0.1 Learning Objectives

By the end of this lecture, you will be able to:

- Understand the normal (Gaussian) distribution and its properties
- Recognize when to use the normal distribution
- Standardize data using z-scores
- Use the standard normal distribution to compute probabilities

0.3 The Normal Distribution

The **normal distribution** (also called **Gaussian distribution**) is one of the most important continuous distributions in probability and statistics.

0.3 The Normal Distribution

The **normal distribution** (also called **Gaussian distribution**) is one of the most important continuous distributions in probability and statistics.

If X has a Normal distribution with mean μ and variance σ^2:

$$X \sim \mathrm{Norm}(\mu, \sigma^2) = \mathrm{N}(\mu, \sigma^2)$$

0.3 The Normal Distribution

The **normal distribution** (also called **Gaussian distribution**) is one of the most important continuous distributions in probability and statistics.

If X has a Normal distribution with mean μ and variance σ^2:

$$X \sim \mathrm{Norm}(\mu, \sigma^2) = \mathrm{N}(\mu, \sigma^2)$$

Probability Density Function:

$$\mathrm{PDF}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

0.4 Normal Distribution Parameters

$X \sim \mathrm{Norm}(\mu, \sigma^2)$ has two parameters:

0.4 Normal Distribution Parameters

$X \sim \text{Norm}(\mu, \sigma^2)$ has two parameters:

Mean (μ):

- Center of the distribution
- The peak of the bell curve
- $E(X) = \mu$

0.4 Normal Distribution Parameters

$X \sim \mathrm{Norm}(\mu, \sigma^2)$ has two parameters:

Mean (μ):

- Center of the distribution
- The peak of the bell curve
- $E(X) = \mu$

Variance (σ^2):

- Spread of the distribution
- Larger variance = wider bell curve
- $\mathrm{Var}(X) = \sigma^2$
- Standard deviation: $\sigma = \sqrt{\mathrm{Var}(X)}$

0.5 Visualizing the Normal Distribution

The normal distribution has a characteristic “bell curve” shape:

0.5 Visualizing the Normal Distribution

0.5 Visualizing the Normal Distribution

The normal distribution has a characteristic "bell curve" shape:

- Symmetric around the mean μ
- Most values cluster near the mean
- Probability decreases as you move away from mean
- Tails extend to $\pm\infty$ (but with very low probability)

