Normal distributions and the Central Limit Theorem

Lecture slides PDF pp. 1–95

Quick exam-prep guide p. 1

Probability and Statistics for the Life Sciences

Measurements vary between individuals, and averages vary between samples. Chapter 6 gives us probability models for both. The distinction matters: the spread of individual observations is not the spread of a sample mean.

These notes follow the current lecture slides in order, combining repeated reveals into one explanation. They supplement the existing four-page review and group problems. Numerical examples use the slides' stated models and data; they are exercises, not claims about current populations.

What you should be able to do

- Translate a continuous probability into an area or a difference of CDF values.
- Standardize a normal variable and find probabilities or percentiles.
- Approximate binomial counts with a normal model when appropriate, including a continuity correction.
- Identify whether a question concerns an individual, a total or a mean, and use the corresponding SD or SE.
- Distinguish exact normal results from CLT approximations and explain the assumptions.

Reading map

Topic	Notes	Slide PDF
Densities and cumulative probabilities	2	2–10
Normal family, tables and percentiles	3–5	11–33
Transformations and normal examples	6–7	34–49
Binomial approximation and fuse example	8–10	50–61
CLT, sampling distributions and song illustration	11–13	62–78
Exact normal sums and standard error	14–15	79–84
Alzheimer's-duration example and synthesis	16–17	85–95

Notation. As in the slides, $N(\mu, \sigma)$ lists the *mean and standard deviation*, not mean and variance. Upper-case $X, \bar{X}, T$ describe random quantities; lower-case $x, \bar{x}, t$ describe observed values. $q = 1 - p$ for Bernoulli/binomial models.

Keep in mind. Links use PDF viewer page numbers, including slide reveals. If a viewer ignores a page link, enter that number or use the section index in the footer.

Continuous probabilities are areas

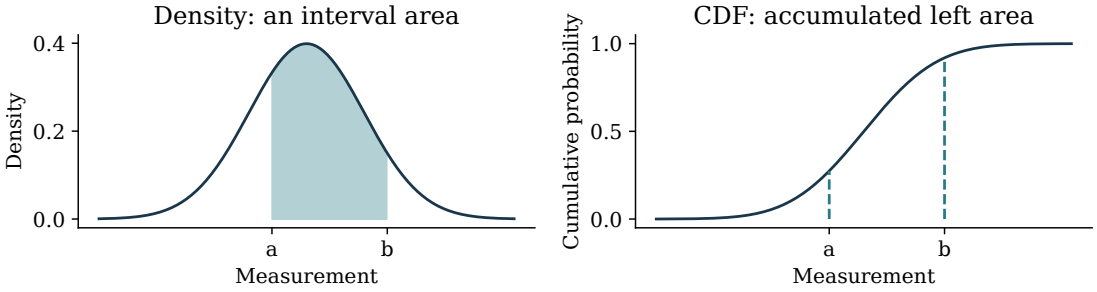
Lecture slides PDF pp. 2–10

Quick exam-prep guide p. 1

A continuous model allows values throughout an interval, such as a concentration or a duration. Its **probability density function** (PDF), $f(x)$, describes how probability is distributed along the measurement scale.

$$f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1, \qquad P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx.$$

You do not need to evaluate these integrals by hand in this course. Their meaning is the area beneath a curve. The density need not be continuous everywhere; the area rules are what define a density.



Height is not probability. $f(x)$ is a density height, whereas $P(a \leq X \leq b)$ is an area. A density may exceed 1 over a short enough interval while its total area remains 1. Its units are reciprocal measurement units, so multiplying by a width produces a unitless probability.

A single point has zero area

For a variable with a density, $P(X = c) = 0$. Consequently, including or excluding a boundary does not change a continuous probability:

$$P(a \leq X \leq b) = P(a < X < b).$$

This does *not* hold for a discrete count with positive point probabilities.

The cumulative distribution function

For any random variable, $F(x) = P(X \leq x)$. For a continuous variable it accumulates density area from the far left up to x:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \qquad P(a < X \leq b) = F(b) - F(a).$$

The CDF rises from 0 to 1 and cannot decrease. Where differentiable, its slope is the density. Density height, cumulative area and interval area answer different questions.

Keep in mind. Sketch the interval or tail first. Then choose the CDF calculation that gives that area.

The normal family and its shape

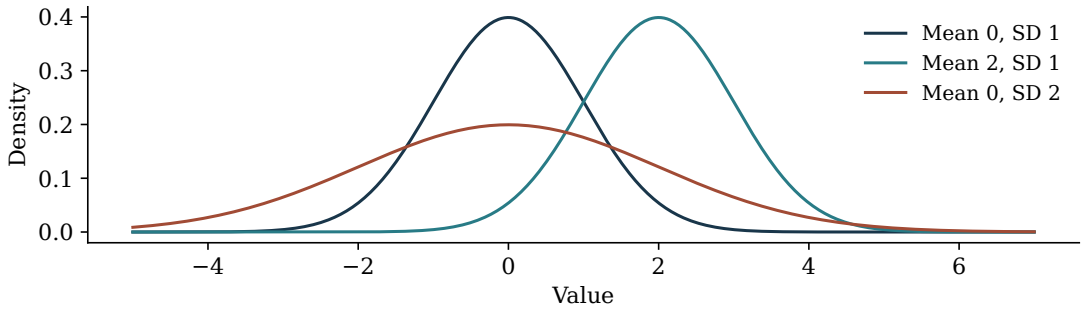
Lecture slides PDF pp. 11–18

Quick exam-prep guide p. 1

The normal, or Gaussian, family gives symmetric bell-shaped densities. In the course notation,

$$X \sim N(\mu, \sigma), \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad \sigma > 0.$$

