Course Handbook: An Introduction to Data Analysis

LEAPS Summer School 2017.

Course co-ordinator, Kevin Ralston, University of Edinburgh

Email: kev.ralston@ed.ac.uk

Acknowledgment: this course pulls together many resources that have been made available online. These resources are the property of the third parties who have generously made access to these sources possible. The SPSS syntax provided on this course is an adapted version of a course created by Professor Paul Lambert and Professor Vernon Gayle. I thank them for making the course available.

Contents

[Course Handbook: An Introduction to Data Analysis 1](#_Toc484446616)

[LEAPS Summer School 2017. 1](#_Toc484446617)

[**Files** 3](#_Toc484446618)

[**Other Resources:** 4](#_Toc484446619)

[**SPSS** 5](#_Toc484446620)

[**Course design** 6](#_Toc484446621)

[**Timetable** 6](#_Toc484446622)

[**Assessment** 6](#_Toc484446623)

[**Lessons** 10](#_Toc484446624)

[**Lesson 1: Research Process** 10](#_Toc484446625)

[**Lesson 2: Data and analysis** 11](#_Toc484446626)

[**Lesson 3: Types of data: exploring data with graphs and numerical summaries** 11](#_Toc484446627)

[**Lesson 4: Association: contingency, correlation, and regression** 12](#_Toc484446628)

[**Lesson 5: Data gathering** 13](#_Toc484446629)

[**Lesson 6: Normal distribution** 14](#_Toc484446630)

[**Lesson 7: Sampling distributions** 14](#_Toc484446631)

[**Lesson 8: Confidence Intervals** 15](#_Toc484446632)

[**Lesson 9: Statistical Significance** 15](#_Toc484446633)

[**Lesson 10, Comparing two groups** 16](#_Toc484446634)

[**Lesson 11:** **Association between categorical variables** 16](#_Toc484446635)

[**Lesson 12:** **Association between quantitative variables** 17](#_Toc484446636)

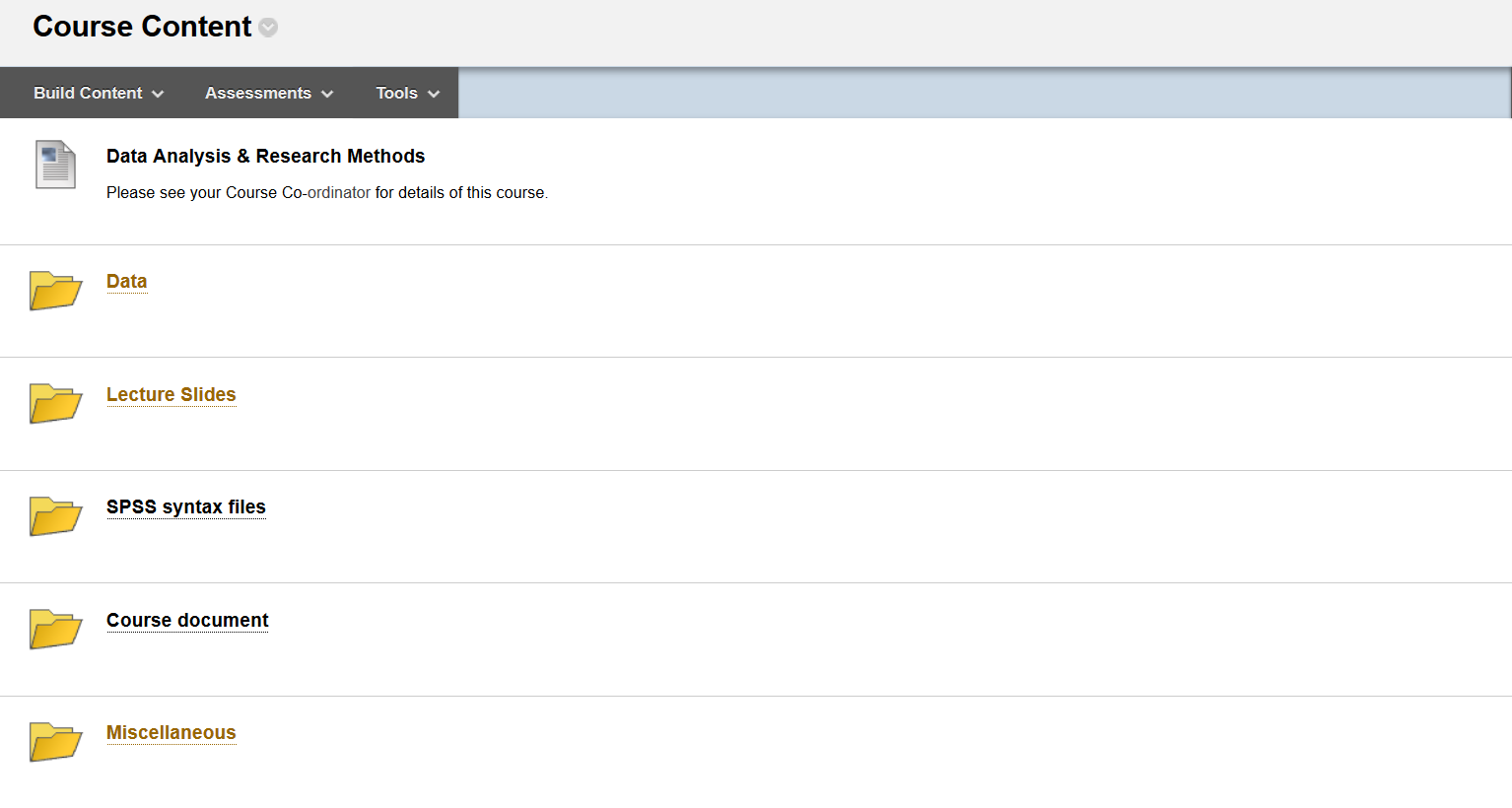
[**Final lesson,** **More on Multivariate analysis** 17](#_Toc484446637)

[**Appendix A** 18](#_Toc484446638)

**Files**

Files for the course are available on LEARN. Files can be also found in a publically accessible Google Drive file:

**Learn file structure**

****

**Data** for the labs can be found in the data folder. The labs will almost always use General Household Survey (GHS) teaching dataset from 1995.

