Assigment 7

laknath Gunathilake

February 27, 2020

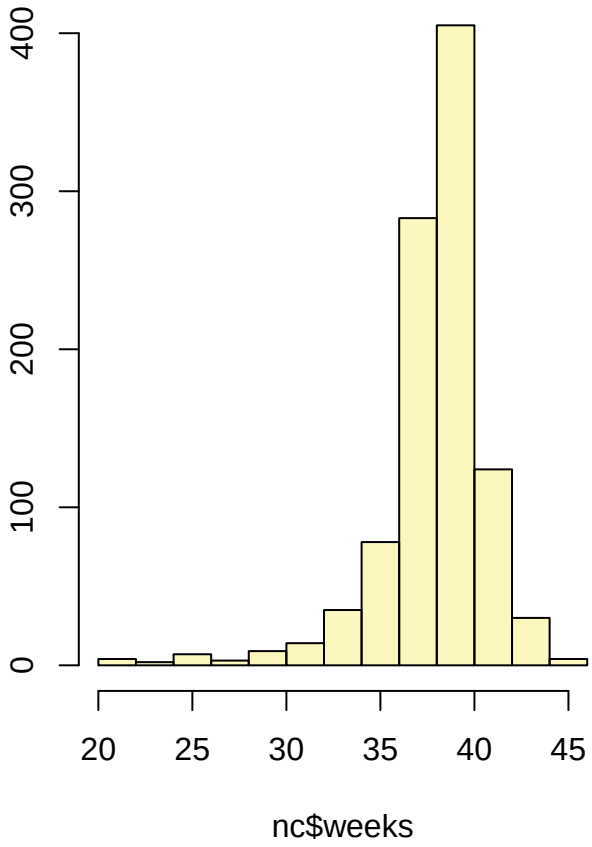
```
download.file("http://www.openintro.org/stat/data/nc.RData", destfile = "nc.RData")
load("nc.RData")
```

On your own

1.Calculate a 95% confidence interval for the average length of pregnancies (weeks) and interpret it in context. Note that since you're doing inference on a single population parameter, there is no explanatory variable, so you can omit the x variable from the function.

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

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




```
## mean = 38.3347 ;  sd = 2.9316 ;  n = 998
## Standard error = 0.0928
```

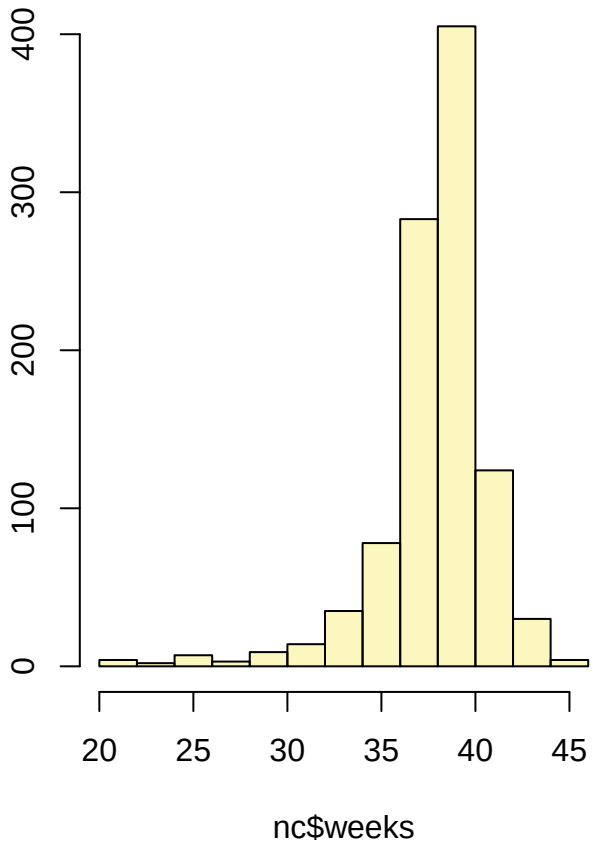
```
## 95 % Confidence interval = ( 38.1528 , 38.5165 )
```

We are 95 percent confident that the average length of a pregnancy is between 38.1528 and 38.5165 weeks

2.Calculate a new confidence interval for the same parameter at the 90% confidence level. You can change the confidence level by adding a new argument to the function: conflevel = 0.90.

