Poverty and Economic Decision-Making

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

Question 1

Load the `poverty.csv` data set. Look at a summary of the `poverty` data set to get a sense of what its variables looks like. Use histograms to examine the univariate distributions of the two financial resources measures: `cash` and `accts_amt`. What can we tell about these variables' distributions from looking at the histograms? Evaluate what the shape of these distributions could imply for the authors' experimental design.

Now, take the *natural logarithm* of these two variables and plot the histograms of these tranformed variables. How does the distribution look now? What are the advantages and disadvantages of transforming the data in this way?

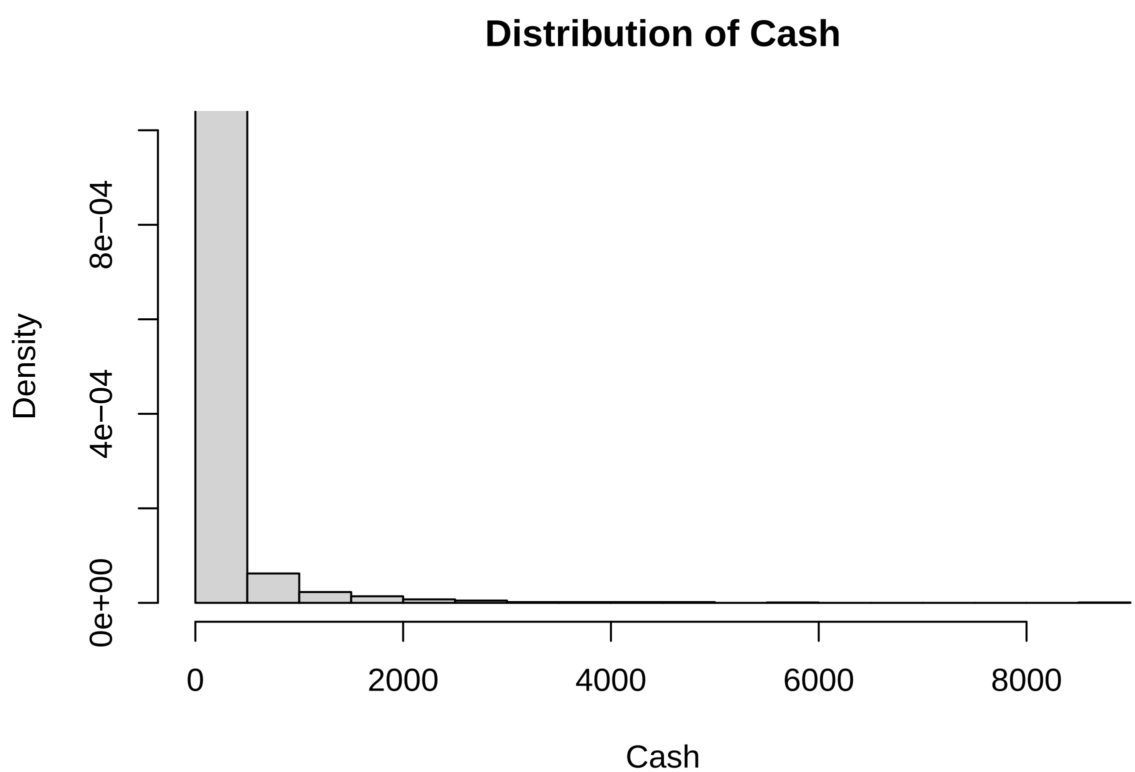
NOTE: Since the natural logarithm of 0 is undefined, researchers often add a small value (in this case, we will use \$1 so that $\log 1 = 0$) to the 0 values for the variables being transformed (in this case, `cash` and `accts_amt`) in order to successfully apply the `log()` function to all values. Be sure to do this recoding only for the purposes of taking the logarithmic transformation – keep the original variables the same.

