《統計學概要》

一、下列是關於條件機率、常態分配及隨機變數之期望值的問題：

(一)假定在一跟肺部有關疾病流行期間，共有3種可能型：A型、B型及C型的肺部疾病。其中染病病人得病機率分別是得A型肺部疾病機率為0.6、B型肺部疾病機率為0.3，而C型肺部疾病機率為0.1。這三型肺部相關疾病皆可能有發燒及咳嗽的症狀，其中A型病人中20%有發燒症狀，B型病人中60%有發燒症狀，而C型病人中80%有發燒症狀；而且50%的A型病人有咳嗽症狀，35%的B型病人有咳嗽症狀，25%的C型病人有咳嗽症狀。現有一染病病人有發燒症狀，請分別算出此病人得這三型肺部疾病的機率並決定這病人最可能得那一型的肺部疾病（5分）

(二)某一大型家電公司其某款售出產品之可用時間服從一平均值為μ（年）且變異數為2.25（年^2）的常態分配。已知此產品可用超過5年的機率為0.025，請算出μ的值及此產品可用時間不超過6個月的機率。（5分）

(三)一家口罩廠商得到一筆從某政局不穩定國家所下總值5千萬元的訂單，給定有0.7的機率此廠商可收到此5千萬元訂單付款，有0.15的機率僅可收到3千萬元付款，有0.1的機率僅可收到1千萬元付款，有0.05的機率收不到任何付款。為保險起見，此廠商投保某一意外險，並先支付1千萬元保費，承保之保險公司將支付此公司應收款項不足的任何差額。如果隨機變數X代表此口罩廠商最終在此投保所花費或賺到的金額。請算出X的期望值E(X)。（5分）

試題評析	(一)本題考貝氏定理，但題目最後的問題並未用到咳嗽的資訊，未知是否出題老師故意，還是題目不清。 (二)本題考常態分配，考生只要清楚瞭解題意，獲得滿分不難。 (三)本題考隨機變數之期望值，考生只要清楚瞭解題意，獲得滿分不難。
考點命中	1.《高點・高上統計學講義》第一回，趙治勳編撰，第四章、第五章。 2.《高點・高上統計學講義》第二回，趙治勳編撰，第七章。

答：

(一)令A,B,C分別表病人得到A型、B型與C型肺部疾病

D表有發燒症狀

$$P(D)=0.6\times0.2+0.3\times0.6+0.1\times0.8=0.38$$

$$P(A|D)=\frac{P(A\cap D)}{P(D)}=\frac{0.6\times0.2}{0.38}=0.3158$$

$$P(B|D)=\frac{P(B\cap D)}{P(D)}=\frac{0.3\times0.6}{0.38}=0.4737$$

$$P(C|D)=1-P(A|D)-P(B|D)=0.2105$$

故病人有發燒症狀下，最有可能得到B型肺部疾病。

(二)令X表可用時間(年)，$X\sim N(\mu,\sigma^2=2.25)$

由題意，$P(X>5)=P(Z>\frac{5-\mu}{\sqrt{2.25}})=0.025$

$$\Rightarrow\frac{5-\mu}{\sqrt{2.25}}=z_{0.025}=1.96\Rightarrow\mu=5-1.96\sqrt{2.25}=2.06\,(\text{年})$$

$$P(X\le0.5)=P(Z\le\frac{0.5-2.06}{\sqrt{2.25}})=P(Z\le-1.04)=0.1492$$

(三)令Y表廠商可收到款項(千萬)

Y=y	5	3	1	0
$f_Y(y)$	0.7	0.15	0.1	0.05

$X=5-Y-1$(千萬)

Y=y	5	3	1	0
X=x	-1	1	3	4
$f_X(x)$	0.7	0.15	0.1	0.05

$E(X)=(-1)0.7+(1)0.15+(3)0.1+(4)0.05=-0.05$(千萬)

