Statistics for Business Students

Student Revision Tips

General Tips

a. Be patient and work slowly and methodically, especially in the early stages when progress may be slow.
b. Do not omit or 'jump around' between chapters; each chapter builds upon knowledge and skills previously gained. You may also find that the Excel applications that you develop require earlier ones in order to work.
c. Try not to compare your progress with others too much. Fastest is not always best!
d. Don't try to achieve too much in one session. Time for rest and reflection is important.
e. Mistakes are part of learning. Do not worry about them. The more you repeat something, the fewer mistakes you will make.
f. Make time to complete the exercises, especially if you are learning on your own. They are your best guide to your progress.
g. The visual walkthrough flash movies have been developed to explore using Excel to solve a particular statistical problem. If you are not sure about the Excel solution then use the visual walkthrough flash movies as a reminder.

Online workbook: Introduction to Excel

The aim of this online workbook is to provide students with an overview of the Excel spreadsheet. The text screenshots and visual walkthrough flash movies use Excel. Like all software packages the best way to learn is by undertaking specific tasks. In this text book we have provided the basic tasks we would expect students to be able to undertake to be able to follow the statistical techniques and their Excel solutions described in later chapters. So read the workbook, undertake the tasks, and access specific Excel resources to broaden your knowledge of Microsoft Excel and its application to solve business related problems.

Online workbook: Numerical skills revision

This online workbook provides the reader with an opportunity to study key numerical skills concepts that we have identified as important to the student to enable them to study later chapters. Again, practice the examples provided and their Excel solutions to familiarise yourself with key mathematical and statistical functions. Furthermore, the workbook contains a refresher on financial mathematics and co-ordinate geometry. Remember that one important aspect of successfully completing a statistics course is practice and then practice again.

Chapter 1: Visualising and presenting data

Chapter 1 is the first main chapter of the book that provides a description of the key methods to visualise and present data. The key revision topics include:

- Why do we tabulate raw data?

- How we can construct tables?
- What are the different types of data?
- Do we have different types of table for different data types?
- How would we convert these tables into graphical form?
- Why are these graphs important?
- How do we create a bar chart and pie chart to represent categorical data? This should include calculating values by hand and using graph paper to construct the chart.
- How do we create a bar chart and pie chart using Excel?
- How do we create a histogram to represent numerical data? This should include calculating values by hand and using graph paper to construct the chart.
- How do we use Excel to generate a histogram?
- What other graphs can Excel create? Practice generating a scatter plot and time series plot. These two plots will be used in later chapters.

Chapter 2: Data descriptors

Chapter 2 is focused on describing techniques that are used to describe data. The graphical methods provide a visual representation but in this chapter we will put number values to the concept of the middle value or how spread out the data is about this middle number. The key revision topics include:

- What do we mean by a middle number/average/central tendency?
- What do we mean by data being dispersed (or spread) about a middle number?
- What are the different types of central tendency (mean, mode, median) and how would you calculate these values for a set of numbers?
- What are the key characteristics of these different measures of central tendency?
- What are the different types of dispersion (or spread) (range, quartile range, variance, standard deviation, quartile range, semi interquartile range) and how would you calculate these values for a set of numbers?
- What are the key characteristics of these different measures of dispersion?
- How would you use Excel to calculate these measures of central tendency and dispersion?
- How do we compare two sets of numbers using the coefficient of variation for a set of numbers?
- How would you compute the coefficient of variation using Excel?
- What do we mean by shape (skewness and kurtosis)?
- How would you use Excel to calculate these measures of shape?
- What do we mean by symmetry and peakedness?
- What methods do we have to put a number to this concept of symmetry and peakedness?
- How would you use Excel to calculate these measures of skewness and peakedness for a set of numbers?
- How would you undertake a five number summary calculation to assess symmetry and the form of the non-symmetry if it exists?

Chapter 3: Probability distributions

The main aim of the chapter is to focus on introducing the student to the concept of a probability distribution. The key revision topics include:

- What are the basic ideas behind the concept of probability?
- What do we mean by the concept of relative frequency and sample probability?
- What are the key probability laws applied to the study of statistics as outlined in this text?
- What do we mean by a relative frequency distribution and its relationship to a probability distribution?
- How do we construct a probability distribution using Excel?
- How do we calculate the expected value and variance for a probability distribution?
- What do we mean by a continuous variable?
- What do we mean by a normal probability distribution?
- What are the characteristics of a normal probability distribution?
- How do we use Excel to calculate probabilities for the normal distribution?
- What do we mean by the standard normal distribution?
- How do we use Excel to calculate probabilities for the standard normal distribution?
- How would I use tables to calculate probabilities for the normal and standard normal distributions? The table method is not covered in the text but you tutor should cover this topic in class (if required).
- How do we use Excel to assess if a distribution is normally distributed (symmetry) using a normal probability plot?
- What other types of continuous variable will we explore in the text (t, F, chi-square)?
- What do we mean by a discrete variable?
- What is a binomial probability distribution?
- What are the characteristics of a binomial probability distribution variable?
- How would we use Excel to calculate binomial probability problems?
- How would I solve binomial probabilities using the binomial distribution probability formula?
- What is a Poisson probability distribution?
- What are the characteristics of a Poisson probability distribution variable?
- How would we use Excel to calculate Poisson probability problems?
- How would I solve Poisson probabilities using the binomial distribution probability formula?
- How would I confirm using Excel that particular distributions can be approximated by other probability distributions?

