## Andrew Nalundasan

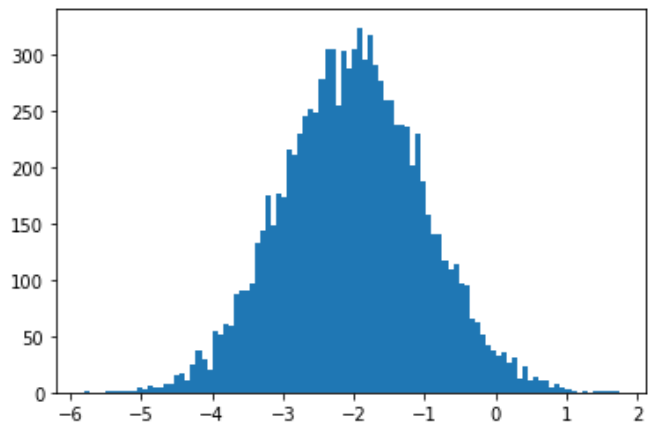
## OMSBA 5067 – ML for Business

## April 3, 2021

## Lab 1

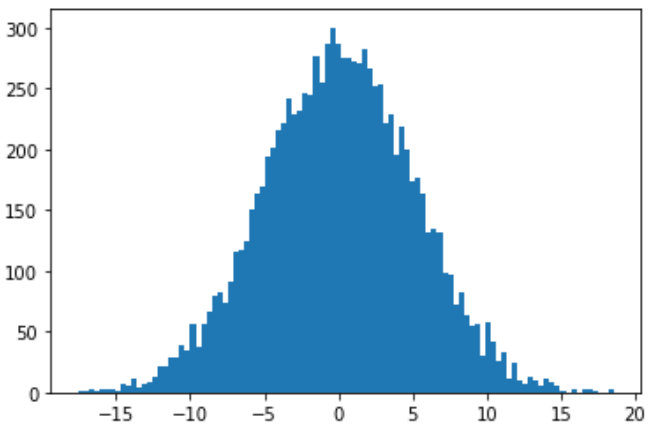
**Step 5a:**

mu = -2 and sigma = 1, obs = 10000



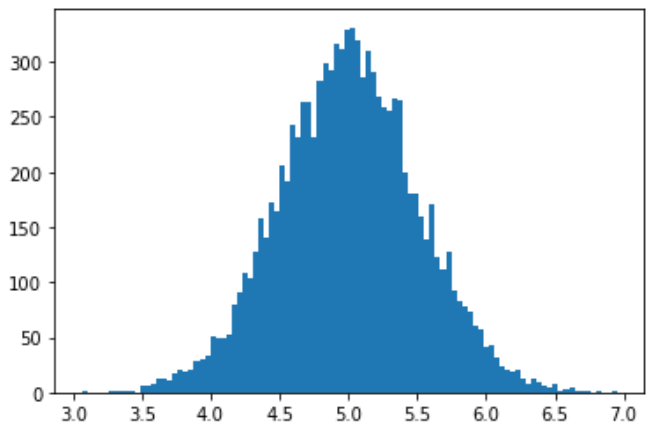
**Step 5b:**

mu = 0 and sigma = 5, obs = 10000

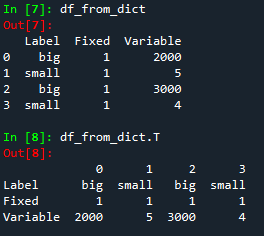


**Step 5c:**

mu = 5 and sigma = 0.5, obs = 10000

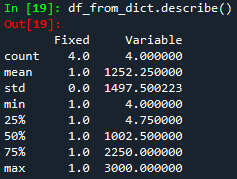


**Step 7a:**



I observed that ‘df\_from\_dict.T’ transposed the rows and columns in the data frame.

**Step 7b:**



I observed that ‘df\_from\_dict.describe()’ provided summary statistics of the data frame.

**Step 9:**

# Generates three sets of numpy normal random numbers with means of -2, 0,

# and +2, and with standard deviations equal to 1. Each set should contain

# 2000 random numbers.

R\_neg2 = np.random.normal(-2, 5, 2000)

R\_zero = np.random.normal(0, 5, 2000)

R\_pos2 = np.random.normal(2, 5, 2000)

# Puts these arrays in a dictionary with these keys: "R\_neg2", "R\_zero", and "R\_pos2".

step\_9\_b = dict()

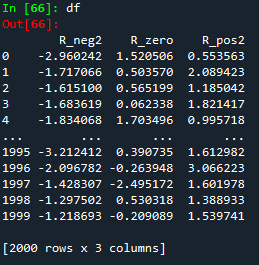
step\_9\_b["R\_neg2"] = np.array(R\_neg2)

step\_9\_b["R\_zero"] = np.array(R\_zero)

step\_9\_b["R\_pos2"] = np.array(R\_pos2)

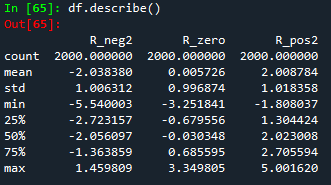
# Converts the dictionary to a DataFrame, name it as "df"

df = pd.DataFrame(step\_9\_b)



# Calculates the statistics of each column (using describe())

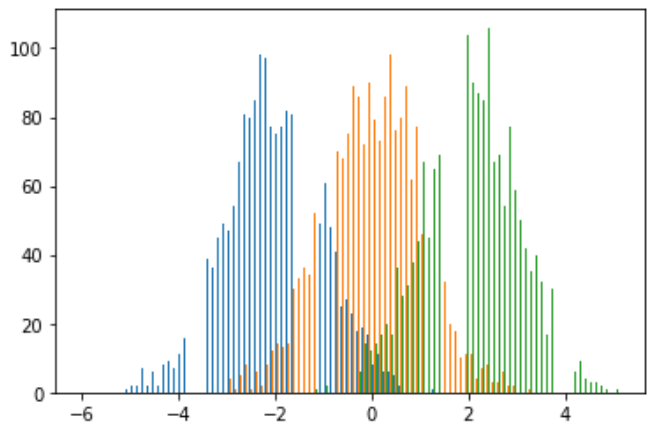
df.describe()



# Plots the histogram of the three sets of numbers on a same graph.

plt.hist([R\_neg2, R\_zero, R\_pos2], 100)

plt.show



all\_low = np.where((df["R\_neg2"]<=-2) & (df["R\_zero"]<=-2) & (df["R\_pos2"]<=-2))

all\_high = np.where((df["R\_neg2"]>=2) & (df["R\_zero"]>=2) & (df["R\_pos2"]>=2))

print("All Low Count: ",len(all\_low[0])," All High Count: ",len(all\_high[0]))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **std. dev = 1** | **std. dev = 2** | **std. dev = 3** | **std. dev = 4** | **std. dev = 5** |
| **all\_low** | 0 | 1 | 18 | 48 | 81 |
| **all\_high** | 0 | 4 | 25 | 43 | 75 |

As we increase the number of standard deviations, we see that both numbers for all\_low and all\_high increase along with the standard deviations. This occurs because when the number of standard deviations increase, the distribution widens, increasing the chances of including values lower than -2 and higher that 2. We see that with 5 standard deviations, this yields the highest amount of all\_low and all\_high. The counts of these numbers will be different each time we run the code since we are working with a random number generator.