

## Assignment-4.R

Purbasha Chatterjee

Thu Oct 19 12:38:00 2017

```
library(Sleuth3)

## Warning: package 'Sleuth3' was built under R version 3.4.2

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.2

##Q1-Ex2.24
wilcox.test(Score ~ Treatment, data = case0101, alternative = "greater")

## Warning in wilcox.test.default(x = c(5, 5.4, 6.1, 10.9, 11.8, 12, 12.3, :
## cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: Score by Treatment
## W = 147.5, p-value = 0.997
## alternative hypothesis: true location shift is greater than 0

library(exactRankTests)

## Warning: package 'exactRankTests' was built under R version 3.4.2

## Package 'exactRankTests' is no longer under development.
## Please consider using package 'coin' instead.

wilcox.exact(Score ~ Treatment, data = case0101, conf.int = TRUE)

##
## Exact Wilcoxon rank sum test
##
## data: Score by Treatment
## W = 147.5, p-value = 0.005547
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
## -6.6 -1.0
## sample estimates:
## difference in location
## -3.45
```

```

t.test(Score~Treatment, data=case0101)

##
## Welch Two Sample t-test
##
## data: Score by Treatment
## t = -2.9153, df = 43.108, p-value = 0.005618
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -7.010803 -1.277603
## sample estimates:
## mean in group Extrinsic mean in group Intrinsic
## 15.73913 19.88333

t.test(Score~Treatment, data=case0101, var.equal = TRUE)

##
## Two Sample t-test
##
## data: Score by Treatment
## t = -2.9259, df = 45, p-value = 0.005366
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -6.996973 -1.291432
## sample estimates:
## mean in group Extrinsic mean in group Intrinsic
## 15.73913 19.88333

ext <- case0101[ which(case0101$Treatment=="Extrinsic") , ]
intr <- case0101[ which(case0101$Treatment=="Intrinsic") , ]

