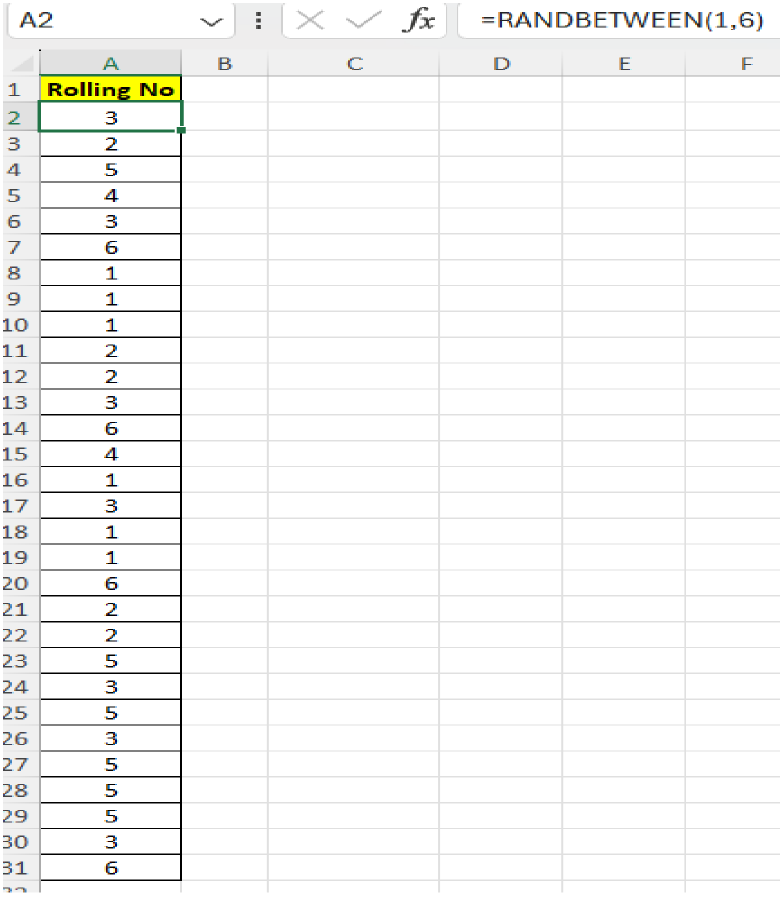
1. **Simulate 30 rolls with =RANDBETWEEN(1,6). What is the probability of rolling a 3 exactly 5 times? (Hint: Use BINOM.DIST**

