100 年公務人員高等考試三級考試試題

類 科：統計
科 目：統計學

一、假設一盒中有紅球 r 個，N-r 個黑球。今隨機自盒中不重複抽取 n 球，並記錄取出紅球的個數為 Y。
㈠試求 Y 之期望值與變異數。
㈡若 $N=10$，$r=3$，$n=4$，求機率至少為 5/9 時，Y 之範圍。

【擬答】：

$r.v.Y$ ：紅球的個數

$\Rightarrow Y \sim HG(N,r,n)$

$$\Rightarrow f(y)=\begin{cases}\dfrac{C_y^r C_{n-y}^{N-r}}{C_n^N}, & y=0,1,2,\cdots,n \\ 0 & ,o.w\end{cases}$$

㈠

(1) $E(Y)=\sum y f(y)=\sum_{y=0}^{n} y\dfrac{C_y^r C_{n-y}^{N-r}}{C_n^N}=\dfrac{nr}{N}$ （須證明）

(2) $E\left[Y(Y-1)\right]=\Sigma y(y-1)f(y)=\sum_{y=0}^{n} y(y-1)\dfrac{C_y^r C_{n-y}^{N-r}}{C_n^N}$

$$=\frac{n(n-1)r(r-1)}{N(N-1)}$$

$\therefore Var(Y)=E(Y^2)-\left[E(Y)\right]^2$

$=E(Y(Y-1))+E(Y)-\left[E(Y)\right]^2$

$=\dfrac{n(n-1)r(r-1)}{N(N-1)}+\dfrac{nr}{N}-(\dfrac{nr}{N})^2$

$=\dfrac{n}{N}r(1-\dfrac{r}{N})(\dfrac{N-n}{N-1})$ (須證明)

㈡ $E(Y)=\dfrac{nr}{N}=\dfrac{4\times 3}{10}=1.2=\mu_y$

$Var(Y)=\dfrac{n}{N}r(1-\dfrac{r}{N})(\dfrac{N-n}{N-1})=\dfrac{4\times 3}{10}(1-\dfrac{3}{10})(\dfrac{10-4}{10-1})$

$=0.56=\sigma_Y^2 \Rightarrow \sigma_Y=\sqrt{0.56}$

由 Chebyshev's Inequality

$P(\left|Y-\mu_y\right|\le k\sigma_Y)\ge 1-\dfrac{1}{k^2}$

$\Rightarrow P(\left|Y-1.2\right|\le \dfrac{3}{2}\times\sqrt{0.56})\ge 1-\dfrac{1}{(\frac{3}{2})^2}=\dfrac{5}{9}$

$\Rightarrow P(-1.12\le Y-1.2\le 1.12)\ge\dfrac{5}{9}$

$\Rightarrow P(0.08\le Y\le 2.32)\ge\dfrac{5}{9}$

∴機率至少為 $\dfrac{5}{9}$ 時，Y 的範圍為 $0.08\le Y\le 2.32$

二、一個二元訊息 X=2 或-2，自 A 傳輸至 B。傳輸過程會受雜訊 N 干擾，於 B 接收之數據為 R=X+N，其中 N 為服從標準常態分布之雜訊。若 R＞0.5，則判定原輸入數據為 2；反之，則判定數據為-2。試求 A 輸入之原數據為 2 時，B 接收數據判定錯誤之機率。

【擬答】：

A 輸入原數據為 2 時，B 接收數據判定錯誤之機率為：當 $x=2 \quad \Rightarrow R \le 0.5$，且 $N \sim N(0.1)$

$\Rightarrow P(2+N \le 0.5) = P(N \le -1.5) = P(Z \le -1.5) = 0.0668$

三、令 $Y_1, Y_2, \cdots Y_n$ 為具獨立同分布、期望值 1/λ 得指數(exponential)隨機變數，試求 λ 的最大概似估計(Maximum :Likelihood Estimate)，並驗證其是否具不偏性(unbiasedness)與一致性(consistency)。

