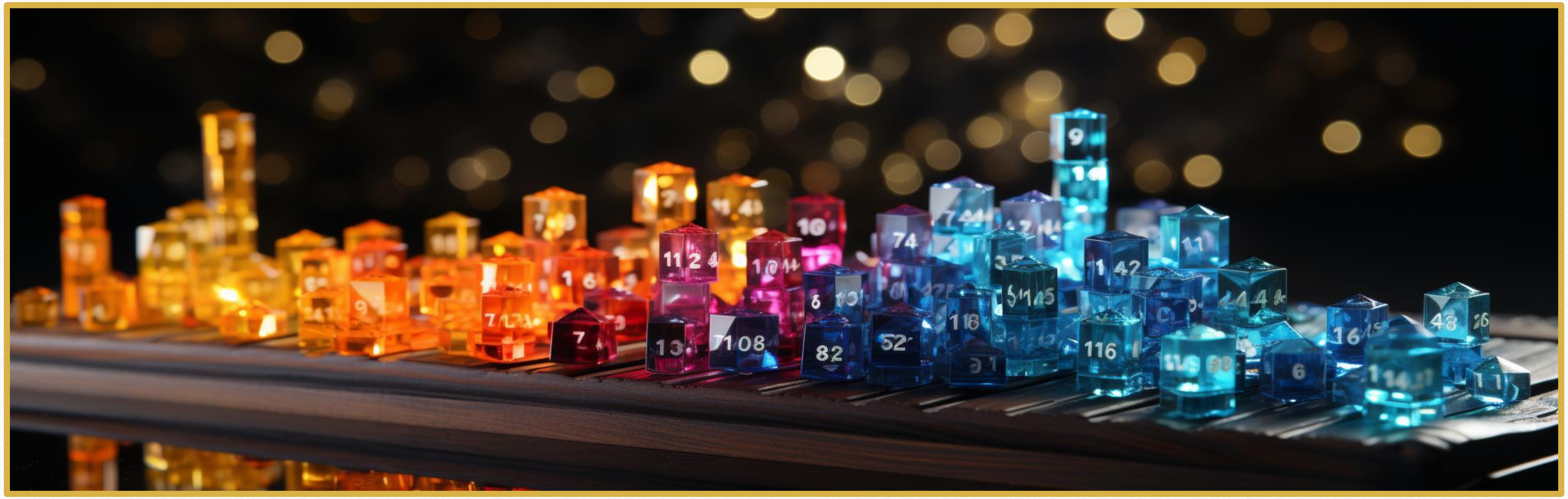
UXG1205 Lecture
10. Normal Distributions

LIN QINJIE

*This set of materials must not be used, shared, uploaded or distributed without permission from Dr Lin Qinjie.

