ST. MARY'S SCHOOL

Medford, Oregon / 2026–2027

Introduction to Statistics

Statistics Workbook

Class investigations, independent practice, homework, and preparation readings.



Materials customized by Jason Javarone for use in the St. Mary's School Applied Math course.

Adapted for St. Mary's School from its AP Statistics materials and Introductory Statistics 2e by Barbara Illowsky and Susan Dean, OpenStax, Rice University. Examples, activities, and sequencing are customized for Applied Math. The original textbook is free at OpenStax.org.

https://openstax.org/details/books/introductory-statistics-2e

OpenStax textbook © 2026 Rice University. This adaptation uses CC BY-NC-SA 4.0. School insignia and previously approved artwork retain their original ownership.

https://creativecommons.org/licenses/by-nc-sa/4.0/

STUDENT EDITION

Using your workbook

In class

On Day 1, complete PR01 and its two self-checks with your teacher. Each day, use the Student Reference for the worked example, then collect data and complete the in-class practice here.

After each class	What to do with the pages
1. Complete HW	Write your name and date on the homework sheet. Do all numbered questions on the front and back. Turn in only this sheet when requested.
2. Read the next lesson's preview	The reading starts on the next sheet. Study the example and display, then answer the two self-checks on the reverse. Keep both pages in your binder.
3. Check your reasoning	Follow the W page number beneath the self-checks to the response guide. Correct your work and bring any remaining question to class. The next entry ticket draws on the previous class and the reading.

The first evening

After Day 1, complete HW01 on W009 and its reverse. Turn the page to PR02 on W011, then answer both checks on the back of that reading. At the start of Day 2, your entry ticket draws on the stopwatch investigation and the reading about distributions.

Know which data to use

Use your class measurements when an activity asks you to collect data. When a problem gives a dataset, use it as printed. If your teacher supplies practice data in place of a live activity, identify the dataset in your answer.

Write a conclusion someone else can follow

Name the group and variable, give the evidence with its units, and explain what the result supports. Show the reasoning behind your calculations.

Supplied school and regional datasets are constructed classroom examples unless identified as documented observations.

DAILY DIRECTORY

Find your classwork and tonight's assignment

Unit 1 • Data and study design

Day	Class topic	Classwork	Homework	Preview	Begins
1	What exactly are we measuring?	W007	W009	PR02	W011
2	What a display makes visible	W013	W015	PR03	W017
3	Compare a typical value and its spread	W019	W021	PR04	W023
4	Compare observations across different groups	W025	W027	PR05	W029
5	Which claims does the study design support?	W031	W033	PR06	W035
6	Checkpoint and revision: describe and compare	W037	W039	PR07	W041

Unit 2 • Prediction and robustness

Day	Class topic	Classwork	Homework	Preview	Begins
7	Reading a scatterplot	W043	W045	PR08	W047
8	Predicting a value and checking the error	W049	W051	PR09	W053
9	Does traffic change the recommendation?	W055	W057	PR10	W059
10	Where does the model have support?	W061	W063	PR11	W065
11	Writing a model recommendation	W067	W069	PR12	W071

Unit 3 • Estimation and confidence

Day	Class topic	Classwork	Homework	Preview	Begins
12	One sample, one estimate	W073	W075	PR13	W077
13	Why sample size changes spread	W079	W081	PR14	W083

Day 1 orientation in class • PR01 • W005

Read across the row for the day you just completed. Submit the HW sheet when requested. Keep the preview and its two self-checks in your binder.

Self-check response guide • W155

Statistical formulas • W159

DAILY DIRECTORY

Find your classwork and tonight's assignment

Unit 3 • Estimation and confidence

Day	Class topic	Classwork	Homework	Preview	Begins
14	What 95% confidence means	W085	W087	PR15	W089
15	Using an interval to make a decision	W091	W095	PR16	W097
16	How far can an estimate reach?	W099	W101	PR17	W103

Unit 4 • Significance and decisions

Day	Class topic	Classwork	Homework	Preview	Begins
17	Simulate a benchmark	W105	W107	PR18	W109
18	How strong and how large?	W111	W113	PR19	W115
19	Two ways a decision can go wrong	W117	W119	PR20	W121
20	Power at a stated alternative	W123	W125	PR21	W127
21	Defend a decision rule	W129	W131	PR22	W133

Unit 5 • Synthesis and revision

Day	Class topic	Classwork	Homework	Preview	Begins
22	Repair a statistical explanation	W135	W137	PR23	W139
23	Write a decision memo from a new case	W141	W143	PR24	W145
24	Independent performance task	W147	W149	PR25	W151
25	Revision and independent transfer	W153	None	None	—

Read across the row for the day you just completed. Submit the HW sheet when requested. Keep the preview and its two self-checks in your binder.

Self-check response guide • W155

Statistical formulas • W159

UNIT 1 / DAY 1 IN CLASS ● PR01 ● PREPARATION READING

Start in class: what does a record describe?

Read together at the first meeting. Nothing was due before Day 1.

Identify a case, variable, and denominator before interpreting a percentage.

Before a team starts serve practice, its coach asks, "What percentage of our practice serves will land in bounds today?" This is a statistical question. Some serves may land in and others out, so the coach needs a set of observations. "Did the first serve land in?" takes just one observation.

Picture a score sheet with one row for every serve. One serve is a case, the event that the row describes. The variable is what we record for that case. Here it is the outcome, with two possible entries: in or out. A player can appear on several rows.

Suppose one player takes 12 indoor practice serves and six land in bounds. To find the proportion in bounds, divide the number that landed in by the total number attempted. Six divided by 12 is 0.50, or 50%. A reader who never saw the score sheet could still understand this sentence: "Six of the player's 12 practice serves today landed in bounds."

A second player takes six serves outdoors. The table below now accounts for 18 attempts. Before we can calculate the combined percentage, we need to know how many of the second player's serves landed in bounds. To compare the players, we would also keep their records separate. The first player took twice as many serves, and the practice conditions differed. A tally of attempts can tell us about those serves. To say anything about a percentage of players, we would have to count players and define what result qualifies a player.

Practice record	What can we count?
First 12 recorded serves	6 in bounds out of 12 attempts
Second player's 6 serves	6 more attempts; in-bounds count not given
All recorded serves	18 attempts; need the in-bounds count from the next six serves to find the combined share

1

UNIT 1 / DAY 1 IN CLASS • PR01 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> A percentage describes the group counted in its denominator.

1. Suppose four of the second player's six serves land in bounds. Write one sentence reporting the combined result. Show what you counted in the numerator and denominator.

2. A teammate describes the combined result as the percentage of the team's players who served in bounds. What did you actually count? How would you define success and count it in a study of players?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Player	Serves in bounds	Total serves
First player	6	12
Second player (check 1)	4	6
Both players		

After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §§1.1–1.2, pp. 5–13. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 1 / DAY 01 • CLASSWORK • T1

Can the class estimate ten seconds?

Define a measurable question and write a claim whose population, variable, and evidence are clear.

Answer question 1 before collecting. Work in pairs. One partner says "start," starts a hidden stopwatch, and stops it when the estimator signals that ten seconds seem to have passed. Round the displayed time to the nearest whole second and record that whole number. Count a recorded 9, 10, or 11 seconds as a hit; every other recorded whole number is a miss. Make ten trials per student, switching roles after ten. Add each pair's hits and trials to the class total once.

Student	1	2	3	4	5	6	7	8	9	10	Hits
Partner A											
Partner B											

1. Before collecting data, write a statistical question this exercise could investigate. What is one case, what will you measure, and in what units?

2. Find the class totals for hits and trials, counting each trial once. Calculate the percentage of trials that were hits. Write a brief interpretation and conclusion that identifies the class activity, the scoring rule, and the group the result describes.

3. State one limit on the conclusion you can draw. Explain how the way we collected the data creates that limit.

1

Name _____ Date _____

UNIT 1 / DAY 01 • INDEPENDENT WORK

A trail counter

A volunteer counted helmet use among 20 cyclists who passed one Greenway entrance between 3 and 4 p.m. on Tuesday. Fifteen wore helmets. These are illustrative counts.

1. Name the case and measured variable. Write a complete quantitative statement about the observed group.

2. Can the volunteer use this count to report helmet use among all local cyclists? Identify one group the count could miss.

Revise: Check the denominator in your helmet-use statement. Rewrite the statement so a reader can tell exactly whom the count describes.

Exit: Of 18 trays returned during one lunch period, 12 held an unopened milk carton. Give the percentage and identify the group it describes.

Name _____ Date _____

UNIT 1 / DAY 1 HOMEWORK • HW01 • PRACTICE

Read the data record before the headline

After Day 1: Homework and reading to prepare for Day 2. Turn in this two-sided HW01 sheet when requested. Keep PR02 on W011 in your binder.

A gym records how people arrive for each of 24 visits between 4 and 5 p.m. on Monday. Nine visits begin with the visitor arriving on foot. A library times 12 checkout transactions; four take more than three minutes. Both records are classroom examples.

Study	One case	Variable	Units or categories
Gym			
Library			

1. Complete the table for each study. Then write one statistical question each study could answer. Make clear which group your question concerns.

1

Name _____ Date _____

UNIT 1 / DAY 1 HOMEWORK • HW01 • PRACTICE CONTINUED

Read the data record before the headline

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR02 on W011 and answer its two self-checks on W012 before Day 2. Keep that sheet in your binder.

2. Find the fraction of gym visits made on foot and the fraction of library transactions that took over three minutes. Use your results to rewrite "37.5% of town residents walk" and "One third of library members are slow."

3. The gym manager wants to know the share of all gym visits made on foot this month. Which visits are missing from Monday's record? Give one reason those visitors' arrival methods might differ.

UNIT 1 / AFTER DAY 1 / FOR DAY 2 ● PR02 ● PREPARATION READING

From individual waits to a distribution

Complete after Day 1 to prepare for Day 2.

Read a dotplot and a histogram, and calculate the mean from individual values.

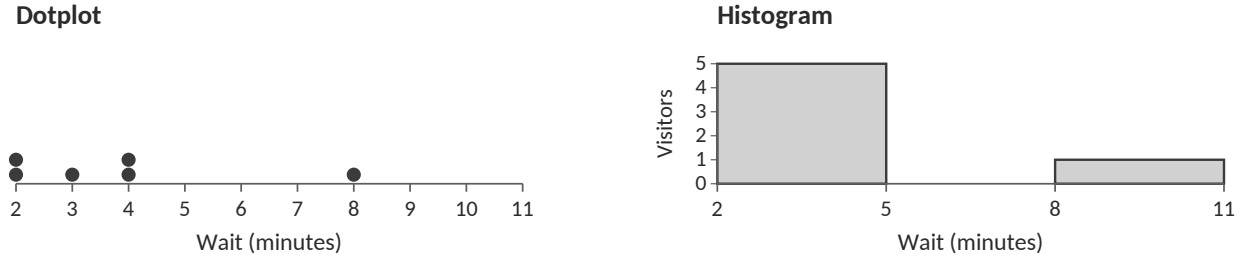
Imagine that you watched six visitors join the library line and recorded each wait in minutes. The waits were 2, 2, 3, 4, 4, and 8. A distribution shows which values occurred and how often. In the dotplot below, read along the bottom to find a time, then count the dots above it. The stack at 4 represents two visitors who each waited four minutes. Six dots account for all six visits.

The histogram shows the same waits grouped into intervals called bins. The first bar counts waits of at least 2 minutes and less than 5. It includes 2, 2, 3, 4, and 4, so it is five visitors tall. An observation on a boundary goes into the bin that begins there. The 8-minute wait therefore belongs in the third bin. The middle bin is empty. The bar heights, 5, 0, and 1, also add to six visitors.

To find the mean, think of sharing all the waiting time equally across the six visits. The total is 2 + 2 + 3 + 4 + 4 + 8 = 23 minutes. Dividing by six gives about 3.8 minutes per visit. The repeated times count each time they occur. Dividing by four, the number of different times on the list, would give the wrong denominator.

Now imagine moving a single dot from 4 to 3. One visit became a minute shorter, so the total falls by a minute and the mean falls by one sixth of a minute. Both positions belong to the first bin, so no bar changes height. The dotplot preserves enough detail to calculate the exact mean. These grouped bars alone leave several different lists of waits possible.

Six library waits



Each dot counts one visitor. Each bar counts the visitors whose waits fall in that interval.

1

UNIT 1 / AFTER DAY 1 / FOR DAY 2 • PR02 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

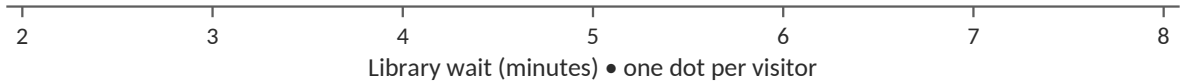
> **KEY IDEA**
>
> A dotplot preserves individual values; a histogram groups values into intervals.

1. Start with the six original waits. A visitor recorded at 2 minutes actually waited 3 minutes. Find the corrected mean in minutes and show how you got it.

