

# Contents

Acknowledgements	vii	Calculating averages	40
Introduction	viii	Mode, median and mean	40
<b>Part 1 Getting started</b>	<b>1</b>	Calculating measures of spread	52
1 What does your assignment ask you to do?	2	Range, interquartile range and standard deviation	52
2 How will you do it?	5	8 What plot should you use?	61
3 Defining your research question	13	Plotting categorical data	63
4 Tips for designing your questionnaire	18	Plotting scale data	66
5 How to enter your data into a spreadsheet	25	<b>Part 3 How do statistical tests work?</b>	<b>76</b>
<b>Part 2 Understanding and describing your data</b>	<b>31</b>	9 What is a statistical hypothesis?	77
6 What type of data do you have? Nominal, Ordinal, Interval or Ratio	32	The null hypothesis and the alternative hypothesis	77
7 Descriptive statistics	38	10 Using probability distributions in statistical tests	81
Population and samples	38	What is probability?	81
		What are degrees of freedom?	84

What about chance?	87	Case study 1: Describing data and looking for an association between variables	108
Significance level ( $\alpha$ )	87	The $\chi^2$ test	119
Two-tailed test	89	Case study 2: Looking for a difference between variables	122
One-tailed test	91	The paired $t$ -test	125
11 Statistical errors and their interpretation	92	Case study 3: Looking for a relationship between variables	126
Type I, Type II and Type III errors	93	The Pearson's correlation	130
<p><math>p</math>-values: a non-technical explanation</p>	96		
Interpreting statistical results	98		
<b>Part 4 What statistical test do you need?</b>	<b>99</b>	<b>Part 5 The statistical process</b>	<b>134</b>
12 The statistical signpost	100	15 You the researcher	135
13 Statistical flowcharts	102	16 You the interpreter	142
Categorical tests	102	Symbols explained	149
Parametric tests	104	References	151
Nonparametric tests	104	Useful resources	153
14 Case studies	108	Index	155

## 1

# What does your assignment ask you to do?

Different tasks require different types of statistical treatment. Before you begin, read your assignment brief carefully to see what exactly you are being asked to do. Then read and re-read your assignment brief, underlining key words.

Take a look at the examples below. Which one is closest to your task?

*Example 1: Have you been asked to perform a statistical calculation or test by hand?*

Calculate the mean ( $\bar{x}$ ) and standard deviation ( $s$ ) of the following dataset:  
33, 78, 56, 44, 82, 63.

*Is the dataset small  
and not entered into a  
spreadsheet?*

## 2

## HOW TO ANALYZE DATA

*Example 2: Have you been asked to analyze a dataset?*

Use the following data to determine if high-jump athletes are significantly taller than long-jump athletes.

*Have you been given the data in electronic form – for example, in a spreadsheet?*

Participant	Jump	Age	Height (m)
1	Long	17	181
2	Long	33	185
3	High	18	193
4	Long	41	183
:	:	:	:
:	:	:	:
29	High	13	178
30	Long	31	184

*The word 'significantly' may indicate the need to perform a statistical test.*

*Example 3: Are you responsible for both collecting the data and performing the data analysis?*

This may apply if you are doing your final project or dissertation.

**What does your assignment ask you to do?**

3

*Example 4: Are you being asked to interpret the results section of a published article?*

Interpreting published results is not an easy task, even for experts in the field!

Your tutor may have given you a journal article and asked you to ...

*appraise / critique / critically appraise*

All these words tell you that you are expected to have an understanding of the author's own statistical process.



**Parts 2, 3 and 4 for more on this.**

Each of these examples requires a different type of statistical analysis. Which example best describes what you have been asked to do?

<b>I have been told to ...</b>	<b>Take a look at</b>	
Perform a statistical calculation or test	Workshop 1 or 2	<b>Pages 6 and 7</b>
Analyze a dataset	Workshop 2 or 3	<b>Pages 7 and 11</b>
Collect a dataset and analyze it	Workshop 3	<b>Page 11</b>
Interpret published results	Part 5	<b>Page 134</b>

# Index

- absolute value, 28
- alternative hypothesis, see hypothesis
- analysis
  - appraise, 4, 142
  - critical or critical analysis, 4, 142
  - critique, 4, 142
- appraise, see analysis
- association, see case study 1
- averages
  - bi-modal, 46
  - central tendency, 40, 104, 137
  - mean, 40–45, 47–51, 59–60, 71, 75, 104, 145, 149
  - median, 40, 44–53, 59–60, 70–71, 75, 104, 145
  - mode, 40, 44–51, 59–60, 75
  - population mean, 41–43, 149
  - sample mean, 41–43, 57, 149
- axis, see plot
- bar chart, see plot
- bi-modal, see averages
- bin, see plot
- box-and-whisker, see plot
- calculation, viii, 6, 8–9, 28–29
- case studies
  - case study 1: association and chi-squared, 108–121
  - case study 2: difference and paired t-test, 122–125
  - case study 3: relationship and Pearson's correlation, 126–133
- categorical, see data
- cell, 25–27, 29
- central tendency, see averages
- chance, see probability
- chart, see plot
- chi square distribution,
  - see probability
- chi squared test, see case study 1, statistical tests
- column, 25–27, 30, 116, 117
- continuous, see data

