

Homework3

Swetha Jayapathy

10/20/2020

R Markdown

```
library(Sleuth3)
library(ggplot2)
```

```
names(ex0332)
```

```
## [1] "College"      "Type"         "InState"      "OutOfState"
```

Question 1a

```
head(ex0332)
```

```
##              College      Type InState OutOfState
## 1      Albany State University Public    5434    17048
## 2    Appalachian State University Public    5175    16487
## 3    Argosy University: Nashville Private  19596    19596
## 4      Brescia University Private   18140    18140
## 5 Central Connecticut State University Public   8055    18679
## 6    Christopher Newport University Public  10084    19306
```

Question 1b :

```
with(ex0332, summary(InState))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3583   6706   13852   17781   22971   97716
```

```
with(ex0332, summary(OutOfState))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4966   15454   18136   22300   24023   97716
```

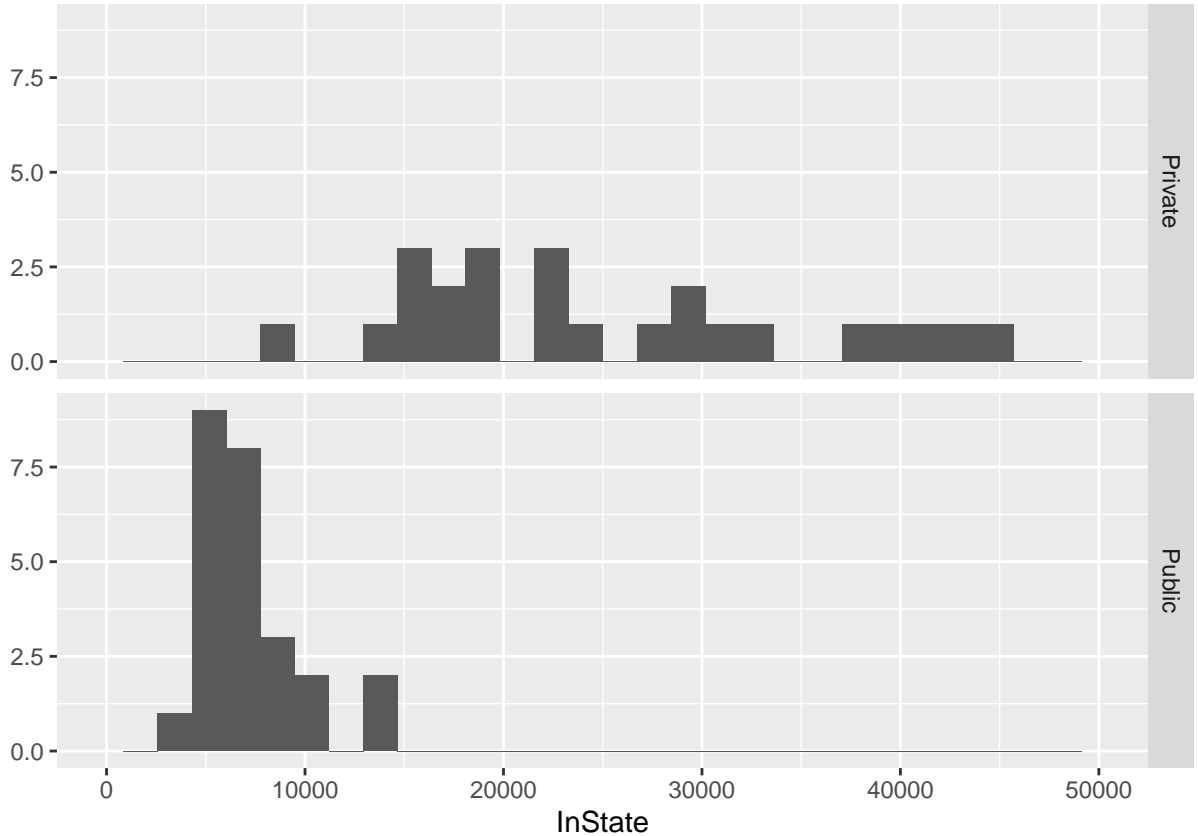
Question 1c :

```
qplot(InState, data=ex0332, geom="histogram", xlim=c(0,50000)) + facet_grid(Type ~ .)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 4 rows containing missing values (geom_bar).
```

