EKT-813: Quantitative Methods
Semester I, 2020

Instructors: Steve Koch (**steve.koch@up.ac.za**) and Jesse Naidoo (**jesse.naidoo@up.ac.za**)
Office Hours: TBA

The point of this course is to prepare you to do empirical work of your own. By itself, this course is not enough to accomplish that goal, but it will give you certain foundational skills: (i) manipulating, summarizing and visualizing data; and (ii) understanding the basics of probability and linear algebra well enough to learn econometrics. We will assume you know calculus, though.

Lectures are from 4.30 PM - 7.30 PM on Monday evenings, in Tukkiewerf 1-37.

Problem Sets: We will post six problem sets (about one every two weeks). ***These are not for credit.*** We will also provide suggested solutions.

Grades will be determined by a weighted average of your scores on the final exam and the midterm. The respective weights will be either (40% midterm, 60% final), or (25% midterm, 75% final) - whichever is in your favor.

Important Dates

Midterm: Monday, March 16th (in class).

Final Exam: TBA.

Textbooks

There is no prescribed text. Everything that we will teach you is available from several different sources (and indeed many of them are available for free online).

Some parts of the first quarter (data wrangling and general "data science" skills) are drawn from Grolemund & Wickham (2017) - which you can read for free online **here**.

Some parts of the second quarter (probability and linear algebra) will follow the treatment in Stachurski (2016). Stachurski has posted many lecture slides and other supporting materials to his **website**.

Other textbooks that you may find useful are

- *Mathematical Statistics with Applications* (7th edition), by Wackerly, Mendenhall, and Sheaffer
- *Mathematical Statistics and Data Analysis* (3rd edition), by John Rice.

Both books cover the same material.

We will partly base our lectures on these texts, but may present the content in a slightly different way in class. Study several different sources (the internet may be useful here) and - most importantly! - do the problems to make sure you understand.

Syllabus

Getting Oriented [1 lecture]

Understanding the source of the data. Data I/O. Data structures and type safety. Tidy data, keys, and duplicates. One-way and two-way tabulations. Getting summary statistics.

References: Grolemund & Wickham (2017) Ch. 4; Wickham (2014).

Data Manipulation [3 lectures]

Selecting and generating ("mutating") variables; filtering observations. Data aggregation (grouped summaries). Recoding variables. Reshaping data ("wide" and "long" formats). Merging (joining) multiple datasets together. Manipulating strings, factors and dates.

References: Grolemund & Wickham (2017) Ch. 3, 5; 12 - 16.

Simulations, Loops, and Functions [1 lecture]

Generating fake data. Iteration: for and while loops; vectorization ("split-apply-combine"). User-defined functions.

References: Grolemund & Wickham (2017) Ch. 17 - 21.

Workflow and Reproducibility [1 lecture]

Automation. Code style. Version control with git. Some principles for organizing your projects.

References: Grolemund & Wickham (2017) Ch. 6, 8; Wilson *et al.* (2017).

Basic Probability [3 lectures]

Axioms for counting; conditional probability and independence. Densities and CDFs. Examples of univariate distributions. Moments. Chebyshev's inequality. Joint and marginal distributions; conditional distributions and densities. Covariance and correlation; finding moments of linear combinations of random variables. Finding the distribution of functions of random variables; order statistics.

References: Stachurski (2016) Ch. 4 - 5; Rice Ch. 1 - 4; Wackerly et al., Ch. 2 - 6.

Large-Sample Approximations [1 lecture]

The classical (frequentist) paradigm. Weak law of large numbers. Central limit theorem. Simple random sampling and the standard error of the mean. Clustered and stratified random sampling.

References: Stachurski (2016) Ch. 6; Rice Ch. 5 , 7; Wackerly et al, Ch. 7, 9.3.

Frequentist Estimation and Inference [2 lectures]

Desirable properties of estimators: lack of bias, efficiency, consistency. Maximum likelihood: principles and some examples. Identification and precision. Sufficiency. Hypothesis testing. Type I and Type II errors; test size and power. Confidence intervals.

References: Stachurski (2016) Ch. 8 - 10; Rice Ch. 8 - 9; Wackerly et al, Ch. 8 - 10.

Basic Linear Algebra [3 lectures]

Systems of linear equations; existence and uniqueness of solutions. Vector spaces, linear independence, basis and dimension. Matrices as linear maps; invertibility, rank, and determinants. Orthogonality. Partitioned matrices. Special types of matrices: symmetric, idempotent, positive and negative definite. Quadratic forms. Eigenvalues and eigenvectors; diagonalization. QR, spectral, and Cholesky decompositions.

References: Stachurski (2016) Ch. 2 - 3.

One class will be taken up by the midterm, one by a pre-midterm review session, and one by a pre-exam review session.

This version: January 29, 2020

References

Grolemund, Garrett, & Wickham, Hadley. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly.

Stachurski, John. 2016. *A Primer in Econometric Theory.* Cambridge, MA: MIT Press.

Wickham, Hadley. 2014. Tidy Data. *Journal of Statistical Software*, **59**(10).

Wilson, Greg, Bryan, Jennifer, Cranston, Karen, Kitzes, Justin, Nederbragt, Lex, & Teal, Tracy K. 2017. Good enough practices in scientific computing. *PLOS Computational Biology*, **13**(6), e1005510.