Five-minute review

1. What does at most 3 mean?

2. n=10,p=.6; mean.

Write your answers before we discuss them.

Normal curves and the empirical rule

What we will learn

Interpret normal models and estimate areas by the empirical rule.

Why does it matter?

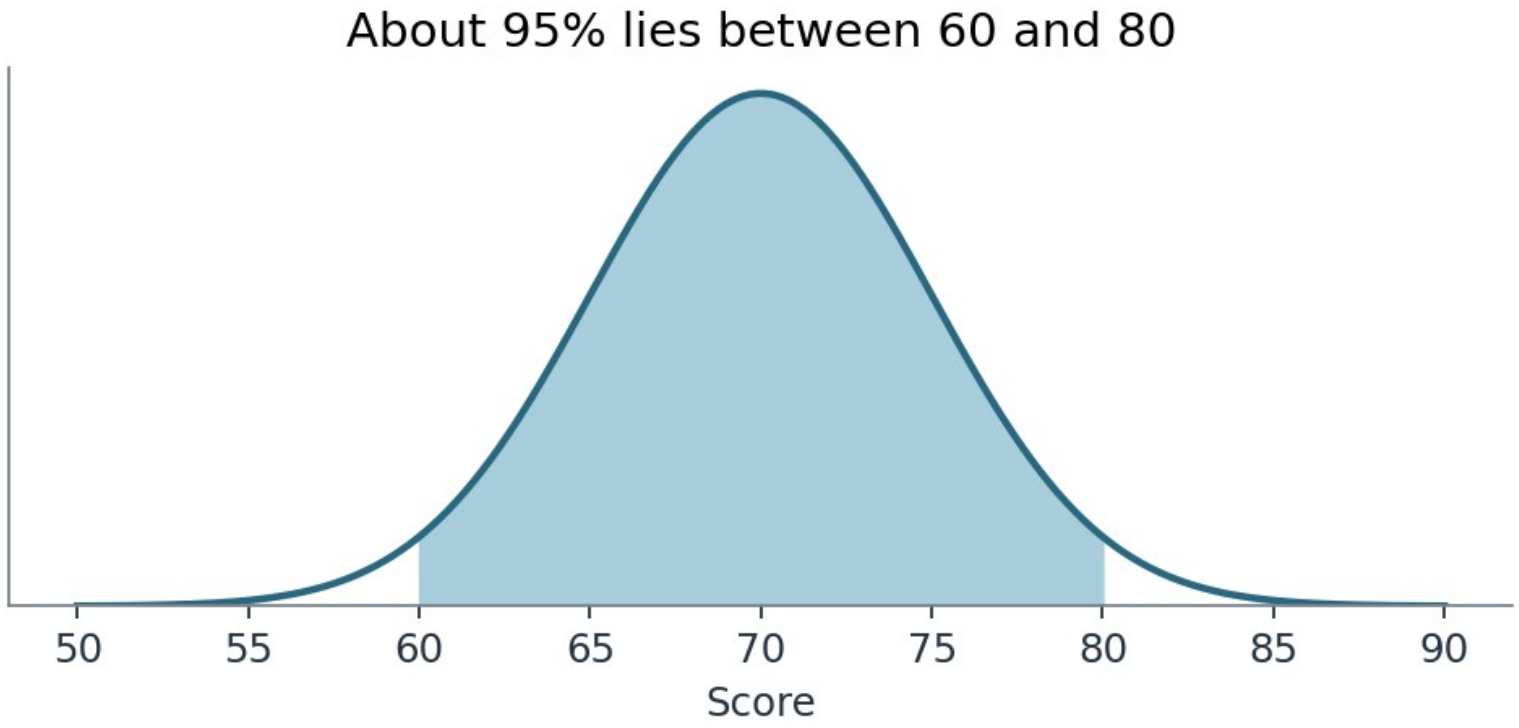
For an approximately normal distribution, the empirical rule gives quick estimates without looking up a table.

Definitions: Normal curves and the empirical rule

Normal distribution: a continuous symmetric bell-shaped probability model specified by mean μ and SD $\sigma>0$.

Empirical rule: about 68%, 95% and 99.7% lie within one, two and three SDs of the mean.

Illustration and interpretation



Notice the pattern

For an approximately normal distribution, the empirical rule gives quick estimates without looking up a table.

Illustrative data / model

Formulas

$$\mu \pm \sigma:\ 68\%, \quad \mu \pm 2\sigma:\ 95\%, \quad \mu \pm 3\sigma:\ 99.7\%$$

Important ideas and notation

A normal distribution is a continuous, symmetric bell-shaped model determined by μ and $\sigma>0$; total area 1.

Approximately 68%, 95%, 99.7% lie within 1, 2, 3 standard deviations of μ. Each tail outside $\pm1\sigma$ is about 16%; outside $\pm2\sigma$ about 2.5%.

