

ETC1010: Introduction to Data Analysis

Week 12

Notes on the final Exam

Lecturer: *Nicholas Tierney*

Department of Econometrics and Business Statistics

✉ nicholas.tierney@monash.edu

June 2020



Well done on your projects!

We did it!

What a semester!

Thank you all for your patience, it's been a hard semester for all of us.

It has been an absolute pleasure to teach you all this semester

You have such a great group of students to work with!

We hope that what we've covered can be useful for you and that you *continue to practice these skills*.

And maybe use these skills in other classes where you are doing data analysis.

Special thanks to the super tutor dream team

Nitika

Sarah

Sherry

Steff

Exam details

- Worth 50% of your final grade
- Delivered online on moodle (short practice example will be made available soon)
- MCQ, TRUE/FALSE, Fill in blanks, short answer.
- **Hurdle requirement:** You must get 40% on the exam to pass the course
- Covers entire span of course except guest lecture

Exam Details

- I'll now talk about the questions in the exam, and some of the concepts you need to be familiar with
- These concepts will help guide what you focus on in the lectures
- Disclaimer: This list is **absolutely not exhaustive** - these are to help give you a sense of what I'm thinking about for each of the questions in the exam.

Tidy data Concepts:

Defining and identifying:

- Variables
- Observations
- Values
- Tidy data

Data Wrangling Concepts:

- Converting "messy" data to tidy data
- Code / key functions to use to convert data into "tidy" data
 - e.g, `pivot_longer`, `pivot_wider`, `separate`, etc.
- Computing summaries using verbs of `dplyr`, `mutate`, `select`, `summarise`, etc.
- Data formats (CSV, HTML, JSON)

Relational data concepts:

- Why do joins of data
- When to do certain types of join
- Predict output of a join
- Sketching out code to summarise data from a join

Data visualisation concepts:

- How the grammar of graphics produces a plot - identify plots produced by which code
- Understand the focus of a given graphic on the data - questions like:
 - "What and how does this graphic make us focus on what feature of the data?"
 - "What do you learn from a graphic?"
- Interpreting a graphic
- Uses of colour
- Hierarchies of data vis

Temporal data

- Extracting and cleaning time information

Workflow

- Filepaths
- The here package
- How data is read into R from certain files in a directory

Missing Values concepts:

- Principles of tidy missing data
- Interpreting graphics of missing data
- Predict output of function on data with missing values
- Imputation - what is it, which methods are good / bad / better

Linear Models Concepts:

- Write down an equation of a model from code output
- The formula $y = 3x + 5$ is a function with input x , and output y , when x is ___ , the output is ___

Linear Models Concepts:

$$\widehat{height}_{in} \sim 3.62 + 0.78Width_{in}$$

- **slope:** For each additional inch the painting is wider, the height is expected to be higher, on average, by 0.78 inches.
- **Intercept:** Paintings that are 0 inches wide are expected to be 3.62 inches high, on average.
- If a painting is 5 inches wide, what is its estimated height?

Linear Models Concepts:

- How to make predictions from a fitted model
- Understand what makes predictions good and bad
- Measurements of model fit:
 - R^2 - what it is - what values mean good/bad fit?
 - **A good idea to bring in an equation of R^2**
- Residuals - What do we expect to see
- Centering variables
- Think about how you can improve fit of models to your data

Programming concepts:

- Why write functions
- How to write a function in R
- How to take existing code and turn it into a function
- Identify potential mistakes in provided code
- Understand what map does

Networks Concepts:

- From an association / correlation matrix, which are most or least related?
- Understand how you can convert a numeric matrix into a binary association matrix
- Understand how a correlation matrix (or other association) of data relates to a provided network

Remember: Not an exhaustive list

These concepts guide what to focus on.

The readings **provide great information that will certainly help improve your understanding**

Remember that these are to help you focus, I can't give you the exam, but I can help tell you what is important.

Disclaimer: This list is not exhaustive - these are to help give you a sense of what I'm thinking about for each of the questions in the exam.

How to study for the exam.

One approach, the Feynman technique, which boils down to:

If you want to understand something well, try to explain it simply.

The Feynman Technique

1. Write the name of the concept at the top of a blank piece of paper.
2. Explain the concept as if you were teaching it to someone else
 - In writing
 - Talking out aloud to a room
 - Talking to a person IRL / zoom
3. Identify knowledge gaps - loop back to your explanation and expand.
4. Challenge yourself to reduce the complexity / jargon of the language

Feynman Technique

Some resources on Feynman technique:

- [Video - 2 minutes](#)
- [Video - 6 minutes](#)
- [Blog post](#)

How to study for the exam.

- When taking practice exam, make sure you don't have the answers / don't check them immediately
- Work through examples and exam on an answer until you are confident you've given it your best go
- Write your own exam questions and share with friends

Exam technique

- Peruse (read through carefully) the entire exam before starting
- Rank questions in terms of difficulty for you
- Complete easy questions first
- Make sure you've read the entire exam before starting. Your brain starts ticking away and working in the background.

What Now?

Join the R community



image from Alison Horst

What Now?

- Consider writing up your project as a blog post using [blogdown](#) - See [Alison Hill's blogpost on creating your own blog](#)
- Consider sharing your project talk with the [Melbourne R user network \(MelbURN\)](#) or [R-Ladies Melbourne](#)
- Learning more about git and github with [happygitwithr](#)
- Join twitter and partake in discussions on #rstats

Major

- Where do you go from here, if you are a **business analytics major**
- Courses using R:
 - **core**
 - ETC2420: randomization and simulation to understand uncertainty, and a little about Bayesian models
 - ETC3250: data mining, computationally intensive approach to fitting models
 - **electives**
 - ETC3555, ETC3580: advanced statistical models, and advanced machine learning methods
 - ETC3550, ETF3500, ETX2250: forecasting and multivariate analysis, data analytics

Masters

Monash is launching a new masters program, the Masters of Business Analytics.

You can learn more about the new masters program at this link:

<https://www.monash.edu/business/master-of-business-analytics>

(check out the video too, there are some familiar faces there!)

Assessment marks

- Assignment 2 marks are finalised, you will have until June 22 to propose changes
- ED assessment and remaining marks will be uploaded and shared with you via moodle

Final Exam

For exam prep, the week before the exam, we will have **three consultation times**. (these will be posted on the course site)

Final Exam

The exam from last year is posted.

Not all topics are the same this year, and this should be clear when you look at the questions.

Music from the semester

Normally I play music during the lecture at the start of the lecture, and during group activities.

If you want to listen to any of the music played this semester you can search for "ETC1010" on spotify, or click on these links:

[ETC1010 Openers](#)

[ETC1010 Closers](#)

Course evaluation

Please complete the course evaluation.

This is a new version, new material, new format. The course will evolve, and your help in improving it is greatly appreciated.