

# EPSY 8266 Lab 1: Introduction to R, lavaan, and regression

Chris Desjardins

1/27/2019

For the lab assignments, please include all R code and all R output that you used to answer a question. This may include graphs (e.g., Q7 - Q9). Please try and be concise, but complete in answering each question.

*The purpose of this assignment is to get you a little more familiar with R, running code, interpreting output, and modifying code.*

## Lab Data Set

We will continue to explore the Holzinger & Swineford (1939) data for this lab. As a reminder, the Holzinger & Swineford (1939) data set was described in the first in-class activity. You can read it into R directly from the course website or you can download the data set from the course website and then read it into R afterwards. The first way was shown in the first in-class activity.

To read in data the second way, you should set your current working directory to the directory (or folder) where you saved the data. You can set the directory through the RStudio menu: **Session -> Set Working Directory -> Choose Directory**.

After you've set your current working directory, you can read in the data into R as follows:

```
hs.data <- read.csv("HolzingerSwineford1939.csv", header = TRUE)
```

## Learning R

Below is some code that will help you answer the next six questions. You will need to modify the commands to answer the next few questions. I recommend running this code to understand what it does if you're unfamiliar with R.

```
# Structure of the data
str(hs.data$id)
typeof(hs.data$id)

# Subsetting the data
subset(hs.data, id == 4)
subset(hs.data, school == "Pasteur" & gender == "Male" & agey == 14)

# Summarizing the data
summary(hs.data$straight)
table(hs.data$school)
aggregate(cubes ~ school, data = hs.data, FUN = mean)
cor(hs.data$visual, hs.data$cubes)
```

**Q1.** What type of a variable is gender in R? What type of a variable is sentence in R? Are these variables the same type or not? How does this refer to their scale of measurement?

**Q2.** What are the minimum and maximum scores on the sentence test?

Q3. What is the mean on the `sentence` test at Grant-White? What is the mean on the `sentence` test at Pasteur?

Q4. What is the correlation between the `sentence` and `wordc` tests?

Q5. What school does student with ID 90 attend and what is their gender?

Q6. How many male and female students are there in the data set?

## Initial Data Analysis

Recall, you can obtain a histogram, stem-and-leaf plot, and scatter plot by running the following code:

```
hist(hs.data$visual)
stem(hs.data$visual)
plot(visual ~ cubes, hs.data)
```

Again, you will need to modify this code to answer the next two questions.

Q7: Briefly describe the distribution of the `sentence` test. Specifically, (a) comment on the shape, (b) the presence of outliers, and (c) whether you think a transformation would be necessary to make it normal.

Q8: Create a scatter plot of `sentence` against `wordm` and describe the nature of the relationship. Specifically, comments on (a) the direction and (b) magnitude of the relationship, (c) whether it's linear or not, and the (d) presence of outliers.

## Multiple Regression

Let's say you are interested in predicting `sentence` given `wordc` and `wordm`. You therefore decide to fit a multiple regression where you regress `sentence` onto `wordc` and `wordm`. For this lab, you will use both the `lm` function and the `sem` function in the `lavaan` package. To use the `sem` function, you first need to load `lavaan` by typing the following.

```
library(lavaan)
```

Next, you need to fit two multiple regression models. Please fit one model using the `lm` function and one model using the `sem` function and save them as `fit.lm` and `fit.sem`. Refer back to the course notes or the in-class assignment for how to do this. The following commands will be help to answer the remaining questions for this lab. You do not need to modify this code.

```
# View modeling assumptions
plot(fit.lm)
plot(fit.lm, which = 4) # Cook's distance

# Get the output from your model
summary(fit.lm)
summary(fit.sem)
summary(fit.sem, standardized = TRUE) # this argument will give you the beta weights

# Confidence intervals
confint(fit.lm)
summary(fit.sem, ci = TRUE)

# Obtain the residuals and predicted values
resid(fit.lm)
resid(fit.sem)
```

```

predict(fit.lm)
lavPredict(fit.sem)

# Obtain fit measures from lavaan
summary(fit.sem, fit.measures = TRUE)

```

Q9: Please comment on the appropriateness of your statistical model that you saved as `fit.lm`. Specifically, do you see (a) evidence of non-constant variance, (b) outliers, (c) influential points, and (d) any gross deviations of normality for your residuals?

Q10: Please report the (a) unstandardized estimates from the regression model you fit for `wordc` and `wordm`, (b) comment on their significance assuming  $\alpha = .05$ , and (c) interpret them.

Q11: Are the results the same using the `lm` and `sem` functions? (yes/no)

Q12: Please (a) report and (b) interpret the standardized estimates (i.e., the beta weights) for `wordc` and `wordm`. HINT: I have shown you one way to do this above and the other way I have shown you in our class notes

Q13: (a) Please report the 95% confidence intervals for unstandardized estimates. (b) How do these confidence intervals relate to the statistical significance of these parameters?

Q14: (a) What is unusual about the fit measures for the `fit.sem` model? (b) Please report the fit statistics (RMSEA, TLI, and CFI) and (b) the degrees of freedom of your model. For (a), if you know why this is the case, please report why otherwise feel free to take a guess.

Q15: (a) Are your residuals the same between `fit.lm` and `fit.sem`? (b) What are these residuals from `fit.sem`? HINT: This relates to one of the other names for SEM.

Q16: Like the residuals, the predicted values differ for these two fitted models, too. (a) What are the predicted values for `fit.sem`? In other words, when you run `lavPredict(fit.sem)` what does R give you back? (b) Why does it look like this? HINT: Read the description in the help page (`?lavPredict`).