**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% |

**Ans:**

data={'Name of Company' : ['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers','Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','SunMicrosystems','Travelers','US Airways','Warner-Lambert'],

‘Measure X’ : [24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,

35.00] }

df=pd.DataFrame(data)

df

sns.boxplot(data['Measure X'])

🡪 Morgan Stanley is a outlier company present in the given data.

mu=df['Measure X'].mean()

print('mu = ',mu )

sig2=df['Measure X'].var()

print('sig2 = ',sig2 )

sig=df['Measure X'].std()

print('sig = ',sig )

mu = 33.27133333333333

sig2 = 287.1466123809524

sig = 16.945400921222028



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.

**Ans:**  Here Q1 = 5 ,Q2 = 7 ,Q3 = 12

IQR = Q3 – Q1 = 5 – 12 = 7

IQR range covered by the central 50% of the data (between Q1 and Q3) is 7 units.

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

**Ans:** There is presence of outlier towards the right hence it is positively skewed.

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?

**Ans**: Then there would not have skewness.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans :** Bars showing highest frequency is the mode of the dataset.

1. Comment on the skewness of the dataset.

**Ans :** Data is positively skewed

1. 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.

**Ans:** From both the graphs we get to see that data is positively skewed.

1. 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.)

**Ans:** Let P(at least one in five calls is misdirected) be denoted as P(at least one). Then, the complement probability is the probability that none of the five attempts is misdirected.

Given that one in 200 calls is misdirected, the probability of a single call being misdirected is 1/200​. Therefore, the probability that a single call is not misdirected is 1−1/200​.

Since the calls are assumed to be independent, the probability that none of the five attempts is misdirected is (1−1/200)^5.

Now, the complement probability is simply P(none), and the probability that at least one in five attempts is misdirected is 1−P(none).

P(at least one)=1−(1−1/200)^5 =1−(199/200)^5 = 0.0248

Therefore, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately 0.0248 or 2.48%.

Top of Form

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?

**Ans:** $2000 have the highest probability , so most likely monetary outcome of the business

ventureis $2000 .

1. Is the venture likely to be successful? Explain

**Ans:** The success of the venture is subjective and depends on the definition of success. If we consider success as having a positive monetary outcome, then we can analyze the probability distribution**.**

P(Success) = P(0) + P(1000) + P(2000) + P(3000)

P(Success) = 0.2 + 0.2 + 0.3 + 0.1 = 0.8

The cumulative probability of success is 0.8, which means that there is an 80% chance of a positive monetary outcome. Therefore, based on the given probability distribution, the venture is likely to be successful, considering success as having a positive monetary outcome.

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

**Ans:** The long-term average earnings, also known as the expected value or mean. Expected Value =E(X)=∑​*xi*​⋅*P*(*xi*​)

E(X)=(−2000\*0.1)+(−1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)

=−200−100+0+200+600+300=800

So, the long-term average earnings for business ventures of this kind is $800 (to the nearest $1,000). This means that, on average, the business is expected to earn $800 per venture in the long run.

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

**Ans:** The measure of risk in a probability distribution is often quantified usingthe standard deviation.

*σ*=sqrt[∑(*xi*​−*μ*)^2⋅*P*(*xi*​)​]

In this case, we've already calculated the expected value to be $800. *σ*=sqrt((−2000−800)2⋅0.1)+((−1000−800)2⋅0.1)+((0−800)2⋅0.2)+((1000−800)2⋅0.2)+((2000−800)2⋅0.3)+((3000−800)2⋅0.1)​

*σ*≈sqrt(12002⋅0.1)+(2002⋅0.1)+(8002⋅0.2)+(2002⋅0.2)+(12002⋅0.3)+(22002⋅0.1)​

*σ*≈sqrt(1440000)+(40000)+(128000)+(40000)+(432000)+(4840000)​

*σ*≈sqrt(6776000​)

*σ*≈2604.85

So, the standard deviation is approximately $2604.85 (to the nearest $1,000). This is a measure of the spread or variability of the returns, indicating the degree of risk involved in the business venture. The larger the standard deviation, the higher the risk.

Top of Form