

Analysis of Variance

Dr. Lee Fawcett



Introduction

ANOVA is an important technique for interpreting the overall variability within an experiment or other study.

It allows an assessment of the significance of the effects of various ‘factors’ and permits an estimate to be made of underlying random variability.

The simplest version is **one-way ANOVA**, in which we compare several population means in a single comparative experiment.
This can be regarded as an extension of the two-sample t -test.

Introduction

The term analysis of variance is often a source of confusion: In spite of its name, ANOVA is concerned with differences between **means** of groups, not differences between variances.

Introduction

The term analysis of variance is often a source of confusion: In spite of its name, ANOVA is concerned with differences between **means** of groups, not differences between variances.

The name analysis of variance comes from the way the procedure uses variances to decide whether the means are different.

Illustrative examples

1. In an experiment to determine the relative effectiveness of four brands of sun cream (Miksa *et al.*, 2016), six volunteers each tried four creams (A, B, C and D) on various parts of their legs.

They were then subjected to ultra-violet radiation and an estimate of the degree of burning was made (higher figures indicate greater burning).

The results are given below, along with some summaries:

							Total	mean	s.d.
A	8.6	8.8	7.9	6.4	6.8	8.1	46.6	7.77	0.969
B	5.9	6.2	6.6	6.6	7.4	6.3	39.0	6.50	0.514
C	6.7	9.0	8.4	8.6	6.9	7.2	46.8	7.80	0.982
D	8.2	7.4	8.4	6.9	7.5	7.5	45.9	7.65	0.554

Illustrative examples

2. The following measurements were made of the width (in mm) of cuckoo's eggs laid in the nest of various types of birds.

Robin	16.0	16.3	15.9	17.0	17.1	16.0	16.6	16.4
	16.9	16.4	16.1	17.0	17.2	16.2	16.9	15.2
Hedge sparrow	17.0	16.1	16.9	16.9	15.8	16.7	17.3	17.0
	17.5	17.5	16.2	16.5	16.2	17.1		
Wren	15.0	16.0	16.0	15.7	16.2	15.9	15.9	16.0
	16.2	16.0	15.5	16.0	15.9	15.5	15.9	
Reed Warbler	16.9	17.1	17.0	16.1	16.5	16.5	16.1	16.5
	17.9	16.5						
Yellowhammer	16.0	16.0	16.1	17.2	16.5	17.0	17.0	17.0
	16.8							

Illustrative examples

Some summaries are given below:

	n_i	T_i	$\bar{x}_i$	s_i
Robin	16	263.2	16.45	0.548
Hedge sparrow	14	234.7	16.76	0.534
Wren	15	237.7	15.85	0.311
Reed Warbler	10	167.1	16.71	0.538
Yellowhammer	9	149.6	16.62	0.482

Illustrative examples

Example **1** is an example of a **planned experiment**, while example **2** is an example of an **observational study**.

Illustrative examples

Example **1** is an example of a **planned experiment**, while example **2** is an example of an **observational study**.

An observational study is an investigation in which processes are observed in their natural environment without direct intervention by the investigator. It is difficult to draw authoritative conclusions because extraneous random variation may obscure or even distort the effects of primary interest.

Illustrative examples

Example **1** is an example of a **planned experiment**, while example **2** is an example of an **observational study**.

An observational study is an investigation in which processes are observed in their natural environment without direct intervention by the investigator. It is difficult to draw authoritative conclusions because extraneous random variation may obscure or even distort the effects of primary interest.

A planned experiment is an investigation in which the sources of extraneous random variation are, as far as possible, identified beforehand and accommodated within the design and analysis of the experiment.

Illustrative examples

Example **1** is an example of a **planned experiment**, while example **2** is an example of an **observational study**.

An observational study is an investigation in which processes are observed in their natural environment without direct intervention by the investigator. It is difficult to draw authoritative conclusions because extraneous random variation may obscure or even distort the effects of primary interest.

A planned experiment is an investigation in which the sources of extraneous random variation are, as far as possible, identified beforehand and accommodated within the design and analysis of the experiment.

Both **1** and **2** may be analysed using **one-way ANOVA**.

One-way ANOVA

Method outline

To answer questions such as those posed in the previous examples, it is tempting to look at all pairs of groups and perform a two-sample t -test each time (multiple t -tests).

Method outline

To answer questions such as those posed in the previous examples, it is tempting to look at all pairs of groups and perform a two-sample t -test each time (multiple t -tests).

However, though this is sometimes done in disciplines such as biology and psychology, this is not a valid procedure:

Method outline

To answer questions such as those posed in the previous examples, it is tempting to look at all pairs of groups and perform a two-sample t -test each time (multiple t -tests).

