Question 1.
simulate 30 rolls with =RANDBETWEEN(1,6). What is the probability of rolling a 3 exactly 5 times? (Hint: Use BINOM.DIST)

Answer:
Given:
A die is rolled 30 times using `=RANDBETWEEN(1,6)`
Probability of getting a 3 in one roll = $\frac{1}{6}$
Number of trials (n) = 30
Number of times 3 occurs (x) = 5

To find:
Probability of rolling a 3 exactly 5 times.

Solution (Using Binomial Distribution)

This is a binomial experiment because:

- Number of trials is fixed (30)
- Each roll is independent
- Probability of success (getting 3) is constant

So, we use BINOM.DIST.

Formula (Theory):

$$p(x)=x)=\binom{n}{x}p^{x}\,(1-p)^{n-x}$$

Substituting values:

$$P(X=5)=\binom{30}{5}\left(\frac{1}{6}\right)^{5}\left(\frac{5}{6}\right)^{25}$$

Solution (In Excel)

Excel formula:

```
=BINOM.DIST(5,30,1/6,FALSE)
```

Final Answer

The probability of rolling a 3 exactly 5 times in 30 rolls is approximately:

0.129

Question 2.
Generate 100 values in Excel using the continuous uniform distribution RAND() and plot a histogram. Describe the shape of the distribution.

Answer:

Step 1: Generate random values

In Excel, the formula

```
=RAND()
```

is entered in cell A1 and dragged down up to A100.
This generates 100 random values between 0 and 1, following a continuous uniform distribution.

Step 2: Plot the histogram

The generated data (A1:A100) is selected and a Histogram chart is created using the Insert → Chart option in Excel.

Step 3: Description of the distribution

The histogram shows that:

- Values are spread across the entire interval from 0 to 1
- Frequencies of different ranges are almost equal
- No specific peak or skewness is observed

Shape of the Distribution

The distribution is uniform in shape.

This is because, in a continuous uniform distribution, all values within the given range have equal probability of occurring.

Answer

The histogram obtained from RAND() values is approximately flat, indicating a uniform distribution.

Question 3.
A dataset has a mean of 50 and a standard deviation of 5. What percentage of values lie between 45 and 55 if the data follows a normal distribution?

Answer:

Given:
Mean (μ) = 50
Standard deviation (σ) = 5

To find:
Percentage of values lying between 45 and 55.

Solution

The given data follows a normal distribution.

- Lower limit = 45 = $\mu - 1\sigma$
- Upper limit = 55 = $\mu + 1\sigma$

So, the range 45 to 55 represents mean ± 1 standard deviation.

According to the Empirical Rule (68–95–99.7 rule), about 68% of the data in a normal distribution lies within ±1 standard deviation from the mean.

Answer

Approximately 68% of the values lie between 45 and 55.

Question:4

What is the concept of standardization (z-score), and why is it important in data analysis? Explain the formula and how standardization transforms a dataset.

Answer:

Concept of Standardization (Z-score)

Meaning of Standardization

Standardization is the process of converting raw data values into z-scores.
A z-score shows how many standard deviations a data value is away from the mean.

Formula of Z-score

$$z=\frac{x-\mu}{\sigma}$$

Where:

- X = original data value
- μ = mean of the dataset
- σ = standard deviation of the dataset

How Standardization Transforms a Dataset

When a dataset is standardized:

- The mean becomes 0
- The standard deviation becomes 1
- Original units are removed
- All values are expressed on the same scale

Each value is transformed into a z-score representing its relative position in the dataset.

Importance of Standardization in Data Analysis

Standardization is important because:

- It allows comparison between datasets with different units or scales
- It helps identify outliers
- It is required for many statistical and machine learning methods
- It makes data easier to interpret using the normal distribution

Answer

Standardization converts raw data into z-scores using the mean and standard deviation. It transforms the dataset to a common scale with mean 0 and standard deviation 1, making comparison and analysis easier.

Question 5.

What is Kurtosis and their type?

Answer:

Kurtosis

Meaning of Kurtosis

Kurtosis is a statistical measure that describes the shape of a distribution, especially the peakedness or flatness of the data compared to a normal distribution.

It tells us how data values are concentrated around the mean.

Types of Kurtosis

There are three types of kurtosis:

1. Mesokurtic

- This is the normal distribution
- It has moderate peak and tails
- Kurtosis value ≈ 3

Example:
Height of students in a large population

2. Leptokurtic

- Distribution is more peaked
- Has heavy tails
- Kurtosis value greater than 3

Example:
Income distribution where extreme values occur frequently

3. Platykurtic

- Distribution is flatter
- Has light tails
- Kurtosis value less than 3

Example:
Uniformly distributed data

Answer

Kurtosis measures the peakedness of a distribution. The three types of kurtosis are Mesokurtic, Leptokurtic, and Platykurtic, depending on whether the distribution is normal, sharply peaked, or flat.

Question 6.
Explain why the uniform distribution is a good model for the outcome of rolling a fair die.

Answer:

Uniform Distribution and Fair Die

Explanation

The uniform distribution is a good model for the outcome of rolling a fair die because each possible outcome has an equal chance of occurring.

A fair die has six possible outcomes: 1, 2, 3, 4, 5, and 6. Since the die is fair, the probability of getting any one of these numbers is the same, that is 1/6.

In a uniform distribution, all outcomes within a given range have equal probability, which matches the behavior of a fair die.

Answer

The uniform distribution is suitable for modeling a fair die because all outcomes from 1 to 6 are equally likely, satisfying the condition of equal probability required for a uniform distribution.

Question 7.
Use Excel to compute the probability of getting at least 8 successes in 15 trials with success probability 0.5

Answer:

Given

Number of trials (n) = 15
Probability of success (p) = 0.5
“At least 8 successes” means 8 or more successes.

Solution

This is a binomial distribution problem because:

- Number of trials is fixed
- Each trial is independent
- Probability of success remains constant

To find probability of at least 8 successes, we calculate:

P(X≥8)=1−P(X≤7)P(X \ge 8) = 1 - P(X \le 7)P(X≥8)=1−P(X≤7)

Excel Method

In Excel, use the cumulative binomial distribution:

```
=1-BINOM.DIST(7,15,0.5,TRUE)
```

Result

The computed probability is approximately:

0.696

Answer

The probability of getting at least 8 successes in 15 trials with success probability 0.5 is approximately 0.696.

Question 8.
How does log transformation help in stabilizing variance and making data more normally distributed?

Answer:

Log Transformation

Meaning of Log Transformation

Log transformation is a data transformation technique in which the logarithm of each data value is taken.
It is commonly used when data is highly skewed or shows large variation.

How Log Transformation Stabilizes Variance

In many datasets, variance increases as the values increase (heteroscedasticity).
Log transformation:

- Compresses large values more than small values
- Reduces the effect of extreme values
- Makes the variance more constant across the data range

This helps in stabilizing variance.

How Log Transformation Makes Data More Normally Distributed

Right-skewed data often has a long tail on the higher side.
Log transformation:

- Pulls in the right tail
- Reduces skewness
- Makes the distribution more symmetric

As a result, the data becomes closer to a normal distribution.

Importance in Data Analysis

Log transformation is important because:

- It improves model performance
- It helps meet assumptions of statistical tests
- It makes patterns easier to interpret

Answer

Log transformation helps stabilize variance by reducing the impact of large values and makes data more normally distributed by reducing skewness. This improves the reliability and interpretation of statistical analysis.