STATISTICS · STUDY GUIDE

Exam 4 Study Guide

This guide covers chapters 7 and 8, the bridge from describing data to making inferences. You will use the normal curve, the empirical rule and the z-table forward and backward. Then you will meet the idea the rest of the course rests on: a statistic has its own distribution, centered on the truth with spread σ ÷ √n, and the Central Limit Theorem tells you its shape. Work one chapter at a time, try each worked example with the steps covered, and finish with the ten self-test items. The cram sheet at the end holds every formula.

WHAT THIS GUIDE COVERS

- Chapter 7 · the normal model, area as probability, the empirical rule, z-scores and the standard normal, left, right and between areas from the table, percentiles and cutoffs, and symmetric middle intervals.
- Chapter 8 · population, sample and sampling distributions, the mean and standard error of $\bar{x}$, one individual versus a sample mean, the Central Limit Theorem, the standard error, and the sampling distribution of a proportion.

CHAPTER 7

The Normal Distribution

Read it in full: **Chapter 7 in the textbook**. Practice it in the lab: **Area Under the Curve** and **Working Backward**.

What to know

A **normal distribution** is a smooth, symmetric, single-peaked model, the bell curve. It is fixed by two parameters: the mean μ (center and peak) and the standard deviation σ (spread). Changing μ slides the curve; changing σ makes it taller and narrower or shorter and wider. We write $X \sim N(\mu, \sigma^2)$. Mean, median and mode coincide at μ, and the curve bends at μ ± σ.

For a continuous variable, **area under the curve is probability**; the total area is 1 and the height is density, not probability. The chance of any single exact value is 0, so $P(X < a)$ and $P(X \le a)$ are

equal. The model is an approximation: it fits heights and measurement errors well, and it fails on strongly skewed or two-peaked data such as income.

The **empirical rule** says about 68%, 95% and 99.7% of a normal distribution lie within 1, 2 and 3 SD of μ (exactly 68.27%, 95.45% and 99.73%). Halving each band gives the blocks 34%, 13.5%, 2.35% and 0.15% on each side. For heights with μ = 64.0 and σ = 2.5 inches, about 2.5% are taller than 69.0 and about 81.5% are between 59.0 and 66.5. For unknown shapes, fall back on Chebyshev: at least 75% within 2σ.

The **standard normal** $Z \sim N(0, 1)$ is the one curve every normal is converted to with $z = (x - \mu) \div \sigma$. Subtracting the mean recenters; dividing by σ converts to standard deviations, which have no units. Back from z: $x = \mu + z\sigma$. Standardizing rescales a distribution but never changes its shape: skewed data stay skewed.

The z-table gives $\Phi(z) = P(Z < z)$, the area to the left, with tenths in the row and hundredths in the column. Three question types cover everything: left tail $P(X < a) = \Phi(z)$; right tail $P(X > a) = 1 - \Phi(z)$; between $P(a < X < b) = \Phi(z_b) - \Phi(z_a)$. Symmetry gives $\Phi(-z) = 1 - \Phi(z)$. With μ = 72 and σ = 8, a score of 86 has z = 1.75 and Φ(1.75) = 0.9599, so about 4.0% score above 86.

