

# Regression

AU STAT-615

Emil Hvitfeldt

2021-1-20

# Welcome

- Introductions
- Syllabus
- Material
- Questions

# Questions

You are encouraged to ask questions when you have them rather than wait for me to ask for questions. If you have a question, chances are that something else has a question.

# Attendance and Camera

Both highly encouraged

Due to COVID-19 both should be practiced according to what makes you safe

# About Me

- Data Analyst at Teladoc Health
- R package developer, about 10 packages on CRAN (textrecipes, themis, paletteer, prismatic, textdata)
- Co-author of "Supervised Machine Learning for Text Analysis in R" to be published soon

# Regression

American University 615

---

This website contains most of the information and material that will be used for one of the sections of the course [Regression](#) at [American University](#).

The navigation bar contains information about the [syllabus](#), [schedule](#), [readings](#), [labs](#) and [assignments](#).

## License

---



This online work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International](#). Visit [here](#) for more information about the license.

## Colophon

---

This book was written in [RStudio](#) using [distill](#). The complete source is available on [GitHub](#). All packages versions are being handled with [renv](#).

# Syllabus

# **An Introduction to Statistical Learning with Applications in R**

Applied Linear Regression Models 4th edition or 5th edition

You can either buy the 4th or the 5th edition, BUT DO NOT BUY BOTH



# Syllabus

Come to me before it is too late

I'm here to help, my main goal for this course is to make you succeed

# Late assignment

There are some (limited) late penalties

It is more important for me that you turn something in then that you give up. You will always get points (sometimes reduced) for late assignments

Contact me if you are having a hard time or need to turn in late

# Lecture

40 + 15 + 40 + 15 + 40

# Labs

A hands-on section where we work together on the implementation side in R

These should be turned in WITH explanatory text.

# Assignments

There will be 10 assignments

It Will contain a mix of conceptual/statistical questions and practical coding exercises about the weekly topic

# Midterm

We will have a midterm halfway through the course

Will be a multi-day "take-home" test

The Midterm will cover multiple weeks of material

# Final Project

The project will be a document(20%) and a presentation given to the class(5%)

# Slack

Discussion place and questions



# Material 1/2

- Introduction to R, notation, motivation, and examples. [1.1-1.2]
- Linear regression: model, estimation, inference, prediction. Regression and correlation. R2. [Chap. 1-2]
- Regression diagnostics: non-normality, nonlinearity, heteroscedasticity [Chap. 3]
- Simultaneous estimation. Other regression models [Chap. 4]
- Multiple regression. Matrix approach (Stat-615). Analysis of variance. Analysis of residuals. Partial correlation and multiple correlation coefficient. [Chap. 5-6]

# Material 2/2

- Model building. Model selection and validation. Extra sum of squares. [Chap. 7-8]
- Regression diagnostics-II. Influential observations and outliers. Effect of multicollinearity. Robust regression. Ridge regression [Chap. 9-10]
- Regression diagnostics-III. Symptoms and remedies. Transformation of variables. Missing data. Analysis of covariance. Comparison of regression lines. [Chap. 10]
- Dummy variables and related methods [Chap. 11]
- (If time permits) Nonlinear relations. Logistic regression. [Chap. 13-14]