

Midterm Review

Key Concepts from Lessons 1–9

Emile Latour

2026-02-11

Today's Agenda

- Quick overview of exam structure
- Key concepts review
 - Study design
 - Probability and Bayes' Theorem
 - Distributions (focus on R functions!)
 - Sampling distributions and CLT
 - Confidence intervals
 - Hypothesis testing
- Your questions

The midterm opens TODAY at 3:00 PM and closes Monday at 11:00 PM







Exam Logistics

Exam Details

Format:

- 8 parts, 120 points total
- Designed for 3-5 hours completion
- Mix of conceptual questions, R code, and interpretation

Resources allowed:

-  Course materials (slides, homework, textbook)
-  R documentation
-  Your notes
-  Other students
-  AI assistants
-  Online help forums

Submit both `.qmd` and `.html` files!

What's Covered

Part 1: Study design (10 pts)

Part 2: Descriptive stats & viz (15 pts)

Part 3: Probability (15 pts)

Part 4: Distributions (20 pts)

Part 5: Sampling distributions & CLT (15 pts)

Part 6: Confidence intervals (15 pts)

Part 7: Hypothesis testing (20 pts)

Part 8: Integration (10 pts)

Key datasets: `nhanes.samp` (you've used this in HW!)

General Tips

Before you start:

- Read all instructions carefully
- Set up your workspace (load packages, data)
- Check that code chunks run

While working:

- Show your work and reasoning
- Provide interpretations in complete sentences
- Use comments in your R code
- Check that your document renders frequently
- **Save regularly**

Before submitting:

- Render one final time
- Check that all answers are complete
- Verify both files are being submitted

Quick Concepts Review

Study Design: The Basics

Experiments vs. Observational Studies

Experiment:

- Researchers assign treatments
- Random assignment
- Can establish causation
- Example: Drug trial with treatment vs. placebo

Observational Study:

- Researchers observe without interfering
- No random assignment
- Can only show association
- Example: Reviewing medical records

Why does this matter?

Only randomized experiments can establish causal relationships!

Sampling Concepts

Key terms:

- **Population vs. Sample**
- **Simple random sample** (each individual has equal chance)
- **Convenience sample** (easily accessible - often biased)

Confounding variables:

- Associated with both explanatory and response variables
- Can make it look like there's a causal relationship when there isn't
- Example: Exercise and cardiovascular health might both be related to socioeconomic status

R Skills Review: Data Wrangling & Visualization

dplyr: The grammar of data manipulation

dplyr provides a consistent set of **verbs** for data manipulation:

Function	What it does
<code>filter()</code>	Keep rows that meet conditions
<code>select()</code>	Keep or drop columns
<code>mutate()</code>	Create or modify columns
<code>arrange()</code>	Sort rows
<code>group_by()</code>	Group data for summaries
<code>summarize()</code>	Calculate summary statistics

These verbs can be chained together with `%>%` for powerful data transformations.

ggplot2: The Basic Template

The ggplot2 pattern

```
1 library(ggplot2)
2
3 ggplot(data = dataset,
4        aes(x = variable, y = variable)) +
5   geom_*() +                # Choose your geometry
6   labs(title = "...",      # Add labels
7        x = "...",
8        y = "...")
```

Key geoms you need to know:

```
1 # Histogram (one continuous variable)
2 geom_histogram(bins = 20)
3
4 # Boxplot (continuous by categorical)
5 geom_boxplot()
6
7 # Scatterplot (two continuous variables)
8 geom_point()
```

ggplot2: Example

```
1 library(ggplot2)
2 library(oibistat)
3 data("nhanes.samp")
4
5 # Histogram
6 ggplot(nhanes.samp, aes(x = Height)) +
7   geom_histogram(bins = 20) +
8   labs(title = "Distribution of Height",
9         x = "Height (cm)",
10        y = "Count")
11
12 # Boxplot by group
13 ggplot(nhanes.samp,
14        aes(x = Gender, y = Weight, fill = Gender)) +
15   geom_boxplot() +
16   labs(title = "Weight by Gender",
17        x = "Weight (kg)",
18        y = "Age (years)")
```

Remember: Start with `ggplot()`, then add layers with `+`

dplyr + rstatix: Summary Statistics

The pattern: `group_by()` then summarize

```
1 library(dplyr)
2 library(rstatix)
3
4 dataset %>%
5   group_by(grouping_variable) %>%
6   get_summary_stats(numeric_variable, type = "mean_sd")
```

Example:

```
1 # Summary stats by one group
2 nhanes.samp %>%
3   group_by(Gender) %>%
4   get_summary_stats(Height, type = "mean_sd")
5
6 # Summary stats by multiple groups
7 nhanes.samp %>%
8   group_by(Gender, SmokeNow) %>%
9   get_summary_stats(Weight, type = "mean_sd")
```

dplyr only: Summary Statistics

The pattern: `group_by()` then summarize

```
1 library(dplyr)
2
3 dataset %>%
4   group_by(grouping_variable) %>%
5   summarise(...)
```