**SPSS syntax files:** contains example programming files for each lesson

**Lecture slides:** self-explanatory!

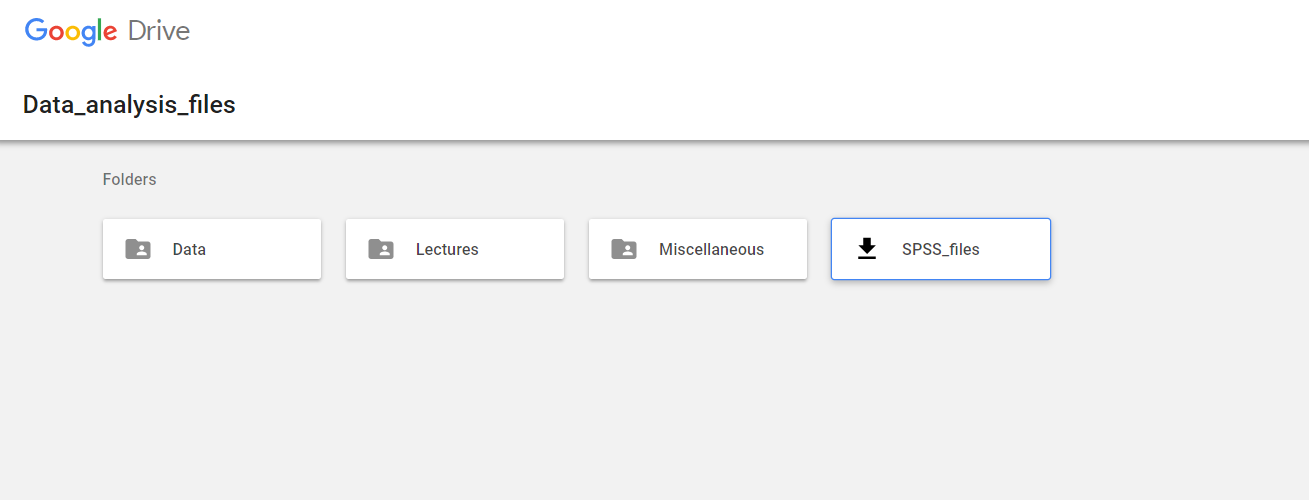
**Course document:** contains this document.

**Miscellaneous:** we will put anything else we come across in here

**Google drive:**

<https://drive.google.com/open?id=0Bw5ocrZBPpjRenZTZ25Kd1ItZ1U>

**Google drive:**

****

SPSS syntax programming files can be found in the file: **SPSS\_files**

The General Household Survey, teaching dataset, along with a PDF codebook can be found in the file: **GHS\_data**

Access to Google Drive may work best through browsers Chrome or Explorer (I tend to use Firefox, because of its open source credentials but this may not work).

The GHS data and SPSS programming syntax files can also be downloaded here:

<http://www.restore.ac.uk/Longitudinal/workshop_materials.html>

This is from a postgraduate course in longitudinal data analysis (LDA). This is a bit more advanced than what we are learning about on this course. The LDA course included an introductory lab (Lab0.sps) with more basic operations. I have taken the short labs on this course from this file. Feel free to work through this and the rest of this course if you want to see some more advanced things.

**Other Resources:**

There are a large number of resources available. For example, you can find textbooks, web-pages, and YouTubes.

There are whole online courses you can sign up to for free on places like Coursera:

<https://www.coursera.org/courses?languages=en&query=data+analysis>

People have written courses you can access e.g. here is another introductory course in analysis using SPSS from Kent State University in the USA, have a look you might find it useful:

<http://libguides.library.kent.edu/SPSS/home>

I recommend checking out the resources specified for a lesson. If you do not connect with the suggested resources it is worth spending some time to find a version of things you do connect with.

There are several open source textbooks that might be of use to you.

Here is a great free text-book introduction to statistics, take a look:

<https://openstax.org/details/introductory-statistics>

Here is another one:

<http://onlinestatbook.com/Online_Statistics_Education.pdf>

Professor Andy Field has written books that students seem to like. E.g. Discovering Statistics Using SPSS, is in a 4th Edition. An earlier edition can be downloaded in many places like here:

<http://www.soc.univ.kiev.ua/sites/default/files/library/elopen/andy-field-discovering-statistics-using-spss-third-edition-20091.pdf>

**SPSS**

Why SPSS? People argue about which software is ‘best’ for analysis. SPSS is a general purpose package often taught in the social sciences. It combines data management capability with a user interphase that does not tend to frighten people too much.

SPSS is often used by government researchers and health researchers in the UK. The SPSS book on Data Management I link to was written by someone working as an actuary in insurance (I notice the link does not work sometimes so will put a copy of the book in the *Misc* folder on LEARN and Google Drive).

Many analysts do not use SPSS, however. I usually use Stata, a package written for economists. But there are many, many packages, and wherever you go as a data analyst someone will always tell you there is a better analysis software that you should be using instead (usually because it’s the one they use).

One of the pieces of coursework is to do a small piece of replicable analysis. You can do this using SPSS. SPSS is available on University managed computers in Labs. You can find a Lab to work in if you like. You can work on this in the Library, where there are computers Labs available. You can download SPSS to a laptop or personal computer.

