Project 2
Lewis Collum
Updated: March 30, 2020

1

We want to see if the mean whiteness of high hydrogen peroxide samples, is greater than the mean whiteness of low hydrogen peroxide samples.

$$H_A : \mu_H > \mu_L$$

$$H_0 : \mu_H \leq \mu_L$$

Python

```
import pandas
import scipy.stats as stats

data = pandas.read_csv('data/1.csv')

def whitenessFromPeroxideLevel(level):
    rowIndices = data['Hydrogen Peroxide'] == level
    rows = data[rowIndices]
    return rows['Whiteness'].tolist()

whiteness = {
    'lowPeroxide': whitenessFromPeroxideLevel('low'),
    'highPeroxide': whitenessFromPeroxideLevel('high')
}

_, p = stats.ttest_ind(
    whiteness['highPeroxide'],
    whiteness['lowPeroxide'])

print(f"\[\\textnormal{{p-value}} = {round(p/2, 3)}\]")
```

$$\textnormal{p-value} = 0.027$$

Conclusion

Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** We are 95% confident that higher level results in whiter garments.

2

$$H_A : \mu_{20} < \mu_{10}$$

$$H_0 : \mu_{20} \geq \mu_{10}$$

Python

This function is used in the following questions.

```
def sliceByUniquesInColumn(dataframe, sliceColumn, onEachSlice):
    slices = {}
    for unique in dataframe[sliceColumn].unique():
        matchingRowIndices = dataframe[sliceColumn] == unique
        slices[unique] = onEachSlice(dataframe[matchingRowIndices])
    return slices
```

```
import pandas
import scipy.stats as stats
from p2 import sliceByUniquesInColumn

percentByPressure = sliceByUniquesInColumn(
    dataframe = pandas.read_csv('data/2.csv'),
    sliceColumn = 'Roller Pressure',
    onEachSlice = lambda slice: list(slice['Percent Pickup']))

_, p = stats.ttest_ind(
    percentByPressure[10],
    percentByPressure[20])

print(f"\[\\textnormal{{p-value}} = {round(p/2, 5)}\]")
```

$$\textnormal{p-value} = 0.00303$$

Conclusion

Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** We are 95% confident that more dense fabric is less absorbent.

3

$$H_A : \mu_{\textnormal{new}} > \mu_{\textnormal{standard}}$$

$$H_0 : \mu_{\textnormal{new}} \leq \mu_{\textnormal{standard}}$$

Python

```
import pandas
import scipy.stats as stats
from p2 import sliceByUniquesInColumn

percentByPressure = sliceByUniquesInColumn(
    dataframe = pandas.read_csv('data/3.csv'),
    sliceColumn = 'Procedure',
    onEachSlice = lambda slice: list(slice['Breaking Strength']))

_, p = stats.ttest_ind(
    percentByPressure['new'],
    percentByPressure['standard'])

print(f"\[\\textnormal{{p-value}} = {round(p/2, 5)}\]")
```

$$\textnormal{p-value} = 0.00326$$

Conclusion

Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** We are 95% confident that the new procedure has a larger breaking strength on average than the standard procedure.

4

$$H_A : \mu_1 \neq \mu_2$$

$$H_0 : \mu_1 = \mu_2$$

Python

```
import pandas
import scipy.stats as stats
from p2 import sliceByUniquesInColumn

percentByPressure = sliceByUniquesInColumn(
    dataframe = pandas.read_csv('data/4.csv'),
    sliceColumn = 'Joystick',
```

```
    onEachSlice = lambda slice: list(slice['Mean Error']))

_, p = stats.ttest_ind(
    percentByPressure[1],
    percentByPressure[2])

print(f"\[\\textnormal{{p-value}} = {round(p, 5)}\]")
```

$$\text{p-value} = 0.3042$$

Conclusion
Since the two-tailed p-value is greater than α (0.05), **we fail to reject the null hypothesis.**

5

$$H_A : \mu_{\text{after}} < \mu_{\text{before}}$$

$$H_0 : \mu_{\text{after}} \geq \mu_{\text{before}}$$

Python

```
import pandas
import scipy.stats as stats
from p2 import sliceByUniquesInColumn

percentByPressure = sliceByUniquesInColumn(
    dataframe = pandas.read_csv('data/5.csv'),
    sliceColumn = 'Green Management Procedures',
    onEachSlice = lambda slice: list(slice['Damaged Inventory (%)']))

_, p = stats.ttest_ind(
    percentByPressure['Before'],
    percentByPressure['After'])

print(f"\[\\textnormal{{p-value}} = {round(p/2, 5):f}\]")
```

$$\text{p-value} = 0.000030$$

Conclusion
Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** We are 95% confident that green management techniques have significantly improved practices.

6

$$H_A : \mu < 0.5\%$$

$$H_0 : \mu \geq 0.5\%$$

Python

```
import pandas
import scipy.stats as stats

csv = pandas.read_csv('data/6.csv')
percentChangeInWeight = list(csv['Weight Gain (%)'])

_, p = stats.ttest_1samp(
    percentChangeInWeight,
    0.5)

print(f"\[\\textnormal{{p-value}} = {round(p/2, 5):f}\]")
```

$$\text{p-value} = 0.000010$$

Conclusion
Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** We are 95% confident that this material experiences less than 0.5% average weight gain for this type of diffusion.

7

$$H_A : \mu > 3.50$$

$$H_0 : \mu \leq 3.50$$

Python

```
import pandas
import scipy.stats as stats

csv = pandas.read_csv('data/7.csv')
densities = list(csv['Densities'])

_, p = stats.ttest_1samp(densities, 3.5)

print(f"\[\\textnormal{{p-value}} = {round(p/2, 5):f}\]")
```

$$\text{p-value} = 0.028790$$

Conclusion
Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** We are 95% confident that the average density is larger than 3.50.

8

$$H_A : p < 0.10$$

$$H_0 : p = 0.10$$

$$\alpha = 0.05$$

Python

```
import pandas
import scipy.stats as stats

p = stats.binom.cdf(
    k = 23,
    n = 324,
    p = 0.1)

print(f"\[\\textnormal{{p-value}} = {round(p, 3)}\]")
```

$$\text{p-value} = 0.045$$

Conclusion
Since the one-tailed p-value is less than α (0.05), **we reject the null hypothesis.** The evidence suggests that the positive result being incorrect is less than 10%.

9

$$H_A : p < 0.05$$

$$H_0 : p = 0.05$$

$$\alpha = 0.05$$

Python

```
import pandas
import scipy.stats as stats

defectiveAcceptance = 0.05
sampleSize = 200
defective = 8

p = stats.binom.cdf(
    k = defective,
    n = sampleSize,
    p = defectiveAcceptance)

print(f"\[\\textnormal{{p-value}} = {round(p, 3)}\]")
```

$$\textnormal{p-value} = 0.327$$

Conclusion

Since the one-tailed p-value is greater than α (0.05), **we fail to reject the null hypothesis.** There is no evidence to show that fewer than 5% are defective.

10

$$H_A : p_H > p_L$$

$$H_0 : p_H = p_L$$

$$\alpha = 0.05$$

Python

```
import pandas
import statsmodels.stats.proportion as proportion

_, p = proportion.proportions_ztest(
    count = [13, 20],
    nobs = [62, 70])

print(f"\[\\textnormal{{p-value}} = {round(p, 3)}\]")
```

$$\textnormal{p-value} = 0.314$$

Conclusion

Since the one-tailed p-value is greater than α (0.05), **we fail to reject the null hypothesis.** There is no evidence to show that larger percentage of insulators will break down at higher temperatures.