2. Would the original histogram reveal the correction in question 1? Identify the relevant bin and explain what its bar counts.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



Library wait (minutes) • one dot per visitor

After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §1.1, pp. 5–6; §2.2, pp. 75–78; §2.5, pp. 98–100. Examples and data are constructed for this course.

Name _____ Date _____

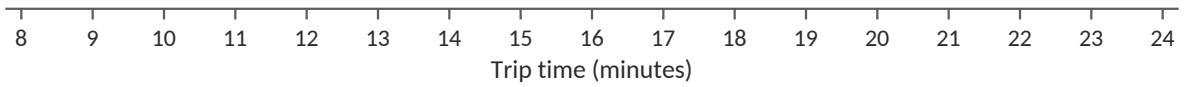
UNIT 1 / DAY 02 • CLASSWORK • T2

Build two displays from one sample

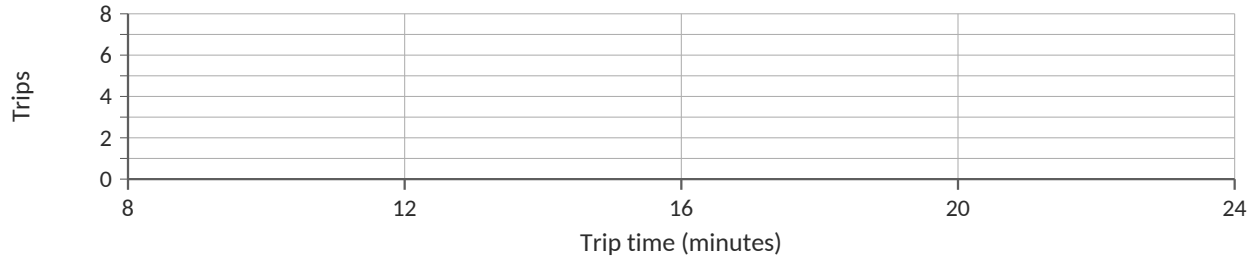
Build a dotplot and histogram from the same quantitative data, then describe a feature supported by the display.

Use these ten supplied practice shuttle times, in minutes: 8, 9, 9, 10, 10, 11, 11, 13, 19, 20. Draw the dotplot and histogram in separate spaces. In a bin such as [8,12), include values of at least 8 and less than 12. Apply that rule to every bin.

Dotplot



Histogram



1. Draw a labeled dotplot. Identify the number of trips at 9 minutes and the number taking at least 13 minutes.

2. Draw a histogram with bins [8,12), [12,16), [16,20), and [20,24). Identify the bin for the 20-minute trip and explain why it belongs there.

3. Write a statement about a cluster or gap you can locate more precisely on the dotplot. How much detail did grouping remove?

1

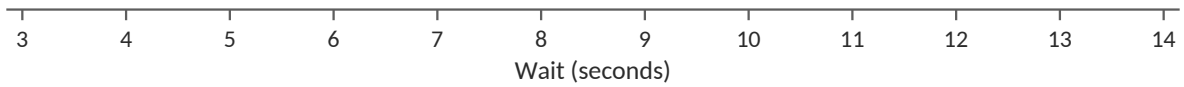
Name _____ Date _____

UNIT 1 / DAY 02 • INDEPENDENT WORK

Waiting for a print job

Eight illustrative print waits, in seconds, are 3, 4, 4, 5, 5, 6, 7, 14.

Dotplot



1. Plot one dot per print job. Then count the waits in [3,7), [7,11), and [11,15). Check that the number of dots and the sum of the three bin counts both equal eight.

2. Find the mean wait for the eight print jobs and show your calculation. Then write a sentence about the distribution that the mean alone does not convey.

Revise: After feedback, choose one answer above. Rewrite one sentence using a number, a unit, or a clearer description of whom the data represent.

Exit: Four different events fell into the histogram bin [10,15) minutes. Could all four have lasted exactly 12 minutes? What else is possible?

Name _____ Date _____

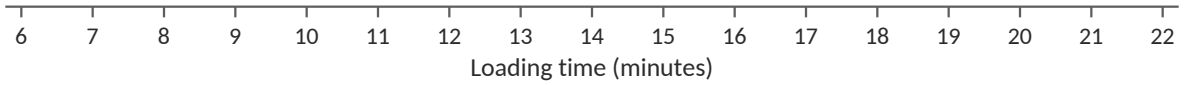
UNIT 1 / DAY 2 HOMEWORK • HW02 • PRACTICE

Display a set of orchard loading times

After Day 2: Homework and reading to prepare for Day 3. Turn in this two-sided HW02 sheet when requested. Keep PR03 on W017 in your binder.

Nine illustrative loading times in minutes are 6, 7, 7, 7, 8, 9, 10, 16, 18.

Dotplot



1. Draw a labeled dotplot. Annotate one stack and one gap with sentences that identify what the marks represent.

1

Name _____ Date _____

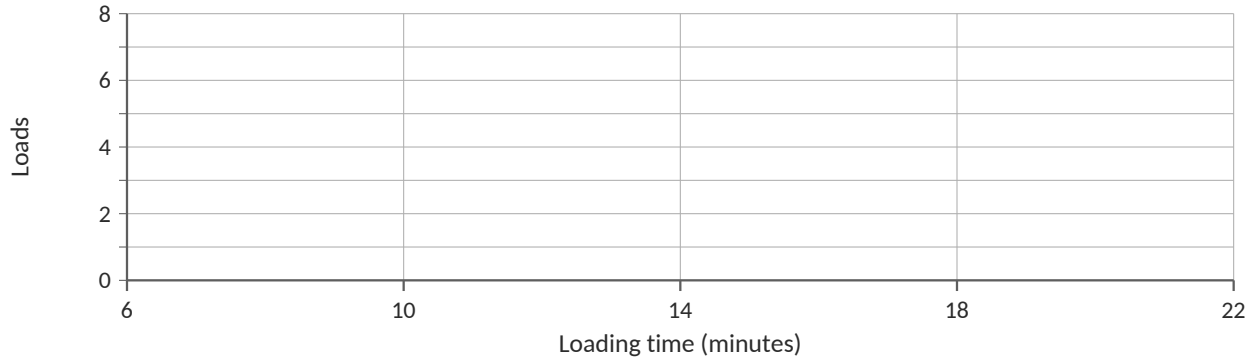
UNIT 1 / DAY 2 HOMEWORK • HW02 • PRACTICE CONTINUED

Display a set of orchard loading times

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR03 on W017 and answer its two self-checks on W018 before Day 3. Keep that sheet in your binder.

Histogram



2. Use the histogram axes on this page. Draw a histogram with bins [6,10), [10,14), [14,18), and [18,22). Label it in minutes and loads. Do the 16- and 18-minute loads fall in the same bin? Use the bin boundaries to explain.

3. Write two sentences describing these loads to a scheduler. Explain what you would lose if the only reported number were the mean.

UNIT 1 / AFTER DAY 2 / FOR DAY 3 ● PR03 ● PREPARATION READING

Why one long wait changes the mean

Complete after Day 2 to prepare for Day 3.

Distinguish the mean from the median and explain their different responses to an extreme value.

Five people wait at a ticket booth for 2, 3, 3, 4, and 8 minutes. How long is a typical wait? If we share their 20 minutes equally among five people, the result is the mean, 4 minutes. If we line the waits up from shortest to longest and pick the middle position, the result is the median. The third wait is 3 minutes. Those answers describe different features of the same line.

Suppose the longest wait should have been recorded as 18 minutes. The total rises by 10 minutes and the number of people stays at five. The mean therefore rises by 10 ÷ 5 = 2 minutes, from 4 to 6. Look at the two rows in the table. The value in the third position stays at 3 minutes even as the longest wait increases.

The median is called resistant because moving an already extreme value farther away can leave the middle untouched. A median can change when a corrected observation moves past the middle position, so put the observations back in order after a correction. For an even number of waits, average the two middle values. For example, the median of 2, 3, 5, and 10 minutes is halfway between 3 and 5, or 4 minutes.

A visitor asking how long people typically wait might prefer the median when a few waits are unusually long. A manager who wants the total waiting time for all five visitors can recover it by multiplying the mean by five. Neither summary on its own tells us how different one person's wait could be from another's. The first and third quartiles mark off the middle half of the ordered waits. If that half runs from 3 to 5 minutes at one booth and from 2 to 8 at another, its width is 2 minutes at the first booth and 6 at the second. This width is the interquartile range (IQR).

Wait record	Ordered times (minutes)
Observed waits	2, 3, 3, 4, 8
Change the longest	2, 3, 3, 4, 18

1

UNIT 1 / AFTER DAY 2 / FOR DAY 3 • PR03 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
> The mean uses every value; the median depends on the middle position after sorting.

1. Start with the original five waits. If the longest wait was actually 13 minutes, by how much would the mean change? Give the revised mean in minutes.

2. Consider a different correction to the original list. One of the 3-minute waits should be 9 minutes. Find the new median and explain why it changes in this case.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Wait record	Sum (minutes)	Number of waits	Mean (minutes)	Median (minutes)
Original		5		
Check 1: 8 becomes 13		5		
Check 2: one 3 becomes 9		5		

After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §§2.3–2.7, pp. 86–110. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 1 / DAY 03 • CLASSWORK • T2

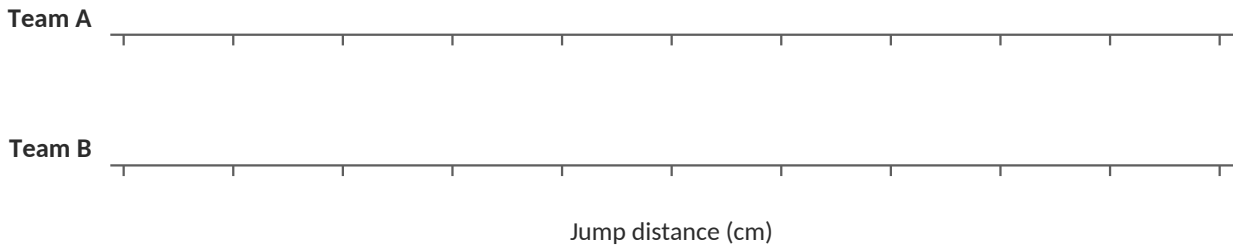
How consistent are two jump sets?

Choose summaries for two distributions and support a comparison with both center and spread.

Use two teams, A and B, with one frog and one launch lane per team. Within a team, keep the frog, surface, and launcher unchanged for eight jumps. Measure the straight-line distance from the front toes at takeoff to the front toes at first landing, to the nearest centimeter. Record all eight distances, including short jumps, in your team's row and copy the other team's row. If the class cannot collect jumps, your teacher will give you two practice sets to enter in the table. Compare the jump sets; differences in frogs or launchers prevent attributing a difference to just one cause.

Jump distance (cm)	1	2	3	4	5	6	7	8
Team A								
Team B								

Choose a common scale that includes all the jumps. Label the tick marks below.



1. Make a dotplot for each team using the same distance scale. Is there a jump far from the others? Explain how that affects which summaries you would compare.

2. For each team, put its eight distances in order. Find the median, the median of each half (Q1 and Q3), and the IQR. Compare both teams in centimeters and give one limit of the comparison.

3. Leave the recorded results intact. Imagine Team A's longest jump was 16 cm longer. Predict, then check, how much its mean, median, and IQR would change.

1

Name _____ Date _____

UNIT 1 / DAY 03 • INDEPENDENT WORK

Two trail crews

Illustrative packing times in minutes: Crew A, 4, 5, 5, 6, 7, 8, 10, 23; Crew B, 5, 6, 6, 7, 8, 9, 10, 11.

1. Compute the median and IQR for both crews using the median-of-halves rule. Which crew has the shorter typical time?

2. Write a standalone comparison of typical time and spread. Explain what the 23-minute pack tells the next crew lead.

Revise: Read your crew comparison without the prompt. Make clear which crew had the shorter median and whether it also had the smaller IQR.

Exit: For seven deliveries, the longest time increases by 14 minutes and stays longest. What happens to the mean and median? Give the reason.

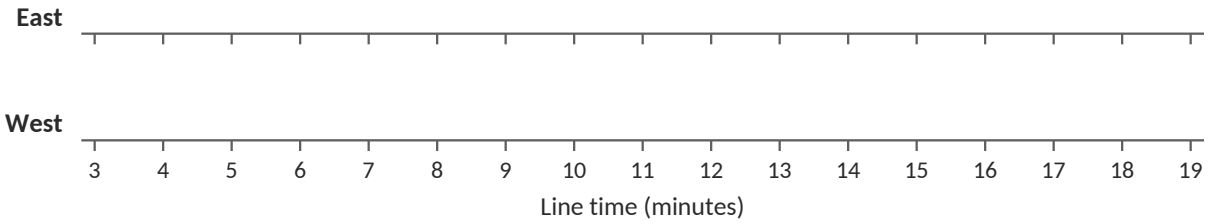
Name _____ Date _____

UNIT 1 / DAY 3 HOMEWORK • HW03 • PRACTICE

Choose a summary that fits the question

After Day 3: Homework and reading to prepare for Day 4. Turn in this two-sided HW03 sheet when requested. Keep PR04 on W023 in your binder.

