PSY121- Statistics in Social Sciences

Measures of Distribution: Variance, Standard Deviation, Standard Error, Skewness, and Kurtosis

Let's remember what SD is ☺ *(standard deviation)*

- Standart sapma, bir veri setindeki değerlerin ortalamadan ne kadar sapma gösterdiğini ölçer. Yani, verilerin "dağılımının genişliğini" gösterir.
- Küçük standart sapma → Veriler ortalamaya yakın, daha tutarlı.
- Büyük standart sapma → Veriler ortalamadan çok uzak, daha dağınık.
- Formula?

X= each data value

$\bar{x}$ = sample mean

n = sample size

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

Variance

Variance/Varyans (σ^2)

- Varyans, standart sapmanın karesi olarak düşünülebilir. Yani, verilerin ortalamadan ne kadar “kareler halinde” sapma gösterdiğini ölçer.
- **Variance can be considered as the square of the standard deviation. In other words, it measures how much the data deviate from the mean in squared units.**
- **Formula:**

$$Variance = \frac{\sum (X_i - \bar{X})^2}{N}$$

Xi = Gözlenen puan (e.g. Math score)
$\bar{X}$ = Mean
N= number of sample/population

Standard Error ($S_{\bar{X}}$ or SE)

- Standart hata, bir örneklem ortalamasının gerçek popülasyon ortalamasından ne kadar sapabileceği!
  - Yani, bir örneklemden bulduğumuz ortalama ile gerçek ortalama arasındaki belirsizliğin ölçüsüdür.
  - Elimizdeki **örnek ortalama**, gerçek **evren ortalamasına (popülasyon ortalamasına)** ne kadar yakın ya da uzak olabilir — bunu söyler.
- Örneklem büyüdükçe, standart hata küçülür (daha güvenilir tahmin). Bu nedenle, araştırmalarda ÖRNEKLEM BÜYÜKLÜĞÜNÜ ARTTIRMAK!!
- Formula?

$$S_{\bar{X}} = \frac{SD}{\sqrt{n}}$$

SD = örneklemin standart sapması
n = örneklem büyüklüğü

Example: Sınıfta 25 öğrencinin notları üzerinden SD ≈ 1.58 ise,

$$SE = \frac{1.58}{\sqrt{25}} = \frac{1.58}{5} \approx 0.316$$

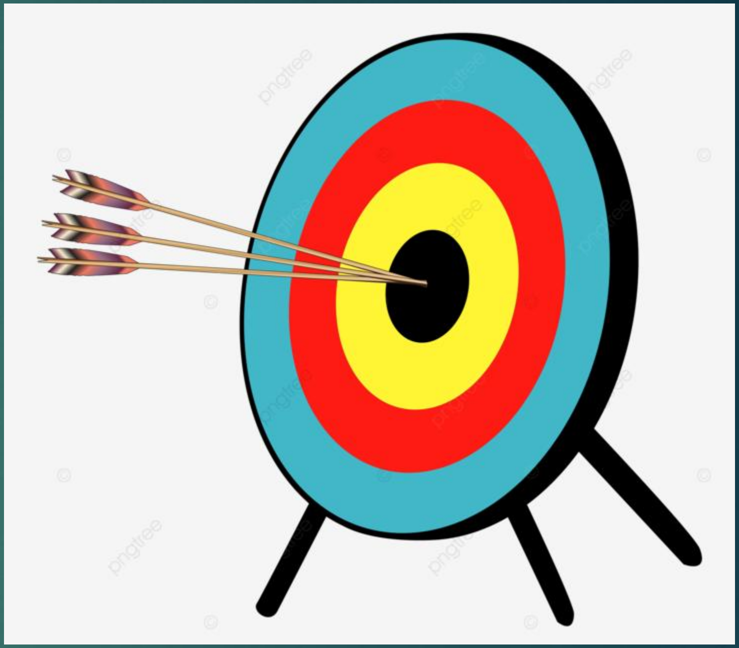
→Bu, örneklem ortalamasının gerçek sınıf ortalamasından yaklaşık ±0.316 puan sapabileceğini gösterir.

Standard Error ($S_{\overline{X}}$ or **SE**)

- Example: Let's say we want to find out the **average sleep duration** of all students at Çağ University.
- But we can't ask everyone, so we take a **sample of 100 students** instead.
- The average sleep time of these 100 students turns out to be **7.2 hours**.
  - If we took another group of 100 students, their average might be **6.9 hours**.
    So, the **sample means** vary a little from one sample to another.
- The amount of this variation is what we call the **standard error**.

Standard Error ($S_{\overline{X}}$ or **SE**)

- Imagine you're shooting arrows at a target.
- **Standard deviation:** How close the arrows are to each other.
- **Standard error:** How close the **average landing point** of the arrows is to the **real center** of the target.