Example:

```
1 # Summary stats by one group
2 nhanes.samp %>%
3   group_by(Gender) %>%
4   summarise(n = sum(!is.na(Height)),
5             mean = mean(Height, na.rm = TRUE),
6             sd = sd(Height, na.rm = TRUE))
7
8 # Summary stats by multiple groups
9 nhanes.samp %>%
10  group_by(Gender, SmokeNow) %>%
11  summarise(n = sum(!is.na(Weight)),
12            mean = mean(Weight, na.rm = TRUE),
13            sd = sd(Weight, na.rm = TRUE))
```

Categorical Data: janitor::tabyl()

The pattern for frequency tables

```
1 library(janitor)
2
3 # One-way table
4 dataset %>%
5   tabyl(variable)
6
7 # Two-way table with row percentages
8 dataset %>%
9   tabyl(row_variable, column_variable) %>%
10  adorn_totals(where = "row") %>%
11  adorn_percentages("row") %>%
12  adorn_pct_formatting(digits = 1)
```

Example:

```
1 # Two-way table
2 nhanes.samp %>%
3   tabyl(Gender, SmokeNow) %>%
4   adorn_totals(where = "row") %>%
5   adorn_percentages("row") %>%
6   adorn_pct_formatting(digits = 1) %>%
7   adorn_ns()
```


Probability

Probability Notation Quick Reference

Basic probabilities:

- **Marginal:** $P(A)$ - probability of A alone
- **Joint:** $P(A \text{ and } B)$ - probability of both
- **Conditional:** $P(A \mid B)$ - probability of A given B

Key formula:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Independence:

Events A and B are independent if: $P(A \text{ and } B) = P(A) \times P(B)$

Probability Rules to Remember

General Addition Rule:

If A and B are any two events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

We subtract $P(A \text{ and } B)$ because those outcomes were counted twice.

General Multiplication Rule:

$$P(A \text{ and } B) = P(A \mid B) \times P(B)$$

This connects **joint** and **conditional** probabilities, and leads directly to Bayes' Theorem!

Special case: Independent events

If A and B are independent: $P(A \text{ and } B) = P(A) \times P(B)$

Bayes' Theorem: The Pattern

You'll almost always use it in this form:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$
$$= \frac{P(B \mid A) \times P(A)}{P(B \mid A) \times P(A) + P(B \mid A^c) \times P(A^c)}$$

In medical testing context:

$$P(\text{Disease} \mid \text{Test}+) = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}$$

Pro tip: Calculate numerator and denominator separately, then divide!

Bayes' Theorem: Step-by-Step

Given: Sensitivity = 0.90, Specificity = 0.85, Prevalence = 0.02

```
1 # Step 1: Set up
2 sensitivity <- 0.90      # P(Test + | Disease)
3 specificity <- 0.85      # P(Test - | No Disease)
4 prevalence <- 0.02      # P(Disease)
5
6 # Step 2: Calculate P(Test + | No Disease)
7 p_pos_given_healthy <- 1 - specificity
8 p_pos_given_healthy
```

[1] 0.15

```
1 # Step 3: Numerator (true positives)
2 numerator <- sensitivity * prevalence
3 numerator
```

[1] 0.018

```
1 # Step 4: Denominator (all positives)
2 denominator <- sensitivity * prevalence + p_pos_given_healthy * (1 - prevalence)
3 denominator
```

[1] 0.165

```
1 # Step 5: PPV
2 ppv <- numerator / denominator
3 ppv
```

[1] 0.1090909

Distributions: The R Functions

The Four Functions Pattern

Every distribution in R has 4 functions:

d-functions: `dbinom()`, `dnorm()`, `dpois()`

- “density” or “probability mass”
- **Exactly** x
- Returns: $P(X = x)$

p-functions: `pbinom()`, `pnorm()`, `ppois()`

- “cumulative probability”
- **At most** x (or **at least** with `lower.tail = FALSE`)
- Returns: $P(X \leq x)$ or $P(X > x)$

