

## An Overview: Writing an Empirical Social Science Paper - Quiz

### Question 1

True or False: The information presented in this lecture is only applicable to social science papers and not to all scientific papers.

True

False



#### Explanation

As Prof. Duflo mentions in the introduction, the information presented in these lectures are for the most part transferable to all scientific papers.

## What to Include in the Introduction - Quiz

### Question 1

True or False: When writing a paper, it is always important to foreshadow without revealing the results in the introduction.

True

False ✓

#### Explanation

A unique trait to an economics paper compared to many other types of academic papers is that one reveals what the results will be in the introduction. This is not standard across all academic papers however and is often not the case in other disciplines.

### Question 2

Select all that apply: In an introduction one tries to:

Inform the reader of past related literature and explain how you are adding to the literature

Explain what question it will be that you're answering in your paper

Show tables and graphs of what your results will be

Explain how you will do your analysis



## Examples of Good Introductions, Part I - Quiz

### Question 1

True or False: When writing an introduction it is acceptable to be repetitive in order to be very explicit with your goal for the paper.

True

False



#### Explanation

In the example given in lecture, what the author was trying to do was repeated and even emphasized using "...-" in order to drive the point of the paper. Someone reading your paper is not as experienced on the topic as you may be so it can be good to be very clear on your topic.

### Question 2

The literature review included in the introduction should:

Show what it is you are bringing to the table

Show why your topic is interesting

Show what has already been answered with respect to your topic

All of the above



#### Explanation

One must include past papers related to your topic in order to show what has already been proven and then use it to show why your additional layer to the topic is interesting. One does not want to pick a topic that has never been researched before because it is not very interesting. However they don't want to pick an overused topic that has been researched many times before.

## Examples of Good Introductions, Part II - Quiz

### Question 1

According to Professor Duflo, an effective introduction is usually around \_\_\_\_\_ pages.

2

✓ Answer: 2

2

#### Explanation

Professor Duflo mentions that the introduction should be around two pages in lecture. This is dependent on the topic and length of the paper, but can be used as a guide.

### Question 2

When foreshadowing one's results in the introduction, one should:

Explicitly state many of the results found, including tables of numerical results

Explicitly state the main results of the paper in some detail in one paragraph

Explicitly state the main results of the paper as well as an aspect of the mechanism which leads to the result in a few paragraphs

Explicitly state the main results of the paper shortly in about one sentence



#### Explanation

It is necessary to answer the main question in about a paragraph based on the results one found. It also beneficial, however, to include information on the mechanism which have produced the big results in order for other researchers to have a basis and potentially continue the research done in your paper. Tables would be too much detail.

## The Data Section, Part I - Quiz

### Question 1

True or False: In order to publish a paper one must include a link to the data used to reach the results.

True

False



#### Explanation

Although it is good practice to include a link to one's data, it is optional. There may be situations where the data is confidential or not able to be published for some reason. It is, however, becoming increasingly more popular to release data sets, a good development for transparency purposes.

### Question 2

Select the best answer: When the data used for a paper was produced using web scrapping one should

State that web scrapping was done and include a link to the code that was used to produce the data.

State that web scrapping was done and include the websites used for the process.

State that web scrapping was done and include the code that was used to produce the data.

Simply state that web scrapping was done and continue with the analysis.



#### Explanation

When detailing where the data came from one must state that it was gathered through web scrapping if this is the case. It is not necessary to show the code that was used to reach the results in the paper itself but is good practice to include a passage to reach the code. If the website used changes often and may cause different results to be reached, this should be stated in the paper and a link to the data set which was used for the purposes of your paper can still be included for reproducibility.

## The Data Section, Part II - Quiz

### Question 1

The ideal complement between the text and the tables in a paper is

To have tables that are easily readable, with less information, which will be explained in the text

To have both the tables and the text be as self-explanatory as possible without overlapping information

To have a table with an abundance of information so that the text can simply revert to the tables

To have both the tables and the text be self-explanatory, regardless of producing repeated information



### Question 2

1.0/1.0 point (graded)

Select all that apply: A good data table

Uses scientific notation

Does not include an abundance of significant digits

Variable names are self explanatory

Has the complete summary statistic taken from the console



#### Explanation

When creating a table you want it to be as readable and self-explanatory as possible. Scientific notation is avoided seeing how it can cause confusion. The amount of significant digits included should be cut down for readability purposes. Two is normally a good amount. Although you may be very comfortable with the variables used in your regression, most likely the readers will not be. In order for a table to be self-explanatory, the variable names must be as self-explanatory as possible. Acronyms should be avoided. Although R makes it easy to output a summary statistics for variables, the output from the console is not formatted correctly for a paper. There are programs within R, however, that will output nicely formatted summary statistics tables.

