Examples: *t*-based Inference for One Sample or Paired Samples

Statistical Sleuth Sections 2.1, 2.2, 2.5

Introduction and Goals

The goals for this handout are:

1. to practice calculations and interpretations of *t*-based inferences (confidence intervals and hypothesis tests) for either:
    * a single mean; or
    * a difference in means with *paired* data
2. to see how those calculations can be done using the `t.test` function in R

We will work with two example studies that look different on the surface, but where the same statistical tools can be used.

Example Studies

Study 1 (Body Temperatures)

It's generally believed that the average body temperature is 98.6 degrees Farenheit (37 degrees Celsius - but we will work in units of degrees Farenheit). Let's investigate whether this is true using measurements of the temperatures of 130 adults. These data are derived from a 1992 study published in the Journal of the American Medical Association [1]. The study describes the study participants as volunteers who were "healthy men and women aged 18 through 40 years". The researchers recorded each subject's body temperature (by an oral thermometer) and heart rate.

Study 2 (Twins Study on Schizophrenia)

Description taken from the Statistical Sleuth [2]:

"In a 1990 article, researchers reported the results of a study tht controlled for genetic and socioeconomic differences by examining 15 pairs of monozygotic ["identical"] twins, where one of the twins was schizophrenic and the other was not [3]. The twins were located through an intensive search throughout Canada and the United States. . . . The researchers used magnetic resonance imaging to measure the volumes (in cm^3) of several regions and subregions inside the twins' brains."

We will examine whether the data provide evidence of a difference in volumes of the left hippocampus between the unaffected and the affected individuals.

First Thing To Do: Get Familiar with the Data

First we need to load a few R packages to extend R's functionality

```
library(readr)
library(dplyr)
library(ggplot2)
library(mosaic)
```

Study 1 (Body Temperatures)

```
body_temps <- read_csv("http://www.evanlray.com/data/jse/normtemp/bodytemp.csv")
```

```
## Parsed with column specification:
## cols(
##   temp = col_double(),
##   sex = col_character(),
##   hr = col_integer()
## )
```

```
head(body_temps)
```

```
## # A tibble: 6 x 3
##    temp sex      hr
##   <dbl> <chr> <int>
## 1  96.3 Male     70
## 2  96.7 Male     71
## 3  96.9 Male     74
## 4  97   Male     80
## 5  97.1 Male     73
## 6  97.1 Male     75
```
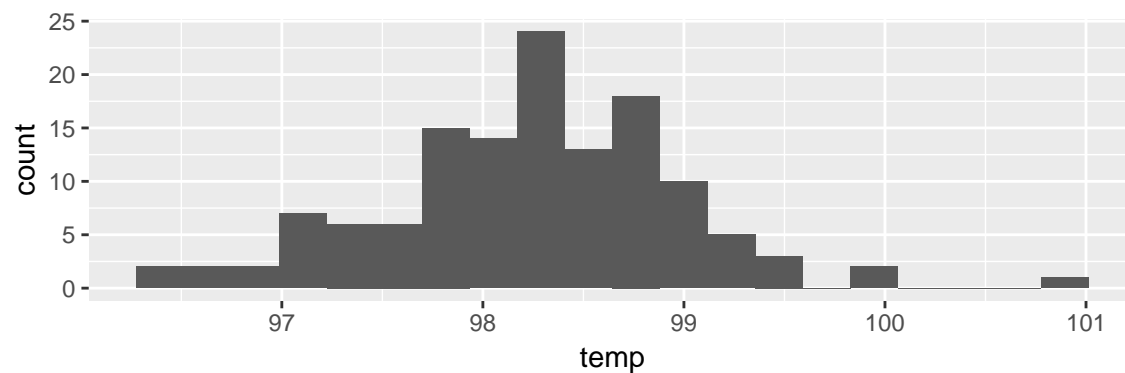
```
dim(body_temps)
```

```
## [1] 130   3
```

```
body_temps %>%
  summarize(
    mean_temp = mean(temp),
    median_temp = median(temp),
    sd_temp = sd(temp)
  )
```

```
## # A tibble: 1 x 3
##   mean_temp median_temp sd_temp
##       <dbl>       <dbl>   <dbl>
## 1      98.2        98.3   0.733
```

```
ggplot(data = body_temps, mapping = aes(x = temp)) +
  geom_histogram(bins = 20)
```


Things to think about:

Is the distribution of values in the sample approximately symmetric and unimodal?

Are there any outliers? If so, how serious are they?

Study 2 (Twins Study on Schizophrenia)

I'm going to ask you to repeat these steps in R in groups, so you have some practice working in R.

So that we can refer to them later and you can check your work in R, here are the results:

Here's the first 6 rows of the data set:

```
## # A tibble: 6 x 3
##    Unaffected Affected difference
##         <dbl>    <dbl>      <dbl>
## 1        1.94     1.27       0.67
## 2        1.44     1.63      -0.190
## 3        1.56     1.47       0.09
## 4        1.58     1.39       0.19
## 5        2.06     1.93       0.13
## 6        1.66     1.26       0.400
```
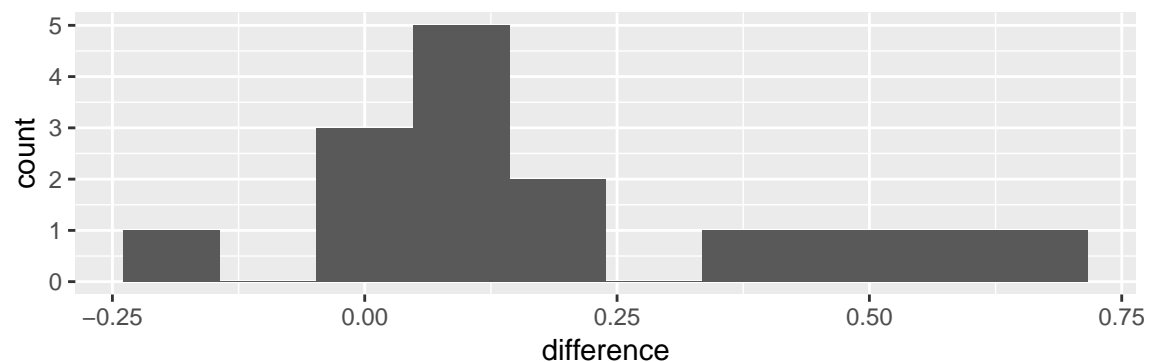
Here's the number of rows (observational units) and columns (variables) in the full data set:

```
## [1] 15  3
```