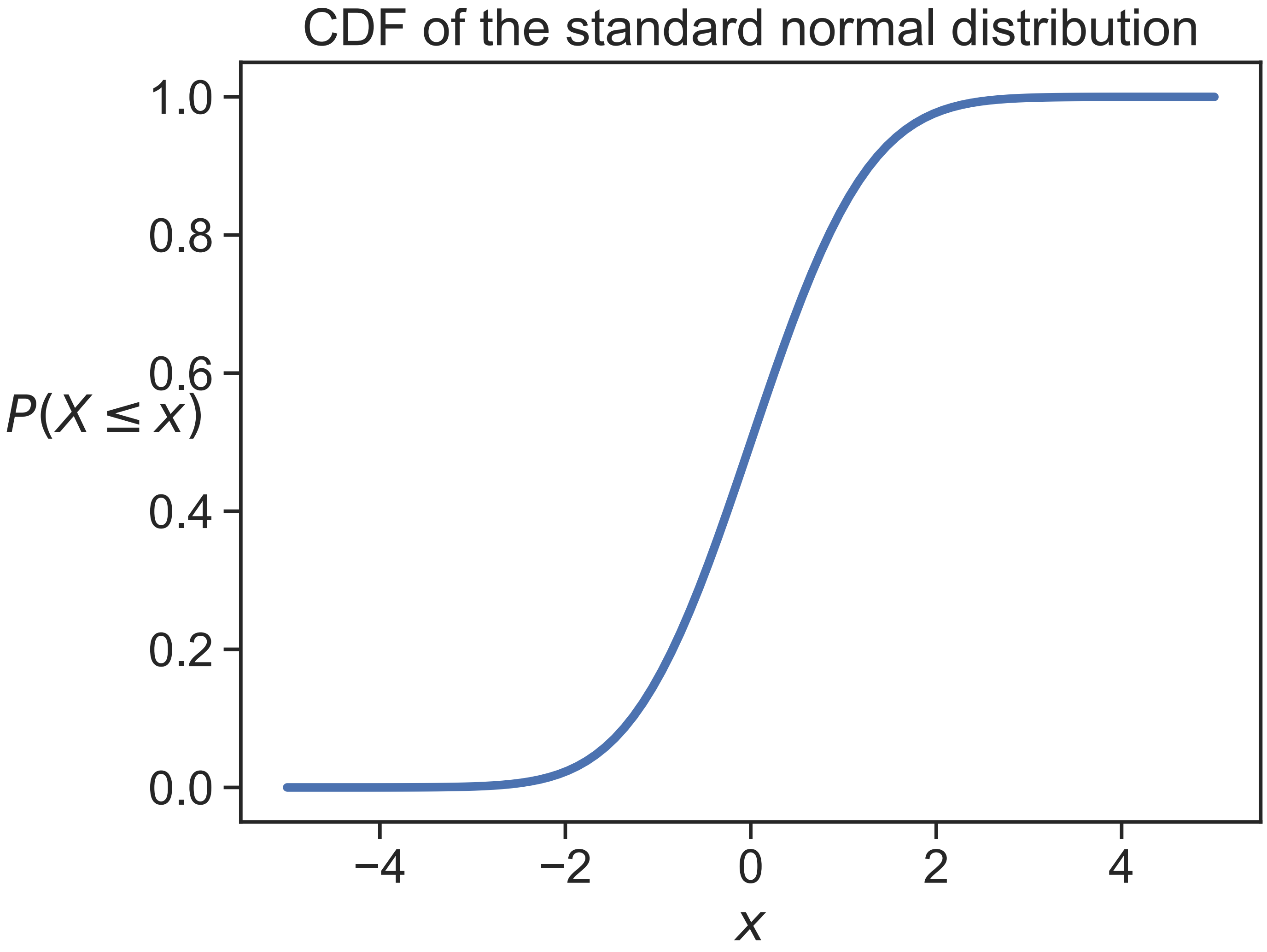
0.5 Visualizing the Normal Distribution

0.5 Visualizing the Normal Distribution

The normal distribution has a characteristic “bell curve” shape:

- Symmetric around the mean μ
- Most values cluster near the mean
- Probability decreases as you move away from mean
- Tails extend to $\pm\infty$ (but with very low probability)

0.5 Visualizing the Normal Distribution



1. Visualizing Data Distributions

1.1 Median, Quantiles, and Quartiles

Median: 50% of the observations of X are less than or equal to $\mathrm{med}(X)$

1.1 Median, Quantiles, and Quartiles

Median: 50% of the observations of X are less than or equal to $\mathrm{med}(X)$

