**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Classwork 6**

**From Exploring Variation to Explaining Variation**

**Textbook Chapter: 4**

**Advanced Learning Objectives:**

* Identify Outcome and Explanatory variables
* Distinguish between Experimental vs. Descriptive research design
* Practice representing relationships among variables
* Discuss sources of variation
* Define Type I Error

**Recap: What we’ve done so far**

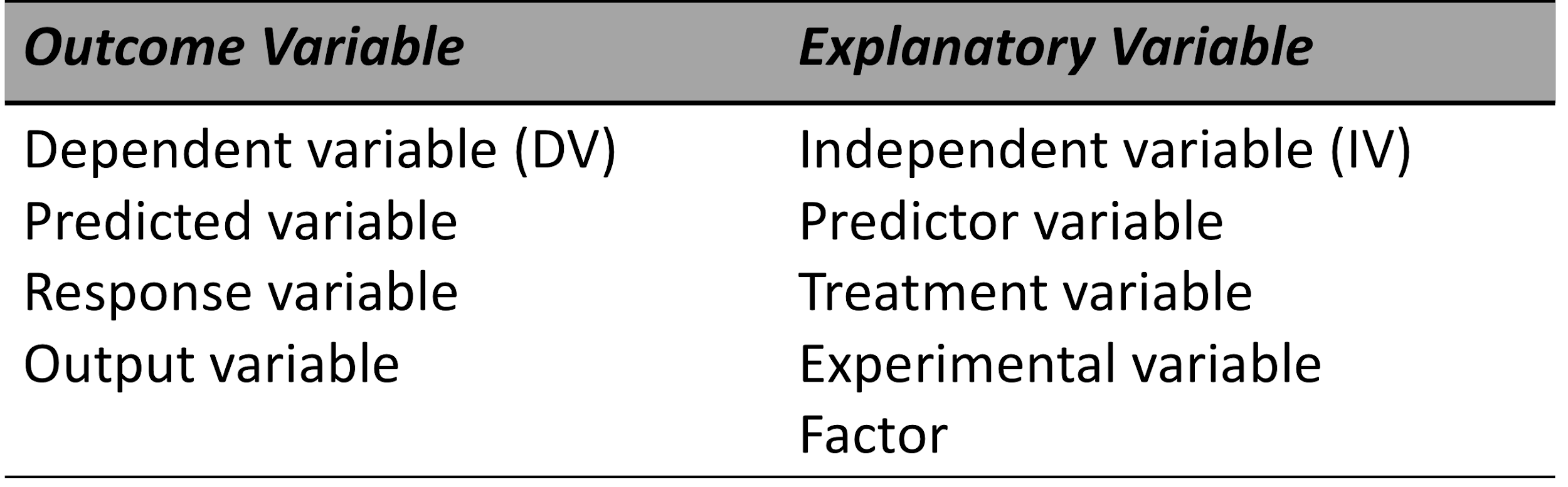
* What is data? How do we represent data?
* What are variables?
  + Different types of variables: Quantitative vs. Qualitative
* How can we explore variables?
  + Visual summaries (histograms, bar plots, etc.)
  + Numeric summaries (min, max, mean, median, quartiles, etc.)

We measure variables and *explore* their variation because we would like to *explain* *why* they vary the way they do. We achieve this by *modeling* variation, and then comparing and evaluating models.

Using statistical models helps us do three useful things:

* Helps us understand the DGP (the processes that cause variation)
* Helps us make predictions
* Helps us improve the functioning of complex systems

**More Ways to Classify Variables**



Outcome and Explanatory variables can be either \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ or \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ .

What is an outcome variable?

What is an explanatory variable?

