Applied Statistics

Lecture Notes

Metin Oktay Yılmaz

Systems Design Master
FH Kärnten
Austria

Contents

1 Design of Experiment

07.03.2023 - Lecture 1 - EDV1

What is a *Design of Experiment*?

It is a straightforward method which:

- is efficient - that is, it has small number of tests
- is matched to the questions to be answered
- results in the highest possibşe precision

Example: How do you determine the weight of two objects?

Method 1: 2 measurements: First put object A on a scale and measure its weight. Then put object B on the scale and measure its weight.

Method 2: 2 measurements: Put object A and B at the same time on the scale and measure their combined weight. Then take one of the objects off of the scale and measure the weight of the object. Then subtract the two values to determine the weight of other object.

Method 2 is preffered for the following reasons:

- Measurements never give the true values since they contain an uncertainty (s). This phenomena is not unavoidable.
- The uncertainties (errors) occur randomly: Probability that the measurements are greater than the real value is equal to the probability that measurements are lower than the real value.

Uncertainty of Method 1: Chance that measured value is greater than A (that is, A + s) is the same as it is lower than A (that is, A - s).

Expected uncertainty (Variance) s ***of A:***

$$0.5(A + s - A)^2 + 0.5(A - s - A)^2 = s^2$$

Result: Weight of A is known to $\pm s$

Uncertainty of Method 2: Now we have 4 possibilities for the outcome of the measurement:

- $A + B$ and $A - B$ are both too big
- $A + B$ and $A - B$ are both too low
- $A + B$ is too low, $A - B$ is too big
- $A + B$ is too big $A - B$ is too low

Expected uncertainty (Variance) of A:

$$= 0.25 * \left(\frac{(A + B + s) + (A - B + s)}{2} - A\right)^2$$

$$+0.25 * \left(\frac{(A+B-s)+(A-B-s)}{2} - A\right)^2$$

$$+0.25 * \left(\frac{(A+B-s)+(A-B+s)}{2} - A\right)^2$$

$$+0.25 * \left(\frac{(A+B+s)+(A-B-s)}{2} - A\right)^2$$

$$= 0.5s^2$$

So, the expected uncertainty of A is now only $0.5s^2$. The weight of A is known to $\pm s/\sqrt{2}$.

Method 1 is so-called one-factor-at-a-time measurement; in Method 2, every weight was effectively measured twice. Therefore we can say that in o DoE, in a statistical design experiment all factors are varied simultaneously; not one factor at a time.

Why would you *design an experiment*?

- Method gives us a structured plan.
- Statistical experimental designs are aligned to statistical analysis(and the used software).
- Statistical experimental designs are much more efficient.
- Because of the structure plan, we are forced to get organized.

What is an *efficient experiment*?

- Efficient experiment gets the required information at the least expenditure of resources.

***required*:** Not too much, not too little. Just right.

***least resources*:** Money, human resources, time-to-market.

***experiment and test*:** There is a big difference between test and experiment. A test is required to determine if some thing or things work or not. Experiment connects the "*if*" of testing to determine "*why.*" Experiment is a structured set of coherent tests that are analysed as a whole to gain understanding of the process. Only with understanding, we are able to control the process. Knowing the functional relationship between input and output, we are able to develop robust products and processes with no suprises.

Prerequisites for a good experimentation are:

- Knowledge of the process
- Having clear goals and objectives
- A response variable or variables (or namely outputs)

Knowledge of the process:

- Prior knowledge from university or job training
- simulation of the process
- small preliminary experiments and/or tests
- other accumulated data

Having clear goals and objectives:

- know the difference between goal and objectives. A goal is the destination that we will arrive after passing objectives.

A good response variable must be:

- quantitative - that is, measurable
- precise - that is, aiming for a low range of deviation from target
- meaningul - that is, related to the customer's requirements

2 Something

Something something - arrived 16 minutes late.

We take the time from the first exp. That would be minutes. Best set of factors with "one-factor-at-a-time" is:

$t = 130$ mins; $T = 225$ °C

If I start now and check the time at 200 °C.

