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

Ann Maharaj

Experimental Design: Completely Randomised Design

Introduction

- The term completely randomised design is synonymous with independent random sampling from several populations when each population is identified as a population of responses under a particular treatment.
- Suppose that out of k treatments to be studied in an experiment, Treatment 1 is applied to n_1 units, Treatment 2 to n_2 units . . . , Treatment k to n_k units.
- In a completely randomised design, n_1 selected at random from the $n = n_1 + n_2 + ... + n_k$ units are to receive Treatment 1, n_2 units randomly selected from the remaining lot are to receive Treatment 2, and continue this way, Treatment k is to be applied to the remaining k units.
- The special case of this design is the comparison of two treatments which is the comparison of two population means discussed in a previous module.

Introduction

- 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.

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

- One factor (Independent variable): Fabric (4 treatments)
- Experimental units: Dresses
- Response (dependent variable): Burn time
- We want to test there is no difference in degree of flammability for the four fabrics.

Data Structure

The data structure for the response measurements are represented by the format shown below, where y_{ij} is the i^{th} observation on Treatment j.

Table: Data Structure for the Completely Randomised Design

	Treatment 1	Treatment 2				Treatment k	
	y_{11}	y_{12}	.	.	.	y_{1k}	
	y_{21}	y_{22}	.	.	.	y_{2k}	
	.	.	.	.	.	.	
	.	.	.	.	.	.	
	.	.	.	.	.	.	
	$y_{n_1 1}$	$y_{n_2 2}$	.	.	.	$y_{n_k k}$	
Means	$\overline{y}_1$	$\overline{y}_2$	.	.	.	$\overline{y}_k$	$\overline{y}$

j^{th} treatment mean: $\overline{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$, $j = 1, 2, ..., k$

Grand mean: $\overline{y} = \frac{1}{k} \sum_{j=1}^{k} \overline{y}_j$

Partition of Total Variation

Given the basic decomposition

$$(y_{ij} - \overline{y}) = (\overline{y}_j - \overline{y}) + (y_{ij} - \overline{y}_j)$$

It can be shown that taking the squares on both side and summing each side over $i = 1, 2, \ldots, n_j$ and $j = 1, 2, ..., k$ provides the decomposition

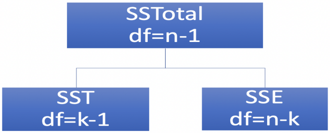
$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \overline{y})^2 = \sum_{j=1}^{k} n_j(\overline{y}_j - \overline{y})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \overline{y}_j)^2$$

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

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

$$SSTotal = SST + SSE$$

Partition of Total Variation



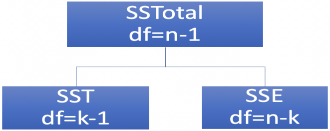
- Total Sum of Squares: $SSTotal = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \overline{y})^2$
  has degrees of freedom $df = \sum_{j=1}^{k} n_j - 1 =$ n - 1.

  $SSTotal$ is a measure of the dispersion of all individual data values in relation to the mean $\overline{y}$ of the entire data set. $\overline{y}$ is referred to as the grand mean.

- Treatment Sum of Squares: $SST = \sum_{j=1}^{k} n_j(\overline{y}_j - \overline{y})^2$
  has degrees of freedom $df = k - 1$.

  SST is a measure of dispersion of the group (treatment) means from the grand mean. It is the variation due to treatments and it is also referred to as between-group variation.

Partition of Total Variation



- Error Sum of Squares: $SSE = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$ has degrees of freedom $df = \sum_{j=1}^{k} n_j - k = n - k$.

  SSE is an aggregated measure of how much data values vary from the mean of their own sampled group. It is the variation due to random sampling (unexplained variation) and it is also referred to as within-group variation.

- SSTotal, SST and SSE all measure the amount of variation with in the data.

- The mean square (variance) associated with a component is Mean Square = Sum of Squares/df.
  - Mean Square Treatment: MST = SST/(k-1)
  - Mean Square Error: MSE = SSE/(n-k)

Population Model

- To implement a formal statistical test for no difference among treatment effects, we have to have a population model for the experiment.
- To this end we assume that the response measurements with the j^{th} treatment constitute a random sample from a normal population with a mean of μ_j and a common variance σ^2.
- The samples are assumed to be mutually independent.

