Practice 2

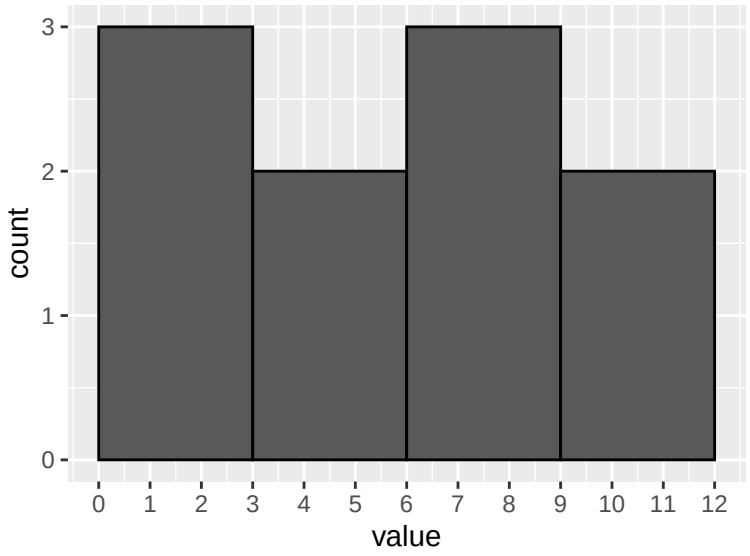
Elena Tuzhilina

January 17, 2023

For the following data `score` represents the test score, `prep_time` represents the preparation time (in hours) and `attend` represents if a person attended the lectures.

score	prep_time	attend
1	0	none
5	5	some
10	10	all
9	14	all
4	3	some
7	5	all
11	14	all
8	8	all
3	6	some
2	5	none

1. Draw a histogram for `score` using cutoffs 0,3,6,9,12.



2. What is the score range?

```
## range = 10
```

3. What are the means and standard deviations of `score` and `prep_time`?

```
## score mean = 6
```

```
## score sd = 3.496029
```

```
## prep_time mean = 7
```

```
## prep_time sd = 4.546061
```

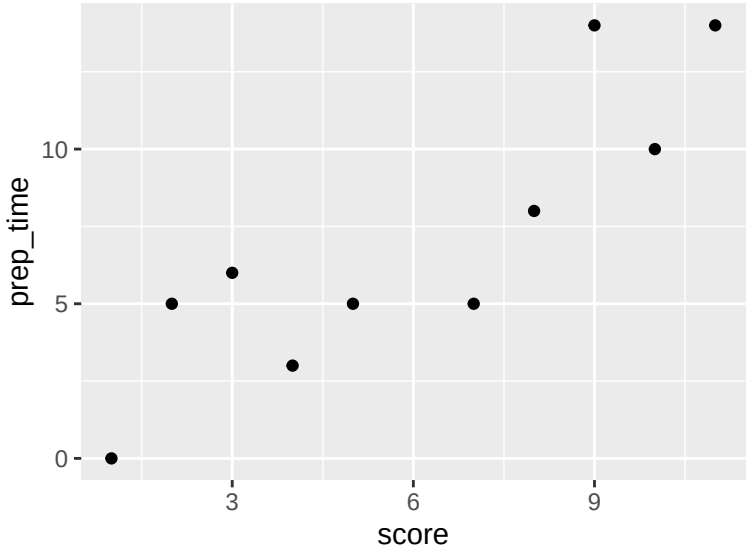
4.You got 13 score on the test. Find z-score.

```
## [1] 2.002271
```

5. If the histogram was symmetric and bell-shaped, would it be a good z-score?

```
## Yes, we are 2 sd above the average (and in top 2.5%)
```

6. Draw `score` vs `prep_time` scatteplot. Do you think there is a relationship?




```
## looks like a positive trend
```

7. What is the correlation between `score` and `prep_time`? Does this confirm your scatterplot findings.

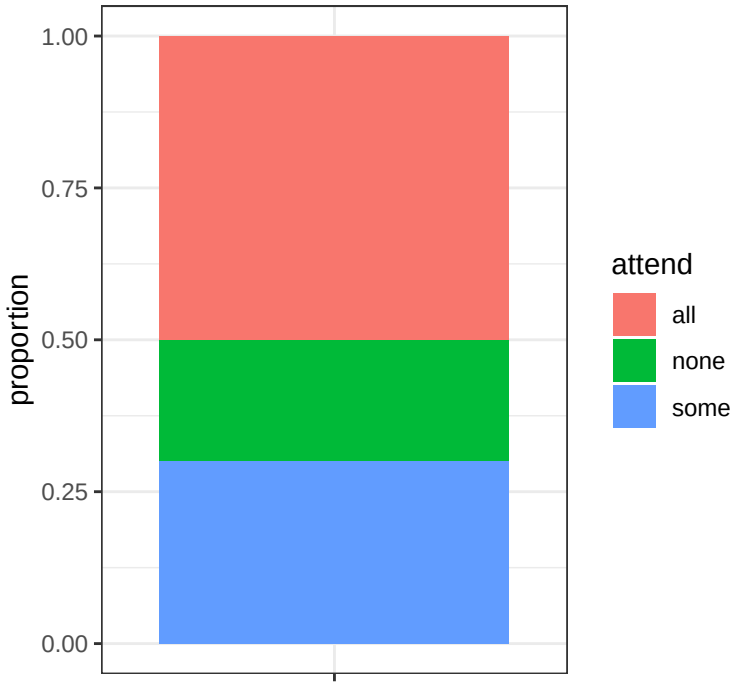
```
## correlation = 0.8668997
```

```
## the correlation is close to 1 and positive => positive dependence
```

8. Compute the distribution table (relative frequencies) for the `attend` variable.

```
##
##  all none some
##  0.5  0.2  0.3
```

9. Draw a stacked diagram for `attend`.



10. Find the average value of `score` and `prep_time` for each category of `attend`. (Try to interpret these values :)

attend	score_mean	score_prep_time
all	9.0	10.200000
none	1.5	2.500000
some	4.0	4.666667

```
## people who attended many lectures have higher test scores
```

```
## people who attended less lectures tend to prepare less
```