Illustrative snack line times, in minutes: East, 3, 4, 4, 5, 6, 7, 8, 19; West, 4, 5, 5, 6, 7, 8, 9, 10.



1. Sketch two aligned dotplots. Compute each median and IQR, then write a comparison of typical time and variability for these observed visits.

1

Name _____ Date _____

UNIT 1 / DAY 3 HOMEWORK • HW03 • PRACTICE CONTINUED

Choose a summary that fits the question

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR04 on W023 and answer its two self-checks on W024 before Day 4. Keep that sheet in your binder.

2. Suppose the longest East wait had been 35 minutes instead of 19. Predict how the mean, median, and range would change. Use the original and revised sums to check the two means.

3. Two other lines have roughly symmetric waits and no unusual extremes. Line A has mean 12 and SD 2 minutes. Line B has mean 15 and SD 3 minutes. Compare their typical waits and variability. Why are mean and SD useful here?

UNIT 1 / AFTER DAY 3 / FOR DAY 4 ● PR04 ● PREPARATION READING

A fair comparison needs a scale

Complete after Day 3 to prepare for Day 4.

Calculate and interpret a standardized score using the correct group's mean and standard deviation.

At a meet, one runner finishes an event in 22 minutes and another finishes a different event in 15 minutes. The second runner spent fewer minutes on the course. Which runner finished farther below the average time for their own event? We need to know how much finishing times typically vary in each event before we can answer.

The first event has a mean of 24 minutes and a standard deviation (SD) of 1 minute. Standard deviation measures how spread out values are around the mean. The first runner's time is 22 − 24 = −2 minutes from the mean. With one minute to each SD, that is two SDs below the event mean. The second event has a mean of 18 minutes and an SD of 3 minutes. The second runner's time is 15 − 18 = −3 minutes from the mean, exactly one SD below it.

The signed number of SDs is the standardized score, or z-score. Subtract the group mean from the observed value and divide by that same group's SD. The two runners have z-scores of −2 and −1. Minutes cancel when you divide minutes by minutes. On the standardized scale the first runner is farther below the event mean, even though the second runner finished in less time and had a larger raw gap in minutes.

A positive z-score places a time above its event mean. A score of zero places it right at the mean. If every time is the same, the SD is zero and a z-score cannot be calculated. To turn a z-score into a ranking, we need more information. For example, knowing a time is two SDs below the mean does not tell us exactly how many runners finished faster. Their actual times or a suitable model of the distribution could answer that question.

Runner	Finish	Event mean	Event SD
1	22 min	24 min	1 min
2	15 min	18 min	3 min

1

UNIT 1 / AFTER DAY 3 / FOR DAY 4 • PR04 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

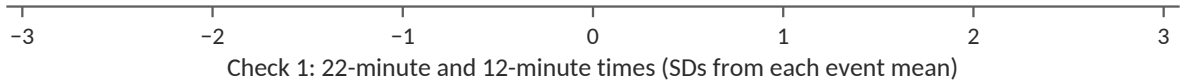
> **KEY IDEA**
>
> A z-score gives the signed distance from a group's mean in units of that group's standard deviation.

1. A third runner completes the second event in 12 minutes. How does that runner compare with the 22-minute runner in the first event on the standardized scale? Show your reasoning.

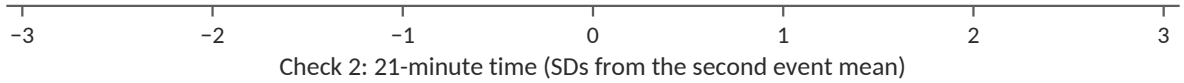
2. Another runner finishes the second event in 21 minutes. A teammate says the time was three SDs above average. Explain how far above the event mean it actually was.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



Check 1: 22-minute and 12-minute times (SDs from each event mean)



Check 2: 21-minute time (SDs from the second event mean)

After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §2.7, pp. 107–110; §6.1, pp. 336–338. Examples and data are constructed for this course.

Name _____ Date _____

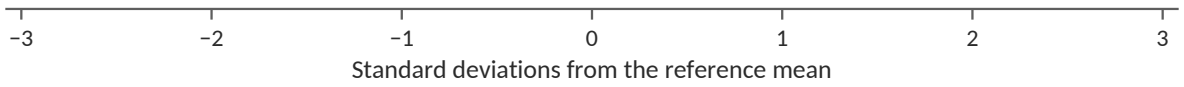
UNIT 1 / DAY 04 • CLASSWORK • T3

Sort the finishes two ways

Calculate and interpret a signed standardized score using the mean and SD of the relevant group.

The table gives one entrant's packing time, the mean time for that event, and its standard deviation (SD). In all three events, a shorter time is faster.

Event	Entrant time (min)	Event mean (min)	Event SD (min)	Time − mean (min)	z-score
North	8	12	2		
South	7	10	3		
West	12	9	1.5		



Standard deviations from the reference mean

1. Rank the three entrants by clock time, fastest first. Complete the difference and z-score columns and mark the three scores on the number line.

2. Who finished farthest below their own event mean in SD units? Which entrants were farthest from their own means in either direction? Report any tie.

3. A student says, "West was two minutes slower than its event mean because its z-score is +2." Check both the number of minutes and the number of SDs.

1

Name _____ Date _____

UNIT 1 / DAY 04 • INDEPENDENT WORK

Which instrument price stands out?

A music school is considering a $3,000 piano, a $550 guitar, and a $600 drum set. Comparable piano prices have mean $4,000 and SD $2,500. For guitars the mean is $500 and SD $200. For drum sets they are $700 and $100. The prices come from an OpenStax textbook exercise.

Data adapted from OpenStax, Introductory Statistics 2e, §2.7, homework exercise 104, pp. 146–147. Access for free at openstax.org.

1. Find the signed z-score for each price. Which price is farthest from its own group mean in dollars? Which is farthest in standard deviations?

2. Interpret the drum set's score for the music director. Can the director conclude from this score alone that the price is in the cheapest 10% of drum sets? Explain.

Revise: Check that your comparison names each instrument's own reference group and distinguishes a dollar difference from a distance in SDs.

Exit: A delivery takes 23 minutes on a route with mean 29 and SD 3 minutes. Give z and interpret it in context.

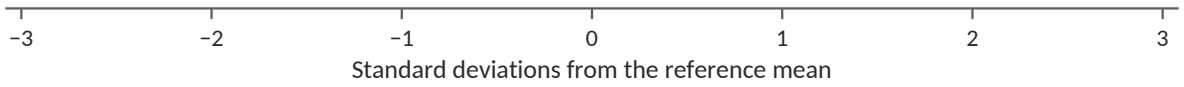
Name _____ Date _____

UNIT 1 / DAY 4 HOMEWORK • HW04 • PRACTICE

Absolute and relative comparisons

After Day 4: Homework and reading to prepare for Day 5. Turn in this two-sided HW04 sheet when requested. Keep PR05 on W029 in your binder.

Two relay legs were timed. Group A's runner took 21 seconds (group mean 25, SD 2 seconds). Group B's runner took 19 seconds (group mean 22, SD 3 seconds). Lower times are faster.



1. Calculate and sketch both z-scores on one standardized axis. Identify the shorter recorded time and the time farther below its own group mean in SD units. Explain why the answers differ.

1

Name _____ Date _____

UNIT 1 / DAY 4 HOMEWORK • HW04 • PRACTICE CONTINUED

Absolute and relative comparisons

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR05 on W029 and answer its two self-checks on W030 before Day 5. Keep that sheet in your binder.

2. A third runner's z-score is +1.5 in Group B. Find that runner's time in seconds and explain what the positive score means in this event.

3. A coach says z = −2 means the runner was 2% faster than average. Diagnose the error, calculate the actual percent below A's mean, and supply a replacement statement.

UNIT 1 / AFTER DAY 4 / FOR DAY 5 ● PR05 ● PREPARATION READING

Two different uses of chance

Complete after Day 4 to prepare for Day 5.

Distinguish random selection from random assignment and state what each supports.

A school help-desk manager wants to know the average wait during a typical shift. Each shift supervisor records the average wait for that shift. The manager draws 30 shifts at random from the complete term roster and collects their averages. Chance decides which shifts enter the study. This is random selection. The list from which the shifts are drawn is the sampling frame.

The results can help estimate the average of the recorded shift averages across the term. Why does the roster matter? If evening shifts were accidentally left off the list, no evening shift could be drawn. The result could still describe the listed shifts, but long evening queues would go uncounted. Check who or what had a chance to enter the study.

Now suppose the question is whether a new sign helps visitors finish a task faster. Thirty volunteers take part. The manager shuffles fifteen "old sign" cards and fifteen "new sign" cards and gives one to each volunteer. Everyone attempts the same task under the same conditions except for the sign. This is random assignment. It decides which sign a volunteer sees. It happens after the volunteers have joined.

Imagine instead letting visitors choose their signs. Experienced visitors might prefer the familiar sign and finish quickly because they know the task. That would make the sign's effect hard to separate from experience. Shuffling the cards gives experience a chance to be distributed across both groups. The groups may still differ by chance. A comparison can provide evidence about the signs' effects for these volunteers. People who chose to volunteer may differ from other visitors, so extending the result to everyone takes additional evidence.

Study	What chance decides	What the comparison addresses
Shift roster	Which 30 shifts enter the sample	Waiting times among the shifts on the roster
Sign trial	Which sign each volunteer uses	How the signs affect the timed task for these volunteers

1

UNIT 1 / AFTER DAY 4 / FOR DAY 5 ● PR05 ● KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
> Selection concerns who is studied; assignment concerns which condition each participant receives.

1. For the help-desk study, name the units that were randomly selected and the full group their results can describe. Could the study show that a new sign caused shorter waits? Explain.

2. Suppose the volunteers choose a sign instead of drawing a card. Give one way their choice could produce a time difference even if the signs work equally well.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Study	How cases entered	How groups were assigned
Shift roster		
Sign trial		

After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §1.2, pp. 18–25; §1.4, pp. 34–37. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 1 / DAY 05 ● CLASSWORK ● T1

Three studies of morning travel

Decide whom a study can describe and whether a difference can reasonably be attributed to a treatment.

Compare three studies of morning travel. In Study A, 42 of 60 volunteers answering a campus poll prefer an earlier shuttle. In B, staff randomly select 80 students from the full roster. All answer, and 48 prefer an earlier shuttle. In C, staff randomly assign 40 volunteer riders to receive one of two reminders and record whether each rider arrives on time.

Study	How did people enter the study?	Was a treatment assigned? How?
A		
B		
C		

1. For A and B, calculate the sample preference percentage. State which larger group, if any, each sample can reasonably represent, and explain how selection supports or limits that use.

2. In C, distinguish the group who entered the study from the process that assigned reminders. What kind of comparison does the design support?

3. Suppose riders chose which reminder to receive and those choosing the earlier reminder arrived sooner. Name a factor that could affect both their choice of reminder and their arrival time. Explain how it could affect each.

1

Name _____ Date _____

UNIT 1 / DAY 05 • INDEPENDENT WORK

Does a packing card change speed?

Illustrative observation: staff who elect to use a checklist finish in 8 minutes on average, compared with 11 minutes among staff who decline. A new trial randomly assigns volunteer staff to use the checklist or the usual process.

1. Write a defensible claim about the original observation and explain how one specific confounder could produce the difference.

2. What new claim could the trial address? Explain what you still need before reporting that the checklist improves speed for all staff.

Revise: Revisit one claim about the checklist. Identify the comparison behind it and make its wording fit how staff entered the groups.

Exit: A random sample of drivers reports coffee use and commute time. Does this establish that coffee changes commute time? Name a specific reason.

Name _____ Date _____

UNIT 1 / DAY 5 HOMEWORK • HW05 • PRACTICE

Repair two overclaims

After Day 5: Homework and reading to prepare for Day 6. Turn in this two-sided HW05 sheet when requested. Keep PR06 on W035 in your binder.

An open poll finds that 90 of 120 respondents support a longer library day. In a separate study, 30 volunteers are randomly assigned to one of two sorting instructions. The group using the revised instructions takes two fewer minutes on average.

1. Repair "75% of all patrons support the change." Explain the selection concern and propose a stronger sample design.

1

Name _____ Date _____

UNIT 1 / DAY 5 HOMEWORK • HW05 • PRACTICE CONTINUED

Repair two overclaims

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR06 on W035 and answer its two self-checks on W036 before Day 6. Keep that sheet in your binder.

2. Repair "The new instructions save every worker two minutes." State what random assignment contributes and what the observed mean difference establishes.

3. Design a study to estimate the effect of the revised sorting instructions for a workforce. How would you choose workers for the study? How would you assign them to instructions?

UNIT 1 / AFTER DAY 5 / FOR DAY 6 • PR06 • PREPARATION READING

A comparison that can stand on its own

Complete after Day 5 to prepare for Day 6.

Combine center, spread, context, and limits in a short statistical conclusion.

