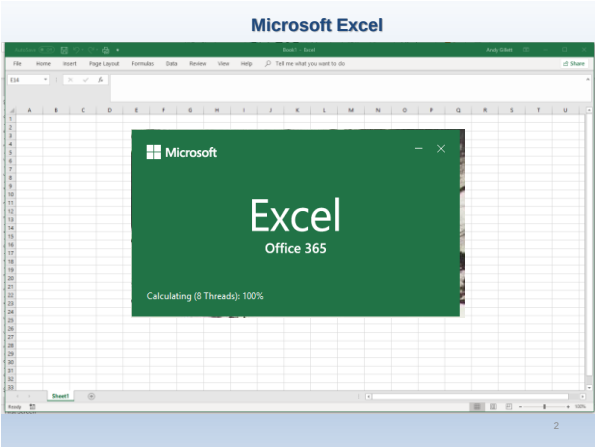
# Research Methods

## *Introduction to Statistics with EXCEL*

Andy Gillett

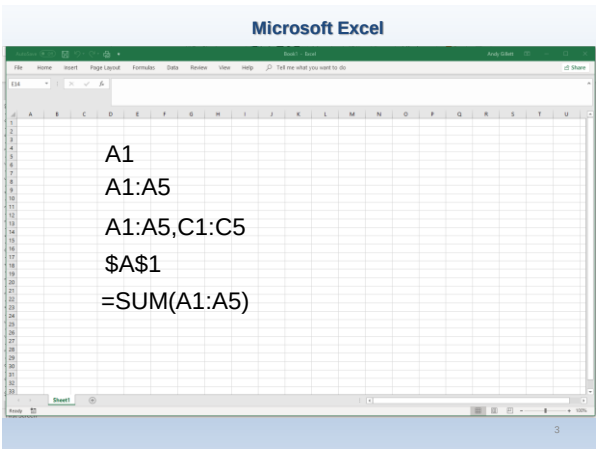
1

1



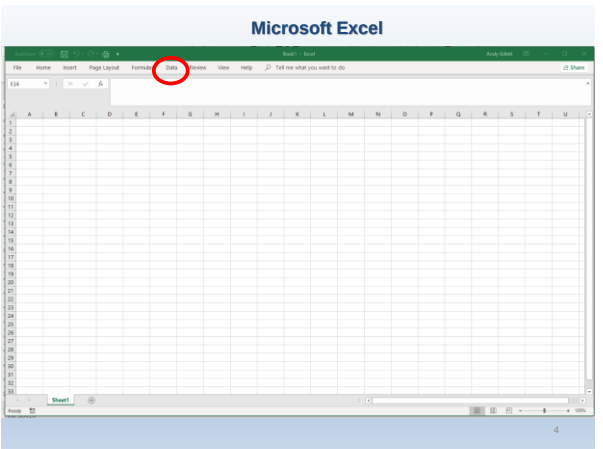

2

2



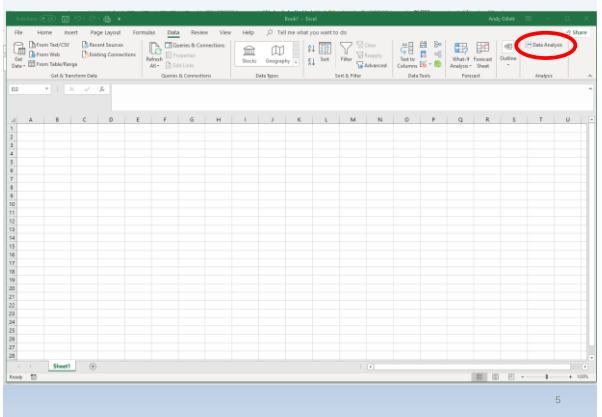

3

3



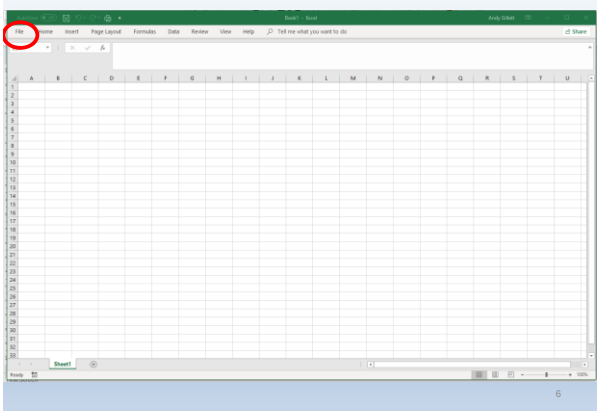

4

4



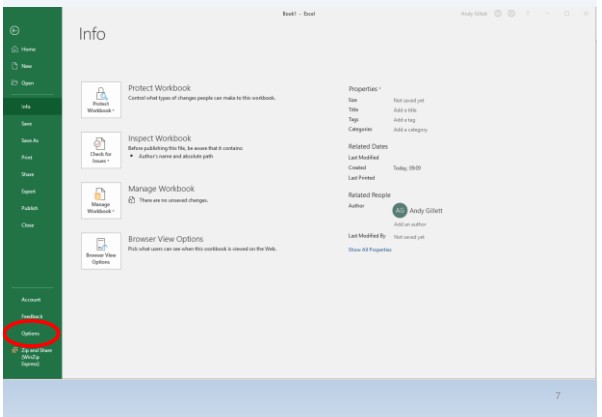
5

5



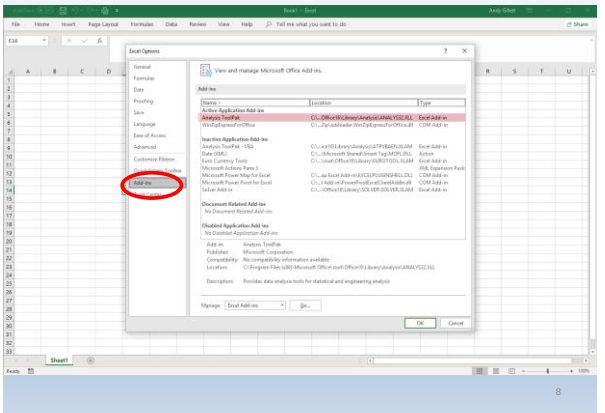
6

6



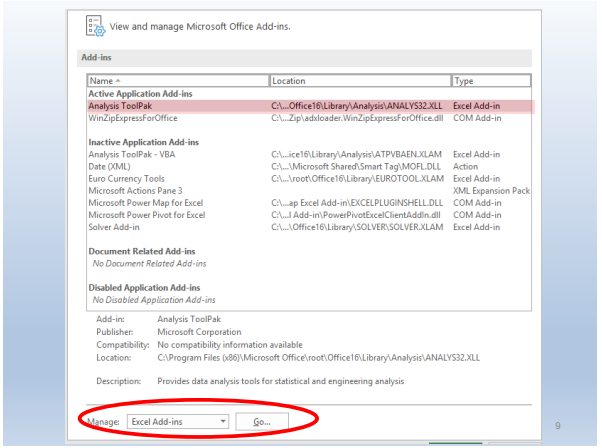

7

7



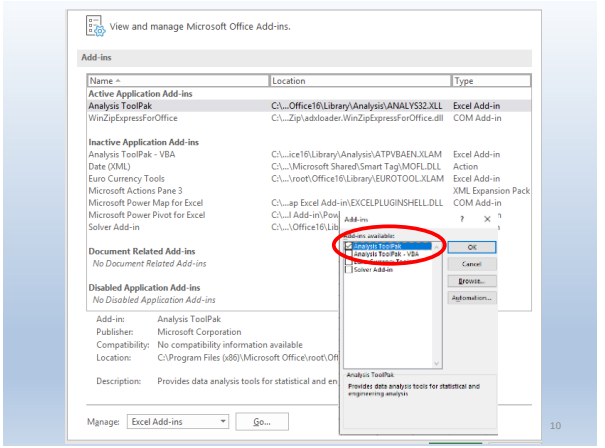

8

8



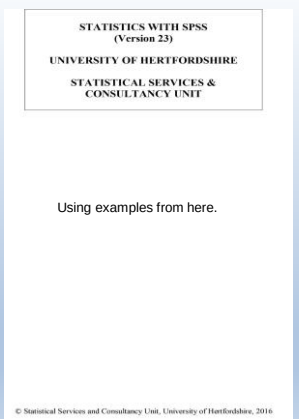

9

9




10

10


STATISTICS WITH SPSS
(Version 23)

UNIVERSITY OF HERTFORDSHIRE

STATISTICAL SERVICES &
CONSULTANCY UNIT

Using examples from here.



11

11

## Describing Data using Statistics

| Descriptive | Inferential |
|---|---|
| Description | Inference |
| Sample | Population |

"When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have."

(Popham & Sirotnik, 1973: 40)

12

12

## Describing Data using Statistics

**Description**

- ❑ Central Tendency
  - Mean, Median, Mode
- ❑ Dispersion
  - Range, Standard Deviation, IQR

**Inference**

- ❑ Exploring Relationships
  - Correlation
  - Regression
- ❑ Comparing Groups
  - t-Test
  - ANOVA

13

13

## Describing Data using Statistics

**Description:**

- The central tendency, which is the common, middle or average value.
- The dispersion, which is how the data values are spread, or dispersed, around the central tendency.

14

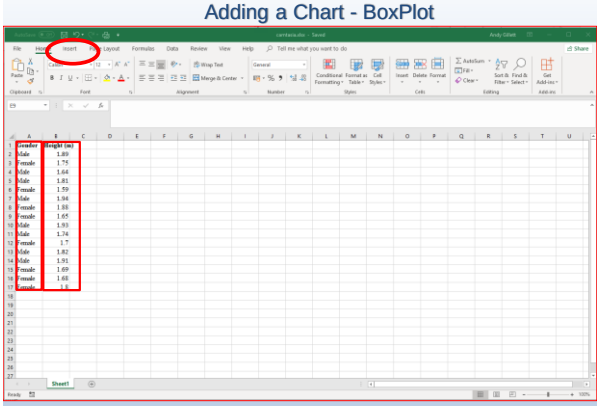
14

| Gender | Height (m) |
|---|---|
| Male | 1.89 |
| Female | 1.75 |
| Male | 1.64 |
| Male | 1.81 |
| Female | 1.59 |
| Male | 1.94 |
| Female | 1.88 |
| Female | 1.65 |
| Male | 1.93 |
| Male | 1.74 |
| Female | 1.70 |
| Male | 1.82 |
| Male | 1.91 |
| Female | 1.69 |
| Female | 1.68 |
| Female | 1.80 |

