Analyzing the Stroop Effect

Perform the analysis in the space below. Remember to follow the instructions and review the project rubric before submitting. Once you've completed the analysis and write up, download this file as a PDF or HTML file and submit in the next section.

(1) What is the independent variable? What is the dependent variable?

Variables

In this instance of the Stroop experiment, the **independent variable** is whether all the words in a reading list are *congruent* or *incongruent*. The words are names of common colors such as “red” or “blue”. Congruent words are printed in a color that matches the word meaning. For example, red and blue are congruent. Incongruent words are printed in a color different from than what the word means: green and blue are incongruent. This variable is categorical in nature and can adopt one of two possible values, congruent or incongruent. The **dependent variable** is the amount of time it takes to recite the colors that the words on the list are printed in. This is a numerical variable; samples are recorded as ratio data.

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

Hypotheses

The null hypothesis is that the population (true) mean time to recite the congruent list, equals the true mean time to recite the incongruent list, i.e., $H_0 : \mu_X = \mu_Y$, where x **and** y **are the populations of congruent and incongruent recital times, respectively**. The alternative hypothesis is that the mean recitation time for the incongruent list is larger ($H_A : \mu_y > \mu_x$).

Statistical Test

We propose evaluating the hypotheses using a **one-tailed, paired-sample *t*-test**. on the differences $d_i = y_i - x_i$. A paired *t*-test is appropriate, because each subject recited the congruent and incongruent lists. This experiment conforms to *within-subject* design. The resulting sample pairs are dependent. Within-subject experimentation is beneficial in controlling for individual subject peculiarities such as reading skill, visual acuity and color perception. Within-subject experimentation is more efficient, less time-consuming and less expensive overall.

Assumptions

In using *t*-testing, one assumption is that the subjects approximate a **random sampling** of the population of interest. For example, to evaluate these hypotheses across all age groups, we would not limit our sampling to schoolchildren. No subject selection details were provided with the sample data; therefore, it is assumed that the subjects do represent the population of interest. For a paired *t*-test, it is assumed that **carryover effects and treatment order effects were negligible**. For example, reading the congruent list first did not sharpen the subjects' focus--or fatigue them--as to

alter their subsequent performance reciting the incongruent list.

(3) Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability. The name of the data file is 'stroopdata.csv'.

In [33]:

```
import pandas as pd

# Read the data into a DataFrame
stroop_data = pd.read_csv('stroopdata.csv')

# Add a deltas column.
x = 'Congruent'
y = 'Incongruent'
stroop_data.loc[:,'Dyx'] = stroop_data.loc[:,y] - stroop_data.loc[:,x]

# Calculate some descriptive statistics
stroop_desc = stroop_data.describe()

# Append some other useful statistics
stroop_desc.loc['median'] = stroop_data.median()
stroop_desc.loc['IQR'] = stroop_desc.loc['75%'] - stroop_desc.loc['25%']

# Let iPython pretty-print the results
stroop_desc
```

Out[33]:

	Congruent	Incongruent	Dyx
count	24.000000	24.000000	24.000000
mean	14.051125	22.015917	7.964792
std	3.559358	4.797057	4.864827
min	8.630000	15.687000	1.950000
25%	11.895250	18.716750	3.645500
50%	14.356500	21.017500	7.666500
75%	16.200750	24.051500	10.258500
max	22.328000	35.255000	21.919000
median	14.356500	21.017500	7.666500
IQR	4.305500	5.334750	6.613000

Summary Statistics

There are $n = 24$ pairs of samples in this data set. The congruent data have a sample mean $\bar{x} = 14.05$ sec with standard deviation $S_X = 3.56$ sec. The incongruent data have $\bar{y} = 22.02$ and $S_Y = 4.80$ sec.

(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.

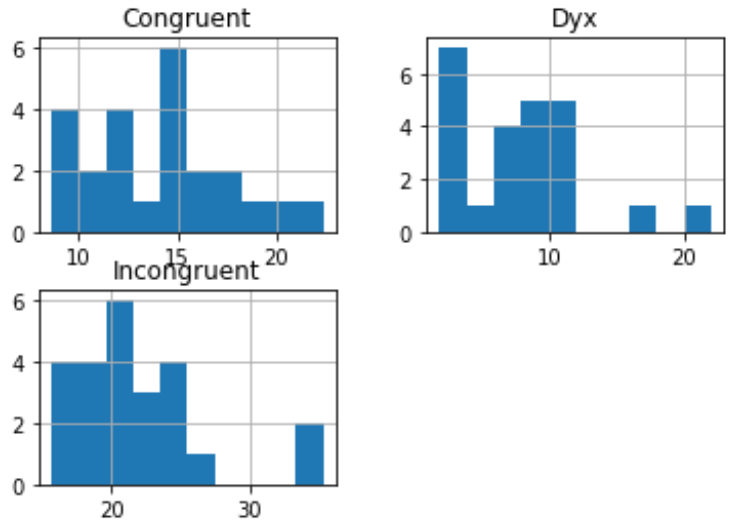
In [37]:

