

# LT2 and LT3 Worksheet

## Preliminaries

For this worksheet, I would like you to conduct and interpret multiple regression and moderation analyses using what you have learned from LT2 and LT3.

*As always, please write your answers and code in an R notebook and submit it as either a .pdf or .rmd (notebook) file in the Moodle submission point.*

Before we start, let's load the necessary packages: "tidyverse", "probemod", "ggplot2", "readr", "lm.beta", "car", "modplot" and "Hmisc".

```
library(tidyverse)
etc
```

## Multiple Regression

To work on testing and interpreting multiple regression, we are going to use data from Atir, Rosenzweig, and Dunning (2015) published in Psychological Science. The dataset can be downloaded from here: <https://www.dropbox.com/s/etld1w5868kxoyt/overclaiming.csv>. The file is also available in the PB130 R project on the github under the folder worksheets → LT2 and LT3.

As a first step in the analysis, download the file and save it to a **logical** place where it can be easily located. I would recommend that you save it in your PBS folder under a subfolder called Worksheets. Now load the file by clicking on the dataframe name in the files window where you saved the protest data and click import data. A new screen will appear with the dataframe. Press "update" in the top right corner and then when it loads click import. You can also copy the R code produced in the bottom right corner and paste that code in an R chunk on your worksheet. For example:

```
protest <- read_csv("Worksheet 3/overclaiming.csv")
View(protest)
```

## Study Description

This study investigated the relationship between self-reported knowledge and overclaiming of knowledge among 202 adults, and whether that relationship still existed when controlling for genuine expertise.

Valuing expertise is important for modern life. When people have a problem, they need to know who to turn to for a solution to their problem. For example, when people get sick, they know that a doctor is an expert in the field of medicine and can help them get better. In general, experts simply know more about a topic than do non-experts. However, experts may be vulnerable to a particular problem of knowing so much. They may have the illusion that they know more about a topic than they actually do. This particular type of overconfidence is called overclaiming. Essentially, overclaiming occurs when people claim that they know something that is impossible to know, such as claiming to know the capital of Sharambia (a country that doesn't actually exist).

To test if experts are susceptible to overclaiming, Atir, Rosenzweig, and Dunning (2015) recruited 202 individuals from an online participant pool. They first asked participants to complete a measure of self-reported knowledge and a measure of overclaiming. The self-reported knowledge questionnaire asked people

to indicate their level of knowledge in the area of personal finance on a scale of 1 to 7 (1 = no knowledge, 7 = full knowledge). This is the “self\_knowledge” variable in the dataset.

The overclaiming task asked participants to indicate how much they knew about 15 terms related to personal finance (e.g., home equity). Included in the 15 items were three terms that do not actually exist (e.g., annualized credit). Thus, overclaiming occurred when participants said that they were knowledgeable about the non-existent terms. The proportion of overclaims made is the “overclaiming” variable in the dataset.

Finally, participants completed a test of financial literacy called the FINRA. Whereas the earlier questionnaires measured self-reported knowledge, the FINRA measured actual knowledge. This is the “FINRA” variable in the dataset.

## Activity

Now I want you to build two linear models that test a) the simple relationship between self-reported knowledge and overclaiming and b) the multiple relationship between a combination of self-reported knowledge and genuine expertise and overclaiming. The first simple regression will tell us if self-reported knowledge predicts overclaiming and the second multiple regression will tell us whether self-reported knowledge still predicts overclaiming after controlling for genuine expertise. Our theory here is that when genuine expertise is eliminated as a possible confounding variable (i.e., controlled for in the multiple regression model), self-reported knowledge would still predict overclaiming.

Before we delve into this topic, let's just clarify the research question and null hypothesis we are testing:

*Research Question* - Does self-reported knowledge predict overclaiming controlling for genuine expertise?

*Null Hypothesis* - The relationship between a combination of self-reported knowledge and genuine expertise and overclaiming will be zero.

