5.3 统计量及其分布

1. 在一本书上我们随机地检查了 10 页，发现每页上的错误数为：

$$4 \quad 5 \quad 6 \quad 0 \quad 3 \quad 1 \quad 4 \quad 2 \quad 1 \quad 4$$

试计算其样本均值、样本方差和样本标准差。

解：

$$\bar{x} = \frac{4+5+6+0+3+1+4+2+1+4}{10} = 3$$

$$s^2 = \frac{1}{10-1}\sum_{i=1}^{10}(x_i-\bar{x})^2 = \frac{34}{9} \approx 3.778$$

$$s = \sqrt{\frac{34}{9}} \approx 1.944$$

□

2. 证明：对任意常数 c, d，有

$$\sum_{i=1}^{n}(x_i-c)(y_i-d) = \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) + n(\bar{x}-c)(\bar{y}-d)$$

证明：

根据性质 $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \bar{x}$, $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \bar{y}$ 可得

$$\begin{aligned}
\text{右边} &= \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) + \sum_{i=1}^{n}(\bar{x}-c)(\bar{y}-d) \\
&= \sum_{i=1}^{n}\left[(x_i-\bar{x})(y_i-\bar{y}) + (\bar{x}-c)(\bar{y}-d)\right] \\
&= \sum_{i=1}^{n}(x_iy_i - \bar{x}y_i - \bar{y}x_i + \bar{x}\bar{y} + \bar{x}\bar{y} - \bar{x}d - \bar{y}c + cd) \\
&= \sum_{i=1}^{n}x_iy_i - \bar{x}\sum_{i=1}^{n}y_i - \bar{y}\sum_{i=1}^{n}x_i + n\bar{x}\bar{y} + n\bar{x}\bar{y} - n\bar{x}d - n\bar{y}c + ncd \\
&= \sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y} - n\bar{y}\bar{x} + n\bar{x}\bar{y} + n\bar{x}\bar{y} - \sum_{i=1}^{n}x_id - \sum_{i=1}^{n}y_ic + \sum_{i=1}^{n}cd \\
&= \sum_{i=1}^{n}(x_iy_i - x_id - y_ic + cd) \\
&= \sum_{i=1}^{n}(x_i-c)(y_i-d) = \text{左边}
\end{aligned}$$

□

3. 设 $x_1, x_2, \cdots, x_n$ 和 $y_1, y_2, \cdots, y_n$ 是两组样本观测值，且有如下关系：

$$y_i = 3x_i - 4, i = 1, 2, \cdots, n$$

试求样本均值 $\bar{x}$ 和 $\bar{y}$ 间的关系以及样本方差 s_x^2 和 s_y^2 间的关系。

解：

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(3x_i - 4) = 3 \cdot \frac{1}{n}\sum_{i=1}^{n} x_i - 4 = 3\bar{x} - 4$$

$$\begin{aligned} s_y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(3x_i - 4 - (3\bar{x} - 4))^2 = \frac{1}{n-1}\sum_{i=1}^{n}[3(x_i - \bar{x})]^2 \\ &= 9 \cdot \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9s_x^2 \end{aligned}$$

□

5. 从同一总体中抽取两个容量分别为 n, m 的样本，样本均值分别为 $\bar{x}_1, \bar{x}_2$，样本方差分别为 s_1^2, s_2^2，将两组样本合并，其均值、方差分别为 $\bar{x}, s^2$，证明：

$$\bar{x} = \frac{n\bar{x}_1 + m\bar{x}_2}{n+m}$$

$$s^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-1} + \frac{nm(\bar{x}_1 - \bar{x}_2)^2}{(n+m)(n+m+1)}$$

证明：

$$\bar{x} = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_{1_i} + \sum_{i=1}^{m} x_{2_i}\right) = \frac{n\bar{x}_1 + m\bar{x}_2}{n+m}$$

$$\begin{aligned} s^2 &= \frac{1}{n+m-1}\left(\sum_{i=1}^{n}(x_{1i} - \bar{x})^2 + \sum_{j=1}^{m}(x_{2j} - \bar{x})^2\right) \\ &= \frac{1}{n+m-1}\left(\sum_{i=1}^{n}\left(x_{1i} - \frac{n\bar{x}_1 + m\bar{x}_2}{n+m}\right)^2 + \sum_{i=1}^{n}\left(x_{2j} - \frac{n\bar{x}_1 + m\bar{x}_2}{n+m}\right)^2\right) \\ &= \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-1} + \frac{nm(\bar{x}_1 - \bar{x}_2)^2}{(n+m)(n+m+1)} \end{aligned}$$

□

8. 设 $x_1, x_2, \cdots, x_n$ 是来自 $U(-1, 1)$ 的样本，试求 $E(\bar{x})$ 和 $\mathrm{Var}(\bar{x})$。

