**STAT 40001/STAT 59800 Statistical Computing Fall 2020**

**Lab-12**

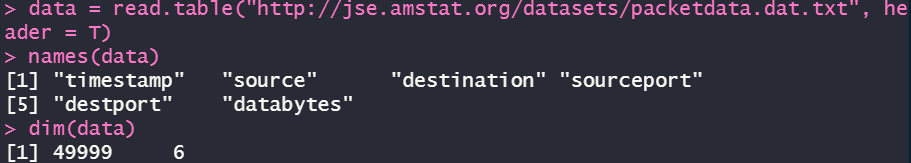
1. The internet traffic data are available in the link below

<http://www.amstat.org/publications/jse/datasets/packetdata.dat.txt>

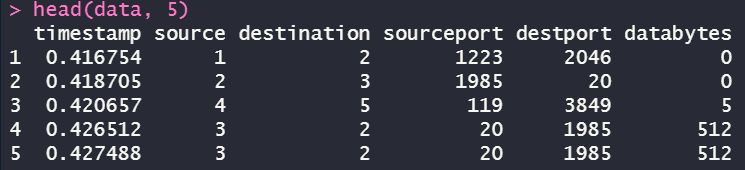
The description of the variables can be obtained below

<https://ww2.amstat.org/publications/jse/datasets/packetdata.txt>

1. Identify the variables and the dimension of the dataset.



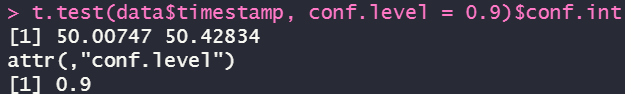
1. Print first five observation of the dataset.



1. Is there any missing value?



1. Construct a 90% confidence interval for the average timestamp.



1. The website below provides the dataset related to the study of the maternal smoking and infant health.

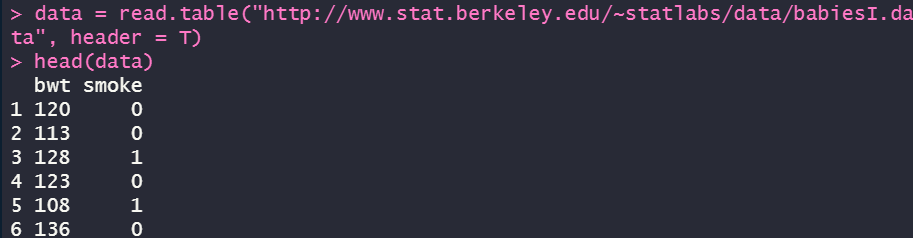
<http://www.stat.berkeley.edu/~statlabs/data/babiesI.data>