```

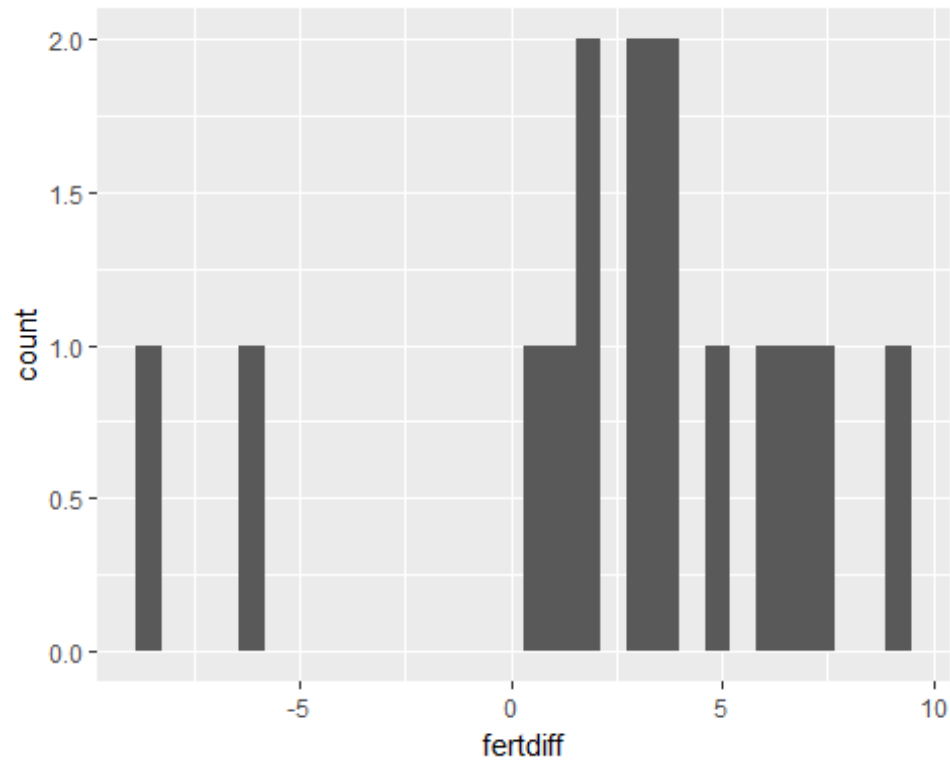
In rank sum test we find the 95% CI range between 1.0 to 6.6, whereas in case of t-test-the 95% CI lies between 1.3 to 7. Thus, we observe there is 0.1 decrement in range that is t-test range difference is 5.7 and rank sum-test difference is 5.6. Additionally, we can observe that in rank sum test the lower interval is 1 as compared to 1.3 in t-test and upper interval 6.6 as compared to 7 in t-test. Thus, low interval values are considered here. It seems to make a slight difference over the range, depending upon the experiment.

```

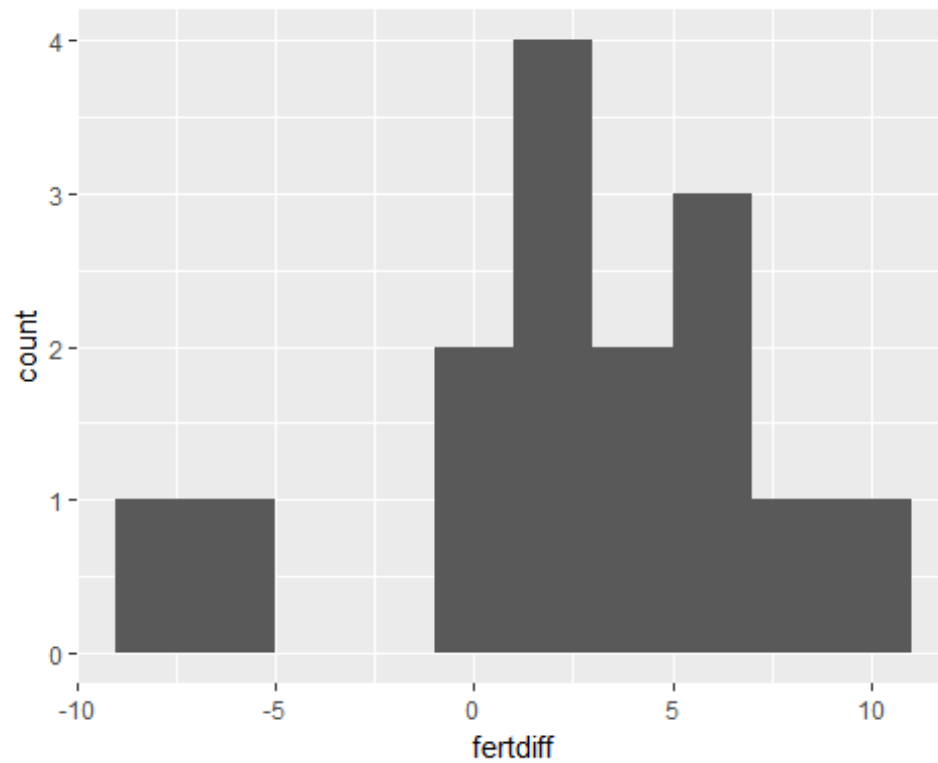
##Q2-Ex2.28
fertdiff <- ex0428$Cross-ex0428$Self