Ans:-

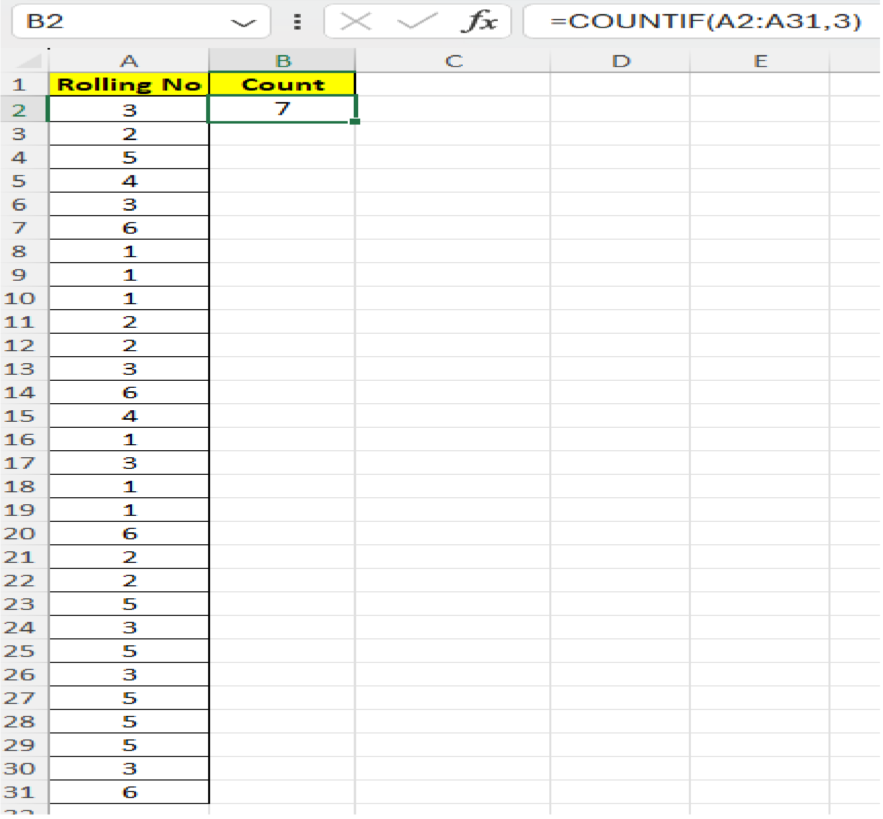
Step- 1



A2 =RANDBETWEEN(1,6)

	A
1	Rolling No
2	3
3	2
4	5
5	4
6	3
7	6
8	1
9	1
10	1
11	2
12	2
13	3
14	6
15	4
16	1
17	3
18	1
19	1
20	6
21	2
22	2
23	5
24	3
25	5
26	3
27	5
28	5
29	5
30	3
31	6

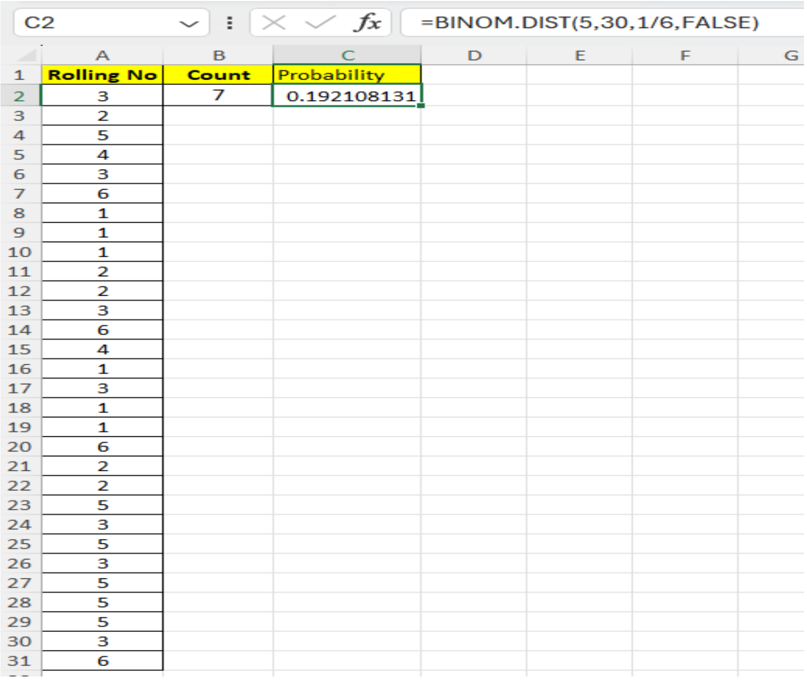
Step- 2



B2 =COUNTIF(A2:A31,3)

	A	B
1	Rolling No	Count
2	3	7
3	2	
4	5	
5	4	
6	3	
7	6	
8	1	
9	1	
10	1	
11	2	
12	2	
13	3	
14	6	
15	4	
16	1	
17	3	
18	1	
19	1	
20	6	
21	2	
22	2	
23	5	
24	3	
25	5	
26	3	
27	5	
28	5	
29	5	
30	3	
31	6	