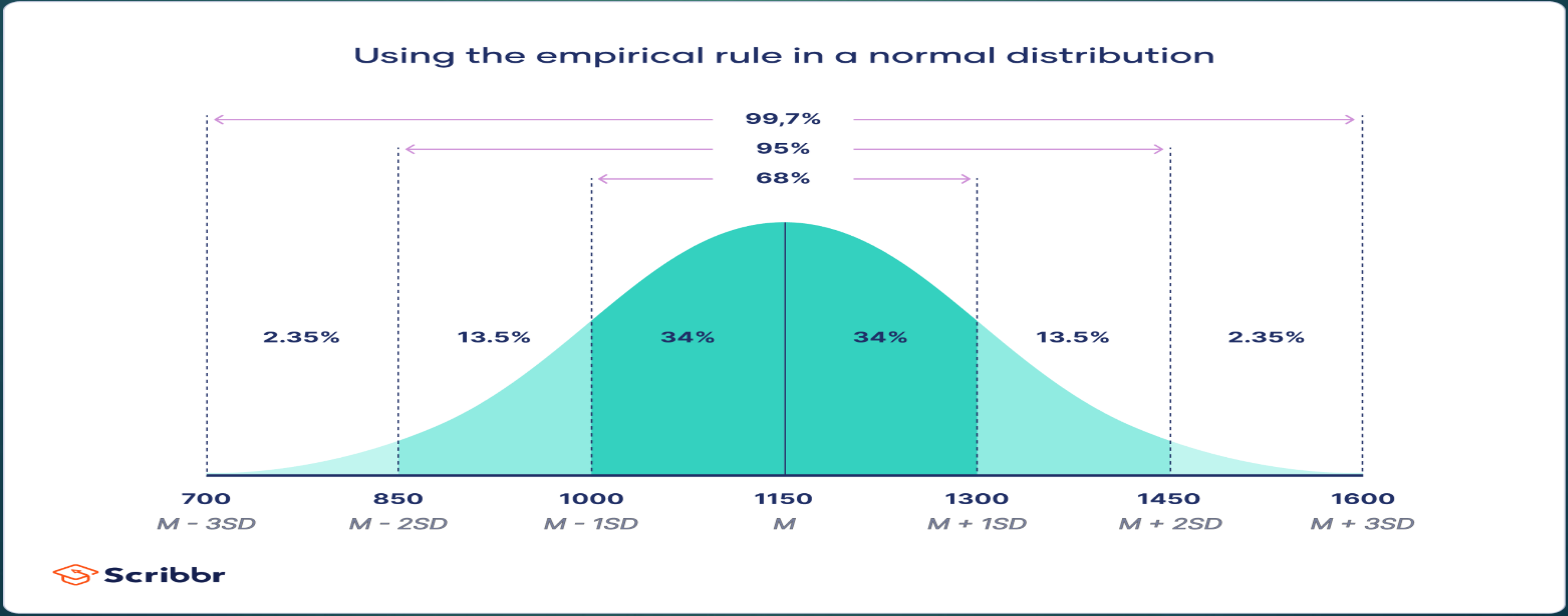


What is Normal Distrubution?

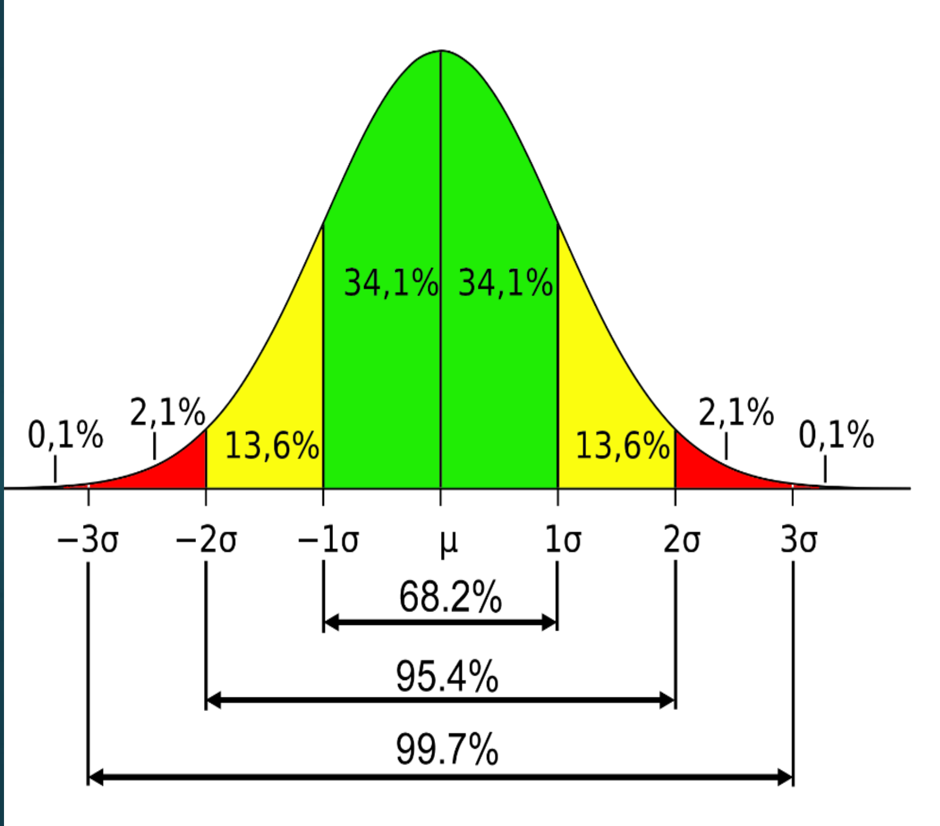
- Generally, it is accepted that the majority of variables observed in the universe display a distribution similar to a bell curve.
- The bell-shaped curve formed by the data related to these variables is called the **normal distribution curve**, and the distribution it represents along the horizontal axis is referred to as the **normal distribution**.
- The **standard normal distribution** has a **mean of 0** and a **standard deviation of 1**.
- The **mode**, **mean**, and **median** are **equal** to each other.
- The **curve is symmetric** with respect to the **vertical axis**.

