Practice 12

Elena Tuzhilina

April 4, 2023

Question 1

You are given information about 15 Titanic passengers.

	PassengerId	Sex	Age	Class	Survived
773	773	female	57	2	no
698	698	female	NA	3	yes
652	652	female	18	2	yes
548	548	male	NA	2	yes
890	890	male	26	1	yes
875	875	female	28	2	yes
392	392	male	21	3	yes
788	788	male	8	3	no
330	330	female	16	1	yes
183	183	male	9	3	no
680	680	male	36	1	yes
560	560	female	36	3	yes
104	104	male	33	3	no
136	136	male	23	2	no
37	37	male	NA	3	yes

1. Compute the contingency table for Sex and Class variables.
2. Compute marginal frequencies for Sex and Class variables.
3. Compute joint probabilities for P(Sex = . . . , Class = . . .). Six probabilities in total.
4. Compute joint probabilities for P(Class = . . . |Sex = . . .). Six probabilities in total.
5. Draw a stacked barplot. Do you think there is an association between these two variables?

Question 2

You are given Sex vs Class contingency table for all Titanic passengers, and you want to test these two variables for the independence.

```
tab = table(data$Class, data$Sex)
kable(tab)
```

female	male
94	122
76	108
144	347

1. State H_0 and H_a.

2. Find expected and observed counts for this table.

3. Find the test statistic.

4. Find the p-value.

5. Draw the conclusion at significance level 0.01.

Question 3

The iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris (150 flowers in total). The species are Iris setosa, versicolor, and virginica.

Here are summary statistics for Sepal and Petal lengths.

```
mean(Sepal.Length)
```

```
## [1] 5.843333
```

```
mean(Petal.Length)
```

```
## [1] 3.758
```

```
sd(Sepal.Length)
```

```
## [1] 0.8280661
```

```
sd(Petal.Length)
```

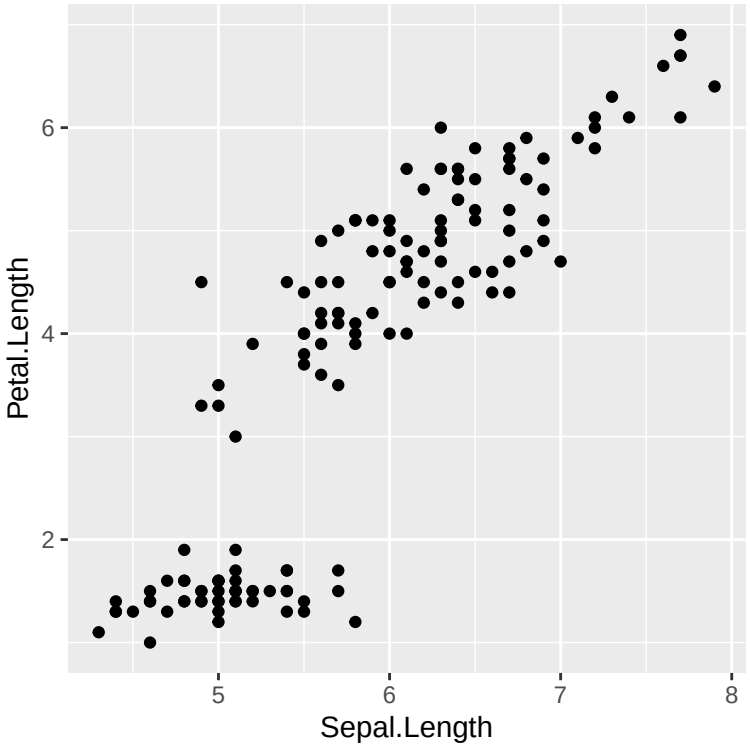
```
## [1] 1.765298
```

```
cor(Sepal.Length, Petal.Length)
```

```
## [1] 0.8717538
```

1. Do you think there is an association between Sepal and Petal lengths?

2. You want to fit the regression line to the following scatterplot plot.



State the regression line equation.

3. Find regression coefficients.

4. What is the interpretation of the regression coefficients?

5. Check if point $Sepal.Length = 6$ and $Petal.Length = 3$ lies on the regression line.

6. Find the residual for point $Sepal.Length = 6$ and $Petal.Length = 3$. Does this point lie below or above the regression line?

7. Check that $Sepal.Length = \bar{x}$ and $Petal.Length = \bar{y}$ point lies on the regression line.

8. Use the regression line to predict the value of Petal length if Sepal length is 6 and 7.

9. Add the regression line to the scatterplot.

10. Find TSS from the provided information.

11. Find R^2 from the provided information. Do you think linear model fits the data well?

12. Find RSS from the provided information.

13. Find ESS from the provided information.