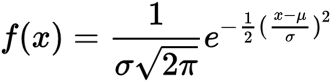
1. What is statistics
2. Types of statistics
3. Population vs sample
4. Data types
5. Levels of data
6. Parameter vs statistics
7. Frequency table
8. Frequency distribution table
9. Bar plot
10. Pie chart
11. Histogram
12. Distribution plot
13. Central tendency
14. Mean-median-mode
15. Mean vs median
16. Outlier
17. Types of distribution
18. Neg skew-pos skew-no skew
19. Range- mean deviation – absolute mean deviation
20. Variance
21. Standard deviation
22. How to find outlier
23. Percentile -Quartile
24. Box plot
25. Univariate-bivariate-multivariate
26. Scatter plot
27. Covariance
28. Covariance matrix
29. Correlation
30. Difference between covariance and correlation

**Normal distribution:**

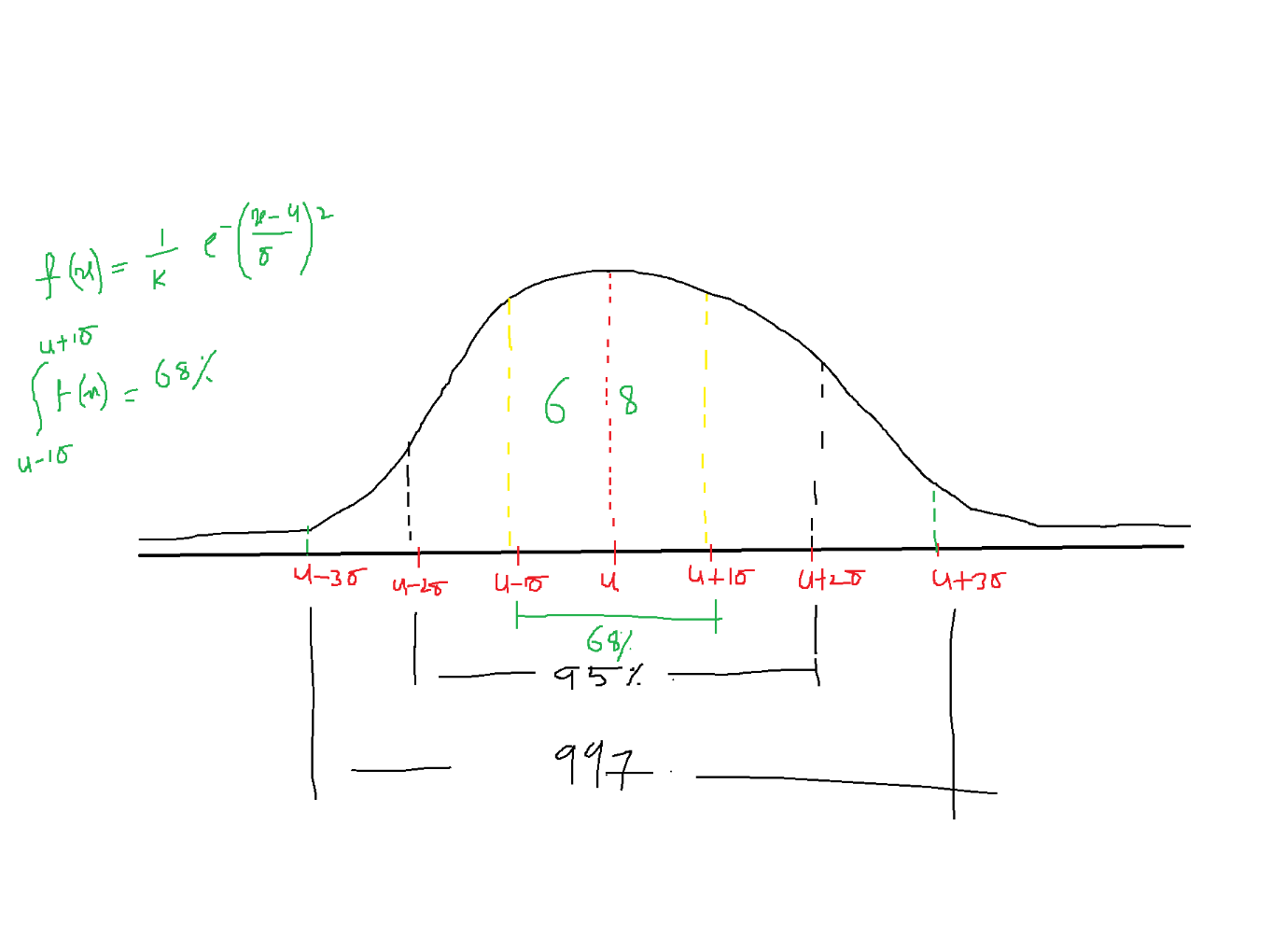
1. **Bell shaped curve**
2. **Above mean and below mean equal amount of data distributed**
3. **Mean=Medin= Mode (No skew)**
4. **Formulae**

****

If you have time

Explore Normal scale vs log scale

Empirical Rule:



Empirical rule:

