APPENDIX 2

Software and Practice Corner

Since you are reading this chapter, you have probably finished reading the book. You might be ready for the next stage of your learning. The best way to learn statistics is to practise data analysis. You need a dataset and statistical software. Unfortunately for everyone, SPSS, the most ubiquitous statistical software, is rather expensive. Luckily, thanks to the generosity of the Free Software Foundation, we now have free statistical software that rivals SPSS. **PSPP**, available free to download, is designed to resemble SPSS. They are compatible with each other, and you can open any file you save in PSPP via SPSS and vice versa. Because of the similar interfaces, you should feel equally comfortable working in either of them. We have included a link for PSPP. The practice database is a PSPP file (saved as a .sav file). The statistical analyses are also performed in PSPP. If you have access to SPSS, courtesy of an institutional or personal subscription, you are welcome to use SPSS to work on the practice file. Those familiar with SPSS should be able to find the similarities between the two. Please bear in mind that PSPP does not offer the same functionality and statistical analyses options as SPSS. The limitations will become obvious as we attempt the practice files. The main disadvantage is the limited option for displaying charts compared to SPSS, but it should prove adequate for most of our everyday analyses.

Please follow the link below to visit the GNU operating system website and follow the link for '**downloading PSPP**' to download compatible versions of PSPP. The software is available to download from multiple servers:

www.gnu.org/software/pspp/#mission-statement

The software is being regularly updated but may contain bugs. It may be best not to have the most recent version. We have used version 1.2.0 that is available to download from the following link:

https://sourceforge.net/projects/pspp4windows/

For an introductory overview of PSPP, you may find the following resources useful. It will prove productive to become familiar with the software by going through these resources before attempting the practice questions. A brief introduction is also given in the next few pages.

Good luck!

An Introduction to PSPP

How to analyse data using PSPP:

www.youtube.com/watch?v=zSDHo8qFmZo

Another introductory lesson to PSPP:

www.youtube.com/watch?v=pK606kWar58

A comprehensive introduction to PSPP:

https://garyfisk.com/pspp/

Another PSPP resource from the California State University:

http://ssric.org/node/699

Let us familiarise ourselves with the different menu options in **PSPP**. A detailed description will take a lot of space and is unnecessary. You may choose instead to get acquainted yourself by clicking the different tabs and working through the software, but a brief introduction will be useful for the beginner. Once you open the practice file, the screen should look like Figure A2.1.

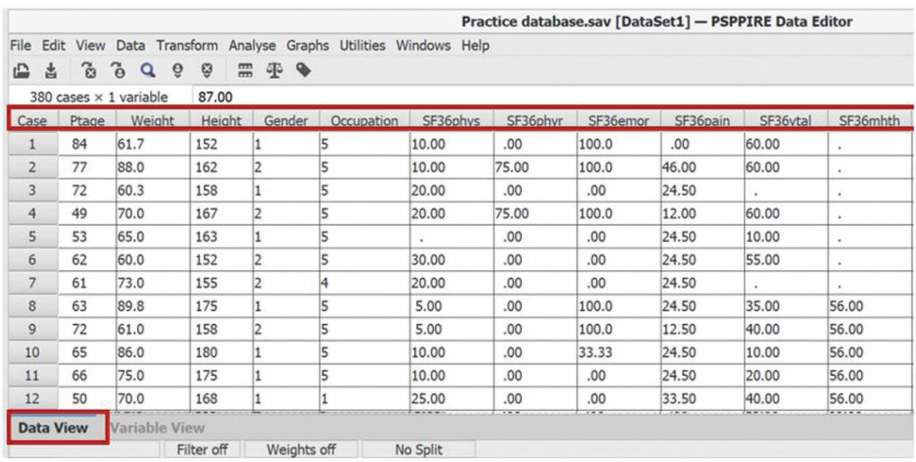


Figure A2.1 Data view page in PSPP.

This is the **Data editor** screen (Data View on the bottom left-hand side of the screen). The marked top row shows the names of the variables. This is the default data page view. If you click on the **Variable View** tab next to the data view tab, the screen will change to Figure A2.2.

Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1	Ptage	Numeric	5	0	Age	None	None	5	Right	Scale	Input
2	Weight	Numeric	4	1	Weight:	None	None	7	Right	Scale	Input
3	Height	Numeric	4	1	Height:	None	None	6	Right	Scale	Input
4	Gender	Numeric	1	0	Sex	{1, Male }...	None	4	Right	Nominal	Input
5	Occupatic	Numeric	1	0	Current C	{1, Heavy Manu	None	4	Right	Nominal	Input
6	SF36phys	Numeric	5	2	Pre-op Pl	None	None	8	Right	Scale	Input
7	SF36phyr	Numeric	5	2	Pre-op R(	None	None	8	Right	Scale	Input
8	SF36emo	Numeric	5	2	Pre-op R(	None	None	8	Left	Scale	Input
9	SF36pain	Numeric	5	2	Pre-op B(	None	None	8	Left	Scale	Input
10	SF36vtal	Numeric	5	2	Pre-op V	None	None	8	Left	Scale	Input
11	SF36mhth	Numeric	5	2	Pre-op M	None	None	8	Left	Scale	Input
12	SF36soci	Numeric	5	2	Pre-op S(	None	None	8	Left	Scale	Input

Data View | Variable View

Filter off | Weights off | No Split

Figure A2.2 Variable View page in PSPP.

The screen shows the details of the different variables. **Value labels** should be assigned to variables when applicable. To assign values, click on the value labels tab of the respective variable. A pop-up screen will appear where you can assign values.

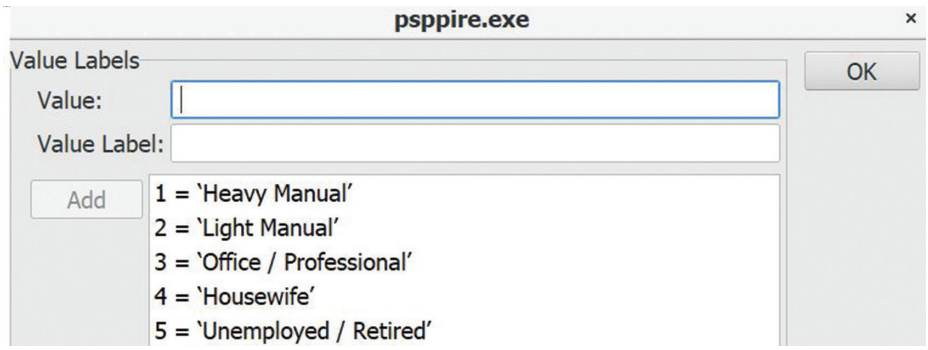


Figure A2.3 Value labels assigned to Occupation.

Figure A2.3 shows the values that we assigned to **occupation**. The variable view also shows the '**measure**' column (Figure A2.2). Just like in SPSS, PSPP divides the data into three types: scale (any measurement), ordinal or nominal. Most of the data in this set are of the scalar variety, but we also have nominal data (gender, occupation) (Figure A2.4). If you scroll down the list, you will find that **y5satisfaction** is indicated as an ordinal variable.

The variable view tab also has a column labelled '**Type**' next to the name of the variable (Figure A2.2). If you double-click on a cell under the Type column, a pop-up screen will appear with options to select the type of variables (Figure A2.5). We are only interested in the *numeric* variety, but there are options for special types, i.e. *date, comma, dot, scientific notation, string* etc. All the variables we enter in our database will be regular numbers, these can be assigned as '*Numeric*' variables. There is an option to assign names as '*String*' variables. They are treated as qualitative variables.

Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure
1	Ptage	Numeric	5	0	Age	None	None	5	Right	Scale
2	Weight	Numeric	4	1	Weight:	None	None	7	Right	Scale
3	Height	Numeric	4	1	Height:	None	None	6	Right	Scale
4	Gender	Numeric	1	0	Sex	{1, Male }...	None	4	Right	Nominal
5	Occupatic	Numeric	1	0	Current G	eavy Manual}...	None	4	Right	Nominal

Figure A2.4 Types of variables.

The most important tabs are the ones labelled '**Analyse**' and '**Graphs**' at the top of the screen in both Data or Variable view. If you click '**Analyse**' a smaller pop-up screen appears with options for various statistical tests. Clicking on '**Graphs**' similarly brings forth another pop-up screen with options for *Scatterplot*, *Histogram* or *Barchart* to display data.

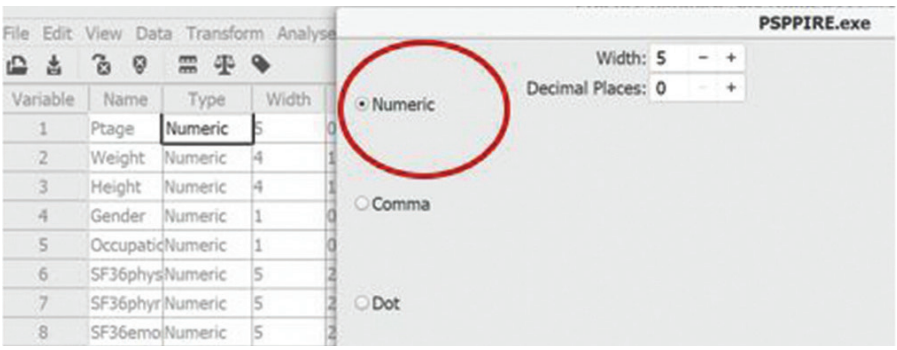


Figure A2.5 Options to choose types of variables in PSPP.

Analyse> Descriptive Statistics: There are four options for descriptive statistics when the analyse tab is clicked, these are: **Frequencies**, **Descriptives**, **Explore**, **Crosstabs**. Unlike SPSS, the **Frequencies** tab is rather useful in PSPP. Clicking this tab will open a pop-up window with options to choose **Statistics**, **Charts** and **Frequency tables**.

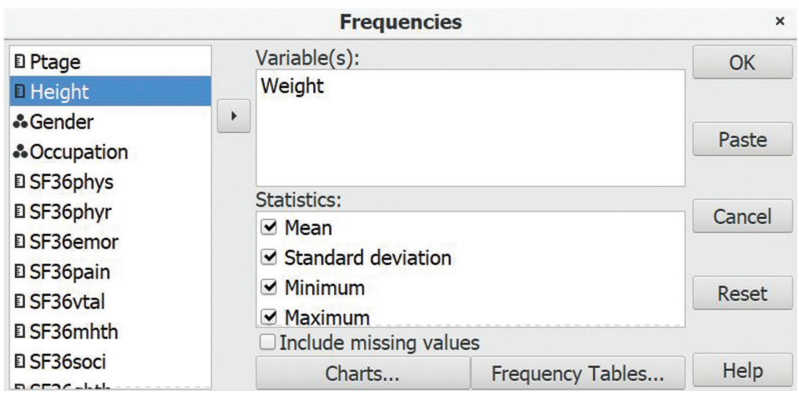


Figure A2.6 Analyse > Descriptive Statistics > Frequencies

Figure A2.6 shows the pop-up window for Frequencies. **Weight** is the variable chosen for analysis. The Statistics menu allows one to measure several statistical parameters, *Mean, Standard deviation, Minimum* and *Maximum* has been chosen. Clicking on the Chart menu opens another pop-up window with options to choose the display of descriptive statistics in *histogram, bar chart* or *pie chart*. The **Frequency table** tab (Figure A2.6) has the option for frequency analysis of each value of the chosen variable as well as of the missing values. The **Descriptives** option is similar to the Frequencies option but does not offer the option to display charts or frequencies.

The **Explore** option offers the same options for data analysis already found in frequencies with the additional advantage of subdivision into categories (Figure A2.7). In the example, weight has been selected as the dependent variable to explore. Gender has been selected as the factor. Clicking on the statistics tab will open another pop-up window with options for descriptives, extremes or percentiles. If we click on descriptives, the results of descriptive statistics for weight will be displayed in a separate output viewer window grouped by gender.

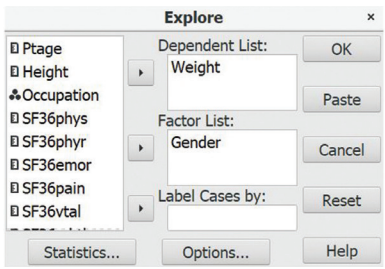


Figure A2.7 Explore option in PSPP.

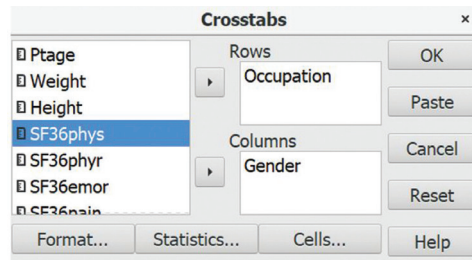


Figure A2.8 Crosstabs option in PSPP.

The **Crosstabs** option creates a contingency table for nominal or ordinal data. Figure A2.8 shows the crosstabs pop-up window, gender has been selected as the column variable and occupation as the row variable. Clicking on the **Statistics** tab will allow one the option to choose the relevant statistics. We are mainly interested in the Chi-Squared statistics.