Answer 1

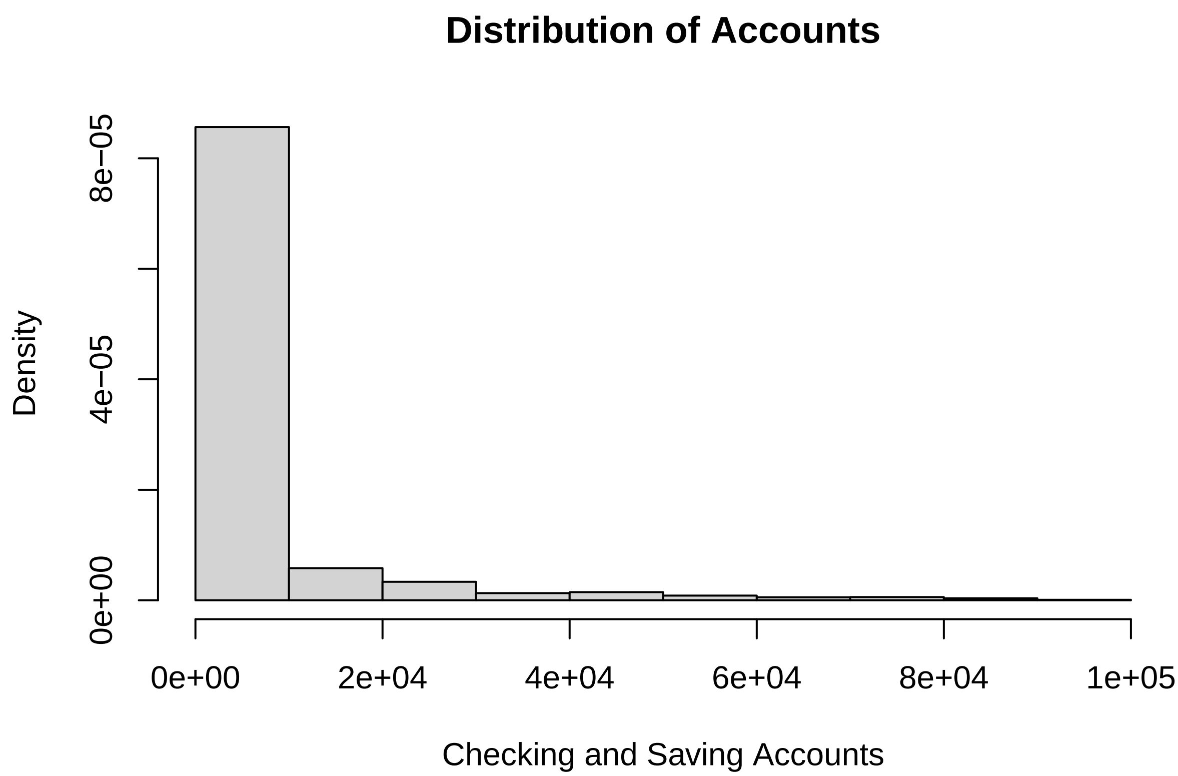
```r
poverty <- read.csv("poverty-1.csv")

summary(poverty)
```

```
##   treatment              cash           accts_amt       stroop_time   
##  Length:2670        Min.   :   0.0   Min.   :    0   Min.   :5.356  
##  Class :character   1st Qu.:  15.0   1st Qu.:  176   1st Qu.:7.436  
##  Mode  :character   Median :  49.5   Median : 1000   Median :7.564  
##                     Mean   : 169.0   Mean   : 6211   Mean   :7.545  
##                     3rd Qu.: 136.2   3rd Qu.: 5000   3rd Qu.:7.692  
##                     Max.   :9000.0   Max.   :95000   Max.   :8.517  
##                     NA's   :226      NA's   :433                    
##  income_less20k  
##  Min.   :0.0000  
##  1st Qu.:0.0000  
##  Median :0.0000  
##  Mean   :0.4139  
##  3rd Qu.:1.0000  
##  Max.   :1.0000  
## 
```

