Introductory Statistics with Excel

Andrew Parnell

University College Dublin

Class 5 - Statistical modelling and Analysis of Variance

Learning outcomes

In this class we will cover

- What is a statistical model?
- Analysis of Variance (ANOVA)
- 1-way ANOVA
- 2-way ANOVA

Moving from hypothesis testing to statistical modelling

- Many people see statistics as just hypothesis testing and confidence intervals
- Most academic statisticians actually work on **statistical modelling** rather than ever running hypothesis tests (confidence intervals are used by all of us)
- A statistical model is a way of linking the data to a theoretical construct (i.e. our ideas about how the experiment worked) and to test them using statistical thinking (usually confidence intervals)
- Most statistical models can be written as:

$$\text{Data} = \text{Structure} + \text{Residuals}$$

- The **data** is what we have observed, the **structure** is a hypothesised way the experiment works, and the **residuals** are the leftover 'random' variation

Example: cows

Reminder:

169.6	142	103.3	111.6	123.4	143.5	155.1
101.7	170.7	113.2	130.9	146.1	169.3	155.5

- Our statistical model might be:

$$\text{Data} = \text{Overall mean} + \text{Residuals}$$

- The overall mean here might be a value that we specify (e.g. 120kg) or something we try and estimate from the data
- The residuals need to be assigned a probability distribution, such as the normal distribution. This will have a mean (usually set to 0) and a standard deviation which again we can either specify or estimate from the data
- For example, if we had specified that the overall mean was 120kg then our first data point would be:

$$169.6 = 120 + 49.6$$

so the residual value is 49.6. We could do this repeatedly and estimate the standard deviation from the set of residual values

More complex statistical models

- If we have multiple groups, e.g. the Clenbuterol data, we might have a more complicated structural part:

$$\text{Data} = \text{Mean associated with group} + \text{Residuals}$$

- Depending whether the data was in e.g. run 1 or run 2, there might be a different group mean
- We might further decide to give different probability distributions to the residual variation in the different groups

Links with t-tests and ANOVA

- The 1-sample t-test (e.g. testing cow's milk yield is 120kg or not) is an example of a statistical model with one overall group mean
- The independent samples t-test we met earlier is an example of a statistical model with two different group means
- The Analysis of Variance (ANOVA) model we are about to meet is an example of a statistical model with multiple different group means (at least 2)

1-way Analysis of Variance

ANOVA

ANalysis Of VAriance is a versatile statistical tool for analysing how the mean value of a **quantitative response** variable is related to one or more **categorical explanatory** factors.

1-way ANOVA

Example: reduction in blood pressure.

- A researcher wishes to compare three different regimes for their effectiveness in lowering blood pressure in patients diagnosed with high blood pressure.
- The recommended regimes are
  1. Medication
  2. Exercise
  3. Diet
- Fifteen patients have volunteered to take part in the experiment.

1-way ANOVA

- The BP in each patient was recorded after they had followed their recommended regime:

Medication	Exercise	Diet
10	6	5
12	8	9
9	3	12
15	0	8
13	2	4

- Specific question: **is there a difference in the mean reduction in blood pressure for the 3 regimes?**

1-way ANOVA

- A t test is used to compare 2 population means when independent random samples are available from the populations.
- Given 3 independent samples, t tests could be used to conduct 3 tests (1 v 2, 2 v 3 and 3 v 1) but this causes problems, e.g. is the type 1 error rate still 5%? Should the groups all have the same residual probability distribution?
- The ANalysis Of VAriance (ANOVA) statistical model (and its associated F-**test**) provides a test that the means are different based on the assumption that the residual standard deviations in each group are all equal.

Experimental design

- The simplest experimental design is called a **completely randomised experimental design**.
- This is an experimental design under which independent random samples of experimental units are selected for each treatment.
- *Blood pressure example*: each patient diagnosed with high blood pressure is randomly assigned to one of the regimes
  i.e. 5 patients in each regime.

