Data Science 1
Experimental Design and Analysis of Variance
Randomised Block Design

Ann Maharaj

Experimental Design: Randomised Block Design

Introduction

- A randomised block design is often used instead of a completely randomised design in studies where there is extraneous variation among the experimental units that may influence the response.
- A significant amount of the extraneous variation may be removed from the comparison of treatments by partitioning the experimental units into fairly homogeneous subgroups or blocks.
- So just as we can pair like subjects or experimental units to improve the procedure of taking independent samples when comparing two treatments (populations), we can also arrange or *block* subjects into homogenous groups of size k when comparing k treatments (populations).
- Then if each treatment is applied to exactly one unit in the block and if comparisons are drawn between treatment responses from the same block extraneous variability should be greatly reduced. This is the concept underlying the randomised block design.

Introduction

- 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

- Two factors (independent variables): Type of tool (4 treatments) and Degree of hardness of the material (5 blocks).
- Experimental units: Pieces of material of each type.
- Response dependent variable: Cutting time
- 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.

Introduction

- Randomisation is a basic part of the block design.
- b blocks each consisting of (partitioned into) k experimental units.
- k treatments are randomly assigned in to the experimental units with each block.
    - Once the grouping of experimental subjects in blocks is accomplished, randomly select one subject from the first block to receive Treatment 1, one of the remaining subjects is randomly selected to receive Treatment 2 and so.
    - The same procedure is repeated with a new randomisation for each of the remaining blocks.
- This means that typically after the runs in one block have been conducted, then move to another block.
    - This results in restriction on randomisation because randomisation is only within blocks.
    - However, whenever possible, it is also preferable to select a random order in which to run all of the experiments rather than to proceed block by block.
- Data within a block are dependent on each other. When $k = 2$, the randomised block design becomes the paired two sample case.

Data Structure

Designating the measurement corresponding to Block *i* and Treatment *j* by y_{ij}, the data structure of a randomised block design with *b* blocks (rows) and *k* treatments (columns) is shown below.

Table: Data Structures of the Randomised Block Design

	Treatment 1	Treatment 2				Treatment k	Block Means
Block 1	y_{11}	y_{12}	.	.	.	y_{1k}	$\overline{y}_{1.}$
Block 2	y_{21}	y_{22}	.	.	.	y_{2k}	$\overline{y}_{2.}$
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
Block b	y_{b1}	y_{b2}	.	.	.	y_{bk}	$\overline{y}_{b.}$
Treatment Means	$\overline{y}_{.1}$	$\overline{y}_{.2}$				$\overline{y}_{.k}$	$\overline{y}_{..}$

j^{th} treatment mean (column) : $\overline{y}_{.j} = \frac{1}{b}\sum_{i=1}^{b} y_{ij}$, $j = 1, 2, \ldots, k$

i^{th} block mean (row) : $\overline{y}_{i.} = \frac{1}{k}\sum_{j=1}^{k} y_{ij}$, $i = 1, 2, \ldots, b$

Grand mean: $\overline{y}_{..} = \frac{1}{k}\sum_{j=1}^{k} \overline{y}_{.j} = \frac{1}{b}\sum_{i=1}^{b} \overline{y}_{i.}$

Partition of Total Variation

Given the basic decomposition

$$(y_{ij} - \overline{y}_{..}) = (\overline{y}_{.j} - \overline{y}_{..}) + (\overline{y}_{i.} - \overline{y}_{..}) + (y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})$$

the following decomposition in sums of square can be obtained

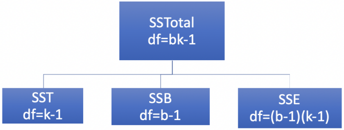
$$\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij} - \overline{y}_{..})^2 = b\sum_{j=1}^{k}(\overline{y}_{.j} - \overline{y}_{..})^2 + k\sum_{i=1}^{b}(\overline{y}_{i.} - \overline{y}_{..})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2$$