Quantile: q% of the observations of X are less than or equal to $Q_q(X, q)$

1.1 Median, Quantiles, and Quartiles

Median: 50% of the observations of X are less than or equal to $\mathrm{med}(X)$

Quantile: q% of the observations of X are less than or equal to $Q_q(X, q)$

Quartile: 25% of the observations of X are less than or equal to the **first quartile**

1.1 Median, Quantiles, and Quartiles

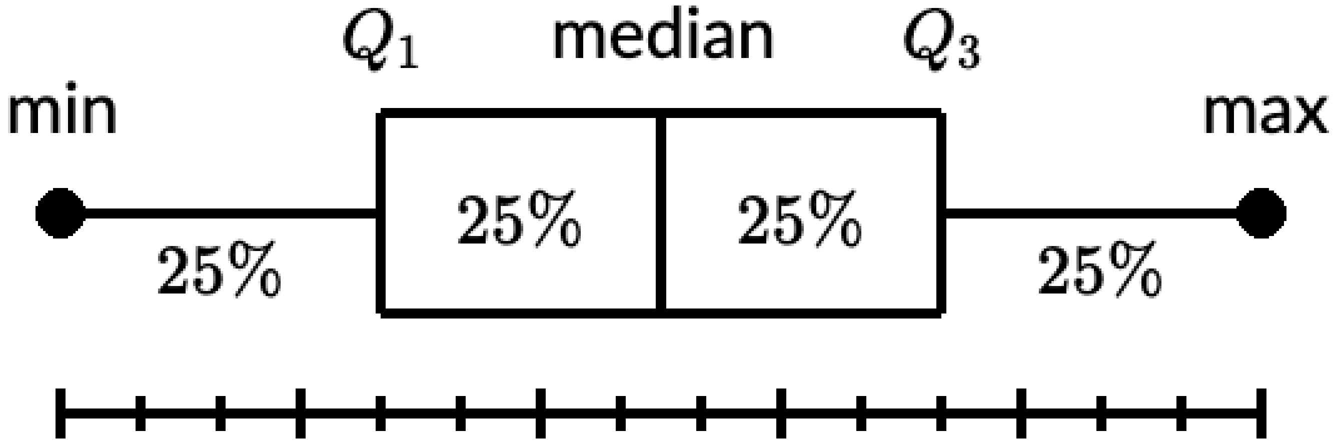
Median: 50% of the observations of X are less than or equal to $\mathrm{med}(X)$

Quantile: q% of the observations of X are less than or equal to $Q_q(X, q)$

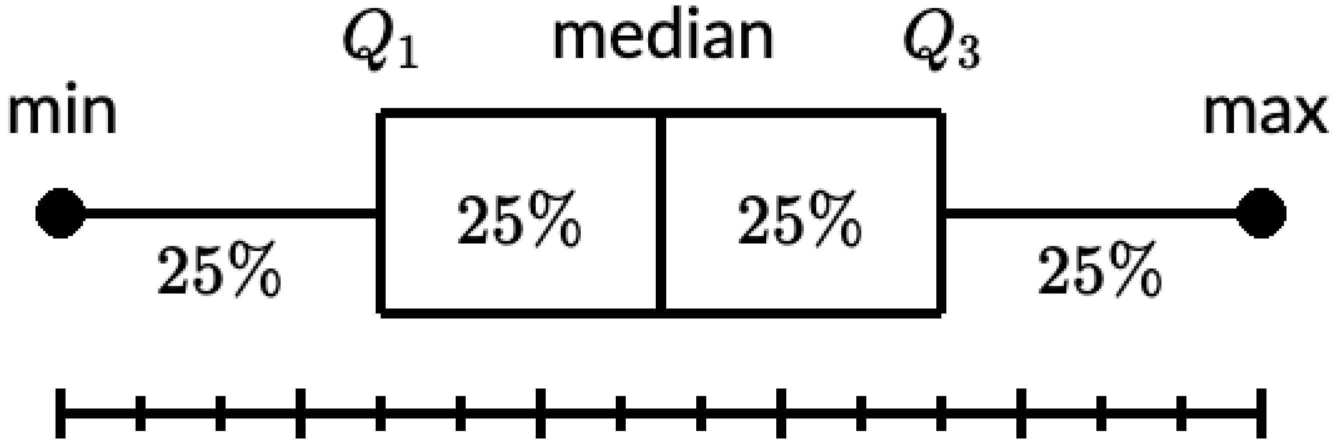
Quartile: 25% of the observations of X are less than or equal to the **first quartile**