There are several steps needed to conduct this analysis:

### Task 1

Use the `rcorr()` function to generate the correlation coefficient matrix. Comment on the correlation coefficients and their associated p values.

```
rcorr(as.matrix(overclaiming[,c("??", "??", "??")], type="??"))
```

### Task 2

Then, use the `lm()` function to build a simple regression model for self-knowledge (variable name self-knowledge) and overclaiming (variable name overclaiming). It is a two-parameter model because we have an intercept (b0) and slope (b1) for the predictor variable. Write the code to build this linear model and save it as a new R object called “simple.model”. In the same chunk, also use the `lm.beta()` function to call the standardized estimates. Don't forget to also request the summary for the simple.model object.

Comment on the meaning of the intercept (b0), the unstandardised and standardised slope estimates (b1), the standard error for the slope, the t-ratio for the slope, and the p value for the slope.

```
simple.model <- ??  
summary(??)  
lm.beta(??)
```

### Task 3

Now use the `Boot()` function to create 5000 resamples with replacement, estimating the linear model parameters on each occasion, and save them in a new R object called “simple.boot”. Then, use the `confint()` function to request the bootstrap confidence intervals for the estimates. Comment on the bootstrap confidence interval of the slope estimate from the output.

```
simple.boot <- Boot(??, f=coef, R = ??)
confint(??, level = ??, type="norm")
```

### Task 4

Now build a multiple regression model to ascertain whether self-reported knowledge (variable name self-knowledge) still predicts overclaiming (variable name overclaiming) while controlling for genuine expertise (variable name FINRA). Use the `lm()` function and write the code to build this linear model and save it as a new R object called “multiple.model”. In the same chunk, also use the `lm.beta()` function to call the standardized estimates. Don't forget to also request the summary for the multiple.model object.

Comment on the meaning of the intercept ( $b_0$ ), the unstandardised and standardised slope estimates ( $b_1$  and  $b_2$ ), the standard errors for the slopes, the t-ratios for the slopes, and the p values for the slopes.

```
multiple.model <- ??
summary(??)
lm.beta(??)
```

### Task 5

Now use the `Boot()` function to create 5000 resamples with replacement, estimating the linear model parameters on each occasion, and save them in a new R object called “multiple.boot”. Then, use the `confint()` function to request the bootstrap confidence intervals for the estimates. Comment on the bootstrap confidence interval of the slope estimate from the output.

```
multiple.boot <- Boot(??, f=coef, R = ??)
confint(??, level = ??, type="norm")
```

### Task 6

Finally, I want you to provide an answer to the research question:

Can we reject the null hypothesis that the multiple relationship is zero?

Does self-reported knowledge predict overclaiming controlling for genuine expertise?

Make sure you justify your answers using the model coefficients and confidence intervals.

## Moderation

### Introduction

To work on testing and interpreting moderation, we are going to use data from Garcia, Branscombe, and Ellemers (2010) published in the European Journal of Social Psychology. The data file is named protest and

can be downloaded from here: <https://www.dropbox.com/s/06a446cixj2cgep/protest.csv>. The file is also available in the PB130 R project on the github under the folder worksheets -> LT2 and LT3.

As a first step in the analysis, download the file and save it to a **logical** place where it can be easily located. I would recommend that you save it in your PBS folder under a subfolder called Worksheets. Now load the file by clicking on the dataframe name in the files window where you saved the protest data and click import data“. A new screen will appear with the dataframe. Press “update” in the top right corner and then when it loads click import. You can also copy the R code produced in the bottom right corner and paste that code in an R chunk on your worksheet. For example:

```
protest <- read_csv("Worksheet 3/protest.csv")
View(protest)
```

## Study Description

In this study, 129 female adults received a written account of the fate of a female attorney - named Catherine - who lost a promotion to a less qualified male as a result of discriminatory actions of the senior partners. After reading this story, the participants were given a description of how Catherine responded to this act of discrimination. Those randomly assigned to the no protest condition (coded protest = 0 in the data file) learned that though very upset by the decision, Catherine decided not to take action against this discrimination and continued at the firm. The remainder of the participants, assigned to a protest condition (coded protest = 1 in the data file), were told that Catherine approached the partners with the request that they reconsider the decision, while giving various explanations as to why the decision was unfair. Following this, the participants were asked to respond to six questions evaluating Catherine (e.g., Catherine has many positive traits“, “Catherine is a person I would like to be friends with“). Their responses were aggregated into a measure of liking, such that participants with higher scores liked her relatively more (variable “liking” in the data file).