There are two variables

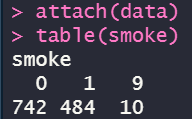
bwt=Birth weight in ounces

smoke=Smoking status of mother 0=not now, 1=yes now, 9=unknown

1. Import the data set in R

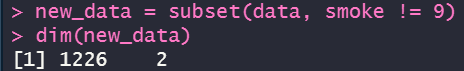


1. How many observations have smoking status unknown?

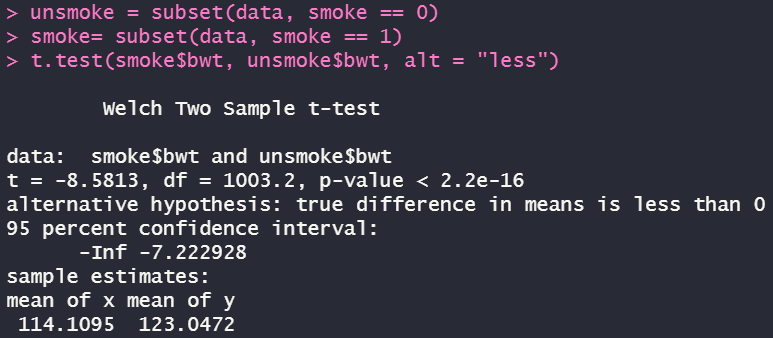


*(There are 10 observations have smoking status unknown)*

1. CLEAN data set by removing subjects whose smoking status is unknown.

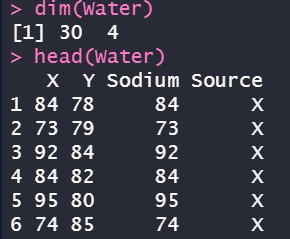


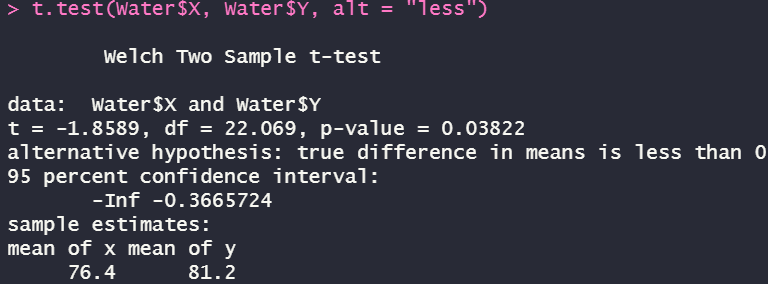
1. Do we have evidence to prove that the newborn baby will have significantly low weight for a smoker mom than for a non-smoker mom?



*(From above the p-value is pretty tiny that we can in favor of the alternative hypothesis that newborn baby will have significantly low weight)*

1. A bottled water company acquires its water from two independent sources, X and Y. Data set *water* in *PASWR* contains the sodium content in each brand of water. Is there statistical evidence to suggest that the average sodium content in the water from source X is less than the average sodium content in water from source Y?





*(based on the result that we got above, the p-value is less than the significance level which means that we can say that source X has less average sodium content than source Y)*

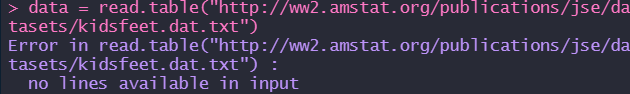
1. Link below provides the foot measurements (in cm) for fourth grade children.

<http://ww2.amstat.org/publications/jse/datasets/kidsfeet.dat.txt>

The description of the variables is provided at

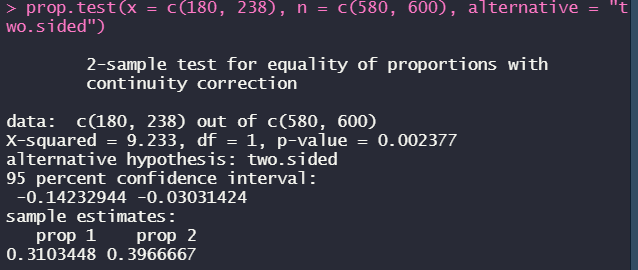
<http://ww2.amstat.org/publications/jse/datasets/kidsfeet.txt>

Do you enough evidence that on average boys foots are longer than of the girls?



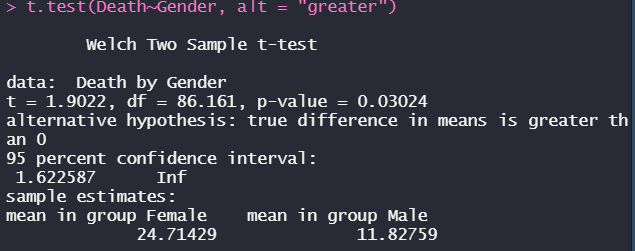
*(somehow it keeps throwing me this error and I have no idea of how to proceed under this condition, but the following processes should be extracting two parts of data for Boys and Girls, then use the “t.test(boys$Length of longer foot, girls$Length of longer foot, alt = “greater”)” to determine if we can follow or reject the null hypothesis based on the p-value we got, if it’s larger than 0.05, then we have strong evidence to follow the null hypothesis and can’t say that the boys foots are longer than those of the girls, or vice versa)*

1. The Pew Research Group conducted a poll in which they asked, “Are you in favor of, or opposed to, executing persons as a general policy when the crime was committed while under the age of 18?” Of the 580 Catholics surveyed, 180 indicated they favored capital punishment; of the 600 seculars (those who do not associate with a religion) surveyed, 238 favored capital punishment. Is there a significant difference in the proportion of individuals in these groups in favor of capital punishment for persons under the age of 18?



*(p-value is pretty small which means that we don’t have strong evidence to support the null hypothesis such that there’s a significant evidence in the proportion of individuals)*

1. Are Female Hurricanes Deadlier than Male Hurricanes? The data set contains archival data on actual fatalities caused by hurricanes in the United States between 1950 and 2012 is provided with this Lab. Please note that two deadliest hurricanes (hurricane Katrina in 2005 (1833 deaths) and Audrey in 1957 (416 deaths) were removed from the data set). Perform the appropriate test.

  
*(the p-value is smaller than 0.05 which means that we can support the alternative hypothesis that Female Hurricanes are deadlier)*