Lab #3: Inference for Numerical Data

Name

Date of lab session

Lab report

```
download.file("http://www.openintro.org/stat/data/nc.RData", destfile = "nc.RData")
load("nc.RData")
```

Load data:

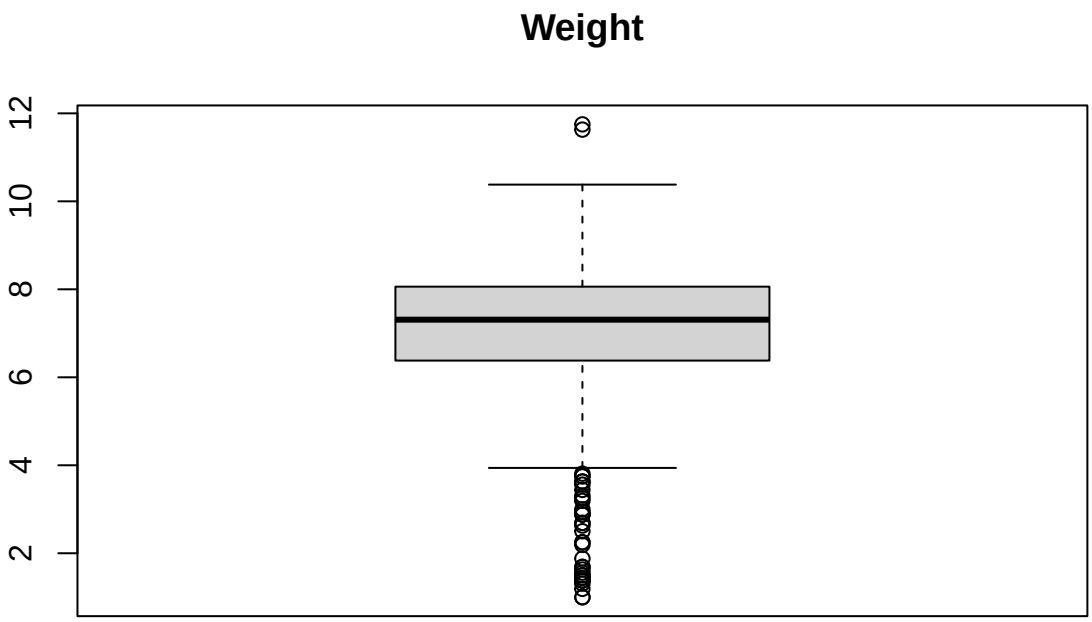
```
set.seed(543)
```

Set a seed:

Exercises:

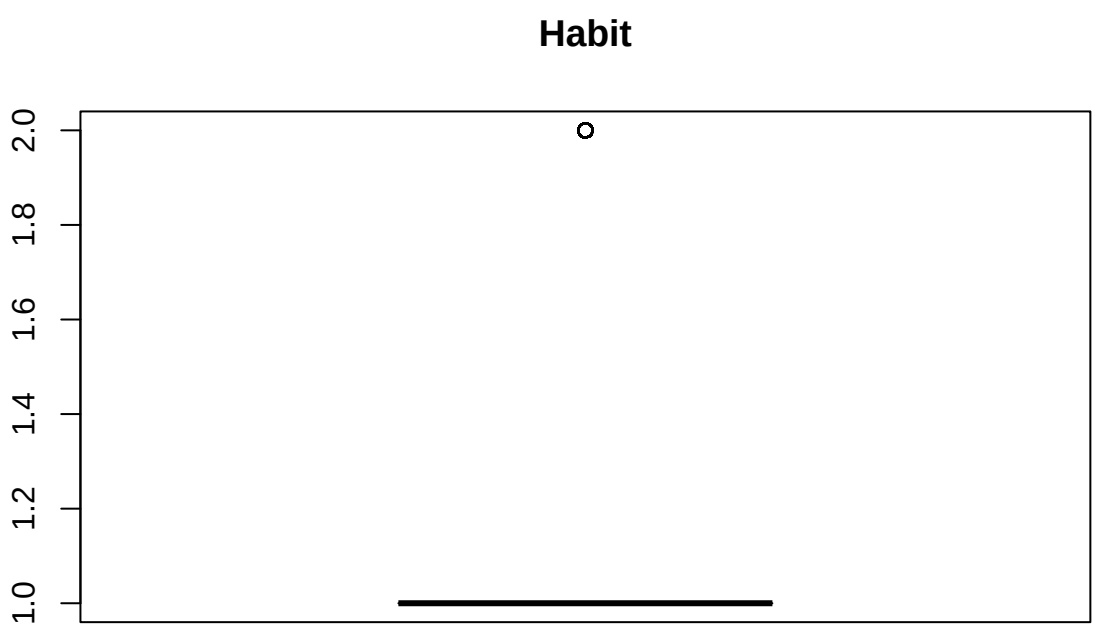
Exercise 1: The cases in this data set are mothers. There are 1000 cases in this data set.

```
boxplot(nc$weight, main = "Weight")
```



