**Topics: Descriptive Statistics and Probability**

1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data = {"Company\_name":['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers','Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways','Warner-Lambert'],"Measure\_x":[0.2423,0.2553,0.2541,0.2414,0.2962,0.2825,0.2581,0.2439,0.4026,0.3295,0.9136,0.2599,0.3942,0.2671,0.35]}

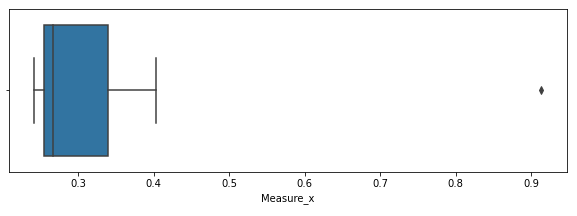
data = pd.DataFrame(data)

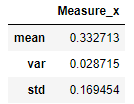
fig = plt.figure(figsize=(10,3))

sns.boxplot(x="Measure\_x",data=data);

data[["Measure\_x"]].agg(["mean","var","std"])

Boxplot







Answer the following three questions based on the box-plot above.

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

IQR = 12 – 5 = 7

This value implies that the middle 50% of values in the dataset have a spread of 7

1. What can we say about the skewness of this dataset?

X variable has a positive skewness

1. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

In this case, there is no outlier and no skewness.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie? Between range 4 to 8
2. Comment on the skewness of the dataset. There is value 25 which is an outlier and because of that data is positively skewed.
3. Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset. Using boxplot, we won’t be able to get any idea about the mode whereas histogram won’t tell us about the median of the data.
4. AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that “could happen.” Suppose that one in 200 long-distance telephone calls is misdirected. What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5, p = 1/200, q = 199/200, k = 0

at least one in five attempted telephone calls reaches the wrong number = 1 - none of the call reaches the wrong number

= 1 - P(0)

= 1 - ⁵C₀ (1/200)⁰ (199/200)⁵⁻⁰

= 1 - (199/200)⁵

= 0.0248

from scipy.stats import binom

result = 1 - binom.pmf(k=0, n=5, p=1/200)

print("Binomial Probability: ",np.round(result,4))

1. Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

|  |  |
| --- | --- |
| x | P(x) |
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

1. What is the most likely monetary outcome of the business venture? $2000 as it has maximum probability as 0.3
2. Is the venture likely to be successful? Explain

venture is likely to be successful as Expected value is positive which is $800

1. What is the long-term average earning of business ventures of this kind? Explain

Expected value = ∑ X P(X) = 800

long-term average earning of business ventures = $800

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Var (X) = E(X²) – [E(X)]² = 2800000 - 800² = 2160000 ( Quite High)

SD = √Var = approx $1470

As Variability is Quite high hence Risk is high