

Chapter 3

The Normal Distribution

Intended Learning Outcomes

- Understand what a density curve is and its properties
- Describe a Normal distribution using $N(\mu, \sigma^2)$
- Apply the 68–95–99.7 rule
- Compute and interpret z -scores
- Use the standard Normal table to find proportions
- Find the k th percentile of a Normal distribution (inverse Normal)

PART 1

Density Curves

What is a smooth mathematical model for data, and how do we describe it?

Continuous Distributions: Motivation

Most real-world measurements (like screen time, testosterone levels, highway speeds) can take any value within a range, not just whole numbers. These are called **continuous distributions**.

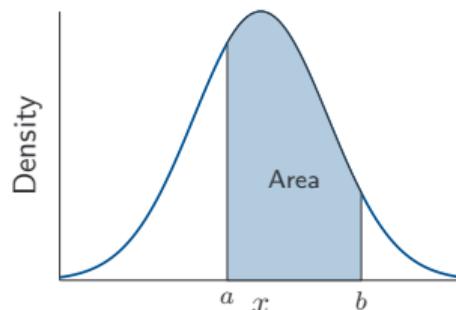
- Discrete: countable outcomes (e.g., number of downloads of an app)
- Continuous: any value in an interval (e.g., daily screen time in hours)

Density Curves: Definition

Density Curve

A **density curve** is a curve that:

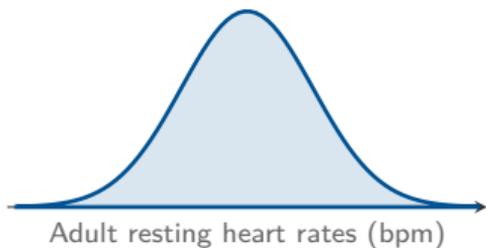
- is always on or above the horizontal axis (i.e. non-negative)
 - has a total area of exactly **1** underneath it
-
- A density curve is a mathematical model for the distribution of a continuous random variable.
 - The **area** under the curve above any range of values equals the **proportion** (or probability) of observations in that range.



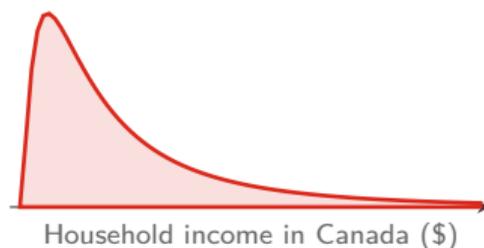
Area between a and b = proportion of observations falling in $[a, b]$.

A variety of density curves exist in the world

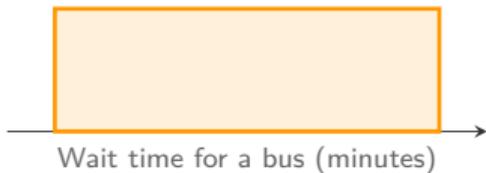
Bell-shaped



Right-skewed



Uniform

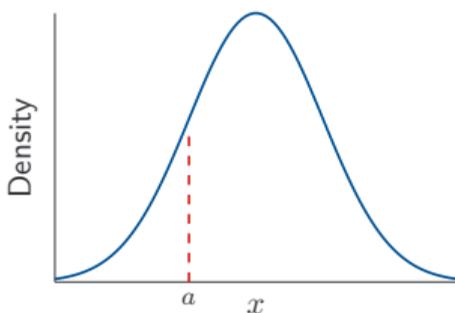


Bimodal



Important note regarding continuous distributions

Principle: For continuous distributions, the probability of observing any single exact value is zero.



The probability of X taking any specific value a is the area under the curve at exactly a , which is zero because it's just a line with no width.

We only have non-zero probabilities for intervals of values, which correspond to areas under the curve.

Probabilities of Intervals for continuous distributions

As a consequence of the previous principle, it follows that the following expressions are equivalent for any values a and b :

- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$
- $P(X \leq b) = P(X < b)$
- $P(X \geq a) = P(X > a)$

From Data to Mathematical Models

In chapter 2, we have described data using **statistics** computed from a sample:

- sample mean \bar{x}
- sample standard deviation s
- the five-number summary

But to understand the bigger picture, we need to describe the **entire population** or the **theoretical model** (like a density curve) that generates the data. To do this, we need a different kind of number.

A sample yields STATISTICS:

- A mean \bar{x} that varies between samples
- A standard deviation s that varies between samples
- Numbers that describe *the data we collected*

A population/model has PARAMETERS:

- A fixed center μ (the true mean)
- A fixed spread σ (the true SD)
- Numbers that describe *the entire population or model*

PART 2

The Normal Distribution

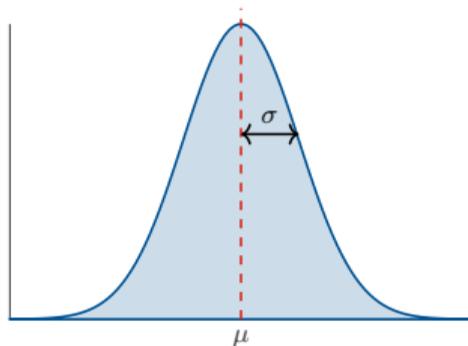
The bell curve: parameters, shape, and the 68–95–99.7 rule

Normal Distributions

Normal Distribution

A **Normal distribution** is described by a symmetric, bell-shaped density curve. It is completely determined by its mean μ and standard deviation σ :

$$X \sim N(\mu, \sigma^2)$$



Normal distributions describe many real measurements: screen time, testosterone levels, caffeine content in coffee.

Notation for Normal Distributions

The Standard Notation:

$$X \sim N(\mu, \sigma^2)$$

- **Read as:** " X follows a Normal distribution with mean μ and variance σ^2 ."
- **The tilde (\sim):** Means "*is distributed according to.*"
- **Note:** the second parameter is the variance σ^2 , not the standard deviation σ . There are two equivalent ways to express the same distribution (e.g. $N(100, 15^2)$ and $N(100, 225)$ are the same distribution).

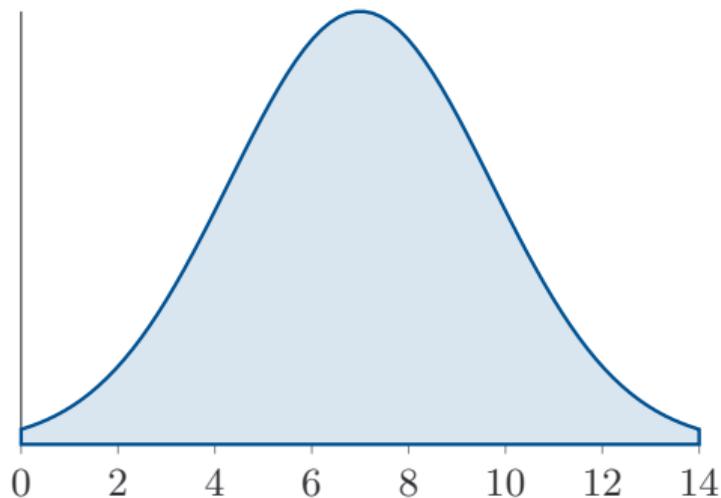
Theoretical Parameters vs. Sample Estimates:

- μ and σ are fixed parameters that define the theoretical model.
- In practice, these are often estimated from data using the sample mean (\bar{x}) and sample standard deviation (s).

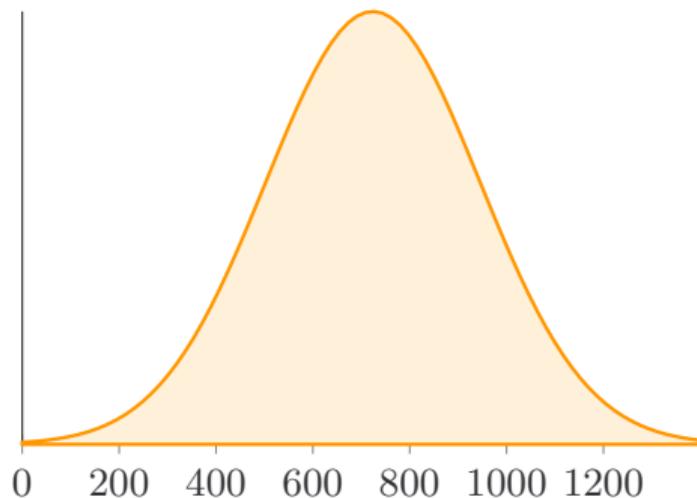
Why Is the Normal Distribution So Common?

