



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

200
1822 - 2022

Analysis of Variance (ANOVA)

Man Geni Addysg Uwch yng Nghymru
Where Higher Education Began in Wales



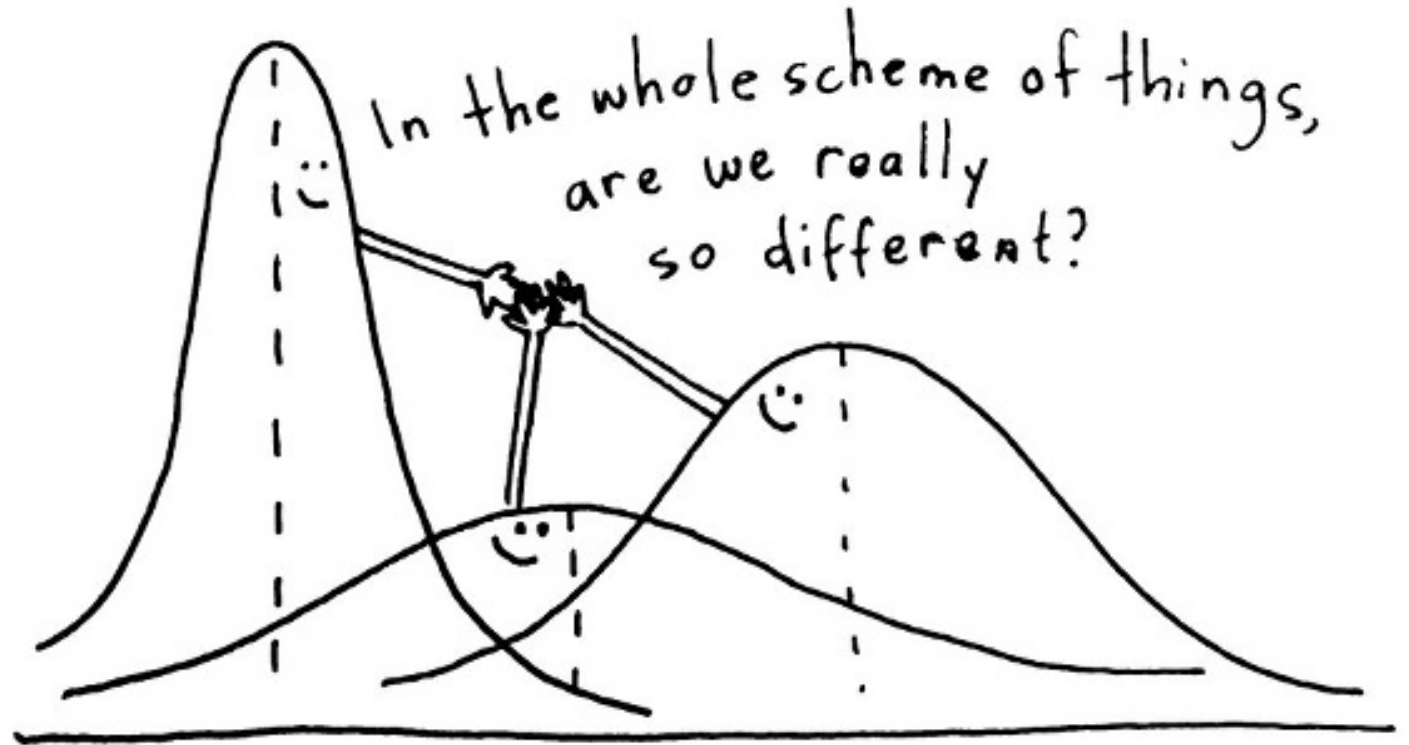
Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

200
1822 - 2022

Man Geni Addysg Uwch yng Nghymru
Where Higher Education Began in Wales

Objectives

- Topics covered
 - ANOVA
 - Why Use ANOVA?
 - Post hoc Analyses
 - The Scheffé Method



ANOVA(Analysis Of VAriance)

An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.

Variance among observations identified , indicate whether observed differences between group means are probably real or merely due to chance

Why Use ANOVA?

