Homework #09

Robert Campbell

26 Apr 2021

Chapter 09

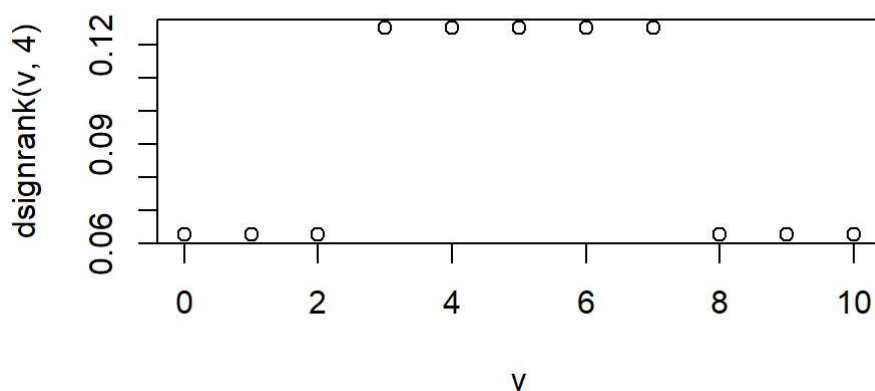### Problem 02

```
2*sum(dsignrank(601:1000,40))
```

```
## [1] 0.009381349
```

Problem 04

```
x <- c(-1,0,2,8)
mu <- 3
xdif <- x-mu
xrank<-rank(abs(xdif))
#sum(xrank[xdif > 0])
v<-0:10
plot(v,dsignrank(v,4))
```



```
2*(sum(dsignrank(0:4, 4)))
```
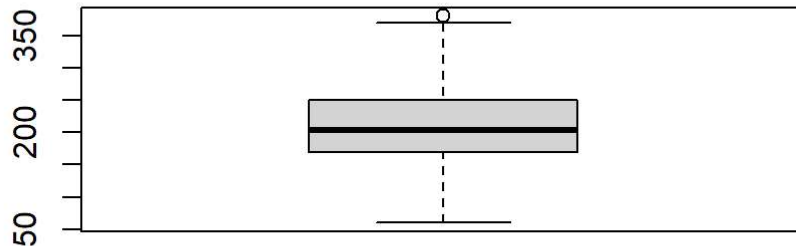
```
## [1] 0.875
```

```
wilcox.test(x, mu=mu)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  x
## V = 4, p-value = 0.875
## alternative hypothesis: true location is not equal to 3
```

Problem 06

a. The data seems reasonably symmetric.

```
weight<- fosdata::weight_estimate$mean200
boxplot(weight)
```


b. The weight that was estimated by

the population was not close to the actual weight of the object.
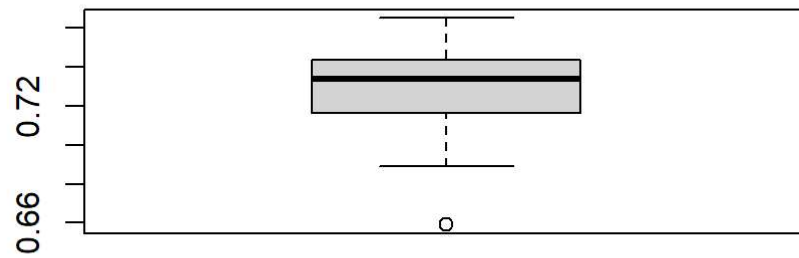
```
wilcox.test(weight, mu=200)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  weight
## V = 1708, p-value = 0.2052
## alternative hypothesis: true location is not equal to 200
```
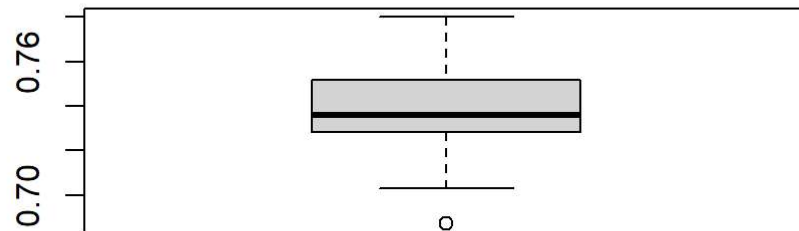
Problem 09

a.

```
sparrow<-Sleuth3::ex0221
boxplot(sparrow$Humerus[sparrow$Status == "Perished"])
```

```
boxplot(sparrow$Humerus[sparrow$Status == "Survived"])
```

b.

```
wilcox.test(Humerus ~ Status, data=sparrow)
```

```
##
##   Wilcoxon rank sum test with continuity correction
##
## data:  Humerus by Status
## W = 331, p-value = 0.1718
## alternative hypothesis: true location shift is not equal to 0
```