However, though this is sometimes done in disciplines such as biology and psychology, this is not a valid procedure:

- If there are lots of groups to compare, there will be lots of t -tests, and so there is the possibility of a **high number of “false positives”** – such a procedure is “less powerful” at detecting any real differences
- Tests would not be mutually independent

Method outline

The idea behind ANOVA is to obtain two estimates of the underlying variance:

- One based on the variability *between* sample means (known as the **between treatments mean square**)
- The other based on the average variability *within* samples (known as the **residual mean square**)

Method outline

The idea behind ANOVA is to obtain two estimates of the underlying variance:

- One based on the variability *between* sample means (known as the **between treatments mean square**)
- The other based on the average variability *within* samples (known as the **residual mean square**)

The first is, essentially, the sample variance of the group means. The second is the average of the group variances, with more weight given to bigger samples.

Method outline

It turns out that if there are no real differences between our groups, then the **variance ratio**

$$VR = \frac{\text{Between treatments mean square}}{\text{Residual mean square}}$$

should be close to one.

If the groups really are different, then this ratio should be bigger than one. The more differences between groups there are, the bigger the variance ratio will get.

Test structure

Suppose we have k populations and hence k samples.

In population i we have random variable $X_i \sim N(\mu_i, \sigma_i^2)$ and there are n_i observations: $x_{i1}, x_{i2}, \dots, x_{in_i}$ (and $N = \sum n_i$).

The **assumptions** are as for the two-sample t -test, i.e.

- **Normality:**

Test structure

Suppose we have k populations and hence k samples.

In population i we have random variable $X_i \sim N(\mu_i, \sigma_i^2)$ and there are n_i observations: $x_{i1}, x_{i2}, \dots, x_{in_i}$ (and $N = \sum n_i$).

The **assumptions** are as for the two-sample t -test, i.e.

- **Normality**: Histograms, Q-Q plots

Test structure

Suppose we have k populations and hence k samples.

In population i we have random variable $X_i \sim N(\mu_i, \sigma_i^2)$ and there are n_i observations: $x_{i1}, x_{i2}, \dots, x_{in_i}$ (and $N = \sum n_i$).

The **assumptions** are as for the two-sample t -test, i.e.

- **Normality**: Histograms, Q-Q plots
- **Independence**:

Test structure

Suppose we have k populations and hence k samples.

In population i we have random variable $X_i \sim N(\mu_i, \sigma_i^2)$ and there are n_i observations: $x_{i1}, x_{i2}, \dots, x_{in_i}$ (and $N = \sum n_i$).

The **assumptions** are as for the two-sample t -test, i.e.

- **Normality**: Histograms, Q-Q plots
- **Independence**: Good experimental design!

Test structure

Suppose we have k populations and hence k samples.

In population i we have random variable $X_i \sim N(\mu_i, \sigma_i^2)$ and there are n_i observations: $x_{i1}, x_{i2}, \dots, x_{in_i}$ (and $N = \sum n_i$).

The **assumptions** are as for the two-sample t -test, i.e.

- **Normality**: Histograms, Q-Q plots
- **Independence**: Good experimental design!
- **Homogeneity** (common variance, i.e. $\sigma_i^2 = \sigma^2$ for all i):

Test structure

Suppose we have k populations and hence k samples.

In population i we have random variable $X_i \sim N(\mu_i, \sigma_i^2)$ and there are n_i observations: $x_{i1}, x_{i2}, \dots, x_{in_i}$ (and $N = \sum n_i$).

The **assumptions** are as for the two-sample t -test, i.e.

- **Normality**: Histograms, Q-Q plots
- **Independence**: Good experimental design!
- **Homogeneity** (common variance, i.e. $\sigma_i^2 = \sigma^2$ for all i):
Check that no standard deviation is more than twice any other

Test structure

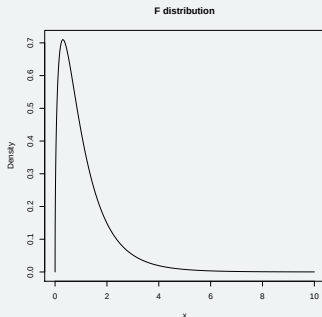
The **null hypothesis** is

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

and we assess this in the usual way – via a p -value associated with the test statistic

$$VR \sim F_{k-1, N-k},$$

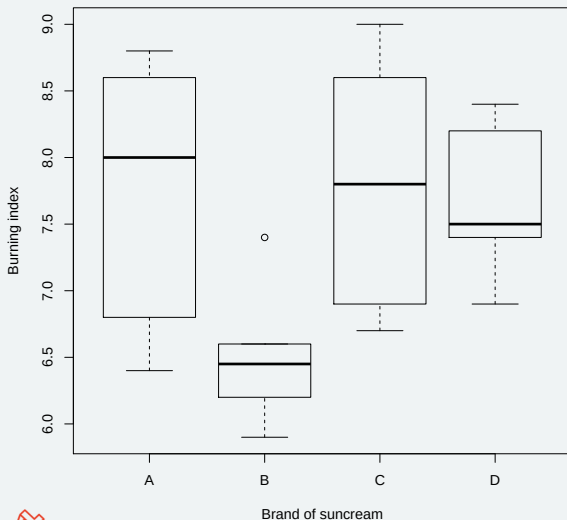
where the F distribution looks like that shown in the plot below.



