《統計學概要》

本試題可能使用之參考值如下：

$z_{0.05}=1.645$，$z_{0.1}=1.28$，$\chi^2_{1,0.05}=3.84$，$\chi^2_{2,0.05}=5.99$，$\chi^2_{3,0.05}=7.81$

一、統計學班上共100名學生，老師宣布期中考之平均分數75分，標準差10分。

(一)若成績不是常態分配，試問介於55分至95分的同學至少幾人？(5分)

(二)若成績不是常態分配，試問高於95分的同學至多幾人？(5分)

(三)若成績之分配為左偏，你的分數75分，試問多數同學的分數比你的高或低？為什麼？(5分)

(四)從全班學生中以抽後不放回的方式隨機抽取30名，令 $\overline{X}$ 代表其平均成績，試求隨機變數 $\overline{X}$ 的變異數。(5分)

試題評析	本題涉及柴夫比雪夫不等式，敘述統計學及抽樣分配之基礎運算。
考點命中	(一)(二)《高上統計學講義》第一回，趙治勳編撰，第四章，4.5節。 (三)《高上統計學講義》第一回，趙治勳編撰，P.16<重要補充>。 (四)《高上統計學講義》第三回，趙治勳編撰，頁第七章，7.2節。

答：

令 X 表期末考成績，$X\sim(\mu=75,\sigma^2=10^2)$

(一)由柴比雪夫不等式

$$P(55<X<95)=P(|X-75|<20)=P(|X-75|<2(10))\geq 1-\frac{1}{2^2}=0.75$$

故期末考成績介於55分至95分間之同學至少為$100(0.75)=75$位

(二)由單邊柴比雪夫不等式

$$P(X>95)=P(X>75+20)\leq\frac{10^2}{10^2+20^2}=0.2$$

故期末考成績高於95分之同學至多為$100(0.2)=20$位

(三)因為左偏分配下，平均數會小於中位數，我考的分數剛好為平均數75分，表示比我高分的人超過50%，因此多數同學的分數比我高。

(四) $V(\overline{X})=\dfrac{\sigma^2}{n}\dfrac{N-n}{N-1}=\dfrac{10^2}{30}\dfrac{100-30}{100-1}=2.3569$ 其中$\dfrac{N-n}{N-1}$為有限母體校正因子

二、隨機變數X之機率密度函數為 $f(x)=ce^{-(x+2)}$，$x\geq -2$。令 x_0 代表該機率分配的中位數。令Bernoulli(p)代表成功機率為p之伯努利分配。假設隨機變數Y在給定X之下的條件機率分配為：

Y服從 $\begin{cases}\text{Bernoulli}(0.8)，若X\geq x_0\\ \text{Bernoulli}(0.6)，若X<x_0\end{cases}$。

(一)試求c。(5分)

(二)試求X的期望值。(5分)

(三)試求X的中位數。(5分)

(四)試求Y的邊際機率分配。(5分)

(五)試求X與Y的共變異數(covariance)。(5分)

試題評析	本題涉及隨機變數與期望值。
考點命中	《高上統計學講義》第一回，趙治勳編撰，第四章。

答：

(一)根據機率公理假設

$$\int_{-2}^{\infty} ce^{-(x+2)}dx = 1$$

$$\therefore \int_{-2}^{\infty} ce^{-(x+2)}dx = c\int_{-2}^{\infty} e^{-(x+2)}dx = c = 1 \qquad \Rightarrow c = 1$$

(二) $E(X) = \int_{-2}^{\infty} xe^{-(x+2)}dx = 1\text{-}2 = \text{-}1$

(三) x_0 為隨機變數 X 之中位數

$$F_X(x_0) = \int_{-2}^{x_0} e^{-(x+2)}dx = 1 - e^{-(x_0+2)} = 0.5$$

$$x_0 = -1.3069$$

(四) $f_Y(y|x) = \begin{cases} (0.8)^y(0.2)^{1-y}, y = 0,1, x \ge x_0 \\ (0.6)^y(0.4)^{1-y}, y = 0,1, x < x_0 \end{cases}$

$$f_Y(y) = P(Y = y) = P(X \ge x_0)P(Y = y | X \ge x_0) + P(X < x_0)P(Y = y | X < x_0)$$

$$= \frac{1}{2}(0.8)^y(0.2)^{1-y} + \frac{1}{2}(0.6)^y(0.4)^{1-y},\ \ y = 0,1$$

$Y = y$	0	1
$f_Y(y)$	0.3	0.7

$Y \sim Ber(0.7)$

(五) $E(Y) = 0.7(1-0.7) = 0.21$，$E(X) = \text{-}1$

$$f_{XY}(x,y) = f_X(x)f_Y(y|x) = \begin{cases} e^{-(x+2)}(0.8)^y(0.2)^{1-y}, y = 0,1, x \ge x_0 \\ e^{-(x+2)}(0.6)^y(0.4)^{1-y}, y = 0,1, x < x_0 \end{cases}$$

$$E(XY) = \int_{-2}^{x_0} x(1)e^{-(x+2)}(0.6)dx + \int_{x_0}^{\infty} x(1)e^{-(x+2)}(0.8)dx = -0.6307 \quad \text{其中 } x_0 = -1.3069$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \text{-}0.4207$$

