《統計學》

一、兩隨機變數X, Y的聯合機率密度函數為 $f(x,y)=ax^2y, 0<x<y<1$。

(一)請算出a使得 $f(x,y)$ 符合聯合機率密度函數的要求。(9分)

(二)請分別算出X, Y的邊際機率密度函數(marginal probability density function) $f_X(x), f_Y(y)$，及條件機率密度函數(conditional probability density function) $f_{X|Y}(x|y), f_{Y|X}(y|x)$。(16分)

試題評析	本題為二元機率函數之性質，邊際機率函數與條件機率函數之計算題。
考點命中	《高點・高上統計學講義》第一回，第五章，趙治勳編撰。

答：

(一)根據機率公理假設

$$\int_0^1\int_0^y ax^2y\,dx\,dy=1$$

$$\int_0^1\int_0^y ax^2y\,dx\,dy=a\int_0^1\frac{1}{3}y^4dy=\frac{a}{15}=1 \quad \Rightarrow a=15$$

(二) $$f_X(x)=\int_x^1 15x^2y\,dy=\frac{15}{2}x^2(1-x^2), 0<x<1$$

$$f_Y(y)=\int_0^y 15x^2y\,dx=5y^4, 0<y<1$$

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{15x^2y}{\frac{15}{2}x^2(1-x^2)}=\frac{2y}{1-x^2}, 0<x<y<1$$

$$f_{Y|X}(y|x)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{15x^2y}{5y^4}=\frac{3x^2}{y^3}, 0<x<y<1$$

二、假設 $X_1, X_2, \ldots, X_n$ 為幾何分配的Geo(p)隨機樣本。

(一)求p的最大概似估計(Maximum likelihood estimator)，$\hat{p}$。(10分)

(二)令 $\theta=\frac{1}{p}$，統計量 $\frac{1}{\hat{p}}$ 是否為 θ 的充分統計量？是否為 θ 的最小變異不偏估計量(minimum variance unbiased estimator)？(15分)

試題評析	本題為MLE與UMVUE之相關考題。
考點命中	《高點・高上統計學講義》第二回，第十章，趙治勳編撰。

答：

(一) $L(p)=f(x_1,x_2,\ldots,x_n;p)=p^n(1-p)^{\sum x_i-n}$

$\ln L(p)=n\ln p+(\sum x_i-n)\ln(1-p)$

$$\frac{\partial \ln L(p)}{\partial p}=\frac{n}{p}-\frac{\sum x_i-n}{1-p}=0 \qquad 且 \frac{\partial^2 \ln L(p)}{\partial p^2}<0$$

$$\hat{p}=\frac{n}{\sum x_i}=\frac{1}{\bar{x}}$$

$\therefore \hat{p}=\frac{1}{\bar{X}}$ 為 p 之MLE

(二) $f_X(x;p)=p(1-p)^{x-1}=1\times\frac{p}{1-p}\times e^{\ln(1-p)\times x}$ 為指數族

$w(p)=\ln(1-p)$ 為線性獨立且至少包含一個一維的矩形

故 $\sum_{i=1}^{n} X_i$ 為 p 之完備充分統計量

由c.s.s之不變性，$\therefore \frac{1}{\hat{p}}=\bar{X}$ 為 $\theta=\frac{1}{p}$ 之完備充分統計量

$$E(\frac{1}{\hat{p}})=E(\bar{X})=\frac{1}{p}$$

由Lehmann-Scheffe可得，$\frac{1}{\hat{p}}$ 為 $\theta=\frac{1}{p}$ 之UMVUE

三、某校今年派出9名代表參加科學知識競賽，得分及敘述性統計如下：

73	81	74	87	90	89	65	79	88

樣本平均數＝80.67，樣本標準差＝8.6747，去年的平均成績為75.5。

欲比較今年的代表隊成績是否比去年好，在5%顯著水準下，考慮兩種方法檢定今年成績是否有顯著進步。檢定結果應包括檢定假說、檢定統計量、臨界值以及檢定結果的詮釋。

(一)假設成績服從常態分配，請完成今年平均成績是否有顯著性進步的檢定。（10分）

(二)假設成績不服從常態分配但大致上對稱，請以符號檢定（sign test）及符號等級檢定（signed rank test），分別完成今年的成績中位數是否有顯著性進步的檢定。（20分）

