Exam Revision Guide

Analysing Data 2020

03/04/2020

## What You Need to Know

To be successful on the exam, you will need to be familiar with all of the vocabulary, concepts, and analyses listed below. The exam will draw on both the **lectures** and the **tutorials** from the entire module. The practicals are also listed when they provide additional explanations/practice for the same material, but there will be no questions drawn from the practicals only that are not also covered in the lectures/tutorials. As a general rule, if you were well-prepared for each of the weekly quizzes, you will be well-prepared for the exam.

## Code on the Exam

#### Included

The exam will include two specific cases where you must demonstrate knowledge of R.

First, you must be able to **read and interpret the lm() function**. This is the only function we will ask you about on the exam. You must know:

* How the function works (i.e. what arguments it takes)
* How to correctly write a formula as the input for lm()
* How the formula in lm() corresponds to the elements of the linear model

Second, you must be able to **read and interpret R output** for all of the analyses that we have covered in this module, as detailed in the [list of topics](#topics-to-revise). Aside from lm() mentioned above, the exam will **not** ask you how to produce these analyses using R.

#### NOT Included

Aside from these specific cases, the exam will not include questions about writing or reading R code. You will not be expected to memorise R code verbatim, and the exam includes no questions about remembering specific functions, proofreading code, etc.

## Revision Tips

### Practice definitions

The glossary below is provided to help you get started on your revision. Fill in the box below each concept or term with its definition and everything you know about it. Then, as you revise, you can correct or add to this information until you have a detailed guide to the topics on this module.

### Test yourself

You should test your knowledge regularly as you are revising. You can do this however you find most useful. For instance, write out flashcards by hand, or [make them online](https://quizlet.com/en-gb). Use the datasets we’ve looked at this term to run analyses in R and practice interpreting the output. Quiz your friends using your glossary.

### Short and frequent

Just like with learning R, small amounts of regular practice will help you consolidate what you’re learning and avoid fatigue. Try to focus on just one topic or idea for a little while (20-30 minutes), then quiz yourself. Don’t try to cram it all in at once!

### Use feedback

Look through your quizzes several times as you are revising. When you begin, use them to identify things you don’t know, or find confusing, and focus on those topics. Then, once you’ve done some revision, look through the questions and answers again and see if you understand them better.

### Ask questions

All of the lecturers on this module will be happy to explain in more depth if you’re confused. Post your questions on [the Discussion board](https://canvas.sussex.ac.uk/courses/9242/discussion_topics) and we’ll get back to you as soon as we can.

## Topics to Revise

The following list of topics are drawn from the lectures, tutorials, and practicals throughout the term. To be well prepared for the exam, you should revise all of the following.

### Scientific Method and Study Design

Resources: [Week 1](https://canvas.sussex.ac.uk/courses/9242/pages/week-1) lecture

* The Scientific Method
* Predictions and hypotheses
* Reliability, validity, and replication
* Establishing causality

### Samples, Populations, and Distributions

Resources: [Week 2](https://canvas.sussex.ac.uk/courses/9242/pages/week-2) lecture, [Week 4](https://canvas.sussex.ac.uk/courses/9242/pages/week-4) tutorial

* What a distribution is, and how to interpret graphs of distributions
* The properties and characteristics of:
  + Normal distributions
  + Sampling distributions
  + Means, standard deviations, and standard errors
  + Scaling and centring
* The central limit theorem and how the sampling distribution of the mean is related to its population distribution

### Estimates and Confidence Intervals

Resources: [Week 3](https://canvas.sussex.ac.uk/courses/9242/pages/week-3) lecture and [Week 4](https://canvas.sussex.ac.uk/courses/9242/pages/week-4) tutorial

* Point and interval estimates
* What confidence intervals are and how they are interpreted
* How confidence intervals are constructed
* The characteristics of the sampling distribution and the *t*-distribution

### Testing Hypotheses

Resources: [Week 4](https://canvas.sussex.ac.uk/courses/9242/pages/week-4) lecture

