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

The independent variable is the type of test, i.e. Congruent word or Incongruent word.
The dependant variable is the time that the participants took in the tests to recognize the color/word.

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

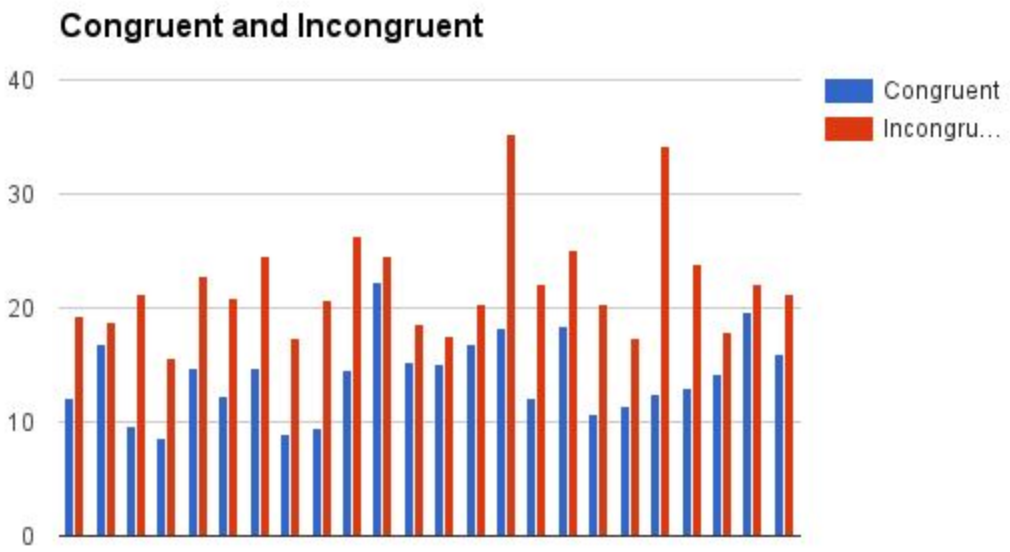
Since the samples are paired, i.e for each individual sample. there is a corresponding unique pair in the other sample we can use the following set of hypotheses:
H0: The null hypotheses that states that the intervention (congruent vs incongruent words) does not have any effect on the dependent variable, the time to recognize the word.
H1: The alternate hypotheses that states that there is a significant difference in the sample after the intervention.
The suitable type of test for this kind of data would be a dependent t-test. This test would help me find if there is a significant difference between the sample from pre and post intervention. If there is, then we can accept the alternate hypotheses and reject the null hypotheses.

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

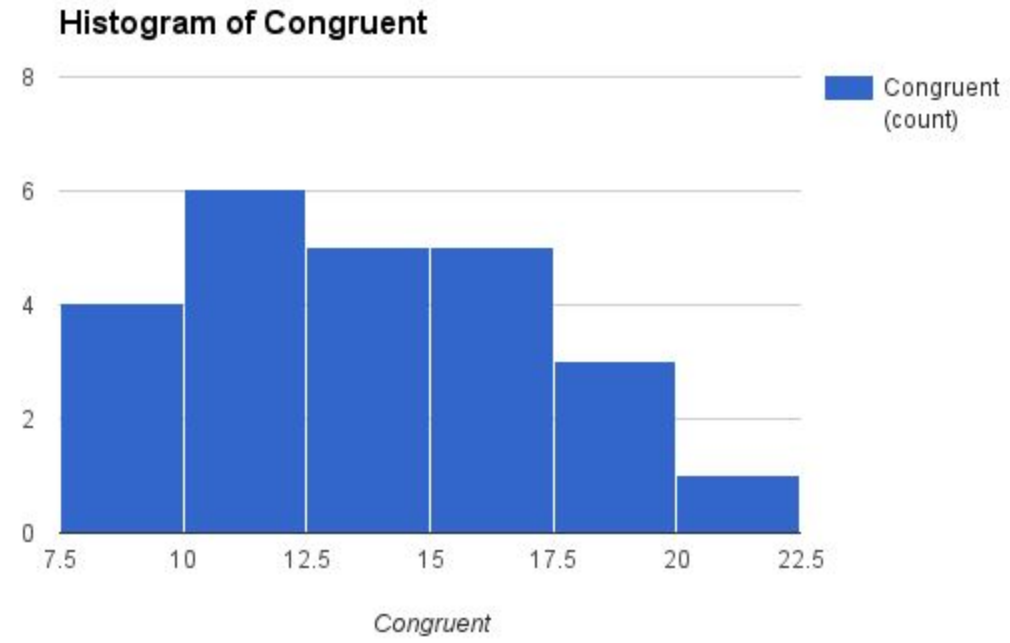
Statistic	**Congruent Word**	**Incongruent Word**
Mean	14.3565	21.0175
Standard Deviation	3.559357958	4.797057122
Variance	12.66902907	23.01175704
Median	14.3565	21.0175