**Identify the outcome variable and the explanatory variable**

1. Students were interested in whether faculty parking lots on campus have more expensive cars than student parking lots. A large team of students went to all of the parking lots on a campus and took pictures of each car. Back in class, they looked up each car's current value. They then determined the median car value for each parking lot. Students wanted to compare the values of cars in faculty lots versus student lots.
2. Geologists wanted to examine the effect of acid rain on rock formations. Random samples of rocks from the same region were either treated with low or high levels of acid rain. Geologists compared changes in mass in the rocks treated with low levels of acid rain to changes in mass in the rocks treated with high levels of acid rain.
3. A study was designed to test whether or not adding extra Vitamin A to the diet increases the ability to see in dim light. One hundred participants were randomly assigned to two groups. Half were given a diet with extra Vitamin A, and the other group was not asked to change their diet. A dim-light reading test was conducted on all participants at the end of the study. The researcher wanted to compare scores on the dim-light reading test of those who received extra vitamin A to scores on those who were not asked to change their diet.

We can examine variation with any combination of variable types.

|  |  |  |
| --- | --- | --- |
|  | **Quantitative Outcome** | **Categorical Outcome** |
| **Categorical**  **Explanatory** |  |  |
| **Quantitative Explanatory** |  |  |

**Research Designs: Experiments vs. the Rest**

What types of relationships are **descriptive**?

What is the only research strategy that can determine a **causal relationship?**

Fill in the table to compare the types of **research studies**:

|  |  |  |
| --- | --- | --- |
|  | **Variables** | **Relationships** |
| **Category I:**  Descriptive study |  |  |
| **Category II:**  Correlational study |  |  |
| **Category III:**  Experimental, Quasi-experimental, Non-experimental Studies |  |  |

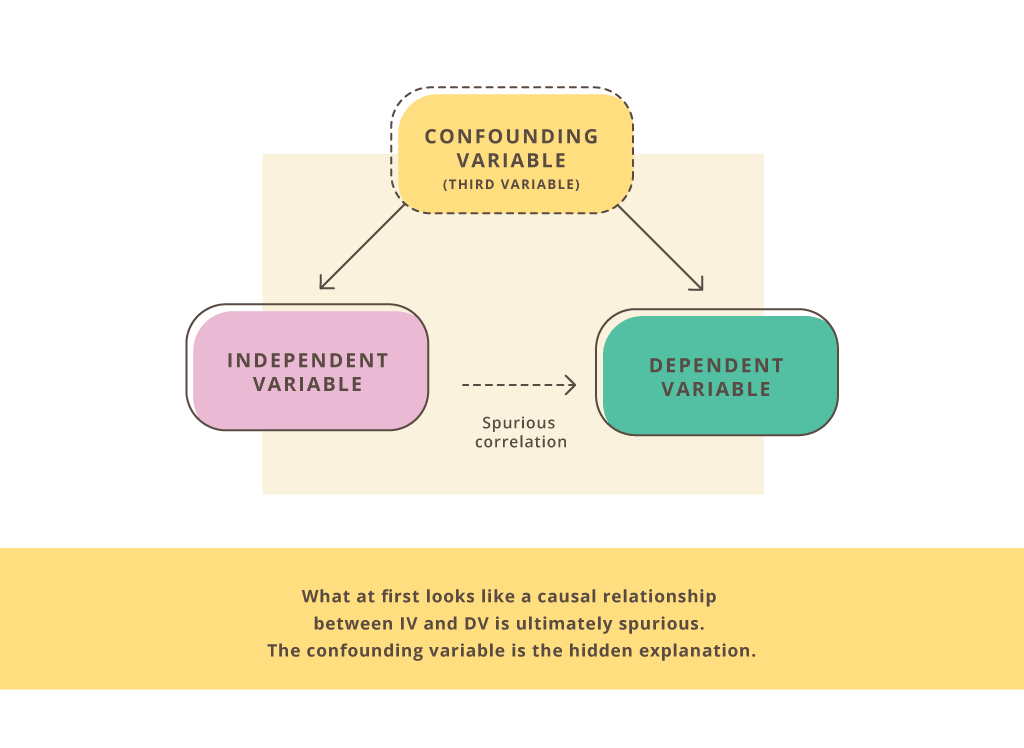
Fill in the table to compare the types of **category III studies**:

|  |  |  |
| --- | --- | --- |
|  | **Variables** | **Relationships** |
| **Experimental:** |  |  |
| **Quasi-experimental:** |  |  |
| **Non-experimental:** |  |  |

**Why CAN’T the other strategies determine causation?**

The **third-variable problem** means an \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_variable might be causing both

variable 1 and variable 2 in a study. Sometimes, this also called a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ variable.