Principle: When many small, independent factors contribute to a measurement, the result tends to follow a Normal distribution.

Screen Time (hours/day)



Testosterone (ng/dL)

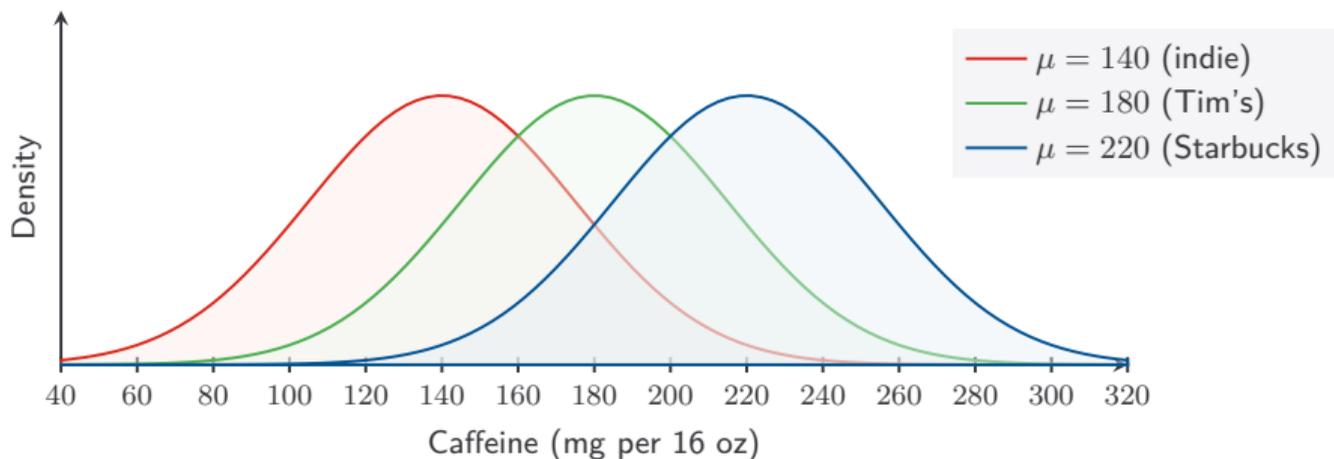


The Effect of μ

Example 3.3

Context: A food scientist measures the caffeine content (in mg per 16 oz cup) at three coffee chains. All three chains have similar cup-to-cup variability ($\sigma = 35$ mg), but different average caffeine levels.

Task: Look at the three curves below. What changes when μ changes? What stays the same?

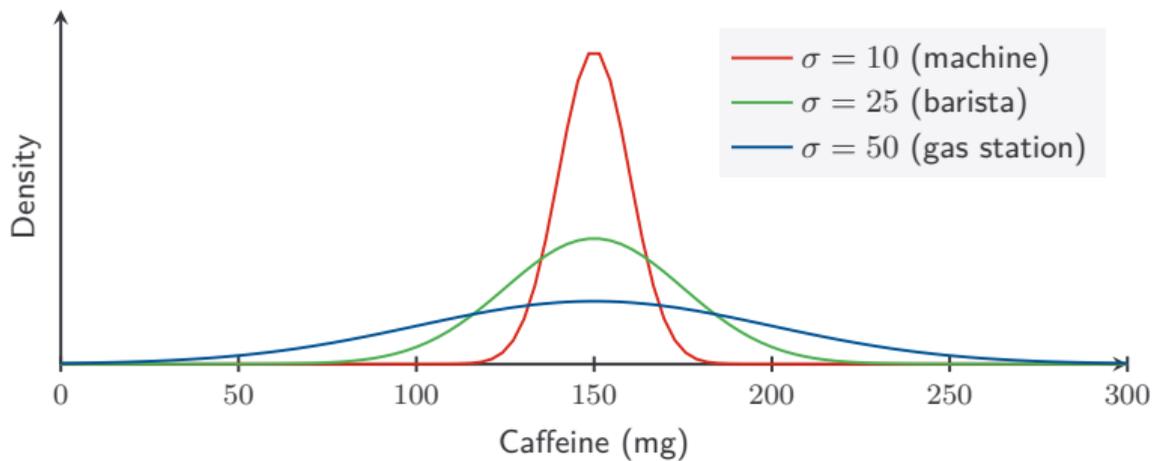


The Effect of σ

Example 3.4

Context: Three different espresso drinks all have the same average caffeine content ($\mu = 150$ mg), but vary in consistency.

Task: Look at the three curves. What changes when σ changes? Which espresso drink is most consistent?

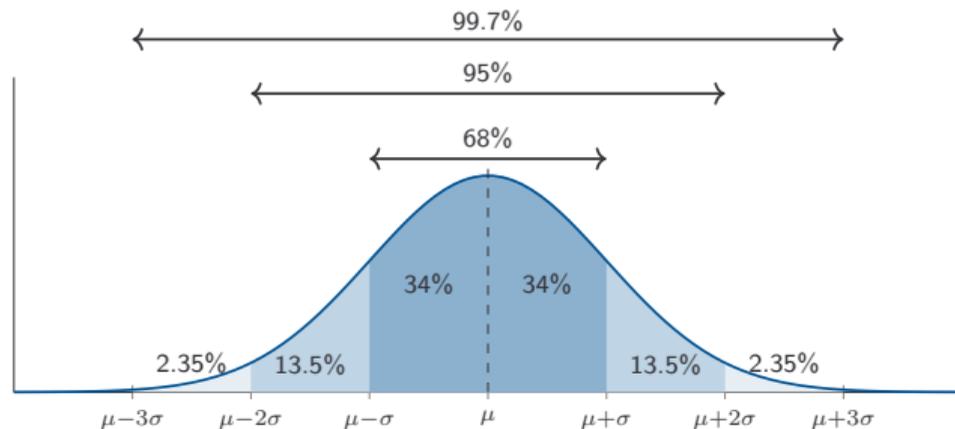


The Empirical Rule

The 68–95–99.7 Rule

In a Normal distribution $N(\mu, \sigma^2)$:

- Approximately **68%** of observations fall within $\mu \pm \sigma$
- Approximately **95%** of observations fall within $\mu \pm 2\sigma$
- Approximately **99.7%** of observations fall within $\mu \pm 3\sigma$



PART 3

Standardizing and z -Scores

How many standard deviations from the mean

Standardizing: From Any Normal to One Table

Principle: All Normal distributions are the same if we measure in units of σ from the mean μ . This is called **standardizing**.

The z -Score

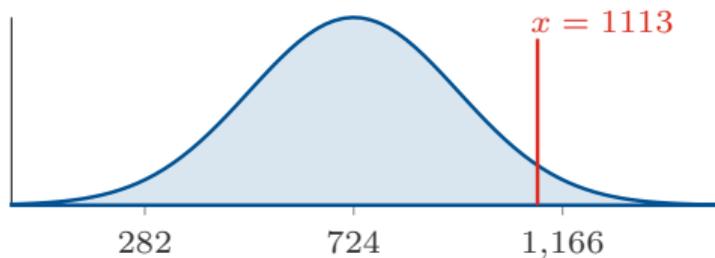
If x is an observation from a distribution with mean μ and standard deviation σ , its **standardized value** (z -score) is:

$$z = \frac{x - \mu}{\sigma}$$

- z tells you **how many standard deviations** x is from the mean
- $z > 0$: above the mean; $z < 0$: below; $z = 0$: at the mean

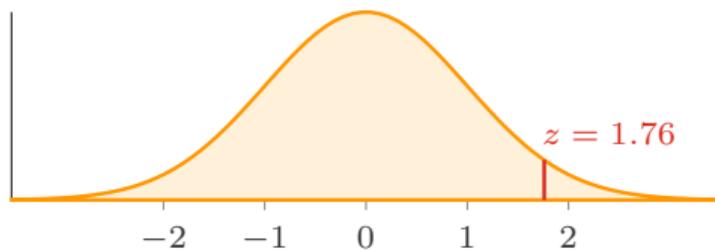
Standardizing: Visualized

Original: $N(724, 221^2)$



↓ $z = \frac{x - \mu}{\sigma}$

Standardized: $N(0, 1)$



Computing z -Scores

Example 3.6

Context: Testosterone levels among men aged 19–29 are approximately $N(724, 221^2)$ ng/dL.