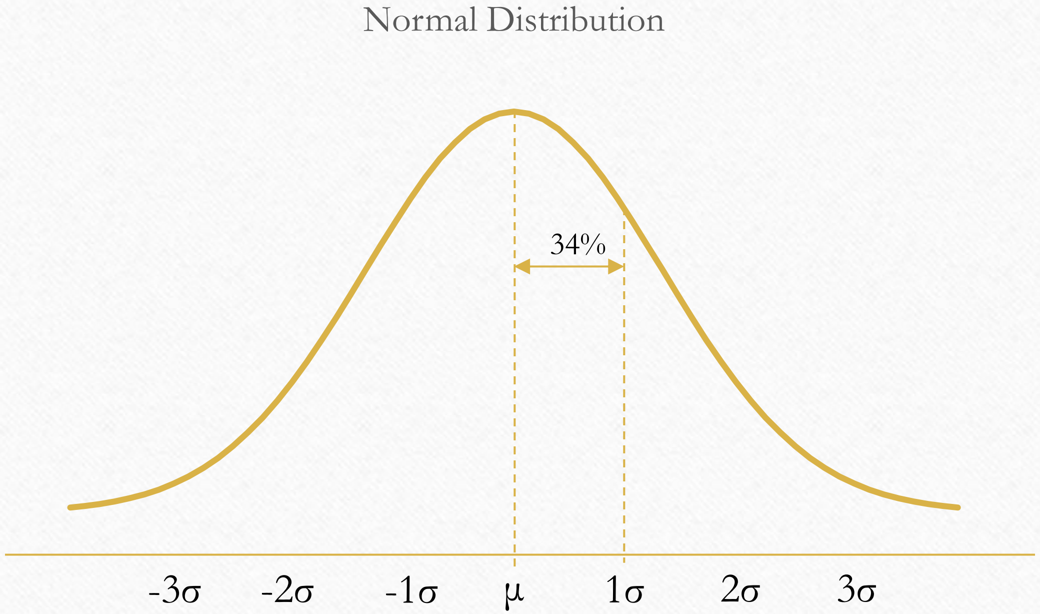
Outline

- Normal Distributions
- Empirical Rule



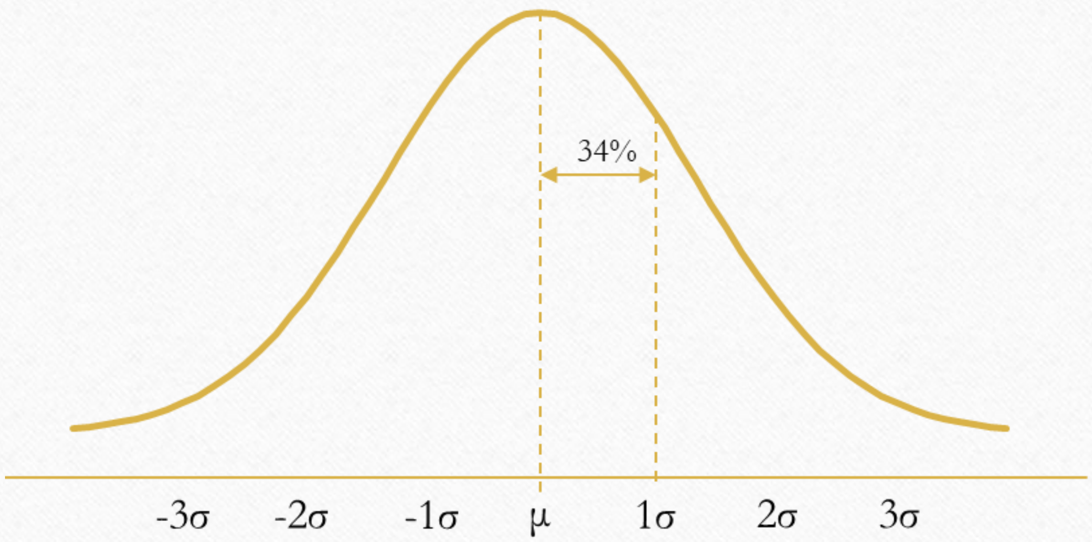
Normal Distributions

- Normally distributed random variables
- Parameters

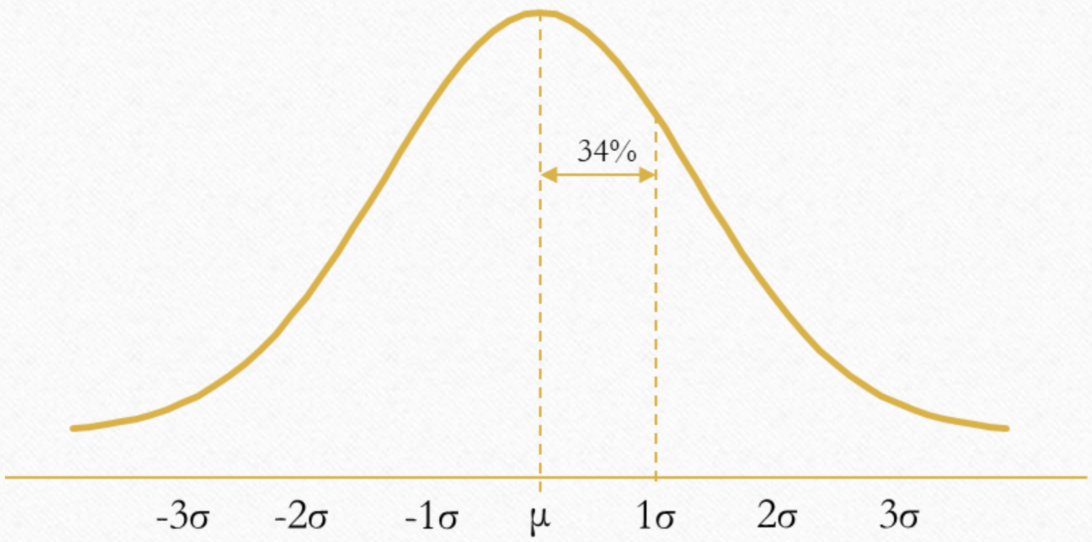


- The **normal distribution** is a type of probability distribution characterized by its bell-shaped and symmetrical curve, often referred to as the "bell curve."
- This distribution is unimodal, with a single peak.
- The characteristics of a normal curve are defined by two key parameters: the **mean** (μ) and the **standard deviation** (σ).

- Recap:
  - The mean represents the weighted average of the dataset.