Population model for comparing k treatments:

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

where $\mu = \frac{\sum_{j=1}^{k} \mu_j}{k}$ = overall mean

$\beta_j = \mu_j - \mu = j^{th}$ treatment effect, $\sum_{j=1}^{k} \beta_j = 0$

ε_{ij} = random error of the observation from the $(i, j)^{th}$ cell.
Assume ε_{ij} are independently and identically distributed as $N(0, \sigma^2)$.

Hypothesis Testing

$H_0 : \beta_1 = \beta_2 = ...\beta_k = 0$

All k treatment effects are equal to zero which implies there is

- no treatment effect
- no variation in means between the treatments
- All k population means are equal.

H_1: Not all the β_j are equal to zero, $j = 1, 2, ..., k$.

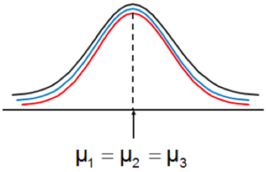
At least one of the treatments effects is not zero which implies

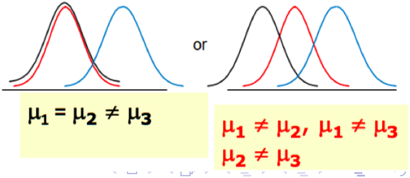
- at least one population mean is different: i.e., there is a treatment effect.
- it does not mean that all population means are different; some pairs may be the same.

Hence the null and alternative hypotheses can be expressed as

$H_0 : \mu_1 = \mu_2 = ... = \mu_k$

H_1: not all μ_j are equal; $j = 1, 2, ..., k$.





Hypothesis Testing

- 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.

- The residual mean square, $MSE = \frac{SSE}{n-k}$ which can be shown to be an unbiased estimator of the population variance σ^2, can be used as a yardstick for determining how large a treatment mean square should be before it indicates significant differences.

- Using statistical distribution theory, it can be shown that under the null hypothesis, the ratio of the treatment mean square and the residual mean square, i.e.,

$$F = \frac{MST}{MSE}$$

  follows an F-distribution with (k-1) and (n-k) degrees of freedom

- It is customary to represent the sum of squares, corresponding degrees of freedom, mean squares and F-test statistic in an Analysis of Variance table (ANOVA).

ANOVA Table

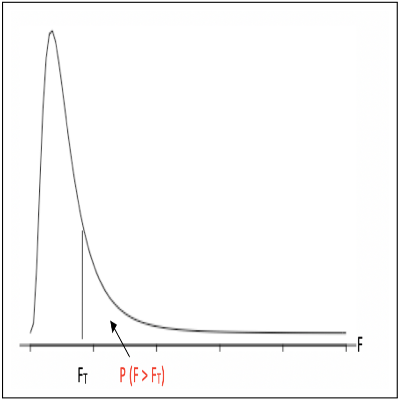
Table: ANOVA Table with F-Ratio

Source	Sum of Squares	df	Mean Square	F_T
Treatments	SST	$k-1$	$MST = \frac{SST}{k-1}$	$F_T = \frac{MST}{MSE}$
Error	SSE	$n-k$	$MSE = \frac{SSE}{n-k}$	
Total	$SSTotal$	$n-1$		

- Because there just one factor (independent variable), we refer to this setup as a One-Way ANOVA.
- Total = Treatments + Error
- The sums of squares always add up.
  - SSTotal = SST + SSE
- Degrees of freedom always add up.
  - $n-1 = (k-1) + (n-k)$

Hypothesis Testing

1. Hypotheses:
   $H_0 : \mu_1 = \mu_2 = ... = \mu_k$
   H_1: not all μ_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



Multiple Comparisons: Tukey's Test

- The F test is only the initial step in our analysis that determines if significant differences exist among treatment means.
- The next step is to detect if there are differences between the two treatments in every pair. This is called post-hoc analysis.
- 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 to use 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 $(\bar{y}_j - \bar{y}_{j'})$ is normally distributed with a mean of $(\mu_j - \mu_{j'})$ and variance $\sigma^2 \left(\frac{1}{n_j} + \frac{1}{n_{j'}} \right)$, where σ^2 is unknown and is estimated by MSE.
- Hence, it can be shown that

$$T_{HSD} = \frac{(\bar{y}_j - \bar{y}_{j'}) - (\mu_j - \mu_{j'})}{\sqrt{MSE \left(\frac{1}{n_j} + \frac{1}{n_{j'}} \right)}}$$

  follows a t-distribution with $df = n - k$
- T_{HSD} is referred to as *Tukey's Honestly Significant Difference* test statistic.