A pickup log records how long people waited on 12 morning pickups and 12 afternoon pickups. Someone reading the log announces, "Mornings are better." A reader who did not see the log would have to guess what "better" means. Are morning waits typically shorter? Are they more predictable? Those are two different questions.

For a typical wait, compare medians. The morning median is 8 minutes and the afternoon median is 10. For consistency, compare interquartile ranges (IQRs). The IQR is the third quartile minus the first quartile. It measures the width of the middle half of a group. Here the morning IQR is 2 minutes and the afternoon IQR is 5. The middle half of the morning waits occupies the narrower interval.

Here is a comparison someone could read on its own: "Among the 24 recorded pickups, the median morning wait was 8 minutes, two minutes shorter than the afternoon median of 10 minutes. Morning waits were also more consistent through the middle half, with an IQR of 2 minutes compared with 5 minutes for afternoon waits." Both sentences put the groups, the measurement, and the evidence in the same place.

Does the log show that scheduling the same pickup in the morning would shorten the wait? Traffic and staffing may differ by time of day. The log can describe the two groups it recorded, but it does not isolate the effect of the schedule. Nor does the smaller morning IQR tell us about the single longest wait. To answer that question, ask for the maximum in each group.

Pickup group	Median duration	IQR
Morning	8 min	2 min
Afternoon	10 min	5 min

1

UNIT 1 / AFTER DAY 5 / FOR DAY 6 • PR06 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
> Name both groups, cite the relevant summaries, and keep the conclusion within the study's scope.

1. Use the pickup data to answer someone who asks, "Was a typical wait shorter in the morning, and by how much?" Write a sentence that makes sense without the question.

2. A second log has the same medians, but its morning IQR is 6 minutes and its afternoon IQR is 3 minutes. Compare the typical waits and the consistency of the waits in this second log.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Pickup group	Median in both logs (min)	First log IQR (min)	Second log IQR (min)
Morning			
Afternoon			

After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §1.4, pp. 34–37; §§2.3–2.5, pp. 86–100. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 1 / DAY 06 • CLASSWORK • T1–T3

After the checkpoint: diagnose the first error

Independently use distribution summaries, relative standing, and study design to write and revise a supported claim.

TODAY'S INDIVIDUAL TASK

Complete the assessment your teacher gives you first. The sample begins on A003 in Sample Assessments. After you turn it in, use this page for feedback.

Use the separate Unit 1 checkpoint first. After feedback, consider a volunteer survey of 16 trail users: 12 report weekend use. A draft says, "Three-fourths of local residents visit on weekends."

1. Identify the first unsupported step and rewrite the claim with the correct group, outcome, and denominator.

2. Choose a sentence from your checkpoint. Revise it, or identify a detail that already makes its meaning clear. If you do not have the checkpoint, explain why the survey above tells us what volunteers reported but cannot confirm their actual visits.

1

Name _____ Date _____

UNIT 1 / DAY 06 • INDEPENDENT WORK

After feedback: fresh transfer

For eight observed bins packed by each crew, Cedar's median packing time was 9 minutes (IQR 2) and Pine's was 11 (IQR 4). In another reference set for Pine, mean time was 12 and SD was 2.5 minutes.

1. Write a standalone comparison of typical time and spread. State a claim about the crews that would require more information.

2. A Pine bin took 7 minutes. Calculate and interpret its standardized score using the supplied reference set. Why not use an IQR here?

Revise: Check one comparison above. Add the group and units to the evidence you used, then state what the evidence supports.

Exit: A random sample of 40 parcels from a large depot log includes 6 damaged parcels. Write a sample result and its defensible population use.

Name _____ Date _____

UNIT 1 / DAY 6 HOMEWORK • HW06 • PRACTICE

Revise and transfer

After Day 6: Homework and reading to prepare for Day 7. Turn in this two-sided HW06 sheet when requested. Keep PR07 on W041 in your binder.

Use your marked checkpoint and these fresh illustrative bicycle-service records. A downtown repair took 26 minutes in a shop group with mean 20 and SD 3. A campus repair took 30 minutes in a group with mean 25 and SD 5.

1. Choose a claim from your checkpoint and clarify its evidence or scope. If your paper is unavailable, revise "The survey shows three quarters of local residents visit on weekends" using the volunteer survey in Day 6 classwork: 12 of 16 volunteer trail users reported weekend use.

1

Name _____ Date _____

UNIT 1 / DAY 6 HOMEWORK • HW06 • PRACTICE CONTINUED

Revise and transfer

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR07 on W041 and answer its two self-checks on W042 before Day 7. Keep that sheet in your binder.

2. Compute both z-scores. State which repair took longer on the clock and which stood farther above its own group mean. Show how you decided.

3. Design a graphical comparison of repair times by shop. Identify the variables to record and explain why two standardized scores cannot supply the full display.

UNIT 2 / AFTER DAY 6 / FOR DAY 7 • PR07 • PREPARATION READING

Two measurements, one case

Complete after Day 6 to prepare for Day 7.

Read paired values on a scatterplot and distinguish an association from a causal explanation.

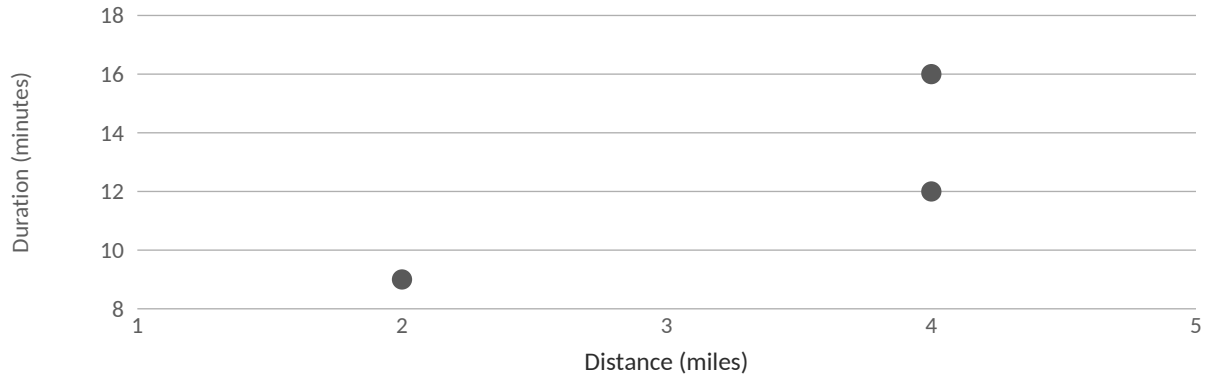
A shuttle dispatcher records two measurements for each trip: its distance in miles and its duration in minutes. A scatterplot keeps the two numbers from a single trip together. On the graph below, point to 2 on the distance axis, then move up to the plotted point. Read across to the time axis. The point (2, 9) represents one 2-mile trip that took 9 minutes.

Two more trips appear at (4, 12) and (4, 16). Both covered four miles, so the points sit one above the other. Their vertical separation is 16 − 12 = 4 minutes. Try reading that distance as four miles and notice the problem: to measure the gap you moved vertically, along the time scale. Sorting all the distances and all the times separately before plotting could pair a distance with the wrong trip.

When there are many trips, first describe the overall direction. A positive association means longer trips tend to take more time. Then inspect the form: does the pattern follow a roughly straight line, bend, or show little pattern? Consider how tightly the points follow that pattern and whether any sit unusually far away. With these three trips we can describe the records we see, but more trips would help judge how consistent the pattern is.

Next ask what could explain the differences. A longer route might have more stops. Stops take time, so the observed increases could reflect both distance and stops. The vertical gap between the two four-mile trips gives the dispatcher a reason to investigate. Recording the number of stops for each future trip would make it possible to compare trips with similar stop counts.

Three paired trip records



Each point represents one trip. The two four-mile trips lasted different amounts of time.

UNIT 2 / AFTER DAY 6 / FOR DAY 7 • PR07 • KEEP YOUR RESPONSES

2

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

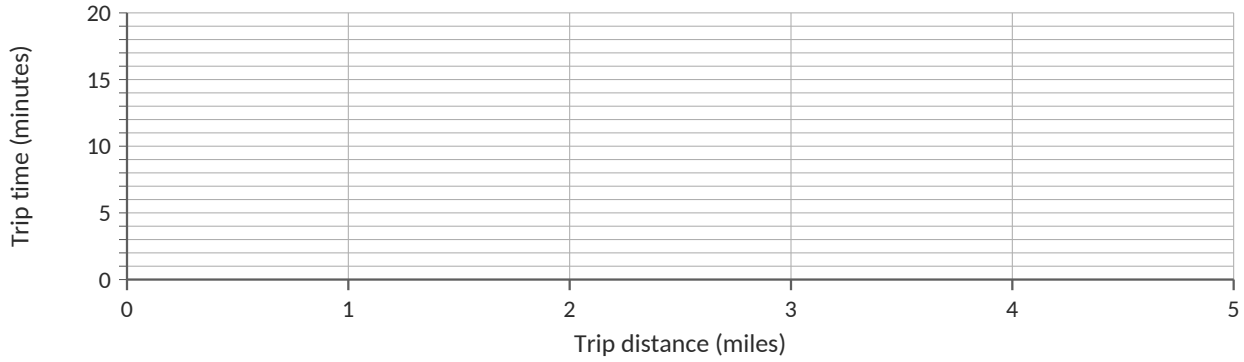
> **KEY IDEA**
>
> Each scatterplot point pairs two measurements from the same case.

1. A fourth point is added at (4, 14). Describe the trip in a sentence. Where would its point fall relative to the two existing four-mile trips?

2. A dispatcher says that doubling the route distance always doubles travel time. Use two trips on the graph to check that claim. Name one other measurement that could help explain the result.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



After you try both questions, check your reasoning on W155. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §12.2, pp. 620–623; §1.4, pp. 34–37. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 2 / DAY 07 • CLASSWORK • T4

How long does it take to sort a batch?

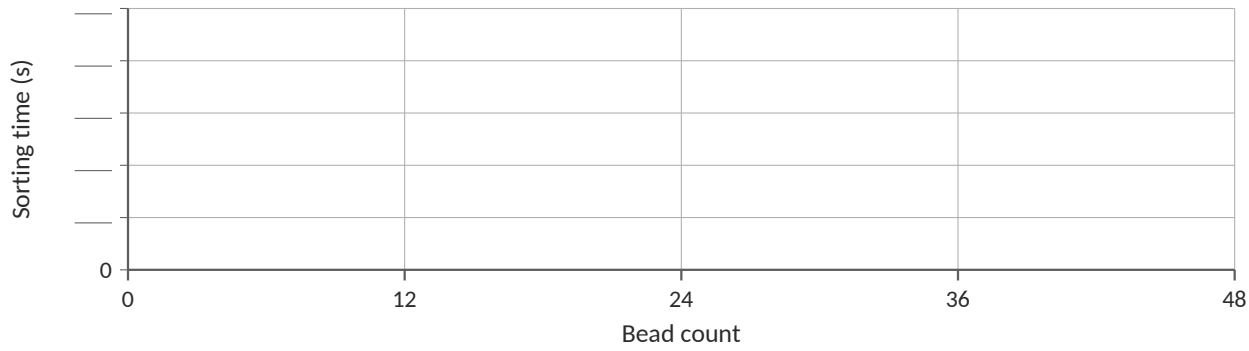
Describe a relationship between two measurements and name one reason it might arise.

Work with a partner. Prepare batches of 12, 24, and 36 beads, half of each color, and mix each batch. Shuffle three cards labeled 12, 24, and 36 to choose the testing order. One teammate sorts all three batches; the other times each trial from "go" until every bead is in its color pile. Record seconds to the nearest whole second. Copy one other team's three times into the last column. Save this table for Day 8. If your teacher supplies data instead, label your table "supplied data."

Beads in batch	Our team: time (s)	Other team: time (s)
12		
24		
36		

Plot key Our team's symbol: _____ Other team's symbol: _____

Label equal steps on the time axis to fit the class data.



1. Choose and label a seconds scale that includes all six times. Plot one point per team–batch trial, using different symbols for the two teams and a key. If points coincide, label that position with both teams. Label one point with its team, bead count, and time. For the shared class plot, contribute only your own team's three trials.

2. Do sorting times generally rise with batch size? Support your answer with two observations. At one batch size, how far apart were the two teams' times?

3. Which feature of the procedure reduces one source of variation? Which feature still limits the scope of the conclusion?