q-functions: `qbinom()`, `qnorm()`, `qpois()`

- “quantile”
- What value gives this probability?
- Returns: value of x

r-functions: `rbinom()`, `rnorm()`, `rpois()`

- “random”
- Generate random samples
- Returns: random values

Binomial Functions in R

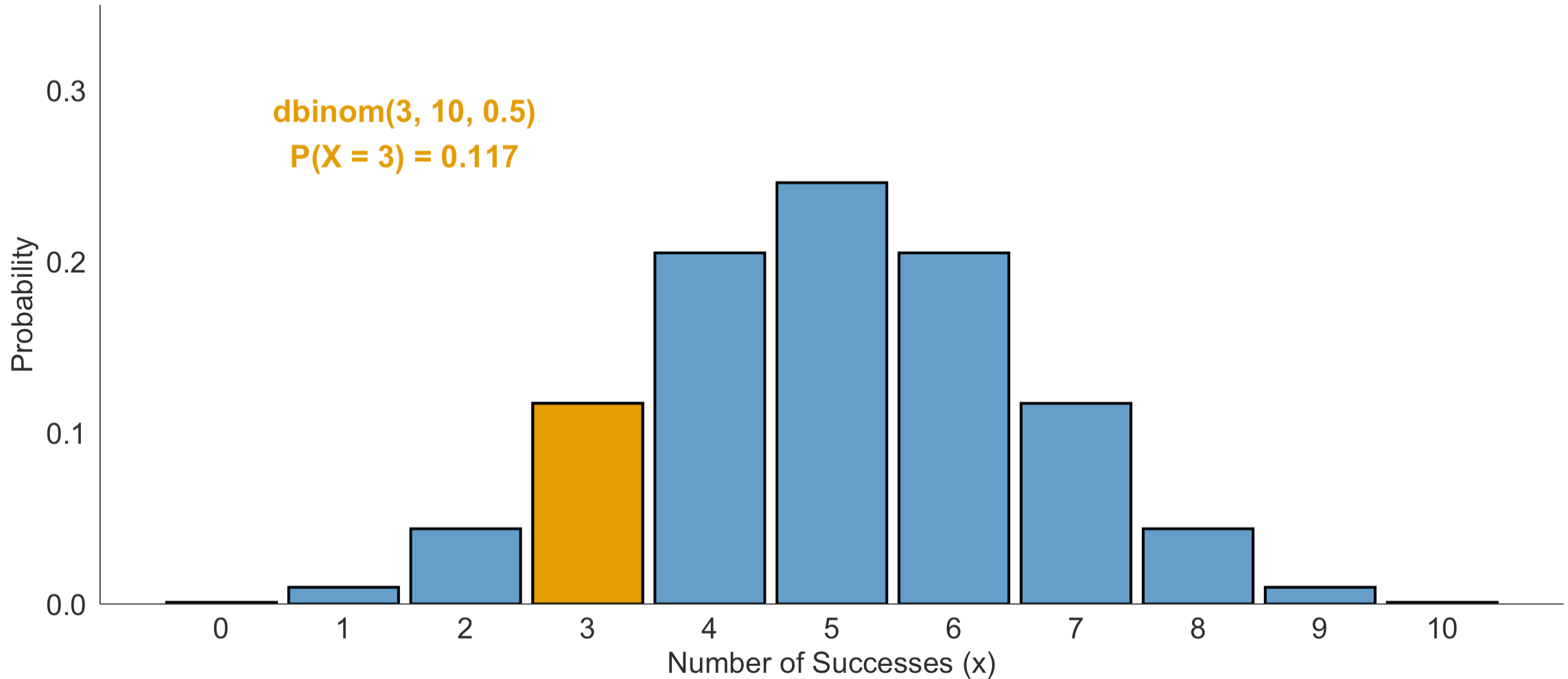
The four binomial functions:

```
1 # d = probability of EXACTLY x successes
2 dbinom(x = 3, size = 10, prob = 0.5)
3 # "What's P(X = 3)?"
4
5 # p = cumulative probability (at most x OR at least x)
6 pbinom(q = 3, size = 10, prob = 0.5)
7 # "What's P(X ≤ 3)?" (default)
8
9 pbinom(q = 3, size = 10, prob = 0.5, lower.tail = FALSE)
10 # "What's P(X > 3)?"
11
12 # q = quantile (what value gives this probability?)
13 qbinom(p = 0.25, size = 10, prob = 0.5)
14 # "What value has 25% of the distribution below it?"
15
16 # r = random samples
17 rbinom(n = 100, size = 10, prob = 0.5)
18 # "Give me 100 random draws"
```


