《統計學》

一、行政院主計總處對家戶所得與消費支出之關係，作了一些研究。將所得與消費支出分別以X與Y表示。已知(X,Y)之聯合動差母函數（joint moment generating function）為 $M_{(X,Y)}(t_1,t_2)=\exp\left\{2t_1+t_2+0.5t_1t_2+2.5t_1^2+2t_2^2\right\}$。

請回答下列問題：（每小題10分，共40分）

(一)求出給定消費支出Y＝1之下，所得X之條件機率密度函數。

(二)求出條件期望值E(X|Y)之變異數Var[E(X|Y)]。

(三)求出消費支出為1之下，所得大於2之條件機率P(X＞2|Y＝1)。

(四)令隨機變數Z＝$(Y-1)^2$，求出Z之機率密度函數。

試題評析	本題有關二元常態分配之動差母函數、邊際分配及條件分配，雖然考古題中甚少出題，但老師在課堂上都已經強調該分配之重要性。
考點命中	《高點．高上統計學講義》第二回，趙治勳編撰，第八章。

答：

$$M_{XY}(t_1,t_2)=e^{2t_1+\frac{5}{2}t_1^2+1t_2+\frac{4}{2}t_2^2+(\frac{1}{4\sqrt{5}})(\sqrt{5})(2)t_1t_2},t_1\in R,t_2\in R$$

由 m.g.f.之唯一性

$$(X,Y)\sim N_2(\mu_1=2,\mu_2=1,\sigma_1^2=5,\sigma_2^2=4,\rho=\frac{1}{4\sqrt{5}})$$

可知 $X\sim N(\mu_1=2,\sigma_1^2=5)$ 與 $Y\sim N(\mu_2=1,\sigma_2^2=4)$

(一) $X|Y=1\sim N(2,\frac{79}{16})$

其中 $E(X|Y=1)=2+\frac{1}{4\sqrt{5}}\frac{\sqrt{5}}{4}(1-1)=2$

$$V(X|Y=1)=5[1-(\frac{1}{4\sqrt{5}})^2]=\frac{79}{16}$$

$$f_{X|Y}(x|y=1)=\frac{1}{\sqrt{2\pi}\sqrt{\frac{79}{16}}}e^{-\frac{(x-2)^2}{2\times\frac{79}{16}}},-\infty<x<\infty$$

(二) $E(X|Y)=2+\frac{1}{4\sqrt{5}}\frac{\sqrt{5}}{4}(Y-1)=2+\frac{Y-1}{16}$

$$V[E(X|Y)]=V[2+\frac{Y-1}{16}]=\frac{1}{16^2}V(Y)=\frac{1}{16^2}(4)=\frac{1}{64}$$

(三) 由(一)得知 $X|Y=1\sim N(2,\frac{79}{16})$

$$P(X>2|Y=1)=P(Z>\frac{2-2}{\sqrt{\frac{79}{16}}})=P(Z>0)=0.5$$

(四) $Y\sim N(1,4)\Rightarrow\frac{Y-1}{2}\sim N(0,1)\Rightarrow\frac{(Y-1)^2}{4}\sim\chi^2_{(1)}$

$$Z=(Y-1)^2=4\times\frac{(Y-1)^2}{4}\sim Gamma(\alpha=\frac{1}{2},\lambda=\frac{1}{2}\frac{1}{4}=\frac{1}{8})$$

$$f_Z(z)=\frac{(\frac{1}{8})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})}z^{\frac{1}{2}-1}e^{-\frac{1}{8}z}=\frac{1}{2\sqrt{2\pi}}z^{-\frac{1}{2}}e^{-\frac{z}{8}},0<z<\infty$$

二、針對家戶之所得（以X表示）與消費支出（以Y表示）之關係，考慮建立下列迴歸模型：$Y_i=\beta_1\sqrt{X_i}+\varepsilon_i,\quad i=1,2,...,n$；其中$\varepsilon_1,\varepsilon_2,\varepsilon_3,...,\varepsilon_n$為相互獨立且具常態分配$N(0,\sigma^2)$之隨機變數。

請回答下列問題：

(一)求出β_1之最小平方估計量$\hat{\beta}_1$。

(二)求出題(一)中，$\hat{\beta}_1$之變異數$Var(\hat{\beta}_1)$。

(三)欲檢定$H_0:\beta_1=0$ vs. $H_1:\beta_1\neq 0$，已收集下列資料

X_i	4	9	4	16	1
Y_i	2	5	2	10	1

請利用收集之資料，以0.05之顯著水準，檢定上述之假設。

（註：若令T(d)為具有自由度為d之t分配的隨機變數，則已知P(T(3)＜2.353)＝0.95，P(T(3)＜3.182)＝0.975，P(T(4)＜2.132)＝0.95，P(T(4)＜2.776)＝0.975

(四)若真實之迴歸模型為$Y_i=\beta_0+\beta_1\sqrt{X_i}+\varepsilon_i$，請用題(一)之$\hat{\beta}_1$去估計真實模型中之$\beta_1$，求出其偏誤（bias）。

試題評析	本題有關迴歸分析，雖然題目較難，但所有題型皆在統計人員之迴歸分析考卷上出現過，只要有在準備迴歸分析之考生，要拿高分並不困難。
考點命中	《迴歸分析》，高點文化出版，趙治勳編著，第二篇。

