# Workshop Handout

*Amanda Montoya*

*May 14, 2017*

**Practice 1**

1. Create a script file, include your name, the date, and "Practice 1" in the comments
2. Create V1, V2, and V3 (on next slide) as separate vectors. What data type is each vector?
3. Can you make V1, V2, and V3 into a matrix where each variable is a column? each variable is a row? Can you figure out how to make this into a dataframe?
4. Create a list with your first name, last name, last 4 digits of your phone number, and your favorite number.
5. Use the `seq` function to create a vector of the years from your birth to 2015 (if your birthday has not happened yet) and 2016 (if your birthday has already occured in 2017). What is the length of this vector? It should be your age.

---

| V1 | V2 | V3 |
|---|---|---|
| 19 | 4.0 | "A" |
| 20 | 3.6 | "A-" |
| 18 | 4.0 | "A" |
| 23 | 2.2 | "C" |
| 22 | 3.0 | "B" |
| 19 | 3.4 | "B+" |

**Practice 2**

All questions use egmat

```
shortvec <- seq(1:10)
egmat <- matrix(1:100, ncol = 5)
```

1. Create a logical matrix from egmat which indicates which elements are greater than 12 AND less than or equal to 8.
2. Select the elements from egmat which are NOT greater than 54 OR greater than 98.
3. Try out this code, what is it doing? What would happen if we reordered shortvec? Try it out.

```
egmat[egmat != shortvec]
```

```
##  [1]  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27
## [18]  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44
## [35]  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61
## [52]  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78
## [69]  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95
## [86]  96  97  98  99 100
```

4. Challenge: select all even values from `egmat` which are not multiples of 3.

### Using Dataframes

The first thing you'll need to do is read in a dataframe from some outside source. You can do your data maintenance in SPSS or Excel then same the dataset as txt, csv, etc.

```r
#Locate file on your computer
vaughndatabig <- read.csv(file = "C:\\Users\\Akmontoya\\Desktop\\UW Mediation in R\\Vaughn2017Study2.cs
#head(vaughndatabig)
#summary(vaughndatabig)

vaughndata <- read.csv(file = "C:\\Users\\Akmontoya\\Desktop\\UW Mediation in R\\Vaughn2017Study2Small.
#head(vaughndata)
#foot(vaughndata)
#names(vaughndata)
```

Head and foot give the first and last 6 rows of a dataset. Good to check that things read in reasonably.

```r
#str(vaughndata)

#summary(vaughndata)
```

### Linear Models

```r
vaughnlm <- lm(relativepromotion ~ nscond,
               data = vaughndata)
vaughnlm
```

```
##
## Call:
## lm(formula = relativepromotion ~ nscond, data = vaughndata)
##
## Coefficients:
## (Intercept)       nscond
##      -2.138        1.335
```

```r
summary(vaughnlm)
```

```
##
## Call:
## lm(formula = relativepromotion ~ nscond, data = vaughndata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1968 -0.8635 -0.1968  0.8032  5.4685
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.1379     0.2222  -9.621   <2e-16 ***
## nscond        1.3347     0.1401   9.525   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.571 on 501 degrees of freedom
## Multiple R-squared:  0.1533, Adjusted R-squared:  0.1516
## F-statistic: 90.73 on 1 and 501 DF,  p-value: < 2.2e-16
```

```r
#str(vaughnlm)
```

### Practice 3

1. Plot autonomy against relative promotion. Does it look like there might be some relationship between the two? Does this relationship look linear? Try using the `with()` function.
2. Make a new linear model predicting relative promotion from autonomy. Is there a significant relationship between autonomy and relative promotion? Write a sentence describing your results, like you might in a paper.
3. Save the fitted and residual values as part of the dataframe. Make sure to make new variable names, so you don't save over previous variables.
4. Plot the residuals in at least 3 different ways. Do you see any evidence of assumption violations?
5. Plot the original data for autonomy and relative promotion. Add the fitted line over this plot.
6. Challenge: Repeated this exercise but including nscond as a predictor. What additional plots might you make now that you have two predictors? Do both predictors significantly predict the outcome? Can you find a nice way to plot the relative promotion, condition, and autonomy all together?

### Moderation Code

```r
vaughnlm.gend <- lm(relativepromotion ~ DichGend*nscond, data = vaughndata)
summary(vaughnlm.gend)
```

```
##
## Call:
## lm(formula = relativepromotion ~ DichGend * nscond, data = vaughndata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3885 -0.7940 -0.0551  0.7416  5.3444
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.14673    0.68544  -0.214 0.830583
## DichGend        -1.33477    0.43713  -3.054 0.002384 **
## nscond          -0.06526    0.43238  -0.151 0.880097
## DichGend:nscond  0.93520    0.27563   3.393 0.000747 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.539 on 495 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.1741, Adjusted R-squared:  0.1691
## F-statistic: 34.79 on 3 and 495 DF,  p-value: < 2.2e-16
```