【擬答】：

$$(Y_1, Y_2, \ldots\ldots, Y_n) \overset{iid}{\sim} \exp(\lambda) \Rightarrow E(Y) = \frac{1}{\lambda} \Rightarrow f(y) = \lambda e^{-\lambda y};\ y>0$$

(一) 1. $L(\lambda) = \prod_{i=1}^{n} f(y_i;\lambda) = \lambda e^{-\lambda y_1} \times \lambda e^{-\lambda y_2} \times \ldots\ldots \times \lambda e^{-\lambda y_n} = \lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}$

2. $\ln L(\lambda) = \ln[\lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}] = n\ln\lambda - \lambda \sum_{i=1}^{n} y_i$

3. $\dfrac{\partial \ln L(\lambda)}{\partial \lambda} = \dfrac{n}{\lambda} - \sum_{i=1}^{n} y_i$

令 $\left.\dfrac{\partial \ln L(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}=0} \Rightarrow \dfrac{n}{\hat{\lambda}} - \sum_{i=1}^{n} y_i = 0$

$\Rightarrow \dfrac{n}{\hat{\lambda}} = \sum_{i=1}^{n} y_i \Rightarrow \hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} y_i} = \dfrac{1}{\overline{Y}}$，且 $\left.\dfrac{\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right|_{\lambda=\frac{1}{\overline{Y}}} = -\dfrac{n}{\overline{Y}^2} < 0$

所以 λ 的最大概似估計式 $\hat{\lambda}_{MLE} = \dfrac{1}{\overline{Y}}$

(二) $(Y_1, Y_2, \ldots\ldots, Y_n) \overset{iid}{\sim} \exp(\lambda)$

$\Rightarrow Y_1 + Y_2 + \ldots\ldots + Y_n \quad \underset{\sim}{iid}(\alpha = n, \lambda)$

$\Rightarrow \overline{Y} = \dfrac{\sum_{i=1}^{n} y_i}{n} \sim \Gamma(\alpha = n, n\lambda)$，令 $X = \overline{Y}$

$\Rightarrow f(x) = \dfrac{(n\lambda)^n}{\Gamma(n)} x^{n-1} e^{-n\lambda x}$；$x > 0$

1. $E(\hat{\lambda}) = E\left(\dfrac{1}{x}\right) = \int_0^\infty \dfrac{1}{x} f(x)\, d\overline{y} = \int_0^\infty \dfrac{1}{x} \dfrac{(n\lambda)^n}{\Gamma(n)} x^{n-1} e^{-n\lambda x} dx$

$= \dfrac{\Gamma(n-1)}{\Gamma(n)} (n\lambda) \int_0^\infty \dfrac{(n\lambda)^{n-1}}{\Gamma(n-1)} x^{(n-1)-1} e^{-n\lambda x} dx = \dfrac{\Gamma(n-1)}{\Gamma(n)} (n\lambda)$

$= \dfrac{\Gamma(n-1)}{(n-1)\Gamma(n-1)} n\lambda = \dfrac{n}{n-1} \lambda$

$\Rightarrow \lim_{n\to\infty} \dfrac{n}{n-1} \lambda = \lambda$

所以 $\dfrac{1}{y}$ 為漸近不偏估計式

2. $E(\hat{\lambda}^2) = E[(\dfrac{1}{x})^2] = \int_0^\infty \dfrac{1}{x^2} f(x)\, dx = \int_0^\infty \dfrac{1}{x^2} \dfrac{(n\lambda)^n}{\Gamma(n)} x^{n-1} e^{-n\lambda x} dx$

$$= \frac{\Gamma(n-2)}{\Gamma(n)}(n\lambda)^2\int_0^\infty \frac{(n\lambda)^{n-2}}{\Gamma(n-2)}\, x^{(n-2)-1}e^{-n\lambda x}dx$$