Exercise 2:

```
boxplot(nc$habit, main = "Habit")
```



Weight is numerical data while habits are categorical. We can also see that there is a distribution of the data points for weight, but habit only has two possible data points so the plot does not even look like a box plot as there are no quartiles.

```
by(nc$weight, nc$habit, length)
```

Exercise 3:

```
## nc$habit: nonsmoker
## [1] 873
## ------------------------------------------------------------
## nc$habit: smoker
## [1] 126
```

```
by(nc$weight, nc$lowbirthweight, length)
```

```
## nc$lowbirthweight: low
## [1] 111
## ------------------------------------------------------------
## nc$lowbirthweight: not low
## [1] 889
```

```
by(nc$weight, nc$gender, length)
```

```
## nc$gender: female
## [1] 503
## ------------------------------------------------------------
## nc$gender: male
## [1] 497
```

```
by(nc$weight, nc$whitemom, length)
```

```
## nc$whitemom: not white
## [1] 284
## ------------------------------------------------------------
## nc$whitemom: white
## [1] 714
```

```
by(nc$weight, nc$marital, length)
```

```
## nc$marital: married
## [1] 386
## ------------------------------------------------------------
## nc$marital: not married
## [1] 613
```

```
by(nc$weight, nc$premie, length)
```

```
## nc$premie: full term
## [1] 846
## ------------------------------------------------------------
## nc$premie: premie
## [1] 152
```

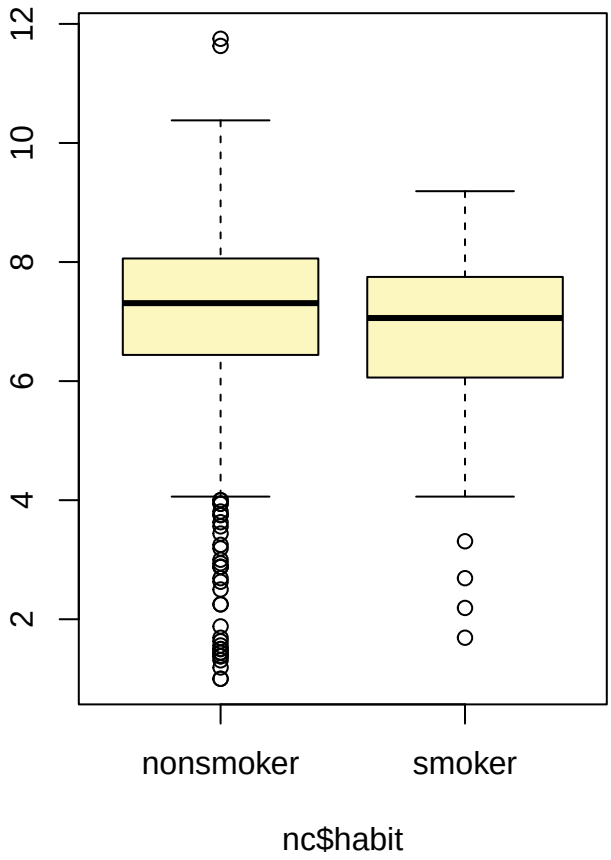
Given that the sample is random. This confirms independence. All possible groups have n>30 meaning they are normal, therefore the conditions for inferences are satisfied.

