，其休閒娛樂支出（Y）會變動多大？即 β 之估計值為何？（10 分）

(二) 試求變異數 σ^2 之估計值為何？（10 分）

(三) 在顯著水準 $\alpha=0.05$ 之下，「勞工的所得越高，其休閒娛樂支出也越高」的說法是否能被接受？請列出虛無假設、對立假設、檢定統計量、拒絕域和結論。（10 分）

考點命中	《高點．高上統計學講義》第八回，蘇建郎編撰，9-2簡單線性迴歸分析，頁16。

答：

(一) $\hat{\beta}_1=r\frac{s_Y}{s_X}=0.5\times\frac{2}{10}=0.1$

估計勞工所得增加1千元，平均休閒娛樂支出多0.1千元。

(二) $SSTO=\sum_{i=1}^{101}(Y_i-\overline{Y})^2=(n-1)s_Y^2=100\times 4=400$

$SSR=r^2\times SSTO=0.25\times 400=100$

$SSE=SSTO-SSR=300$

$MSE=\dfrac{SSE}{n-2}=\dfrac{300}{99}=3.0303$ 為變異數 σ^2 估計值

(三)① $\begin{cases}H_0:\rho=0\\H_1:\rho\neq 0\end{cases}$ $\alpha=0.05$

② $t^*=\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\dfrac{0.5\sqrt{99}}{\sqrt{1-0.25}}=5.74$

③ $\alpha=0.05$，$C=\{t^*:|t^*|>t_{\alpha/2}(n-2)=t_{0.025}(99)=1.98\}$

$t^*=5.74\in C$，$reject\ H_0$

④故在顯著水準 $\alpha=0.05$ 的情況下，根據樣本資料顯示，我們有證據說 $\rho\neq 0$，即勞工所得越高，休閒娛樂支出也越高。