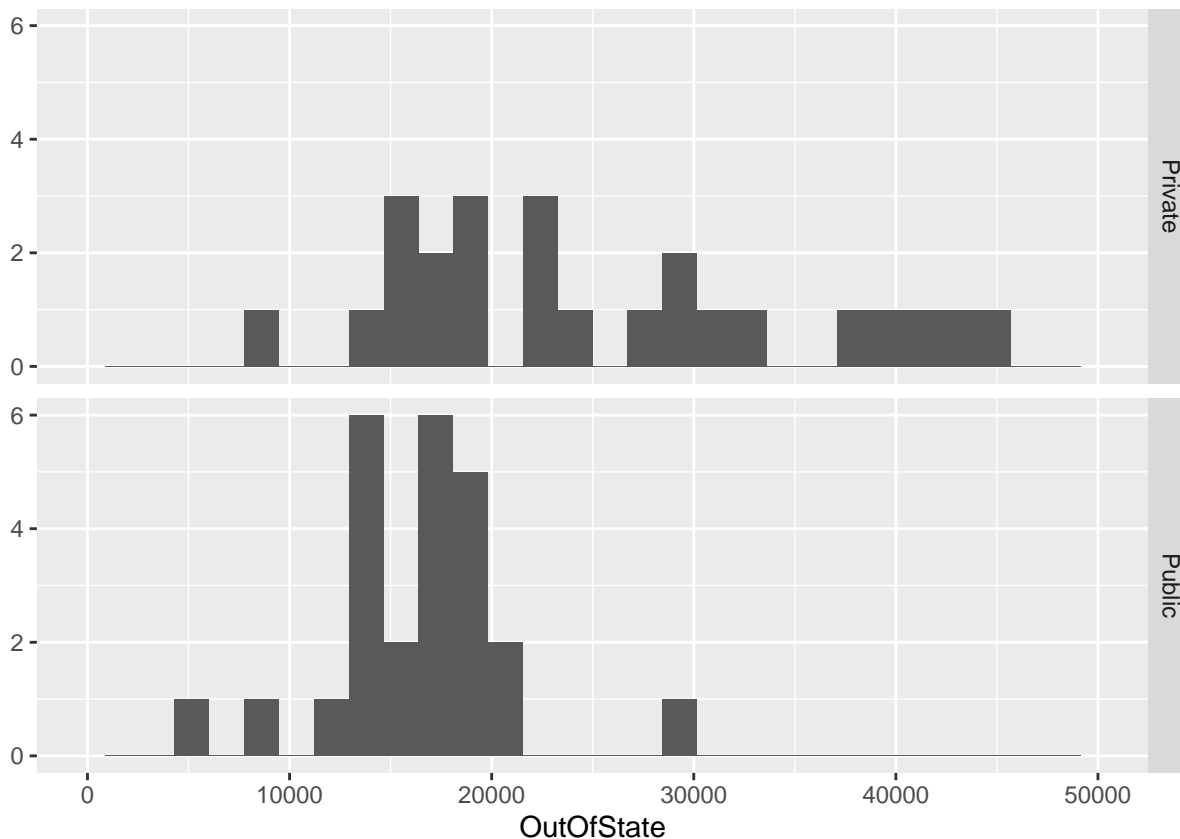


```
qplot(OutOfState, data=ex0332, geom="histogram", xlim=c(0,50000)) + facet_grid(Type ~ .)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 4 rows containing missing values (geom_bar).
```



Question 1d

```
#Summary of Public College
log_OutOfState<-log(ex0332$OutOfState)
with(ex0332, summary(log_OutOfState[Type=="Public"]))
```