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

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




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

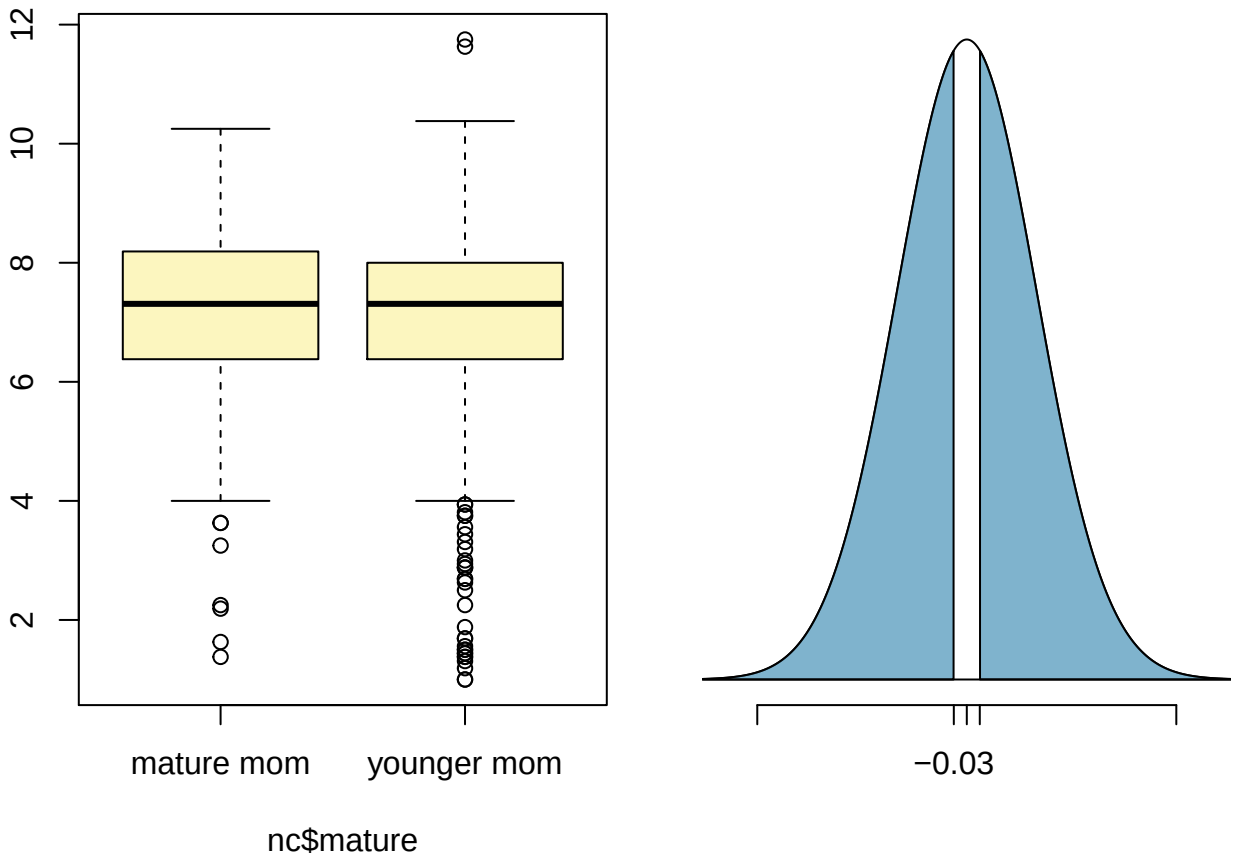
3.Conduct a hypothesis test evaluating whether the average weight gained by younger mothers is different than the average weight gained by mature mothers.

```
inference(y=nc$weight,x = nc$mature,est = "mean",type = "ht",null = 0,alternative = "twosided",method =
```

```
## Response variable: numerical, Explanatory variable: categorical
## Difference between two means
## Summary statistics:
## n_mature mom = 133, mean_mature mom = 7.1256, sd_mature mom = 1.6591
## n_younger mom = 867, mean_younger mom = 7.0972, sd_younger mom = 1.4855

## Observed difference between means (mature mom-younger mom) = 0.0283
##
## H0: mu_mature mom - mu_younger mom = 0
```

```
## HA: mu_mature mom - mu_younger mom != 0
## Standard error = 0.152
## Test statistic: Z =  0.186
## p-value =  0.8526
```



Since the p value is over 0.05 we cannot reject the null hyptothesis that the mean difference is 0

4.Now, a non-inference task: Determine the age cutoff for younger and mature mothers. Use a method of your choice, and explain how your method works.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
nc%>%select(mature,mage)%>%group_by(mature)%>%summarize(min(mage),max(mage))
```

```
## # A tibble: 2 x 3
##   mature      `min(mage)` `max(mage)`
##   <fct>             <int>       <int>
## 1 mature mom           35          50
```

