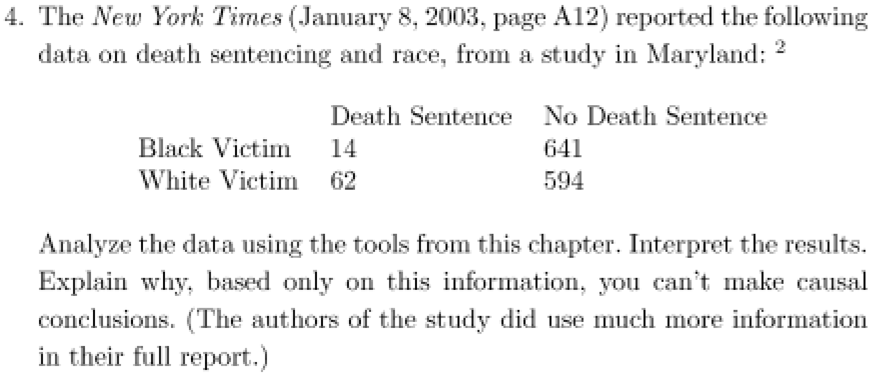
Exercises

Dalia Breskuviene

April 11, 2022

Exercise 15.6.4



4. The *New York Times* (January 8, 2003, page A12) reported the following data on death sentencing and race, from a study in Maryland: [2]

	Death Sentence	No Death Sentence
Black Victim	14	641
White Victim	62	594

Analyze the data using the tools from this chapter. Interpret the results. Explain why, based only on this information, you can't make causal conclusions. (The authors of the study did use much more information in their full report.)

Using the data in the exercise I want to investigate hypothesis if death sentence and victims skin color are related:

H0: Death Sentence and Victims Colour are indeperdent

H1: Death Sentence and Victims Colour are deperdent. Null hypothesis is rejected

To test the hyposesis I will calculate p-value, which is a measure of the evidence against H0. The smaller the p-value, the stronger the evidence against H0.

There are many ways to test hypothesis above. I will try several of them. Firstly, I will do Pearsons X^2 test to check independence between the features:

```
## U statistic: 32.103709343626
```

```
## p-value 1.5e-08
```

```
## [1] "Dependent: Null hypothesis is rejected"
```

Other way is to check independence is the likelihood ratio test(LRT):

```
## LRT statistics: 34.5335058946819
```

```
## p-value: 4.1897680880254e-09
```

```
## [1] "Dependent: Null hypothesis is rejected"
```

Hence both Pearson's statistic and LRT leads to the refusal of unassosiation of the victim's skin color and death sentence.

It can be concluded that Victims' Colour of the skin is associated with Death Sentence (reject H0). However, I would like to know how strong or weak dependency is. To investigate this - odds ratio needs to be calculated.

```
## odds ratio: 0.209249660308993
```

In cases where victim was white skined the desicion of Death Sentense where 5 times as likely.

```
## Odds ratio confidence interval: 0.1145330518385060.382295063622048
```

```
## with a certainty of 95%, death sentence is sentenced
##        from 3 to 9 times more often if the victims skin color was white.
```

This means that, with a certainty of 95%, death sentence is sentenced from 3 to 9 times more often if the victims' skin color was white. We can see that the skin colour of the victim is assotiated to the death sentence, however we can't state that it is a cause of the decision because of the lack of information. There can be other factors contributing to both - maybe the death sentence is given to people killing rich victims. Being white and and rich might be more likely than being black and rich, hence the weight is more on the white victim side.