Visual: dbinom() - EXACTLY

dbinom: Probability of EXACTLY x successes

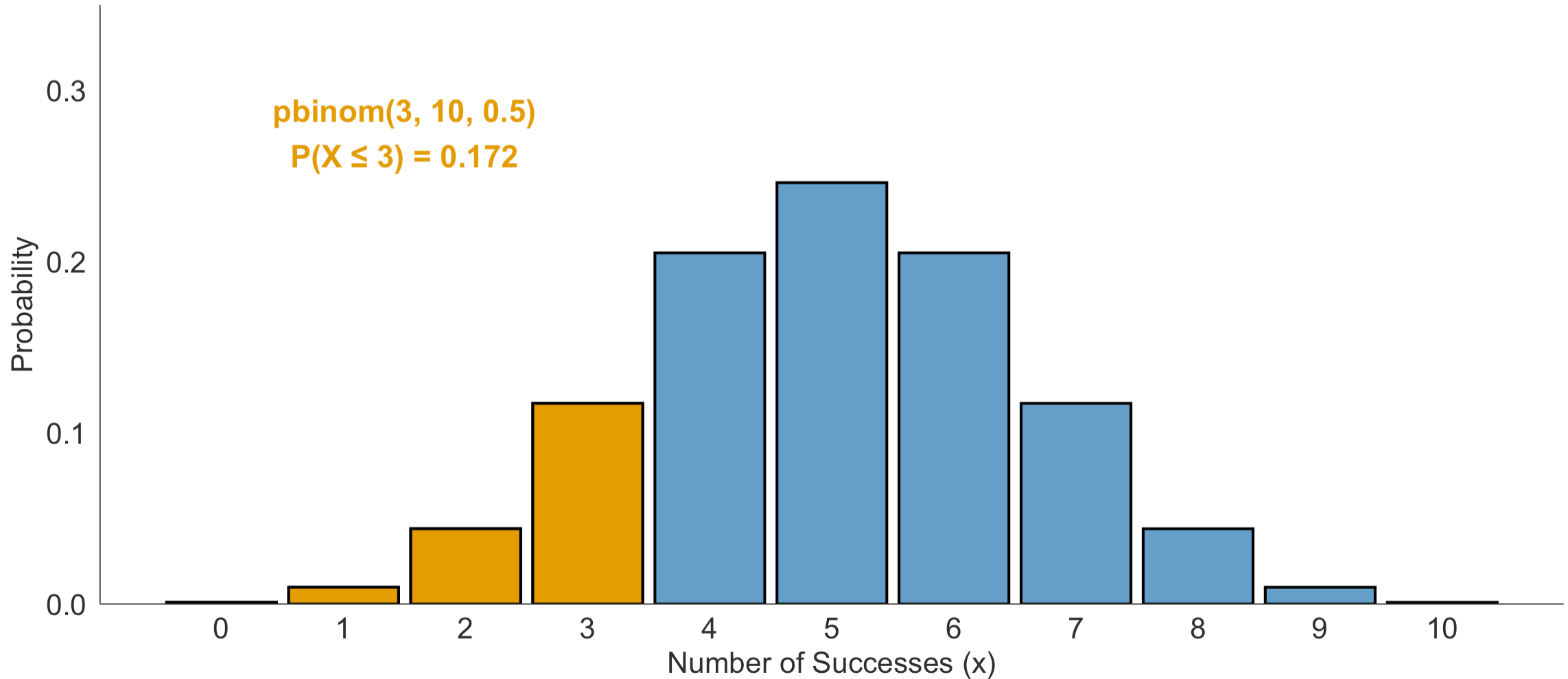
Binomial with $n = 10$, $p = 0.5$



Visual: pbinom() - AT MOST (default)

