Homework of Model Design

sword

2019-09-21

Contents

1 Exercise in Chapter I

1.1 Exercise 1.3

First of all, we need to recognize what the problem is. The problem we encounter is to study the relationship between the growth of flowers in the garden and the conditions of sunlight, water, fertilizer and soil. The flower here should be limited to one certain type of garden flowers, satisfying the characteristic of large number of easy cultivation. Meanwhile, the individuals we select to test must be in the same condition.

Then, we set out to select the response variable. There are plenty of attributes that can be used to describe the growth of flowers. Here go some easy to measure and compare, including flowering rate, crown diameter and flower lasting days.

Now we should determine factors, levels, and range. We choose four factors to study, and the levels or ranges are listed as below.

- Sunshine of 50% shade net or natural light.
- Moisture of watering once, twice or three times a day(30 ml for each time).
- Fertilizer of urea, super phosphate or potassium sulphate.
- Soil condition of acidic soil, neutral soil or alkaline soil.

1.2 Exercise 1.9

We will illustrate the significance of randomization by the following reasons.

- Randomization guarantees the normality of error. The randomized test randomly selects test sequence and experimental condition so that the errors in all directions are all the same, thereby satisfying the normality.
- Randomization reduces the interference of hidden variables. The samples are randomly assigned to make the effects of implicit variables among the groups consistent, which enhances the comparability. Bias can be effectively controlled as well.
- Randomization strengthens the robustness of the model. The randomized test largely suppresses the heteroscedasticity of errors and makes the anti-interference ability of the model stronger.

2 Exercise in Chapter II

2.1 Exercise 2.7

We can conduct a paired sample t-test and P-value is defined as follows.

$$\text{P-value} = P\{t(9) > t_0\}$$

where $t(9)$ stands for t-distribution with degree of freedom of 9.

Hence the series of P-value with different t_0 given is shown in the table below.

Table 1: P-value of t_0 given

	t_0	P-value
(a)	2.31	0.0231
(b)	3.60	0.0029
(c)	1.95	0.0415
(d)	2.19	0.0271

2.2 Exercise 2.16

Suppose the breaking strength of a fiber obeys $\mathrm{N}(\mu\,, 3^2)$.

(a) $H_0\,:\,\mu < 150$ v.s. $H_1\,:\,\mu \geq 150$

(b) We can conduct a U-test with MATLAB and the result is displayed as below.

Table 2: U-test for breaking strength

U-stat	P-value	h	ci
-0.83	0.7977	0	$[146.28\,, \infty)$

Since P-value is larger than 0.05, we cannot reject the null hypothesis under confidence level of 5%. The conclusion is that the breaking strength of a fiber is less than 150 psi.

(c) The P-value is, just as above, 0.7977.

(d) One of the 95% confidence interval is $[146.28\,, \infty)$.

2.3 Exercise 2.26

(a) Use MATLAB, and the result is shown below.

Table 3: F-test for the equality of two variance of type1 and type2

F-stat	df1	df2	P-value	h
0.9782	9	9	0.9744	0

Since P-value is larger than 0.05, we convince that the two variances are equal.

(b) The result is shown as below.

Table 4: t-test for burning times

t-stat	df	P-value	h
0.0480	17.9978	0.9622	0

Since P-value is 0.9622, larger than 0.05, we convince that the mean burning times are equal.

(c) The normality assumption makes us able to determine the distribution of testing statistics, otherwise, we cannot construct a rejecting region.

We can conduct kstest to test the normality assumption of both types of flares, and the result is displayed as below.

Table 5: ks-test for both normality assumption

	ks-stat	P-value	h
Type1	0.1737	0.8474	0
Type2	0.1263	0.9908	0

With what is shown above, we convince that the two normality assumptions are both true.

2.4 Exercise 2.37

(a) The data indicates that both solutions do NOT have the same mean etch rate since we have the following test table.

Table 6: t-test for etch rate

t-stat	df	P-value	h	ci
-2.9584	7	0.0212	1	$[-0.7422\,, -0.0828]$

(b) The 95% confidence interval, as shown above, is $[-0.7422\,, -0.0828]$.

(c) The result is displayed as below.

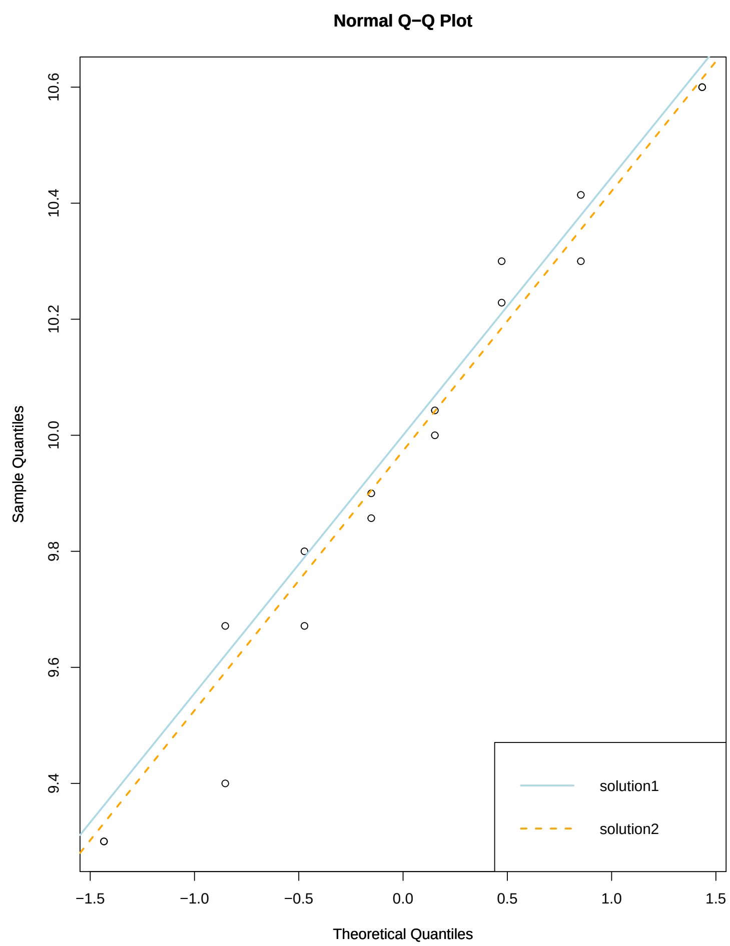

Figure 1: normal probability of two solutions

Since two lines are approximately parallel, the two solutions data have equal variance.