第二章 【p25 复习题】

1. 设总体X服从P( λ )分布，试写出样本$X_1$、$X_2$、⋯、$X_n$的联合分布律。

解：$P\{X_i=x_i\}=p(x_i)=\dfrac{\lambda^{x_i}}{x_i!}e^{-\lambda}$，

$$P\{X_1=x_1, X_2=x_2, \cdots, X_n=x_n\}=\prod_i p(x_i)=\frac{\lambda^{\sum_i x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda}。$$

2. 设总体X服从N(0,1)分布，试写出样本$X_1$、$X_2$、⋯、$X_n$的联合分布密度。

解：$p(x_i)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x_i^2}{2}}$， $p^*(x_1, x_2, \cdots, x_n)=\prod_i p(x_i)=\left(\dfrac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_i x_i^2}{2}}$。

3. 设总体X服从N( μ , σ $^2$ )分布,试写出样本$X_1$、$X_2$、⋯、$X_n$的联合分布密度。

解：$p(x_i)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$，

$$p^*(x_1, x_2, \cdots, x_n)=\prod_i p(x_i)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_i \frac{(x_i-\mu)^2}{2\sigma^2}}。$$

画直方图的SAS程序：

```
data ex; input x @@;
cards;
70 72 94 24 68 57 90 185 95 93
109 64 58 79 40 118 84 70 99 132
154 100 77 34 68 26 48 87 85 95
123 105 107 55 45 73 109 58 101 134
94 94 62 156 61 84 77 123 135 40
107 79 131 72 66 30 44 141 98 100
90 78 44 50 58 60 76 78 92 101
62 152 97 81 54 98 75 118 130 90
115 136 100 80 69 98 84 25 179 97
76 56 73 43 22 82 60 68 160 139
;
proc gchart;
vbar x/type=pct; run;
```

计算单变量数字特征的SAS程序：

1）不需要输入频数的程序

```
data ex; input x @@;
cards;
1 2 2 3 3 3 4 5 6 7 8
;
proc univariate vardef=n; run;
```

2）需要输入频数的程序

```
data ex; input x f @@; (f 是频数)
cards;
5.5 4 7.5 11 9.5 17 11.5 23
13.5 18 15.5 14 17.5 10 19.5 3
;
proc univariate vardef=n; var x; freq f; run;
```

计算协方差及相关系数的 SAS 程序

```
data ex; input x y @@;
cards;
1.58 180 9.98 28 9.42 25
1.25 117 0.3 165 2.41 175
11.01 40 1.85 160 6.04 120 5.92 80
;
proc corr cov vardef=n; run;
```

【p30 复习题】

1. 观测5头基础母羊的体重(单位：kg)分别为53.2,51.3,54.5,47.8,50.9，试计算这个样本观测值的数字特征：(1)样本总和，(2)样本均值，(3)离均差平方和，(4)样本方差，(5)样本标准差，(6)样本修正方差，(7)样本修正标准差，(8)样本变异系数，(9)众数，(10)中位数，(11)极差，(12)75%分位数。答案：⑴257.7，⑵51.54，⑶25.972，⑷5.194，⑸2.279，⑹6.493，⑺2.548，⑻4.944%，⑼无，(10)51.3，(11)6.7，(12)53.2。

2. 观测100支金冠苹果枝条的生长量(单位：cm)得到频数表如下：

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 组下限 | 19.5 | 24.5 | 29.5 | 34.5 | 39.5 | 44.5 | 49.5 | 54.5 | 59.5 |
| 组上限 | 24.5 | 29.5 | 34.5 | 39.5 | 44.5 | 49.5 | 54.5 | 59.5 | 64.5 |
| 组中值 | 22 | 27 | 32 | 37 | 42 | 47 | 52 | 57 | 62 |
| 频数 | 8 | 11 | 13 | 18 | 18 | 15 | 10 | 4 | 3 |

试计算这个样本观测值的数字特征：(1)样本总和，(2)样本均值，(3)离均差平方和，(4)样本方差，(5)样本标准差，(6)样本修正方差，(7)样本修正标准差，(8)样本变异系数，(9)众数，(10)中位数，(11)极差，(12)75%分位数。答案：⑴3950，⑵39.5，⑶10275，⑷102.75，⑸10.137，⑹103.788，⑺10.188，⑻25.792%，⑼37,42，(10)39.5，(11)40，(12)47。