Use the rule only for approximately normal data. Check histogram, outliers and context; symmetric does not necessarily mean normal.

Example 5.1.1

Normal model $\mu=70$, $\sigma=5$. Estimate $P(60 \le X \le 80)$.

Solution:

Within 2 SD: ≈95%.

Example 5.1.2

Same model, P(X>75).

Solution:

Above +1 SD: $\frac{100-68}{2} \approx 16\%$.

Example 5.1.3 • A normal model for test scores

If μ=70 and σ=5, roughly what percentage is above 80?

Score interval	Approx. percent
65 to 75	68%
60 to 80	95%
55 to 85	99.7%

Illustrative classroom dataset

Example 5.1.3 • Solution

If $\mu=70$ and $\sigma=5$, roughly what percentage is above 80?

Solution:

About 2.5%, half of the 5% outside two SDs.

Individual practice

1. $\mu=100$, $\sigma=10$; within 1 SD interval.

2. $P(X<80)$ for that model.

3. Would strongly right-skewed waiting times fit the rule?

Group activity

For $\mu=50$, $\sigma=4$, shade areas 46–54, 42–58, and beyond 62.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. 90–110; about 68%.

2. About 2.5%.

3. Not without evidence of approximate normality.

Group solution

Solution:

68%, 95%, and right tail ≈.15%.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Why is P(X=70)=0 in a continuous model?

2. A normal model has μ=100 and σ=10. Estimate the percentage between 80 and 120.

Show your reasoning. Submit before leaving.

Five-minute review

1. Why is $P(X=70)=0$ in a continuous model?

2. $\mu=100$, $\sigma=10$; within 1 SD interval.

Write your answers before we discuss them.

Z-scores and relative standing

What we will learn

Standardize observations and compare relative standing.

Why does it matter?

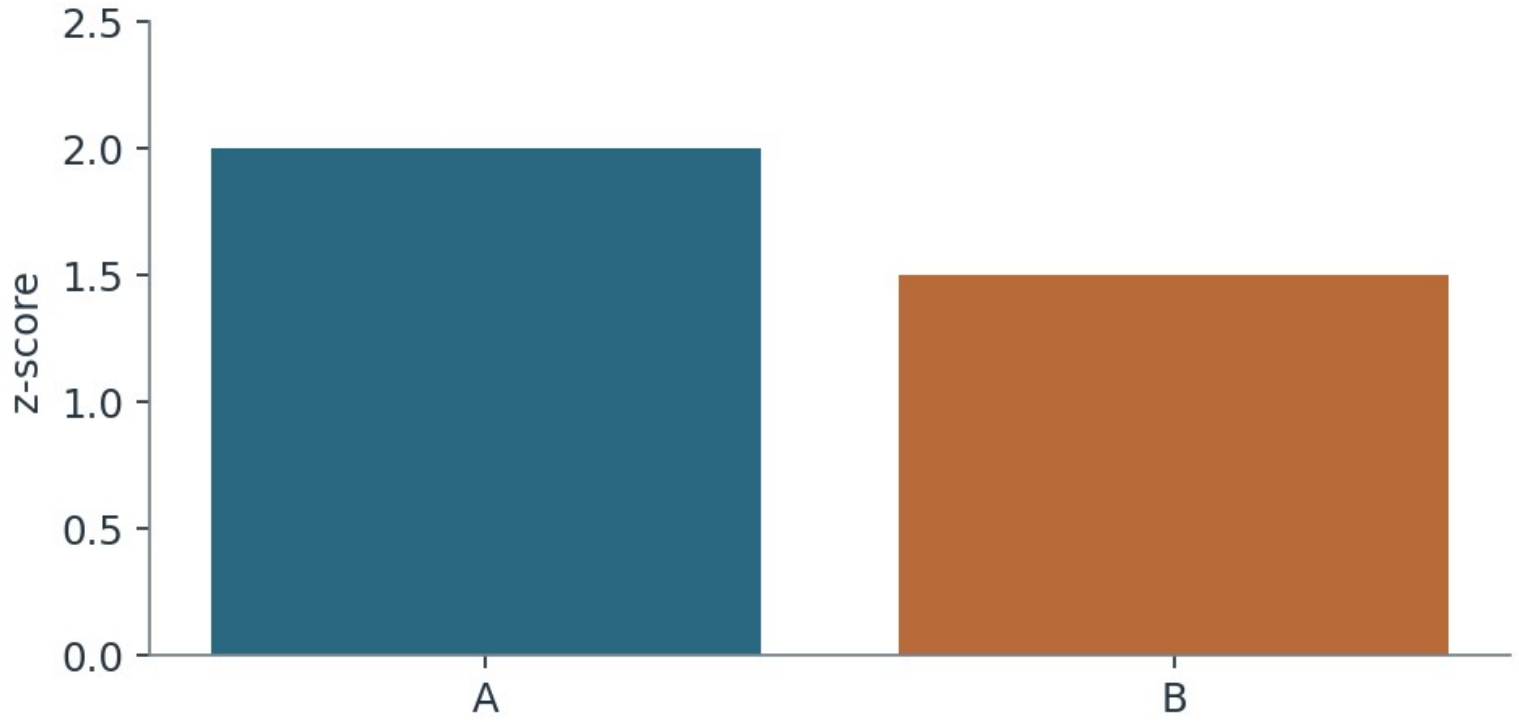
Raw scores can be misleading when the tests have different centers and spreads. Standardize to compare relative position.