  - The standard deviation quantifies the dispersion or variability within the dataset.
  - A larger standard deviation indicates greater variability or dispersion in the data points.

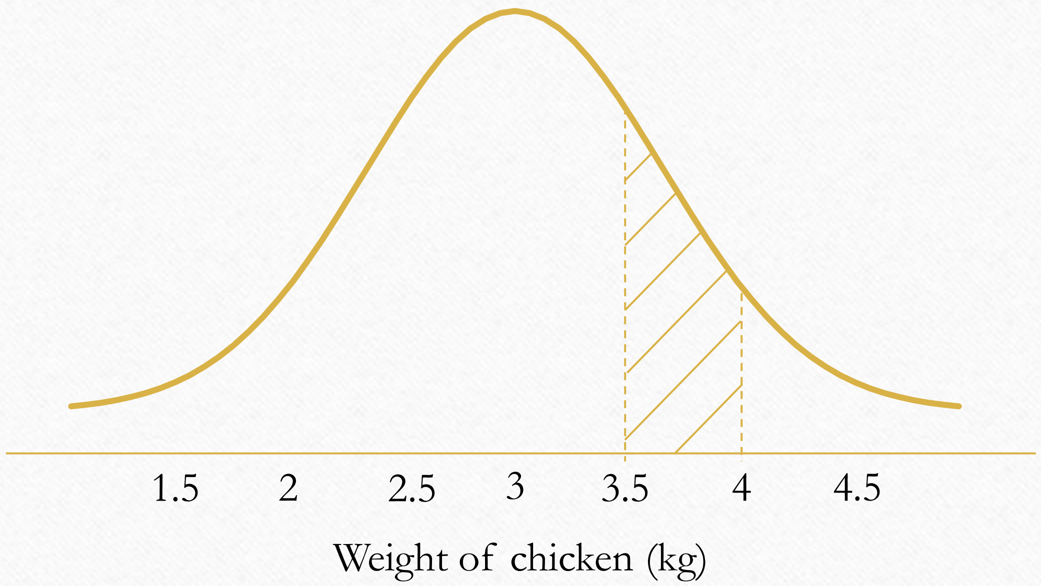


- The **normal distribution** is commonly applied to measurements such as heights, lengths, and weights, which possess an infinite number of potential values within any given range.
- The distribution is continuous. (Different from binomial distribution which is discrete.)
- Therefore, the normal distribution is depicted with a continuous curve rather than discrete bars or points.