該廠商在此投保之期望花費為50萬元

二、下列是關於資料集中趨勢、分散趨勢及相對位置之統計量的問題：

(一)某公司欲由其A分公司及B分公司所推薦之兩人擇一來晉升主管職位。這兩家分公司員工績效成績資訊如下：

	A分公司	B分公司
績效成績84分在該分公司之z分數	0.5	3
績效成績72分在該分公司之z分數	−2.5	−1
該推薦員工在其所在分公司之績效成績	90	82.5

人事主管決定用z分數（z score）來決定晉升人選，請問是那一家分公司推薦的員工得以晉升並說明原因。（5分）

(二)一組樣本數為10且以華氏溫標為單位之溫度資料 $x_1,\ldots,x_{10}$ ，其重要資訊如下：

第30百分位數（30th percentile）	平均值	四分位距（interquartile range）	變異數	變異係數（coefficient of Variation）
74.3	81.5	16.2	76.5	0.11

若以攝氏溫標為單位，即資料為 $y_i=\frac{5}{9}(x_i-32)$， $i=1,\ldots,10$ ，請算出關於攝氏溫標資料 $y_1,\ldots,y_{10}$ 的5個統計量：第30百分位數、平均值、四分位距、變異數及變異係數。（10分）

(三)一組資料其頻率分布圖呈現鐘形分布，利用經驗法則（empirical rule）可得知區間[12,22]包含大約此組資料的95%資料。請利用經驗法則得到包含大約此組資料的68%資料之區間。（5分）

試題評析	(一)本題考Z分數，考生只要清楚瞭解題意，獲得滿分不難。 (二)本題考資料轉換後對各種統計量之影響，獲得滿分不難。 (三)本題考經驗法則，獲得滿分不難。
考點命中	《高點‧高上統計學講義》第一回，趙治勳編撰，第三章。

答：

(一) A公司

已知 $\frac{84-\mu}{\sigma}=0.5\cdots\cdots\cdots(1)$

$\frac{72-\mu}{\sigma}=-2.5\cdots\cdots\cdots(2)$

(1)(2)聯立方程可得 $\mu=82,\sigma=4$

故A公司推薦員工之z分數為 $z_A=\frac{90-82}{4}=2$

B公司

已知 $\frac{84-\mu}{\sigma}=3\cdots\cdots\cdots(1)$

$\frac{72-\mu}{\sigma}=-1\cdots\cdots\cdots(2)$

(1)(2)聯立方程可得 $\mu=75,\sigma=3$

故B公司推薦員工之z分數為 $z_B=\frac{82.5-75}{3}=2.5$

由於 $z_B>z_A$，故B公司推薦員工得以晉升

(二) 第30百分位數：$\frac{5}{9}(74.3-32)=23.5$

平均數：$\bar{y}=\frac{\sum y_i}{n}=\frac{\sum\frac{5}{9}(x_i-32)}{n}=\frac{5}{9}(\bar{x}-32)=\frac{5}{9}(81.5-32)=27.5$

四分位距：$IQR_Y=\frac{5}{9}IQR_X=\frac{5}{9}(16.2)=9$

變異數：$S_Y^2=\frac{\sum(y_i-\bar{y})^2}{n-1}=\frac{\sum[\frac{5}{9}(x_i-32)-\frac{5}{9}(\bar{x}-32)]^2}{n-1}=(\frac{5}{9})^2\frac{\sum(x_i-\bar{x})^2}{n-1}$

$=(\frac{5}{9})^2S_X^2=(\frac{5}{9})^2 76.5=23.611$

變異係數：$C.V._Y=\frac{S_Y}{\bar{y}}=\frac{\sqrt{23.611}}{27.5}=0.1767$

(三) 由經驗法則得知

$\begin{cases}\mu-2\sigma=12\\ \mu+2\sigma=22\end{cases}\Rightarrow\mu=17,\sigma=2.5$

根據經驗法則，將有68%資料落在區間 $(\mu-1\sigma,\mu+1\sigma)=(17-1(2.5),17+1(2.5))=(14.5,19.5)$。