Getting SPSS onto a personal computer:

<http://www.ed.ac.uk/information-services/computing/desktop-personal/software/main-software-deals/spss/getting-spss/spss-non-managed>

A book on SPSS programming:

<http://www.spsstools.net/en/spss-programming-book/>

A website where fully annotated SPSS code is available for a variety of analyses:

<https://stats.idre.ucla.edu/other/annotatedoutput/>

A list of statistical packages:

<https://en.wikipedia.org/wiki/List_of_statistical_packages>

**Course design**

This course is designed so that each lesson introduces resources to enable understanding of key concepts in data analysis. Most lessons will include presentations and have readings and lab material suggested. This can be worked through during time available in class and as homework. There is a lot to learn about data analysis. We could spend the rest of our lives studying data analysis and still not know all there is to know. There is not time for you to become ‘expert’ in everything you are introduced to on this course. The aim is to introduce you to concepts and resources and to get you to think about how to apply these in a small piece of research. This is the one step on a data analysis expedition that might last the rest of your life.

**Timetable**

|  |  |  |
| --- | --- | --- |
| 1,Tues 6th June | Introduction  A look at the ‘research process’ and starting with the software (SPSS). Syntax and read data into the package. |  |
| 2, Fri 9th June | Data, what are data and where to find it! |  |
| 3, Tues, 13th  June | Exploring data with graphs and numerical summaries |  |
| 4, Fri  16th June | Association: contingency, correlation, and regression |  |
| 5, Tues  20th June | Gathering data |  |
| 6, Fri 23rd June | Probability distributions (normal) |  |
| 7, Tues  27th June | Sampling distributions | Hand in course work 1 |
| 8, Fri  30th June | Statistical inference: confidence intervals |  |
| 9, Tues  4th July | Statistical inference: significance tests |  |
| 10, Fri  7th July | Project work |  |
| 11, Tues  11th July | Comparing two groups |  |
| 12, Fri  14th July | Project work |  |
| 13, Tues  18th July | Categorical association and regression | Hand in course work 2 |
| 14, Fri  21st July | Class test |  |

**Assessment**

Assessment will comprise of three items, two pieces of coursework and a class test:

30% Coursework, write a research proposal

40% Coursework, write a short piece of replicable analysis

30% Class test

**Course Work 1 write a research proposal, 30% of grade.**

Write a research proposal. In reference to the course document on *Writing a research project proposal* (appendix a, below, and also on Google Drive)*,* produce a research proposal of up to 1000 words.

*This should include a research question and the reasons why it is important to study the issue you highlight. You may want to consider where and how you would find/collect data on the subject. You may also want to suggest the type of analysis you might do. You may want to consider the ethics of your research. Look at related literature. You may want to look at other research that has been undertaken in the area and how your work would fit in with this. Does your research answer questions other research does not? Does your research aim to replicate something that has been done previously? If so is this important. Find similar research in the newspapers, at least search on google scholar and mention what other researchers in the field have found.*

You *should* be critically aware of your research proposal. On your proposal, comment on its strengths and weaknesses. Where would you foresee problems arising? What are the difficulties?

A *research proposal* is a concise and coherent summary of your proposed *research*. It sets out the central issues or questions that you intend to address. It outlines the general area of study within which your research falls, referring to the *current state of knowledge* and any recent debates on the topic.

Lots of advice is ‘out there’ on writing a research proposal. Have a search (I recommend using Duck Duck Go, it doesn’t track you and commercialise your data).

Here is a useful resource on formulating research questions.

* <http://www.socscidiss.bham.ac.uk/research-question.html>

Below is the University of Birmingham’s criteria for those applying to do higher degrees. You might want to use this as a template. Although, we do not expect you to produce as comprehensive a document as someone applying to undertake a Doctorate in law! (http://www.birmingham.ac.uk/schools/law/courses/research/research-proposal.aspx)

A research proposal should normally include the following information:

**1. Title**

This is just a tentative title for your intended research. You will be able to revise your title during the course of your research if you are accepted for admission.

**2. Abstract**

The proposal should include a concise statement of your intended research of no more than 100 words. This may be a couple of sentences setting out the problem that you want to examine or the central question that you wish to address.

**3. Research Context**

You should explain the broad background against which you will conduct your research. You should include a brief overview of the general area of study within which your proposed research falls, summarising the current state of knowledge and recent debates on the topic. This will allow you to demonstrate a familiarity with the relevant field as well as the ability to communicate clearly and concisely.

**4. Research Questions**

The proposal should set out the central aims and questions that will guide your research. Before writing your proposal, you should take time to reflect on the key questions that you are seeking to answer. Many research proposals are too broad, so reflecting on your key research questions is a good way to make sure that your project is sufficiently narrow and feasible (i.e. one that is likely to be completed with the normal period for a MJur, MPhil or PhD degree).

You might find it helpful to prioritize one or two main questions, from which you can then derive a number of secondary research questions. The proposal should also explain your intended approach to answering the questions: will your approach be empirical, doctrinal or theoretical etc?

**5. Research Methods**

The proposal should outline your research methods, explaining how you are going to conduct your research. Your methods may include visiting particular libraries or archives, field work or interviews.

Most research is library-based. If your proposed research is library-based, you should explain where your key resources (e.g. law reports, journal articles) are located (in the Law School’s library, Westlaw etc). If you plan to conduct field work or collect empirical data, you should provide details about this (e.g. if you plan interviews, who will you interview? How many interviews will you conduct? Will there be problems of access?). This section should also explain how you are going to analyse your research findings.

**6. Significance of Research**

The proposal should demonstrate the originality of your intended research. You should therefore explain why your research is important (for example, by explaining how your research builds on and adds to the current state of knowledge in the field or by setting out reasons why it is timely to research your proposed topic).

**7. Bibliography**

The proposal should include a short bibliography identifying the most relevant works for your topic.

**Course Work 2, write a short piece of replicable analysis, 30% of grade.**

Write a syntax (SPSS programming) file that reads in organises, and does a piece of analysis on data. You can use the General Household Survey (GHS) file supplied for this course. You can download another data file from the web and do some analysis that might interest you. If you do this you will have to do some work to get the file into SPSS if it is not an SPSS file. Or you can input data yourself and analyse it.

