Statistics: the science of decisions project

Stefan Dulman

February 6, 2016

Question 1: What is our independent variable? What is our dependent variable?

The independent variable is the number of words matching their ink color. In the first data set, all the words match their color, in the second experiment none of them does. The independent variable is set before the experiment and not modified by the actual experiment run. The dependent variable is the recorded reaction time. This variable is influenced by the mismatch between the words and their displayed colors. Based on the recorded differences, various hypothesis can be tested with respect to the influence the independent variable has.

Question 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 null hypothesis is that the two recorded time response series are samples from the same population and the mismatch between them is just the effect of chance. An alternative hypothesis could be that the two recorded time series come from different populations (because our brain uses two different mechanisms for processing images and text, with different response times). If subscripts $_c$, $_i$ stand for "congruent", respectively "incongruent", the sample mean is denoted by m and the population mean by μ, then:

$$H_0 : \mu_c = \mu_i$$
$$H_1 : \mu_c \neq \mu_i$$

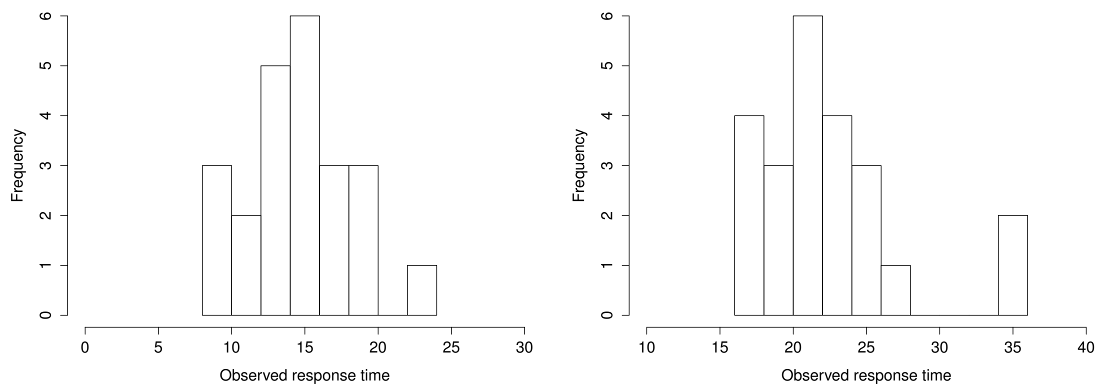
I will perform a dependent two-tailed t-test with a confidence level of say 5%. This choice is justified by the following reasons:
(i) the mean and standard deviations of the populations are unknown;
(ii) the set of participants is the same in the two conditions (24 participants);
(iii) the experiment is of type "repeated measures design", where the set of participants is first asked to perform a task with one experimental setup and then asked to perform the same task in a different experimental setup;
(iv) two-tailed version of the test was chosen because, without any other information, the population means can be in any kind of relationship towards each other.

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

The data consists of 24 readings for both the congruent and incongruent conditions. The mean values in the two sets are: $m_c = 14.05$ and $m_i = 22.02$. Additionally, the median values are $med_c = 14.36$ and $med_i = 21.02$.

The standard deviations of the samples are: $s_c = 3.56$, $s_i = 4.80$.

Question 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.



(a) Histogram of congruent condition (b) Histogram of incongruent condition

Figure 1: Histograms of experiment data

I noticed that there is a big difference in the distribution of the two data sets. All the incongruent condition data is above the median value of the congruent data. I also noticed that the distribution of the incongruent data is a bit more compact, if one makes abstraction of the outlier points (the whiskers in the graph extend to a data point maximum 1.5 interquartile ranges from the box).

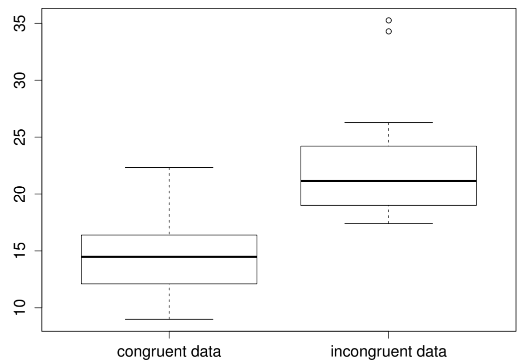


Figure 2: Boxplot of experiment data

Question 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?

First, I performed the difference between the two data sets. This leads to a new data set with a mean value $m_d = -7.96$ and a standard deviation of $s_d = 4.86$. Next, I computed the t-value as:

$$t = \frac{\mu_c - \mu_i}{\frac{s_d}{\sqrt{n}}} = -8.03$$

Table 1: α values for t-test with 23 degrees of freedom

Confidence level	two-tails critical stat. val.
5%	± 2.069
1%	± 2.807
0.1%	± 3.768

Back to the original questions: "what is the confidence level/critical statistic value?". I listed these values in the table above. For example, for a confidence level of 5%, the critical statistic values are ± 2.069.

Next, I checked the values in the t-table for various confidence levels. Notice that the obtained t-value is way out in the critical zone, meaning that I reject the null-hypothesis. In other words, there is a significant difference between the two data sets and this difference is almost impossible by pure chance. The p-value corresponding to the computed t-value is less than 0.00001.

Looking at the data from the experiment (for example the box plot graph) one notices the obvious difference in the quartile values of the two distributions. The difference is confirmed by the t-test. The t-value is so large that the null-hypothesis would be rejected even when considering the single-tail test, with the alternate hypothesis $H_2 : \mu_c < \mu_i$.

Secondly, while taking the test, it became obvious that I required significantly more concentration for the incongruent test, hence the time response was larger. This observation is also in line with the analysis above.

Question 6: 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?

I think the Stroop effect can be explained also by the fact that text recognition is triggered in our brain and takes priority over the color recognition. If the texts in the incongruent test would be replaced with rectangles then much faster response would be observed. As a human, when seeing the colored word, the first response in my head is the written text, which I need to consciously ignore and force my attention on the color. I think this is exactly what the theory of subsumption architecture organization of our brain is about (see the book "Understanding intelligence" of Rolf Pfeifer).

In line with the above theory, any mechanism that needs conscious overriding would take more time. For example, stopping a falling object comes natural and very fast (think of a rolling ball of a table). If the ball is hot, the mechanism would be overridden by our perception of heat, so we would decide comparably fast not to stop the ball, because it would burn our hand. Let it fall. But if we need conscious override for a higher goal - say the hot ball would fall on the head of an unaware toddler, we would go forward and stop it, even if it burns our hand. The fine point here, is that the deciding to override the learnt order of mechanisms will occur slower. We would stop the hot ball, but after a longer period of "thinking".

Another somewhat comic example is asking people to ride a bike with reversed command for the steering wheel. Observing the reply of our own reinforcement learning mechanism in our head to counteract falling, then consciously reversing it, takes so much time that any untrained person falls off the bike. See an youtube video here:
`https://www.youtube.com/watch?v=uJ424nF4N4U`