

# Basic Course on R: Hypothesis Testing and Confidence Intervals 1 Practical

Elizabeth Ribble\*

28 Oct - 1 Nov 2019

## Contents

1	Baby Data	2
---	-----------	---

---

\*emcclel3@msudenver.edu

# 1 Baby Data

1. Read in the data “R\_data\_January2015.csv” with a header and row names from the first column. Assign it to the object `babydata` and allow strings be converted to factors. Attach the data to the environment.
2. What are the dimensions of `babydata`? What is the class? Answer these questions separately with two functions and then together with one function.
3. Answer the following questions pertaining to the variable `SAH`:
  - (a) What are the 20% quantiles of `SAH`?
  - (b) What are the mean, median, variance and standard deviation of `SAH`?
  - (c) Create a stem and leaf plot of `SAH`.
  - (d) Create a histogram and a horizontal boxplot of `SAH` in one graphics window where the plot of the histogram is above the boxplot.
  - (e) Utilize all 3 graphs to describe the shape of the distribution of `SAH`.
  - (f) Log-transform `SAH` (assign it to `logSAH`).
  - (g) What are the 20% quantiles of `logSAH`?
  - (h) What are the mean, median, variance and standard deviation of `logSAH`?

- (i) Create a stem and leaf plot of `logSAH`.
  - (j) Create a histogram and a horizontal boxplot of `logSAH` in one graphics window where the plot of the histogram above the boxplot.
  - (k) Utilize all 3 graphs to describe the shape of the distribution of `logSAH`.
  - (l) What did the log transformation do to the values of `SAH`?
  - (m) Take a random sample of size 50 from `logSAH` and make a histogram. Does this distribution have a similar shape compared to that of all `logSAH` values?
  - (n) Take a random sample of size 50 with replacement from `logSAH` and make a histogram. Does this distribution have a similar shape compared to that of all `logSAH` values?
4. Answer the following questions pertaining to the variable `medication`:
- (a) Use a function to create frequency table of the number of mothers taking medication and not taking medication.
  - (b) Calculate the percent of the mothers who are taking medication; what is the percentage?
5. Answer the following questions pertaining to the variable `educational_level`:
- (a) Create a frequency table of the number of mothers in each education level.



6. Answer the following questions pertaining to the variable **Status**:

- Are the average **SAH** values for the two levels of **Status** (normal brain development or intellectual disability) different? Formulate a hypothesis, test it, and make a decision about whether or not you can reject the null hypothesis. Can you use a *t*-test (either on the raw or log-transformed data)? Why or why not (hint: check distributions with plots)?
- What is the fold change of log-**SAH** between the 2 groups? Calculate it two ways: use the output from the previous test and also use the data itself (function **mean** plus logical operators).
- Make a boxplot of the **SAH** values of the 2 groups and calculate the fold change of **SAH** between the 2 groups. Does the difference seem clinically relevant? Why or why not?