- Second quartile = median
- Third quartile = 75th percentile = $Q_{75}(X)$
- **Inter-Quartile Range (IQR)** = $Q_{75}(X) - Q_{25}(X)$

1.2 Visualization: Boxplot

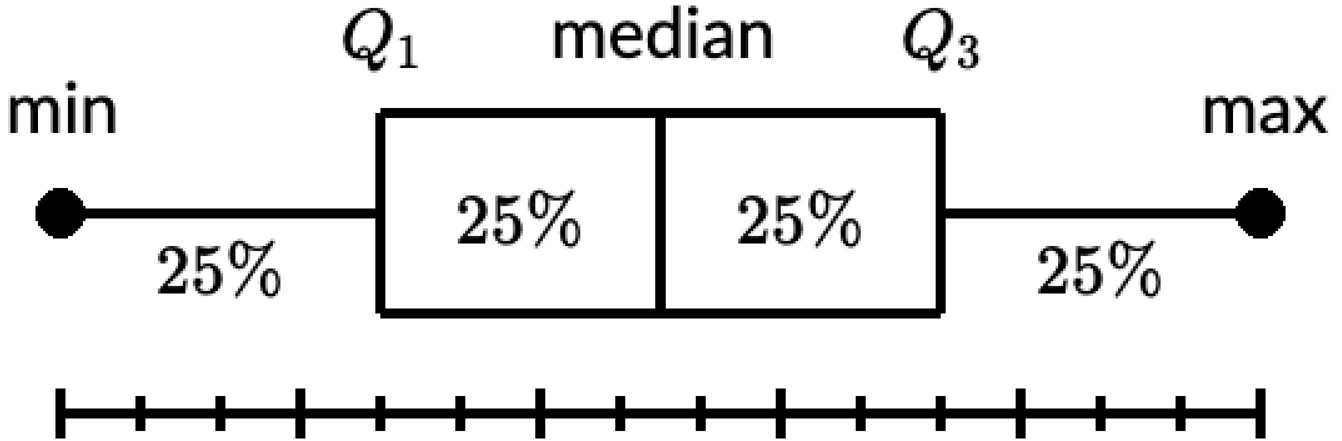


1.2 Visualization: Boxplot



Inter-Quartile Range = $Q_{75}(X) - Q_{25}(X)$

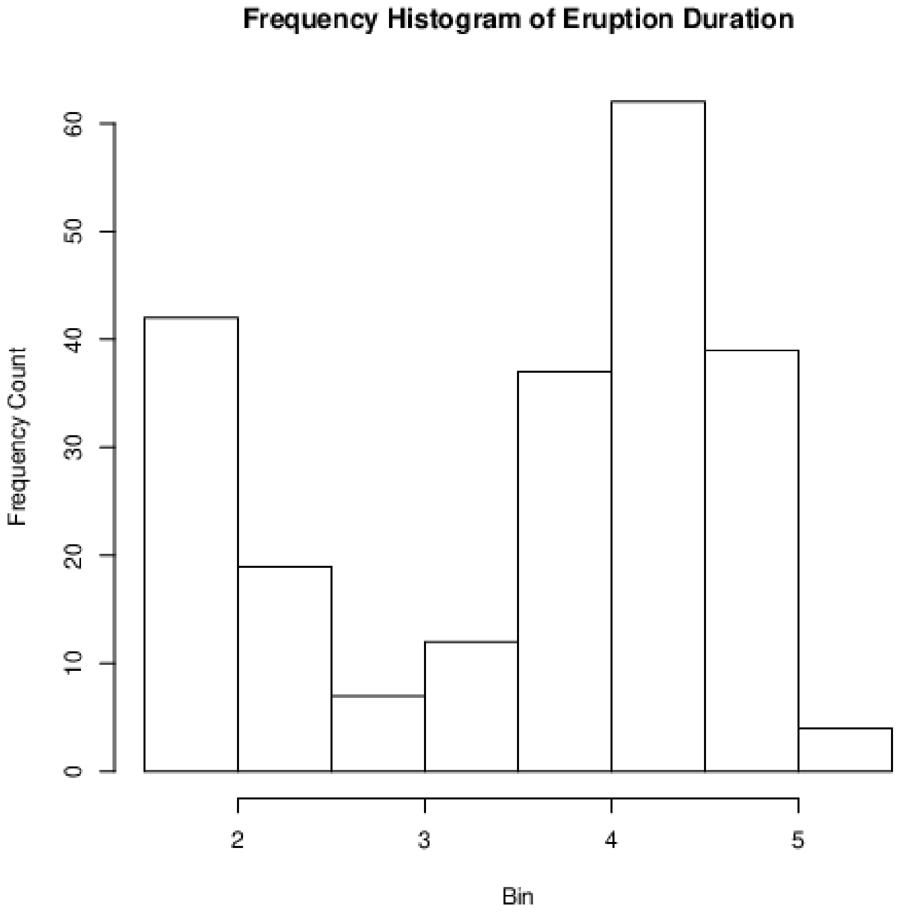
1.2 Visualization: Boxplot



Inter-Quartile Range = $Q_{75}(X) - Q_{25}(X)$

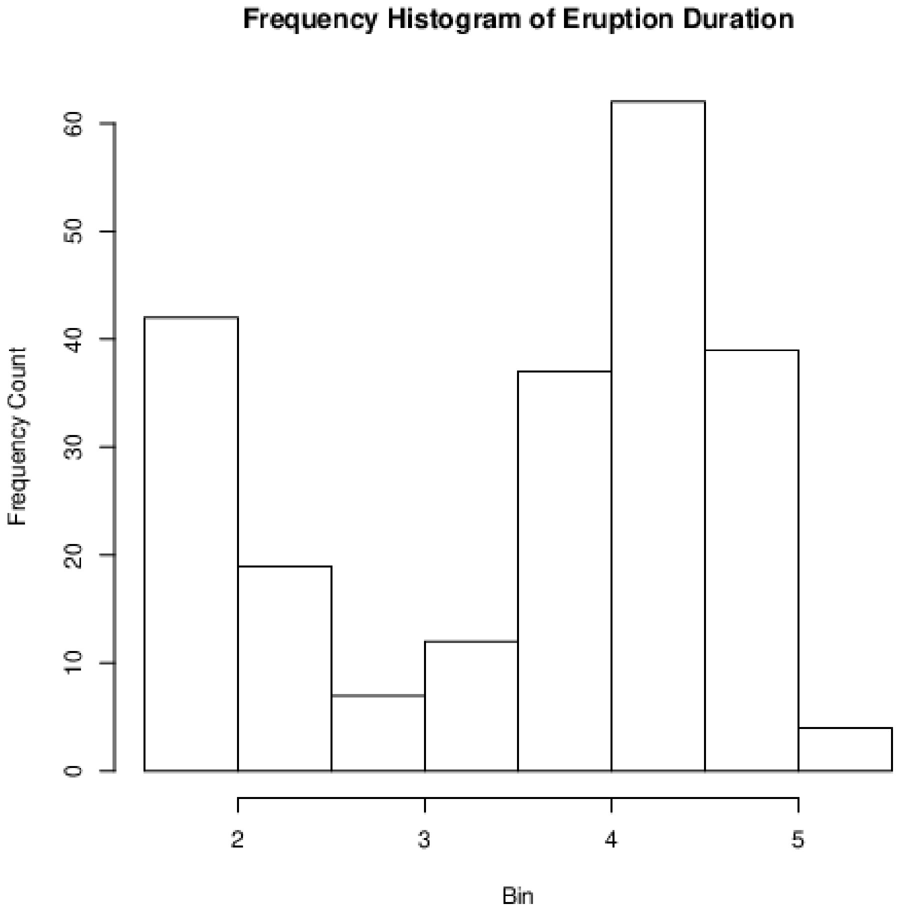
Boxplots show the distribution of data through quartiles and identify outliers.

1.3 Visualization: Histogram



A **histogram** displays the distribution of data by dividing it into bins and showing the frequency of observations in each bin.