Step- 3



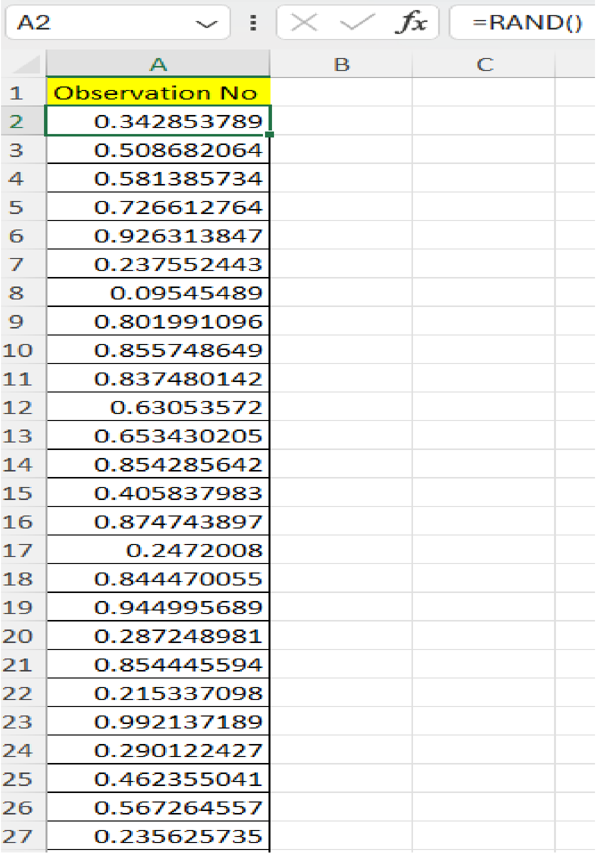
C2 =BINOM.DIST(5,30,1/6,FALSE)

Rolling No	Count	Probability
3	7	0.192108131
2		
5		
4		
3		
6		
1		
1		
1		
2		
2		
3		
6		
4		
1		
3		
1		
1		
6		
2		
2		
5		
3		
5		
3		
5		
5		
5		
3		
6		

The probability of rolling a 3 exactly 5 times in 30 dice rolls is 0.1921 (19.21%), calculated using the binomial distribution.