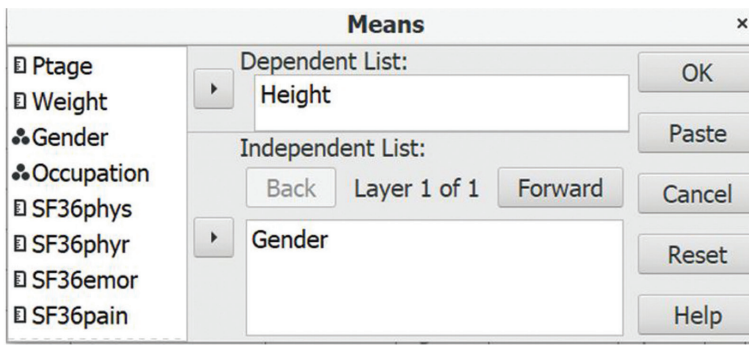


Figure A2.9 Analyse > Compare means > Means

Analyse> **Compare means**: this option offers parametric tests, including the *t*-test and ANOVA. **Analyse**> **Compare means**> **Means** allows the opportunity to compute the mean and standard deviation for a variable grouped into categories (Figure A2.9). In this example, we wished to investigate *height* (dependent variable) grouped by *gender* (independent variable). The drop-down menu for **Analyse** also offers separate options for correlation, regression and non-parametric analysis. The **Transform**> **Compute** tab is useful to transform a variable into a different one. *We shall learn how to use it to log transform a variable later in this appendix.*

If you wish to learn more about PSPP, read the reference manual. Choose **Help**> **Reference manual** from the menu. Most of the options of PSPP are available via the command options at the top of the screen and in the pop-up boxes. Additional options are available via the syntax command option (you can find this by clicking **File**>**New**>**Syntax**, a new pop-up window will appear offering you the option to write the syntax). This is similar to writing in programming language and may occasionally be necessary.

Now that you are familiar with the interface, it is time to begin data analysis. The included dataset is a fictional one. It is of a group of patients who had treatment for an illness. The researchers collected demographic data, pre-treatment general health score (**SF-36**), subsequent follow-up data at years one and five, and patient satisfaction with treatment at year five. SF-36 has several components and they are separately listed. You will need to consult the database to answer the first section of the practice corner. For the second section, you do not need the database.

We highly recommend that you take full advantage of the practice corner. We consider the practice corner as being complementary to the learning objectives of the book chapters. Almost all the statistical concepts and assumptions introduced in the book have been further elaborated here with practical examples. It should prove useful to reinforce the learning gained from the printed book. The questions are generally arranged in the same order as that of the book chapters. We hope that the exercise will help make you familiar with the different functions of PSPP and the commonly performed statistical tests.

Finally, we would also like to emphasise the cardinal rule of statistical tests, that before taking a test result for granted you should check whether the assumptions that underpin them have been satisfied or not. This is a mistake that many of us non-statisticians are guilty of committing when conducting statistical tests. The importance of checking underlying assumptions can not be over-emphasised because the validity of the tests are wholly dependent on them being true.

The answers to the practice corner questions are given in the latter part of this appendix.

Practice Corner: Questions

Q1. Let us begin with an analysis of the demographic data in the practice dataset.

a. How would you objectively assess if data distribution for age, height and weight followed the normal distribution?

b. What would be the appropriate measure of data distribution for the variables age, height and weight and why? Report the appropriate values.

c. What is your estimate of the likely range within which the true population mean of these three variables would be found?

d. What proportion of the population is likely to be shorter than 180cm?

Q2. How can you investigate if there was a difference in age between men and women?

Q3. What would be the appropriate significance test to perform to look for a difference in physical functioning score between pre-treatment (variable name: SF36phys) and one year after treatment (y1SF36phys)?

Q4. What non-parametric test would you perform to assess if treatment made any difference to SF36phys a year following treatment?

Q5. Assuming that data were normally distributed how would you find out if pre-treatment SF36 general health (SF36ghth) varied due to the occupation of the patient?

Q6. What would be the most suitable graphical plot to display vitality data over five years (SF36vtal)? Could you analyse data separately according to gender? (*You may wish to use syntax code to answer this question. If you find it too difficult, check the answer and try again with a different variable).*

Q7. Did treatment make any difference to bodily pain (SF36pain) over five years? At what stage, if any, was the difference significant?

Q8. How would you find out if there was any correlation between height and weight?

Q9. If you find a correlation between height and weight, can you quantify this relation further by using a regression equation? What would be the predicted weight of the patient who has a height of 167cm?

Q10. Write out the regression equation by adding age to the regression analyses. How does it affect the predictive power of the model?

Q11. Was there a difference in patient satisfaction (y5satisfac) between men and women?

Q12. How can you investigate if patients who were mentally distressed (SF36mhth) before initiation of treatment had lower satisfaction with treatment than those who were not? (*You may wish to transform SF36mhth; SF36mhth score 0–50 is known to be indicative of mental distress*).

Q13. Can you investigate if age (Ptage), occupation and pre-treatment mental health state (SF36mhth) could be used to predict post-treatment patient satisfaction?

You will not need to consult the database to answer the rest of the practice corner questions.

Q14. Researchers investigated if plant-based diets were effective in reducing dyslipidaemia. What was the primary outcome measure of the study? What kind of statistical test would be appropriate for this research?

Q15. Authors conducted a prospective non-inferiority randomised controlled trial (RCT) to determine the validity of the superstition that utterance of the word 'quiet' in a clinical setting increases workload. Twenty-nine days were assigned in which staff were to say 'Today will be a quiet day' randomly and 32 days were assigned in which staff were to refrain from saying the word 'quiet'. The primary composite outcome was 'clinical episodes'. A difference of 30 clinical episodes was considered as the margin of non-inferiority [1].

a. What was the null hypothesis of this trial? What was the alternative hypothesis?

A mean 139.0 clinical episodes occurred on control days compared with 144.9 on days when the experimental intervention was uttered; a difference of 5.9 episodes (95% confidence interval – 12.9 to 24.7).

b. How will you interpret the results?

Q16. Researchers conducted an open-label RCT to assess the effectiveness of convalescent plasma in the management of Covid-19 in adults. The primary outcome was all-cause mortality or progression to severe disease at 28 days post-enrolment. Four hundred and sixty four adults were recruited; 235 were assigned to convalescent plasma with the best standard of care (intervention arm) and 229 to the best standard of care only (control arm) [2].

a. What would be your null hypothesis?

b. Since this was an open-label trial what was the risk of detection bias with regards to the stated outcomes?

In the intervention group (n = 235), seven did not receive it, of the 228 who received the intervention, one patient was lost to follow-up and two more patients did not receive the full treatment dose. In the control group (n = 229), three patients received plasma instead, four participants withdrew consent, one was later lost to follow-up.

c. How many patients would qualify for intention-to-treat analysis and how many patients would qualify for per-protocol analysis?

The primary composite outcomes were as follows:

Intervention group: n = 235; all-cause mortality at 28 days or progression to severe disease: n = 44. Control group: n = 229; mortality or disease progression: n = 41.

d. Can you perform an appropriate statistical test to investigate for any differences in outcome?

e. What would be your preference for conveying risk of events in the intervention groups, risk or odds?

f. If you describe risk ratio in the intervention group would this be adjusted or unadjusted?

Secondary outcomes included total hospital stay in days and total days of respiratory support, the results are, intervention group: median total hospital stay in days (interquartile range (IQR)): 14 (10–19); control group: median total hospital stay in days (IQR): 13 (10–18)).

g. What would be an appropriate significance test to assess for a difference in hospital stay?

Q17. Researchers conducted an RCT to determine women's satisfaction with pain relief during labour using patient-controlled analgesia with remifentanil compared to epidural analgesia. This was an equivalence trial. The primary outcome was satisfaction with pain relief, measured on a visual analogue scale and expressed as area under the curve (AUC); a higher AUC represented higher satisfaction with pain relief. Ten per cent reduction on the visual analogue scale for satisfaction with pain relief was the equivalence margin [3].

a. What were the null and the alternative hypotheses for this study?

Pain relief was ultimately used in 65% in the remifentanil group and 52% in the epidural analgesia group. Of women primarily treated with remifentanil, 13% converted to epidural analgesia, while in women primarily treated with epidural analgesia 1% converted to remifentanil.

b. How would you analyse these results?

The area under the curve for total satisfaction with pain relief was 30.9 in the remifentanil group versus 33.7 in the epidural analgesia group. The AUC for satisfaction with pain relief was 25.6 in the remifentanil group versus 36.1 in the epidural analgesia group (mean difference −10.4, −13.9 to −7.0).

c. What is your conclusion concerning the null hypothesis?

Q18. In a recent study, researchers investigated if antibiotics prophylaxis was able to prevent maternal infection after operative vaginal birth. This was a blinded RCT. The authors reported that a confirmed or suspected infection rate in the group allocated to the antibiotic treatment was 11% (180/1619) and the group allocated to placebo was 19% (306/1606) [4].

a. What was the risk ratio of infection in the antibiotic allocated group?

b. What is the number needed to treat (NNT) to prevent maternal infection after operative vaginal birth if treated with prophylactic antibiotics compared to placebo?

c. What are the odds ratios of infection in the antibiotic allocated group?

Q19. Researchers assessed whether resistant hypertension is an independent predictor for all-cause mortality in individuals with type 2 diabetes. Based on baseline blood pressure (BP), they categorised participants as normotensive, untreated hypertensive, controlled hypertensive, uncontrolled hypertensive, or resistant hypertensive. Kaplan–Meier analysis was used to assess all-cause mortality. Figure Q19.1 shows the unadjusted cumulative survival figures for different categories of hypertension [5].

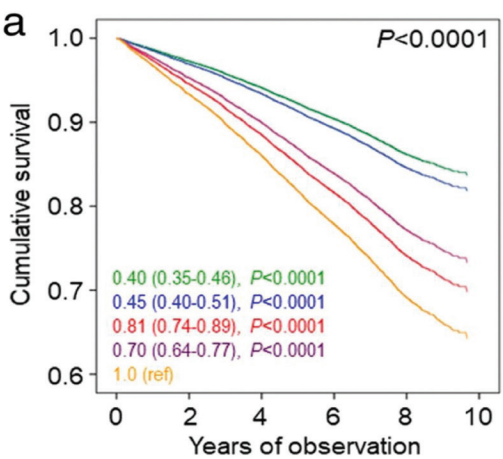


Figure Q19.1 Unadjusted cumulative survival figures for different categories of hyptertension. © BMC Medicine, reproduced under CC BY 2.0 [5].

Green – normotensive
Blue – untreated hypertensive
Red – controlled hypertensive
Purple – uncontrolled hypertensive
Yellow – resistant hypertensive

a. How would you interpret the survival analysis figure? (*ignore the CI margin*)

b. What was the hazard of interest in this study?

c. Can you please calculate the median survival rate?

d. How would you find out if the differences between the apparent survival curves were significant?

e. How could you confirm if there was a trend in increasing mortality according to the different categories of hypertension?

f. These patients would have several co-morbidities and these factors would probably influence the all-cause mortality. How can you analyse the results by adjusting for the effect of the other risk factors?

Q20. Let us consider an area with a high prevalence of the Coronavirus. Let us consider the prevalence of viral infection to be 10%, the sensitivity of the coronavirus testing kit to be 70% and the specificity to be 90%. If an individual is randomly tested for the virus and tests positive:

a. What is the probability that the person has contracted the Coronavirus?

b. What is the probability of contracting the virus even if the test is negative?

Q21. Researchers undertook a systematic review to investigate the sensitivity and specificity of superficial wound cultures compared to deep tissue cultures in lower extremity wounds. They found the pooled sensitivity for superficial wound swab was 49% and specificity was 62%. How likely is it that the results of superficial wound culture would influence the post-test probability of wound infection?

Q22. Given the prostate cancer prevalence among men of 0.8%, if the sensitivity of the Prostate Specific Antigen (PSA) test is 90% and the specificity is 85%:

a. What is the probability that a man with a positive PSA test result has prostate cancer?

b. What are the pre- and post-test odds of prostate cancer in someone with a positive test?

Q23. Researchers conducted a systematic review and meta-analysis to investigate the effect of topical Tranexamic acid (TXA) on blood loss and rates of transfusion after total hip and knee replacement surgery. The review included 14 RCTs. They were able to pool the results from nine trials for meta-analysis. The Chi-Squared value for heterogeneity was 3.80, the I^2 value was 0.

a. What is your conclusion regarding the evidence for heterogeneity in the meta-analysis?

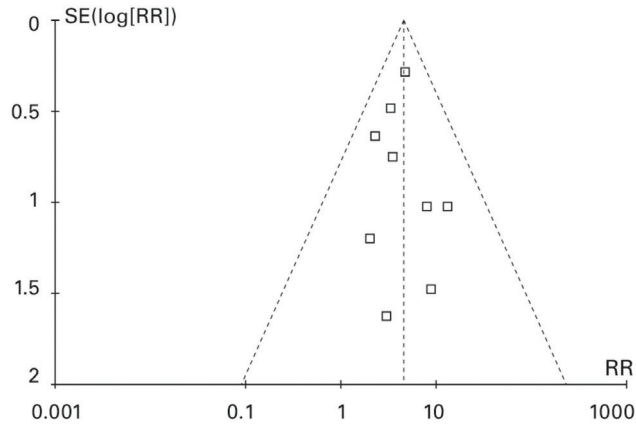


Figure Q23.1 The Funnel plot.

