Chapter 6 Normal probabilities

Detailed notes pp. 2–7

Lecture slides PDF pp. 2–49

Use this guide to recall the main ideas after reading the detailed notes. Follow a link when you cannot explain a formula's meaning or conditions. This is a study guide, not a statement about permitted exam aids.

A density gives areas; a CDF accumulates them

$$F(x) = P(X \le x), \qquad P(a < X \le b) = F(b) - F(a).$$

For a variable with a density, $P(X = c) = 0$; strict and inclusive endpoints give the same probability. Density height $f(x)$ is not itself probability.

Course notation and normal shape

$X \sim N(\mu, \sigma)$ uses **mean and SD**, with $\sigma > 0$. The normal curve is symmetric about μ. About 68%, 95%, 99.7% lie within 1, 2, 3 SDs. These are normal-model properties, not rules for every distribution.

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1), \qquad \Phi(z) = P(Z \le z).$$

Question	Calculation
Below b	$\Phi((b - \mu)/\sigma)$
Above a	$1 - \Phi((a - \mu)/\sigma)$
Between a and b	$\Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma)$
Lower r percentile	$\mu + \sigma\Phi^{-1}(r)$

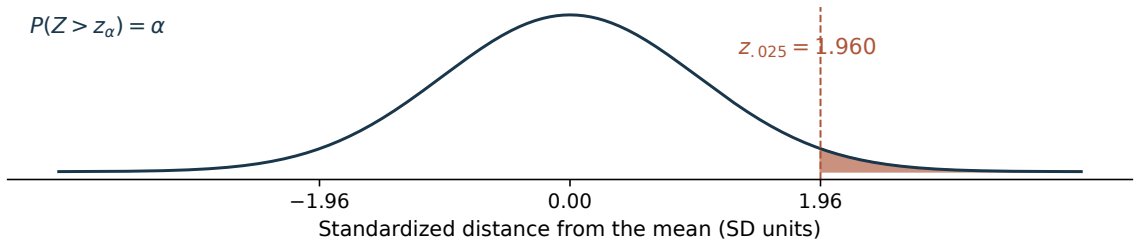
Tails, critical values and transformations

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

Useful upper-tail cutoffs: $z_{.05} = 1.645$, $z_{.025} = 1.960$, $z_{.005} = 2.576$. A central 95% interval uses two 2.5% tails; a one-sided 5% cutoff uses one 5% tail.

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

Normal inputs stay normal under a nonconstant linear transformation.



The critical value marks a boundary; the tail area is a probability.

Keep in mind. Sketch the event first, label the random quantity, then standardize. A percentile asks for a cutoff, not for another probability.

Chapter 6 Binomial approximation

Detailed notes pp. 8–10
Lecture slides PDF pp. 50–61

Choose the event counted as success

For independent trials with common success probability p,

$$B \sim \text{Binomial}(n, p), \quad q = 1 - p, \quad \text{E}(B) = np, \quad \text{SD}(B) = \sqrt{npq}.$$

If B counts defects, p must be the defect probability, not the reliability.

Check before using the normal curve

The course guideline is $np > 5$ **and** $nq > 5$. Both matter; large n alone is insufficient when p is near 0 or 1. The approximation may still be poor in extreme tails. With conditions adequate, use $Y \sim N(np, \sqrt{npq})$.

Convert integer bars to continuous boundaries

Integer count event	Corrected normal event
$a \le B \le b$	$a - .5 < Y < b + .5$
$B \le a$	$Y < a + .5$
$B \ge a$	$Y > a - .5$
$B < a$	$Y < a - .5$
$B > a$	$Y > a + .5$

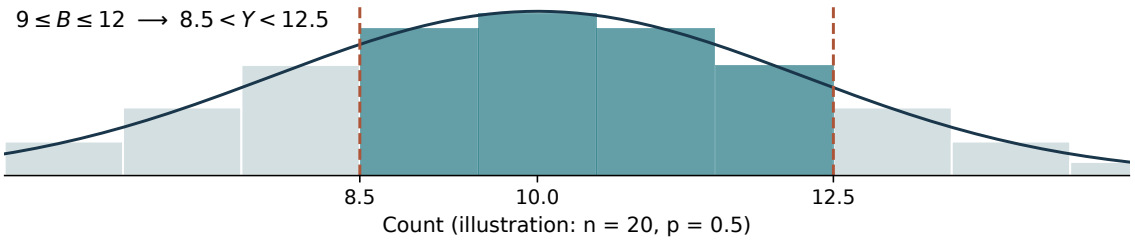
Then subtract np and divide by $\sqrt{npq}$. For instance,