三、櫻花村於過去40個月內發生交通事故的資料如下：

交通事故次數	0	1	2	3	4	4
月份個數	10	15	11	3	1	0

(一)試估計平均每個月發生的交通事故次數。（4分）

(二)在顯著水準 $\alpha = 5\%$ 之下，檢定每個月發生的交通事故次數是否服從卜瓦松（Poisson）分配。（16分）

試題評析	本題涉及卡方適合度檢定。
考點命中	《高上統計學講義》第四回，趙治勳編撰，頁78。

答：

令 X 表發生交通事故之次數

(一) $\bar{x}=\dfrac{(0)(10)+(1)(15)+\cdots+(4)(1)}{40}=1.25$ 次

(二)

x	0	1	2	3	>4
月份個數	10	15	11	3	1
期望次數	11.460	14.325	8.953	3.731	1.531

H_0 為真下, $X\sim Poisson(\hat{\lambda}=1.25)$

因此 $x=0$ 時之期望次數為 $40P(X=0)=40(\dfrac{e^{-1.25}1.25^{o}}{0!})=11.46$ 以此類推。

發現x=3,4以上之期望次數不大於5

因此合併組後如下:

x	0	1	2	3以上
月份個數	10	15	11	4
期望次數	11.460	14.325	8.953	5.262

H_0 :資料來自波瓦松分配 vs H_1 : not H_0

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

R.R.: Reject H_0 at $\alpha=0.05$ if $\chi^{2*}>\chi^2_{0.05(2)}=5.99$

$\because \chi^{2*}=0.9885$ $\therefore$ Don't reject H_0

我們沒有足夠證據去推論資料不是來自於波瓦松分配。

四、一盒中有四顆彈珠，其中θ顆為紅色，4−θ顆為白色。王小明想檢定 H_0：θ=2 vs. H_1：θ≠2。王小明的檢定方法是先以抽後放回的方式抽出兩顆彈珠；其次，若抽出的兩顆同色，就拒絕 H_0，而若抽出的兩顆為不同色，就不拒絕 H_0。

(一)試求該檢定方法的型一誤差（Type I error probability）。（5分）

(二)試求該檢定方法在θ=1時的型二誤差（Type II error probability）。（5分）

(三)若王小明以抽後放回的方式抽100次，結果60次為紅色，40次為白色。在顯著水準α=10%之下，檢定 H_0：θ=2 vs. H_1：θ≠2。（10分）

試題評析	本題涉及母體成功比例之假設檢定。
考點命中	《高上統計學講義第四回》，趙治勳編撰，頁30。

答：

令 Y 表抽出的兩顆彈珠中紅色之個數, $Y\sim Bin(2,\dfrac{\theta}{4})$

$H_0:\theta=2 vs H_1:\theta\neq 2$

(一) $\alpha=P(Y=0 or Y=2\mid\theta=2)=1-P(Y=1\mid\theta=2)=1-\binom{2}{1}(\frac{1}{2})^1(\frac{1}{2})^{2-1}=0.5$

(二) $\beta=P(Y=1\mid\theta=1)=\binom{2}{1}(\frac{1}{4})^1(\frac{3}{4})^{2-1}=\frac{3}{8}$

(三)

令 X 表抽出紅色彈珠

母體: $X \sim Ber(p=\frac{\theta}{4})$

樣本: $X_1, X_2, \cdots, X_{100} \overset{iid}{\sim} Ber(p=\frac{\theta}{4})$

點估計: $\hat{p} \underset{byC.L.T.}{\sim} N(p, \frac{p(1-p)}{100})$

$H_0: \theta = 2 vs H_1: \theta \neq 2$

等價於 $H_0: p = 0.5 vs H_1: p \neq 0.5$

T.S.: $Z = \dfrac{\hat{p}-0.5}{\sqrt{\dfrac{(0.5)(0.5)}{100}}} \underset{byC.L.T.}{\sim} N(0,1)$

R.R.: Reject H_0 at $\alpha = 0.1$ if $|Z^*| > z_{0.05} = 1.645$

$\because Z^* = \dfrac{\frac{60}{100}-0.5}{\sqrt{\dfrac{(0.5)(0.5)}{100}}} = 2$ $\quad\quad \therefore reject H_0$

我們有足夠證據去推論盒中之紅色彈珠數不為兩個

五、某餐廳過去8年之年營業額（萬元）如下：

年	2006	2007	2008	2009	2010	2011	2012	2013
營業額	36	40	42	45	46	49	54	56

(一)試以迴歸模型估算此時間數列的線性趨勢。（12分）

(二)試以(一)估算所得之線性趨勢預測2014年的營業額。（3分）

試題評析	本題涉及簡迴歸估計。
考點命中	《迴歸分析熱門題庫》書，高點出版，趙治勳編撰，頁2-19。

答：

令 X 表年度, Y 表營業額(萬元)

(一)

$\sum X_i = 16076, \sum X_i^2 = 32304764, \sum Y_i = 368, \sum Y_i^2 = 17254, \sum X_i Y_i = 739612$

$SS_{XY} = \sum X_i Y_i - \dfrac{\sum X_i \sum Y_i}{n} = 116$, $SS_X = \sum X_i^2 - \dfrac{(\sum X_i)^2}{n} = 42$

假設迴歸模型: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

(1) $\hat{\beta}_1 = \dfrac{SS_{XY}}{SS_X} = 2.7619$, $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = \text{-}5504.038$

$\therefore \hat{y} = \text{-}5504.038 + 2.7619x$

(2) $\hat{y} = \text{-}5504.038 + 2.7619(2014) = 58.429$