15

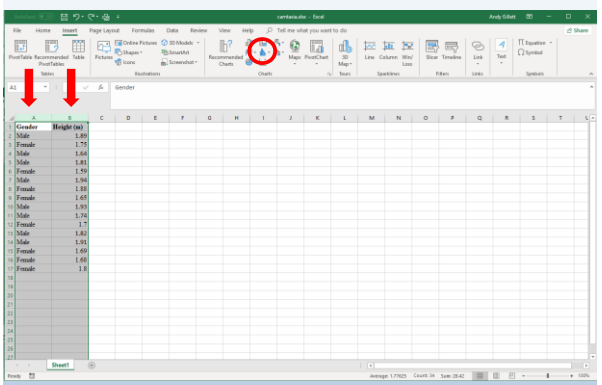
15

## Adding a Chart - BoxPlot



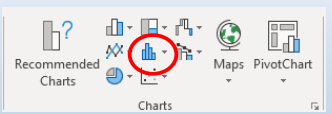
16

16



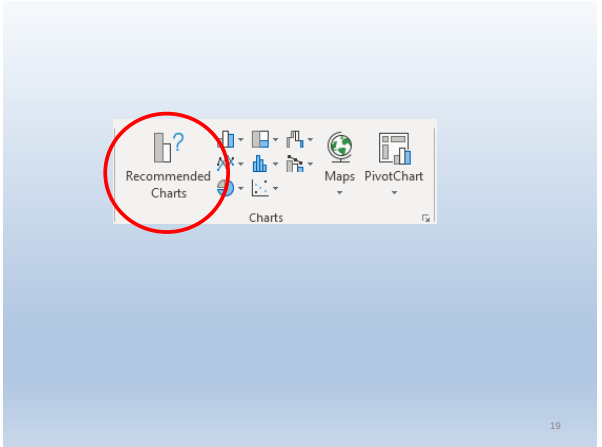
17

17

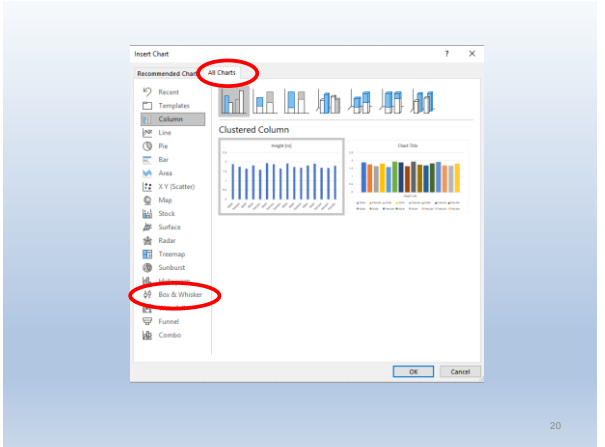



18

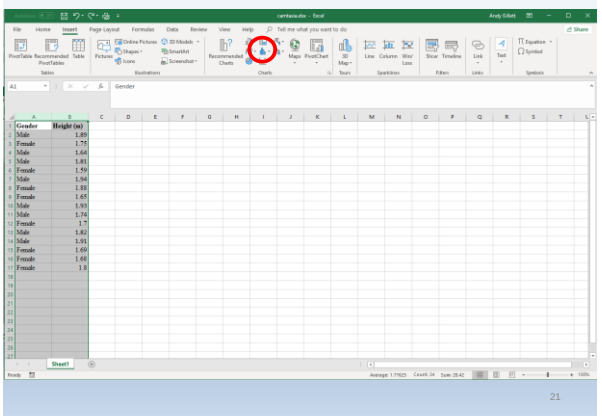
18



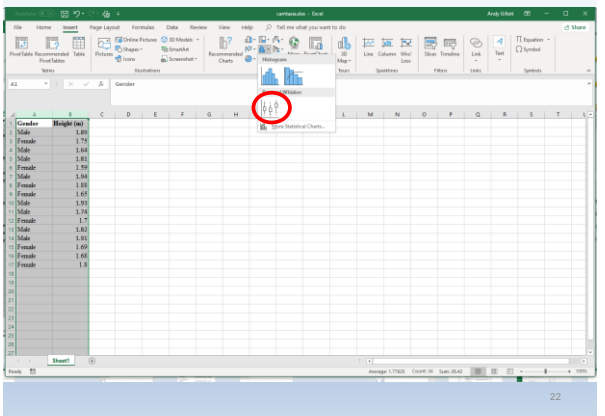

19



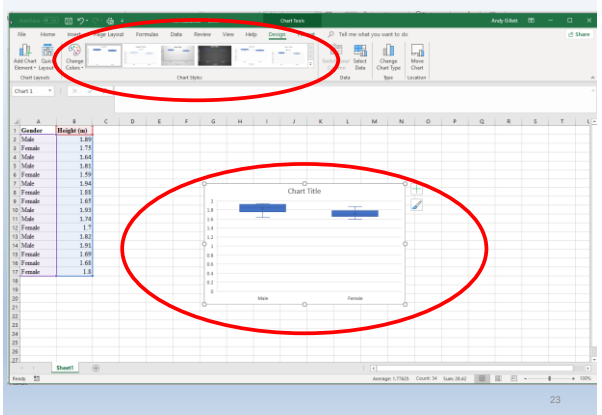

20



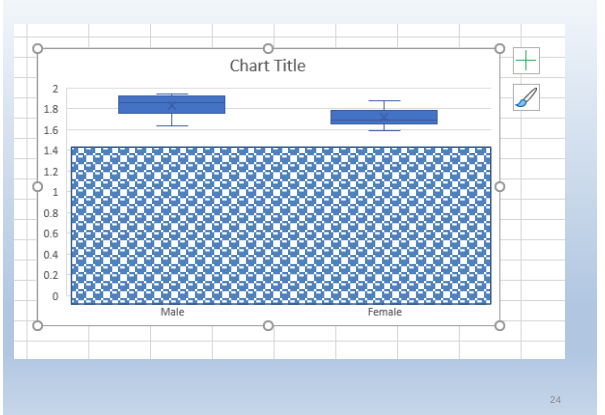
21



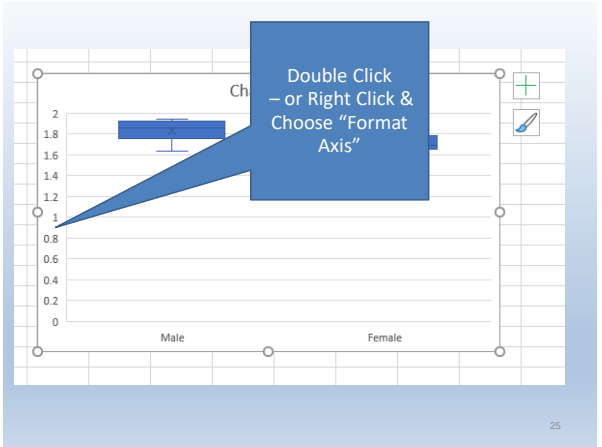
22



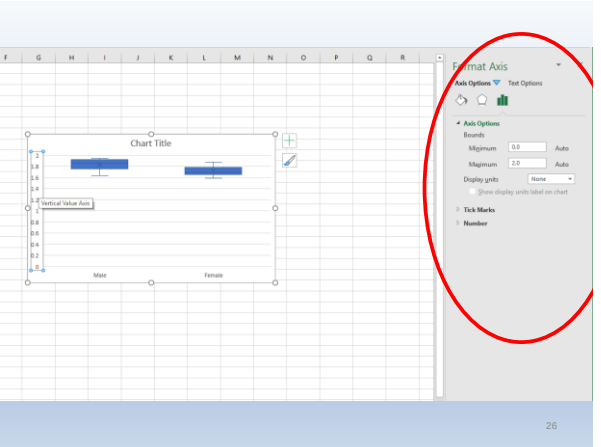

23



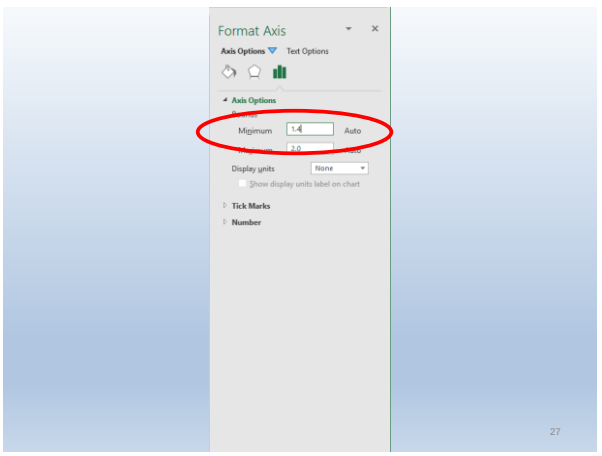

24



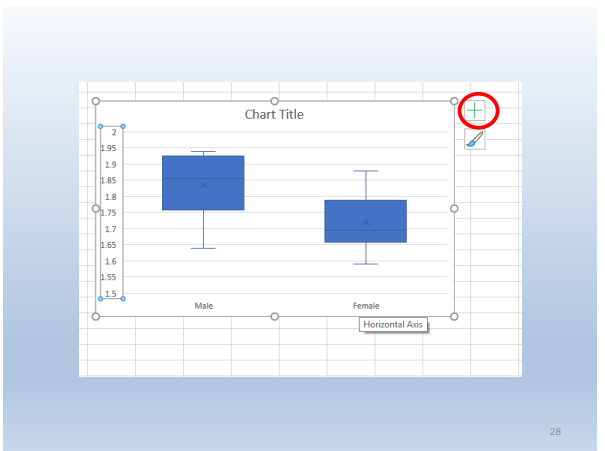

25




26

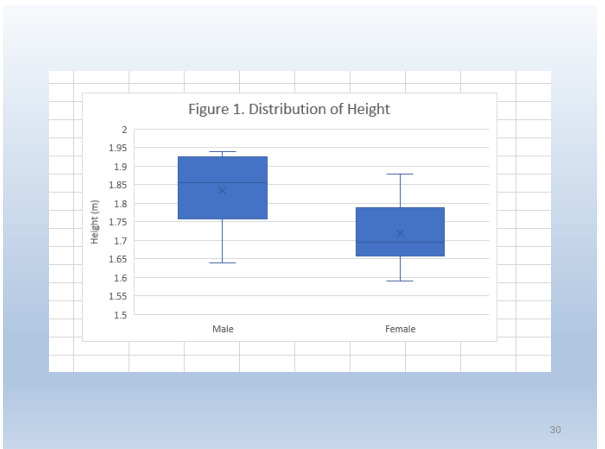



27

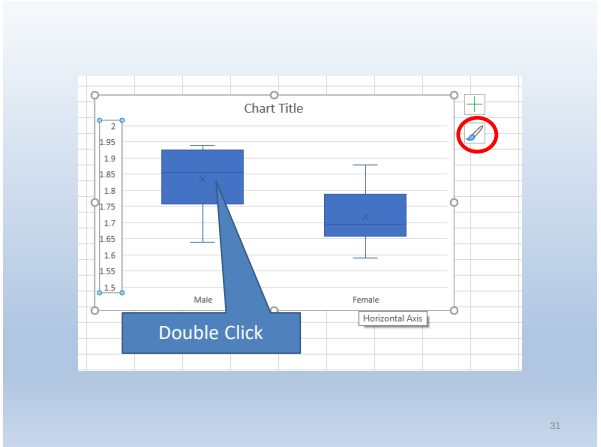



28

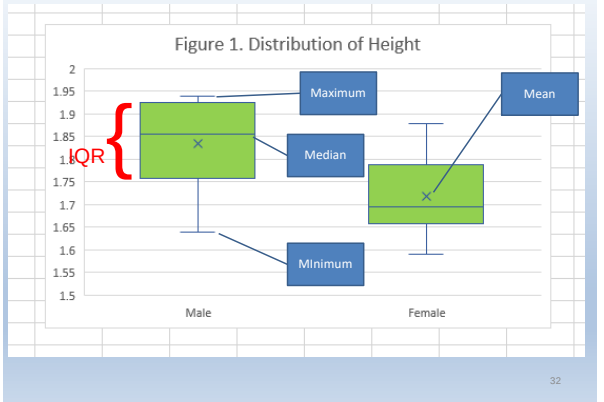

29



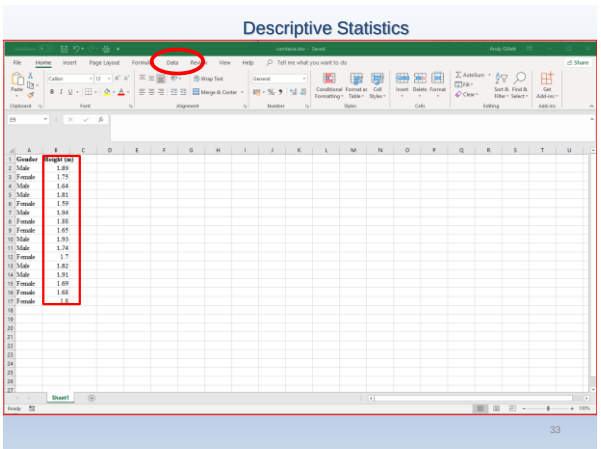

30



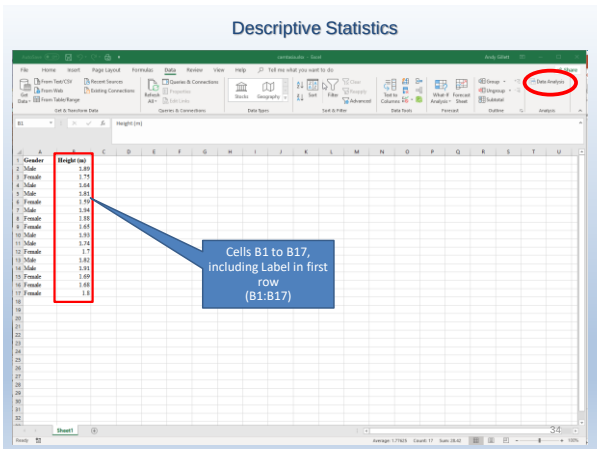

31



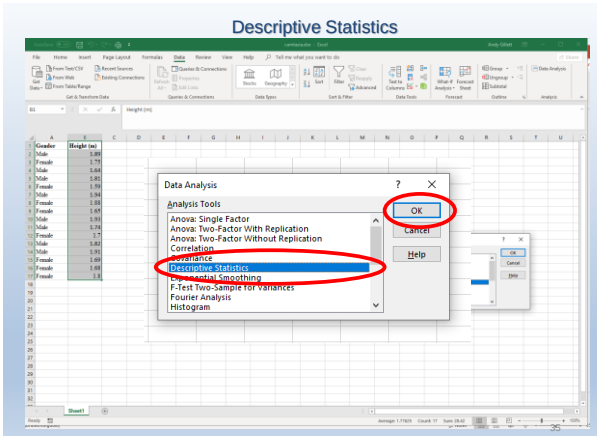

32



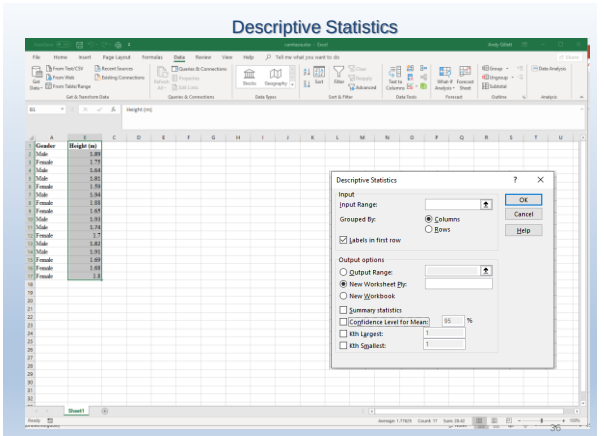

33




34

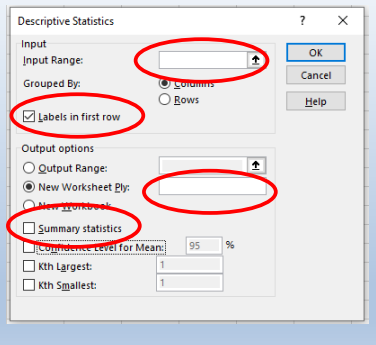



35



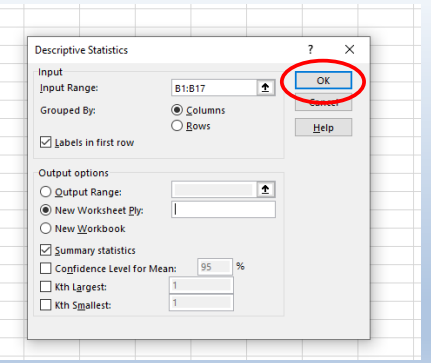

36

## Descriptive Statistics



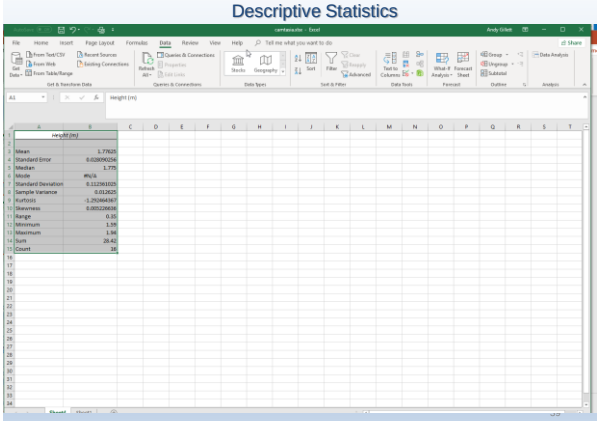

37

## Descriptive Statistics




38

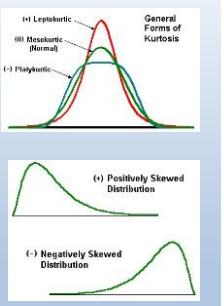
## Descriptive Statistics



39

## Descriptive Statistics

| | A | B | C |
|---|---|---|---|
| 1 | *Height (m)* | | |
| 2 | | | |
| 3 | Mean | 1.77625 | |
| 4 | Standard Error | 0.028090256 | |
| 5 | Median | 1.775 | |
| 6 | Mode | #N/A | |
| 7 | Standard Deviation | 0.112361025 | |
| 8 | Sample Variance | 0.012625 | |
| 9 | Kurtosis | -1.292464367 | |
| 10 | Skewness | 0.005226636 | |
| 11 | Range | 0.35 | |
| 12 | Minimum | 1.59 | |
| 13 | Maximum | 1.94 | |
| 14 | Sum | 28.42 | |
| 15 | Count | 16 | |
| 16 | | | |



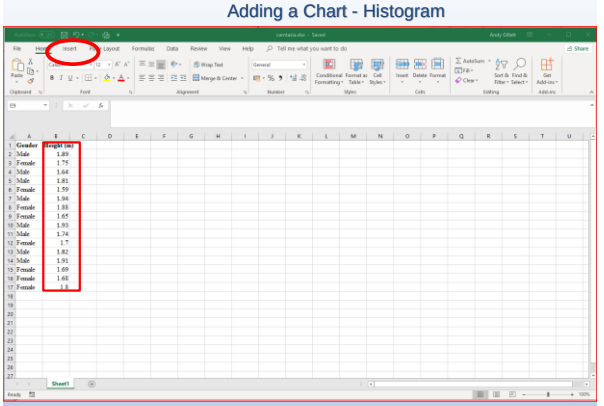