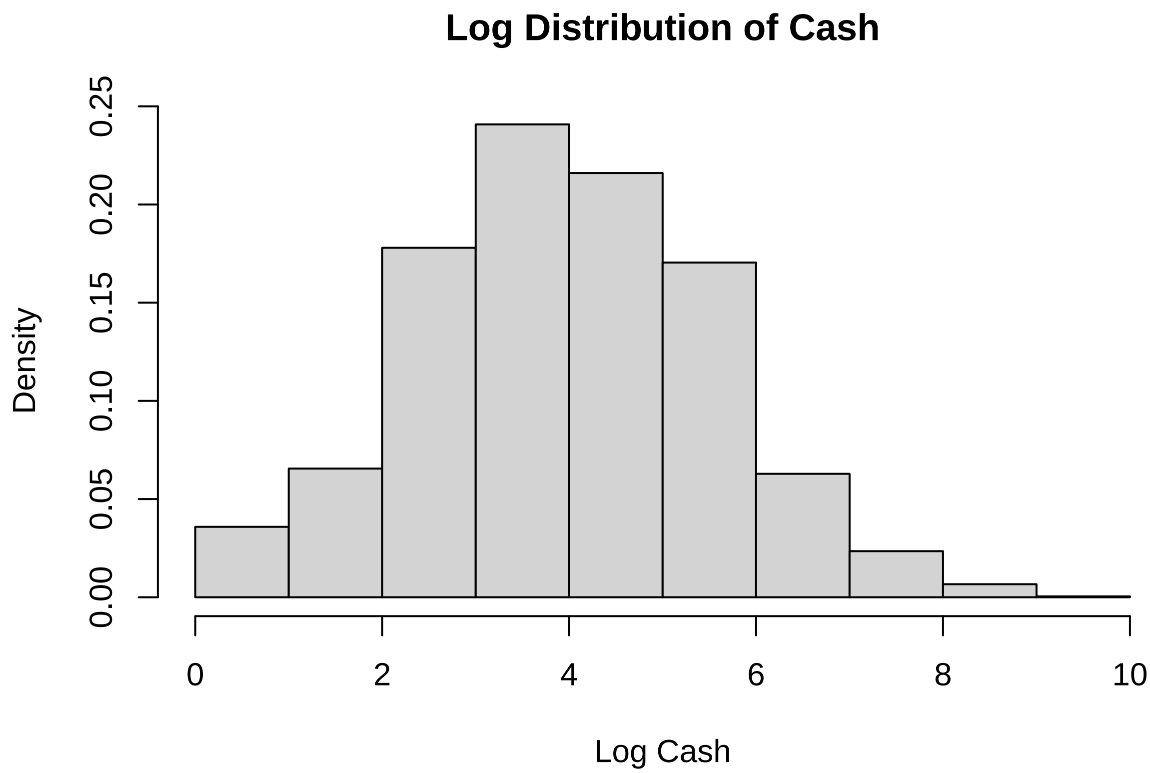
```r
hist(x = poverty$cash, freq = FALSE, ylim = c(0, .001),
     xlab = "Cash", main = "Distribution of Cash")
```