Definitions: Z scores and relative standing

Z-score: the signed number of standard deviations a value lies from the mean.

Standardization: converting observations to a common scale centered at zero and measured in SD units.

Illustration and interpretation



Notice the pattern

Raw scores can be misleading when the tests have different centers and spreads. Standardize to compare relative position.

Illustrative data / model

Formulas

$$z = \frac{x - \mu}{\sigma}, \quad x = \mu + z\sigma$$

Important ideas and notation

For population model $z=(x-\mu)/\sigma$. For descriptive sample standardization use $(x-\bar{x})/s$. A z-score has no units.

$z>0$ above mean; $z<0$ below mean; abs(z) gives distance in SD units. Rearrange $x=\mu+z\sigma$.

Z-scores compare standardized position; a higher score is not necessarily better for quantities such as race time.

Example 5.2.1

Class A score 85, $\mu=75$, $\sigma=5$; B score 90, $\mu=84$, $\sigma=4$. Compare the two scores using z-scores.

Solution:

$z_A=2$, $z_B=1.5$. A is higher relative to its class.

Example 5.2.2

$\mu=50$, $\sigma=8$, $z=-1.5$. Find the original value x.

Solution:

$x=50-12=38$.

Example 5.2.3 • Comparing scores from different tests

Which score is farther above its own mean in SD units?

Class	Score	Mean	SD
A	85	75	5
B	90	84	4

Illustrative classroom dataset

Example 5.2.3 • Solution

Which score is farther above its own mean in SD units?

Solution:

A: z=2 versus B: z=1.5.

Individual practice

1. $x=70$, $\mu=60$, $\sigma=5$.

2. $x=18$, $\mu=24$, $\sigma=3$.

3. $\mu=100$, $\sigma=12$, $z=.5$.

Group activity

Two runners: A 48s, $\mu=52$s, $\sigma=2$s; B 45s, $\mu=48$s, $\sigma=3$s.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. $z=2$.

2. $z=-2$.

3. 106.

Group solution

Solution:

$z_A = -2$

$z_B = -1$

A is faster relative to group, even though raw time is slower.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. What is the z-score at the mean?

2. A score is 92 in a model with mean 80 and SD 6. Find and interpret z.

Show your reasoning. Submit before leaving.

Five-minute review

1. What is the z-score at the mean?

2. $x=70$, $\mu=60$, $\sigma=5$.

Write your answers before we discuss them.

Normal probabilities and z-tables

What we will learn

Calculate left, right and interval probabilities using a cumulative z-table.

Why does it matter?