40

## Descriptive Statistics

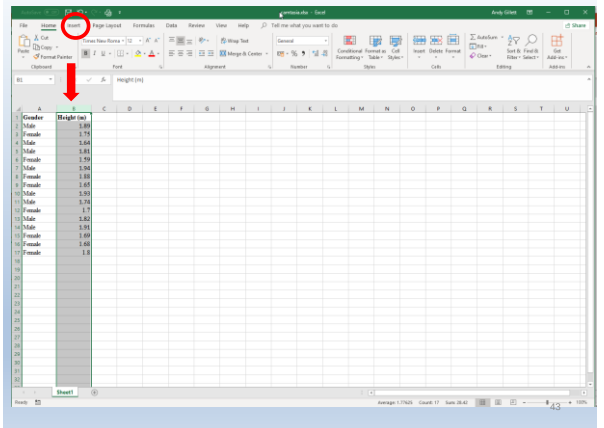
| Gender | Height (m) | | | | |
|---|---|---|---|---|---|
| Male | 1.89 | | | | |
| Female | 1.75 | | 1st Quartile | =QUARTILE(B2:B17,1) | 1.6875 |
| Male | 1.64 | | 3rd Quartile | =QUARTILE(B2:B17,3) | 1.8825 |
| Male | 1.81 | | IQR | =(F4-F3) | 0.195 |
| Female | 1.59 | | | | |
| Male | 1.94 | | | | |
| Female | 1.88 | | | | |
| Female | 1.65 | | | | |
| Male | 1.93 | | | | |
| Male | 1.74 | | | | |
| Female | 1.7 | | | | |
| Male | 1.82 | | | | |
| Male | 1.91 | | | | |
| Female | 1.69 | | | | |
| Female | 1.68 | | | | |
| Female | 1.8 | | | | |

41

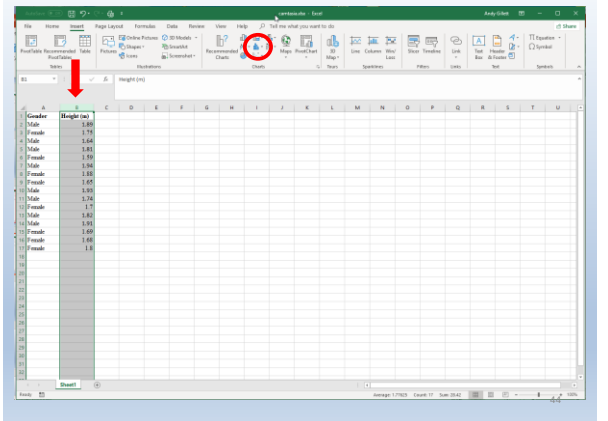
## Adding a Chart - Histogram



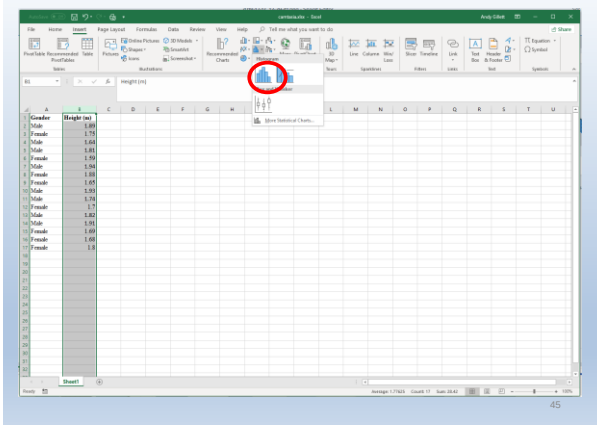
42



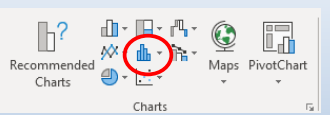

43



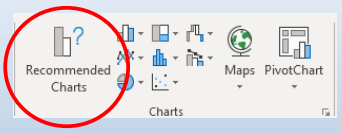

44



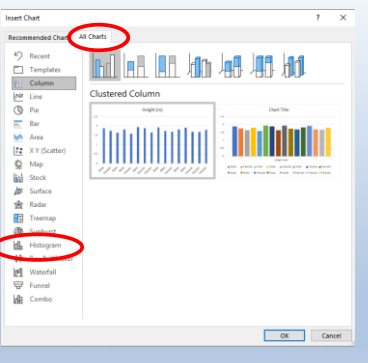

45



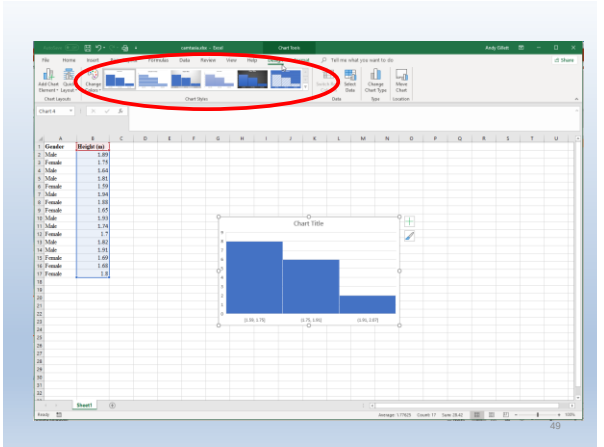

46



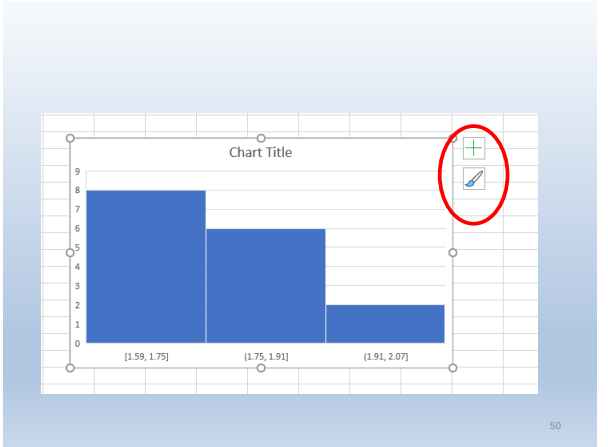

47



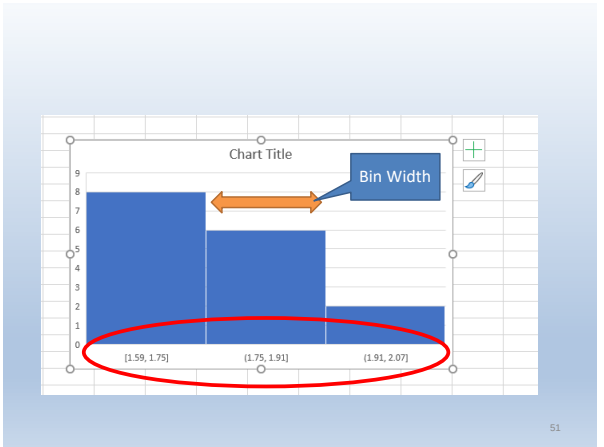

48



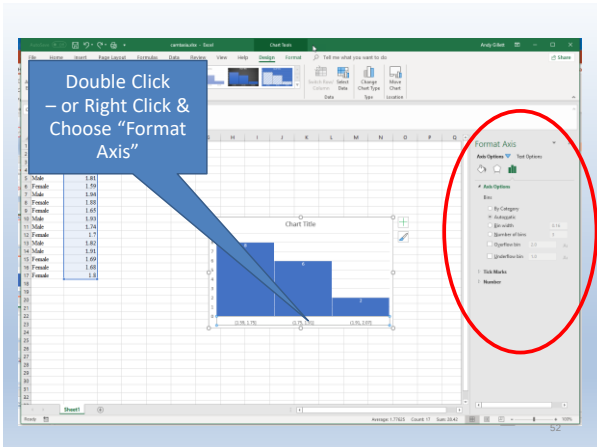

49



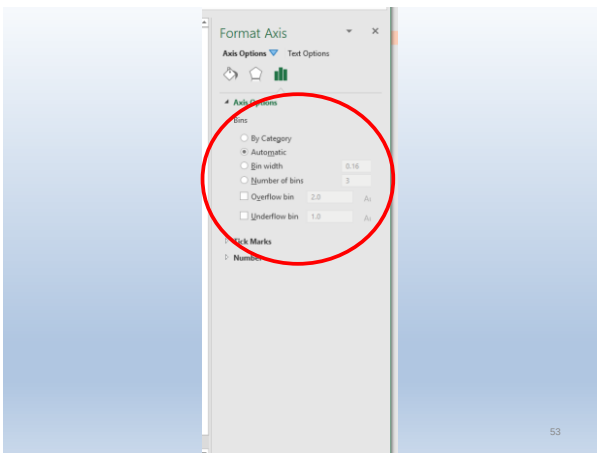

50




51




52




53


54

## Describing Data using Statistics

**Description**

- Central Tendency
- Dispersion

**Inference**

– Exploring Relationships
- Correlation
- Regression

– Comparing Groups
- t-Test
- ANOVA

55

55

## Describing Data using Statistics

**Description**

**Inference**

– Exploring Relationships

– Comparing Groups

"When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have."

(Popham & Sirotnik, 1973: 40)

56

56

## Describing Data using Statistics

**Description**

**Inference**

– Exploring Relationships

– Comparing Groups

"When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have."