Write a short report on the analysis you did. Present the analysis and offer an interpretation of the analysis presented. Your report should be less than 2000 words. Headings you might want to use to structure your report are below. Your report should refer to other research and literature from the field in which you are researching. You should address how your work relates to what has already been done?

|  |  |
| --- | --- |
| What is the problem? | [Introduction](http://abacus.bates.edu/%7Eganderso/biology/resources/writing/HTWsections.html#introduction) |
| How did I solve the problem? | [Materials and Methods](http://abacus.bates.edu/%7Eganderso/biology/resources/writing/HTWsections.html#methods) |
| What did I find out? | [Results](http://abacus.bates.edu/%7Eganderso/biology/resources/writing/HTWsections.html#results) |
| What does it mean? | [Discussion](http://abacus.bates.edu/%7Eganderso/biology/resources/writing/HTWsections.html#discussion) |
|  | References |
| See, how to write a paper in a scientific format:  http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWsections.html | |

This assessment is about learning to go through the process of undertaking an analysis in a systematic and efficient way. Getting data into SPSS is hard, as is organising it. Running a statistical analysis is easy, running a justifiable analysis, however, requires knowledge and understanding. Creating new research, evaluating and analysing are complex skills.

Here are resources on reading data into SPSS:

<http://libguides.library.kent.edu/SPSS/CreateData>

http://libguides.library.kent.edu/SPSS/ImportData

Here is a resource with syntax for different types of analysis using SPSS:

<https://stats.idre.ucla.edu/other/annotatedoutput/>

**Lessons**

**Lesson 1: Research Process**

What we will cover today:

* We will learn about the research process.
  + Processes of producing robust, replicable research
* We will learn about research proposals
  + Criteria for writing proposals.
  + Issues around generating research proposals.
  + The first piece of coursework
* Practical:
  + How to get data into and out of SPSS

Some multiple choice questions.

Ice breaker: Is anyone interested in data analysis? Opinion line task.

Introduction to the research process, short lecture

Introduce the first piece of course work: Produce a research proposal

**Lab:** Lesson1, getting data into SPSS, using Syntax files and reading in data!

Exercise in creating data, **lets do this!**

<http://libguides.library.kent.edu/SPSS/CreateData>

<http://www.stattutorials.com/SPSS/TUTORIAL-SPSS-Create-a-data-set.htm>

Exercise in reading data into SPSS from other file formats such as Excel and text files. This may be useful if you have data in other formats. You do not need to do this now, but it is there if you need it.

<http://libguides.library.kent.edu/SPSS/ImportData>

Here is a useful resource on formulating research questions.

* <http://www.socscidiss.bham.ac.uk/research-question.html>

Here is a **free** book on programming in SPSS, which you may find useful, although all the programming you need is given to you for this course, its good to have resources and multiple ways to check what you are doing (and also SAS, R, and Python!)

* <https://developer.ibm.com/predictiveanalytics/wp-content/uploads/sites/48/2015/04/Programming-and-Data-Management-for-IBM-SPSS-Statistics-23.pdf>

Paper on undertaking reproducible data analysis, the software these analysts use is Stata, but the principles apply to SPSS and any other analysis package.

Lambert and Gayle (2017) The Workflow: A Practical Guide to Producing Accurate, Efficient, Transparent and Reproducible Social Survey Data Analysis, *NCRM Working Paper*

* <http://eprints.ncrm.ac.uk/4000/1/The%20Workflow%20A%20Practical%20Guide%20to%20Producing%20Accurate%2C%20Efficient%2C%20Transparent%20and%20Reproducible%20Social%20Survey%20Data%20Analysis%20.pdf>

**Lesson 2: Data and analysis**

Today we will:

* Think a bit about data and where to find it.
* Work through an online introduction to data analysis and SPSS
* Do some lab work on coding data
* Introduce the second piece of course work

**Group task:** discuss a subject/piece of research you might be interested in doing!

**Additional material:**

Have a look at these web pages on data and data sources:

What are data and where to find it?

<http://elon.libguides.com/data>

Some data sources:

<http://elon.libguides.com/c.php?g=331454&p=2224517>

<http://elon.libguides.com/c.php?g=331454&p=2227698>

Work through this course, it introduces data analysis and SPSS

Learning to analyse quants data:

<http://archive.learnhigher.ac.uk/analysethis/main/quantitative.html>

A tutorial on SPSS syntax is below:

<http://libguides.library.kent.edu/SPSS/Syntax>

**Lab:** Lesson2 this is a lab on recoding variables, data operationalisation tasks which all analyses need to do.

**Lesson 3: Types of data: exploring data with graphs and numerical summaries**

Today we will:

* Talk about types of data
* Read about data and run some simple simulations
* Lab on analytic techniques

**Additional Material:**

Reading on variables:

[*http://survivestatistics.com/variables/*](http://survivestatistics.com/variables/)

Describing the centre of quants data

[*http://survivestatistics.com/mean-median/*](http://survivestatistics.com/mean-median/)

Some simple simulations describing data:

Describing the spread

Mean and median film ratings:

<https://economicsnetwork.ac.uk/statistics/film_ratings.htm>

Skew, median and mean

<https://economicsnetwork.ac.uk/statistics/skew1.htm>

Skew a different distribution (poisson)

<https://economicsnetwork.ac.uk/statistics/poisson_mmm.htm>

Article on skew and kurtosis

<https://www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics>

**Lab:** Have a look at some data, reading it in beginning to analyse data. There are some analyses demonstrated that are possible for the different types of variable we have been introduced to today.

**Lesson 4: Association: contingency, correlation, and regression**

What we will do today:

* Learn about chi square and categorical association
* Learn about correlation and metric associations
* Learn about regression

**Lecture:** on categorical associations, metric (linear, quantitative –all the same thing different words) associations and regression.

**Lab:** taking a look at a range of test statistics available and thinking about how these might apply to an analysis you want to do.

Additional suggested material, take a look if you have time in class, if not, take a look when you have a moment or feel you need to brush up/extend/expand your understanding.

