max project
maximising the value of survey data in adult social care

QORU
quality and outcomes of person-centred care research unit

Exploring the differences between respondent groups

Clara Heath

University of Kent

LSE The London School of Economics and Political Science

University of Oxford

MAX Toolkit webinar

2nd December 2016

Purpose of the webinar

Provide a brief overview of the variables in the ASCS and PSS SACE

Introduce you to independent t-tests and analysis of variance (ANOVA)

Demonstrate how you can carry out t-tests and ANOVAs in Excel and use the findings to explore the differences between respondent groups

Why spend time on analysis?

'Descriptives' (e.g. frequencies of responses) are commonly used to report survey data but cannot be used to guide local decision-making.

> *It's not about presenting a survey; it's about what do people need to know to develop the business* [Manager]

Statistical tests can be used to explore **relationships between survey variables** and **differences between respondent groups**, and can be supplemented with qualitative data (e.g. respondent comments).

- **Make sense of reported outcomes**
- **Highlight local variations + issues**
- **Inform local performance improvements**
- **Minimise need for further local research**

Analysis tools in the MAX toolkit

The MAX analysis and interpretation guide and associated tools can help users to move beyond the 'descriptives' and conduct more focused – and potentially useful – analyses.

Analysis	No. of Variables	Tests hypotheses
Cross-tabulations	2	No
Chi square	2	Yes
Independent t-tests*	2	Yes
Analysis of variance*	2+	Yes

* These tests can be conducted in a blank Excel sheet. Instructions on how to this are provided in the MAX toolkit.

ASCOF indicators in the ASCS

Data from the ASCS currently populates 8 indicators in the Adult Social Care Outcomes Framework (ASCOF)

ASCOF variables	Question
1A: Social care-related quality of life	3a-9a
1B: The proportion of people who use services who have control over their daily life	3a
1I: Proportion of people who use services and their carers, who reported that they had as much social contact as they would like	8a
1J: Adjusted social care-related quality of life	See 1A
3A: Overall satisfaction of people who use service with their care and support	1
3D: The proportion of people who use services and carers who find it easy to find information about services	12
4A: The proportion of people who use services who feel safe	7a
4B: The proportion of people who use services who say that those services have made them feel safe and secure	7b

Explanatory variables in the ASCS

The ASCS also contains a number of variables that can help LAs to further explore ASCOF indicators. These include:

Explanatory variables	Questions
User characteristics (e.g. age, gender, ethnicity)	Data Return
Self-perceived health	13
Pain + discomfort / anxiety + depression	14
Abilities (ADLs and IADLs)	15-16
Self-perceived design of home	17
Getting around outside of home	18
Receipt of practical help (beyond those provided by LA)	19
Purchase additional / top up care	20

Existing sources of information (e.g. LA records) can also be used to supplement ASCS data.

ASCOF indicators in the PSS SACE

Data from the PSS SACE currently populates 5 indicators in the Adult Social Care Outcomes Framework (ASCOF)

ASCOF variables	Question
1D: Carer-reported quality of life	7-12
1L: Proportion of people who use services and their carers, who reported that they had as much social contact as they would like	11
3B: Overall satisfaction of carers with social services	4
3C: The proportion of carers who report that they have been included or consulted in discussions about the person they care for	18
3D: The proportion of people who use services and carers who find it easy to find information about services	16

Explanatory variables in the PSS SACE

The PSS SACE also contains a number of variables that can help LAs to further explore ASCOF indicators. These include:

Explanatory variables	Questions
Characteristics of carer	
Employment status	19/20
Length of time spent caring	21
Amount of time spent caring each week	22
Self-reported health	24
Age, Gender, Ethnicity	26-28
Characteristics of cared for person	
Age	1
Health conditions	2
Living situation	3

Exploring group differences

Respondent groups – can be characterised by a wide range of features, including:

- **User characteristics** (e.g. age, gender, employment)
- **Survey responses**(e.g. ASCOF indicators, satisfaction)

Potential value of analysis – can help you to **understand data** and **identify which groups report**:

- **Good outcomes + why** → share practice
- **Poor outcomes + why** → inform commissioning of new services; design and delivery of existing services

Examples of LA analysis [ASCS]

Analysis of group differences noted in Adult Social Care Survey reports from 5 LAs during earlier document review.

	Age	Gender	Ethnicity	Primary Client Group	Service Type
1A: SCRQoL	3	1	1	5	3
1B: Control	2	1		2	2
3A: Satisfaction	1			4	1
3D: Finding Info	1			3	1
4A: Safety	2			4	2
4B: Service impact on safety	1			2	1
TOTAL	**10**	**2**	**1**	**20**	**10**

Some of the managers and commissioners interviewed for the project also conducted their own group-level analysis.

Examples of LA analysis [PSS SACE]

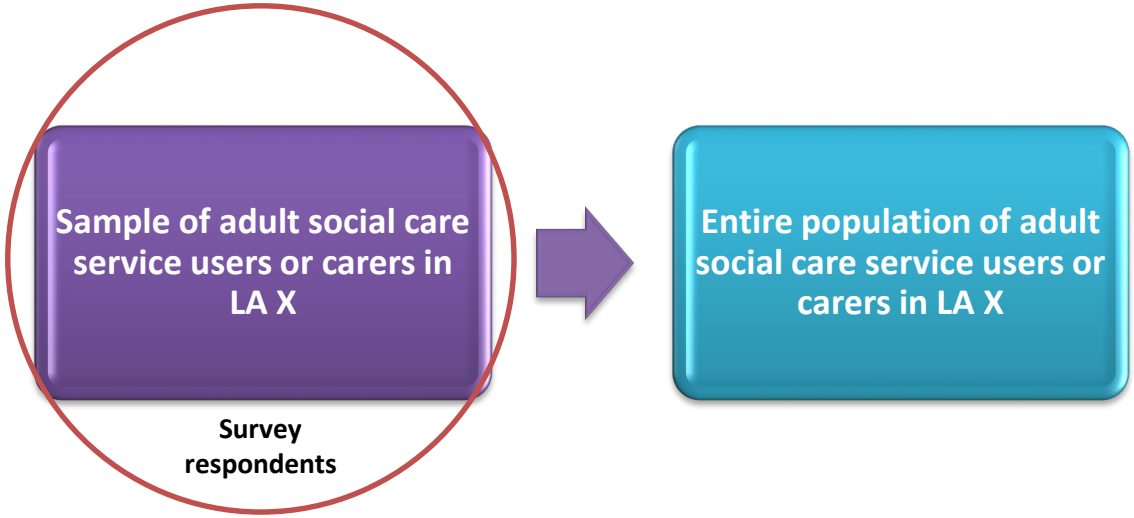
Analysis of group differences noted in Carers' Survey reports from 6 LAs during earlier document review.