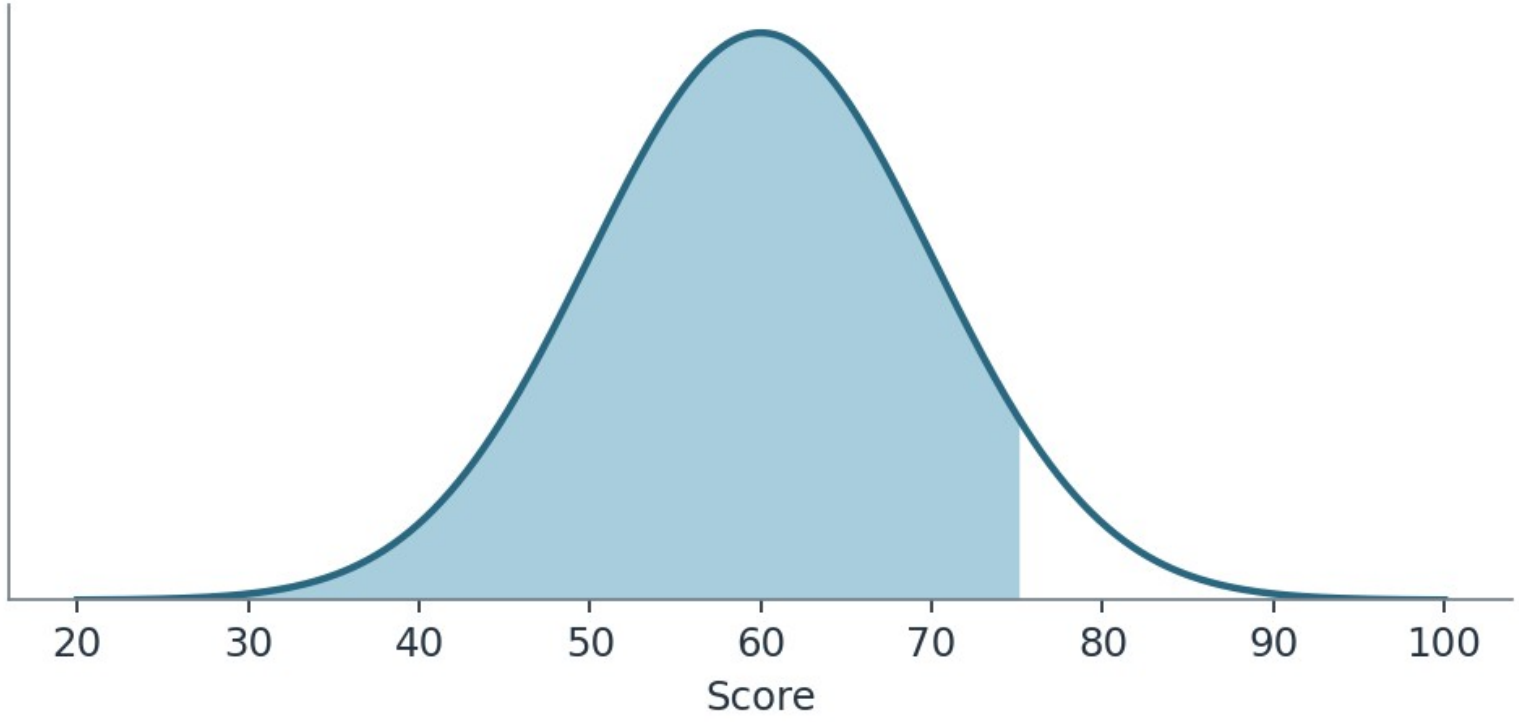
For a normal model, standardize the score and use the area to the left. A right-tail probability is its complement.

Definitions: Normal probabilities and z tables

Standard normal distribution: normal model with mean 0 and SD 1.

Cumulative normal probability $\Phi(z)$: the area to the left of z under the standard normal curve.

Illustration and interpretation



Notice the pattern

For a normal model, standardize the score and use the area to the left. A right-tail probability is its complement.

Illustrative data / model

Formulas

$$P(Z > z) = 1 - \Phi(z)$$

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

Important ideas and notation

$\Phi(z)=P(Z\le z)$, where Z is standard normal. Use a left-tail table, spreadsheet NORM.S.DIST(z,TRUE), or R pnorm(z).

Right tail = $1-\Phi(z)$; interval = $\Phi(z_2)-\Phi(z_1)$. Convert X boundaries to z first.

For a normal distribution, probabilities are areas, not heights. Label units and shade the intended region.

Example 5.3.1

$\mu=60$, $\sigma=10$. $P(X \leq 75)$.

Solution:

$z=1.5$

$\Phi(1.5)=.9332$.

Example 5.3.2

P(50<X<75) for same model.

Solution:

z bounds −1,1.5

.9332−.1587=.7745.

Example 5.3.3 • From a raw score to an area

What fraction is above 75?

Quantity	Value
Mean	60
SD	10
Cutoff	75
Left area	0.9332

Illustrative classroom dataset

Example 5.3.3 • Solution

What fraction is above 75?

Solution:

$1 - 0.9332 = 0.0668$ (6.68%).

Individual practice

1. $P(Z>1.2)$.

2. $P(Z<-.5)$.

3. $P(-1<Z<1)$.

Group activity

Model $\mu=100$, $\sigma=15$. Estimate $P(85<X<130)$ with technology.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. $1-.8849=.1151$.

2. .3085.

3. .6827 approximately.

Group solution

Solution:

$$\Phi(2)-\Phi(-1)=.97725-.15866\approx.81859.$$

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. If a z-table gives area left, how do you find area right?

2. For a normal model with $\mu=50$ and $\sigma=5$, find $P(X>55)$, given $\Phi(1)=0.8413$.

Show your reasoning. Submit before leaving.

Five-minute review

1. If a z-table gives area left, how do you find area right?

2. $P(Z>1.2)$.

Write your answers before we discuss them.

Inverse normal and percentiles

What we will learn

Find normal cutoffs from given areas.

Why does it matter?