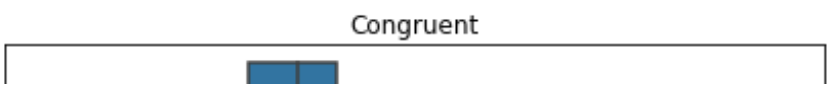
```
%matplotlib inline

import matplotlib.pyplot as plt
import seaborn as sns

# Frequency histogram plot
fig0 = plt.figure()
stroop_data.hist()

# Box-and-whisker plot
fig1 = plt.figure()

ax0 = fig1.add_subplot(211)
sns.boxplot(x=x, data=stroop_data)
ax0.set_title(x)
ax0.set_xlabel('')
ax0.set_xlim(0, 40)

ax1 = fig1.add_subplot(212)
sns.boxplot(x=y, data=stroop_data)
ax1.set_title(y)
ax1.set_xlabel('Time (sec)')
ax1.set_xlim(0, 40)

plt.tight_layout()

# Scatter plot
fig2 = plt.figure()
ax0 = fig2.add_subplot(111)
plt.scatter(stroop_data.loc[:,x], stroop_data.loc[:,y])
ax0.set_xlabel(x)
ax0.set_ylabel(y)
```

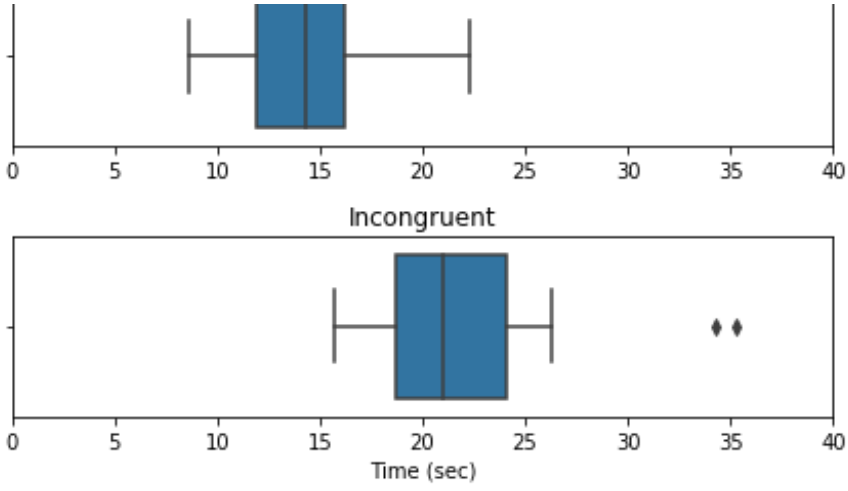
Out[37]:

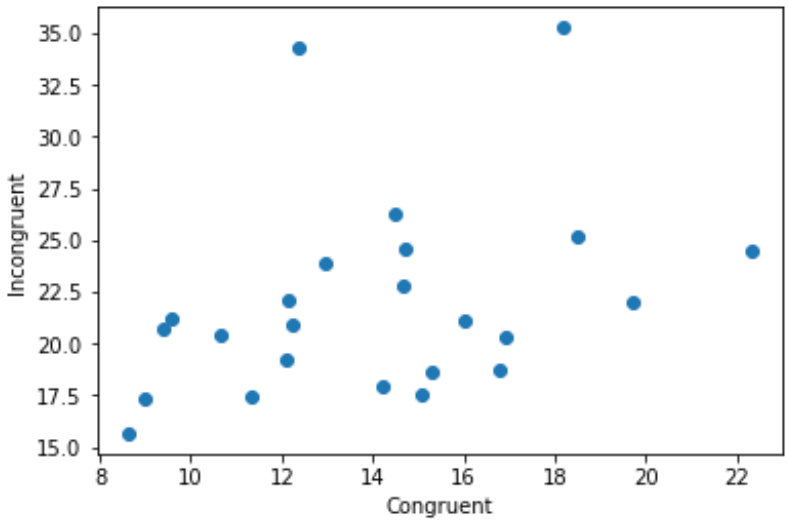
```
Text(0,0.5,'Incongruent')
```

```
<matplotlib.figure.Figure at 0x7fa2c066a9e8>
```









The histograms suggest that the underlying distributions may be positively skewed. This makes sense, since time cannot be less than zero, but it can be arbitrarily large. The box-and-whisker plots indicate that two of the incongruent times may be outliers. In the absence of other information, the accepted definition of an outlier is a sample value that lies more than 1.5 IQRs beyond Q_1 or Q_3. The box-and-whisker plots also show that the Q_1-to-Q_3 ranges don't overlap at all, suggesting these populations may be significantly different. Lastly, the scatter plot suggests positive correlation between the congruent and incongruent times, although that is not the focus of this analysis.