ANOVA terminology

- The response variable is the variable being measured in the experiment:
  drop in BP
- Factors are variables applied to the object or individual in the experiment:
  regime
- Factor levels are the values of the factor used in the experiment:
  medication, exercise, diet
- Treatments are the factor-Level combinations used
  same as factor levels here!
- Experimental units are the objects on which the response variables and factor levels are measured/recorded:
  patients

ANOVA terminology

- We are interested in assessing the the effect of the factors on the response variable:
  effect of regime on drop in BP
- Quantitative factors are measured on a numerical scale:
  age, daily medication dose
- Qualitative factors are not measured on a numerical scale: regime
- Reminder: a designed experiment is one for which the researcher controls the specifications of the treatments and the method of assigning the experimental units to each treatment:
  completely randomized design
- An observational study is one in which the researcher simply observes the treatments and the response on a sample of experimental units:
  e.g. examine a doctor's records and record regime and drop in BP for each patient

1-way ANOVA

- In general we want to compare the means of k populations.

	1	2	3	...	k
Population mean	pop mean 1	pop mean 2	pop mean 3	...	pop mean k
Population st dev	pop sd 1	pop sd 2	pop sd 3	...	pop sd k
Sample size	n_1	n_2	n_3	...	n_k
Sample mean	sample mean 1	sample mean 2	sample mean 3	...	sample mean k
Sample st dev	sample st dev 1	sample st dev 2	sample st dev 3	...	sample st dev k

- Formally, in ANOVA, we test:

$$H_0 : \text{pop mean } 1 = \text{pop mean } 2 = \text{pop mean } 3 = \ldots = \text{pop mean } k$$

against the alternative

$$H_A : \text{ at least one mean is different from the others.}$$

1-way ANOVA

Example

- For the BP experiment:

Medication	Exercise	Diet
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
sample mean 1 = 11.8	sample mean 2 = 3.8	sample mean 3 = 7.6
sd 1 = 2.4	sd 2 = 3.2	sd 3 = 3.2

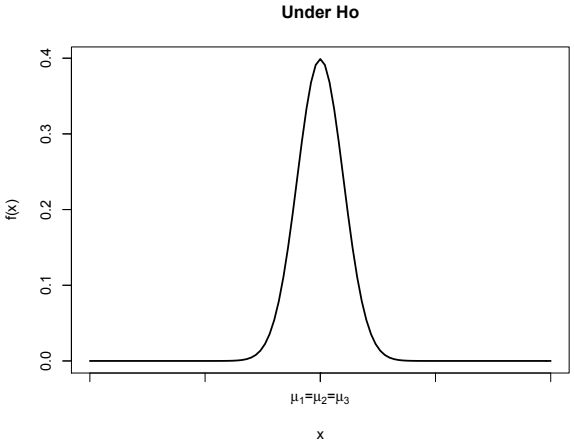
- Is there evidence in the data which suggests that the average BP under the 3 regimes differs?
- Certainly the 3 sample means *vary* – but this will happen even if H_0 is true!
- Do the 3 sample means *vary enough* to suggest that there is a difference in the true underlying population means?

1-way ANOVA: assumptions

- Assumptions are similar to those for the 2-sample t-test.

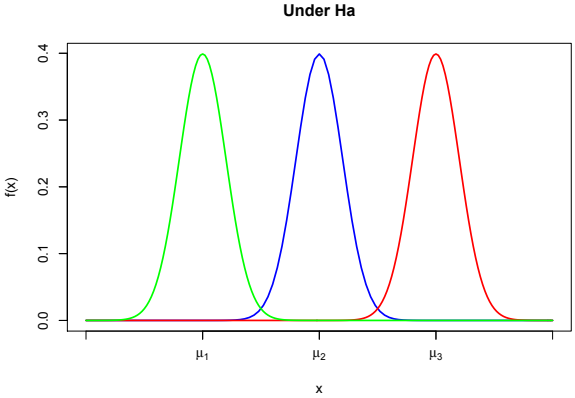
1. The samples are independent random samples.
2. The sample means are approximately normally distributed within each population.
3. All populations have the **same** standard deviation.

1-way ANOVA: assumptions



- Under H_0 all populations have the same mean.
  Given the assumptions, all have the same residual standard deviation.

1-way ANOVA: assumptions



- Under H_a all populations have different means.
  Given the assumptions, all have the same residual standard deviation.