Exercise 4: Null Hypothesis: Average weight of babies of smokers = average weight of babies of non smokers Alternative Hypothesis: Average weight of babies of smokers not = average weight of babies of non smokers

```
inference(y = nc$weight, x = nc$habit, est = "mean", type = "ci", null = 0,
          alternative = "twosided", method = "theoretical")
```

Exercise 5:

```
## Response variable: numerical, Explanatory variable: categorical
## Difference between two means
## Summary statistics:
## n_nonsmoker = 873, mean_nonsmoker = 7.1443, sd_nonsmoker = 1.5187
## n_smoker = 126, mean_smoker = 6.8287, sd_smoker = 1.3862
```




```
## Observed difference between means (nonsmoker-smoker) = 0.3155
##
## Standard error = 0.1338
## 95 % Confidence interval = ( 0.0534 , 0.5777 )
```

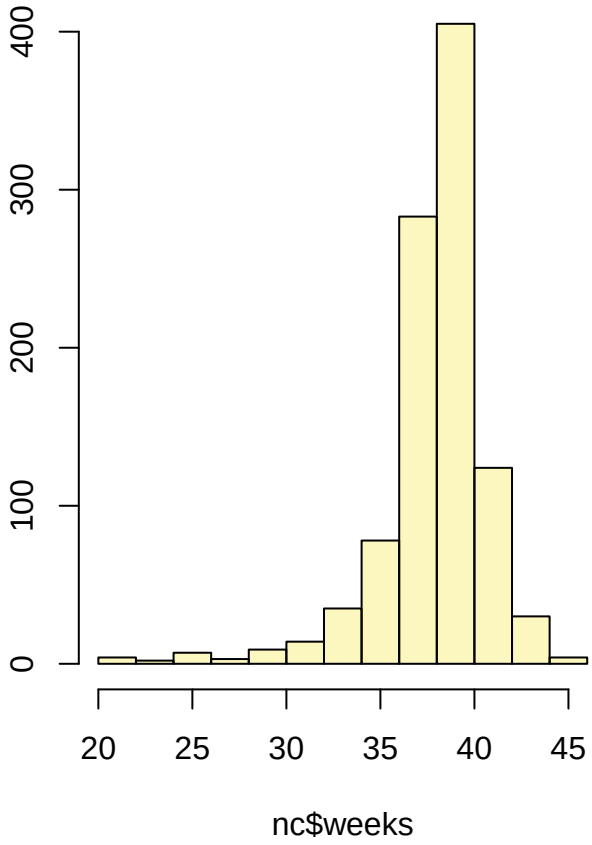
We are 95% confident that the mean weight of the babies of the non-smokers are anywhere from 0.0534 to 0.5777 pounds heavier than babies of smokers. We can most likely say we can reject the null, as we are 95% confident that the difference in means is greater than zero, not zero.