試題評析	本題為無母數統計學中的符號檢定與符號等級檢定，跟講義上的例題幾乎一模一樣。
考點命中	《高點・高上統計學講義》第三回，第十三章，例1、2，趙治勳編撰。

答：

(一)母體：$X\sim N(\mu,\sigma^2)$

樣本：$X_1,X_2,\ldots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$

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

$H_0:\mu\le 75.5$ vs $H_1:\mu>75.5$

T.S.: $T=\frac{\bar{X}-75.5}{S/\sqrt{9}}\sim t_{(8)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $T^*>t_{(8)0.05}=1.8595$

$\because T^{*}=\dfrac{80.67-75.5}{8.6747/\sqrt{9}}=1.788 \qquad \therefore$ don't reject H_0

我們沒有足夠證據去推論今年平均成績有進步。

(二)因為母體接近對稱，故假設母體中位數(η)等於平均數75.5

符號檢定

$H_0:\eta\le 75.5$ vs $H_1:\eta>75.5$

Data：

73	81	74	87	90	89	65	79	88
-	+	-	+	+	+	-	+	+

因此 $T_0=6$, n=9

T.S.: $T\sim Bin(n=9,p=\frac{1}{2})$

R.R.: Reject H_0 at $\alpha=0.05$ if $\alpha>p-value$

$\because p-value=\sum_{t=6}^{9}\binom{9}{t}(\frac{1}{2})^9=0.2539 \qquad \therefore$ Don't reject H_0

我們沒有足夠證據推論今年平均成績有進步。

符號等級檢定

$H_0:\eta\le 75.5$ vs $H_1:\eta>75.5$

Data：

73	81	74	87	90	89	65	79	88
-	+	-	+	+	+	-	+	+
2	5	3	6	9	8	1	4	7

因此 $W^{+*}=39$, n=9

T.S.: W^{+}

R.R.: Reject H_0 at $\alpha=0.05$ if $W^{+*}\ge\frac{9(9+1)}{2}-W_{0.05}=45-8=37$

$\because W^{+*}=39 \qquad \therefore$ reject H_0

我們有足夠證據推論今年平均成績有進步。

四、欲比較4種環保材質吸管的瑕疵率，研究部門以單因子（one-way）隨機實驗各進行5次檢測，實驗結果的敘述性統計如下：

Treat i	1	2	3	4
平均數($\bar{y}_{i\cdot}$)	15.6	13.8	7.2	11.4

$\bar{y}_{\cdot\cdot}^2=144$、$\sum_{i=1}^{4}\bar{y}_{i\cdot}^2=615.6$、$\sum_{j=1}^{5}\bar{y}_{\cdot j}^2=757.5$、$\sum_{i=1}^{4}\sum_{j=1}^{5}y_{ij}^2=3{,}320$，其中 y_{ij} 為第i種材質在第j次實驗的瑕疵率檢測值，$\bar{y}_{i\cdot}=\sum_{j=1}^{5}y_{ij}/5$、$\bar{y}_{\cdot j}=\sum_{i=1}^{4}y_{ij}/4$、$\bar{y}_{\cdot\cdot}=\sum_{i=1}^{4}\sum_{j=1}^{5}y_{ij}/20$

(一)請完成變異數分析（ANOVA）表，以適當的符號定義參數，寫出檢定4種材質瑕疵率有差異的假說，並在5%顯著水準下提出你的檢定結論。（10分）

(二)假設第1種環保材質是市面上常用的材質，其他3種是新材質，請寫出常用材質與新材質的

平均瑕疵率是否有差異，$\mu_1=\frac{1}{3}(\mu_2+\mu_3+\mu_4)$，的對比假說，並在5%顯著水準下提出你的檢定結論。（10分）

試題評析	本題為抽樣分配，相關內容都在課本中。
考點命中	1.《高點・高上統計學講義》第三回，第十二章例3，趙治勳編撰。 2.《高點・高上統計學講義》補充題庫，第二十二週例10，趙治勳編撰。

答：

(一)

ANOVA TABLE				
source	SS	d.f.	MS	F
材質	198	3	66	F^*=4.364
Error	242	16	15.125	
Total	440	19		

假設：$Y_{ji}=\mu+\alpha_j+\varepsilon_{ji}$，$\varepsilon_{ji}\overset{iid}{\sim}N(0,\sigma^2)$，$j=1,2,3,4$，$i=1,2,3,4,5$

