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

   Each roll has:

- Probability of getting a 3 = $p = \frac{1}{6}$
- Number of trials = $n = 30$
- We want exactly 5 times → $x = 5$
- 1−p = 5/6

  The probability of rolling a 3 exactly 5 times in 30 rolls is about 19.21%.

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

   Describe the Shape of the Distribution

- The histogram appears approximately flat / rectangular
- No strong peak or skewness
- All intervals have roughly equal frequency
- Small ups and downs occur due to random sampling

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?**

- Mean $\mu = 50$
- Standard deviation $\sigma = 5$

$$45 = 50 - 1\sigma$$
$$55 = 50 + 1\sigma$$

   So, the range **45 to 55** is **within ±1 standard deviation** of the mean.
   About **68%** of values lie within **±1 standard deviation** of the mean.
   Approximately 68% of the values lie between 45 and 55.

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.**

   Standardization is the process of converting data values into z-scores, which tell us how far a value is from the mean, measured in standard deviations.
   In simple terms, a z-score shows a value's relative position within a dataset.

   Formula: $z = (x-\mu)/\sigma$

   Where:

- x= individual data value
- μ= mean of the dataset
- σ= standard deviation

After standardization:

- The mean becomes 0
- The standard deviation becomes 1
- Original units are removed (unit-free scale)

Example:

If the mean score is 50 and standard deviation is 5:

- A value of 55 →

$$z = \frac{55 - 50}{5} = 1$$

This means the value is 1 standard deviation above the mean.

Why Standardization Is Important in Data Analysis

1. Comparison Across Different Scales
   Allows fair comparison between datasets with different units (e.g., marks vs. salaries).
2. Outlier Detection
   Large positive or negative z-scores indicate unusual or extreme values.
3. Used in Statistical Methods
   Required in many techniques such as:
   - Normal distribution probabilities
   - Regression analysis
   - Machine learning algorithms (KNN, SVM, PCA)
4. Improves Model Performance
   Prevents variables with larger scales from dominating analysis.

5. What is Kurtosis and their type?

Kurtosis is a statistical measure that describes the shape of a distribution, specifically the peakedness and tail thickness compared to a normal distribution.
In simple words, kurtosis tells us how heavy or light the tails of a distribution are.

Types of Kurtosis
There are three main types:

1. Mesokurtic

- Kurtosis ≈ 3 (or excess kurtosis = 0)
- Shape similar to a normal distribution
- Moderate peak and tails
  Example: Normal distribution

2. Leptokurtic

- Kurtosis > 3 (positive excess kurtosis)
- Sharp peak with heavy tails
- More extreme values (outliers)
  Example: Financial return data

3. Platykurtic

- Kurtosis < 3 (negative excess kurtosis)
- Flat peak with light tails
- Fewer extreme values

Example: Uniform distribution

Kurtosis Formula (Population)

$$\text{Kurtosis} = \frac{1}{n}\sum\left(\frac{x-\mu}{\sigma}\right)^4$$

- Often reported as Excess Kurtosis = Kurtosis − 3
- Normal distribution has excess kurtosis = 0

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

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

Explanation

- A fair die has six outcomes: 1, 2, 3, 4, 5, and 6.
- Since the die is fair, no number is favored over another.
- Therefore, each outcome has an equal probability:

$$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$$

Why Uniform Distribution Fits

- Equal likelihood: Uniform distribution assumes all outcomes are equally likely, which matches a fair die.
- No skewness: The distribution is symmetric with no bias toward any value.
- Consistent frequencies: Over many rolls, each number appears about the same number of times.

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

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

1. Stabilizing Variance

In many datasets, variability increases as values get larger (called heteroscedasticity).

How log helps:

- The log function compresses large values more than small ones
- Reduces the influence of extreme values
- Makes the spread of data more uniform across levels

Example:

- Original values: 10, 100, 1000

- Log values: 1, 2, 3
  Large gaps shrink → variance becomes more stable

  2. Making Data More Normally Distributed
  Many real-world datasets have a long right tail (positive skew).
  How log helps:
- Pulls in the right tail
- Reduces skewness
- Makes the distribution more symmetric and bell-shaped

  3. Improves Statistical Modeling
- Many statistical methods assume normality and constant variance
- Log-transformed data better meets these assumptions
- Leads to more reliable regression and hypothesis testing results

  4. Key Insight
  Log transformation converts multiplicative relationships into additive ones, simplifying patterns and reducing unequal spread.