Problem 1

A store has a featured new seasonal product. Let N be the random variable which designates the number of customers who come to the store during the season, where N ~ Poi (A). It is estimated that the probability of a customer buying this new product is B and this independently from one customer to another.

LETTRE	DÉFINITION	VALEUR
A	$11 + Z_5 + Z_7$	24
B	$\left(\sum_{i=1}^{7} Z_i\right)/60$	8/15
C	$12 + Z_2 + Z_4$	25
D	$\max_i \{Z_i\} - \min_i \{Z_i\}$	9

a) It is assumed here that the store has unlimited stock of this product. Let X and Y be the random variables such that X = the number of customers who buy the product; Y = the number of customers who do not buy the product. Are the variables X and Y independent? To justify.

b) The store has a profit of C $ for each unit sold. Each unsold unit is expected to be put into storage for the next year at a cost of D $. Determine the value of the number of stored units n that the store should have to maximize its average profit.

Problem 2

This problem is a case study which consists of an analysis of data taken from Motor Trend magazine for 1975. It was collected as a result of a study of the fuel efficiency of various models and makes of vehicles, particularly their fuel consumption.
The data. The data to be analyzed consists of a sample of 160 observations on five variables measuring a number of characteristics of the vehicles in the study. Table 1 below shows the different variables of the study (column number in the file, symbol, name, and description). The symbols and column numbers are as they appear in your custom data set.

Col. n°	Symbole (Nom dans le fichier)	Description
1	Y (mpg)	La consommation du véhicule (en milles par gallon)
2	X_1 (horsepower)	La puissance du moteur du véhicule (en livres par pied)
3	X_2 (weight)	Le poids du véhicule (en livres)
4	X_3 (origin)	Le code du pays d'origine du véhicule (**1** : États-Unis et **0** : Autres pays)

Tableau 1 : Les variables de l'analyse.

The goal is to analyze this data in order to determine the possible links between different variables, and to determine a statistical model allowing to describe and predict the consumption of a vehicle from the most relevant characteristic.

a) We want to check whether the consumption of a vehicle is significantly different depending on the country of origin. For this we can consider the consumption variable (mpg) divided into two groups according to the code of the country of origin (origin) and make a comparison of the two groups in terms of mean, symmetry and variability. To do this, perform the following analyzes and give a brief conclusion:
    - two juxtaposed histograms, and two juxtaposed Tukey (or "Box Plot") diagrams
    - a table of descriptive statistics by group: mean, quartiles, standard deviation, confidence interval for the average
    - a hypothesis test on the equality of variances for the two groups;
    - a hypothesis test on the equality of means for the two groups.

Phase 2

In this phase, we are interested in determining a model to explain consumption according to the various factors considered. To do this, we consider simple regression models by considering consumption (mpg) as the dependent variable Y.

b) We consider the following six models:

Modèle 1 : $Y = \beta_0 + \beta_1 X_1 + \varepsilon;$ **Modèle 2 :** $Y = \beta_0 X_1^{\beta_1} e^{\varepsilon};$ **Modèle 3 :** $Y = \beta_0 e^{\beta_1 X_1 + \varepsilon};$

Modèle 4 : $Y = \beta_0 + \beta_1 X_2 + \varepsilon;$ **Modèle 5 :** $Y = \beta_0 X_2^{\beta_1} e^{\varepsilon};$ **Modèle 6 :** $Y = \beta_0 e^{\beta_1 X_2 + \varepsilon},$

$$\beta_1 \text{ and } \beta_2 \text{ are parameters and } \varepsilon \text{ is a random error}$$

Notice: The coefficients β0 and β1 as well as the error ε are not the same from one model to another.

For each of the six models above:

- Make the adjustment (i.e. get the table of regression coefficients, the analysis of variance table).
- Test the significance of the model and perform a residual analysis (normality, homoscedasticity, outliers, etc.)
- In conclusion: make a comparison and say which of the six models is better than the others. Justify your choice by specifying the criteria used.

c) Based on the best model you obtained in b), calculate a forecast interval for the consumption (mpg) of a vehicle with the following characteristics: power (X1 = 120), weight (X3 = 2200). Briefly comment on the result.