三、一公司利用簡單線性迴歸模式來建立其銷售額（Y）與所進行的廣告次數（X）之間的關係。給定下列10組資料 $(y_1,x_1),\ldots,(y_{10},x_{10})$，其中 y_i 為第i個月之銷售額，而 x_i 為第i個月所進行的廣告次數，根據所得之資訊：10筆銷售額資料 $y_1,\ldots,y_{10}$ 其平均值為130而變異數為1748，10筆廣告次數資料 $x_1,\ldots,x_{10}$ 其平均值為14，此外，銷售額資料 $y_1,\ldots,y_{10}$ 與廣告次數資料 $x_1,\ldots,x_{10}$ 其共變異數為316，利用最小平方法（least squares method）所得估計迴歸關係式為 $Y=60+5X$。

(一)若第3個月的銷售額為 $y_3=88$ 而所進行的廣告次數為 $x_3=8$，試算出此月銷售額的配適值（fitted value）及殘差值（residual）。（5分）

(二)算出判定係數（coefficient of determination）R^2 及銷售額資料 $y_1,\ldots,y_{10}$ 與廣告次數資料 $x_1,\ldots,x_{10}$ 的相關係數。（5分）

(三)在顯著水準 $\alpha=0.05$，利用t檢定法檢定是否廣告次數會影響銷售額，即檢定迴歸之線性關係式的斜率是否為0。（5分）

試題評析	本題考簡迴歸分析，只要考生對計算公式夠熟悉，獲得滿分不難。
考點命中	《高點・高上迴歸分析講義》第一回，趙治勳編撰，第二章。

答：

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

(一) $\hat{y}_3 = 60 + 5 \times x_3 = 60 + 5 \times 8 = 100$

$e_3 = y_3 - \hat{y}_3 = 88 - 100 = -12$

(二) $R^2 = \dfrac{SSR}{SST} = \dfrac{\hat{\beta}_1 SS_{XY}}{SS_Y} = \dfrac{\hat{\beta}_1 (n-1) S_{XY}}{(n-1) S_Y^2} = \dfrac{5(10\text{-}1)316}{(10-1)1748} = 0.9039$

$r_{XY} = (sign\hat{\beta})\sqrt{R^2} = +\sqrt{0.9039} = 0.9507$

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

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

R.R.: Reject H_0 at $\alpha = 0.05$ if $|T^*| > t_{(8)0.025} = 2.306$

$\because T^* = 8.674 \qquad \therefore reject H_0$

結論：我們有足夠證據去推論廣告次數會影響銷售額。

四、下列是關於母體比率其估計量之抽樣分配及母體平均之檢定的問題：

(一)欲判斷某生產線每日平均產量μ是否不少於一定數量，對此生產線隨機取得16筆日產量資料且此16筆日產量資料其標準差為3。利用此組樣本及t檢定法檢定此生產線每日平均產量μ是否不少於70，即檢定 $H_0: \mu \ge 70$ 對 $H_1: \mu < 70$，所得之p值（p value）為0.01。在顯著水準 $\alpha = 0.1$，請利用此組樣本及t檢定法檢定每日平均產量μ是否不少於69，即檢定 $H_0: \mu \ge 69$ 對 $H_1: \mu < 69$。（10分）

(二)針對母體比率p，p≤0.5，利用樣本數為1600之隨機樣本所得之樣本比率統計量 $\bar{P}$ 來估計p，可得抽樣誤差 $|\bar{P} - p|$ 在0.015內的機率約為0.8664。請算出p的可能值及當樣本數為2500時，$\bar{P}$ 大於0.22的機率。（10分）

(三)欲比較A廠牌儀器的平均壽命 μ_1 與B廠牌儀器的平均壽命 μ_2 之差異。隨機抽樣7台A廠牌儀器，此7台A廠牌儀器壽命平均值為10，而隨機抽樣9台B廠牌儀器，此9台B廠牌儀器壽命平均值亦為10。這16台儀器之壽命變異數為4。在顯著水準 $\alpha = 0.01$，請利用t檢定法檢定 $H_0: \mu_1 - \mu_2 \ge 2$ 對 $H_1: \mu_1 - \mu_2 < 2$。（10分）

試題評析	本題考抽樣分配與母體平均數之假設檢定，相關題型也有出現在考古題，只是題量較多，考驗考生之作答速度。
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表每日產量