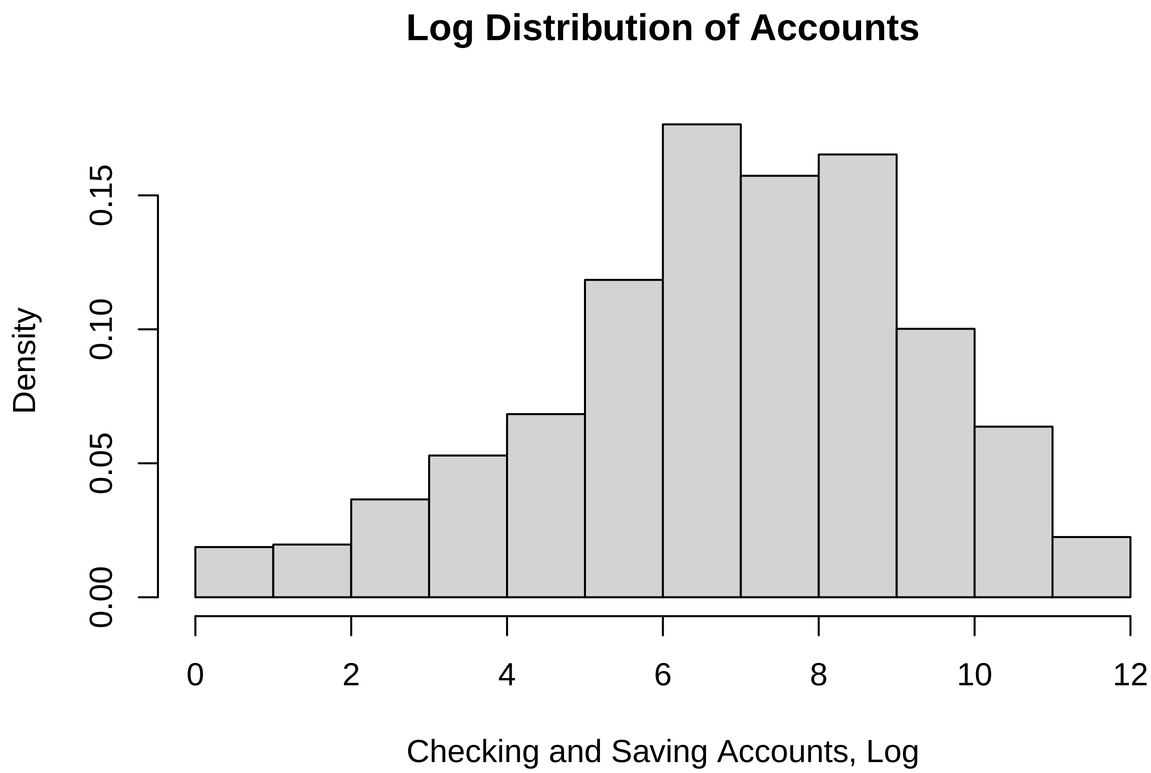
```r
hist(x = poverty$accts_amt, freq = FALSE,
     xlab = "Checking and Saving Accounts", main = "Distribution of Accounts")
```




```r
poverty$log_cash <- log(poverty$cash)

poverty$log_accts <- log(poverty$accts_amt)

hist(x = poverty$log_cash, freq = FALSE,
     xlab = "Log Cash", main = "Log Distribution of Cash")
```




```r
hist(x = poverty$log_accts, freq = FALSE,
     xlab = "Checking and Saving Accounts, Log", main = "Log Distribution of Accounts")
```



Through the two histograms we can see that the largest density occurs in the lowest bins containing responses with low amounts of either cash or account savings. There is an immediate dropoff from the initial brackets, outlining that the majority of respondents do not hold a large amount of savings in either form.

When analyzing logarithmic function of these variables we see that the distribution is still skewed towards a common value in the middle. This verifies that the distribution is focused on one range of variables.

Question 2

Now, let's examine the primary outcome of interest for this study– the effect of a change in financial situation (in this case, getting paid on payday) on economic decision-making and cognitive performance. Begin by calculating the treatment effect for the `stroop_time` variable (a log-transformed variable of the average response time for the stroop cognitive test), using first the mean and then the median. What does this tell you about differences in the outcome across the two experimental conditions?

Secondly, let's look at the relationship between finanical circumstances and the cognitive test variable. Produce two scatter plots side by side (hint: use the par(mfrow)) before your plot commands to place graphs side-by-side), one for each of the two experimental conditions, showing the bivariate relationship between your *log-transformed* `cash` variable and the amount of time it took subjects to complete the stroop cognitive test administered in the survey (`stroop_time`). Place the `stroop_time` variable on the y-axis. Be sure to title your graphs to differentiate between the `Before Payday` and `After Payday` conditions. Now do the same, for the *log-transformed* `accts_amt` variable.

Briefly comment on your results in light of the hypothesis that changes in economic circumstances will influence cognitive performance.