What is Normal Distrubution?

- Scores tend to cluster around the center.
- The right and left sides of the normal distribution curve extend infinitely and never touch the horizontal axis.
- In other words, the curve is **asymptotic** and takes values in the range **(-∞, +∞)**. – sonsuz bir dağılım gibi…



The Empirical Rule: How Scores Are Distributed in a Normal Curve



- In a **normal distribution**, data are distributed symmetrically around the mean according to the **Empirical Rule (68–95–99.7 Rule):**
- About **68%** of the data fall **within ±1 standard deviation** from the mean.
- About **95.4%** of the data fall **within ±2 standard deviations** from the mean.
- About **99.7%** of the data fall **within ±3 standard deviations** from the mean.

The Empirical Rule: How Scores Are Distributed in a Normal Curve

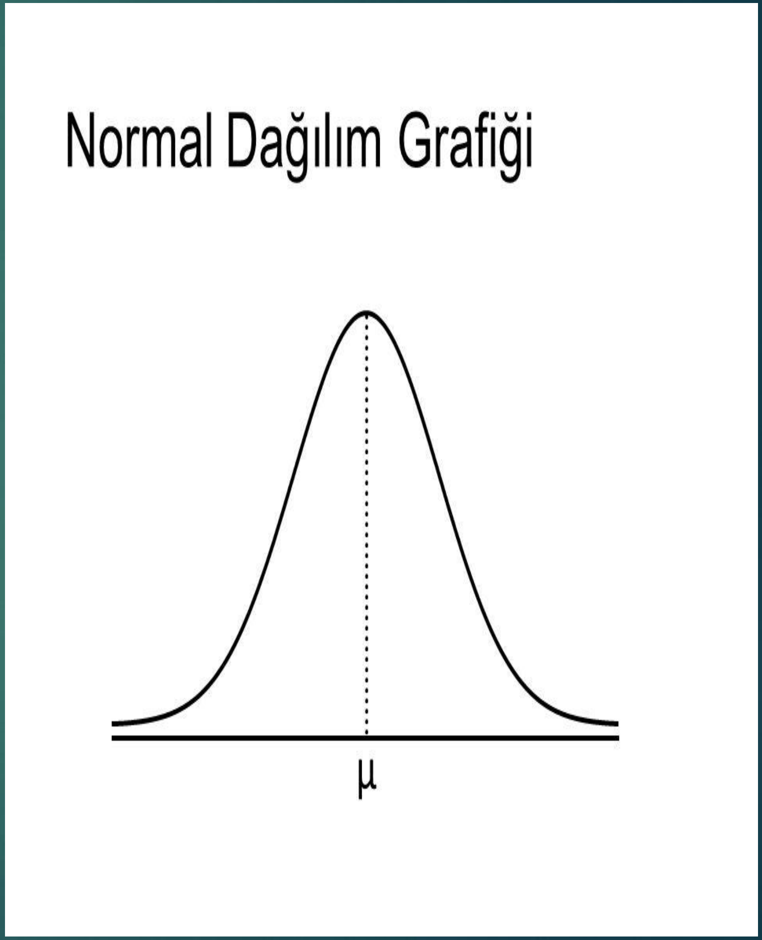
It means that in a **normal distribution** (the bell-shaped curve):

- **68%** of all scores are **close to the mean**, within **one standard deviation** (SO, most people have average scores).
- **95.4%** of all scores are within **two standard deviations**, meaning almost everyone's score is fairly typical.
- **99.7%** of all scores are within **three standard deviations**, so **very few people** have extremely high or low scores.