\*\* Empirical rule valid if data follows normal distribution only

Example: Data distribution based on empirical rule

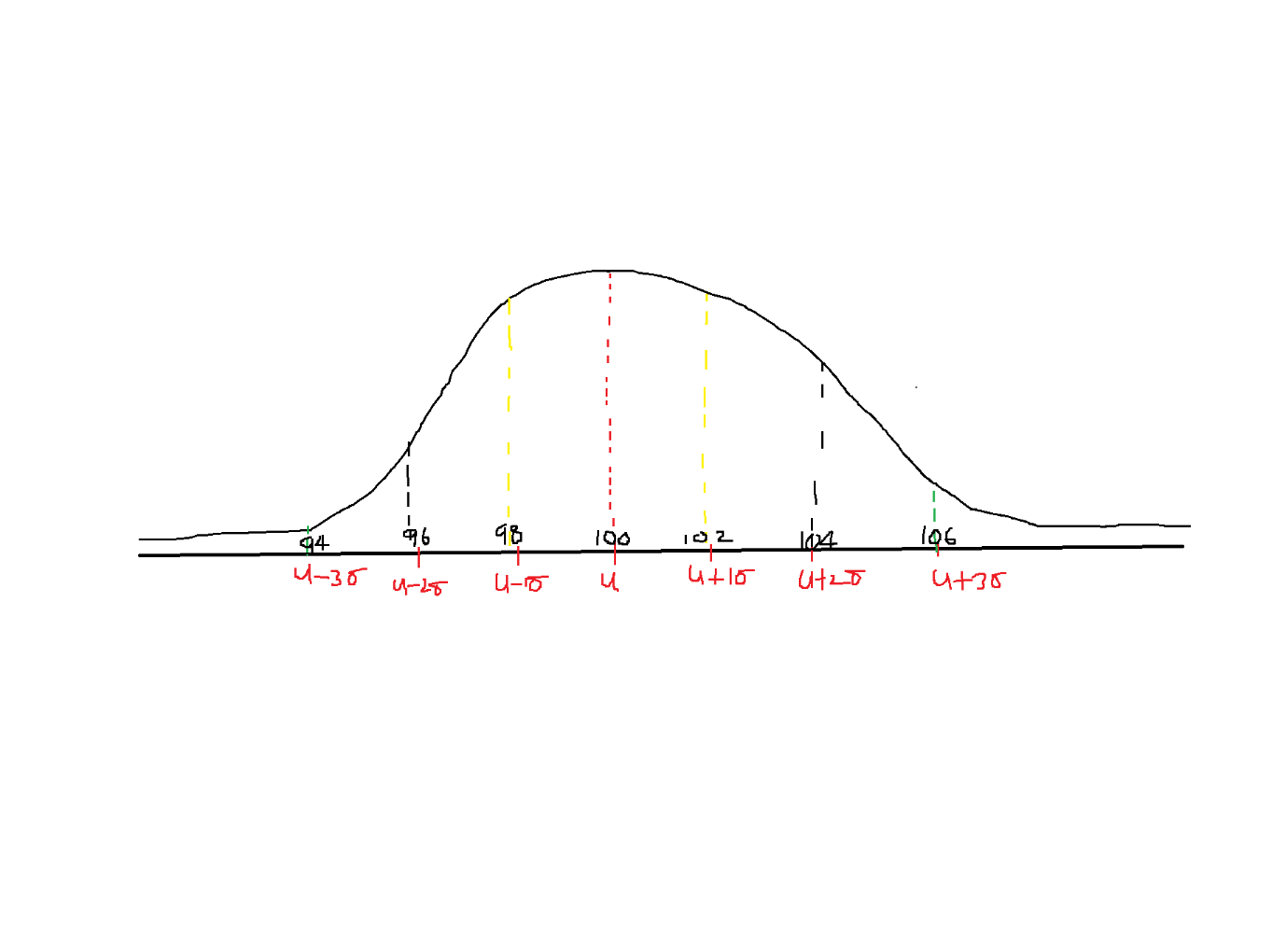
Assume that in India the petrol rates are on of average 100 rs

It might vary by 2rs state by state

Apply Empirical rule on this:

