习题 A

1．以下是某工厂通过抽样调查得到的 10 名工人一周内生产的产品数

149　156　160　138　149　153　153　169　156　156

试由这一些样本数据构造经验分布函数并作图.

解 （1）经验分布函数为

$$F_n(x)=\begin{cases}0, x<138\\0.1, 138\le x<149\\0.3, 149\le x<153\\0.5, 153\le x<156\\0.8, 156\le x<160\\0.9, 160\le x<169\\1, x\ge 169\end{cases}$$

（2）经验分布函数图像（略）.

2．为研究某厂工人生产某种产品的能力，我们随机调查了 20 位工人某天生产的该种产品的数量，数据如下

160　196　164　148　170　175　178　166　181　162

161　168　166　162　172　156　170　157　162　154

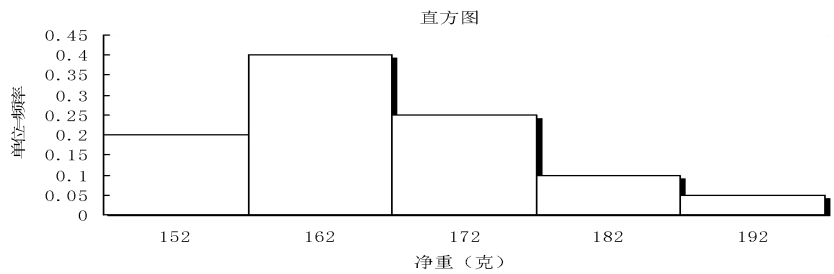
（1）构造该批数据的频率分布表（分成 5 组）

（2）画出直方图.

解（1）数据中的最小值是 148，最大值是 196．这 20 个数据就散布在闭区间[148，196]中．取一个略大一点的区间（147,197],它的端点都是整数．我们将（147，197]五等分，排在下表的第一列．计算数据落入各段的个数 n_i，填入第二列．计算出数据落入各段的频率 $f_i=\dfrac{n_i}{n}$，依次填入第三列．最后将各列之和填入最后一行，得到如下的频率分布表.

零件重量	发生次数 n_i	发生频率 $f_i=\dfrac{n_i}{n}$
（147,157]	4	20%
（157,167]	8	40%
（167,177]	5	25%
（177,187]	2	10%
（187,197]	1	5%
总数	20	100%

（2）直方图如下



3．设 $X_1,\cdots,X_n$ 为来自自均匀分布 U(-1，1)的样本，试求 $E(\bar{X})$和$D(\bar{X})$.

解 因为 $E(X_i)=\frac{-1+1}{2}=0, D(X_i)=\frac{[1-(-1)]^2}{12}=\frac{1}{3}$, 所以

$$E(\bar{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}0=0\ ;$$

$$D(\bar{X})=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\left(\frac{1}{n}\right)^2\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{3}=\frac{1}{3n}.$$

4．设 $X_1,\cdots,X_n$ 是来自 $\chi^2(m)$ 分布的样本.求样本均值 $\bar{X}$ 的期望与方差.

解 因为 $E(X_i)=m, D(X_i)=2m$, 所以

$$E(\bar{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}m=m\ ;$$

$$D(\bar{X})=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\left(\frac{1}{n}\right)^2\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}2m=\frac{2m}{n}.$$

5．在总体 N(7.6，4)中抽取容量为 n 的样本，如果要求样本均值落在（5.6，9.6）内的概率不小于 0.95，则 n 至少为多少？

解 由总体 X~N(7.6，4)可知，$\bar{X}\sim N(7.6,\frac{4}{n})$，要 $P\{5.6<\bar{X}<9.6\}\geq 0.95$，即要

$$\Phi\left(\frac{\sqrt{n}(9.6-7.6)}{2}\right)-\Phi\left(\frac{\sqrt{n}(5.6-7.6)}{2}\right)=\Phi(\sqrt{n})-\Phi(-\sqrt{n})=2\Phi(\sqrt{n})-1\geq 0.95$$

从而要 $\Phi(\sqrt{n})\geq 0.975=\Phi(1.96)\Rightarrow n\geq 1.96^2=3.8416$，所以 n 至少为 4.