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

$$(\overline{y}_j - \overline{y}_{j'}) \pm t_{\alpha/2m, n-k} \sqrt{MSE\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)}$$

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

- 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 independent 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.

Levene's Test for Homogeneity of Variance

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$
H_1: Not all of the σ_j^2 are equal, $j = 1, 2, ..., k$

To test the null hypothesis of equal variances:

- Determine the absolute value of the differences between each value of each treatment group and the mean or the trimmed mean or the median of the group; using the median provides a robust measure of variation.
- Perform a one-way ANOVA on these absolute differences.
- The initial data structure and transformed data structure required for Levene's Test are given on the next slide.

Table: Data Structure for the Completely Randomised Design

	Treatment 1	Treatment 2				Treatment k
	y_{11}	y_{12}	.	.	.	y_{1k}
	y_{21}	y_{22}	.	.	.	y_{2k}
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	$y_{n_1 1}$	$y_{n_2 2}$	.	.	.	$y_{n_k k}$
Means	$\overline{y}_1$	$\overline{y}_2$	.	.	.	$\overline{y}_k$

Table: Transformed Data Structure for Levene's Test

Treatment 1	Treatment 2				Treatment k
$\lvert y_{11} - \overline{y}_1 \rvert$	$\lvert y_{12} - \overline{y}_2 \rvert$	.	.	.	$\lvert y_{1k} - \overline{y}_k \rvert$
$\lvert y_{21} - \overline{y}_1 \rvert$	$\lvert y_{22} - \overline{y}_2 \rvert$	.	.	.	$\lvert y_{2k} - \overline{y}_k \rvert$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$\lvert y_{n_1 1} - \overline{y}_1 \rvert$	$\lvert y_{n_2 2} - \overline{y}_2 \rvert$	.	.	.	$\lvert y_{n_k k} - \overline{y}_k \rvert$

Example

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 and are in the file fabric.csv.

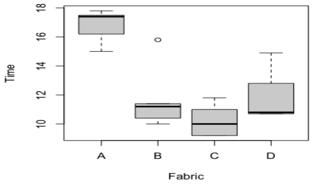
Fabric A	Fabric B	Fabric C	Fabric D
17.8	11.2	11.8	14.9
16.2	11.4	11.0	10.8
17.5	15.8	10.0	12.8
17.4	10.0	9.2	10.7
15.0	10.4	9.2	10.7

1. Test the null hypothesis of no difference in degree of flammability of the four fabrics. Use the 5% level of significance.
2. If the null hypothesis in Part (1) is rejected, perform post-hoc analysis, using Tukey's multiple comparisons tests.
3. Perform diagnostic checks on the assumptions of normality and homogeneity of variances.

Example - Part 1

It is clear from the boxplots there are some differences in the degree of flammability of the four fabrics.

1. Hypotheses:
   $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
   H_1: Not all the μ_j are equal
   $j = 1, 2, 3, 4$
2. Level of Significance: $\alpha = 0.05$
3. Test Statistic: $F_T = 13.89$
4. P-value: p-value $= 0.0001$
5. Decision Rule: Reject H_0 if
   $p - value < \alpha$
6. Since $p - value < 0.05$, reject H_0 at the 5% level and indeed any reasonable level of significance.

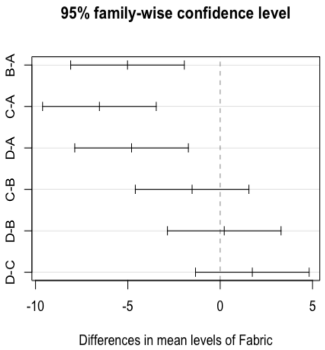



```
> #One-way ANOVA
> fab.aov<- aov(Time~Fabric)
> summary(fab.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
Fabric       3 120.50   40.17   13.89 0.000102 ***
Residuals   16  46.26    2.89
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclude there are significant differences in degree of flammability of the four fabrics.

Example - Part 2

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

Fit: aov(formula = Time ~ Fabric)

$Fabric
     diff        lwr        upr     p adj
B-A -5.02 -8.09676 -1.94324 0.0013227
C-A -6.54 -9.61676 -3.46324 0.0000851
D-A -4.80 -7.87676 -1.72324 0.0019981
C-B -1.52 -4.59676  1.55676 0.5094118
D-B  0.22 -2.85676  3.29676 0.9968426
D-C  1.74 -1.33676  4.81676 0.3968476

> plot(fab.tukey)
```