Correlation simulation:

<http://rlanders.net/correlation.html>

Another correlation simulation:

<http://shiny.albany.edu/stat/corrsim/>

Correlation in plain English:

<http://neoacademic.com/2009/09/10/a-plain-english-explanation-of-correlation/>

Lab on correlation and scatterplot

Predicting an outcome

http://www.statstutor.ac.uk/resources/uploaded/PearsonCorrelation/Tutorial/pearsons.htm

|  |  |
| --- | --- |
| correlation: | https://www.youtube.com/watch?v=372iaWfH-Dg |
| intro to regressions: | https://www.youtube.com/watch?v=ocGEhiLwDVc |

We want to say something about the magnitude of an association:

* **Exactly** –**1.** A perfect downhill (negative) linear relationship
* –**0.70.** A strong downhill (negative) linear relationship
* –**0.50.** A moderate downhill (negative) relationship
* –**0.30.** A weak downhill (negative) linear relationship
* **0.** No linear relationship
* **+0.30.** A weak uphill (positive) linear relationship
* **+0.50.** A moderate uphill (positive) relationship
* **+0.70.** A strong uphill (positive) linear relationship
* **Exactly +1.** A perfect uphill (positive) linear relationship

**Lesson 5: Data gathering**

What we will do today

* Lecture on types of sampling.
* Exercise on Surveys and Survey Methods
* Opinion line fun
* Free time to work on course work

Work through this online from the section headed **Survey** to **Plus & Minus of Survey Methods**

<https://www.socialresearchmethods.net/kb/survey.php>

Take a look at online data collection facilities.

<https://www.onlinesurveys.ac.uk/>

<https://www.surveymonkey.co.uk/>

Here is a book, from the American Statistical Association, which introduces surveys and sampling have a read:

<https://www.whatisasurvey.info/overview.htm>

**Lab:** more data coding and analysis options.

*You don’t need to know how to do all of the things you see in these labs, you don’t need to understand every possible part of analysis. You just need to be aware that there are different options available and to be help make choices about which to use at which time. The resources in this course introduce you to a range of options and get you to start to think about applying and interpreting some of these.*

**Lesson 6: Normal distribution**

Talk on distributions including standard errors

Read more about sd and variance and also wider concepts if you have time/energy

**Lab:** Lesson 6 – univariate inferential statistics

**Additional material:**

A simple introduction and exercise on standard deviation:

<http://www.mathsisfun.com/data/standard-deviation.html>

A short explanation of probability distributions:

<http://stattrek.com/probability-distributions/probability-distribution.aspx>

A chapter on different distributions, the heading ‘typical shapes’ introduces the normal distribution:

<https://www.boundless.com/statistics/textbooks/boundless-statistics-textbook/frequency-distributions-4/>

Before the next lesson on sampling distributions here is an article on, central limit theorem:

<https://economicsnetwork.ac.uk/statistics/sample_means_households.htm>

Some online lectures on probability distributions, we might watch these as a class:

A short lecture on the probability distribution of a coin toss:

<https://www.khanacademy.org/math/statistics-probability/random-variables-stats-library/discrete-and-continuous-random-variables/v/discrete-probability-distribution>

Probability distribution on dice rolls:

<https://www.khanacademy.org/math/precalculus/prob-comb/independent-events-precalc/v/events-and-outcomes-2>

**Lesson 7: Sampling distributions**

A bit different today: A tail of 3 distributions, a class exercise on sampling.

Central limit theorem – again, if you haven’t looked at this, check it out:

<http://survivestatistics.com/sampling-distribution-central-limit-theorem/>

A simulation of central limit theorem:

<https://economicsnetwork.ac.uk/statistics/sample_means_households.htm>

**Lab:** Lesson 7 – bivariate inferential statistics

**Lesson 8: Confidence Intervals**

Lecture, introduction to confidence intervals.

Online exercise, on calculating confidence intervals. Do the calculation again using a t-distribution you can find the t-distribution from the link below and plug this into the calculation:

* <http://www.mathsisfun.com/data/confidence-interval.html>

Explanation of when to use a t distribution or a normal (z) distribution:

* <http://onlinestatbook.com/2/estimation/mean.html>

t-distribution table:

* <https://www.medcalc.org/manual/t-distribution.php>

Confidence interval calculator which uses a t-distribution (you can check -or cheat- using this):

<http://www.sample-size.net/confidence-interval-mean/>

**Lab:** Lesson 8, multivariate approach to data analysis, a regression analysis. This is just a short lab. We talked about regression in Lesson 4. You can look back at lesson 6 at confidence intervals again, now we have talked about them. Think about what this would mean for your own research.

**Lesson 9: Statistical Significance**

Lecture, statistical significance, p-values

Reading and some questions on significance testing:

<http://onlinestatbook.com/2/logic_of_hypothesis_testing/significance.html>

Reading and some questions on type I and type II errors:

<http://onlinestatbook.com/2/logic_of_hypothesis_testing/errors.html>

Steps in hypothesis testing:

<http://onlinestatbook.com/2/logic_of_hypothesis_testing/steps.html>

An SPSS example of correlation and its significance test – *we did this before but now we know more!*

<http://libguides.library.kent.edu/SPSS/PearsonCorr>

An SPSS chi-square example – *we also did this before but now we know more!*

<http://libguides.library.kent.edu/SPSS/ChiSquare>

**Lab:** Lesson9, multivariate analysis techniques

**Lesson 10, Comparing two groups**

Lecture on comparing means. T-test.