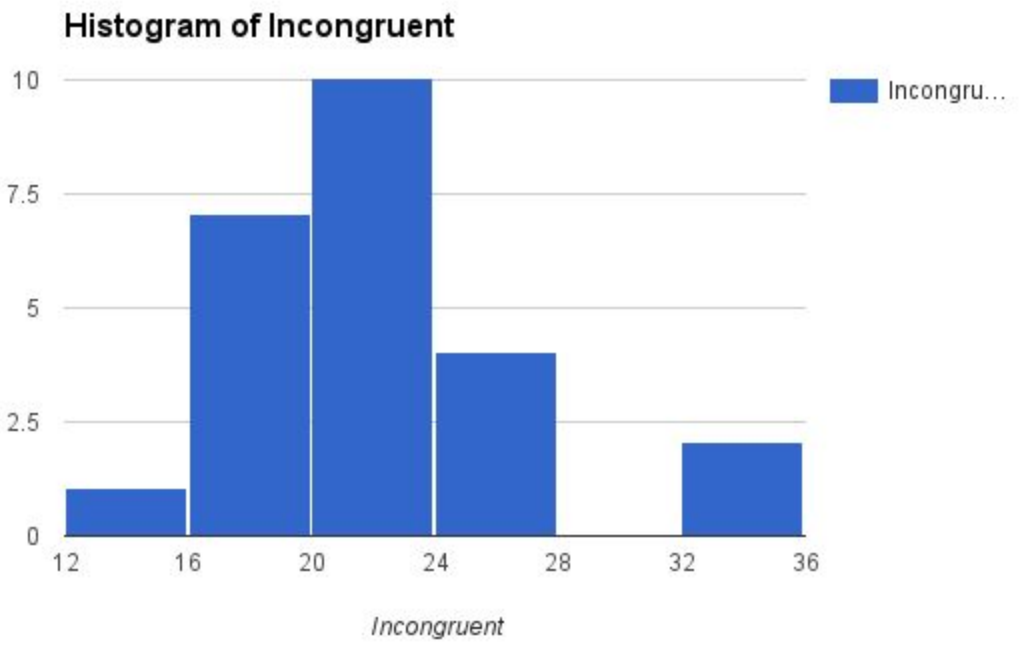
The mode cannot be determined for the dataset since none of the values repeat.

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.**



The above visualization shows the variation in time between congruent and incongruent times. It clearly indicates that in all the cases, the incongruent time is higher than congruent time. This indicates that when names of the color are written in colors other than what the word says, it will most likely take longer for people to recognize it correctly.





The above two histograms indicate the time data. We can see the most between 10-12.5 was the most common congruent time and 20-24 was the most common incongruent time.
We can also infer from the histogram that 20-22.5 was the least common congruent time and 12-16 was the least common incongruent time. The interesting thing about this is that it can be inferred that most people took at least 16 sec for incongruent test and very few people took more than 20 secs for congruent test.

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?**

Assuming P-critical value of 0.05 and a 2-tailed test for a degree of freedom df = 24 if we get a t value < -2.0639 or > 2.0639, we can reject the null hypotheses

t is calculated as t = mean of differences/ (standard deviation of differences/sqrt(n))

t = -7.364875/ (5.767400818/sqrt(25))
t = -6.384916908 which is less than -2.0639

This means that we can reject the null hypotheses and confidently say that incongruent word-color pair has a negative impact on the participant's reading time.

Reference:

http://www.statisticslectures.com/topics/dependentsamplest/#video