	Age		Employ. status	Ethnicity	Health		Length of time caring	No of hrs caring pw
	Carer	CFP			Carer	CFP		
1D: Carer reported QOL	2	1		2	2	2	1	3
3B: Satisfaction	2	1	1		2		1	1
3C: Included or consulted	1	1	1	1	1	2	1	1
3D: Ease of finding info	1	1	1		1	2	1	
TOTAL	**6**	**4**	**3**	**3**	**6**	**6**	**4**	**5**

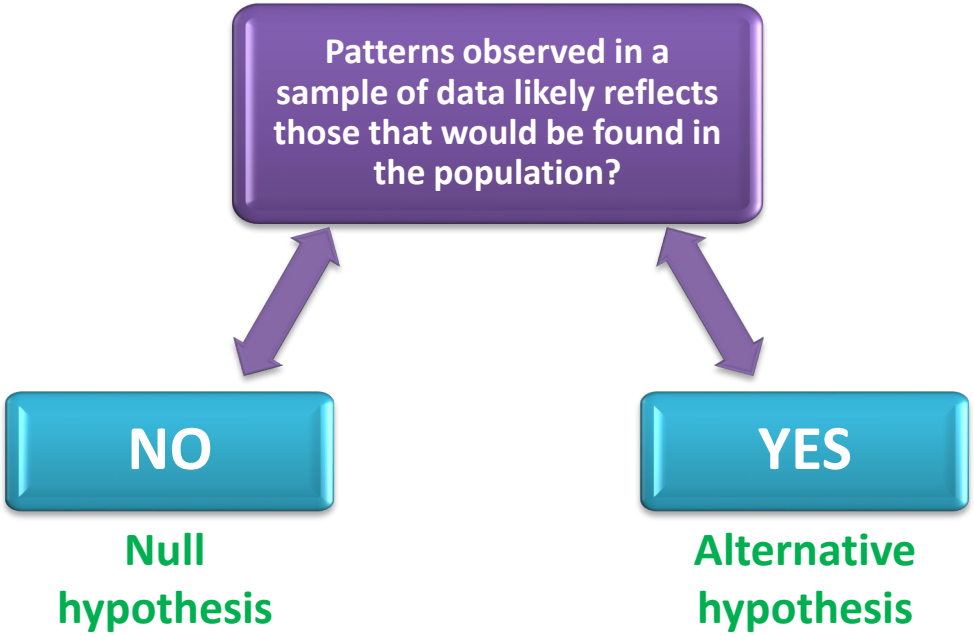
Examples indicate that LA are interested in group differences. Analysis where conducted, however, was generally limited to cross-tabulations.

Hypothesis testing

Involves using **inferential statistical tests** such as **t-tests** or **ANOVA** to determine whether ***the patterns observed in a sample of data are likely to reflect those that would be found in the population***.



Tests explore two hypotheses – the **null hypothesis** and the **alternative hypothesis** – and determines which one should be accepted as true.



Potentially very useful exploring group differences and for improving the local relevance and value of ASCS and PSS SACE data.

Using inferential tests to explore group differences

Independent t-tests

- Explore the differences between **two independent groups** on the same continuous variable
- For example, differences in SCRQoL scores between men and women

Analysis of variance (ANOVA)

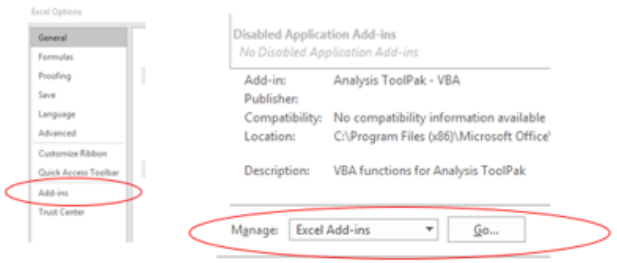
- Explore the differences between **more than two independent groups** on the same continuous variable
- For example, differences in SCRQoL scores between groups classified by satisfaction with services and support

Getting started with analysis

Programme Requirements

Microsoft Excel 2010 or later + Analysis ToolPak. See **installing the analysis toolpak and real statistics resource pack** in MAX toolkit.

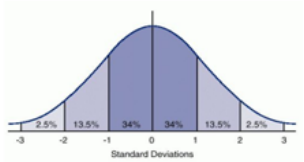
1. Click the **File** tab, and then click **Options**.
2. Click **Add-Ins**, and then in the **Manage** box, select **Excel Add-ins** and then select **Go**.



Knowledge Requirements

Understanding of basic statistical terminology. See **getting started with statistics** in MAX toolkit.

central limit theorem)[2] that a distribution will fall symmetrically around the mean to produce a **bell-shaped curve** if a sufficient sample is drawn. This means that most values will be grouped near the centre of the distribution and the remaining values will tail off away from the mean in equal measures.



The **normal distribution** which produces a bell-shaped curve and plots the percentage of the data-set that should fall within a given range.

INDEPENDENT T-TESTS

Assumptions

Criteria	Details
DV can be measured on a continuous scale	SCRQoL, Carer-QOL and age are all continuous variables.
IV is categorical and independent	With the exception of SCRQoL, Carer QOL and age, all variables in surveys are categorical.
Independence of observations	All responses to survey are independent (i.e. respondents provide one response for each question)
Normal distribution	Observations are normally distributed. Tested during Step 2.
Homogeneity of variances	Variation in each group is approximately equal. Tested during Step 3.

Example question

Do overall social care-related quality of life (SCRQoL) scores (ASCOF 1A) differ between men and women?

Dependent Variable

- **Social care-related quality of life** (SCRQoL)
- Variable is measured on a continuous scale (0 – 24)

Independent Variable

- **Gender** (men | women)
- Variable is categorical and independent

Conducting t-tests in Excel

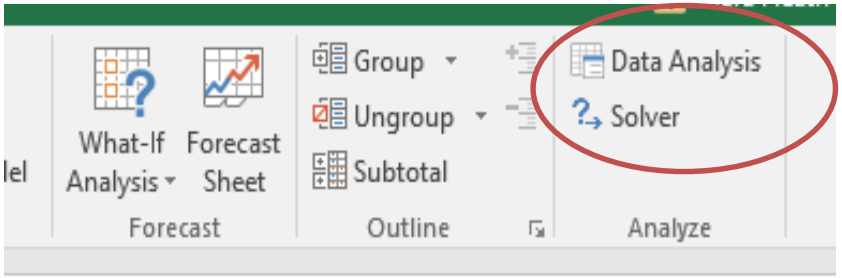
Step 1: copy + paste relevant data from your NHS Digital data return into a blank Excel sheet.

Put data in separate columns

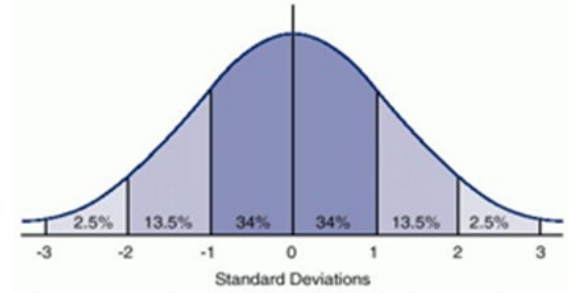
Men	Women
6	8
7	8
7	8
8	8
8	9
8	9

Include labels for the groups

Step 2: check whether data is **normally distributed** [go to Data tab].



The Normal Distribution [1/3]



The **normal distribution** which produces a bell-shaped curve and plots the percentage of the data-set that should fall within a given range.

Also known as the **bell-shaped curve**. A frequency distribution of a set of independent, randomly generated variables where:

- most values are grouped near the centre
- remaining values tail off away from the centre in equal measures
- mean, median and mode are the same

The Normal Distribution [2/3]

Based on **central limit theorem** which states that the averages (mean) of a number of variables will become normally distributed if the sample is sufficiently large.