1.3 Visualization: Histogram

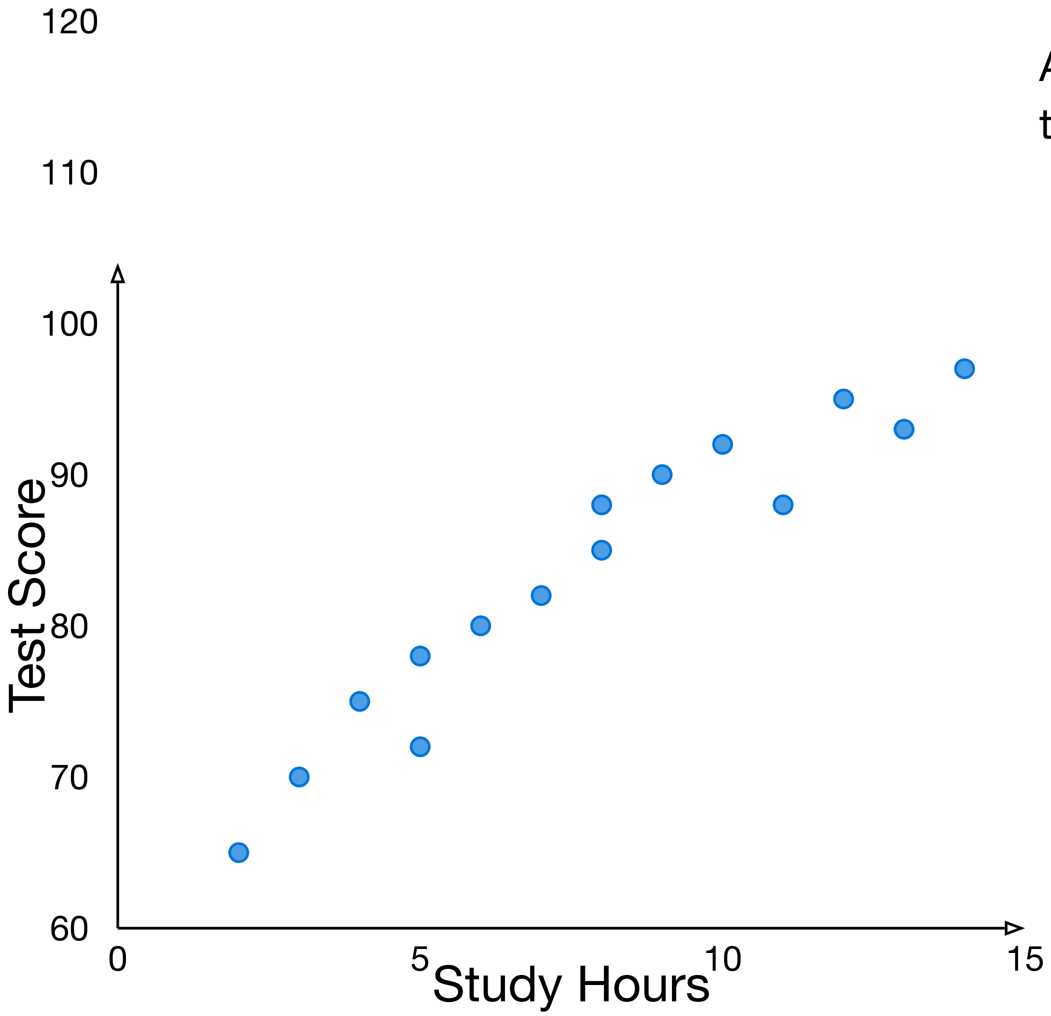


A **histogram** displays the distribution of data by dividing it into bins and showing the frequency of observations in each bin.

Key features:

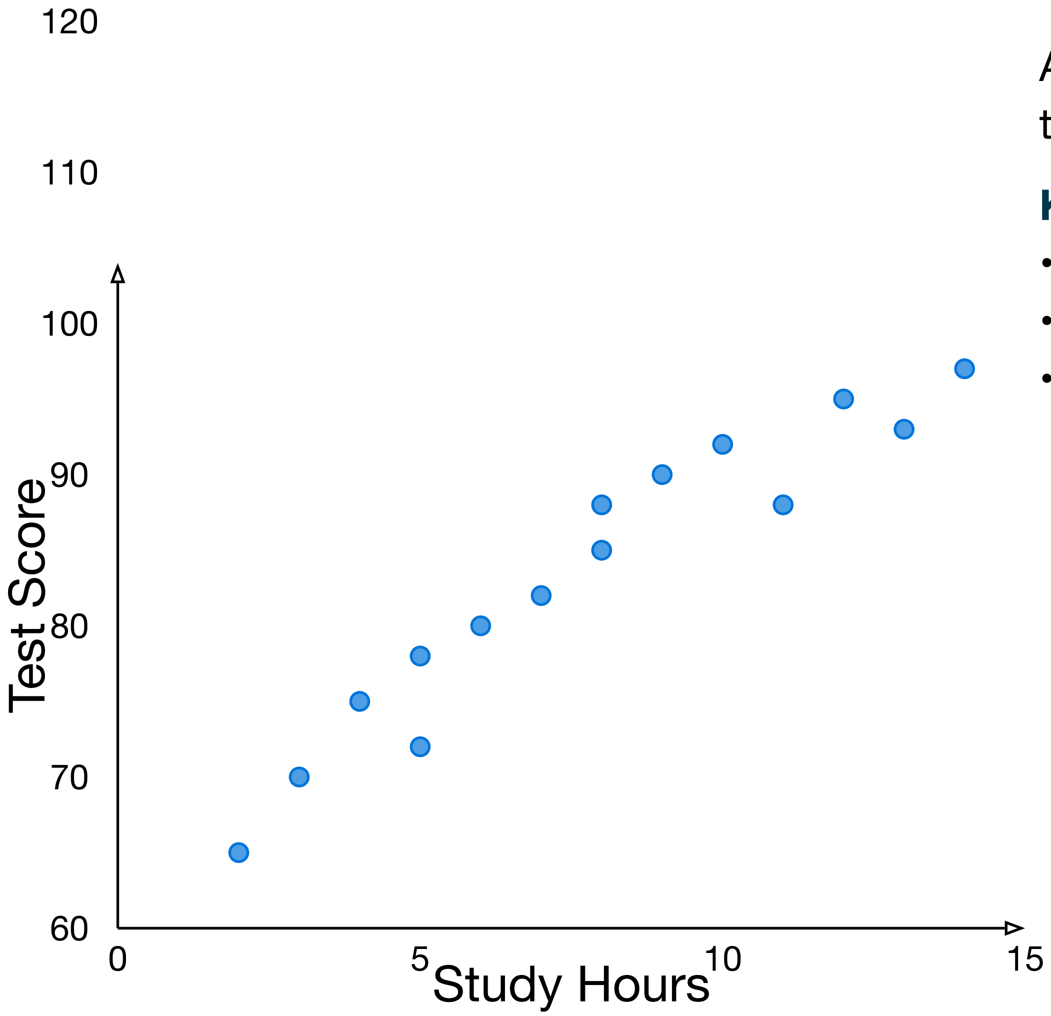
- X-axis: range of values (bins)
- Y-axis: frequency or count
- Shape reveals the distribution pattern

1.4 Visualization: Scatter Plot



A **scatter plot** displays the relationship between two continuous variables.