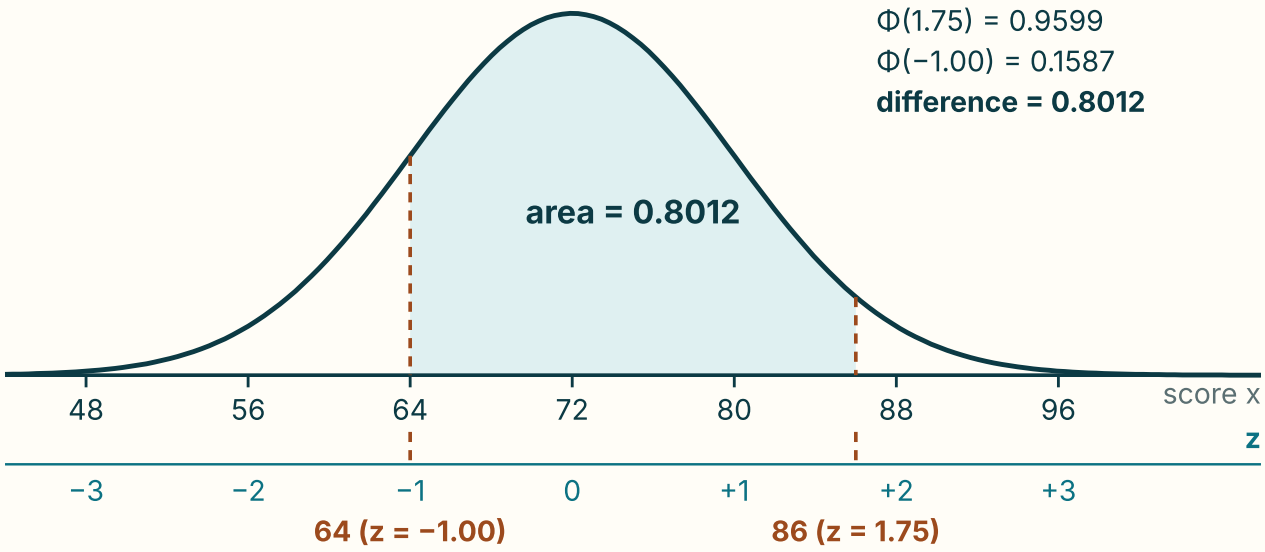
Always sketch and shade before computing, and check against the empirical rule. An area beyond z = 2 should be near 2.5%, and an area below the mean is 0.5000 exactly. Key values: Φ(1.00) = 0.8413, Φ(1.645) = 0.9500, Φ(1.96) = 0.9750, Φ(2.576) = 0.9950.

Working backward goes area → z → x. The kth percentile has k% below it, so "top 10%" means a left-tail area of 0.90. Find the closest area in the body of the table, read z, then use $x = \mu + z\sigma$. The 90th percentile of the exam is 72 + 1.28(8) = 82.24. Percentiles below 50 have negative z: the 25th percentile is 72 + (−0.67)(8) = 66.64. Dropping the minus sign is the classic error.

A symmetric middle interval splits the leftover area evenly. The middle 95% uses z = ±1.96, giving 56.32 to 87.68 for the exam. This is the ancestor of the confidence interval, but it is a different object: it describes where 95% of individuals fall when μ and σ are known, not a range for an unknown parameter.

LOCK THESE IN

- Normal is fixed by μ and σ; area under the curve is probability.
- 68, 95, 99.7 within 1, 2, 3 SD; blocks 34, 13.5, 2.35, 0.15 each side.
- $z = (x - \mu) \div \sigma$; $x = \mu + z\sigma$.
- Left: $\Phi(z)$. Right: $1 - \Phi(z)$. Between: $\Phi(z_b) - \Phi(z_a)$.
- $\Phi(-z) = 1 - \Phi(z)$, and $P(X < a) = P(X \leq a)$ for a continuous variable.
- Key z values: 1.645 (top 5%), 1.96 (middle 95%), 2.576 (middle 99%), 1.28 (90th percentile).
- Top k% means a left-tail area of $1 - k/100$.
- Standardizing never makes skewed data normal.



Area is probability. For scores with μ = 72 and σ = 8, the share between 64 and 86 is $\Phi(1.75) - \Phi(-1.00) = 0.9599 - 0.1587 = 0.8012$.

Formulas and when to use them

Quantity	Formula	When to use it
z-score	$z = (x - \mu) \div \sigma$	Convert a value to standard units
Back to x	$x = \mu + z\sigma$	Percentiles and cutoffs
Left tail	$P(X < a) = \Phi(z_a)$	"Below," "less than," "at most"
Right tail	$P(X > a) = 1 - \Phi(z_a)$	"Above," "more than," "at least"
Between	$\Phi(z_b) - \Phi(z_a)$	"Between a and b"
Symmetry	$\Phi(-z) = 1 - \Phi(z)$	Tables with only positive z
Middle C%	$z = \pm z^*$ with $(1 - C)/2$ in each tail	Symmetric intervals
Chebyshev	at least $1 - 1 \div k^2$	Shape unknown or not normal

Values worth knowing by heart

z	Area to the left Φ(z)	Use
0.00	0.5000	The mean splits the curve in half
±0.674	0.2500 and 0.7500	First and third quartiles
1.00	0.8413	Gives 68.27% within ±1
1.28	0.8997 (nearest to 0.90)	90th percentile
1.645	0.9500	Top 5% cutoff; middle 90%
1.96	0.9750	Middle 95%
2.00	0.9772	Gives 95.45% within ±2
2.326	0.9900	99th percentile
2.576	0.9950	Middle 99%
3.00	0.9987	Gives 99.73% within ±3

Worked examples

WORKED EXAMPLE 1 · THE EMPIRICAL RULE WITH NO TABLE

Newborn weights are about normal with μ = 3,400 g and σ = 500 g. What percent weigh more than 4,400 g, and what percent weigh between 2,900 and 4,400 g?