c. No, there is not enough evidence to conclude that there is significant difference in Humerus length between the two populations.
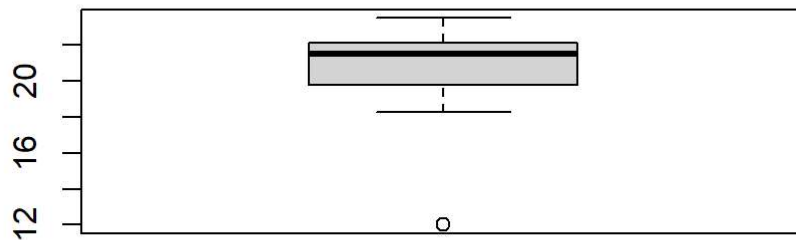
Problem 10

a. Yes, there is significant difference in the height of the flowers.

```
flower<- Sleuth3::ex0428
wilcox.test(flower$Cross, flower$Self)
```
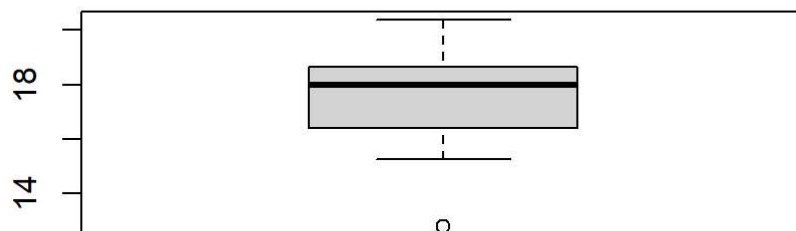
```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  flower$Cross and flower$Self
## W = 185.5, p-value = 0.002608
## alternative hypothesis: true location shift is not equal to 0
```

b. Both datasets are skew with one significant outlier.

```
boxplot(flower$Cross)
```



```
boxplot(flower$Self)
```


Problem 11

a. Hypothesis: There will be significant difference in antibody levels between the sample with malaria symptoms and the sample without.

P = 2.127e-05, reject the null, there is significant difference in antibody levels.

```
malaria <- ISwR::malaria
wilcox.test(ab ~ mal, data=malaria)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  ab by mal
## W = 1533.5, p-value = 2.127e-05
## alternative hypothesis: true location shift is not equal to 0
```

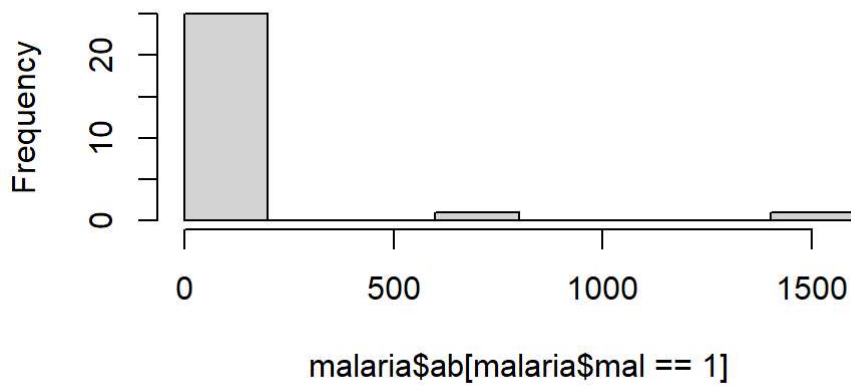b. No, the data is not normal and does not meet the assumptions of the t.test

```
hist(malaria$ab[malaria$mal == 0])
```

Histogram of malaria$ab[malaria$mal == 0]
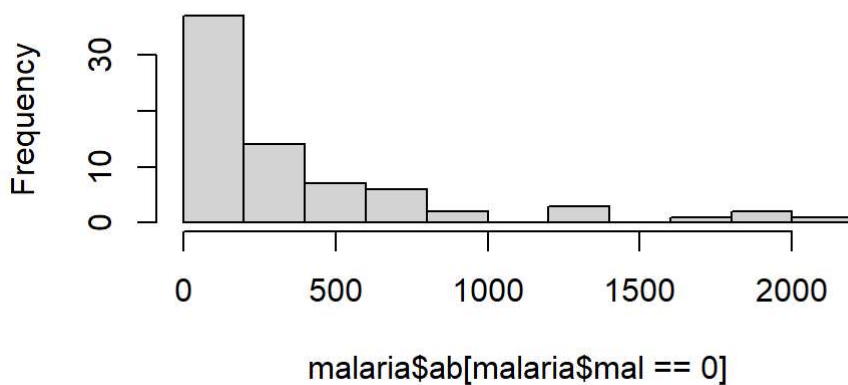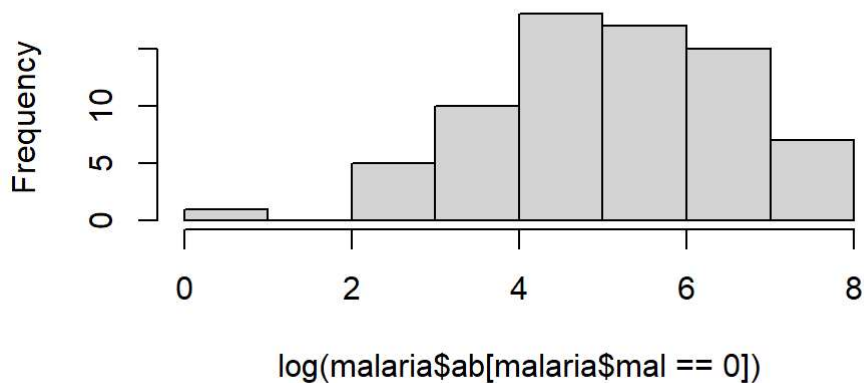


```
hist(malaria$ab[malaria$mal == 1])
```