pbinom with lower.tail = TRUE (default): AT MOST x successes

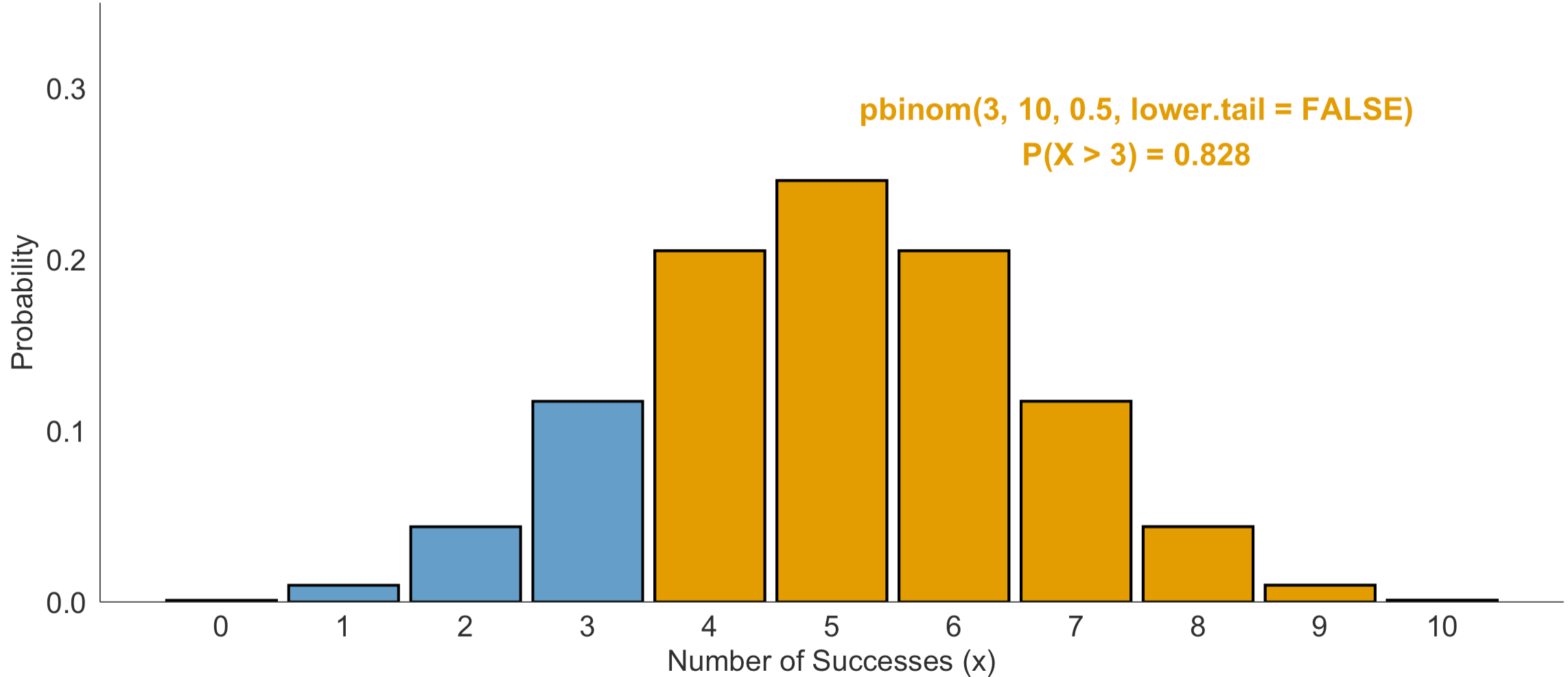
Binomial with $n = 10$, $p = 0.5$



Visual: pbinom(..., lower.tail = FALSE) - MORE THAN

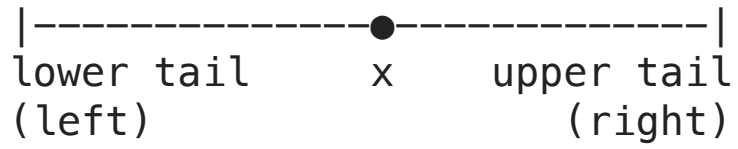
pbinom with lower.tail = FALSE: MORE THAN x successes

Binomial with $n = 10$, $p = 0.5$



The lower.tail = TRUE/FALSE Concept

Visual Guide



lower.tail = TRUE → gives you $P(X \leq x)$ — everything to the **LEFT** of (and including) x

lower.tail = FALSE → gives you $P(X > x)$ — everything to the **RIGHT** of x

Quick decision rule:

- Question has "**less than**" or "**at most**" → `lower.tail = TRUE` (default)
- Question has "**greater than**" or "**at least**" → `lower.tail = FALSE`

Common Confusion: "At least 5" (Discrete Distributions)

DISCRETE distributions only!

This slide applies to **binomial and Poisson** (discrete). For Normal (continuous), $P(X \geq 5) = P(X > 5)$ - no adjustment needed!

Question: What is $P(X \geq 5)$ for a binomial distribution?

Step 1: Translate to something the computer understands

- "At least 5" = "5 or more" = "greater than 4"
- So: $P(X \geq 5) = P(X > 4)$

Step 2: Code it

```
1 # Correct!
2 pbinom(q = 4, size = 10, prob = 0.5, lower.tail = FALSE)
3
4 # Also correct (but slightly more work)
5 1 - pbinom(q = 4, size = 10, prob = 0.5)
```

Practice: Binomial Questions

A vaccine is 75% effective. You vaccinate 20 people.

Which R function do you use?

Question 1: What's the probability that *exactly* 15 are protected?

1 # Your answer:

Question 2: What's the probability that *at most* 12 are protected?

1 # Your answer:

Question 3: What's the probability that *at least* 18 are protected?

1 # Your answer:

Question 4: What's the expected number protected?

1 # Your answer:

Practice: Binomial Answers (1/2)

A vaccine is 75% effective. You vaccinate 20 people.

Question 1: What's the probability that *exactly* 15 are protected?

```
1 # Question 1: Exactly 15
2 dbinom(x = 15, size = 20, prob = 0.75)
[1] 0.2023312
```

Question 2: What's the probability that *at most* 12 are protected?

```
1 # Question 2: At most 12
2 pbinom(q = 12, size = 20, prob = 0.75)
[1] 0.1018119
```

```
1 # OR: pbinom(q = 12, size = 20, prob = 0.75, lower.tail = TRUE)
```

Practice: Binomial Answers (2/2)

Question 3: What's the probability that *at least* 18 are protected?

```
1 # Question 3: At least 18 (same as > 17)
2 pbinom(q = 17, size = 20, prob = 0.75, lower.tail = FALSE)
[1] 0.09126043

1 # OR: 1 - pbinom(q = 17, size = 20, prob = 0.75)
```