STEP 1 · BUILD THE RULER

μ ± 1σ: 2,900 to 3,900. μ ± 2σ: 2,400 to 4,400. μ ± 3σ: 1,900 to 4,900.

STEP 2 · ABOVE 4,400

4,400 is μ + 2σ. Outside ±2σ lies 5%, half on each side: **2.5%**.

STEP 3 · SPLIT AT THE MEAN

From 2,900 (μ − 1σ) to 3,400 is 34%. From 3,400 to 4,400 (μ + 2σ) is 47.5%.

STEP 4 · ADD

34% + 47.5% = **81.5%**.

CHECK

The answer is above the 68% inside ±1σ and below the 95% inside ±2σ, as it should be.

WORKED EXAMPLE 2 · LEFT, RIGHT AND BETWEEN WITH THE TABLE

Reaction times are normal with μ = 250 ms and σ = 40 ms. Find P(X < 300), P(X > 300) and P(220 < X < 300).

STEP 1 · STANDARDIZE 300

z = (300 − 250) ÷ 40 = 50 ÷ 40 = 1.25.

STEP 2 · LEFT TAIL

Φ(1.25) = **0.8944**.

STEP 3 · RIGHT TAIL

1 − 0.8944 = **0.1056**.

STEP 4 · STANDARDIZE 220

z = (220 − 250) ÷ 40 = −30 ÷ 40 = −0.75, and Φ(−0.75) = 0.2266.

STEP 5 · BETWEEN

0.8944 − 0.2266 = **0.6678**, about 66.8%.

WORKED EXAMPLE 3 · A CUTOFF FOR THE TOP 15%

For the same reaction times (μ = 250, σ = 40), what time marks the slowest 15%?

STEP 1 · LEFT-TAIL AREA

The slowest 15% sit above the cutoff, so 85% lie below it: area = 0.8500.

STEP 2 · REVERSE LOOKUP

Φ(1.03) = 0.8485 and Φ(1.04) = 0.8508. The closer is 1.04, off by 0.0008.

STEP 3 · UN-STANDARDIZE

x = 250 + (1.04)(40) = 250 + 41.6 = **291.6 ms**.

CHECK

The cutoff is above the mean and about one SD up, which fits an area of 0.85 between Φ(1.00) = 0.8413 and Φ(1.50) = 0.9332.

WORKED EXAMPLE 4 · A MIDDLE 90% INTERVAL AND A Z COMPARISON

Find the middle 90% of reaction times (μ = 250, σ = 40). Then compare a time of 310 ms here with a sprint time of 11.2 s where μ = 12.0 and σ = 0.4 (lower is faster).

STEP 1 · TAILS

10% left over, 5% in each tail, so z = ±1.645.

STEP 2 · BOUNDS

250 ± 1.645(40) = 250 ± 65.8: **184.2 to 315.8 ms**.

STEP 3 · TWO Z-SCORES

Reaction: (310 − 250) ÷ 40 = 1.50. Sprint: (11.2 − 12.0) ÷ 0.4 = −2.00.

STEP 4 · READ THEM

The reaction time is 1.50 SD slower than average. The sprint is 2.00 SD faster than average, the more unusual performance.

TRAPS PEOPLE FALL INTO

- **Treating curve height as probability.** Only area is probability. A single exact value has probability 0.
- **Using the empirical rule on skewed data.** It holds for approximately normal data only.
- **Reading a right tail straight from the table.** The table gives left areas. Subtract from 1.
- **Turning "top 10%" into an area of 0.10.** The cutoff has 90% below it. Look up 0.90.
- **Dropping the minus sign.** A percentile below 50 has a negative z and lands below the mean.
- **Using 2 for the exact 95% middle.** The exact value is 1.96. 2 gives 95.45%.
- **Believing standardizing fixes shape.** z-scores of skewed data are still skewed.
- **Mixing "middle 95%" with "95% confidence."** One describes individuals with known μ and σ; the other is about estimating an unknown parameter.