```
## 2 younger mom          13          34
```

My method provides the minimum and and maximum age values for each group (mature and young moms) This indicates that the age cutoff for young moms is 35 (young mom is if they are between 13 and 34) and mature mom if they are between 35 and 50

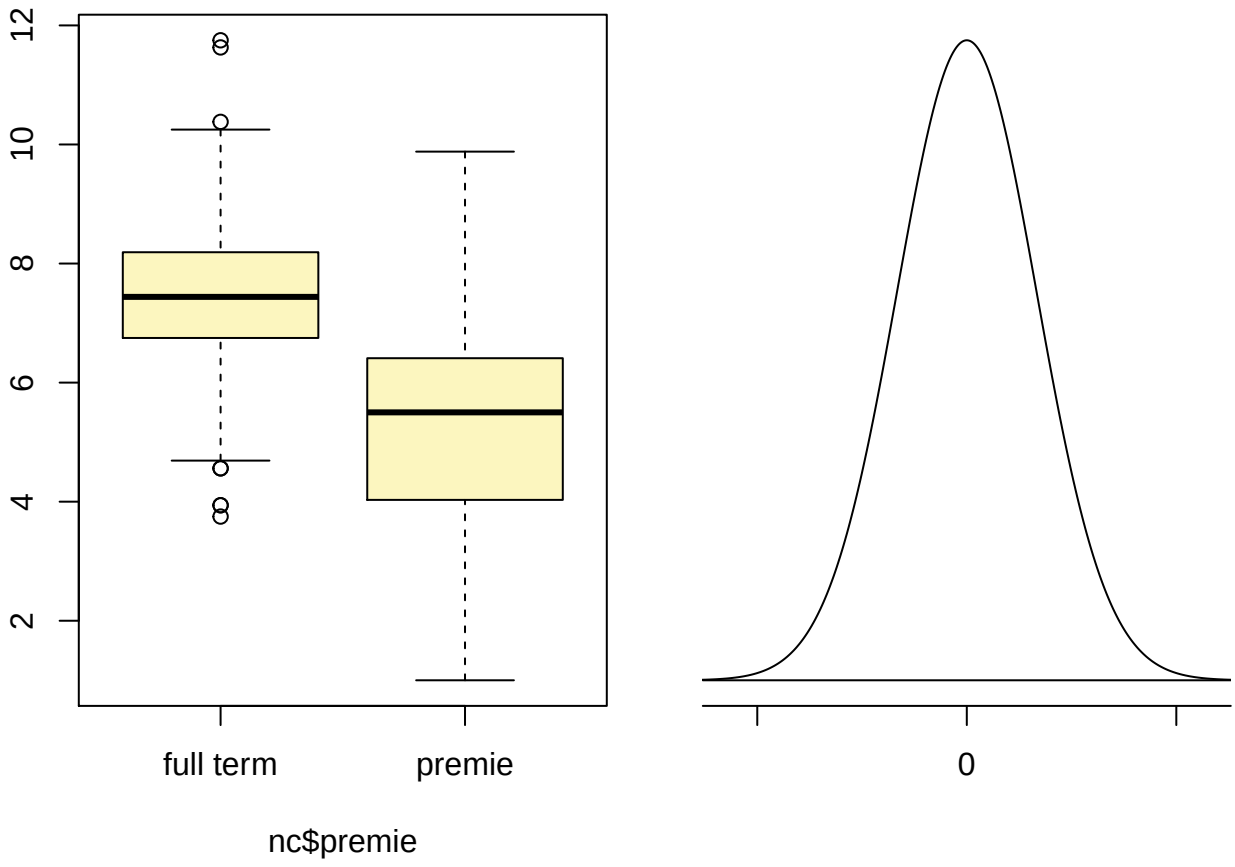
5.Pick a pair of numerical and categorical variables and come up with a research question evaluating the relationship between these variables. Formulate the question in a way that it can be answered using a hypothesis test and/or a confidence interval. Answer your question using the inference function, report the statistical results, and also provide an explanation in plain language.

Research Question Is the average Birthweight of premature babies different from the ones who were due in full term

```
inference(y=nc$weight,x = nc$premie,est = "mean",type = "ht",null = 0,alternative = "twosided",method =
```

```
## Response variable: numerical, Explanatory variable: categorical
## Difference between two means
## Summary statistics:
## n_full term = 846, mean_full term = 7.4594, sd_full term = 1.075
## n_premie = 152, mean_premie = 5.1284, sd_premie = 1.9696
```

```
## Observed difference between means (full term-premie) = 2.331
##
## H0: mu_full term - mu_premie = 0
## HA: mu_full term - mu_premie != 0
## Standard error = 0.164
## Test statistic: Z =  14.216
## p-value =  0
```



Since the p values is less than 0.05 we reject the hypothesis and conclude that that the average difference of birthweight between premature and full term babies is not 0.