The parameter μ is both the mean and the center of symmetry. The parameter σ is the SD; increasing it spreads the curve out and lowers its peak while preserving total area 1. The formula describes the shape; solving probability questions requires areas, not memorizing the density formula.



The empirical rule applies to a normal shape

Interval	Approximate normal probability
$\mu \pm \sigma$	.6827
$\mu \pm 2\sigma$	.9545
$\mu \pm 3\sigma$	.9973

Half the area lies on each side of μ. These percentages are properties of a normal curve, not universal percentages for every population.

Why a normal model appears here

The slides begin with a binomial count $B \sim \text{Binomial}(1000, 1/3)$. Computing $P(100 \le B \le 350)$ exactly means adding 251 binomial probabilities. A smooth normal curve can approximate such a histogram when both expected successes and failures are sufficiently large. We develop that approximation on pp. 8–10, then generalize the idea to sums through the CLT.

Keep in mind. A bell curve is a model to justify, not a default shape for every biological measurement. Bounded, skewed or multimodal data may require care.

Standard normal probabilities and tables

Lecture slides PDF pp. 19–31

Quick exam-prep guide p. 1

The **standard normal variable** $Z \sim N(0,1)$ has mean 0 and SD 1. Its cumulative distribution function is denoted by the capital Greek letter Phi:

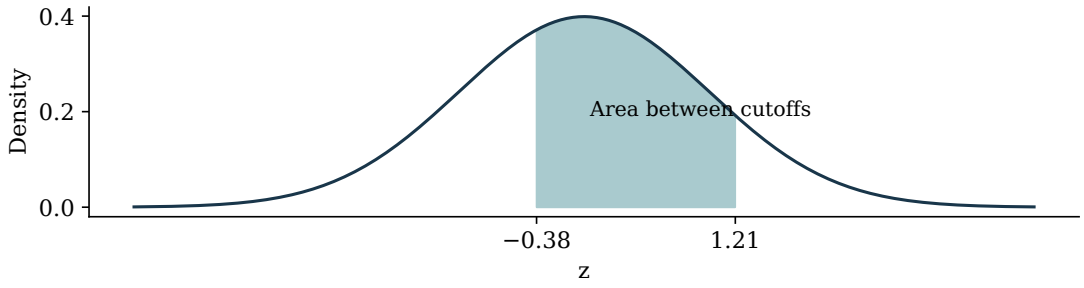
$$\boxed{\Phi(z) = P(Z \le z).}$$

The input z is a fixed cutoff; Z is random. The usual course table gives the cumulative area to the left. Read its headings before using a different table. There is no elementary algebraic formula for this cumulative area, so we use a table or numerical software.

Three area rules

$$\begin{aligned} P(Z \ge a) &= 1 - \Phi(a), \\ P(a \le Z \le b) &= \Phi(b) - \Phi(a), \\ \Phi(-z) &= 1 - \Phi(z) \qquad \text{(symmetry)}. \end{aligned}$$

For example, the slides use $\Phi(.44) \approx .6700$ to obtain $P(Z < -.44) \approx .3300$.



Slide example: an interval crossing zero

Find $P(-.38 \le Z \le 1.21)$.

1. The upper cutoff accumulates $\Phi(1.21) \approx .8869$.
2. The lower cutoff accumulates $\Phi(-.38) = 1 - \Phi(.38) \approx 1 - .6480 = .3520$.
3. Subtract the lower cumulative area from the upper:

$$P(-.38 \le Z \le 1.21) \approx .8869 - .3520 = \boxed{.5349}.$$

The unrounded numerical result is about .5349; the displayed result follows the slide table rounding.

Keep in mind. A table value is not automatically the requested probability. Determine whether you need a left area, a complement, or a difference.

Percentiles reverse the probability question

Lecture slides PDF pp. 32–33
Quick exam-prep guide p. 1

A probability question starts with a cutoff and asks for an area. A percentile question starts with an area and asks for the cutoff.