Comparing means:

<http://onlinestatbook.com/2/tests_of_means/testing_means.html>

T-test example:

<http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1438>

Paired t-test example:

<http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1439>

Regression analysis UCLA annotates explanation, we introduced regression early on, we looked at it a bit in the lab, it is time to think about it a bit more:

<https://stats.idre.ucla.edu/spss/output/regression-analysis/>

**Lab:** Lesson10, Regression analysis, another look at specifying a regression.

*The regression uses dummy category coding. This is used for categorical variables. One category from the variable is omitted and the result from the other categories are interpreted in contrast to the category that is omitted (see the link below, this is quite technical, I’ll talk about it in the lecture for lesson 12).*

<https://stats.idre.ucla.edu/spss/faq/how-do-i-interpret-theparameter-estimates-for-dummy-variables-in-regression-or-glm/>

**Lesson 11:** **Association between categorical variables**

Lecture: Chi-square again, recap and some more issues, now we know more

Online lab work:

<http://libguides.library.kent.edu/SPSS/ChiSquare>

Readings:

Chi square distribution:

<http://onlinestatbook.com/2/chi_square/distribution.html>

Contingency table:

<http://onlinestatbook.com/2/chi_square/contingency.html>

**Lab:** Lesson11, syntax file, categorical associations, Gamma, Phi, Cramers V and Eta

**Lesson 12:** **Association between quantitative variables**

Talk on regression analysis

Regression:

<http://onlinestatbook.com/2/regression/multiple_regression.html>

<https://stats.idre.ucla.edu/spss/output/regression-analysis/>

Lab: Lesson 12 more regression examples.

**Final lesson,** **More on Multivariate analysis**

This course is just the start: We have seen OLS, Logit, Poisson, M-logit, O-Logit models. There are loads of other techniques we’ve not been able to cover. For example, longitudinal models, event history, fixed-effects models and multi-level models, plus many more.

We’ll talk a bit more about this. We will think about the class test. We will talk about the course.

**Appendix A**

**Writing a research project proposal**

This document was created for a course in Data Analysis at the Leaps Summer School. It outlines issues analysts need to think about when generating ideas for research. It is an adapted version of material created at Sheffield Hallam University, (<http://www.socscidiss.bham.ac.uk/>).

The original document was produced to support students undertaking final year dissertations in social science subjects. The issues around formulating research projects are relevant for any piece of empirical work.

You have only a short time to write your proposal. You will need to make decisions about which parts to focus your time on and what the most important parts to include in your proposal are.

This document gives a lot of guidance which should help you to make decisions. The purpose of this task is to get you to start to think about the formulating research. It is the first step on a process which could lead to you becoming an independent researcher and analyst.

You are not expected to come up with a proposal that will change the world. We want you to engage with the challenge and with wider work, in the area you choose, and to start to think about how you would approach the problems you identify.

**Avoid too broad a topic**

Avoid too broad a topic or one that is overly ambitious: it is better to find a thoroughly researched and argued answer to a small question than to fail to find the answer to one which is too big or diffuse.  
Your main interest in the topic may be:

* An area of social life.
* A type of method that you would like to use.
* A body of theory that you are interested in exploring.

Bringing all three of these together is a way of narrowing the focus of the dissertation into a manageable project.

**Start writing at the beginning of the project**

Many people find it useful to keep a research notebook in which you can record:

* Your initial ideas.
* Points from the literature or other sources that you are consulting.
* Your observations and impressions.
* New ideas as they develop.
* Problems that you come across.

**Keep records of your reading at the preliminary stage**

Keep an accurate record of the bibliographical details of all the material that you read - doing this as you progress will save an enormous amount of time at the end of the project.

* Check that you know the referencing system specified by your course and get into the habit of using it.
* Always write the full bibliographic details on the top of any photocopies or print-outs.
* Keep a running bibliography in alphabetical order as a computer file or card index.
* Note in your research diary when you read a particular source.
* Use software that is available to you to manage your references, e.g. Ref Works.

Note: When you refer directly to the work of an author in your dissertation, it is especially important to record the details precisely, to ensure that it is accurately referenced and to avoid the risk of plagiarism (for more information see the section on **plagiarism**).

**Do lots of reading**

A final year project, like many other forms of assessments, needs to be located within the existing literature in that area. In order to do this you need to do lots of reading! Typically you will read:

* Classic studies in your chosen area.
* Recent studies published as books or journal articles.
* Research methods and methodology texts.
* Social theory relevant to your approach.
* Primary literature sources.

The typical length of a bibliography for a dissertation would include anything between 25-50 references. For example, there are, to varying degrees, references to:

* Theoretical/conceptual material.
* Methods/methodologies.
* Policy/practitioner literature.

**Be organised and keep notes**

The process of thinking about the dissertation topic and methods is an evolving one. It may help to get some form of personal recording of the ideas, links and resources that you come across in the initial thinking and information-gathering stages. Do not simply rely on your memory to store all the strands of information you come across. A key part of success in dissertation-writing is being organised and systematic in your approach and the earlier you can adopt this, the better.  
This type of note-taking may link into the writing of other learning logs or personal development planning you are doing already within your degree.   
You might want to keep a record of:

* Questions or ideas that interest you.
* Possible ways of researching these.
* References to follow up at a later stage.
* Sources of information that you have found useful.
* Notes on articles and papers you have read or programmes you have seen or heard.

You can use library resources and the Internet to discover primary and secondary material and to find out what critical and scholarly material is available to inform your study.

**Recap**

*The first task is to establish your overall area of interest.* It will be important that the topic you choose can interest you enough to sustain your commitment. Write down the reasons for your interest. These may provide pointers to the sort of questions you want to ask within the study. As you go on, you may find that your interest begins to take shape, or changes a little, as you focus more closely on it.

**Clarifying your Ideas**

* What is the overall area of your interest? Write a paragraph that would give someone else a clear picture of the issues.
* How has your interest developed over time? Can you identify incidents or experiences that have generated your interest? These may be personal or professional, or to do with current work priorities.
* Are there any key writers who have shaped your interest or whose views conflict with yours?
* Where would you like the work to lead in the longer term? Is this research connected with work you currently do or would like to do at some stage?
* Does anyone else have an interest in the topic you choose for your study? This may not be a problem but it is important to recognise if there are others with interests in the work.