6. 设 $x_1,\ldots,x_{16}$ 是来自 $N(\mu,\sigma^2)$ 的样本，经计算 $\bar{x}=9, s^2=5.32$, 试求 $P\{|\bar{x}-\mu|<0.6\}$.

（$tcdf(1.04,15)=0.8426$, 其中 $tcdf(x,n)$ 表示自由度为 n 的 t 分布的分布函数在 x 处的分布函数值）

解 由正态总体抽样分布理论可知 $\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t(n-1)$，现 n=16, $s^2=5.32$,所以

$$P\{|\bar{x}-\mu|<0.6\}=P\{\frac{\sqrt{n}\,|\bar{x}-\mu|}{S}<\sqrt{\frac{16}{5.32}}\times 0.6\}=P\{\frac{\sqrt{n}\,|\bar{x}-\mu|}{S}<1.04\}$$

$$= tcdf(1.04,15) = 0.8426 .$$

7. 设 $X \sim N(\mu,\sigma^2)$, μ 未知,且 σ^2 已知, $X_1,\cdots,X_n$ 为取自此总体的一个样本, 指出下列各式中哪些是统计量, 哪些不是, 为什么?

(1) $X_1 + X_2 + X_n - \mu$　(2) $X_n - X_{n-1}$　(3) $\dfrac{\bar{X}-\mu}{\sigma}$　(4) $\displaystyle\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$

解　因为 $X_1 + X_2 + X_n - \mu$ ，$\dfrac{\bar{X}-\mu}{\sigma}$，$\displaystyle\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$ 都是样本 $X_1,\cdots,X_n$ 和 μ 的函数，而是未知参数，所以它们都不是统计量，但 $X_n - X_{n-1}$ 是样本 $X_1,\cdots,X_n$ 的函数且不含未知参数，故 $X_n - X_{n-1}$ 是统计量.

8．设总体 $X\sim N(10,9)$, $X_1,\cdots,X_6$ 是它的一个样本, $Z=\displaystyle\sum_{i=1}^{6}X_i$,(1)写出 Z 的概率密度; (2) 求 P(Z>41).

解　（1）由正态随机变量的线性组合仍然服从正态分布可知，$Z=\displaystyle\sum_{i=1}^{6}X_i \sim N(60,54)$，所以其概率密度函数为　$f(x)=\dfrac{1}{6\sqrt{3\pi}}e^{-\frac{(x-60)^2}{108}}, -\infty < x < +\infty$；

（2）$P\{Z>41\}=1-P\{Z\le 41\}=1-\Phi\left(\dfrac{41-60}{\sqrt{54}}\right)=1-\Phi(-2.59)=\Phi(2.59)=0.9952$.

9．设从总体 $X\sim N(\mu,\sigma^2)$ 中抽取容量为 18 的样本，μ, σ²未知 ，

(1)求 $P(S^2/\sigma^2 \leqslant 1.2052)$,其中 $S^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\bar{X})^2$；(2) 求 $D(S^2)$.

（$chi2cdf(20.4884,17)$=0.7500 其中 $chi2cdf(x,n)$ 表示自由度为 n 的 χ^2 分布的分布函数在 x 处的分布函数值）

解　（1）由正态总体抽样分布知识可得 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，现 n=18,故

$$P\{\frac{S^2}{\sigma^2}\le 1.2052\}=P\{\frac{(n-1)S^2}{\sigma^2}\le 17\times 1.2052\}=P\{\frac{(n-1)S^2}{\sigma^2}\le 20.4884\}$$

$$= chi2cdf(20.4884,17)=0.7500；$$

（2）因为 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，故

$$\frac{(n-1)^2}{\sigma^4}D(S^2)=D\left(\frac{(n-1)S^2}{\sigma^2}\right)=2(n-1)\Rightarrow D(S^2)=\frac{2\sigma^4}{17}(n=18)$$

