Probability and Statistics for Data Science

Data Science Course

Dr. Ariel Mantzura

2025-02-09

Home Exam

1. Download Credit data set.
2. Calculate correlation heatmap for each value of “Region” variable separately.
3. What can you learn from the heat map.
4. Plot a box plot of the “Income” variable or each value of the variable “Cards”.
5. What can you learn from comparing the boxplots.

Home Exam

- Suppose a store has three types of customers:
  - Regular customers (60%)
  - Premium customers (30%)
  - VIP customers (10%)
- The probability that a customer makes a purchase given their category is:
  - Regular: 0.4
  - Premium: 0.7
  - VIP: 0.9
- What is the overall probability that a randomly chosen customer makes a purchase?

Home Exam

- A bank is assessing whether a loan applicant will default on their loan. Based on historical data:
    - 5% of all loan applicants default on their loans.
    - 80% of those who default had a low credit score.
    - 30% of all applicants have a low credit score.
- If a randomly chosen applicant has a low credit score, what is the probability that they will default?

Home exam

- Simulate a 10X1000 matrix with $x_1, x_2, \ldots, x_{10} \sim exp(2)$
- Calculate a biased and unbiased estimator for $Var(X) = \frac{1}{\lambda^2}$
- Average on the results for both methods.
- print the results.
- Write an expression for the difference.

Home exam

- Simulate 1000 samples of size 100 from the $exp(3)$ distribution.
- What is the mean $E(X)$ of $exp(3)$? What is the variance and sd?
- Test the following two sided hypotheses.

$$H_0 : \mu_0 = E[exp(3)]$$

$$H_1 : \mu_0 \neq E[exp(3)]$$

- Use both methods learned in class.
- Calculate the proportion of times that H_0 was rejected.

Home exam - extra

- Write the pmf of $X_1, \ldots, X_n$ given Y. Write the pmf if given Y, $X_1, \ldots, X_n$ are independent.
- Write the density of two independent normal variables.
- Assume that X_1, X_2 have a multivariate Normal distribution:

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

  and that

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Prove that X_1 and X_2 are independent?