1.4 Visualization: Scatter Plot

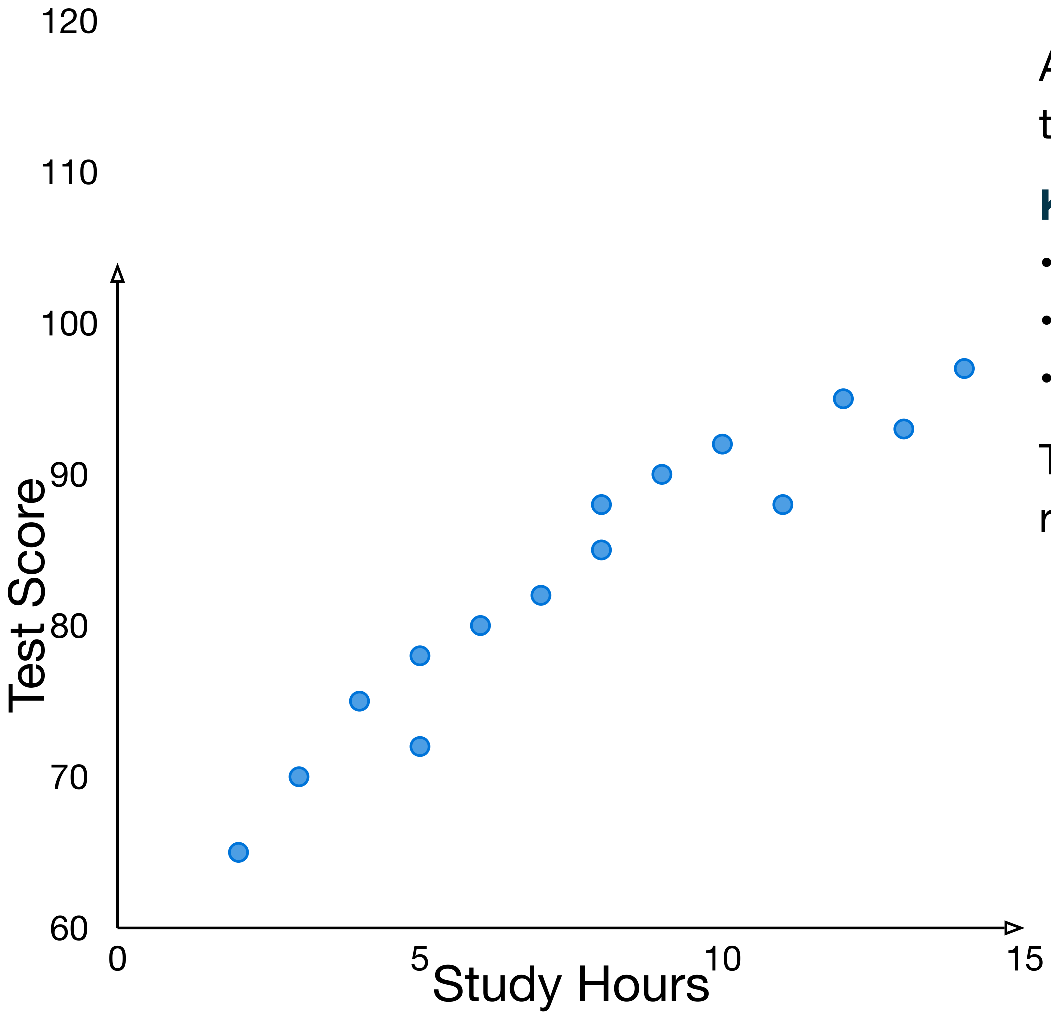


A **scatter plot** displays the relationship between two continuous variables.

Key features:

- Each point represents one observation
- Shows correlation patterns
- Useful for identifying trends and outliers

1.4 Visualization: Scatter Plot



A **scatter plot** displays the relationship between two continuous variables.