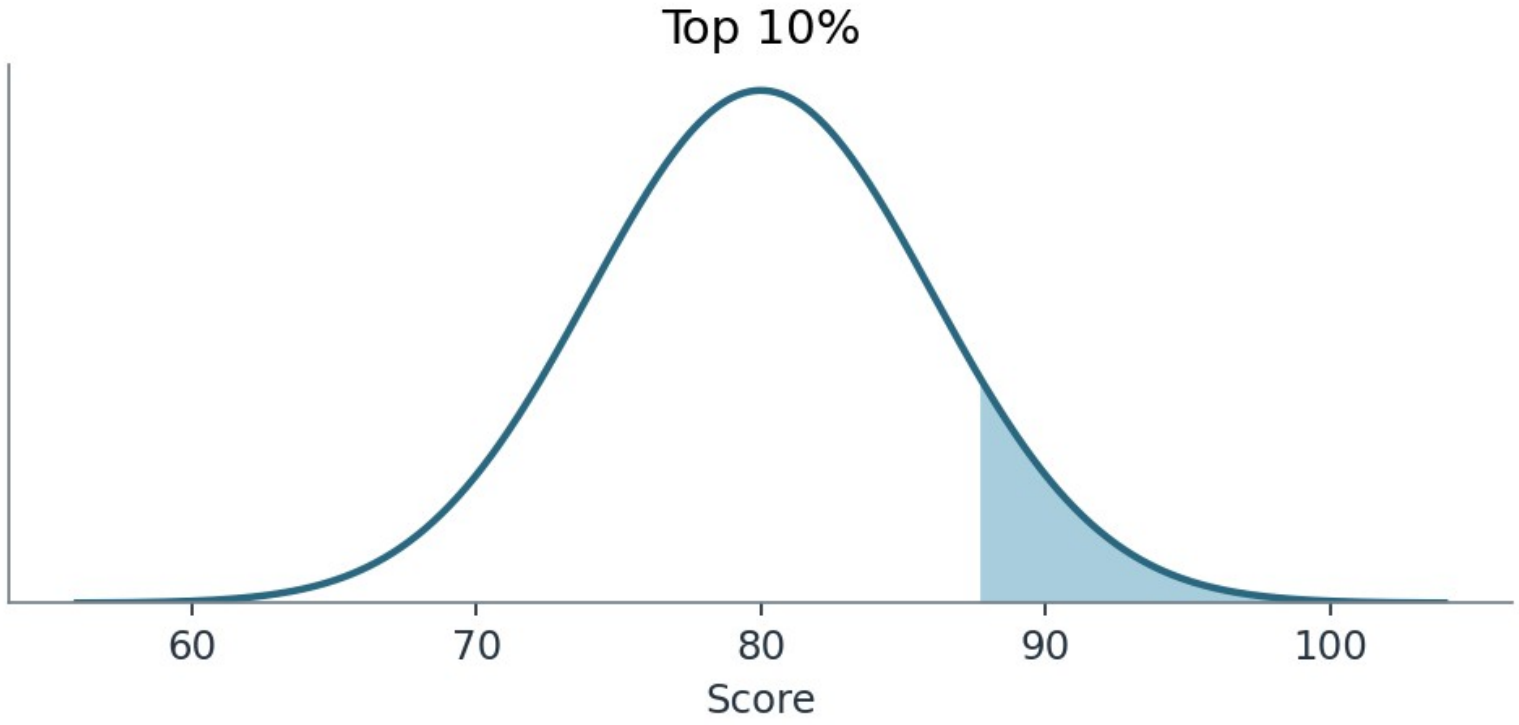
Instead of starting with a raw score, start with the desired area. For the top 10%, 90% lies below the cutoff.

Definitions: Inverse normal and percentiles

Percentile cutoff: a value with a stated proportion of the distribution at or below it.

Inverse normal: finds a z-score from its specified left-tail probability.

Illustration and interpretation



Notice the pattern

Instead of starting with a raw score, start with the desired area. For the top 10%, 90% lies below the cutoff.

Illustrative data / model

Formulas

$$z = \Phi^{-1}(p), \quad x = \mu + z\sigma$$

Important ideas and notation

Inverse normal finds z from a left-tail area: $z=\Phi^{-1}(p)$, $0<p<1$. Use R qnorm(p) or spreadsheet NORM.S.INV(p).

Convert to original units with $x=\mu+z\sigma$. A top-tail fraction r corresponds to left-tail area $1-r$.

Check direction: top 10% cutoff is above the mean; lower 10% below. For central areas, split the excluded area equally between tails.

Example 5.4.1

Scores $\mu=80$, $\sigma=6$. Find top 10% cutoff.

Solution:

$z \approx 1.2816$

$x \approx 87.69$.

Example 5.4.2

Find central 90% for $\mu=50$, $\sigma=4$.

Solution:

$z=\pm 1.6449$

$x\approx 43.42$ to 56.58.