$\Rightarrow$ Total SS $=$ Treatment SS $+$ Block SS $+$ Residual or Error SS

That is, the total variation is partitioned into variation due to treatments, blocks and due to the residual or error variation.

$$SSTotal = SST + SSB + SSE$$

Partition of Total Variation


SSTotal
df=bk-1

SST
df=k-1

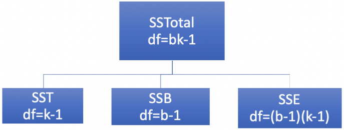
SSB
df=b-1

SSE
df=(b-1)(k-1)

- Total Sum of Squares: $SSTotal = \sum_{i=1}^{b} \sum_{j=1}^{k} (y_{ij} - \overline{y}_{..})^2$
  has degree of freedom $df = bk - 1$
  $SSTotal$ is a measure of the dispersion of all individual data values in relation to the mean $\overline{y}_{..}$ (grand mean) of the entire data set.

- Treatment Sum of Squares: $SST = b \sum_{j=1}^{k} (\overline{y}_{.j} - \overline{y}_{..})^2$
  has degree of freedom $df = k - 1$
  SST is a measure of dispersion of the treatment means from the grand mean. It is the variation due to treatments and it is also referred to as between treatment variation.

Partition of Total Variation



SSTotal
df=bk-1

SST
df=k-1

SSB
df=b-1

SSE
df=(b-1)(k-1)

- Block Sum of Squares: $\text{SSB} = k\sum_{i=1}^{b}(\bar{y}_{i.} - \bar{y}_{..})^2$ has degree of freedom $df = b - 1$
  SSB is a measure of dispersion of the block means from the grand mean. It is the variation due to blocks and it is also referred to as between block variation.

- Error Sum of Squares: $\text{SSE} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ has degrees of freedom $df = (b-1)(k-1)$.
  SSE is an aggregated measure of how much data values vary from the mean of their own treatment and the mean of their own block. It is the variation due to random sampling (unexplained variation).

Mean Squares

Mean Square = Sum of Squares/df

MST = SST/(k-1)

MSB = SSB/(b-1)

MSE = SSE/(b-1)(k-1)

Tests for treatment effects and for block effects can now be performed by comparing the corresponding mean square with the yardstick of the residual mean square *MSE* which can be shown to be an unbiased estimator of the common population variance σ^2.

Population Model for a Randomised Block Design

- To implement a formal statistical test for treatment effects, and block effects we require a population model for the experiment.

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

where $i = 1, 2, ..., b$ and $j = 1, 2, ..., k$

$\mu = \frac{\sum_{j=1}^{k} \mu_{.j}}{k} = \frac{\sum_{i=1}^{b} \mu_{i.}}{b} =$ overall mean
$\beta_j = \mu_{.j} - \mu$: j^{th} treatment effect; $\sum_{j=1}^{k} \beta_j = 0$
$\tau_i = \mu_{i.} - \mu$: i^{th} block effect: $\sum_{i=1}^{b} \tau_i = 0$
$\varepsilon_{ij} =$ random error of the observation from the $(i, j)^{th}$ cell.
Assume ε_{ij} are independently and identically distributed as $N(0, \sigma^2)$.

- The model is additive because within a fixed block, the block effect is fixed; for a fixed treatment, the treatment effect is fixed across blocks; this implies the blocks and treatments do not interact.

Hypothesis Testing

- Testing treatment effects:
  $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$
  All k treatment effects are equal to zero which implies there is no treatment effect and hence the treatment population means are all equal.

  H_1: Not all the β_j are equal to zero, $j = 1, 2, ..., k$.
  At least one of the treatments effects is not zero which implies here is a treatment effect and hence at least one treatment population mean is different.

  Hence the null and alternative hypotheses can be expressed as
  $H_0 : \mu_{.1} = \mu_{.2} = ... = \mu_{.k}$
  H_1: not all $\mu_{.j}$ are equal; $j = 1, 2, ..., k$.
- To test the null hypothesis, we observe that the treatment mean square, $MST = \frac{SST}{k-1}$
  - is expected to be small when the population means are all equal.
  - is likely to be large when the means differ markedly.
- It can be shown that under the null hypothesis, $F_T = \frac{MST}{MSE}$ follows an F-distribution with $(k-1)$ and $(b-1)(k-1)$ degrees of freedom.