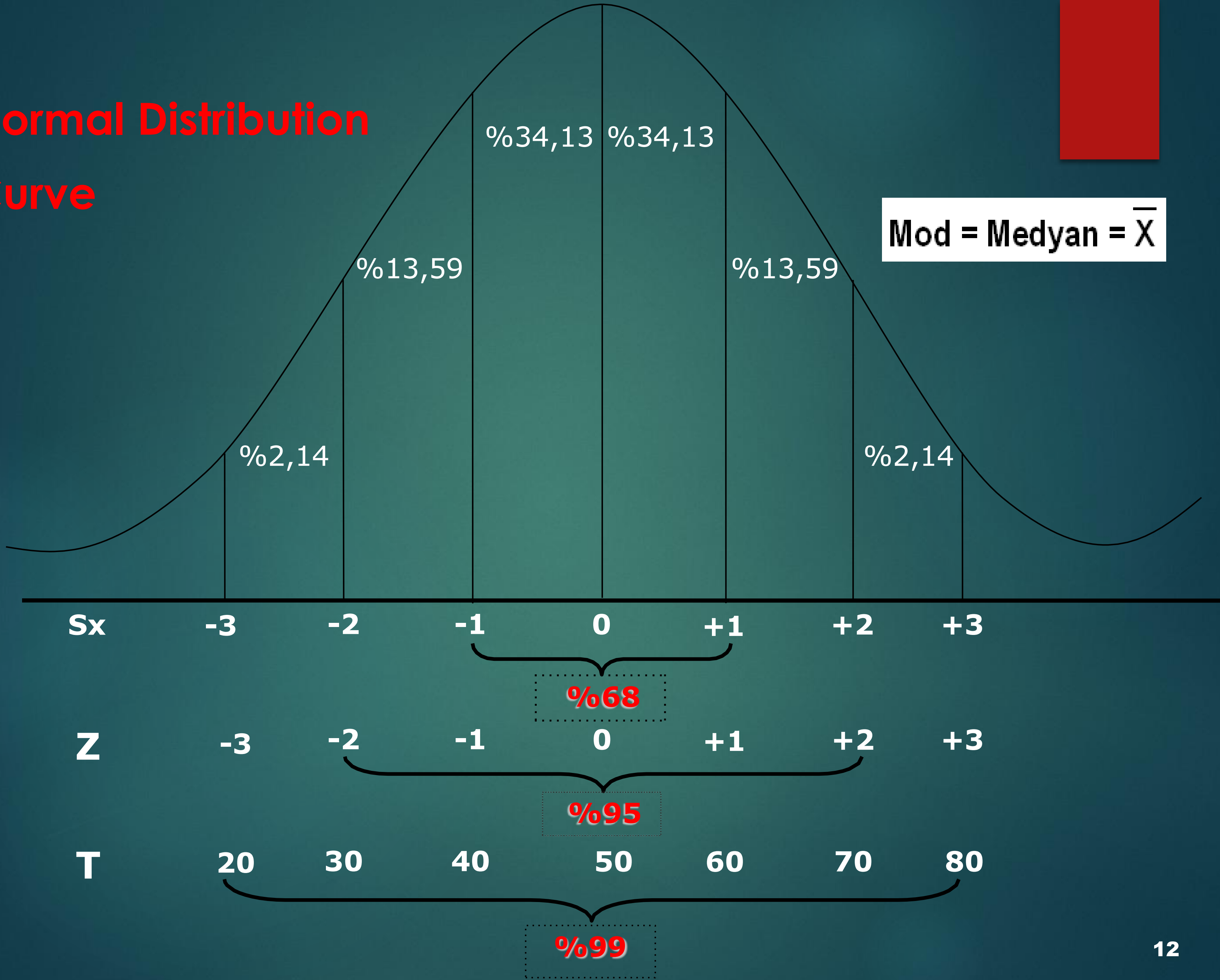
Testing the Normality of a Distribution

- Examining the distribution by plotting its graph.

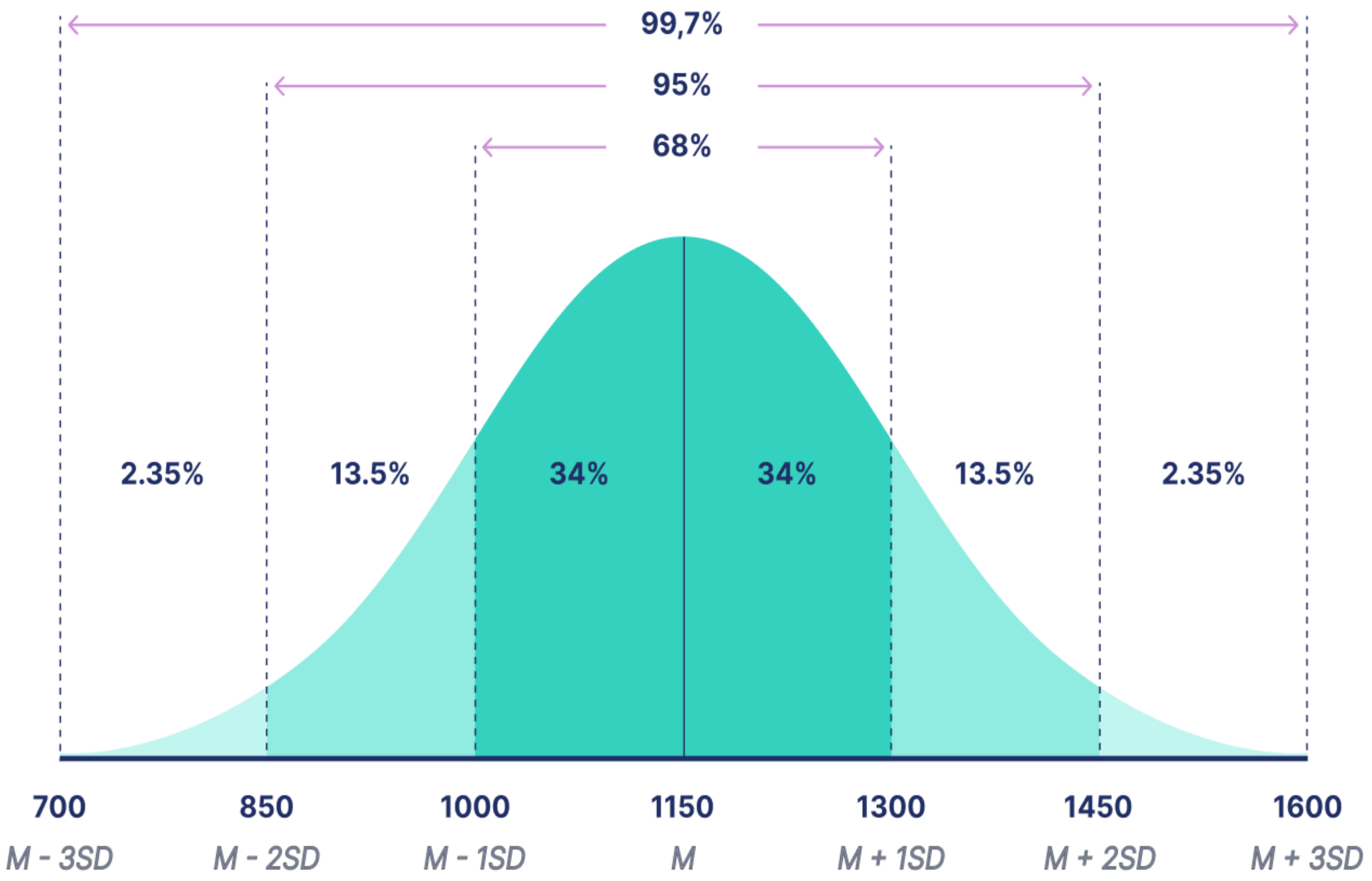
- Analyzing the mean, median, and mode values. In a normal distribution, these values are equal.
- The closer these statistics are to each other, the more the distribution approaches normality; the further they are from each other, the more skewed the distribution becomes.