Example 5.4.3 • Finding a cutoff for the top group

What raw score is the cutoff?

Quantity	Value
Mean	80
SD	6
Left area	0.90
z cutoff	1.2816

Illustrative classroom dataset

Example 5.4.3 • Solution

What raw score is the cutoff?

Solution:

$80 + 6(1.2816) \approx 87.69.$

Individual practice

1. Find z for 50th percentile.

2. Find z for 2.5th percentile.

3. $\mu=70$, $\sigma=5$; 97.5th percentile.

Group activity

Set symmetric limits for central 95% of a normal model $\mu=20$, $\sigma=2$.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. 0.

2. -1.960.

3. 79.8.

Group solution

Solution:

$20 \pm 1.96 \times 2 = 16.08$ to 23.92.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. What left-tail area marks the top 5%?

2. A normal model has $\mu=40$, $\sigma=5$. The 90th percentile of Z is 1.2816. Find the cutoff.

Show your reasoning. Submit before leaving.

Five-minute review

1. What left-tail area marks the top 5%?

2. Find z for 50th percentile.

Write your answers before we discuss them.

Sampling distributions and the central limit theorem

What we will learn

Distinguish individual observations and sample means, and use standard error.

Why does it matter?

Individual scores vary more than the averages of independent random samples. Larger samples generally produce more stable estimates.

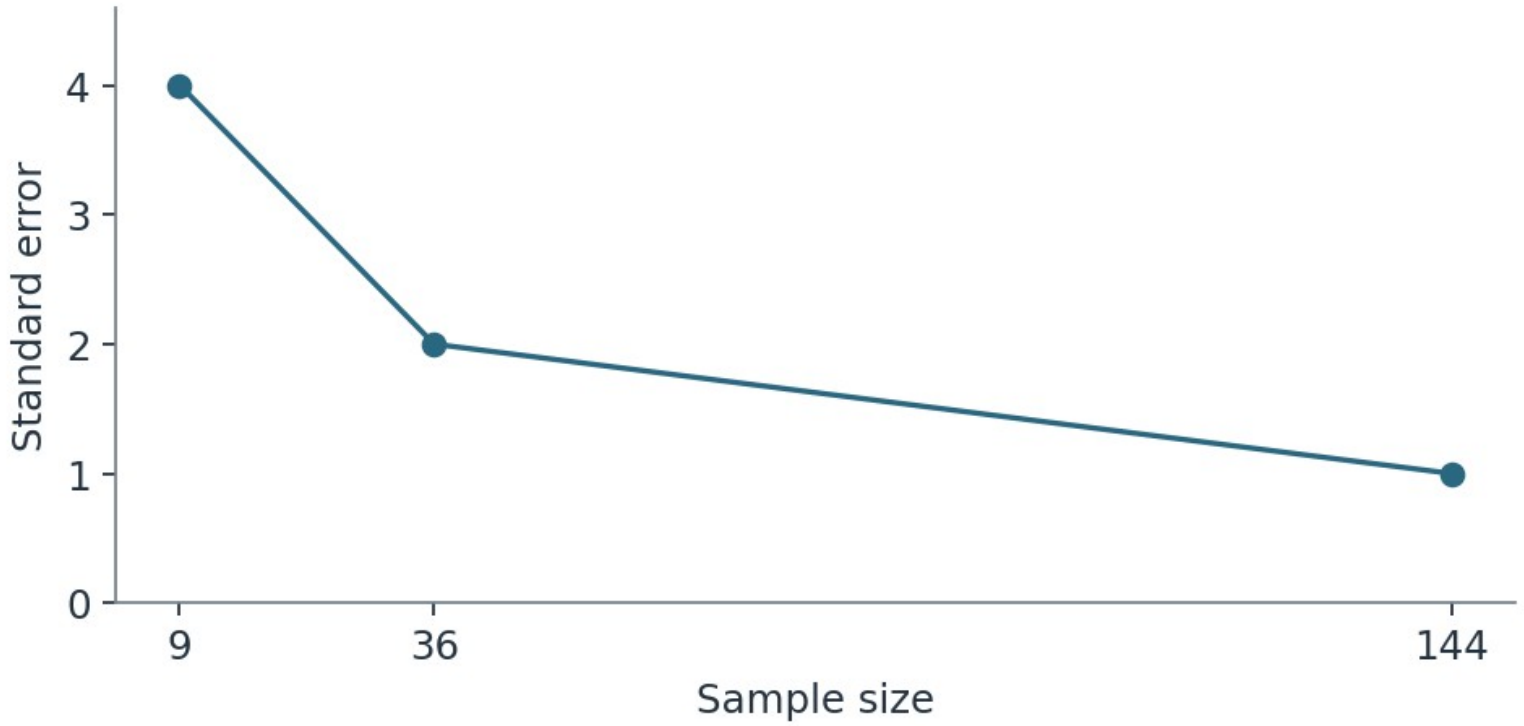
Definitions: Sampling distributions and the central limit theorem

Sampling distribution: distribution of a statistic over repeated random samples of the same size.

