

Worksheet 12

Author:

Discussants:

=====

This is an optional worksheet for those students who are interested in practicing the material covered in the last week of class. You do not need to turn it in since you also have to complete your class final project, although I have added a hand in link on moodle for those students who would like feedback (and who would like it mentioned on your eval).

```
library(knitr)
# makes sure the code is wrapped to fit when it creates a pdf
opts_chunk$set(tidy.opts = list(width.cutoff = 60))
set.seed(1) # set the random number generator to always give the same sequence of random numbers
```

Please answer the following Lock5 questions examining a single mean:

Lock5 exercise 6.76: Find the endpoints (quantile function values) of a t-distribution with 5% beyond them in each tail if the sample has a size of $n = 10$ (and the underlying distribution of the data is relatively normal so $n = 10$ is a large enough sample size). Hint: all quantile functions start with the letter q , and then have the distribution name, so for quantiles of the t distribution we use the `qt()` function and we specify the area we want and the degrees of freedom (df), i.e., `qt(area, df = degrees_of_freedom)`.

Answer:

Lock5 exercise 6.106: Find a 95% confidence interval for μ using the sample results $\bar{x} = 12.7$, $s = 5.6$, and $n = 30$. Assume the underlying distribution is relatively normal so using the t-distribution is valid here. Hint: as usual, compute the t^* quantile value, then compute the standard error and then the confidence interval.

Answer:

Lock5 exercise 6.136: Run a hypothesis test to test: $H_0 : \mu = 15$ vs. $H_A : \mu > 15$ using the sample results $\bar{x} = 91.7$, $s = 12.5$, $n = 30$. Hint: compute the t-statistic and the use `pt()` to find the appropriate p-value.

Answer:

Lock5 exercise 6.142: The Autistic Brain Autistic children often have a small head circumference at birth, followed by a sudden and excessive increase in head circumference during the first year of life. A recent study examined the brain tissue in autopsies of seven autistic male children between the ages of 2 and 16.

The mean number of neurons in the prefrontal cortex in non-autistic male children of the same age is about 1.15 billion. The prefrontal cortex is the part of the brain most disrupted in autism, as it deals with language and social communication. In the sample of seven autistic children, the mean number of neurons in the prefrontal cortex was 1.94 billion with a standard deviation of 0.50 billion. The values in the sample are not heavily skewed. Use the t-distribution to test whether this sample provides evidence that autistic male children have more neurons (on average) in the prefrontal cortex than non-autistic children. (This study indicates that the causes of autism may be present before birth.) Hint: write down the null and alternative hypotheses, compute the statistic of interest and then calculate a p-value using the `pt()` function.

Answers:

Please answer the following Lock5 questions examining hypothesis tests comparing differences in two means

Lock5 exercise 6.250: Test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$ using the sample results $\bar{x}_1 = 56$, $s_1 = 8.2$ with $n_1 = 30$ and $\bar{x}_2 = 51$, $s_2 = 6.9$ with $n_2 = 40$. Compute the statistic of interest and then calculate a p-value using the `pt()` function.

Answers:

Lock5 exercise 6.258: Stomach Bacteria and Irritable Bowel Syndrome Studies are finding that bacteria in the stomach are essential for healthy functioning of the human body. One study compared the number of unique bacterial genes in stomachs of healthy patients and those of patients with irritable bowel syndrome (IBS). For healthy patients, we have $\bar{x} = 564$ million with $s = 122$ million and $n = 99$. For those with IBS, we have $\bar{x} = 425$ million with $s = 127$ million and $n = 25$. Both distributions appear to be approximately normally distributed. Test to see if people with IBS have, on average, significantly fewer unique bacterial genes in their stomachs. Show all details, including giving the degrees of freedom used.

Answers:

Questions Comparing different hypothesis tests

In the following exercises we will use several different types of tests to examine whether students' pulse rates are higher when they are taking a quiz compared to when they are listening to a lecture. We will assume that the mean pulse rate is higher when taking a quiz compared to when listening to a lecture (since quizzes are more stressful). A data set was collected from 10 students when they took a quiz and when the same 10 students were listening to a lecture. The code to download the data set is below.

```
pulse_data <- read.csv("http://www.lock5stat.com/datasets/QuizPulse10.csv")
```

Exercise 1: For the first exercise let's treat the two samples as independent. Start by stating the null and alternative hypotheses in symbols. Then apply the following steps to conduct a two-sample t-test:

- compute the mean and standard deviation (or variance) for the quiz pulse rates
- compute the mean and standard deviation (or variance) for the lecture pulse rates

- c) compute the standard error
- d) compute the t-statistic
- e) compute the p-value
- f) draw a conclusion with $\alpha = .05$

State the null and alternative hypotheses here:

```
# a) compute the mean and standard deviation (or variance)
# for the quiz pulse rates
```

```
# b) compute the mean and standard deviation (or variance)
# for the lecture pulse rates
```

```
# c) compute the standard error
```

```
# d) compute the standard error
```

```
# e) compute the p-value
```

Answer:

Question 2: For the second exercise let's use a paired sample t-test to see if this would change our conclusion. To do this create a new vector that has for each student, the pulse rates during a quiz minus the pulse rate during the lecture. Let's call the mean of this vector of differences \bar{x}_d . State the null and alternative hypothesis and then complete the following steps to do a hypothesis test:

- a) plot the difference of pulse rates using the stripchart(diff_means_vector) function
- b) compute the standard error of the difference of scores
- c) compute the t-statistic
- d) compute the p-value
- e) draw a conclusion with $\alpha = .05$

State the null and alternative hypotheses here:

```
# a) plot the data using the stripchart() function:
```

```
# b) compute the standard error of the difference of scores
```

```
# c) compute the t-statistic
```

```
# d) compute the p-value
```

Answer:

Question 3: One more method we have used to test whether two means are different is resampling methods. Let's again treat the two samples as if they were independent and run a resampling test to assess whether there are differences in the quiz vs. lecture pulse rates by doing the following steps:

- a) compute the observed statistic
- b) create a null distribution (i.e., combined the data, repeat 10,000 times: shuffle the combined data, create null statistics)
- c) compute the p-value
- d) draw a conclusion with $\alpha = .05$

Answer:

Question 4: R actually has built in functions to do t-tests which is called `t.test()`. If we have two samples called `sample1` and `sample2`, we can do a two independent sample t-test using `t.test(sample1, sample2)`. The resulting p-value is for a two-tailed test. If we want the p-value for a right-tailed test we can set the argument to the function `alternative = "greater"` i.e., `t.test(sample1, sample2, alternative = "greater")`. We can run a paired t-test by setting the the argument to the function `paired = TRUE`.

Use R's `t.test()` function to calculate p-values on the pulse data using an independent sample t-test and a paired sample t-test. Compare the results to the results you got using your own t-test. Do you notice any differences? What might be causing these differences (hint: R is doing something slightly more sophisticated in one case).

Answer:

Reflection

How did this worksheet go? Are the concepts around parametric hypothesis tests for mean making sense?