ANALYSIS OF VARIANCE

1 Way analysis of variance (ANOVA)

- this is a very frequently used statistical technique in the social sciences and in research generally
- as with the t test for comparing means, anova is used to compare means of different groups / variables, but the t test can only handle comparing means of 2 groups whereas ANOVA can handle any number of groups
- ANOVA can allow us to deal simultaneously with 2 or more independent variables, allowing us to measure the effect of each variable separately but also allowing us to measure the interaction between the effects of 2 or more variables

1 Way analysis of variance (ANOVA)

- The null hypothesis for one way ANOVA:
  - H_0: There is no difference between the n groups being compared
  - Or $H_o: \mu_1 = \mu_2 = \cdots = \mu_n$
- Alternative hypothesis
  - H_A: There is no difference between the n groups being compared
  - Or $H_A: \mu_1, \mu_2, \dots, \mu_n$ are not all equal

1 Way analysis of variance (ANOVA)

MACQUARIE University SYDNEY·AUSTRALIA

- Assumptions made for ANOVA:
  - the populations from which the samples were drawn are normally distributed
  - the populations all have the same standard deviation
  - the samples selected from these populations are independent and random

- using the data from the samples and these assumptions we can calculate a **test statistic TS which we denote by F** this test statistic measures the extent to which the sample means are not all equal

1 Way analysis of variance (ANOVA)

MACQUARIE University SYDNEY·AUSTRALIA

- if the null hypothesis is true then the test statistic has a known distribution called the F distribution. We can obtain critical values (CV) and P-values for this distribution.
- Excel and various statistical software packages have built in functions / tools for these calculations. More on this later.
- The F distribution has 2 parameters, df1 and df2.

1 Way analysis of variance (ANOVA)

THE PROCESS FOR TESTING THE NULL HYPOTHESIS

- 1) compute the test statistic TS from the sample
- 2) identify the parameters df1, df2 of the F distribution this test statistic would have if the null hypothesis is true
- 3) decide on the level of significance alpha for the test
- 4) compute the critical value CV for the F distribution that corresponds to the significance level alpha (excel has a function that can do this)
- 5) check whether the TS is higher than the CV and if so reject the Null Hypothesis and if not don’t reject it **OR**
- 6) compute the pvalue for the TS obtained from the F distribution (excel has a built in function that can do this)
- 7) check if the pvalue is lower than the significance level and if so reject the Null Hypothesis and if not then don’t reject it

Hypothesis test for 1 way ANOVA

- An example:
- M = 3 treatments and N = 4 replications for each treatment
- This is an experiment involving the mean yields of a particular crop grown in soil treated with 3 different types of fertilizer

Treatment A	48	49	50	49
Treatment B	47	49	48	48
Treatment C	49	51	50	50

Hypothesis test for 1 way ANOVA

- Summary
  - It is tedious to calculate this by hand and very easy to make a numerical error
  - Much better to use some statistical software package or excel to do the calculations
  - This test statistic has a distribution called the F distribution with DF1=M-1 and DF2 = M(N-1)
  - We can calculate the pvalue of the TS and given the level of significance alpha we can calculate the critical value of the TS
  - This allows us to perform the Hypothesis testing for 1 way ANOVA