母體：$X \sim N(\mu,\sigma^2)$ 假設 $X \sim N$

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

點估計：$\bar{X} \sim N(\mu, \dfrac{\sigma^2}{16})$

$H_0 : \mu \ge 70$ vs $H_1 : \mu < 70$

T.S.：$T = \dfrac{\bar{X} - 70}{S/\sqrt{16}} = \dfrac{\bar{X} - 70}{3/\sqrt{16}} \sim t_{(n-1=16)}$

$p-value=0.01=P(T_{(15)}\le T)\Rightarrow T=\dfrac{\bar{X}-70}{3/\sqrt{16}}=-t_{(16)0.01}=-2.602$

$\Rightarrow \bar{X}=70-2.602\dfrac{3}{\sqrt{16}}=68.0485$

$H_0:\mu\ge 69$ vs $H_1:\mu<69$

T.S.：$T=\dfrac{\bar{X}-69}{S/\sqrt{16}}\sim t_{(n-1=16)}$

R.R.：Reject H_0 at $\alpha=0.1$ if $T^*<-t_{(15)0.1}=-1.341$

$\because T^*=\dfrac{68.0485-69}{3/\sqrt{16}}=-1.2687$ $\quad\therefore don't\,reject H_0$

結論：我們沒有足夠證據去推論每日平均產量小於69。

(二) 母體：$X\sim Ber(p)$

樣本：$X_1,X_2,\cdots,X_{1600}\overset{iid}{\sim}Ber(p)$

點估計：$\bar{P}=\hat{p}=\dfrac{\sum X_i}{1600}\underset{by C.L.T.}{\sim}N(p,\dfrac{p(1-p)}{1600})$

$P(|\bar{P}-p|\le 0.015)=P(|Z|\le\dfrac{0.015}{\sqrt{\dfrac{p(1-p)}{1600}}})=0.8664$

$\Rightarrow\dfrac{0.015}{\sqrt{\dfrac{p(1-p)}{1600}}}=z_{0.0668}=1.5\Rightarrow p=0.2$

樣本數2500下，

點估計：$\bar{P}=\hat{p}=\dfrac{\sum X_i}{2500}\underset{by C.L.T.}{\sim}N(p=0.2,\dfrac{p(1-p)}{2500}=\dfrac{0.2(0.8)}{2500}=0.000064)$

$P(\bar{P}>0.22)=P(Z>\dfrac{0.22-0.2}{\sqrt{0.000064}})=P(Z>2.5)=0.0062$

(三) $\bar{x}_{混}=\dfrac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}=\dfrac{7\times10+9\times10}{7+9}=10$

$S^2{}_{混}=\dfrac{\sum(n_i-1)S_i^2+\sum n_i(\bar{x}_1-\bar{x}_{混})^2}{n_1+n_2-1}=6\Rightarrow\sum(n_i-1)S_i^2=60$

$S_p^2=\dfrac{\sum(n_i-1)S_i^2}{n_1+n_2-2}=\dfrac{60}{7+9-2}=4.2857$

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

假設(1) $X_1\sim N, X_2\sim N$
(2) $X_1\perp X_2$
(3) $\sigma_1^2=\sigma_2^2=\sigma^2$

樣本：$X_{11},X_{12},\cdots,X_{17}\overset{iid}{\sim}N(\mu_1,\sigma^2)\qquad X_{21},X_{22},\cdots,X_{29}\overset{iid}{\sim}N(\mu_2,\sigma^2)$

點估計：$\bar{X}_1-\bar{X}_2\sim N(\mu_1-\mu_2,\sigma^2(\dfrac{1}{7}+\dfrac{1}{9}))$

$H_0:\mu_1-\mu_2\ge 2$ vs $H_1:\mu_1-\mu_2<2$

T.S.：$T=\dfrac{(\overline{X}_1-\overline{X}_2)-(2)}{\sqrt{S_p^2(\frac{1}{7}+\frac{1}{9})}}\sim t_{(7+9-2=14)}$