ggplot(data=ex0428, aes(fertdiff))+geom_histogram()

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

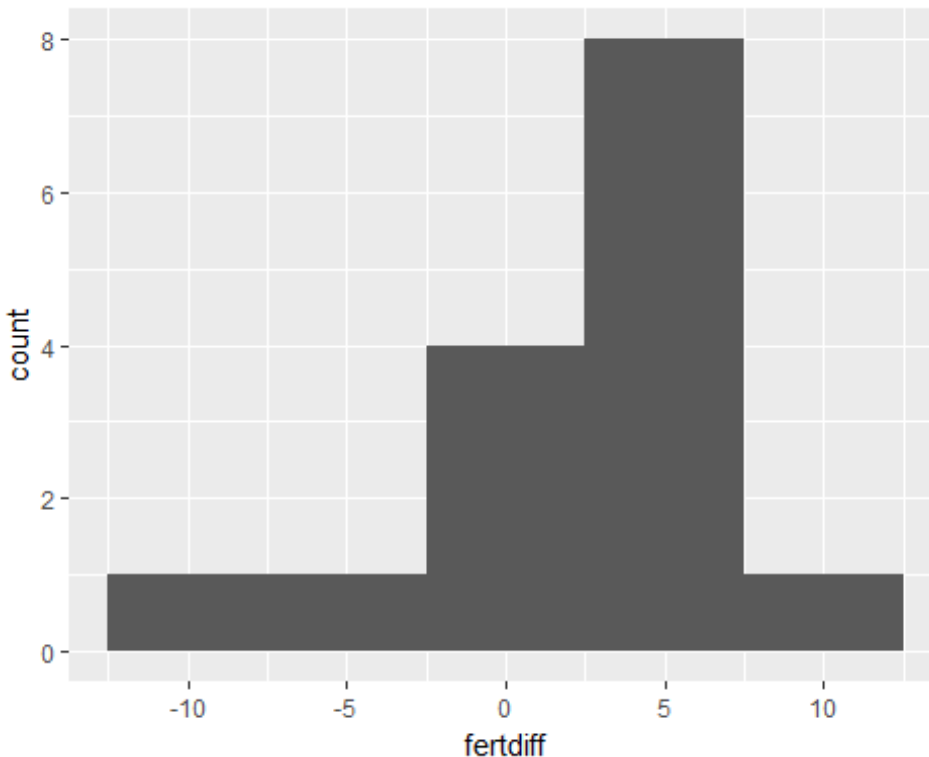
```



```
ggplot(data=ex0428, aes(fertdiff))+geom_histogram(binwidth = 2)
```



```
ggplot(data=ex0428, aes(fertdiff))+geom_histogram(binwidth = 5)
```



```
b) ##paired t-test for the hypothesis
t.test(ex0428$Cross, ex0428$Self, paired = TRUE)

##
## Paired t-test
##
## data: ex0428$Cross and ex0428$Self
## t = 2.1474, df = 14, p-value = 0.04976
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.003165044 5.230168289
## sample estimates:
## mean of the differences
##                2.616667
```