1 way ANOVA in Excel

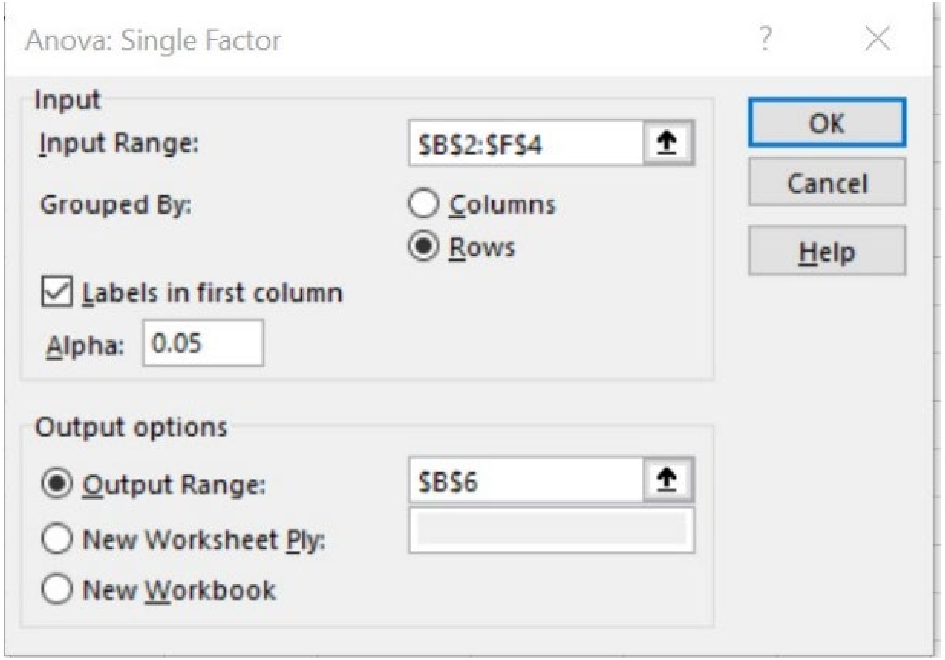
We put the data into excel as shown in the below screenshot

	A	B	C	D	E	F
1						
2		Treatment A	48	49	50	49
3		Treatment B	47	49	48	48
4		Treatment C	49	51	50	50

We choose "anova: single factor" from "data analysis"

A dialogue box appears and it looks like this

1 way ANOVA in Excel



1 way ANOVA in Excel

	A	B	C	D	E	F	G	H
6		Anova: Single Factor						
7								
8		SUMMARY						
9		*Groups*	*Count*	*Sum*	*Average*	*Variance*		
10		Treatment A	4	196	49	0.666667		
11		Treatment B	4	192	48	0.666667		
12		Treatment C	4	200	50	0.666667		
13								
14								
15		ANOVA						
16		*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
17		Between Groups	8	2	4	6	0.022085	4.256495
18		Within Groups	6	9	0.666667			
19								
20		Total	14	11				

Experimental design issues

- in setting up the study / research / data collection to compare groups “experimental units” (people or other things) are randomly assigned to different “treatments” or groups. This is known as a randomised design where the units should be as homogeneous as possible and the treatments allocated at random. We want to achieve a set of observations that are **independent and with constant variance** if we can.

- However, when faced with a set of secondary data in particular, we cannot always be sure that these assumptions are met and so we need to check these

F-test for variance

MACQUARIE University SYDNEY·AUSTRALIA

- ANOVA assumes that data samples are drawn from populations with equal variances. We can use a F-test to test this assumption. In Data analysis of Excel, select ’F test two sample for variances’. This is a two-sample test.

1 Factor ANOVA and multiple linear regression

- The F test result is often provided as a by-product of the estimation result for OLS model.
- We can create a multiple regression by turning the categorical factor into a set of “indicator variables”

Two way analysis of variance without replication

MACQUARIE University SYDNEY·AUSTRALIA

- We may want to analyze the impact of two factors or variables at the same time.
- Although one-way ANOVA can be applied to each factor individually, simultaneuous analysis of two factors help to reduce the overall error rate.
- Analysis for two factors is called two-way ANOVA.
- In two way ANOVA, data should be balanced, i.e. all cells must have the same number of observations.

Two way analysis of variance with replication

- When we have more than one factor, factors may interact and produce 'interaction effect'.
- Some examples of interaction effect include:
  - Interaction between adding sugar to coffee and stirring the coffee. Neither of the two individual variables has much effect on sweetness but a combination of the two does.
  - Interaction between smoking and inhaling asbestos fibres: Both raise lung carcinoma risk, but exposure to asbestos multiplies the cancer risk in smokers and non-smokers. Here, the joint effect of inhaling asbestos and smoking is higher than the sum of both effects.
- When experiments are designed for two factors and have more than one replication, we can evaluate the significance of the interaction effect between the two factors.