Answer 2

```r
before <- subset(poverty, subset = (treatment == "Before Payday"))
after <- subset(poverty, subset = (treatment == "After Payday"))

#stroop_diff <- mean(after$stroop_time) - mean(before$stroop_time) * 100

stroop_diff <- (mean(after$stroop_time) - mean(before$stroop_time)) * 100
cat("SATE Mean:")
```

```
## SATE Mean:
```

```r
stroop_diff
```

```
## [1] 1.142325
```

```r
cat("\n")
```

```r
stroop_diff_median <- (median(after$stroop_time) - median(before$stroop_time)) * 100
cat("SATE Median:")
```

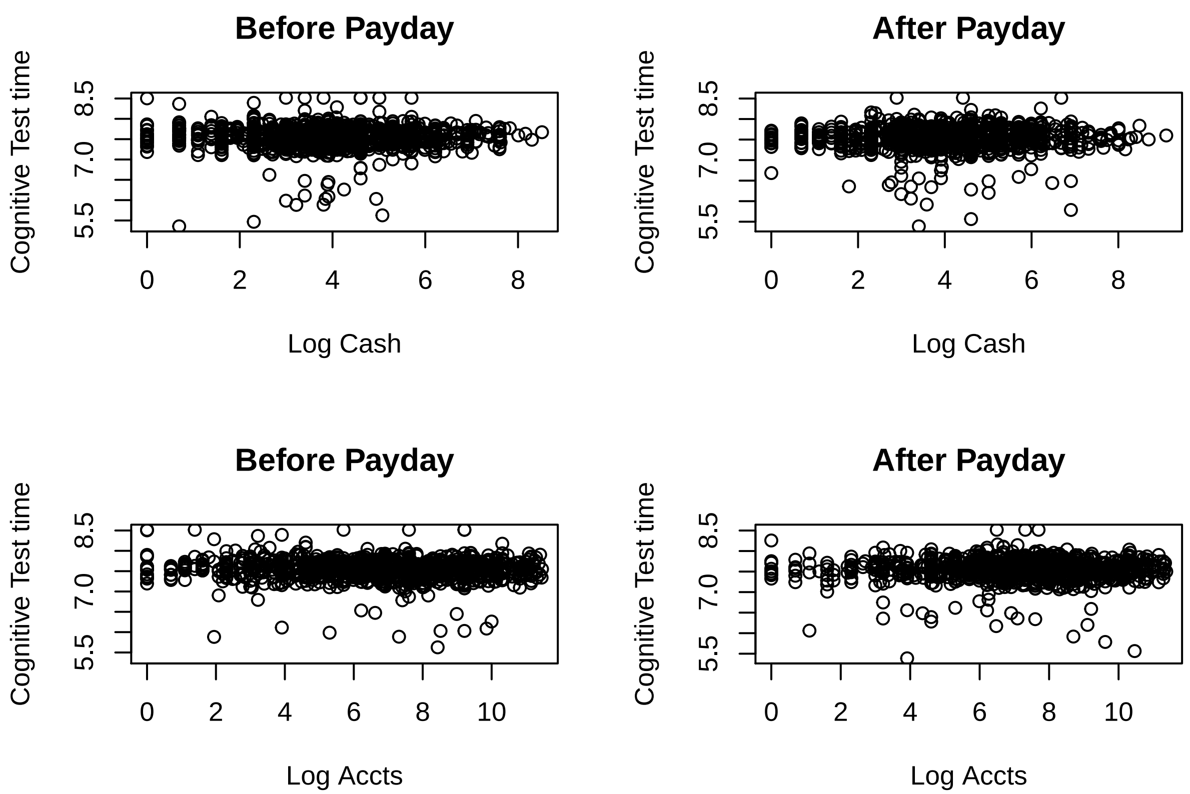
```
## SATE Median:
```

```r
stroop_diff_median
```

```
## [1] 1.430528
```