Chapter 7 self-test · ten items

1. Scores are normal with μ = 100 and σ = 15. About what percent fall between 70 and 115?

A. 68%

B. 95%

C. 47.5%

D. 81.5%

D. 81.5% 70 is μ − 2σ and 115 is μ + 1σ: 47.5% + 34% = 81.5%.

2. What does the z-table entry $\Phi(z)$ give?

A. The area to the right of z
B. The area to the left of z
C. The height of the curve at z
D. The area between $-z$ and z

B. The area to the left of z Standard tables report cumulative left-tail area, $P(Z < z)$.

3. X is normal with $\mu = 50$ and $\sigma = 5$. What is $P(X > 58)$?

A. 0.9452
B. 0.1600
C. 0.0548
D. 0.4364

C. 0.0548 $z = (58 - 50) \div 5 = 1.60$, $\Phi(1.60) = 0.9452$, and the right tail is 0.0548.

4. X is normal with $\mu = 20$ and $\sigma = 4$. What is $P(16 < X < 26)$?

A. 0.9332
B. 0.0919
C. 0.7745
D. 0.2255

C. 0.7745 $z = -1.00$ and 1.50. $\Phi(1.50) - \Phi(-1.00) = 0.9332 - 0.1587 = 0.7745$.

5. $\Phi(1.20) = 0.8849$. What is $\Phi(-1.20)$?

A. 0.1151
B. 0.8849
C. −0.8849
D. 0.3849

A. 0.1151 By symmetry, $\Phi(-z) = 1 - \Phi(z) = 1 - 0.8849 = 0.1151$.

6. What z-score marks the top 5% of a normal distribution?

A. 1.96
B. −1.645
C. 0.05
D. 1.645

D. 1.645 The cutoff has 95% below it, and $\Phi(1.645) = 0.9500$.

7. $\mu = 500$ and $\sigma = 100$. What value is the 10th percentile?

A. 628
B. 490
C. 372
D. 400

C. 372 The 10th percentile has $z \approx -1.28$, so $x = 500 + (-1.28)(100) = 372$.

8. Which two values enclose the middle 95% when μ = 40 and σ = 5?

A. 35.0 and 45.0
B. 30.0 and 50.0
C. 31.8 and 48.2
D. 30.2 and 49.8

D. 30.2 and 49.8 40 ± 1.96(5) = 40 ± 9.8. Using 1.645 gives the middle 90%.

9. A right-skewed variable is converted to z-scores. What is the shape of the z-scores?

A. Still right-skewed
B. Exactly normal
C. Symmetric but flat
D. Left-skewed

A. Still right-skewed Standardizing subtracts and divides by constants. It changes center and scale, never shape.

10. For a continuous normal variable, how does P(X ≤ 30) compare with P(X < 30)?

A. P(X ≤ 30) is larger
B. They are equal
C. P(X < 30) is larger
D. It depends on σ

B. They are equal A single point has zero area, so including or excluding 30 changes nothing.

CHAPTER 8

Sampling Distributions and the Central Limit Theorem

Read it in full: **Chapter 8 in the textbook**. Practice it in the lab: **Central Limit Theorem** and **Standard Error**.

What to know

Keep three distributions apart. The **population distribution** is the spread of individual values in the whole group (μ and σ). The **sample data** are the n values you collected ($\bar{x}$ and s), and their histogram looks like the population. The **sampling distribution** is the distribution of a statistic, such as $\bar{x}$, across every possible sample of size n. One dot in it is one whole sample reduced to one number. You never see it; you derive it.

Because the sample is random, a statistic is a **random variable**, and its spread is **sampling variability**, the ordinary fact that honest samples disagree. Build one by hand: from the population 2, 4, 6, 8 ($\mu = 5$, $\sigma^2 = 5$), the 16 samples of size 2 with replacement give means from 2 to 8, peaked at 5, with mean exactly 5 and variance 2.5. Even a flat population gives a peaked distribution of $\bar{x}$.

For a simple random sample, $\mu_{\bar{x}} = \mu$, so $\bar{x}$ is an **unbiased** estimator: its errors center on zero, though any one $\bar{x}$ almost surely misses. And $\sigma_{\bar{x}} = \sigma \div \sqrt{n}$, given independence (with replacement, or n under about 10% of N). The population shape affects only the shape of $\bar{x}$'s distribution, never its center or spread. A normal population gives an exactly normal $\bar{x}$ for any n.

Heart rates with $\mu = 74$ and $\sigma = 12$: for $n = 36$, $\sigma_{\bar{x}} = 12 \div 6 = 2$, and $P(\bar{x} > 77) = P(Z > 1.50) = 0.0668$. One adult above 77 is common ($z = 0.25$, about 40%), but a mean of 36 above 77 is rare. Never divide by $\sqrt{n}$ when the question is about one individual: find the noun, then pick σ or $\sigma \div \sqrt{n}$.

