

Statistics for Genomic Data Science

Instructor: [Jeff Leek](#)

Class Website:

- [Statistics for Genomic Data Science](#)
- [Genomic Data Science Specialization](#)

Resources:

- [Installing R for Windows](#)
- [Installing R for Mac](#)
- [Installing RStudio Mac](#)
- [Rstudio, R project, and Bioconductor](#)
- [Rstudio's cheatsheets](#)
- [PH525x Statistics for Genomics Book](#)

Course Materials

Week/Lecture	Lecture	Video	Notes	Code
0/1	Welcome		Google Slides pdf	
0/2	What is statistics?		Google Slides pdf	
0/3	Finding statistics you can trust		Google Slides pdf	
0/4	Getting help		Google Slides pdf	
0/5	What is data?		Google Slides pdf	
0/6	Representing data		Google Slides pdf	
—	—	—	—	—
1/1	Week 1 Introduction		Google Slides pdf	
1/2	Reproducible research		Google Slides pdf	
1/3	Achieving reproducible research		Google Slides pdf	NA
1/4	R markdown		html	R markdown R code
1/5	The three tables in genomics		Google Slides pdf	
1/6	The three tables in genomics (in R)		html	R markdown R code

Week/Lecture	Lecture	Video	Notes	Code
1/7	Experimental Design: variability, replication, and power		Google Slides pdf	NA
1/8	Experimental Design: confounding and randomization		Google Slides pdf	
1/9	Exploratory Analysis		Google Slides pdf	
1/10	Exploratory Analysis in R		html	R markdown R code
1/11	Data transforms		html	R markdown R code
1/12	Clustering		Google Slides pdf	
1/13	Clustering in R		html	R markdown R code
—	—	—	—	—
2/1	Week 2 Introduction		Google Slides pdf	
2/2	Dimension reduction		Google Slides pdf	
2/3	Dimension reduction (in R)		html	R markdown R code
2/4	Pre-processing and normalization		Google Slides pdf	
2/5	Quantile normalization (in R)		html	R markdown R code
2/6	The linear model		Google Slides pdf	html R markdown R code
2/7	Linear models with categorical covariates		Google Slides pdf	
2/8	Adjusting for covariates		Google Slides pdf	html R markdown R code
2/9	Linear regression in R		html	R markdown R code
2/10	Many regressions at once		Google Slides pdf	
2/11	Many regressions in R		html	R markdown R code
2/12	Batch effects and confounders		Google Slides pdf	
2/13	Batch effects in R		html	R markdown R code
—	—	—	—	—

Week/Lecture	Lecture	Video	Notes	Code
3/1	Week 3 Introduction		Google Slides pdf	
3/2	Logistic regression		Google Slides pdf	
3/3	Regression for counts		Google Slides pdf	
3/4	GLMs in R		html	R markdown R code
3/5	Inference		Google Slides pdf	
3/6	Null and alternative hypotheses		Google Slides pdf	
3/7	Calculating statistics		Google Slides pdf	
3/8	Comparing models		Google Slides pdf	html R markdown R code
3/9	Calculating statistics in R		html	R markdown R code
3/10	Permutation		Google Slides pdf	
3/11	Permutation in R		html	R markdown R code
3/12	P-values		Google Slides pdf	
3/13	Multiple testing		Google Slides pdf	
3/14	P-values and multiple testing in R		html	R markdown R code
—	—	—	—	—
4/1	Week 4 Introduction		Google Slides pdf	
4/2	Gene set analysis		Google Slides pdf	
4/3	More enrichment		Google Slides pdf	
4/4	Gene set analysis in R		html	R markdown R code
4/5	The process for RNA-seq		Google Slides pdf	
4/6	The process for Chip-Seq		Google Slides pdf	

Week/Lecture	Lecture	Video	Notes	Code
4/7	The process for DNA methylation		Google Slides pdf	
4/8	The process for GWAS/WGS		Google Slides pdf	
4/9	Combining data types (eQTL)		Google Slides pdf	
4/10	eQTL in R		html	R markdown R code
4/11	Researcher degrees of freedom		Google Slides pdf	Interesting example
4/12	Inference vs. prediction		Google Slides pdf	
4/13	Knowing when to get help		Google Slides pdf	R markdown
4/14	Course Wrap-up		Google Slides pdf	
—	—	—	—	—

Course R package

You can get all of the code used in the class by installing the R package:

```
source("http://bioconductor.org/biocLite.R")
biocLite("devtools") # only if devtools not yet installed
biocLite("jtleek/genstats", ref="gh-pages")
```

You can see the list of lecture notes and open them using the vignette command:

```
library(genstats)
vignette(package="genstats")
vignette("01_13_clustering")
```

Miscellaneous

Feel free to submit typos/errors/etc via the github repository associated with the class: https://github.com/jtleek/genstats_site

This web-page is modified from Andrew Jaffe's [Summer 2015 R course](#), which also has great material if you want to learn R.

This page was last updated on 2015-09-07 07:37:18 Eastern Time.