Standard error: SD of the sampling distribution of a statistic.

Central limit theorem: under independent identical sampling with finite variance, the distribution of the standardized sample mean approaches normality as n increases.

Illustration and interpretation



Notice the pattern

Individual scores vary more than the averages of independent random samples. Larger samples generally produce more stable estimates.

Illustrative data / model

Formulas

$$E(\bar{x}) = \mu, \quad SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

Important ideas and notation

The sampling distribution describes a statistic across repeated random samples of fixed size. For independent observations, $E(\bar{x})=\mu$ and $SE(\bar{x})=\sigma/\sqrt{n}$.

The central limit theorem: with independent identically distributed observations and finite variance, sample means approach normality as n grows; no universal n=30 guarantee.

Using the method correctly

If population is normal, sample mean is normal at any n. For simple random samples without replacement from a finite population, account for dependence when sampling fraction is large.

Example 5.5.1

$\mu=80$, $\sigma=12$, independent $n=36$. Find the mean and standard error of the sampling distribution, then estimate $P(\bar{x}>84)$.

Solution:

Mean of $\bar{x}=80$

$$SE=\frac{12}{6}=2$$

if normal approximation is valid, $P(\bar{x}>84)=P(Z>2)\approx.0228$.

Example 5.5.2

Compare n=25 and n=100, σ=10. Compare the standard errors.

Solution:

SEs 2 and 1

fourfold n halves SE.

Example 5.5.3 • Averages from repeated samples

What happens to SE when sample size quadruples?

Sample size	Population SD	Standard error
9	12	4
36	12	2
144	12	1

Illustrative classroom dataset

Example 5.5.3 • Solution

What happens to SE when sample size quadruples?

Solution:

It is halved because $SE = \sigma/\sqrt{n}$.

Individual practice

1. $\mu=40$, $\sigma=8$, $n=16$: SE?

2. What is $E(\bar{x})$ when $\mu=40$?

3. If $\sigma=20$, SE=4, find n.

Group activity

Simulate 200 samples of size 5 and 200 of size 30 from a fair die; plot sample means.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. 2.

2. 40.

3. 25.

Group solution

Solution:

Results vary

both centered near 3.5

larger n has smaller spread, often more normal shape.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Is the standard error the SD of individual observations?

2. A population has SD 18. For independent samples of size 81, find the standard error of the mean.

Show your reasoning. Submit before leaving.

Five-minute review

1. Is the standard error the SD of individual observations?

2. $\mu=40$, $\sigma=8$, $n=16$: SE?

Write your answers before we discuss them.

When is a normal model appropriate?

What we will learn

Use shape, context and empirical percentages to assess whether a normal model is reasonable.

Why does it matter?

A normal model is an approximation. Shape and unusual values matter, not just the mean and standard deviation.

