Lab 2 Report

From

Gokul Kaisaravalli Bhojraj
Id: 80789
Program: Business Intelligence
e-mail: h18gokka@du.se

1) **Consider Example 3 Lecture 3 (Dated 23 April). Create a treatment allocation layout for an RBD to assign 3 treatments on 3 workers working 2 days on each on the machines.**

   Question: A factory manager had to decide which one among the three available brands of a machine should be bought. Manager got one machine from each brand, for a few days, for testing. One machine needs one operator and produces identical items, but their production speeds differ.

   Solution:

RBD (**Randomized Block Design)** which can be defined as: With a randomized block design, the experimenter divides subjects into subgroups called **blocks**, such that the variability within blocks is less than the variability between blocks. Then, subjects within each block are randomly assigned to treatment conditions. In other words, deals with one factor applied on a population which is non-homogenous in one aspect (block).

In our case, workers are assigned to blocks, the three treatments. Each treatment is a homogeneous experience, that is why we use RBD.

R Output:

```
  plots block treatments
1   101     1          A
2   102     1          B
3   103     1          C
4   201     2          B
5   202     2          A
6   203     2          C
7   301     3          B
8   302     3          C
9   303     3          A
```

So, in this way the manager can assign the machines to the workers. Where **1**, **2** and **3** are the blocks. **A**, **B** and **C** are the workers.

2) **Consider the Latin Square Design of Lecture 3. Create a layout for an LSD to allocate 3 treatments on two blocking factors (worker and working slots of the day) each having 3 levels.**

   Question: A factory manager had to decide which one among the three available brands of a machine should be bought. Manager got one machine from each brand, for a few days, for testing. One machine needs one operator and produces identical items, but their production speeds differ.

   Solution:

LSD (**Latin-Square Design**)– Latin square allow for two blocking factors. In other words, these designs are used to simultaneously control (or eliminate) two sources of nuisance variability. Put it simple, the idea behind this is to test two variables multiple times. We can check that when we run the code line and we obtain the $sketch:

R Output:

```
$sketch
     [,1] [,2] [,3]
[1,] "B"  "C"  "A"
[2,] "A"  "B"  "C"
[3,] "C"  "A"  "B"
```

We can see below that the three treatments are assigned to the two blocking factors and all have 3 levels.

R Output:

```
   plots row col treatments
1  1001   1   1          B
2  1002   1   2          C
3  1003   1   3          A
4  2001   2   1          A
5  2002   2   2          B
6  2003   2   3          C
7  3001   3   1          C
8  3002   3   2          A
9  3003   3   3          B
```

So, in this way the manager can assign the machines to the workers by using two blocki ng factors. The two blocking factors are

- `Row`
- `Col`

Where **1**, **2** and **3** are the blocks. **A**, **B** and **C** are the workers.

3) **Paints used on marking highways must be very durable. In one experiment, paints from four different suppliers (labelled A, B, C, D) were tested on six different highway sites (1, 2, …, 6). After a considerable length of time, the average wear for the samples at six sites was as recorded (see data on Learn, in Lab2.xlsx file). What kind of experimental design is this? What can you say about the relative resistance to wear for the four paints?**

Excel sheet content:

Sites	Paints				
	A	B	C	D	
1	69	59	55	70	
2	83	65	65	75	
3	74	64	59	74	
4	61	52	59	62	
5	78	71	67	74	
6	69	64	58	74	

Solution:

Here, the type of experimental design used is **completely random design (CRD).** Because, each supplier is randomly assigned to a site and they have the same probability of being assigned to a site. The table below shows the average wear for the samples:

R Output:

```
   A-Paint B-Paint C-Paint D-Paint
2     69      59      55      70
3     83      65      65      75
4     74      64      59      74
5     61      52      59      62
6     78      71      67      74
7     69      64      58      74
```

Next, we run the anova function. Anova can be defined as: Analysis of variance, a statistical method in which the variation in a set of observations is divided into distinct components.

R Output:

```
Analysis of Variance Table

Response: val
          Df Sum Sq Mean Sq F value    Pr(>F)    
tm_vector  3 665.13 221.708  20.387 1.503e-05 ***
Blocks     5 568.71 113.742  10.459 0.0001808 ***
Residuals 15 163.12  10.875                      
---
```

We can see there is a significant difference between the paints. Furthermore, we proceed to summary () and we obtain the following:

R Output:

```
lm(formula = val ~ tm_vector + Blocks)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2917 -2.1042 -0.0833  1.5833  6.7083 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       68.8750     2.0194  34.106 1.25e-15 ***
tm_vectorPaint-B  -9.8333     1.9039  -5.165 0.000115 ***
tm_vectorPaint-C -11.8333     1.9039  -6.215 1.65e-05 ***
tm_vectorPaint-D  -0.8333     1.9039  -0.438 0.667849    
Blocks2            8.7500     2.3318   3.752 0.001922 ** 
Blocks3            4.5000     2.3318   1.930 0.072765 .  
Blocks4           -4.7500     2.3318  -2.037 0.059696 .  
Blocks5            9.2500     2.3318   3.967 0.001240 ** 
Blocks6            3.0000     2.3318   1.287 0.217759    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.298 on 15 degrees of freedom
Multiple R-squared:  0.8832,  Adjusted R-squared:  0.821 
F-statistic: 14.18 on 8 and 15 DF,  p-value: 1.032e-05
```

When we look at beta values, we compare them to the reference, which is “A Paint”, we can see that “B Paint” lasts approximately almost 10 months less than A, while C lasts almost 12 months less than A. D may last almost 1 month less than A, but if we look at its P-value we can see there is no significance.

Hence, by looking at the wearing time of all the 4 paints we can conclude that **“A Paint”** is the best option.

4) **Suppose, you want to test 3 treatments by using a CRD. How many individuals (or experimental units) do you need to recruit (in total) to ensure that any effect of size 20% of the standard deviation should be detected with a power of 0.9 at 5% level of significance?**

Solution:

For this question we will use a power analysis. Power analysis can be defined as: Power analysis is an important aspect of experimental design. It allows us to determine the sample size required to detect an effect of a given size with a given degree of confidence. Conversely, it allows us to determine the probability of detecting an effect of a given size with a given level of confidence, under sample size constraints. If the probability is unacceptably low, we would be wise to alter or abandon the experiment.

R Output:

```
     Balanced one-way analysis of variance power calculation

              k = 3
              n = 106.455
              f = 0.2
      sig. level = 0.05
           power = 0.9

NOTE: n is number in each group
```

n equals approximately **106** individuals and that is the number of individuals that we need to recruit to ensure that any effect of size 20% of the standard deviation should be detected with a power of 0.9 at 5% level of significance.

****