Week 0 Activity 4

Checking which distribution fits the data



**Kabir Maniar**

# INTRODUCTION

In this activity we are going to find out which **distribution function** fits our **data**. We are going to use the following dataset. [Click here to open.](https://docs.google.com/spreadsheets/d/1a_8TIZIptTHlkISMhAuY_P_ZeLjeZS9HiUYXoYV31lE/edit?usp=sharing) This dataset contains information about cars released from 1970 to 1982. It contains data about the cars weight, horsepower, acceleration, mpg and country of origin.

We are trying to find the distribution of the Milage(MPG) of the car.

# PROCEDURE

1. Collect the data and plot it
2. Check which distribution does the data fit
3. Prove that the data follows the following distribution.

# DATA

[Week 0 Activity 4](https://docs.google.com/spreadsheets/u/1/d/1fUVs1ahoW2yqx2eXZjAyWNfSqhvoJfhShD99Ck0dFmI/edit)

# RESULTS

The range of our data is from 9 to 47.

So, we must find out the probability of cars manufactured in this time period (1970-1982)  
Let **X** be the discrete random variable whose value is the number of cars with that MPG.

|  |  |
| --- | --- |
| X | Calculated Probability |
| 9 | 0.002512562814 |
| 10 | 0.005025125628 |
| 11 | 0.01005025126 |
| 12 | 0.01507537688 |
| 13 | 0.05025125628 |
| 14 | 0.05025125628 |
| 15 | 0.0527638191 |
| 16 | 0.04522613065 |
| 17 | 0.03768844221 |
| 18 | 0.06030150754 |
| 19 | 0.05025125628 |
| 20 | 0.05025125628 |
| 21 | 0.03266331658 |
| 22 | 0.03266331658 |
| 23 | 0.04271356784 |
| 24 | 0.03768844221 |
| 25 | 0.04271356784 |
| 26 | 0.04773869347 |
| 27 | 0.03768844221 |
| 28 | 0.03266331658 |
| 29 | 0.03266331658 |
| 30 | 0.02763819095 |
| 31 | 0.03266331658 |
| 32 | 0.0351758794 |
| 33 | 0.02010050251 |
| 34 | 0.02512562814 |
| 35 | 0.01005025126 |
| 36 | 0.02261306533 |
| 37 | 0.01507537688 |
| 38 | 0.01256281407 |
| 39 | 0.007537688442 |
| 40 | 0.005025125628 |
| 41 | 0.002512562814 |
| 42 | 0 |
| 43 | 0.005025125628 |
| 44 | 0.007537688442 |
| 45 | 0 |
| 46 | 0.002512562814 |

So, we must find out the probability of cars manufactured in this time period (1970-1982)  
Let **X** be the discrete random variable whose value is the number of cars with that MPG.  
If we plot this we get this sort of graph.

A graph with a line

Description automatically generated

This is the histogram of the MPG Values:

A graph of a number of people

Description automatically generated

As you can see it looks like a Beta Distribution. We can try to find the Alpha and Beta values with trial and error, which comes out to be α=2.1 and β=4.1  
The PMF of the **Beta Distribution** with these values looks like this:

A graph with a blue line

Description automatically generated  
As you can see our calculations follow this trend and we can get an approximate probability with this function.