# Statistics and probability exam task

Artem Moskovets, SMD-2

Report Bug · Request Feature

## About the project

This project aims to discover correlations and patterns from the survey, and get evaluation for statistics and probability course 😃

## Prerequirements

Packages from requirementns.txt should be installed.

## Built with

- Pandas
- Matplotlib
- Scipy

## Predictions

Three predictions were made in this project:

1. Responders with non-binary gender disagree more with statement, that changing gender is pathological, comparing with male and female responders. (Columns A02 and B11)

2. Responders with non-binary gender more often wondered about their gender-identity, comparing with male and female responders. (Columns A02 and B12)

3. Responders with non-binary gender more often were assigned as wrong gender, comparing with male and female responders. (Columns A02 and B13)

## First view of the data

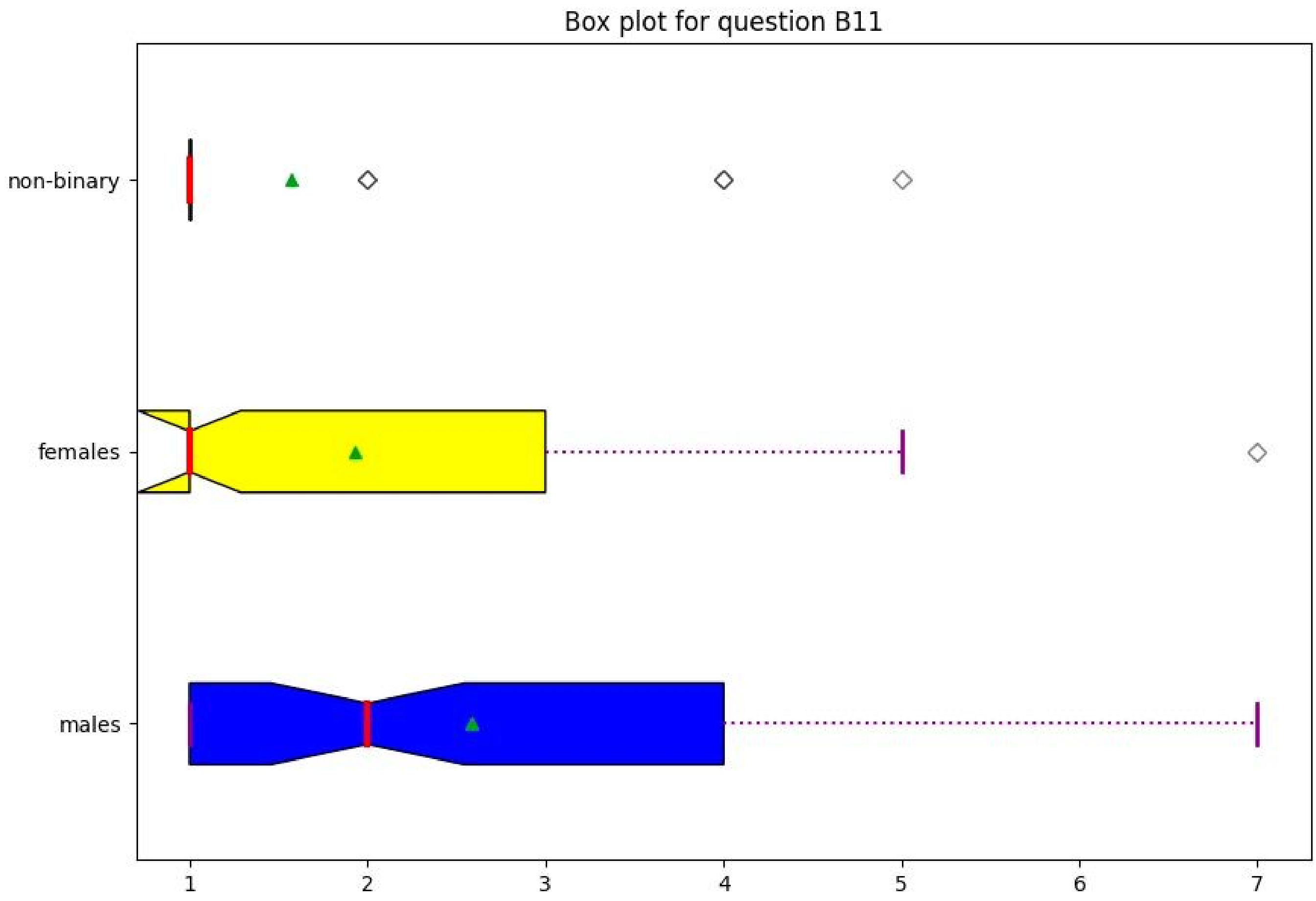
Before visualising, I decided to see most important metrics just with a text:

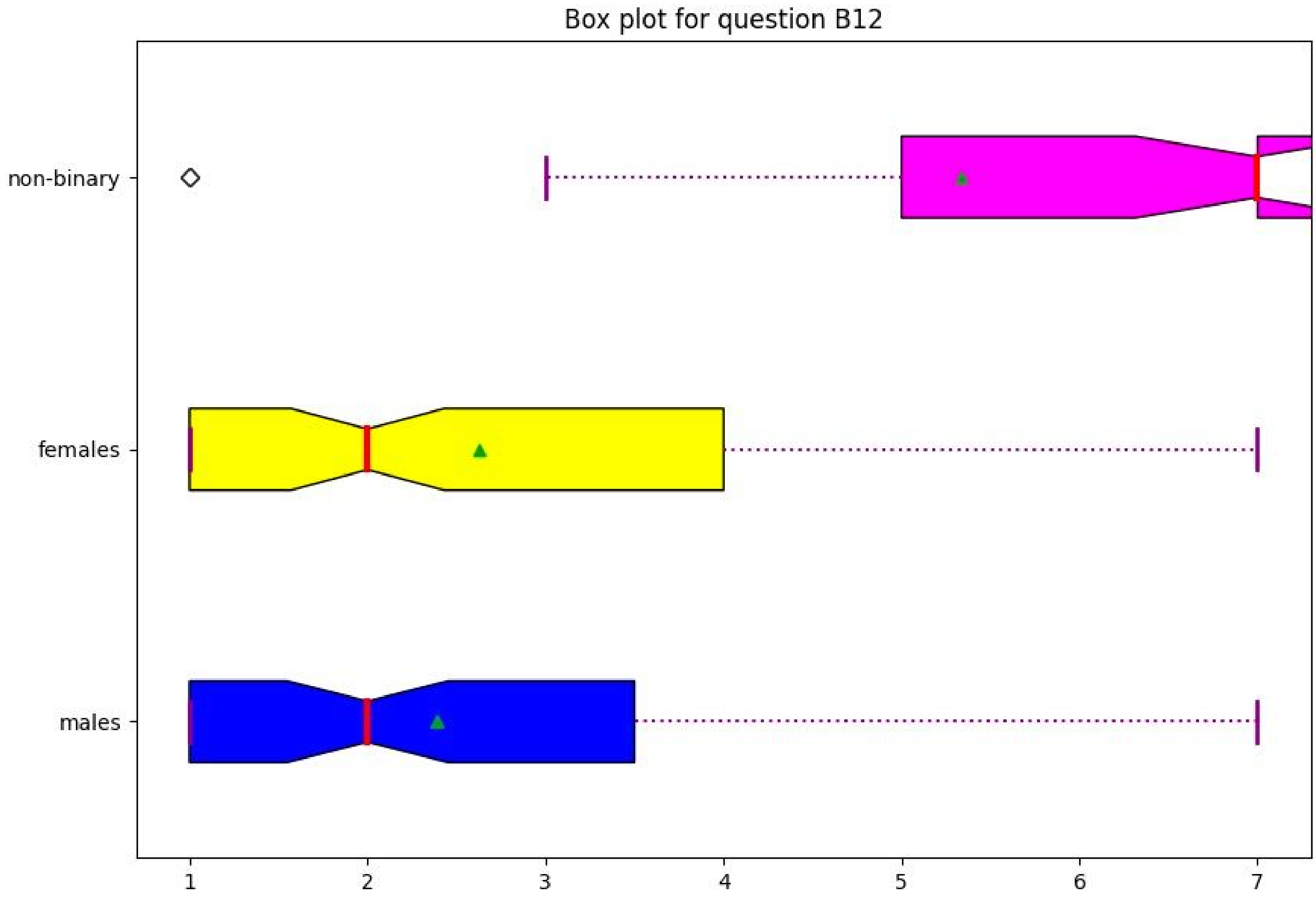
```
amount of males in the survey: 75
amount of females in the survey: 120
amount of non-binary in the survey: 21

mean in the B11 column for males = 2.587
mean in the B11 column for females = 1.925
mean in the B11 column for non-binary = 1.571

mean in the B12 column for males = 2.387
mean in the B12 column for females = 2.625
mean in the B12 column for non-binary = 5.333
mean in the B13 column for males = 1.293
mean in the B13 column for females = 1.433
mean in the B13 column for non-binary = 4.19
```