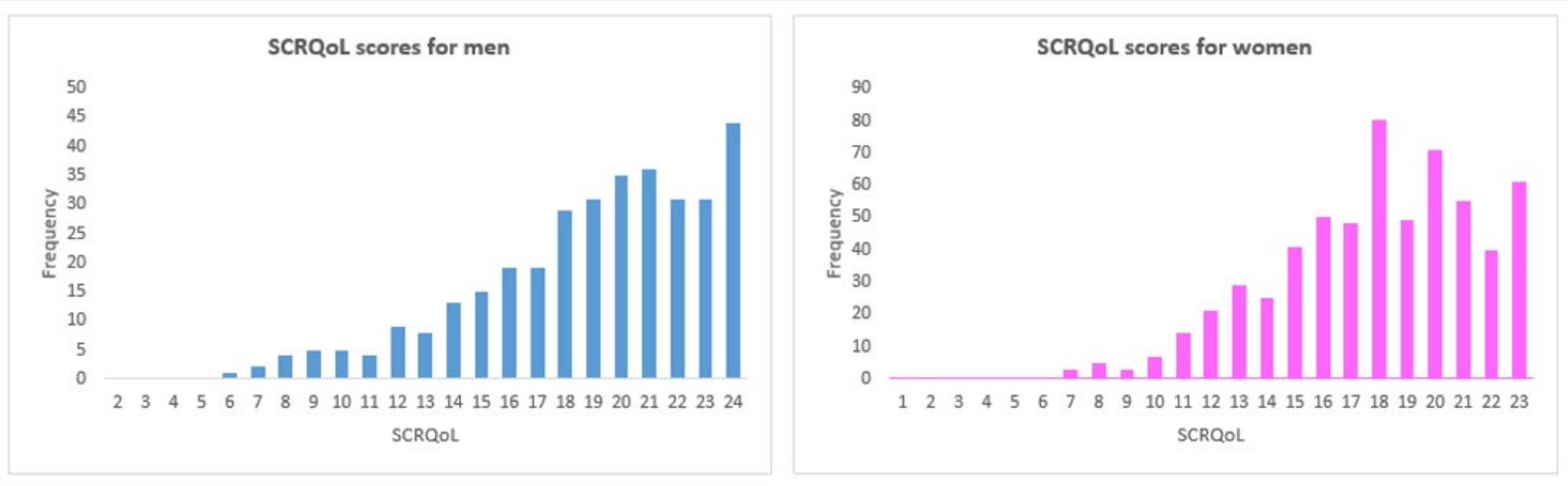
An important concept in **inferential statistics** as parametric tests which compare sample means (e.g. t-tests, ANOVA) assume that data is normally distributed.

	Descriptive	Inferential
Dataset	Population	Sample
Purpose	Describe data	Make predictions

The Normal Distribution [3/3]

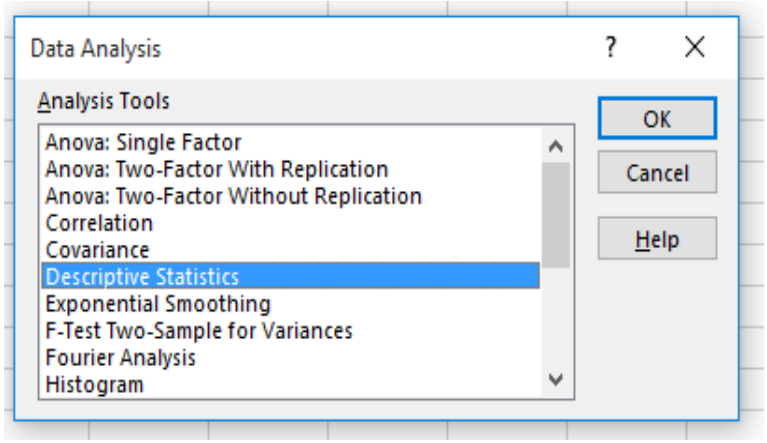
Current sample: 984 respondents [341 (men) + 603 (women)]

Distribution: SCRQoL scores (0-24)



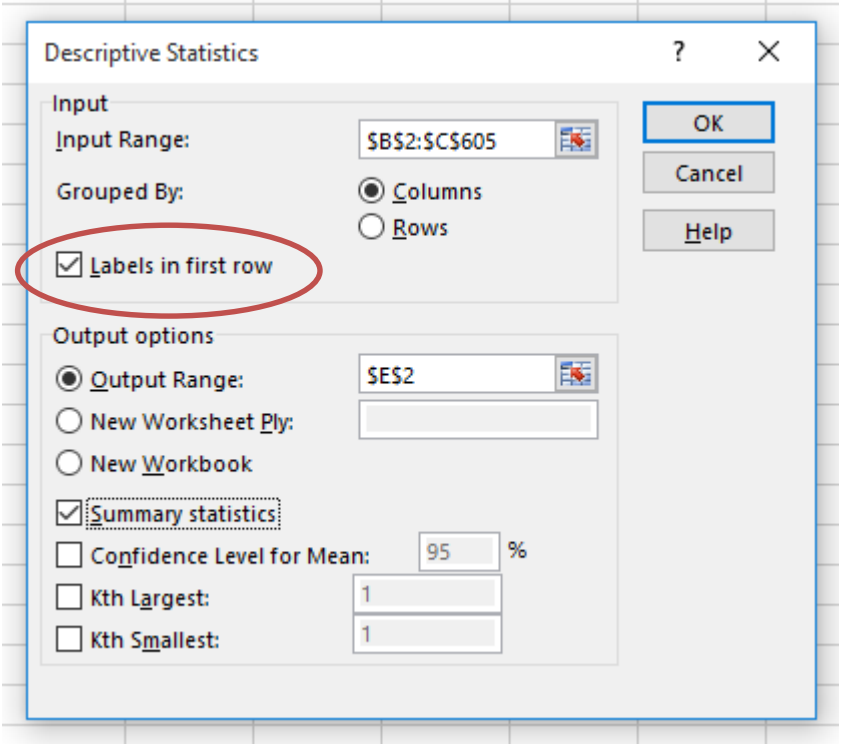
Histogram shows that data are not normally distributed but descriptive measures can be used to establish whether parametric or non-parametric statistical tests should be used.

Select **Descriptive Statistics** and press OK.



Complete fields in Descriptives Statistics window.

Remember to **select labels in first row**



Look at **skewness** in the output data.