The p-value is less than 0.05 and hence we can reject the null hypothesis.

```
c) ##95% CI for additive tretment effect
```

```
wilcox.exact(ex0428$Cross,ex0428$Self, conf.int = TRUE, paired=TRUE)
```

Exact wilcoxon signed rank test

```
data: ex0428$Cross and ex0428$Self
v = 96, p-value = 0.04126
alternative hypothesis: true mu is not equal to 0
95 percent confidence interval:
 0.50 5.19
sample estimates:
```

(pseudo)median  
3.125

Thus, the 95% confidence interval lies between 0.5 to 5.2.

Although it is a paired test, I tried to run the below test to understand the results.

```
wilcox.exact(ex0428$Cross,ex0428$Self, conf.int = TRUE)
```

```
##
## Exact Wilcoxon rank sum test
##
## data: ex0428$Cross and ex0428$Self
## W = 185.5, p-value = 0.001715
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
##  1.63 4.87
## sample estimates:
## difference in location
##                3.31
```

d) In the histogram, certain outliers can be observed. The interval between 2.5-5, a moderate skewness could be observed. This is because of the presence of certain difference points which are far from the mean value.

e) ## signed rank test to find a two-sided p-value

```
wilcox.test(ex0428$Cross,ex0428$Self, paired = TRUE, alternative = "two.sided")
```

```
##
## Wilcoxon signed rank test
##
## data: ex0428$Cross and ex0428$Self
## V = 96, p-value = 0.04126
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.exact(ex0428$Cross,ex0428$Self, paired = TRUE, alternative = "two.sided")
```

```
##
## Exact Wilcoxon signed rank test
##
## data: ex0428$Cross and ex0428$Self
## V = 96, p-value = 0.04126
## alternative hypothesis: true mu is not equal to 0
```