(Popham & Sirotnik, 1973: 40)

57

57

## Describing Data using Statistics

**Relationships:**

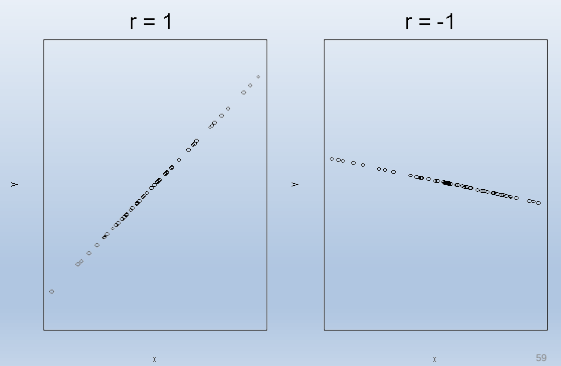
- The association between two variables.
- The correlation between two variables.

Correlation/correlate; Association/associate; Relationship/relate; Covariance/covariation/covary

58

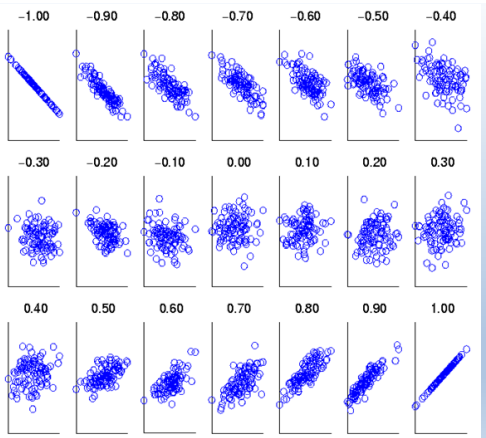
58

## Correlation Coefficient (r)




59

59




60

60

## Correlation Coefficients

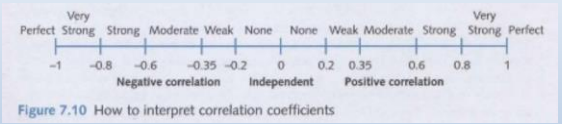



Figure 7.10 How to interpret correlation coefficients

(Saunders & Lewis, 2012: 183)



61

## Correlation

There is much interest in whether there is a relationship between pupils' ability at sport and their maths ability. Accordingly a study was carried out in which both sport and maths ability were rated on each of 10 pupils sampled at random from a class of first year pupils in a secondary school. Sport ability was rated on a scale of 1 to 20. Maths ability was measured by a pencil and paper test on a scale from 1 to 30.

Statistical Services and Consultancy Unit (2016: 33-46)



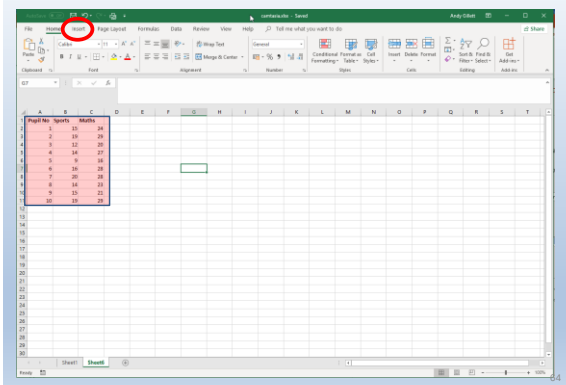
62

## Correlation

| Pupil No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sports score | 15 | 19 | 12 | 14 | 9 | 16 | 20 | 14 | 15 | 19 |
| Maths score | 24 | 29 | 20 | 27 | 16 | 28 | 28 | 23 | 21 | 29 |

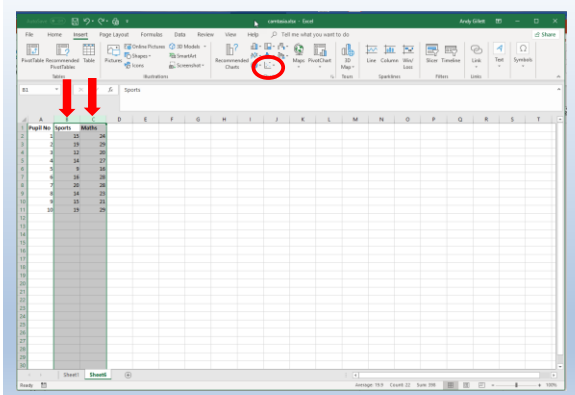


63



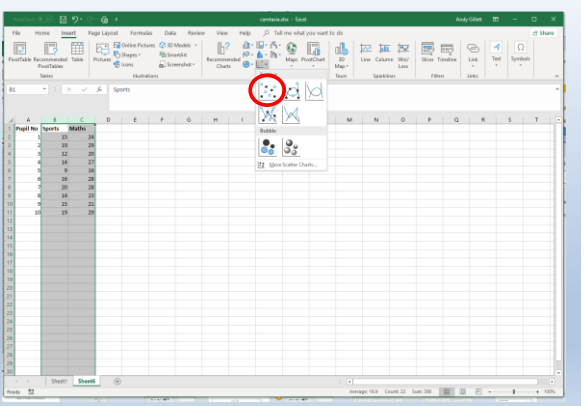


64



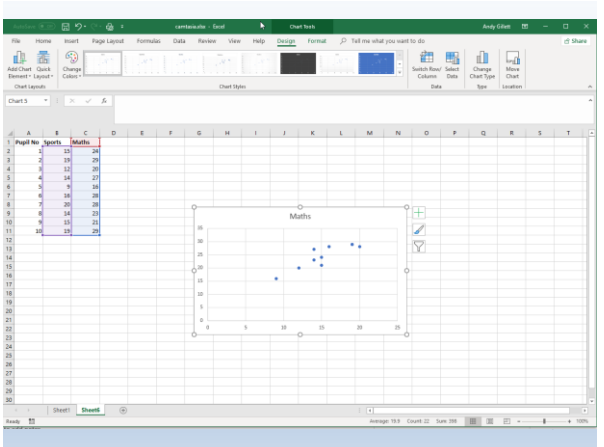


65

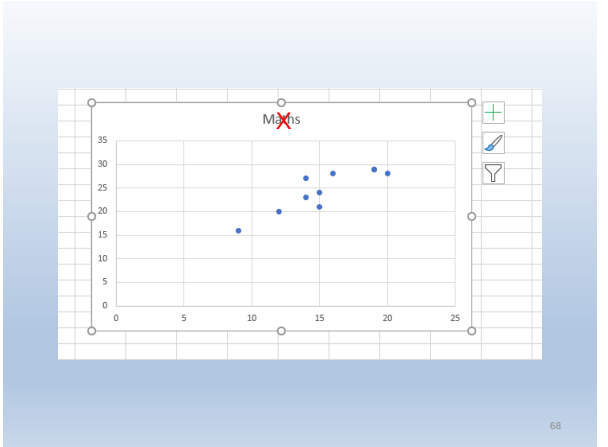




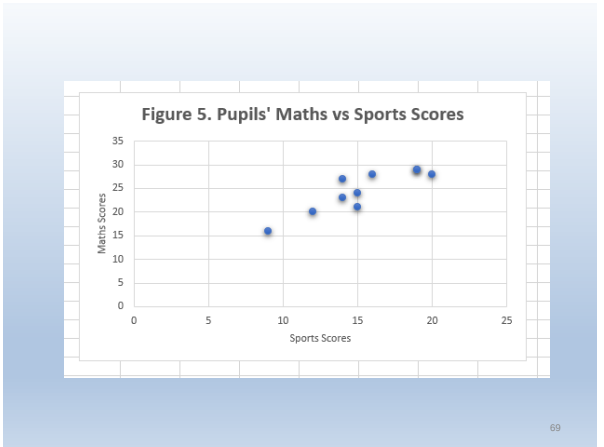
66



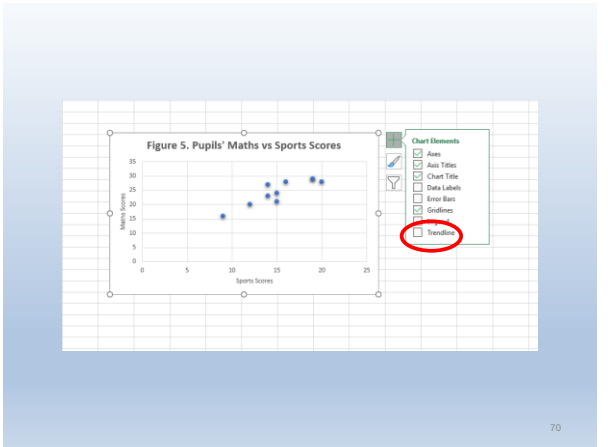

67



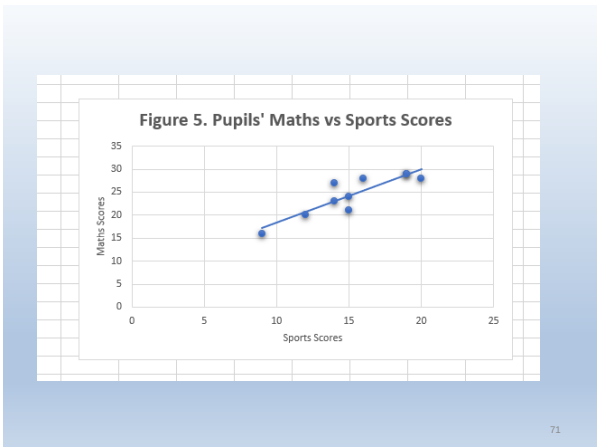

68



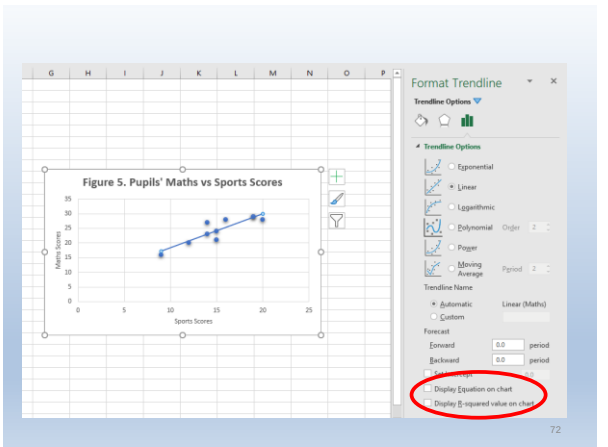

69



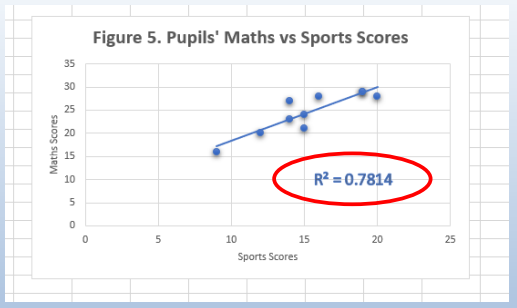

70




71



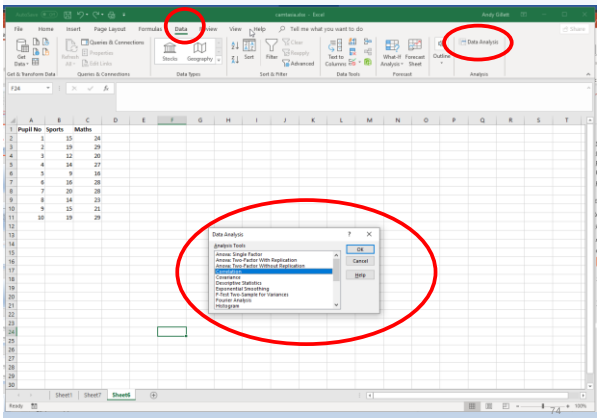

72




r (Correlation Coefficient) = √0.7814 = 0.884 = Very Strong

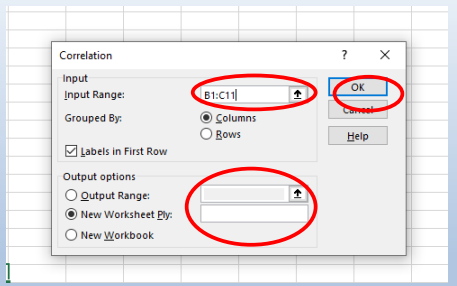
73

73



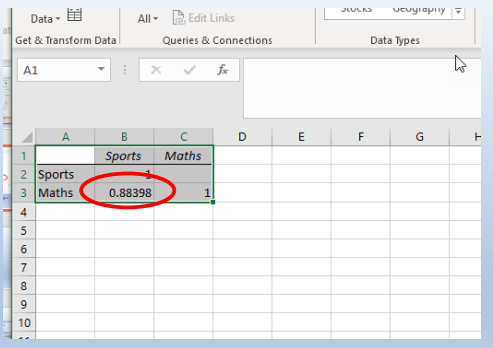

74

74




75

75




76

76

The correlation, r, has the value 0.884.

The *p*-value is the probability that the correlation found between the two variables could be a chance result. As *p* here is less than 0.05 the correlation is unlikely to be a chance result.

The conclusion is that there is a relationship between SPORT scores and MATHS scores in the wider population of first year secondary school pupils. The results can be generalised.

77

77

## Regression

A study looked at the possibility of predicting students' future reading ability (READING) from their intelligence score (IQ).
Ten children were included in the investigation, selected at random from a primary school class. The children's IQ scores were obtained when they entered the school and the corresponding reading score was obtained at the end of the first year. The reading score was the result out of 100 on a test.

