

Gov50 Section 4

Background

Do changes in one's financial circumstances affect one's decision-making process and cognitive capacity? In an experimental study, researchers randomly selected a group of US respondents to be surveyed before their payday and another group to be surveyed after their payday. Under this design, the respondents of the **Before Payday** group are more likely to be financially strained than those of the **After Payday** group. The researchers were interested in investigating whether or not changes in people's financial circumstances affect their decision making and cognitive performance. Other researchers have found that scarcity induce an additional mental load that impedes cognitive capacity. This exercise is based on:

Carvalho, Leandro S., Meier, Stephen, and Wang, Stephanie W. (2016). "Poverty and economic decision-making: Evidence from changes in financial resources at payday." *American Economic Review*, Vol. 106, No. 2, pp. 260-284.

In this study, the researchers administered a number of decision-making and cognitive performance tasks to the **Before Payday** and **After Payday** groups. We focus on the *numerical stroop task*, which measures cognitive control. In general, taking more time to complete this task indicates less cognitive control and reduced cognitive ability. They also measured the amount of cash the respondents have, the amount in their checking and saving accounts, and the amount of money spent. The data set is in the CSV file **poverty.csv**. The names and descriptions of variables are given below:

Name	Description
<code>treatment</code>	Treatment conditions: Before Payday and After Payday
<code>cash</code>	Amount of cash respondent has on hand
<code>accts_amt</code>	Amount in checking and saving accounts
<code>stroop_time</code>	Log-transformed average response time for cognitive stroop test
<code>income_less20k</code>	Binary variable: 1 if respondent earns less than 20k a year and 0 otherwise

Question 1: Using `summary()`

First, read in the experiment's data from `poverty.csv` into an object named `dat`, using `read_csv()`.

We may think that a respondent's financial situation is important in this experiment. Let's get some summary statistics for `cash` and `accts_mt` to help us understand these variables better.

Use the `summary()` function to learn about the distribution of cash on hand and money in the bank in our sample. Instead of using the pipe `|>`, be sure to use the `$` to tell R which column you want to summarize. You can use the following format: `summary(dataset$variable)`

What do you notice about these variables that may be important for the study? Do you think the mean or the median is a better summary statistic of the central tendency of these variables?

Question 2: Proportion table

Now let's look at the outcome variable `stroop_time`. Use `case_when` to break this continuous variable into three categories : under 7 seconds, between 7 and 7.5 seconds, and more than 7.5 seconds. Then create a table that shows both the counts and the proportions of the sample that fall into these three categories.

- Hint: First create a table of counts. Then use `mutate()` to add a column of proportions.

Question 3: Examining and plotting variables with multiple groupings

Next let's look at the distribution of cash on hand by treatment status and income.

First, use `mutate()` to create a new variable for the income indicator `income_less20k` that assigns more informative labels as the variable values. Replace the numerical values with "Lower income" and "Higher income". Then create a table that shows average cash on hand by each combination of the income indicator and the treatment assignment.

Then create a bar plot of the data in the table. The y-axis should show cash on hand and the x-axis should show the income indicator. The plot should contain double bars, with one bar corresponding to before payday and one bar corresponding to after payday. * Hint: To create the double bars, use `geom_col(position = "dodge")`. You will have to set the `fill()` argument in the `aes()` function.

Question 4: Missing data

Now let's look to see how our average treatment effect and case count changes based on what we decide to do with missing data.

First make a table that shows the mean outcome and case count for `strop_time` by treatment status.

Then replicate the original table dropping all cases that have missing values using `drop_na()`.

Finally, replicate the original table dropping all cases that have missing values for the `cash` variable using a combination of `filter()` and `is.na()`. For this third table, you will need to use the `!` operator to filter your data down to the set of observations for which `cash` is not missing. Remember - we can use the `!` operator to negate logical statements (e.g., `!=` means "not equal to").

Compare the average treatment effects and the case counts across the three tables. What do you notice?

Question 5: BONUS QUESTION - Examining central tendency, range and spread by subgroup

Let's take another look at the distribution of financial resources by income level. Create a table that groups by the low income indicator variable `income_less20k`. Look at the table at the beginning of this activity to get more information about this variable. This indicator variable should be in the rows and you should have the following summary statistics for `accts_amt` in the columns:

- mean
- median
- standard deviation
- minimum value
- maximum value

After creating the table, plot the distribution of `accts_amt` for each income group using `geom_histogram()` and `facet_wrap()`. Add vertical lines for the mean and median on each graph using `geom_vline()` - be sure to make the mean and median lines different colors.

Does the table or the plot give you more detailed information about the distribution of `accts_amt`? When might it be helpful to provide more detailed information and when might it be helpful to provide less detailed information to your audience?