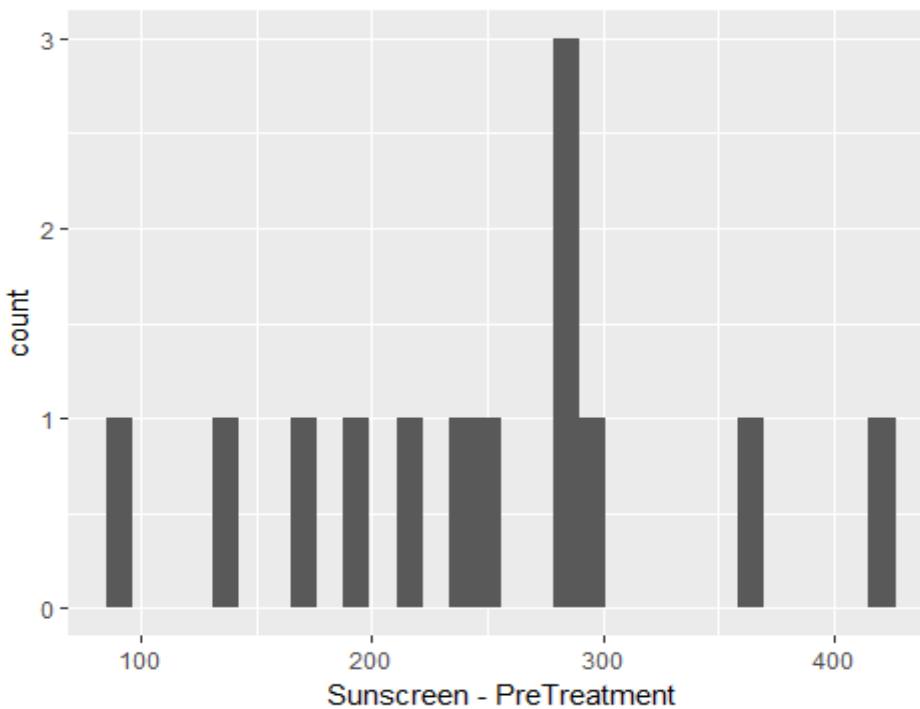
The p-value is 0.04126, which is less than 0.05. Hence, we can reject the null hypothesis.

```
##Q3-Ex2.30
t.test(ex0430$Sunscreen, ex0430$PreTreatment, paired=TRUE)

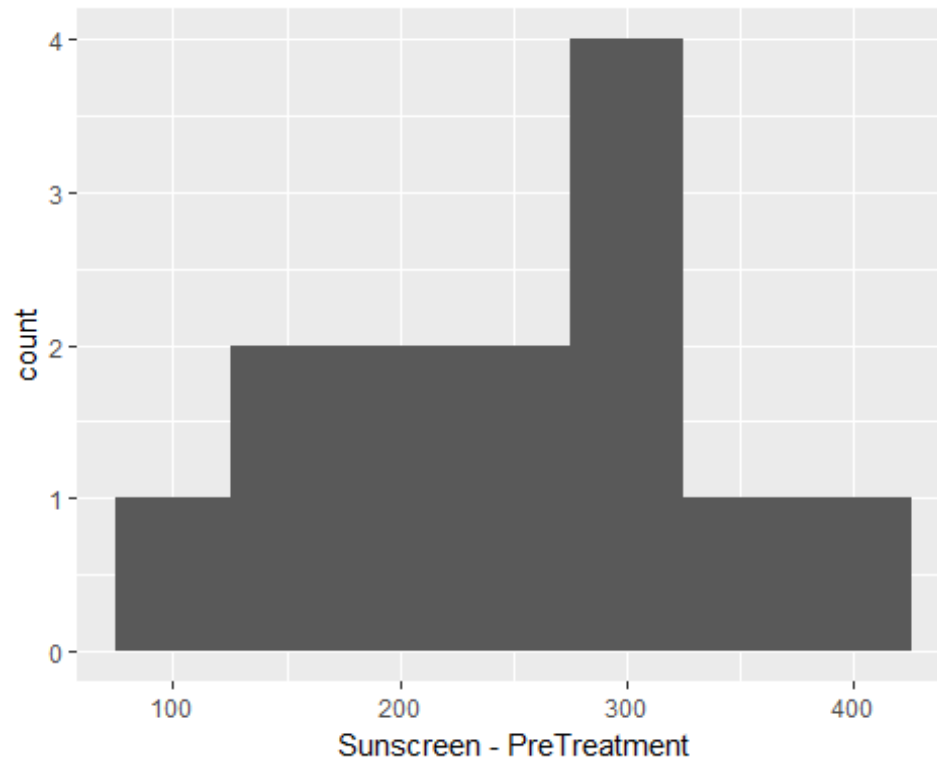
##
## Paired t-test
##
## data: ex0430$Sunscreen and ex0430$PreTreatment
## t = 10.015, df = 12, p-value = 3.525e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 194.1042 302.0497
## sample estimates:
## mean of the differences
## 248.0769

ggplot(ex0430, aes(Sunscreen-PreTreatment))+geom_histogram()

