回归分析-第四次上机

南开大学统计与数据科学学院，马东升

2024 年 11 月 19 日

1 习题 6.4

1.1 题目

用工程 A,B,C 生产同一型号的电池，为比较其质量，从各厂的产品中随机抽取 6 只电池，经测试寿命（h）如下表 1。

表 1: 电池寿命表

生产厂	电池寿命					
A	40	48	38	42	45	46
B	26	34	30	28	32	33
C	39	40	48	50	50	52

（1）将数据表成方差分析模型 (6.1.2) 的形式；

注，6.1.2 式为：$\begin{cases} y_{ij} & = \mu_i + e_{ij} \\ e_{ij} & \overset{IID}{\sim} N(0,\sigma^2) \end{cases}, i = 1, ..., a; j = 1, ..., n_i$

（2）在显著性水平 $\alpha = 0.05$ 下检验三厂生产的电池的平均寿命有无显著差异？列出方差分析表；

（3）记 μ_A, μ_B, μ_C 分别为三厂生产的电池的平均寿命，写出均值之差 $\mu_A - \mu_B, \mu_A - \mu_C, \mu_B - \mu_C$ 的 95% 同时置信区间。

1.2 源代码（R）

```r
LevelA <- c(40,48,38,42,45,46)
LevelB <- c(26,34,30,28,32,33)
LevelC <- c(39,40,48,50,50,52)

# 组合数据
data <- data.frame(
  Value = c(LevelA, LevelB, LevelC),
  Group = factor(rep(c("LevelA", "LevelB", "LevelC"), each = 6))
)
data

# 单因素方差分析
anova_result <- aov(Value ~ Group, data = data)

# 显示结果
summary(anova_result)
```

```

# 定义函数计算均值差的置信区间
mean_diff_ci <- function(A, B, anova_result, alpha = 0.05) {
  # 计算均值
  mean1 <- mean(A)
  mean2 <- mean(B)

  # 提取 MSE（组内均方误差）
  mse <- summary(anova_result)[[1]]["Residuals", "Mean Sq"]

  # 样本大小
  n1 <- length(A)
  n2 <- length(B)

  # 计算标准误差
  se_diff <- sqrt(mse * (1/n1 + 1/n2))

  # 自由度
  df <- summary(anova_result)[[1]]["Residuals", "Df"]

  # t 分布的临界值
  t_crit <- qt(1 - alpha / 6, df)

  # 计算均值差和置信区间
  mean_diff <- mean1 - mean2
  lower_ci <- mean_diff - t_crit * se_diff
  upper_ci <- mean_diff + t_crit * se_diff

  # 返回结果
  return(data.frame(
    Mean_Difference = mean_diff,
    Lower_CI = lower_ci,
    Upper_CI = upper_ci
  ))
}

result <- mean_diff_ci(LevelA, LevelB, anova_result)
print(result)

result <- mean_diff_ci(LevelA, LevelC, anova_result)
print(result)

result <- mean_diff_ci(LevelB, LevelC, anova_result)
print(result)
```

1.3 统计分析结论

1.3.1 第（1）问

$$
\begin{cases}
40 & = \mu_A + e_{A1} \\
48 & = \mu_A + e_{A2} \\
38 & = \mu_A + e_{A3} \\
42 & = \mu_A + e_{A4} \\
45 & = \mu_A + e_{A5} \\
46 & = \mu_A + e_{A6} \\
26 & = \mu_B + e_{B1} \\
34 & = \mu_B + e_{B2} \\
30 & = \mu_B + e_{B3} \\
28 & = \mu_B + e_{B4} \\
32 & = \mu_B + e_{B5} \\
33 & = \mu_B + e_{B6} \\
39 & = \mu_C + e_{C1} \\
40 & = \mu_C + e_{C2} \\
48 & = \mu_C + e_{C3} \\
50 & = \mu_C + e_{C4} \\
50 & = \mu_C + e_{C5} \\
52 & = \mu_C + e_{C6} \\
e_{ij} & \overset{IID}{\sim} N(0, \sigma^2), i = A, B, C; j = 1, 2, 3, 4, 5, 6
\end{cases}
$$

1.3.2 第（2）问

方差分析表如表 2。

表 2: 电池寿命表

方差来源	平方和	自由度	均方	F 值	P 值
工厂 SSA	855.1	2	427.6	23.25	$\mathbf{2.54e-05}$
误差 SSE	275.8	15	18.4		
总和 SST	1,130.9	17			

由于 P 值为 $2.54e-05 < 0.05$，则在显著性水平 $\alpha = 0.05$ 下三厂生产的电池的平均寿命有显著差异。

1.3.3 第（3）问

三个区间分别为：

$$[\bar{y_A} - \bar{y_B} - t_{1-\frac{\alpha}{6}}(15)\hat{\sigma}\sqrt{\frac{1}{6}+\frac{1}{6}}, \bar{y_A} - \bar{y_B} + t_{1-\frac{\alpha}{6}}(15)\hat{\sigma}\sqrt{\frac{1}{6}+\frac{1}{6}}]$$

$$[\bar{y_A}-\bar{y_C}-t_{1-\frac{\alpha}{6}}(15)\hat{\sigma}\sqrt{\frac{1}{6}+\frac{1}{6}},\bar{y_A}-\bar{y_C}+t_{1-\frac{\alpha}{6}}(15)\hat{\sigma}\sqrt{\frac{1}{6}+\frac{1}{6}}]$$

$$[\bar{y_B}-\bar{y_C}-t_{1-\frac{\alpha}{6}}(15)\hat{\sigma}\sqrt{\frac{1}{6}+\frac{1}{6}},\bar{y_B}-\bar{y_C}+t_{1-\frac{\alpha}{6}}(15)\hat{\sigma}\sqrt{\frac{1}{6}+\frac{1}{6}}]$$

其中 $\bar{y_A},\bar{y_B},\bar{y_C}$ 分别是三个水平观测值的均值，$\hat{\sigma}=\frac{SSE}{15}$，$\alpha=0.05$。
用软件计算得到三者的 95% 同时置信区间：

$$\mu_A-\mu_B\in[5.997483,19.33585]$$

$$\mu_A-\mu_C\in[-10.00252,3.33585]$$

$$\mu_B-\mu_C\in[-22.66918,-9.330816]$$