解：

设随机变量 $X \sim U(-1,1)$，则

$$E(\bar{x}) = EX = \frac{-1+1}{2} = 0$$

$$\mathrm{Var}(\bar{x}) = \frac{\mathrm{Var}(X)}{n} = \frac{\frac{(-1-1)^2}{12}}{n} = \frac{1}{3n}$$

□

9. 设总体二阶矩存在，$x_1, x_2, \cdots, x_n$ 是样本，证明 $x_i - \bar{x}$ 与 $x_j - \bar{x}$ $(i \neq j)$ 的相关系数为 $-(n-1)^{-1}$。

证明：

根据样本均值的性质，$E(x_i - \bar{x}) = E(x_j - \bar{x}) = 0$。

设随机变量 X 表示从总体中抽出的一个样本，则 EX^2 存在。

$$E(x_i - \bar{x})(x_j - \bar{x}) = E(x_i x_j - \bar{x} x_i - \bar{x} x_j + \bar{x}^2) = Ex_i x_j - \bar{x} Ex_i - \bar{x} Ex_j + \bar{x}^2$$

将 x_i 与 x_j 看作独立的两次抽样，则 $x_i, x_j \overset{\text{i.i.d.}}{\sim} X$，所以 $Ex_i x_j = Ex_i Ex_j = (EX)^2$，$Ex_i = EX, Ex_j = EX$。所以

$$E(x_i - \bar{x})(x_j - \bar{x}) = (EX)^2 - 2\bar{x}EX + \bar{x}^2 = \frac{1}{1-n} = -(n-1)^{-1}$$

□

10. 设 $x_1, x_2, \cdots, x_n$ 为一个样本，$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 是样本方差，试证：

$$\frac{1}{n(n-1)}\sum_{i<j}(x_i - x_j)^2 = s^2$$

证明：

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i<j}(x_i - x_j)^2 = \frac{1}{n(n-1)}\sum_{i<j}(x_i - \bar{x} + \bar{x} - x_j)^2 \\
&= \frac{1}{n(n-1)}\sum_{i<j}[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - x_j) + (\bar{x} - x_j)^2] \\
&= \frac{1}{n(n-1)} \cdot \frac{1}{2} \sum_{i=1,2,\cdots,n;j=1,2,\cdots,n} [(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - x_j) + (\bar{x} - x_j)^2] \\
&= \frac{1}{2n(n-1)}\left[n\sum_{i=1}^{n}(x_i - \bar{x})^2 + 0 + n\sum_{j=1}^{n}(x_j - \bar{x})^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2
\end{aligned}$$

□

11. 设总体 4 阶中心距 $\nu_4 = E[x - E(x)]^4$ 存在，试证：对样本方差 $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$，有

$$\operatorname{Var}(s^2) = \frac{n(\nu - \sigma^4)}{(n-1)^2} - \frac{2(\nu_4 - 2\sigma^4)}{(n-1)^2} + \frac{\nu - 3\sigma^4}{n(n-1)^2}$$

其中 σ^2 为总体 X 的方差。

证明：

$$\begin{aligned}
\text{右边} &= \frac{n^2\nu_4 - n^2\sigma^4 - 2n\nu_4 + 4n\sigma^4 + \nu_4 - 3\sigma^4}{n(n-1)^2} \\
&= \frac{\nu_4(n^2 - 2n + 1) - \sigma^4(n^2 - 4n + 3)}{n(n-1)^2} \\
&= \frac{\nu_4(n-1)^2 - \sigma^4(n-1)(n-3)}{n(n-1)^2} \\
&= \frac{\nu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)} \\
\text{左边} &= E(s^2)^2 - (Es^2)^2 = Es^4 - (Es^2)^2 = Es^4 - \sigma^4
\end{aligned}$$

实在证明不出来了。 □

12. 设总体 X 的 3 阶矩存在，若 $x_1, x_2, \cdots, x_n$ 是取自该总体的简单随机样本，$\bar{x}$ 为样本均值，s^2 为样本方差，试证：$\operatorname{Cov}(\bar{x}, s^2) = \dfrac{\nu_3}{n}$，其中 $\nu_3 = E[x - E(x)]^3$。

证明：

$$E\bar{x} = EX, Es^2 = \operatorname{Var} X$$

$$E(\bar{x}s^2) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

$$\operatorname{Var}\bar{x} = \frac{\operatorname{Var} X}{n}, \operatorname{Var} s^2 = \operatorname{Var}\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

也证明不出来了。 □

15. 从指数总体 $\operatorname{Exp}(\frac{1}{\theta})$ 抽取了 40 个样品，试求 $\bar{x}$ 的渐近分布。

解：

设随机变量 X 表示从总体中抽出的一个样本，则

$$EX = \frac{1}{\frac{1}{\theta}} = \theta,\ \operatorname{Var} X = \frac{1}{\left(\frac{1}{\theta}\right)^2} = \theta^2$$

