

```
## Error in opts_chunk$set(cache = TRUE, eval = TRUE,  
include = TRUE, echo = TRUE, : argument is missing, with  
no default
```

Welcome!

Objectives: Linear statistical models are the foundation for most of applied statistics. We will develop statistical computation skills (R programming) and mathematical skills (working with matrices) while studying data analysis using linear models.

Pre-requisites: We will assume familiarity with material in STATS 250. All course notes and labs are at

open.umich.edu/find/open-educational-resources/statistics

If you have a different background (AP Statistics, STATS 280, or some other introductory statistics class) you should check the STATS 250 notes and if necessary come for help in office hours.

Let's get started

We will work through a data analysis using a linear model, and then study the math and stats so that (i) we can command the computer to generate what we want; (ii) we can interpret what the computer tells us.

- Obtain the data from the internet
- Install R (www.r-project.org) and Rstudio (www.r-project.org)
- Read the data into R
- Plot the data
- Develop a model
- Estimate parameters and test hypotheses of interest
- Interpret the results

The two rising stars in statistical computing are R and Python (<http://r4stats.com/articles/popularity/>). Generally, R is preferred for data analysis, and Python for larger programming projects.

Let's get started

We will work through a data analysis using a linear model, and then study the math and stats so that (i) we can command the computer to generate what we want; (ii) we can interpret what the computer tells us.

- Obtain the data from the internet
- Install R (www.r-project.org) and Rstudio (www.r-project.org)
- Read the data into R
- Plot the data
- Develop a model
- Estimate parameters and test hypotheses of interest
- Interpret the results

The two rising stars in statistical computing are R and Python (<http://r4stats.com/articles/popularity/>). Generally, R is preferred for data analysis, and Python for larger programming projects.

We live in an era of abundant data. Learn R!

Case study: Are people healthier in booms or busts?

- Is population health **pro-cyclical** (improving in business cycle booms) or **counter-cyclical** (improving in recessions), or neither?
- **Life expectancy at birth** combines instantaneous death rates at all ages and is a basic measure of current population health.
- USA data for 1933–2015 are in the file `life_expectancy.txt` on the course GitHub repository `github.com/ionides/401w18/01` or the website `ionides.github.io/401w18/01`.

```
# The United States of America, Life expectancy at birth.  
# Downloaded from Human Mortality Database on 30 Oct 2017.  
# HMD request that you register at http://www.mortality.org  
# if you use these data for research purposes.
```

Year	Female	Male	Total
1933	62.78	59.17	60.88
1934	62.34	58.34	60.23

- Note: `#` denotes a comment in R, so the first four text lines will be ignored when we read in the data.

Read the data into R and then inspect it

```
L <- read.table(file="life_expectancy.txt",header=TRUE)
```

Question: Why should we prefer to use the command line form of R rather than a menu option, say in R Commander?

Now, let's check on the data. To see the first three rows,

```
L[1:3,]
```

```
##   Year Female  Male Total
## 1 1933  62.78 59.17 60.88
## 2 1934  62.34 58.34 60.23
## 3 1935  63.04 58.96 60.89
```

Here, we're using **matrix indexing**. $L[i, j]$ is the row i column j entry of L . Also, $1:3$ is the sequence 1,2,3 and the blank space after the comma in $L[1:3,]$ requests all the rows for the specified columns.