Sun cream example

```
1 > suncream
2   A   B   C   D
3 1 8.6 5.9 6.7 8.2
4 2 8.8 6.2 9.0 7.4
5 3 7.9 6.6 8.4 8.4
6 4 6.4 6.6 8.6 6.9
7 5 6.8 7.4 6.9 7.5
8 6 8.1 6.3 7.2 7.5
9
10 > suncream2 = stack(suncream)
11 > head(suncream2)
12   values ind
13 1    8.6   A
14 2    8.8   A
15 3    7.9   A
16 4    6.4   A
17 5    6.8   A
18 6    8.1   A
19
20 > boxplot(suncream2[,1]~suncream2[,2], main="", xlab="Brand of suncream", ylab="
    Burning index")
21 > boxplot(suncream2$values~suncream2$ind, main="", xlab="Brand of suncream", ylab="
    Burning index")
```

Sun cream example



Sun cream example

```
1 > fit = aov(suncream2[,1] ~ suncream2[,2])
2 > summary(fit)
3      Df Sum Sq Mean Sq F value Pr(>F)
4 suncream2[, 2] 3  6.981  2.3271   3.763 0.0272 *
5 Residuals    20 12.368  0.6184
6 ---
7 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We compare $VR = 3.763$ to the $F_{3,20}$ distribution:

Sun cream example

```
1 > fit = aov(suncream2[,1] ~ suncream2[,2])
2 > summary(fit)
3      Df Sum Sq Mean Sq F value Pr(>F)
4 suncream2[, 2] 3  6.981  2.3271   3.763 0.0272 *
5 Residuals    20 12.368  0.6184
6 ---
7 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We compare $VR = 3.763$ to the $F_{3,20}$ distribution:

- Reject H_0 at the 5% level of significance ($p = 2.72\%$)

Sun cream example

```
1 > fit = aov(suncream2[,1] ~ suncream2[,2])
2 > summary(fit)
3      Df Sum Sq Mean Sq F value Pr(>F)
4 suncream2[, 2] 3  6.981  2.3271   3.763 0.0272 *
5 Residuals    20 12.368  0.6184
6 ---
7 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We compare $VR = 3.763$ to the $F_{3,20}$ distribution:

- Reject H_0 at the 5% level of significance ($p = 2.72\%$)
- There is evidence to suggest the (average) effect of the sun creams is not the same

Sun cream example

```
1 > fit = aov(suncream2[,1] ~ suncream2[,2])
2 > summary(fit)
3      Df Sum Sq Mean Sq F value Pr(>F)
4 suncream2[, 2] 3  6.981  2.3271   3.763 0.0272 *
5 Residuals    20 12.368  0.6184
6 ---
7 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We compare $VR = 3.763$ to the $F_{3,20}$ distribution:

- Reject H_0 at the 5% level of significance ($p = 2.72\%$)
- There is evidence to suggest the (average) effect of the sun creams is not the same
- But: Where do the differences lie?

One-way ANOVA: Multiple comparisons

If a one-way ANOVA suggests a rejection of H_0 – that all population means are the same – the obvious question to ask is:

“Are all the means different to each other? If not, where do the differences lie?”

- Many follow-up procedures have been devised; the most popular is **Tukey's Highest Significant Difference** method

One-way ANOVA: Multiple comparisons

If a one-way ANOVA suggests a rejection of H_0 – that all population means are the same – the obvious question to ask is:

“Are all the means different to each other? If not, where do the differences lie?”

- Many follow-up procedures have been devised; the most popular is **Tukey’s Highest Significant Difference** method
- Tukey’s method is based on the calculation of a ‘yardstick’, and a comparison of pairwise sample means to this yardstick
- As always, we can let our favourite software take care of all the calculations – we just need to interpret the output

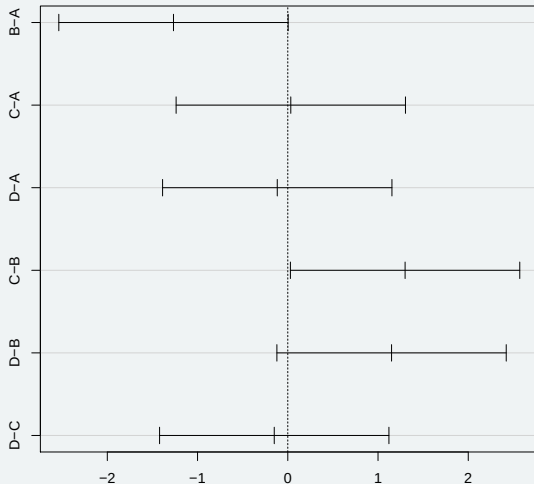
Sun cream example

```
1 > (tukeytest = TukeyHSD(fit))
2   Tukey multiple comparisons of means
3     95% family-wise confidence level
4
5 Fit: aov(formula = suncream2[, 1] ~ suncream2[, 2])
6
7 $`suncream2[, 2]`
8      diff          lwr          upr      p adj
9 B-A -1.26666667 -2.53745423 0.004120898 0.0509352
10 C-A  0.03333333 -1.23745423 1.304120898 0.9998530
11 D-A -0.11666667 -1.38745423 1.154120898 0.9938573
12 C-B  1.30000000  0.02921243 2.570787565 0.0438088
13 D-B  1.15000000 -0.12078757 2.420787565 0.0849351
14 D-C -0.15000000 -1.42078757 1.120787565 0.9871822
```