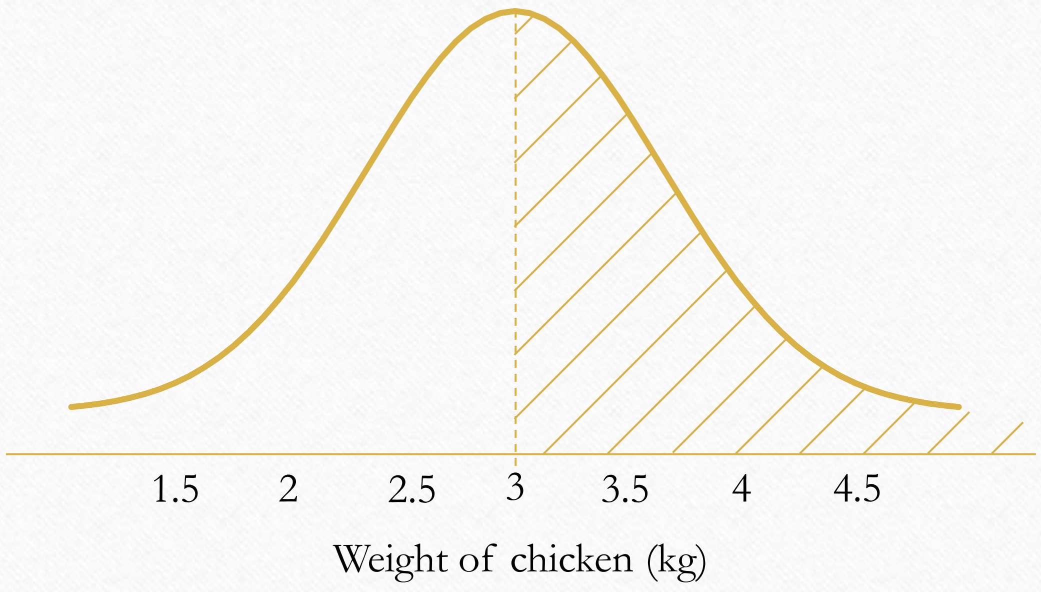
- Continuous distributions don't allow for the probability of a specific, singular outcome.
  - For e.g., $P(X = 2.5)$
- Rather, we discuss the **cumulative probability**, representing the likelihood of a span of values.
  - For e.g., $P(X > 2.5)$

- Due to the alien invasion, everywhere is a mess. So you now need to catch your own chicken for lunch.
- You are handed a normal distribution graph showing the weight of chicken (X).



- From the distribution, what does the shaded area represent?
  A. The probability that the chicken is between 3.5kg and 4kg.
  B. The probability that the chicken is above 3.5kg.
  C. The probability that the chicken is above 4kg.

- Another example.
- The above shaded region represents $P(X > 4)$.



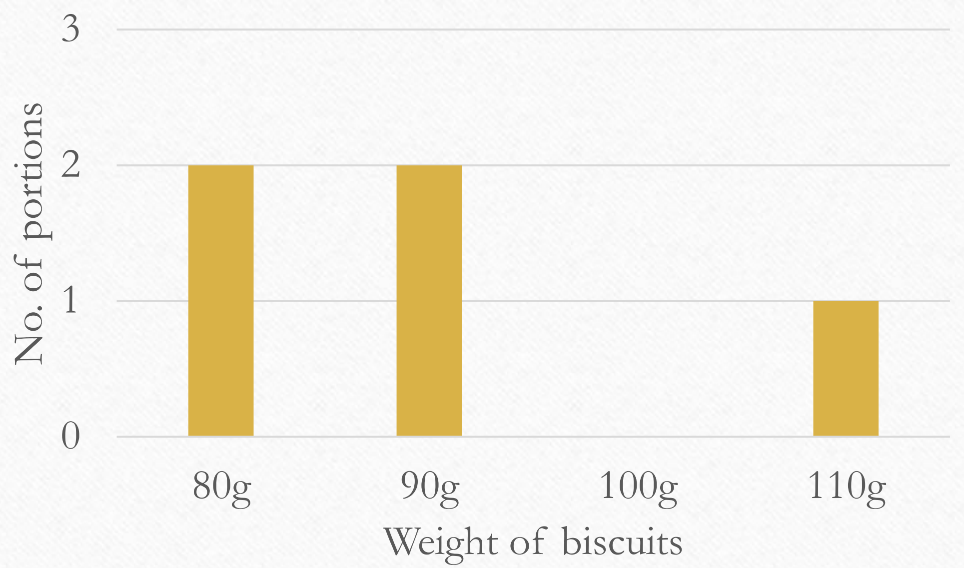
- Keep in mind, the area below the curve of any continuous probability distribution equals 1.
- Given the symmetrical nature of this curve, there's a 50% probability that a chicken chosen at random will weigh more than 3kg.

- The concept of the **normal distribution** was introduced in the 1800s by a mathematician named Gauss, leading to its alternate name, the **Gaussian distribution**.
- While every normal distribution has a bell-like shape, it's essential to note that not every bell-shaped curve qualifies as a normal distribution. (For e.g., logistic distribution)
- It's called "normal" because most common natural phenomena can be modelled using normal distribution. For e.g.,
  - Time taken to commute to work
  - Marks or grades in a large class.
  - Height of trees at the park

- Next concept – the **Central Limit Theorem (CLT)**
- The CLT proposes that when you take multiple sample means from a population, they will generally form a normal distribution, regardless of whether the original population distribution was normal or not.

