Hypothesis Testing

Brayden Yates

University of the Cumberlands

MSDS-530: Fundamentals of Data Science

Dr. Jason Turner

10 February 2024

Hypothesis Testing

For this project, I chose a dataset of particular interest to me and my personal goals. This dataset is the Reddit Survey on Financial Independence. Those in the FIRE (Financial Independence, Retire Early) community are typically frugal individuals who very closely track all aspects of their personal finances. This means the dataset may be more complete and have more precision than other personal-finance related datasets.

For my one-sample t-test, I chose to analyze luxury expenses and necessary expenses. Here is the code needed for this analysis:

```
import pandas as pd

import numpy as np

from scipy.stats import ttest_1samp

df = pd.read_csv('reddit_finance.csv')

df

df = df.dropna(subset=['2020_lux_exp', '2020_necessities_exp'])

luxuries = df['2020_lux_exp'].astype(float)

necessary_expenses = df['2020_necessities_exp'].astype(float)

t_statistic, p_value = ttest_1samp(luxuries,

popmean=necessary_expenses.mean())

print("t-statistic:", t_statistic)

print("p-value:", p_value)
```

```
t-statistic: -0.09388963223932237
p-value: 0.9252079658303848
```

Based on these results, we can derive some interesting insights. The mean of luxury expenses is lower than the mean of the necessary expenses, but the magnitude of this is small. With regards to the p-value, a higher p-value indicates that there is weak evidence against the null hypothesis. Our P-value of 0.925 indicates that there is a very high probability of observing this if the null hypothesis is true. We are unable to reject the null hypothesis, meaning that there is not enough evidence to indicate that luxury expenses differ significantly from necessary expenses for the entire populations. This also may indicate that individuals in the FIRE community tend to balance luxury and essential spending equally in this dataset.

We can examine these two attributes with more precision by evaluating them in the context of a paired sample test. Here is the code for that test:

```
t_statistic, p_value = ttest_rel(luxuries, necessary_expenses)

print("t-statistic:", t_statistic)

print("p-value:", p_value)
```

```
t-statistic: -0.11681845617899224

p-value: 0.9070178063872212
```

If we examine this data, we see in the t-statistic that luxury expenses do tend to be lower than necessary expenses, typically by a larger margin than in the one sample test. However, it is also indicated by the p-value of 0.907 that there is a significantly high probability of seeing this if the null hypothesis were true—

that there is no significant difference between luxury expenses and necessary expenses. With this analysis, we are unable to reject this null hypothesis.

For our independent two sample t-test, I used the following code:

```
fin_indy_yes = df[df['fin_indy'] == 'Yes']

fin_indy_no = df[df['fin_indy'] == 'No']

gross_income_yes = fin_indy_yes['2020_gross_inc'].dropna()

gross_income_no = fin_indy_no['2020_gross_inc'].dropna()

necessary_expenses_yes =
fin_indy_yes['2020_necessities_exp'].dropna()

necessary_expenses_no = fin_indy_no['2020_necessities_exp'].dropna()

t_stat_gross_income, p_value_gross_income =
ttest_ind(gross_income_yes, gross_income_no)

t_stat_necessary_expenses, p_value_necessary_expenses =
ttest_ind(necessary_expenses_yes, necessary_expenses_no)

print("Independent Samples T-test Results for 2020 Gross Income:")

print("t-statistic:", t_stat_gross_income)

print("p-value:", p_value_gross_income)



print("\nIndependent Samples T-test Results for 2020 Necessary
Expenses:")
```

```
print("t-statistic:", t_stat_necessary_expenses)

print("p-value:", p_value_necessary_expenses)
```

```
Independent Samples T-test Results for 2020 Gross Income:
t-statistic: 4.405783957323624
p-value: 1.1205089638084743e-05

Independent Samples T-test Results for 2020 Necessary Expenses:
t-statistic: -0.30951032414298646
p-value: 0.7569715610579175
```

With regards to 2020 gross income, the t-statistic is 4.40. This indicates there is a significant difference in 2020 gross income between financially independent individuals and those who are not. The p-value is very low—1.121*10^-5. This indicates there is strong independence against the null hypothesis. Because of this, we can reject the null hypothesis concluding that there is a significant difference in 2020 income between those who are and are not Financially Independent. With regards to necessary expenses, we have a t-statistic of -0.31 and a very high p-value of 0.75. This indicates that there is not a significant difference between financially-independent and non-financially-independent individuals. However, the high p-value shows that there is not enough evidence to conclude that there is a significant difference in 2020 necessary expenses between financially-independent and financially-dependent individuals.

I had a few assumptions before beginning this analysis. First, I assumed that those in the FIRE community had very low luxury expenses when compared to their essential expenses. The data shows that this is untrue. A likely reason for this could be that the income of individuals in the FIRE community is able to support this amount of luxury spending, while still pursuing goals of early retirement. A more likely conclusion is that individuals in this community tend to be very frugal and spend little money on essentials. Members of this community may also categorize things as luxuries that non-FIRE community members would consider to be essentials or some other category. I also assumed that gross income would

be dramatically reduced when one is financially-independent as compared to those who are not. Financially independent individuals have enough in savings to make working unnecessary. This means that they are free to pursue any other interest they like, even if that means a reduction in income. The data agreed with this assumption.

References

Utkarsh Singh. (n.d.). *Reddit Survey on Financial Independence*. Kaggle.

https://www.kaggle.com/datasets/utkarshx27/reddit-survey-on-financial-independence