$$= \frac{\Gamma(n-2)}{\Gamma(n)}(n\lambda)^2 = \frac{\Gamma(n-2)}{(n-1)(n-2)\Gamma(n-2)}(n\lambda)^2$$

$$= \frac{n^2}{(n-1)(n-2)}\lambda^2$$

$$\Rightarrow Var(\hat{\lambda}) = E(\hat{\lambda}^2) - [E(\hat{\lambda})]^2$$

$$= \frac{n^2}{(n-1)(n-2)}\lambda^2 - \left(\frac{n}{n-1}\lambda\right)^2$$

$$= \frac{n^2}{(n-1)^2(n-2)}\lambda^2，且 \lim_{n\to\infty} Var(\hat{\lambda}) = 0$$

捫所以 $\hat{\lambda} = \frac{1}{y}$ 具有一致性

四、若 Y_1, Y_2 具獨立同分布在 $(\theta, \theta+1)$ 之均勻分布(uniform distribution)，檢定 $H_0:\theta=0$ vs. $H_a:\theta=0.5$，考慮拒絕域 $Y_1+Y_2>c$。

㈠試求 c 值，使其型一誤差(Type I error probability)為 0.02。

㈡試求㈠之檢定的檢定力(power)。

【擬答】：

$$(Y_1,Y_2)\overset{iid}{\sim}U(\theta,\theta+1)$$

$$\begin{cases} H_0:\theta=0 \\ H_a:\theta=0.5 \end{cases}$$

拒絕域：$Y_1+Y_2>C$

㈠當 H_0 為真時，即 $\theta=0$

$$\Rightarrow (Y_1,Y_2)\overset{iid}{\sim}U(0,1)$$

$$\Rightarrow \begin{cases} f(y_1)=1, 0\le y_1\le 1 \\ f(y_2)=1, 0\le y_2\le 1 \end{cases}$$

因為 y_1, y_2 獨立 $\Rightarrow f(y_1,y_2)=f(y_1)f(y_2)=1, 0\le y_1\le 1, 0\le y_2\le 1$

$$\Rightarrow \alpha = P(\text{Re}\,H_0 | H_0\text{為真}) = P(Y_1+Y_2>C|\theta=0)$$

$$= P(Y_1+Y_2>C) = \int_{C-1}^{1}\int_{C-y_2}^{1} f(y_1,y_2)dy_1dy_2$$

$$= \int_{C-1}^{1}\int_{C-y_2}^{1} |dy_1dy_2 = \int_{C-1}^{1}(1-C+y_2)dy_2$$

$$= \left[y_2 - Cy_2 + \frac{1}{2}y_2^2\right]\Bigg|_{y_2=C-1}^{y_2=1}$$

$$= \frac{1}{2}C^2 - 2C + 2 = 0.02$$

$$\Rightarrow C = \frac{9}{5} \text{或} \frac{11}{5}(\text{不合})$$

所以 $C = \frac{9}{5} = 1.8$

㈡當 H_0 為假，即 $\theta=0.5$

$$\Rightarrow (Y_1,Y_2)\overset{iid}{\sim}U(0.5,1.5)$$

$$\Rightarrow \begin{cases} f(y_1)=1, \quad 0.5\le y_1\le 1.5 \\ f(y_2)=1, \quad 0.5\le y_2\le 1.5 \end{cases}$$