1-way ANOVA

- Use ANOVA to test H_0 (all population means equal).
- We do this using an F**-test**:

$$F = \frac{\text{Variation between sample means}}{\text{Variation within groups}}$$

- The numerator is 0 if all means are equal – gets larger the more spread out the means are.
- If that variation is 'large', evidence that at least one of the means is different from the others $\Rightarrow$ reject H_0.
- The denominator provides a guide for determining whether the numerator is 'large'.
- Comparing variation between group means to the variation within groups; hence "analysis of variance".

1-way ANOVA

The results are usually presented in a table with the following values:

- **SSTr**: the sum-of-squares associated with the treatment. This is a measure of how variable the means are between the groups
- **SSR**: the sum-of-squares associated with the residuals. This is a measure of how variable the residuals are within a group
- **MSTr**: the SSTr value corrected for the number of groups
- **MSR**: the SSR value corrected for the number of observations

The ratio of these last two provides the F-statistic

The ANOVA table

Source	df	SS	MS	F
Treatment	k - 1	SSTr	MSTr $= \frac{\text{SSTr}}{k-1}$	$F = \frac{\text{MSTr}}{\text{MSR}}$
Residual	$n-k$	SSR	MSR $= \frac{\text{SSR}}{n-k}$	
Total	n - 1	SSTo		

SS and MS stand for sum-of-squares and mean-square respectively.

1-way ANOVA

- We won't go into the details of how these values are created, but the table you get from Excel gives:

Source	df	SS	MS	F
Treatment	2	160.13	80.07	9.17
Residual	12	104.80	8.73	
Total	14	264.93		

- The residual standard deviation is obtained from $\sqrt{\text{MSR}} = 2.95$.
- Excel also gives you the critical value for the F-test and the p-value, here 0.004
- The p-value is far smaller than 0.05 so we would reject H_0 and conclude that, at the 5% level there is at least one different mean

1-way ANOVA – DIY!

Example: Journal of Adolescent Research 1992

- Is musical preference associated with reckless behaviour?
- Musical Preference:
  Acoustic/Pop, Mainstream Rock, Hard Rock, Heavy Metal
- Reckless Behaviour:
  How many times in the past year have you indulged in reckless behaviour?
- Here reckless behaviour is defined as 'speeding over 80mph'.

1-way ANOVA

- Cutting to the chase, here is the ANOVA table:

Source	df	SS	MS	F	p
Treament	3	12.85	4.2833	5.09	0.0029
Residual	76	63.9	0.8408		
Total	79	76.75			

- For significance level 5% the mean number of occasions driving over 80mph differs for the groups defined in terms of musical preference.

Comparing means using ANOVA

1-way ANOVA: which means differ?

- If ANOVA fails to accept H_0 we are likely to be interested in **which** means differ.
- To compare specific means, either with a test or confidence interval, we need the standard error of difference SED.
- For means of samples of sizes n_1 and n_2:

$$\text{SED} = \sqrt{\text{MSR}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- The SED makes use of MSR to estimate the residual standard deviation, thereby using more information than a standard t-test between groups.

1-way ANOVA: which means differ?

Example: Is musical preference associated with reckless behaviour?

- In this case there are 20 subjects in each treatment group, so:

$$\text{SED} = \sqrt{\frac{2 \times 0.84079}{20}} = 0.28996$$

- To test at significance level 1% with $df = 76$, the t critical value (from Excel - `T.INV.2T(0.01, 73)`) is 2.642.
- Hence, the population means differ at significance level 1% if:

$$\frac{|\text{ sample mean 1} - \text{sample mean 2 }|}{\text{SED}} \quad > \quad \text{Critical value}$$

1-way ANOVA: which means differ?

- It is useful when comparing means to define the least significant difference LSD.
- We define the least significant difference LSD to be:

$$\text{LSD} = \text{Critical value} \times \text{SED}$$

- Hence sample means 1 and 2 differ at the 1% level if:

$$|\ \text{sample mean 1} - \text{sample mean 2}\ | > \text{LSD}$$

1-way ANOVA: which means differ?

Example: Is musical preference associated with reckless behaviour?

- We previously inferred that the means differed for the groups defined in terms of musical preference. Which means actually differ?

	Acoustic/ Pop	Mainstream Rock	Hard Rock	Heavy Metal
Sample means	2.50	2.70	2.75	3.55

- Using a 1% significance level

$$\text{LSD} = \text{critical value} \times \text{SED} = 2.642 \times 0.290 = 0.766$$

- So number of reckless events on Heavy Metal differs significantly from the others (all differences ≥ 0.8).
- There are no significant differences among the other 3 groups (all differences ≤ 0.25).