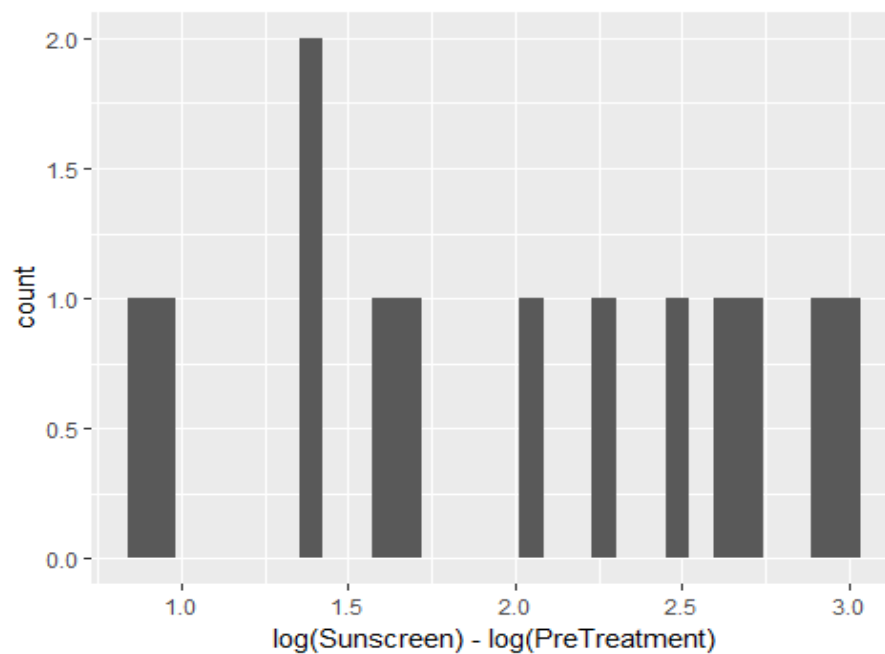
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



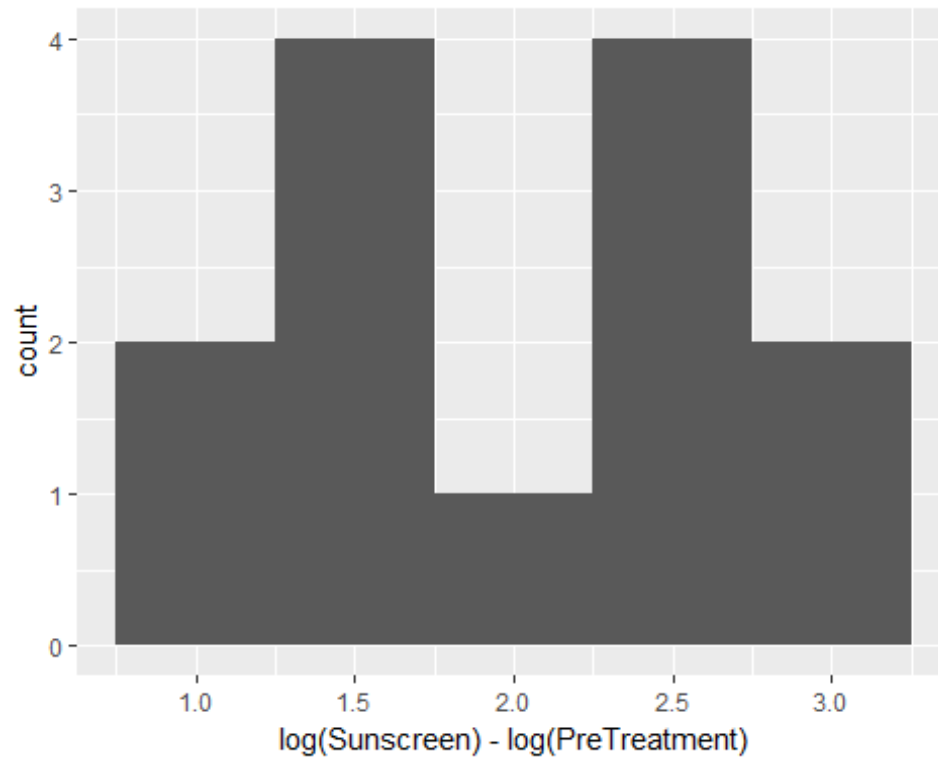
```
ggplot(ex0430, aes(Sunscreen-PreTreatment))+geom_histogram(binwidth = 50)
```



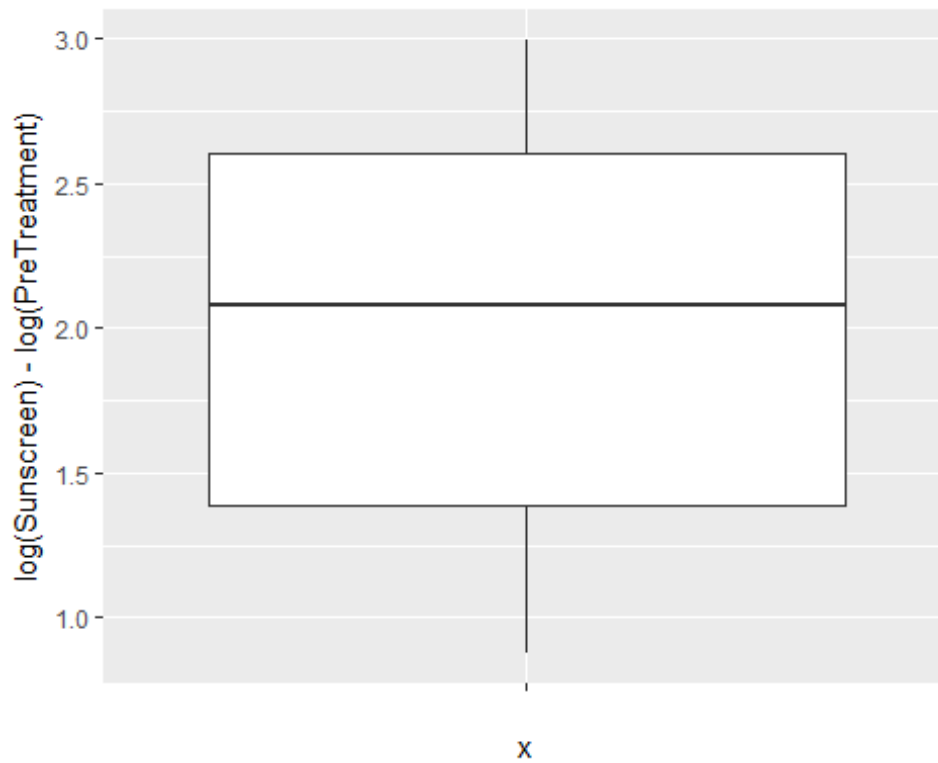
```
ggplot(ex0430, aes(log(Sunscreen)-log(PreTreatment)))+geom_histogram()
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggplot(ex0430, aes(log(Sunscreen)-log(PreTreatment)))+geom_histogram(binwidth
= 0.5)
```