Reproduced with permission and © of The British Editorial Society of Bone & Joint Surgery [6].

b. Figure Q23.1 displays a funnel plot; is there any evidence of publication bias?

c. Figure Q23.2 displays the meta-analysis of studies investigating the effect of topical TXA in total knee replacement. How would you interpret the results? (the outcome of interest was blood transfusion).

	Control		Topical TXA			Risk ratio	Risk ratio
Study or subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Fixed, 95% CI**	**M-H, Fixed, 95% CI**
Canata et al[38] 2012	2	32	1	32	4.4%	2.00 [0.19 to 20.97]	
Wong et al[11] 2010	5	35	4	64	12.4%	2.29 [0.66 to 7.97]	
Ishida et al[35] 2011	1	50	0	50	2.2%	3.00 [0.13 to 71.92]	
Sa-Ngasoongsong et al[36] 2011	10	45	6	90	17.5%	3.33 [1.29 to 8.59]	
Roy et al[31] 2012	7	25	2	25	8.8%	3.50 [0.80 to 15.23]	
Seo et al[34] 2012	47	50	10	50	43.8%	4.70 [2.69 to 8.22]	
Sa-Ngasoongsong et al[37] 2013	8	24	1	24	4.4%	8.00 [1.08 to 59.13]	
Georgiadis et al[33] 2013	4	51	0	50	2.2%	8.83 [0.49 to 159.80]	
Alshryda et al[39] 2013 (TRANX-K)	13	78	1	79	4.4%	13.17 [1.76 to 98.24]	
Total (95% CI)		**390**		**464**	**100.0%**	**4.51 [3.02 to 6.72]**	
Total events	97		25				

Heterogeneity: Chi² = 3.80, df = 8 (P = 0.87); I² = 0%
Test for overall effect: Z = 7.38 (P < 0.001)

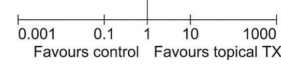


Figure Q23.2 The Forest plot.

Reproduced with permission and © of The British Editorial Society of Bone & Joint Surgery [6].

Practice Corner: Answers

A1.

a. One would assume that everyday continuous variables like age, height, weight etc should be normally distributed in a large database, but it would be best to confirm this assumption. This can be performed in several ways. A simple option is to visually analyse a histogram. We go to the menu, choose **Graphs>histogram**, tick the box for **Display normal curve>OK**. We can select this function for **ptage, weight** and **height**.

The results will appear in the PSPP output viewer in a separate window (Figure A2.10). PSPP will compute the mean value, SD and the number of cases and will display the histogram. On visual inspection, it appears that **height** is the variable most symmetrically distributed, and the distribution of **age** appears reasonably normal. The distribution of age does have a left-sided tail, although not very prominent. The graph indicates that most patients were around 64 years old with a height of 167cm (mean). **Mean** and **SD** are suitable measures of data spread for **age** and **height**. The distribution of **weight** does have a more extended right-sided tail. The graph suggests that most patients weighed around the mean value of 76kg, but there were some overweight patients in the group as well. However, none of the three graphs indicates a large departure from the assumption of normality. It would be reasonable to present the data with parametric measures: **mean** and **SD** (included in the figure).

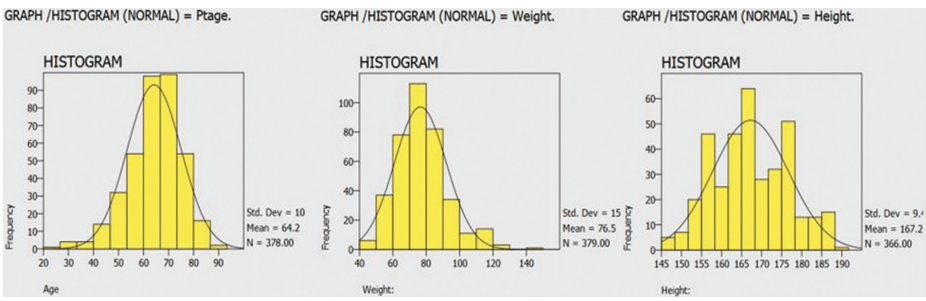


Figure A2.10 PSPP output viewer. Histogram with superimposed normal distribution curve for age, height and weight.

Most statisticians advise that visual inspection of a histogram is a better measure of data distribution than undertaking statistical tests. However, we have included a statistical test just so that you are aware of the test and know how to use it when necessary.

This test is known as the **Kolmogorov-Smirnov (KS) test for normality**. The null hypothesis of the test is that data follow a normal distribution. The test calculates to

what extent (% of cases) data deviates from a normal distribution curve. If the null hypothesis were not false, the deviation would be small, the p-value would be large, and we would accept the null hypothesis. If it were false, the differences would be more substantial, and the p-value would be low. To perform the KS test for age, weight and height, we choose **Analyse>Non-parametric statistics>1 sample KS test**. We click **Normal** for Test distribution box. Table A2.1 shows the results.

One-Sample Kolmogorov-Smirnov Test

		Age	*Height:*	*Weight:*
N		378	366	379
Normal Parameters	*Mean*	64.25	167.21	76.49
	Std. Deviation	10.80	9.46	15.57
Most Extreme Differences	*Absolute*	.08	.08	.09
	Positive	.04	.08	.09
	Negative	-.08	-.06	-.05
Kolmogorov-Smirnov Z		1.57	1.46	1.66
Asymp. Sig. (2-tailed)		.009	.019	.005

Table A2.1 KS test for normality for age, height and weight.

According to the test, there is strong evidence against the null hypothesis ($p < 0.05$). This happened because the test is too powerful when the sample is large. We saw from visually inspecting the histograms that departure from normality was minimal (*this can be further confirmed later when we compare the difference between mean and median, a marker of how skewed the data distributions are*). This is a good example of why visual inspection may be more appropriate to assess data distribution rather than conducting the KS test.

Let us assume we were convinced that at the very least weight was not normally distributed and required further consideration. In this case, we would present the median and the interquartile range as the appropriate measures of data distribution. Let us explore how to transform a variable. We could attempt to log-transform weight. This would reduce the variations and create a more normal variable. We choose the command **Transform>Compute**. A new pop-up screen comes up at this stage (Figure A2.11). We have to choose a name for the transformed variable, let's choose **Lnwt** as the target variable. Next, we need to write the formula. The formula for log transformation is **LN(variable name)> LN(Weight)**. We write the formula in the window under '*numeric expression*'. Click OK and a new variable is created.

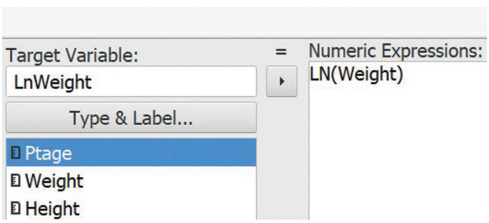


Figure A2.11 Log transformation of weight.

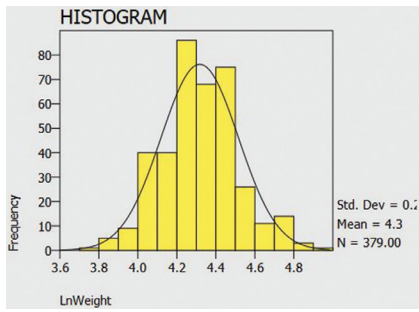


Figure A2.12 Histogram of Lnwt.

Let us check the distribution of the new variable with a histogram (Figure A2.12). The new distribution appears more symmetrical but one could re-run the KS test to double-check it. Please note that the mean value of the transformed variable is smaller than the mean value of the original. This is because it is log-transformed data. *This exercise was included for the sake of learning but is not necessary for this question. It is a diversion from which we shall now go back on track.*

b. Since the variables were reasonably normally distributed, mean and SD would be the appropriate measures of data distribution. We can find mean value and the SD by selecting **Analyse>Descriptive statistics>Explore**. The **Explore** window will pop-up. We find age, weight and height and click on the arrow to send these variables onto the dependent list. Next, click on **Statistics**, a smaller pop-up window will appear. We check on the **Descriptives** box, click on **Continue**, then **OK**. The resulting output will display the required information (Figure A2.13). Note the difference between mean and median for each variable that indicates that departure from normality was minimal.

			Statistic	*Std. Error*
Weight:	Mean		76.71	.82
	95% Confidence Interval for Mean	Lower Bound	75.10	
		Upper Bound	78.31	
	5% Trimmed Mean		75.91	
	Median		75.00	
	Variance		243.14	
	Std. Deviation		15.59	
	Minimum		44.00	
	Maximum		140.00	
	Range		96.00	
	Interquartile Range		18.25	
	Skewness		.81	.13
	Kurtosis		.94	.25
Age	Mean		64.20	.57
	95% Confidence Interval for Mean	Lower Bound	63.08	
		Upper Bound	65.33	
	5% Trimmed Mean		64.68	
	Median		65.00	
	Variance		119.02	
	Std. Deviation		10.91	
	Minimum		23.00	
	Maximum		90.00	
	Range		67.00	
	Interquartile Range		14.00	
	Skewness		-.67	.13
	Kurtosis		.64	.25
Height:	Mean		167.24	.50
	95% Confidence Interval for Mean	Lower Bound	166.27	
		Upper Bound	168.22	
	5% Trimmed Mean		167.17	
	Median		168.00	
	Variance		89.47	
	Std. Deviation		9.46	
	Minimum		145.00	
	Maximum		189.00	
	Range		44.00	
	Interquartile Range		15.00	
	Skewness		.10	.13

Figure A2.13 Output of descriptive statistics.

c. We can never be sure regarding the value of the true population mean. However, the respective 95% CI of the mean value represents a plausible range within which the true population mean is likely to be found. Please note that when descriptive statistics are discussed there is no need to estimate the 95% CI; CI is an inferential statistic. In this instance, we are attempting to estimate the unknown true population mean from our limited sample and as such CI is necessary.

d. To find out the proportion of the population shorter than 180cm in height we need to find the *z*-value of 180cm. The formula for *z* score is:

$$z = \frac{x - \mu}{\sigma}$$

$x = 180$; $\mu = 167.24$; $\sigma = 9.46$; $z = 1.348$

A *z*-value of 1.35 means that a person with a height of 180cm is 1.35 SD away from the mean value; quite tall compared to the mean height. If we consulted a statistics table, we would find that a *z*-value of 1.35 has a P lower (lower tail) of 0.9114 and P upper (upper tail) of 0.0885. This means that a person with a height of 180cm is taller than 91.14% of the population and shorter than 8.85% of the same. Therefore, 91.14% of the population would be shorter than 180cm in height.

A2.

We have already seen that **ptage** is normally distributed. We want to test for differences in age between men and women. Data are independent, continuous, normally distributed. The independent samples *t*-test is the appropriate statistical test to perform. We go to **Analyse>Compare means>Independent samples t-test**, a new pop-up window appears. We choose **ptage** as the Test variable and **gender** as the grouping variable. A new window appears asking us to define the groups. We assign 'male' as group 1 and 'female' as group 2 (Figure A2. 14).

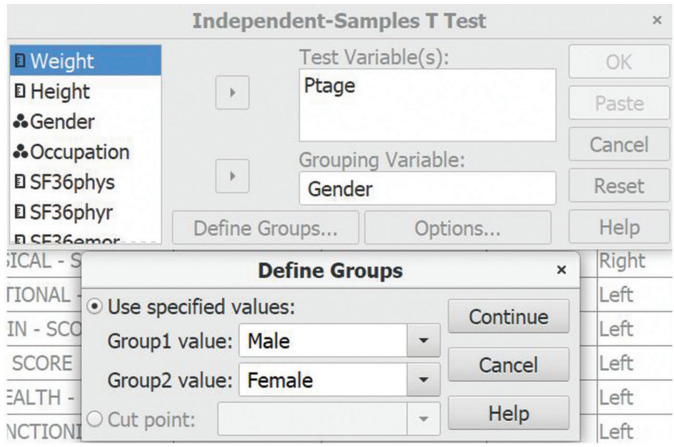


Figure A2.14 Independent samples *t*-test.

Group Statistics

	Sex	N	Mean	Std. Deviation	S.E. Mean
Age	Male	158	62.42	10.32	.82
	Female	220	65.56	10.96	.74

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Age	Equal variances assumed	.14	.705	-2.81	376.00	.005	-3.14	1.12	-5.33	-.94
	Equal variances not assumed			-2.84	349.86	.005	-3.14	1.10	-5.31	-.96

Figure A2.15 Output tables for independent samples *t*-test.

We click **OK** and the results are displayed in the output screen (Figure A2.15). We should recall that for the *t*-test to be valid the variances between the groups should be similar. We find from the first table that the SDs of male and female age were similar, therefore equal variances assumption is valid. The top table also informs us that the mean age among men and women was 62.42 and 65.56 respectively. We now look at the first row of the bottom table, the mean difference is −3.14, women were slightly older than men. The *t*-statistic was −2.81, which is significantly different for 376 degrees of freedom (*df*) (p = 0.005).

Our conclusion is: *Men and women were significantly different in age (t-statistic −2.81, df 376, p = 0.005).*

A3.

Physical functioning score is not truly interval data but can be considered as a continuous variable. If the data were normally distributed the appropriate significance test would be the paired samples *t*- test. We should be able to recall that for a *t*-test to be valid two assumptions need to be valid. Let us assess the distribution of data (Figure A2.16). The histograms indicate that both values are asymmetrical in distribution but pre-op score (**SF36phys**) is more so. Pre-op physical functioning data were right-skewed, and Year 1 score (**y1SF36phys**) was left-skewed. The shapes of the distributions mean that the bulk of the patients in the pre-treatment stage had a lower physical functioning score, but following treatment, the trend somewhat reversed. This is encouraging news.

