

Basic Course on R: Hypothesis Testing and Confidence Intervals 2 Practical

Elizabeth Ribble*

18-24 May 2017

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. Answer the following questions pertaining to the variables `vitaminB12` and `homocysteine`:
 - (a) What are the Pearson and Spearman correlations between `vitaminB12` and `homocysteine`? Are they similar? Formulate a hypothesis, do a test, and make a decision as to whether either the Pearson or Spearman correlation is statistically significant. Include a scatterplot of `homocysteine` versus `vitaminB12` to support your findings.
 - (b) Plot a histogram of each variable to decide whether the Pearson correlation is appropriate to use. Is it?
 - (c) Does the correlation improve after a log transformation of both variables? Make plots and do a test on the appropriate (Spearman or Pearson - depends on distribution!) correlation to answer this question.
 - (d) Let’s see what happens when we “categorize” a continuous variable. Cut `vitaminB12` into 4 groups, where the breaks are the 5 quantile points of `vitaminB12`. Make sure you include the lowest breakpoint by specifying “`incl=TRUE`”. Assign the output to `catB12`. What are the levels of this new variable?

- (e) Using the log-transformed variable from part (c), assess how the log of **homocysteine** and **catB12** relate. Make a boxplot of log-**homocysteine** for each level of **catB12**.
- (f) Are the means of log-**homocysteine** equal across all levels of **catB12**? Formulate a hypothesis, test it, and make a decision for statistical significance.
- (g) Now let's see if log-**homocysteine** varies on both **smoking** and **catB12** levels. Make a boxplot of log-**homocysteine** for all combinations of the 2 categories. Then formulate a hypothesis, test it, and make a decision for statistical significance on both categorical variables.
3. We suspect that people who drink alcohol (**alcohol** is yes) might also be smokers (**smoking** is yes). Formulate a hypothesis, test it using the appropriate test, and make a decision about statistical significance. [Hint: use **table**.]