Question 4: What's the expected number protected?

```
1 # Question 4: Expected value
2 n <- 20
3 p <- 0.75
4 expected <- n * p
5 expected
[1] 15

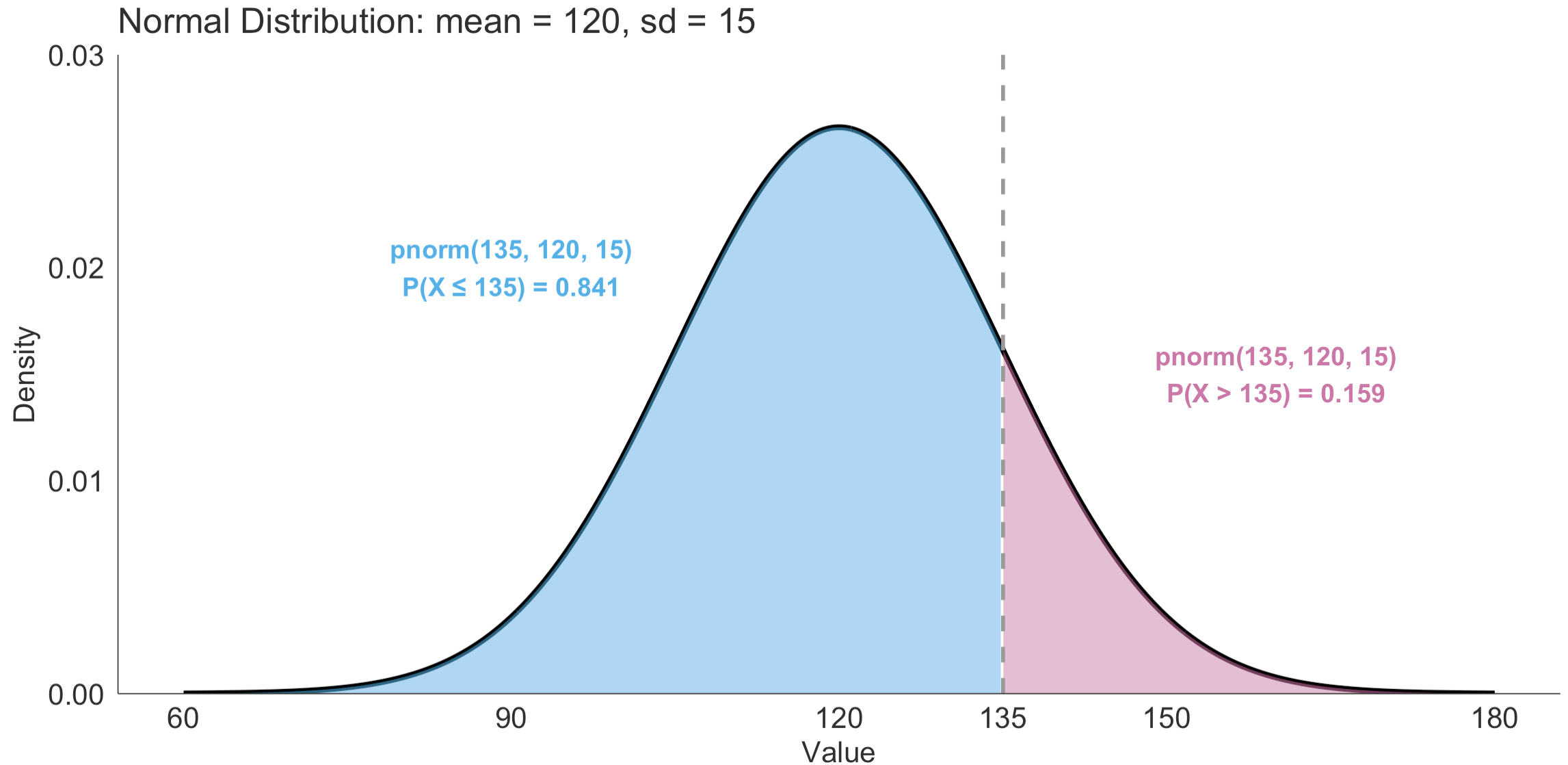
1 # Bonus: standard deviation
2 variance <- n * p * (1 - p)
3 st_dev <- sqrt(variance)
4 st_dev
[1] 1.936492
```


Normal Distribution Functions

The same pattern applies!

```
1 # d = density (height of curve at x)
2 dnorm(x = 120, mean = 100, sd = 15)
3 # Rarely used in practice
4
5 # p = cumulative probability
6 pnorm(q = 120, mean = 100, sd = 15)
7 # "What proportion have values ≤ 120?"
8
9 pnorm(q = 120, mean = 100, sd = 15, lower.tail = FALSE)
10 # "What proportion have values > 120?"
11
12 # q = quantile (most common!)
13 qnorm(p = 0.95, mean = 100, sd = 15)
14 # "What value has 95% below it?" (95th percentile)
15
16 # r = random samples
17 rnorm(n = 100, mean = 100, sd = 15)
18 # "Give me 100 random values from this distribution"
```

Visual: Normal Distribution



Practice: Normal Distribution

Blood pressure is Normal with mean = 120 mmHg, sd = 15 mmHg.

Which R function?

Question 1: What proportion have BP above 135?

1 # Your answer:

Question 3: What BP value is the 90th percentile?

1 # Your answer:

Question 2: What proportion have BP between 110 and 130?

1 # Your answer:

Question 4: If we want the middle 95%, what are the cutoffs?