Find: the z -score for each of the following testosterone levels.

(a) A man has 950 ng/dL.

(b) A man has 400 ng/dL.

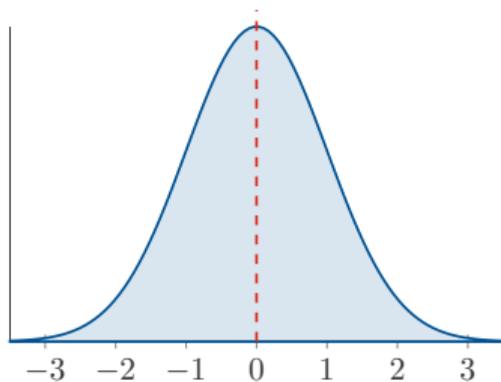
The Standard Normal Distribution

Standard Normal Distribution

The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1:

$$Z \sim N(0, 1)$$

$$Z \sim N(0, 1)$$



Why the Standard Normal?

Principle: Every Normal distribution can be converted to $N(0, 1)$ by standardizing.

This means we can compute any normal probability using just the distribution of the standard normal.

Without standardizing:

- Need a separate table for every (μ, σ) combination
- Testosterone: table for $N(724, 221^2)$
- Screen time: table for $N(7, 2.7^2)$
- Caffeine: table for $N(188, 36^2)$

⋮

- Infinitely many tables

With standardizing:

- Convert any x to $z = \frac{x - \mu}{\sigma}$
- Look up $\Phi(z)$ in one standard Normal table
- Works for any Normal distribution
- One table handles every problem

PART 4

Calculating Probabilities of the Normal Distribution

Using standard normal table to find exact areas under the curve

The standard normal distribution

- Any $N(\mu, \sigma^2)$ can be standardized to $N(0, 1)$ using $z = \frac{x-\mu}{\sigma}$
- So we only ever need one distribution to compute probabilities from normal random variables (the standard Normal distribution)

For example, if we want to calculate $P(X \leq 1113)$ where $X \sim N(724, 221^2)$, we can rewrite this in terms of Z :

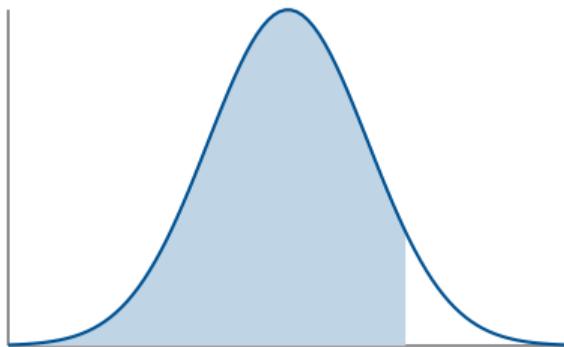


The Φ Notation

$\Phi(z)$: *Phi of z*

We denote $\Phi(z)$ for the probability that a standard Normal random variable is less than or equal to z :

$$\Phi(z) = P(Z \leq z) \quad \text{where } Z \sim N(0, 1)$$



STANDARD NORMAL TABLE

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Decoding the standard normal table

Given a value of z , the table gives $\Phi(z) = P(Z \leq z)$

- Row:

- Column:

The table covers $z = 0.00$ to $z = 3.49$.

Reading the Standard Normal Table

Example 3.9

Find: Find the probability $P(Z \leq 1.76)$ in the standard normal table.

Find: $\Phi(1.76) = P(Z \leq 1.76)$.

Row 1.7, column .06:

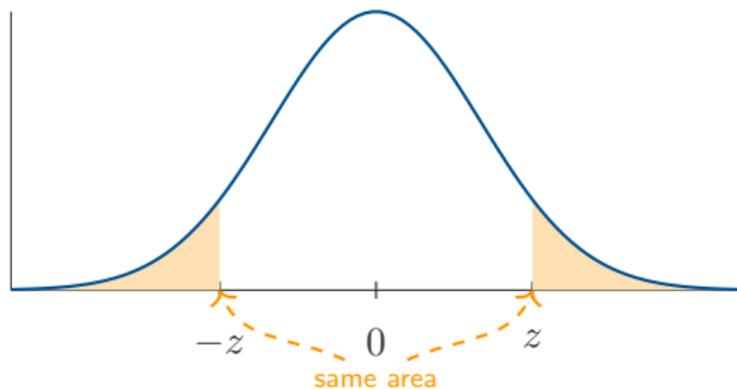
$$\Phi(1.76) =$$

STANDARD NORMAL TABLE

<i>z</i>04	.05	.06	.07	.08	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.59382	.9394	.9406	.9418	.9429	...
1.69495	.9505	.9515	.9525	.9535	...
1.79591	.9599	.9608	.9616	.9625	...
1.89671	.9678	.9686	.9693	.9699	...
1.99738	.9744	.9750	.9756	.9761	...
		⋮	⋮	⋮	⋮	⋮	⋮

Using the Table When z Is Negative

Our table only lists $\Phi(z)$ for $z \geq 0$. The curve's symmetry handles negative z .

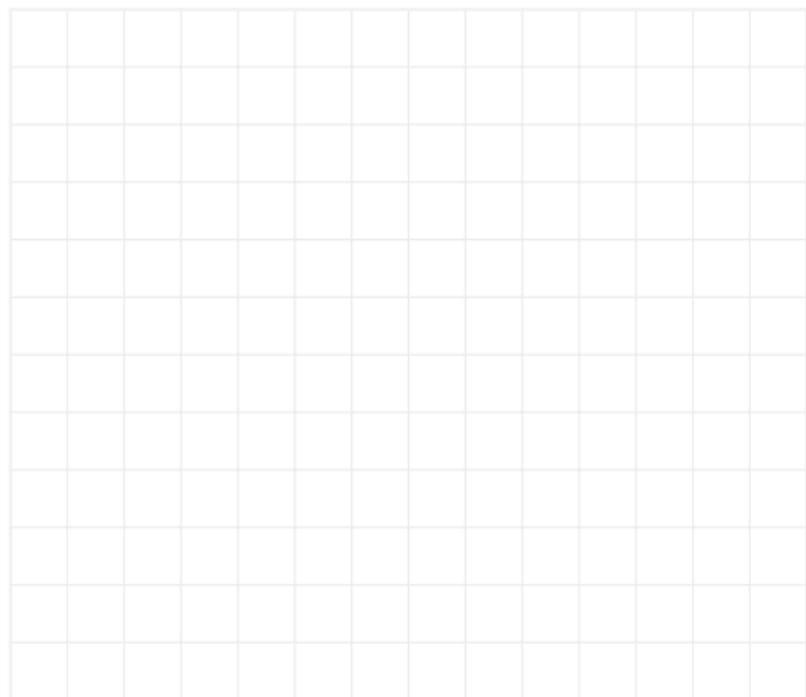
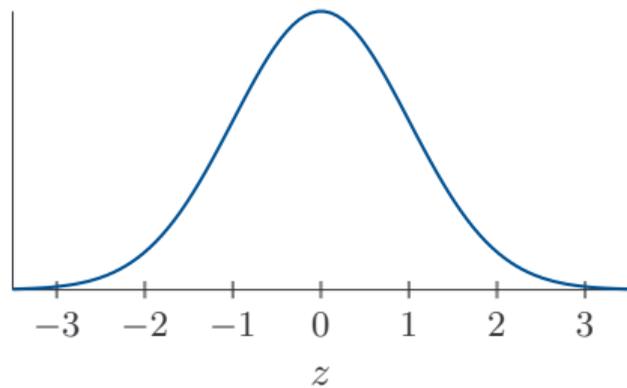


Symmetry Rule

For any $z > 0$:

Quick Example: $\Phi(-1.50)$

Find: Find $\Phi(-1.50)$.

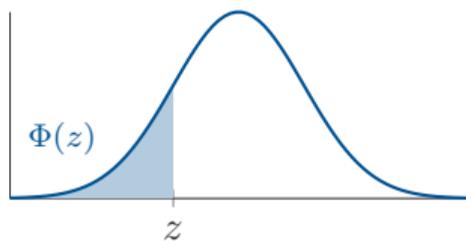
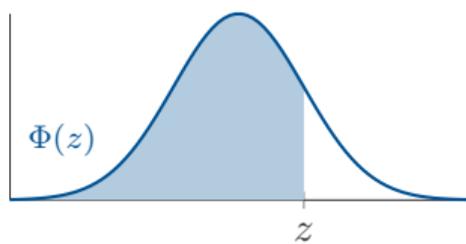


Three Types of Normal Calculations

When working with the Normal distribution, there are three main types of probability calculations:

- **Left-tail:** $P(X \leq x)$: the proportion of observations less than or equal to a value.
- **Right-tail:** $P(X \geq x)$: the proportion of observations greater than or equal to a value.
- **Between:** $P(a \leq X \leq b)$: the proportion of observations between two values.