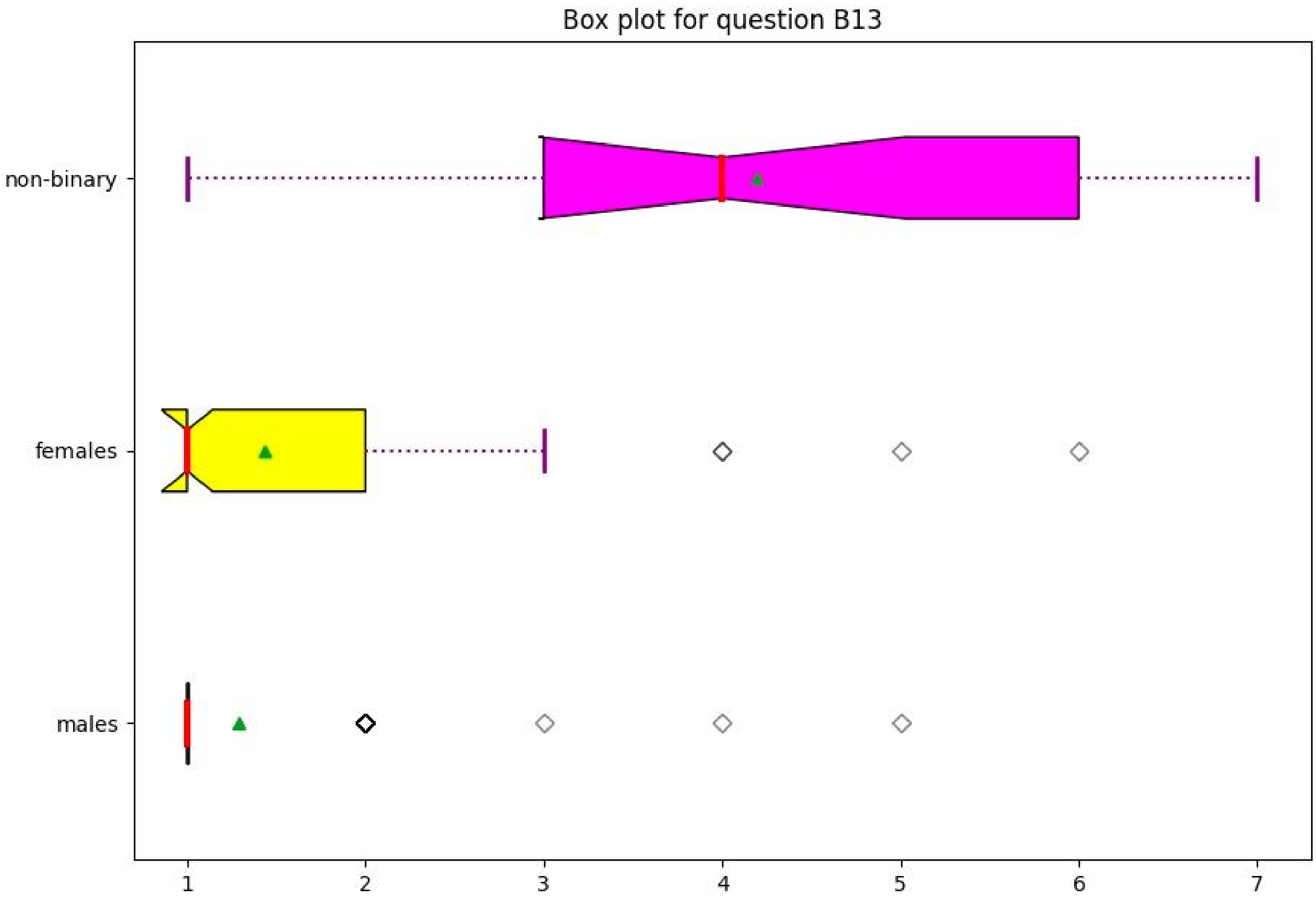
Even here we can see that my assumptions were right

## Visualising the data

To visualise the data, I decided to use boxplots:


Box plot for question B11
non-binary
females
males
1
2
3
4
5
6
7


Box plot for question B12
non-binary
females
males
1
2
3
4
5
6
7



## Prove the predictions

To prove these predistions, I used Kruskal–Wallis H test:

```
B11: KruskalResult(statistic=15.317488299926785,
pvalue=0.00047189967404393635)
B12: KruskalResult(statistic=26.517104443076377,
pvalue=1.745355633681029e-06)
B13: KruskalResult(statistic=51.69910168479064, pvalue=5.938581341979491e-
12)
```

In each case we can see large discrepancy among rank sums, which represents high H-score, and tiny p-value, that corresponds to tiny influence of random in our samples.

## Conclusions

So, all three predictions are true, according to the Kruskal–Wallis H test.

Thank you for attention!