



数据科学基础

Foundations of Data Science

假设检验练习

陈振宇

南京大学智能软件工程实验室

www.iselab.cn

假设检验和区间估计

例 7.13: 南京某区 16 座预售楼盘均价如下表 (单位: 元/平方米)

7800	7400	7300	6700	7000	7800	8200	7400
8300	6800	7700	7700	7400	12900	7500	7700

试推断南京市某区楼盘价格与媒体公布的 7900 元/平方米是否相符 (显著性水平为 $\alpha = 0.05$)?

请分别采用区间估计, Z检验, t检验, 符号检验和秩和检验对本题做一下对比分析

区间估计

考虑到 σ^2 未知，且有 $\bar{X} = 7850, S^2 = 2005333, S = 1416$

于是有 $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$

即有 $P\{-t_{\frac{\alpha}{2}}(n-1) < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}}(n-1)\} = 1 - \alpha$

则 μ 的一个置信水平为 $1 - \alpha$ 的置信区间为 $(\bar{X} \pm \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1))$

代入数据计算可得置信区间为 $(7070.7, 8629.3)$

又 7900 位于置信区间内。则推断符合。

Z检验

设 $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

拒绝域为 $|Z| \geq Z_{\frac{\alpha}{2}} = 1.96$

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} = -0.14124$$

$|Z| < 1.96$ 未落入拒绝域

故判断符合

t检验

设 $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

拒绝域为 $|t| \geq t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(15) = 2.1315$

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = -0.14124$$

$|t| < 2.1315$ 未落入拒绝域

故判断符合



符号检验

$$H_0 : p = 0.5 \quad H_1 : p \neq 0.5$$

$$B(3; 16, 0.5) = \sum_{i=0}^3 C_{16}^i (0.5)^{16} = 0.0213 < 0.025$$

落入拒绝域，故拒绝原假设。即认为判断不符合。

秩和检验

将原整体视为两个整体，上方8个样本为一组，下方8个样本为一组。

将7900重复8个为一组作为对照组，将以上两组分别与该组进行秩排序

第一组：7800 7400 7300 6700 7000 7800 8200 7400

第二组：8300 6800 7700 7700 7400 12900 7500 7700

其中第一组与对照组排序后对照组统计量 $T = 92$

根据 $\alpha = 0.05, n_1 = 8, n_2 = 8$, 查表的 $T_1 = 49, T_2 = 87, T = 92 > 87$

故拒绝假设 H_0 ，即判断不符合。

卡方拟合优度检验

例 7.19: 将 84 个 Etruscan 人的男子头颅的最大宽度 (毫米) 如下。

141 148 132 138 154 142 150 146 155 158 150 140 147 148 144 150 149 145 149 158 140
143 141 144 144 126 140 144 142 141 140 145 135 147 146 141 136 140 146 142 137 137
148 154 137 139 143 140 131 143 141 149 148 135 148 152 143 144 141 143 147 146 152
150 132 142 142 143 153 149 146 149 138 142 149 142 137 134 144 146 147 140 142 145

现问在显著性水平 $\alpha = 0.1$ 下检验假设

$$H_0: \text{总体服从正态分布 } f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$

请采用3组不同的分组方式对本题进行卡方拟合优度检验对比分析



卡方检验

由极大似然估计可得 $\hat{\mu} = 143.76$ $\hat{\sigma} = 5.96$

由正态分布概率公式得 $\hat{p}_i = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

1. 将 $x < 135, 135 \leq x < 138, 138 \leq x < 142, 142 \leq x < 146, 146 \leq x < 150, 150 \leq x < 154, x \geq 154$, 进行分组

第一组 5个 $\hat{p}_1 = \Phi\left(\frac{135 - 143.76}{5.96}\right) = 1 - 0.9292 = 0.0708$

则 $\frac{f_1^2}{n\hat{p}_1} = 4.2037$, 以此类推

计算得 $X^2 = 85.15 - 84 = 1.15 < X_{0.1}^2(7 - 2 - 1) = 7.779$, 故符合

2. 将 $x < 136, 136 \leq x < 139, 139 \leq x < 143, 143 \leq x < 147, 147 \leq x < 151, 151 \leq x$, 进行分组

计算得 $X^2 = 86.42 - 84 = 2.42 < X_{0.1}^2(3) = 6.251$, 故符合

3. 将 $x < 136, 136 \leq x < 139, 139 \leq x < 142, 142 \leq x < 147, 147 \leq x < 151, 151 \leq x$, 进行分组

计算得 $X^2 = 85.38 - 84 = 1.3 < X_{0.1}^2(3) = 6.251$, 故符合