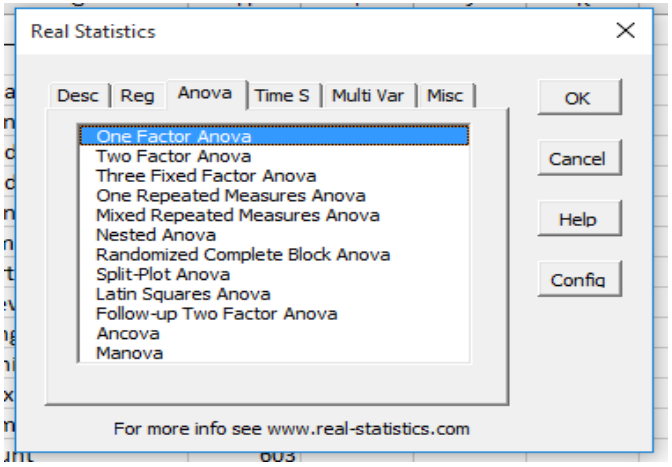
Men		*Women*	
Mean	18.92962	Mean	18.8408
Standard Error	0.2222	Standard Error	0.14693
Median	20	Median	19
Mode	24	Mode	19
Standard Deviation	4.103195	Standard Deviation	3.608019
Sample Variance	16.83621	Sample Variance	13.0178
Kurtosis	0.208988	Kurtosis	-0.15424
Skewness	-0.85272	Skewness	-0.57416
Range	18	Range	16
Minimum	6	Minimum	8
Maximum	24	Maximum	24
Sum	6455	Sum	11361
Count	341	Count	603

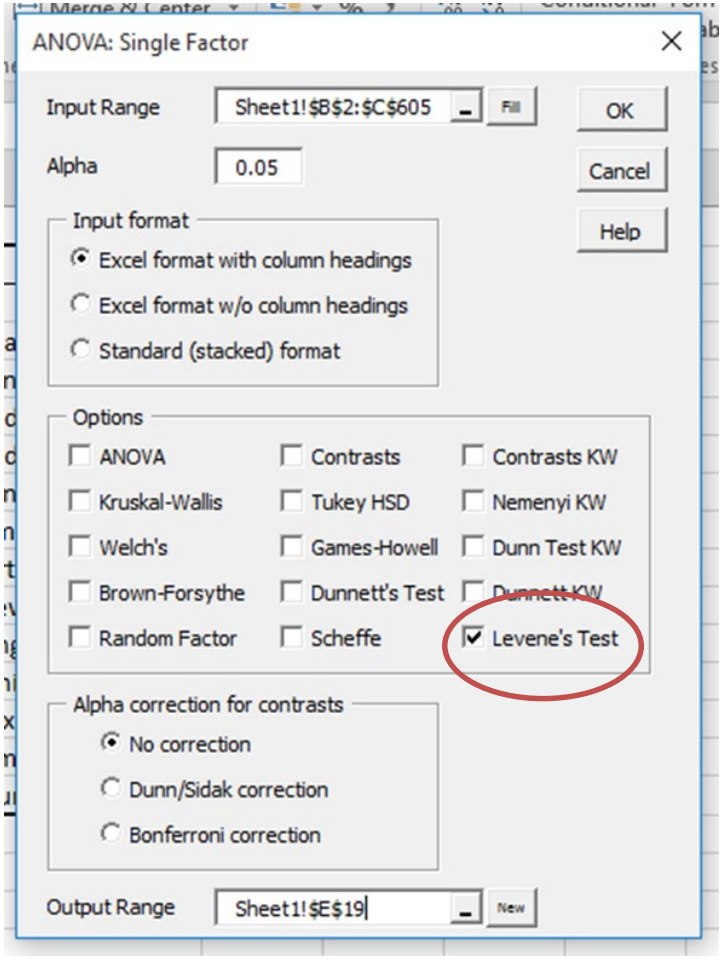
Skewness measures the symmetry of the frequency distribution. Perfect symmetry = 0. See **Getting Started with Statistics** for further information.

Data is not normally distributed so assumption is violated. In this instance, you will need to run the non-parametric equivalent [**a Mann Whitney U Test**].

Most, if not all, respondents generally report good quality of life. These distributions will therefore be negatively skewed.

Step 3: check whether the group variances are **approximately equal**. Open **Real Statistics window** – CTRL + M – select **One Factor ANOVA**





Complete fields in **ANOVA: single factor window**.

Remember to **select Levene's Test**

Look at **means p-value** in the output data.

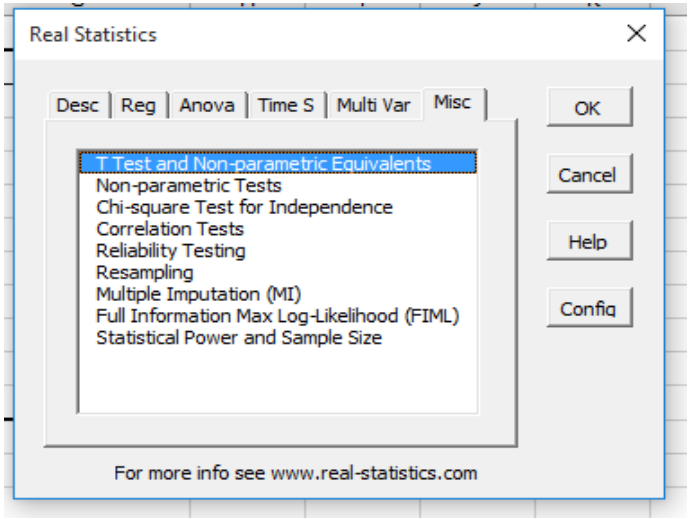
Levene's Tests	
type	p-value
means	0.016965
medians	0.036295
trimmed	0.018075

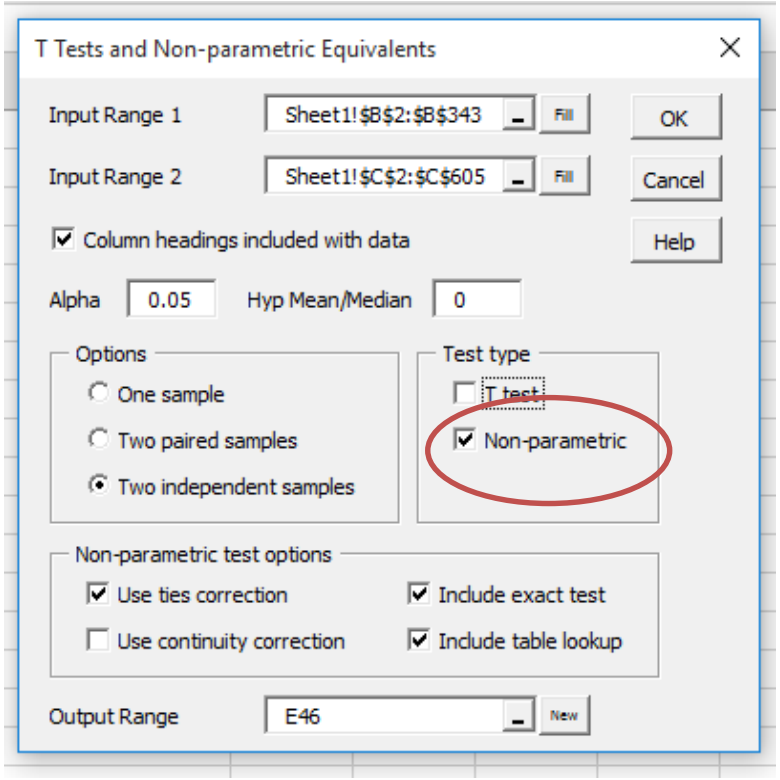
P-value is less than 0.05. Variance in groups is not equal so assumption is violated.

Overview of t-test options

	Normal Distribution		Homogeneity of Variance	
Test	**Yes**	**No**	**Yes**	**No**
T- test with equal variances assumed	✓		✓	
T-test with equal variances not assumed	✓			✓
Mann Whitney U Test		✓	✓	✓

Step 4: run **t-test** [**Mann Whitney U-test**]. Open **Real Statistics window** – CTRL + M – click on the **Misc** tab and select **T-test and Non-parametric equivalents**.





Complete fields in **T-test + non-parametric equivalents window**.

Remember to **select Non-parametric**

Look at **p-value** in the output data.

