HW3_Statistics

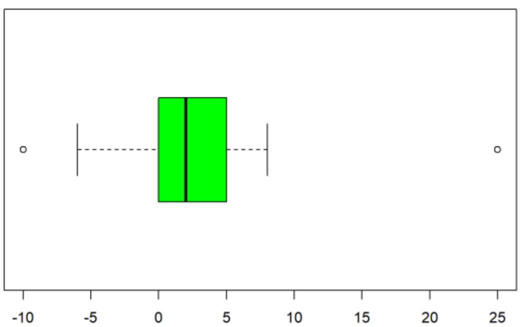
Lilit Hovsepyan

04 10 2019

Problem 1.

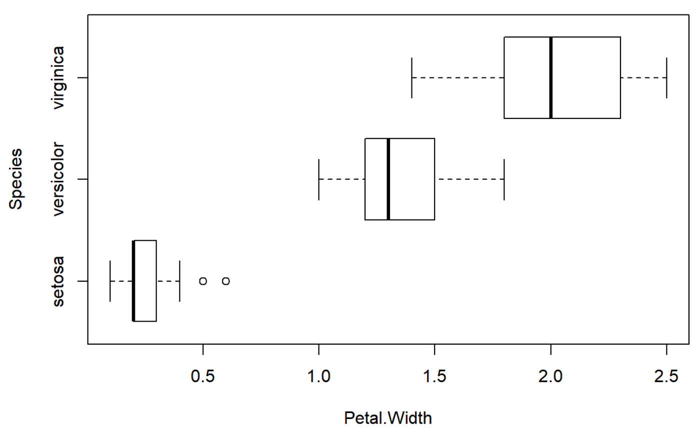
c.

```
a <- c(25,-10,3,1,2,8,4,0,-1,7,7,2,-1,2,-6,5,0)
boxplot(a, horizontal = T, col = "green")
```



d.

```
boxplot(Petal.Width~Species, data=iris, horizontal = T)
```



Problem 2.

b.

```
f<-function(x,y){
  quantile(x,y)
}
f(c(190,2,-3,4,1,2,-3,4,1,2,-3,4),0.15)
## 15%
##  -3
```

Problem 3.

c.

```
x<-rnorm(100,0,1)
y<-quantile(x,seq(0.1,0.9,by=0.1))

alpha<-seq(0.01,0.99,by=0.01)
qqplot(alpha,y)

alphadiv<-1-alpha
par(new=T)
qqplot(alphadiv,y)
```

Problem 4.

b.1

```
x<-rexp(200,3)
y<-rexp(400,0.2)
qqplot(x,y)
abline(0,1, col="blue")
```

b.2

```
qqexp <- function(x){
  qqplot(quantile(x,seq(0.1,0.9,by=0.1)),qexp(seq(0.1,0.9,by=0.1),1))
}
```

```
qqexp(c(5,8,6,9,7,-3,-6,-2,8,0,1,0,1,8,9,-2,-1))



qqunif <- function(x){
  qqplot(quantile(x,seq(0.1,0.9,by=0.1)),qunif(seq(0.1,0.9,by=0.1)))
}
qqunif(c(5,8,6,9,7,-3,-6,-2,8,0,1,0,1,8,9,-2,-1))
```

Problem 5.

c.1

```
head(mtcars,3)
```

c.2,3

```
mtcars.new<-mtcars[,c(1,2,3,4,5,6,7,10,11)]
head(mtcars.new,3)
```

c.4

```
cor.math<-cor(mtcars.new)
```

c.5

cyl-mpg > -0.85 (strong, negative, linear relationship)

disp-mpg > -0.84 (strong, negative, linear relationship)

hp-mpg > -0.77 (strong, negative, linear relationship)

wt-mpg > -0.86 (strong, negative, linear relationship)

disp-cyl > 0.90 (strong, positive, linear relationship)

hp-cyl > 0.83 (strong, positive, linear relationship)

wt-cyl > 0.78 (strong, positive, linear relationship)

hp-cyl > 0.79 (strong, positive, linear relationship)

wt-disp > 0.89 (strong, positive, linear relationship)

hp-carb > 0.75 (strong, positive, linear relationship)

c.6

```
heatmap(cor.math)
```

d.2

```
x<-c(-2,0,4)
y<-c(2,0,100)
cor(x,y,method="spearman")
```

d.4

```
x<-seq(1,50,by=1)
y<-x^4
Pearson<-cor(x,y)
Spearman<-cor(x,y,method="spearman")
Pearson
Spearman
```

d.5

 # ol=10 case

```
x<-seq(1,50,by=1)
ol=10
a<-seq(1,50,1)
a[2]<-ol
y<-a
y
scatterplot(x,y)
plot(x,y,main="scatterplot")
pearson<-cor(x,y)
spearman<-cor(x,y,method="spearman")
c(pearson, spearman
```

 # ol=100 case

```
x<-seq(1,50,by=1)
ol=100
a<-seq(1,50,1)
a[2]<-ol
```

```
y<-a
y
plot(x,y,main="scatterplot")
pearson<-cor(x,y)
spearman<-cor(x,y,method="spearman")
c(pearson, spearman
```

 # ol=1000 case

```
x<-seq(1,50,by=1)
ol=1000
a<-seq(1,50,1)
a[2]<-ol
y<-a
y
plot(x,y,main="scatterplot")
pearson<-cor(x,y)
spearman<-cor(x,y,method="spearman")
c(pearson, spearman
```

Here we see that in case of ol=10 the pearson and spearman correlation coefficients are almost the same (P: 0.9970585, S: 0.9969507), but after changing that outlier by 100 and then by 1000 we see the tendency of decreasing of Pearsons correlation coefficients is faster than Spearman corr coefs. (in ol=100 case: P: 0.6442164, S: 0.8870588, in ol=1000 case: P: -0.1318401 S: 0.8870588). Which means that Pearson correlation is sensitive to outliers (in the 3rd case Spearman corr coef is similar with the 2nd case, which means that it is starting from some point is not sensitive to outliers anymore).

d.6.1

```
install.packages("MASS")
library(MASS)
help(Animals)
```

Here we have the average body and brain weights (correspondingly in kg and g) for 28 species of land animals. Our variables are numeric (ratio scale)

d.6.2

```
head(Animals,3)
tail(Animals,3)
```

d.6.3

```
cor(Animals$body,Animals$brain)
cor(Animals$body,Animals$brain,method="spearman")
```

d.6.4

It is because of the sensitiveness of Pearson correlation to the outliers. An here in our dataset we have outliers.