- **Comparison of Multiple Groups:**

- ANOVA is beneficial when you have more than two groups to compare.

- Instead of conducting multiple t-tests (which can increase the chance of making Type I errors), ANOVA allows you to test several groups in a single analysis.

- **Control for Type I Error:**

- multiple comparisons (multiple t-tests),  likelihood of making Type I error.

Why Use ANOVA?--Type I errors and Multiple t Tests

- With multiple t Tests, probability (p) of a Type I error is

$$p = 1 - (1 - \alpha)^c$$

Where

α is the alpha level

c is the number of comparisons

- The formula to determine the number of pairwise comparisons (c) that can be made from k groups is given by:

$$c = \frac{k \times (k-1)}{2}$$

Where, k is the number of groups.

Why Use ANOVA?--Type I errors and Multiple t Tests

Assume Group size(k) = 3

$$c = \frac{k \times (k-1)}{2} = \frac{3 \times (3-1)}{2} = 3$$

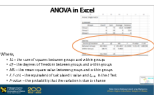
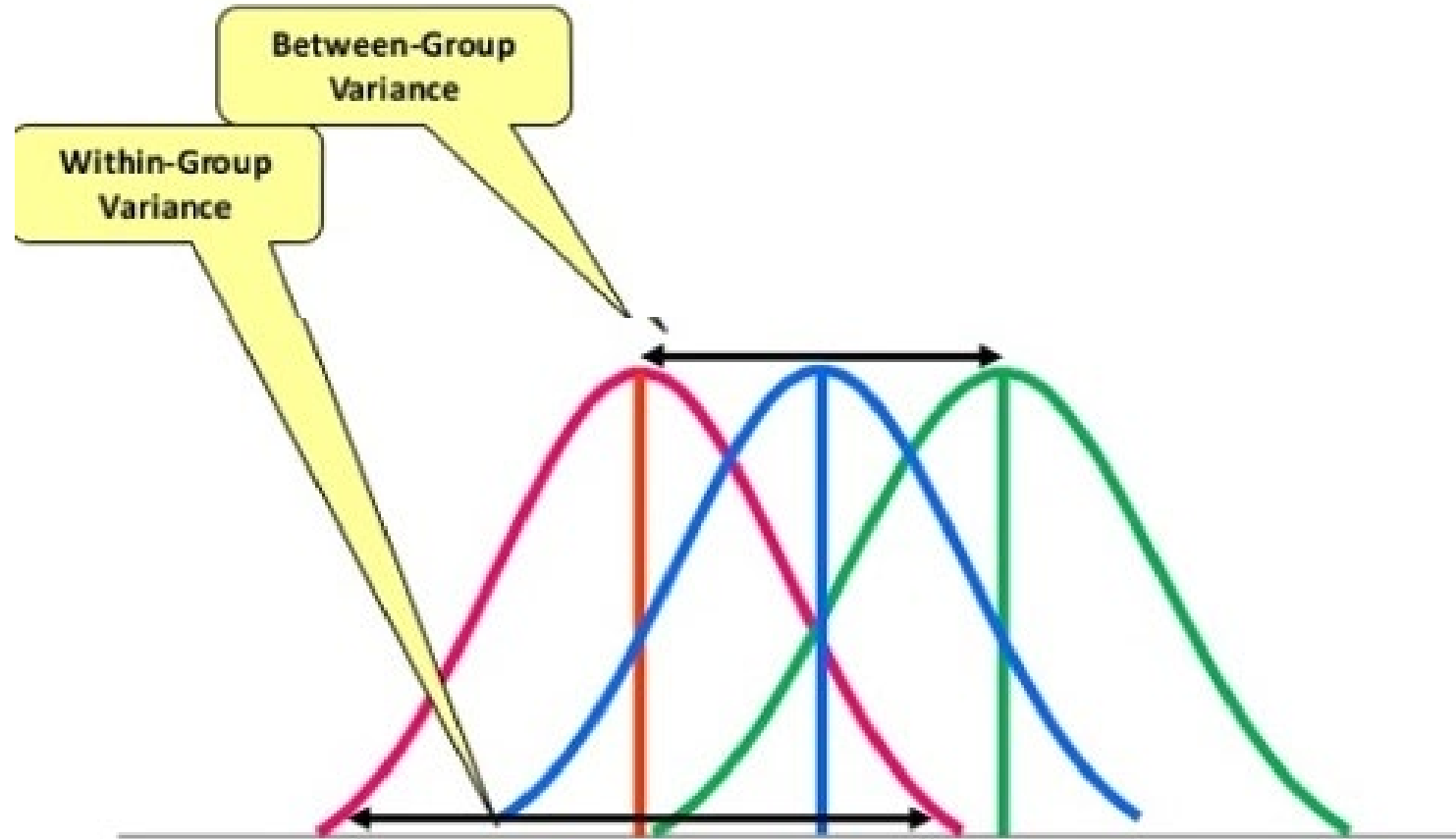
If α (*significance level*) is 0.05 $p = 1 - (1 - \alpha)^c$