Name _____ Date _____

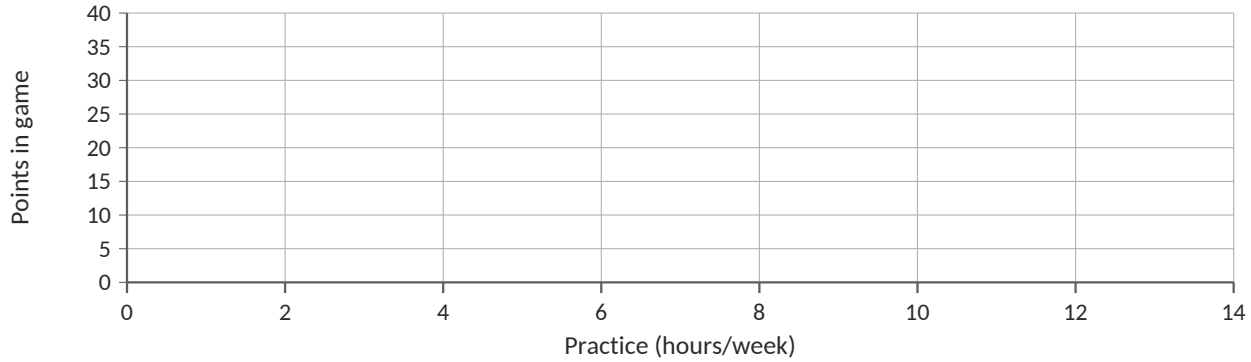
UNIT 2 / DAY 07 • INDEPENDENT WORK

2

Basketball practice and scoring

Adapted from OpenStax Try It 12.5. In this textbook example, a high-school basketball player records weekly jump-shot practice and points scored in a game: (5, 15), (7, 22), (9, 28), (10, 31), (11, 33), (12, 36). Each pair gives practice hours first and game points second. These are supplied textbook data, not St. Mary's team records.

Data adapted from OpenStax, Introductory Statistics 2e, §12.2, Try It 12.5, Table 12.2, pp. 621–622. Access for free at openstax.org.



1. Plot all six pairs with practice hours on the horizontal axis. Describe the direction and form. Support your description with two observations.

2. A teammate says these records prove that adding practice hours will increase this player's next-game score. Explain why that claim goes beyond the data and name one other influence worth recording.

Revise: Replace any claim that practice "proved" or "guaranteed" a higher score with a statement supported by these six records.

Exit: Two observed four-mile shuttle trips took 12 and 17 minutes. In one sentence, describe what this tells a dispatcher about using distance alone.

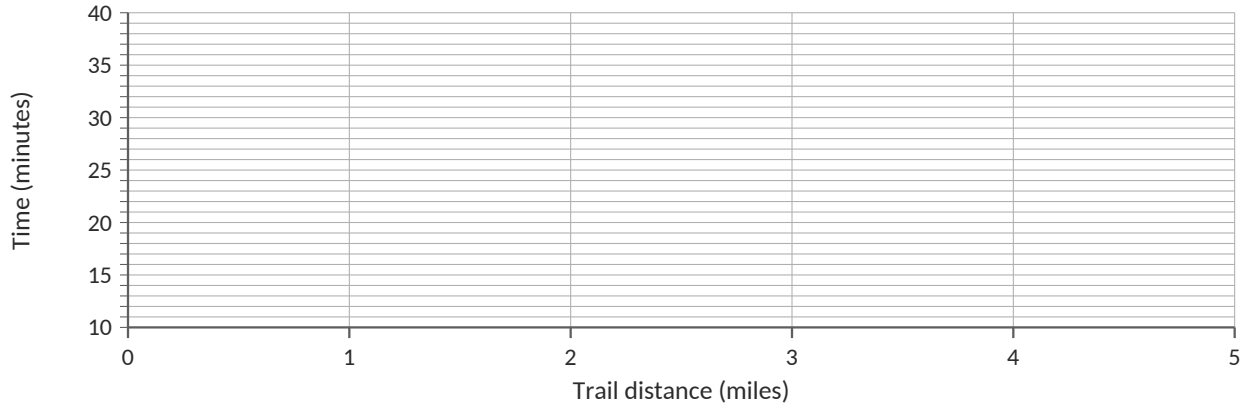
Name _____ Date _____

UNIT 2 / DAY 7 HOMEWORK • HW07 • PRACTICE

Trail counter reports

After Day 7: Homework and reading to prepare for Day 8. Turn in this two-sided HW07 sheet when requested. Keep PR08 on W047 in your binder.

The illustrative (trail miles covered, minutes spent checking counters) pairs are (1, 12), (2, 17), (2, 23), (3, 24), (4, 31), (4, 37). The crews chose existing routes.



1. Make a labeled scatterplot. Circle one point and translate it into a sentence that identifies the case and both measurements.

Name _____ Date _____

UNIT 2 / DAY 7 HOMEWORK • HW07 • PRACTICE CONTINUED

2

Trail counter reports

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR08 on W047 and answer its two self-checks on W048 before Day 8. Keep that sheet in your binder.

2. Describe the overall pattern and the variation at a fixed distance. Write for a reader who has never seen the graph.

3. A supervisor says walking one extra mile necessarily adds six minutes. Explain why the evidence is insufficient and propose one useful additional measurement.

UNIT 2 / AFTER DAY 7 / FOR DAY 8 • PR08 • PREPARATION READING

A prediction and its residual

Complete after Day 7 to prepare for Day 8.

Use a supplied linear model and interpret observed minus predicted in context.

A print-shop worker is waiting for a batch of labels. For jobs between three and seven labels, a supplied model predicts the printing time in seconds from the number of labels, k. Its rule is predicted seconds = 4 + 2k. The slope, 2 seconds per label, means that adding a label adds two seconds to the prediction. The constant 4 is the value the rule gives at zero labels, outside the range of jobs described. It would take evidence from the printer to interpret it as a measured startup time.

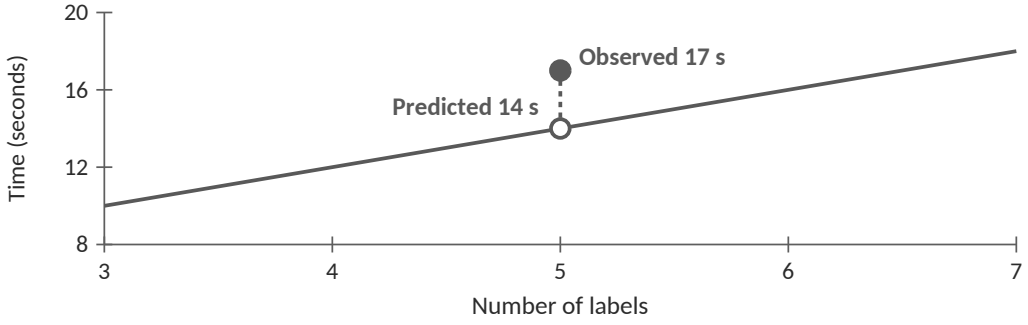
To predict the time for five labels, use k = 5. The rule gives 4 + 2 × 5 = 14 seconds. On the graph below, find 5 on the horizontal axis and move up until you reach the line. The open circle at 14 seconds marks this prediction. A five-label job was actually timed at 17 seconds. The filled point for that observation belongs directly above the open circle because both describe five labels.

The difference between an observed outcome and the prediction for that same input is a residual. Subtract in that order: 17 − 14 = +3 seconds. The positive sign tells the worker the job took three seconds longer than predicted. If another five-label job finished in 12 seconds, its residual would be 12 − 14 = −2 seconds. Its point would sit below the line.

The slope and the residual answer different questions. The slope describes how the prediction changes when the input increases by a label. A residual describes the model's error on one particular job. Its units are seconds. To decide if the model is useful for planning, the worker would compare the sizes and directions of residuals across many jobs.

Labels	Predicted time	Observed time	Observed − predicted
5	14 s	17 s	+3 s

One five-label print job



UNIT 2 / AFTER DAY 7 / FOR DAY 8 • PR08 • KEEP YOUR RESPONSES

2

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> A residual compares an observation with the model's prediction for that same input.

1. The worker queues six labels. Predict the printing time. How much longer is that prediction than the one for five labels?

2. The six-label job in question 1 actually finishes in 13 seconds. Find its signed residual and write a sentence explaining what the sign tells the worker.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Labels	Predicted seconds	Observed seconds	Observed − predicted (s)
5		17	
6		13	

After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §12.1, pp. 618–620; §12.3, pp. 623–626. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 2 / DAY 08 ● CLASSWORK ● T4

Check yesterday's bead times

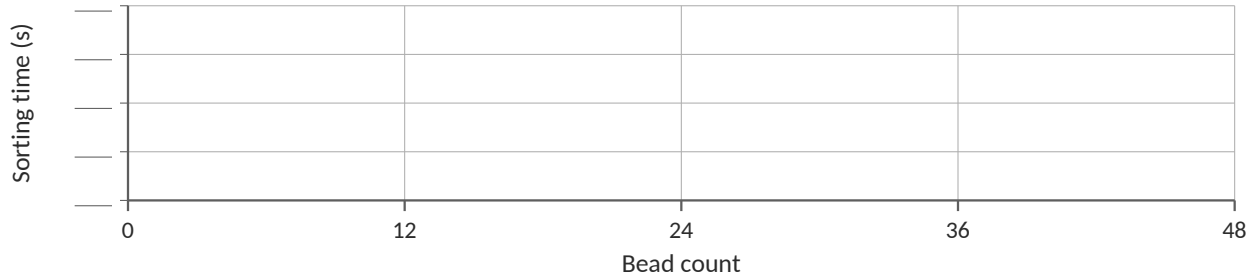
Use a given line for an in-range prediction and interpret the sign and units of a residual.

Use one batch from your Day 7 bead-sorting table. If that table is unavailable, use the supplied record: Team X sorted 24 beads in 17 seconds. For this exercise, use predicted sorting seconds = 3 + 0.55b, where b is bead count and the intended range is 12–36 beads. This is a given comparison model, not a line fitted to your class's results. Fill the row below with your chosen batch.

Beads	Observed time (s)	Predicted time (s)	Residual (s)

Plot key Observed time: _____ Predicted time: _____

Choose a time scale that distinguishes the actual and predicted times.



1. Predict your batch's time and complete the row. Choose and label a seconds scale that contains both times. At the batch's bead count, mark the observed and predicted times with different symbols and a key. Calculate the signed residual in seconds.

2. Interpret the slope, then predict the change in sorting time when the batch increases from 12 to 24 beads.

3. A student reports "the error is 0.55." Explain what additional information would be needed to find an individual error.

Name _____ Date _____

UNIT 2 / DAY 08 • INDEPENDENT WORK

2

Printing team programs

A print shop uses predicted seconds = 7 + 1.5p for jobs with 2–18 pages, where p is page count. A 10-page job actually takes 19 seconds.

1. Predict the time and calculate the signed residual. Describe the model's error for this job.

2. How much longer does the model predict for 14 pages than for 10 pages? Interpret without implying every pair of jobs will differ by that amount.

Revise: Check your signed error. Write a sentence that names the job and says how many seconds the model missed, and in which direction.

Exit: A line predicts 18 minutes for a five-mile trip. The trip took 14 minutes. Give its residual and interpret it.

Name _____ Date _____

UNIT 2 / DAY 8 HOMEWORK • HW08 • PRACTICE

Checking a bike-route model

After Day 8: Homework and reading to prepare for Day 9. Turn in this two-sided HW08 sheet when requested. Keep PR09 on W053 in your binder.

A supplied line for observed routes of 2–8 miles predicts cycling minutes = 6 + 3d, where d is miles. Three illustrative records (miles, minutes) are (2, 13), (5, 19), and (7, 30).

Miles	Observed time (min)	Predicted time (min)	Residual (min)
2	13		
5	19		
7	30		

1. Complete the table. Identify which ride has the largest error in magnitude and say whether it took more or less time than predicted.

Name _____ Date _____

UNIT 2 / DAY 8 HOMEWORK • HW08 • PRACTICE CONTINUED

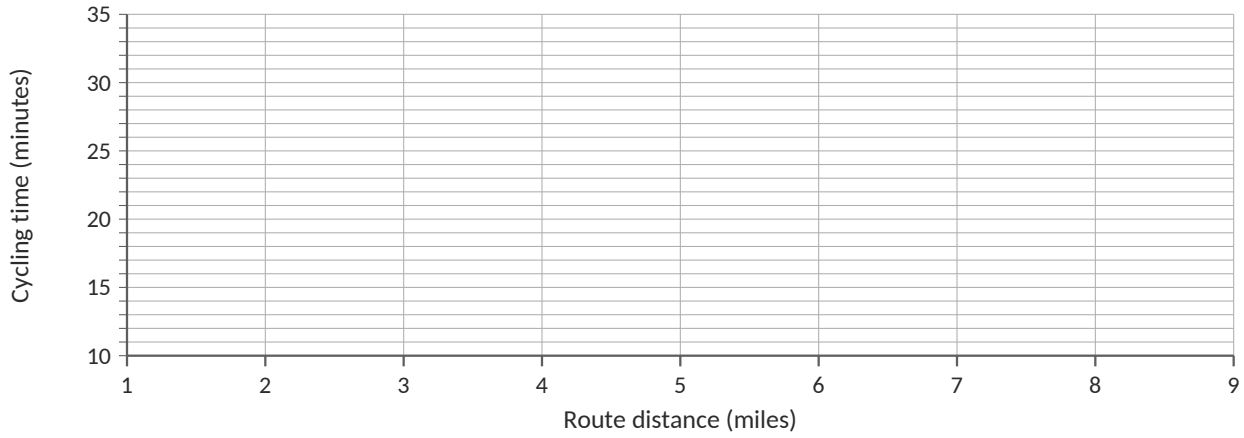
2

Checking a bike-route model

Continue the same assignment. Use both sides of this sheet for the data and your work.