On your own:

```
inference(y = nc$weeks, est = "mean", type = "ci", null = 0,
          alternative = "twosided", method = "theoretical")
```

1:

```
## Single mean
## Summary statistics:
```



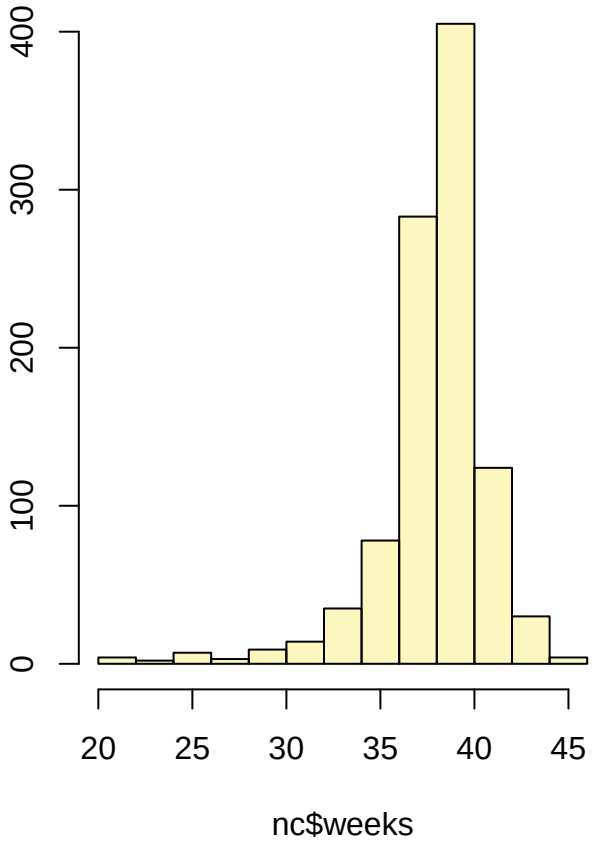

```
## mean = 38.3347 ;  sd = 2.9316 ;  n = 998
## Standard error = 0.0928
## 95 % Confidence interval = ( 38.1528 , 38.5165 )
```

We are 95% confident that the mean length of pregnancies are between 38.1528 and 38.5165 weeks.

```
inference(y = nc$weeks, est = "mean", type = "ci", null = 0,
          alternative = "twosided", method = "theoretical", conflevel = 0.90)
```

2:

```
## Single mean
## Summary statistics:
```



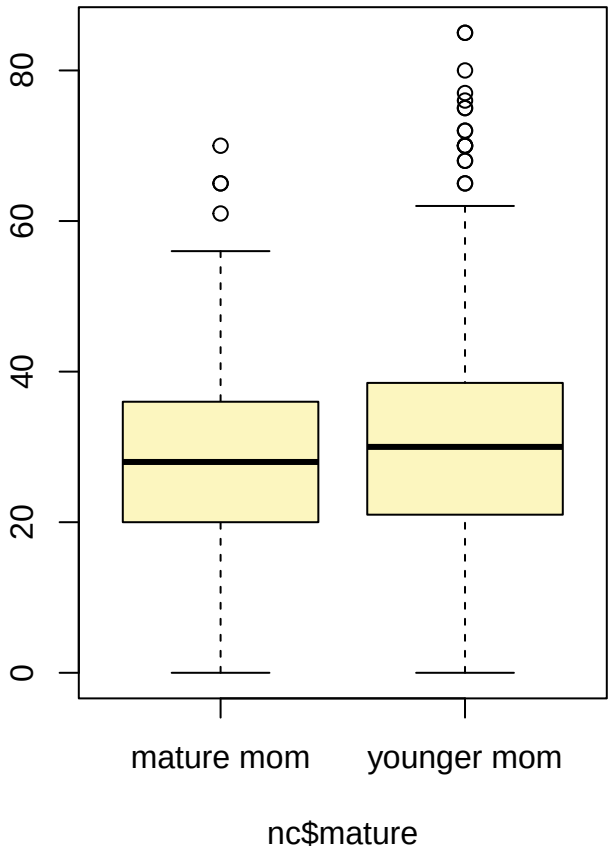

```
## mean = 38.3347 ;  sd = 2.9316 ;  n = 998
## Standard error = 0.0928
## 90 % Confidence interval = ( 38.182 , 38.4873 )
```

We are 90% confident that the mean length of pregnancies are between 38.182 and 38.4873 weeks.

