# Psy/Educ 6600: Chapter 3

SUMMARY DESCRIPTIVE STATISTICS

Your Name

Spring 2018

## Contents

# PREPARATION

## Packages

- Make sure the packages are **installed** *(Package tab)*

```
library(tidyverse)     # Loads several very helpful 'tidy' packages
library(readxl)        # Read in Excel datasets
library(furniture)     # Nice tables (by our own Tyson Barrett)
library(psych)         # Lots of nice tid-bits
```

# SECTION C

## Import Data, Define Factors, and Compute New Variables

- Make sure the **dataset** is saved in the same *folder* as this file
- Make sure the that *folder* is the **working directory**

  NOTE: I added the second line to convert all the variables names to lower case. I still kept the `F` as a capital letter at the end of the five factor variables.

```
data_clean <- read_excel("Ihno_dataset.xls") %>%
  dplyr::rename_all(tolower) %>%
  dplyr::mutate(genderF = factor(gender,
                                 levels = c(1, 2),
                                 labels = c("Female",
                                            "Male"))) %>%
  dplyr::mutate(majorF = factor(major,
                                levels = c(1, 2, 3, 4,5),
                                labels = c("Psychology",
                                           "Premed",
                                           "Biology",
                                           "Sociology",
                                           "Economics"))) %>%
  dplyr::mutate(reasonF = factor(reason,
                                 levels = c(1, 2, 3),
                                 labels = c("Program requirement",
                                            "Personal interest",
                                            "Advisor recommendation"))) %>%
  dplyr::mutate(exp_condF = factor(exp_cond,
                                   levels = c(1, 2, 3, 4),
                                   labels = c("Easy",
                                              "Moderate",
                                              "Difficult",
                                              "Impossible"))) %>%
  dplyr::mutate(coffeeF = factor(coffee,
                                 levels = c(0, 1),
                                 labels = c("Not a regular coffee drinker",
                                            "Regularly drinks coffee")))  %>%
  dplyr::mutate(hr_base_bps = hr_base / 60) %>%
  dplyr::mutate(anx_plus = rowsums(anx_base, anx_pre, anx_post)) %>%
  dplyr::mutate(hr_avg = rowmeans(hr_base + hr_pre + hr_post)) %>%
  dplyr::mutate(statDiff = statquiz - exp_sqz)
```

##Question C-1/3. Descriptive Statistics -full-

Use the `psych::describe()` function to find the ~~mode~~, **median** and **mean**, as well as the ~~**range**, semi-interquartile range, *unbiased* **variance*z, and*~~ unbiased*** standard deviation** for each of the *quantitative variables* in Ihno's data set.

> Make sure to use a `dplyr::select(var1, var2, ..., var12)` step to select only the variables of interest.

```
# Descriptive Stats: all quant vars
```

## Question C-4 Boxplots

### (a) Boxplot

Create a plot for the `statquiz` variable using a `geom_boxplot()` layer.

> Make sure to specify the astheticis in `ggplot(aes(...))`. Since you want to plot the entire sample together, set `x = "Full Sample"` and `y = continuous_var`

```
# Boxplot: statquiz
```

### (b) Boxplots -by- a Factor

Create a plot for the `statquiz` variable by `majorF`.

> You may choose to (1) split the x-axis with the `x = grouping_var` option in the asthetics, (2) specify a variable to fill in the boxes with color with the `fill = grouping_var`, or (3) make seperate panels by adding a `facet_grid(. ~ grouping_var` layer.

```
# Boxplot: statquiz, by majorF
```

### (c) Boxplot -for- a Subset

Use a `dplyr::filter()` step filter the subjects in the dataset to create a **Boxplot** for the `statquiz` variable for just the `female Biology` majors.

> Make sure to use `==` instead of `=` to test for equality within the filter step. Make sure the use a `&` symbol to require multiple conditions.

```
# Boxplot: statquiz, for a subset
```

### (d) Boxplots -by- a Factor and -for- a Subset

Use `dplyr::filter()` to create a SIDE-by-SIDE Boxplots for the `statquiz` variable that compares the `female Psychology` majors to the `female Biology` majors.

> A helpful symbol-set is `%in%`, which tests if the thing before it (`majorF`) is included in the concatinated list of options (eg. `c("Biology", "Psychology")`) that comes after it. Make sure the use a `&` symbol to require multiple conditions.

```
# Boxplot: statquiz, by a factor, for a subset
```

### Question C-5. Boxplots -for- Repeated Measures

Create Boxplots for both baseline and prequiz **anxiety**, so that they appear side-by-side on the same graph.

> Some data manipulations is needed to "stack" the two variables (baseline and pre-test) into a single variable. This is done with with the `tidyr::gather(key = "new_key", value = "new_value", old_var_1, old_var_2, ...)` function.

```
# Boxplot: anxiety, compare two repeated measures
```

## Question C-6. Descriptive Statistics -by- a Factor

Use `furniture::table1()` to find the *mean* and *standard deviation* for each of the *quantitative variables* separately for the `male` and `female` ecomonics majors. **Make sure to include the `na.rm = FALSE` option in the `table1()` function to include all ten ecomonics majors.**

> Make sure to (1) use the `dplyr::filter(grouping_var == "Value")` to reduce to just Economics majors and (2) use the `dplyr::group_by(grouping_var)` step to seperate the Males and Females, before the `furniture::table1()` step.

```
# Descriptive Stats: all quant vars, by genderF
```