KEEP THE NEXT SHEET

Read PR09 on W053 and answer its two self-checks on W054 before Day 9. Keep that sheet in your binder.



2. Use the axes on this page. Draw the model line from 2 to 8 miles and mark the three observed rides. At five miles, draw the vertical gap between the point and the line. Use the point's position to explain the sign of its residual.

3. Interpret the slope, predict the change from three to six miles, and explain why that difference would be a poor promise to an individual cyclist.

UNIT 2 / AFTER DAY 8 / FOR DAY 9 • PR09 • PREPARATION READING

What does the model leave out?

Complete after Day 8 to prepare for Day 9.

Explain model specification and test whether a practical decision changes across reasonable models.

Two four-mile trips appear in a school shuttle log. The ordinary-traffic trip took 12 minutes. The heavy-traffic trip took 17. A dispatcher wants to schedule another four-mile trip during heavy traffic. Can the dispatcher allow 15 minutes? Looking up the most recent trip gives one observation. For a prediction about a new trip, we need to decide which information the model should use.

Model specification means choosing the inputs and the mathematical rule used for prediction. A distance-only model gives every four-mile trip the same predicted time. A model that also uses traffic can give ordinary and heavy trips different predictions at the same distance. The two records illustrate why traffic might matter. We would need many more trips to judge whether including it improves predictions.

Suppose a distance-only model based on a larger log predicts 14 minutes at four miles. A second model based on that log predicts 12.5 minutes in ordinary traffic and 16.5 minutes in heavy traffic. For the proposed heavy-traffic trip, 14 minutes fits the 15-minute slot. The second prediction exceeds the limit by 1.5 minutes. Holding the trip and limit fixed lets us attribute the change in the scheduling decision to our choice of model.

Now allow 18 minutes for the same trip. Both predictions fit. The decision is robust to switching between these two models at that limit, although their predictions still differ. To choose between models for future trips, compare their prediction errors on trips that were set aside when fitting them. Extra detail is worthwhile if it leads to better predictions on those trips.

Four-mile trip	Observed time	Distance-only prediction	Prediction with traffic
Ordinary traffic	12 min	14 min	12.5 min
Heavy traffic	17 min	14 min	16.5 min

UNIT 2 / AFTER DAY 8 / FOR DAY 9 • PR09 • KEEP YOUR RESPONSES

2

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> A decision is robust to a model choice when it stays the same under the alternatives checked.

1. Why does the distance-only model give the same prediction for the two four-mile trips? What do the recorded 12 and 17 minutes tell you?

2. An ordinary four-mile trip has a 13-minute slot. Use both models' predictions to tell the dispatcher whether the decision is robust.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Model	Ordinary four-mile trip: prediction (min)	Fits the 13-minute slot?
Distance only		
With traffic		

After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. Foundation: OpenStax, Introductory Statistics 2e, §12.3 (pp. 623–630) and §1.4 (pp. 34–37). Model-choice extension and classroom data developed for this course.

Name _____ Date _____

UNIT 2 / DAY 09 • CLASSWORK • T5

Dispatch a 3.5-mile shuttle

Compare two plausible model specifications and identify a decision that depends on which one is used.

Let d be trip distance in miles. Predict travel time in minutes using all trips together (pooled): 1.91 + 3.45d; ordinary traffic: 4.33 + 2d; heavy traffic: 9.67 + 2d. The recorded distances were 2–4 miles in ordinary traffic and 3–5 in heavy traffic. The dispatcher allows the trip if the predicted time is 15 minutes or less.

Specification	Predicted time at 3.5 miles	Within 15 minutes?
Pooled		
Ordinary traffic		
Heavy traffic		

1. Calculate all three predictions at 3.5 miles and make a scheduling decision for each. Why is this distance suitable for comparing the fits?

2. A dispatcher uses the pooled line to schedule a heavy-traffic trip of 3.5 miles. By how many minutes does the heavy-traffic prediction differ? What trend do all three lines predict as distance increases?

3. Describe a small collection of new trip data that would help compare the pooled and traffic-specific models. Explain why choosing the model with the shorter prediction is not sufficient.

Name _____ Date _____

UNIT 2 / DAY 09 • INDEPENDENT WORK

2

Preparing order bags

For orders of 2–8 bags, a shop predicts preparation time in minutes with three illustrative models. Using b for bag count, the rules are pooled = 2 + 1.6b, prepacked = 1 + b, and custom = 4 + 2b. All order types were observed throughout that range. A ten-minute slot is available.

1. At four bags, compute each prediction and decide whether a slot of ten minutes or less is adequate on each model.

2. Tell the scheduler what remains true about how the predicted time changes with order size and which specific recommendation changes.

Revise: Revise your scheduling advice so a worker can tell which kind of four-bag order is predicted to exceed the ten-minute slot.

Exit: Two models for a three-mile run predict 12 minutes pooled and 17 in heavy traffic. A 15-minute slot is available. Is the scheduling decision robust? Explain.

Name _____ Date _____

UNIT 2 / DAY 9 HOMEWORK • HW09 • PRACTICE

A crew plans its unloading time

After Day 9: Homework and reading to prepare for Day 10. Turn in this two-sided HW09 sheet when requested. Keep PR10 on W059 in your binder.

For deliveries of 2–8 crates to either site type, the illustrative time predictions in minutes are pooled = 3 + 1.8c, ramp = 2 + 1.2c, and stairs = 5 + 2c. Here c is crates. The crew has a 13-minute slot.

Crate count	Pooled (min)	Ramp (min)	Stairs (min)
3			
5			

1. Make a prediction table for three and five crates using all three models. Circle every prediction that exceeds the slot.

Name _____ Date _____

UNIT 2 / DAY 9 HOMEWORK • HW09 • PRACTICE CONTINUED

2

A crew plans its unloading time

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR10 on W059 and answer its two self-checks on W060 before Day 10. Keep that sheet in your binder.

2. At three crates, do all three models give the same decision about the 13-minute slot? Repeat for five crates. State when the decision is robust to the choice of model.

3. One five-crate ramp delivery took ten minutes. Calculate its residual from both the pooled and ramp predictions and explain why one errand cannot settle model choice.

UNIT 2 / AFTER DAY 9 / FOR DAY 10 • PR10 • PREPARATION READING

An equation can go farther than the evidence

Complete after Day 9 to prepare for Day 10.

Identify extrapolation and check whether a model applies to the proposed group and conditions.

A school mailroom recorded packing times for jobs containing 4 to 10 envelopes. Its model predicts minutes as 2 + 0.5e, where e is the number of envelopes. For six envelopes, the prediction is 2 + 0.5 × 6 = 5 minutes. The request falls between the smallest and largest recorded jobs. Using the pattern within its observed input range is called interpolation. The log need not contain a job of every possible size between the endpoints.

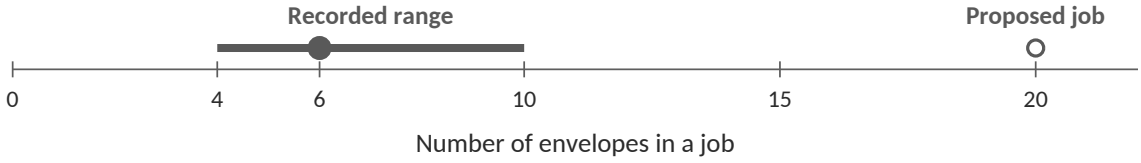
What if a teacher brings in twenty envelopes? Substitution gives 2 + 0.5 × 20 = 12 minutes. This prediction extends the pattern ten envelopes beyond the largest recorded job. That is extrapolation. Perhaps workers have to refill the label tray after ten envelopes. If so, the next ten envelopes involve an extra step that the log could never have revealed. Timing several comparable large jobs would let the mailroom check its prediction.

Sometimes the range printed at the top of a report belongs to several groups combined. Suppose hand-sealed jobs in a second log contain 4–8 envelopes and machine-sealed jobs contain 6–12. To predict a hand-sealed job, look only at the hand-sealed records. Ten envelopes is inside the combined range yet outside the observed hand-sealed range.

Before recommending a schedule, name the proposed job, find the matching group, and check its smallest and largest inputs. Then ask whether anything important about the process changed. A new sealing machine can change the time for even a six-envelope job. A matching range answers one question about the evidence. Comparable working conditions and the size of past prediction errors also matter.

Envelope count	Prediction	Distance from the measured range
6	5 minutes	Within the observed 4–10 envelopes
20	12 minutes	10 envelopes beyond the largest observed job

Jobs in the mailroom log



UNIT 2 / AFTER DAY 9 / FOR DAY 10 • PR10 • KEEP YOUR RESPONSES

2

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

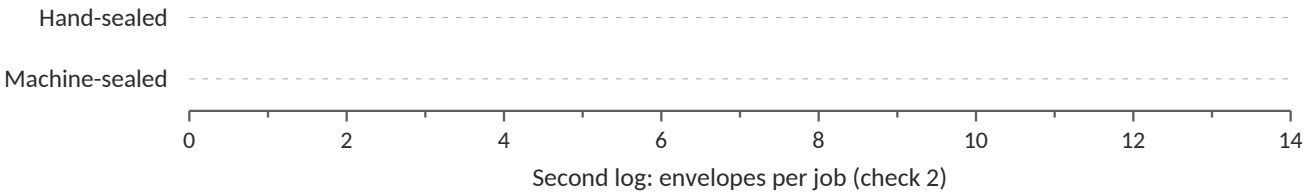
> **KEY IDEA**
>
> A calculable prediction is not necessarily supported by the observed range or conditions.

1. Use the first mailroom's model to predict the time for eight envelopes. Explain why the 12-minute prediction for twenty envelopes needs different evidence.

2. In the second log, ten-envelope jobs are proposed for both sealing methods. Which prediction requires extrapolation? Give the observed range for each method.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §12.5, pp. 635–636. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 2 / DAY 10 • CLASSWORK • T5

Select an appropriate dispatch claim

Check the observed range and conditions before defending a forecast.

Here d is distance in miles, and the lines predict trip time in minutes. In ordinary traffic, the observed range is 2–4 miles and the line is 4.33 + 2d. In heavy traffic it is 3–5 miles and the line is 9.67 + 2d. The full log covers 2–5 miles. All observations are fair-weather daytime trips.

1. Use the heavy-traffic line to predict the time for a 2.5-mile trip. Was that distance observed in heavy traffic? Explain why checking only the range for all trips could mislead the dispatcher.

2. Compare two four-mile ordinary-traffic trips, both in daylight. One takes place in fair weather, the other during an ice storm. Which features of each forecast match the recorded conditions?

3. Write a response to a dispatcher requesting a nine-mile heavy-traffic prediction. Include one number, the limitation, and a targeted data request.

Name _____ Date _____

UNIT 2 / DAY 10 • INDEPENDENT WORK

2

A printer's production estimate

In this supplied classroom case, a printer recorded single-sided jobs of 4–16 pages and double-sided jobs of 6–12 pages. A double-sided model predicts seconds = 8 + 1.5p, where p counts printed pages, not sheets. All runs used ordinary paper.

1. Evaluate using the double-sided fit for a 14-page double-sided job. Report the prediction and the range comparison.

2. Evaluate the same fit for a ten-page double-sided job on thick card. What data would help you check that estimate?

Revise: Revise a forecast above so a reader can distinguish the number the equation gives from the cases actually used to build it.

Exit: Trips observed in light rain range from 1 to 4 miles. Dry trips range from 2 to 8. Is a six-mile light-rain estimate supported by the rain data? Explain.

Name _____ Date _____

UNIT 2 / DAY 10 HOMEWORK • HW10 • PRACTICE

Scheduling field visits

After Day 10: Homework and reading to prepare for Day 11. Turn in this two-sided HW10 sheet when requested. Keep PR11 on W065 in your binder.

A technician's line for paved routes is predicted minutes = 5 + 2d, observed at 2–6 miles. For gravel routes the line is 7 + 3d, observed at 4–8 miles. Data came from clear summer mornings.

Route request	Prediction (min)	Inside its road type's observed range?
5 miles, paved		
3 miles, gravel		
9 miles, paved		

1. Complete the table for the three requests. Use the equation and the observed distance range for each request's road surface.

Name _____ Date _____

UNIT 2 / DAY 10 HOMEWORK • HW10 • PRACTICE CONTINUED

2

Scheduling field visits

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR11 on W065 and answer its two self-checks on W066 before Day 11. Keep that sheet in your binder.

2. A three-mile gravel forecast is inside the combined 2–8-mile range. Draft a two-sentence correction for someone planning the technician's day.

3. A colleague wants to promise a five-mile paved visit in snow takes 15 minutes. Identify the unchecked condition and design a focused collection to evaluate the promise.