The **Central Limit Theorem**: for independent draws from **any** population with a finite σ, $\bar{x}$ is approximately Normal(μ, $\sigma \div \sqrt{n}$) when n is large, and the sum is approximately Normal($n\mu$, $\sigma\sqrt{n}$). It does not make the population or your data normal, it is not the law of large numbers (which says $\bar{x}$ goes to μ), and it fails without finite variance (the Cauchy distribution). $n \geq 30$ is a rule of thumb: about 5 to 10 is enough for symmetric populations, and strong skew may need 100 to 300 or more.

In the book's skewed call example, $\mu = 4.0$ and $\sigma = 6.0$ minutes with $n = 100$: SE = 0.60, $z = (5.0 - 4.0) \div 0.60 \approx 1.67$, so $P(\bar{x} > 5.0) \approx 0.0475$. The same 4.8% is the chance the 100 calls total more

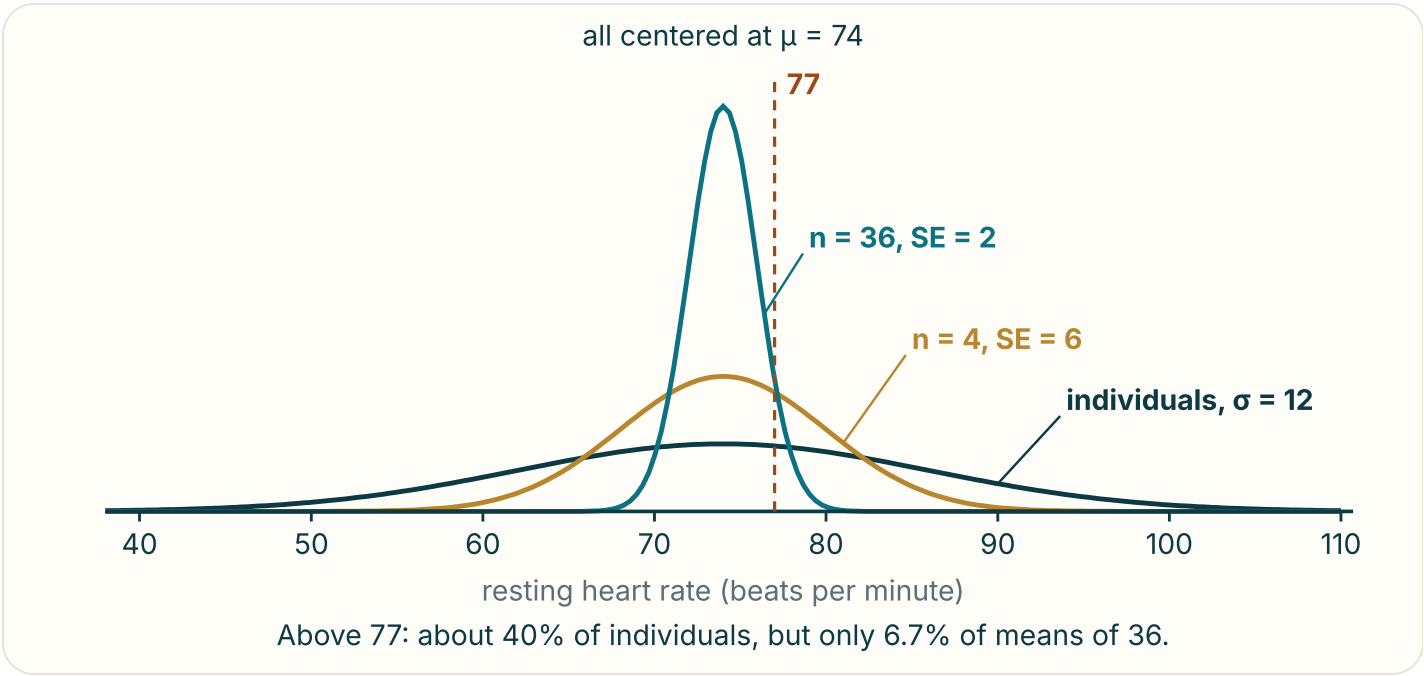
than 500 minutes. Judge whether the CLT applies from n and the severity of skew, not from your sample histogram.

