【統計、經建行政】

《統計學》

試題評析
今年普考統計學考題內容涵蓋了多項分配、分割定理、常態分配及其性質、區間估計、檢定、無母數檢定等，各單元考題分布還算平均。考題難易適中，但部份考題計算需時較長，部份查表值也需靠內外插法才能求出；除此之外，考生只要掌握基本題型，應能獲致高分。一般考生考60分以上不是問題，程度好的考生甚至可拿90分以上。

一、若棒球名投手王建民去年投球紀錄為投出直球、下墜球與上飄球之比率為28:34:38。今欲觀測其未來投出10球中，直球、下墜球與上飄球之球數，則：
(一)請定義必要之隨機變數並列出其衍生的樣本空間(induced sample space)。(5分)
(二)請推導(Derive)其機率函數(probability function)，並敘明該機率模式之名稱。(10分)
(三)請問其恰投出2次直球、3次下墜球與5次上飄球之機率為何？(5分)
(四)請問其投出直球之期望數(expected number)與變異數(variance)為何？(5分)
(五)若已知投手王建民投出直球、下墜球與上飄球為好球之機率分別為0.88，0.82與0.8，則請問其投出好球之機率為何？(5分)

答：

(一)令 X_1 表王建民未來投出10球中直球之球數
令 X_2 表王建民未來投出10球中下墜球之球數
令 X_3 表王建民未來投出10球中上飄球之球數
樣本空間為 $\Omega=\{(x_1,x_2,x_3)\mid x_1+x_2+x_3=10;x_i=0,1,\cdots,10;i=1,2,3\}$

(二)1. $f(x_1,x_2,x_3)=\dfrac{10!}{x_1!x_2!x_3!}(0.28)^{x_1}(0.34)^{x_2}(0.38)^{x_3}$，

其中 $x_1+x_2+x_3=10;x_i=0,1,\cdots,10;i=1,2,3$

2.（X_1,X_2,X_3）$\sim M_3(10;\dfrac{28}{100},\dfrac{34}{100},\dfrac{38}{100})$ 為三項分配

(三) $f(2,3,5)=\dfrac{10!}{2!3!5!}(0.28)^2(0.34)^3(0.38)^5=0.0615$

(四) $\because X_1\sim B(10,0.28)$
$\therefore E(X_1)=10(0.28)=2.8$
$Var(X_1)=10(0.28)(1-0.28)=2.016$

(五)令A表王建民投出直球之事件
令B表王建民投出下墜球之事件
令C表王建民投出上飄球之事件
令G表王建民投出上飄球之事件
則

$P(G) = P(A) \bullet P(G|A)+P(B) \bullet P(G|B)+P(C) \bullet P(G|C)$
$= 0.28 \bullet 0.88+0.34 \bullet 0.82+0.38 \bullet 0.8 = 0.8292$

二、若考試成績X服從平均數(mean)與μ與標準差(standard deviation)為σ之常態分配(normal distribution)，則：

(一)請列出隨機變數(random variable)X之機率密度函數(probability density function)。（2分）

(二)若評審依下表之等級(Grades)評定考試成績，試求各等級成績之百分比(percentage)？（7分）

Scores X	$X \geq \mu+3\sigma$	$X \in [\mu+2\sigma, \mu+3\sigma]$	$X \in [\mu+\sigma, \mu+2\sigma]$	$X \in [\mu, \mu+\sigma]$	$X \in [\mu-\sigma, \mu]$	$X \in [\mu-2\sigma, \mu-\sigma]$	$X < \mu-2\sigma$
Grades	A^+	A	A^-	B	C	D	F

(三)令Z=(X-μ)/σ，試分別求Z之機率密度函數、期望數E(Z)與變異數V(Z)為何？（10分）

(四)若考生總數為100,000人，試分別求各等級成績之考生人數為何？（7分）

(五)若考試成績欲轉換為常態分配平均數70與標準差10之常態分數，則：

1.請定義必要之隨機變數(random variable)並列出相對應之機率密度函數。（3分）

2.試求各等級成績之界限(boundary)為何？（7分）

3.試分別求中央50%與80%成績之界限(boundary)為何？（4分）

答：

(一)若X服從常態分配N(μ,σ^2),則其p.d.f爲：

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \text{ , } x\text{、} \mu \in \mathrm{R} \text{ , } \sigma > 0$$

(二)成績爲F的人佔全部考生的

$P(X < \mu-2\sigma) = P(Z < -2)$=1-0.9772=0.0228

成績爲D的人佔全部考生的

$P(\mu-2\sigma \leq X \leq \mu-\sigma) = P(-2 \leq Z \leq -1) = P(1 \leq Z \leq 2) = 0.9772-0.8413 = 0.1359$