1-way ANOVA: which means differ?

Example: music preference/reckless behaviour.

- Given that the Heavy Metal group differs significantly from the others we may wish to construct a 95% confidence interval for the mean of the Heavy Metal group.
- The mean for this group is 3.55, with standard error:

$$\sqrt{\frac{\text{MSR}}{20}} = \sqrt{\frac{0.841}{20}} = 0.205$$

- The t critical value, for 95% confidence, with df $= 76$ is 1.992 (Excel - `T.INV.2T(0.05, 76)`)
- Hence the CI of interest is:

$$3.55 \pm 1.992 \times 0.20504 = 3.55 \pm 0.4084 = (3.14,\ \ 3.96) \text{ events}$$

ANOVA with randomised block designs

The Randomised Block Design

- Usually an investigator wishes to establish treatment effects (i.e. differences in treatment means) by rejecting H_0.
- It is possible that H_0 is not rejected even when treatment means do differ
- This is a Type II error, which occurs with probability β.
- The power of a test (for specified α), is the probability that H_0 is rejected so:
  - when H_0 is true, power $= \alpha$.
  - when H_0 is false, the power will depend e.g. on how different the means are, and so power $= 1 - \beta$.

Blocking to increase power

- Power (the probability of detecting differences) is greater when there is less variation among the experimental units i.e. when σ^2 is small.
- In such a case, most of the variation can be attributed to differences between the treatment means and a large F value is more likely.
- Blocking compares values for different treatments measured on similar experimental units.
- *Similar* means similar with respect to characteristics which influence the response.
- *Eg:* when comparing effect of dietary regimes on weight loss, use individuals of same gender and similar weight.

Randomised Block Design

- Say we wish to conduct an experiment to compare k treatment means.

1. Divide the experimental units into groups of k units, so that units within a group are as similar as possible.
2. Within each group randomly allocate treatments so that each unit in a group receives a different treatment.

- The groups are called blocks.
- Allocation of units to blocks is non-random.
- In each block allocation of units to treatments is random.

Randomised block design

- 40 volunteers have agreed to take part in an evaluation of 4 computer training courses.
- Each volunteer will be sent to one of the 4 courses and asked to complete a questionnaire at the end of the course.
- From each questionnaire a general satisfaction score on a scale of 0 to 100 will be computed.
- Possible ways of deciding which volunteer attends which course include:
  - completely randomised design.
  - randomised block design.

Experimental Design

- For a completely randomised design the volunteers are listed in random order. The top 10 are sent on course 1, the next 10 on course 2, etc.
- For a randomised block design the volunteers are grouped into blocks of size 4, such that members of a block are 'similar'.
- Blocks could be constructed on the basis of the individuals previous computer experience; the 4 most experienced in the first block, the next 4 most experienced in the next block, etc.
- The randomised block design could attribute part of the total variation in satisfaction scores to the different experience levels of the volunteers.
- This would reduce residual variation and make it more likely that differences between the satisfaction scores for different courses are detected as significant if they truly exist.

Randomised block designs and ANOVA

Example: computer course evaluation

- Block and Treatment Means:

Block	Tr 1	Tr 2	Tr 3	Tr 4	Mean
1	52.1	51.3	57.9	56.9	54.550
2	59.9	58.8	62.7	63.4	61.200
3	63.3	63.8	64.3	67.8	64.800
4	68.9	69.3	68.3	74.5	70.250
5	68.5	76.4	75.0	73.8	73.425
6	73.2	80.9	80.5	80.1	78.675
7	83.8	84.0	85.6	85.1	84.625
8	85.0	83.4	88.7	91.1	87.050
9	92.8	90.4	91.5	95.0	92.425
10	93.3	95.5	96.6	98.8	96.050
Mean	74.08	75.38	77.11	78.65	76.305

Randomised block designs and ANOVA

- The resulting ANOVA table is below:

Source	DF	SS	MS	F	p-value
Treatment	3	119.5	39.8	**10.0**	0.0001
Block	9	6875.1	763.9	**191.4**	$<$0.0001
Residual	27	107.8	4.0		
Corrected Total	39	7102.4			

- We are not going to cover the maths of where these values come from!