**Narrowing down the Focus!**

* What are the questions to which you want to find answers in your research? You might have a hypothesis – i.e. a belief about something (founded upon evidence) which has never been fully tested, proved or disproved. You may, on the other hand, want to couch your interest in terms of an exploration of issues, attitudes or experiences, or as a question. Write a list of all the questions you want to answer and group them into priorities or hierarchies and show the connections between them. At this stage you may want to do some weeding out of overlapping or less relevant questions. It is helpful to list your questions and then to answer why you want to know the answer and how it will help you to pursue your overall enquiry.
* Where is any work currently being done in this area? Can you identify any specialist collections of literature? Are there particular people associated with them?
* What do you know about what is currently known, written about or researched in the area?
* How are you going to track down the research and theory to support your study? Talk to tutors on the course to see if anyone can help.

**Formulating the Research Question**

**Introduction**

In the previous section we talked about ways to define your topic, but there is a difference between a topic and a question. You may have found your topic, but within that topic you must find a question, which identifies what you hope to learn. Finding a question sounds serendipitous, but research questions need to be shaped and crafted. This section examines the factors that go into creating a good research question, dividing this X factor into six categories.

[**Watch video on formulating a good research question (.wmv)**](http://www.socscidiss.bham.ac.uk/videos/v5-formulatinggoodqs.wmv)

This video clip contains comments from the following academics:

* Malcolm Todd  
  Sociology
* Shawna McCoy  
  Criminology
* Christopher Crowther-Dowey  
  Criminology
* Iain Garner  
  Psychology
* Kevin Bonnett  
  Sociology

[**Download Case Study 4 - Formulating the research question: youth justice policy and intervention**](http://www.socscidiss.bham.ac.uk/Documents/Case%20Study%204.docx)

What is a good research question?

It is important to start your thinking about the dissertation with a question rather than simply a topic heading. The question sets out what you hope to learn about the topic. This question, together with your approach, will guide and structure the choice of data to be collected and analysed.

Some research questions focus your attention onto the relationship of particular theories and concepts: **'how does gender relate to career choices of members of different religions?'** Some research questions aim to open an area to let possible new theories emerge: **'what is going on here?'** is the most basic research question in exploratory research. For an undergraduate dissertation, your question **needs to be more targeted than either of these**.  
Creating a research question is a task. Good research questions are formed and worked on, and are rarely simply found. You start with what interests you, and you refine it until it is workable.

There is no recipe for the perfect research question, but there are bad research questions. The following guidelines highlight some of the features of good questions.

**Top Tips:**

* Relevant.
* Manageable in terms of research and in terms of your own academic abilities.
* Substantial and with original dimensions.
* Consistent with the requirements of the assessment.
* Clear and simple.
* Interesting.

**Relevant**

The question will be of academic and intellectual interest to people in the field you have chosen to study. The question arises from issues raised in the literature or in practice.  
You should be able to establish a clear purpose for your research in relation to the chosen field. For example, are you filling a gap in knowledge, analysing academic assumptions or professional practice, monitoring a development in practice, comparing different approaches or testing theories within a specific population?

**Manageable**

You need to be realistic about the scope and scale of the project. The question you ask must be within your ability to tackle. For example, are you able to access people, statistics, or documents from which to collect the data you need to address the question fully? Are you able to relate the concepts of your research question to the observations, phenomena, indicators or variables you can access? Can this data be accessed within the limited time and resources you have available to you?  
Sometimes a research question appears feasible, but when you start your fieldwork or library study, it proves otherwise. In this situation, it is important to write up the problems honestly and to reflect on what has been learnt. It may be possible, with your supervisor, to develop a contingency plan to anticipate possible problems of access.

**Substantial and (within reason) original**

The question should not simply copy questions asked in other final year modules, or modules previously undertaken. It shows your own imagination and your ability to construct and develop research issues. And it needs to give sufficient scope to develop into a dissertation.

**Consistent with requirements of the assessment**

The question must allow you the scope to satisfy the learning outcomes of the course.  
For example, you can choose to conduct a theoretical study, one that does not contain analysis of empirical data. In this case, it will be necessary for you to think carefully before making such a choice. You would be required to give an account of your methodology, to explain why theoretical analysis was the most appropriate way of addressing the question and how you have gone about using theoretical models to produce new insights about the subject.

**Clear and simple**

The complexity of a question can frequently hide unclear thoughts and lead to a confused research process. A very elaborate research question, or a question which is not differentiated into different parts, may hide concepts that are contradictory or not relevant. This needs to be clear and thought-through, but it is one of the hardest parts of your work.  
Equally, you may want to begin with your literature review and data collection and you may feel tempted to 'make do' with a broad and vague research question for the moment. However, a muddled question is likely to generate muddled data and equally muddled analysis.  
If you create a clear and simple research question, you may find that it becomes more complex as you think about the situation you are studying and undertake the literature review. Having one key question with several sub-components will guide your research here.

**Interesting**

This is essential. The question needs to intrigue you and maintain your interest throughout the project. There are two traps to avoid.

* Some questions are convenient - the best you can come up with when you are asked to state a question on a form, maybe – or perhaps the question fits in with your units so you decide it will suffice.
* Some questions are fads - they arise out of a particular set of personal circumstances, for example a job application. Once the circumstances change you can lose enthusiasm for the topic and it becomes very tedious.

Make sure that you have a real, grounded interest in your research question, and that you can explore this and back it up by academic and intellectual debate. It is your interest that will motivate you to keep working and to produce a good dissertation.

**Student voice**   
  
It’s not an easy task formulating a research question. Here one student talks about the difficulties she had:  
  
I knew what I wanted to write about but I couldn’t get a question to match. My original question was too vague and unanswerable. In terms of tightening it up, I knew I wanted to link disability to employment. I tried to get a question from that but it was a descriptive question that I ended up scrapping on the advice of the supervisor, he told me it wasn’t any good as a question.   
  
(Todd, Bannister and Clegg, 2004, p340)  
  
This student did eventually come up with a workable question and went on to complete her dissertation. She was not afraid to call on the support of her supervisor and was willing to listen to his advice as to what would and wouldn’t work.