10. $X_1,X_2,\cdots,X_{10}$ 为总体 $N(0,0.3^2)$ 的一个样本, 求:

(1) $E(\sum_{i=1}^{10}X_i^2),D(\sum_{i=1}^{10}X_i^2)$；　(2) $P\{\sum_{i=1}^{10}X_i^2>1.44\}$

解　(1) 由题设条件可知 $\frac{X_1}{0.3},\frac{X_2}{0.3},\cdots,\frac{X_{10}}{0.3}$ 相互独立，且都服从标准正态分布，由 χ^2 分布的结构性定义可知，$\frac{\sum_{i=1}^{10}X_i^2}{0.3^2}\sim\chi^2(10)$，从而 $E\left(\frac{\sum_{i=1}^{10}X_i^2}{0.3^2}\right)=10,D\left(\frac{\sum_{i=1}^{10}X_i^2}{0.3^2}\right)=20$，再由数学期望与方差的性质可得 $E(\sum_{i=1}^{10}X_i^2)=10\times0.3^2=0.9,D(\sum_{i=1}^{10}X_i^2)=20\times0.3^4=0.162$；

(2) 由(1)知 $P\{\sum_{i=1}^{10}X_i^2>1.44\}=P\{\frac{\sum_{i=1}^{10}X_i^2}{0.3^2}>16\}=1-chi2cdf(16,10)=0.10$.

习题 B

1. 设总体密度函数为

$$f(x,\theta)=\begin{cases}3x^2, & 0<x<1\\ 0, & 其它\end{cases}$$

现从该总体抽得一个容量为 5 的样本, 试计算 $P\{X_{(1)}\le 1/2\}$.

解 $P\{X_{(1)}\le1/2\}=1-P\{X_{(1)}>1/2\}=1-P\{X_1>1/2,X_2>1/2,\cdots X_5>1/2\}$

$$=1-P\{X_1>1/2\}P\{X_2>1/2\}\cdots P\{X_5>1/2\}=1-[P\{X_1>1/2\}]^5$$

$$=1-\left(\int_{1/2}^1 3x^2\mathrm{d}x\right)^5=1-\left(1-\frac{1}{8}\right)^5=1-\left(\frac{7}{8}\right)^5.$$

2. （1）设 $X_1,X_2,\cdots,X_6$ 来自总体 $N(0,1),Y=\left(X_1+X_3+X_5\right)^2+\left(X_2+X_4+X_6\right)^2$,

试确定常数 C 使 CY 服从 χ^2 分布.

（2）设 $X_1,X_2,\cdots,X_5$ 来自总体 $N(0,1),Y=\dfrac{(X_1+X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$,试确定常数 C 使 CY 服从 t 分布.

(3) 已知 $X\sim t(n)$,证明 $X^2\sim F(1,n)$.

解 （1）由题设条件可知 $\dfrac{X_1+X_3+X_5}{\sqrt{3}},\dfrac{X_2+X_4+X_6}{\sqrt{3}}$ 相互独立且都服从标准正态分布，故由卡方分布的结构性定义可知 $\dfrac{1}{3}Y=\left(\dfrac{X_1+X_3+X_5}{\sqrt{3}}\right)^2+\left(\dfrac{X_2+X_4+X_6}{\sqrt{3}}\right)^2\sim\chi^2(2)$，所以，当 C=1/3 时 CY 服从 χ^2 分布；

（2）由题设条件可知 $\dfrac{X_1+X_2}{\sqrt{2}}\sim N(0,1),X_3^2+X_4^2+X_5^2\sim\chi^2(3)$，且两者相互独立，由 t 分布的结构性定义可知 $\dfrac{\sqrt{6}}{2}Y=\dfrac{(X_1+X_2)/\sqrt{2}}{\sqrt{(X_3^2+X_4^2+X_5^2)/3}}\sim t(3)$，所以，当 $C=\dfrac{\sqrt{6}}{2}$ 时，CY 服从 t 分布；

（3）证明 由 $X\sim t(n)$知，存在 $U\sim N(0,1),V\sim\chi^2(n)$，且 U 与 V 相互独立，使得

$X=\dfrac{U}{\sqrt{V/n}}\sim t(n)$，因此 $U^2\sim\chi^2(1)$ 且与 V 独立，由 F-分布的结构性定义可知

$X^2=\dfrac{U^2/1}{V/n}\sim F(1,n)$.