Statistical Services and Consultancy Unit (2016: 36-42)

78

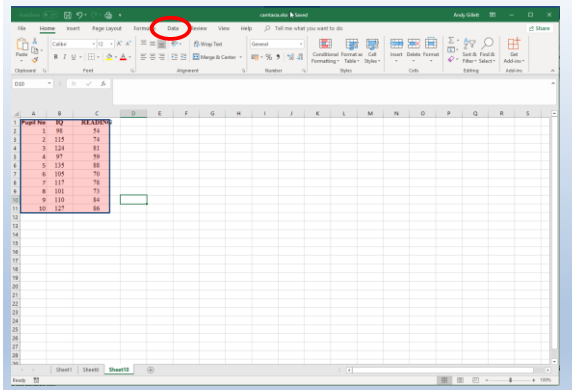
78

## Regression

| Pupil No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| IQ Score | 98 | 115 | 124 | 97 | 135 | 105 | 117 | 101 | 110 | 127 |
| Reading | 54 | 74 | 81 | 59 | 88 | 70 | 78 | 73 | 84 | 86 |

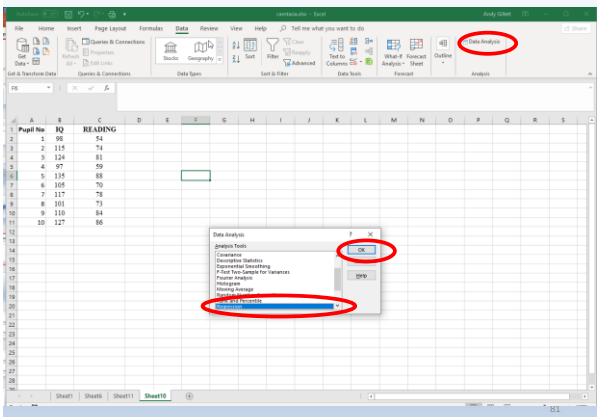
79

79



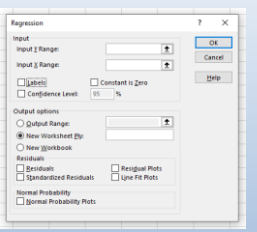
80

80



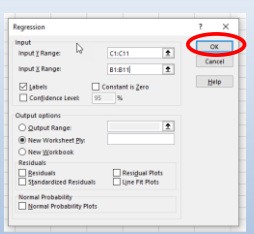
81

81



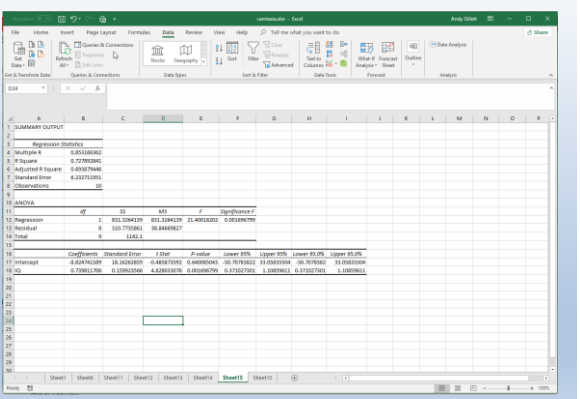
82

82



83

83



84

84

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.853166362 |
| R Square | 0.727892841 |
| Adjusted R Square | 0.693879446 |
| Standard Error | 6.232711951 |
| Observations | 10 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 831.3264139 | 831.3264139 | 21.40018202 | 0.001696799 |
| Residual | 8 | 310.7735861 | 38.84669827 | | |
| Total | 9 | 1142.1 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* |
|---|---|---|---|---|---|
| Intercept | -8.824741589 | 18.16262859 | -0.485873592 | 0.640085043 | -50.70783822 |
| IQ | 0.739811706 | 0.159923566 | 4.626033076 | 0.001696799 | 0.371027301 |

Very Strong Correlation (0.85)

0.0016 < 0.05 therefore significant

Reading = -8.82 + 0.74 x IQ

85

85

## Regression

The regression equation

"READING = −8.82 + 0.740 × IQ"

can be obtained by equating the dependent variable **Reading** to the *CONSTANT* value −8.825, plus the independent variable **IQ** multiplied by its coefficient 0.740.

86

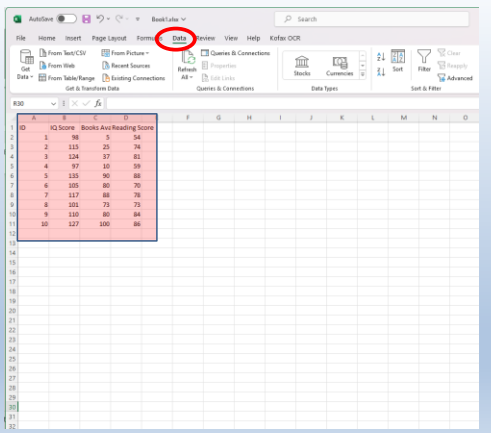
86

## Example

Since $p$ is very small, below 0.05, the relationship is unlikely to be a chance result. Hence it is concluded that there is a relationship between IQ and reading score for primary school children. Hence, we can reject $H_0$ in favour of $H_1$. For every primary school child whose IQ is known the reading score can be predicted.

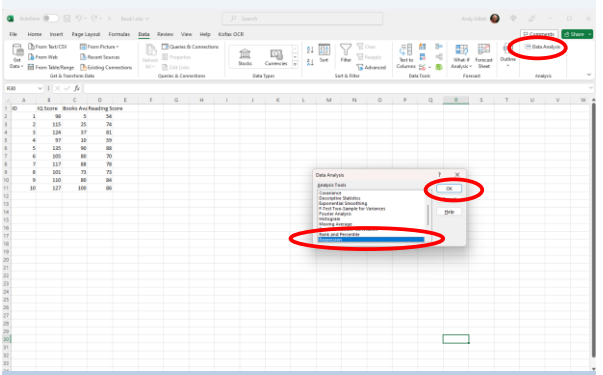
87

87



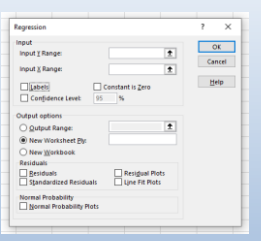
88

88



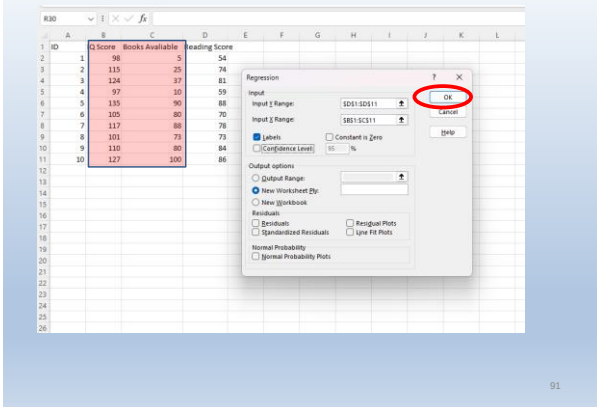
89

89



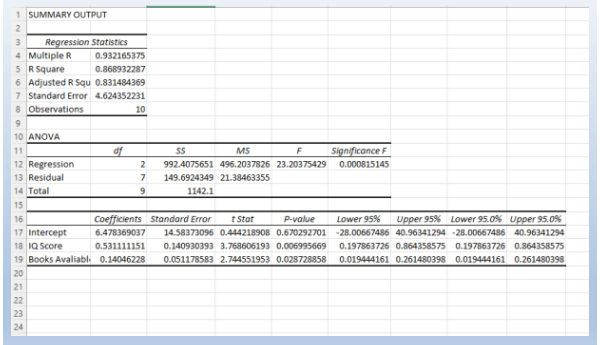

90

90



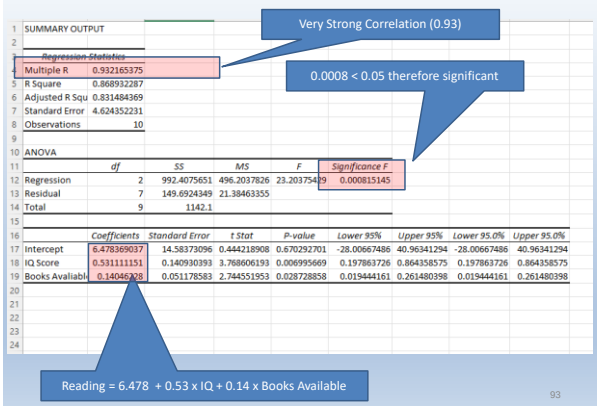

91

91



| | | | | | |
|---|---|---|---|---|---|
| SUMMARY OUTPUT | | | | | |
| *Regression Statistics* | | | | | |
| Multiple R | 0.932165375 | | | | |
| R Square | 0.868932287 | | | | |
| Adjusted R Squ | 0.831484369 | | | | |
| Standard Error | 4.624352231 | | | | |
| Observations | 10 | | | | |
| ANOVA | | | | | |
| | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 992.4075651 | 496.2037826 | 23.20375429 | 0.000815145 |
| Residual | 7 | 149.6924349 | 21.38463355 | | |
| Total | 9 | 1142.1 | | | |
| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 6.478369037 | 14.58373096 | 0.444218908 | 0.670292701 | -28.00667486 | 40.96341294 | -28.00667486 | 40.96341294 |
| IQ Score | 0.531111151 | 0.140930393 | 3.768606193 | 0.006995669 | 0.197863726 | 0.864358575 | 0.197863726 | 0.864358575 |
| Books Avaliabl | 0.14046228 | 0.051178583 | 2.744551953 | 0.028728858 | 0.019444161 | 0.261480398 | 0.019444161 | 0.261480398 |

92




93

## Multiple Regression

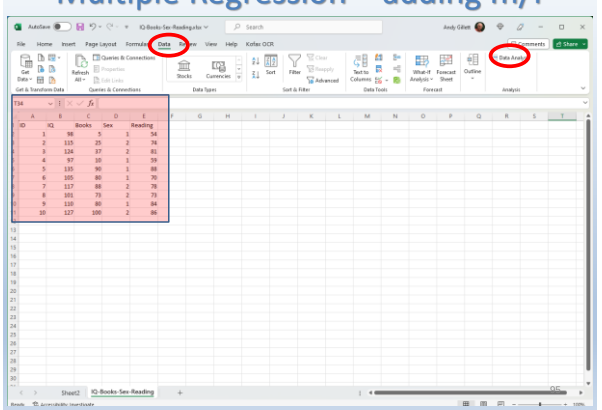
The regression equation

"READING = 6.478 + .140 × Books + .531 x IQ "

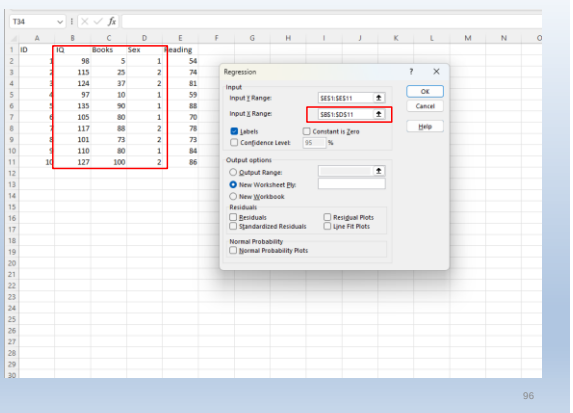
can be obtained by equating the dependent variable **Reading** to the *CONSTANT* value 6.478, plus the independent variable **IQ** multiplied by its coefficient .531 and the **Number of Books** multiplied by .140.

94

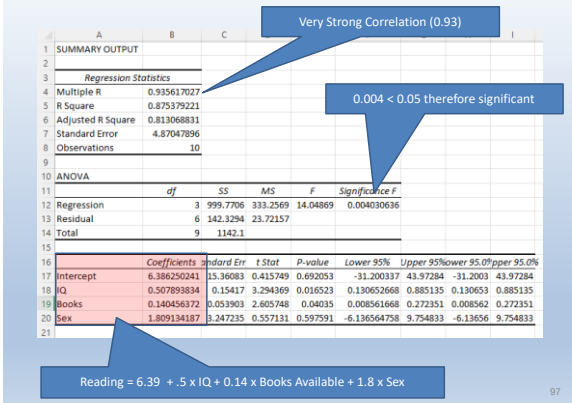
## Multiple Regression – adding m/f



95




96




| SUMMARY OUTPUT | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Regression Statistics* | | | | | | | | |
| Multiple R | 0.935617027 | | | | | | | |
| R Square | 0.875379221 | | | | | | | |
| Adjusted R Square | 0.813068831 | | | | | | | |
| Standard Error | 4.87047896 | | | | | | | |
| Observations | 10 | | | | | | | |
| ANOVA | | | | | | | | |
| | *df* | *SS* | *MS* | *F* | *Significance F* | | | |
| Regression | 3 | 999.7706 | 333.2569 | 14.04869 | 0.004030636 | | | |
| Residual | 6 | 142.3294 | 23.72157 | | | | | |
| Total | 9 | 1142.1 | | | | | | |
| | *Coefficients* | *Standard Err* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 6.386250241 | 15.36083 | 0.415749 | 0.692053 | -31.200337 | 43.97284 | -31.2003 | 43.97284 |
| IQ | 0.507893834 | 0.15417 | 3.294369 | 0.016523 | 0.130652668 | 0.885135 | 0.130653 | 0.885135 |
| Books | 0.140456372 | 0.053903 | 2.605748 | 0.04035 | 0.008561668 | 0.272351 | 0.008562 | 0.272351 |
| Sex | 1.809134187 | 3.247235 | 0.557131 | 0.597591 | -6.136564758 | 9.754833 | -6.13656 | 9.754833 |



97

## Multiple Regression

The regression equation

Reading = 6.39 + .5 x IQ + 0.14 x Books Available + 1.8 x Sex

Note that Sex is a categorical variable with two levels (Males = 1 & Females = 2).