Summaries of differences in volume of the left hippocampus (unaffected - affected):

```
## # A tibble: 1 x 3
##    mean_difference median_difference sd_difference
##              <dbl>             <dbl>         <dbl>
## 1            0.199              0.11         0.238
```

A histogram of differences in volume of the left hippocampus (unaffected - affected):

Things to think about:

Is the distribution of values in the sample approximately symmetric and unimodal?

Are there any outliers? If so, how serious are they?

Hypothesis Test: How much evidence that the parameter is not equal to a given value?

Study 1 (Body Temperatures)

What is/are the population parameter(s) of interest?

What would the null and alternative hypotheses be for a relevant hypothesis test?

What is the sample statistic to use for the test, and its interpretation?

If the null hypothesis is true, what is the sampling distribution of the statistic above?

What is the p-value for the test?

```
t.test(~ temp, data = body_temps)
```

```
##
##  One Sample t-test
##
## data:  temp
## t = 1527.9, df = 129, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   98.12200 98.37646
## sample estimates:
## mean of x
##   98.24923
```

What is the interpretation of this p-value in context (interpret the definition of the p-value in this context)?

What is the strength of evidence against the null hypothesis provided by the data? Use this scale (See Section 2.5):

p-value	Strength of evidence against H_0	Compare to...
0.10 or less	Some evidence; not conclusive	Probability of 4 heads in a row is 0.0625
0.05 or less	Moderate amount of evidence	Probability of 5 heads in a row is 0.03125
0.01 or less	Strong evidence	Probability of 7 heads in a row is 0.007813
0.001 or less	Very strong evidence	Probability of 10 heads in a row is 0.0009766

What is the scope of the conclusions for the results above?

Study 2 (Twins Study on Schizophrenia)

What is/are the population parameter(s)?

What would the Null and Alternative hypotheses be for a relevant hypothesis test?

What is the sample statistic to use for the test, and its interpretation?

If the null hypothesis is true, what is the sampling distribution of the statistic above?

What is the p-value for the test? Use R to find it; your result should be 0.006062.

What is the interpretation of this p-value in context (interpret the definition of the p-value in this context)?

What is the strength of evidence against the null hypothesis provided by the data? Use the scale from the previous page.

What is the scope of the conclusions for the results above?

Confidence Interval: What is a range of plausible values for the parameter?

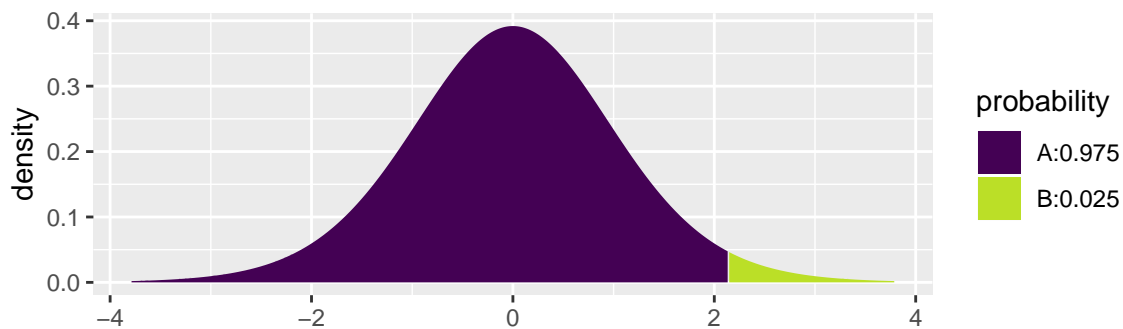**Study 1 (Body Temperatures)**

Find a 95% confidence interval for the population mean body temperature using `t.test`

```
t.test(~ temp, data = body_temps)
```

```
##
##   One Sample t-test
##
## data:  temp
## t = 1527.9, df = 129, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   98.12200 98.37646
## sample estimates:
## mean of x
##   98.24923
```

Find a 95% confidence interval for the population mean body temperature using the formula. The critical value t^* can be found with `qt` or `xqt`.

```
xqt(0.975, df = 14)
```



```
## [1] 2.144787
```

What is the interpretation of your confidence interval?

Study 2 (Twins Study on Schizophrenia)

Find a 95% confidence interval for the mean difference in volume of the left hippocampus for twins who are unaffected by schizophrenia and who are affected by schizophrenia using `t.test`. (You can use the same R output as for the hypothesis test above.)

Find a 95% confidence interval for the population mean body temperature using the formula. The critical value t^* can be found in R with `qt` or `xqt`.

References

1. Mackowiak, P. A., Wasserman, S. S., and Levine, M. M. (1992), "A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich," Journal of the American Medical Association, 268, 1578-1580.

2. Ramsey, F., and Schafer, D. (2012). The statistical sleuth: a course in methods of data analysis. Cengage Learning.

3. Suddath, Richard L., et al. (1990), "Anatomical abnormalities in the brains of monozygotic twins discordant for schizophrenia," New England Journal of Medicine 322.12, 789-794.