Chapter 4: Sampling distributions

In the earlier chapters we have been concerned with calculating probabilities based on a specific data value. In this and later chapters we will explore methods that are used to solve problems involving assessing the probability of a specific average occurring. The key revision topics include:

- Why sample?
- What are the key types of sampling that can be performed?

- What type of sampling is performed within the analysis of data using statistical tests?
- What are the different types of error associated with sampling?
- What do we mean by sampling from a population?
- What do we mean by a sample being representative of a population?
- What do we mean by unbiased and biased estimates?
- How would you confirm using Excel that the sample mean is an unbiased estimator of the population mean (assume all possible sample points of size 2 are sampled)?
- How would you confirm using Excel that the sample mean is an unbiased estimator of the population mean (assume all possible sample points of size 2 are sampled)?
- How would you confirm using Excel the relationship between the standard error of the estimate and the population standard deviation and sample size?
- How would you use Excel to undertake sampling of size n from a normal distribution to confirm the effect on the distribution shape as sample size increases?
- What do we mean by the central limit theorem?
- How would you use Excel to calculate the probability that a sample average (mean, proportion) as a particular value (NORM.DIST, NORM.S.DIST)?
- What do we mean by sampling from a finite population?
- What is the sampling from a finite population correction factor?
- How would you use Excel to sample from a non-normal distribution (Data Analysis ToolPak)?

Chapter 5: Point and interval estimates

Chapter 5 provides the student with an introduction to elementary inferential statistics. The key revision topics include:

- What do we mean by a point estimate of the population mean and proportion?
- What do we mean by an interval estimate of the population mean and proportion?
- How would we calculate a point estimate of the population mean (or proportion, variance, standard deviation) using the sample?
- How would we calculate an interval estimate using the Z and T distributions?
- How would we use Excel to calculate the required point and interval estimates?
- How would we calculate an appropriate sample size for a particular confidence interval (assuming normal distribution)?

Chapter 6: An introduction to hypothesis testing

Chapter 6 provides the student with an introduction to parametric hypothesis testing. The key revision topics include:

- What do we mean by a hypothesis statement?
- What do we mean by parametric and non-parametric data variables?
- What do we mean by a significance level?
- What do we mean by one and two tail tests?
- What are the key types of error associated with hypothesis testing?
- What do we mean by a p-value and critical test statistic?

Students will need to make sure that they practice reproducing the solutions by hand with tables of critical statistics. The text uses Excel to provide the critical test statistics (p-value, critical test value) and includes the appropriate set of critical tables in the appendices.

Chapter 7: Parametric hypothesis testing

Chapter 7 provides the student with an introduction to parametric hypothesis testing. The key revision topics include:

- How would we use Excel to apply a one sample Z test to a data sample (mean, proportion)?
- How would we use Excel to apply a one sample T test to a data sample (mean, proportion)?
- How would we use Excel to apply a two sample Z test for a population mean and proportion?
- How would we use Excel to apply a two sample Z test for a population mean assuming independent samples with equal variances?
- How would we use Excel to apply a two sample Z test for a population mean assuming independent samples with unequal variances?
- How would we use Excel to apply a two sample Z test for a population mean assuming dependent samples?
- How would we confirm that two population variances are equal?
- How would we use Excel to conduct the variance ratio test?

Chapter 8: Chi-square and non-parametric hypothesis testing

In Chapter 8 we introduced the student to tests that require the underlying population to be normally distributed or assumed to be approximately normally distributed where the data type is numerical. In this chapter we will explore the chi-square test for category data and a range of non-parametric tests which make no assumption about the form of the population distribution but can be used with ordinal (or ranked) data variables. The key revision topics include:

- What is a category variable?
- How do we measure independence (or association) between two category variables (or at least one category variable)?
- How would we use the chi-square test to test for association?
- How would we apply the chi-square test to test 2 population proportions when we have independent and dependent variables?
- What do we mean by a non-parametric test?
- How would we apply the sign test to assess the sample median against the population median (p = 0.5)?
- How would we use Excel to apply the sign test?
- How would we apply the Wilcoxon signed rank sum test (or matched pairs) to two dependent samples?
- How would we use Excel to apply the Wilcoxon signed rank sum test?

- How would we apply the Mann-Whitney test to two independent samples?
- How would we use Excel to apply the Mann-Whitney test?

Students will need to make sure that they practice reproducing the solutions by hand with tables of critical statistics. The text uses Excel to provide the critical test statistics (p-value, critical test value) and includes the chi-square critical table in the appendices.