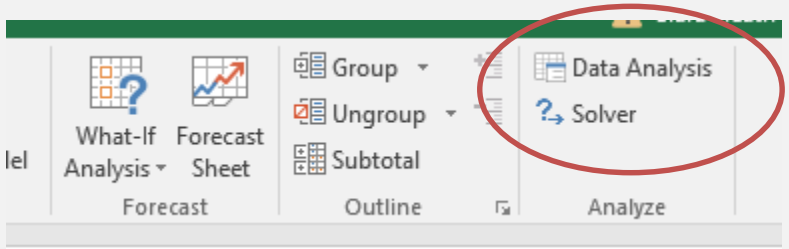
Unless you have specified the direction of the difference between your groups, you will look at the **two tail value**

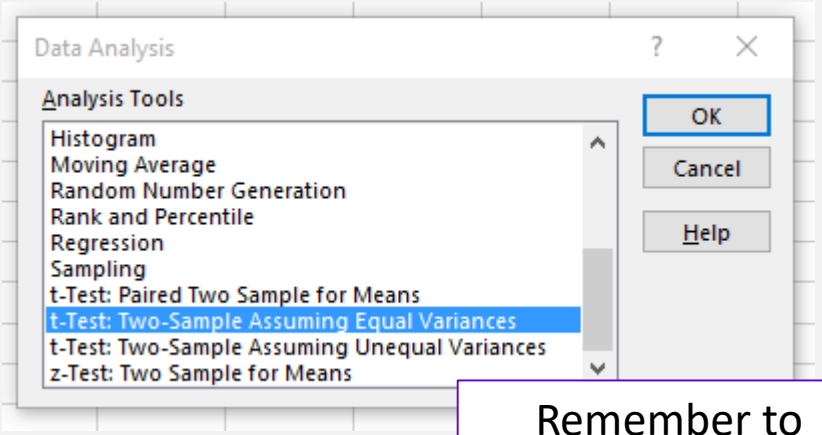
Mann-Whitney Test for Two Independent Samples		
	Men	Women
count	341	603
median	20	19
rank sum	165680.5	280359.5
U	98253.5	107369.5
	one tail	two tail
alpha	0.05	
U	98253.5	
mean	102811.5	
std dev	4008.417	ties
z-score	1.137107	
effect r	0.03701	
U-crit	96218.24	94955.14799
p-value	0.127747	0.255493399
sig (norm)	no	no

P-value is more than 0.05. Differences between groups is not statistically significant [**Note**: this is confirmed in last row].

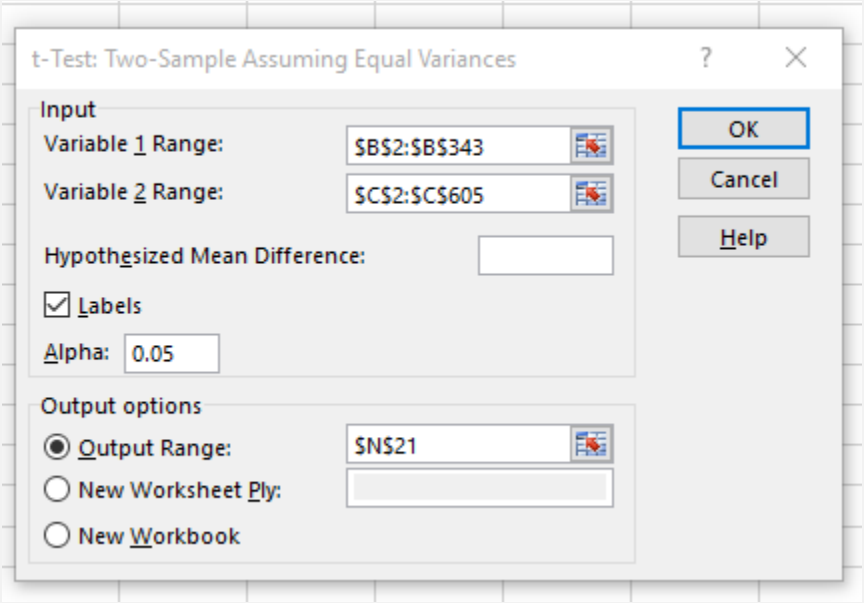
Conducting parametric t-tests in Excel

Parametric versions of t-tests can be conducted using the Analysis Toolkpak.





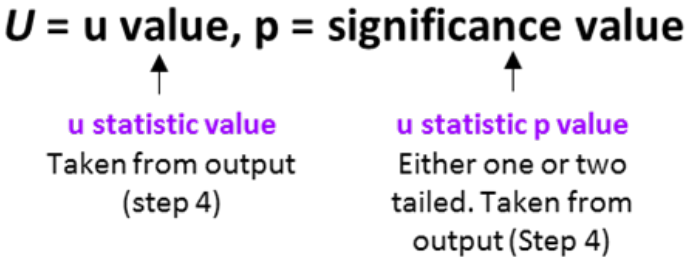
Remember to **select** the most appropriate t-test [see Step 3]



See the **step-by-step instructions** for further guidance on how to do this

Reporting results of t-test analysis

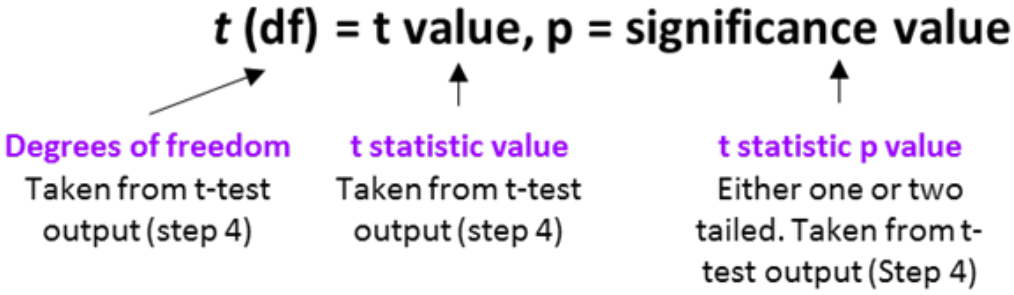
The usual format for reporting the results of a **Mann Whitney u-test** is:



You should also include the **median** for each group. For example,

> ***Social care-related quality of life (SCRQoL) scores [ASCOF 1A] for men (Mdn = 20) did not differ significantly from women (Mdn = 19) (U = 94955.47, p = 0.26).***

The usual format for reporting the results of a **t-test** is:



You should also include the **mean** and **standard error** for each group. For example,

Men who responded to the adult social care survey did not report significantly different social care-related quality of life (SCRQoL) [ASCOF 1A] (M = 18.9, SE = 0.22) than women (M = 18.8, SE = 0.15), (t (634) = 0.33, p = 0.74).

ANALYSIS OF VARIANCE (ANOVA)

Assumptions

Criteria	Details
DV can be measured on a continuous scale	SCRQoL, Carer-QOL and age are all continuous variables.
IV is categorical and independent	With the exception of SCRQoL, Carer QOL and age, all variables in surveys are categorical.
Independence of observations	All responses to survey are independent (i.e. respondents provide one response for each question)
Normal distribution	Observations are normally distributed. Tested during Step 2.
Homogeneity of variances	Variation in each group is approximately equal. Tested during Step 3.