```
ggplot(ex0430, aes(" ", log(Sunscreen) - log(PreTreatment)))+geom_boxplot()
```





```
wilcox.exact(ex0430$Sunscreen,ex0430$PreTreatment,paired = TRUE, conf.int = TRUE)
```

```
## Exact Wilcoxon signed rank test
```

```
data: ex0430$Sunscreen and ex0430$PreTreatment  
V = 91, p-value = 0.0002441  
alternative hypothesis: true mu is not equal to 0  
95 percent confidence interval:  
 192.5 300.0  
sample estimates:  
(pseudo)median  
 248.75
```

```
wilcox.exact(ex0430$Sunscreen-ex0430$PreTreatment,conf.int=TRUE)
```

```
##  
## Exact Wilcoxon signed rank test  
##  
## data: ex0430$Sunscreen - ex0430$PreTreatment  
## V = 91, p-value = 0.0002441  
## alternative hypothesis: true mu is not equal to 0  
## 95 percent confidence interval:  
## 192.5 300.0  
## sample estimates:  
## (pseudo)median  
## 248.75
```

```
exp(wilcox.test(log(ex0430$Sunscreen)-log(ex0430$PreTreatment),conf.int=TRUE)  
$conf.int)
```

```
## [1] 4.618802 12.247449  
## attr(,"conf.level")  
## [1] 0.95
```

##[ 4.62 , 12.25 ]With 95% confidence, we estimate the median person will experience an SPF from 4.6 to 12.2. Note: the paired t-Test gives a similar answer. Using this test has little practical effect on the analysis.

```
exp(t.test(log(ex0430$Sunscreen)-log(ex0430$PreTreatment))$conf)
```

```
## [1] 4.760755 11.426252  
## attr(,"conf.level")  
## [1] 0.95
```

##[ 4.76 , 11.43 ]There may be confounding depending on the details of the experiment, but the fact that pairs consist of the same individual strengthen the conclusion

- The paired t-test gives a p-value less than 0.05, allowing us to reject the null hypothesis. The 95% CI lies between 194.1042 to 302.0497. The mean difference is 248.0769
- Log transformation was applied to normalize the outliers. But it does not seem to make a great effect. From the box plot it can be observed that the median of the log difference is around 2.1.
- The signed rank test has been used to generate the confidence interval. The 95% CI value changes to 192.5 to 300.0 as compared to 194.1042 to 302.0497 in t-test. The mean difference observed here is 248.75. The paired test behaves in the same manner for the difference of sunscreen and pre-sunscreen treatment effect.
- The tests were applied over the log transformed values to observe their behavior. It was observed that 95% CI value lies between 4.62 to 12.25- when signed rank test was applied, which means that median person will experience the SPF in the given range. In t-test the 95% CI lies between 4.760755 to 11.426252. The paired t-Test generates a similar answer, so it could be said that it has hardly made any effect on the analysis.
- We can say that there might be some confounding variables depending on the experiment. If the pairs use the same individual, then it would fortify our conclusion.