In this case the regression weighting (1.8) is added to the predicted Reading value. So for females 1.8 (1.8 x 1) would be added and for males 3.6 (1.8 x 2) would be added.



98

## Multiple Regression: Relative Importance

Which variables are important?

| | *Coefficients* | *Standard Err* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 6.386250241 | 15.36083 | 0.415749 | 0.692053 | -31.2003 | 43.97284 | -31.2003 | 43.97284 |
| IQ | 0.507893834 | 0.15417 | 3.294369 | 0.016523 | 0.130653 | 0.885135 | 0.130653 | 0.885135 |
| Books | 0.140456372 | 0.053903 | 2.605748 | 0.04035 | 0.008562 | 0.272351 | 0.008562 | 0.272351 |
| Sex | 1.809134187 | 3.247235 | 0.557131 | 0.597591 | -6.13656 | 9.754833 | -6.13656 | 9.754833 |

We cannot compare these coefficients as they use different scales. We need to standardize them – put them on the same scale. We will then have the standardised regression coefficients, which we can compare.



99

## Multiple Regression: Relative Importance

Which variables are important?

=STANDARDIZE(B2:B11,AVERAGE(B2:B11),STDEV(B2:B11))

| ID | IQ | Books | Sex | Reading | IQ Standardised | Books Standardised | Sex Standardised | Reading Standardised |
|---|---|---|---|---|---|---|---|---|
| 1 | 98 | 5 | 1 | 54 | -1.146945897 | -1.503912381 | -0.948683298 | -1.837551562 |
| 2 | 115 | 25 | 2 | 74 | 0.161650093 | -0.944837146 | 0.948683298 | -0.062139425 |
| 3 | 124 | 37 | 2 | 83 | 0.854436205 | -0.609392006 | 0.948683298 | 0.559254823 |
| 4 | 97 | 10 | 1 | 59 | -1.223922132 | -1.364143572 | -0.948683298 | -1.393698528 |
| 5 | 135 | 90 | 1 | 88 | 1.701174787 | 0.872157366 | -0.948683298 | 1.180649072 |
| 6 | 105 | 80 | 1 | 70 | -0.608112254 | 0.592619748 | -0.948683298 | -0.417221852 |
| 7 | 117 | 88 | 2 | 78 | 0.315602562 | 0.816249842 | 0.948683298 | 0.292943003 |
| 8 | 101 | 73 | 2 | 73 | -0.916071193 | 0.396943416 | 0.948683298 | -0.150910032 |
| 9 | 110 | 80 | 1 | 84 | -0.223231081 | 0.592619748 | -0.948683298 | 0.825566644 |
| 10 | 127 | 100 | 2 | 86 | 1.085364909 | 1.151694983 | 0.948683298 | 1.003107858 |

Run the regression function as before.



100

## Multiple Regression: Relative Importance

Which variables are important?

| SUMMARY OUTPUT | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Regression Statistics* | | | | | | | | |
| Multiple R | 0.935617027 | | | | | | | |
| R Square | 0.875379221 | | | | | | | |
| Adjusted R Square | 0.813068831 | | | | | | | |
| Standard Error | 0.432355373 | | | | | | | |
| Observations | 10 | | | | | | | |
| ANOVA | | | | | | | | |
| | *df* | *SS* | *MS* | *F* | *gnificance F* | | | |
| Regression | 3 | 7.878413 | 2.626138 | 14.04869 | 0.004031 | | | |
| Residual | 6 | 1.121587 | 0.186931 | | | | | |
| Total | 9 | 9 | | | | | | |
| | *Coefficients* | *Standard Err* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -1.38519E-17 | 0.136723 | -1E-16 | 1 | -0.33455 | 0.334549 | -0.33455 | 0.334549 |
| IQ Standardised | 0.585713812 | 0.177792 | 3.294369 | 0.016523 | 0.150671 | 1.020756 | 0.150671 | 1.020756 |
| Books Standardised | 0.446036475 | 0.171174 | 2.605748 | 0.04035 | 0.027189 | 0.864884 | 0.027189 | 0.864884 |
| Sex Standardised | 0.084642546 | 0.151926 | 0.557131 | 0.597591 | -0.28711 | 0.456392 | -0.28711 | 0.456392 |



101

## Multiple Regression: Relative Importance

Which variables are important?

| | *Coefficients* |
|---|---|
| Intercept | -1.38519E-17 |
| IQ Standardised | 0.585713812 |
| Books Standardised | 0.446036475 |
| Sex Standardised | 0.084642546 |

You now have the standardised regression coefficients, and can now compare them.

IQ: 0.586
Sex: 0.085
Books Available: 0.446

And you can see that the most important predictor is **IQ**, followed closely by **Books Available**. **Sex** is not contributing very much.



102

## Describing Data using Statistics

**Description**

**Inference**

- Exploring Relationships
- Comparing Groups

"When an individual uses descriptive statistics, he talks about the data he has; but with inferential statistics, he talks about data he does not have."

(Popham & Sirotnik, 1973: 40)

103

103

## Comparing Groups (2 groups)

An experiment is carried out to see if announcing the items on a conveyor belt as they pass helps children remember the items better. Twenty-two children, all aged 12 years, were randomly allocated; 10 to the announcement group, and the remaining 12 to the no-announcement group. The children were shown the items on the conveyor belt with or without announcing each item as they passed. The number of items recalled out of each of the ten letters in the list was recorded for each child:

104

104

## Comparing Groups

The researchers wanted to find out if the announcement had any impact on the number of items recalled.

| | Items Recalled | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Announcement:** | 5 | 7 | 8 | 9 | 5 | 4 | 9 | 8 | 7 | 4 | | |
| **No announcement:** | 3 | 8 | 5 | 6 | 3 | 5 | 9 | 3 | 3 | 4 | 7 | 4 |

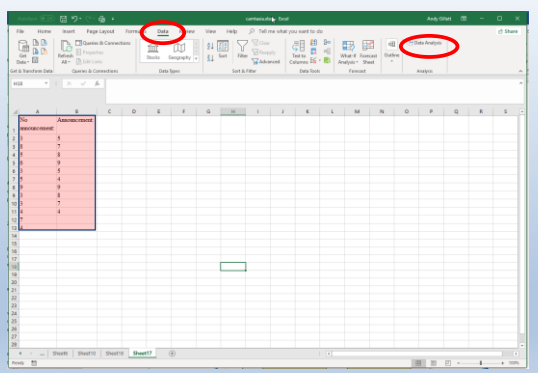
105

105


106

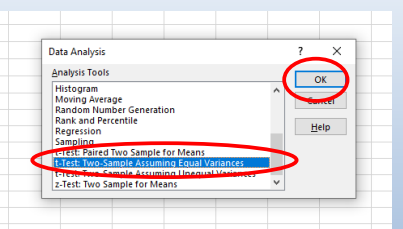
106

## t-Test



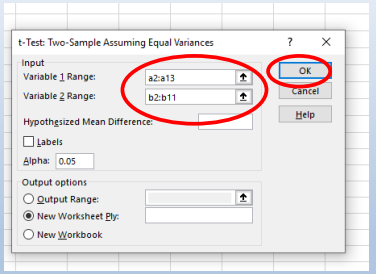
107

107



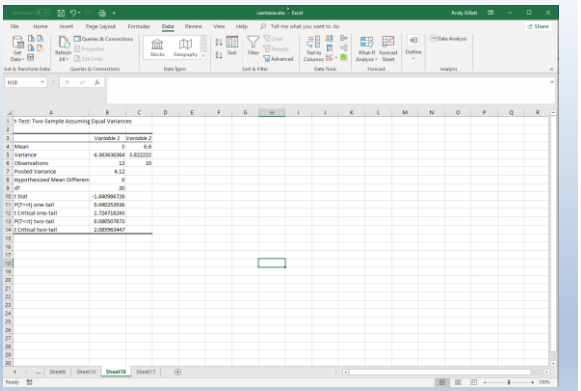

108

108




109

109



110

110

| t-Test: Two-Sample Assuming Equal Variances | | |
|---|---|---|
| | *Variable 1* | *Variable 2* |
| Mean | 5 | 6.6 |
| Variance | 4.363636364 | 3.822222222 |
| Observations | 12 | 10 |
| Pooled Variance | 4.12 | |
| Hypothesized Me | 0 | |
| df | 20 | |
| t Stat | -1.840986726 | |
| P(T<=t) one-tail | 0.040253936 | |
| t Critical one-tail | 1.724718243 | |
| P(T<=t) two-tail | 0.080507873 | |
| t Critical two-tail | 2.085963447 | |

P = 0.08 > 0.05 Therefore not significant

The conclusion is that there is no benefit to the wider population from the use of announcements.

111

111

## Comparing Groups (3 groups or more)

An experiment was carried out in which a sample of adults was selected and randomly allocated to four groups. All were asked to consume amounts of alcohol as follows:

Group 1: NO ALCOHOL
Group 2: ONE UNIT OF ALCOHOL
Group 3: TWO UNITS OF ALCOHOL
Group 4: THREE UNITS OF ALCOHOL

after which they were asked drive a car simulator and perform an emergency stop at 40 miles per hour. The distance to stopping the simulator in metres by each group member is shown below.

112

112

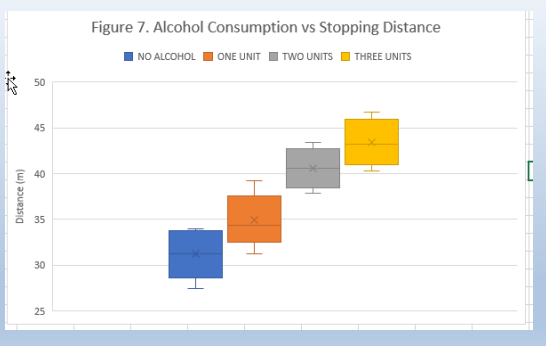
## Comparing Groups (3 groups or more)

The results were as follow.

| | | | | | |
|---|---|---|---|---|---|
| NO ALCOHOL | 31.2 | 27.4 | 29.8 | 33.5 | 34.0 |
| ONE UNIT | 31.2 | 33.7 | 34.3 | 39.2 | 36.0 |
| TWO UNITS | 42.1 | 40.6 | 43.4 | 37.9 | 39.0 |
| THREE UNITS | 46.7 | 45.1 | 43.2 | 41.7 | 40.3 |

113

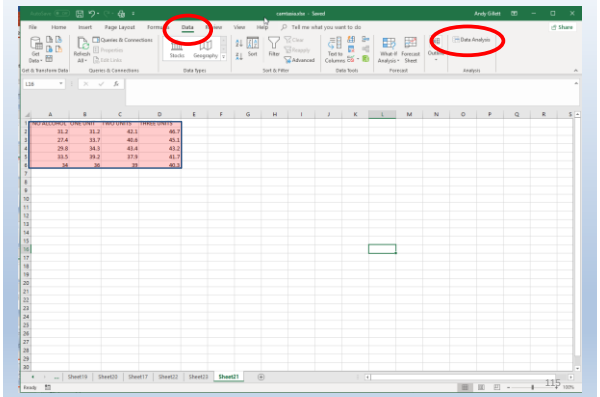
113



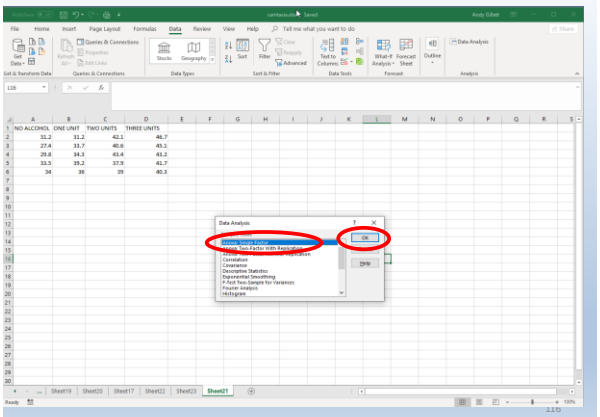

114

114

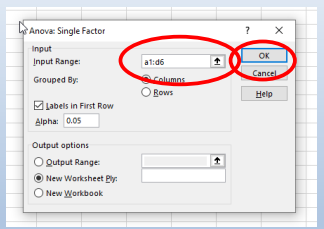
## ANOVA



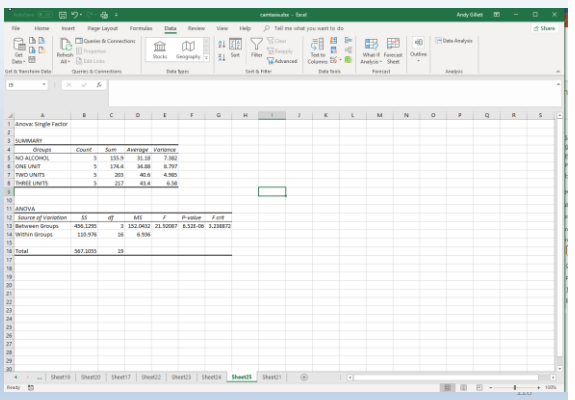
115



116




117



118

Anova: Single Factor

SUMMARY

| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| NO ALCOHOL | 5 | 155.9 | 31.18 | 7.382 |
| ONE UNIT | 5 | 174.4 | 34.88 | 8.797 |
| TWO UNITS | 5 | 203 | 40.6 | 4.985 |
| THREE UNITS | 5 | 217 | 43.4 | 6.58 |

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 456.1295 | 3 | 152.0432 | 21.92087 | 6.52E-06 | 3.238872 |
| Within Groups | 110.976 | 16 | 6.936 | | | |
| Total | 567.1055 | 19 | | | | |

$p = 6.52E\text{-}06 = 6.52 \times 10^{-6} = 0.00000652 < 0.05$
Therefore significant

The conclusion is that there is a significant difference between the distance it takes to stop a car simulator according to the amount of alcohol consumed.