Sun cream example

```
1 > plot(tukeytest)
```

95% family-wise confidence level



Conclusions

- We rejected H_0 at the 5% level of significance

Conclusions

- We rejected H_0 at the 5% level of significance
- There was evidence to suggest a difference (on average) between the four different sun creams
- In fact, the only significant difference was between creams B and C; no real differences were found between any other pairs

Conclusions

- We rejected H_0 at the 5% level of significance
- There was evidence to suggest a difference (on average) between the four different sun creams
- In fact, the only significant difference was between creams B and C; no real differences were found between any other pairs
- Anything else we might have done?

Randomised Blocks Designs

Illustrative example

In this example, not only have we got different treatments, but each treatment is split up into **blocks**.

Illustrative example

In this example, not only have we got different treatments, but each treatment is split up into **blocks**.

The data below are the results of a **randomised block design** and give the yields (in kilograms) of ten strains of carrots grown in four different plots.

		Plot				Treatment means
		1	2	3	4	
Treatment (strain of carrot)	A	30.0	28.0	28.3	31.7	29.5
	B	33.5	29.6	31.9	29.8	31.2
	C	32.8	29.0	31.2	25.8	29.7
	D	22.2	25.2	24.7	22.3	23.6
	E	30.2	33.1	30.1	27.8	30.3
	F	30.5	30.0	28.5	31.8	30.2
	G	32.6	31.7	30.1	30.8	31.3
	H	30.0	29.7	27.6	29.5	29.2
	I	36.7	35.7	35.6	32.4	35.1
	J	28.5	35.0	32.0	28.1	30.9
Block means		30.7	30.7	30.0	29.0	30.1

Illustrative example

The analysis is exactly the same as the one-way analysis, but this time we have two sets of population means to check:

treatment means and **block means**.

Illustrative example

The analysis is exactly the same as the one-way analysis, but this time we have two sets of population means to check:

treatment means and **block means**.

We want to test for any differences in mean yield between strains of carrot, i.e.

$$H_0 : \mu_A = \mu_B = \cdots = \mu_J.$$

Illustrative example

The analysis is exactly the same as the one-way analysis, but this time we have two sets of population means to check:

treatment means and **block means**.

We want to test for any differences in mean yield between strains of carrot, i.e.

$$H_0 : \mu_A = \mu_B = \cdots = \mu_J.$$

However, we should also check to see if the blocks have had any effect on mean yield, and so we also test

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

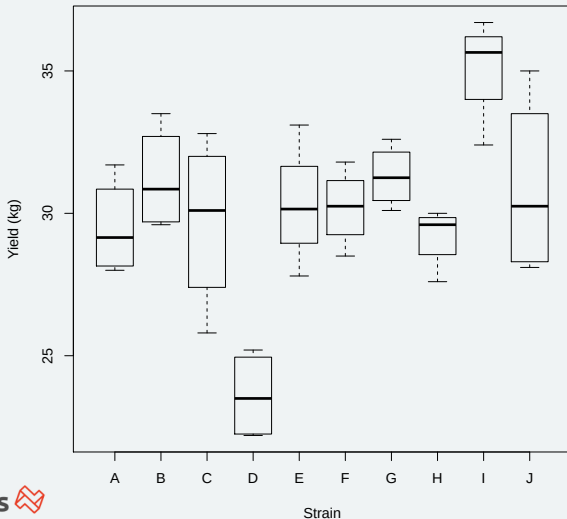
Assumptions are checked in the usual way.

Illustrative example

```
1 > head(carrots)
2   Yield Strain Plot
3   1  30.0     A    1
4   2  28.0     A    2
5   3  28.3     A    3
6   4  31.7     A    4
7   5  33.5     B    1
8   6  29.6     B    2
9
10 > carrots$Plot = as.factor(carrots$Plot)
```

Illustrative example

```
1 > boxplot(carrots$Yield~carrots$Strain, xlab="Strain", ylab="Yield (kg)")
```



Illustrative example

```
1 > aggregate(carrots$Yield, list(carrots$Strain), FUN=mean)
2   Group.1      x
3   1      A 29.5
4   2      B 31.2
5   3      C 29.7
6   4      D 23.6
7   5      E 30.3
8   6      F 30.2
9   7      G 31.3
10  8      H 29.2
11  9      I 35.1
12 10      J 30.9
13
14 > aggregate(carrots$Yield, list(carrots$Strain), FUN=sd)
15   Group.1      x
16  1      A 1.710750
17  2      B 1.852926
18  3      C 3.030951
19  4      D 1.572683
20  5      E 2.171021
21  6      F 1.363818
22  7      G 1.086278
23  8      H 1.086278
24  9      I 1.867262
25 10      J 3.246537
```

