Stat 412/512

Welcome (Back!)

Jan 5 2015

Charlotte Wickham

stat512.cwick.co.nz

Hello (again), my name is

Charlotte

Your TAs are:
Ben Brintz
Katie Eng
Joe Maurer

stat512.cwick.co.nz

Use canvas!
https://oregonstate.instructure.com/

Textbook: The Statistical Sleuth by F. Ramsey and D. Shafer, 3rd Ed. The 2nd Edition is also fine.

Assessment

Homework × 6	30%
Data Analyses × 2	20%
Quizzes × 3	20%
Final Exam	20%
Participation	5%
Regression in your field	5%

tentative schedule posted online

Homeworks & Labs

Labs are on Wednesdays.

Homeworks/DA **released** Weds @ noon on class webpage

Homeworks/DA **due** Weds @ noon on canvas.

Quizzes

First quiz scheduled for Fri Jan 23rd **in class**

The remaining quizzes will be online like in ST411/511.

Participation

Form a study group, between 3 and 6 people.

Meet at least three times during the quarter.

Maintain a study journal that records the dates you met and topics you discussed.

Each member of the group will submit this journal on canvas before the final exam.

Regression in your field

A regression in your field assignment will be provided in week 4 and due (tentatively) in week 7.

It will involve a short (less than one page) written report on a journal article you select.

Academic Integrity

If I say something is “individual” I mean it.

Identical (or almost identical) homeworks or data analyses will result immediately in an **Academic Dishonesty Report**

Read the syllabus for other examples.

It’s disrespectful to me, to your fellow students, and it devalues the worth of an Oregon State University degree.

If something outside school is impacting your ability to complete your work, talk to me.

Didn’t take ST411/511 with me

http://stat511.cwick.co.nz

Get RStudio (if you haven’t used it before)

http://stat512.cwick.co.nz/installingRlocally.html

Get friendly with someone in lab who can explain “Compiling an RStudio notebook”

If you haven’t used ggplot2 before, have a look at:

http://stat511.cwick.co.nz/lectures/07-some-r.pdf

In general: my code is a one way of doing things. As long as you get things done, I don’t care what code you use.

Getting help

Ask a question on the canvas discussion board.

Ask a question in lab.

Go to your TAs office hours. TBA

Come to my office hours / help sessions.

TBA

Questions on syllabus?

What is multiple regression?

Simple linear regression

$\mu\{ Y \mid X\} = \beta_0 + \beta_1 X$

A single response as a linear function of a single explanatory variable.

Multiple linear regression

$\mu\{ Y \mid X_1, X_2, X_3, ...\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ...$

A single response as a linear function of a many explanatory variables.

+ some assumptions

Multiple linear regression is a fundamental method in statistics

If you understand multiple linear regression, lot's of other statistical models/methods are just extensions:

Mixed models/Hierarchical models: β's aren't fixed numbers but have distributions.
Quantile regression: change RHS to quantiles (instead of mean)
Generalized linear models (poisson regression, logistic regression): subpopulations aren't Normal, you model a parameter of their distribution (that may not be the mean)
Lasso and ridge regression: estimate the β's in a way that penalizes them for being big
Generalized additive models: non-linear LHS + penalization

And some of the things we saw in ST511 are special cases:

two sample t-test
models in one-way ANOVA
simple linear regression

Broad overview of ST512

Statistical Sleuth:

Chapters 9 & 10. What kind of questions can be answered with multiple regression, and how do you answer them?

Chapter 11. Checking & refining regression models

Chapter 13 & 14. Two way classifications (special cases of multiple regression)

Chapter 12. Model selection (notice this is out of order)

Chapter 15 & 16. Serial correlation and repeated measures (dealing with violations of independence)

Wednesday

Review

Simple linear regression:

- model for the mean
- interpreting intercept and slope
- assumptions and residuals
- R output

Types of statistical inference

The **response** variable is the measurement we are interested in explaining or predicting.

Y

The **explanatory** variable is the measurement we want to use to explain or predict the response.

X

The **simple linear regression** model

Parameters

$$\mu\{Y|X\} = \beta_0 + \beta_1 X$$

Intercept Slope

The mean response as a function of the explanatory variable is a straight line.

Describes the relationship between the response and explanatory variable with **two** parameters.

Intercept and Slope

The **intercept** gives the mean response at an explanatory value of zero.

The **slope** gives the **change in the mean response** for a **1 unit change** in the explanatory variable.