SMP Re-exam

Laurits Ivar Anesen

August 2020

Note: Please state your answers in the attached notebook ("SMP-exam.ipynb"). If you hand in any exercise by hand, please state so in the notebook under the respective assignment (e.g. state "# on paper" in the code block). Please make sure that you hand in in .pdf format (you can save the notebook as .html and then convert/print this to .pdf). Before handing in, change the name of the file to your name. Please make sure that your line of reasoning is apparent either by means of explanation or in terms of code or in terms of both.

Assignment 1 (20%)

Let X denote a continuous stochastic variable with the following probability density function (pdf):

$$f(x) = \begin{cases} \frac{1}{e^q} e^x & \text{for } x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Where e is Euler's number (2.7182)

a. Find the value q that makes f a valid probability density function.

b. Find the cumulative distribution function.

c. Determine the probability of $P(X < 0)$ and $P(X > 2)$.

d. Determine the expected value and the variance of X.

Assignment 2 (10%)

Event A occurs with a probability of 0.2 given that event B also occurs, and 0.3 if event B does not occur. Event B occurs with the probability of 0.4.

a. What is $P(A)$?

b. What is $P(B|A)$?

c. What is $P(B|\overline{A})$?

Assignment 3 (15%)

a. In a kindergarten a bag contains 10 red trains and 20 blue trains. A kid takes 5 random trains from the bag. What is the probability that there is at least 3 blue trains among them?

b. A website has 5000 events in one minute. It can only manage 100 events a second, otherwise it gets overloaded. What is the probability that the server gets overloaded during any second?

c. Two gamblers decides to play a game against each other 10 times. If player A wins more than 6 out of 10 games, player B has to give him a price. They are equally good at the game. What is the probability that player A wins a price? What is the probability that player A wins the price among the first 8 games?

Assignment 4 (20%)

A study was conducted to compare the lifespan of married men with men who are single. 100 singles and 100 married men were randomly selected. The data is displayed in ”lifespan_civil_status.xlsx”.

a. Determine estimates for the quartiles, average lifespan, standard deviation and variance of both civil status.

b. Setup 95% confidence intervals for the mean of both civil status’, and accompany the intervals with plots that display the rejection region

c. Is it reasonable to assume that the lifespan of both civil status is normally distributed? Explain using plots and discuss skewness and kurtosis.

d. Setup a 99% confidence interval for the mean lifespan difference between the two civil status and accompany the intervals with plots that display the rejection region.

e. Is there *significant* evidence to support the claim that the mean lifespan differ based on your civil status?

f. Is there evidence to support the claim that the standard deviations of the two model civil status differ *significantly*?

Assignment 5 (15%)

The dataset ”cars.csv” contains information about cars from a online auction in North America. 2498 cars are listed together with 12 features for each car. The objective is to determine whether color, brand and state (the location in which the car is being available for purchase) is independent of each other.

a. Create a contingency table, placing color on the vertical axis and brand on the horizontal axis, and test whether color is independent on the brand.

b. Create a contingency table, placing color on the vertical axis and state on the horizontal axis, and test whether color is independent on the state.

c. Create a contingency table, placing brand on the vertical axis and state on the horizontal axis, and test whether brand is independent on the state.

Assignment 6 (10%)

a. Using the dataset from Assignment 5, create 5 equally sized bins from the data in the price column such that you transform price into a categorical variable.

b. Based on this new categorical price variable, does the color significantly influence the price?

Assignment 7 (10%)

An experiment was conducted with 300 paired datapoints. The summary of the data follows:

$$\sum_{i=1}^{n} x_i = 731 \qquad \sum_{i=1}^{n} x_i^2 = 6363$$

$$\sum_{i=1}^{n} y_i = 588 \qquad \sum_{i=1}^{n} y_i^2 = 11545$$

$$\sum_{i=1}^{n} y_i x_i = 8730$$

Fit a simple linear regression model between x and y by finding the estimates of intercept and slope.