The **standard error** is the SD of a statistic's sampling distribution: $SE(\bar{x}) = \sigma \div \sqrt{n}$, estimated by $s \div \sqrt{n}$. SD describes individuals and does not shrink with n; SE describes an estimate and falls as $1 \div \sqrt{n}$. Quadruple n to halve the SE. To reach a target SE, $n = (s \div SE)^2$, rounded up. Population size N does not appear, as long as the sample is a small fraction.

A **proportion** is a mean of 0s and 1s: $\hat{p} = x \div n$, $\mu_{\hat{p}} = p$, and $SE(\hat{p}) = \sqrt{[p(1 - p) \div n]}$. The shape is approximately normal when the **success-failure condition** holds: $np \geq 10$ and $n(1 - p) \geq 10$. Variability peaks at $p = 0.5$, the safe planning value. A margin of error is about $2 \times SE$ and covers random sampling error only, never bias.

LOCK THESE IN

- One dot in a sampling distribution is one whole sample.
- $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma \div \sqrt{n}$ (independence: n under 10% of N).
- One person: divide by σ. A sample mean: divide by $\sigma \div \sqrt{n}$.
- CLT: $\bar{x} \approx \text{Normal}(\mu, \sigma \div \sqrt{n})$ for large n, any shape with finite σ.
- The CLT never makes the population or the data normal.
- SE shrinks as $1 \div \sqrt{n}$; SD does not shrink with n.
- Proportions: $\mu_{\hat{p}} = p$; $SE = \sqrt{[p(1 - p) \div n]}$; need $np \geq 10$ and $n(1 - p) \geq 10$.
- Margin of error ≈ $2 \times SE$ and ignores bias.



Averages vary less than individuals. The sampling distribution of $\bar{x}$ keeps the center $\mu = 74$ but narrows to $\sigma \div \sqrt{n}$: 6 for $n = 4$ and 2 for $n = 36$. A threshold that 40% of people pass is cleared by only 6.7% of sample means.

Formulas and when to use them

Quantity	Formula	When to use it
Center of $\bar{x}$	$\mu_{\bar{x}} = \mu$	Always; $\bar{x}$ is unbiased
Spread of $\bar{x}$	$\sigma_{\bar{x}} = \sigma \div \sqrt{n}$	Independent draws
z for a mean	$z = (\bar{x} - \mu) \div (\sigma \div \sqrt{n})$	Questions about a sample mean
z for one value	$z = (x - \mu) \div \sigma$	Questions about one individual
Sum of n draws	Normal($n\mu$, $\sigma\sqrt{n}$)	Totals, by the CLT
Estimated SE	$s \div \sqrt{n}$	σ unknown
n for a target SE	$n = (s \div SE)^2$, round up	Planning precision
Finite population correction	$\times \sqrt{[(N - n) \div (N - 1)]}$	Sample over 10% of N
Proportion center and SE	$\mu_{\hat{p}} = p$; $\sqrt{[p(1 - p) \div n]}$	Categorical data
Success-failure	$np \geq 10$ and $n(1 - p) \geq 10$	Before using a normal model for $\hat{p}$
Margin of error (rough)	$\approx 2 \times SE$	Random error only

Shape of the sampling distribution of $\bar{x}$

Population	Center and spread	Shape of $\bar{x}$
Normal	μ and $\sigma \div \sqrt{n}$	Exactly normal for any n
Symmetric, not normal	μ and $\sigma \div \sqrt{n}$	Normal by about n = 5 to 10
Moderately skewed	μ and $\sigma \div \sqrt{n}$	Normal by about n = 30
Strongly skewed	μ and $\sigma \div \sqrt{n}$	May need 100 to 300 or more
No finite variance	Undefined	The CLT fails

Worked examples

WORKED EXAMPLE 1 · A SAMPLING DISTRIBUTION BUILT BY HAND

A population holds the values 1, 3, 5. List every sample of size 2 drawn with replacement, and find the mean and variance of $\bar{x}$.

STEP 1 · PARAMETERS

$\mu = 9 \div 3 = 3$. Squared deviations 4, 0, 4 sum to 8, so $\sigma^2 = 8 \div 3 \approx 2.667$.

STEP 2 · THE 9 SAMPLES

(1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5), each with probability 1/9.

STEP 3 · THEIR MEANS

$\bar{x}$ values: 1 (1 way), 2 (2 ways), 3 (3 ways), 4 (2 ways), 5 (1 way).

STEP 4 · MEAN OF $\bar{X}$

$\Sigma\bar{x} \div 9 = 27 \div 9 =$ **3**, exactly μ.