Randomised block designs and ANOVA

- Conclude: there is very strong evidence ($p = 0.0001$) that mean satisfaction scores depend on the training course.
- Conclude: block means differed significantly ($p < 0.0001$), so blocking did succeed in reducing the residual (error) variation.
- Warning: if blocking did not reduce residual error, the randomised block design would be less powerful that the completely randomised design.
- The loss of degrees of freedom would make critical values larger, and F values would not have increased sufficiently to make up for this.

Randomised block designs and ANOVA

- As with a completely randomised block experiment, use a LSD to compare particular treatment means.
- Treatment means average l values so

$$SED = \sqrt{\frac{2\text{MSE}}{l}} \quad \text{and as before} \quad \text{LSD} = \text{critical value} \times \text{SED}.$$

- For the training course example with 5% significance level and critical value $= =$ 2.052 (`T.INV.2T(0.05, 27)`):

$$\begin{aligned} \text{LSD} &= \text{critical value} \times \text{SED} \\ &= 2.052 \times \sqrt{\frac{2 \times 3.991741}{10}} \\ &= 2.052 \times 0.8935 \\ &= 1.834 \end{aligned}$$

Randomised block designs and ANOVA

- The treatment means were 74.08, 75.38, 77.11 and 78.65, and the LSD is 1.834.
- So course 3 significantly different from 1, 4 different from 1, 4 different from 2, 3 different from 2
- 2 and 1 not significantly different from each other, similarly 3 and 2, 4 and 3 (at the 5% significance level).

Multiple comparisons

- When H_0: all means are equal is rejected by the F test we have used the LSD (based on a t) test to see which means differ.
- With $k = 4$, for example, this involves 6 different tests: tr1 mean vs tr2, tr1 vs tr3, etc, etc
- When H_0 is true and we make many independent comparisons we will reject H_0 much more often than the significance level that we set (e.g. 5%)
- We can use a **multiple comparisons procedure** to reduce this frequency.
- The most common is that of **Bonferroni** correction where we reduce the significance level by dividing it by the number of tests, e.g. test at the $0.05/6 = 0.83\%$ level instead of the 5% level

2-way ANOVA

Two-way ANOVA

- Two-way ANOVA examines the effect of 2 categorical factors on the mean of a quantitative response variable.

Examples

- effect of variety (3 levels) and planting density (10, 20, 30 thousand plants per hectare) on yield of tomato plants.

- effect of investment manager (3 levels) and investment type (4 levels) on 5 year yield on a €10,000 investment.
- Effect of

Factorial experimental designs

- Effect of two or more treatments explored simultaneously.
- Every level of each treatment is studied under the conditions of every level of all other treatments.
- Two, three, four, ..., n treatments can be studied in same experiment.
- Here we'll consider only 2.

Advantages of factorial experimental designs

- Need 2 completely randomised designs to analyse effects of two treatments.
- Factorial design can analyse the effect of 2 treatments together, saving time and effort.
- Can be use to control confounding variables, by building such variables into the experimental design.
- Increased power: effects of second variable removed from the SSR.
- Perhaps most importantly: can explore possible interaction between two treatment variables.

Factorial experimental designs: example

- The time required to make coke from coal was examined in a 2×3 experiment with 18 observations.
- The effect of oven width and oven temperature on the process are of interest.
- Oven width has three levels: 4, 8 and 12 inches.
- Oven temperature has two levels: 1600^oF and 1900^oF.

width	temp	time
4	1600	3.5
4	1600	3.0
4	1600	2.7
4	1900	2.2
4	1900	2.3
4	1900	2.4
8	1600	7.1
⋮	⋮	⋮

Factorial experimental designs: example

- The intersection of each level of each treatment gives rise to **cells**: in the example, there are 3 observations in each of 6 cells.
- One treatment is arbitrarily designated the **row treatment**: here let's make the the oven width.
- The other treatment is the **column treatment**: here that's oven temperature.
- For this module, only consider designs with equal numbers of observations in each cell.

Two-way ANOVA

- Factorial designs with 2 factors are analysed using a two-way ANOVA.
- Two-way ANOVA tests the following hypotheses:

Row effects:	H_0 :	Row means are all equal.
	H_A :	At least one row mean is different.
Column effects:	H_0 :	Column means are all equal.
	H_A :	At least one column mean is different.
Interaction effects:	H_0 :	The interaction effects are 0.
	H_A :	An interaction effect is present.