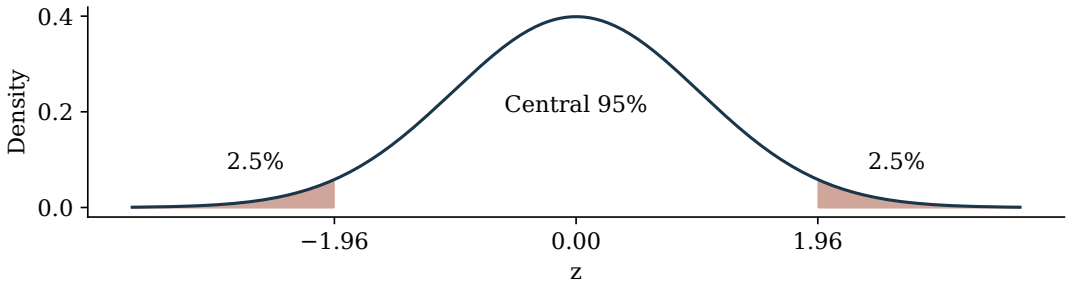
$$P(Z \le z) = r \quad \Longrightarrow \quad z = \Phi^{-1}(r), \qquad 0 < r < 1.$$

The inverse symbol means "find the input having this CDF value"; it does not mean $1/\Phi(r)$.

The course uses upper-tail critical values

$$\boxed{P(Z > z_\alpha) = \alpha, \qquad \Phi(z_\alpha) = 1 - \alpha.}$$

Thus $z_{.025} \approx 1.960$ is the 97.5th percentile, because .025 of the area lies to its right. The subscript is a *tail probability*, not the percentile itself.



Left cumulative area	Cutoff	Upper-tail notation
.10	−1.282	$z_{.90}$
.90	1.282	$z_{.10}$
.95	1.645	$z_{.05}$
.975	1.960	$z_{.025}$
.995	2.576	$z_{.005}$

Central intervals split the remaining area

To enclose $1 - \alpha$ centrally, put $\alpha/2$ in each tail:

$$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha.$$

For 95% central coverage, $\alpha = .05$, so the endpoints are ± 1.96. A one-sided upper 5% cutoff instead uses 1.645. This distinction will reappear in confidence intervals and tests.

Keep in mind. First label the desired left area or the upper-tail area. That prevents using 1.645 where 1.960 is needed.

Changing units and standardizing

Lecture slides PDF pp. 34–41
Quick exam-prep guide p. 1

For constants a, b and any variable with the needed moments,

$$\mathrm{E}(aX+b) = a\,\mathrm{E}(X) + b, \quad \mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X), \quad \mathrm{SD}(aX+b) = |a|\,\mathrm{SD}(X).$$

Adding a constant shifts the center without changing spread. Multiplication changes spread by the absolute value of the multiplier. If X is normal and $a \neq 0$, the transformed variable is also normal:

$$X \sim N(\mu, \sigma) \quad \Longrightarrow \quad aX + b \sim N(a\mu + b, |a|\sigma).$$

For $a = 0$, the result is the constant b, not a nondegenerate normal variable.

Slide example: Celsius to Fahrenheit

If $C \sim N(5, 10)$ and $F = \frac{9}{5}C + 32$, then

$$\mathrm{E}(F) = \tfrac{9}{5}(5) + 32 = 41, \qquad \mathrm{SD}(F) = \tfrac{9}{5}(10) = 18, \qquad F \sim N(41, 18).$$

The shift 32 changes the mean, not the SD.

Standardization puts values on the same scale

For $X \sim N(\mu, \sigma)$, subtract its mean and divide by its SD:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

For a particular value c, its **z-score** $(c - \mu)/\sigma$ tells us how many SDs it lies above or below the mean. The units cancel.

$$P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right),$$
$$P(X \geq a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right),$$
$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Reverse the transformation for a percentile

If $P(X \leq c) = r$, then

$$\boxed{c = \mu + \sigma\Phi^{-1}(r).}$$

This changes a standard-normal cutoff back into the original measurement units.

Keep in mind. Standardizing any variable gives center 0 and SD 1 when the moments exist. It gives a normal distribution only when a normal model or an approximation is justified.

Worked normal examples from the slides

Lecture slides PDF pp. 42–49
Quick exam-prep guide p. 1

A numerical interval

Let $X \sim N(4, 3)$, so the SD is 3. Find $P(3 \le X \le 7.5)$.

$$z_L = \frac{3-4}{3} = -\frac{1}{3}, \qquad z_U = \frac{7.5-4}{3} = \frac{7}{6}.$$

Therefore $P(3 \le X \le 7.5) = \Phi(7/6) - \Phi(-1/3) \approx \boxed{.5089}$. The slides report .5083 after rounding the cutoffs to $-.33$ and 1.17 and using table values. Both calculations follow the same method; keep precision until the final step when using software.

Ponderal-index model for newborns

The slide exercise models full-term newborn ponderal index by $X \sim N(2.4, .45)$, in the units printed on the slides, kg/m^3. Use these stipulated values for the exercise.

(1) What percentage is below 1.75?

$$\begin{aligned} P(X < 1.75) &= \Phi\left(\frac{1.75-2.4}{.45}\right) \\ &= \Phi(-1.4444\ldots) \approx .0743. \end{aligned}$$

Thus about **7.43%** under the stated model. The slide's two-decimal table calculation uses -1.44 and gives .0749, or 7.49%.