Example question

Do overall social care-related quality of life (SCRQoL) scores (ASCOF 1A) differ by rating of satisfaction with services?

Dependent Variable

- **Social care-related quality of life** (SCRQoL)
- Variable is measured on a continuous scale (0 – 24)

Independent Variable

- **Satisfaction with services**
  - Extremely satisfied
  - Very satisfied
  - Quite satisfied
  - Neither satisfied nor dissatisfied
  - Quite dissatisfied
  - Very dissatisfied
  - Extremely dissatisfied
- Variable is categorical and independent

Conducting ANOVA in Excel

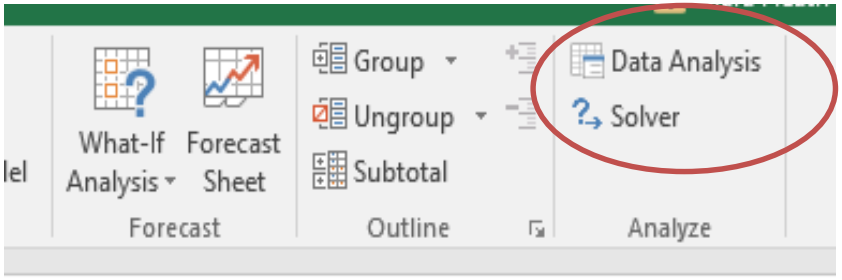
Step 1: copy + paste relevant data from your NHS Digital data return into a blank Excel sheet.

Put data in separate columns

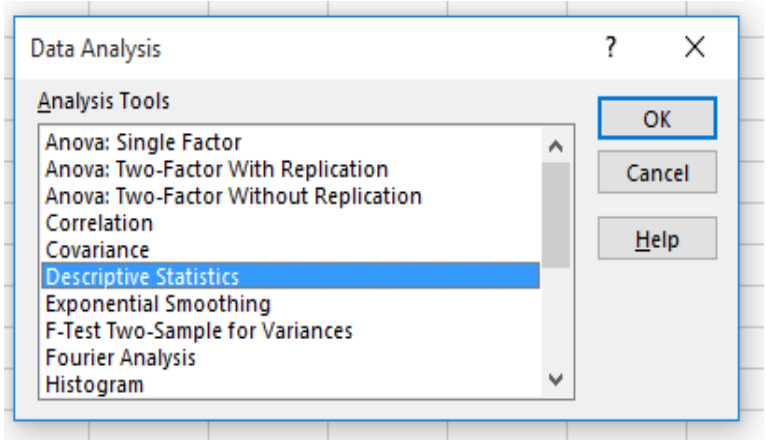
Extremely/ very satisfied	Quite satisfied	Neither	Quite, very or extremely dissatisfied
10	7	8	8
11	8	8	9
12	9	9	9
12	9	10	9
12	9	10	10
12	9	11	11

Include labels for the groups

Step 2: check whether data is **normally distributed** [go to Data tab].

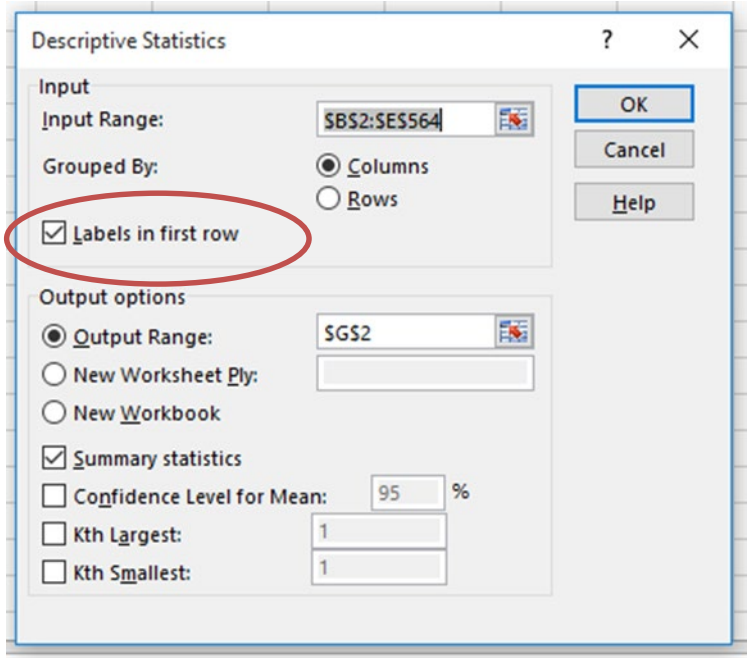


Select **Descriptive Statistics** and press OK.



Complete fields in Descriptives Statistics window.

Remember to **select labels in first row**



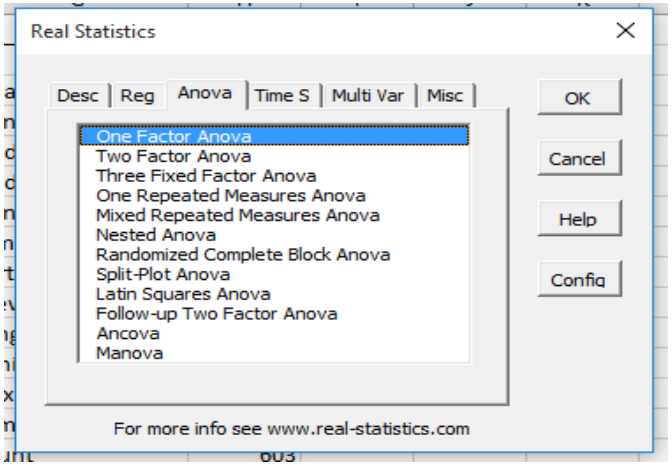
Look at **skewness** in the output data.