1 # Your answer:

Practice: Normal Answers (1/2)

```
1 mu <- 120
2 sigma <- 15
3
4 # Question 1: Above 135
5 pnorm(135, mean = mu, sd = sigma, lower.tail = FALSE)
```

```
[1] 0.1586553
```

```
1 # OR
2 1 - pnorm(135, mean = mu, sd = sigma, lower.tail = TRUE)
```

```
[1] 0.1586553
```

```
1 # Question 2: Between 110 and 130
2 pnorm(130, mean = mu, sd = sigma) - pnorm(110, mean = mu, sd = sigma)
```

```
[1] 0.4950149
```

Practice: Normal Answers (2/2)

```
1 # Question 3: 90th percentile
2 qnorm(0.90, mean = mu, sd = sigma)
[1] 139.2233
```

```
1 # Question 4: Middle 95% (2.5th to 97.5th percentiles)
2 qnorm(0.025, mean = mu, sd = sigma) # Lower
[1] 90.60054
```

```
1 qnorm(0.975, mean = mu, sd = sigma) # Upper
[1] 149.3995
```

Sampling Distributions & CLT

Central Limit Theorem: What You Need to Know

The CLT in plain language

For a random sample of size n from ANY population with mean μ and SD σ :

The sampling distribution of \bar{X} is approximately normal when n is large (≥ 30)

With:

- Mean = μ
- Standard Error = σ / \sqrt{n}

Why this matters:

1. We can use normal distribution tools even if data aren't normal
2. Larger samples \rightarrow smaller standard error \rightarrow more precise estimates
3. This is the foundation for confidence intervals and hypothesis tests!

Standard Error vs. Standard Deviation

Don't confuse these!

Standard Deviation (s or σ):

- Measures spread of *individual observations*
- Describes variability in the data
- Formula: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Standard Error (SE):

- Measures uncertainty of the *sample mean*
- Describes variability of \bar{x} across (theoretical) samples
- Formula: $SE = \frac{s}{\sqrt{n}}$

Key insight: SE gets smaller as n increases, but s stays roughly the same!

Example: CLT in Action

Population: Mean = 5.1 hours, SD = 1.9 hours, right-skewed

Sample: $n = 40$

What's the sampling distribution of \bar{X} ?

```
1 # Parameters
2 mu <- 5.1
3 sigma <- 1.9
4 n <- 40
5
6 # Sampling distribution
7 mean_xbar <- mu
8 se_xbar <- sigma / sqrt(n)
9
10 mean_xbar
```

```
[1] 5.1
```

```
1 se_xbar
```

```
[1] 0.3004164
```

```
1 # Example probability about the sample mean, P(Xbar > 5.6)
2 pnorm(5.6, mean = mean_xbar, sd = se_xbar, lower.tail = FALSE)
```

```
[1] 0.04802059
```

Sampling distribution: approximately Normal(mean = μ , sd = $\frac{s}{\sqrt{n}}$). Even though the original data are right-skewed!



Confidence Intervals

Confidence Interval Formula

General form

point estimate \pm critical value $\times SE$

For a mean:

$$\bar{x} \pm t^* \times \frac{s}{\sqrt{n}}$$

Where:

- \bar{x} = sample mean
- t^* = critical value from t-distribution with $df = n - 1$
- s = sample standard deviation
- n = sample size

Calculating a 95% CI in R (1/2)

```
1 # Sample data
2 n <- 25
3 xbar <- 14.2
4 s <- 1.8
5 confidence_level <- 0.95
6
7 # Step 1: Find critical value
8 t_star <- qt((1 + confidence_level) / 2, df = n - 1)
9 # OR: qt(0.975, df = 24)
10 t_star
```

```
[1] 2.063899
```

```
1 # Step 2: Calculate standard error
2 SE <- s / sqrt(n)
3 SE
```

```
[1] 0.36
```

Calculating a 95% CI in R (2/2)

```
1 # Step 3: Calculate margin of error
2 margin_error <- t_star * SE
3 margin_error
```

```
[1] 0.7430035
```

```
1 # Step 4: Build the interval
2 lower <- xbar - margin_error
3 upper <- xbar + margin_error
4
5 c(lower, upper)
```

```
[1] 13.457 14.943
```

Interpreting Confidence Intervals

Correct interpretation

"We are 95% confident that the population mean is between 13.5 and 14.9."

What this means:

- If we repeated this process many times (new samples, new CIs)
- About 95% of those intervals would contain the true μ
- We don't know if *this specific interval* contains μ , but we're using a reliable method

Common mistakes - DON'T SAY:

- ❌ "There's a 95% probability that μ is in this interval"
- ❌ "95% of the data are in this interval"
- ❌ "We're 95% sure the sample mean is in this interval"

What Affects CI Width?

$$\text{Margin of error} = t^* \times \frac{s}{\sqrt{n}}$$

To get a narrower CI:

1. **Increase sample size** ($n \uparrow \rightarrow \text{SE} \downarrow \rightarrow \text{narrower CI}$)
2. **Decrease confidence level** ($95\% \rightarrow 90\% \rightarrow \text{smaller } t^* \rightarrow \text{narrower CI}$)
 - But this means less confidence!
3. **Reduce variability** ($s \downarrow \rightarrow \text{SE} \downarrow \rightarrow \text{narrower CI}$)
 - Often not under our control