Normal Distribution Curve

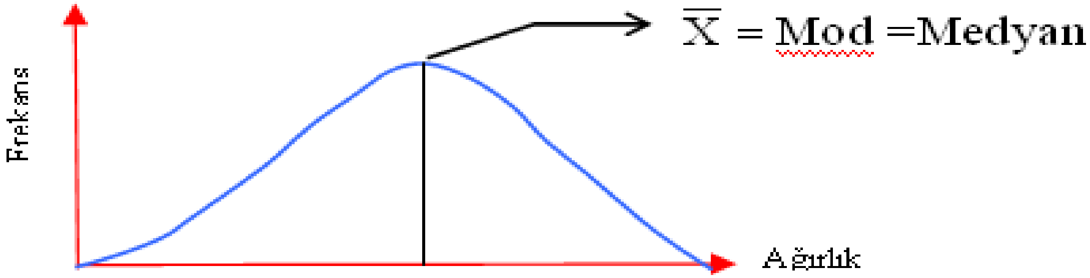


Using the empirical rule in a normal distribution

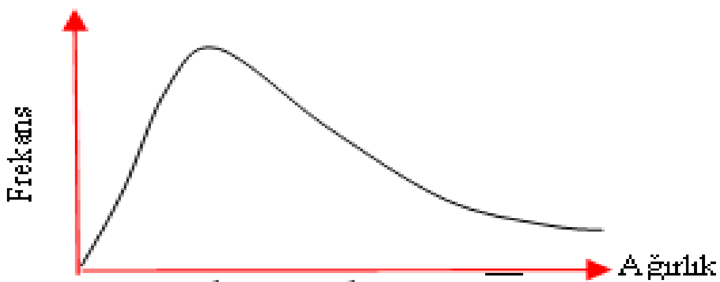


Scribbr

Simetrik dağılım



Simetrik olmayan dağılımlar

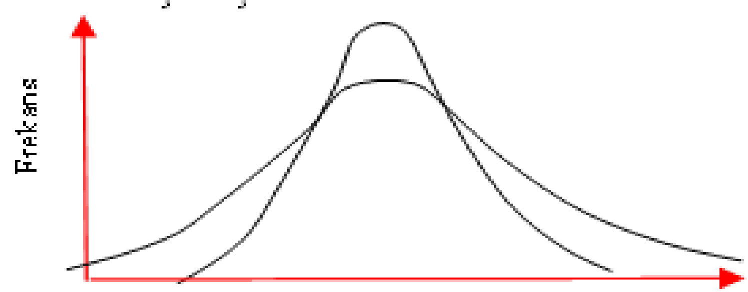


Mod <Medyan < $\overline{X}$

$\overline{X}$ < Medyan <Mod

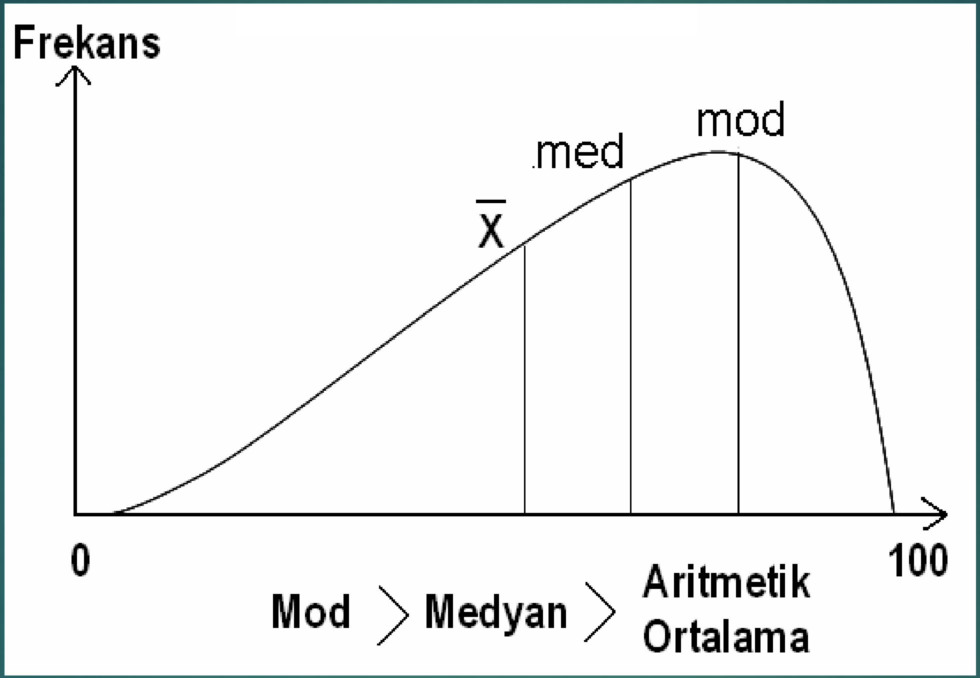
İki dağılımın merkezi eğilimlerinin karşılaştırılması

Değişimleri aynı, mod, medyan ve ortalamaları farklı



Değişimleri farklı, mod, medyan ve ortalamaları aynı

Skewness in a Distribution: Negatively Skewed (Left-Skewed Distribution)

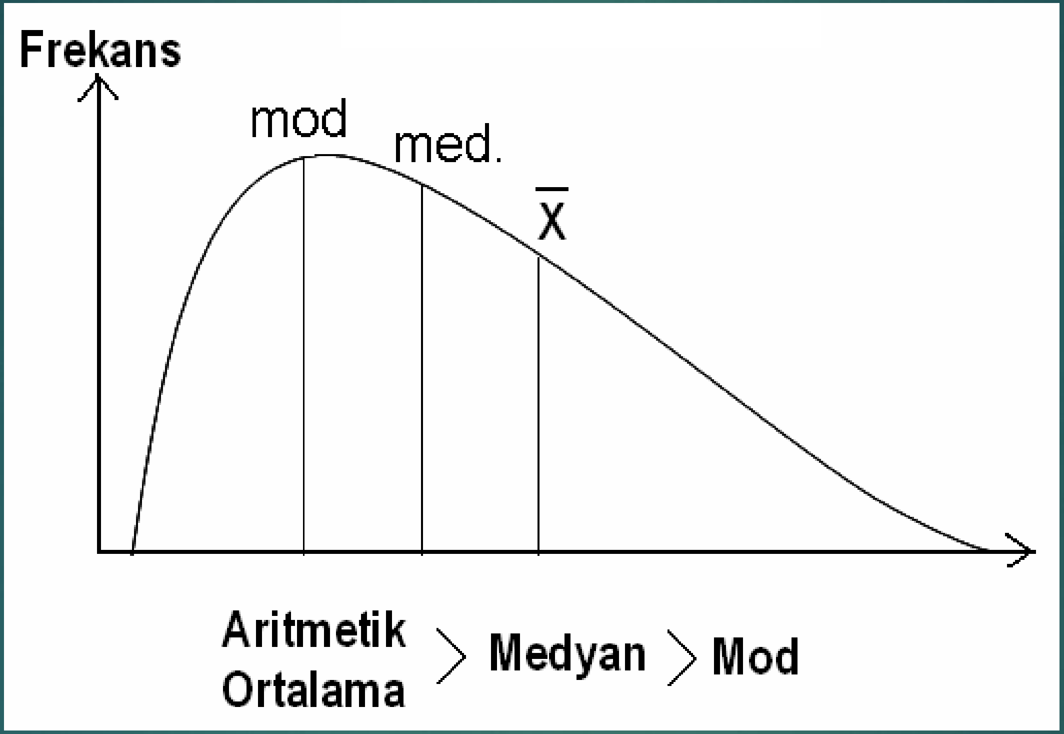


- Most of the scores are **clustered on the right side** of the distribution.

  Negatively skewed (left-skewed): Class performance is high.

