

Design of Experiments (DOE)

- experimentation is the process of proposing and verifying hypotheses
- a hypothesis is a statement that can be tested
 - e.g. "an alphabetically ordered menu item list can be scanned more quickly than a randomly organised list"
- the experimenter manipulates one or more "variables", or "factors", to observe the effect these changes have on one or more other "variables" or "factors"
- DOE is a procedure for planning the experiment so that the data obtained can be analysed to yield valid and objective conclusions.

Note: source for this material is the *e-Handbook of statistical methods*, available at <http://www.itl.nist.gov/div898/handbook/index.htm>

COMP106 - lecture 21 - p.1/15

Elements of DOE

- the experiment is seen as a "black box", where only input and output are considered
- the input are the *process variables*
 - these are the factors that the experimenter chooses to manipulate
 - for example: screen size, input method (menu vs command language) etc.
 - these are often called *independent* variables
- the output are the *response variables*
 - these are the result of the experiment
 - for example: typing speed, number of mistakes, etc.
 - these are often called *dependent* variables, as they should change with changes in the process variables

COMP106 - lecture 21 - p.2/15

Elements of DOE (2)

- sometimes *co-factors* are also considered
 - these are uncontrolled variables, that might however influence the output
 - for example: different machines, ambient factors, etc.
- and one needs to keep in mind that *errors* can occur:
 - any observations, or response variable, is supposed to include some error (or noise)
 - statistical methods are needed to collect enough evidence to distinguish between the real output and the error

COMP106 - lecture 21 - p.3/15

How to organise an experiment

- first: list all possible process variables, or input factor, that can influence the result
- decide the *treatments* to use in the experiment
 - a treatment is basically the combination of the process variables that one decides to use
 - for instance, with two input factors, one can study each of them alone and the combination between the two
 - with three input factors, one can study each of them alone, all the possible pairs, and the combination of the three of them together
 - the more cross-factors are considered, the better the experiment is

COMP106 - lecture 21 - p.4/15

How to organise an experiment (2)

- the simplest model for an experiment is the *linear* model, that can be summarised with the formula (for two input factors):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \text{error}$$

- where Y is the output, X_1 and X_2 are the two input factors, and all β are the parameters that one wants to find out (that is how the input influence the output)
- for three input factors you will have:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3 + \text{error}$$

COMP106 - lecture 21 - p.5/15

When to use DOE?

- **comparative** experiment:
 - choosing between alternatives
 - we want to make a choice between different values of one of the input factors
 - for instance, we want to compare menu vs command language
 - this is the main factor under study, although there might be other factors to be included in the experiment
- **screening** experiment:
 - selecting the key factor that influences the response
 - some of the input factors may have little impact on the response, while others may be crucial
 - we want to determine these important factors, and screen them out from the less important ones

COMP106 - lecture 21 - p.6/15

When to use DOE? (2)

- **modelling** experiments:
 1. Hitting a target
 - we want to obtain a certain value for the output variable
 - for example, reducing to zero the mistakes the user does(!)
 - rather than working in an *ad hoc* manner, until we reach the right combination of input factors
 - one can fit a model estimated from a small experiment and use this model to determine the necessary adjustments to hit the target
 2. Maximising or minimising a response:
 - similar to the previous one
 - in this case we want, for example to minimise the number of errors

COMP106 - lecture 21 - p.7/15

When to use DOE? (3)

3. Reducing variation:
 - again similar to hitting a target
 - we want to reduce disparities among data
 - e.g. have that all users do the same numbers of errors
4. Making a Process Robust
 - minimize the influence of ALL external factors on the final result
 - a robust system should have good performance under extreme conditions

COMP106 - lecture 21 - p.8/15

Checklist for successful DOE

- Check performance of measurement devices first.
- Keep the experiment as simple as possible.
- Check that all planned runs are feasible.
- Watch out for process drifts and shifts during the run.
- Avoid unplanned changes.
- Allow some time (and back-up) for unexpected events.
- Maintain effective ownership of each step in the experimental plan.
- Preserve all the raw data—do not keep only summary averages!
- Record everything that happens.
- Reset equipment to its original state after the experiment.

COMP106 - lecture 21 - p.9/15

Assumptions for DOE

1. the measurement system is capable
 - that is the measurement devices can measure all changes the experimenter hopes to see
 - a capable measurement system is:
 - accurate: there is agreement between a measurement made on an object and its true (target or reference) value
 - unbiased: it is accurate *on average*, over a large number of measurements
2. the process is stable
 - there are no "special causes" that create variations in the results
 - (normal, predictable causes are OK)

COMP106 - lecture 21 - p.10/15

Assumptions for DOE (2)

3. the response variables behaviour can be described by a simple model (e.g. linear)
4. the model *residuals* are well behaved
 - residuals are estimates of the differences between the observed and the predicted results
 - the predicted results depend on the choice of the model that was assumed would apply
 - as in all problems involving measuring errors, the assumption is that
 - *one expects residuals to be (roughly) normal and (approximately) independently distributed with a mean of 0 and some constant standard deviation*
 - this can be checked with the usual graphical methods
 - histograms
 - normal probability plots

COMP106 - lecture 21 - p.11/15

Setting up the experiment

- the first step in organising an experiment is choosing the process and response variables
 - include all important factors
 - check there are not impossible combinations (e.g. using mouse and keyboard at the same time)
- then one needs to choose the *ranges* or levels for each factor
 - i.e.: the set of values that the factor can have
 - one can choose extreme factors
 - so you make sure you catch all possibilities
 - but this could lead to unfeasible experiments, or a complicated model for the response variables behaviour
 - or stick to a list of the most probable/feasible ones
 - the most popular DOE only consider two levels (e.g. small screen/large screen)

COMP106 - lecture 21 - p.12/15

Setting up the experiment (2)

- then, one should decide on a schedule for *randomisation*
 - this is extremely important
 - one needs to be sure that one run of the experiment is not influenced by the previous run, nor influences the following run
 - for example, in an experiment with real user completing a task
 - participants should be randomly selected
 - then they should be randomly assigned to the level groups (e.g. the ones using the small screen and the ones using the large screen)
 - each participant receives the same instructions and procedure

COMP106 - lecture 21 - p.13/15

How many runs of the experiment?

- in a completely randomised experiment:
 - if there are k process variables, or factors
 - and there are L levels for each factors
 - and there should be n replications per level
 - then the number of runs is $N = k * L * n$
- for example:
 - two factors: screen size and input method ($k = 2$)
 - two levels for each ($L = 2$):
 - screen size = Large and Small
 - input method = Mouse and Keyboard
 - three runs for each level ($n = 3$)
 - $N = k * L * n = 2 * 2 * 3 = 12$

COMP106 - lecture 21 - p.14/15

Which experimental design?

- it depends on:
 - the type of experiment
 - the amount of resources available
 - the control one wants to have over making wrong decisions (*Hypothesis tests*)
- for example:

Number of factors	Comparative experiment	Screening experiment	Modelling experiment
1	Single factor randomised		
2 to 4	Randomised block	Full or fractional factorial	Central composite
5 or more	Randomised block	Fractional factorial	Screen first to reduce the number of factors

COMP106 - lecture 21 - p.15/15