MATH205, Fall 2022
Midterm (Simulations), Section 050L
Monday, Oct 24, 2:30 pm

Instructions

Your solution needs to be submitted on Canvas (Assignment "Midterm - Simulations"), including:

- An R script of the work
- Screenshots of the results and the figures

Problem 1 (20pts)

In the library MASS, there is a dataset *Melanoma* which contains data on 205 patients in Denmark with malignant melanoma..

(i) Create a histogram to describe the distribution of tumour thickness in mm

(ii) Use *which.max* to find the survival time in days since the operation of the patient with the highest age

(ii) Make a scatter plot to visualize the relationship between two variables the survival time in days since the operation and tumour thickness.

(iii) The function

$$boxplot(y \sim grp)$$

allows us to create a comparative boxplot that describe a (continuous) variable y across different subgroups according to the (discrete) grouping variables grp.

Use this function to produce a comparative boxplot that represent the relationship between the survival time in days since the operation and indicator of ulceration.

The plots should have clear titles and all axes labeled.

Problem 2 (20pts)

Let X be a discrete random variable with the following probability mass function table

x	1	2	3
p(x)	0.32	0.25	0.43

(a) Simulate a dataset of $n = 5000$ random draws from the distribution.

(b) Compute the mean, the median and the standard deviation of the dataset

(b) Produce a bar plot of the dataset.