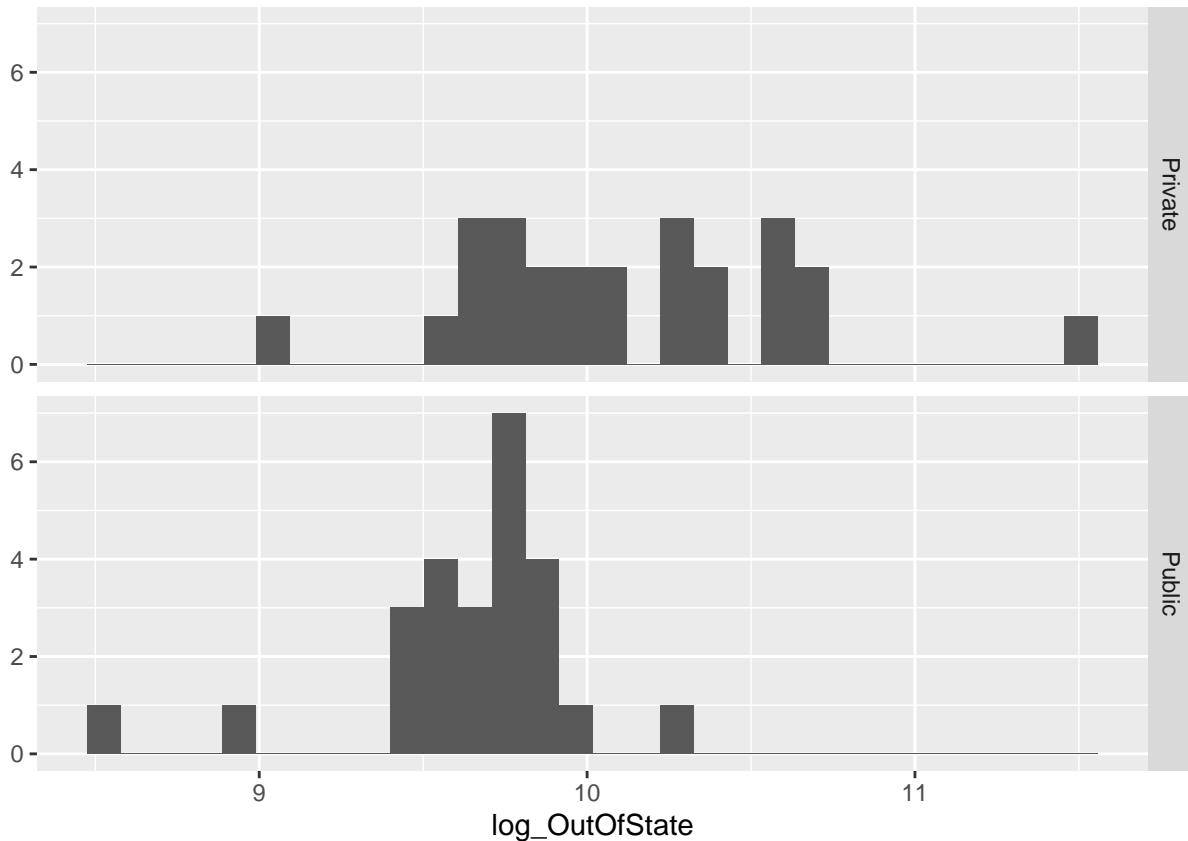
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  8.510   9.524   9.710   9.647   9.808  10.280
```

```
#Summary of Private College
log_OutOfState<-log(ex0332$OutOfState)
with(ex0332, summary(log_OutOfState[Type=="Private"]))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  9.082   9.797  10.051  10.125  10.415  11.490
```

```
qplot(log_OutOfState, data=ex0332, geom="histogram") + facet_grid(Type ~ .)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Question 1e

Normality : This assumption states that the values are normally distributed. We can see from the graphs of the untransformed data of both the samples that it is very skewed and therefore it is not reasonable to apply this assumption. Whereas the distribution of the transformed data is normally distributed and hence it would be easy to apply our assumption.

Equal Variance : As per the distribution and the box plot of untransformed data, equal variance does not apply due to its varying sizes. However, when the data is transformed, the box plots looks quite similar and hence equal variance makes sense

Independence : Data in the two samples are collected randomly which results in random sampling. Data in both the groups are independent of each other. Therefore Independence assumption holds good in both untransformed data as well as transformed data.

Question 1f

```
t.test(log_OutOfState~Type, data=ex0332, var.equal=TRUE,
       alternative="greater")
```

```
##
## Two Sample t-test
##
## data: log_OutOfState by Type
## t = 4.0051, df = 48, p-value = 0.0001073
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 0.2779993      Inf
```

```
## sample estimates:
## mean in group Private mean in group Public
##          10.124980          9.646683
```

It can be seen from the above 2 sample t-test that the mean of Out of State tuition fees is greater at Private college than that in Public college (p-value = 0.0001073).

Question 1g

```
exp(10.124980)/exp(9.646683)
```

```
## [1] 1.613325
```

Statistical Conclusion We have strong evidence that the median of Private Universities is higher than the median of Public Universities for Out of state tuition fees (p-value = 0.0001073).

Question 1h

```
t.test(log_OutOfState~Type, data=ex0332, var.equal=TRUE)
```

```
##
## Two Sample t-test
##
## data: log_OutOfState by Type
## t = 4.0051, df = 48, p-value = 0.0002145
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2381828 0.7184111
## sample estimates:
## mean in group Private mean in group Public
##          10.124980          9.646683
```

```
#exp(c(0.2381828, 0.7184111))
exp(0.2381828)
```

```
## [1] 1.268941
```

```
exp(0.7184111)
```

```
## [1] 2.051172
```

The 95% confidence interval for the difference in population mean ranges from 1.268941 to 2.051172.

Question 1i

Statistical Conclusion:

It is evident that the log mean of Private Universities is higher than the log mean of Public Universities for Out of state tuition fees. The estimated ratio of median in Private college to the median in Public college for Out of State tuition fees is 1.613325(p-value = 0.0001073) with 95% confidence interval in the range 1.268941 to 2.051172.