《統計學(B)-經建》

一、令$X_1,\ldots,X_6$表示服從指數分配（Exponential distribution）Exp(θ)的隨機樣本，其機率密度函數為$f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}},\ x>0$每小題8分，共24分）

（一）計算$E(X_1)$。（E：Expectation）（須列出計算過程）

（二）令$Y=\sum_{i=1}^{6}X_i$，求隨機變數Y的動差母函數（moment generating function），並回答Y的機率分配名稱。（須列出計算過程）

（三）令$\bar{Y}=\frac{\sum_{i=1}^{6}X_i}{6}$，求隨機變數$\bar{Y}$的動差母函數（moment generating function），並回答$\bar{Y}$的機率分配名稱。（須列出計算過程）

試題評析	本題考的是 Gamma 分配及其分配，以及動差母函數的唯一性，考生必須能看到動差母函數就知道隨機變數是什麼分配。

答：

(一)$E(X_1)=\int_0^\infty x\cdot\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\frac{1}{\theta}\cdot\int_0^\infty x^{2-1}\cdot e^{-\frac{x}{\theta}}dx=\frac{1}{\theta}\cdot\theta^2\Gamma(2)=\theta$

(二)由於 $\{X_i\}_{i=1}^6\overset{i.i.d.}{\sim}exp(\theta)$

$$M_X(t)=E(e^{tX})=(1-\theta t)^{-1},t<\frac{1}{\theta}$$

$$M_Y(t)=E(e^{tY})=E\left(e^{t\sum_{i=1}^6X_i}\right)=\prod_{i=1}^{6}E(e^{tX_i})=(1-\theta t)^{-6},t<\frac{1}{\theta}$$

由動差母函數知 $Y\sim Gamma(\alpha=6,\beta=\theta)$

(三)$M_{\bar{Y}}(t)=E\left(e^{t\bar{Y}}\right)=E\left(e^{t\frac{1}{6}Y}\right)=M_Y\left(\frac{1}{6}t\right)=\left[1-\theta\left(\frac{1}{6}t\right)\right]^{-6}=\left(1-\frac{\theta}{6}t\right)^{-6},t<\frac{6}{\theta}$

由動差母函數知 $\bar{Y}\sim Gamma\left(\alpha=6,\beta=\frac{\theta}{6}\right)$

二、考慮簡單線性迴歸模型，$y_i=\beta_0+\beta_1x_i+\varepsilon_i$，$\varepsilon_i\overset{iid}{\sim}N(0,\sigma^2)$，$\hat{\beta}_0$和$\hat{\beta}_1$分別為$\beta_0$與$\beta_1$之最小平方估計式，計算下列各子題：（每小題10分，共20分）

（一）計算$Cov(\bar{y},\hat{\beta}_1)$，其中$\bar{y}$為反應變數Y之平均數，Cov是指共變異數（Covariance）。

（二）計算$Var(\hat{\beta}_0+0.8\hat{\beta}_1)$，Var是指變異數（Variance）。

試題評析	本題為簡單迴歸係數性質的推導，過程繁雜需特別注意。

答：

(一)

$$\begin{aligned}Cov(\bar{Y},\hat{\beta}_1)&=Cov\left(\beta_0+\beta_1\bar{X}+\sum_{i=1}^{n}\frac{1}{n}\epsilon_i,\sum_{i=1}^{n}\frac{(X_i-\bar{X})}{SS_X}Y_i\right)\\&=Cov\left(\beta_0+\beta_1\bar{X}+\sum_{i=1}^{n}\frac{1}{n}\epsilon_i,\sum_{i=1}^{n}\frac{(X_i-\bar{X})}{SS_X}(\beta_0+\beta_1X_i+\epsilon_i)\right)\end{aligned}$$