Illustrative example

```
1 > aggregate(carrots$Yield, list(carrots$Plot), FUN=mean)
2   Group.1      x
3   1      1 30.7
4   2      2 30.7
5   3      3 30.0
6   4      4 29.0
7
8 > aggregate(carrots$Yield, list(carrots$Plot), FUN=sd)
9   Group.1      x
10  1      1 3.803799
11  2      2 3.221111
12  3      3 2.970597
13  4      4 3.130495
```

Illustrative example

Some comments so far

- Yields for strain D seem very low, and for strain I quite high, relative to yields for all other strains (which all seem similar)

Illustrative example

Some comments so far

- Yields for strain D seem very low, and for strain I quite high, relative to yields for all other strains (which all seem similar)
- There does not appear to be a difference in average yields across the different plots (blocks)

Illustrative example

Some comments so far

- Yields for strain D seem very low, and for strain I quite high, relative to yields for all other strains (which all seem similar)
- There does not appear to be a difference in average yields across the different plots (blocks)
- Most boxplots indicate reasonable symmetry (and perhaps **Normality**) in the yield distributions, perhaps with the exception of strains A, D, G and I – although this is difficult to tell with such small samples

Illustrative example

Some comments so far

- Yields for strain D seem very low, and for strain I quite high, relative to yields for all other strains (which all seem similar)
- There does not appear to be a difference in average yields across the different plots (blocks)
- Most boxplots indicate reasonable symmetry (and perhaps **Normality**) in the yield distributions, perhaps with the exception of strains A, D, G and I – although this is difficult to tell with such small samples
- There could be an issue with assuming equal variances across the 10 strains, with some standard deviations more than twice some others

Illustrative example

```
1 > fit = aov(carrots$Yield ~ carrots$Strain+carrots$Plot)
2 > summary(fit)
3
4      Df Sum Sq Mean Sq F value    Pr(>F)
5 carrots$Strain  9  287.7    31.96   8.335 7.93e-06 ***
6 carrots$Plot    3   19.4     6.47   1.686  0.193
7 Residuals      27  103.5     3.83
8 ---
9 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Illustrative example

The ANOVA table tells us that:

- We **reject** the 'main' null hypothesis at the 0.1% level of significance...

Illustrative example

The ANOVA table tells us that:

- We **reject** the 'main' null hypothesis at the 0.1% level of significance...
- ... there is **overwhelming evidence** to reject the claim that the average yields are the same across all strains of carrot

Illustrative example

The ANOVA table tells us that:

- We **reject** the 'main' null hypothesis at the 0.1% level of significance...
- ... there is **overwhelming evidence** to reject the claim that the average yields are the same across all strains of carrot
- We **retain** the null hypothesis associated with the blocking by plot...

Illustrative example

The ANOVA table tells us that:

- We **reject** the 'main' null hypothesis at the 0.1% level of significance...
- ... there is **overwhelming evidence** to reject the claim that the average yields are the same across all strains of carrot
- We **retain** the null hypothesis associated with the blocking by plot...
- ... effects of the strain of carrots aside, the plots themselves have had no impact upon the average yield of the carrots

Illustrative example

The ANOVA table tells us that:

- We **reject** the 'main' null hypothesis at the 0.1% level of significance...
- ... there is **overwhelming evidence** to reject the claim that the average yields are the same across all strains of carrot
- We **retain** the null hypothesis associated with the blocking by plot...
- ... effects of the strain of carrots aside, the plots themselves have had no impact upon the average yield of the carrots

So we must now investigate further, to see where the differences in yields lie!

Illustrative example

