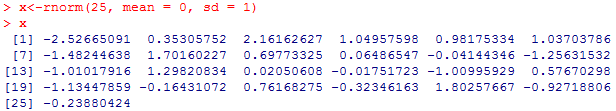
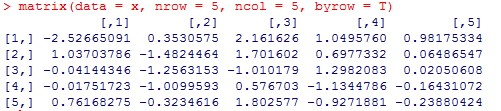
1. **Data generation and matrix indexing:**
2. **Generate a vector with 25 elements and each element independently follows a normal distribution (with mean =0 and sd=1);**

Below screen shot shows the code to generate a random normal distribution with mean = 0 and sd = 1

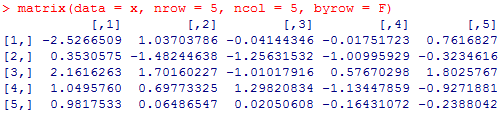


1. **Reshape this vector into a 5 by 5 matrix in two ways (arranged by row and column);**

By row:

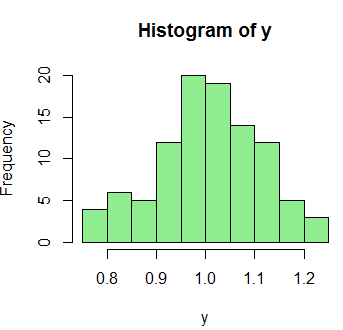


By column:



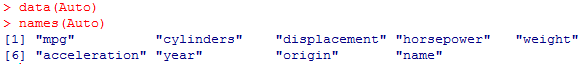
1. **Similarly, generate another vector with 100 elements and plot its histogram**





1. The above plot is a histogram of the frequencies of the vectors generated using *rnorm*. It can be observed that the histogram is bell shaped with mean as 1 and sd as 0.1.
2. **Upload the Auto data set, which is in the ISLR library. Understand information about this data set by either ways we introduced in class (like “?Auto” and names(Auto))**

After installing the package ISLR, we load the ISLR library which contains the *Auto* dataset. Below screenshot shows the variables in the dataset.



1. **Make a scatterplot between two of the following variables (try to plot all scatterplots in one figure; hint: use pairs() command): “mpg”, “displacement”, “horsepower”, “weight”, “acceleration”. By observing the plots, do you think the two variables in each scatterplot are *correlated*? If so, how?**

The code to create the matrix of scatterplot is:



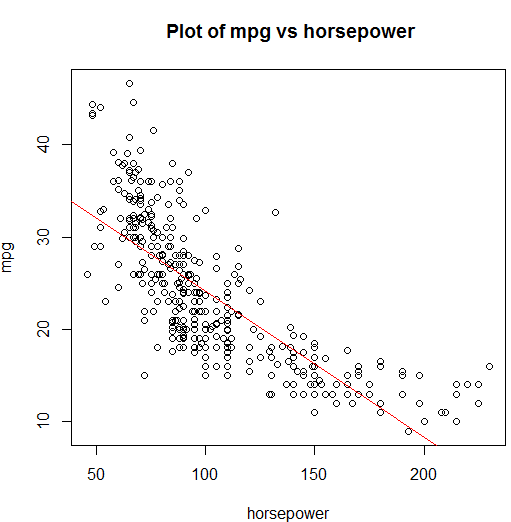
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | From the matrix plot, we can see correlation between the various variables. We can see correlation in the below table.   |  |  | | --- | --- | | **Correlation between**  **variables** | **Type of correlation** | | *mpg vs displacement* | Negative | | *horsepower vs mpg* | Negative | | *mpg vs weight* | Negative | | *horsepower vs weight* | Positive | | *horsepower vs acceleration* | Negative | | *displacement vs weight* | Positive | | *displacement vs horsepower* | Positive | |

There are two types of correlation, positive and negative correlation. A positive correlation occurs when increase in one variable increases the other variable increases. A negative correlation occurs when increase in one variable increases the other variable decreases.

1. **Draw a line on the scatterplot of mpg vs. horsepower to represent relationship between the two variables.**

Below is the code to create a scatterplot and a line to fit the points between mpg and horsepower:





The red line is the linear fit for the points between mpg and horsepower. This line fits the points in accordance to the linear regression.

1. **Is there a better way to represent their relationship rather than the linear model you just drew?**

The blue line is a better way to represent the linear model. The blue line corresponds to a polynomial regression between the two variables.