$$= Cov\left(\sum_{i=1}^{n}\frac{1}{n}\epsilon_i, \beta_1 + \sum_{i=1}^{n}\frac{(X_i-\bar{X})}{SS_X}\epsilon_i\right)$$
$$= \sum_{i=1}^{n}\frac{1}{n}\frac{(X_i-\bar{X})}{SS_X}Var(\epsilon_i) + \sum_{i=1}^{n}\sum_{i\neq j}\frac{1}{n}\frac{(X_i-\bar{X})}{SS_X}Cov(\epsilon_i,\epsilon_j)$$
$$= \sum_{i=1}^{n}\frac{(X_i-\bar{X})}{SS_X}\frac{\sigma^2}{n} = \frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{(X_i-\bar{X})}{SS_X} = 0$$

(二)

$$Var(\hat\beta_0 + 0.8\hat\beta_1) = Var(\hat\beta_0) + 0.64Var(\hat\beta_1) + 2\cdot 0.8\cdot Cov(\hat\beta_0,\hat\beta_1)$$
$$= \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{SS_X}\right) + 0.64\cdot\frac{\sigma^2}{SS_X} + 1.6Cov(\bar{Y}-\hat\beta_1\bar{X},\hat\beta_1)$$
$$= \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{SS_X}\right) + 0.64\cdot\frac{\sigma^2}{SS_X} + 1.6\left[Cov(\bar{Y},\hat\beta_1) - \bar{X}Var(\hat\beta_1)\right]$$
$$= \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{SS_X}\right) + 0.64\cdot\frac{\sigma^2}{SS_X} + 1.6\left[-\bar{X}\frac{\sigma^2}{SS_X}\right]$$
$$= \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{SS_X}+\frac{0.64}{SS_X}-1.6\frac{\bar{X}}{SS_X}\right)$$
$$= \sigma^2\left[\frac{1}{n}+\frac{(\bar{X}-0.8)^2}{SS_X}\right]$$

三、有關於汽車碳氫化合物排放量（克/英里）的研究，記錄碳氫化合物排放量Y（100克/公里），和相對應的累積里程數X（以1000公里為單位）。初步整理樣本資料如下所示：

$n=11$，$\sum_{i=1}^{n}x_i = 190.2356$，$\sum_{i=1}^{n}y_i = 212.9375$，$\sum_{i=1}^{n}x_i^2 = 4086.6461$，$\sum_{i=1}^{n}y_i^2 = 4152.344$，$\sum_{i=1}^{n}x_iy_i = 3808.8281$。

使用以上資料回答下列問題，請詳細將所使用之公式及計算過程列出。（每小題9分，共36分）

(一)計算最小平方迴歸線。（計算至小數點後4位數）

(二)迴歸判定係數（R^2）為何？

(三)顯著水準為0.05，檢定迴歸斜率是否顯著異於0.16。

(四)在x＝25時，求對應之反應變數Y預測值的95%預測區間。

（$t_{10,0.025}=2.228$，$t_{9,0.025}=2.262$，$t_{10,0.05}=1.812$，$t_{9,0.05}=1.833$）

試題評析	迴歸直線計算與後續係數檢定及預測區間計算，過去也考過類似的題目，計算量大需特別注意。

答：

(一)

$$\hat\beta_1 = \frac{n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2} = \frac{11\times 3808.8281 - 190.2356\times 212.9375}{11\times 4086.6461 - 190.2356^2} = 0.1585$$

$$\hat\beta_0 = \bar{Y} - \hat\beta_1\bar{X} = \frac{212.9375}{11} - 0.1585\times\frac{190.2356}{11} = 16.6168$$