- critical, *see* analysis
- critical analysis, *see* analysis
- critical region, 88, 90–91
- critical value, 88, 90–91, 97–98, 118
- critique, *see* analysis
- data
  - categorical, 33, 36, 51, 63, 100–103
  - continuous, 37, 69, 102
  - dataset, 2, 9–10, 38, 59, 66, 138, 141, 145
  - dependent variable, 61
  - discrete, 37, 102
  - expected value, 115, 117, 150
  - independent variable, 61, 86
  - interval, 32, 34–36, 48, 51, 59, 75, 101
  - nominal, 32–33, 36, 44, 51, 59, 75, 101–102
  - nonparametric, 51, 101–102, 104, 106
  - normally distributed, 48, 50–51, 59, 75, 87, 100–101, 104
  - observed value, 117, 150
  - ordinal, 32–33, 36, 45–48, 51, 59, 75, 101–102
  - outliers, 11, 61, 70–71, 136
  - paired, 123
    - parametric, 51, 101–102, 104–105
    - ratio, 32, 35–36, 48, 51, 59, 75, 101
    - scale, 21, 36, 51, 59, 66, 100–101
    - skewed, 49–51
    - symmetric, 48, 67
    - unpaired, 114
  - dataset, *see* data
  - degrees of freedom, 82, 84–86, 118, 150
  - demographic questions, *see* questionnaire
  - dependent variable, *see* data
  - descriptive statistics, *see* statistics
  - diagram, *see* plot
  - difference, 8–9, 95, 99–101, 105–106, 108, 141
    - see* case study 2
  - discrete, *see* data
  - dispersion, *see* spread
  - errors
    - type I, 93–94
    - type II, 94–95
    - type III, 96
  - estimate, 39, 57, 58
  - expected value, *see* data

frequency table, see plot  
 function, 9, 28, 60, 82

histogram, see plot

hypothesis

- alternative hypothesis, 77, 79–80, 88, 91, 93–96, 114, 123, 128, 150
- null hypothesis, 77–78, 84, 88–91, 93–98, 114, 123, 128, 150
- statistical hypothesis, 77–80

independent variable, see data

inferential statistics, see statistics

interpreting results, see results

interquartile range, see spread

interval, see data

line graphs, see plot

maximum, 52, 60, 70

mean, see averages

median, see averages

minimum, 52, 60, 70

mode, see averages

nominal, see data

nonparametric, see data

normal distribution, see probability

normally distributed data, see data

null hypothesis, see hypothesis

observed value, see data

one tail, see statistical tests

online questionnaire, see questionnaire

operations, 27–28

ordinal, see data

outliers, see data

paired, see data

paired t-test, see statistical tests

parameter, see population

parametric, see data

pearson's correlation, see case study 3,  
 statistical tests

pie chart, see plot

pilot, see questionnaire

plot

axis, 61, 73

bar chart, 63, 65–66, 75, 113

bin, 69

box-and-whisker, 66, 70–71, 75

chart, 61

cross plot, 67

diagram, 61



- frequency table, 63–64, 67, 69, 75
- histogram, 69, 75
- line graphs, 72–73, 75
- pie chart, 63–64, 75
- scatter plot, 74, 127
- stem and leaf, 68
- population
  - parameter, 39, 149
  - population mean, see averages
  - population size, 39, 56, 149
  - population standard deviation, see spread
  - population variance, see spread
- power
  - statistical power, 95, 104
- probability
  - chance, 78–80, 87–89, 91
  - chi square distribution, 82, 118
  - normal distribution, 82
  - probability distribution function, 82–84
  - t distribution, 82
  - p-values, 96–98, 125, 140, 150
- probability distribution functions, see probability
- p-values, see probability
- qualitative, see questionnaire
- quantitative, see questionnaire
- quartile, see spread
- questionnaire
  - demographic, 20, 111
  - online questionnaire, 19
  - pilot, 20, 109–110
  - qualitative, 19, 111
  - quantitative, 19
  - survey, 24
- range, see spread
- ratio, see data
- relationship, see case study 3
- reporting results, see results
- research question, 13–18, 20, 24, 63, 77, 100, 108, 135, 143–145
- results
  - interpreting results,  $x$ - $x_i$ , 4, 76, 134, 138–140
  - reporting results,  $x$ , 44, 48, 50–51, 59, 111, 133, 138–140
- row, 25–27, 30, 116–117
- sample, 38–39, 77, 92
- sample mean, see averages

- sample size, 39, 104, 136, 149
- sample standard deviation, see spread
- sample variance, see spread
- scale, see data
- scatter plot, see plot
- significance, 87–89, 94–98, 118, 138, 140, 145, 150
- skewed data, see data
- spread
  - dispersion, 52
  - interquartile range, 52–53, 59–60, 70, 75, 145
  - quartile, 52–53, 60, 70
  - range, 52, 59–60, 75, 137
  - standard deviation (population and sample), 52–60, 75, 145, 149
  - variance (population and sample), 56, 59–60, 149
- spreadsheet, x, 9, 19, 25–30
- square, 8, 55–57
- square root, 8, 55–57
- standard deviation, see spread
- statistics
  - descriptive statistics, xi, 11, 31, 38–60, 101
  - inferential statistics, xi, 11–12, 77, 101
  - interpreting statistics, see interpreting results
  - statistical flowchart, see statistical signpost
  - statistical hypothesis, see hypothesis
  - statistical signpost
    - categorical, 100–103
    - nonparametric, 101, 104, 106
    - parametric, 101, 104–105
    - statistical flowchart, 102, 107
  - statistical tables, 88, 118, 125
  - statistical test
    - chi squared test, 103, 115, 118–119, 121, 150
    - one tailed test, 91
    - paired t-test, unpaired t-test, 103, 105–106, 136
    - pearson’s correlation, 105, 129–133, 150
    - two-tailed test, 89–90
  - stem and leaf, see plot
  - sum, 28, 42
  - survey, see questionnaire
  - symbols, 28, 149–150
  - symmetric, see data

t distribution, see probability

test statistic, 89, 97–98, 119–120, 124

true zero, 34–35

t-test, see statistical tests

two tailed, see statistical tests

type I error, see error

type II error, see error

type III error, see error

unpaired, see data

variance, see spread