# DOE – An example of
# Two-Factor Experimental Design with Replication

In the last blog on “DOE – Two-factor factorial design”, we have discussed the statistical concepts and equations for the two-factor experimental design with replications. Now we illustrate these concepts with a simple statistical design of experiments.

Hypothetically we conducted a series of dirt removal experiments to study the effects of two detergent brands X and Y, and the effects of warm and hot washing temperatures, with the other factors such as the period of washing and the rotary speeds being kept constant. The dirt removed measured in mg in this $2^2$ (two levels and two factors) factorial experiments are tabulated as below:

| Detergent X | | Detergent Y | |
|---|---|---|---|
| Warm | Hot | Warm | Hot |
| 14 | 18 | 17 | 21 |
| 16 | 19 | 19 | 20 |
| 13 | 17 | 20 | 18 |
| 12 | 19 | 17 | 22 |
| 17 | 20 | 22 | 23 |

From the above table, we have factors $r$ = 2, level $c$ = 2, replication $n'$ = 5 and total number of data $n$ = 20.

If we were to apply the equations listed in my previous blog on “DOE – Two-factor factorial design”, we would get the same outcome as the outputs generated by any statistical software.   Hence, we can use the Microsoft Excel spreadsheet for its output. First of all, we would re-tabulate the data in the following manner:

| Detergent | Warm | Hot |
|---|---|---|
| X | 14 | 18 |
| X | 16 | 19 |
| X | 15 | 17 |
| X | 12 | 19 |
| X | 17 | 20 |
| Y | 17 | 21 |
| Y | 19 | 20 |
| Y | 20 | 18 |
| Y | 17 | 22 |
| Y | 22 | 23 |

The ANOVA: Two-Factor with replication in the Excel Data Analysis Tools was used and its output is given below:

Anova: Two-Factor With Replication

| SUMMARY | Warm | Hot | Total |
|---|---|---|---|
| X | | | |
| Count | 5 | 5 | 10 |
| Sum | 74 | 93 | 167 |
| Average | 14.8 | 18.6 | 16.7 |
| Variance | 3.7 | 1.3 | 6.23 |
| | | | |
| Y | | | |
| Count | 5 | 5 | 10 |
| Sum | 95 | 104 | 199 |
| Average | 19.0 | 20.8 | 19.9 |
| Variance | 4.5 | 3.7 | 4.54 |
| | | | |
| Total | | | |
| Count | 10 | 10 | |
| Sum | 169 | 197 | |
| Average | 16.9 | 19.7 | |
| Variance | 8.54 | 3.57 | |

ANOVA

| Source of Variation | SS | df | MS | F | P-value | F crit |
|---|---|---|---|---|---|---|
| Brand | 51.2 | 1 | 51.2 | 15.515 | 0.001 | 4.494 |
| Temperature | 39.2 | 1 | 39.2 | 11.879 | 0.003 | 4.494 |
| Interaction | 5 | 1 | 5 | 1.515 | 0.236 | 4.494 |
| Within error | 52.8 | 16 | 3.3 | | | |
| | | | | | | |
| Total | 148.2 | 19 | | | | |

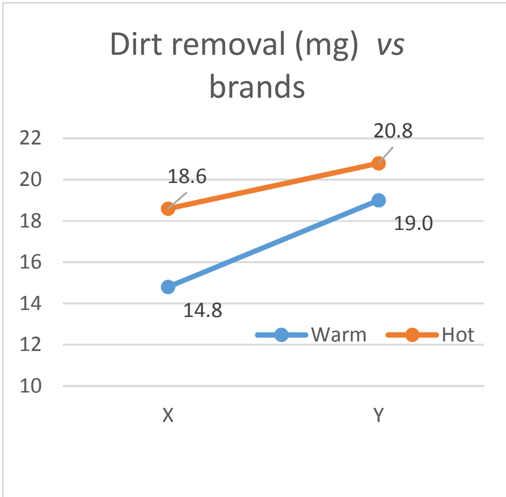
To interpret the results, we start by testing whether there is an interaction effect between the factor *A* (detergent brand) and factor *B* (washing temperature). If the interaction effect is found to be significant, we have to be cautious in the interpretation of any significant main effects. On the other hand, if the interaction effect is not significant, we can then focus on the main effects, i.e. potential differences in detergent brand (factor *A*) and potential differences in washing temperatures (factor *B*).