Mean petrol =

Standard deviation =



In india 68% states have petrol rates between 98 rs to 102 rs

In india 95% states have petrol rate between 96 to 104rs

In India 99.7% (almost all the states) have petrol rate between 94 to 106rs

Which means in India less than 94 rs and more than 106s

In india min petrol rate is 94 rs

And maximum petrol rate is 106rs

3\*sigma ========= oulier

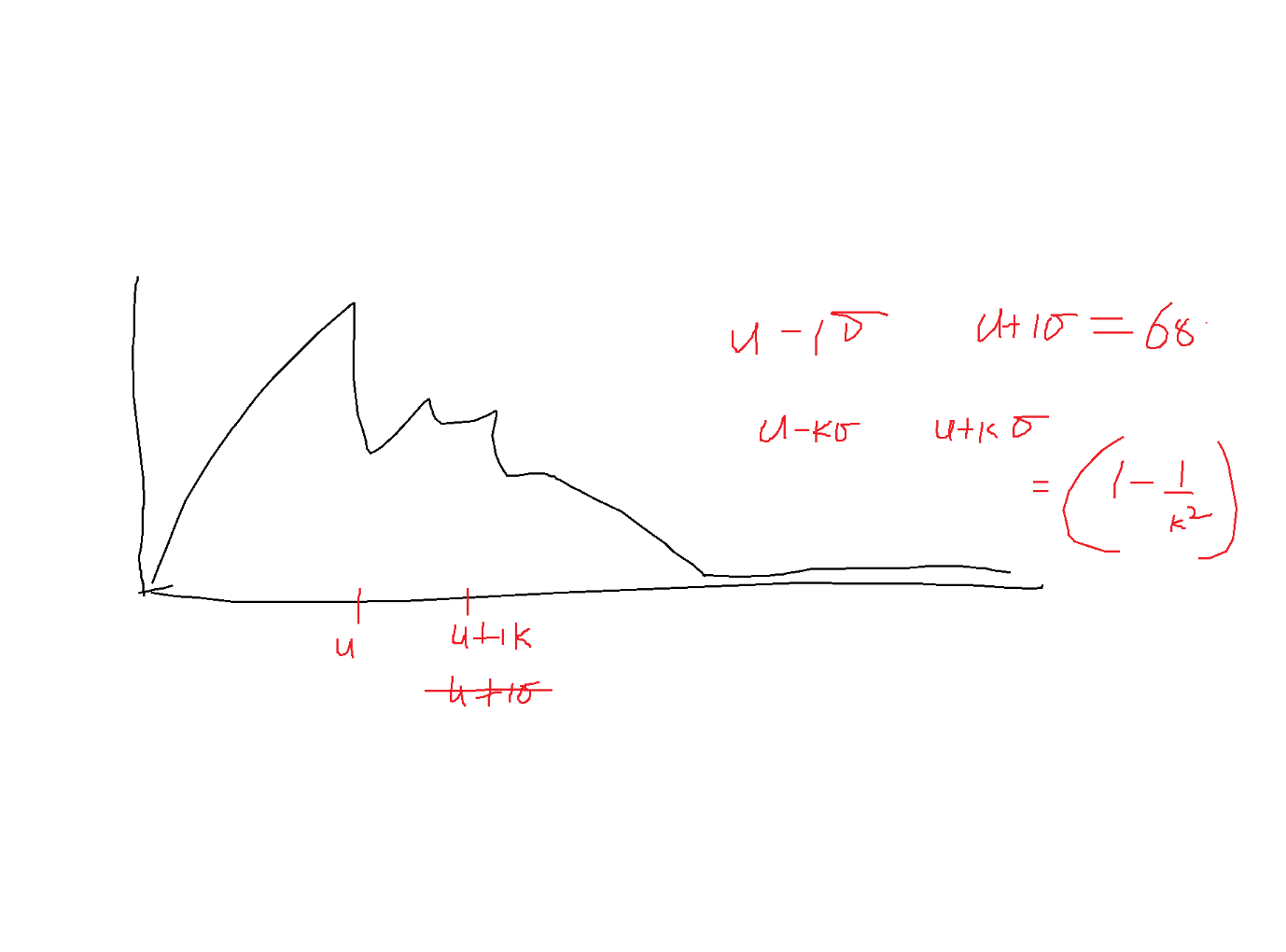
Q3+3\*IQR ============== > huge oulier

-3 to 3 range

Empirical rule valid only for Normal distribution

What happens if data does not follows normal distribution

**Chebyshev theorem or Chebyshev Inequality:**

****

**K=2**

**For K =2 standard deviations it will cover 75% of data**

**When data follow normal distribution : 2 standard deviations ==== > 95%**

**When data does nor follow ND : 2 standard deviation ======== > 75% (1-1/k^2)**

**When data follow normal distribution : 3 standard deviations ==== > 99.7%**

**When data does nor follow ND : 3 standard deviation ======== > approx. : 90%**

**Chebyshev Inequality/ Theorem**

Suppose k=2

If data not follows normal distribution, the data coverage between 2 standard deviations is

Suppose k=3

If data not follows normal distribution, the data coverage between 3 standard deviations is

**Draw back in Chebyshev :**

**K=1**

| **std** | **Normal distribution (Empirical)** | **Not Normal( Chebyshev)**  **1-1/k^2** |
| --- | --- | --- |
| **1** | **68%** | **-------- draw back ---------- when k=1, 0% coverage** |
| **2** | **95%** | **75%** |
| **3** | **99.7** | **88.8%** |

**What is difference between Empirical rule and Chebyshev rule**

Chebyshev's inequality is more general, stating that a minimum of just 75% of values must lie within two standard deviations of the mean and 88.89% within three standard deviations for a broad range of different probability distributions.

**so do we use this inequality to make the distribution normal?**

**This is not a conversion of data to normal**

**This will happen only morning time**

**what about toher distributions**

**Empirical ========== ND ( Formulae)**

**Chebsyehv ========= Foramula is there are not**

**Studying ======== It has a formulae ===== how much data coverage**

**Area a under the curve**

Every distribution ========= Formulae is important