(2) Which value has 10% of the population below it? Write the required probability before finding the cutoff:

$$P(X < c) = .10 \quad \Longrightarrow \quad \Phi\left(\frac{c-2.4}{.45}\right) = .10.$$

The standard-normal cutoff is $\Phi^{-1}(.10) \approx -1.28155$. Undo standardization:

$$c = 2.4 + .45(-1.28155) \approx \boxed{1.8233 \text{ kg/m}^3}.$$

Using the slide's rounded -1.28 gives 1.824 kg/m^3.

Interpret the direction before calculating

The lower tenth percentile must be below the mean 2.4. A positive z-score here would locate an upper percentile. A negative z-score is a location, not a negative probability.

Keep in mind. These are probabilities and percentiles for individual measurements. No $\sqrt{n}$ appears because the question is not about an average of n newborns.

Approximating a binomial count

Lecture slides PDF pp. 50–54
Quick exam-prep guide p. 2

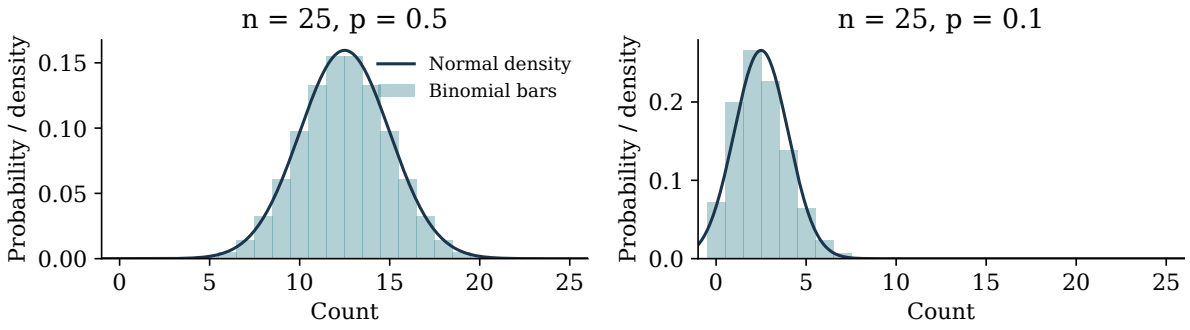
Let B count successes in n independent trials with common success probability p. Set $q = 1 - p$. Then

$$B \sim \text{Binomial}(n, p), \qquad \text{E}(B) = np, \qquad \text{Var}(B) = npq.$$

A useful approximating continuous variable has the same center and spread:

$$\boxed{Y \sim N(np, \sqrt{npq}).}$$

This matches the heights of much of the binomial histogram when it is not too skewed. It does not make the count continuous: B still takes only integers from 0 to n.



Check both expected counts

The slides use $np > 5$ and $nq > 5$. These are approximation guidelines, not a guarantee of exact accuracy, especially in extreme tails.

- For $n = 25, p = .5$, both expected counts are 12.5; the histogram is symmetric.
- For $n = 25, p = .1$, $np = 2.5$ is small; the histogram is skewed and a normal approximation is less reliable.

If the model itself is wrong (dependent trials or unequal success probabilities), large n does not repair it.

Source formula clarification

For $n = 25, p = .5$, the approximating density is

$$f(y) = \frac{1}{2.5\sqrt{2\pi}} \exp\left[-\frac{(y - 12.5)^2}{12.5}\right].$$

Slide PDF p. 53 omits the minus sign in the exponent. A normal density decreases away from its center; the negative sign is essential.

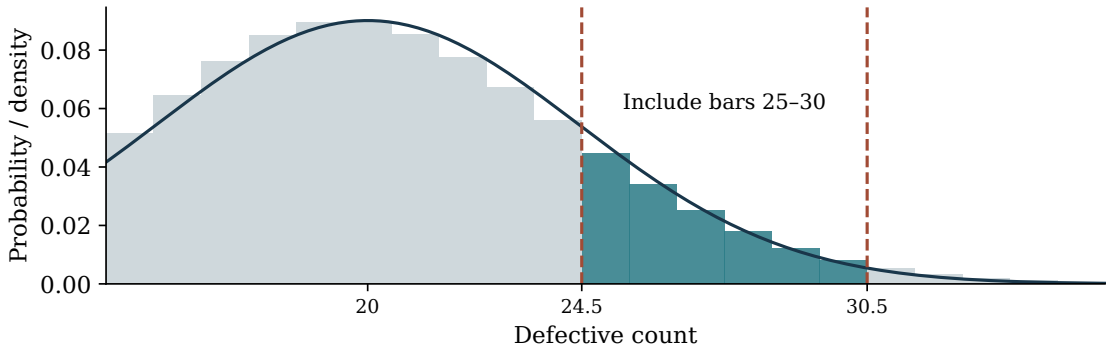
Keep in mind. Use p for the event being counted. If a trial works with probability .98 but B counts failures, use $p = .02$.