3. 观测10个华农本地早桔果横径X（单位：厘米）与果重Y（单位：克）如下：

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 果横径$x_i$ | 7.0 | 6.5 | 5.8 | 4.1 | 5.5 | 6.7 | 6.3 | 4.2 | 6.1 | 5.1 |
| 果重$y_i$ | 117 | 96 | 79 | 44 | 62 | 106 | 88 | 48 | 85 | 55 |

试计算这个样本观测值的协方差矩阵与相关系数矩阵。

答：$\begin{pmatrix} 0.9061 & 21.84 \\ 21.84 & 562 \end{pmatrix}$，$\begin{pmatrix} 1 & 0.9768 \\ 0.9768 & 1 \end{pmatrix}$。

4. 对棉纱支数X与断点拉力Y（单位：克）进行36次观测得到的数据如下：

| 拉力 \ 支数 | | 41 | 43 | 45 | 47 | 49 | 51 |
|---|---|---|---|---|---|---|---|
| 6.25 | 重 | 1 | 2 | 2 | 1 | 0 | 0 |
| 5.75 | 复 | 0 | 1 | 3 | 4 | 2 | 3 |
| 5.25 | 数 | 0 | 3 | 5 | 7 | 1 | 1 |

求这个样本观测值的协方差矩阵与相关系数矩阵。

答：$\begin{pmatrix} 6.2840 & -0.1466 \\ -0.1466 & 0.1364 \end{pmatrix}$，$\begin{pmatrix} 1 & -0.1584 \\ -0.1584 & 1 \end{pmatrix}$。

5. 设总体X的数学期望为E(X)、方差为D(X)，取自总体的样本容量为n，试证明：

(1)E($\overline{X}$)=E(X)；(2)D($\overline{X}$)=D(X)。

证：设样本为$X_1$、$X_2$、…、$X_n$，则$X_1$、$X_2$、…、$X_n$相互独立且与总体同分布，$E(X_1)=E(X_2)=\cdots=E(X_n)=E(X)$，$D(X_1)=D(X_2)=\cdots=D(X_n)=D(X)$，

$$E(\overline{X})=E\left(\frac{1}{n}\sum_i X_i\right)=\frac{1}{n}E\left(\sum_i X_i\right)=\frac{1}{n}\sum_i E(X_i)=\frac{1}{n}\sum_i E(X)=\frac{1}{n}\bullet nE(X)=E(X);$$

$$D(\overline{X})=D\left(\frac{1}{n}\sum_i X_i\right)=\frac{1}{n^2}D\left(\sum_i X_i\right)=\frac{1}{n^2}\sum_i D(X_i)=\frac{1}{n^2}\sum_i D(X)=\frac{1}{n^2}\bullet nD(X)=\frac{1}{n}D(X)。$$

6. 设总体X的数学期望为E(X)、方差为D(X)，取自总体的样本容量为n，试计算：

(1)$E(\sum_i X_i)$；(2) $E(\sum_i X_i^2)$；(3)$E(2\sum_{i<j}\sum X_iX_j)$；(4)$E(\sum_i X_i)^2$。

解：(1)$E(\sum_i X_i)=\sum_i E(X_i)=\sum_i E(X)=nE(X)$；

(2)

$$E(\sum_i X_i^2)=\sum_i E(X_i^2)=\sum_i\left\{D(X_i)+[E(X_i)]^2\right\}=\sum_i\left\{D(X)+[E(X)]^2\right\}=n\left\{D(X)+[E(X)]^2\right\};$$

$$(3)E(2\sum_{i<j}\sum X_iX_j)=2\sum_{i<j}\sum E(X_iX_j)=2\sum_{i<j}\sum E(X_i)E(X_j)=2\sum_{i<j}\sum E(X)E(X)$$

$$=2\sum_{i<j}\sum[E(X)]^2=n(n-1)[E(X)]^2;$$

$$(4)E(\sum_i X_i)^2=E(\sum_i X_i^2+2\sum_{i<j}\sum X_iX_j)=n\left\{D(X)+[E(X)]^2\right\}+n(n-1)[E(X)]^2。$$