答：

令$V_i=\sqrt{X_i}$

模型：$Y_i=\beta_1\sqrt{X_i}+\varepsilon_i=\beta_1V_i+\varepsilon_i,\varepsilon_i\overset{iid}{\sim}N(0,\sigma^2)$

(一) $\hat{\beta}_1=\frac{\sum_{i=1}^{n}V_iY_i}{\sum_{i=1}^{n}V_i^2}$

(二) $V(\hat{\beta}_1)=V(\frac{\sum_{i=1}^{n}V_iY_i}{\sum_{i=1}^{n}V_i^2})=\frac{V(\sum_{i=1}^{n}V_iY_i)}{[\sum_{i=1}^{n}V_i^2]^2}=\frac{\sum_{i=1}^{n}V_i^2V(Y_i)}{[\sum_{i=1}^{n}V_i^2]^2}=\frac{\sum_{i=1}^{n}V_i^2\sigma^2}{[\sum_{i=1}^{n}V_i^2]^2}=\frac{\sigma^2}{\sum_{i=1}^{n}V_i^2}$

(三) $\sum V_i=12$， $\sum V_i^{\ 2}=34$， $\sum Y_i=20$， $\sum Y_i^{\ 2}=134$， $\sum V_iY_i=64$

$\hat{\beta}_1=\frac{\sum_{i=1}^{n}V_iY_i}{\sum_{i=1}^{n}V_i^2}=\frac{64}{34}=1.8824$ $\therefore \hat{y}=1.8824v$

$V_i=\sqrt{X_i}$	2	3	2	4	1
Y_i	2	5	2	10	1
$\hat{Y}_i$	3.7648	5.6472	3.7648	7.5296	1.8824
e_i	-1.7648	-0.6472	-1.7648	2.4704	-0.8824

$\hat{\sigma}^2=MSE=\frac{SSE}{n-1}=\frac{\sum e_i^{\ 2}}{n-1}=\frac{13.5294}{4}=3.38235$

(注意：模型沒有截距項，自由度為 n-1)

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

T.S.: $T=\frac{\hat{\beta}_1-0}{\sqrt{\frac{MSE}{\sum V_i^2}}}\sim t_{(4)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $|T^*|>t_{(4)0.025}=2.776$

$\because T^*=\frac{1.8824-0}{\sqrt{\frac{3.38235}{34}}}=5.9682$ $\therefore$ reject H_0

我們有足夠證據去推論 $\beta_1\neq 0$

(四) 真實模型：$Y_i=\beta_0+\beta_1\sqrt{X_i}+\varepsilon_i=\beta_0+\beta_1V_i+\varepsilon_i$

偏誤 $=E(\hat{\beta}_1)-\beta_1=\beta_0\frac{\sum_{i=1}^{n}V_i}{\sum_{i=1}^{n}V_i^2}+\beta_1-\beta_1=\beta_0\frac{\sum_{i=1}^{n}V_i}{\sum_{i=1}^{n}V_i^2}=\frac{12}{34}\beta_0=\frac{6}{17}\beta_0$

其中 $E(\hat{\beta}_1)=E(\frac{\sum_{i=1}^{n}V_iY_i}{\sum_{i=1}^{n}V_i^2})=\frac{\sum_{i=1}^{n}V_iE(Y_i)}{\sum_{i=1}^{n}V_i^2}=\frac{\sum_{i=1}^{n}V_i(\beta_0+\beta_1V_i)}{\sum_{i=1}^{n}V_i^2}=\beta_0\frac{\sum_{i=1}^{n}V_i}{\sum_{i=1}^{n}V_i^2}+\beta_1$

三、為了降低流行性感冒對民眾造成的傷害和損失，政府從民國106年10月1日起開始針對特定對象，提供免費流感疫苗接種服務。假設已知臺北市某區之衛生所平均要等10分鐘，才會有一位民眾上門接受免費疫苗接種服務。

請回答下列問題：（每小題10分，共20分）

（一）求出某日該衛生所從早上8點整上班到下午5點整下班都沒有民眾上門接種免費疫苗之機率。

（二）令變數S為該衛生所從早上8點上班後，直到等到第100位民眾上門接種免費疫苗所需等候時間（單位：小時），求出變數S之變異數Var(S)。

試題評析	本題有關 Poisson 分配與 Gamma 分配之計算題，考生們在作答相關題型時，單位時間一定要判斷清楚。
考點命中	《高點・高上統計學講義》第二回，趙治勳編撰，第六章及第七章。

答：

假設滿足 Poisson 過程

(一)

令 $N(t)$ 表在 t 倍 10 分鐘內民眾上門接受免費疫苗接種服務之人數， $N(t) \sim Poisson(\lambda = 1)$

9 小時=> t =54 倍 10 分鐘

$$P(N(t=54)=0) = \frac{e^{-1\times 54}(1\times 54)^0}{0!} = e^{-54}$$

(二)

依題意， $S \sim Gamma(\alpha = 100, \lambda = 6)$ (單位時間：1 小時)

$$V(S) = \frac{\alpha}{\lambda^2} = \frac{100}{6^2} = \frac{25}{9} \sim Gamma(\alpha = 100, \lambda = 6)$$