Hypothesis Testing

- Testing block effects:
  $H_0 : \tau_1 = \tau_2 = ... = \tau_b = 0$
  All b block effects are equal to zero which implies there is no block effect hence the block population means are all equal.

  H_1: Not all the τ_i are equal to zero, $i = 1, 2, ..., b$.
  At least one of the block effects is not zero which implies there is a block effect and hence at least one block population mean is different.

  Hence the null and alternative hypotheses can be expressed as
  $H_0 : \mu_{1.} = \mu_{2.} = ... = \mu_{b.}$
  H_1: not all $\mu_{i.}$ are equal; $i = 1, 2, ..., b$.

- To test the null hypothesis, we observe that the block mean square, $MSB = \frac{SSB}{b-1}$
  - is expected to be small when the block effects are equal to zero.
  - is likely to be large when the block effects are markedly. different from zero

- It can be shown that under the null hypothesis, $F_B = \frac{MSB}{MSE}$

  follows an F-distribution with $(b-1)$ and $(b-1)(k-1)$ degrees of freedom.

Two-Way ANOVA Table

Table: Two-Way ANOVA Table

Source	Sum of Squares	df	Mean Square	F-Ratio
Treatments	SST	k -1	$MST = \frac{SST}{k-1}$	$F_T = \frac{MST}{MSE}$
Blocks	SSB	b -1	$MSB = \frac{SSB}{b-1}$	$F_B = \frac{MSB}{MSE}$
Error	SSE	$(b-1)(k-1)$	$MSE = \frac{SSE}{(b-1)(k-1)}$	
Total	$SSTotal$	$bk-1$		

Because there are two factors (treatments and blocks), we refer to this setup as a Two-Way ANOVA.

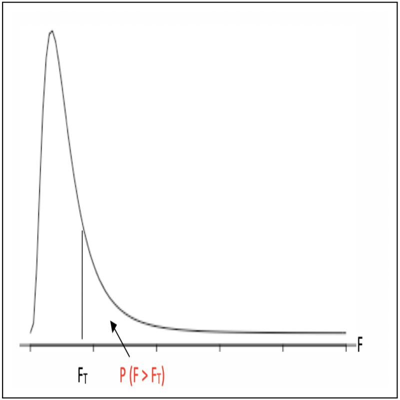
Features of Two-Way ANOVA

- Total = Treatments + Blocks + Error
- The sums of squares always add up
  - $SSTotal = SST + SSB + SSE$
- Degrees of freedom always add up.
  - $bk - 1 = (k - 1) + (b - 1) + (b - 1)(k - 1)$
- The denominators of the F-Ratios are always the same but the numerators are different.

Hypothesis Testing

Testing Treatment Effects:

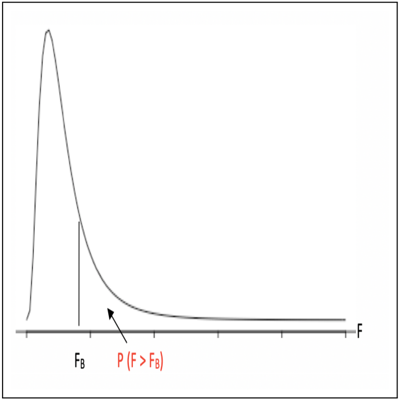
1. Hypotheses:
   $H_0 : \mu_{.1} = \mu_{.2} = ... = \mu_{.k}$
   H_1: not all $\mu_{.j}$ are equal;
   $j = 1, 2, ..., k$.
2. Level of Significance: α
3. Test Statistic: $F_T = \frac{MST}{MSE}$
4. P-value: p-value $= P(F > F_T)$
5. Decision Rule: Reject H_0 if
   $p - value < \alpha$
6. Conclusion