Exercise 15.6.5:

Analyze the data on the variables Age and Financial Status

Two rows with unknown age or financial status were removed from dataset.

```
##                          35-54 55 and over under 35
## better than a year agor    26          11       34
## same as a year ago         23          37       16
## worse than a year ago      17          22       21
```

I would like to test a hypothesi:

H0: Age and Financial status are independent variables

H1: Age and Financial status are dependent variables

```
## LRT statistics: 22.0637145702841
```

```
## p-value: 0.000194651577572258
```

```
## [1] "Dependent: Null hypothesis is rejected"
```

```
## U statistic: 20.6793143448602
```

```
## p-value 0.000366559
```

```
## [1] "Dependent: Null hypothesis is rejected"
```

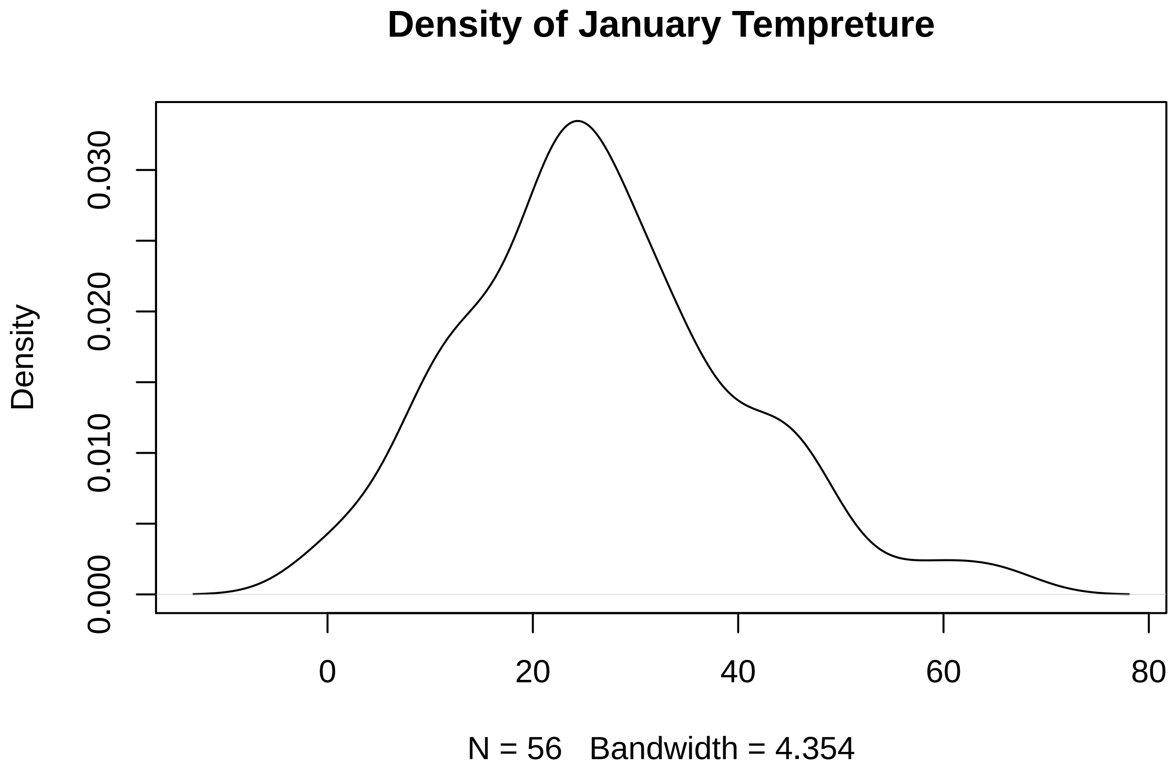
Likelihood ratio test and Pearson X^2 test show that age and financial status are dependent.

Exercise 15.6.6:

Estimate the correlation between temperature and latitude.