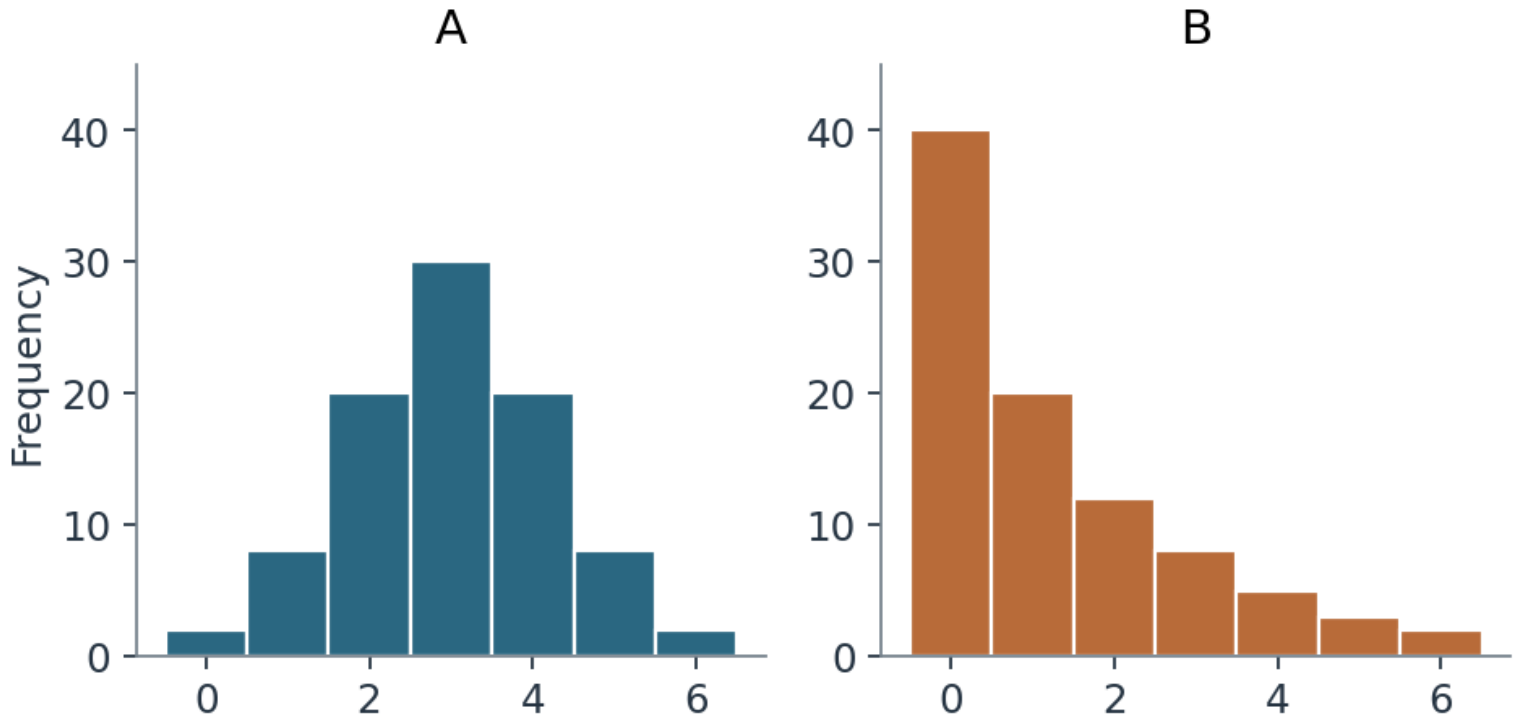
Definitions: When is a normal model appropriate

Model suitability: whether a probability model reasonably describes data for the intended analysis.

Skewness: asymmetry with a longer tail in one direction.

Unimodal distribution: a distribution with one main peak.

Illustration and interpretation



Notice the pattern

A normal model is an approximation. Shape and unusual values matter, not just the mean and standard deviation.

Illustrative data / model

Important ideas and notation

A normal model is continuous, symmetric and unimodal. Its mean and standard deviation determine its location and spread.

Inspect a histogram or dot plot for symmetry, a single central peak, gaps and extreme outliers. Compare observed fractions within 1, 2 and 3 SDs to 68%, 95% and 99.7%.

Using the method correctly

These are diagnostic clues, not a proof of normality. Strong skew, multiple clusters or a hard boundary near the center can make normal-model estimates misleading.

Example 5.6.1

In a sample of 100, 68 values are within one SD, 94 within two and 99 within three. The histogram is symmetric and unimodal. Comment.

Solution:

The fractions are close to the empirical rule

a normal model may be reasonable, though not proven.

Example 5.6.2

Commute times are mostly 5–20 minutes with a few above 90. Should we automatically use the normal model?

Solution:

No. A long right tail and extreme observations suggest a skewed distribution.

Example 5.6.3 • Does the bell curve fit?

Which histogram is more plausibly normal?

Bin	A	B
1	2	40
2	8	20
3	20	12
4	30	8
5	20	5
6	8	3
7	2	2

Illustrative classroom dataset

Example 5.6.3 • Solution

Which histogram is more plausibly normal?

Solution:

A

it is symmetric and mound-shaped, unlike the right-skewed B.

Individual practice

1. A bell-shaped distribution has mean 50 and SD 5. Estimate the fraction between 40 and 60.

2. A histogram has two separated peaks. Is a single normal model likely adequate?

3. In a sample of 200, 130 are within one SD. Calculate the observed percentage.

Group activity

Compare two illustrative distributions: A has bin counts 2, 8, 20, 30, 20, 8, 2 in equal-width bins; B has counts 40, 20, 12, 8, 5, 3, 2. Sketch histograms and justify which is more suitable for a normal model.

Show your calculation or graph. Write a sentence explaining the result in context.

Check your answers

1. About 95%, if the model is appropriate.

2. Usually not; investigate separate groups or other models.

3. 65%; compare with 68% together with the shape and context.

Group solution

Solution:

A is symmetric and unimodal, so is the better candidate. B has a long right tail. Neither histogram alone proves a particular distribution.

Compare your conclusion with the question. Check your units and explain what the numbers represent.

Exit ticket • 5 minutes

1. Why should we check the shape before using a normal probability table?

2. House prices show many moderate values and a few extremely expensive properties. Is a symmetric normal model automatically reasonable?

Show your reasoning. Submit before leaving.