Hypothesis Testing

Testing Block Effects

1. Hypotheses:
   $H_0 : \mu_{1.} = \mu_{2.} = ... = \mu_{b.}$
   H_1: both all $\mu_{i.}$ are equal;
   $i = 1, 2, ..., b$
2. Level of Significance: α
3. Test Statistic: $F_B = \frac{MSB}{MSE}$
4. P-value: p-value $= P(F > F_B)$
5. Decision Rule: Reject H_0 if
   $p - value < \alpha$
6. Conclusion



Multiple Comparisons: Tukey's Test

- As before the F test for treatment effects is only the initial step in our analysis that determines if significant differences exist among treatment means.

- The next step is to perform post-hoc analysis, i.e., to detect if there are differences between the two treatments in every pair.

- Given we have k treatments, the test for difference between and two treatments, say, Treatment j and Treatment $j^{'}$ is
  $H_0 : \beta_j - \beta_{j'} = 0$ versus $H_1 : \beta_j - \beta_{j'} \neq 0$.
  which is equivalent to
  $H_0 : \mu_{.j} - \mu_{.j'} = 0$ versus $H_1 : \mu_{.j} - \mu_{.j'} \neq 0$.

- To obtain a test statistic, we need the sampling distribution of the difference between treatment means $(\overline{y}_{.j} - \overline{y}_{.j'})$ which is an estimator of $(\mu_{.j} - \mu_{.j'})$.

Multiple Comparisons: Tukey's Test

- It can be shown that $(\overline{y}_{.j} - \overline{y}_{.j'})$ is normally distributed with a mean of $(\mu_{.j} - \mu_{.j'})$

  and variance $\sigma^2 \left(\frac{1}{b} + \frac{1}{b}\right) = \sigma^2 \left(\frac{2}{b}\right)$, where σ^2 is unknown and is estimated by MSE.

- Hence, it can be shown that *Tukey's Honestly Significant Difference* test statistic is

$$T_{HSD} = \frac{(\overline{y}_{.j} - \overline{y}_{.j'}) - (\mu_{.j} - \mu_{.j'})}{\sqrt{MSE\left(\frac{2}{b}\right)}}$$

follows a t-distribution with $df = (b-1)(k-1)$.

Multiple Comparisons: Tukey's Test

- It can be shown that a $100(1-\alpha)\%$ confidence interval for each of the m number of multiple pairwise differences $\mu_{.j} - \mu_{.j'}$ is given by

$$(\bar{y}_{.j} - \bar{y}_{.j'}) \pm t_{\alpha/2m,(b-1)(k-1)} \sqrt{MSE\left(\frac{2}{b}\right)}$$

  where $t_{\alpha/2m,(b-1)(k-1)}$ is the upper $\alpha/2m$ quantile of the t-distribution with $df = (b-1)(k-1)$.

- Using this procedure, the probability of all m statements being correct is at least $(1-\alpha)$. Dividing α by $2m$ adjusts for the probability of the Type 1 error since m pairwise comparisons are required.

- The p-value of each test can also be obtained by also adjusting for the probability of Type 1 error.

Assumptions and Diagnostic Checks

Assumptions

1. Random Sampling: the observations should be from random samples drawn from the dependent populations of interest.
2. Population Normality: The populations from which the samples are drawn should be normal.
3. Homogeneity of Variance: variances of populations should be equal.

Diagnostic Checks

- We perform the normality checks on the residuals of the estimated model examining a normal Q-Q plot and the Shapiro-Wilk test for normality.
- We perform homogeneity of variances checks on the residuals by examining a residual plot and also using Levene's test. Note that other types of tests may also be use for the randomised block design. Refer to Bhandary and Dai (2013) in the Reference Section.