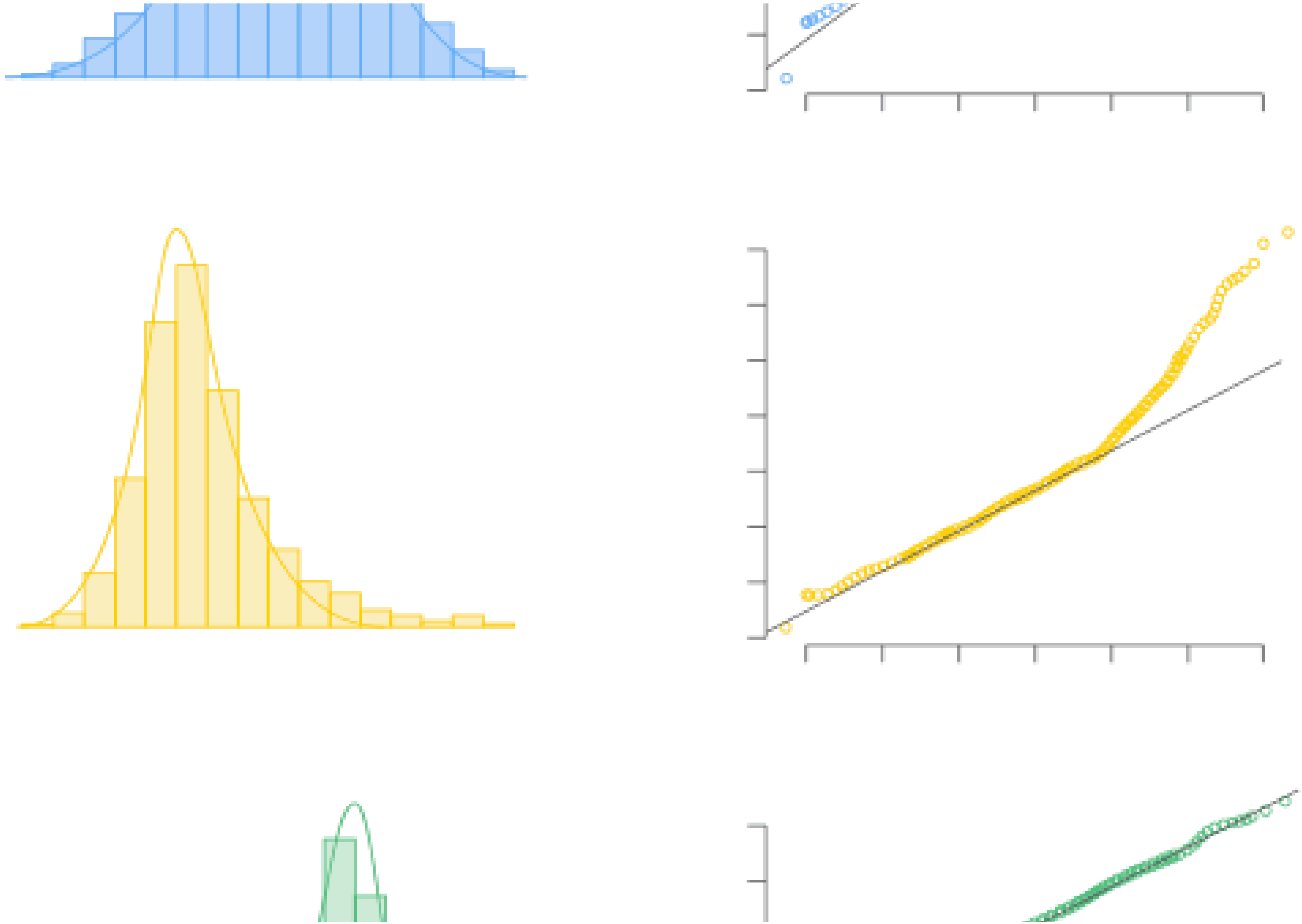
Key features:

- Each point represents one observation
- Shows correlation patterns
- Useful for identifying trends and outliers

This example shows study hours vs. test scores, revealing a positive correlation.

1.5 Quantile-Quantile Plot

Right-skewed data



2. Back to the Normal Distribution

2.1 Why is the Normal Distribution Important?

2.1 Why is the Normal Distribution Important?

The Central Limit Theorem:

The sum of many independent and identically distributed (i.i.d.) random variables approaches a Normal distribution, regardless of the original distribution.

2.1 Why is the Normal Distribution Important?

The Central Limit Theorem:

The sum of many independent and identically distributed (i.i.d.) random variables approaches a Normal distribution, regardless of the original distribution.

This means that many real-world phenomena that result from combining many small independent effects follow a normal distribution:

- Heights of people
- Measurement errors
- Test scores (under certain conditions)
- Stock market returns (approximately)

2.2 Example: Heights

Suppose adult male heights in the US follow a normal distribution with:

- Mean: $\mu = 70$ inches
- Standard deviation: $\sigma = 3$ inches

2.2 Example: Heights

Suppose adult male heights in the US follow a normal distribution with:

- Mean: $\mu = 70$ inches
- Standard deviation: $\sigma = 3$ inches

So $X \sim \mathrm{Norm}(70, 9)$ where X is height in inches.

2.2 Example: Heights

Suppose adult male heights in the US follow a normal distribution with:

- Mean: $\mu = 70$ inches
- Standard deviation: $\sigma = 3$ inches

So $X \sim \mathrm{Norm}(70, 9)$ where X is height in inches.

What is the probability that a randomly selected adult male is between 67 and 73 inches tall?

2.3 Computing Probabilities: The Challenge

For the normal distribution, we cannot compute the CDF in closed form.

2.3 Computing Probabilities: The Challenge

For the normal distribution, we cannot compute the CDF in closed form.

$$\text{CDF}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

2.3 Computing Probabilities: The Challenge

For the normal distribution, we cannot compute the CDF in closed form.

$$\mathrm{CDF}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

This integral has no elementary closed-form solution!

2.3 Computing Probabilities: The Challenge

For the normal distribution, we cannot compute the CDF in closed form.

$$\mathrm{CDF}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

This integral has no elementary closed-form solution!

Solutions:

- Use numerical methods (computers do this)
- Use standardization + lookup tables
- Use statistical software (R, Python, etc.)

