Regression Exercise

## Note

*This document gives you a practice example. It does not show you an example of simple and multiple linear regression. If you need a more in-depth guide, then go to the file Regression-Guide.docx.*

## Running our Simple Linear Regression

In R it is easier to run the model and then check our assumptions when running a simple linear Regression.

The syntax for running the model is the following:

model\_name <- lm(DV ~ predictor, data = our\_dataframe)  
  
summary(model\_name)

The code lm(DV ~ predictor, data = our\_dataframe) runs our linear regression, whereas the code summary(model\_name) prints out the results in a readable manner. We can substitute any name in place of model\_name.

Let’s run a linear regression using the wellbeing\_df data frame. We will predict Wellbeing from Age scores and call it model3.

model3 <-

What are your results?

## Assumptions

Now we can check our assumptions. There are several assumptions to running a linear regression.

* The outcome/DV is a interval/ratio level data
* The predictor variable is interval/ratio or categorical (with two levels)
* All values of the outcome variable are independent (i.e., each score should come from a different participant)
* The predictors have non-zero variance
* The relationship between outcome and predictor is linear
* The residuals should be normally distributed
* There should be homoscedasticity (homogeneity of variance, but for the residuals)

The first three assumptions we can check just by examining the data (and we have met these assumptions).

Finish the code here to check each of these assumptions

check\_model(model\_name) #this will enable us to check assumptions 4 and 5  
  
  
check\_normality(model\_name) #this will enable us to check normality assumption  
  
  
check\_heteroscedasticity(model\_name) #this will enable us to check homoscedasticity

## Run Power Analysis

Run a post-hoc power analysis to estimate the power of the study.

effect\_size <- sqrt(adjusted r2) #you will find the adjusted r2 in your model results  
  
pwr.f2.test(u = number of predictors, #number of predictors  
 v = degrees of freedom, #check your model results  
 f2 = effect\_size,   
 sig.level = .05)

# Running a Multiple Linear Regression

Let’s keep the same model used in the last example, but let’s add two more predictors (Depression and Fatigue) by running a multiple linear regression. The syntax for running this test in R is very straightforward and similar to the simple linear regression syntax

model\_name <- lm(DV ~ predictor1 + predictor2 + predictor3, data = our\_dataframe)  
  
summary(model\_name)

As you can see, all we need to do to run the model is literally add another predictor to our formula. We still do the same steps by first saving this model to a variable and then using the summary() to make the results readable.

Using this syntax, run a multiple linear regression with Age, Depression, and Fatigue as predictors and call it model4.

model4 <-

What are your results?

## Check Our Assumptions

The assumptions for multiple linear regression are the same as simple linear regression, but we also have to check for multicollinearity

check\_model(model\_name) #this will enable us to check assumptions 4 and 5  
  
  
check\_normality(model\_name) #this will enable us to check normality assumption  
  
  
check\_heteroscedasticity(model\_name) #this will enable us to check homoscedasticity   
  
check\_collinearity(model\_name) #this will enable us to check whether we have multicollinearity

## Power Analysis

Run a post-hoc power analysis to estimate the power of the study.

#f2 = effect\_size = sqrt(adjusted R2)  
  
effect\_size <- sqrt(adjusted r2)  
  
pwr.f2.test(u = number of predictors, #number of predictors  
 v = degrees of freedom,   
 f2 = effect\_size,   
 sig.level = .05)

# Wrap Up

There you have it! That’s how to run linear and multiple linear regression in R.