When using either median or mean we see that their is a rise in stroop test after payday occurs.

```r
par(mfrow=c(2,2))

plot(x = before$log_cash, y = before$stroop_time,
     xlab = "Log Cash",
     ylab = "Cognitive Test time",
     main= "Before Payday")

plot(x = after$log_cash, y = after$stroop_time,
     xlab = "Log Cash",
     ylab = "Cognitive Test time",
     main= "After Payday")

plot(x = before$log_accts, y = before$stroop_time,
     xlab = "Log Accts",
     ylab = "Cognitive Test time",
     main= "Before Payday")

plot(x = after$log_accts, y = after$stroop_time,
     xlab = "Log Accts",
     ylab = "Cognitive Test time",
     main= "After Payday")
```



We see that in both distributions the after payday cognitive test times are *slightly* lower than the test times before payday.This makes sense because as people have more available money, they likely have less cognitive stress on themselves.

Question 3

Now, let's take a closer look at whether or not the `Before Payday` versus `After Payday` treatment created measurable differences in financial circumstances. What is the effect of payday on participants' financial resources? To help with interpretability, use the original variables `cash` and `accts_amt` to calculate this effect. Calculate both the mean and median effect. Does the measure of central tendency you use affect your perception of the effect?

Answer 3

```
treatment_effect_cash_mean <- mean(before$cash, na.rm = TRUE) - mean(after$cash, na.rm = TRUE)

treatment_effect_cash_median <- median(before$cash, na.rm = TRUE) - median(after$cash, na.rm = TRUE)

treatment_effect_accts_mean <- mean(before$accts_amt, na.rm = TRUE) - mean(after$accts_amt, na.rm = TRU

#treatment_effects_accts_median <- median(before$accts_amt, na.rm = TRUE) - median(after$accts_amt, na.

treatment_effect_accts_median <- median(before$accts_amt, na.rm = TRUE) - median(after$accts_amt, na.rm
```

```r
cat("Treatment Cash Mean is: ", treatment_effect_cash_mean, "\n")
```

```
## Treatment Cash Mean is:  -36.77511
```

```r
cat("Treatment Cash Median is: ", treatment_effect_cash_median, "\n")
```

```
## Treatment Cash Median is:  -10
```

```r
cat("Treatment Accts Mean is: ", treatment_effect_accts_mean, "\n")
```

```
## Treatment Accts Mean is:  251.9087
```

```r
cat("Treatment Accts Median is: ", treatment_effect_accts_median, "\n")
```

```
## Treatment Accts Median is:  -450
```

