

2023 01 05 VB-STA5 Exam in Statistics

Thursday 5th of January.

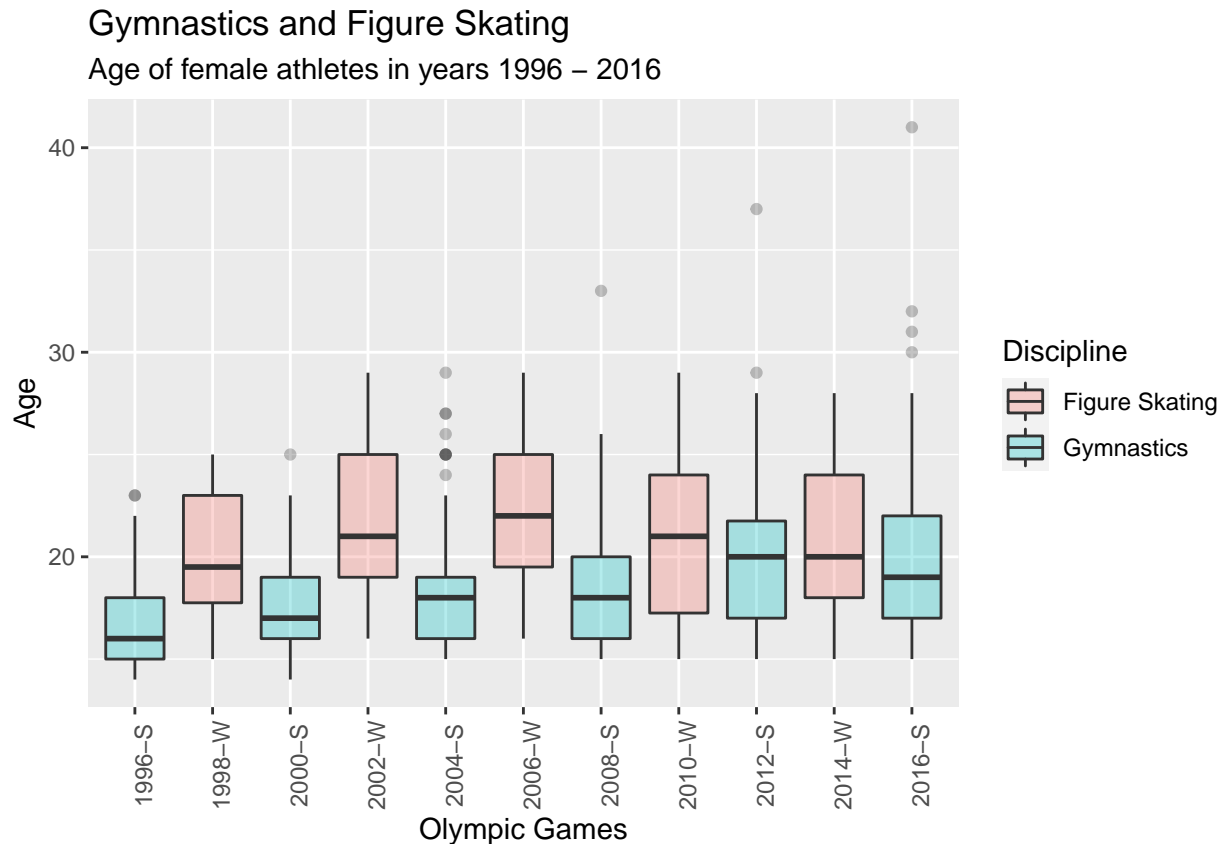
The exam set consists of 3 main exercises with 9 sub exercises in total.

Each sub exercise is weighted equally when grading the hand-ins.

1. Gymnastics and figure skating.

Dataset *data/gym_figskate.csv* contains information about Olympic athletes in Figure Skating and Gymnastics in years 1964 to 2016.

a) Recreate the plot.



b) Describe the plot.

c) Is there significant difference in between average height of male figure skaters competing in 1972 Sapporo Olympics and 2002 in Salt Lake City. Conduct a suitable statistical test.

2. Candles Market

Dataset *candles_revenue.csv* contains information about candle market revenue around the world in Euros. Dataset *population_2020.csv* contains information about world population in 2020.

a) Join the two datasets.

b) Calculate revenue per capita in year 2020. Present in descending order five countries with highest revenue per capita in Europe in format presented below.

The example shows five countries with highest revenue per capita in Asia in ascending order

| Country | Revenue per capita in EUR |
|----------------------|---------------------------|
| Japan | 4.292695 |
| Brunei Darussalam | 4.320168 |
| United Arab Emirates | 4.511445 |
| Singapore | 6.129297 |
| Qatar | 7.920696 |

- c) In 2015 Yankee Candle company conducted a survey of random candle users in selected Nordics (Denmark, Finland, Iceland, Norway, and Sweden). The purpose of the survey was to evaluate customer needs. They survey encompassed 1000 people. Is the survey distribution following the revenue distribution for these countries in 2015?

| Country | n |
|---------|-----|
| Denmark | 256 |
| Finland | 179 |
| Iceland | 98 |
| Norway | 193 |
| Sweden | 274 |

3. Students expenses.

Dataset *UniversityStudentsMonthlyExpenses.csv* contains information about monthly expenses of randomly sampled students in U.S.A. in the 2000s.

- Create a multiple regression model to predict students Monthly Expenses and tune it.
- Is the model valid.
- Predict monthly expenses for student 17721.

| Student | 17721 |
|-----------------------|--------|
| Gender | Female |
| Age | 22 |
| Study Year | 2 |
| Living | Home |
| Scholarship | No |
| Part time job | Yes |
| Transport | Car |
| Smoking | Yes |
| Drinks | Yes |
| Games and Hobbies | No |
| Cosmetics, Self-care | Yes |
| Monthly Subscriptions | Yes |