Questions For Investigation

As a general note, be sure to keep a record of any resources that you use or refer to in the creation of your project. You will need to report your sources as part of the project submission.

1. What is our independent variable? What is our dependent variable?

Independent Variable - Words Condition congruent / incongruent

Dependent Variable - Time taken to name the ink colors in equally-sized lists

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

1. Ho - NULL HYPOTHESES - Whether no difference in time to name the ink colors for congruent as well as incongruent words. ($\mu_i - \mu_c = 0$)
2. Ha - ALTERNATE HYPOTHESES - Whether there is a time difference (+ve or -ve) to name the ink colors for congruent as well as incongruent words. ($\mu_i - \mu_c \neq 0$)

μ_i - Population Mean of the time taken by the subjects to say incongruent words.
μ_c - Population Mean of the time taken by the subjects to say congruent words.

Dependent T-Test (also called the paired t-test or paired-samples t-test) would be the right choice for the statistical test as:

1. **Dependent t-test compares the means of two related groups to determine whether there is a statistically significant difference between these means.**
2. **The dependent t-test can look for "differences" between means when participants are measured on the same dependent variable under two different conditions as our test use the same participants under two different conditions.**
3. **We are not given any parameters related to population which rules out the z - test.**

3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

```
#3.py
import math
import pandas

csv = pandas.read_csv('./stroopdata.csv')

cong = csv['Congruent']
incong = csv['Incongruent']

congMean = cong.mean()
incongMean = incong.mean()

congStd = cong.std()
incongStd = incong.std()

print("Congruent Mean", congMean)
print("Incongruent Mean", incongStd)
print("Congruent Deviation", congStd)
print("Incongruent Deviation", incongStd)

// Output
Congruent Mean 14.051124999999999
Incongruent Mean 4.797057122469138
Congruent Deviation 3.5593579576451955
Incongruent Deviation 4.797057122469138
```

Congruent Mean 14.051124999999999
Incongruent Mean 4.797057122469138
Congruent Deviation 3.5593579576451955
Incongruent Deviation 4.797057122469138

4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.
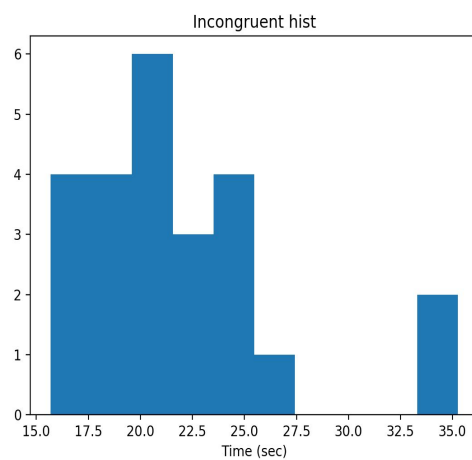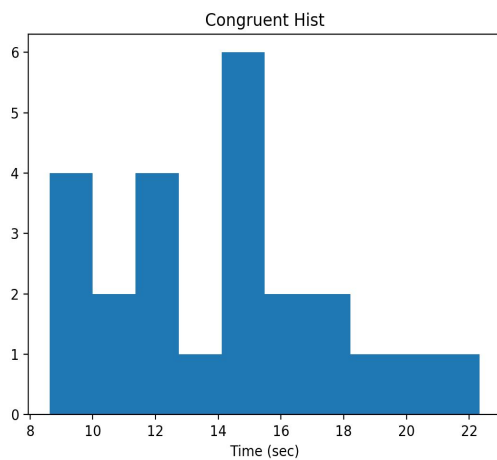
```
#4.py
import matplotlib.pyplot as plt
import pandas

csv = pandas.read_csv('./stroopdata.csv')

cong = csv['Congruent']
incong = csv['Incongruent']

plt.hist(cong)
plt.title('Congruent Hist')
plt.xlabel('Time (sec)')
plt.show()


plt.hist(incong)
plt.title('Incongruent hist')
plt.xlabel('Time (sec)')
plt.show()
```


Looking at the plot one can clearly say, subjects time varies a lot in the incongruent histogram (15.5 - 35) than congruent (8 - 22). Also the mean has a significance difference, incongruent is around 22 - 23 and for congruent is somewhere between 14 - 15.

5. Now, perform the statistical test and report your results. What is your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?

```python
#5.py
import matplotlib.pyplot as plt
import pandas
from scipy.stats import t
import math

csv = pandas.read_csv('./stroopdata.csv')


cong = csv['Congruent']
incong = csv['Incongruent']


variance = (cong - incong).std()
sample_size = len(cong)
degrees_of_freedom = sample_size - 1

point_estimate = incong.mean() - cong.mean() # ui - uc

t_statistic = point_estimate / ( variance / math.sqrt(sample_size) )
t_critical = t.ppf(.95, degrees_of_freedom) # 95% confidence, 23 degrees of freedom

print("tCritical : ", t_critical, ". tStatistic: ", t_statistic, ". Point Estimate: ",
point_estimate)

print("sample_size : ", sample_size, "variance", variance)


// Output
tCritical :  1.7138715277470473 . tStatistic:  8.020706944109957 . Point Estimate:
7.964791666666667
sample_size :  24 variance 4.864826910359056
```

With our confidence level of 95% and 23 degrees of freedom :
tCritical : 1.7138715277470473

We reject the null hypothesis as our
tStatistical (8.020706944109957) > tCritical (1.7138715277470473).

Yes, results do match our expectation of congruent task takes less time than incongruent.

REFERENCE LINKS :

https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php

http://www.statisticshowto.com/standardized-test-statistic/

https://statistics.laerd.com/statistical-guides/dependent-t-test-statistical-guide.php