【p37 复习题】

1. 设总体的X服从N(1,4)分布，样本容量为16，均值为$\overline{X}$，试求：

① P{|X−0.5|<1}；② P{|$\overline{X}$ −0.5|<1}。

解：P{|X−0.5|<1}= P{−0.5<X<1.5}

$$=P\left\{\frac{-0.5-1}{2}<\frac{X-1}{2}<\frac{1.5-1}{2}\right\}\approx F0,1(0.25)-F0,1(-0.75)=0.3721;$$

P{|$\overline{X}$ −0.5|<1}= P{−0.5<$\overline{X}$ <1.5}

$$=P\left\{\frac{-0.5-1}{1/2}<\frac{\overline{X}-1}{1/2}<\frac{1.5-1}{1/2}\right\}\approx F0,1(1)-F0,1(-3)=0.8413。$$

2. 若总体X～N(μ, σ $^2$)有两个容量都是n、均值分别为$\overline{X}_1$ 与$\overline{X}_2$ 的独立样本，且P{|$\overline{X}_1-\overline{X}_2$|>σ}≈0.01，试求n。

解：P{|$\overline{X}_1-\overline{X}_2$|>σ}=1−P{|$\overline{X}_1-\overline{X}_2$|<σ}

$$=1-P\left\{\frac{|\overline{X}_1-\overline{X}_2|}{\sqrt{\frac{2\sigma^2}{n}}}<\sqrt{\frac{n}{2}}\right\}\approx 0.01,\ \sqrt{\frac{n}{2}}=2.58,\ n\approx 14。$$

3. 设有A、B两种牌号的电子元件，平均使用寿命为4000与4500小时，标准差为300与200小时。如果自A种牌号的电子元件中抽出100件作为随机样本进行测试，自B种牌号的电子元件中抽出50件作为随机样本进行测试，那么B种牌号的电子元件较A种牌号的电子元件，其平均寿命至少超出450小时的概率是多少？

解：P{|$\mu_B-\mu_A$|>450}=1− P{|$\mu_B-\mu_A$|<450}

$$=1-P\left\{\frac{\mu_B-\mu_A}{\sqrt{\frac{(300)^2}{100}}+\sqrt{\frac{(200)^2}{50}}}<\frac{-50}{\sqrt{1700}}\right\}\approx 1-F0,1(-1.21)=0.8869。$$

4. 设有A、B两种牌号的电子元件，平均使用寿命的标准差为40与50小时。如果自A种牌号的电子元件中抽出8件作为随机样本进行测试，自B种牌号的电子元件中抽出16件作为随机样本进行测试，那么A种牌号的电子元件较B种牌号的电子元件，其平均寿命的方差大过1倍的概率是多少？

解：$P\left\{\frac{S_A^2}{S_B^2}>2\right\}$=$P\left\{\frac{8S_A^2}{16S_B^2}>\frac{2\times 8}{16}\right\}$=$P\left\{\frac{8S_A^2\div 7}{16S_B^2\div 15}>\frac{15}{7}\right\}$

$$=P\left\{\frac{S^{*2}_A\div(40)^2}{S^{*2}_B\div(50)^2}>\frac{15\times(50)^2}{7\times(40)^2}\right\}=P\{F(7,15)>3.35\},$$

因为$F_{0.975}(7,15)=3.29$, $F_{0.99}(7,15)=4.14$, 所以$P\left\{\frac{S_A^2}{S_B^2}>2\right\}<0.025$。

5. 求相继出生的200个婴儿中，男孩的比例在43%至57%之间的概率。

解：$P\{86<\sum_i X_i<114\}\approx P\left\{\frac{85.5-100}{\sqrt{50}}<\frac{\sum_i X_i-100}{\sqrt{50}}<\frac{114.5-100}{\sqrt{50}}\right\}$

$\approx$F0,1(2.05)– F0,1(–2.05)=0.9596。

6. 选举结果表明，某候选人获得选票总数的65%。如果任取两个独立的随机样本，每个样本由200票组成，试求两样本中投票选举这个候选人的比例相差超过10%的概率。

解：$P\left\{\frac{\left|\sum_i X_i-\sum_i Y_i\right|}{200}>0.1\right\}=1-P\left\{\frac{\left|\sum_i X_i-\sum_i Y_i\right|}{200}\le 0.1\right\}$

$$\approx 1-P\left\{\frac{-0.1\times 200-0.5}{200}\le\frac{\sum_i X_i-\sum_i Y_i}{200}\le\frac{0.1\times 200+0.5}{200}\right\}$$

$$=1-P\left\{\frac{-0.1025}{\sqrt{0.002275}}\le\frac{\overline{X}-\overline{Y}}{\sqrt{0.002275}}\le\frac{0.1025}{\sqrt{0.002275}}\right\}$$