UNIT 2 / AFTER DAY 10 / FOR DAY 11 • PR11 • PREPARATION READING

From a prediction to a recommendation

Complete after Day 10 to prepare for Day 11.

Write a decision memo that separates a model prediction from a guarantee for one outcome.

A maintenance crew needs to know whether a four-mile trip can fit a 14-minute slot. The trip model predicts 13 minutes, based on trips of 2–5 miles under comparable conditions. A memo saying "It works" leaves the crew guessing. What trip? What limit? How close is the prediction to that limit? A reader who never saw the question should be able to recover all three from the answer.

Start with the comparison. Four miles is inside the recorded range. The prediction leaves 14 − 13 = 1 minute before the deadline. That supports penciling the trip into the schedule. To decide how much to rely on that schedule, examine how closely predictions have matched actual trips.

For example, suppose another comparable trip was predicted to take 12 minutes and actually took 15. Its residual is observed minus predicted, or 15 − 12 = +3 minutes. A positive residual matters to a dispatcher because the trip took longer than planned. If the proposed four-mile trip ran 3 minutes over its prediction, it would take 16 minutes and miss the slot by 2. One previous error shows how a miss could happen. It does not tell us how often one will happen.

A useful memo ties the action to the stakes. A crew with a flexible arrival time might keep the slot. A crew trying to reach a gate before it locks might ask for a longer slot and check how often comparable trips ran late. Give the prediction, identify its support in the log, and explain the uncertainty that matters to the person making the decision.

Memo detail	What the reader can establish
Four-mile trip	Within the measured 2–5-mile range
Prediction 13 min; slot 14 min	One minute between the prediction and the limit

UNIT 2 / AFTER DAY 10 / FOR DAY 11 • PR11 • KEEP YOUR RESPONSES

2

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> A recommendation needs the decision rule, evidence, and a limitation that could change the action.

1. Write a two-sentence recommendation to the crew if the gate locks after 14 minutes. Include the distance, observed range, prediction, and evidence about a possible delay.

2. The same crew next asks about a seven-mile trip. What additional observations would make a recommendation more defensible? Explain why.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Decision	Evidence	Limit of the evidence

After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §12.3, pp. 623–626; §12.5, pp. 635–636. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 2 / DAY 11 • CLASSWORK • T4–T5

Revise after the individual task

Defend an in-range model decision with a prediction, a checked error, and a specific limit.

> **TODAY'S INDIVIDUAL TASK**
>
> Complete the assessment your teacher gives you first. The sample begins on A005 in Sample Assessments. After you turn it in, use this page for feedback.

After you turn in your model memo, review a different case. A route model predicts 14 minutes for a three-mile trip. A three-mile run took 12 minutes. The recorded trips ranged from 2 to 5 miles.

1. A draft says, "It is off by two, so the model is bad." Identify two pieces of information a new reader needs and revise the claim.

2. If a rider asks about an eight-mile trip, write one limit and request one specific observation that would reduce uncertainty.

Name _____ Date _____

UNIT 2 / DAY 11 • INDEPENDENT WORK

2

A new print-room check

A printer model for 3–15 single-sided pages gives predicted seconds = 6 + 1.2p. An observed ten-page job took 20 seconds. A coordinator proposes an 18-second maximum.

1. Give the prediction and residual for the observed job. Explain which direction the line missed.

2. The coordinator wants to promise that every ten-page job will finish within 18 seconds. Write two sentences using both the model prediction and the observed job to evaluate that promise.

Revise: Check your scheduling recommendation. Make clear what 18 seconds measures and what happened in the observed ten-page job.

Exit: One job had a residual of −2 minutes from a prediction of 11 minutes. State the observed time and interpret the sign.

Name _____ Date _____

UNIT 2 / DAY 11 HOMEWORK • HW11 • PRACTICE

Revise and transfer

After Day 11: Homework and reading to prepare for Day 12. Turn in this two-sided HW11 sheet when requested. Keep PR12 on W071 in your binder.

Use feedback on today's memo. A new illustrative model for 4–20 jars gives predicted labeling seconds = 5 + 0.6j, with j in jars. One observed 15-jar run took 12 seconds.

1. Revise one sentence from your model memo and explain what you clarified. If the memo is unavailable, revise this claim: "It is off by two, so the model is bad." A three-mile trip took 12 minutes; the model predicted 14 minutes.

Name _____ Date _____

UNIT 2 / DAY 11 HOMEWORK • HW11 • PRACTICE CONTINUED

2

Revise and transfer

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR12 on W071 and answer its two self-checks on W072 before Day 12. Keep that sheet in your binder.

2. Compute and interpret the new residual. Would a 14-second maximum for all 15-jar runs follow from this example? Explain.

3. For a 15-jar job, a second line predicts 17 seconds when labels are applied by hand. Which prediction fits the 14-second slot? A time of exactly 14 seconds fits. What would you measure to decide which line to use?

UNIT 3 / AFTER DAY 11 / FOR DAY 12 • PR12 • PREPARATION READING

From one sample to a distribution of estimates

Complete after Day 11 to prepare for Day 12.

Distinguish an individual outcome, a sample proportion, and the population proportion.

A snack bin holds 8 pretzel packets and 12 cracker packets. The question is what proportion of all the packets are pretzels. We can count the whole bin, so we know the population proportion is 8/20 = 0.40. We will keep that value fixed and see what a student would learn from sampling just five times.

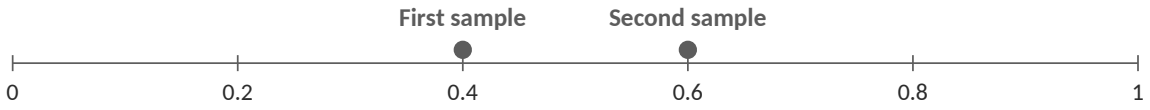
A student draws a packet at random, records its type, returns it, and mixes before the next draw. With a fair draw each time, returning the packet restores the original 8 pretzels among 20 packets. A pretzel can even be drawn twice. Suppose the five outcomes are pretzel, cracker, cracker, pretzel, cracker. The sample proportion is the number of pretzel draws divided by all draws in that sample, or 2/5 = 0.40. It happens to equal the population proportion.

The second team gets three pretzel draws out of five, giving 3/5 = 0.60. Place one dot at 0.40 for the first team's complete sample and one at 0.60 for the second team's complete sample. A third team that also gets 0.60 would stack its own dot at 0.60. Counting the dots tells us how many samples the teams took. Reading a dot's position tells us the proportion of pretzel draws in one sample.

Each team can get a different mix even though it used the same bin. This is sampling variability. The theoretical distribution of the sample proportion over all such random samples of five is called its sampling distribution. The teams' plot, with enough independent repetitions, lets us approximate that distribution. In a real survey we usually have just one estimate and an unknown population share.

What is counted?	Pretzels	Total count	Share
Whole bin	8	20	0.40
First sample	2	5	0.40
Second sample	3	5	0.60

Two samples from the same bin



Pretzel share in one sample of five draws

UNIT 3 / AFTER DAY 11 / FOR DAY 12 • PR12 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

3

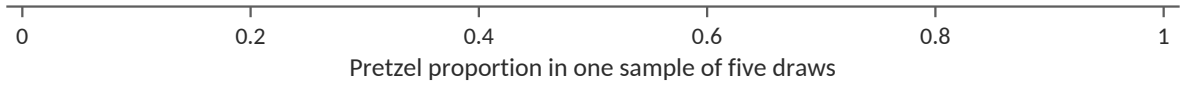
> **KEY IDEA**
>
> One dot in a sampling display represents one complete sample's estimate.

1. A dot is at 0.60 on the five-draw plot. How many draws gave a pretzel? Can you tell whether three different pretzel packets were drawn? Explain.

2. Suppose the third team gets three pretzels in five draws. Describe the dotplot for all three teams, and state what stays at 0.40 in the bin.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



Pretzel proportion in one sample of five draws

After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. OpenStax, Introductory Statistics 2e, §§1.1–1.2 (pp. 6–9, 19–20) and §8.3 (pp. 420–421). The snack-bin activity uses constructed data.

Name _____ Date _____

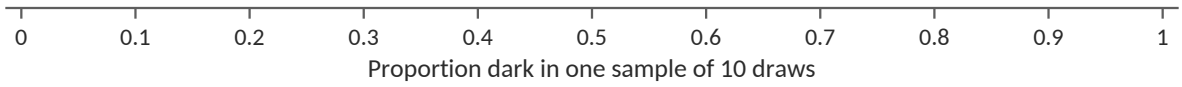
UNIT 3 / DAY 12 • CLASSWORK • T6

Build a class sampling plot

Turn individual yes/no outcomes into a sample proportion and explain what one dot in a plot of samples represents.

Work in pairs with 12 dark and 8 light beads of equal size. Predict the dark share before drawing. Then each partner takes a sample of ten draws. For every draw, mix the bag, draw without looking, record the color, and return the bead before the next draw. Keep the two samples separate. If you cannot draw beads, your teacher will supply counts to use instead.

Sample	Dark draws	Light draws	Total draws	Fraction dark
Partner 1			10	
Partner 2			10	
Whole class				



Proportion dark in one sample of 10 draws

1. Before drawing, what proportion of dark draws do you expect in the long run, given the bag's contents? Then calculate each partner's sample proportion. How do the two observed proportions compare?

2. On the 0–1 scale, add one dot for each completed ten-draw sample. Stack dots for equal proportions. What would a single class average hide?

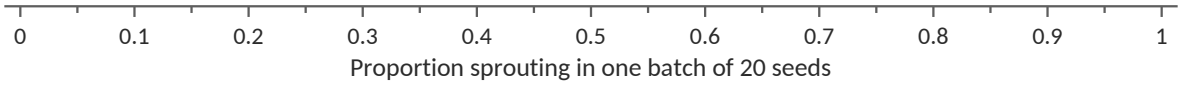
3. Add the dark counts from all completed samples and divide by the total number of draws. Compare this pooled proportion with 0.60. Which number describes the draws, and which describes the bag?

Name _____ Date _____

UNIT 3 / DAY 12 • INDEPENDENT WORK

Germination batches

A simulator assigns each seed an independent 0.70 chance to sprout. Four samples each contain 20 seeds. Their sprout counts are 11, 15, 15, and 18.



Proportion sprouting in one batch of 20 seeds

1. For each batch, divide the number of seeds that sprouted by 20. Plot the four proportions. How many batches have a proportion of 0.75?

2. Does the largest estimate show that the simulator's sprouting probability increased? Explain using the figures.

Revise: Revisit your statement about the largest estimate. Name what one dot represents and the number of seeds behind it.

Exit: In a plot of ten-draw team samples, one dot lies at 0.30. What produced that dot?

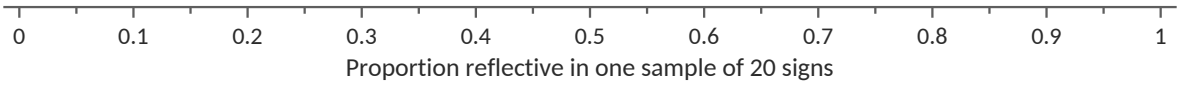
Name _____ **Date** _____

UNIT 3 / DAY 12 HOMEWORK • HW12 • PRACTICE

Repeated samples of trail signs

After Day 12: Homework and reading to prepare for Day 13. Turn in this two-sided HW12 sheet when requested. Keep PR13 on W077 in your binder.

In this model, 45% of signs have a reflective strip. Six independent samples of 20 signs produce counts of 6, 10, 8, 13, 10, and 7 reflective signs.



1. Calculate all six estimates. Draw and label their dotplot. Write a caption about the tallest stack a reader can understand without this prompt.

Name _____ Date _____

UNIT 3 / DAY 12 HOMEWORK • HW12 • PRACTICE CONTINUED

Repeated samples of trail signs

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR13 on W077 and answer its two self-checks on W078 before Day 13. Keep that sheet in your binder.

2. Pool the six samples to estimate the share of reflective signs. Separately, subtract the smallest sample proportion from the largest. Explain what the pooled estimate and the range each describe.

3. A student says a pooled 45% guarantees the next 20 signs will include nine reflective signs. Repair the claim, using two of the six samples as evidence.

UNIT 3 / AFTER DAY 12 / FOR DAY 13 • PR13 • PREPARATION READING

Why sample size changes precision

Complete after Day 12 to prepare for Day 13.

Compare sampling variability at two sample sizes without promising a better result from every larger sample.

Suppose a delivery service models each delivery as independently on time with probability 0.80. To simulate a working day, generate ten outcomes, divide the on-time count by ten, and record one dot for that sample proportion. Repeat with many new sets of ten outcomes. Now repeat the experiment with sets of forty. Both sets of dots describe the same 80% model. Only the number of deliveries per sample has changed.

The table gives illustrative summaries. For ten deliveries, the middle half of the sample proportions extends from 0.70 to 0.90. Its width is 0.90 − 0.70 = 0.20, or 20 percentage points. For forty deliveries, the middle half extends from 0.75 to 0.85, a width of 10 percentage points. A narrower middle half tells us that the central estimates cluster more tightly. It tells us nothing about the exact endpoints of all the estimates.