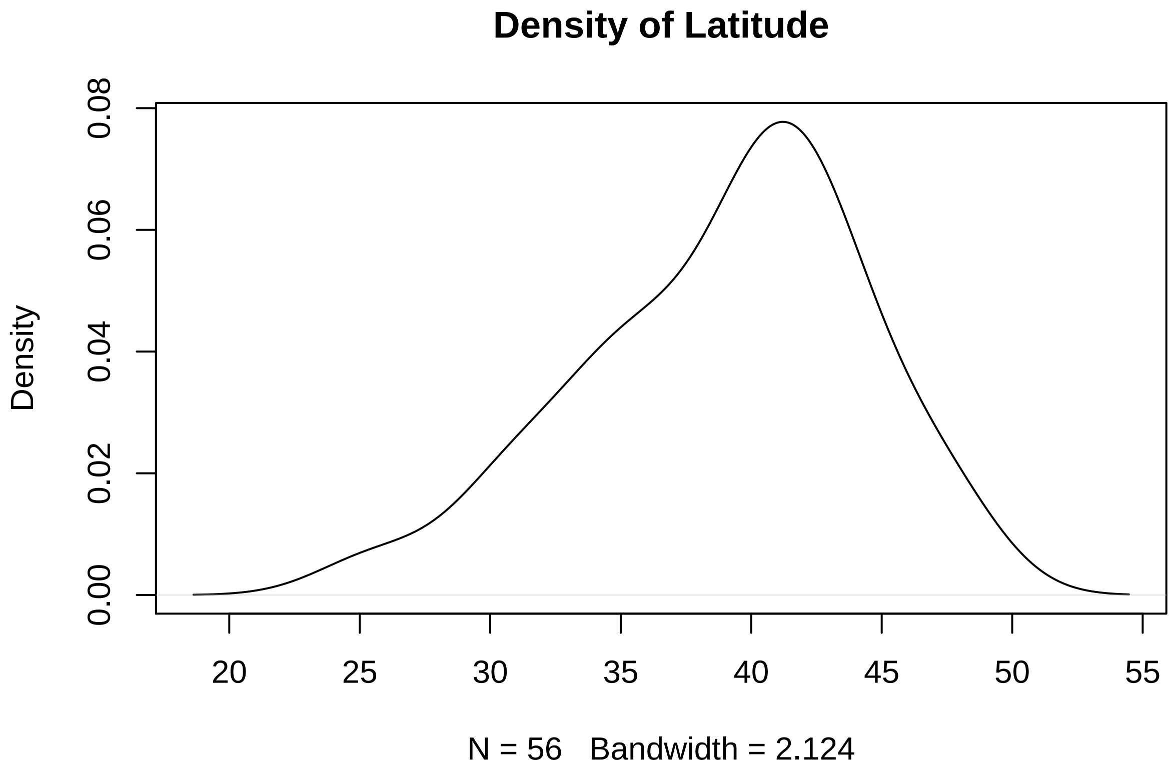
Use the correlation coefficient. Provide estimates, tests, and confidence intervals

In our case both variables - tempreture and latitude - are continuous variables.



```
##
##  Shapiro-Wilk normality test
##
## data:  temp_table$JanTemp
## W = 0.97823, p-value = 0.4041
```

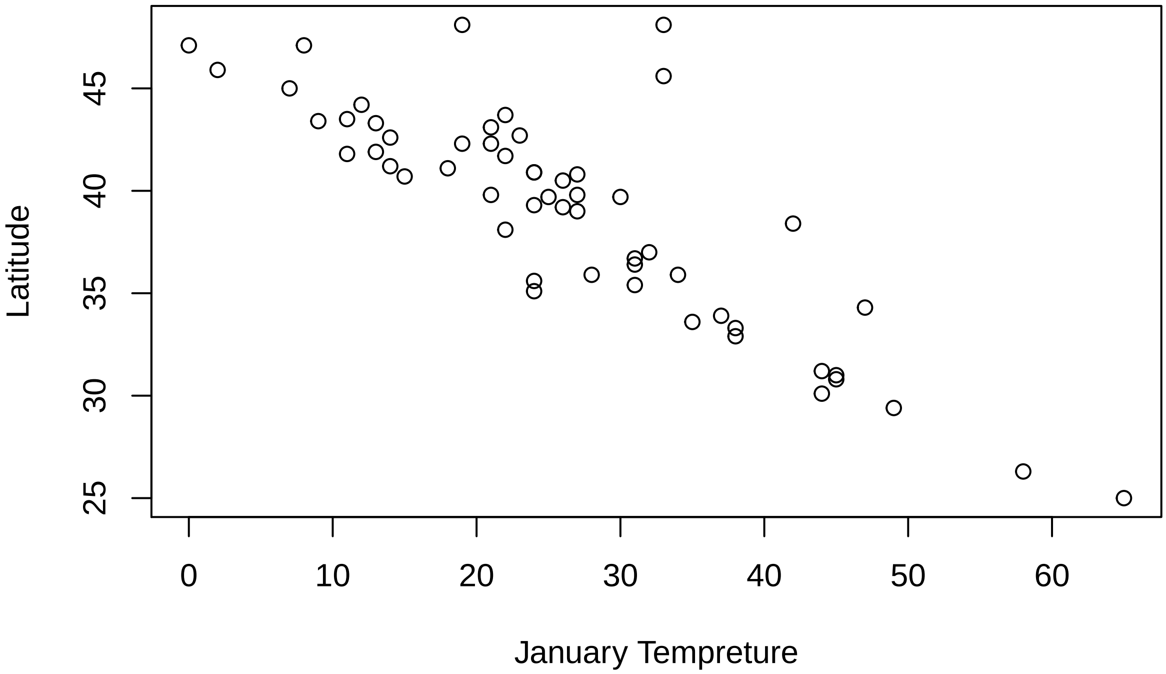
From the output of Shapiro-Wilk normality test on January temperature we see that, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.



```
##
##  Shapiro-Wilk normality test
##
## data:  temp_table$Lat
## W = 0.96823, p-value = 0.1458
```