$\approx$1–(F0,1(2.15)– F0,1(–2.15))=0.0316。

7. 设总体$X\sim N(\mu,\sigma^2)$分布，样本容量为n，均值为$\overline{X}$，方差为$S_n^2$，$X_{n+1}$与$X_1$、$X_2$、…、$X_n$相互独立且分布相同，试论述$\frac{X_{n+1}-\overline{X}}{S_n}\sqrt{\frac{n-1}{n+1}}$的分布。

解：$X_{n+1}\sim N(\mu,\sigma^2)$，$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$，$X_{n+1}$与$\overline{X}$相互独立，根据可加性，

$X_{n+1}-\overline{X}\sim N(0,\sigma^2+\frac{\sigma^2}{n})$，而$\frac{nS_n^2}{\sigma^2}\sim\chi^2(n-1)$，因此

$$\frac{X_{n+1}-\overline{X}}{S_n}\sqrt{\frac{n-1}{n+1}}=\frac{\dfrac{X_{n+1}-\overline{X}}{\sqrt{\sigma^2+\dfrac{\sigma^2}{n}}}}{\sqrt{\dfrac{nS_n^2}{\sigma^2(n-1)}}}\sim t\,(N-1)。$$

8. 设总体 X～N(μ$_1$,σ$^2$)分布，X～N(μ$_2$,σ$^2$)分布，X 的样本容量为 m，Y 的样本容量为 n，两样本相互独立，均值为$\overline{X}$及$\overline{Y}$，离均差平方和为 SSX 及 SSY，a 及 b 是两个常数，试论述$\dfrac{a(\overline{X}-\mu_1)+b(\overline{Y}-\mu_2)}{S_w\sqrt{\dfrac{a^2}{m}+\dfrac{b^2}{n}}}$的分布，式中的$S_w=\sqrt{\dfrac{SSX+SSY}{m+n-2}}$。

解：$a(\overline{X}-\mu_1)+b(\overline{Y}-\mu_2)\sim N(0,\left(\dfrac{a^2}{m}+\dfrac{b^2}{n}\right)\sigma^2)$，$\dfrac{SSX+SSY}{\sigma^2}\sim\chi^2(m+n-2)$，因此

$$\frac{a(\overline{X}-\mu_1)+b(\overline{Y}-\mu_2)}{S_w\sqrt{\dfrac{a^2}{m}+\dfrac{b^2}{n}}}=\frac{\dfrac{a(\overline{X}-\mu_1)+b(\overline{Y}-\mu_2)}{\sqrt{\left(\dfrac{a^2}{m}+\dfrac{b^2}{n}\right)\sigma^2}}}{\sqrt{\dfrac{SSX+SSY}{\sigma^2\,(m+n-2)}}}\sim t(m+n-2)。$$

9. 设(X,Y)服从 N($\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho$)分布，($X_i$,$Y_i$)是这个二维总体的样本，容量为 n, 若$\overline{X}$及$\overline{Y}$是样本均值，$S_1^2$及$S_2^2$是样本方差，R 是样本的相关系数，$\overline{X}-\overline{Y}$与$S_1^2+S_2^2-$ 2R$S_1S_2$相互独立，试论述$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2+S_2^2-2RS_1S_2}{n-1}}}$的分布。

解：($X_i$,$Y_i$)～N($\mu_1-\mu_2,\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2$)，$\overline{X}\sim N(\mu_1,\dfrac{\sigma_1^2}{n})$，$\overline{Y}\sim N(\mu_2,\dfrac{\sigma_2^2}{n})$，

$\text{cov}(\overline{X},\overline{Y})=\dfrac{1}{n^2}\text{cov}(\sum_i X_i,\sum_i Y_i)=\dfrac{1}{n^2}(n\rho\sigma_1\sigma_2)$，

$D(\overline{X}-\overline{Y})=\dfrac{1}{n}(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2)$，$\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\dfrac{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}{n})$，

当$\dfrac{S_1^2+S_2^2-2RS_1S_2}{\frac{1}{n}(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2)}\sim\chi^2(n-1)$时，$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2+S_2^2-2RS_1S_2}{n-1}}}\sim t\,(n-1)$。