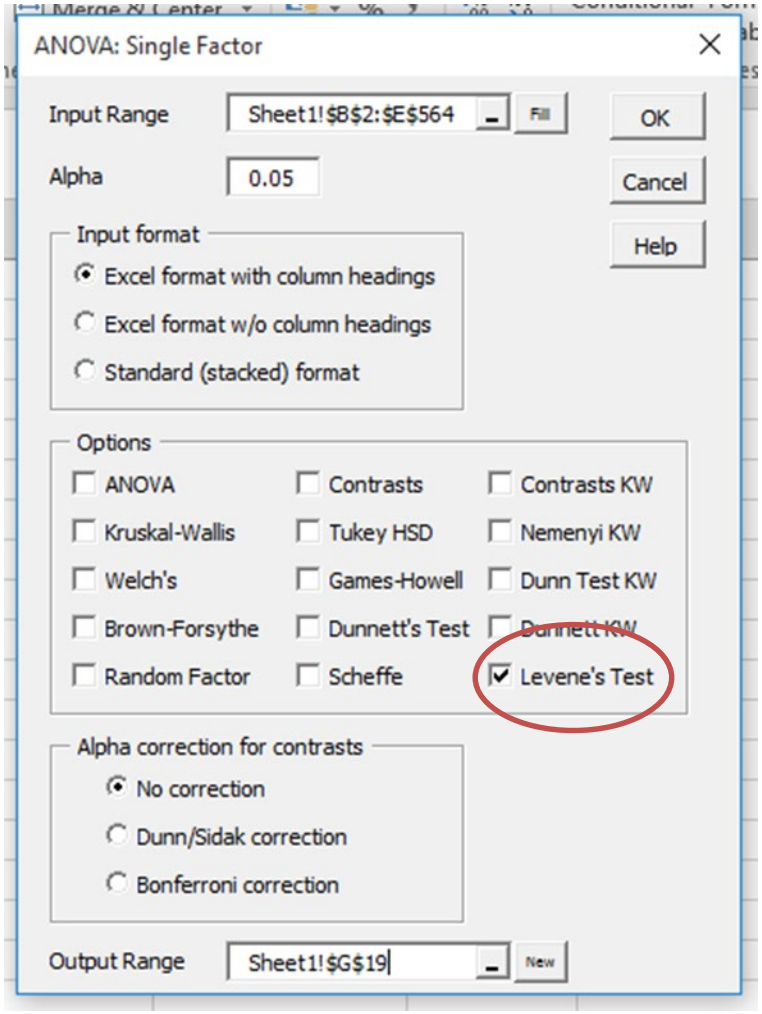
Extremely/very satisfied		Quite satisfied		Neither		Quite, very or extremely dissatisfied	
Mean	20.19	Mean	17.46	Mean	16.77	Mean	14.62
Standard Error	0.12	Standard Error	0.23	Standard Error	0.56	Standard Error	0.71
Median	21.00	Median	18.00	Median	17.00	Median	13.50
Mode	24.00	Mode	18.00	Mode	15.00	Mode	13.00
Standard Deviation	2.96	Standard Deviation	3.70	Standard Deviation	4.20	Standard Deviation	4.63
Sample Variance	8.73	Sample Variance	13.68	Sample Variance	17.60	Sample Variance	21.46
Kurtosis	-0.16	Kurtosis	-0.36	Kurtosis	-0.79	Kurtosis	-0.65
Skewness	-0.62	Skewness	-0.34	Skewness	-0.29	Skewness	0.38
Range	14	Range	17	Range	16	Range	17
Minimum	10	Minimum	7	Minimum	8	Minimum	7
Maximum	24	Maximum	24	Maximum	24	Maximum	24
Sum	11348	Sum	4523	Sum	939	Sum	614
Count	562	Count	259	Count	56	Count	42

Data is not normally distributed so assumption is violated. In this instance, you will need to run the non-parametric equivalent [**a Kruskal Wallis Test**].

Most, if not all, respondents generally report good quality of life. These distributions will therefore be negatively skewed.

Step 3: check whether the group variances are **approximately equal**. Open **Real Statistics window** – CTRL + M – select **One Factor ANOVA**





Complete fields in **ANOVA: single factor window**.

Remember to **select Levene's Test**

Look at **means p-value** in the output data.

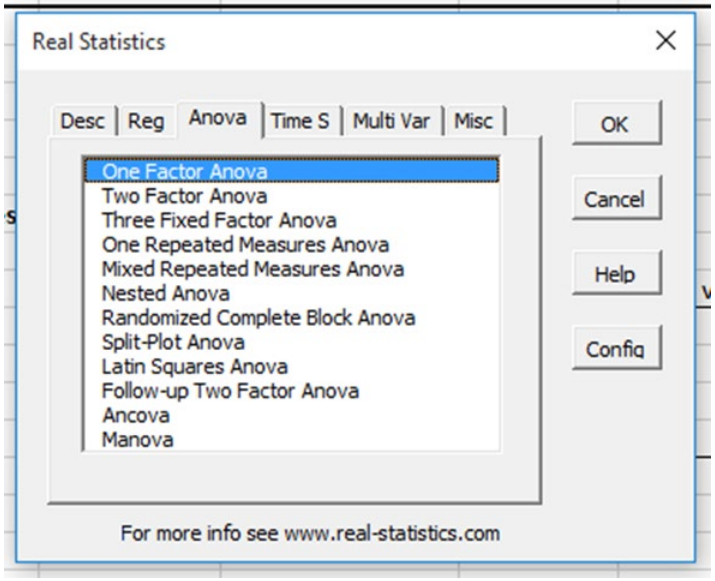
Levene's Tests	
type	p-value
means	0.00
medians	0.00
trimmed	0.00

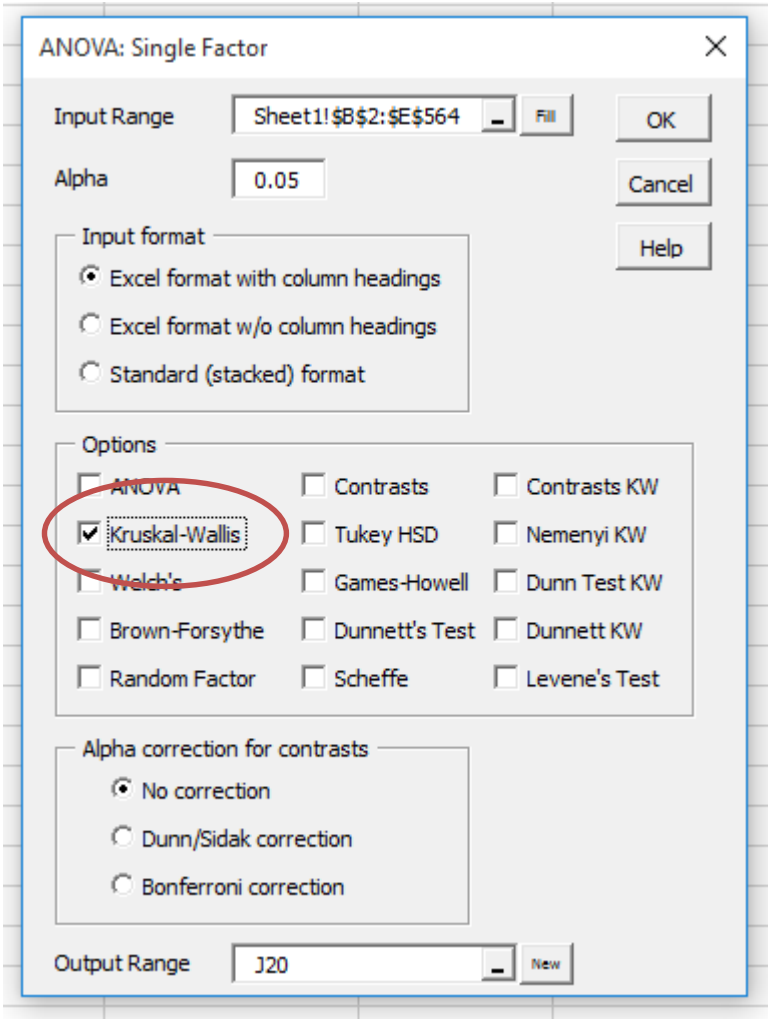
P-value is less than 0.05. Variance in groups is not equal so assumption is violated.

Overview of ANOVA options

	Normal Distribution		Homogeneity of Variance	
Test	**Yes**	**No**	**Yes**	**No**
Single factor ANOVA	✓		✓	
Kruskal Wallis Test		✓	✓	✓

Step 4: run **t-test** [**Kruskal-Wallis test**]. Open **Real Statistics window** – CTRL + M – click on the **Anova** tab and select **One Factor Anova**.





Complete fields in **ANOVA: Single Factor window**.