R.R.：Reject H_0 at $\alpha=0.01$ if $T^*<-t_{(14)0.01}=-2.624$

$QT^*=\dfrac{(10-10)-(2)}{\sqrt{4.2857(\frac{1}{7}+\frac{1}{9})}}=-1.917\qquad \therefore don'trejectH_0$

結論：我們沒有足夠證據去推論 $\mu_1-\mu_2<2$ 。

五、欲比較三種不同品牌的電池其平均壽命是否一致，每種品牌電池各隨機抽測5顆電池並記錄其壽命。這三組樣本其壽命的平均值分別為3、4.5及3.6，而其壽命的標準差分別為1.5、1及1.5。在顯著水準 $\alpha=0.05$ ，請利用單因子變異數分析法（one-way ANOVA）來檢定這三種品牌電池平均壽命是否一致，即檢定：

$H_0:\mu_1=\mu_2=\mu_3$ 對 $H_1:\mu_1$ 、 μ_2 及 μ_3 並不完全相等。

其中 μ_1 、 μ_2 及 μ_3 分別是這三種品牌電池的平均壽命。（10分）

試題評析	本題考變異數分析，相關題型也有出現在考古題，獲得滿分不難。
考點命中	《高點・高上統計學講義》第三回，趙治勳編撰，第十二章。

答：

n_i	5	5	5
$\overline{y}_i$	3	4.5	3.6
s_i	1.5	1	1.5

$SSE=\sum(n_i-1)S_i^2=(5-1)(1.5^2+1^2+1.5^2)=22$

$SSR=\sum n_i(\overline{Y}_i-\overline{\overline{Y}})^2=5[(3-3.7)^2+(4.5-3.7)^2+(3.6-3.7)^2]=5.7$

ANOVA TABLE				
Source	SS	d.f.	MS	F
品牌	5.7	2	2.85	F^*=1.5548
Error	22	12	1.833	
Total	27.5	14		

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

T.S.：$F=\dfrac{MSR}{MSE}\sim F_{(2,12)}$

R.R.：Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(2,12)}=3.8853$

$QF^*=1.5548\qquad \therefore don'trejectH_0$

結論：我們沒有足夠證據去推論三個品牌之平均壽命是不盡相等的。

六、欲調查某種家電產品市場是否由四家廠商平均瓜分，即使用此四家廠商之人數比率 p_1 、 p_2 、 p_3 及 p_4 皆為0.25。一組隨機樣本其樣本數為n之問卷結果其資訊如下：

使用家電廠商A人數比率為27%，使用家電廠商B人數比率為25%，使用家電廠商C人數比率為26%，使用家電廠商D人數比率為22%。根據上述問卷結果，利用卡方檢定（chi-squared test）

來檢定市場是否由此四家廠商平均瓜分，即檢定：

$H_0: p_1 = p_2 = p_3 = p_4 = 0.25$ 對 H_1：並非 $p_1 = p_2 = p_3 = p_4 = 0.25$ 。

在顯著水準 $\alpha = 0.01$，結果為拒絕 H_0 。試求出n最小可能值。（已知n為100的倍數）（10分）

試題評析	本題考卡方適合度檢定，結合計算樣本數，出的不錯的考題，考生需要瞭解題意，由卡方適合度檢定之檢定統計量著手以計算樣本數。
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：

卡方適合度檢定之檢定統計量為

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(0.27n - 0.25n)^2}{0.25n} + \frac{(0.25n - 0.25n)^2}{0.25n} + \frac{(0.26n - 0.25n)^2}{0.25n} + \frac{(0.22n - 0.25n)^2}{0.25n}$$

$$= n[\frac{(0.27 - 0.25)^2}{0.25} + \frac{(0.25 - 0.25)^2}{0.25} + \frac{(0.26 - 0.25)^2}{0.25} + \frac{(0.22 - 0.25)^2}{0.25}]$$

$$= 0.0056n$$

已知拒絕 H_0，故 $\chi^2 = 0.0056n > \chi^2_{(3)0.01} = 11.3449 \Rightarrow n > 2025.875$

樣本數 n 之最小可能值為2100（注意：題目有提醒 n 為100的倍數）。