119

119

## Comparing Groups

An experiment was carried out in which a sample of adults was selected and randomly allocated to four groups. All were asked to consume amounts of alcohol as follows:

Group 1: NO ALCOHOL
Group 2: ONE UNIT OF ALCOHOL
Group 3: TWO UNITS OF ALCOHOL
Group 4: THREE UNITS OF ALCOHOL

after which they were asked drive a car simulator and perform an emergency stop at 40 miles per hour. The distance to stopping the simulator in metres by each group member is shown below.

Is this the same for both males and females?

120

120

## Comparing Groups (4 groups and M/F)

The results were as follows.

| | | | | | | |
|---|---|---|---|---|---|---|
| MALES | NO ALCOHOL | 31.2 | 27.4 | 29.8 | 33.5 | 34.0 |
| | ONE UNIT | 31.2 | 33.7 | 34.3 | 39.2 | 36.0 |
| | TWO UNITS | 42.1 | 40.6 | 43.4 | 37.9 | 39.0 |
| | THREE UNITS | 46.7 | 45.1 | 43.2 | 41.7 | 40.3 |
| FEMALES | NO ALCOHOL | 35.1 | 32.3 | 35.4 | 38.1 | 35.9 |
| | ONE UNIT | 32.8 | 35.9 | 36.1 | 40.9 | 38.2 |
| | TWO UNITS | 38.1 | 35.9 | 40.2 | 39.8 | 37.6 |
| | THREE UNITS | 39.8 | 31.7 | 41.2 | 39.1 | 37.9 |

Stopping Distance in Metres

121

121

## Three Questions

1. Is there a difference in the stopping distance according to the amount of alcohol consumed?
2. Is there a difference in the stopping distance between male and females?
3. Is the effect of amount of alcohol consumed on stopping distance different for males and females?

122

122

## EXCEL - Data

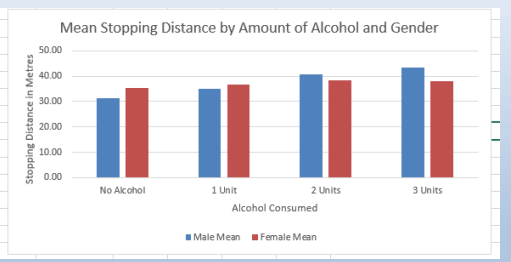
B16

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Gender | No Alcohol | 1 Unit | 2 Units | 3 Units | | |
| 2 | Male | 31.20 | 31.20 | 42.10 | 46.70 | | |
| 3 | Male | 27.40 | 33.70 | 40.60 | 45.10 | | |
| 4 | Male | 29.80 | 34.30 | 43.40 | 43.20 | | |
| 5 | Male | 33.50 | 39.20 | 37.90 | 41.70 | | |
| 6 | Male | 34.00 | 36.00 | 39.00 | 40.30 | | |
| 7 | Female | 35.10 | 32.80 | 38.10 | 39.80 | | |
| 8 | Female | 32.30 | 35.90 | 35.90 | 31.70 | | |
| 9 | Female | 35.40 | 36.10 | 40.20 | 41.20 | | |
| 10 | Female | 38.10 | 40.90 | 39.80 | 39.10 | | |
| 11 | Female | 35.90 | 38.20 | 37.60 | 37.90 | | |
| 12 | | | | | | | |
| 13 | | | | | | | |

123

123

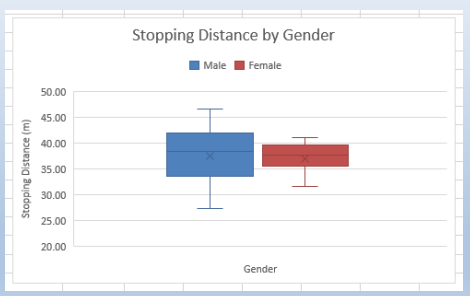
## Descriptive Statistics




124

124

## Total (M&F) Means




125

125

## Descriptive Statistics

| *Male* | | *Female* | |
|---|---|---|---|
| Mean | 37.515 | Mean | 37.1 |
| Standard Error | 1.221631243 | Standard Error | 0.620780493 |
| Median | 38.45 | Median | 37.75 |
| Mode | 31.2 | Mode | 35.9 |
| Standard Deviation | 5.463301007 | Standard Deviation | 2.776214765 |
| Sample Variance | 29.84765789 | Sample Variance | 7.707368421 |
| Kurtosis | -0.984023157 | Kurtosis | -0.542471441 |
| Skewness | -0.132899632 | Skewness | -0.442046056 |
| Range | 19.3 | Range | 9.5 |
| Minimum | 27.4 | Minimum | 31.7 |
| Maximum | 46.7 | Maximum | 41.2 |
| Sum | 750.3 | Sum | 742 |
| Count | 20 | Count | 20 |
| Confidence Level(95. | 2.556903578 | Confidence Level(95 | 1.299308505 |

126

126

## Descriptive Statistics

| | No Alcohol | 1 Unit | 2 Units | 3 Units |
|---|---|---|---|---|
| Mean | 33.27 | 35.83 | 39.46 | 40.67 |
| Standard Error | 1.00244701 | 0.944463 | 0.707766 | 1.31327 |
| Median | 33.75 | 35.95 | 39.4 | 40.75 |
| Mode | #N/A | #N/A | #N/A | #N/A |
| Standard Deviation | 3.17001577 | 2.986656 | 2.238154 | 4.152924 |
| Sample Variance | 10.049 | 8.920111 | 5.009333 | 17.24678 |
| Kurtosis | -0.06924912 | -0.52334 | -0.22062 | 1.742686 |
| Skewness | -0.46548337 | 0.216964 | 0.291447 | -0.82207 |
| Range | 10.7 | 9.7 | 7.5 | 15 |
| Minimum | 27.4 | 31.2 | 35.9 | 31.7 |
| Maximum | 38.1 | 40.9 | 43.4 | 46.7 |
| Sum | 332.7 | 358.3 | 394.6 | 406.7 |
| Count | 10 | 10 | 10 | 10 |
| Confidence Level(95. | 2.26769268 | 2.136525 | 1.601079 | 2.970823 |



127

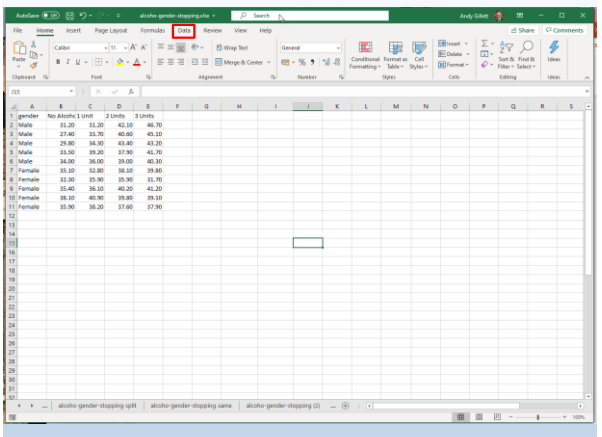
## ANOVA

One-Way Between-Groups ANOVA
Single independent variable.

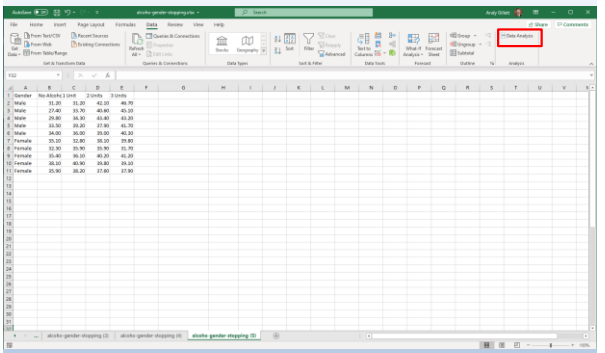
Two-Way Between-Groups ANOVA
Factorial/Multifactorial ANOVA
2 independent variables - two factors
Two factor design: A & B
Each factor can have different values - levels of the factor