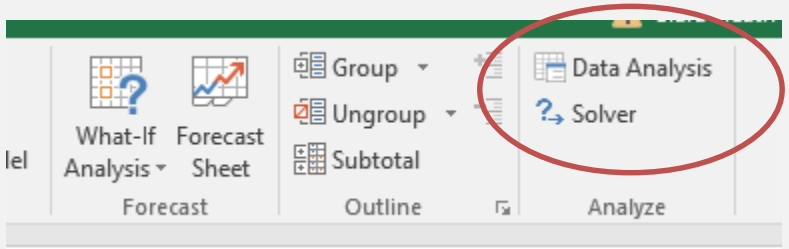
Look at **p-value** in the output data.

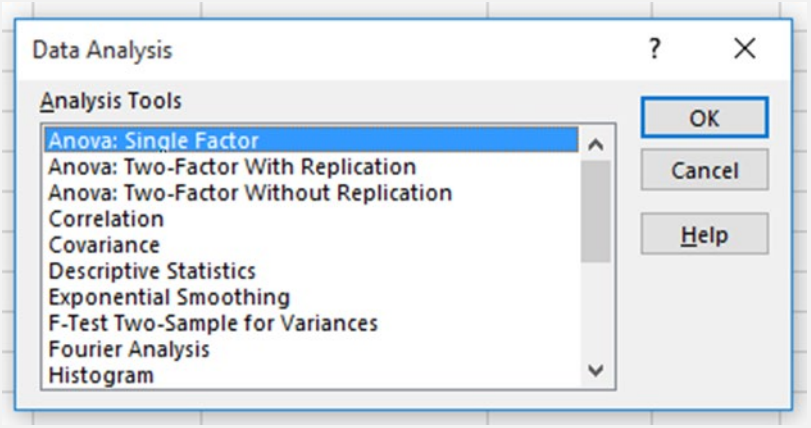
Kruskal-Wallis Test					
	Extremely/very	Quite satisfied	Neither	Quite, very or extremely dissatisfied	
median	21	18	17	13.5	
rank sum	305066.5	90581	17997.5	9095	
count	562	259	56	42	919
r^2/n	165597098.6	31679218.38	5784107.254	1969500.595	205029924.8
H-stat					150.01
H-ties					151.21
df					3.00
p-value					0.00
alpha					0.05
sig					yes

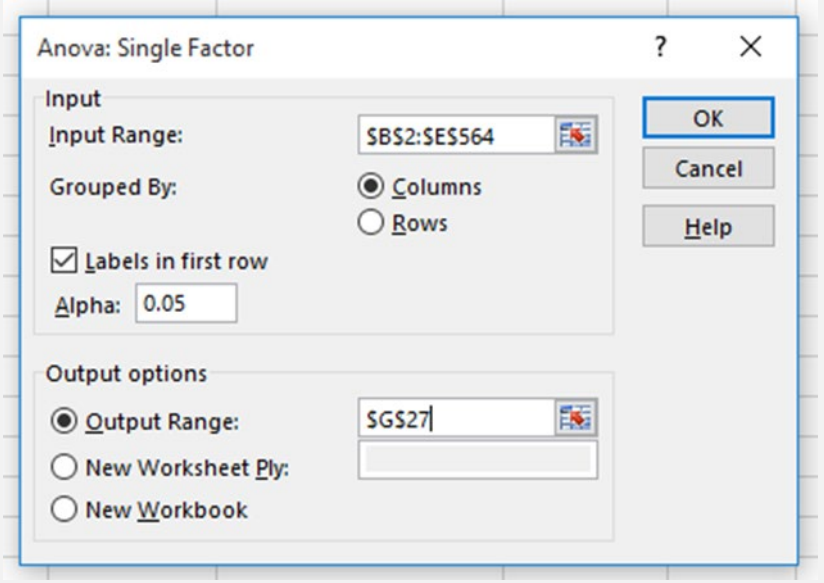
P-value is less than 0.05. Differences between groups is statistically significant [**Note**: this is confirmed in last row].

Conducting parametric ANOVA in Excel

Parametric versions of ANOVA can be conducted using the Analysis Toolkpak.



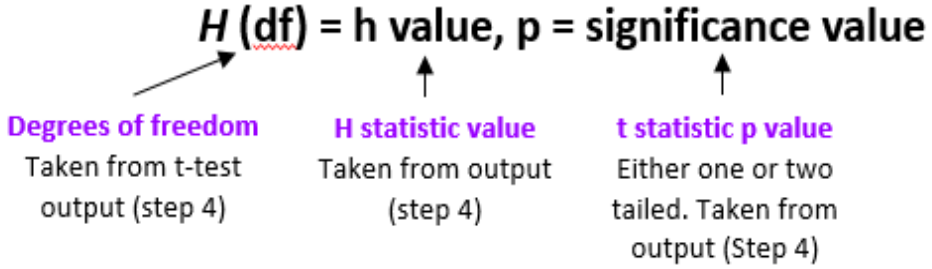




See the **step-by-step instructions** for further guidance on how to do this

Reporting results of ANOVA

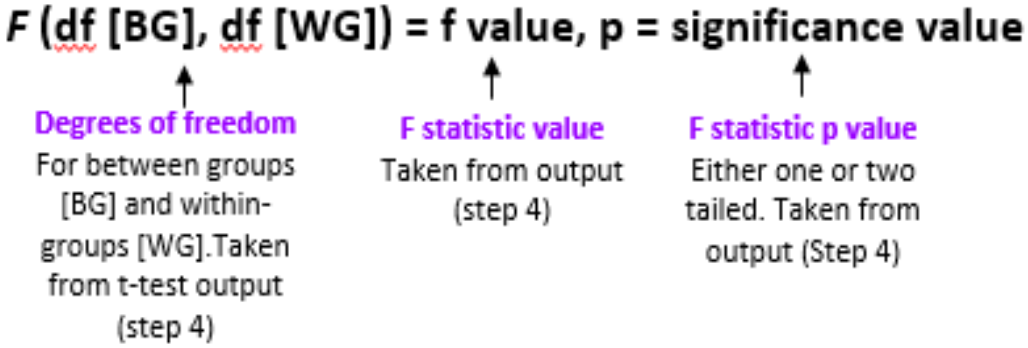
The usual format for reporting the results of a **Kruskal Wallis test** is:



Try to report in everyday terms. For example,

Social care related quality of life (SCRQoL) [ASCOF 1A] is significantly affected by satisfaction with services, H (3) = 150.01, p = 0.00.

The usual format for reporting the results of a **single factor ANOVA** is:



Try to report in everyday terms. For example,

Social care related quality of life (SCRQoL) [ASCOF 1A] is significantly affected by ratings of satisfaction with services, *F* (3,915) = 74, p = 0.00.

Conducting post-hoc analysis

Post-hoc t-tests can be conducted to **compare individual groups** (e.g. extremely / very satisfied vs. quite satisfied) and may help you to further **understand statistically significant differences**.

Guidance on how to conduct post-hoc t-tests is provided in step-by-step instructions, available in the MAX toolkit.

Further Information

To find out more about the MAX project, download the reports on earlier research activities or access the MAX toolkit:

Website: www.maxproject.org.uk

Email: maxproject@kent.ac.uk

Disclaimer

Department of Health and Social Care disclaimer: The MAX toolkit and website are based on independent research commissioned and funded by the NIHR Policy Research Programme (Maximising the value of survey data in adult social care (MAX) project and the MAX toolkit implementation and impact project). The views expressed on the website and in publications are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or its arm's length bodies or other government departments.