Levene's Test for Homogeneity of Variance

$H_0 : \sigma^2_{.1} = \sigma^2_{.2} = ... = \sigma^2_{.k}$
H_1: Not all of the $\sigma^2_{.j}$ are equal, $j = 1, 2, ..., k$

To test the null hypothesis of equal variances:

- For each combination of block i and each treatment j, determine

  $|y_{ij} - \overline{y}_{.j} - \overline{y}_{i.} + \overline{y}_{..}|$

- Perform a one-way ANOVA on these absolute measures.

- The initial data structure and transformed data structure required for Levene's Test are given on the next slide.

Table: Data Structures of the Randomised Block Design

	Treatment 1	Treatment 2				Treatment k	Block Means
Block 1	y_{11}	y_{12}	.	.	.	y_{1k}	$\bar{y}_{1.}$
Block 2	y_{21}	y_{22}	.	.	.	y_{2k}	$\bar{y}_{2.}$
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
Block b	y_{b1}	y_{b2}	.	.	.	y_{bk}	$\bar{y}_{b.}$
Treatment Means	$\bar{y}_{.1}$	$\bar{y}_{.2}$				$\bar{y}_{.k}$	$\bar{y}_{..}$

Table: Transformed Data Structure for Levene's Test

	Treatment 1	Treatment 2				Treatment k
Block 1	$\lvert y_{11} - \bar{y}_{.1} - \bar{y}_{1.} + \bar{y}_{..} \rvert$	$\lvert y_{12} - \bar{y}_{.2} - \bar{y}_{1.} + \bar{y}_{..} \rvert$	.	.	.	$\lvert y_{1k} - \bar{y}_{.k} - \bar{y}_{1.} + \bar{y}_{..} \rvert$
Block 2	$\lvert y_{21} - \bar{y}_{.1} - \bar{y}_{2.} + \bar{y}_{..} \rvert$	$\lvert y_{22} - \bar{y}_{.2} - \bar{y}_{2.} + \bar{y}_{..} \rvert$	.	.	.	$\lvert y_{2k} - \bar{y}_{.k} - \bar{y}_{2.} + \bar{y}_{..} \rvert$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Block b	$\lvert y_{b1} - \bar{y}_{.1} - \bar{y}_{b.} + \bar{y}_{..} \rvert$	$\lvert y_{b2} - \bar{y}_{.2} - \bar{y}_{b.} + \bar{y}_{..} \rvert$	.	.	.	$\lvert y_{bk} - \bar{y}_{.k} - \bar{y}_{b.} + \bar{y}_{..} \rvert$

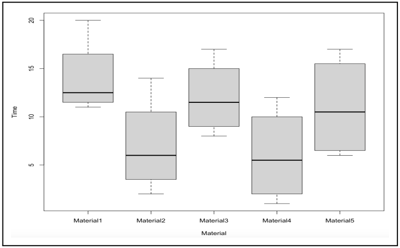
Example

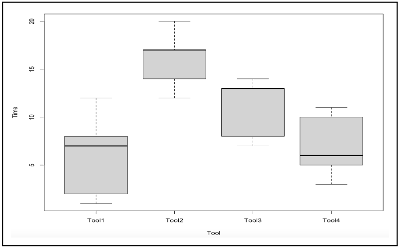
The cutting speeds of four types of tools (treatments) are being compared in an experiment. Five materials of varying degrees of hardness are to be used as experimental blocks. The data pertaining to measurements of cutting time in second is given below and in the file tools.csv.

	Tool 1	Tool 2	Tool 3	Tool 4
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

1. At the 5% significance level, test it there is no difference between the mean cutting times of the four tools and if there is a block effect.
2. If necessary, perform multiple comparison tests.
3. Perform relevant diagnostic checks for normality and homogeneity of variances.
4. Explain why the randomised block design used in this experiment is preferable to the completely randomised design

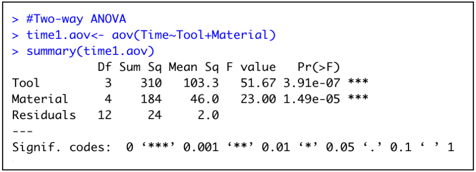
Example





- There do appear to be differences in the cutting times amongst the four types of tools.
- This is also the case for some types of materials.
- The focus is on determining if there are differences between tool types with extraneous variation as a result of material type removed.