128

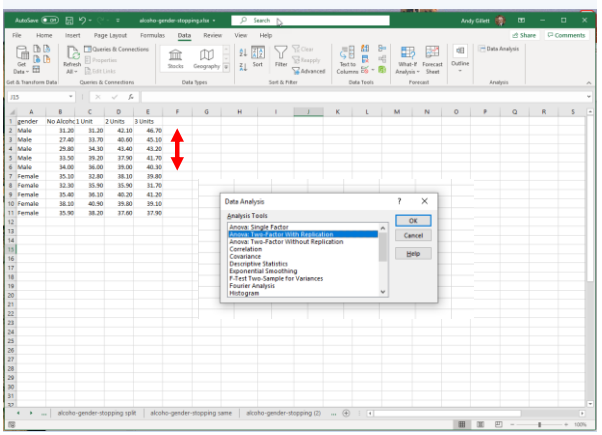


129

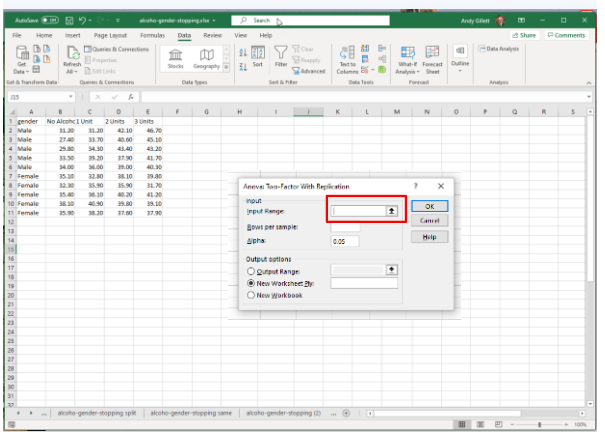




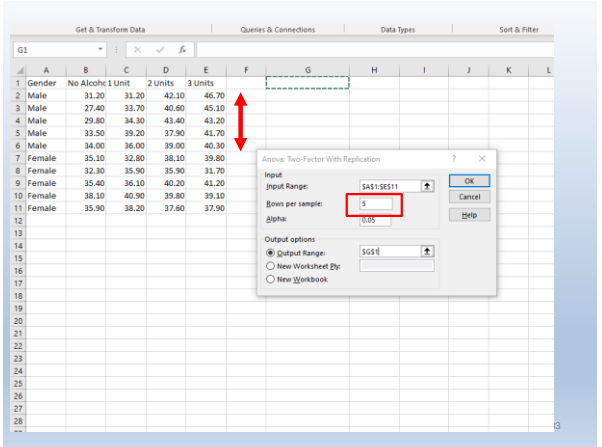
130



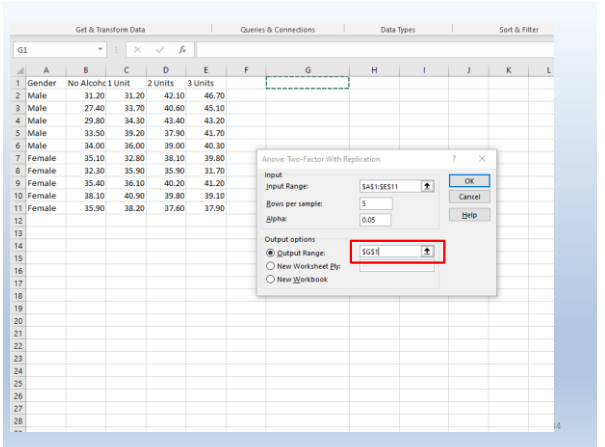

131



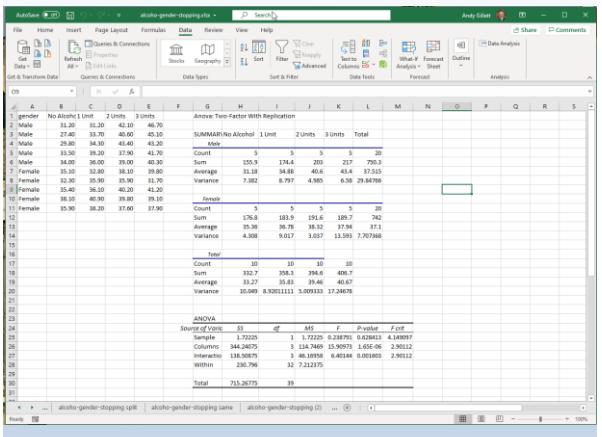

132




133




134



135

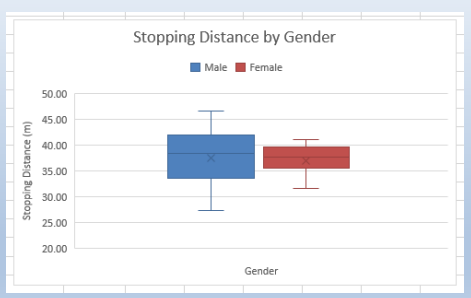
Variance = $SD^2$

Anova: Two-Factor With Replication

| SUMMARY | No Alcohol | 1 Unit | 2 Units | 3 Units | Total | |
|---|---|---|---|---|---|---|
| *Male* | | | | | | |
| Count | 5 | 5 | 5 | 5 | 20 | |
| Sum | 155.9 | 174.4 | 203 | 217 | 750.3 | |
| Average | 31.18 | 34.88 | 40.6 | 43.4 | 37.515 | |
| Variance | 7.382 | 8.797 | 4.985 | 6.58 | 29.84766 | SD = 5.5 |
| *Female* | | | | | | |
| Count | 5 | 5 | 5 | 5 | 20 | |
| Sum | 176.8 | 183.9 | 191.6 | 189.7 | 742 | |
| Average | 35.36 | 36.78 | 38.32 | 37.94 | 37.1 | |
| Variance | 4.308 | 9.017 | 3.037 | 13.593 | 7.707368 | SD = 2.8 |
| *Total* | | | | | | |
| Count | 10 | 10 | 10 | 10 | | |
| Sum | 332.7 | 358.3 | 394.6 | 406.7 | | |
| Average | 33.27 | 35.83 | 39.46 | 40.67 | | |
| Variance | 10.049 | 8.92011111 | 5.009333 | 17.24678 | | |

136

## Total Means (SD)



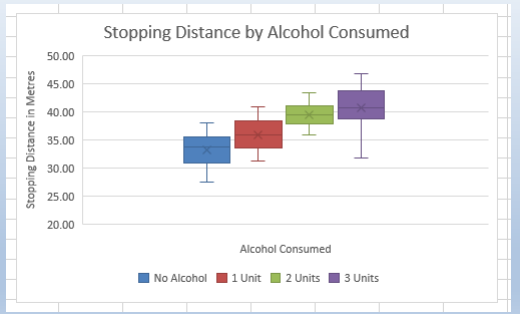

137

Anova: Two-Factor With Replication

| SUMMARY | No Alcohol | 1 Unit | 2 Units | 3 Units | Total |
|---|---|---|---|---|---|
| *Male* | | | | | |
| Count | 5 | 5 | 5 | 5 | 20 |
| Sum | 155.9 | 174.4 | 203 | 217 | 750.3 |
| Average | 31.18 | 34.88 | 40.6 | 43.4 | 37.515 |
| Variance | 7.382 | 8.797 | 4.985 | 6.58 | 29.84766 |
| *Female* | | | | | |
| Count | 5 | 5 | 5 | 5 | 20 |
| Sum | 176.8 | 183.9 | 191.6 | 189.7 | 742 |
| Average | 35.36 | 36.78 | 38.32 | 37.94 | 37.1 |
| Variance | 4.308 | 9.017 | 3.037 | 13.593 | 7.707368 |
| *Total* | | | | | |
| Count | 10 | 10 | 10 | 10 | |
| Sum | 332.7 | 358.3 | 394.6 | 406.7 | |
| Average | 33.27 | 35.83 | 39.46 | 40.67 | |
| Variance | 10.049 | 8.92011111 | 5.009333 | 17.24678 | |

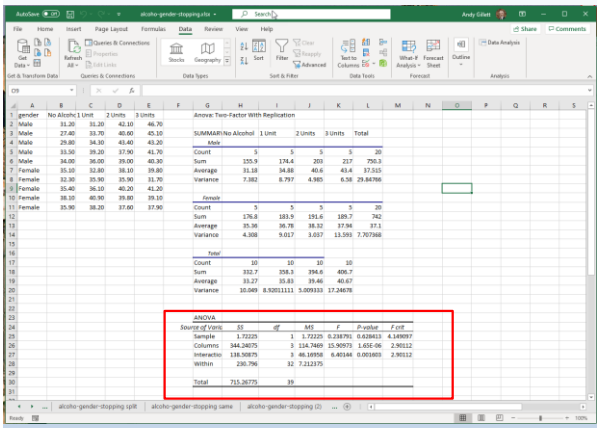
138

## Total – Means (SD)




139

139



140

## Main Effects

ANOVA

| Source of Variation | SS | df | MS | F | P-value | F crit |
|---|---|---|---|---|---|---|
| Sample | 1.72225 | 1 | 1.72225 | 0.238791 | 0.628413 | 4.149097 |
| Columns | 344.24075 | 3 | 114.7469 | 15.90973 | 0.000002 | 2.90112 |
| Interaction | 138.50875 | 3 | 46.16958 | 6.40144 | 0.001603 | 2.90112 |
| Within | 230.796 | 32 | 7.212375 | | | |
| | | | | | | |
| Total | 715.26775 | 39 | | | | |

Sample = Gender – not significant (p=0.63)

Columns = Amount of Alcohol Consumed – significant (p=0.000002 <0.05)

141

141

## Interaction

ANOVA

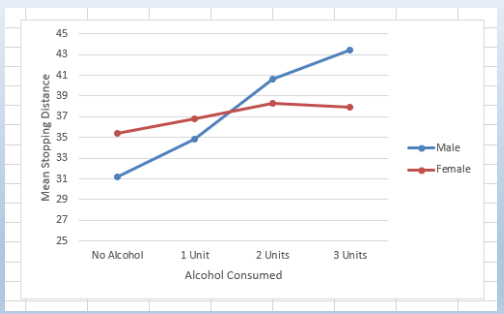
| Source of Variation | SS | df | MS | F | P-value | F crit |
|---|---|---|---|---|---|---|
| Sample | 1.72225 | 1 | 1.72225 | 0.238791 | 0.628413 | 4.149097 |
| Columns | 344.24075 | 3 | 114.7469 | 15.90973 | 0.000002 | 2.90112 |
| Interaction | 138.50875 | 3 | 46.16958 | 6.40144 | 0.001603 | 2.90112 |
| Within | 230.796 | 32 | 7.212375 | | | |
| | | | | | | |
| Total | 715.26775 | 39 | | | | |

Interaction = whether the influence of alcohol on stopping distance depends on whether you are male or female – significant (p=0.0016)

142

142

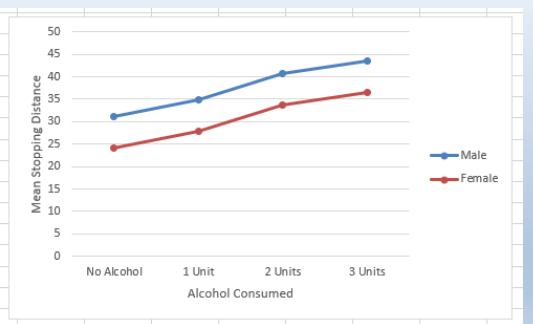
## Means Plots




143

143

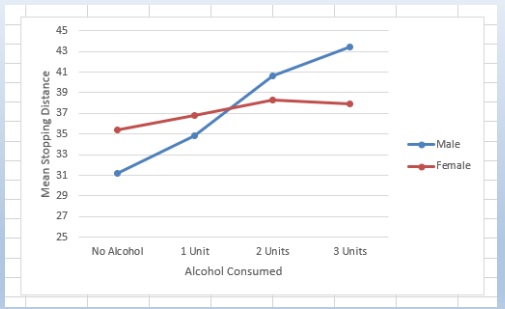
## Means Plots – No Interaction




144

144

## Means Plots - Interaction




145

145

## M/F Differences per Treatment Group

| | | | | | |
|---|---|---|---|---|---|
| 30 | t-Test: Two-Sample Assuming Equal Variances | | | | |
| 31 | | | | | |
| 32 | | No Alcohol | One Unit | Two Units | Three Units |
| 33 | Mean | 31.18 | 34.88 | 40.6 | 43.4 |
| 34 | Variance | 7.382 | 8.797 | 4.985 | 6.58 |
| 35 | Observations | 5 | 5 | 5 | 5 |
| 36 | Pooled Variance | 5.845 | 8.907 | 4.011 | 10.0865 |
| 37 | Hypothesized Mea | 0 | 0 | 0 | 0 |
| 38 | df | 8 | 8 | 8 | 8 |
| 39 | t Stat | -2.73 | -1.01 | 1.80 | 2.72 |
| 40 | P(T<=t) one-tail | 0.01 | 0.17 | 0.05 | 0.01 |
| 41 | t Critical one-tail | 1.86 | 1.86 | 1.86 | 1.86 |
| 42 | P(T<=t) two-tail | 0.03 | 0.34 | 0.11 | 0.03 |
| 43 | t Critical two-tail | 2.31 | 2.31 | 2.31 | 2.31 |
| 44 | | | | | |

146

146

**Anova: Single Factor**

**Males**

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 456.1295 | 3 | 152.0432 | 21.92087 | 0.000007 | 3.238872 |
| Within Groups | 110.976 | 16 | 6.936 | | | |
| | | | | | | |
| Total | 567.1055 | 19 | | | | |

**Females**

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 26.62 | 3 | 8.873333 | 1.184888 | 0.346793 | 3.238872 |
| Within Groups | 119.82 | 16 | 7.48875 | | | |
| | | | | | | |
| Total | 146.44 | 19 | | | | |

Males:
p = < 0.05
Therefore significant

The conclusion is that for males there is a significant difference between the distance it takes to stop a car simulator according to the amount of alcohol consumed.

147

147

## Interaction

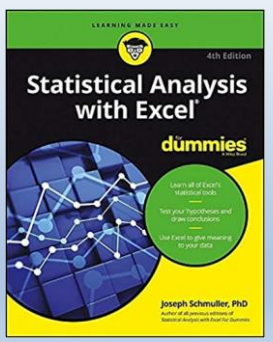
The conclusion is that the effect of alcohol on car stopping distance was different for male subjects than for females.

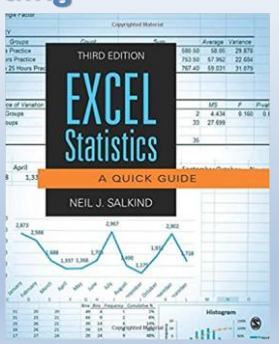
*A two-way between-groups analysis of variance was conducted to explore the impact of gender and amount of alcohol consumed on a car stopping distance. There was a significant interaction effect between the amount of alcohol consumed and the gender of the person driving the car on the distance needed to stop the car (F3, 32) = 6.4, p<0.05).*

148

148

## Reading

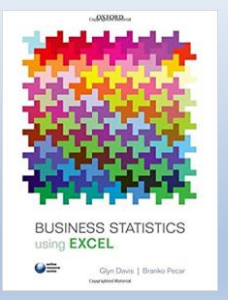


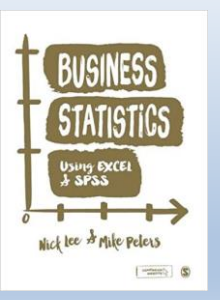




149

149

## Reading




150

150

# References & Further Reading


Anderson, D., Sweeney, D. & Williams, T. (2001). *Modern business statistics with Microsoft Excel*. New York: Cengage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: Routledge.

Davis, G. & Pecar, B. (2013). *Business statistics using Excel*. Oxford: Oxford University Press.

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford: Oxford University Press

Hand, D. J. (2008). *Statistics: A very short introduction*. Oxford: Oxford University Press.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle.

Jelen, B. (2013). *Excel 2013 charts and graphs*. Indianapolis: Que.

Keller, G. (2000). *Applied statistics with Microsoft Excel*. London: Duxbury.

Lee, N. & Peters, M. (2015). *Business statistics using Excel and SPSS*. London: Sage.

Popham, W. J. , & Sirotnik, K. A. (1973). *Educational statistics: Use and interpretation* (2nd ed.). New York: Harper & Row.

Salkind, N. J. (2011*). Statistics for people who (think they) hate statistics* (4th ed.). London: Sage.

Salkind, N. J. (2015). *Excel statistics: A quick guide* (3rd ed.). London: Sage.

Saunders, M., Lewis, P. and Thornhill, A. (2016). *Research methods for business students* (7th ed.). London: FT/Prentice Hall.

Schmuller, J. (2016). *Statistical analysis with Excel for dummies* (4th ed). London: Wiley.

Shavelson, R. (1988). *Statistical reasoning for the behavioral sciences.* Boston: Allyn and Bacon.

Speigelhalter, D. (2020). *The art of statistics: Learning from data*. London: Pelican.

Statistical Services and Consultancy Unit (2016). *Statistics with SPSS* (Version 23). Hatfield: University of Hertfordshire.




151