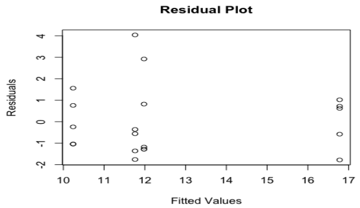


There are significant differences in the degree of flammability between

- Fabrics A and B
- Fabrics A and C
- Fabrics A and D

Example - Part 3

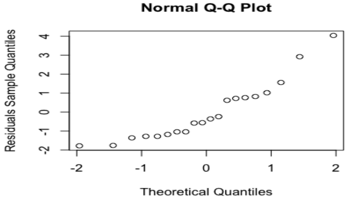
```
> #residuals and fitted values of model
> res  <- fab.aov$residuals
> fit<- fab.aov$fitted.values
> #residual plot to assess constant variance assumption
> plot(fit,res, main="Residual Plot",xlab= "Fitted Values", ylab="Residuals")
```




```
> #Test for constant variance
> leveneTest(fab.aov)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.1788 0.9092
      16
```

- The scatterplot of residuals versus fitted values does not display a fan-shaped pattern; hence this is which an indication of constant variances.
- The results of Levene's test indicate that the assumption of constant variances is valid.

Example - Part 3




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

        Shapiro-Wilk normality test

data:  res
W = 0.88926, p-value = 0.02606
```

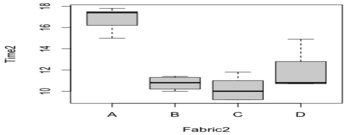
- The Normal QQ plot does show some deviation from the diagonal.
- The Shapiro-Wilk test indicates that the assumption of normality of the errors is not valid.

We next remove the outlier to determine if there are any changes to the outcome of the tests.

Example - without outlier

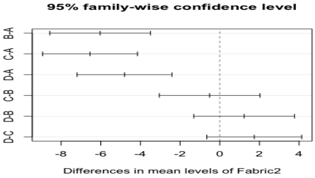
On removal of the outlier, we observe:

- While the F-value is much larger and the corresponding p-value is much smaller, there is no change in the outcome of the F-test.
- While the confidence intervals are narrower and p-values associated with significant difference are much smaller, there is no change in the outcome of the Tukey HSD test.



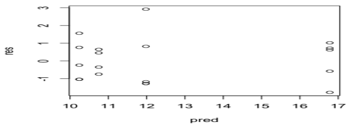

```
> #Multiple Comparisons
> fabb.tukey <- TukeyHSD(fabb.aov)
> fabb.tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Time2 ~ Fabric2)

$Fabric2
     diff        lwr       upr     p adj
B-A -6.03 -8.5684842 -3.491516 0.0000297
C-A -6.54 -8.9333059 -4.146694 0.0000057
D-A -4.80 -7.1933059 -2.406694 0.0001910
C-B -0.51 -3.0484842  2.028484 0.9368468
D-B  1.23 -1.3084842  3.768484 0.5203006
D-C  1.74 -0.6533059  4.133306 0.1991623

> plot(fabb.tukey)
```

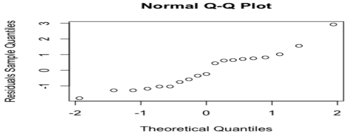
```
> #Analysis without outlier
> #One-way ANOVA
> fabb.aov<- aov(Time2~Fabric2)
> summary(fabb.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
Fabric2      3 130.72   43.57   25.28 4.08e-06 ***
Residuals   15  25.86    1.72
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Example-without outlier

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






```
> #Test for constant variance
> leveneTest(fabb.aov)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.3139 0.8151
      15
```

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

        Shapiro-Wilk normality test

data:  res
W = 0.93832, p-value = 0.2459
```

- The residual plot the results of Levene's test indicates that the assumption of constant variances is valid.
- The Shapiro-Wilk test indicates that the assumption of normality of the errors is now valid, and the deviation from the diagonal in the Normal Q-Q plot is less evident.