```
1 > fit = aov(carrots$Yield~carrots$Strain)
2 > (tukeytest = TukeyHSD(fit))
3 Tukey multiple comparisons of means
4 95% family-wise confidence level
5
6 Fit: aov(formula = carrots$Yield ~ carrots$Strain)
7
8 $'carrots$Strain'
9      diff      lwr      upr      p adj
10 B-A  1.7   -3.18287942  6.5828794 0.9687681
11 C-A  0.2   -4.68287942  5.0828794 1.0000000
12 D-A -5.9  -10.78287942 -1.0171206 0.0087628
13 E-A  0.8   -4.08287942  5.6828794 0.9998894
14 F-A  0.7   -4.18287942  5.5828794 0.9999639
15 G-A  1.8   -3.08287942  6.6828794 0.9556236
16 H-A -0.3   -5.18287942  4.5828794 1.0000000
17 I-A  5.6    0.71712058 10.4828794 0.0149321
18 J-A  1.4   -3.48287942  6.2828794 0.9914560
19 C-B -1.5   -6.38287942  3.3828794 0.9862283
20 D-B -7.6  -12.48287942 -2.7171206 0.0003631
21 E-B -0.9   -5.78287942  3.9828794 0.9997090
22 F-B -1.0   -5.88287942  3.8828794 0.9993216
23 G-B  0.1   -4.78287942  4.9828794 1.0000000
24 H-B -2.0   -6.88287942  2.8828794 0.9186875
25 I-B  3.9   -0.98287942  8.7828794 0.2091322
26 J-B -0.3   -5.18287942  4.5828794 1.0000000
27 D-C -6.1  -10.98287942 -1.2171206 0.0061004
28 E-C  0.6   -4.28287942  5.4828794 0.9999903
29 F-C  0.5   -4.38287942  5.3828794 0.9999980
30 G-C  1.6   -3.28287942  6.4828794 0.9788231
31 H-C -0.5   -5.38287942  4.3828794 0.9999980
32 I-C  5.4    0.51712058 10.2828794 0.0211333
33 J-C  1.2   -3.68287942  6.0828794 0.9972209
34 E-D  6.7    1.81712058 11.5828794 0.0020076
```

Conclusions

- Average yields are significantly different between strain D and every other strain

Conclusions

- Average yields are significantly different between strain D and every other strain
- Three main groupings have emerged, **within which** average yields *are not* significantly different:

Conclusions

- Average yields are significantly different between strain D and every other strain
- Three main groupings have emerged, **within which** average yields *are not* significantly different:
 - Strains (B, G, I, J)
 - Strains (A, B, C, E, F, G, H, J)
 - Strain (D)

One/Two-way ANOVA

We will now break for a problem session covering this section.

Please go to **Problem Session 1** of **Lesson 6**.

See you in a bit!

Factorial Experiments

Illustrative example

The weight gain for male rats under six feeding treatments was studied in a **completely randomised experiment**.

The factors were:

(3 levels)	Source of protein:	1 = Beef	2 = Cereal	3 = Pork
(2 levels)	Level of protein:	1 = High	2 = Low	

The experimenter is particularly interested in comparing mean weight gains among sources of protein and how these comparisons differ between levels of protein.

Illustrative example

Note that there are **six possible source/level combinations**: Beef/High, Beef/Low, Cereal/High, Cereal/Low, Pork/High and Pork/Low.

Illustrative example

Note that there are **six possible source/level combinations**: Beef/High, Beef/Low, Cereal/High, Cereal/Low, Pork/High and Pork/Low.

Ten rats were randomly allocated to each of these source/level combinations:

		Source		
		Beef	Cereal	Pork
Level	High	10	10	10
	Low	10	10	10

Illustrative example

Such a design falls into the realms of **Two-way ANOVA**.

- We have **two factors** that might impact upon the main **response variable**

Illustrative example

Such a design falls into the realms of **Two-way ANOVA**.

- We have **two factors** that might impact upon the main **response variable**
 - There might be differences in average weight gain depending on the **source** of protein

Illustrative example

Such a design falls into the realms of **Two-way ANOVA**.

- We have **two factors** that might impact upon the main **response variable**
 - There might be differences in average weight gain depending on the **source** of protein
 - There might also be differences in average weight gain depending on the **level** of protein

Illustrative example

Such a design falls into the realms of **Two-way ANOVA**.

- We have **two factors** that might impact upon the main **response variable**
 - There might be differences in average weight gain depending on the **source** of protein
 - There might also be differences in average weight gain depending on the **level** of protein
- This leads to two null hypotheses:
 - Source: $H_0 : \mu_B = \mu_C = \mu_P$

Illustrative example

Such a design falls into the realms of **Two-way ANOVA**.

- We have **two factors** that might impact upon the main **response variable**
 - There might be differences in average weight gain depending on the **source** of protein
 - There might also be differences in average weight gain depending on the **level** of protein
- This leads to two null hypotheses:
 - Source: $H_0 : \mu_B = \mu_C = \mu_P$
 - Level: $H_0 : \mu_H = \mu_L$

Illustrative example

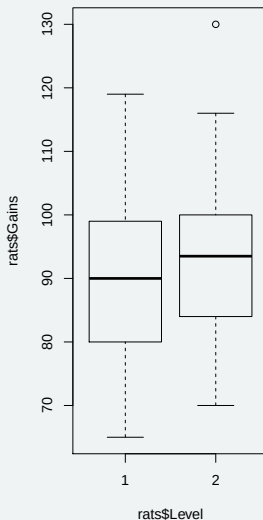
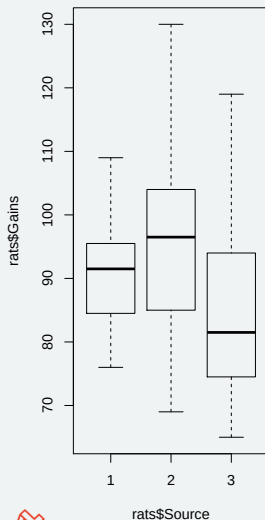
Such a design falls into the realms of **Two-way ANOVA**.

- We have **two factors** that might impact upon the main **response variable**
 - There might be differences in average weight gain depending on the **source** of protein
 - There might also be differences in average weight gain depending on the **level** of protein
- This leads to two null hypotheses:
 - Source: $H_0 : \mu_B = \mu_C = \mu_P$
 - Level: $H_0 : \mu_H = \mu_L$
- We might also want to check for an **interaction**: for example, any dependence of weight gain on source of protein might itself depend on the level of protein given
 - Interaction:
 H_0 : There is no interaction between source and level

Checking Normality