Trade-off: Precision vs. Confidence

Hypothesis Testing

The Hypothesis Testing Framework

Six steps:

1. **State hypotheses:** H_0 and H_A
2. **Set significance level:** Usually $\alpha = 0.05$
3. **Check assumptions:** Independence, normality/large sample
4. **Calculate test statistic:** $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
5. **Find p-value:** Probability of seeing data this extreme if H_0 true
6. **Make conclusion:** Reject H_0 if p-value $< \alpha$

Writing Hypotheses

Template

$H_0 : \mu = \mu_0$ (null value)

$H_A : \mu \neq \mu_0$ (two-sided) OR $\mu > \mu_0$ OR $\mu < \mu_0$ (one-sided)

Example: Is mean body temperature different from 98.6°F?

- $H_0 : \mu = 98.6$
- $H_A : \mu \neq 98.6$

In words:

- H_0 : The population mean body temperature is 98.6°F
- H_A : The population mean body temperature is not 98.6°F

P-values: What They Mean

Definition

The p-value is the probability of observing data as extreme as (or more extreme than) what we observed, **assuming** H_0 is true.

Interpretation guide:

- p-value < 0.05 → Strong evidence against H_0 → Reject H_0
- p-value \geq 0.05 → Insufficient evidence → Fail to reject H_0

What p-values are NOT

- ✗ NOT the probability that H_0 is true
- ✗ NOT the probability of making an error
- ✗ NOT a measure of effect size or importance

One-Sample t-test in R

```
1 # Using raw data
2 t.test(data$variable,
3         mu = null_value,
4         alternative = "two.sided") # or "less" or "greater"
```

Example:

```
1 # Test if mean pulse rate is different from 72 bpm
2 data("nhanes.samp")
3
4 t.test(nhanes.samp$Pulse,
5         mu = 72,
6         alternative = "two.sided")
```

Understanding t.test() Output

One Sample t-test

```
data:  nhanes.samp$Pulse
t = 2.1383, df = 170, p-value = 0.03392
alternative hypothesis: true mean is not equal to 72
95 percent confidence interval:
 72.15999 76.00376
sample estimates:
mean of x
 74.08187
```

What to report:

- **t-statistic:** How many SEs away from null value
- **p-value:** Evidence against H_0
- **95% CI:** Range of plausible values for μ
- **Sample mean:** Our point estimate

Connection: CIs and Hypothesis Tests

Key relationship

For a two-sided test at $\alpha = 0.05$:

If the 95% CI does NOT contain μ_0 , then we reject $H_0 : \mu = \mu_0$

If the 95% CI DOES contain μ_0 , then we fail to reject H_0

Both methods use the same math - just different framing!

- CI: "What values are plausible for μ ?"
- Hypothesis test: "Is this specific value plausible?"

Final Reminders

Statistical Significance vs. Practical Significance

Don't confuse these!



Statistical significance ($p < 0.05$):

- Evidence that there's a real effect
- Doesn't tell you if the effect matters

Practical/Clinical significance:

- Effect is large enough to matter in real-world terms
- You must judge this based on context

Example: A drug lowers blood pressure by 0.5 mmHg ($p = 0.001$)

- Statistically significant?  Yes
- Clinically meaningful?  No (too small to affect patient outcomes)

Common Mistakes to Avoid

1. **Forgetting to check assumptions** before running tests
2. **Confusing standard deviation and standard error**
3. Using `lower.tail = TRUE` when you need `FALSE` (or vice versa)
4. **Misinterpreting p-values and confidence intervals**
5. **Forgetting to include** `na.rm = TRUE` when there are missing values
6. **Not rendering your document before submitting**

Pro tip: When using `pbinom()` or `ppois()`, sketch a quick picture to check if you want the left or right tail!

Exam Strategy

Time management:

- Don't get stuck on one question
- Move on and come back if needed
- Leave buffer time for rendering and checking

Interpretation questions:

- Write in complete sentences
- Reference numbers from your calculations
- Connect to context (not just abstract statistics)

Coding questions:

- Show your work (don't just give answers)
- Use comments to explain your logic
- Check that your code actually runs!

Resources

During the exam:

- Lecture slides (especially the summary slides)
- Homework solutions (you've seen similar questions!)
- R help: `?t.test`, `?pnorm`, etc.

After you submit:

- The exam will close Monday at 11 PM
- Solutions will be available after
- We'll get feedback to you within a week
- Remember: This is a learning experience!

Questions?

Your Turn!

What questions do you have?





- Concepts that are still unclear?
- R functions you're unsure about?
- Specific exam questions you want to discuss?

Remember:

- Exam opens today at 3 PM
- You have until Monday at 11 PM
- You've got this! 💪

Good luck!

You're well prepared:

-  You've done the homework
-  You've come to class
-  You've practiced with R
-  You know where to find help

Trust your preparation and think carefully about each question.