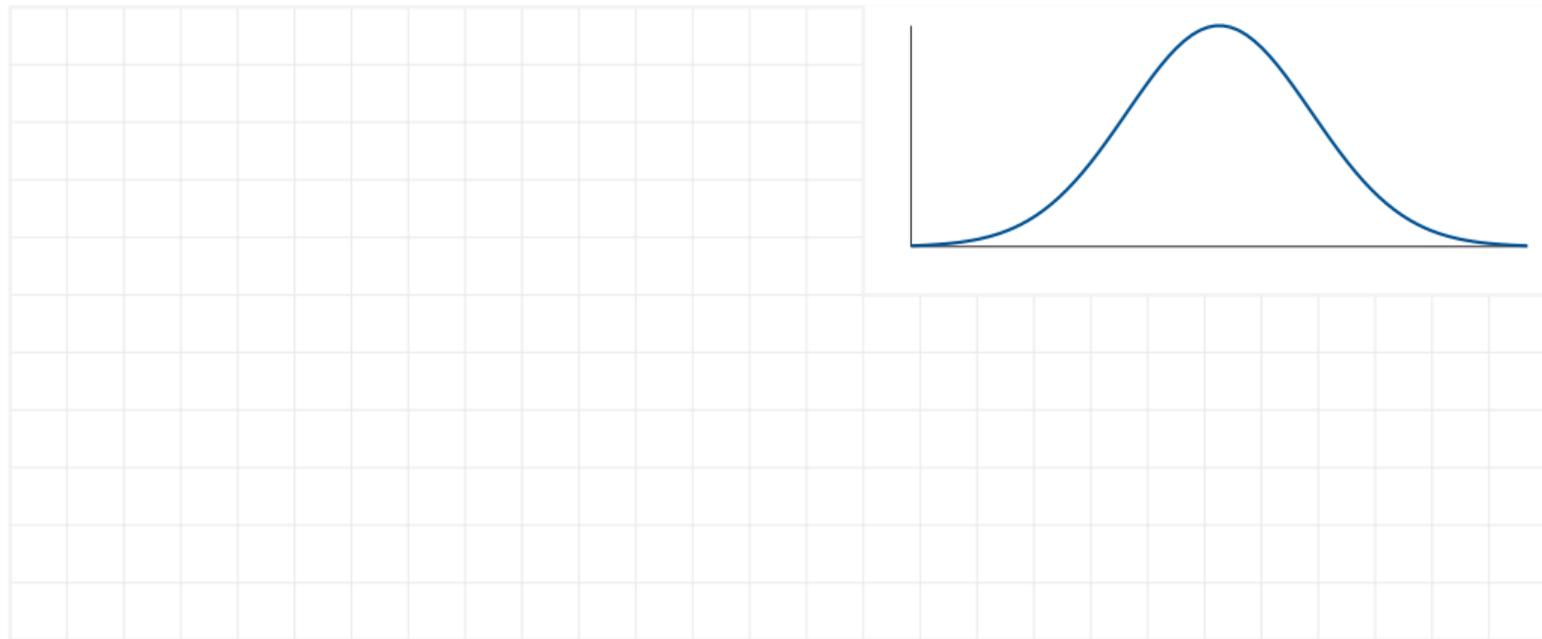
Left Tail: $P(Z \leq z)$



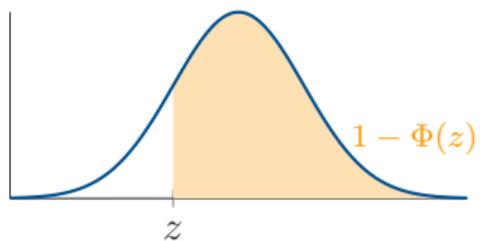
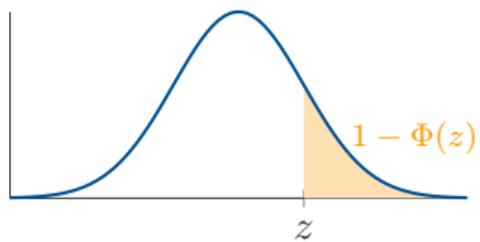
Screen Time: Under 4 Hours per Day

Context: Daily smartphone screen time among college students is approximately $N(7, 2.7^2)$ hours.

Find: What proportion of students use their phone less than 4 hours per day?



Right Tail: $P(Z \geq z)$

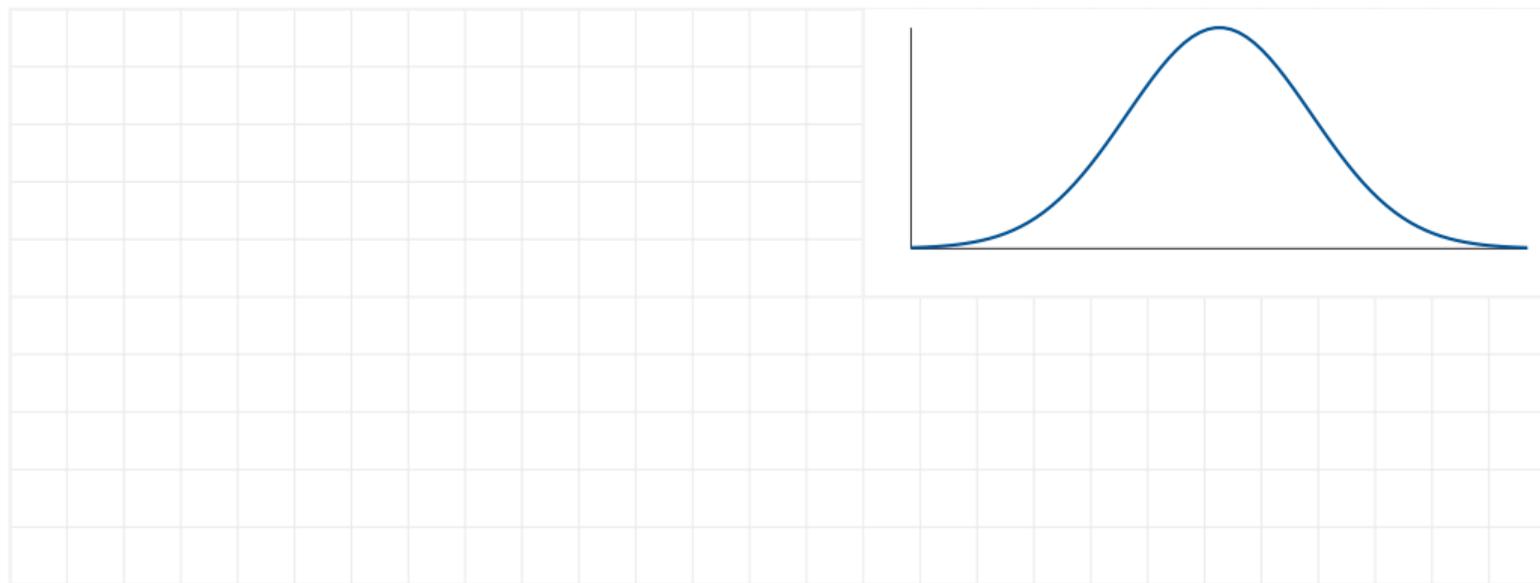


Screen Time: Over 12 Hours per Day

Example 3.10

Context: Daily smartphone screen time among college students is approximately $N(7, 2.7^2)$ hours.