In addition to this measure of liking, each participant was scored on the Modern Sexism Scale, used to measure how pervasive a person believes sex discrimination is in society. The higher a person’s score, the more pervasive they believe sex discrimination is in society (variable sexism in the data file). The focus of the study was to assess extent to which the action of Catherine affected perceptions of her — specifically how much participants liked her - and whether the size of such an effect depends on a person’s beliefs about the pervasiveness of sex discrimination in society. As we saw in LT3, this is a classic question of moderation.

## Activity

Before we investigate this moderation, though, let’s just clarify the research question and null hypothesis you will be testing:

*Research Question* - Is the effect of whether Catherine protested or not on the extent to which she is liked moderated by beliefs about the pervasiveness of sex discrimination in society?

*Null Hypothesis* - The interaction of Catherine’s decision to protest and perceptions of sex discrimination will be zero.

To test this research question, there are several tasks needed:

### Task 1

First, you are going to use the `lm()` function to build a three-parameter linear model. It is a three-parameter model because we have three coefficients in the linear model when running moderation analysis; (1) the predictor (protest), (2) the moderator (sexism), and (3) the predictor\*moderator interaction. The outcome is liking. Write the code to build this linear model and save it as a new R object called “mod.model”. Don’t forget to also request the summary for the mod-model object.

```
mod.model <- ??  
summary(??)
```

## Task 2

Now use the `Boot()` function to create 5,000 resamples with replacement, estimating the linear model parameters on each occasion, and save them in a new R object called “mod.boot”. Then, use the `confint()` function to request the bootstrap confidence intervals for the estimates. Comment on the bootstrap confidence interval of the b3 or interaction term estimate from the output.

```
mod.boot <- Boot(??, f=coef, R = ??)  
confint(??, level = ??, type="norm")
```

## Task 3

Lets now probe the interaction. Becuase the moderator is continous (i.e., sexism is measured on a Liktert scale), we need to calculate the conditional effect of protest condition on liking at various levels sexism. You will recall we can do this in a couple of ways. First, lets use pick-a-point to ascertain the conditional effect at low (-1 SD), moderate (mean), and high (+1 SD) sexism. To do so, use the `pickapoint` function we worked with in LT3. Comment on the conditional effects at low, moderate, and high sexism. Where in the range of sexism does protest predict liking and where does it not?

```
pickapoint(??, dv='??', iv='??', mod='??')
```

## Task 4

A better approach to `pickapoint` is the Johnson Neyman technique, which uses the critical value of t (1.96) to yield an region of significance for the moderator at which the conditional effect of the predictor on the outcome is and isn’t significant. Three outcomes are possible using this function:

- 1) A single value of the moderator that the conditional effect is either significant above, or below, but not both.
- 2) Two values of the moderator that the conditional effect is significant between, or either side of, but not both.
- 3) No value of the moderator indicating that the conditional effect is either significant across the entire range of the moderator or is not significant anywhere in the range of the moderator.

Write the code below to run the JN technique on the protest data using the `jnt()` function. Comment on the value that is output from the analysis.

```
jnt(??, predictor = "??", moderator = "??", alpha = ??)
```

## Task 5

The JN value(s) is informative information, but nevertheless quite difficult to visualise so finally I would like you to write the code to plot this moderation using the `modplot()` function. Once this is complete, comment on the plot and nature of the moderation.

```
modplot(??, predictor = "??", moderator = "??", alpha = ??, jn = TRUE)
```

### Task 6

Finally, provide an answer to the research question - can we reject the null hypothesis that the interaction of Catherine's decision to protest and perceptions of sex discrimination will be zero? Make sure you justify your answer using both the linear model and the probing of the interaction.