From the output of Shapiro-Wilk normality test we can see that, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.

In this case we can apply correlation test to check if variables are dependent.



```
##
##  Pearson's product-moment correlation
##
## data:  x and y
## t = -11.759, df = 54, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9084073 -0.7530196
## sample estimates:
##        cor
## -0.8480352
```

From the Pearson's correlation test we can conclude that variables are dependent.

Exercise 15.6.7:

Test whether calcium intake and drop in blood pressure are associated.

I would like to know if calcium intake and drop in blood pressure are associated. As Calcium intake is discrete and blood pressure changes is continuous variable, I need to test hypothesis:

H0: F1 = . . . =Fn

H1: not H0,

where F1, … Fn are Fi(z)=P(Z<=z|Y=i) (Fi is CDF of Z (blood pressure changes) and Y is (Calcium intake)

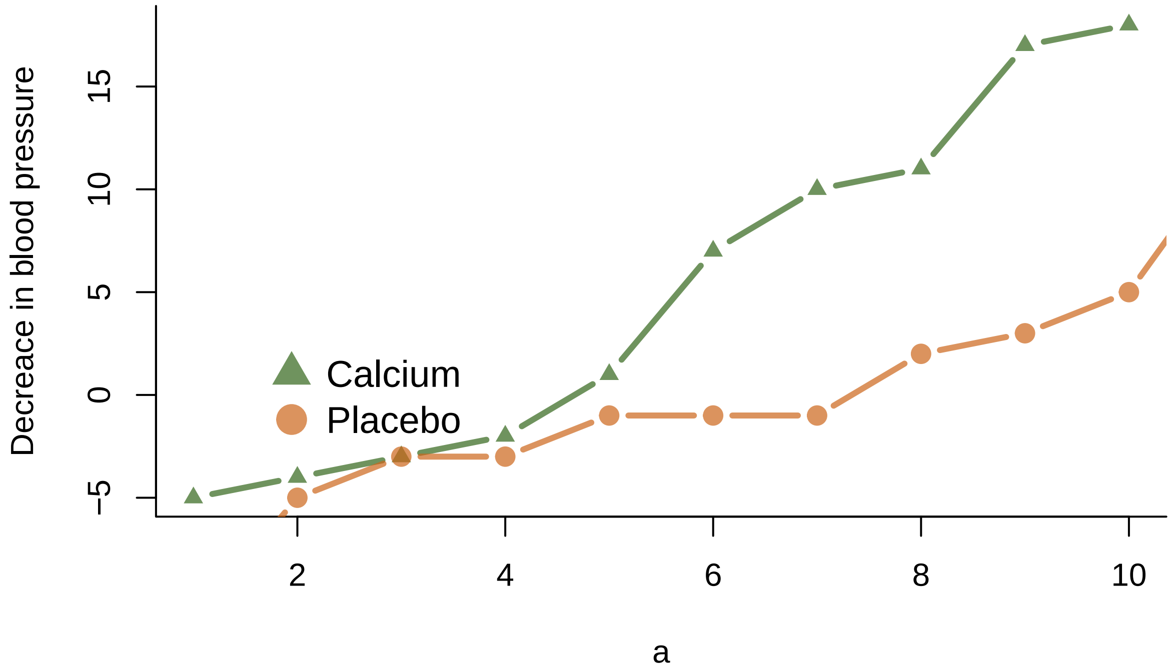
To do that I will use the two sample Kolmogorov-Smirnov test:

```
## Warning in ks.test(x, y): cannot compute exact p-value with ties
```