* Types of hypotheses, including their definitions and characteristics
* The definitions and characteristics of the null and alternative hypotheses
* The definition of the *p*-value, where it comes from, and what it does and does not tell you
* How to use *p*-values and significance levels to assess hypotheses

### Measures of Association

Resources: [Week 5](https://canvas.sussex.ac.uk/courses/9242/pages/week-5) lecture, [Week 6](https://canvas.sussex.ac.uk/courses/9242/pages/week-6) tutorial and practical

For both correlation and chi-squared tests:

* What kind of data each test typically requires
* What the null hypothesis of each test is
* How to read tables and graphs of the data
* How to read and interpret output from rcorr(), cor.test(), and chisq.test()
* How to correctly report the results of these tests

### Comparing Two Means

Resources: [Week 6](https://canvas.sussex.ac.uk/courses/9242/pages/week-6) lecture, [Week 7](https://canvas.sussex.ac.uk/courses/9242/pages/week-7) practical and tutorial

Revise how the t-test works, including:

* What kind of design it’s appropriate for
* What the null and alternative hypotheses of this test are
* How to interpret graphs of the data
* What *t* represents and how to interpret it
* How *t* and *r* are related, as described in the lecture
* How to read and interpret output from t.test() and lm()
* How to correctly report the results from t.test()

### The Linear Model

Resources: [Week 8](https://canvas.sussex.ac.uk/courses/9242/pages/week-8) and [Week 9](https://canvas.sussex.ac.uk/courses/9242/pages/week-9) lectures; [Week 9](https://canvas.sussex.ac.uk/courses/9242/pages/week-9) and [Week 10](https://canvas.sussex.ac.uk/courses/9242/pages/week-10) tutorial and practicals

* The general model equation
* The linear model equation, with one or multiple predictors
* What each element of the linear model equation means, especially beta values for each predictor
* How to use the linear model to make predictions
* How to create and interpret a linear model in R, including:
  + Specifying the correct formula and using the lm() function
  + The beta values associated with each predictor and the intercept, standardised and unstandardised
  + *t* and its associated *p* (for each predictor)
  + R2, adjusted R2, *F*-ratio and its associated *p* (for model fit)
  + R2 change, *F*-change and its associated *p* (for comparing models)

### Effect Sizes and Conclusions

Resources: [Week 10](https://canvas.sussex.ac.uk/courses/9242/pages/week-10) lecture; [Week 11](https://canvas.sussex.ac.uk/courses/9242/pages/week-11) tutorial and practical

* What an effect size is and why they are useful
* How to use and interpret the correlation coefficient *r* as an effect size
* How to use and interpret Cohen’s *d*, including:
  + How to calculate *d*
  + What constitutes small, medium, and large values
* The difference between *p*-values and effect sizes

## Glossary

To get started on your revision, try writing down everything you know about the following terms and concepts. Then, check your memory against the lectures and tutorials. If you’re correct, well done! If not, revise the materials, or ask for clarification on the Discussion board.

* Steps of the scientific method
* Reliability
* Validity
* Replication
* Replication crisis
* How causality is established
* Distribution
* Normal distribution
* Sampling distribution
* Mean
* Standard deviation
* Standard error
* Confidence interval
* Scaling
* Centring
* The Central Limit Theorem
* Point estimate
* Interval estimate
* *t*-distribution
* Degrees of freedom
* Conceptual hypothesis
* Operational hypothesis
* Statistical hypothesis
* Null hypothesis
* Alternative hypothesis
* Significance (alpha) level
* *p*-value
* Correlation coefficient *r*
* Scatterplot
* Chi-squared test
* Bar plot
* *t*-test
* Means plot
* Error bars
* General model equation
* Linear model equation
* Linear model with a categorical predictor
* Linear model with a continuous predictor
* *b*0
* *b*1
* Unit change
* Standardised beta
* R2
* Adjusted R2
* *F*-ratio
* *F*-change
* R2 change
* Signal-to-noise ratio
* Effect size
* Cohen’s *d*
* Statistical significance
* Standarisation
* Randomisation

## Good luck!