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

For a proportion event, write $\widehat{p} = B/n$ and convert to integer counts first. "At least c" means $B \ge \lceil nc \rceil$; the continuity correction is half a *count*.

Use the existing worked example to check the workflow

The fuse example on detailed p. 10 proceeds through variable, conditions, corrected boundaries, z-scores and area. Recreate those steps before checking its answer. Numerical rounding and error from the normal approximation are different sources of discrepancy.



Include the full bars for the integer outcomes in the event.

Keep in mind. The normal approximation does not change a discrete count into a continuous observation. The half-count correction matches the bars included in the event.

Chapter 6 Means totals and standard errors

Detailed notes pp. 11–17
Lecture slides PDF pp. 62–95

Select the scale before the formula

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

Quantity	Mean	SD	Standardized form
X_i	μ	σ	$(X_i-\mu)/\sigma$
$T=\sum X_i$	$n\mu$	$\sigma\sqrt{n}$	$(T-n\mu)/(\sigma\sqrt{n})$
$\bar{X}=T/n$	μ	$\sigma/\sqrt{n}$	$(\bar{X}-\mu)/(\sigma/\sqrt{n})$

Exact moments, separate shape. The mean/variance formulas do not by themselves imply normality.

- **Exact normal:** independent normal observations give normal totals and means for any n.
- **CLT approximation:** iid observations with finite positive variance give approximately normal standardized totals/means for sufficiently large n. The course's $n \geq 30$ is a heuristic; strong skew may need more.
- **Finite population:** without-replacement draws are dependent. A small sampling fraction may justify approximate independence; otherwise use the Chapter 7 correction when applicable.

Independent sums and differences

For independent variables, $\mathrm{Var}(\sum a_i X_i)=\sum a_i^2\sigma_i^2$. In particular, $\mathrm{Var}(X-Y)=\sigma_X^2+\sigma_Y^2$. Add variances, not SDs. Normal inputs make the linear combination exactly normal.

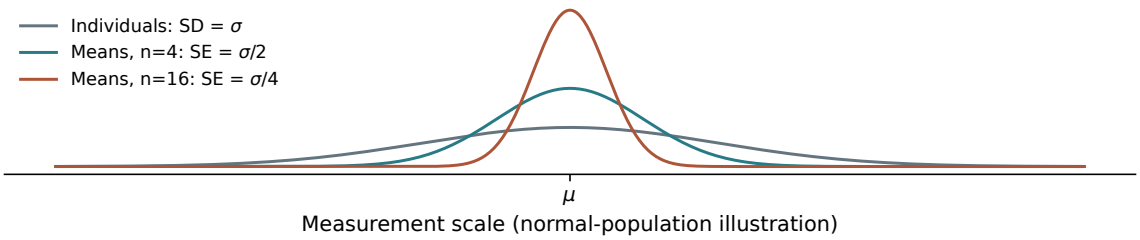
SE and unbiasedness

$\mathrm{SE}(\bar{X})=\sigma/\sqrt{n}$ measures repeated-sample variability of the mean. It is neither individual SD nor the realized error $\bar{x}-\mu$. Quadruple n to halve SE. Unbiasedness means $\mathrm{E}(\bar{X})=\mu$, not that every estimate is correct.

For a normal-based target $P(|\bar{X}-\mu|\leq d)\geq 1-\alpha$ with known σ,

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{d}\right)^2; \quad \text{round up.}$$

This is approximate assurance when normality comes from the CLT.



Averaging narrows the sampling distribution, not the population.

Keep in mind. Attempt the existing cell/fertilizer questions on p. 3 independently; released solutions follow on p. 4. Explain why the selected SD or SE matches the question.