Chapter 9: Linear correlation and regression analysis

Chapter 9 introduces the students to the concept of linear correlation and regression analysis. The key revision topics include:

- How can we represent the possible relationship visually between two data sets?
- How would we use Excel to generate an appropriate scatter plot or time series plot?
- Why taking care with the axes scales important when using a scatter plot (or time series plot) to assess visually a potential relationship?
- What are the key methods to assess the correlation between two data sets involving numerical and/or ordinal data?
- How would we assess the significance of these values of correlation?
- How would we use Excel to calculate the appropriate correlation coefficient and assess the significance level?
- What method will we use to fit a line to the data set?
- What are the regression assumptions?
- What tests would we use to assess the reliability of the regression line fit?
- How would we use Excel formulas (sum, count, multiplication, and so on) to fit a linear regression model to a data set and assess the reliability/significance of the line fit?
- How would we use Excel functions (slope, intercept, rsq, t.dist, t.inv, f.dist, f.inv, and so on) to fit a linear regression model to a data set and assess the reliability/significance of the line fit?
- How would we use Excel Data Analysis ToolPak to fit a linear regression model to a data set and assess the reliability/significance of the line fit?
- What do we mean by a non-linear relationship between two variables?
- Can I identify the form of the non-relationship and name the curves?
- How would I use Excel add trend line to fit a non-linear curve to a data set?
- What do we mean by multiple regression models?

Students will need to make sure that they practice reproducing the solutions by hand with tables of critical statistics. The text uses Excel to provide the critical test statistics (p-value, critical test value) and includes the appropriate set of critical tables in the appendices.

Chapter 10: Introduction to time series data, long term forecasts and seasonality

Chapter 10 continues the process of fitting a relationship between two data values but in this case one of the variables is time. The key revision topics include:

- What do we mean by a time series?

- What do we mean by a stationary and non-stationary time series?
- What do we mean by a seasonal pattern?
- What do we mean by univariate and multivariate methods?
- Why taking care with the axes scales important when using a scatter plot (or time series plot) to assess visually a potential relationship?
- What do we mean by trend extrapolation and be able to calculate values by hand and using Excel?
- What do we mean by seasonal models and be able to calculate values by hand and using Excel?
- What do we mean by forecasting errors and be able to calculate the values of specific error terms by hand and by using Excel?
- What is the prediction interval and how is it calculated?

Chapter 11: Short and medium term forecasts

Chapter 10 continues the process of fitting a relationship between two data values but in this case one of the variables is time, however, in this chapter we are concerned with short-term and medium-term forecasts only. The key revision topics include:

- What do we mean by moving average (and time series smoothing) and be able to calculate values by hand and using Excel?
- What do we mean by exponential smoothing and how do we calculate the values by hand and using Excel?
- How is seasonality treated by using exponential smoothing principles?
- How to use Holt-Winters method?
- How to handle prediction intervals?

Online chapters: Want to learn more

These online chapters extend virtually all of the chapters from the textbook with incremental material to help students learn more, if they wish so. The key files included are:

- A1Wa Histograms with unequal class intervals
- A1Wb Frequency polygon
- A1Wc Cumulative frequency curve
- A1Wd Superimposing two sets of data onto one graph
- A1We Using Excel to create a box plot
- A2Wa Inferring the population skewness value from the sample
- A2Wb Inferring the population kurtosis value from the sample
- A2Wc Chebyshevs Theorem simulation
- A2Wd Measures of average and dispersion for a frequency distribution
- A2We Statsmark frequency distributions Generating a grouped frequency distribution from raw data using SPSS
- A3Wa The probability laws
- A3Wb Probability Distributions and Approximations
- A3Wc Other useful probability distributions
- A4Wa Use SPSS to demonstrate the Central Limit Theorem

- A5Wa Bootstrapping
- A6Wa Common assumptions about data
- A6Wb Three Common P-Value Mistakes You'll Never Have to Make
- A6Wc Comparing population variances - Variance Ratio F test and Levenes test
- A6Wd Statistical power and type II error
- A7Wa Factorial Experiments workbook
- A7Wb Welch ANOVA test
- A8Wa Two sample Z tests using Excel
- A8Wb Chi square goodness of fit
- A8Wb Chi square test for one population variance
- A8Wd Chi square test for normality
- A9Wa Testing the significance of linear correlation between the two variables
- A9Wb Testing the significance of Spearman rank correlation coefficient
- A9Wc The use of t-test to test whether the predictor variable is a significant contributor
- A9Wd The use of F-test to test whether the predictor variable is a significant contributor
- A9We Confidence interval estimate for slope
- A9Wf Autocorrelation
- A9Wg Standard error for the autocorrelation function
- A9Wh Significance of the autocorrelation coefficients and evaluation
- A9Wi Partial autocorrelation function
- A9Wj Error and residual inspection
- A9Wk Non-linear regression analysis
- A9Wl Multiple regression analysis
- A10Wm Linear regression and Durbin Watson test for autocorrelation
- A10Wn Regression goodness of fit
- A10Wa Other type of trends
- A10Wb Index numbers
- A11Wa Compare different equations when forecasting with exponential smoothing
- A11Wb Excel ETS seasonal exponential smoothing

It is important to emphasise that these chapter either provide more advanced material to the topics from the textbook, or they provide deeper explanation of some of the topics covered in the textbook.