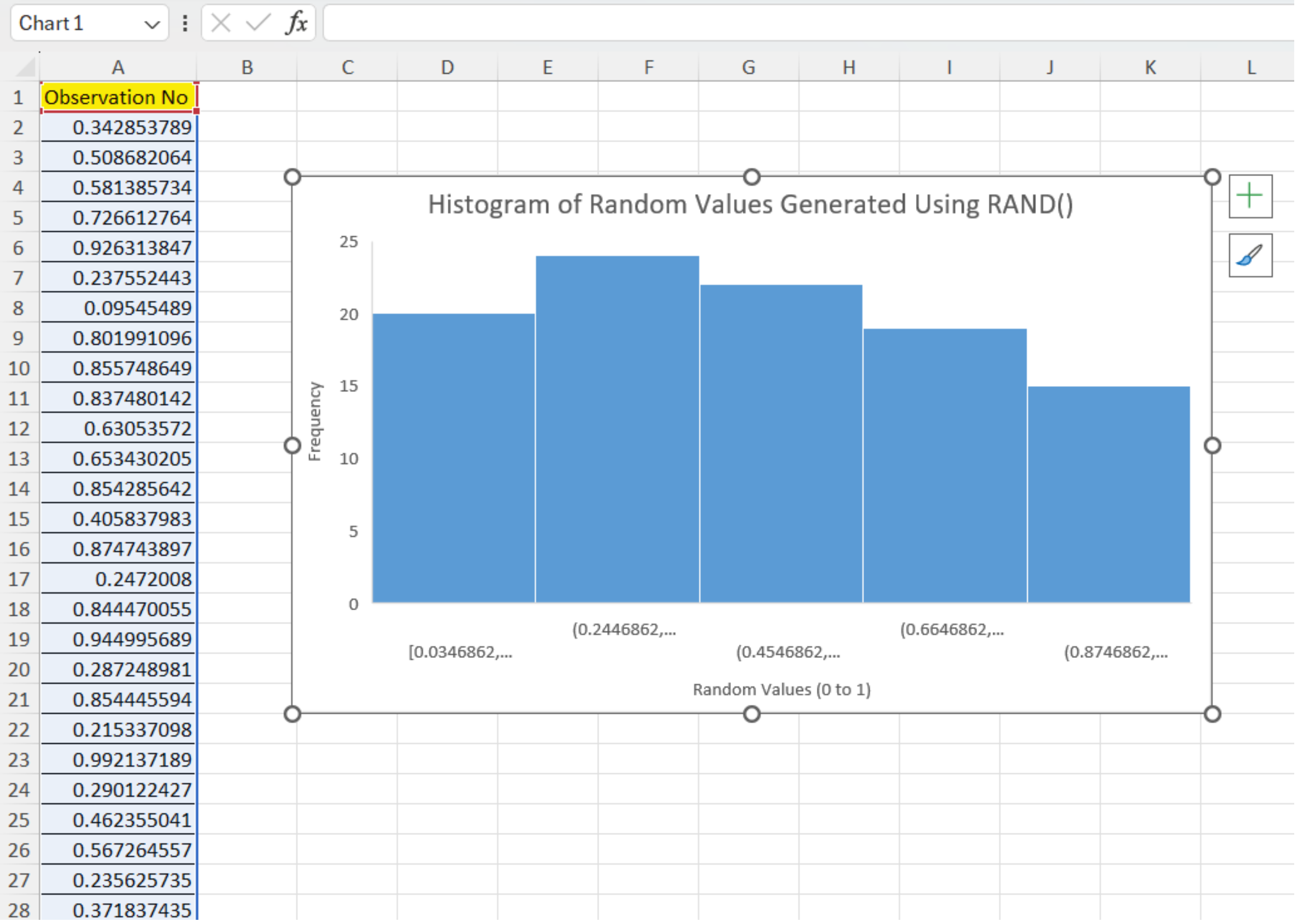
2. **Generate 100 values in Excel using the continuous uniform distribution RAND() and plot a histogram. Describe the shape of the distribution.**

Ans:-



A2 =RAND()

Observation No
0.342853789
0.508682064
0.581385734
0.726612764
0.926313847
0.237552443
0.09545489
0.801991096
0.855748649
0.837480142
0.63053572
0.653430205
0.854285642
0.405837983
0.874743897
0.2472008
0.844470055
0.944995689
0.287248981
0.854445594
0.215337098
0.992137189
0.290122427
0.462355041
0.567264557
0.235625735



	A
1	Observation No
2	0.342853789
3	0.508682064
4	0.581385734
5	0.726612764
6	0.926313847
7	0.237552443
8	0.09545489
9	0.801991096
10	0.855748649
11	0.837480142
12	0.63053572
13	0.653430205
14	0.854285642
15	0.405837983
16	0.874743897
17	0.2472008
18	0.844470055
19	0.944995689
20	0.287248981
21	0.854445594
22	0.215337098
23	0.992137189
24	0.290122427
25	0.462355041
26	0.567264557
27	0.235625735
28	0.371837435

The histogram of the 100 values generated using the RAND() function is approximately uniform (rectangular) in shape, indicating that the values are evenly distributed between 0 and 1. This confirms the characteristics of a continuous uniform distribution.

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?**

Ans:-

- Mean μ = 50
- Standard deviation σ = 5
- The interval 45 to 55 is μ ± 1σ

Using the Empirical Rule (68–95–99.7 rule):

Normal distribution with a mean of 50 and a standard deviation of 5. According to the Empirical Rule, approximately 68% of the values lie within one standard deviation of the mean.The range 45 to 55 represents 50±5 about 68% of the data falls within this interval.