Why the continuity correction is half a count

Lecture slides PDF pp. 55–56

Quick exam-prep guide p. 2

A binomial probability $P(B = k)$ occupies one histogram bar of width 1 centered at integer k. Its edges are $k - .5$ and $k + .5$. A normal area should cover the same bars.



For **integer cutoffs** a, b and $Y \sim N(np, \sqrt{npq})$,

Count event	Matching continuous area
$a \le B \le b$	$a - .5 < Y < b + .5$
$B \le a$	$Y < a + .5$
$B \ge a$	$Y > a - .5$
$B < a$	$Y < a - .5$
$B > a$	$Y > a + .5$

For instance, $B < 25$ means $B \le 24$, so its upper corrected boundary is 24.5, not 25.5.

Then standardize the corrected boundaries

$$P(a \le B \le b) \approx \Phi\left(\frac{b + .5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - .5 - np}{\sqrt{npq}}\right).$$

For noninteger cutoffs, first determine the included integers. For a proportion event, first write it as a count event: if $\hat{p} = B/n$, then $\hat{p} \ge c$ means $B \ge \lceil nc \rceil$. Apply the half-count correction on that count scale.

Keep in mind. The correction is .5 of one count, not .5 of a proportion or one-half of a standard deviation.

Fuse reliability worked example

Lecture slides PDF pp. 57–61

Quick exam-prep guide p. 2

Question from the slides. The reliability of an electrical fuse is .98. A random sample of 1,000 fuses is tested. Find the approximate probability of between 25 and 30 defectives, inclusive. Assume independent fuses with a common failure probability.

1. Define the random variable and check the model

Let B be the number of defective fuses. Reliability is the chance of working, so the failure probability is .02:

$$B \sim \text{Binomial}(1000, .02), \qquad np = 20 > 5, \quad nq = 980 > 5.$$

The approximating normal variable has mean 20 and SD $\sqrt{19.6} = 4.42719$:

$$Y \sim N(20, 4.42719).$$

2. Include the full bars for counts 25 through 30

$$P(25 \le B \le 30) \approx P(24.5 < Y < 30.5).$$

The standardized boundaries are

$$z_L = \frac{24.5 - 20}{\sqrt{19.6}} = 1.01645, \qquad z_U = \frac{30.5 - 20}{\sqrt{19.6}} = 2.37171.$$

3. Subtract cumulative areas

$$P(25 \le B \le 30) \approx \Phi(2.37171) - \Phi(1.01645) \approx \boxed{.1459}.$$

Using the slide's two-decimal table cutoffs gives $.9911 - .8461 = .1450$.

Rounding error and approximation error are different

The exact binomial calculation is

$$\sum_{k=25}^{30} \binom{1000}{k} (.02)^k (.98)^{1000-k} \approx .1419.$$

Keeping more digits makes the *normal calculation* more precise, but does not guarantee closer agreement with the exact binomial probability. Here the rounded-table result happens to be closer. This clarifies the final statement on slide pp. 60–61.

Return to the chapter's opening motivation

For $B \sim \text{Binomial}(1000, 1/3)$, the same procedure uses mean 1000/3, SD $\sqrt{1000(1/3)(2/3)}$, and corrected boundaries $99.5, 350.5$ for $100 \le B \le 350$. There is one common workflow, regardless of whether success is called a defect, a response or another binary outcome.

Keep in mind. State the event and the approximation before the decimal answer. They show what the number means and how it was obtained.

From Bernoulli sums to the CLT

Lecture slides PDF pp. 62–66

Quick exam-prep guide p. 3

A binomial count can be written as a sum of independent indicators:

$$B = X_1 + \cdots + X_n, \qquad X_i \sim \text{Bernoulli}(p).$$

Each X_i is 1 for a success and 0 otherwise. Since $\mathrm{E}(X_i) = p$ and $\mathrm{Var}(X_i) = pq$, the sum has mean np and variance npq. The previous approximation asks whether many small random contributions together have a nearly normal shape.

The Central Limit Theorem

Suppose $X_1, \ldots, X_n$ are independent and identically distributed (iid), with common mean μ and finite, positive variance σ^2. Then, as n grows,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \quad \text{approaches } N(0,1) \text{ in distribution.}$$

For sufficiently large n, this gives the practical approximation

$$\boxed{T = \sum_{i=1}^{n} X_i \overset{\text{approx}}{\sim} N(n\mu, \sigma\sqrt{n}).}$$

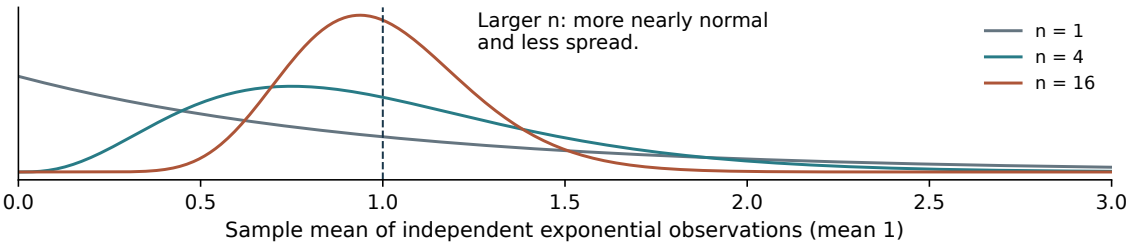
The theorem concerns the distribution of a *standardized sum*. It does not say that the original observations become normal.