Basic idea of DoE: paradigm change

Target oriented method instead of one-factor-at-a-timefor onderstanding (or optimising) the process or product.

Slogan: Investigate many factors $x_1, x_2, x_3, ..., x_n$ (input variables) with few levels instead of few factors with many levels. Therefore we get the following DoE procedure:

1 Recognise and describe the problem

2 Decide on the input parameters (factors) and their levels

3 Choose the appropriate response variable or variables

4 Design the experiment

5 Perform the experiment and measure the outcome of each run

6 Evaluate the dataset

7 Finalize and document the results

3 Factorial Designs

Let's say we have two factors x_1, x_2. We now choose the levels: Two levels for x_1, three levels for x_2. Altogether we have 6 combinations, which means 6 runs. This is called a $2 * 3$ factorial design. In general we have more than 2 factors: $x_1, x_2, x_3...x_p$

$$\begin{gathered} x_1\text{: } l_1 \text{ levels} \\ x_2\text{: } l_2 \text{ levels} \\ x_3\text{: } l_3 \text{ levels} \\ \cdots \\ x_p\text{: } l_p \text{ levels} \end{gathered}$$

We have in total:

$$l_1 * l_2 * ... * l_p$$

factorial design. A speecial case is that each factor is varied on exactly two levels. This is called 2^p factorial design.

Example:

$p = 2$ (2 factors) at 2 levels $\rightarrow 2^2$ factorial design

num	factor x_1	factor x_2	mass flow	temperature	depos. rate	notation
1	-	-	1000 sccm	680 °C	140 nm/min	y_1
2	+	-	1500 sccm	680 °C	372 nm/min	y_2
3	-	+	1000 sccm	750 °C	428 nm/min	y_3
4	+	+	1500 sccm	750 °C	500 nm/min	y_4

TODO: Make a massfow-temperature cartesian graph and point out the deposition rates.

Main effect mass flow:

TODO correct this:

$$\frac{500+372}{2} - \frac{2}{1}$$

Main effect temperature:

$$\frac{428+500}{2} - \frac{140+372}{2} = 208$$

If we change the temperature from 680 °C to 750 °C, the deposition rate changes by 208 nm/min, averaged over the mass flow.

3) What is the effect of the temperature on the deposition rate, if we change the temperature from °C to 750 °C and differentiate between the massflow.

Interaction between temperature and massflow (aka interaction temperature * massflow)

$$\frac{128-288}{2} = -80 = \frac{(y_4-y_2)-(y_3-y_1)}{2} = \frac{y_1+y_4}{2} - \frac{y_2+y_3}{2}$$

If we change the temperature from 680 °C to 750 °C and differentiate between the massflow, the deposition rate is changed by -80 nm/min

4) What is the effect of massflow on the deposition rate if we change the massflow from 1000 sccm to 1500 sccm and differentiate between the temperature?

$$\frac{72-232}{2} = -80 = \frac{(y_4-y_3)-(y_2-y_1)}{2} = \frac{y_1+y_4}{2} - \frac{y_2+y_3}{2}$$

If we change the massflow from 1000 sccm to 1500 sccm and differentiate between the temperature, the deposition rate is again changed by -80 nm/min.

If the temperature range of 680 °C and 750 °C and the mass flow range of 1000 sccm and 1500 sccm, the temperature has the strongest effect on the deposition rate, followed by the mass flow and the interaction.

We can now visualize the results:

Main effect plots:

TODO: Draw two cartesian graphs. one is massflow [sccm] - deposition rate [nm/min] the other is temperature [°C] - deposition rate [nm/min]

First plot:
1000 sccm - 281 nm/min
1500 sccm - 436 nm/min
Main effect mass flow: $\Delta = 152$

Second plot:
680 °C - 256 nm/min
750 °C - 464 nm/min
Main effect temperature: $\Delta = 208$

there are many graphs drawn here. got lost.
cornerstone software is used to analyse the dataset.