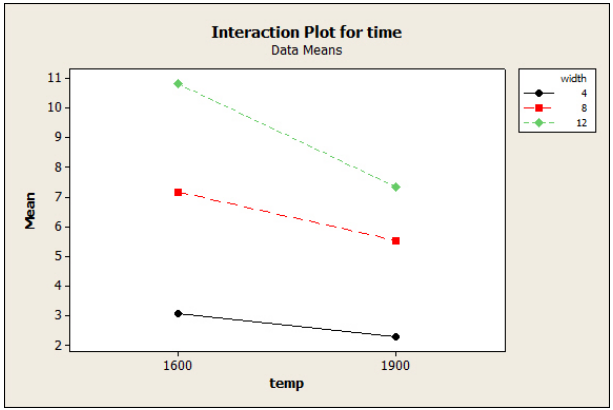
Two-way ANOVA: main effects and interactions

- The row effects and column effects are often called main effects.
- Essentially effect = mean difference between levels.
- An interaction occurs when the effect of one treatment varies according to the levels of the other treatment.
- In a factorial design, interaction occurs when the pattern of cell means in one row (going across the columns) varies from the cell mean pattern in another row.
- Similarly, interaction occurs when the pattern of cell means differs between columns.

Graphically depicting interactions

- Denote the no. of levels in the row treatment by R, and similarly in the column treatment by C.
- In the coking experiment $R = 3$ and $C = 2$.
- Can graphically examine factorial design experiment data by plotting cell means within each row (say) and connecting them with a line.
- If the R lines are parallel, no interaction is indicated.

Graphically depicting interactions



- If the lines are parallel there is likely to be no interaction. Looks like there might be interaction here.

Interpreting effects in the presence of interaction

- When interaction effects are significant, main effects are confounded and **should not be analysed in the usual manner**.
- It is not possible in such a case to state that row or column effects are significant as the main effect of one treatment varies with the levels of the other main effect.

Two-way ANOVA table: coking example.

Width effect:	H_0 :	Row means are all equal.
	H_A :	At least one row mean is different.
Temp. effects:	H_0 :	Column means are all equal.
	H_A :	At least one column mean is different.
Interaction effects:	H_0 :	The interaction effects are 0.
	H_A :	An interaction effect is present.

- To test these three hypotheses we, as usual, determine a F test statistic for each.

Computing a two-way ANOVA table

Source	**df**	**SS**	**MS**	**F**	**p**
Row factor	$Ro-1$	SSRo	$\frac{\text{SSRo}}{Ro-1}$	$\frac{\text{MSRo}}{\text{MSR}}$	
Column factor	$C-1$	SSC	$\frac{\text{SSC}}{C-1}$	$\frac{\text{MSC}}{\text{MSR}}$	
Interaction	$(Ro-1)\times(C-1)$	SSI	$\frac{\text{SSI}}{(Ro-1)\times(C-1)}$	$\frac{\text{MSI}}{\text{MSR}}$	
Residual	$Ro\times\times C\times(n-1)$	SSR	$\frac{\text{SSE}}{Ro\times C\times(n-1)}$		
Total	$Ro\times Cn-1$	SST			

Two-way ANOVA: coking example.

Source	DF	SS	MS	F	P
width	2	123.1	61.6	222.1	<0.001
temp	1	17.2	17.2	62.1	<0.001
Interaction	2	5.7	2.9	10.3	0.003
Error	12	3.3	0.3		
Total	17	149.4			

- First, test for interaction:

$$F = \frac{\text{MSI}}{\text{MSE}} = \frac{2.8506}{0.2772} = 10.28$$

- At the 5% level, the critical value is 3.885 (`F.INV(0.95, 2, 12)`).
- Hence, there is very strong evidence of a width $\times$ temperature interaction – the 'no interaction' null hypothesis is rejected at the 5% level.

Two-way ANOVA: coking example.

