NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

Independent Samples T Tests Using Python (ipynb)

F. A. Rosales
Department of Mathematics, Far Eastern University
2021044631@feu.ph / francesr1001 (github.com)

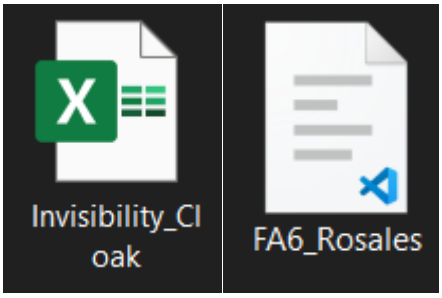
Abstract

Creating a driven data that will help the user analyze the existing code of the said data. As this paper aims to get the independent samples t-test for Invisibility Cloak data set, it requires a multiple and complex formula and packages in Python to gather the data asked such as; validity, mode, mean, median, standard deviation, variance, skewness, standard error skewness, kurtosis, and other quantile and percentile. Therefore, by the step-by-step and introductions of this paper would guilde you to unravel and understand the Independent Sample T-Test for Invisibility Cloak.

File Used:

Excel File: Invisibility_Cloak.xlsx

Python File: FA6_Rosales.ipynb



Using the two file, which I manipulate and use my data. As mentioned, I also use multiple of packages that I will discuss later.

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

List of Sections:

The ipynb file consist of 5 sections, named below.

Introductions: Data of Invisibility Cloak Excel

Cloak Summary Data: List of Data of Cloak

Mischief Summary Data: List of Data of Mischief

Welch Two Sample: Sample T-Test

Mean Differences: Differences of our Mean

In reference in FA6_Rosales.ipynb

Introduction

Introducing the main data, "Invisibility_Cloak.xlsx" consist of the required Data of Cloak, and Mischief.

	Participant	Cloak	Mischief
0	1	0	3
1	2	0	1
2	3	0	5
3	4	0	4
4	5	0	6
5	6	0	4
6	7	0	6
7	8	0	2
8	9	0	0
9	10	0	5
10	11	0	4
11	12	0	5
12	13	1	4
13	14	1	3
14	15	1	6
15	16	1	6
16	17	1	8
17	18	1	5
18	19	1	5
19	20	1	4
20	21	1	2
21	22	1	5
22	23	1	7
23	24	1	5

Panda a package is used in this section that will extract the file excel.

ASSUMPTION#1:

Since the data in our Mischief are greater, then we can get a list of data there that would be greater than Cloak.

ASSUMPTION#2:

The quantile 1 of Cloak would be Zero since the preceding are zeros. Possibly or probably, quartiles 3 would be 1 since everything is also 1.

In reference in FA6_Rosales.ipynb

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

Cloak Summary Data

This section on the other hand introduces the data of Cloak in the existing excel data.

	Cloak Summary Data	Data	Score
0		Valid	24.000
1		Mode	0.000
2		Mean	0.500
3		Median	0.500
4		Standard Deviation	0.500
5		Variance	0.250
6		Skewness	0.000
7		Standard Error of Skewness	0.472
8		Kurtosis	-2.000
9		Q1	0.000
10		Q2	0.500
11		Q3	1.000
12		D9	1.000
13		P95	1.000

Panda is also used here, to extract values of Column Cloak.

Numphy to be able to manipulate numerical data.

Such as helps to get the following

```
from scipy.stats import skew, kurtosis,
iqr, scoreatpercentile
```

```
mean_value = np.mean(list_num)
median_value = np.median(list_num)
mode_value = list_num.mode().iloc[0]
std = np.std(list_num)
variance = np.var(list_num)
skewness_value = skew(list_num)
kurtosis_value = kurtosis(list_num)

q1 = np.percentile(list_num, 25)
q2 = np.percentile(list_num, 50)
q3 = np.percentile(list_num, 75)

iqr_value = iqr(list_num)

d9 = scoreatpercentile(list_num, 90)
p95 = scoreatpercentile(list_num, 95)
```

The formula and equation used in python to gather our data is used.

Name Cloak as list_num

In reference in FA6_Rosales.ipynb

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

Mischief Summary Data

This section on the other hand introduces the data of Mischief in the existing excel data.

	Mischief Summary Data	Data	Score
0		Valid	24.000
1		Mode	5.000
2		Mean	4.375
3		Median	5.000
4		Standard Deviation	1.821
5		Variance	3.318
6		Skewness	-0.482
7		Standard Error of Skewness	0.472
8		Kurtosis	0.089
9		Q1	3.750
10		Q2	5.000
11		Q3	5.250
12		D9	6.000
13		P95	6.850

Followed by the summary of cloak, the same packages can be also used to gather these data.

```
list_num2 = data_sheet["Mischief"]
mean_value2 = np.mean(list_num2)
median_value2 = np.median(list_num2)
mode_value2 = list_num2.mode().iloc[0]
std2 = np.std(list_num2)
variance2 = np.var(list_num2)
skewness_value2 = skew(list_num2)
kurtosis_value2 = kurtosis(list_num2)
q1_2 = np.percentile(list_num2, 25)
q2_2 = np.percentile(list_num2, 50)
q3_2 = np.percentile(list_num2, 75)
iqr_value2 = iqr(list_num2)

d9_2 = scoreatpercentile(list_num2, 90)
p95_2 = scoreatpercentile(list_num2, 95)
```

The formula and equation used in python to gather our data is used.

We name Mischief as list_num2.

In reference in FA6_Rosales.ipynb

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

COMPARISION & ANALYSIS

ASSUMPTION#3

As assumed, since we said the data of Mischief are indeed greater than in Cloak.

