

## Statistical data analysis, Assignment 6

---

**Problem 1.** An experiment is conducted to explore the relationship between the electrical conductivity (EC) of the soil and the yield of watermelons (kg/ha). Below are the summary data:

EC level	1.6	3.8	6.0
$n_i$	4	4	4
$\bar{Y}_i$	59.5	55.4	50.7
$S_i^2$	6.7	7.1	6.3

- (a) Constructed the appropriate analysis of variance table.
- (b) At  $\alpha = 0.05$ , test the hypothesis that there are no effects due to the EC of the soil.

**Problem 2.** The following is a statistical summary table of the difference test scores of the five groups of students after participating in the competition:

Source	SS	DF	MS	F
Between	40	(a)	(b)	(c)
Within	(d)	60	5.00	

- (a) Complete (a), (b), (c), (d).
- (b) What is the number of all observations? And, what is the value of total sum of square?
- (c) At significant level  $\alpha = 0.01$ , is there a difference in the treatment mean?

**Problem 3. (R practice)** Professor Samuel Oak is studying the discharge current of Pikachu. After research and measurement many times, the discharge data obtained are ordered as follows:

1, 3, 7, 8, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 29, 32, 35, 40, 42, 45, 50 (kA)

- (a) For drawing Q-Q plot, we need to convert this data into the standard normal quantile, that is, assume that this data obeys Normal distribution, standardizing this data, denoted  $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$  orderly. and compare it with  $Z_{(k)}$  with  $P(Z < Z_{(k)}) = (k - 0.5)/n$ , where  $Z \sim N(0, 1)$ . Please obtaining  $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ .
- (b) Draw the two Q-Q plot of the discharge variable side-by-side to check normality by using the results obtained from part(a) and *qqnorm*, *qqline* (note. In comparison, the samples inputed in *qqnorm* or *qqline* should be  $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$  in R , respectively. If discharge obeys Normal distribution, the two plots should be nearly identical.

**Problem 4. (R practice)** For iris data, explore the relationship between Petal Length and Species:

- (a) Draw the boxplots for Sepal.Length to Species. Comment on the difference in Sepal.Length mean across different species.
- (b) At level 0.05, is there a difference in the species group mean? Please report the ANOVA table and explain what statistics you used to do your conclusion.
- (c) Please check these assumption: normality and homogeneity of variance.