[Download Case Study 5 Devising research questions from a real geographical and social situation (.docx)](http://www.socscidiss.bham.ac.uk/Documents/Case%20Study%205.docx)

**Moving into action**

* By now you should be doing lots of reading in the area. Make sure you note, either on computer or on index cards, anything you read that is relevant to your study. Can you map out the contemporary debates and critiques in the area? Are there any recent legal or policy changes of significance? What are the main practice issues to consider?
* Where (i.e. in what settings) does the work you are interested in take place? What access do you have to it? Will there be ethical issues? How might you be able to negotiate access? What obstacles are there? While it is early days to be specific about you data collection, it is important to know that you are on a course which will yield data, rather than a series of negative responses.
* What sort of time scales are you going to need to do the sort of research you are planning? How much time have you got? Are your plans unrealistic?
* Having thought about these things, try narrowing down your ideas again to the sort of research you can do.
* Make a list of the skills and knowledge you bring to the research task. Do you like interviewing? Will you be able to have the interviews transcribed? Are you keen to do surveys? Remember that you will need to have a reasonable sample to undertaken meaningful quantitative analysis.
* Are there sources of secondary data that you could access?
* Are there possibilities for documentary analysis?

So far, we have considered a number of issues relevant to developing an appropriate research methodology for your dissertation. The chart below should help you to synthesise your thinking to date. Work through each of the boxes but be prepared to revisit this at different stages of the dissertation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Look at the template below and consider each of the sections. | | | | |
| **Research Question** | **Data Sources and Methods** | **Justification** | **Practicalities (e.g. resources and skills)** | **Ethical Issues** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

[Download this template (.docx)](http://www.socscidiss.bham.ac.uk/Documents/Doc%2014%20Template%20for%20Research%20Questions.docx)

**Summary**

Good research questions are:

* **Relevant:** Arising from issues raised in literature and/or practice, the question will be of academic and intellectual interest.
* **Manageable:** You must be able to access your sources of data (be they documents or people), and to give a full and nuanced answer to your question.
* **Substantial and original:** The question should showcase your imaginative abilities, however far it may be couched in existing literature.
* **Fit for assessment:** Remember, you must satisfy the learning outcomes of your course. Your question must be open to assessment, as well as interesting.
* **Clear and simple:** A clear and simple research question will become more complex as your research progresses. Start with an uncluttered question then unpeel the layers in your reading and writing.
* **Interesting:** Make your question interesting, but try to avoid questions which are convenient or flashy. Remember, you will be thinking about this question for an entire year.

**Key Questions**

* What aspect do you find the most interesting about your chosen field or topic?
* Is there 'room' for investigation in this sub-topic area?
* Have you tried formulating questions in different ways?
* Are you happy with your questions? (You will be the one working on them!)

**Further reading**

**BRYMAN, A. (2004).** Social Research Methods. 2nd ed., Oxford, Oxford University Press, chapter 2  
**CRESSWELL, J. W. (2003).** Research Design: Qualitative, Quantitative and Mixed Method Approaches. London, Sage, chapter 6  
**PUNCH, K. F. (1998).** Introduction to Social Research – Quantitative and Qualitative Approaches. London, Sage, chapter 4

**Acknowledgment**

I acknowledge fully those responsible for the original production of the resource on which this was based:

**Dr Malcolm Todd** - coordinating author and project leader of this resource. He is the author of a number of academic journal articles, co-authored and edited books.

**Julia Waldman** - Learning and Teaching Adviser, School of Social Sciences, University of Southampton.

**Ian Baker** - He has taught English Literature at Sheffield Hallam University since 1968.

**Dr Jenny Blain** - Senior Lecturer in Sociology at Sheffield Hallam University.

**Christopher Winch** - Professor of Philosophy of Education at University College Northampton.

**Andrew Pilkington** - Primary teaching and research specialism is within the Sociology of Race and Ethnicity.

**Dr Anne Hollows** - Principal Lecturer in Social Work at Sheffield Hallam University's Health and Social Care Research Centre.

**Sue Hemmings** - Staff Tutor at The Open University in East Anglia.

**Prof Ann Macaskill** - Chairs Sheffield Hallam University Research Ethics Committee.

**Dr John Steel -** Research focus of new learning technologies in higher education, with emphasis on the social and cultural aspects of their use.

**Dr Ruth McManus** - Lecturer in Sociology at the University of Canterbury, Christchurch, New Zealand. Honours Degree in Sociology at Glasgow University, MA in Women's Studies at UKC and PhD in Sociology at Massey University, NZ.

**Technical Authors**

**Dr Darren Marsh** - He is interested in the application of new technologies for learning and their relationship to the development of scholarship.

**Richard McCarter** - Multimedia courseware developer

**David Jackson** - Multimedia courseware developer

The production of this resource has been a truly collective effort. Colleagues have been extremely generous with their time and commitment to this project. My thanks to all of them:

**University of Birmingham**

* Andrew Harrison
* Helen Howard
* Darren Marsh
* David Mills
* Anthony Rosie

**University of Southampton**

* Julia Waldman

**Sheffield Hallam University**

* Peter Ashworth
* Phil Bannister
* Jenny Blain
* Kevin Bonnett
* Sue Clegg
* Christopher Crowther-Dowey
* Iain Garner
* Anne Hollows
* Alison Hudson
* Elizabeth Lawrence
* Richard McCarter
* Alan McGauley
* Shawna McCoy
* Anne Macaskill
* Richard Mather
* Grant Naylor
* Karen Smith
* Gary Taylor
* Sarah Ward

**Students who took part in development and evaluation of this website:**

* Rachel Eames
* Joanna Melvin
* Robin Waldman
* Elizabeth Washington
* Jemma Watchman

**Other Institutions**

* Paul Bridgen, University of Southampton
* Sue Hemmings, The Open University
* Janet Morton, University of Leeds
* Andy Pilkington, University College Northampton
* Beverley Searle, University of Durham
* John Steel, University of Sheffield
* Faith Tucker, University College Northampton
* Christopher Winch, University College Northampton
* Martin Wolf, University of Warwick