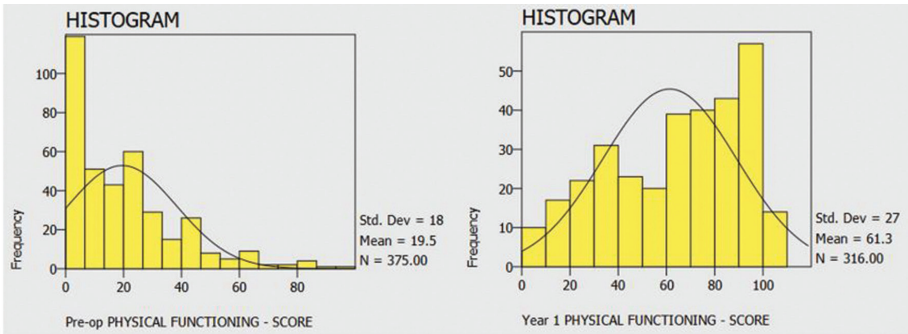


Figure A2.16 Histogram of Pre-op and Year 1 physical functioning.

For paired samples *t*-test it is not the variables themselves but the paired difference that should be normally distributed. To assess this, we first need to find the paired difference. We can use the **Syntax** command to do this. We go to **File**>**New**>**Syntax**, a new window of **Syntax editor** opens. We type the code shown in Figure A2.17. The first line is a general instruction to the software to compute, the second line specifies that we want PSPP to subtract SF36phys from y1SF36phys. Next, we click **Run**> **All**. A new variable **sub_1** is created by the software. Next, we choose this variable to display the histogram. The resulting graph is displayed in Figure A2.18.

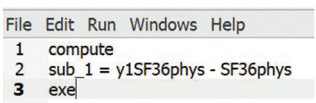


Figure A2.17 Syntax for subtraction.

sub_1

N	*Valid*	313
	Missing	67
Mean		42.57
Std Dev		27.37
Kurtosis		-.28
Skewness		-.27
Percentiles	50 (Median)	45.00

Figure A2.19 Sub_1 statistics.

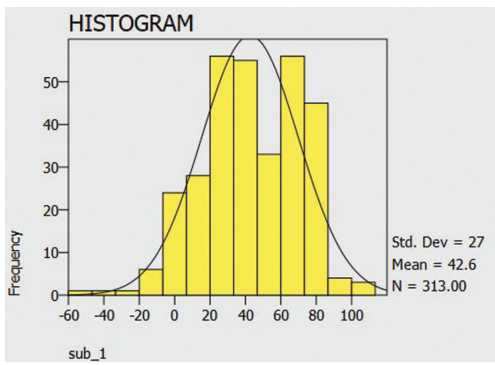


Figure A2.18 Histogram of sub_1.

The histogram suggests that the paired differences are reasonably normally distributed. We have previously seen that one can conduct the KS test to investigate if data are normally distributed or not. We also saw previously that the test can be too sensitive for large samples and may indicate strong evidence against the normal distribution even if the departure from normality was minimal.

Another objective measure of assessing data distribution is to measure skewness and kurtosis. Skewness is a measure of the symmetry of the data distribution. Symmetrical data has a skewness value of '0'. The more asymmetrical the data the further the value from '0'. Positive values of skewness indicate right-skewed distribution and vice versa. Kurtosis is a measure of the flatness or the spreading out of the data values. The more spread the data, the flatter the peak of data distribution.

A kurtosis value of 3 indicates normally distributed data. PSPP has a kurtosis value of '0' for normal distribution; positive kurtosis value indicates a right-tailed data and negative kurtosis value indicates a left-tailed data. To assess skewness and kurtosis we go to **Analyse>Descriptive Statistics>Descriptives**, we choose **sub_1** and transfer it to the variables window and click *mean, SD, median, skewness* and *kurtosis* in the **Statistics** dialogue box. Figure A2.19 shows the results. There was minimal difference between mean and median and skewness and kurtosis values were near to 0.

The second assumption is that the variances were equal. Equal variances assumption is not required for paired samples *t*-test. However, should we be interested, a rough guide of variance is the SD. For the equal variances assumption to be valid the size of the SD of one group should not be more than double that of the other. We observe that SD of the pre-op score is 17.16 and that of year 1 score is 27.86. There was more variability in SF-36 score a year following treatment compared to before treatment, but it is within the margin of allowance. The assumption of equal variances would have been valid.

Let us proceed with the paired samples *t*-test. The test is available under the commands **Analyse>Compare means>Paired samples t-test**. Our null hypothesis is that the mean of the paired differences is '0'. The results are as follows (Figure A2.20).

Paired Sample Statistics

		Mean	N	Std. Deviation	S.E. Mean
Pair 1	Pre-op PHYSICAL FUNCTIONING - SCORE	18.72	313	17.16	.97
	Year 1 PHYSICAL FUNCTIONING - SCORE	61.29	313	27.86	1.57

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Pre-op PHYSICAL FUNCTIONING - SCORE & Year 1 PHYSICAL FUNCTIONING - SCORE	313	.34	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Pre-op PHYSICAL FUNCTIONING - SCORE - Year 1 PHYSICAL FUNCTIONING - SCORE	-42.57	27.37	1.55	-45.62	-39.53	-27.52	312	.000

Figure A2.20 Output of paired samples *t*-test.

Please note that the output table identifies the variables according to their label (for example SF36phys is the variable name, the variable is labelled as 'Pre-op physical functioning').

Of the three output tables, the top one shows the mean and the SD. The results confirm that physical functioning score improved from 18.72 before treatment to 61.29 at year 1 following treatment.

The middle output table shows the degree of correlation between each pair of samples.

The bottom output table demonstrates the calculations of the test statistic. The mean difference between pairs of values is −42.57, 95% CI of the difference does not contain the null value (95% CI −45.62 to −39.53). The probability of observing a test statistic of −27.52 for 312 degrees of freedom is very slim ($p < 0.001$) by chance alone. Treatment resulted in a significant improvement in physical functioning.

A4.

Our data are paired. Wilcoxon matched-pair signed rank test is the appropriate non-parametric test to conduct. Since it is a non-parametric test, we do not need to test any assumptions. Our null hypothesis is that the median of the differences between the paired observations is 0. There is some advantage to performing a non-parametric test in this scenario as the scores are not perfect measurements, it is not interval data but ranks and non-parametric tests investigate the differences in ranks. From the top menu, we choose **Analyse>Non-parametric statistics>2 related samples**. The test pair will be **SF36phys** and **y1SF36phys**. We choose the Wilcoxon test (Figure A2.21).

Ranks

		N	Mean Rank	Sum of Ranks
Pre-op PHYSICAL FUNCTIONING - SCORE - Year 1 PHYSICAL FUNCTIONING - SCORE	Negative Ranks	290	158.74	46036.00
	Positive Ranks	15	41.93	629.00
	Ties	8		
	Total	313		

Test Statistics

	Pre-op PHYSICAL FUNCTIONING - SCORE - Year 1 PHYSICAL FUNCTIONING - SCORE
Z	-14.74
Asymp. Sig. (2-tailed)	.000

Figure A2.21 Output of Wilcoxon matched-pair signed rank test.

Each patient's pre-op score is paired with the same patient's year 1 score and the differences noted. In 290 cases out of 313, the year 1 score was higher and in 15 instances it was lower. When the scores are tied (8), they are ignored. The negative and positive signs are ignored, and the differences ranked in order of their absolute values and summed. The sum of the negative ranks (**T negative = 46036**) is greater than the sum of the positive ranks (**T positive = 629**).

If our null hypothesis was not false, T positive and T negative would be similar. If it was false there would be a difference between the two scores and the sums would be different. The more different the two sums, the more different the two scores. The test statistic W is the smaller of the two sums (W = 629). We expect a certain value of W when the null hypothesis is not false. If it was false and there was indeed a difference between the paired observations, W would be smaller than our expected value of W. We want to know what are the chances that our observed W could be smaller than the expected W by chance alone if the null hypothesis was not false. The p-value of the test statistic is < 0.001. The chances are very slim. The median of the differences between the paired observations is not 0.

We can conclude that there was a significant difference in physical functioning score a year following treatment compared to that before the initiation of treatment. We should report the results as the following:

> *The difference between pre-op physical functioning score and year 1 physical functioning score was significant (Wilcoxon matched-pair signed rank test, $n = 313$, $W = 629$, $p < 0.001$).*

A5.

Since we are assuming that data are normally distributed a parametric test would be appropriate. Occupation has five categories; we have to test more than two means. We should perform the One-way ANOVA test. The easy option is to choose One-way ANOVA from the menu. We go to **Analyse>Compare means>One way ANOVA**. A new pop-up window appears (Figure A2.22). We choose **SF36ghth** as the Dependent variable and **occupation** as the factor. The limitation of using this method is that PSPP does not give us any option for any post-hoc tests which are essential to find out which group is different. To enable PSPP to perform post hoc tests we have to go to the Syntax mode. We choose **File>New>Syntax**, a new window opens (Figure A2.23).

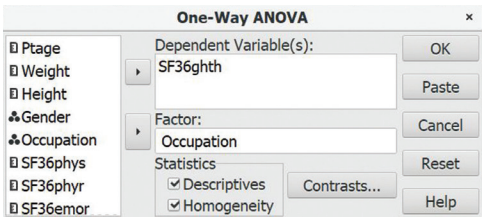


Figure A2.22 One-way ANOVA.

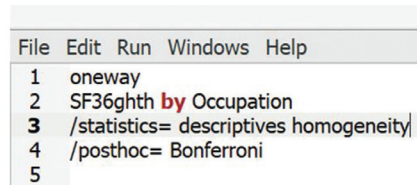


Figure A2.23 One-way ANOVA.

The first line instructs the software to conduct the one-way ANOVA test. The second line specifies the dependent variable (SF36ghth) and the factor (by Occupation). The third and fourth line instructs PSPP to display descriptives and homogeneity statistics and to perform the Bonferroni test. The results are seen in several tables. Figure A2.24 is the first output table and shows descriptive statistics. The light manual group had the lowest general health score (65.14) and housewives had the highest score (72.03). The next table (Figure A2.25) shows the results of the ANOVA test.

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Pre-op GENERAL HEALTH - SCORE	*Heavy Manual*	22	68.82	15.29	3.26	62.04	75.60	40.00	97.00
	Light Manual	29	65.14	22.03	4.09	56.76	73.52	20.00	95.00
	Office / Professional	46	70.96	20.61	3.04	64.84	77.08	25.00	100.00
	Housewife	30	72.03	19.15	3.50	64.88	79.18	25.00	100.00
	Unemployed / Retired	250	69.27	19.63	1.24	66.82	71.71	15.00	100.00
	Total	377	69.35	19.64	1.01	67.36	71.34	15.00	100.00

Figure A2.24 Descriptive statistics, one-way ANOVA.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Pre-op GENERAL HEALTH - SCORE	*Between Groups*	857.14	4	214.28	.55	.697
	Within Groups	144148.64	372	387.50		
	Total	145005.78	376			

Figure A2.25 Output of one-way ANOVA.

These are the same tables that would be reproduced if we chose the One-way ANOVA option via the analyse menu. The between-groups sum of squares is smaller than the within-groups sum of squares. The *F*-statistics is low (0.55). There is no evidence

against the null hypothesis to suggest that there was any difference in pre-treatment general health between the different categories of occupation (p = 0.697). The last table (Figure A2.26) would indicate whether there was a difference between the groups.

Multiple Comparisons (Pre-op GENERAL HEALTH - SCORE)

	(I) Current Occupation:	(J) Current Occupation:	Mean Difference (I - J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Bonferroni	Heavy Manual	Light Manual	3.68	5.57	1.000	-12.04	19.40
		Office / Professional	-2.14	5.10	1.000	-16.55	12.27
		Housewife	-3.22	5.53	1.000	-18.82	12.39
		Unemployed / Retired	-.45	4.38	1.000	-12.81	11.91
	Light Manual	Heavy Manual	-3.68	5.57	1.000	-19.40	12.04
		Office / Professional	-5.82	4.67	1.000	-19.00	7.36
		Housewife	-6.90	5.13	1.000	-21.37	7.58
		Unemployed / Retired	-4.13	3.86	1.000	-15.03	6.77
	Office / Professional	Heavy Manual	2.14	5.10	1.000	-12.27	16.55
		Light Manual	5.82	4.67	1.000	-7.36	19.00
		Housewife	-1.08	4.62	1.000	-14.12	11.97
		Unemployed / Retired	1.69	3.16	1.000	-7.23	10.61
	Housewife	Heavy Manual	3.22	5.53	1.000	-12.39	18.82
		Light Manual	6.90	5.13	1.000	-7.58	21.37
		Office / Professional	1.08	4.62	1.000	-11.97	14.12
		Unemployed / Retired	2.77	3.80	1.000	-7.98	13.51
	Unemployed / Retired	Heavy Manual	.45	4.38	1.000	-11.91	12.81
		Light Manual	4.13	3.86	1.000	-6.77	15.03
		Office / Professional	-1.69	3.16	1.000	-10.61	7.23
		Housewife	-2.77	3.80	1.000	-13.51	7.98

Figure A2.26 Post-hoc test, one-way ANOVA.