  Mean < Median < Mode.
- The questions and the test are easy ☺

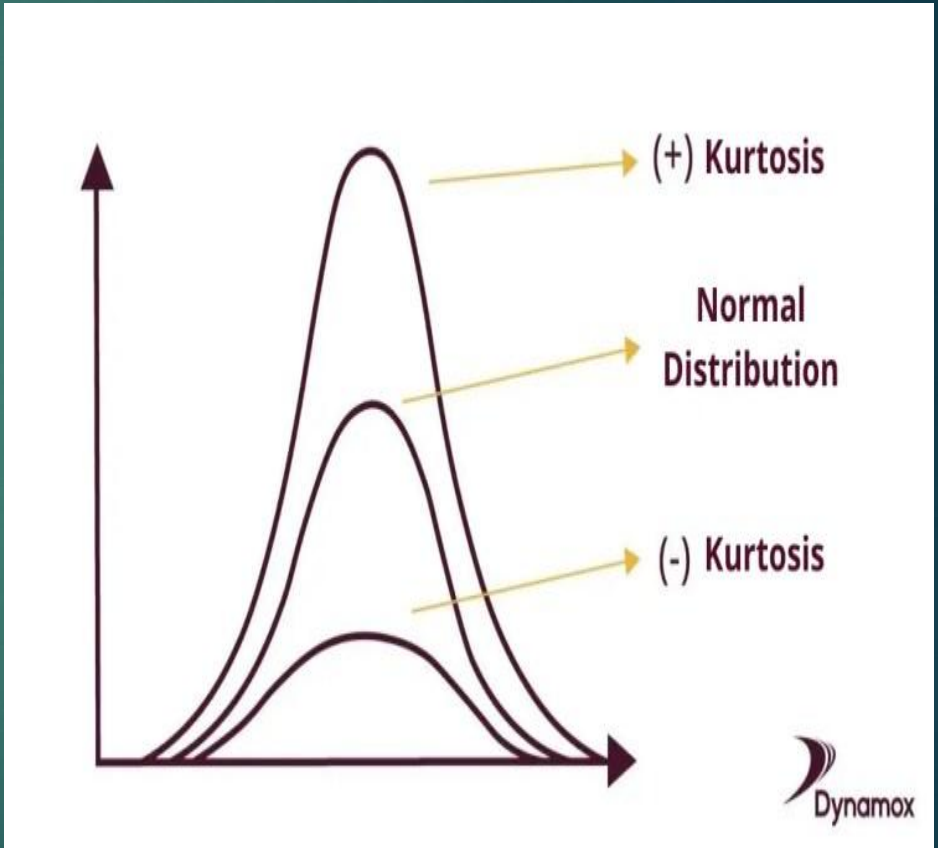
Skewness in a Distribution: Positively Skewed (Right-Skewed Distribution)



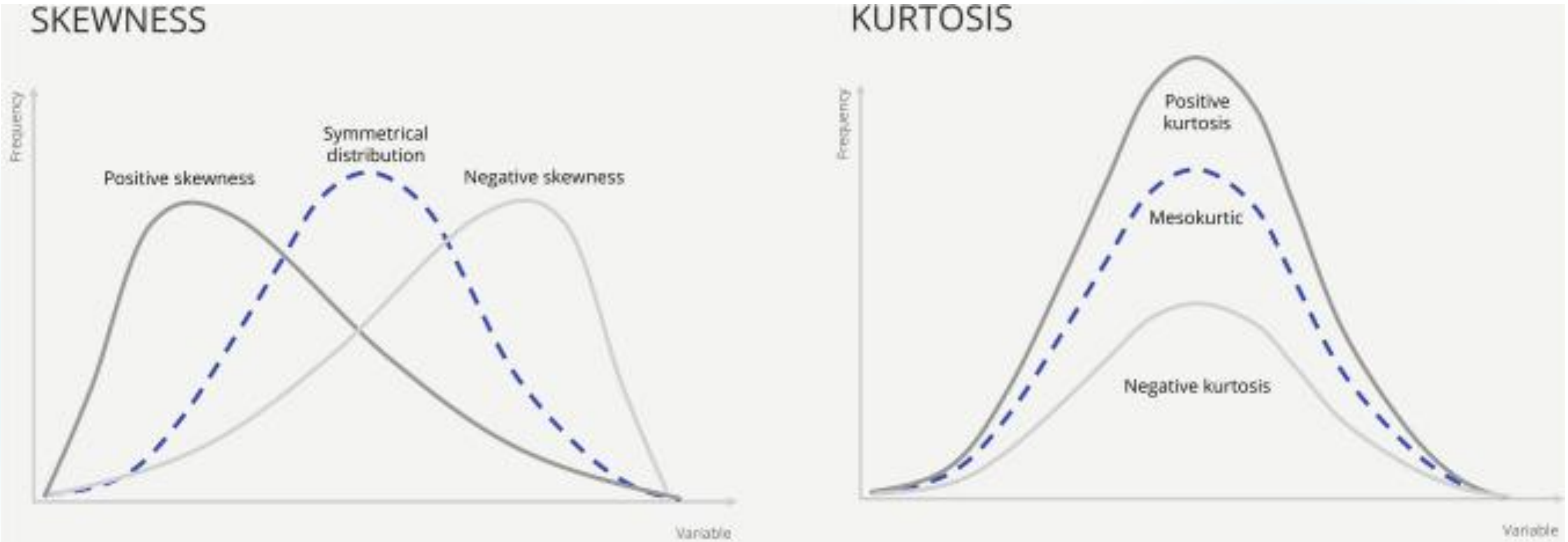
- Most of the scores are **clustered on the left side** of the distribution. **Positively skewed (right-skewed):** <u>Class performance is low.</u>
- **Mode < Median < Mean.**
- **The questions and the test are difficult!!!**

Kurtosis

- Kurtosis describes the **shape** of a distribution in terms of how **peaked or flat** it is compared to a normal distribution.
- In a **perfectly normal distribution**, the kurtosis value is **zero**.
- A **positive kurtosis** means the distribution is **more peaked** than normal — most of the data are close to the mean, but there are also more extreme values…
- A **negative kurtosis** means the distribution is **flatter** than normal — data are more spread out.