成績爲C的人佔全部考生的

$P(\mu-\sigma \leq X \leq \mu) = P(-1 \leq Z \leq 0) = P(0 \leq Z \leq 1) = 0.8413-0.5 = 0.3413$

成績爲B的人佔全部考生的

$P(\mu \leq X \leq \mu+\sigma) = P(0 \leq Z \leq 1) = 0.3413$

成績爲A^-的人佔全部考生的

$P(\mu+\sigma \leq X \leq \mu+2\sigma) = P(1 \leq Z \leq 2) = 0.1359$

成績爲A的人佔全部考生的

$P(\mu+2\sigma \leq X \leq \mu+3\sigma) = P(2 \leq Z \leq 3) = 0.9987-0.9772 = 0.0215$

成績爲A^+的人佔全部考生的

$P(\mu+3\sigma \leq X) = P(3 \leq Z) = 1-0.9987 = 0.0013$

(三)

1. $Z = \frac{X-\mu}{\sigma}$，則稱Z爲標準常態分配，簡記爲Z~N(0,1)。且 $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

2. $M_Z(t)=\int_{-\infty}^{\infty}\left(e^{tz}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\right)dz=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(z^2-2tz\right)}dz=\frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}t^2}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(z-t)^2}dz$

$$=\frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}t^2}\cdot\sqrt{2\pi}=e^{\frac{1}{2}t^2}$$

$$M'_Z(t)=te^{\frac{1}{2}t^2}$$

$$M''_Z(t)=(t^2+1)e^{\frac{1}{2}t^2}$$

$$E(Z)=M'_Z(0)=0$$

$$E(Z^2)=M''_Z(0)=(0^2+1)e^{\frac{1}{2}(0)^2}=1$$

$$Var(Z)=E(Z^2)-[E(Z)]^2=1-0^2=1$$

(四)

Grades	A^+	A	A^-	B	C	D	F
各等級成績所佔百分比	0.0228	0.1359	0.3413	0.3413	0.1359	0.0215	0.0013
各等級成績考生人數	2280	13590	34130	34130	13590	2150	130

(五)

1.(1)X~N(μ,σ^2)

$$\Rightarrow Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

$$\Rightarrow 10Z=10(\frac{X-\mu}{\sigma})\sim N(0,10^2)$$

$$\Rightarrow 10Z+70=10(\frac{X-\mu}{\sigma})+70\sim N(70,10^2)$$

令常態分數M=$10Z+70=10(\frac{X-\mu}{\sigma})+70$即爲所求

(2)因爲$M\sim N(70,10^2)$，所以由(一)可知M之pdf爲

$f_M(m)=\frac{1}{10\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{m-70}{10})^2}$　其中$m\in$R

2.當 $X=\mu-2\sigma \Rightarrow M=10(\frac{\mu-2\sigma-\mu}{\sigma})+70=50$

當 $X=\mu-\sigma \Rightarrow M=10(\frac{\mu-\sigma-\mu}{\sigma})+70=60$

當 $X=\mu \Rightarrow M=10(\frac{\mu-\mu}{\sigma})+70=70$

當 $X=\mu+\sigma \Rightarrow M=10(\frac{\mu+\sigma-\mu}{\sigma})+70=80$

當 $X=\mu+2\sigma \Rightarrow M=10(\frac{\mu+2\sigma-\mu}{\sigma})+70=90$

當 $X=\mu+3\sigma \Rightarrow M=10(\frac{\mu+3\sigma-\mu}{\sigma})+70=100$

故新的常態分數M各等級成績之界限如下：

M	$M\geq100$	$90\leq M\leq100$	$80\leq M\leq90$	$70\leq M\leq80$	$60\leq M\leq70$	$50\leq M\leq60$	$M\leq50$
Grades	A^+	A	A^-	B	C	D	F

3.(1)令中央50%的界限各為a、b，亦即

$$P(a\leq M\leq b)=50\%=P(\frac{a-70}{10}\leq\frac{M-70}{10}\leq\frac{b-70}{10})\cong P(-0.675\leq Z\leq0.675)$$

$$\Rightarrow\begin{cases}\frac{a-70}{10}\cong-0.675\\ \frac{b-70}{10}\cong0.675\end{cases}\Rightarrow\begin{cases}a\cong63.25\\ b\cong76.75\end{cases}$$