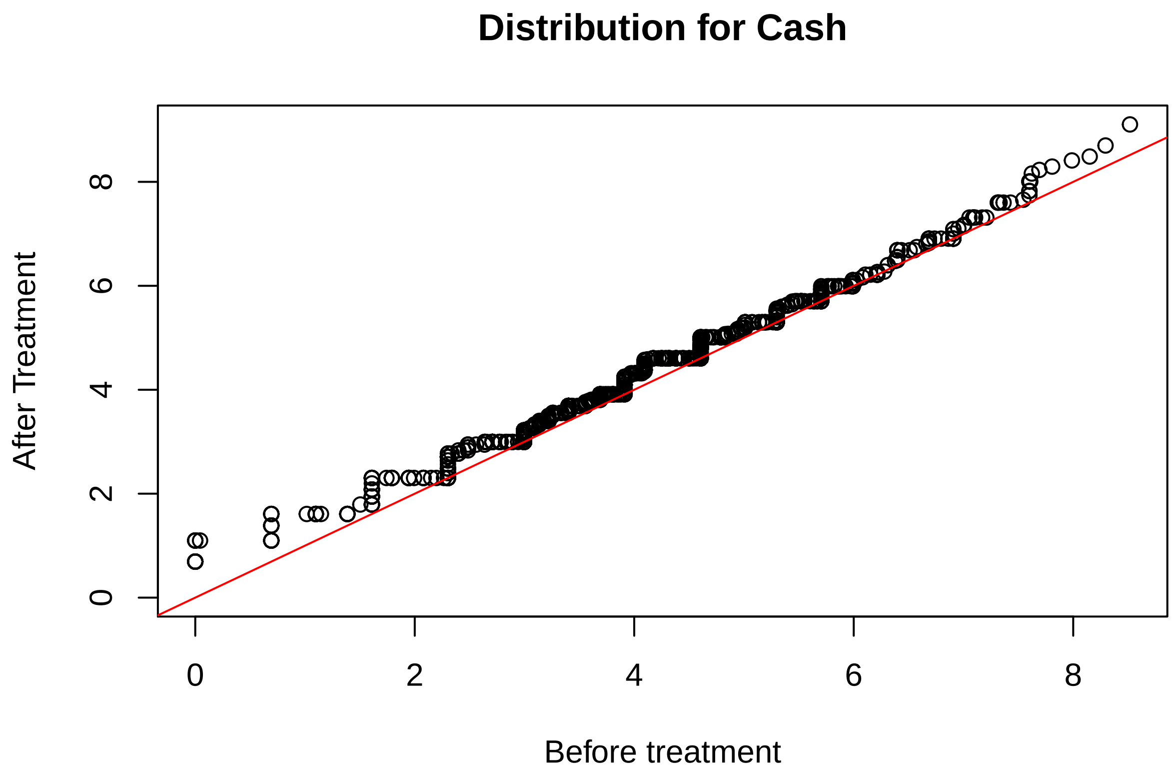
This shows that the outliers have quite a large effect on the central tendency in the mean calculation as their is a large gap between the mean and the median numbers. Additionally this shows that the treatment does have quite a large affect on the before-after dynamic of these variables.

Question 4

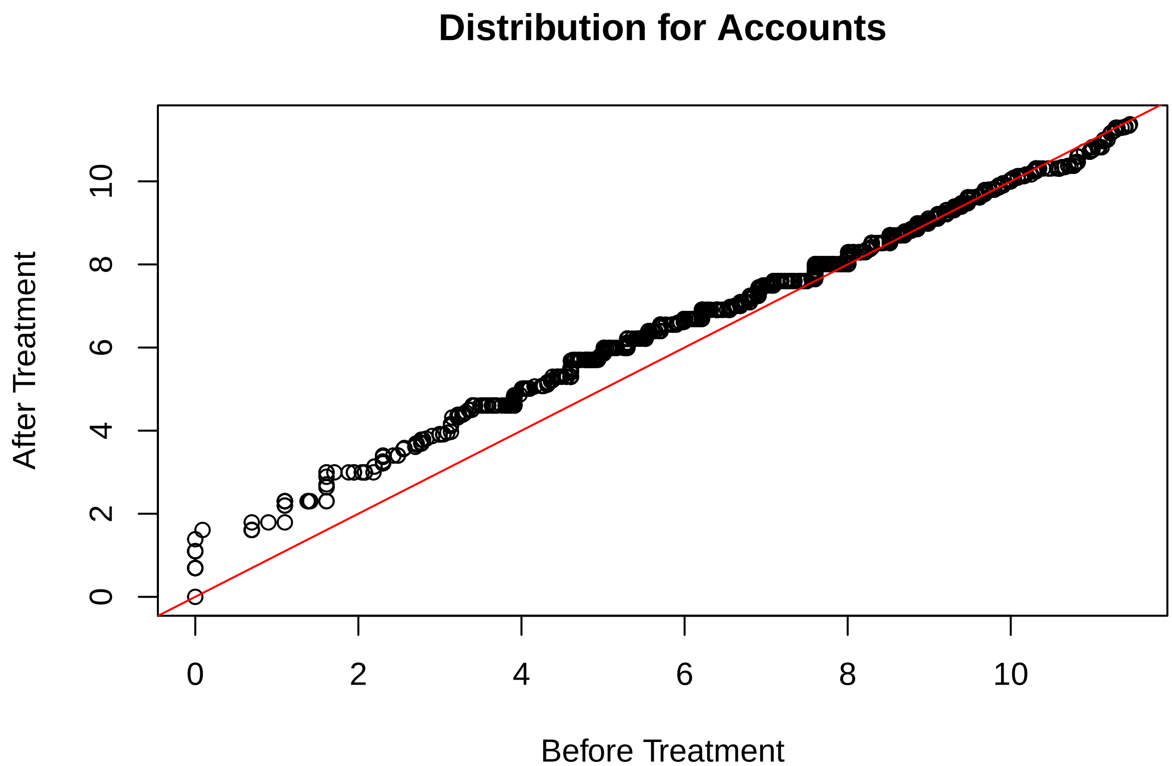
Compare the distributions of the `Before Payday` and `After Payday` groups for the *log-transformed* `cash` and `accts_amt` variables. Use quantile-quantile plots to do this comparison, and add a 45-degree line in a color of your choice (not black). Briefly interpret your results and their implications for the authors' argument that their study generated variation in financial resources before and after payday. When appropriate, state which ranges of the outcome variables you would focus on when comparing decision-making and cognitive capacity across these two treatment conditions.