Normal Distributions

- Here's an example to illustrate the CLT.
- Imagine you and 4 other survivors are selecting randomly from 5 biscuit ration portions. 2 portions weigh 80 grams, 2 are 90g, and 1 is 110g.



- Every evening, the 5 of you randomly select from the biscuit ration portions for dinner. By computing the weighted average, we determine the **expected value**:

$$\mu = \left(\frac{2}{5} \times 80\right) + \left(\frac{2}{5} \times 90\right) + \left(\frac{1}{5} \times 110\right) = 90 \text{ grams}$$

- This value signifies the average amount of biscuits you will consume for dinner over an extended period.

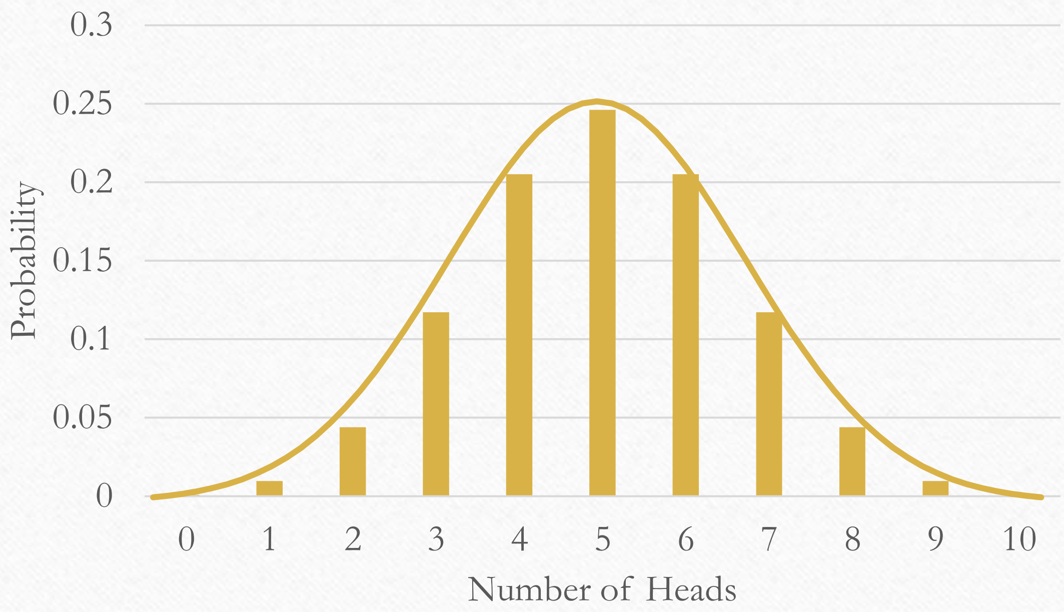
- If you look at the biscuits you receive weekly as a group, it's a set with 7 outcomes.
  - $n = 7$
- Within this group, you would be able to calculate the mean. This number (mean) represents the average weight of biscuits you receive daily for a week.
- The CLT suggests that if you chart the average weight of your biscuit rations each week, over a long time, this chart would align with a normal distribution centred around our initial expected value of 90.

- To illustrate in a table:

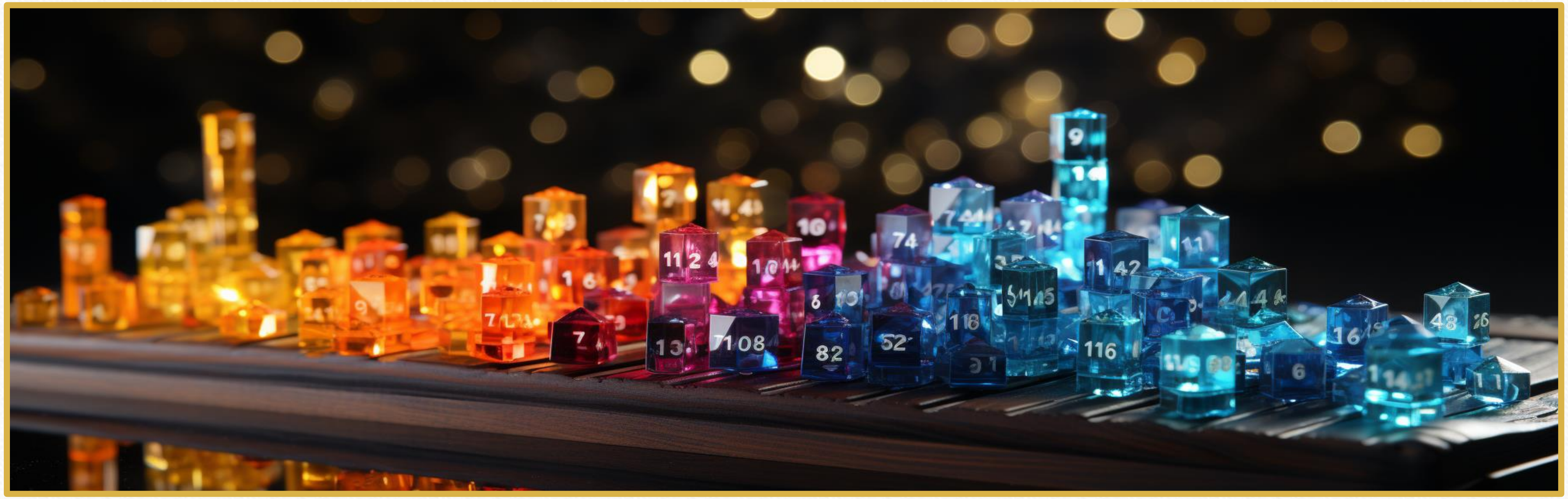
Week	Biscuit Weights (7 days)	Weekly Average
Week 1	{110, 110, 80, 80, 90, 80, 110}	94.29
Week 2	{80, 90, 90, 110, 110, 80, 110}	95.71
Week 3	{110, 90, 110, 110, 110, 110, 80}	102.86
Week 4	{90, 80, 90, 80, 80, 110, 80}	87.14
Week 5	{80, 90, 90, 80, 90, 90, 80}	85.71
Week 6	{90, 90, 80, 80, 80, 80, 80}	82.86
Week 7	{80, 80, 90, 90, 90, 80, 80}	84.29
Week 8	{80, 80, 110, 80, 110, 90, 110}	94.29
Week 9	{80, 90, 110, 90, 90, 80, 90}	90
Week 10	{90, 110, 90, 80, 90, 80, 80}	88.57

And many more weeks continued beyond this.

- The CLT is very important in statistics. It is relevant in nearly all scenarios with repeated selections, even if the original distribution of those selections isn't normal.
- Whether we chart the total grams of biscuits consumed monthly, the average yearly biscuit rations, or many other situations unrelated to biscuits, we would still observe the normal distribution.



- Remember the binomial distribution graph we plotted in the previous topic?
- The bars represent a binomial distribution, while the line depicts a normal distribution.
- The binomial distribution can be well approximated by a normal distribution!



Empirical Rule

- Simple rule to find probabilities in a normal distribution

Empirical Rule

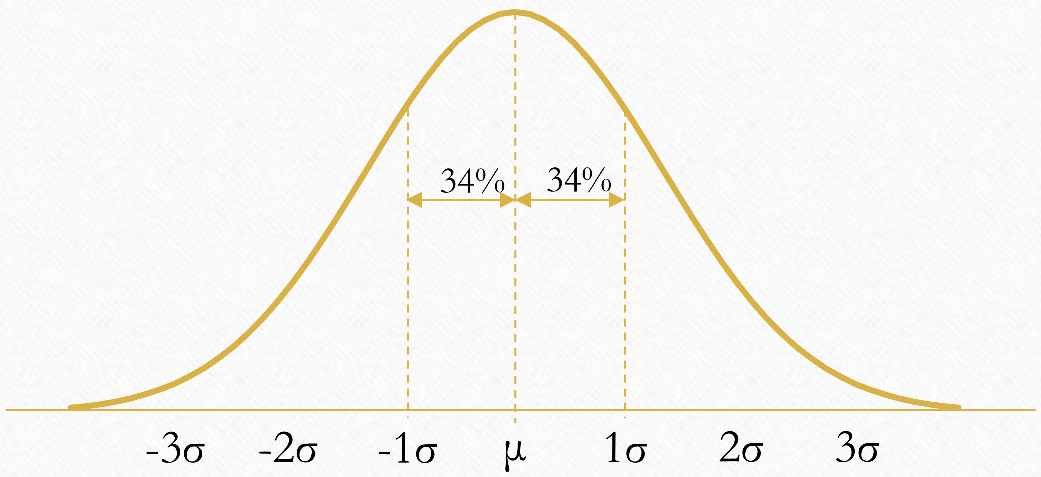
- You need to head out of the safe house with Billy to complete an anti-alien mission. 👽
- Since communication is key, he needs your help to figure out the likely battery life of the walkie-talkies.
- First, you'll have to familiarize yourself with the **Empirical Rule**.