(5) Now, perform the statistical test and report the results. What is the 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?

In [38]:

```
from scipy import stats

def print_tp(t, p, df):
    '''Print right-tailed t-statistic and p value in APA style'''
    print("t({0:.0f}) = {1:.2f},  p = {2:1.2e}, right".format(df, t, p))
    return None

# Deg. of freedom
df = stroop_desc.loc['count','Congruent'] - 1

# Data lists
x = stroop_data.loc[:,'Congruent']
```

```
y = stroop_data.loc[:,'Incongruent']
sample_diffs = stroop_data.loc[:,'Dyx']

# Calculate t-critical
alpha = .0005
t_crit = stats.t.ppf(1-alpha, df)
print("At alpha = {0:.4f}, t* = {1:.3f}".format(alpha, t_crit))

# Calculate the paired (relative) t-statistic and p-value
t, p = stats.ttest_rel(y, x)
print_tp(t, p, df)

# Calculate the confidence interval
mean_diff = sample_diffs.mean()
std_err = stats.sem(sample_diffs)
margin_of_err = t_crit * std_err
lwr_bound = mean_diff - margin_of_err
upr_bound = mean_diff + margin_of_err
print("99.95% CI [{0:.2f}, {1:.2f}]".format(lwr_bound, upr_bound))



# Recalculate t and p, excluding the paired samples that include an outlier
y-value
mask = (
    stroop_data.loc[:,'Incongruent'] <
    stroop_desc.loc['75%','Incongruent'] +
    1.5 * stroop_desc.loc['IQR','Incongruent']
)
t_new, p_new = stats.ttest_rel(
    stroop_data.loc[:,'Congruent'].loc[mask],
    stroop_data.loc[:,'Incongruent'].loc[mask],
)
print(' ')
print("Excluding pairs with `y` outliers:")
print_tp(t_new, p_new, df)
```

```
At alpha = 0.0005, t* = 3.768
t(23) = 8.02,  p = 4.10e-08, right
99.95% CI [4.22, 11.71]

Excluding pairs with `y` outliers:
t(23) = -9.54,  p = 4.38e-09, right
```

Test Results

At the $p = 0.9995$ confidence level, the critical value is $t^* = 3.768$.

The t-statistic is $t(23) = 8.02; p = 4.10x10^{-8}$. This t-statistic lies in the critical region. Therefore, we reject the null hypothesis. With 99.95% confidence, the true mean time to recite the incongruent list is *greater* than the time to recite the congruent list.

Digging Deeper

Possible Reasons for the Stroop Effect

The Wikipedia page on Stroop effect cites four theories for the performance difference between

The Wikipedia page on Stroop effect cites four theories for the performance difference between congruent and incongruent word lists.

1. Processing speed - word processing is faster than color processing in the brain. Since the incongruent list forces the subjects to perform color processing, it takes longer.
2. Selective attention - color recognition requires more attention than reading and therefore takes longer.
3. Automaticity - color recognition is a less routine task than reading and therefore takes longer.
4. Parallel distributed processing - when two pathways in the brain are energized simultaneously, the stronger pathway interferes with the weaker one. The color recognition pathway is weaker than the reading pathway.

Possible Extensions to this Experiment

1. Design a third *minimally-interfering* list of words whose meanings avoid color associations (e.g., house, music), to see if the incongruent recitation times can be reduced.
2. Test whether an *attention hog*, such as balancing a pencil in the palm, can increase the recitation time for the congruent list.
3. Design a sample of poor readers, such as first graders, to test whether *reducing reading automaticity* increases the recitation time for the congruent list.

References

1. https://en.wikipedia.org/wiki/Stroop_effect
2. https://faculty.washington.edu/chudler/java/ready.html
3. https://review.udacity.com/#!/rubrics/71/view
4. https://seaborn.pydata.org/generated/seaborn.boxplot.html
5. https://www.datacamp.com/community/tutorials/seaborn-python-tutorial#show
6. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html
7. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html
8. https://mkaz.tech/code/python-string-format-cookbook/
9. https://matplotlib.org/users/tight_layout_guide.html