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

**Solution:**

This problem is based on the **Binomial Distribution** because:

1. There are a fixed number of rolls (30).
2. Each roll is independent.
3. Each roll has only two outcomes — “Success” (getting a 3) or “Failure” (not getting a 3).
4. The probability of success stays constant at .

**Formula:**

**Given:**

**Step-by-step:**

**In Excel:**  
=BINOM.DIST(5, 30, 1/6, FALSE)

**Final Answer:**

So, the probability of rolling a 3 exactly 5 times in 30 rolls is **19.21%**.

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

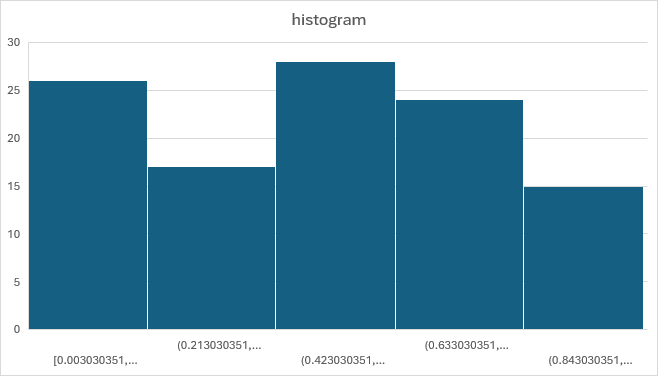
**Solution:**

The RAND() function in Excel generates numbers between **0 and 1**, where each value in this range is equally likely.  
This follows a **Continuous Uniform Distribution**.

**In Excel:**=RAND().

**Explanation:**  
A uniform distribution means all values occur with equal probability.  
So, when you plot the histogram, it should look **flat** — meaning no particular number range appears more often than another.

**Final Answer:**  
The histogram is **roughly flat or rectangular**, showing that all values between 0 and 1 occur with almost equal frequency.



**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?**

**Solution:**

This question uses the **Empirical Rule** (68–95–99.7 Rule) for Normal Distribution.

**Empirical Rule:**

* 68% of data lies within ±1 standard deviation.
* 95% lies within ±2 standard deviations.
* 99.7% lies within ±3 standard deviations.

**Given:**

* Mean (μ) = 50
* Standard deviation (σ) = 5

**Range:**  
45 = 50 - 5 = μ - 1σ  
55 = 50 + 5 = μ + 1σ

So, the range from 45 to 55 covers **one standard deviation** from the mean.

**Final Answer:**  
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.**

**Solution:**

**Concept:**  
Standardization (or z-score) converts raw data into a common scale that shows how far each value is from the mean, measured in standard deviations.

**Formula:**

Where:  
= standardized value  
= raw score  
= mean  
= standard deviation

**Why It’s Important:**

1. It allows comparison of values from different datasets or scales.
2. It helps identify **outliers** (values far from the mean).
3. It’s essential for many statistical and machine learning methods.

**Transformation Effect:**  
After standardizing:

* The new **mean becomes 0**.
* The new **standard deviation becomes 1**.

**Final Answer:**  
Standardization transforms data into z-scores, making them comparable and centered around zero.

**5. What is Kurtosis and its types?**

**Solution:**

**Concept:**  
Kurtosis measures how heavy or light the tails of a distribution are compared to a normal distribution — in simple terms, it shows **how peaked or flat** a distribution is.

**Types of Kurtosis:**

1. **Mesokurtic (K = 3)**
   * Same as a normal curve.
   * Moderate tails and peak.
2. **Leptokurtic (K > 3)**
   * Sharper peak and heavier tails.
   * Indicates more outliers.
3. **Platykurtic (K < 3)**
   * Flatter peak and lighter tails.
   * Indicates fewer outliers.

**Final Answer:**  
Kurtosis describes the "tailedness" of a distribution.  
Types are **Mesokurtic**, **Leptokurtic**, and **Platykurtic**.

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

**Solution:**

A **uniform distribution** means every possible result has an equal chance of happening.

**For a fair die:**

* Outcomes: 1, 2, 3, 4, 5, 6
* Each outcome has the same probability:

Since all outcomes are equally likely, this matches the definition of a **discrete uniform distribution**.

**Final Answer:**  
A fair die follows a **discrete uniform distribution** because each of its six faces has an equal probability of appearing.

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

**Solution:**

We are finding:

where .

**In Excel:**  
=1 - BINOM.DIST(7, 15, 0.5, TRUE)

**Explanation:**  
The cumulative probability up to 7 gives .  
Subtracting this from 1 gives the probability of getting **8 or more** successes.

**Final Answer:**

This makes sense because with , the distribution is symmetrical — half the time you’ll get 7 or fewer successes, and half the time 8 or more.

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

**Solution:**

A **log transformation** is used when data is highly skewed (especially right-skewed).

**1. Stabilizing Variance:**  
Large values get compressed more than small ones.  
For example:

This compression reduces the spread of data, making variance more stable.

**2. Making Data Normal:**  
Right-skewed data has a long tail on the higher side.  
Taking logs “pulls in” the tail, making the data more **symmetric** and closer to a **bell-shaped curve**.

**Final Answer:**  
Log transformation:

1. **Stabilizes variance** by compressing large values.
2. **Makes data more normally distributed** by reducing right-skewness.