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

In this case, the dependent variable in the measured reaction time. The independent variable , is the type of condition (interference) that changes between the two samples : congruent and incongruent condition.

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

The appropriate set of hypotheses for this task is :

H0 (Null hypotheses) : $M_c = M_i$
HA (Alternative hypotheses) : $M_c \neq M_i$

where M_c is the mean for congruent condition samples and M_i is the mean for incongruent conditions.

I understand that these are the appropriate set of hypotheses given the fact that the objective of the test is to understand if there is any difference in reaction between the congruent and the incongruent condition. There are no mentions of expectations about increase or reduction in reaction time for the different conditions so i choose to define the alternative hypotheses neither greater than or less than.

I expect to perform a dependent t-test within-subject since the same subject take the test twice. In this case , it will be a repeated measure design because a measure is taken (reaction time) for the same subject but given different conditions (one measure is with congruent words and colors and the second one is incongruent words and colors). I found this scenario similar to the one we reviewed during the statistics course around use of cell phones keyboards and the impact of changing between alphabetic and qwerty based ones.

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

In the next list we'll report some descriptive statistics on the datasets:

Mc (Congruent test mean) : 14.05
Mi (Incongruent test mean) : 22.01
Point estimate / Average of differences : -7.96
Standard deviation of the differences : 4.86
T-statistics : -8.02
Cohens D : -1.64
Confidence interval 95% -lower- : -10.01
Confidence interval 95% -upper- : -5.91

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.

I like the following visualization because it shown that at any comparison between pair of values (same subject, congruent vs incongruent) there seems to be a proportion in the delay of the reaction no matter the level of reaction of the individual
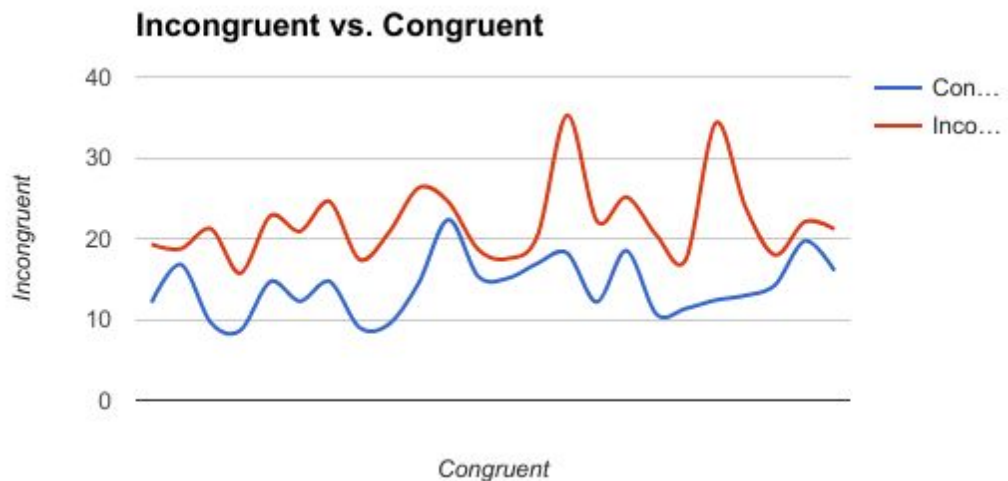
5. Now, perform the statistical test and report your results. What is your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?

Based on information provided in #3 :

T-statistics : -8.02
Alpha level : 0.05
TYpe of test : two tails
T-critical +/- : 2.064
P-value < 0.05

Conclusion :

Conclusion is to reject the null hypothesis since the results are considered statistically significant .

The results match my expectations since , at first sight, you perceive that it won't be the same, in effort, to go through the congruent condition than the incongruent one. Congruent version is very intuitive since words and colors match; move to the incongruent version requires to separate the read comprehension from the visual intuition of the color.

6. Optional: What do you think is responsible for the effects observed? Can you think of an alternative or similar task that would result in a similar effect? Some research about the problem will be helpful for thinking about these two questions!

I guess is a matter of immediate sensorial perception vs. mediate reading comprehension. I guess that any task where immediate comprehension by physical sense (like sounds , or touch or visual) is compared with a mediated version of the same comprehension will show the same reaction time differences. Additional example could be color based controls to manage a machine vs same controls functionality based on named commands; reaction time will increase , i guess probably that the reason why alerts systems uses visual gadgets to accomplish its tasks.

References

Stroop effect definition : https://en.wikipedia.org/wiki/Stroop_effect
Lesson 26, L10a, t-Tests (Keyboards)
P-value : http://www.graphpad.com/quickcalcs/pValue2/