Let us recall that the other assumption of the one-way ANOVA is that the variances of the distributions are equal. This assumption needs to be tested. We can just compare the SDs of different groups to test this assumption. We can also explore variance with the **Levene's test**. Figure A2.22 shows that under the Statistics tab both Descriptives and Homogeneity have been ticked. We also instructed PSPP to undertake this test when we wrote the syntax commands (Figure A2.23). This will instruct PSPP to perform the Levene's test of homogeneity (Figure A2.27). The p-value is very high. Therefore, there is no evidence to reject the null hypothesis that the distributions have equal variance. The assumption is valid.

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
Pre-op GENERAL HEALTH - SCORE	1.04	4	372	.389

Figure A2.27 Test of homogeneity.

The one-way ANOVA option does not allow one to test for the effect of more than a single factor. If we wished to test for the effect of more than a single factor, we have to choose **Analyse** > **Univariate analysis**. Let us assume we wish to investigate the effect of both **occupation** and **gender** on **SF36ghth**. We choose SF36ghth as our dependent variable and gender and occupation as fixed factors. If we click **OK**, the results would be displayed in the output table (Figure A2.28). There is no evidence to reject the null hypothesis of no difference between the different groups. The interaction of occupation and sex did not affect the results; none of the three null hypotheses are rejected.

Tests of Between-Subjects Effects

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Corrected Model	2350.25	9	261.14	.67	.734
Intercept					
Current Occupation:	728.69	4	182.17	.47	.759
Sex	.00	1	.00	.00	1.000
*Current Occupation: * Sex*	1137.41	4	284.35	.73	.571
Error	142655.53	367	388.71		
Total	1958165.00	377			
Corrected Total	145005.78	376			

GLM

GLM SF36ghth BY Occupation Gender.

Figure A2.28 Two-way ANOVA results.

A6.

We need to assess the data distribution first to determine whether a parametric or non-parametric measure of data distribution would be the most suitable plot to display. We assess data distribution utilising the histogram function and find them to be asymmetrical. Therefore, we shall use a box plot. Unfortunately, there is no easy method to generate a box plot in PSPP. We need to employ syntax command. To do this, we go to **File**> **New**>**Syntax**. A new smaller pop-up window, syntax editor will appear. We need to type the following syntax command (Figure A2.29).

File Edit Run Windows Help

```
1  examine
2  variable= SF36vtal y1SF36vtal y5SF36vtal by
3  gender
4  /statistics= descriptives
5  /plot= boxplot
6  /compare=variable
7  /percentile
```

Figure A2.29 Syntax editor screen, boxplot commands.

You will find this command if you search for boxplot in the PSPP manual. The first line is a general instruction. The second line specifies the variables to be displayed. The third line instructs the software to break down the analysis by gender. The fourth line will enable the display of descriptive statistics. The fifth line is an instruction to display boxplot. The compare variable command will instruct a display of the variables in a single box so that we can easily compare them. The percentile command is to enable calculation of the interquartile range (IQR). Next, we choose **Run**>**All** from the top of the pop-up screen. The results will be displayed in the output viewer (Figure A2.30). We have already learnt about boxplot. The bold black horizontal line inside the box is the median, the box shows the IQR, the outliers are indicated by the circles and their case numbers.

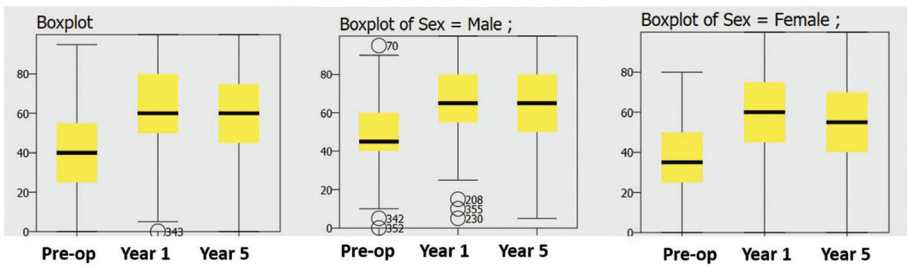


Figure A2.30 Box-plot of pre-op, yr 1 and yr 5 vitality, divided by gender.

The **descriptives** statistics box in the output viewer will also display the descriptive statistics, including the median and the IQR. We can conclude from the display that the median value of pre-op vitality score was 40 (IQR = 25–55), and improved following

treatment to 60 at year 1 (IQR = 50–80), the improvement was maintained at year 5 following treatment (median = 60, IQR = 45–75).

We can also conclude that the trend of improvement appears similar in men and women, although there were more outliers among the men, especially before treatment and at year 1 after treatment.

A7.

Our variable of interest is **SF36pain** which was been measured thrice in the same group of patients; pre-op, one year and five years after treatment. To find out if the treatment made any significant difference to bodily pain (SF36pain) we need to investigate for the differences in bodily pain between the different stages of treatment. We have to test for differences between three scores, pre-op, year 1 post-op and year 5 post-op. If data distribution were normal the appropriate significance test would be the **repeated-measures ANOVA**. If distributional assumptions were not met then the correct option would be to perform the non-parametric **Friedman test**.

As we have learnt before in Chapter 11, for the assumptions made in the repeated-measures ANOVA test to be satisfied it is the residuals and not the actual data that have to be normally distributed. It is only possible to investigate this assumption once the test is performed and the residuals are saved and analysed against predicted values. In fact, SPSS offers the option to plot the residuals. As yet, this option is not available in PSPP. The software continues to be updated and the option may well be available in the near future.

For the sake of learning let us perform a non-parametric test. Because data are related, our test of interest, in this case, is the **Friedman test**.

We select **Analyse>Non-parametric tests>K related samples**. A pop-up window arrives. We transfer the variables onto the test variables box and click to select Friedman test. The output is displayed in Figure A2.31. The Ranks table shows the mean rank for each of the three variables. The test compares the mean ranks between the three groups.

Ranks

	Mean Rank
Pre-op BODILY PAIN - SCORE	1.16
Year 1 BODILY PAIN - SCORE	2.47
Year 5 BODILY PAIN - SCORE	2.37

Test Statistics

N	331
Chi-Square	391.13
df	2
Asymp. Sig.	.000

Figure A2.31 Output of Friedman test.

The test statistics table shows that there was a significant difference between the mean ranks of the related variables. Although the Friedman test informs us that there was an overall difference between the different stages of treatment it does not inform us when the difference was significant. To identify this we need to perform post-hoc tests.

Because we need to perform three sets of post-hoc tests the significance level has to be likewise amended by employing the Bonferroni correction to the standard p-value, $0.05 \div 3 = 0.0167 = 0.017$. The variables are related; we need to perform the Wilcoxon

matched-pair signed rank test in pairs by choosing the respective variables (Figures A2.32, A2.33, A2.34). SPSS offers an easier option.

Pair 1: Pre-op versus year 1

Ranks

		N	Mean Rank	Sum of Ranks
Pre-op BODILY PAIN - SCORE - Year 1 BODILY PAIN - SCORE	Negative Ranks	303	164.58	49868.50
	Positive Ranks	14	38.18	534.50
	Ties	19		
	Total	336		

Test Statistics

	Pre-op BODILY PAIN - SCORE - Year 1 BODILY PAIN - SCORE
Z	-15.11
Asymp. Sig. (2-tailed)	.000

Figure A2.32 Pre-op versus yr 1 bodily pain Wilcoxon matched-pair results output.

Pair 2: Pre-op versus year 5

Ranks

		N	Mean Rank	Sum of Ranks
Pre-op BODILY PAIN - SCORE - Year 5 BODILY PAIN - SCORE	Negative Ranks	323	182.25	58867.00
	Positive Ranks	23	50.61	1164.00
	Ties	25		
	Total	371		

Test Statistics

	Pre-op BODILY PAIN - SCORE - Year 5 BODILY PAIN - SCORE
Z	-15.50
Asymp. Sig. (2-tailed)	.000

Figure A2.33 Pre-op versus yr 5 bodily pain Wilcoxon matched-pair results output.

Pair 3: year 1 versus year 5

Ranks

		N	Mean Rank	Sum of Ranks
Pre-op BODILY PAIN - SCORE - Year 5 BODILY PAIN - SCORE	Negative Ranks	323	182.25	58867.00
	Positive Ranks	23	50.61	1164.00
	Ties	25		
	Total	371		

Test Statistics

	Pre-op BODILY PAIN - SCORE - Year 5 BODILY PAIN - SCORE
Z	-15.50
Asymp. Sig. (2-tailed)	.000

Figure A2.34 Year 1 versus year 5 bodily pain, Wilcoxon matched-pair results output.

The results suggest that patients made a significant improvement in bodily pain one year following treatment. They did not make any significant improvement thereafter. The benefit gained in the first year after treatment was maintained at five years.

A8.

Before we calculate the correlation coefficient, we need to satisfy ourselves that the underlying assumptions have been met to allow us to calculate a valid statistical measure. We have already learnt that for the linear correlation coefficient to be valid data has to be approximately normally distributed. We have seen in the last question that this assumption is true for our data. The variables must be continuous. Although this is not truly interval data it can be considered as continuous data. Further, the relationship should be linear. We need a scatterplot first to check if that is so. This can be displayed by selecting **Graphs>Scatterplot**, select **height** in the *x*-axis and **weight** in the *y*-axis, and click **OK**. The results should look like Figure A2.35.

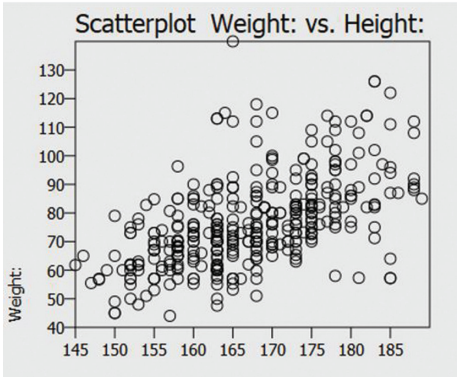


Figure A2.35 Scatterplot of weight/height.

A linear relation is evident although there is no scope in PSPP to insert the line of best fit. It does appear from the charts that there is a reasonable argument for a linear relationship between height and weight. It would be valid to proceed to the calculation of Pearson's correlation coefficient. We choose **Analyse>Bivariate correlations**. A screen will appear (Figure A2.36). We select **weight** and **height**, click for **two-tailed test of significance** and **flag significant correlations**. The results are reproduced in Figure A2.37.

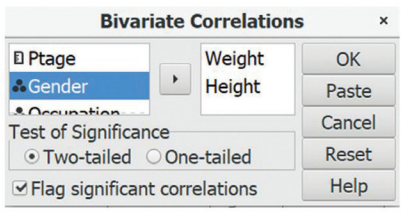


Figure A2.36 Correlation command.

Correlations

		Weight:	Height:
Weight:	*Pearson Correlation*	1.00	.55
	Sig. (2-tailed)		.000
	N	379	366
Height:	*Pearson Correlation*	.55	1.00
	Sig. (2-tailed)	.000	
	N	366	366

Figure A2.37 Pearson Correlation.

The value of Pearson's correlation coefficient is 0.55, and it is highly significant (p-value is < 0.001). The coefficient has a positive value which suggests that there is an increase in weight with an increase in height. We should recall that Pearson's correlation coefficient only measures a linear relationship.

A9.

Our variables in question are continuous. Since there is a single independent variable (**height**) and a single dependent variable (**weight**), a simple linear regression will produce the necessary regression equation.

We go to the data view page, select **Analyse>Regression>Linear.** A new window appears (Figure A2.38). We have to choose weight as the dependent variable and height as the independent variable. PSPP will offer many statistics options. We accept them all.

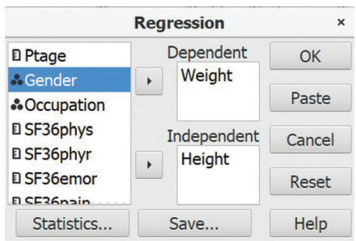


Figure A2.38 Simple Linear regression.

Model Summary (Weight:)

R	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
.55	.30	.30	13.07

ANOVA (Weight:)

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	26535.73	1	26535.73	155.27	.000
Residual	62207.56	364	170.90		
Total	88743.29	365			

Coefficients (Weight:)

	Unstandardized Coefficients		*Standardized Coefficients*		
	B	*Std. Error*	*Beta*	*t*	*Sig.*
(Constant)	-73.99	12.11	.00	-6.11	.000
Height:	.90	.07	.55	12.46	.000

Figure A2.39 Linear regression output.

The output is reproduced in Figure A2.39. The R^2 value is 0.30. The R^2 value indicates that height explains about 30% of the variability of weight in this group of patients. Since we only have a single independent variable, the R^2 value is the square of the correlation coefficient ($R = 0.55$). The middle table is the ANOVA test to investigate for a significant relationship between the variables. The ***F*-statistic** has a significant p-value (< 0.001). The bottom table shows the regression equation in column B. The column contains the *constant* (–73.99) and the *coefficient* (0.90). The coefficient in the column Beta is the **Pearson correlation coefficient** (0.55). From this table, we can write the regression equation:

$$\textbf{Weight} = -73.99 + 0.90 \times \textbf{height}$$

Therefore, someone with a height of 167cm is predicted to have the weight of:

$$-73.99 + (0.90 \times 167) = -73.99 + 150.30 = 76.31\text{kg}$$

Looking back at the database at case number 4, the patient with a height of 167cm had an actual weight of 70kg. Therefore, actual $\mathbf{Y} = 70$ and $\mathbf{Y'} = 76.31$. The residual (difference) $\mathbf{e} = (\mathbf{Y}\text{-}\mathbf{Y'}) = -6.31$.