Empirical Rule

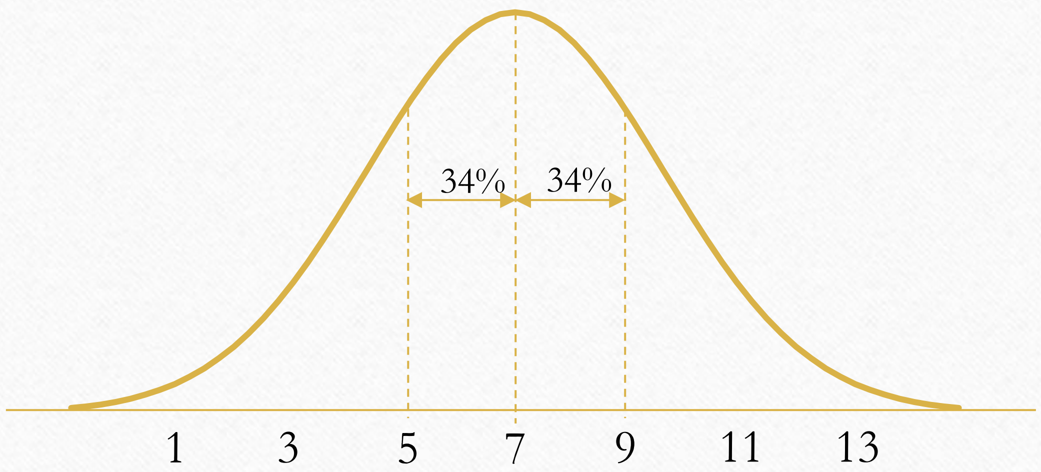
- Most walkie-talkie brands have an average battery life, but the actual battery duration can differ for each individual unit.

- Let X represent the battery life of the walkie-talkie you'll be using. X is a continuous random variable that follows a normal distribution.

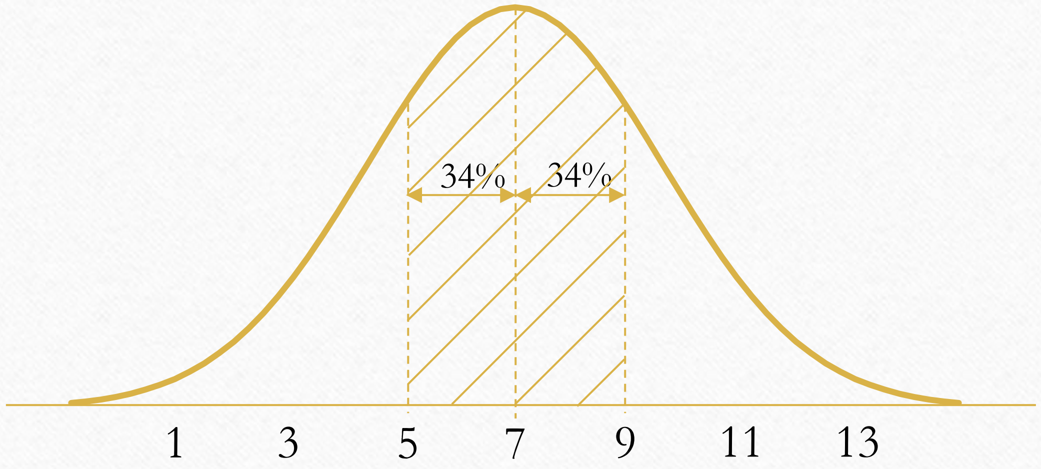
Empirical Rule



- **Empirical rule** says that for a normal distribution,
  - 68% of the data will be within 1 standard deviation of the mean.
  - 95% of the data will be within 2 standard deviations of the mean.
  - 99.7% of the data will be within 3 standard deviations of the mean.
- Another name for this rule is the **68-95-99.7 rule**.