What counts as sufficiently large?

- The course uses $n \geq 30$ as a working guideline. It is not a universal guarantee; strong skew or extreme tails may require much more data.
- For Bernoulli data, check np and nq, even when n is large.
- Independence, the sampling design and finite variance matter. Repeated measurements from the same organism are not automatically independent.
- Sampling without replacement induces dependence. If the sample is a small fraction of a large population, the independent-sampling approximation may be reasonable; Chapter 7 explains the correction for a substantial fraction.

Why the normal approximation is useful

Select the random quantity and justify its distribution before standardizing. Then use the same normal-area tools.



New illustration: means of independent exponential observations become less skewed and more concentrated.

Keep in mind. Exact moment formulas and an approximately normal shape are separate claims. Correct mean and variance alone do not prove normality.

Sampling distributions of totals and means

Lecture slides PDF pp. 67–75

Quick exam-prep guide p. 3

Before a sample is collected, its observations $X_1, \ldots, X_n$ are random. Any quantity calculated from them is a **statistic**, such as

$$T = \sum X_i, \qquad \bar{X} = \frac{1}{n}\sum X_i, \qquad S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2.$$

The probability distribution of a statistic over repeated samples of the same size and design is its **sampling distribution**. It differs from the distribution of individual observations within one sample.

Derive the center and spread

For independent observations with common mean μ and variance σ^2,

$$\mathrm{E}(T) = \sum \mathrm{E}(X_i) = n\mu, \qquad \mathrm{Var}(T) = \sum \mathrm{Var}(X_i) = n\sigma^2,$$
$$\mathrm{E}(\bar{X}) = \mathrm{E}(T)/n = \mu, \qquad \mathrm{Var}(\bar{X}) = \mathrm{Var}(T)/n^2 = \sigma^2/n.$$

The mean formula uses linearity of expectation. The variance-sum step uses independence (zero covariances). Taking square roots gives the SDs.

Random quantity	Center	Variance	SD
One observation X_i	μ	σ^2	σ
Total T	$n\mu$	$n\sigma^2$	$\sigma\sqrt{n}$
Mean $\bar{X}$	μ	σ^2/n	$\sigma/\sqrt{n}$

Apply the CLT to the quantity asked about

Under the iid finite-variance conditions and with an adequate sample size,

$$T \overset{\text{approx}}{\sim} N(n\mu, \sigma\sqrt{n}), \qquad \bar{X} \overset{\text{approx}}{\sim} N(\mu, \sigma/\sqrt{n}).$$

The standardized quantities are actually the same:

$$\frac{T - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Thus an event for a total can be converted into an equivalent event for its average; the probability agrees when both cutoff and scale are converted consistently.

Keep in mind. Larger samples make the total more variable in absolute units but the average less variable. The formulas differ because a total adds n contributions, whereas a mean divides that total by n.

What a sampling distribution looks like

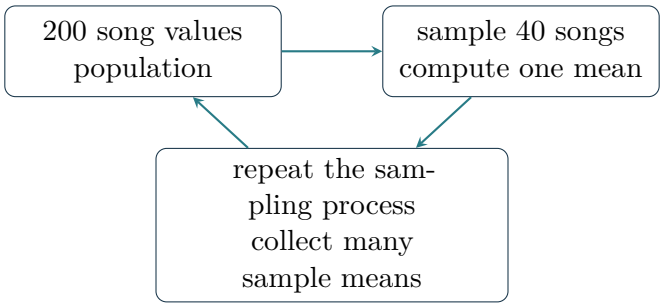
Lecture slides PDF pp. 76–78

Quick exam-prep guide p. 3

Slide illustration. The slides describe the top 200 Taylor Swift songs by estimated Spotify streams as of August 2023. They repeatedly sample 40 songs and calculate the mean stream count. The displayed sample means include 273,692,641 and 267,909,152; 117 repeated samples were used in the source illustration.

Three distributions must be kept distinct

1. **Population distribution:** one stream-count value for each of the 200 songs. Its shape may be skewed.
2. **One sample:** 40 selected songs with 40 stream counts. Compute one mean from this sample.
3. **Sampling distribution:** repeat the sample-selection process and collect one mean from each sample. The histogram of those means can be much more symmetric and concentrated.



Repeating the sampling process is the idea behind a sampling distribution. It is different from simply averaging the same sample repeatedly. The 117 repetitions describe the illustration; the theoretical distribution considers all possible samples under the design.