所以 $\bar{x}$ 的渐近分布为 $N(\theta, \theta^2)$。 □

17. 设 $x_1, x_2, \cdots x_{20}$ 是从二点分布 $b(1,p)$ 抽取的样本，试求样本均值 $\bar{x}$ 的渐近分布。

解：

设随机变量 X 表示从总体中抽出的一个样本，则

$$EX = p,\ \operatorname{Var} X = p(1-p)$$

所以 $\bar{x}$ 的渐近分布为 $N(p, p(1-p))$。 □

23. 设总体 X 服从几何分布，即 $P(X=k) = pq^{k-1}, k = 1, 2, \cdots$，其中 $0 < p < 1, q = 1-p$，$x_1, x_2, \cdots, x_n$ 为该总体的样本，求 $x_{(n)}, x_{(1)}$ 的概率分布。

解：

设总体 X 的概率密度函数为 $p(x)$，分布函数为 $F(x)$，则

$$F(x) = \sum_{k=1}^{\lfloor x \rfloor} pq^{k-1} = \frac{p - pq^{\lfloor x \rfloor}}{1-q}$$

$$p_{x_{(n)}}(x) = \frac{n!}{(n-1)!}[F(x)]^{n-1} p(x) = npq^{\lfloor x \rfloor - 1}\left[\frac{p - pq^{\lfloor x \rfloor}}{1-q}\right]^{n-1}$$

$$p_{x_{(1)}} = \frac{n!}{(n-1)!}[1-F(x)]^{n-1} p(x) = npq^{\lfloor x \rfloor - 1}\left[1 - \frac{p - pq^{\lfloor x \rfloor}}{1-q}\right]^{n-1}$$

□

28. 设总体 X 的分布函数 $F(x)$ 是连续的，$x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 为取自此总体的次序统计量，设 $\eta_i = F(x_{(i)})$，试证：

(1) $\eta_1 \leqslant \eta_2 \leqslant \cdots \leqslant \eta_n$，且 η_i 是来自均匀分布 $U(0,1)$ 总体的次序统计量。

证明：

因为 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$ 且 $F(x)$ 单调，$\eta_i = F(x_{(i)})$，所以 $\eta_1 \leqslant \eta_2 \leqslant \cdots \leqslant \eta_n$。 □

(2) $E(\eta_i) = \dfrac{i}{n+1},\ \operatorname{Var}(\eta_i) = \dfrac{i(n+1-i)}{(n+1)^2(n+2)}, 1 \leqslant i \leqslant n$

证明：

设总体的概率密度函数为 $p(x)$，则 η_i 的分布函数为

$$p_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!}[F(x)]^{i-1}[1-F(x)]^{n-i} p(x)$$

$$\begin{aligned}E(\eta_i) &= \sum_{i=1}^{n} F(x_{(i)}) p_{(i)}(x_{(i)}) \\ &= \sum_{i=1}^{n} F(x_{(i)}) \frac{n!}{(i-1)!(n-i)!} [F(x_{(i)})]^{i-1} [1-F(x_{(i)})]^{n-i} p(x) \\ &= \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} [F(x_{(i)})]^{i} [1-F(x_{(i)})]^{n-i} p(x)\end{aligned}$$

$$\mathrm{Var}(\eta_i) =$$

实在是不会了。 □

(3) η_i 和 η_j 的协方差矩阵为 $\begin{bmatrix} \frac{a_1(1-a_1)}{n+2} & \frac{a_1(1-a_2)}{n+2} \\ \frac{a_1(1-a_2)}{n+2} & \frac{a_2(1-a_2)}{n+2} \end{bmatrix}$，其中 $a_1 = \dfrac{i}{n+1}, a_2 = \dfrac{j}{n+1}$。

证明：

$$E(\eta_i) = \frac{i}{n+1},\ E(\eta_j) = \frac{j}{n+1},\ E(\eta_i \eta_j) =$$

实在是不会了。 □

32. 设总体 X 的密度函数为 $p(x) = \begin{cases} 3x^2, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases}$，$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(5)}$ 为容量为 5 的取自此总体的次序统计量，试证 $\dfrac{x_{(2)}}{x_{(4)}}$ 与 $x_{(4)}$ 相互独立。

证明：

根据相互独立的定理，需要证明 $\forall x, y,\ p_{\frac{x_{(2)}}{x_{(4)}}}(x) \cdot p_{x_{(4)}}(y) = p_{\frac{x_{(2)}}{x_{(4)}}, x_{(4)}}(x, y)$，之后就不会了。 □

35. 对下列数据构造箱线图：

472	425	447	377	341	369	412	419	400	382	366	425	399	398	423
384	418	392	372	418	374	385	439	428	429	428	430	413	405	381
403	479	381	443	441	433	419	379	386	387