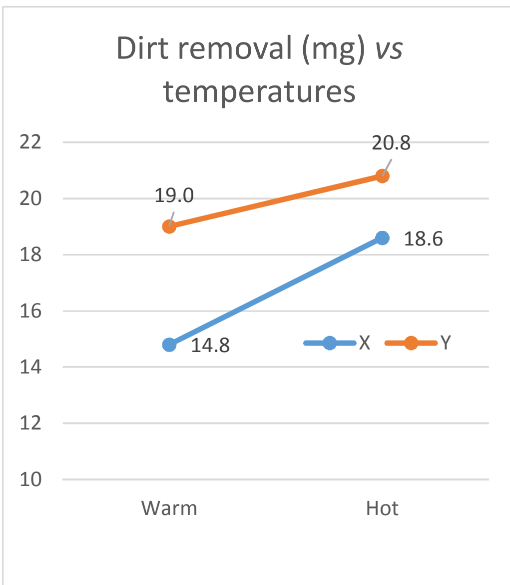
At the 0.05 level of significance to determine whether there is evidence of an interaction, the decision rule is to reject the null hypothesis of no interaction between brand and temperature if the calculated F value is larger than the critical $F$ value of 4.494, the upper-tail critical value from the $F$ distribution with 1 degree of freedom (*df*) in the numerator and 16 degrees of freedom in the denominator. Because $F = 1.515 < F_c = 4.494$, or alternatively because the *P*-value = 0.236 > 0.05, we conclude that we can accept the null hypothesis $Ho$, accepting the fact that there is insufficient evidence of an interaction between these two factors. Now, let us turn our focus on the main effects.

In testing at the 0.05 level of significance for a difference between the two detergent brands (factor *A*), the decision rule is to reject the null hypothesis if the calculated $F$ value exceeds 4.494, the upper-tail critical value from the $F$ distribution with 1 degree of freedom (*df*) in the numerator and 16 degrees of freedom in the denominator. Because $F = 15.515 > F_c = 4.494$, or because the *P*-value = 0.001 < 0.05, we reject $Ho$ and conclude that there is evidence of a difference between the two detergent brands in terms of the average amount of dirt removed. Brand Y is removing more dirt (an average of 19.9mg) than brand X (16.7 mg).

Similarly in testing at the 0.05 level of significance for a difference between the two washing temperatures (factor *B*), the decision rule is to reject the null hypothesis if the calculated $F$ value exceeds 4.494, the upper-tail critical value from the $F$ distribution with 1 degree of freedom (*df*) in the numerator and 16 degrees of freedom in the denominator. Because $F = 11.879 > F_c = 4.494$, or because the *P*-value = 0.003 < 0.05, we reject $Ho$ and conclude that there is evidence of a difference between the two washing temperatures in terms of the average amount of dirt removed. Brand Y is removing more dirt (an average of 19.9mg) than brand X (16.7 mg).

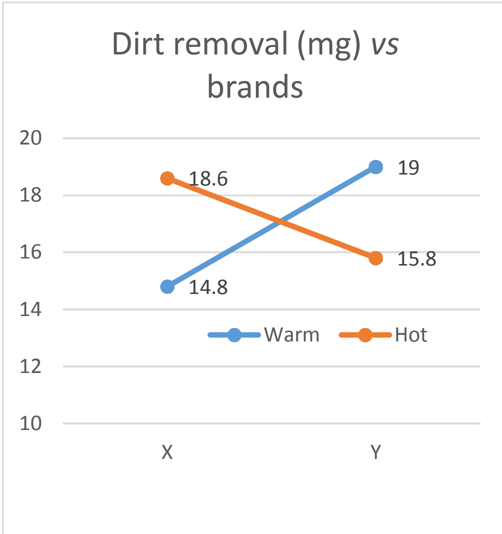
Graphically, we can also present the above results in plots as below:

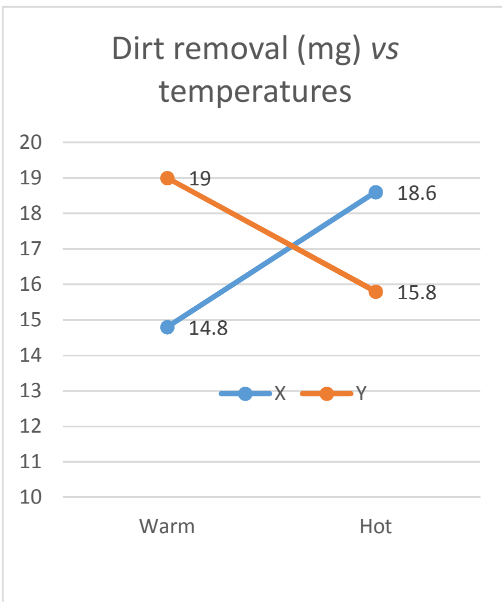



*What would be the interpretation if there was evidence of interaction between the brands and the washing temperatures?*

If there was a pronounced interaction between these two factors, we would conclude that any differences between the brands is different for the washing temperatures. For example, if the average dirt content removed by brand Y at a hot temperature was 15.8 mg instead of 20.8 mg in this example, then we could conclude that brand X was better at a hot temperature (18.6 mg dirt) and brand Y was more dirt efficient at a warm temperature (19.0 mg dirt). The profound interaction effect is seen vividly by the following plots:


Dirt removal (mg) *vs* brands
20
18
16
14
12
10
18.6
19
15.8
14.8
Warm
Hot
X
Y


Dirt removal (mg) *vs* temperatures
20
19
18
17
16
15
14
13
12
11
10
19
18.6
15.8
14.8
X
Y
Warm
Hot