STEP 5 · VARIANCE OF $\bar{X}$

$\Sigma(\bar{x} - 3)^2 \div 9 = 12 \div 9 \approx$ **1.333**, which is $\sigma^2 \div 2 = 2.667 \div 2$. The square-root law falls out of counting.

WORKED EXAMPLE 2 · ONE PERSON VERSUS A SAMPLE MEAN

Test scores are normal with $\mu = 500$ and $\sigma = 100$. Find P(one student scores above 540) and P(the mean of 25 students is above 540).

STEP 1 · ONE STUDENT

$z = (540 - 500) \div 100 = 0.40$, so $P = 1 - 0.6554 =$ **0.3446**.

STEP 2 · SPREAD OF $\bar{X}$

$\sigma_{\bar{x}} = 100 \div \sqrt{25} = 100 \div 5 = 20$.

STEP 3 · STANDARDIZE THE MEAN

$z = (540 - 500) \div 20 = 2.00$.

STEP 4 · THE TAIL

$P(\bar{x} > 540) = 1 - 0.9772 =$ **0.0228**.

READ IT

About a third of students clear 540, but a class average of 25 clears it only about 2% of the time. The population is normal, so $\bar{x}$ is exactly normal here.

WORKED EXAMPLE 3 · THE CLT ON A SKEWED POPULATION

Clinic wait times are right-skewed with $\mu = 30$ and $\sigma = 24$ minutes. For a random sample of $n = 64$ waits, find $P(\bar{x} < 25)$ and the matching statement about the total.

STEP 1 · CHECK

The population is skewed, but $n = 64$ is large, the waits are independent and 64 is a small share of all waits. The CLT applies.

STEP 2 · SE

$24 \div \sqrt{64} = 24 \div 8 =$ **3.0**.

STEP 3 · STANDARDIZE

$z = (25 - 30) \div 3.0 = -1.6667$, which rounds to -1.67.

STEP 4 · THE TABLE

$\Phi(-1.67) =$ **0.0475**.

STEP 5 · AS A TOTAL

A mean under 25 for 64 waits is the same event as a total under 1,600 minutes, so that also has probability 0.0475.

WORKED EXAMPLE 4 · A SAMPLE PROPORTION

Suppose 35% of a large city's adults bike to work at least once a month. In a random sample of 300, what is $P(\hat{p} > 0.40)$?

STEP 1 · CONDITIONS

$np = 300(0.35) = 105$ and $n(1 - p) = 195$, both at least 10. The sample is far under 10% of the city.

STEP 2 · SE

$\sqrt{(0.35 \times 0.65 \div 300)} = \sqrt{0.000758} \approx$ **0.0275**.

STEP 3 · STANDARDIZE

$z = (0.40 - 0.35) \div 0.0275 \approx 1.82$.

STEP 4 · RIGHT TAIL

$1 - \Phi(1.82) = 1 - 0.9656 =$ **0.0344**.

STEP 5 · PLAN A BIGGER STUDY

To cut the SE to 0.02, solve $\sqrt{[p(1 - p) \div n]} = 0.02$: $n = 0.35 \times 0.65 \div 0.02^2 = 568.75$, so **569** adults.

TRAPS PEOPLE FALL INTO

- **Calling the sample histogram the sampling distribution.** Your data show individuals. The sampling distribution shows statistics from many samples.
- **Dividing by $\sqrt{n}$ for one individual.** One person uses σ. Only a sample mean uses $\sigma \div \sqrt{n}$.

- **Saying the CLT makes the data normal.** Your 5,000 incomes stay skewed. Only the distribution of $\bar{x}$ goes normal.
- **Treating n ≥ 30 as a law.** Heavier skew needs more; normal populations need none.
- **Mixing up SD and SE.** Reporting mean ± SE to describe people makes the data look far tidier than they are.
- **Thinking SE depends on population size.** With a small sampling fraction, N drops out.
- **Skipping the success-failure check.** With p near 0 or 1 and small n, the normal model for $\hat{p}$ fails.
- **Trusting the margin of error to cover bias.** It counts random sampling error only.

Chapter 8 self-test · ten items

1. What does one dot in the sampling distribution of $\bar{x}$ represent?

A. One individual in the population
B. The mean of one whole sample
C. One value in your data set
D. The population mean μ

B. The mean of one whole sample A sampling distribution collects a statistic from many samples, so each dot is one sample's mean.