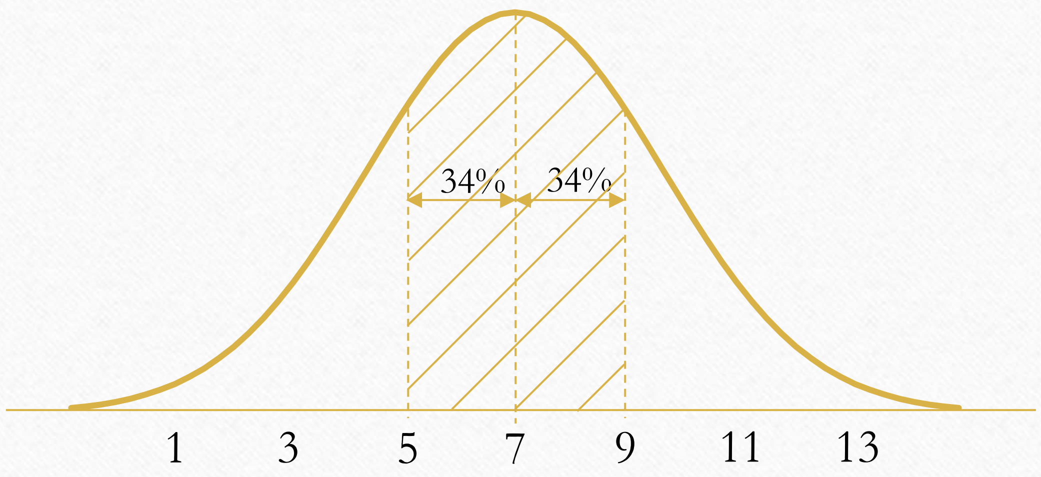


- Let's take an example.
- In the aftermath of the alien invasion, the average weight of a desert cactus is 7 kilograms with a standard deviation of 2 kg.
- Therefore, one standard deviation below the mean is 7−2=5. One standard deviation above the mean is 7+2=9.



- The **Empirical Rule** indicates that the shaded region, between 5 and 9, covers an area of 0.68.
- In other words, we anticipate that 68% of our post-alien invasion desert cacti will have a weight ranging between 5 and 9 kilograms, regardless of the number we observe.

Empirical Rule



- Similarly, 95% of our post-alien invasion desert cacti will have a weight ranging between 3 and 11 kilograms, regardless of the number we observe.
- Similarly, 99.7% of our post-alien invasion desert cacti will have a weight ranging between 1 and 13 kilograms, regardless of the number we observe.

- Suppose any desert cactus weighing over 11 kg could be a surveillance in disguise used by aliens. In that case, you need to remove all those heavy ones, which is represented by the above shaded region.
- Question: What percentage of the cactus will weigh more than 11kg?

- If 95% of the cacti weigh between 3 and 11 kg, then 5% will either weigh more than 11 kg or less than 3 kg.
- Given the symmetric nature of the distribution, we can deduce that 2.5% (Ans) of the cacti will have a weight exceeding 11 kg.

- We can also use the **Empirical Rule** backward to determine if a pattern follows a normal distribution.
  - If the rules are obeyed, then it's a normal distribution. Else, it's not.
- For example, from a survey, 74 out of 85 participants fall within 1 standard deviation of the data. Calculating,

$$\frac{74}{85} = 87.06\%$$

- 87.06% of the data falls within one standard deviation from the average!
- Therefore, it's not a normal distribution.

- Back to combating aliens. Billy gives you a normal distribution diagram representing the battery life of walkie-talkies using data collected for many years. You know that $\mu = 30$ and $\sigma = 4.5$.
- Billy is hoping that there's a 90% chance that the batteries can survive at least 24 hours – i.e. $P(X > 24) = 0.9$. Can you find out whether it's 90%?
- Unfortunately, the **Empirical Rule** can't help you with that. But in the next topic, Z-scores might be able to help you.

The End