3: Null Hypothesis: average weight gained by young mothers = average weight gained by mature mothers Alternative Hypothesis: average weight gained by young mothers not = average weight gained by mature mothers

```
inference(y = nc$gained, x = nc$mature, est = "mean", type = "ci", null = 0,
          alternative = "twosided", method = "theoretical")
```

```
## Response variable: numerical, Explanatory variable: categorical
## Difference between two means
## Summary statistics:
## n_mature mom = 129, mean_mature mom = 28.7907, sd_mature mom = 13.4824
## n_younger mom = 844, mean_younger mom = 30.5604, sd_younger mom = 14.3469
```




```
## Observed difference between means (mature mom-younger mom) = -1.7697
##
## Standard error = 1.2857
## 95 % Confidence interval = ( -4.2896 , 0.7502 )
```

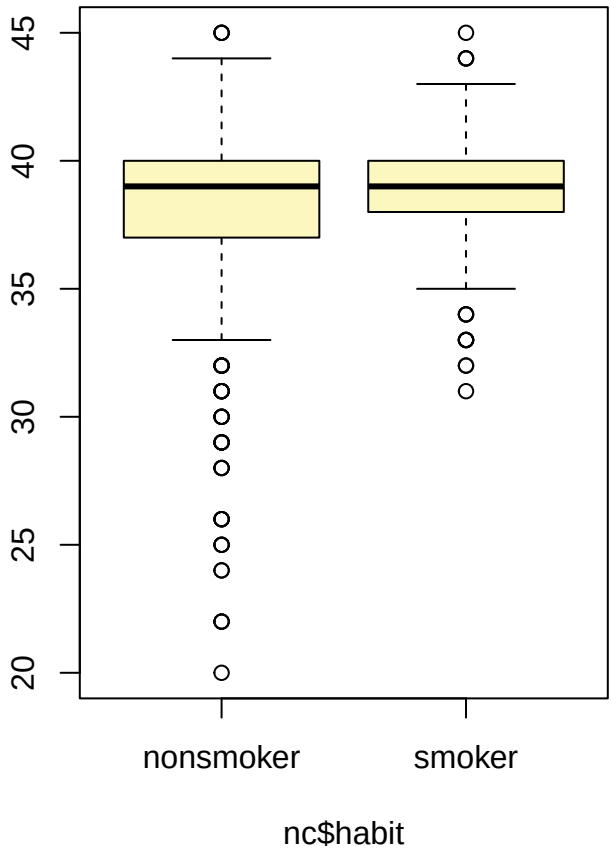
We are 95% confident that the difference in weight gained by mature mothers and young mothers (mature - young)are between -4.2896 and 0.7502 lbs. Cannot reject null, as zero is included in interval.

4: The data set is in numeric order of mage (Mother age) so if you go down the data set, we see the cut off for young vs mature moms at 34(young) and 35(mature).

5: Null Hypothesis: No difference in average length of pregnancies between smokers and non-smokers. Alternative Hypothesis: Difference in average length of pregnancies between smokers and non-smokers.

```
inference(y = nc$weeks, x = nc$habit, est = "mean", type = "ci", null = 0,
          alternative = "twosided", method = "theoretical")
```

```
## Response variable: numerical, Explanatory variable: categorical
## Difference between two means
## Summary statistics:
## n_nonsmoker = 872, mean_nonsmoker = 38.3188, sd_nonsmoker = 2.9936
## n_smoker = 126, mean_smoker = 38.4444, sd_smoker = 2.4676
```




```
## Observed difference between means (nonsmoker-smoker) = -0.1256
##
## Standard error = 0.2421
## 95 % Confidence interval = ( -0.6001 , 0.3488 )
```

Cannot reject null as 0 is included in confidence interval. 95% sure that the difference in average length of pregnancies are between -0.6001 and 0.3488 weeks.