A design qualification

Forty of 200 songs is a 20% sampling fraction. If each sample is drawn without replacement, that fraction is not negligible. With finite-population variance defined using divisor N, the mean's SD is

$$\mathrm{SD}(\bar{X}) = \frac{\sigma}{\sqrt{40}}\sqrt{\frac{200-40}{200-1}} \approx .8967\frac{\sigma}{\sqrt{40}}.$$

The multiplier is the finite-population correction, developed in Chapter 7 notes p. 12. The slide histogram is an empirical illustration, not a proof that observations are iid. The original spreadsheet's sampling procedure and raw values are not reconstructed here.

Keep in mind. Ask what was sampled, how a statistic was computed and what was repeated. A histogram of averages describes averages, not individual songs or patients.

When the normal result is exact

Lecture slides PDF pp. 79–81
Quick exam-prep guide p. 3

If the observations are independent and *already normal*, their sum is exactly normal for any sample size:

$$X_i \sim N(\mu_i, \sigma_i), \quad X_i \text{ independent} \quad \Longrightarrow \quad \sum_{i=1}^{n} X_i \sim N\left(\sum \mu_i, \sqrt{\sum \sigma_i^2}\right).$$

The means and SDs need not be equal, and n can be as small as 2. This is a property of normal variables, not a large-sample approximation.

Random sampling from a normal population

For iid $X_i \sim N(\mu, \sigma)$,

$$\boxed{T \sim N(n\mu, \sigma\sqrt{n}), \qquad \bar{X} \sim N(\mu, \sigma/\sqrt{n}) \quad \text{for every } n.}$$

The center and spread agree with the previous moment calculations. What has strengthened is the shape statement: exact normality replaces approximate normality.

Independent differences use the same reasoning

A difference is a sum with coefficient -1. For independent normal variables,

$$X_1 - X_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right).$$

Variances *add* because the coefficient is squared. More generally, for independent normal X_i,

$$\sum a_i X_i + b \sim N\left(\sum a_i \mu_i + b, \sqrt{\sum a_i^2 \sigma_i^2}\right)$$

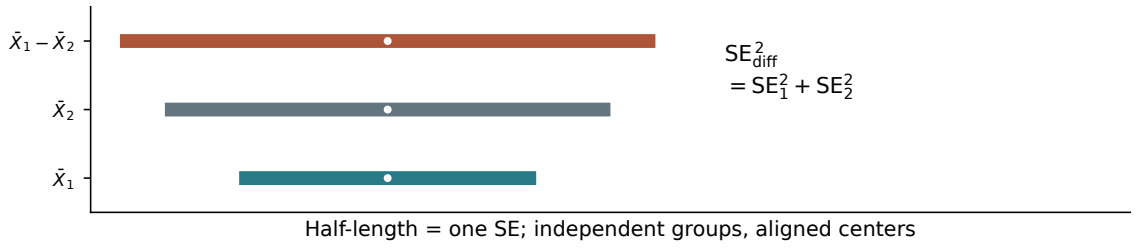
when the resulting variance is positive. This extends the slide's sum rule and connects to the existing group problem comparing fertilizers.

If independence fails

For two variables,

$$\mathrm{Var}(X_1 - X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) - 2\,\mathrm{Cov}(X_1, X_2).$$

A paired before/after measurement on the same organism can have substantial covariance. Dropping it without justification changes the SE. Chapters 8 and 9 will handle such designs by analyzing within-pair differences.



Independence removes covariance. Subtracting means still adds their variance contributions.

Keep in mind. Known means and SDs do not by themselves imply a normal sum. State either independent normal inputs for an exact result, or the conditions for a CLT approximation.

Standard error describes an estimate's variability

Lecture slides PDF pp. 82–84
Quick exam-prep guide p. 3

The **standard error** of an estimator is the SD of its sampling distribution. For the sample mean under independent sampling,

$$\boxed{\mathrm{SE}(\bar{X}) = \mathrm{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}.}$$

This measures the typical scale on which the mean varies between samples of size n. It is not the actual error $\bar{x} - \mu$ of the sample we happened to observe.

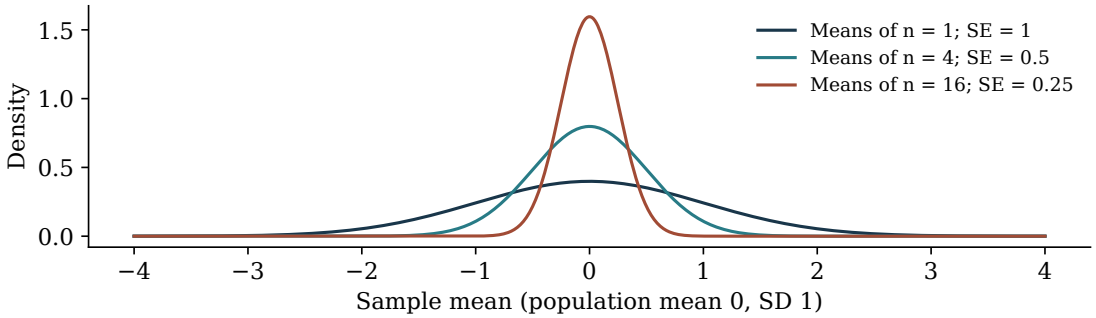


Illustration for independent samples from a normal population with mean 0 and SD 1. The curves are exact normal sampling distributions in this model.