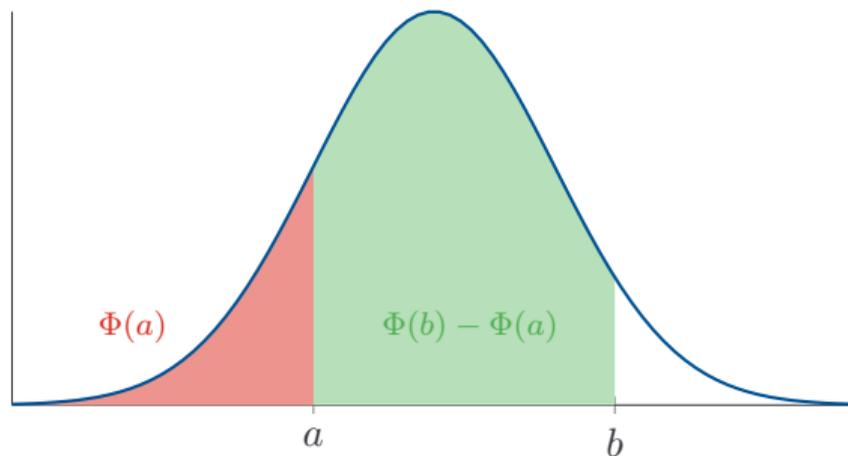
Find: What proportion of students spend more than 12 hours per day on their phone?



Between Two Values: $\Phi(b) - \Phi(a)$

Look up both values, then subtract.

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$



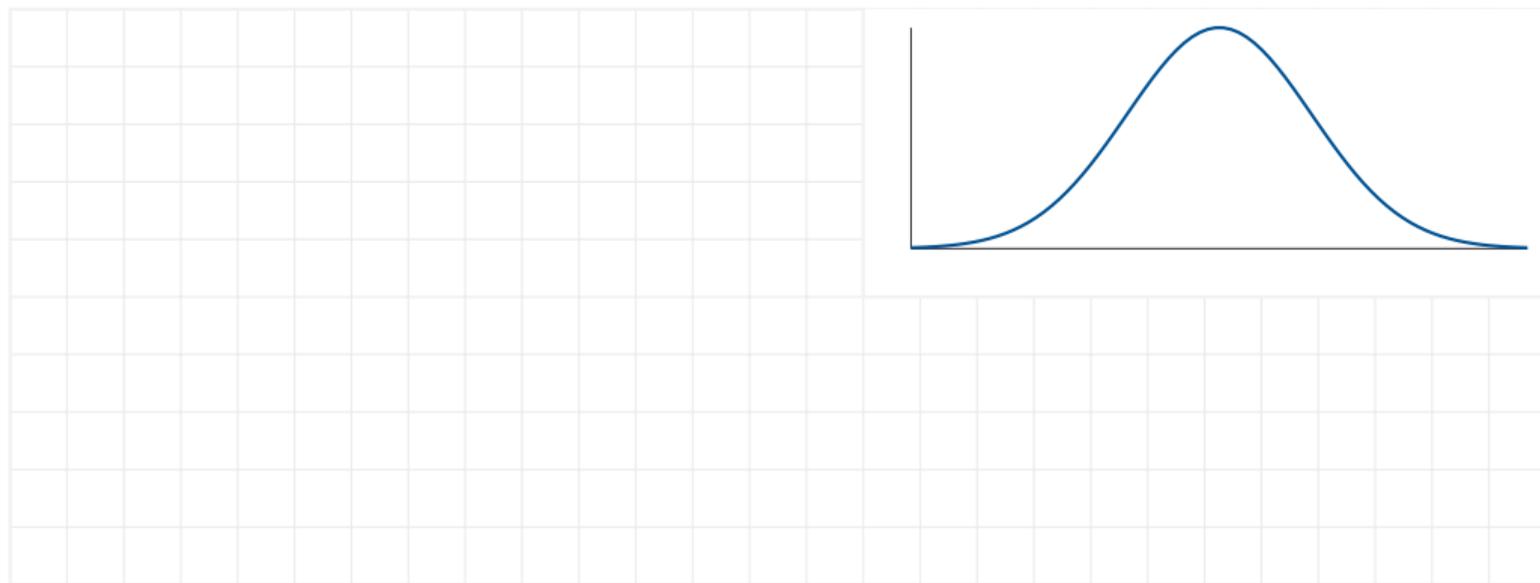
The green area represents $\Phi(b)$: the cumulative probability at b . Subtracting $\Phi(a)$ (red) leaves only the probability between a and b (green).

Screen Time: Between 5 and 10 Hours per Day

Example 3.11

Context: Daily smartphone screen time among college students is approximately $N(7, 2.7^2)$ hours.

Find: What proportion of students spend between 5 and 10 hours per day on their phone?



Three-Step Method: Finding a Probability of the Normal Distribution

Use this when you know a value and want to find what proportion of the distribution is above or below it.

Computing a normal probability

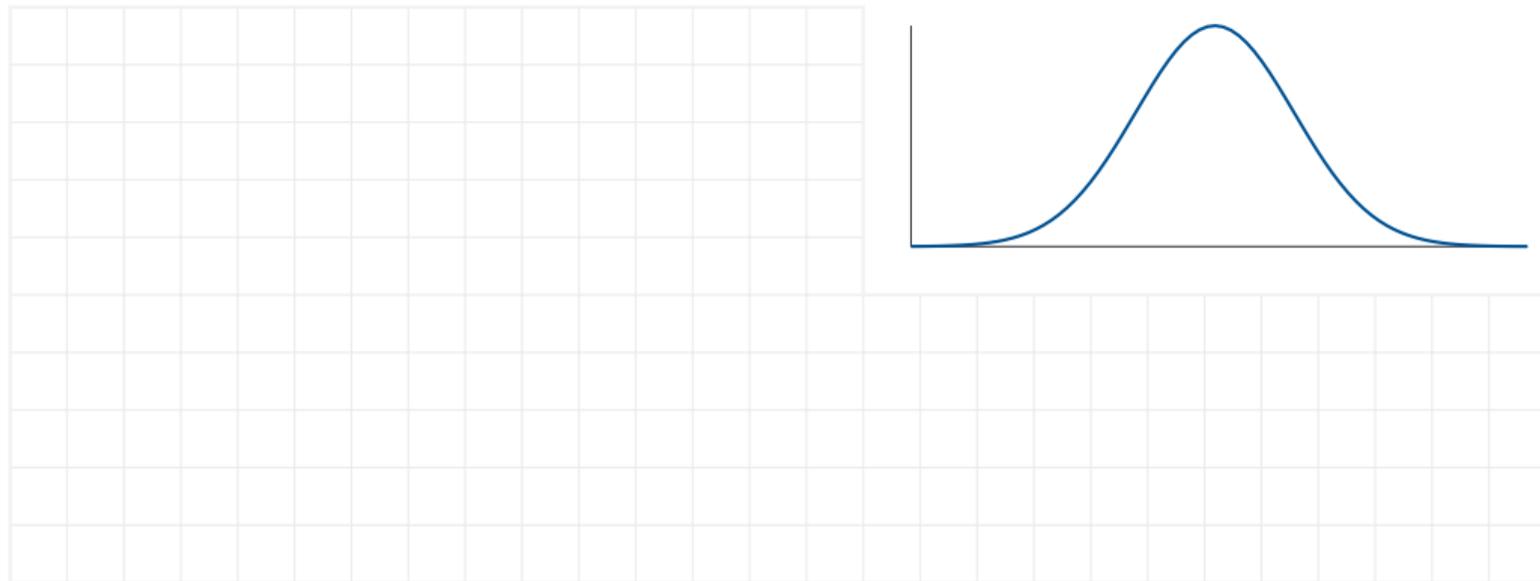
1. **Standardize:** Convert x -values to z -scores.
2. **State and sketch:** Draw the standard Normal curve, mark the value(s) of interest, and shade the area you want.
3. **Identify the type:** Is this a left-tail, right-tail, or between calculation? If any z -score is negative, apply the symmetry rule.
4. **Look up and compute:** Use the standard Normal table to find the required area.

Caffeine in Coffee: Exceeding 250 mg

Example 3.12

Context: A food-safety researcher measured the caffeine content of 16 oz specialty coffees. Cups contain on average 188 mg of caffeine, with a standard deviation of 36 mg, following $N(188, 36^2)$ mg.

Find: What percentage of cups exceed 250 mg (a level associated with anxiety symptoms)?