$$p = 1 - (1 - 0.05)^3 = 1 - (0.95)^3 = 0.14 = 14\%$$

- Chance of making a Type I error increases to 14%

Why Use ANOVA?

- ANOVA examines the overall variance **between groups** and **within groups**



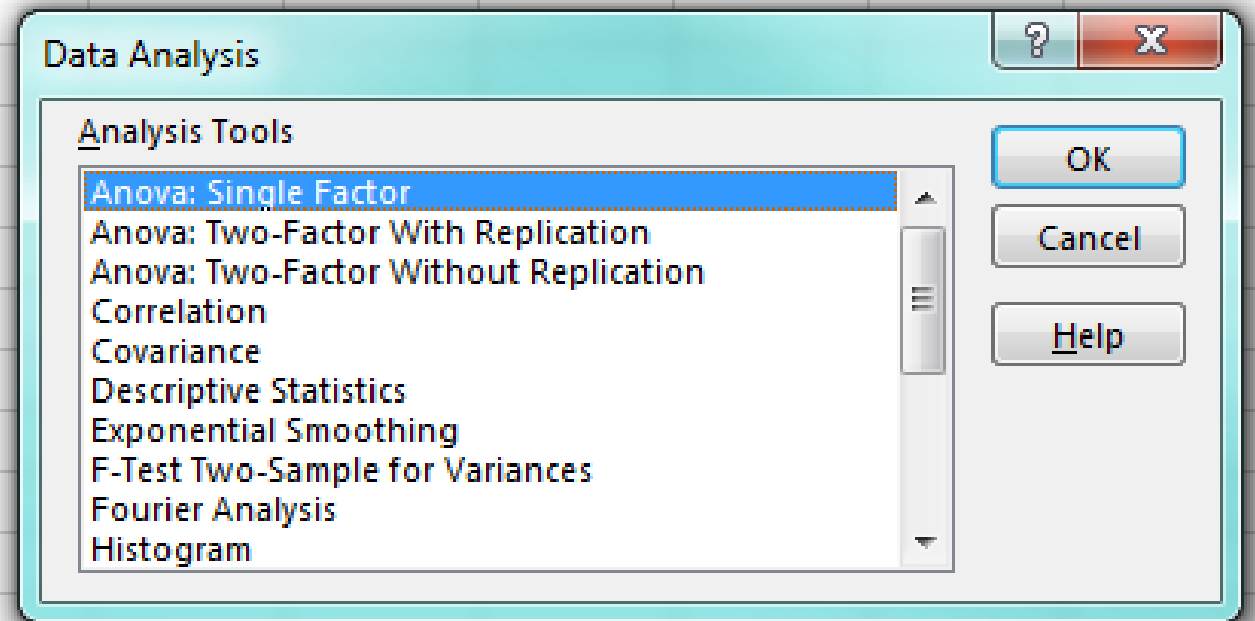
Example

Students from different colleges take the same exam and we want to analyse if one college outperforms the other.