```
1 > head(rats)
2   Level Source Gains
3   1      1      1  109
4   2      2      1  108
5   3      1      2   94
6   4      2      2  130
7   5      1      3   75
8   6      2      3   94
9
10 > rats$Source = as.factor(rats$Source)
11 > rats$Level = as.factor(rats$Level)
12
13 > par(mfrow=c(1,2))
14 > boxplot(rats$Gains~rats$Level)
15 > boxplot(rats$Gains~rats$Source)
```

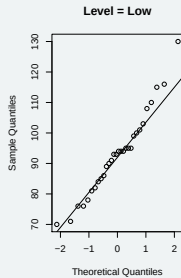
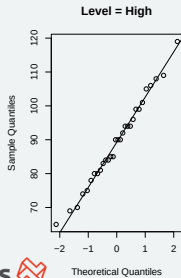
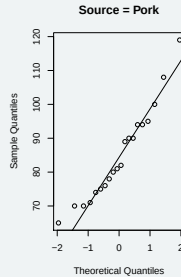
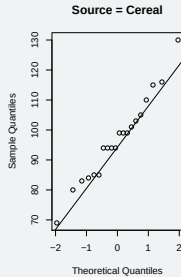
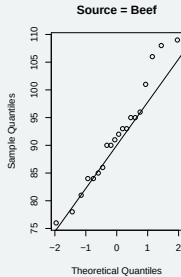
Checking Normality



Checking Normality

```
1 > par(mfrow=c(2,3))
2
3 > qqnorm(rats$Gains[rats$Source==1], main="Source = Beef")
4 > qqline(rats$Gains[rats$Source==1])
5 > qqnorm(rats$Gains[rats$Source==2], main="Source = Cereal")
6 > qqline(rats$Gains[rats$Source==2])
7 > qqnorm(rats$Gains[rats$Source==3], main="Source = Pork")
8 > qqline(rats$Gains[rats$Source==3])
9
10 > qqnorm(rats$Gains[rats$Level==1], main="Level = High")
11 > qqline(rats$Gains[rats$Level==1])
12 > qqnorm(rats$Gains[rats$Level==2], main="Level = Low")
13 > qqline(rats$Gains[rats$Level==2])
```

Checking Normality



Checking equality of variances

```
1 > aggregate(rats$Gains, list(rats$Source), FUN = sd)
2   Group.1      x
3   1      1  9.320464
4   2      2 14.203317
5   3      3 13.926404
6 > aggregate(rats$Gains, list(rats$Level), FUN = sd)
7   Group.1      x
8   1      1 13.00703
9   2      2 13.72572
```

Interpreting R output

```
1 > fit = aov(rats$Gains~rats$Source+rats$Level+rats$Source*rats$Level)
2 > summary(fit)
3
4      Df Sum Sq Mean Sq F value Pr(>F)
5 rats$Source      2   1422    710.9   4.856 0.0115 *
6 rats$Level       1    220    220.4   1.506 0.2251
7 rats$Source:rats$Level 2   1043    521.7   3.564 0.0352 *
8 Residuals      54   7905    146.4
9
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusions

Source of protein

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Level of protein

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Level of protein

- $p = 22.51\%$, so we **retain** H_0

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Level of protein

- $p = 22.51\%$, so we **retain** H_0
- There is no evidence to suggest a difference in mean weight gain between the two levels of protein (High/Low)

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Level of protein

- $p = 22.51\%$, so we **retain** H_0
- There is no evidence to suggest a difference in mean weight gain between the two levels of protein (High/Low)

Interaction

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Level of protein

- $p = 22.51\%$, so we **retain** H_0
- There is no evidence to suggest a difference in mean weight gain between the two levels of protein (High/Low)

Interaction

- $p = 3.52\%$, so we **reject** H_0

Conclusions

Source of protein

- $p = 1.15\%$, so we **reject** H_0 at the 5% level of significance
- There is evidence to suggest a difference in mean weight gains between the three sources of protein (Beef/Cereal/Pork)

Level of protein

- $p = 22.51\%$, so we **retain** H_0
- There is no evidence to suggest a difference in mean weight gain between the two levels of protein (High/Low)