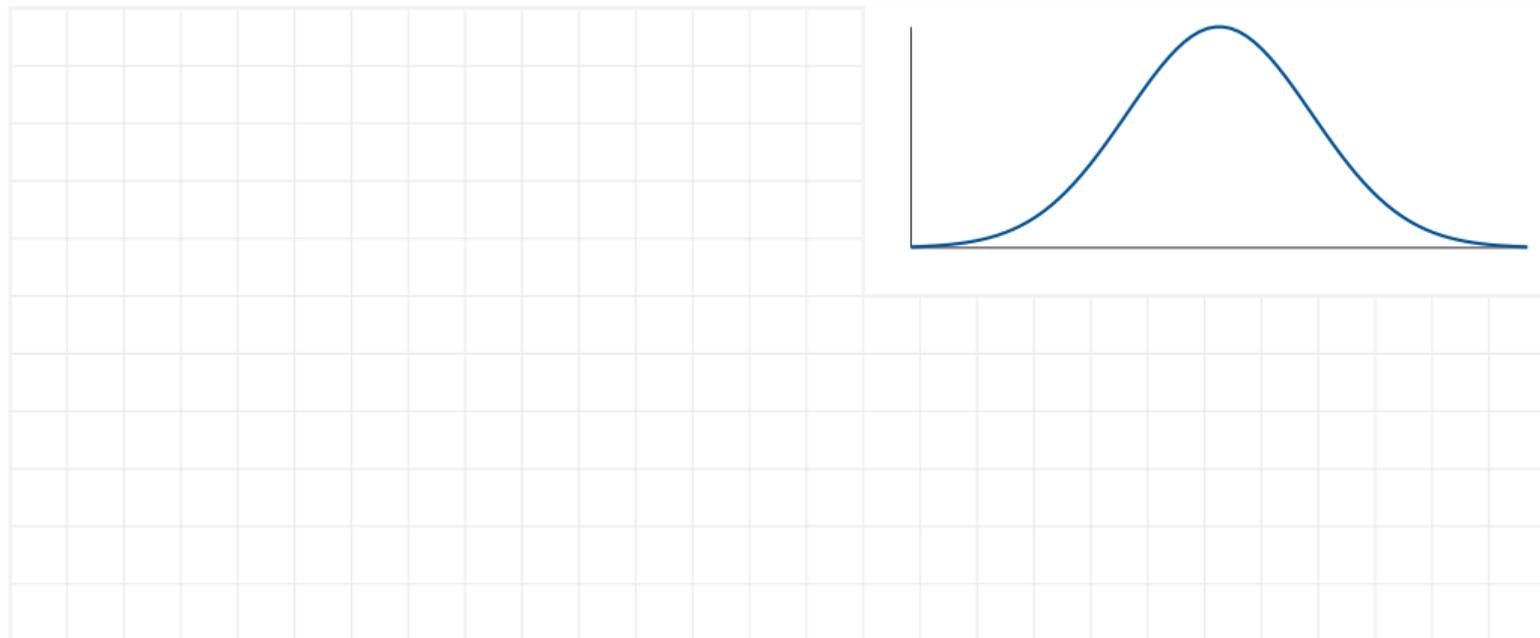


Screen Time: Under 3 Hours per Day

Example 3.13

Context: Daily smartphone screen time among college students is approximately Normally distributed with a mean of 7 hours and a standard deviation of 2.7 hours.

Find: What percentage of students use their phone less than 3 hours per day?

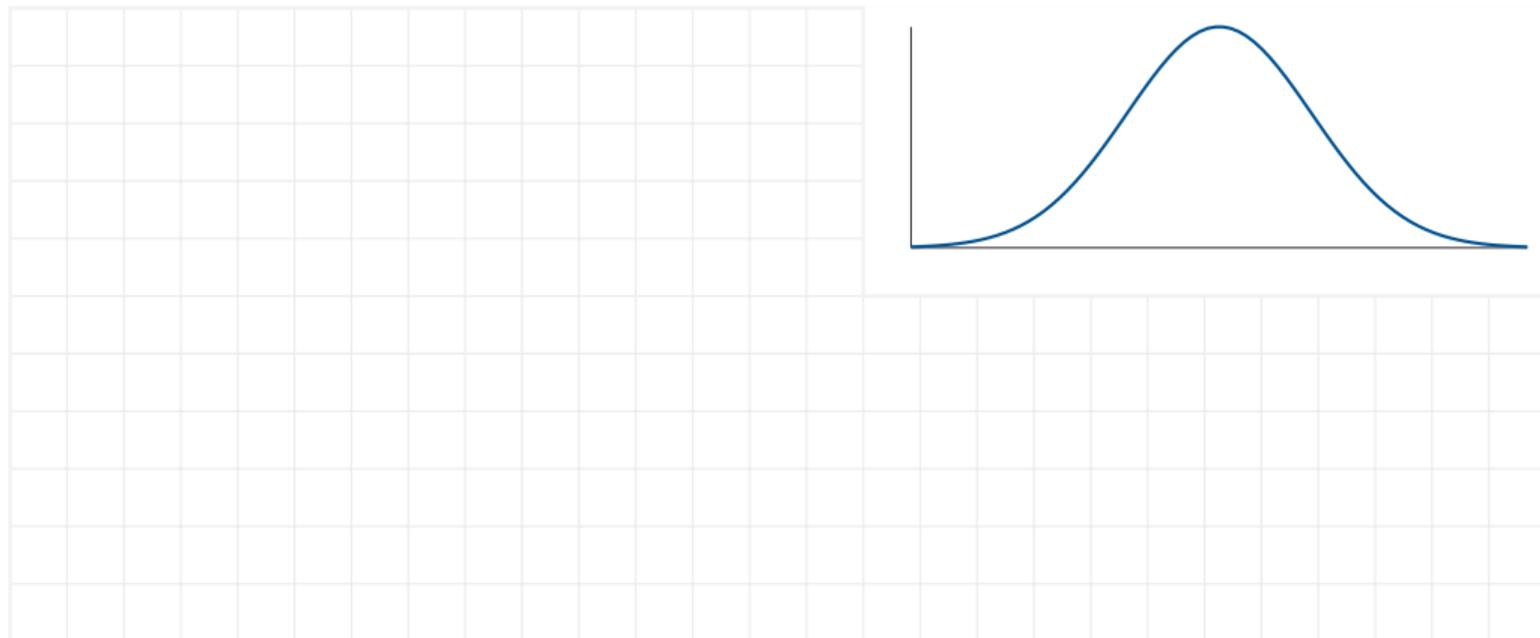


Screen Time: Between 4 and 10 Hours

Example 3.14

Context: Daily smartphone screen time among college students is approximately Normally distributed with a mean of 7 hours and a standard deviation of 2.7 hours.

Find: What percentage of students spend between 4 and 10 hours per day on their phone?



PART 5

Normal Percentile Calculations

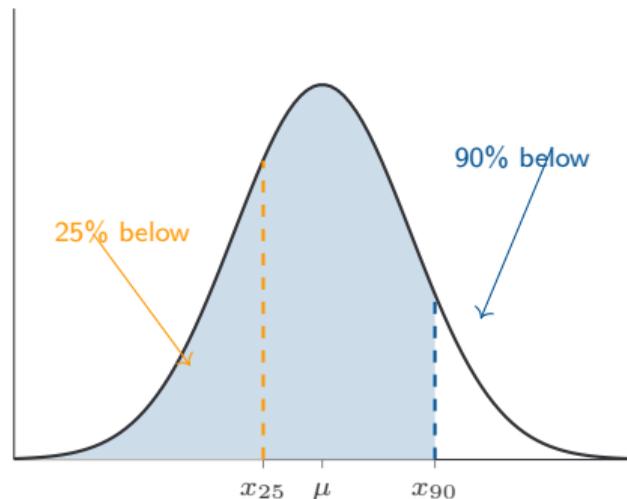
Given a proportion, what value has this much area below it?

Percentiles

Percentile

The k th percentile of a distribution is the value x such that $k\%$ of observations fall at or below x .