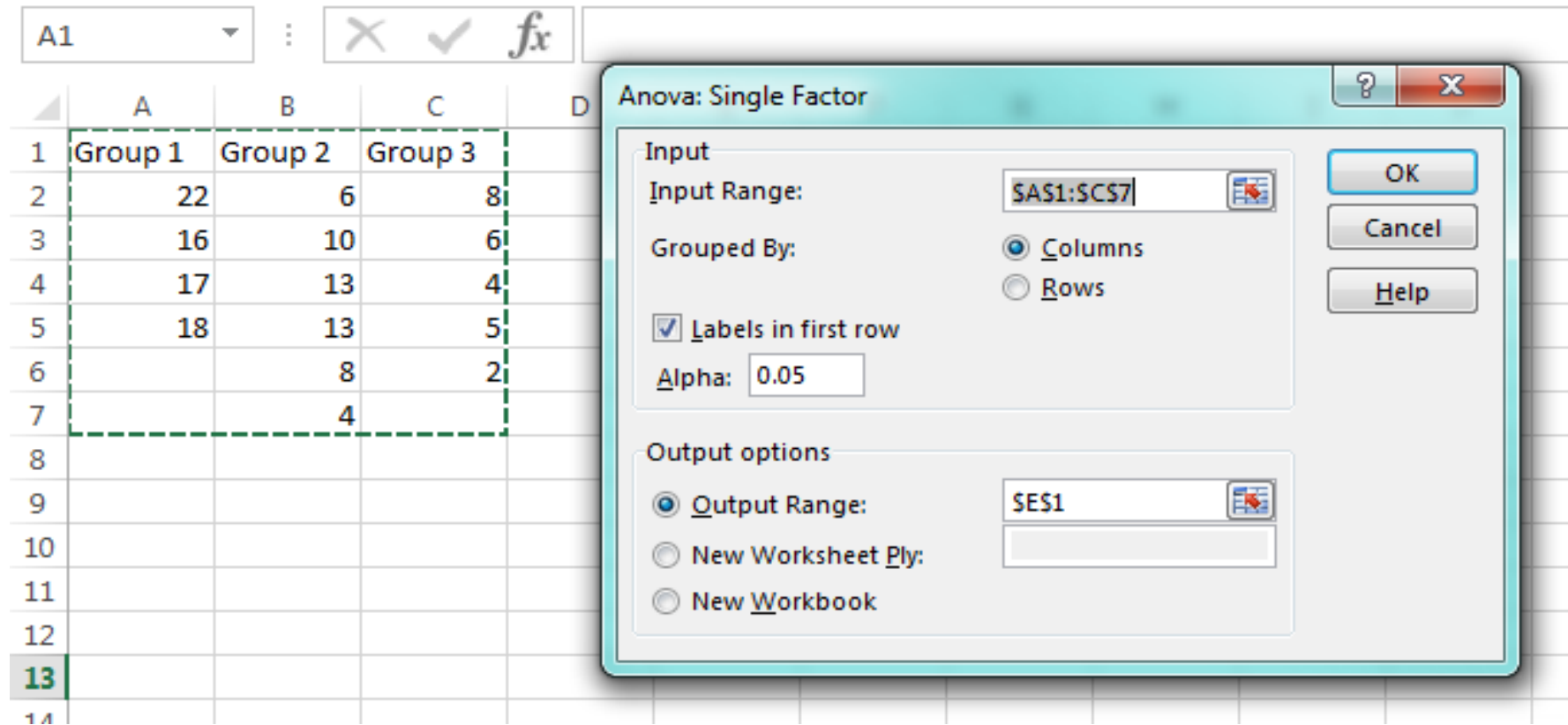
Group 1	Group 2	Group 3
22	6	8
16	10	6
17	13	4
18	13	5
	8	2
	4	

ANOVA in Excel

	A	B	C	D	E	F	G	H	I	J
1	Group 1	Group 2	Group 3							
2	22	6	8							
3	16	10	6							
4	17	13	4							
5	18	13	5							
6		8	2							
7		4								
8										
9										
10										
11										
12										



ANOVA in Excel



The image shows an Excel spreadsheet with data for an ANOVA test. The data is organized into three columns: Group 1, Group 2, and Group 3. The values for Group 1 are 22, 16, 17, and 18. The values for Group 2 are 6, 10, 13, and 13. The values for Group 3 are 8, 6, 4, and 5. The ANOVA: Single Factor dialog box is open, showing the input range as \$A\$1:\$C\$7, grouped by columns, with labels in the first row. The alpha level is set to 0.05. The output range is set to \$E\$1.