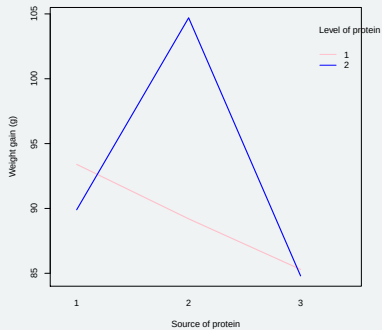
Interaction

- $p = 3.52\%$, so we **reject** H_0
- There is evidence in the data of an interaction between source and level

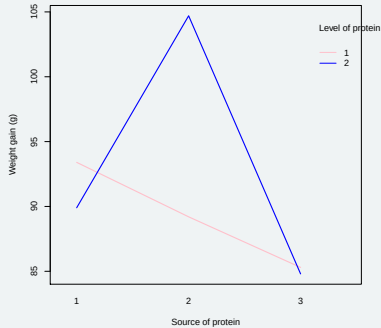
Interaction

```
1 par(mfrow=c(1,1))
2 interaction.plot(x.factor = rats$Source, #x-axis variable
3                 trace.factor = rats$Level, #variable for lines
4                 response = rats$Gains, #y-axis variable
5                 fun = mean, #metric to plot
6                 ylab = "Weight gain (g)",
7                 xlab = "Source of protein",
8                 col = c("pink", "blue"),
9                 lty = 1, #line type
10                lwd = 2, #line width
11                trace.label = "Level of protein")
```

Interaction



Interaction



- Weight gains are similar, irrespective of the level of protein, when the source of protein is meat-based
- Weight gains are considerably higher when the level of protein is high only when the source of protein is cereal

Factorial experiments

We will now break for a problem session covering this section.

Please go to **Problem Session 2** of **Lesson 6**.

See you in a bit!

Conclusions

- 1 ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples

Conclusions

- 1 ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples
- 2 One-way ANOVA

Conclusions

- 1 ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples
- 2 One-way ANOVA
 - ▶ Can be used to analyse data from **planned experiments** and **observational studies**

Conclusions

- ① ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples
- ② One-way ANOVA
 - ▶ Can be used to analyse data from **planned experiments** and **observational studies**
 - ▶ Can be seen as an extension of a two-sample *t*-test for more than two groups

Conclusions

- ① ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples
- ② One-way ANOVA
 - ▶ Can be used to analyse data from **planned experiments** and **observational studies**
 - ▶ Can be seen as an extension of a two-sample t -test for more than two groups
 - ▶ We essentially have **one factor** – the group identifier – which might be affecting the response variable

Conclusions

- ① ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples
- ② One-way ANOVA
 - ▶ Can be used to analyse data from **planned experiments** and **observational studies**
 - ▶ Can be seen as an extension of a two-sample t -test for more than two groups
 - ▶ We essentially have **one factor** – the group identifier – which might be affecting the response variable
 - ▶ Test statistic is the **Variance Ratio** – compares the variance of the group means to the mean of the group variances

Conclusions

- ① ANOVA – looks for **differences between means** of groups. Takes into account the **variability** of samples
- ② One-way ANOVA
 - ▶ Can be used to analyse data from **planned experiments** and **observational studies**
 - ▶ Can be seen as an extension of a two-sample t -test for more than two groups
 - ▶ We essentially have **one factor** – the group identifier – which might be affecting the response variable
 - ▶ Test statistic is the **Variance Ratio** – compares the variance of the group means to the mean of the group variances
 - ▶ Can be followed up with a **Tukey HSD** multiple comparisons procedure to see where differences might lie

③ Two-way ANOVA

③ Two-way ANOVA

- ▶ Useful when there are **two factors** potentially influencing the response variable
- ▶ Could look like a one-way ANOVA, but with a potential **blocking factor**

③ Two-way ANOVA

- ▶ Useful when there are **two factors** potentially influencing the response variable
- ▶ Could look like a one-way ANOVA, but with a potential **blocking factor**
- ▶ Could be a **factorial experiment**, with two main factors of interest (and perhaps an interaction)

③ Two-way ANOVA

- ▶ Useful when there are **two factors** potentially influencing the response variable
- ▶ Could look like a one-way ANOVA, but with a potential **blocking factor**
- ▶ Could be a **factorial experiment**, with two main factors of interest (and perhaps an interaction)

④ Assumptions

③ Two-way ANOVA

- ▶ Useful when there are **two factors** potentially influencing the response variable
- ▶ Could look like a one-way ANOVA, but with a potential **blocking factor**
- ▶ Could be a **factorial experiment**, with two main factors of interest (and perhaps an interaction)

④ Assumptions

- ▶ Samples drawn from **Normal populations** with **equal variances**

③ Two-way ANOVA

- ▶ Useful when there are **two factors** potentially influencing the response variable
- ▶ Could look like a one-way ANOVA, but with a potential **blocking factor**
- ▶ Could be a **factorial experiment**, with two main factors of interest (and perhaps an interaction)

④ Assumptions

- ▶ Samples drawn from **Normal populations** with **equal variances**
- ▶ **Independence** within samples