2023 02 22 VB-STA5 Exam in Statistics

Wednesday 22nd of February.

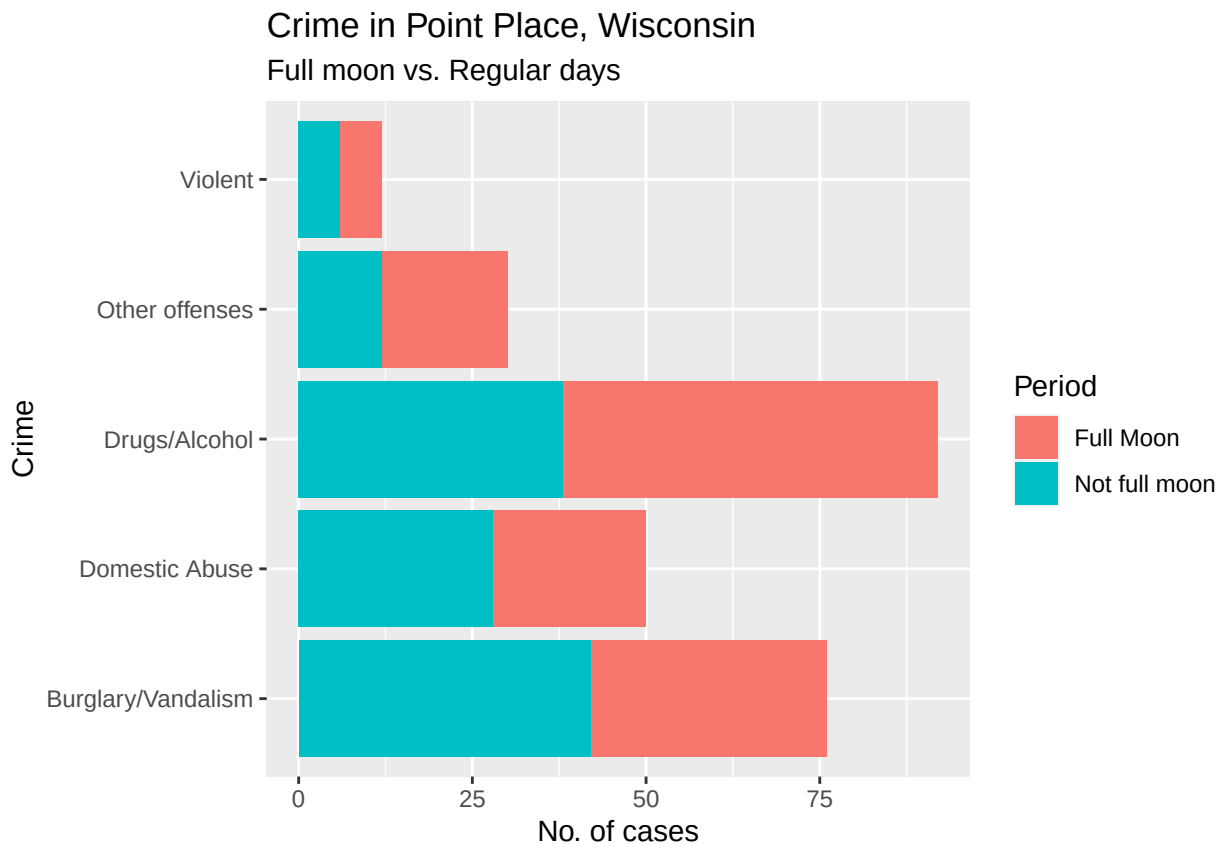
The exam set consists of 3 main exercises with 9 sub exercises in total.

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

1. Full moon influence.

Some people believe that a full moon elicits unusual behavior in people. The dataset *data/full_moon.csv* lists arrests made in a small town during weeks of six full moons and six other randomly selected weeks in the same year.

a) Recreate the plot.



b) Describe the plot, including (but not limited to) comparison of the crime types according to moon phase.

c) Is there statistical relation in between the crimes committed and moon phase? Which test can you use to check it? What are conditions of this test to be valid? Conduct the test and form statistical conclusions.

2. Salmon farming

Dataset *data/salmon.csv* contains information about dangerous chemical compounds found in salmon selected randomly from various salmon farms around the world.

a) Calculate mean amount of highly cancerogenous substances - PCBs (Total PCBs) per country of salmon origin. Present in ascending order in format presented below.

The example shows mean pesticides per country of salmon origin in ascending order

Location	Mean Pesticides Contents
Chile	14.58580
Washington	15.75944
Maine	30.10850
Eastern Canada	33.57167
Western Canada	38.26839
Norway	41.76317
Scotland	46.68493
Faroe Islands	51.53442

b) Is there a statistically significant difference in between total PCBs contents in salmon from Scotland and from Norway? Which test is appropriate to use here? What are its conditions? Conduct the test and form statistical conclusions.

3. Tailor.

A high-end tailor has a detailed database of their customers measurements *data/customer_data.csv*, and a separate one with the shirt size they have been buying *data/customer_sizes.csv*.

a) Join the two datasets.

b) Which tuned multiple regression model would be suitable for predicting new customers shirt size? Create such a multiple regression model.

c) What should be satisfied for the model (3b) to be valid. Check if the model you created in 3b is valid?

d) Predict shirt size of the new customer, Tom:

Tom	measurments
Age	43
Weight	136.25
Height	67.5
Chest	87.6
Waist Size	30.5
Hip	88.6
Thigh	52.0
Knee	34.9
Ankle	22.5
Bicep	27.7
Forearm	27.5
Wrist	18.5