因此迴歸直線為 $\hat{Y} = 16.6168 + 0.1585X$

(二)

$$SST = \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}y_i^2 - \frac{(\sum_{i=1}^{n}y_i)^2}{n} = 4152.344 - \frac{212.9375^2}{11} = 30.3096$$

$$SSR=\hat{\beta}_1{}^2\sum_{i=1}^{n}(x_i-\bar{x})^2=\hat{\beta}_1{}^2\left[\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}\right]=0.1585^2\left(4086.6461-\frac{190.2356^2}{11}\right)=20.0145$$

$$SSE=SST-SSR=30.3096-20.0145=10.2951$$

$$R^2=\frac{SSR}{SST}=\frac{20.0145}{30.3096}=0.6603$$

(三)

$$\hat{\sigma}^2=\frac{SSE}{n-2}=\frac{10.2951}{11-2}=1.1439$$

$$SE(\hat{\beta}_1)=\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}=\sqrt{\frac{1.1439}{796.684}}=0.0379$$

β_1 之 95% 信賴區間為

$[\hat{\beta}_1\mp t_{0.025}(9)\times SE(\hat{\beta}_1)]=[0.1585\mp 2.262\times 0.0379]=[0.0728,0.2442]$

由於 $0.16\in[0.0728,0.2442]$

∴ 迴歸斜率沒有顯著異於0.16

(四)95% 預測區間為

$$\left[\hat{\beta}_0+\hat{\beta}_1\cdot 25\mp t_{0.025}(9)\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(25-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)}\right]$$

$$=\left[16.6168+0.1585\cdot 25\mp 2.262\sqrt{1.1439\left(1+\frac{1}{11}+\frac{\left(25-\frac{190.2356}{11}\right)^2}{796.684}\right)}\right]=[17.9675,23.1911]$$

四、觀察記錄某一城市在最近三個月內(90天)每天汽機車意外事故的次數,其次數分配如下所示:

意外事故次數	0	1	2	3	4
觀察天數	32	34	17	6	1

檢定每天汽機車意外事故次數是否服從波松(Poisson)分配。

(一)寫出虛無假設與對立假設。(5分)

(二)在顯著水準為$\alpha=0.05$時,寫出檢定統計量、棄卻域和結論。(須列出計算過程)(15分)

($\chi^2_{2,0.025}=7.38$,$\chi^2_{3,0.025}=9.35$,$\chi^2_{4,0.025}=11.14$,$\chi^2_{2,0.05}=5.99$,$\chi^2_{3,0.05}=7.81$,$\chi^2_{4,0.05}=9.49$)

波松分配累積機率表

x	λ = E(X)		
	0.5	1.0	2.0
0	0.607	0.368	0.135
1	0.910	0.736	0.406
2	0.986	0.920	0.677
3	0.998	0.981	0.857
4	1.000	0.996	0.947
5	1.000	0.999	0.983
6	1.000	1.000	0.995
7	1.000	1.000	0.999

試題評析	本題為卡方檢定中的適合度檢定,題目不難重點在於合併組別。

答：

（一）

H_0: 事故次數服從波松分配 against H_1: 事故次數不服從波松分配

（二）

$$\hat{\lambda}_{MLE} = \bar{x} = \frac{0 \times 32 + 1 \times 34 + 2 \times 17 + 3 \times 6 + 4 \times 1}{90} = 1$$

將資料整理使得合併後 $e_i \geq 5$

事故次數	天數	p_i	e_i
0	32	0.368	33.12
1	34	$0.736 - 0.368 = 0.368$	33.12
2	17	$0.920 - 0.736 = 0.184$	16.56
≥ 3	7	$1 - 0.920 = 0.080$	7.20

$$\varphi = \sum_{i=1}^{4} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(2)$$

$\alpha = 0.05, RR = \{\varphi \big| \varphi > \chi^2_{0.05}(2) = 5.99\}$

$$\varphi^* = \frac{(32 - 33.12)^2}{33.12} + \frac{(34 - 33.12)^2}{33.12} + \frac{(17 - 16.56)^2}{16.56} + \frac{(7 - 7.20)^2}{7.20} = 0.079 \notin RR$$

∴ 無法拒絕虛無假設

我們沒有足夠證據推論事故次數不服從波松分配