- So, the interaction term is significant. Looking at the cell means gives us an idea of what's happening:

		Temp	
		1600	1900
Width	4	3.07	2.30
	8	7.17	5.53
	12	10.80	7.33

- The difference in time to coking at high and low temperatures increases with oven width.
- **When this is the case the individual tests for the two factors make no sense.**

Two-way ANOVA: tests following a significant interaction

- In the case of a significant interaction, we can compare temperatures at each width, using an LSD.
- Each cell mean is calculated from $n = 3$ observations so

$$\begin{aligned} \text{SED} = \sqrt{\frac{2\text{MSR}}{n}} &= \sqrt{\frac{2 \times 0.2772}{3}} \\ &= 0.4298837 \\ \text{critical value} &= 2.179 \end{aligned}$$

$$\begin{aligned} \text{Therefore LSD} &= (2.179) \times (0.4298837) \\ &= 0.937 \end{aligned}$$

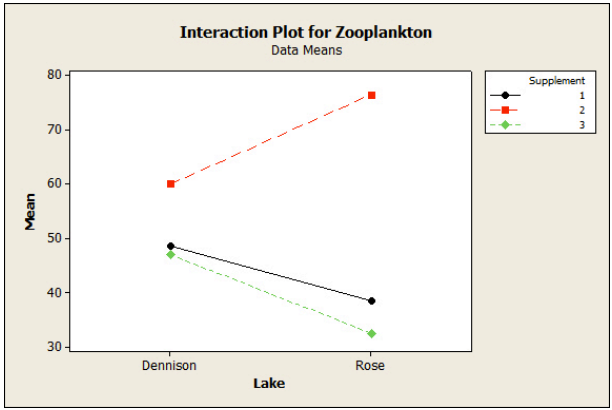
- Only the means at width 4 under both temperatures are not significantly different, at the 5% level.

Two-way ANOVA: example 2

- A study on how zooplankton live in 2 lakes.
- 12 tanks have been set up: 6 with water from Lake Rose, and 6 with water from Lake Dennison.
- One of three nutrient supplements to each tank.
- After 30 days, count the zooplankton in unit volume of water.
- Use two-way ANOVA to test if the population means are equal, or equivalently, to test whether there is significant evidence of interactions and main effects.

Zooplankton	Supplement	Lake
34	1	Rose
43	1	Rose
57	1	Dennison
40	1	Dennison
85	2	Rose
⋮	⋮	⋮

Two-way ANOVA: zooplankton example



Two-way ANOVA: zooplankton example

Source	df	SS	MS	F	p
Supplement	2	1918.50	959.25	9.25	0.015
Lake	1	21.33	21.33	0.21	0.666
Interaction	2	561.17	280.58	2.71	0.145
Residual	6	622.00	103.67		
Total	11	3123.00			

- Test for interaction: $p = 0.145$
- There is no significant evidence for a supplement$\times$lake interaction at the 5% level
- i.e. the effect of supplement is the same for the two lakes.

Two-way ANOVA: zooplankton example

- Since the interaction is not significant we regard it as random error and combine it with SSE.
- Therefore, we also add the interaction df to the error df.
- Therefore need to re-compute all the relevant MS, F test statistics and p values.

Source	df	SS	MS	F	p
Supplement	2	1918.50	959.250	6.49	0.021
Lake	1	21.33	21.333	0.14	0.714
Residual	8	1183.17	147.8963		
Total	11	3123.00			

- At 5% level: significant supplement effect, but no significant lake effect.

Two-way ANOVA: zooplankton example

Supplement	1	2	3
Mean	43.50	68.25	39.75

- Each mean is based on 4 values (2 tanks from each lake) so:

$$\begin{aligned} \text{SED} &= \sqrt{\frac{2\text{MSR}}{4}} = \sqrt{\frac{2 \times 147.896}{4}} \\ &= 8.599 \end{aligned}$$

$$\begin{aligned} \text{Since critical value} &= 2.306 \\ \text{LSD} &= 2.306 \times 8.599 = 19.83 \end{aligned}$$

- The means under both supplement 1 and supplement 3 differ significantly from the mean zooplankton count under supplement 2.

Summary of class 5

- These are all examples of **statistical models**, either with a single mean, or a set of means for each group / treatment / block
- In general ANOVA models assume that all the residuals have the **same** standard deviation
- We run F**-tests** to determine whether the overall means are different, then LSD tests (possibly with multiple comparisons correction) to determine which means are **different**
- We can use 1-way, 2-way, etc, ANOVA to test for **multiple effects and their interactions**