3. 设 $\overline{X}_1$ 和 $\overline{X}_2$ 是来自正态总体 $N(\mu,\sigma^2)$ 容量为 n 的两独立样本 $X_{11},X_{12},\cdots,X_{1n}$ 和 $X_{21},X_{22},\cdots,X_{2n}$ 的样本均值, 试确定 n, 使得这两个样本均值之差超过 σ 的概率大约为 0.01.

解 由正态总体的抽样分布可知 $\overline{X}_1\sim N(\mu,\dfrac{\sigma^2}{n}),\overline{X}_2\sim N(\mu,\dfrac{\sigma^2}{n})$，且相互独立，从而有 $\overline{X}_1-\overline{X}_2\sim N(0,\dfrac{2\sigma^2}{n})$，现要确定 n, 使得 $P\{|\overline{X}_1-\overline{X}_2|>\sigma\}\approx 0.01$，即

$$P\{\frac{\sqrt{n}\,|\overline{X}_1-\overline{X}_2|}{\sqrt{2}\sigma}>\sqrt{\frac{n}{2}}\}=2[1-\Phi(\sqrt{\frac{n}{2}})]\approx 0.01$$

$$\Phi(\sqrt{\frac{n}{2}}) \approx 0.995 = \Phi(2.575) \Rightarrow n = 2\times 2.575^2 = 13.26$$

因此 $n=14$.

4. 设总体 $X \sim N(\mu,\sigma^2)$，从该总体中取一个容量为 $n=16$ 的样本 $(X_1, X_2, \cdots, X_{16})$，求：(1) $P\left\{\frac{\sigma^2}{2} \le \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \le 2\sigma^2\right\}$；(2) $P\left\{\frac{\sigma^2}{2} \le \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \le 2\sigma^2\right\}$.

($chi2cdf(32,16)=0.9900, chi2cdf(8,16)=0.0511$，$chi2cdf(32,15)=0.9936, chi2cdf(8,15)=0.0762$)

解 由正态总体的抽样分布可知

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n), \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1)$$

现样本容量为 $n=16$.

$$(1) P\left\{\frac{\sigma^2}{2} \le \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \le 2\sigma^2\right\} = P\left\{\frac{n}{2} \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \le 2n\right\} = P\left\{8 \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \le 32\right\}$$

$$= chi2cdf(32,16) - chi2cdf(8,16) = 0.9900 - 0.0511 = 0.9389;$$

$$(2) P\left\{\frac{\sigma^2}{2} \le \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \le 2\sigma^2\right\} = P\left\{\frac{n}{2} \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \le 2n\right\} = P\left\{8 \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \le 32\right\}$$

$$= chi2cdf(32,15) - chi2cdf(8,15) = 0.9936 - 0.07562 = 0.9174.$$

5. 从正态总体 $N(0, 0.5^2)$ 中抽取容量为 10 的样本 $X_1, X_2, \cdots, X_{10}$，求：(1) $P\{\sum_{i=1}^{10} X_i^2 > 4\}$；

(2) $P\{\sum_{i=1}^{10}(X_i-\bar{X})^2 > 2.85\}$. $(chi2cdf(16,10)=0.9004, chi2cdf(11.4,9)=0.7507)$