For the regression equation to be valid, we need to satisfy ourselves that our data meet all the required assumptions. We have already established that there is a linear relationship between the independent and the dependent variables. We need to further satisfy ourselves that for each value of the predictor variable, the residuals are normally distributed and that the variance of y is the same for each value of x. It is mandatory

to check that our data met these assumptions which can only be confirmed once the residual values are known.

To check these assumptions, we need the value of the residuals and the predicted values or fitted values. When we click the dialogue box for regression, we shall see a tab *'Save'* next to *'Statistics'*.

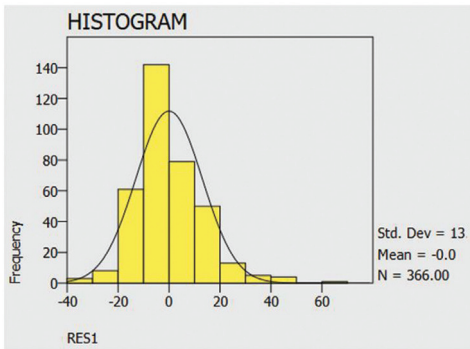


Figure A2.40 Histogram of residuals from weight versus height linear regression.

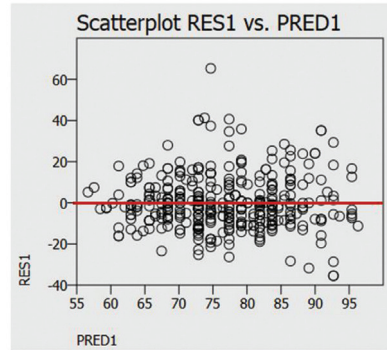


Figure A2.41 Scatterplot of residuals against predicted values.

If we click on the 'Save' tab it opens another window with an option to save both residuals and predicted values. We tick both. Two new variables are created in the dataset: RES1 (residuals) and PRED1 (predicted values or fitted values). We **plot the residuals in a histogram** (Figure A2.40). The distribution appears reasonably close to normal; the mean value is 0. Next, we **plot the residuals against the predicted values** in a scatterplot. If our assumptions are *correct* there will be *no pattern* between the residuals and the fitted values. If a scatterplot is created the residuals should be randomly scattered around 0 for the entire range of the predicted values.

Figure A2.41 demonstrates the scatterplot. A red line has been drawn through 0 for the sake of convenience (this function is not available in PSPP). It shows that the residuals are mostly randomly scattered without any definite pattern around 0. The assumptions were generally true. The regression model is valid.

Please note that another useful measure is to conduct an **inverse normal plot**. It is a scatterplot that compares the data distribution against a normal distribution. The plot is linear if data are normal and curved if they are not. This option is not available in PSPP.

A10.

Before adding **ptage** to the model, one should first establish that there is indeed a linear relationship between age and weight. You should be able to conduct this function now.

The analysis does suggest that there is a linear relationship between age and weight. Pearson's correlation coefficient is −0.23. The negative value indicates that with increasing age, there is a loss of weight.

Model Summary (Weight:)

R	R Square	Adjusted R Square	Std. Error of the Estimate
.56	.31	.31	12.98

ANOVA (Weight:)

	Sum of Squares	df	Mean Square	F	Sig.
Regression	27541.57	2	13770.79	81.77	.000
Residual	60961.95	362	168.40		
Total	88503.53	364			

Coefficients (Weight:)

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	-55.51	13.81	.00	-4.02	.000
Height:	.86	.07	.52	11.63	.000
Age	-.17	.06	-.12	-2.69	.008

Figure A2.42 Output of multiple linear regression.

Let's add age as an independent variable in addition to height and re-run the command. Figure A2.42 shows the results. The p-values for the model as well as for each of the independent variables suggest that they are all statistically significant. The adjusted R^2 value is 0.31. The adjusted R^2 value for our simple linear regression model was 0.30. This would indicate that the addition of age has increased the predictive ability of the model by 1% only.

The new regression equation is:

$$\text{Weight} = -55.51 + 0.86 \times \text{height} - 0.17 \times \text{age}$$

The equation is informing us that if we controlled for height, for people with similar height, there would be 0.17 unit reduction of weight for each unit increase in age. Similarly, for a patient of the same age, there will be 0.86 unit increase in weight for each unit increase in height.

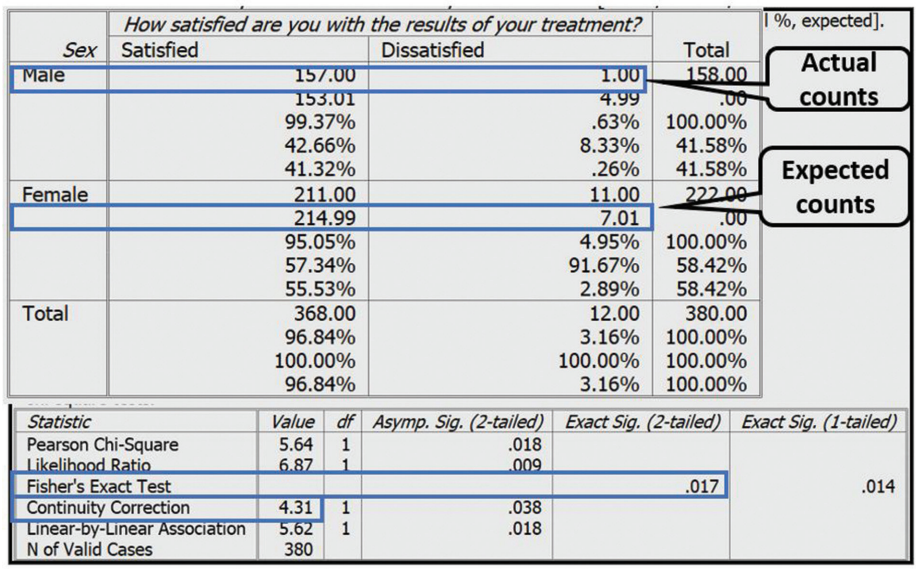
The height of the fourth patient in the database was **167cm**, he was **49 years** old. By applying this equation, we find the predicted weight to be 79.78 kg. His actual weight

was **70 kg**. Therefore, the **residual is −9.78**. Height alone appears to be a better predictor of weight than both height and age combined for this database.

We shall not repeat the exercise of checking assumptions by plotting the residuals in a histogram and the residuals versus fitted values in a scatterplot, but they are both essential for confirmation of model validity.

A11.

This is a test of proportions. We need to perform a Chi-Squared test. The test will compare the observed frequencies against the expected frequencies and calculate the χ2 statistic. We choose **Analyse>Descriptive statistics>Crosstabs**. A new smaller window labelled '**Crosstabs**' will appear. We select **gender** for the rows and **yr5satisfaction** for the columns. Next, we need to click on **Statistics** and select **Chisq**. Next, we click on **Cells** and select *total* and *expected*. The results are displayed in Figure A2.43.



Sex	How satisfied are you with the results of your treatment? Satisfied	Dissatisfied	Total
Male	157.00	1.00	158.00
	153.01	4.99	.00
	99.37%	.63%	100.00%
	42.66%	8.33%	41.58%
	41.32%	.26%	41.58%
Female	211.00	11.00	222.00
	214.99	7.01	.00
	95.05%	4.95%	100.00%
	57.34%	91.67%	58.42%
	55.53%	2.89%	58.42%
Total	368.00	12.00	380.00
	96.84%	3.16%	100.00%
	100.00%	100.00%	100.00%
	96.84%	3.16%	100.00%

Statistic	Value	df	Asymp. Sig. (2-tailed)	Exact Sig. (2-tailed)	Exact Sig. (1-tailed)
Pearson Chi-Square	5.64	1	.018		
Likelihood Ratio	6.87	1	.009		
Fisher's Exact Test				.017	.014
Continuity Correction	4.31	1	.038		
Linear-by-Linear Association	5.62	1	.018		
N of Valid Cases	380				

Figure A2.43 Output of the chi-squared test.

The first row of the top table shows the observed frequencies and the second row the expected frequencies. Among the men, 157 of them were satisfied and a single patient was dissatisfied. The expected counts were 153.01 and 4.99 respectively. Among the women, 211 of them were satisfied and 11 dissatisfied. The expected counts were 214.99 and 7.01 respectively. The lower table shows the statistics.

The Pearson Chi-Squared value is 5.64, the degrees of freedom is 1, the difference in the rate of satisfaction between men and women proved to be significant ($p = 0.017$). Please note that **Fisher's exact test** was performed. We have already learnt that Fisher's exact test is recommended if any of the expected cell counts is < 5; it was the correct test to perform for our data.

We also learnt previously that **Yates' continuity correction** is recommended for a 2x2 contingency table even if sample size requirements are met. We note in Figure A2.43 that Yates' correction yielded a smaller value for the χ2 statistic. The results should be reported as:

There was a significant relation between sex and patient satisfaction (Fisher's Exact test, Pearson Chi-Squared statistic = 5.64 (Yates' continuity correction = 4.31, df 1, p = 0.017).

A12.

Indication of patients' pre-treatment mental state is available in the pre-treatment mental health score (**SF36mhth**). However, we have a problem on our hands, as, currently this score is a scalar one. We need to convert this score to a dichotomous type for the ease of calculation. Mental health score ranges from 0–100. A cut-off value of 50 has been recommended by authors to categorise mental distress. Those who score 0–50 can be categorised as suffering from mental distress and those who score >50 as not being distressed. To categorise the variable, we need to transform it. We choose **Transform>Recode into different variable**. A new smaller window will appear (Figure A2.44).

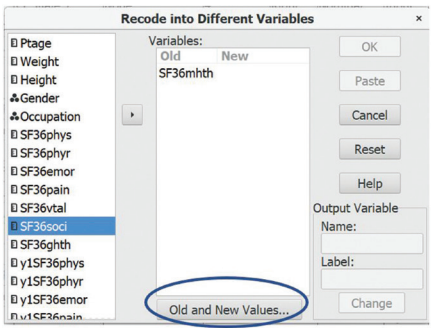


Figure A2.44 Pop-up window for recoding a variable.

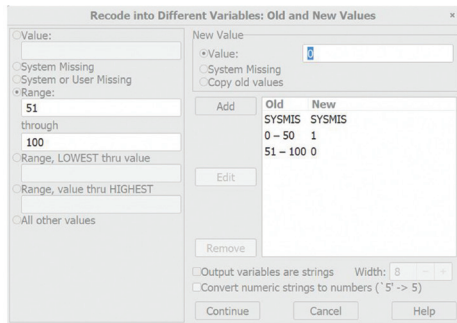


Figure A2.45 Assigning values.

We scroll down on the left-hand side to find **SF36mhth** and transfer this variable onto the *old variables* box. Next, we need to choose the *Old and New Values…* tab at the bottom of the pop-up screen. This will take us to a new screen (Figure A2.45).

We need to assign values for the new variable. Since we are dealing with a range of values, we click on *range* and assign *0–50* as the *old range*, the new value for this range is 1 (**distressed**). We click *add*, next we click on *range* again and assign *51–100* as the *old range* and click *add*, the new value for this range is 0 (**not distressed**). We also need to instruct PSPP how to deal with **missing values**, if there were any. We click on the *System Missing* button on the left and when the **New Value** label is activated, we choose *System Missing*. PSPP will identify missing values and mark them with a dot (**SYSMIS**). We click on **Continue**. This will bring us back to the previous pop-up screen. If we click on **SF36mhth**, a new box named Output variable will become activated. We have to give a name (**Distress**) and label (**Mental health**) to the new variable. Next, we click *Change*. The new variable (**Distress**) will appear in the dataview screen (Figure A2.46).

If we click *OK,* the new variable distress will appear in dataview. We need to go back to the variable view screen to label this new variable correctly. We choose *numeric* as the type of variable, label the values in the value label column (**0 = not distressed, 1= distressed**) and choose *nominal* as the measure. We can have an overview of the new variable if we choose **Analyse>Descriptive Statistics>Freqencies.**

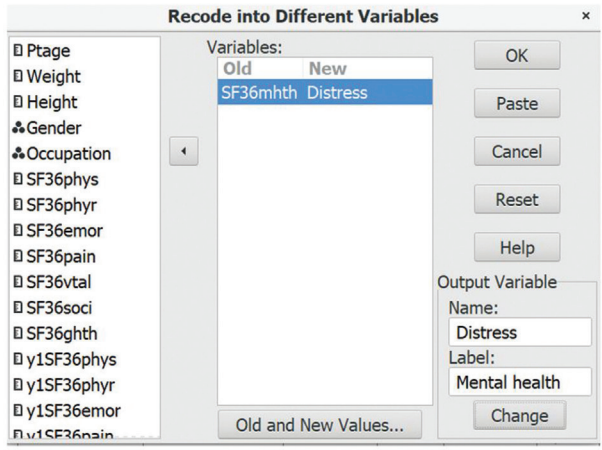


Figure A2.46 Old and new variables.