- $P(X \leq x) = k/100$
- The 50th percentile is the **median**
- The 25th and 75th percentiles are the **quartiles**
- A z -score and a percentile convey the same information in different units

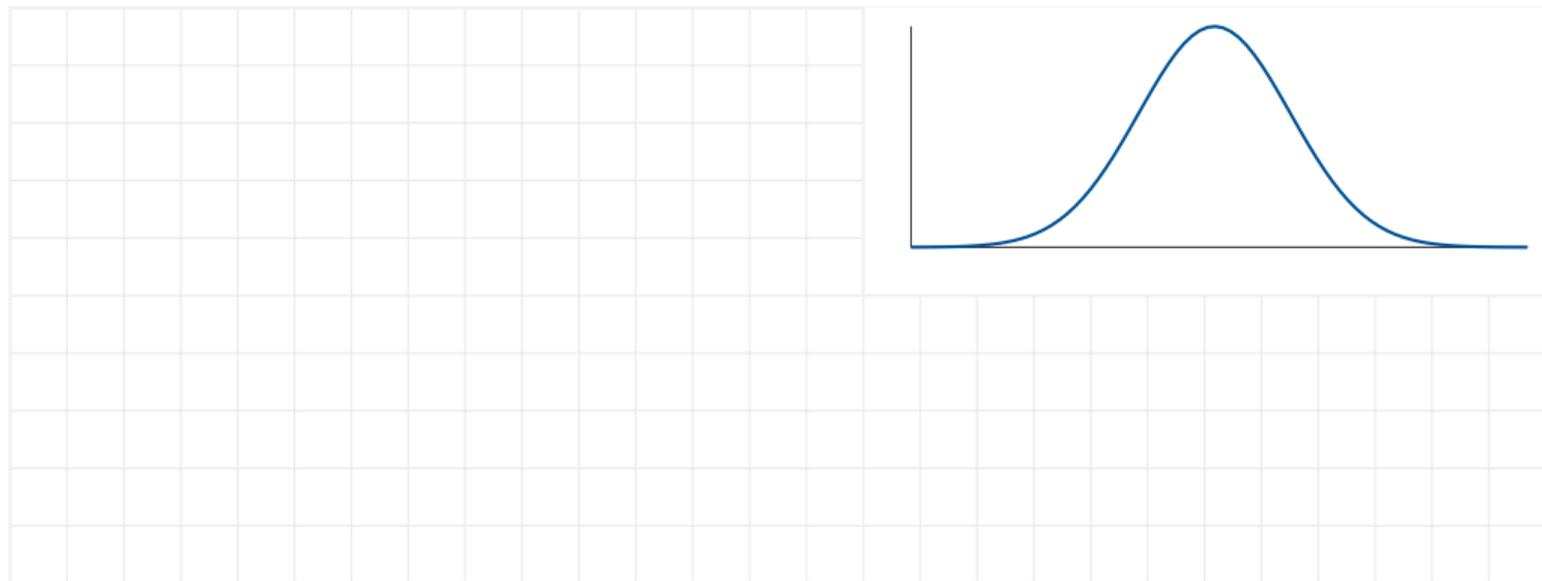


Highway Speed: Fastest 10% of Drivers

Example 3.15

Context: Vehicle speeds on a 100 km/h highway are approximately $N(117, 8^2)$ km/h.

Find: Above what speed are the fastest 10% of drivers?



Three-Step Method: Find a Value Given a Proportion

Use this when you know a percentage (like “the top 10%”) and want to find the cutoff value.

Inverse Normal Calculation

- 1. State and sketch:** Draw the Normal curve and shade the given area. Express it as a **left-tail** cumulative proportion p .
- 2. Use standard normal table backwards:**
If $p > 0.5$: search the body of standard normal table for p and read off z (positive).
If $p < 0.5$: search for $1 - p$ instead, read off the positive z , then negate it.
- 3. Unstandardize:** Convert back to the original scale:

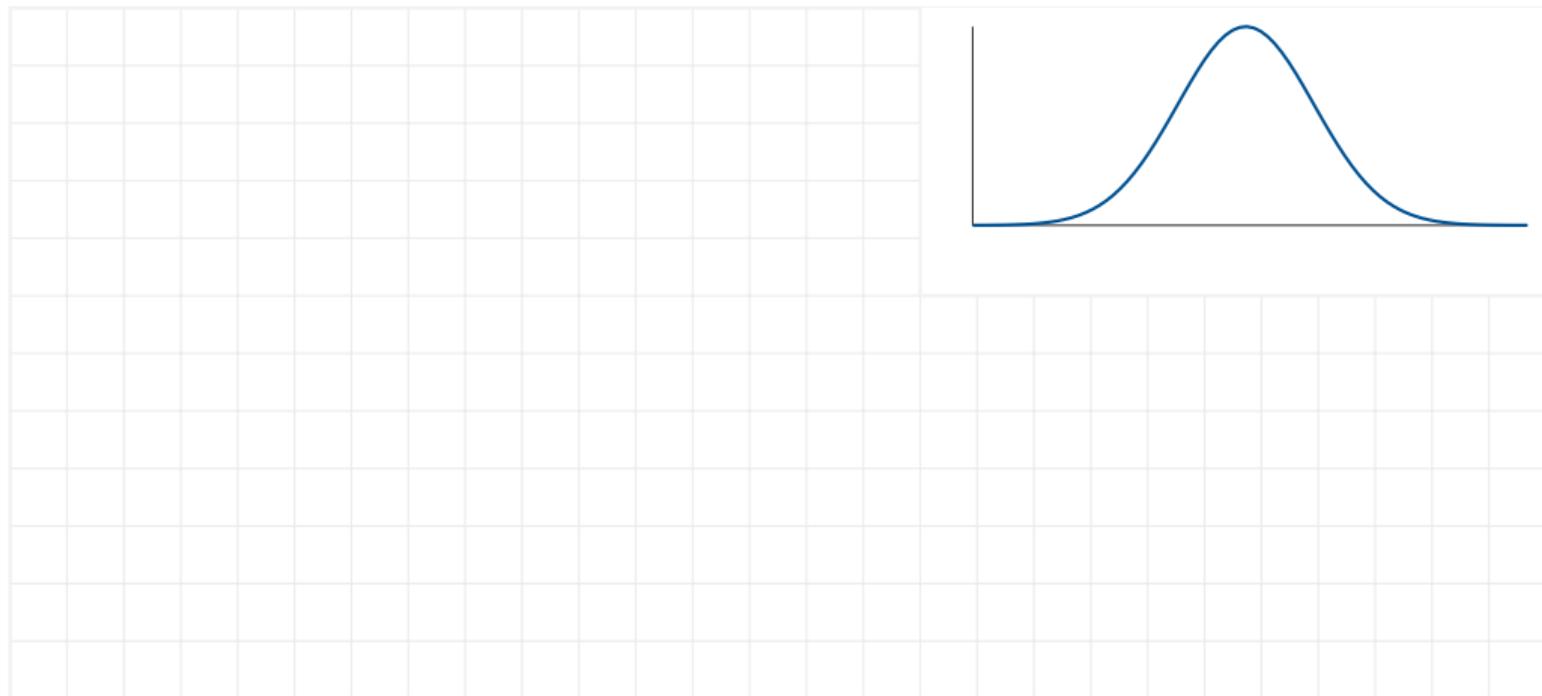
$$x = \mu + z\sigma$$

Highway Speed: Middle 90% of Traffic

Example 3.16

Context: Vehicle speeds follow $N(117, 8^2)$ km/h.

Find: The speeds x_L and x_U that bound the middle 90% of traffic (5% in each tail).



Some caveats (for the final exam and last assignment)

- If a question can be answered using the 68–95–99.7 rule, you don't need to use the standard normal table. You can just give an approximate answer based on the rule.
- The normal table only provides values for z between 0 and 3.49. For z greater than 3.49, the area is so close to 1 that we can treat it as 1 for all practical purposes.
- Similarly, for $z < -3.49$, the area is so close to 0 that we can treat it as 0 for all practical purposes.
- When looking up values in the standard normal table, if the exact value is not listed, choose the closest one. This will give you an approximation of the area.
- If two values are equally close (e.g. attempting to look up $\Phi(1.375)$), you can choose either one (e.g. choose $\Phi(1.37)$ or $\Phi(1.38)$). The difference in area will be very small, and it won't significantly affect your answer.

CHAPTER 3

Summary

Key ideas and formulas from this chapter

Chapter 3 Summary

■ Density Curves

- Area = 1, always ≥ 0
- Area under curve = proportion
- Idealized model for data distributions

■ Normal Distribution

- $N(\mu, \sigma^2)$: symmetric, bell-shaped
- μ controls center, σ controls spread
- 68–95–99.7 rule for quick estimates

■ z -Scores & Calculations

- Standardize: $z = \frac{x - \mu}{\sigma}$
- Unstandardize: $x = \mu + z\sigma$
- $\Phi(z) = P(Z \leq z)$ (standard normal table, $z \geq 0$)
- Symmetry: $\Phi(-z) = 1 - \Phi(z)$
- Right tail: $1 - \Phi(z)$
- Between: $\Phi(b) - \Phi(a)$
- Always draw and shade first

PRACTICE

Problems

Applying the tools from Chapter 3

Practice Problem 1: Density Curves

Example 3.17

Context: The time a student spends waiting for a campus shuttle can be modelled by a density curve that is a straight line from $(0, 0.4)$ to $(5, 0)$ and is 0 elsewhere (shuttles run every 5 minutes).