2.5 The Standard Normal Distribution

The **standard normal distribution** is a special case where:

- $\mu = 0$ (mean is zero)
- $\sigma^2 = 1$ (variance is one)

2.5 The Standard Normal Distribution

The **standard normal distribution** is a special case where:

- $\mu = 0$ (mean is zero)
- $\sigma^2 = 1$ (variance is one)

We write: $Z \sim \mathrm{Norm}(0, 1) = \mathrm{N}(0, 1)$

2.5 The Standard Normal Distribution

The **standard normal distribution** is a special case where:

- $\mu = 0$ (mean is zero)
- $\sigma^2 = 1$ (variance is one)

We write: $Z \sim \mathrm{Norm}(0, 1) = \mathrm{N}(0, 1)$

PDF of standard normal:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

2.5 The Standard Normal Distribution

The **standard normal distribution** is a special case where:

- $\mu = 0$ (mean is zero)
- $\sigma^2 = 1$ (variance is one)

We write: $Z \sim \mathrm{Norm}(0, 1) = \mathrm{N}(0, 1)$

PDF of standard normal:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

CDF of standard normal:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t) dt$$

2.6 Converting to Standard Normal

Any normal distribution can be transformed into a standard normal distribution using **standardization**:

2.6 Converting to Standard Normal

Any normal distribution can be transformed into a standard normal distribution using **standardization**:

$$Z = \frac{X - \mu}{\sigma}$$

2.6 Converting to Standard Normal

Any normal distribution can be transformed into a standard normal distribution using **standardization**:

$$Z = \frac{X - \mu}{\sigma}$$

If $X \sim \mathrm{Norm}(\mu, \sigma^2)$, then $Z \sim \mathrm{Norm}(0, 1)$

2.6 Converting to Standard Normal

Any normal distribution can be transformed into a standard normal distribution using **standardization**:

$$Z = \frac{X - \mu}{\sigma}$$

If $X \sim \mathrm{Norm}(\mu, \sigma^2)$, then $Z \sim \mathrm{Norm}(0, 1)$

This is incredibly useful because:

- We only need tables/functions for ONE distribution ($\mathrm{Norm}(0, 1)$)
- All normal distribution probabilities can be computed using standardization

2.7 Z-Scores

The standardized value $Z = \frac{X-\mu}{\sigma}$ is called a **Z-score**.

2.7 Z-Scores

The standardized value $Z = \frac{X-\mu}{\sigma}$ is called a **Z-score**.

Interpretation:

- Z-score tells us how many standard deviations away from the mean a value is
- $Z = 0$: value equals the mean
- $Z = 1$: value is 1 standard deviation above the mean
- $Z = -2$: value is 2 standard deviations below the mean

2.8 Example: Computing a Z-Score

Recall: Adult male heights $X \sim \mathrm{Norm}(70, 9)$ (mean 70 inches, std dev 3 inches).

2.8 Example: Computing a Z-Score

Recall: Adult male heights $X \sim \mathrm{Norm}(70, 9)$ (mean 70 inches, std dev 3 inches).

What is the Z-score for a height of 76 inches?

2.8 Example: Computing a Z-Score

Recall: Adult male heights $X \sim \mathrm{Norm}(70, 9)$ (mean 70 inches, std dev 3 inches).

What is the Z-score for a height of 76 inches?

$$Z = \frac{X - \mu}{\sigma} = \frac{76 - 70}{3} = \frac{6}{3} = 2$$

2.8 Example: Computing a Z-Score

Recall: Adult male heights $X \sim \mathrm{Norm}(70, 9)$ (mean 70 inches, std dev 3 inches).

What is the Z-score for a height of 76 inches?

$$Z = \frac{X - \mu}{\sigma} = \frac{76 - 70}{3} = \frac{6}{3} = 2$$

A height of 76 inches is 2 standard deviations above the mean.

2.9 Example: Computing Probabilities

Adult male heights: $X \sim \mathrm{Norm}(70, 9)$

What is $P(X \leq 73)$?

2.9 Example: Computing Probabilities

Adult male heights: $X \sim \mathrm{Norm}(70, 9)$

What is $P(X \leq 73)$?

Step 1: Standardize

$$Z = \frac{73 - 70}{3} = 1$$

2.9 Example: Computing Probabilities

Adult male heights: $X \sim \mathrm{Norm}(70, 9)$

What is $P(X \leq 73)$?

Step 1: Standardize

$$Z = \frac{73 - 70}{3} = 1$$

Step 2: Use standard normal CDF

$$P(X \leq 73) = P(Z \leq 1) = \Phi(1)$$

2.10 The Z-Table

Historically, we used **Z-tables** (also called **standard normal tables**) to look up values of $\Phi(z)$.

2.10 The Z-Table

Historically, we used **Z-tables** (also called **standard normal tables**) to look up values of $\Phi(z)$.

The Z-table gives $P(Z \leq z)$ for various values of z.

2.10 The Z-Table

Historically, we used **Z-tables** (also called **standard normal tables**) to look up values of $\Phi(z)$.

The Z-table gives $P(Z \leq z)$ for various values of z.

For example: $\Phi(1) \approx 0.8413$

2.10 The Z-Table

Historically, we used **Z-tables** (also called **standard normal tables**) to look up values of $\Phi(z)$.

The Z-table gives $P(Z \leq z)$ for various values of z.

For example: $\Phi(1) \approx 0.8413$

So $P(X \leq 73) \approx 0.8413$ or about 84%.

2.10 The Z-Table

Historically, we used **Z-tables** (also called **standard normal tables**) to look up values of $\Phi(z)$.

The Z-table gives $P(Z \leq z)$ for various values of z.

For example: $\Phi(1) \approx 0.8413$

So $P(X \leq 73) \approx 0.8413$ or about 84%.

Note: Today, we typically use software rather than tables, but understanding Z-scores is still essential.

2.11 The Standard Normal Table (Z-Table)

The Z-table shows $\Phi(z) = P(Z \leq z)$ for the standard normal distribution:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

For negative z: use symmetry $\Phi(-z) = 1 - \Phi(z)$