Lecture 2

William Cipolli

Colgate University

September 3, 2019

Homework Completed for Today

- Check out the Moodle Page
- Complete Diagnostic 0 (by 8am)
- Install R, LaTeX, and RStudio
- Read "Statistics at Liberal Arts Colleges"

Homework for September 10-17

- Week 2
  - Diagnostic 1 will be posted after class Thursdays 9/5 – due by noon 9/9
  - Listen to the posted NPR podcast – due for class 9/10
- Week 3
  - Diagnostic 2 will be posted after class Thursdays 9/12 – due by noon 9/16
  - Complete Homework 0 (Posted) – due 11:59p 9/17

Q: What stuck out to you in this reading?

My Takeaways – Statistics and the Liberal Arts

- This paper was written in 1989, but it (startlingly) feels current.
- Liberal arts -> good statistics graduate students
- Statistics is by nature interdisciplinary
- Liberal arts and statistics are distinguished by their "concern for the general over the specific",
- "a liberal arts education trains a person to understand the world better" – so does statistics

My Takeaways – What is Statistics?

- **Statistics**, as it is concerned with gathering, organizing, and analyzing data, and with inferring from these data to the underlying reality, is a powerful intellectual method that can be applied in many contexts. In academia, statistics is a part of the curriculum in psychology, sociology, biology, and economics, to name but a few disciplines."[pg 80]
- H.G. Wells anticipated that statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write. We believe this day has come. [page 81] – if it hadn't then, it certainly has.
- Statistics is not mathematics.

My Takeaways – Working with Data

- Preparing a good data set for class takes work (this is actually the worst – you'll see)
- One must find raw data and take the time to understand them
  - We need to understand the question
  - We need to understand the data
  - We need to use the data to build a model that answers the question (this can often happen in a few ways)
- This all involves discussion and effective communication with collaborators – statisticians often need to be a "Jack of all trades"

My Takeaways – Section 4

We don't have to talk about section 4, but if you ever wondered what it's like being me that's it... well, except for the whole phone budget thing. Instead, we started the Data Science Collaboratory at Colgate.

Basic Operations

Operator	Functionality	Example From Notes
$=$	sets an object equal to a value	a=4+5
$<-$	works as an equal sign	a<-4+5
ls()	lists objects in the session	ls()
rm()	removes object(s) from the session	rm("a")

Basic Data Types

Data Type	Description
character	any character or word (string of characters) input
numeric	any real number (integer or double)
logical	a binary datatype with possible values TRUE and FALSE
complex	complex numbers with real and imaginary parts

Classes of Data Types

Data Type	Description
factor	data stored as integers, and have character labels associated with these unique integers.
integer	a whole number (integer)
double	a floating point number (includes integers)

Logical Operators

Operator	Functionality
$x < y$	is x less than y
$x <= y$	is x less than or equal to y
$x > y$	is x greater than y
$x > y$	is x greater than or equal to y
$x == y$	is x equal to y
$x! = y$	is x not equal to y
$x\%in\%y$	is x an element of y

Logical Operators

Operator	Functionality
cond1&cond2	logical and
cond1&&cond2	logical and (in sequence)
cond1|cond2	logical or
cond1||cond2	logical or (in sequence)
!cond	logical not
is.na(x)	is x NA – a missing value, (or nan)
is.nan(x)	is x not a number – i.e., 0/0
is.null(x)	is x NULL –an empty object

Creating Vectors

Operator	Functionality
c(...)	creates a vector of elements
seq(from=x,to=y,by=z)	creates vector from x to y
x:y	equivalent to seq(from=x,to=y,by=1)
rep(x,y)	creates a vector of x repeated y times
order(x)	returns the positions of elements in x in ascending order unless decreasing is set to TRUE

Describing Vectors

Operator	Functionality
length(x)	returns the number of elements in x
x%*%y	returns the dot product of x and y
head(x)	returns a preview of the beginning of x
unique(x)	returns the unique elements in x
summary(x)	returns a summary of x
table(x)	returns a table of frequencies for each observation in x

Subsets of Vectors

Operator	Functionality
which(*logical*)	return the position of the elements in the given *logical* statement is true
subset(*x*,*logical*)	return the elements of *x* where the given *logical* statement is true
any(*logicals*)	returns true if all values are TRUE
all(*logicals*)	returns true if any values are TRUE

Summarize Vectors

Operator	Functionality
min(x)	returns the minimum element in x
max(x)	returns the maximum element in x
sum(x)	returns the sum of all elements in x
cumsum(x)	returns the cumulative sum of x
mean(x)	returns the average value of elements in x
sd(x)	returns the standard deviation of elements in x
quantile(x,probs=y)	returns the y-th percentile of elements in x

Comparing Vectors

Operator	Functionality
intersect(x,y)	returns the common elements of x and y
union(x,y)	returns the elements of x and y combined
setdiff(x,y)	returns the values in x that are not in y

Lists

Operator	Functionality
list(...)	creates a list of elements

Next Time

- Data Frames
- Matrices