因為 y_1,y_2 獨立 $\Rightarrow f(y_1,y_2)=1,\ 0.5\le y_1\le 1.5,\ 0.5\le y_2\le 1.5$

$$\Rightarrow power=1-\beta=P(\text{Re}\,H_0\,|\,H_0\text{為假})$$

$$=P(Y_1+Y_2>\frac{9}{5}\,|\,\theta=0.5)$$

$$=1-P(Y_1+Y_2\le\frac{9}{5})$$

$$=1-\int_{0.5}^{1.3}\int_{0.5}^{1.8-y_2}f(y_1,y_2)dy_1dy_2$$

$$=1-\int_{0.5}^{1.3}\int_{0.5}^{1.8-y_2}1\,dy_1dy_2$$

$$=1-\int_{0.5}^{1.3}(1.3-y_2)dy_2$$

$$=1-\left[1.3y_2-\frac{1}{2}y_2{}^2\right]\Big|_{y_2=0.5}^{y_2=1.3}$$

$$=1-0.32=0.68$$

所以檢定力為 0.68

五、在工廠環境安全調查中發現，工廠 A 在過去 6 週中發生意外事故的次數分別為 16, 9, 17, 19, 24, 8，而工廠 B 在過去 4 週中發生意外之次數則分別為 26, 30, 25, 28。假設工廠每週意外事故之次數服從獨立同分布的卜瓦松分布，在顯著水準 5%下，試求檢定兩工廠之安全環境是否相同時之 p-value(不需化簡)。

【擬答】：

$$(X_1, X_2, \ldots\ldots, X_6)\overset{iid}{\sim}Poi(\lambda_1)\Rightarrow X_1+X_2+\ldots\ldots+X_6=93$$

$$(Y_1, Y_2, Y_3, Y_4)\overset{iid}{\sim}Poi(\lambda_2)\Rightarrow Y_1+Y_2+Y_3+Y_4=109$$

$$\begin{cases}H_0：\lambda_1=\lambda_2\\ H_1：\lambda_1\ne\lambda_2\end{cases}$$

$$\Rightarrow P(\sum_{i=1}^{6}X_i=K\mid\sum_{i=1}^{6}X_i+\sum_{i=1}^{4}Y_i=n)$$

$$=\frac{P(\sum_{i=1}^{6}X_i=K,\sum_{i=1}^{4}Y_i=n-K)}{P(\sum_{i=1}^{6}X_i+\sum_{i=1}^{4}Y_i=n)}=\frac{\dfrac{e^{-6\lambda_1}(6\lambda_1)^K}{K!}\times\dfrac{e^{-4\lambda_2}(4\lambda_2)^{n-K}}{(n-K)!}}{\dfrac{e^{-(6\lambda_1+4\lambda_2)}(6\lambda_1+4\lambda_2)^n}{n!}}$$

$$=\frac{n!}{K!(n-K)!}(\frac{6\lambda_1}{6\lambda_1+4\lambda_2})^K(\frac{4\lambda_2}{6\lambda_1+4\lambda_2})^{n-K}=C_K^n P^K(1-P)^{n-K}$$

其中 $P=\dfrac{6\lambda_1}{6\lambda_1+4\lambda_1}=\dfrac{6}{10}=0.6$〔$H_0：\lambda_1=\lambda_2$〕

令 $W=\sum_{i=1}^{6}X_i\mid\sum_{i=1}^{6}X_i+\sum_{i=1}^{4}Y_i=n$

則 $W\sim b(n=202, p=0.6)$

又 $n=202$,且 $np\ge 5$，則二項分配可由常態分配逼近

其中 $\begin{cases}E(W)=np=202\times0.6=121.2\\ Var(W)=npq=202\times0.6\times0.4=48.48\end{cases}$

$\Rightarrow\begin{cases}H_0：\lambda_1=\lambda_2\\ H_1：\lambda_1\ne\lambda_2\end{cases}$ 即 $\begin{cases}H_0：P=\dfrac{3}{5}\\ H_1：P\ne\dfrac{3}{5}\end{cases}$

檢定統計量 $Z=\dfrac{93-121.2}{\sqrt{48.48}}=-4.05$

$\Rightarrow$p-value＝2P（Z＜－4.05）

六、下表為二元資料配適迴歸模型所得的部分輸出報表：

來源(Source)	自由度(df)	平方和(Sum of squares)
模型(Model)	1	307.247
誤差(Error)	8	63.653
總和(C.Total)	9	307.900

