Mathematical practice final exam 2023

2023-07-10

1.

```
tmp <- c(5, 8, 1)
```

Create the vectors

(a) $(5, 8, 1, 5, 8, 1, \ldots, 5, 8, 1)$ where there are 10 occurrences of 5.

(b) $(5, 5, \ldots, 5, 8, 8, \ldots, 8, 1, 1, \ldots, 1)$ where there are 30 occurrences of 5, 20 occurrences of 8 and 10 occurrences of 1.

2.

Execute the following lines which create two vectors of random integers which are chosen with replacement from the integers $0, 1, \ldots, 999$. Both vectors have length 250.

```
xVec <- sample(0:999, 250, replace=T)
yVec <- sample(0:999, 250, replace=T)
```

(a) Create the vector $(y_2 - x_1, \cdots, y_n - x_{n-1})$.

(b) Pick out the values in yVec which are > 600.

(c) What are the index positions in yVec of the values which are > 600?

(d) Sort the numbers in the vector xVec in the order of increasing values in yVec.

(e) Pick out the elements in yVec at index positions $1, 4, 7, 10, 13, \cdots$

3.

```
X <- c(34, 33, 65, 37, 89, NA, 43, NA, 11, NA, 23, NA)
```

Write a piece of R code to count the number of occurrences of `NA` in `X`?

4.

For this problem we'll use the (built-in) dataset state.x77.

```r
data(state)
state.x77 <- as_tibble(state.x77, rownames  = 'State')
```

a. Select all the states having an income less than 4300, and calculate the average income of these states.

b. Sort the data by income and select the state with the highest income.

c. Add a variable to the data frame which categorizes the size of population: <= 4500 is `S`, $> 4500 $ is `L`.

d. Find out the average income and illiteracy of the two groups of states, distinguishing by whether the states are small or large.

5.

a. Write a function to simulate `n` observations of (X_1, X_2) which follow the uniform distribution over the square $[0, 1] \times [0, 1]$.

b. Write a function to calculate the proportion of the observations that the distance between (X_1, X_2) and the nearest edge is less than 0.25, and the proportion of them with the distance to the nearest vertex less than 0.25.

6.

Mortality rates per 100,000 from male suicides for a number of age groups and a number of countries are given in the following data frame.

```r
suicrates <- tibble(Country = c('Canada', 'Israel', 'Japan', 'Austria', 'France', 'Germany',
'Hungary', 'Italy', 'Netherlands', 'Poland', 'Spain', 'Sweden', 'Switzerland', 'UK', 'USA'),
Age25.34 = c(22,  9, 22, 29, 16, 28, 48,  7,  8, 26,  4, 28, 22, 10, 20),
Age35.44 = c(27, 19, 19, 40, 25, 35, 65,  8, 11, 29,  7, 41, 34, 13, 22),
Age45.54 = c(31, 10, 21, 52, 36, 41, 84, 11, 18, 36, 10, 46, 41, 15, 28),
Age55.64 = c(34, 14, 31, 53, 47, 49, 81, 18, 20, 32, 16, 51, 50, 17, 33),
Age65.74 = c(24, 27, 49, 69, 56, 52, 107, 27, 28, 28, 22, 35, 51, 22, 37))
```

a. Transform `suicrates` into *long* form.

b. Construct side-by-side box plots for the data from different age groups, and comment on what the graphic tells us about the data.

7.

In the data set `pressure`, the relevant theory is that associated with the Claudius - Clapeyron equation, by which the logarithm of the vapor pressure is approximately inversely proportional to the absolute temperature (`temperature + 273`). Transform the data in the manner suggested by this theoretical relationship, plot the data, fit a regression line, and add the line to the graph. Does the fit seem adequate?

```r
data(pressure)
```