Section 5

Section in 30ish Second

▶ Cover some common mistakes in PS1 and PS2

▶ Questions

▶ Regression Practice

    ▶ Available on bCourses as "Section 5 Activity"

Common Mistakes

I only grade one submission for fairness reasons. Here are the primary reasons students have lost points so far:

▶ Forbidden Packages

▶ Not answering all parts of a question

▶ Incorrectly updating rows in a data frame

▶ Incorrect loops

Forbidden Packages

▶ Points have to be taken off for using functions from packages not allowed on the assignment. **Only use packages explicitly allowed on the assignment.**

   ▶ Functions from the following packages are always allowed `stats`, `graphics`, `grDevices`, `utils`, `datasets`, `methods`, `base`.

   ▶ Functions from `ggplot2` are always allowed unless noted otherwise.

▶ If you're not sure if you can use a function or a package, please ask me or post to Ed Discussion.

All Parts of the Question Need to be Answered

▶ I cannot give you points if there is no answer to a part of a question.

▶ Generally speaking, if a question says "Comment" or an a synonym of comment somewhere you need to write some English in addition to the math.

▶ If a question says "Report" you need to output an answer.

Updating Rows in a Data Frame

▶ The usual method is `data$col[i] = update`.

  ▶ `data[["col"]][i] = update` also works and is more general

▶ If you update a vector outside of a data frame, it will not update the data frame column.

▶ Generally speaking, we should always update the column the question asks us to update.

Updating Rows in a Data Frame

```
u = 1:6
v = 1:6
dat = data.frame(u,v)
head(dat)
```

```
  u v
1 1 1
2 2 2
3 3 3
4 4 4
5 5 5
6 6 6
```

```
v[1] = 10
print(v)
```

```
[1] 10  2  3  4  5  6
```

```
head(dat)
```

Incorrect Loops (Part 1)

▶ Not actually looping through data

```
t = c(1,2,3)
# length(t) == 3
for(i in length(t)){
  print(t)
}
```

```
[1] 1 2 3
```

▶ Trying to pass more than one value to a single element of a
vector

```
x = vector(mode = "logical", length = 1)
x[1] = c(1,2,3,4,5)
```

```
Warning in x[1] = c(1, 2, 3, 4, 5): number of items to rep]
multiple of replacement length
```

```
print(x)
```

Incorrect Loops (Part 2)

▶ Not sampling the entire vector with replacement

```
y = vector(mode = "logical", length = 1)
y = mean(sample(1:100, 5, replace = T))
print(y)
```

```
[1] 57.4
```

▶ Hardcoding a variable in the function

```
same_mean_3_times = function(z){
  w = vector(mode = "logical", length = 3)
  for(i in 1:3){
    w[i] = mean(sample(1:100, replace = F))
  }
  return(w)
}
same_mean_3_times(z = 1:5) ## want to get the output c(3,3,
```

```
[1] 50.5 50.5 50.5
```

Incorrect Loops (Part 3)

The following do not lose points but are strongly discouraged. See the Loop Speed handout on bCourses on Ed Discussion for why.

c()

```
v = c()
x = 1:100
for(i in 1:N){
  v = c(v, mean(x))
}
```

append()

```
v = c()
for(i in 1:100000){
  v = append(v, mean(sample(x)))
}
```

Things that will lose points on future assignments

▶ Printing the entirety of a data frame or vector that is large because it makes it hard for me to find your answers

  ▶ use `head(dat)` or `vec[1:6]`

▶ Leaving in the extraneous template code and words

  ▶ e.g. *"Add text/explanation…"*
  ▶ Make sure to delete this before submission

Questions from Lecture

▶ Note we've done multiplication on vectors in R like this

```
a = c(1,2,3)
b = c(2,3,4)
a*b
```

[1] 2 6 12

In Lecture it was mentioned that the inner product is

```
a%*%b
```

```
     [,1]
[1,]   20
```

The difference between these is that the first computes "element-wise" multiplication. The second in the inner product.

▶ Other questions about the course material?

Section 5 Activity

▶ The activity for today is on bCourses.

  ▶ It is drawn from your textbook with some minor modifications.

▶ If/When you finish the "Main" section, throw up your hand(s).

▶ The "Bonus" section provides some additional practice if you finish early.