Histogram of malaria$ab[malaria$mal == 1]

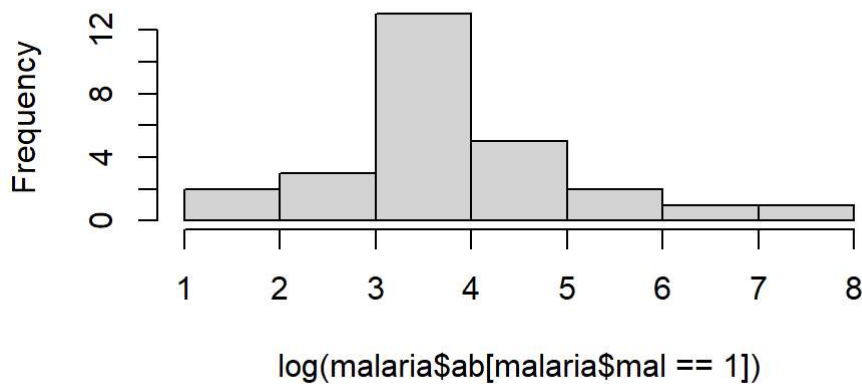c. Taking the log of the data does make it roughly normal, making the t.test more viable.

```
hist(log(malaria$ab[malaria$mal == 0]))
```

Histogram of log(malaria$ab[malaria$mal == 0])


```
hist(log(malaria$ab[malaria$mal == 1]))
```

Histogram of log(malaria$ab[malaria$mal == 1])

d. Like the Wilcoxon test, this does show significance with a p-value of 2.376e-05. However, the p-value from the t.test is larger than the Wilcoxon p-value 2.127e-05, the difference is likely due to the fact that even the log of the data is still not completely normal.

```
t.test(log(ab) ~ mal, data=malaria)
```

```
##
##  Welch Two Sample t-test
##
## data:  log(ab) by mal
## t = 4.6398, df = 52.256, p-value = 2.376e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.7831352 1.9764775
## sample estimates:
## mean in group 0 mean in group 1
##        5.126524        3.746717
```

Problem 15

a. ~16% of the time

```
test<-replicate(10000,{x<-rnorm(100,0.1,1); t.test(x, mu=0)$p.value})
mean(test<0.05)
```

```
## [1] 0.1718
```

b. ~16% of the time

```
test<-replicate(10000,{x<-rnorm(100,0.1,1); wilcox.test(x, mu=0)$p.value})
mean(test<0.05)
```

```
## [1] 0.1642
```

c. t.test: ~88% of the time

```
test<-replicate(10000,{x<-rnorm(1000,0.1,1); t.test(x, mu=0)$p.value})
mean(test<0.05)
```

```
## [1] 0.8853
```

```
 wilcox.test: ~88% of the time
```

```
test<-replicate(10000,{x<-rnorm(1000,0.1,1); wilcox.test(x, mu=0)$p.value})
mean(test<0.05)
```

```
## [1] 0.8708
```

d. Since their accuracy is similar, anecdotally I'd suggest running t.test as wilcox.test took significantly longer with larger n.

Problem 16

a.   i. The true mean is 0
       ii. Type-II
b. Continuing the assumptions from a, a Type-I error has 0% probability of occurring as it cannot happen when the null is false.
c. Type-I error has 0% probability of occurring as it cannot happen when the null is false.
d. Type-II error has ~1.2% probability of occurring in a t.test

```
pvals <- replicate(10000, {
  x <- rt(20, df=86)
  t.test(x, mu=1)$p.value
})
power<- mean(pvals < 0.05)
1 - power
```

```
## [1] 0.0149
```

e. Type-II error has ~1.7% probability of occurring in a wilcox.test

```
pvals <- replicate(10000, {
  x <- rt(20, df=86)
  wilcox.test(x, mu=1)$p.value
})
power<- mean(pvals < 0.05)
1 - power
```

```
## [1] 0.0189
```

f. It looks like t.test is more powerful when used on a t-distribution of data.