2. A population has μ = 80 and σ = 20. For n = 16, what are the center and spread of $\bar{x}$?

A. 80 and 20
B. 5 and 80
C. 80 and 1.25
D. 80 and 5

D. 80 and 5 $\mu_{\bar{x}} = \mu = 80$ and $\sigma_{\bar{x}} = 20 \div \sqrt{16} = 5$.

3. Weights are normal with μ = 170 and σ = 30. What is P(x̄ > 176) for n = 100?

A. 0.4207
B. 0.9772
C. 0.2000
D. 0.0228

D. 0.0228 SE = 30 ÷ 10 = 3, z = (176 − 170) ÷ 3 = 2.00, and 1 − 0.9772 = 0.0228.

4. A population is strongly right-skewed. What does the CLT say about x̄ for large n?

A. The population itself becomes normal
B. The sample data become normal as n grows
C. Its distribution is approximately normal
D. x̄ equals μ exactly once n is large

C. Its distribution is approximately normal Only the distribution of the sample mean approaches normal. The population and the data keep their skew.

5. A study quadruples its sample size. What happens to the SE of x̄?

A. It is cut to a quarter
B. It is cut in half
C. It doubles
D. It stays the same

B. It is cut in half SE = σ ÷ √n, and √4 = 2.

6. A sample of 64 has s = 16. What is the estimated SE of $\bar{x}$?

A. 4
B. 2
C. 0.25
D. 16

B. 2 $s \div \sqrt{n} = 16 \div 8 = 2$.

7. p = 0.5 and n = 100. What is SE($\hat{p}$)?

A. 0.25
B. 0.005
C. 0.50
D. 0.05

D. 0.05 $\sqrt{(0.5 \times 0.5 \div 100)} = \sqrt{0.0025} = 0.05$.

8. Which sample meets the success-failure condition for p = 0.04?

A. n = 300
B. n = 100
C. n = 200
D. n = 50

A. n = 300 np must be at least 10: $300 \times 0.04 = 12$, while $200 \times 0.04 = 8$ falls short.

9. Which statement contrasts SD and SE correctly?

A. SD describes individuals; SE describes an estimate
B. SD shrinks as n grows; SE stays the same size
C. SD and SE become the same once n is large
D. SE describes individuals; SD describes an estimate

A. SD describes individuals; SE describes an estimate The SD settles toward σ as n grows. The SE shrinks as 1 ÷ √n.

10. s = 15. What sample size gives an SE of 1.5?

A. 10
B. 225
C. 100
D. 150

C. 100 n = $(15 \div 1.5)^2 = 10^2 = 100$.

ONE PAGE

Cram sheet

Everything from chapters 7 and 8 in one place. Cover the right column and see if you can rebuild each row.

Formulas

Name	Formula	Note
z-score	$z = (x - \mu) \div \sigma$	$x = \mu + z\sigma$
Areas	Left $\Phi(z)$; right $1 - \Phi(z)$; between $\Phi(z_b) - \Phi(z_a)$	$\Phi(-z) = 1 - \Phi(z)$
Key z	1.28, 1.645, 1.96, 2.326, 2.576	90th, 95th, 97.5th, 99th, 99.5th percentiles
Mean of $\bar{x}$	$\mu_{\bar{x}} = \mu$	Unbiased
SE of $\bar{x}$	$\sigma \div \sqrt{n}$ (or $s \div \sqrt{n}$)	4 × n halves it
z for $\bar{x}$	$(\bar{x} - \mu) \div (\sigma \div \sqrt{n})$	Sample mean questions
Sum	Normal($n\mu$, $\sigma\sqrt{n}$)	CLT for totals
Proportion	$\mu_{\hat{p}} = p$; SE = $\sqrt{[p(1 - p) \div n]}$	$np \geq 10$, $n(1 - p) \geq 10$
Planning n	$n = (s \div SE)^2$	Round up

Rules that are not formulas

Topic	The rule
Empirical rule	68, 95, 99.7; blocks 34, 13.5, 2.35, 0.15.
Area	Area is probability; a single point has probability 0.
Backward	Area → z → x; "top 10%" is a left area of 0.90.
Shape	Standardizing never fixes skew.
Three distributions	Population, sample data, sampling distribution.
Noun check	One person: σ. Sample mean: σ ÷ √n.
CLT	Any shape with finite σ; n ≥ 30 is a rule of thumb.
SD vs SE	SD for people; SE for estimates.

Dr. Garg Learning · CC BY-SA 4.0