## Methods and Results - Quiz

### Question 1

V@Á [•oÁ ] [ :ca Á aÁ ;Á@Á^• |cÁ^&a }ÁÁ Á

- Show how one has interpreted a complicated model
- Draw a connection between the goal of the paper and the results found
- Demonstrate clearly what the regression model used was and its results
- Give a lot of numerical results found in the analysis



#### Explanation

Prof. Duflo mentions multiple times in lectures that the purpose of the results is to link the paper's objective that is most likely previously stated multiple times in the paper to the regression results one finds. This way the regression is what we expected and fits in well with the paper.

## The Conclusion - Quiz

### Question 1

The conclusion is not always necessary for a paper, however it may be helpful to (Select all that apply):

- Explain some of the research implications your paper may have
- Explain some of the policy implications your paper may have
- Show your results once more, including graphics
- Explain the research question



#### Explanation

The conclusion can be useful to close a paper by explaining the results once more and how they can be used for future purposes. It is not a results section however and should not include detailed results to the extent of a graphic. The research question may be reiterated briefly but it is not the place to be fully detailed once again.

## Doing more with Regression Output in R, Part I - Quiz

### Question 1

True or False: Whenever a categorical variable is used within a regression model in R you must explicitly factor the variable before using it in the model equation.

True

False



#### Explanation

If the factors of a variable are naturally distinguishable as factors, for example if they are character values, then R will automatically treat it as a categorical variable and it is not necessary to factor the variable before using it. On the other hand, if it is not recognizable as factors, such as the example in lecture where the levels were integers, then you must specify to R that you want the variable to be treated as a categorical variable with the values as factors.

### Question 2

You are working on a data set called "workers" in R and you write the following command:

```
> attach(workers)
```

What does this command do?

R will assume you are calling variables within workers without you having to specify

R will list all the variables within workers

Workers will be attached to any other data set saved in your workspace

Nothing, this command does not exist in R



#### Explanation

One attaches a data set so that you do not need to keep referring back to which data set you are referring. You can call a variable from that data set without specifying and R will refer to that data set

### Question 3

Which of the following are variables stored within a linear model created in R? Select all that apply.

`coef`

`corr`

`coefci`

`pval`



#### Explanation

Both `coef` and `coefci` are particularly mentioned in lecture. The command `lm()` stores a lot of information about your model you can later reach. `coef` allows you to reach the values of the estimated of your coefficients and `coefci` gives you their confidence intervals. Both the R-squared and the p-values can be seen from the summary output of the linear model.

## Doing more with Regression Output in R, Part II - Quiz

### Question 1

What is the command to add predictions from a linear model to your data frame? Please type in just the command with no parentheses or other punctuation (e.g. hist instead of "hist" or hist()).

✓ **Answer:** add\_predictions

## Doing more with Regression Output in R, Part III - Quiz

### Question 1

What is the following command testing?

```
> linearHypothesis(mod, c("gender = 2"))
```

- Whether the variable "gender" should be included in the linear model "mod".
- Whether the linear model "mod" has a variable named "gender" with two levels.
- Whether a linear model is a good model for "mod" when the coefficient for "gender" is set to two.
- Whether the coefficient "gender" is probabilistically equal to 2.

✓

## Independent Dummy Variables - Quiz

### Question 1

Which of the following variables would be a dummy variable in a linear model? Select all that apply.

 Age Race Gender Political affiliation

✓

#### Explanation

A dummy variable must be used when a categorical variable takes on many levels. Race can take on many different factors: white, black, Asian, and so on. In order to include this in a regression model you would need a dummy variable for each level, giving the value of 1 if the observation has that race and 0 if not. The same can be said for gender and political affiliation. Age, on the other hand, holds a numerical value and can be used directly. If you wish to bin ages and create age categories then one would use a dummy variable.