$H_0:\mu_1=\mu_2=\mu_3=\mu_4$ vs H_1:至少一個$\mu_i\neq\mu_j$

T.S.: $F=\dfrac{MSR}{MSE}\sim F_{(3,16)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(3,16)}=3.24$

$\because F^*$=4.364　　$\therefore reject H_0$

我們有足夠的統計證據去推論材質為影響瑕疵率之重要因素。

(二)點估計：$\bar{Y}_1-\frac{1}{3}(\bar{Y}_2+\bar{Y}_3+\bar{Y}_4)\sim N(\mu_1-\frac{1}{3}(\mu_2+\mu_3+\mu_4),\frac{4}{15}\sigma^2)$

其中 $E[\bar{Y}_1-\frac{1}{3}(\bar{Y}_2+\bar{Y}_3+\bar{Y}_4)]=\mu_1-\frac{1}{3}(\mu_2+\mu_3+\mu_4)$

$V[\bar{Y}_1-\frac{1}{3}(\bar{Y}_2+\bar{Y}_3+\bar{Y}_4)]=\frac{\sigma^2}{5}+\frac{1}{9}(\frac{\sigma^2}{5}+\frac{\sigma^2}{5}+\frac{\sigma^2}{5})=\frac{4}{15}\sigma^2$

$H_0:\mu_1=\frac{1}{3}(\mu_2+\mu_3+\mu_4)$ vs $H_1:\mu_1\neq\frac{1}{3}(\mu_2+\mu_3+\mu_4)$

$\Rightarrow H_0:\mu_1-\frac{1}{3}(\mu_2+\mu_3+\mu_4)=0$ vs $H_1:\mu_1-\frac{1}{3}(\mu_2+\mu_3+\mu_4)\neq 0$

T.S.: $T=\dfrac{\bar{Y}_1-\frac{1}{3}(\bar{Y}_2+\bar{Y}_3+\bar{Y}_4)-(0)}{\sqrt{\frac{4}{15}MSE}}\sim t_{(16)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $T^*>t_{(16)0.05}=1.7459$

$\because T=\dfrac{15.6-\frac{1}{3}(13.8+7.2+11.4)-(0)}{\sqrt{\frac{4}{15}(15.125)}}=2.39$　　$\therefore$ reject H_0

我們有足夠證據去推論常用材質與新材質平均瑕疵率有差異。

附表

符號等級檢定(威爾卡森，臨界點 $W_{n',\alpha}$)

n'：去除等於中位數後的樣本數，α：顯著水準

樣本數 n'

單尾	雙尾	5	6	7	8	9	10	11	12	13	14	15	16
α=.05	α=.10	1	2	4	6	8	11	14	17	21	26	30	36
α=.025	α=.05		1	2	4	6	8	11	14	17	21	25	30
α=.01	α=.02			0	2	3	5	7	10	13	16	20	24
α=.005	α=.01				0	2	3	5	7	10	13	16	19

樣本數 n'

單尾	雙尾	17	18	19	20	21	22	23	24	25	26	27	28
α=.05	α=.10	41	47	54	60	68	75	83	92	101	110	120	130
α=.025	α=.05	35	40	46	52	59	66	73	81	90	98	107	117
α=.01	α=.02	28	33	38	43	49	56	62	69	77	85	93	102
α=.005	α=.01	23	28	32	37	43	49	55	61	68	76	84	92

樣本數 n'

單尾	雙尾	29	30	31	32	33	34	35	36	37	38	39	40
α=.05	α=.10	141	152	163	175	188	201	214	228	242	256	271	287
α=.025	α=.05	127	137	148	159	171	183	195	208	222	235	250	264
α=.01	α=.02	111	120	130	141	151	162	174	186	198	211	224	238
α=.005	α=.01	100	109	118	128	138	149	160	171	183	195	208	221

樣本數 n'

單尾	雙尾	41	42	43	44	45	46	47	48	49	50	39	40
α=.05	α=.10	303	319	336	353	371	389	408	427	446	466	271	287
α=.025	α=.05	279	295	311	327	344	361	379	397	415	434	250	264
α=.01	α=.02	252	267	281	297	313	329	345	362	380	398	224	238
α=.005	α=.01	234	248	262	277	292	307	323	339	356	373	208	221