Data Science 1
Experimental Design and Analysis of Variance
Introduction

Ann Maharaj

Experimental Design

- Careful planning of an experiment is a prerequisite for success when employing statistical principles to draw conclusions.
- A well-designed experiment helps to achieve clear interpretations and to avoid complicated analysis.
- The basic units for which response measurements are collected are called experimental units or subjects; these measurements are values of a dependent variable which is quantitative.
- Distinct types of conditions that are observed or manipulated on the experimental units are called factors. Factors are the independent variables which are qualitative.
- The different modes of presence of a factor are called factor levels.
- The levels of a factor are called treatments, and for two or more factors, each specific combination of the factor levels is called a treatment combination.

Experimental Design

We are going to study three types of designs

1. Completely Randomised Design
2. Randomised Block Design
3. 2-Factor Factorial Design

Example 1: Completely Randomised Design

- As part of a multilab study, four fabrics are tested for flammability at a national standards centre. Random samples of five dresses made from each fabric were selected. Burn times in seconds were recorded after a paper tab was ignited on the hem of each dress. The data are shown below.

	Treatments		
Fabric A	Fabric B	Fabric C	Fabric D
17.8	11.2	11.8	14.9
16.2	11.4	11.0	10.8
17.5	15.8	10.0	12.8
17.4	10.0	9.2	10.7
15.0	10.4	9.2	10.7

- Independent samples
- One factor (Independent variable - qualitative): Fabric (4 treatments)
- Experimental units: 20 Dresses
- Response (dependent variable - quantitative): Burn time
- Number of replicates at each treatment level $= 5$
- We want to test there is no difference in degree of flammability for the four fabrics.

Example 2: Randomised Block Design

- The cutting speeds of four types of tools (treatments) are being compared in an experiment. Five materials of varying degrees of hardness are to be used as experimental blocks. The data pertaining to measurements of cutting time in seconds is given below.

	Treatments			
	Tool 1	Tool 2	Tool 3	Tool 4
Blocks				
Material 1	12	20	13	11
Material 2	2	14	7	5
Material 3	8	17	13	10
Material 4	1	12	8	3
Material 5	7	17	14	6

- Related samples
- Two factors (independent variables-qualitative): Type of tool (4 treatments)
  Degree of hardness of the material (5 blocks).
- Experimental units: 20 Pieces of material.
- Response dependent variable -quantitative: Cutting time
- Number of replicates for each treatment combination $= 1$
- We want test that there is no difference in the cutting speeds of the four types of tools, taking into account the variation in the five material types.

Example 3: 2-Factor Factorial Design

An engineer tests three plate materials for a new battery at three temperature levels ($-9^oC, 21^oC, 68^oC$). Four batteries (replicates) are tested with each combination of plate material and temperature. All 36 tests are run in random order and the lifetimes in hours are recorded.

Table: Lifetime in hours for Battery Design Experiment

	Temperature Level		
	1	2	3
Material Type			
1	130	34	20
	74	80	82
	155	40	70
	180	75	58
2	150	136	25
	188	122	70
	159	106	58
	126	115	45
3	138	174	96
	168	120	104
	110	150	82
	160	139	60

- Independent samples
- Two factors (independent variables - qualitative): Material Type, Temperature Level
- Factor A: Material Type; 3 treatment levels
- Factor B: Temperature Level: 3 treatment levels
- Number of replicates for each combination for each treatment combination $= 4$
- Experimental units: 36 batteries
- Response (dependent variable - quantatative): Lifetime in hours
- We want to test the effect of material type and temperature have on the life of a battery?

Available materials

1. Short videos including demonstrations using R software
2. Accompanying slides
3. R scripts and relevant data sets
4. Exercises
5. Additional material
6. List of references