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.**

Ans:-

Concept of Standardization (Z-score)

Standardization is the process of converting raw data values into z-scores, which show how many standard deviations a value is from the mean of the dataset.

Z-score Formula :- $Z = \frac{X-\mu}{\sigma}$

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

Why Standardization Is Important:-

- It allows comparison of values from different datasets
- It helps identify outliers
- It is essential for normal distribution analysis
- Many statistical and machine learning methods require standardized data

How Standardization Transforms a Dataset

- The mean becomes 0
- The standard deviation becomes 1
- The shape of the distribution remains the same
- Values are expressed on a common scale

5. **What is Kurtosis and their type?**

Ans:-

Kurtosis

Kurtosis is a statistical measure that describes the shape of a distribution, specifically the peakedness and tail thickness compared to a normal distribution.

Types of Kurtosis

1. **Mesokurtic**
   - Kurtosis ≈ 3 (or excess kurtosis = 0)

- Shape similar to the normal distribution
- Moderate peak and tails

2. **Leptokurtic**
    - Kurtosis greater than 3 (positive excess kurtosis)
    - Sharper peak and heavier tails
    - Indicates more extreme values (outliers)
3. **Platykurtic**
    - Kurtosis less than 3 (negative excess kurtosis)
    - Flatter peak and lighter tails
    - Fewer extreme values

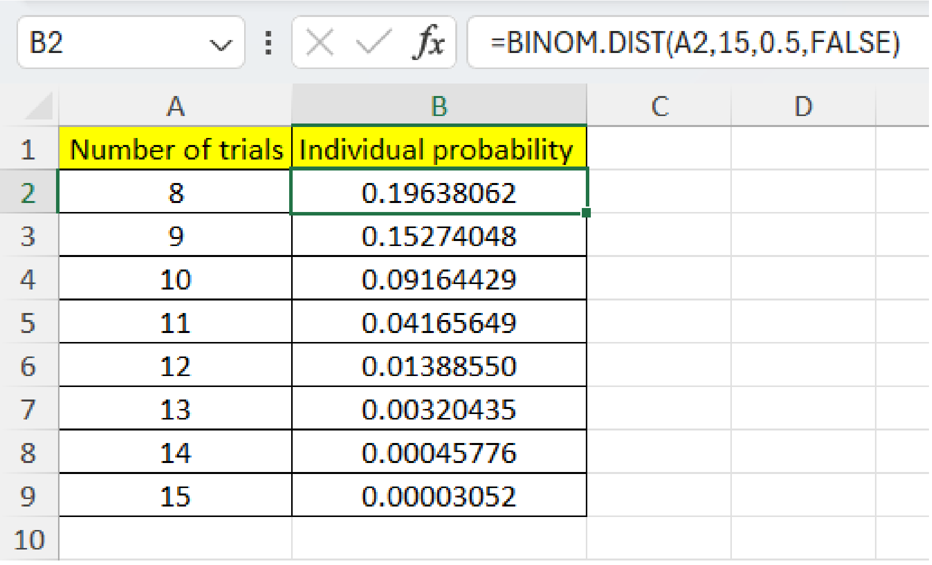
6. **Explain why the uniform distribution is a good model for the outcome of rolling a fair die.**

Ans:-

The uniform distribution is a good model for the outcome of rolling a fair die because each face (1–6) has an equal probability of occurring. Since no outcome is more likely than another, the probabilities are evenly distributed across all possible values. This equal-likelihood property matches the defining characteristic of a uniform distribution, making it an appropriate model for a fair die.