㈠試求判定係數(coefficient of determination) R^2 值。

㈡在那些條件下，可以檢定迴歸模型的顯著性？

㈢假設㈡之條件成立，檢定迴歸模型是否顯著。(α=5%, 臨界值(critical value)=5.32)

【擬答】：

㈠ $R^2=\dfrac{SSR}{SST}=\dfrac{307.247}{370.900}=0.8284$

㈡ $Y_i=\alpha+\beta X_i+\varepsilon_i : i=1,2,\cdots\cdots,n$

⑴ ε_i 為隨機誤差(Random Error Term)，且須滿足以下條件：

① ε_i 為隨機誤差，且 $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

即 ε_i 為常態分配且 $E(\varepsilon_i)=0$ ，$Var(\varepsilon_i)=\sigma^2$

② $Cov(\varepsilon_i,\varepsilon_j)=0 : i\neq j$

即任二組隨機誤差彼此獨立。

③ $Cov(\varepsilon_i,x)=0$

即隨機誤差與 X 變項無關。

⑵ α 、β 為迴歸係數，α 為截距，β 為斜率。

⑶ X_i ：自變項，非隨機變數。

⑷ Y_i ：依變項，為隨機變數。

【參考：林惠玲，陳正倉著統計學方法與應用三版(上、下)：97 年 9 月：雙葉書廊：P327~P238】

㈢ $\begin{cases}H_0:\beta=0\\H_1:\beta\neq 0\end{cases}$

$\alpha=0.05$

拒絕域 $C=\{F\,|F>F_{0.05}(1.8)=5.32\}$

檢定統計量 $F=\dfrac{MSR}{MSE}=\dfrac{307.247/1}{63.653/8}=38.62>5.32$

$\in C\Rightarrow \mathrm{Re}\,H_0$

結論：有證據顯示 $\beta\neq 0$，即迴歸模型達顯著水準

七、下表為一完全隨機設計(complete randomized design)實驗的觀測數據：

處理項 (Treatment)	觀測值 (Observations)			
1	2	1	3	
2	1	5		
3	9	5	6	4
4	3	4	5	

試寫出 ANOVA 表(Analysis of Variance Table)，並做出 F 檢定的結論。

（α=5%,臨界值(critical value)=4.07）

【擬答】：

處理項	觀察值				和
1	2	1	3		$T_1.=6$
2	1	5			$T_2.=6$
3	9	5	6	4	$T_3.=24$
4	3	4	5		$T_4.=12$

$T..=6+6+24+12=48$

$n_1=3,\ n_2=2,\ n_3=4,\ n_4=3,\ N=12$

$\sum\sum X_{ij}{}^2=2^2+1^2+\cdots+5^2=248$

① $SST=\sum\sum X_{ij}{}^2-\frac{T..^2}{N}=248-\frac{48^2}{12}=56$

② $SSR=\sum\frac{T_i.^2}{n_i}-\frac{T..^2}{N}=\left[\frac{6^2}{3}+\frac{6^2}{2}+\frac{24^2}{4}+\frac{12^2}{3}\right]-\frac{48^2}{12}=30$

③ $SSE=SST-SSR=56-30=26$

ANOVA 表

來源	SS	df	MS	F 值
組間	30	3	10	$F=3.08$
組內	26	8	3.25	
總變異	56	11		

$\begin{cases} H_0:\ \mu_1=\mu_2=\mu_3=\mu_4 \\ H_1:\ \mu_i\text{不全相同},\ i=1,2,3,4 \end{cases}$

α=0.05

拒絕域 $C=\{F\,|\,F>F_{0.05}(3.8)=4.07\}$

檢定統計量 $F=3.08<4.07\notin C\Rightarrow not\ \mathrm{Re}\,Ho$

結論：沒有證據顯示 μ_i 不全相同，$i=1,2,3,4$