	A	B	C
1	Group 1	Group 2	Group 3
2	22	6	8
3	16	10	6
4	17	13	4
5	18	13	5
6		8	2
7		4	

Anova: Single Factor

Input
Input Range:
Grouped By: ☒ Columns ☐ Rows
☒ Labels in first row
Alpha:

Output options
☒ Output Range:
☐ New Worksheet Ply:
☐ New Workbook

OK
Cancel
Help

ANOVA in Excel

	A	B	C	D	E	F	G	H	I	J	K
1	Group 1	Group 2	Group 3		Anova: Single Factor						
2	22	6	8								
3	16	10	6		SUMMARY						
4	17	13	4		<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
5	18	13	5		Group 1	4	73	18.25	6.916667		
6		8	2		Group 2	6	54	9	13.6		
7		4			Group 3	5	25	5	5		
8											
9											
10					ANOVA						
11					<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
12					Between Groups	402.9833	2	201.4917	22.23356	9.21E-05	3.885294
13					Within Groups	108.75	12	9.0625			
14											
15					Total	511.7333	14				
16											

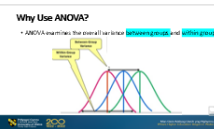
- The first part of the results shows the descriptive statistics for each of the group samples
 - Count, sum, average (mean) and variance (standard deviation squared)

ANOVA in Excel

E	F	G	H	I	J	K	
Anova: Single Factor							
SUMMARY							
Groups	Count	Sum	Average	Variance			
Group 1	4	73	18.25	6.916667			
Group 2	6	54	9	13.6			
Group 3	5	25	5	5			
ANOVA							
Source of Variation	SS	df	MS	F	P-value	F crit	
Between Groups	402.9833	2	201.4917	22.23356	9.21E-05	3.885294	
Within Groups	108.75	12	9.0625				
Total	511.7333	14					

Where,

- *SS* – the sum of squares between groups and within groups
- *df* – the degrees of freedom between groups and within groups
- *MS* – the mean square value between groups and within groups
- *F*, *F crit* – the equivalent of calculated *t* value and t_{crit} in the *t* Test
- *P-value* – the probability that the variation is due to chance



Limitations of One-Way ANOVA

- A one-way ANOVA tells us that at least two groups are different from each other (if null hypothesis is rejected).
- which groups are different.? 
- If our test returns a significant f-statistic, we may need to run a **post-hoc test** to tell us exactly which groups differ in means. Below I have mentioned the steps to perform one-way ANOVA in Microsoft Excel along with a post-hoc test.

Popular post hoc tests include:

- **Tukey's HSD (Honestly Significant Difference):** This test compares all possible pairs of group means and provides a simultaneous test for the difference between each pair.
- **Bonferroni correction:** Adjusts the significance level for each comparison to control the overall family-wise error rate.
- **Scheffé's method:** Provides a conservative approach for testing all possible pairwise comparisons.
- **Dunnett's test:** Used for comparing multiple treatment groups against a control group.

Exercise

- A company is thinking of replacing their current flavour of drink with two new flavours. They do a taste test with each of the new flavours and their current flavour and get their testers to rate the tastes out of ten. These are their results

New flavour 1	New flavour 2	Current flavour
8	2	7
9	3	6
10	4	5
10	2	8
6	1	10
8	3	6
9	2	7

1. State your hypotheses
2. Select α , your significance level
3. Calculate F
4. Find F_{crit}
5. Decide whether or not to reject the null hypothesis
6. If you reject the null hypothesis, calculate the strength of association

Post Hoc Analysis

- In the last problem we discovered that there was a significant difference between the **means**
 - Current flavour: 7 New flavour 1: 8.57 New flavour 2: 2.42
- In Anova method which pairs are statistically different is not clear
- *Post Hoc Analysis* determines pairs which show significant differences
 - Scheffé method (C)
 - Groups of equal or unequal N's