### Probing Interactions in R

```r
comp.mod.coef <- coef(vaughnlm3)
probe.contrast <- c(0, 1, 0, 3.3040)
```

```
theta.est.3.3 <- comp.mod.coef%*%probe.contrast
vcov(vaughnlm3)
```

```
##                    (Intercept)      nscond  competence nscond:competence
## (Intercept)          0.8554742 -0.63048052 -0.17984015        0.12474269
## nscond              -0.6304805  0.51798369  0.12474269       -0.09719396
## competence          -0.1798402  0.12474269  0.04177946       -0.02701433
## nscond:competence    0.1247427 -0.09719396 -0.02701433        0.01963177
```

```
se.theta.3.3 <- sqrt(probe.contrast%*%vcov(vaughnlm3)%*%probe.contrast)
t.theta.3.3 <- theta.est.3.3/se.theta.3.3
vaughnlm3$df.residual
```

```
## [1] 499
```

```
2*(1 - pt(t.theta.3.3, df = vaughnlm3$df.residual))
```

```
##            [,1]
## [1,] 0.02219823
```

### Practice 4

1. Data clean up: Make a new variable DichGender, based on the Gender variable. In the original variable 1 = male, 2 = female, 3 = Other, 4 = Prefer not to answer. In the new variable include all male and female respondents using 1 and 2 codes, recode 3 and 4 as missing.
2. Test if the effect condition depends on Gender.
3. Probe the effect of condition at reported gender values, and probe the effect of gender at the reported condition values.
4. Make some visualizations that clearly convey the results. If you can't get all the results on one plot, can you use the subset function to make two comparable plots?
5. Challenge: Can you repeat the above exercises but use the original Gender variable and include everyone instead?

### Mediation Estimation

```
vaughnlm.cpath <- lm(relativepromotion ~ nscond, data = vaughndata)
coef(vaughnlm.cpath)
```

```
## (Intercept)      nscond
##   -2.137922    1.334709
```

```
cpath <- coef(vaughnlm.cpath)[2]

vaughnlm.apath <- lm(autonomy ~ nscond, data = vaughndata)
coef(vaughnlm.apath)
```

```
## (Intercept)      nscond
##   0.9416406   2.3093634
```

```
apath <- coef(vaughnlm.apath)[2]

vaughnlm.bpath <- lm(relativepromotion ~ nscond+autonomy, data = vaughndata)
coef(vaughnlm.bpath)
```

```
## (Intercept)      nscond    autonomy
##  -2.6156877   0.1629932   0.5073761
```

```
cprimepath <- coef(vaughnlm.bpath)[2]
bpath <- coef(vaughnlm.bpath)[3]

ind.effect <- apath*bpath
```

#### Practice 5

1. Build a function which given the independent, mediator, and outcome variable calculates the indirect effect.
2. Building on your previous function add a part of the output which gives a TRUE/FALSE value for the Causal Steps Test and one for the Joint Significance test.
3. Building on your previous function calculate the standard error of the indirect effect, the Z-value, and p-value for the Sobel test. Check out the norm function (dnorm, pnorm, qnorm, rnorm)

$$Z = \frac{ab}{\sqrt{b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2}}$$

#### Practice 6

1. Build a function takes a random sample with replacement from a dataset and calculates the indirect effect.
2. Use a loop function to calculate a large number of bootstrapped indirect effects (e.g. 1000)
3. Add to the function to calculate a 95% confidence interval using the percentiles of the bootstrapped indirect effects.

4. Can you combine this function with the last so you'll get the results of the Causal Steps, Join Significance, Sobel, and Bootstrap methods all together?

#### Moderated Mediation Estimation

```
vaughnlm.cpath <- lm(relativepromotion ~ nscond, data = vaughndata)
coef(vaughnlm.cpath)
```

```
## (Intercept)      nscond
##   -2.137922    1.334709
```

```
cpath <- coef(vaughnlm.cpath)[2]

vaughnlm.apath <- lm(autonomy ~ nscond*gender, data = vaughndata)
coef(vaughnlm.apath)
```

```
##   (Intercept)        nscond        gender nscond:gender
##     2.3752522     1.3806861    -0.9661252     0.6235727
```

```
a1path <- coef(vaughnlm.apath)[2]
a2path <- coef(vaughnlm.apath)[3]
a3path <- coef(vaughnlm.apath)[4]
```

```
vaughnlm.bpath <- lm(relativepromotion ~ nscond+autonomy, data = vaughndata)
coef(vaughnlm.bpath)
```

```
## (Intercept)      nscond    autonomy
##  -2.6156877   0.1629932   0.5073761
```

```
cprimepath <- coef(vaughnlm.bpath)[2]
bpath <- coef(vaughnlm.bpath)[3]

cond.ind.W1 <- (a1path + a3path*1)*bpath #Indirect for Males
cond.ind.W2 <- (a1path + a3path*2)*bpath #Indirect for Females
```

**Practice 7**

1. Build a function which estimates all the paths in a first stage moderated mediation model.
2. Build a function which calculates the conditional indirect effect, conditional on the coded values of W (1 and 2). Can you make this more general? What if W were coded as different values?
3. Add to your function the ability to calculate the index of moderated mediation.
4. Add to the function to calculate a 95% confidence interval using the percentiles of the bootstrapped index of moderated mediation.
5. Challenge: Can you make a new function which examines second stage moderated mediation?