- (a) How do we verify this is a valid density curve?
- (b) What proportion of students wait less than 2 minutes?
- (c) What is the probability a student waits between 1 and 4 minutes?



Practice Problem 2: The 68–95–99.7 Rule

Example 3.18

Context: Resting heart rate among healthy adults follows $N(70, 10^2)$ bpm.

- (a) What percentage of adults have a resting heart rate between 60 and 80 bpm?
- (b) What percentage of adults have a resting heart rate above 80 bpm?
- (c) What percentage of adults have a resting heart rate below 50 bpm?
- (d) Above what heart rate are the highest 0.15% of adults?



Practice Problem 3: z -Scores

Example 3.19

Context: Testosterone levels among men follow $N(724, 221^2)$ ng/dL. Testosterone levels among women follow $N(33, 12^2)$ ng/dL.

- (a) Find the z -scores for a man at 500, 800, and 1200 ng/dL.
- (b) A woman at 45 ng/dL and a man at 1000 ng/dL. Who has more unusual testosterone?
- (c) What testosterone level corresponds to a z -score of -2 for women?

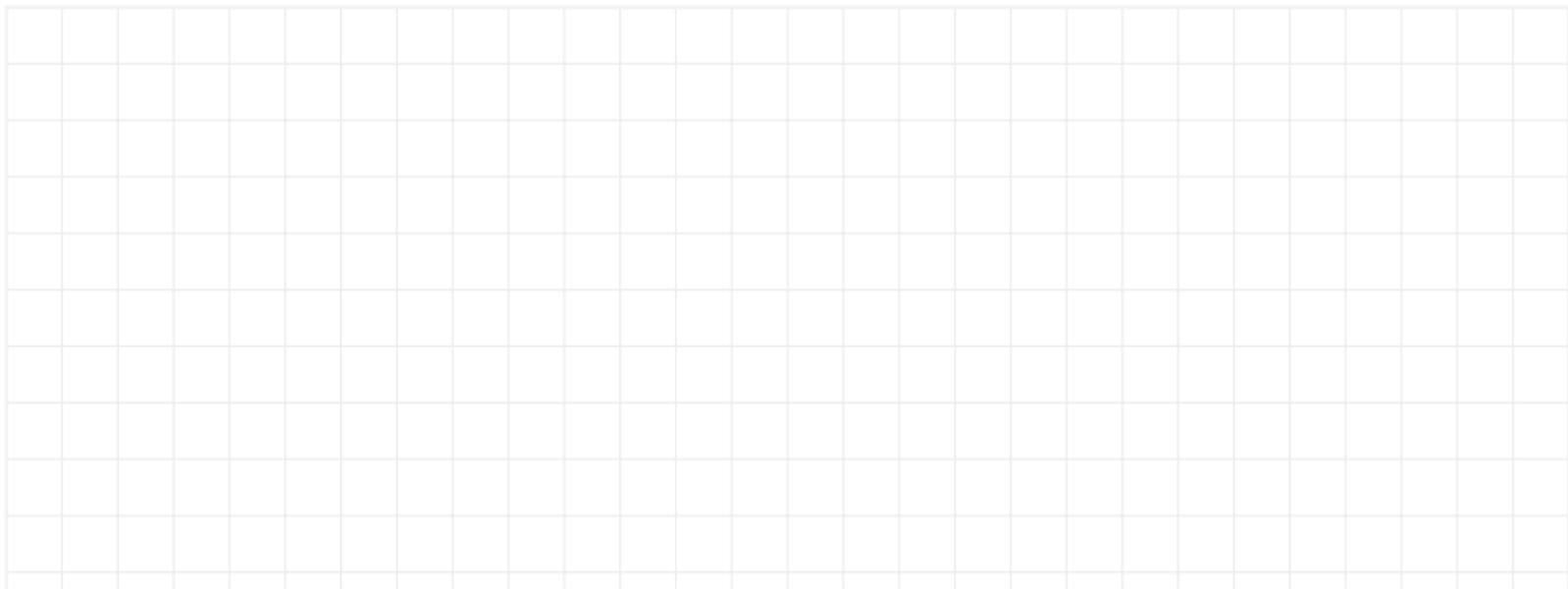


Practice Problem 4: Forward Normal Calculations

Example 3.20

Context: Caffeine content in 16 oz specialty coffees follows $N(188, 36^2)$ mg.

- (a) What proportion of cups contain less than 140 mg of caffeine?
- (b) What proportion contain more than 250 mg?
- (c) What proportion contain between 140 and 224 mg?

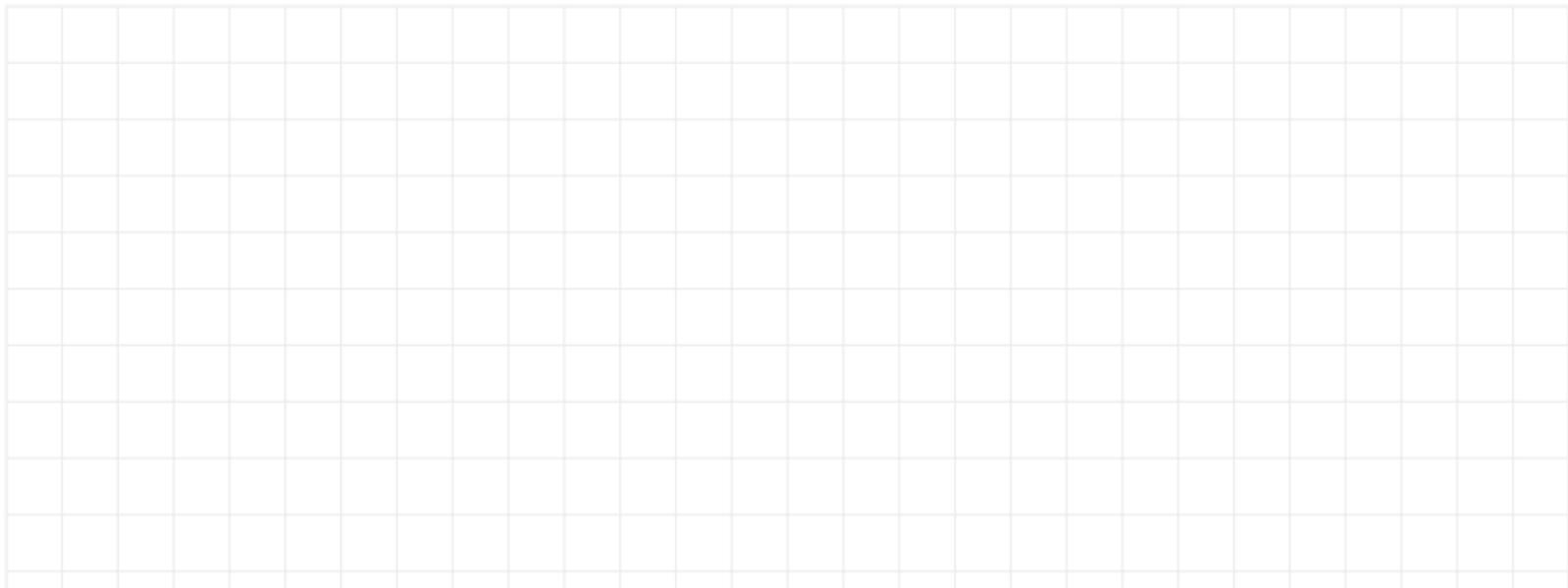


Practice Problem 5: Inverse Normal Calculations

Example 3.21

Context: Vehicle speeds on a 100 km/h highway follow $N(117, 8^2)$ km/h.

- (a) Below what speed are the slowest 5% of drivers?
- (b) What speed is at the 25th percentile?
- (c) Find the speeds that bound the middle 50% of traffic.



Practice Problem 6: Screen Time

Example 3.22

Context: Daily smartphone screen time among college students is approximately Normally distributed with a mean of 7 hours and a standard deviation of 2.7 hours.

- (a) What percentage of students have screen time between 9 and 12 hours per day?
- (b) What is the probability a student has exactly 7 hours of screen time?
- (c) How low would screen time need to be to fall in the lowest 25%?