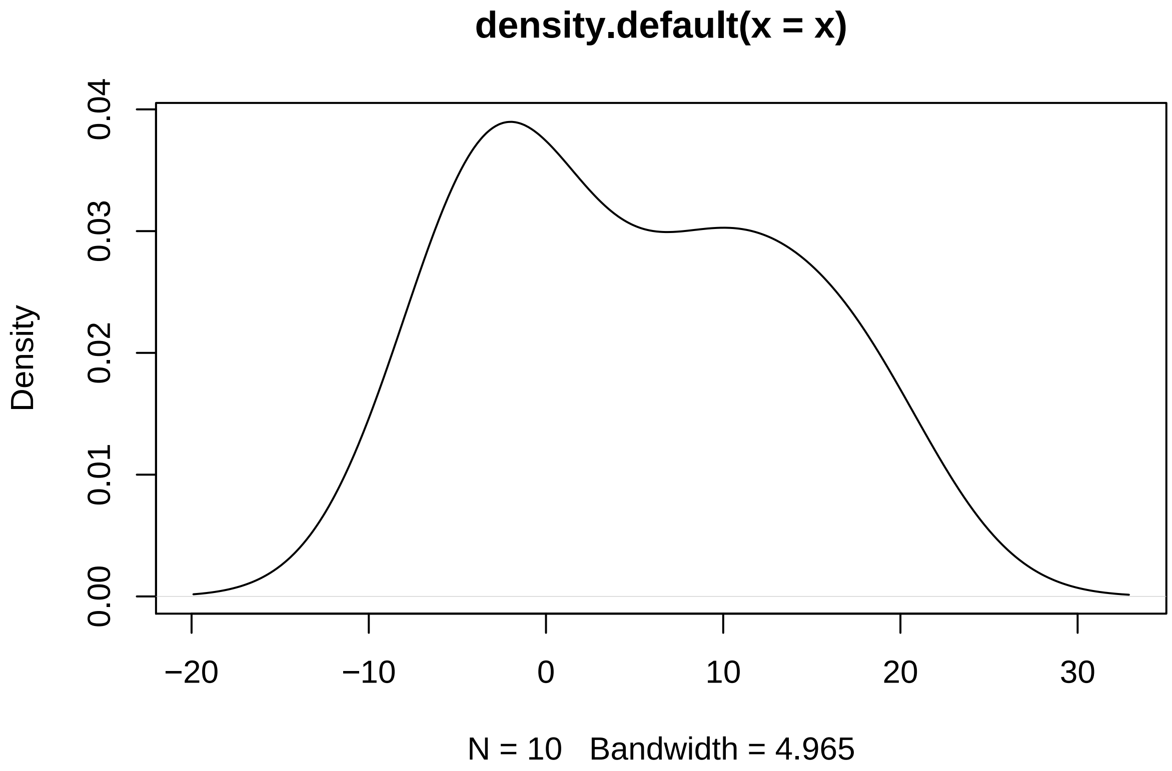
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.40909, p-value = 0.3446
## alternative hypothesis: two-sided
```

Since the p-value is grater than .05, we do not reject the null hypothesis

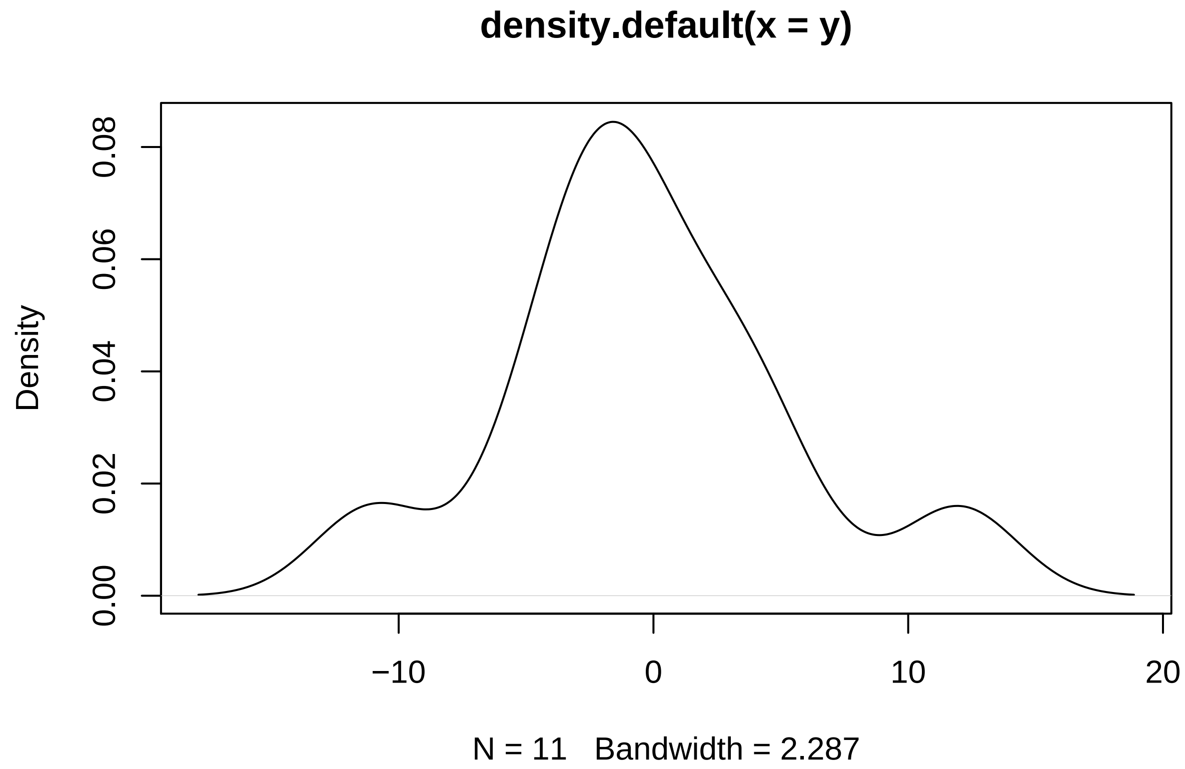
```
a=c(1:length(x))
b=c(1:length(y))

# Make a basic graph
plot( sort(x)~a , type="b" , bty="l" ,  ylab="Decreace in blood pressure" , col=rgb(0.2,0.4,0.1,0.7) , l
lines(sort(y) ~b , col=rgb(0.8,0.4,0.1,0.7) , lwd=3 , pch=19 , type="b" )

legend("bottomleft",
  legend = c("Calcium", "Placebo"),
  col = c(rgb(0.2,0.4,0.1,0.7),
  rgb(0.8,0.4,0.1,0.7)),
  pch = c(17,19),
  bty = "n",
  pt.cex = 2,
  cex = 1.2,
  text.col = "black",
  horiz = F ,
  inset = c(0.1, 0.1))
```



```
plot(density(x))
```



```
plot(density(y))
```


density.default(x = y)
Density
0.00
0.02
0.04
0.06
0.08
−10
0
10
20
N = 11   Bandwidth = 2.287