Example - Part 1



```
> #Two-way ANOVA
> time1.aov<- aov(Time~Tool+Material)
> summary(time1.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
Tool         3    310   103.3   51.67 3.91e-07 ***
Material     4    184    46.0   23.00 1.49e-05 ***
Residuals   12     24     2.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Testing Treatment Effects:

1. Hypotheses:
   $H_0: \mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4}$
   H_1: both all $\mu_{.j}$ are equal; $j = 1, 2, 3, 4$.
2. Level of Significance: $\alpha = 0.05$
3. Test Statistic: $F_T = 51.67$
4. P-value: $p - value \approx 0.000$
5. Decision Rule: Reject H_0 if $p - value < 0.05$
6. SInce the $p - value < 0.05$, reject H_0 at the 5% level and indeed any reasonable level of significance. Conclude that here are significant differences in the cutting times of the four tool types. Hence there is a significant treatment effect.

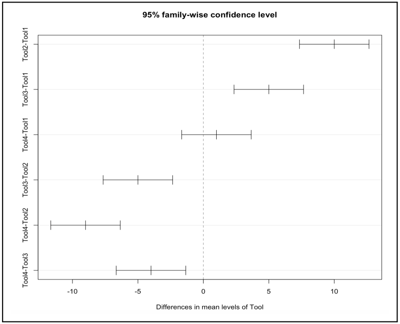
Testing Block Effects

1. Hypotheses:
   $H_0: \mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.} = \mu_{5.}$
   H_1: both all $\mu_{i.}$ are equal; $i = 1, 2, 3, 4, 5$
2. Level of Significance: $\alpha = 0.05$
3. Test Statistic: $F_B = 23.00$
4. P-value: $p - value = \approx 0.000$
5. Decision Rule: Reject H_0 if $p - value < 0.05$
6. SInce the $p - value < 0.05$, reject H_0 at the 5% level and indeed any reasonable level of significance. Conclude that here are significant differences in the mean cutting times for the five material types. Hence there is a significant block effect.

Example - Part 2

- Since we are primarily interested in differences between treatment means we perform multiple comparisons for tool types only.

```
> #Multiple Comparisons
> time1.tukey <- TukeyHSD(time1.aov)
> time1.tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Time ~ Tool + Material)

$Tool
            diff        lwr       upr     p adj
Tool2-Tool1   10   7.344534 12.655466 0.0000006
Tool3-Tool1    5   2.344534  7.655466 0.0005896
Tool4-Tool1    1  -1.655466  3.655466 0.6858866
Tool3-Tool2   -5  -7.655466 -2.344534 0.0005896
Tool4-Tool2   -9 -11.655466 -6.344534 0.0000018
Tool4-Tool3   -4  -6.655466 -1.344534 0.0036697
```



- There are significant differences in mean cutting times for all pairs except between Tool 1 and 4.

Example - Part 3

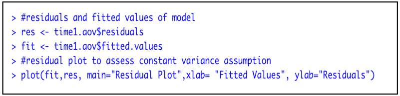
Table: Original Data Structure

	Tool 1	Tool 2	Tool 3	Tool 4		
Material 1	12	20	13	11	14	$\bar{y}_{1.}$
Material 2	2	14	7	5	7	$\bar{y}_{2.}$
Material 3	8	17	13	10	12	$\bar{y}_{3.}$
Material 4	1	12	8	3	6	$\bar{y}_{4.}$
Material 5	7	17	14	6	11	$\bar{y}_{5.}$
	6	16	11	7	10	
	$\bar{y}_{.1}$	$\bar{y}_{.2}$	$\bar{y}_{.3}$	$\bar{y}_{.4}$	$\bar{y}_{..}$	

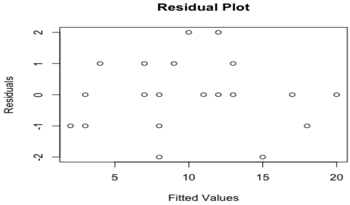
Table: Transformed Data Structure for Levene's Test

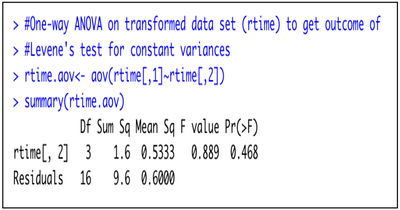
	Tool 1	Tool 2	Tool 3	Tool 4
Material 1	2	0	2	0
Material 2	1	1	1	1
Material 3	0	1	0	1
Material 4	1	0	1	0
Material 5	0	0	2	2

Example - Part 3