Mental health

Value Label	*Value*	*Frequency*	*Percent*	*Valid Percent*	*Cum Percent*
Not-distressed	0	320	84.21	85.79	85.79
Distressed	1	53	13.95	14.21	100.00
	.	7	1.84	Missing	
	Total	380	100.0	100.0	

Mental health

N	*Valid*	373
	Missing	7
Minimum		.00
Maximum		1.00
Sum		53.00

Figure A2.47 Frequency table of distress.

The table confirms that out of a total of 380 cases, there were 7 cases of missing data, 320 patients were not distressed and 53 patients were distressed. If we repeat the Chi-Squared test by choosing the *Crosstabs* option with **Distress** as the row variable and **y5satisfac** as the column variable, the results indicate that:

pre-treatment mental health did not significantly influence post-treatment patient satisfaction (Fisher's exact test, Chi-Squared statistic = 3.72 (Yates' continuity correction = 2.28, p = 0.117).

Please note again that due to low expected counts in one cell Fisher's exact test was performed and Yates' continuity correction calculated since this was a 2 × 2 table.

A13.

Our *independent* variables of interest are **Ptage**, **Occupation** and **SF36mhth**, and the *dependent* variable is the **y5satisfac**. The dependent variable, y5satisfac, is a binary one (category 1 = satisfied, category 2 = dissatisfied). We need to perform a **binary logistic regression** analysis to answer our question.

However, before we conduct our analysis, we need to satisfy ourselves that our dataset meets the assumptions of the test:

1. Our dependent variable is dichotomous, the independent variables are a mix of continuous and categorical variables.

2. We also know that observations are independent and not related.

3. Our sample size is large and we have only three independent variables, therefore sample size assumptions are met.

4. Next, we need to satisfy ourselves that there is no multicollinearity, i.e. the independent variables are not *highly* correlated to each other.

We have two continuous variables, Ptage and SF36mhth and a categorical variable. One way to do this is to calculate the **Variance Inflation factor** (VIF). Fortunately, this option is available in the latest version of PSPP (version 1.4.1). We need to check for multicollinearity between the three independent variables: Ptage, Occupation and SF36mhth. We go back to the linear regression screen, we can choose SF36mhth as our dependent variable and the other two as the independent variables, it does not matter. Next, we click on *statistics*; several options are made available in a separate screen. We choose the **Tol** option, click **OK**. This will give us the *Collinearity statistics* in the form of **Tolerance** and **VIF**, the results are shown in Figure A2.48.

Coefficients (Pre-op MENTAL HEALTH - SCORE)

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	66.77	6.23	.00	10.71	.000		
Current Occupation:	-.21	.91	-.01	-.23	.815	.80	1.26
Age	.08	.11	.04	.72	.470	.80	1.26

Figure A2.48 Correlation between age and SF36mhth.

Tolerance and VIF are both collinearity statistics. Tolerance is a measure of collinearity of the variables. A low tolerance value indicates a perfect linear relation, a low tolerance value of < 0.1 should be investigated further.

VIF is a measure of the effect of collinearity in the model. There is no formal cut-off level but values >2.5 may raise concern. The results would suggest that there is no evidence to suggest age, occupation and pre-treatment mental health are *highly* correlated.

5. Finally, we should also confirm that there is a linear relationship between the logit function of the dependent variable and the independent continuous variable. This can be investigated with the **Box-Tidwell test**. Currently, this option is not available in PSPP.

Let's proceed to **logistic regression**. We choose **Analyse>Regression>Binary Logistic**. We click on **y5satisfac** and transfer it to the *Dependent* variable box. Similarly, we transfer **Ptage**, **Occupation** and **SF36mhth** onto the *Independent* variable box. We click **OK** and the results are displayed. Several tables are produced. Out of 380 cases, there were 9 missing. The Model summary table is displayed in Figure A2.49.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	96.46	.03	.10

Figure A2.49 Model summary of logistic regression.

The **Nagelkerke R square** is a pseudo R square value and indicates that the model has poor ability to explain patient satisfaction five years after treatment (10%). The model explains only 10% of the variability in yr5satisfac. However, without comparing other models one cannot discard it. It may well be that this is the best available model. The Classification table (Figure A2.50) indicates the accuracy of a model at the beginning of model fitting without introducing any predictor variables. Since most patients were satisfied the best strategy for the software at the beginning, without any information of predictor variables being made available, was to treat everyone as satisfied. This led to a sensitivity of 100% and a specificity of 0%, with an overall accuracy of 96.8%. We can therefore observe that even without any predictor variables the accuracy of the model was quite high. It was always going to be difficult to match that!

Classification Table

Observed			Predicted		
			How satisfied are you with the results of your treatment?		
			Satisfied	Dissatisfied	Percentage Correct
Step 1	How satisfied are you with the results of your treatment?	Satisfied	359	0	100.0%
		Dissatisfied	12	0	.0%
	Overall Percentage				96.8%

Figure A2.50 Classification table without any predictive variables.

Figure A2.51 shows the final model with all the included variables. Also note that since we had a categorical variable (occupation) with five different categories, the software automatically created indicator variables for analysis and compared them against a baseline occupation exposure. We did not have to instruct this separately. To enable this function, you have to ensure that value labels are correctly identified and the variable is correctly indicated as ordinal in the measure option.

There does not seem to be an option to perform the **Hosmer and Lemeshow test** in PSPP. This function is available in SPSS. We have already learnt that this test is performed

to assess the goodness of fit of the model. The column **Exp(B)** has been marked with a red border. Exp(B) stands for the **exponential of the estimated regression coefficient**. It produces the OR that adjusts for the influence of the other predictor variables in the model. The column immediately to the left shows the significance of the Wald statistic. None of the independent variables has any significant ability to predict y5satisfac (p-value>0.05). The latest version of PSPP also displays how the dummy variables were coded. This is displayed in Figure A2.52.

		B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
								Lower	Upper
Step 1	Age	-.03	.03	1.35	1	.246	.97	.92	1.02
	Pre-op MENTAL HEALTH - SCORE	-.02	.01	2.91	1	.088	.98	.95	1.00
	Current Occupation:			3.28	4	.512			
	Current Occupation:(1)	-19.97	14890.66	.00	1	.999	.00	.00	+Infinit
	Current Occupation:(2)	-.18	1.16	.02	1	.878	.84	.09	8.12
	Current Occupation:(3)	.36	.89	.17	1	.684	1.44	.25	8.25
	Current Occupation:(4)	1.31	.76	2.96	1	.085	3.70	.83	16.47
	Constant	-.15	1.86	.01	1	.934	.86		

Figure A2.51 Variables in logistic regression and their significance.

		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
Current Occupation:	Heavy Manual	22	1	0	0	0
	Light Manual	29	0	1	0	0
	Office / Professional	46	0	0	1	0
	Housewife	30	0	0	0	1
	Unemployed / Retired	244	0	0	0	0

Figure A2.52 Categorical variables' codings.

We can conclude that age, occupation or pre-operative mental health would not be suitable for predicting post-treatment patient satisfaction.

A14.

Researchers wished to investigate the efficacy of a plant-based diet in reducing serum cholesterol level. The primary outcome measure was a reduction in serum cholesterol level. As we have learnt before this is a surrogate outcome measure. It may have been clinically more relevant to investigate if plant-based diets helped to reduce cardiovascular mortality and/or morbidity etc.

Serum cholesterol is a continuous variable. Therefore, depending on the distribution of data (most likely parametric) the appropriate statistical test would be a test for difference in means. If the data are related (before and after plant-based diet) a related samples *t*-test would be appropriate. If data are independent (plant-based diet group versus non-plant-based diet group), an independent samples *t*-test may be appropriate.

A15.

a. The null hypothesis of the trial would be:

the utterance of the word 'quiet' would adversely affect the workload beyond the non-inferiority margin of 30 clinical episodes (i.e. the workload will be inferior, >30 episodes, one-sided test).

The alternative hypothesis would be:

the utterance of the word 'quiet' would not adversely affect the workload (i.e. the workload will not be inferior beyond the 30 episodes margin, could be similar or better, one-sided test).

b. In this case, we do not require the p-value to reject the null hypothesis. The 95% CI of the difference in clinical episodes is a marker of the evidence for/against it. If the upper bound of this 95% CI is less than the margin of non-inferiority of 30 clinical episodes, then we can consider that non-inferiority was shown. The upper margin was 24.7. There is strong evidence to reject the null hypothesis. We accept the alternative hypothesis that uttering the word 'quiet' did not appear to harm the clinical workload in this study! *(You may wish to re-visit Figure 9.7 for a visual interpretation of the margins of non-inferiority in a RCT*).

A16.

a. The null hypothesis would be: there is no difference in the proportion of participants progressing to severe disease or all-cause mortality 28 days post-enrolment between the intervention arm and control arm.

b. Since this was an open-label trial both the caregivers and the participants were aware of the allocated intervention. This may have affected the level of care or patients' expectations. However, as far the outcomes were concerned these were very objective and therefore unlikely to be affected by detection bias as such.

c. All participants originally allocated would qualify for intention-to-treat (ITT) analysis. If we are to perform ITT, all 235 participants originally allocated to the intervention arm and 229 participants allocated to the control arm should be analysed in ITT irrespective of their received intervention or follow-up status. Since patients were lost to follow-up we need to have a strategy to address missing values. Multiple imputations are currently the most accepted method for the analysis of missing data [7]. When multiple imputations are performed missing data are imputed according to a predictive regression model. Some authors employ a 'modifed ITT' where all patients are followed-up regardless of the treatment they receive but only those with complete follow-up are analysed [8].

Per-protocol analysis will only include participants who complied with the original study protocol. Of the 228 participants in the plasma arm, one was lost to follow-up and two received only a single dose, therefore 225 participants would qualify for per-protocol analysis (226, if the participant lost to follow-up had received the full intervention). In the control arm, 226 complied with the control arm but one was lost to follow-up later, therefore, 225 participants would qualify for per-protocol analysis.

d. An appropriate significance test would be the chi-squared test for association/independence. The null hypothesis would be: the outcomes (28 days all-cause mortality or progression to severe disease) are independent of the type of intervention received.

We do not have individual patient data but we do have the numbers for each group that we can put in any freely available online chi-squared calculators and complete our analysis (Table A2.2) [9]. The $\chi 2$ statistic of 0.0521 is less than the critical value for the $\chi 2$ statistic.

	Dead or severe	Alive or well	*Marginal Row Totals*
Intervention	44 (43.05) [0.02]	191 (191.95) [0]	235
Control	41 (41.95) [0.02]	188 (187.05) [0]	229
Marginal Column Totals	85	379	464 (Grand Total)

Table A2.1 Results of the chi-squared test.

P = 0.819; we can not reject the null hypothesis. The variables are independent, there was no relation between the type of treatment received and 28 days all-cause mortality or progression to severe disease.

e. Since this was a prospective trial, we know the exact proportion of events in the exposed versus the non-exposed group. We can use risk as our outcome measure.

f. Although RCTs adjust for confounding factors, calculation of risk ratio will not be able to account for the effect of other exposure factors, the calculated risk will be unadjusted risk ratio. A better option may be to calculate the OR to account for possible confounders.

g. We need to find out the data distribution to decide which is the appropriate statistical test to perform. We note that the authors displayed non-parametric measures. This would imply that data were not normally distributed. Since the two groups were not related the appropriate statistical test would be the Mann-Whitney U test.

A17.

a. The null hypothesis would be:

the difference in satisfaction with pain relief between the two groups would be greater than 10% (could be in either direction, two-sided test) (i.e. the two groups are not equivalent concerning patient satisfaction, could be better or worse, two-sided test).

The alternative hypothesis would be:

the difference in satisfaction with pain relief, between the two groups, would be less than 10% (the two groups are equivalent concerning patient satisfaction).

b. Intention-to-treat is the gold standard for the analysis of results. However, in this situation analgesia was not universally used and there was substantial cross-over from one arm. ITT may give a pragmatic estimate of the intention of using the intervention but is likely to favour the alternative hypothesis in this scenario. Therefore, as-treated analysis may be a better option.

c. Results can be tested for equivalence by determining whether the upper and lower limits of the 95% confidence interval of the primary endpoint AUC of satisfaction with pain exceeded the equivalence margin of 10% or not.

Since the as-treated analysis showed that the mean difference and the upper limit of the 95% CI exceeded the 10% margin, the two interventions were not equivalent for patient satisfaction; we cannot reject the null hypothesis and we are unable to accept the alternative hypothesis.

A18.

It may be convenient if we created a 2 × 2 contingency table with the results (Table A2.3).

	No infection	Infection	Total
Antibiotics	1439	180	1619
Placebo	1300	306	1606

Table A2.3 Contigency table of infection rate in antibiotics versus the placebo group.

a. Therefore, the **risk of infection** in the prophylactic antibiotic group was: 180 ÷ 1619 = 0.11.

The **risk of infection** in the placebo group was: 306 ÷ 1606 = 0.19

The **risk ratio** of infection in the prophylactic antibiotic group was 0.11 ÷ 0.19 = 0.578

The prophylactic antibiotic group was nearly half as likely (relative risk reduction 42%) to develop infection compared to the placebo group.

b. The absolute risk reduction was 0.19–0.11 = 0.08.

Numbers needed to treat (NNT) was 1 ÷ 0.08 = 12.5.