解（1）由正态总体的抽样分布可知 $4\sum_{i=1}^{10} X_i^2 \sim \chi^2(10)$，因此

$$P\{\sum_{i=1}^{10} X_i^2 > 4\} = P\{4\sum_{i=1}^{10} X_i^2 > 16\} = 1 - P\{4\sum_{i=1}^{10} X_i^2 \le 16\}$$

$$= 1 - chi2cdf(16,10) = 1 - 0.9004 = 0.0996;$$

(2) 由正态总体的抽样分布可知 $4\sum_{i=1}^{10}(X_i-\bar{X})^2 \sim \chi^2(9)$，因此

$$P\{\sum_{i=1}^{10}(X_i-\bar{X})^2 > 2.85\} = P\{4\sum_{i=1}^{10}(X_i-\bar{X})^2 > 11.4\} = 1 - P\{4\sum_{i=1}^{10} X_i^2 \le 11.4\}$$

$$= 1 - chi2cdf(11.4,9) = 1 - 0.7507 = 0.2493.$$

6. 设 $X_1,X_2,\cdots,X_9$ 是来自正态总体 X 的简单随机样本

$$Y_1=\frac{1}{6}(X_1+X_2+\cdots+X_6),\ Y_2=\frac{1}{3}(X_7+X_8+X_9),\mathrm{S}^2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2$$

证明：统计量 $Z=\dfrac{\sqrt{2}(Y_1-Y_2)}{S}$ 服从自由度为 2 的 t 分布.

证明 不妨设总体 $X\sim N(\mu,\sigma^2)$，由正态总体的抽样分布可知

$$Y_1=\frac{1}{6}(X_1+X_2+\ldots+X_6)\sim N(\mu,\frac{\sigma^2}{6}),Y_2=\frac{1}{3}(X_7+X_8+X_9)\sim N(\mu,\frac{\sigma^2}{3})\text{，}$$

$\dfrac{2S^2}{\sigma^2}=\dfrac{\sum\limits_{i=7}^{9}(X_i-Y_2)^2}{\sigma^2}\sim\chi^2(2)$, 且 Y_1,Y_2,S^2 相互独立，因此 $Y_1-Y_2\sim N(0,\frac{1}{2}\sigma^2)$，而

$\dfrac{\sqrt{2}}{\sigma}(Y_1-Y_2)\sim N(0,1)$ 与 $\dfrac{2S^2}{\sigma^2}$ 独立，由 t 分布的定义可知

$$Z=\frac{\sqrt{2}(Y_1-Y_2)}{S}=\frac{\sqrt{2}(Y_1-Y_2)/\sigma}{\sqrt{\left(\dfrac{2S^2}{\sigma^2}\right)\Big/2}}\sim t(n)\ .$$

7. 设总体体 $X\sim N(\mu,\sigma^2)$，从该总体中抽取样本 $X_1,X_2,\cdots,X_n,X_{n+1}$，记 $\bar{X}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$，$S_n^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X}_n)^2,S_n=\sqrt{S_n^2}.$

证明：$\sqrt{\dfrac{n}{n+1}}\cdot\dfrac{X_{n+1}-\overline{X_n}}{S_n}\sim t(n-1).$

证明 由正态总体的抽样分布可知 $\bar{X}_n\sim N(\mu,\dfrac{\sigma^2}{n}),\dfrac{(n-1)S_n^2}{\sigma^2}\sim\chi^2(n-1)$ 且相互独立，由题设条件还可知，它们还与 X_{n+1} 独立，由此可得 $X_{n+1}-\bar{X}_n\sim N(0,\dfrac{n+1}{n}\sigma^2)$，故

$\sqrt{\dfrac{n}{n+1}}\cdot\dfrac{X_{n+1}-\bar{X}_n}{\sigma}\sim N(0,1)$ 与 $\dfrac{(n-1)S_n^2}{\sigma^2}\sim\chi^2(n-1)$ 独立，由 t 分布的定义可知

$$\sqrt{\frac{n}{n+1}}\cdot\frac{X_{n+1}-\bar{X}_n}{S_n}=\frac{\sqrt{\dfrac{n}{n+1}}\cdot\dfrac{X_{n+1}-\bar{X}_n}{\sigma}}{\sqrt{\dfrac{(n-1)S_n^2}{\sigma^2}\Big/(n-1)}}\sim t(n-1)\ .$$