Answer 4

```r
qqplot(before$log_cash, after$log_cash,
       xlab = "Before treatment",
       ylab = "After Treatment",
       main = "Distribution for Cash")
abline(a = 0, b = 1, col = "Red")
```




```r
qqplot(before$log_accts, after$log_accts,
       xlab = "Before Treatment",
       ylab = "After Treatment",
       main = "Distribution for Accounts")
abline(a = 0, b = 1, col = "Red")
```



These plots show that their is a relationship between the amount of money that respondents have in cash or accounts before and after payday. With this information, we can see that the majority of respondents are in the upper section of these graphs, as can be seen by a much thicker distribution. With this knowledge we could test cognitive speed in this range to discover the outcomes of these treatment conditions.

Question 5

In class, we covered the difference-in-difference design for comparing average treatment effects across treatment and control groups. This design can also be used to compare average treatment effects across different ranges of a *pre-treatment variable*- a variable that asks about people's circumstances before the treatment and thus could not be affected by the treatment. This is known as *heterogeneous treatment effects* – the idea that the treatment may have differential effects for different subpopulations. Let's look at the pre-treatment variable `income_less20k`. Calculate the treatment effect of Payday on amount in checking and savings accounts separately for respondents earning more than 20,000 dollars a year and those earning less than 20,000 dollars. Use the original `accts_amt` variable for this calculation. Then take the difference between the effects you calculate. What does this comparison tell you about how payday affects the amount that people have in their accounts? Are you convinced by the authors' main finding from Question 2 in light of your investigation of their success in manipulating cash and account balances before and after payday?

Answer 5

```
# 1 if respnodent earns less than 20K

more_20K <- subset(poverty, subset = (poverty$income_less20k == 0))
```

```r
less_20K <- subset(poverty, subset = (poverty$income_less20k == 1))

more_before <- subset(more_20K, subset = (treatment == "Before Payday"))

more_after <- subset(more_20K, subset = (treatment == "After Payday"))

less_before <- subset(less_20K, subset = (treatment == "Before Payday"))

less_after <- subset(less_20K, subset = (treatment == "After Payday"))

treatment_more <- mean(more_before$accts_amt, na.rm = TRUE) - mean(more_after$accts_amt, na.rm = TRUE)
treatment_more
```

```
## [1] 649.9476
```

```r
treatment_less <- mean(less_before$accts_amt, na.rm = TRUE) - mean(less_after$accts_amt, na.rm = TRUE)
treatment_less
```

```
## [1] -81.0914
```

```r
(difference <- treatment_more - treatment_less)
```

```
## [1] 731.039
```

We see that through this calculation, payday has a much higher impact on the average amount in accounts for those that had more to begin with than it did for those who had less20K. This is likely because they have more money or larger paychecks to handle after payday. In question 2 we discussed that those who had more money in their accounts would have less stress. Combined with the findings here, I do not beleive the data is clear enough to state a conclusion on this.