Skewness and Kurtosis
(Çarpıklık ve Basıklık)



What is Standard Score?
Z and T scores

- A **standard score** is obtained by subtracting the mean from an observed score and then dividing by the standard deviation. It expresses the score in **units of standard deviation**.
- There are two common types of standard scores: **Z scores** and **T scores**.
- A distribution with a **mean of 0.00** and a **standard deviation of 1.00** is called a **unit normal distribution** or **standard normal distribution**.
- When the results of a test are converted to this standard normal distribution, the resulting scores are called **Z scores**.

Z score is…

- A **Z score** is a **standard score** that tells you how far a particular value is from the mean, measured in **standard deviations**.
- Could be NEGATIVE…

Which means;

- Z = 0 → the score is exactly at the mean! ☺
- Z = +1 → the score is **1 standard deviation above the mean**
- Z = -2 → the score is **2 standard deviations below the mean**

$$Z = \frac{X - \overline{X}}{SS}$$

Z : Z-Puanı
X : Bir öğrencinin puanı
$\overline{X}$: Puanlar dağılımının ortalaması
SS : Puanlar dağılımının standart sapması

Let's have an example…

- Sena took her mathematics exam. The class average (mean) was **70 points**, and the standard deviation was **10 points.** Sena scored **85 points** on the exam.

What is Sena's **Z score**?

Formula:

- $Z = \frac{Score - Mean}{Standard\ Deviation}$
- $Z = \frac{85-70}{10} = 1.5$

Interpretation:

- Sena scored **1.5 standard deviations above the class mean**. Her score is higher than most of her classmates.

Let's have another example… ☺

1. According to the table, in which test did Ali perform better?
2. In which two tests are his performances equal?

	1.test	2.test	3.test	4.test	5.test
Ali'nin puanı	44	80	68	44	60
Ortalama	50	80	60	30	50
Std. sapma	6	5	8	14	5

1. Test için Z puanı= ?
2. Test için Z puanı= ?
3. Test için Z puanı= ?
4. Test için Z puanı= ?
5. Test için Z puanı= ?

What about T Score?

- A T score is a **rescaled version of a Z score** that is easier to read.
- Typically, it is set with a **mean of 50** and a **standard deviation of 10**.
- The purpose is to **avoid negative** **or** **decimal numbers** and make scores more interpretable.

$$T = 50 + \left[\frac{X - \overline{X}}{SS}\right].10$$

T : T-Puanı
X : Bir öğrencinin puanı
$\overline{X}$: Puanlar dağılımının ortalaması
SS: Puanlar dağılımının standart sapması

Formülde görüldüğü gibi parantez içerisindeki bölüm Z'ye eşittir. O halde Z puanları hesaplanmış ise, formül aşağıdaki gibi yazılabilir.

$$T_i = 10\,Z_i + 50$$

Case Study: Comparing Exam Performances

Scenario:

- The psychology department at Sunshine High School conducted three different exams for the same group of students: **Memory Test**, **Attention Test**, and **Problem-Solving Test**. The results are normally distributed.

Exam	Class Mean	Standard Deviation	Student's Score (Ali)
Memory	70	8	85
Attention	65	5	72
Problem-Solving	80	10	85

Case Study: Comparing Exam Performances

1. Please, calculate Ali's **Z score** for each test.
2. Convert the **Z scores into T scores** (mean = 50, SD = 10).
3. Determine in which test Ali performed best relative to his classmates.
4. Identify if Ali's performance is equal in any two tests when standardized.
5. **Discuss:** What does this tell you about comparing raw scores across tests with different means and standard deviations?

Exam	Class Mean	Standard Deviation	Student's Score (Ali)
Memory Exam	70	8	85
Attention Exam	65	5	72
Problem-Solving Exam	80	10	85

Case Study: Comparing Exam Performances (Answers)

Memory Test:

- $Z = \frac{85-70}{8} = 1.875\,, T = 50 + (1.875 \times 10) = 68.75$

Attention Test:

- $Z = \frac{72-65}{5} = 1.4\,, T = 50 + (1.4 \times 10) = 64$

Problem-Solving Test:

- $Z = \frac{85-80}{10} = 0.5\,, T = 50 + (0.5 \times 10) = 55$

Case Study: Comparing Exam Performances (Answers)

Analysis:

- Ali performed **best in the Memory Test** relative to his classmates (highest Z and T scores).
- His performance in the **Attention Test** is closer to Memory than Problem-Solving, but **no two tests have exactly the same standardized performance**.
- This shows why **standardizing scores is important**: raw scores alone don't account for differences in difficulty or spread between tests.

Any questions???