Math 207
Introduction to Statistics

Introductory R Exercises

1. **Arithmetic Operations**. Please enter the following commands in R then write down the output that R produces. Explain to yourself what R thinks it is doing.

a) `2.0 + 3.1`

b) `2.0 * 3.1`

c) `3.0 ^ 2.0`

d) `9.0 ^ 0.5`

e) `sqrt(9.0)`

f) `exp(1)`

g) `sin(pi)`

h) `cos(pi/2)`

i) `tan(pi/2)`

2. **Vector Operations**. Try the following in R and explain to yourself what R thinks it is doing.

a) `x <- c(1, 0, -2)`

b) `x`

c) `y <- c(-1, 4.5, 3)`

d) `x + y`

e) `x + 2`

f) `x*y`

g) `x ^ y`

h) `exp(x)`

i) `1:10`

j) `seq(1, 21, by=5)`

3. **Statistics**. Try the following in `R` and explain to yourself what `R` thinks it is doing.

   a) `x <- seq(-5, 5)`

   b) `mean(x)`

   c) `sd(x)`

   d) `var(x)`

   e) `min(x)`

   f) `max(x)`

   g) `range(x)`

   h) `length(x)`

   i) `median(x)`

   j) `sum(x)`

   k) `prod(x)`

   ℓ) Repeat a)– k) using `x <- seq(0, 2, by=0.2)`. Compare your results.

4. **Subsets**. Try the following in `R` and explain to yourself what `R` thinks it is doing.

   a) `a <- seq(10, 100, by = 10)`

   b) `a`

   c) `a[1]`

   d) `a[5]`

   e) `a[0]`

   f) `a[c(2, 4, 6, 8)]`

   g) `a[-c(2, 4, 6, 8)]`

   h) `a[2:3] <- c(-333, -555)`

5. **Booleans**. Try the following in `R` and explain to yourself what `R` thinks it is doing.

a) `a <- seq(10, 100, by = 10)`

b) `a > 30`

c) `a[a > 30]`

d) `a == 50`

e) `a[a == 50]`

f) `a[a != 50]`

g) `a <= 30`

h) `a[a <= 30]`

i) `a[(a >= 50) & (a <= 90)]`

j) `a[(a < 30) | (a > 70)]`

k) `a[(a <= 30) & !(a < 20)]`

6. **Missing Values**. Try the following in `R` and explain to yourself what `R` thinks it is doing.

a) `x <- c(NA, 2, 3)`

b) `x`

c) `x + 8`

d) `x < 5`

e) `is.na(x)`

f) `mean(x)`

g) `mean(x, na.rm = TRUE)`

h) `x[is.na(x)] = 0`

i) `x`

j) `mean(x)`

7. **Experiments**

a) Enter `sample(10, 4)` several times. What is `R` doing?

b) Repeat a) using different values in place of 10 and 4. What happens if you enter `sample(10, 11)`?

c) Enter `sample(c("H", "T"), 5, replace=TRUE)`. What is `R` doing?

d) Use `R` to simulate flipping a coin 100 times. 1000 times.

8. **Graphics**

a) Enter the following and sketch the result.

```
x <- seq(-4,4,by=0.1)
plot(x, dnorm(x), type="l", cex.lab=1.5, cex.axis=1.5, xlim=c(-4,4), ylim=c(0,0.5))
```

b) Explain what the `xlim` and `ylim` parts of the command above mean.

c) Experiment to try to figure out what the `cex.lab` and `cex.axis` parts of the command do.

9. **Graphics II**. Do the following in `R`.

```
install.packages("ggplot2")  # You only need to install this once.
library(ggplot2)             # Do this every time you want to use the package.

?mpg
head(mpg)
summary(mpg)


p <- ggplot(mpg, aes(x = displ, y = hwy))
p <- p + geom_point() # add a plot layer with points
print(p)

p <- ggplot(mpg, aes(x = displ, y = hwy))
p <- p + geom_point(aes(colour = class))
print(p)

p <- ggplot(mpg, aes(x = displ, y = hwy))
p <- p + geom_point(aes(colour = class, size = cyl, shape = drv))
print(p)
```