We also assumed if we get the difference of cloak to mischief it will be NEGATIVE.

ASSUMPTION #4

As data shown below, we can assume we, as well can get the data for Welch Two Sample T-Test easily as we all have the data equality, and we do not need to assume and change blank data as NA.

ASSUMPTION #5

95% confidence would still be used in this data for 1.96 of a 95% confidence interval) to provide a range within which the true difference in means is likely to fall.

	Cloak Summary Data	Data	Score
0		Valid	24.000
1		Mode	0.000
2		Mean	0.500
3		Median	0.500
4		Standard Deviation	0.500
5		Variance	0.250
6		Skewness	0.000
7		Standard Error of Skewness	0.472
8		Kurtosis	-2.000
9		Q1	0.000
10		Q2	0.500
11		Q3	1.000
12		D9	1.000
13		P95	1.000

	Mischief Summary Data	Data	Score
0		Valid	24.000
1		Mode	5.000
2		Mean	4.375
3		Median	5.000
4		Standard Deviation	1.821
5		Variance	3.318
6		Skewness	-0.482
7		Standard Error of Skewness	0.472
8		Kurtosis	0.089
9		Q1	3.750
10		Q2	5.000
11		Q3	5.250
12		D9	6.000
13		P95	6.850

In comparison, the q1 in Cloak is INDEED 0 since the preceding data of our sheet are all zero. Q3 will be 1, while on the other hand since the data in Mischief are random from 1 to 8, then the data would just surround mentioned number.

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

Welch Two Sample t-test

This section on the other hand introduces the data of our Welch Two Sample t-test.

OUTPUT

```
Welch Two Sample t-test:
t = -9.839, df = 23.0, p-value = 2.49e-10
Alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval: (-10.610719, -9.066827)
Sample estimates:
Mean in Cloak: 0.5
Mean in Mischief: 4.375
```

The help of packges

```
from scipy.stats import skew, kurtosis,
iqr, scoreatpercentile
```

helps us to gather the shown data.

In reference in FA6_Rosales.ipynb

The margin of error for the estimated difference in means is computed using the same procedure that is used to get the confidence interval in the Welch Two Sample t-test. It makes use of the **t-statistic**, which is calculated by dividing the observed mean difference by the standard error of the difference. To calculate the margin of error, multiply the standard error by the critical value, which is 1.96 for a 95% confidence interval. This yields a range that represents the likelihood that the true mean difference will lie inside.

```
t_statistic, p_value = ttest_ind(list_num, list_num2, equal_var=False)

ci_low, ci_high = t_statistic - 1.96 * ((list_num.var()/len(list_num)) + (list_num2.var()/len(list_num2)))**0.5, \
                  t_statistic + 1.96 * ((list_num.var()/len(list_num)) + (list_num2.var()/len(list_num2)))**0.5

print("Welch Two Sample t-test:")
print(f"t = {t_statistic:.3f}, df = {min(len(list_num)-1, len(list_num2)-1):.1f}, p-value = {p_value:.2e}")
print("Alternative hypothesis: true difference in means is not equal to 0")
print(f"95 percent confidence interval: ({ci_low:.6f}, {ci_high:.6f})")
print("Sample estimates:")
print(f"Mean in Cloak: {list_num.mean()}")
print(f"Mean in Mischief: {list_num2.mean()}")
```

In reference in FA6_Rosales.ipynb

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

Mean Differences ES

This section on the other hand introduces the data of our Mean Differences.

OUTPUT

```
Mean Difference ES: -3.88 [-4.58, -3.25]
Variance of mean difference: 2.94
P-value of mean difference: 0.0000
```

These are still the packages can be used using stats package.

```
from scipy.stats import skew, kurtosis,
iqr, scoreatpercentile
```

helps us to gather the shown data.

In reference in FA6_Rosales.ipynb

The difference in means divided by the pooled standard deviation is used to get the mean difference effect size, or Cohen's d. Using **bootstrapping**, the confidence interval for the mean difference is computed, yielding a range of tenable values for the actual effect size. The p-value is the probability of finding such a difference in the absence of a true effect. The paired t-test is used to determine whether the mean difference is significantly different from zero.

```
mean_diff = pg.compute_effsize(list_num, list_num2, paired=False, eftype='Cohen')

mean_difference = list_num.mean() - list_num2.mean()

ci_low, ci_high = pg.compute_bootci(list_num - list_num2, func=np.mean, confidence=0.95)

t_statistic, p_value = ttest_rel(list_num, list_num2)

print(f"Mean Difference ES: {mean_difference:.2f} [{ci_low:.2f}, {ci_high:.2f}]")
print(f"Variance of mean difference: {variance_mean_difference:.2f}")
print(f"P-value of mean difference: {p_value:.4f}")
```

In reference in FA6_Rosales.ipynb

NAME: Frances Aneth C. Rosales
APM1111 – STATISTICS
COURSE: BSAM DATSCI (2021)

DATE: November 11, 2023

CONCLUSION

As gathered data computed in python, The Cloak Invisibility Data Set were able to show by the used of essential packages and formula in Python. Results were shown that the Mean Difference of Cloak and Mischief is pertains as -3.88 [-4.58, -3.25]. Such that, the variance of mean difference is 2.94 and P-Value is 0. The assumption highlighted is also shown such as.

Assumption 1 data in our Mischief are greater hence, those data under it are greater than cloak. Another one is the zero of our quartiles. With our data shown t = -9.839, df = 23.0, p-value = 2.49e-10 as Welch Two Sample t-test were also gathered by the use of python and formula.