```
> #residuals and fitted values of model
> res <- time1.aov$residuals
> fit <- time1.aov$fitted.values
> #residual plot to assess constant variance assumption
> plot(fit,res, main="Residual Plot",xlab= "Fitted Values", ylab="Residuals")
```



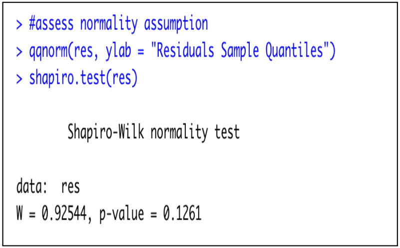


```
> #One-way ANOVA on transformed data set (rtime) to get outcome of
> #Levene's test for constant variances
> rtime.aov<- aov(rtime[,1]~rtime[,2])
> summary(rtime.aov)
           Df Sum Sq Mean Sq F value Pr(>F)
rtime[, 2]  3    1.6  0.5333   0.889  0.468
Residuals  16    9.6  0.6000
```

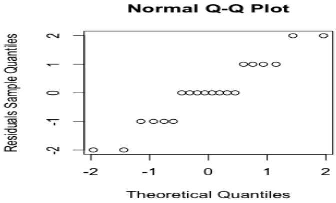
- The results of Levene's test indicates that the assumption of constant variances is valid.

- The scatterplot of residuals versus fitted values displays no particular pattern, an indication of constant variances

Example - Part 3



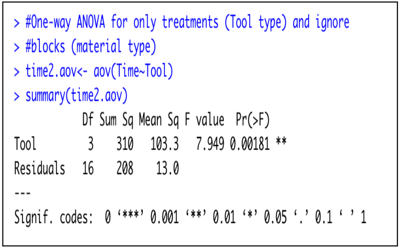
```
> #assess normality assumption
> qqnorm(res, ylab = "Residuals Sample Quantiles")
> shapiro.test(res)

        Shapiro-Wilk normality test

data:  res
W = 0.92544, p-value = 0.1261
```



- The Normal QQ plot doesn't show much deviation from the diagonal.
- The Shapiro-Wilk test indicates that the assumption of normality of the errors is valid.

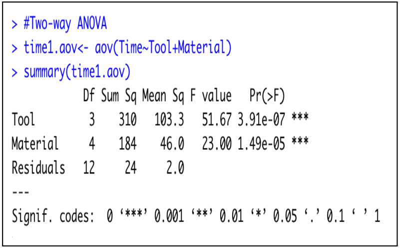
Example - Part 3

Completely Randomised Design



```
> #One-way ANOVA for only treatments (Tool type) and ignore
> #blocks (material type)
> time2.aov<- aov(Time~Tool)
> summary(time2.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)
Tool         3    310   103.3   7.949 0.00181 **
Residuals   16    208    13.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Randomised Block Design



```
> #Two-way ANOVA
> time1.aov<- aov(Time~Tool+Material)
> summary(time1.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
Tool         3    310   103.3   51.67 3.91e-07 ***
Material     4    184    46.0   23.00 1.49e-05 ***
Residuals   12     24     2.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- For the Completely Randomised Design, the mean square error is 13.0 compared to 2.0 for the Randomised Block Design.
- Thus, we can conclude that blocking has been advantageous in reducing the extraneous variability.