What might be a **third variable** that could explain why income (variable 1)affects memory (variable 2)?

In a non-experimental design, we often don’t know if **variable 1 causes variable 2** *or* **if variable 2 causes**

**variable 1** and this is called the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ .*.*

Experimental research can address these limitations with research design:

* Manipulation:
* Control:
  + Random Sampling
  + Random Assignment

In a well-designed experiment, there are **two possible reasons** for the variation we observe between groups:

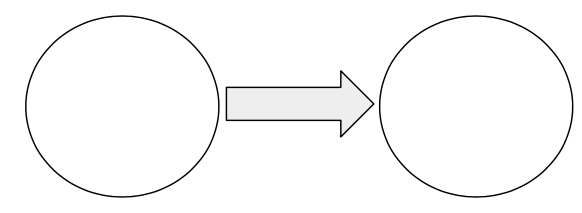
1. Which of the following would show a descriptive relationship, and which would show a causal relationship?

Also, what is the DV? What is the IV?

1. One group of students is randomly assigned to take a test that is timed, and another group of students is randomly assigned to take a test without being timed.
2. A group of students is asked to report whether their parents went to college and what their GPA is.

**Representing Relationships Among Variables: Path Diagrams and Word Equations**

Path Diagrams:



Word Equations:

**DATA = MODEL + ERROR**

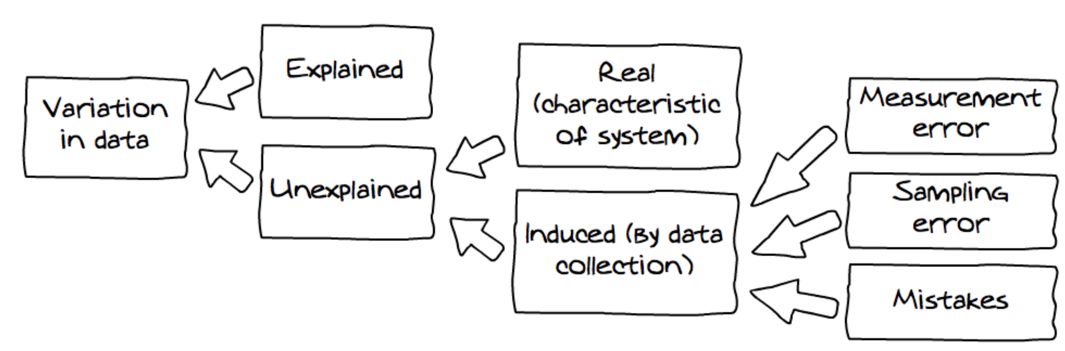
**Null/Empty Model:**

**Research/Explanatory Model:**

“Variation in \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is explained by variation in \_\_\_\_\_\_\_\_\_\_\_\_\_ *plus variation in other stuff*.”

Can we add more variables to our model?

**Sources of Variation**



Even if we successfully find that variable x explains some of the variation in variable y, we will still find variation in each group (although it will be less than the overall variation). What type of variation is this?

**Ruling out Random Chance and Type I Error**

Simply due to the nature of sampling variability, we will always observe some differences in the distributions across groups. How do we know if this difference between experimental and control groups is a real effect and not just due to chance? How distinctive would the difference have to be to convince us that putting smiley faces on receipts actually had an effect on tips?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ error is when we conclude that some variable we manipulated had an effect, when in fact

the observed difference was simply due to random sampling variation.

We can never reduce the chance of Type I error to \_\_\_\_\_\_\_\_\_\_\_\_\_\_; there is always some chance you might be

fooled.

On the other hand, there is also the possibility of Type II Error. What do you think that is?