(2)令中央80%的界限各為c、d，亦即

$$P(c\leq M\leq d)=80\%=P(\frac{c-70}{10}\leq\frac{M-70}{10}\leq\frac{d-70}{10})\cong P(-1.28\leq Z\leq1.28)$$

$$\Rightarrow\begin{cases}\frac{c-70}{10}\cong-1.28\\ \frac{d-70}{10}\cong1.28\end{cases}\Rightarrow\begin{cases}c\cong57.2\\ d\cong82.8\end{cases}$$

三、茲為研究臺北地區房屋售價（單位：新台幣萬元／坪）之實情，乃於該區隨機抽出樣本其分組次數分配表如下：

Class Boundaries	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	4	6	12	15	25	16	12	6	4

（一）若顯著水準(significance level)為5%，試檢定臺北地區房屋售價之母體是否為常態(normal)？（10分）

(二)若臺北地區房屋售價之母體為常態分配下，請列出臺北地區房屋售價變異數之95%信賴區間(confidence interval)估計值(estimate)。(5分)

(三)若臺北地區房屋售價之母體為常態分配下，請列出臺北地區房屋售價平均數之95%信賴區間估計值。(5分)

(四)若臺北地區房屋售價之母體為常態分配下，在顯著水準為5%的情形下，試檢定臺北地區房屋售價平均數是否為550,000？(10分)

答：

(一)

Class Boundaries	10~20	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100	合計
midpoint(m_i)	15	25	35	45	55	65	75	85	95	
Frequency(f_i)	4	6	12	15	25	16	12	6	4	100
$m_i \cdot f_i$	60	150	420	675	1375	1040	900	510	380	5510
$m_i^2 \cdot f_i$	900	3750	14700	30375	75625	67600	67500	43350	36100	339900

由上表可得母體平均數μ及母體變異數σ^2的估計值分別為：

$$\hat{\mu} = \bar{x} = \frac{\sum_i m_i \cdot f_i}{\sum_i f_i} = \frac{5510}{100} = 55.1$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\left[\sum_i m_i^2 \cdot f_i - n(\bar{x})^2\right] = \frac{1}{100-1}\left[339900 - 100(55.1)^2\right] = 366.6566$$

設 X 表示台北地區房屋售價

① H_0: X~ $N(\mu, \sigma^2)$

② H_1: H_0不成立

③ $\alpha = 0.05$

④ C={ $\chi^2 \mid \chi^2 > \chi^2_{0.05}(4) = 9.488$ }，其中$P(\chi^2 > \chi^2_\alpha) = \alpha$

⑤檢定統計量 $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} = 2.1014 \notin C$

X	10~30	30~40	40~50	50~60	60~70	70~80	80~100
O_i	10	12	15	25	16	12	10
H_0成立下之 p_{i0}	0.0857	0.1202	0.1798	0.206	0.1808	0.1215	0.0872
e_i	8.57	12.02	17.98	20.6	18.08	12.15	8.72
O_i - e_i	1.43	-0.02	-2.98	4.4	-2.08	-0.15	1.28

（其中理論次數e_i <5者與鄰組合併）

⑥不拒絕H_0，即在顯著水準0.05下，無足夠證據證明台北地區房屋售價不服從常態分配。

(二) $\left(\dfrac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)=\left(\dfrac{(99)(366.6566)}{\chi^2_{0.025}(99)},\dfrac{(99)(366.6566)}{\chi^2_{0.975}(99)}\right)$

$$=\left(\frac{(99)(366.6566)}{\chi^2_{0.025}(99)},\frac{(99)(366.6566)}{\chi^2_{0.975}(99)}\right)$$

$$\approx\left(\frac{(99)(366.6566)}{128.6020},\frac{(99)(366.6566)}{73.0910}\right)$$

$$\approx(282.2584,496.6275)$$

(三) $\overline{X}\pm Z_{0.025}\dfrac{S}{\sqrt{n}}=55.1\pm 1.96\cdot\dfrac{19.1483}{\sqrt{100}}=(51.3469,58.8531)$ (萬元/坪)

(四)

① $H_0:\mu=55$(萬元/坪)
② $H_1:\mu\neq 55$
③ $\alpha=0.05$
④C={ $Z\,|\,|Z|>Z_{0.025}=1.96$ }其中 $P(Z>Z_\alpha)=\alpha$
⑤ $Z=\dfrac{\overline{X}-55}{S/\sqrt{n}}=\dfrac{55.1-55}{19.1483/\sqrt{100}}=0.0522\notin C$
⑥結論：不拒絕 H_0，亦即在顯著水準 $\alpha=0.05$ 下，我們的資料不足以證明台北地區房屋售價平均數不是每坪550000元。