Keep these three quantities separate

- **Population SD σ:** variability among individual observations.
- **SE $\sigma/\sqrt{n}$:** variability among means from repeated samples.
- **Observed error $\bar{x} - \mu$:** the unknown discrepancy for one realized sample.

When σ is unknown, $s/\sqrt{n}$ estimates the SE. Estimating it introduces uncertainty; Chapter 9 explains why a t distribution becomes relevant under normal sampling.

Why four times the sample size halves SE

$$\mathrm{SE}_{4n} = \frac{\sigma}{\sqrt{4n}} = \frac{1}{2}\frac{\sigma}{\sqrt{n}}.$$

Doubling n only multiplies SE by $1/\sqrt{2} \approx .707$. Better precision has diminishing returns as sample size grows.

Precision is not the same as representativeness

A small SE can describe a precisely estimated mean for the wrong population if sampling is biased. Larger samples reduce random variability under the model; they do not remove selection bias or fix dependence.

Keep in mind. Before standardizing, identify the random variable. An individual uses σ; a mean of n independent observations uses $\sigma/\sqrt{n}$.

Alzheimer's-duration worked example

Lecture slides PDF pp. 85–93

Quick exam-prep guide p. 3

Question from the slides. In the exercise model, disease duration from onset of symptoms until death ranges from 3 to 20 years, with mean 8 and SD 4 years. An administrator randomly selects 30 records of deceased patients from a large medical-center database. Find the probability that the sample mean is below 7 years, the probability it lies within one year of 8, and a sample size making the second probability at least .95.

These are stipulated historical exercise values, not medical guidance or verified population estimates. Assume the sample represents the stated database and its fraction of that database is small. We use the course CLT approximation; the individual durations were not assumed normal.

1. The sample mean is below 7 years

$$\bar{X} \overset{\text{approx}}{\sim} N\left(8, \frac{4}{\sqrt{30}}\right), \qquad \text{SE} = \frac{4}{\sqrt{30}} = .73030 \text{ years.}$$

$$P(\bar{X} < 7) \approx \Phi\left(\frac{7-8}{4/\sqrt{30}}\right) = \Phi(-1.36931) \approx \boxed{.0855}.$$

The slide's rounded -1.37 gives .0853.

2. The sample mean is within one year of 8

"Within one year" means $7 \le \bar{X} \le 9$, not that every sampled patient has a duration in this range:

$$P(7 \le \bar{X} \le 9) \approx \Phi(1.36931) - \Phi(-1.36931) \approx \boxed{.8291}.$$

The slide table calculation gives .8294.

3. Choose the sample size before sampling

We want the central interval 8 ± 1 to contain at least 95% of the approximate sampling distribution. Its half-width in SE units must be at least $z_{.025} \approx 1.96$:

$$\frac{1}{4/\sqrt{n}} \ge 1.96 \quad \Longrightarrow \quad n \ge (1.96 \cdot 4)^2 = 61.4656.$$

Round *up*: $\boxed{n = 62}$. This is the required total sample size, not 62 additional records. Under the normal approximation, $n = 61$ gives about .9491 and $n = 62$ gives about .95099.

Keep in mind. The 95% design target is approximate because the shape was justified by the CLT. The first two moments alone cannot guarantee exactly 95% coverage at finite n.

Unbiasedness and the connection to inference

Lecture slides PDF pp. 94–95

Quick exam-prep guide p. 3

An **estimator** is a statistic used to estimate a population parameter. It is **unbiased** when its expected value equals that parameter. For a random sample with common mean μ,

$$\mathrm{E}(\bar{X}) = \frac{1}{n}\sum \mathrm{E}(X_i) = \mu.$$

This equality does not require normality or large n. It describes average behavior across samples, not a promise that every $\bar{x}$ equals μ.

A common framework for the coming chapters

Chapter	Question being answered
6	Given population parameters, how likely is an individual observation, total or mean to fall in a specified region?
7	Given sample data, what interval estimates a population parameter with a stated procedure's coverage?
8	If a null claim were true, how surprising would the observed discrepancy be?
9	How should inference change when normal-sample variability must be estimated, especially in small samples?

The shared quantity is a discrepancy measured on its relevant variability scale:

$$\frac{\text{random quantity} - \text{its model center}}{\text{its SD or SE}}.$$

The distribution used after this standardization still needs justification. Normal population results can be exact; CLT-based results are approximate; estimated-SE ratios can require t distributions.

Use the existing practice after reviewing the concepts

The existing cell and fertilizer questions are on the concise resource's p. 3. Attempt them before opening the released solutions on p. 4. In each case, name the random quantity and its units, state the model assumptions, derive its center and spread, then choose the required tail or interval.

For a shorter recall pass, use the Chapter 6 exam-prep guide. Use the detailed pages whenever a formula's conditions or meaning are unclear. No new practice set is introduced here.

Keep in mind. A correct decimal without the correct random variable or sampling assumptions does not fully answer a statistical question.