Anova: Single Factor				
SUMMARY				
Groups	Count	Sum	Average	Variance
New flavou	7	60	8.571429	1.952381
New flavou	7	17	2.428571	0.952381
Current flay	7	49	7	2.666667

The Scheffé method of Post Hoc Analysis

- The Scheffé method requires the calculation of C_{obt} which is then compared against C_{crit} (which is calculated, not looked up)

$$C_{obt} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{MS_W \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

Where

- $\bar{X}_1, \bar{X}_2$ are the means of the two groups being compared
- N_1, N_2 are the N s of the two groups
- MS_W is the within-group mean square from your ANOVA

$$C_{crit} = \sqrt{(K - 1)(F_{crit})}$$

- Where, K is the number of groups
- F_{crit} is the critical value of the F test from your ANOVA

The Scheffé method of Post Hoc Analysis

- Current flavour v. flavour 1

$$C_{obt} = \frac{\bar{X}_C - \bar{X}_1}{\sqrt{MS_W \left(\frac{1}{N_C} + \frac{1}{N_1} \right)}} = \frac{7 - 8.57}{\sqrt{1.86 \left(\frac{1}{7} + \frac{1}{7} \right)}} = \frac{-1.57}{\sqrt{1.86 \left(\frac{2}{7} \right)}}$$

$$C_{obt} = \frac{-1.57}{\sqrt{0.53}} = \frac{-1.57}{0.73} = -2.15$$

- The minus sign doesn't matter in this case as we could have calculated $\bar{X}_C - \bar{X}_1$ or $\bar{X}_1 - \bar{X}_C$, giving us

$$C_{obt} = 2.15$$

- Next, compared with the critical value

$$C_{crit} = \sqrt{(K - 1)(F_{crit})} = \sqrt{(3 - 1)(3.55)} = \sqrt{7.1}$$
$$C_{crit} = 2.66$$

$$C_{obt} < C_{crit}$$

Conclusion:

No significant difference between the current flavour and new flavour 1

The Scheffé method of Post Hoc Analysis

- Current flavour v. flavour 2

$$C_{obt} = \frac{\bar{X}_C - \bar{X}_2}{\sqrt{MS_W \left(\frac{1}{N_C} + \frac{1}{N_2} \right)}} = \frac{7 - 2.43}{\sqrt{1.86 \left(\frac{1}{7} + \frac{1}{7} \right)}} = \frac{4.57}{\sqrt{1.86 \left(\frac{2}{7} \right)}}$$

$$C_{obt} = \frac{4.57}{\sqrt{0.53}} = \frac{4.57}{0.73} = 6.26$$

- This now has to be compared with the critical value
 $C_{crit} = \sqrt{(K - 1)(F_{crit})} = \sqrt{(3 - 1)(3.55)} = \sqrt{7.1} = 2.66$

$$C_{obt} > C_{crit}$$

Conclusion: Difference between the current
flavour and new flavour 2

Interpreting the results

	New flavour 1	New flavour 2
Current flavour	$\bar{X}_c < \bar{X}_1$ $C_{obt} < C_{crit}$	$\bar{X}_c > \bar{X}_2$ $C_{obt} > C_{crit}$

- The rating for flavour 1 is higher but not significantly higher than for the current flavour
 - Even though flavour 1 scores higher it is not high enough to justify replacing the current flavour with flavour 1
- The rating for flavour 2 is significantly lower than for the current flavour
 - It would be a very bad idea to replace the current flavour with flavour 2

Note: When conducting Scheffé post-hoc tests there is no increase in the likelihood of making Type I errors with repeated tests

Exercise

Group 1	Group 2	Group 3
22	6	8
16	10	6
17	13	4
18	13	5
	8	2
	4	

- Suppose we have three groups
 - Groups 1 and 2 have been given a treatment of some sort
 - Group 3 is the control group
- Conduct an ANOVA test to identify if there are significant differences between groups
 - If such differences exist, carry out *post hoc* tests to identify which treatments made a significant difference

Summary

- Topics covered
 - Post hoc Analyses
 - The Scheffé Method of Post hoc Analysis