Therefore, for every 13 patients treated with prophylactic antibiotics compared to placebo after operative vaginal birth, a single case of maternal infection was averted.

c. The odds of infection in the antibiotic allocated group was: 180 ÷ 1439 = 0.125

The odds of infection in the placebo allocated groups was: 306 ÷ 1300 = 0.235

The **odds ratios** of infection in the prophylactic antibiotic group was 0.125 ÷ 0.235 = 0.53

As we have discussed before, when events are rare, risk and odds ratios are nearly identical.

A19.

a. This is a survival curve (Figure Q19.1). Maximum follow-up was 10 years. It is not clear what were the numbers at risk at each stage nor the number of censored patients. The curve indicates that there was an apparent difference in survival according to the different categories of hypertension. All-cause mortality appears highest in the resistant hypertensive group (yellow) and lowest in the normotensive (green) group. The 10-year survival in the resistant hypertensive group was around 65% and in the normotensive group around 85%.

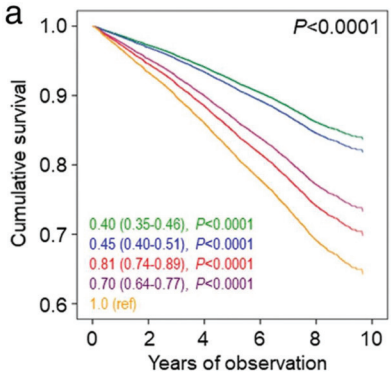


Figure Q19.1 Unadjusted cumulative survival figures for different categories of hyptertension.

Green – normotensive
Blue – untreated hypertensive
Red – controlled hypertensive
Purple – uncontrolled hypertensive
Yellow – resistant hypertensive

b. The hazard of interest was all-cause mortality.

c. Median survival cannot be commented on as survival rate was >50% at 10 years. We cannot predict median survival as it is not possible to extrapolate data beyond the maximum follow-up; we do not know what will happen to the survival curve after 10 years.

d. One would perform a Log-rank test to investigate if the apparent differences in survival rates were significantly different.

e. If the mortality rate in different categories was marked in frequencies, we could investigate this with a chi-squared test for trend.

f. If we perform a Cox regression analysis, we should be able to adjust for the effect of the other risk factors. The authors identified several risk factors for cardiovascular death, as well as co-morbidities and complications and undertook Cox regression. Figure A2.53 displays the results following Cox regression analysis.

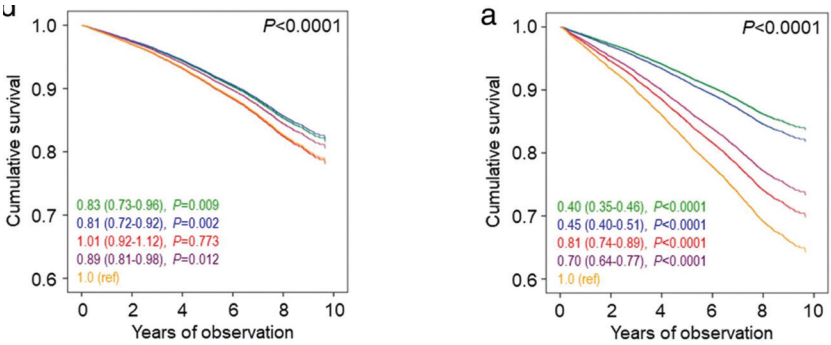


Figure A2.53 Adjusted HR after Cox regression (left), compare with figure Q19.1 unadjusted HR (right).

Compared to the unadjusted hazard ratio in Figure Q19.1, the adjusted HR in Figure A2.53 demonstrates that compared to resistant hypertensive (yellow) there was no difference in survival rate in the controlled hypertensive (red) group (look at the margin of the 95% CI, crossed 1). Normotensives (green), untreated hypertensives (blue) and uncontrolled hypertensives (purple) demonstrated improved survival rate when adjusted for other risk factors.

The margin of difference was much less in the adjusted model compared to the unadjusted one. The 10-year survival probability for the resistant hypertensive group is now just under 80% and that of the normotensive group 83%. This would indicate that although resistant hypertension proved to be an independent predictor of all-cause mortality the relationship was not straightforward and was not solely dependent on control of hypertension in this group of patients.

A20.

Since the incidence of the virus is 10%, for every 100 persons, 10 persons have the virus. The results are broken down in Figure A2.54.

	Test negative		Test positive	
😊	81	True – ve	False + ve	9 😞
😊	3	False – ve	True + ve	7 😞

Figure A2.54 Interpreting the Covid 19 test results, Sensitivity = 70% and specificity = 90% (prevalence = 10%)

a. Therefore, the probability of having contracted the virus when tested positive is:

PPV = TP ÷ TP + FP = 7 ÷ 16 = 43%

b. The probability of not having contracted the virus if tested negative is:

NPV = TN ÷ FN + TN = 81 ÷ 84 = 0.964 = 96.4%

The probability of having contracted the disease if tested negative is very low, so we can use the test to rule out the disease. We cannot use it to rule in the disease since the probability of having contracted the disease even if tested positive is not high.

A21.

To answer this question, we need to ascertain the positive and negative likelihood ratios (LR) of superficial wound infection.

$$\text{LR + ve} = \frac{\text{Sensitivity}}{1 - \text{Specificity (FPR)}}$$

$$\text{and LR} - \text{ve} = \frac{1 - \text{Sensitivity (FNR)}}{\text{Specificity}}$$

We know that sensitivity = 49%, specificity = 62%

Therefore, FPR = 1-Specificity = 38%,

FNR = 1-sensitivity = 51%,

Therefore,

$$\text{LR + ve} = \frac{49}{38} = 1.29, \qquad \text{LR} - \text{ve} = \frac{51}{62} = 0.82$$

The results suggest that a positive test was only marginally more (1.29) likely in someone with infection than someone without. The negative LR was quite high and suggests that the test was not accurate at obtaining a negative result in those without the disease compared to those with the disease. Henceforth, we conclude that the results of this test are unlikely to affect the post-test probability.

A22.

a. We can use the Bayes' theorem to answer this question. We have the following pieces of information:

The prevalence of prostate cancer is 0.8%.

P(Prostate Cancer) = 0.008,

P(no Prostate Cancer) = 0.992

The sensitivity of the PSA test is 90%.

P(test + ve given Prostate Cancer + ve) = 0.90.

The specificity of the PSA test is 85%.

The false-positive rate of the PSA test is: (1-Specificity) = 15%.

P (test + ve if Prostate Cancer – ve) = 0.15

P(test + ve) = P(test + ve if Prostate Cancer + ve) × P(Prostate Cancer) + P(test + ve if Prostate Cancer – ve) × P(no Prostate Cancer)

P (test + ve) = 0.90 × 0.008 + 0.15 × 0.992 = 0.0072 + 0.1488 = 0.156

$$\text{P (Prostate cancer + ve if test + ve} = \frac{\text{P (test + ve if Prostate Cancer + ve)} \times \text{P (Prostate Cancer)}}{\text{P (test + ve)}}$$

$$\text{P (Prostate Cancer + ve given test + ve)} = \frac{0.90 \times 0.008}{0.156} = \frac{0.0072}{0.156} = 0.046 = 4.6\%$$

If you found it hard to get your head around the concept of Bayes' theorem, we can also descriptively approach the problem (Figure A2.55). The calculations are slightly different as these are in full numbers.

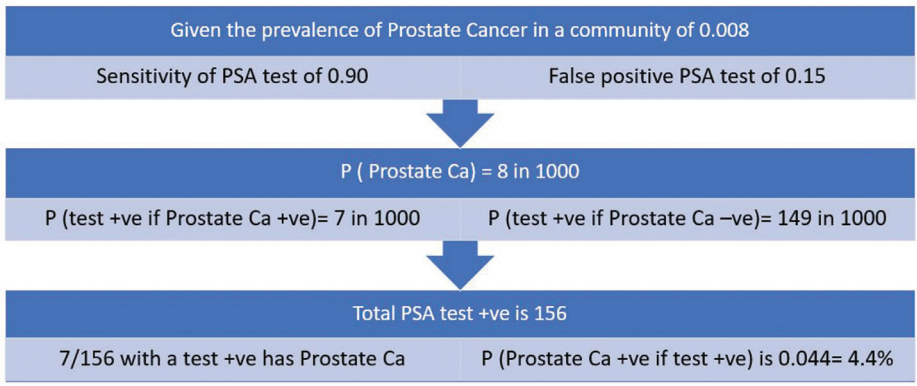


Figure A2.55 Probability of prostate cancer being present if PSA test is positive.

b. The pre-test odds of prostate cancer in a patient was:

$$0.008 \div 0.992 = 0.008$$

The post-test odds of prostate cancer in a patient with a positive test was:

$$0.046 \div 0.954 = 0.048$$

Post-test odds of prostate cancer was six times more likely, but still low.

A23.

a. Before looking at the numbers we need to assess the study population, intervention, co-intervention etc. to assess for clinical heterogeneity. Provided there was sufficient evidence for the presence of clinical heterogeneity, the Chi-Squared statistic and I^2 value would indicate evidence of statistical heterogeneity. The Chi-Squared value was small and for 8 *df,* the p-value was 0.87. Similarly, the I^2 statistic is low. They both suggest a low probability of statistical heterogeneity.

b. The funnel plot is symmetrical with studies scattered on both sides of the pooled treatment effect. This would indicate little evidence of publication bias.

c. Nine studies were included in the meta-analysis, these are displayed in rows. The second and the fourth column show the number of events in the control group and the topical TXA group respectively. The third and the fifth columns show the total participants in the control group and the topical TXA group. The sixth column shows the weight ascribed to each study. Maximum weight was ascribed to **Seo et al**.

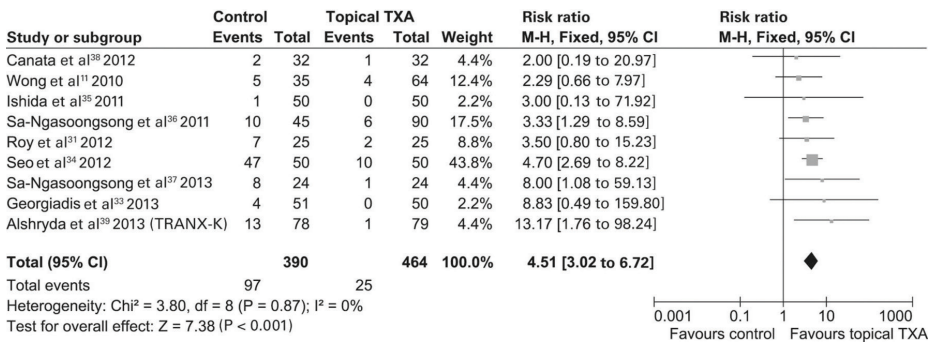


Figure Q23.2 The Forest plot.

Reproduced with permission and © of The British Editorial Society of Bone & Joint Surgery [6].

Their study was not the largest but had the maximum events. Since there was no evidence of heterogeneity the authors performed a fixed-effects model of meta-analysis. The Forest plot in the extreme right shows the point estimates (square block) as well as the 95% CI (the whiskers). The diamond is the summary estimate and favours the topical TXA arm. The ends of the diamond are well away from the line of no effect. We can conclude that the application of topical TXA in total knee arthroplasty led to a significant reduction in blood transfusion.

References

1. Brookfield CR, Phillips PPJ, Shorten RJ. Q fever – the superstition of avoiding the word 'quiet' as a coping mechanism: randomised controlled non-inferiority trial. *BMJ* 2019;**367**:l6446. doi: 10.1136/bmj.l6446.

2. Agarwal A, Mukherjee A, Kumar G et al; PLACID Trial Collaborators. Convalescent plasma in the management of moderate covid-19 in adults in India: open label phase II multicentre randomised controlled trial (PLACID Trial). *BMJ* 2020; **371** :m3939. doi: 10.1136/bmj.m3939.

3. Freeman LM, Bloemenkamp KW, Franssen MT et al. Patient-controlled analgesia with remifentanil versus epidural analgesia in labour: randomised multicentre equivalence trial. *BMJ* 2015; **350**:h846. doi: 10.1136/bmj.h846.

4. Knight M, Chiocchia V, Partlett C et al; ANODE collaborative group. Prophylactic antibiotics in the prevention of infection after operative vaginal delivery (ANODE): a multicentre randomised controlled trial. *Lancet* 2019;**393**(10189):2395–2403. doi: 10.1016/S0140-6736(19)30773-1.

5. Solini A, Penno G, Orsi E et al. Is resistant hypertension an independent predictor of all-cause mortality in individuals with type 2 diabetes? A prospective cohort study. *BMC Med* 2019;**17**:83. doi:10.1186/s12916-019-1313-x.

6. Alshryda S, Sukeik M, Sarda P et al. A systematic review and meta-analysis of the topical administration of tranexamic acid in total hip and knee replacement. *Bone Joint J* 2014;**96-B**(8):1005–1015. doi:10.1302/0301-620X.96B8.33745.

7. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol* 2017;**17**(1):162. doi:10.1186/s12874-017-0442-1.

8. Del Re AC, Maisel NC, Blodgett JC et al. Intention-to-treat analyses and missing data approaches in pharmacotherapy trials for alcohol use disorders. *BMJ Open* 2013;**3**:e003464. doi: 10.1136/bmjopen-2013-003464.

9. https://www.socscistatistics.com/tests/ [last accessed 8 February 2021].