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

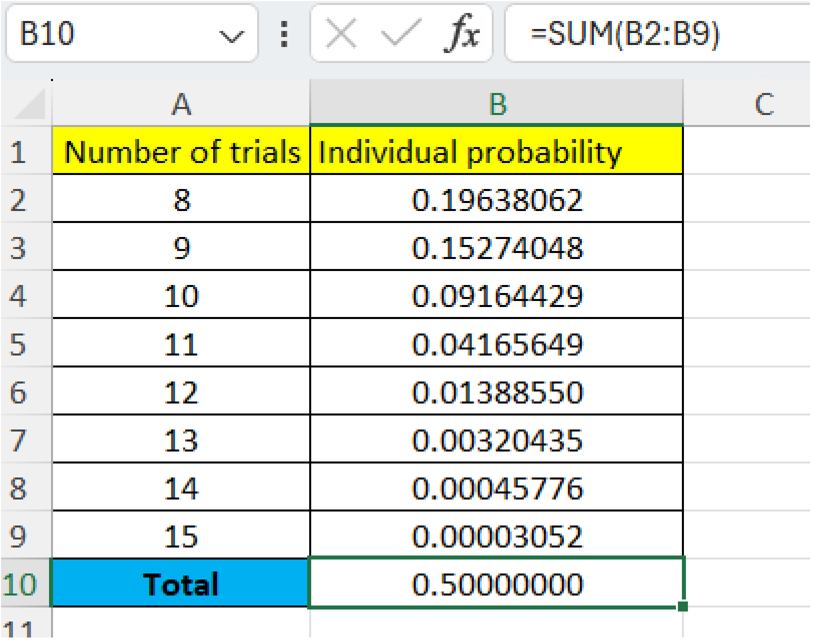
Ans:-



B2

=BINOM.DIST(A2,15,0.5,FALSE)

	A	B	C	D
1	Number of trials	Individual probability		
2	8	0.19638062		
3	9	0.15274048		
4	10	0.09164429		
5	11	0.04165649		
6	12	0.01388550		
7	13	0.00320435		
8	14	0.00045776		
9	15	0.00003052		
10				



B10 =SUM(B2:B9)

	A	B	C
1	Number of trials	Individual probability	
2	8	0.19638062	
3	9	0.15274048	
4	10	0.09164429	
5	11	0.04165649	
6	12	0.01388550	
7	13	0.00320435	
8	14	0.00045776	
9	15	0.00003052	
10	**Total**	0.50000000	
11			

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

Ans:-

1. Variance stabilization

Some datasets have heteroscedasticity, which means the variance of the data changes with the mean. For example:

- If you measure incomes, small incomes might vary by a few dollars, while large incomes can vary by thousands.
- Or in counts of events (like bacteria growth), variance often increases as the mean increases.

In such cases, standard statistical methods (like regression or t-tests) that assume constant variance can give misleading results.

Why log helps

The log function compresses large values more than small values.

Mathematically: **Y = log(X)**

- When X is large, a small percentage change in X corresponds to a smaller absolute change in log(X).
- When X is small, the change is relatively bigger.

This reduces the spread of large values, making variance more uniform across the range of the data.

Example:

X	log(X)
10	2.30
100	4.61
1000	6.91

- The difference between 10 → 100 is 90 in original scale, but only ~2.3 in log scale.
- The difference between 100 → 1000 is 900 in original scale, but only ~2.3 in log scale.

So, the variance at higher X values shrinks, stabilizing it across the dataset.

2. Making data more normally distributed

Many real-world datasets are positively skewed: they have a long right tail (like income, population counts, etc.).

- Log transformation pulls in the right tail, compressing extreme large values.
- Small values remain largely unchanged.

This can make a skewed distribution more symmetric, which is closer to the normal distribution.

Graphical intuition:

- Original data: right-skewed (long tail to the right)
- After log: data "folds" the tail inwards → roughly bell-shaped

3. Mathematical intuition

If X is log-normally distributed, meaning log(X) is normally distributed:-

$$\mathbf{X \sim \text{LogNormal}\ (\mu,\sigma^2) \Longrightarrow \log(X) \sim N\ (\mu,\sigma^2)}$$

So taking log transforms a multiplicative model (errors multiply) into an additive one (errors add), which is easier to handle with standard statistical techniques.