Why does sample size matter? Changing one delivery from late to on time changes a ten-delivery estimate by 1/10, or 10 percentage points. The same change in forty deliveries shifts its estimate by only 1/40, or 2.5 percentage points. Across independent samples, random differences in individual deliveries typically have less effect on the proportion when we average more outcomes. We call that increased precision.

A larger sample can still be farther from 80% on a particular run. Also, compare what was sampled. If the service records only dry-weather deliveries, adding thousands more dry-weather deliveries cannot tell us how often drivers are on time in storms. Precision about the sampled trips and a fair representation of all trips are separate requirements.

Sample size	Middle half of simulated on-time shares
10 deliveries	0.70 to 0.90
40 deliveries	0.75 to 0.85

UNIT 3 / AFTER DAY 12 / FOR DAY 13 • PR13 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

3

> **KEY IDEA**
>
> Larger comparable random samples usually give more concentrated estimates, not guaranteed accuracy.

1. Use the table to compare the two middle-half widths. Give their difference in percentage points and interpret it for the delivery service.

2. In one run, 8 of 10 deliveries and 28 of 40 deliveries are on time. Which estimate is closer to the 80% model rate? Does the result contradict the table?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Deliveries per sample	Lower quartile	Upper quartile	Width (percentage points)
10			
40			

After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. OpenStax, Introductory Statistics 2e, §1.2 (pp. 18–25) and §8.3 (pp. 420–422). The delivery model and summaries were developed for this course.

Name _____ Date _____

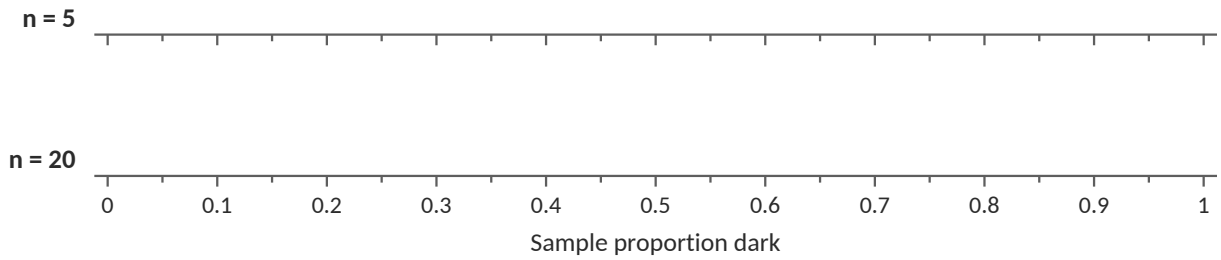
UNIT 3 / DAY 13 • CLASSWORK • T6

Compare two class sampling plots

Compare variability at two sample sizes and explain why increasing the sample size cannot correct selection bias.

Mix 12 dark and 8 light beads of the same size. Draw without looking, returning and remixing the bead after every draw. Record the number of dark beads in the first 5 draws and in all 20 draws. Each team posts its two counts. Use the class results or your teacher's supplied team counts.

Our sample	Dark draws	Total draws	Dark share
Draws 1–5		5	
Draws 1–20		20	



1. Divide your two counts by 5 and 20. How far is each dark share from 60%, in percentage points? Copy all teams' shares onto the two class plots. Stack one dot per team on each plot.

2. Find the range of the team proportions on each plot. Which varied more in this class, or were the ranges equal? Which sample size would you expect to give less variation over many repetitions?

3. Another class draws only from the top layer of an unmixed bag. Why might 100 draws from that layer still estimate the whole bag poorly?

Name _____ Date _____

UNIT 3 / DAY 13 • INDEPENDENT WORK

Who wants more bike racks?

A community college parking office surveys 2,000 students who own parking passes. A second team randomly samples 120 students from a much larger complete enrollment roster; all answer. Both ask whether students want more bike racks.

1. Which survey better supports a claim about all enrolled students? Identify a possible reason the other result could differ.

2. If the second team increases its random sample from 120 to 480, predict what happens to its typical sampling variability. What will still vary?

Revise: Underline the group each survey could reach. Check that your conclusion refers to that group.

Exit: Each simulated trial has a 30% chance of success and is independent of the others. One group repeatedly takes samples of 20 and another takes samples of 80. Which group's sample percentages should vary less? Around what value should each group's percentages cluster?

Name _____ **Date** _____

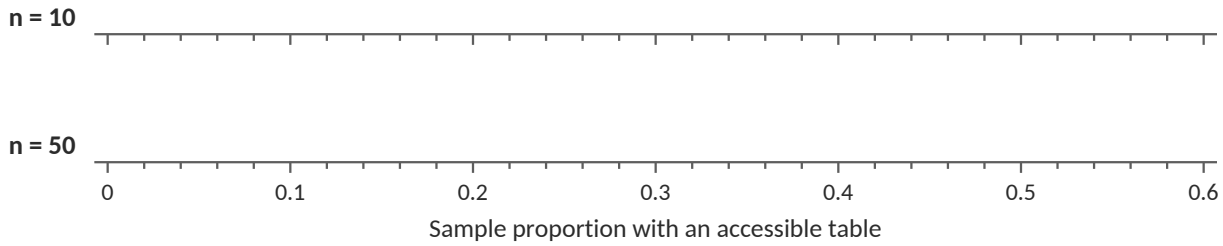
UNIT 3 / DAY 13 HOMEWORK • HW13 • PRACTICE

Trailhead features

After Day 13: Homework and reading to prepare for Day 14. Turn in this two-sided HW13 sheet when requested. Keep PR14 on W083 in your binder.

In a simulation, each trailhead independently has a 30% chance of having an accessible picnic table. Each row shows six simulated samples of the same size. Each entry counts the trailheads with an accessible table in one sample.

Trailheads per sample	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
10	1	2	3	4	4	5
50	11	13	15	16	17	19



1. Divide each count by the sample size for its row. Plot one dot per sample on the appropriate graph, using the same horizontal scale for both. Compare the two ranges in percentage points.

Name _____ Date _____

UNIT 3 / DAY 13 HOMEWORK • HW13 • PRACTICE CONTINUED

Trailhead features

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR14 on W083 and answer its two self-checks on W084 before Day 14. Keep that sheet in your binder.

2. A student says every larger sample must be closer to 30%. Find one sample of 10 that is closer to 30% than a sample of 50. Use those two results to correct the claim.

3. A volunteer counts 800 trailheads from a list of popular accessible locations. Explain one way this list could distort the estimate. Describe a better selection rule.

UNIT 3 / AFTER DAY 13 / FOR DAY 14 • PR14 • PREPARATION READING

Confidence belongs to the procedure

Complete after Day 13 to prepare for Day 14.

Explain what changes across confidence intervals and what a confidence level means.

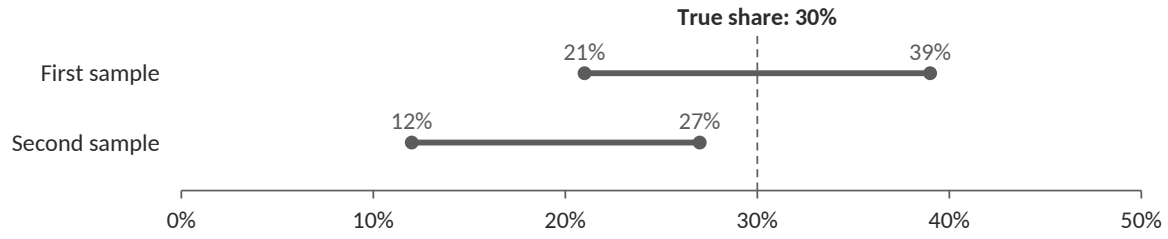
If we knew that exactly 30% of the beads in a bin were yellow, we could test how well a method estimates that share. We can do this in a simulation because we choose the true proportion ourselves. Take a random sample, calculate its yellow percentage, and use a supplied procedure to give an interval of plausible values for the bin's yellow share. Repeat the entire procedure with a fresh random sample of the same size.

The first interval runs from 21% to 39%. On the display, follow that segment until it crosses the dashed line at 30%. The true share lies between its endpoints, so the interval captures the true share. The second interval runs from 12% to 27%. Its upper end falls three percentage points short of the line. It misses. Both samples came from the same bin. A different group of draws changed the estimate and the endpoints.

A 95% confidence procedure aims to capture the true population proportion in about 95 out of every 100 intervals over many comparable repetitions, provided its assumptions hold. The first two samples here captured it only once. Two intervals are far too few to judge the long-run performance of that procedure. Even in a larger batch, the observed percentage can differ from 95%.

In a field survey, we can calculate the endpoints from our sample, but the true population share is usually unknown. Imagine covering the dashed line with a slip of paper. The two intervals remain visible, yet we can no longer check which ones capture the true share. Confidence describes how often the method succeeds across repeated samples. Individual beads only have outcomes such as yellow or another color.

Two interval estimates for the yellow-bead share



UNIT 3 / AFTER DAY 13 / FOR DAY 14 • PR14 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

3

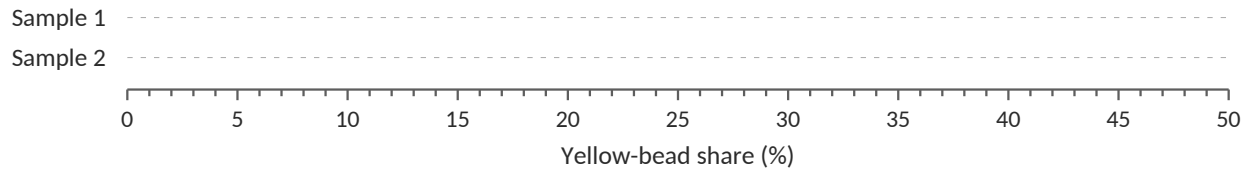
> **KEY IDEA**
>
> The population value stays fixed while intervals change from sample to sample.

1. Which simulated interval missed the true yellow share, and by how many percentage points did its nearest endpoint miss? How can you tell here?

2. A class runs 25 fresh samples through the same 95% procedure. Of the 25 intervals, 23 contain 30%. What percentage captured the truth in this batch? Explain why that result is possible.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



After you try both questions, check your reasoning on W156. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, Chapter 8 introduction, pp. 405–406; §8.1, pp. 407–410. Examples and data are constructed for this course.

Name _____ Date _____

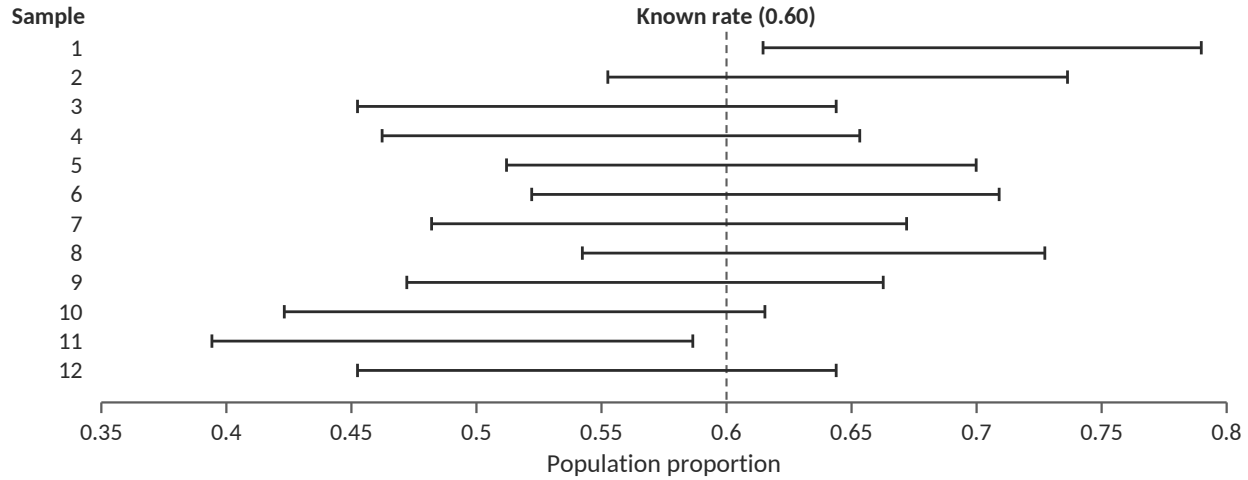
UNIT 3 / DAY 14 • CLASSWORK • T6

Count the captures

Read a repeated-interval display and explain coverage as a property of the sampling and interval procedure.

Each horizontal line is a 95% interval from one simulated sample of 100. The population proportion is 0.60. An interval captures 0.60 if it includes that value, even at an endpoint. For the second set, use 12 more intervals from the same simulation. If the class skips that run, use this supplied result: all 12 intervals in the second set captured 0.60.

Twelve simulated intervals



1. Identify one sample whose interval misses 0.60. Explain how its endpoints show the miss. What fraction of all 12 intervals capture 0.60?

2. Combine both sets. Divide the total number of captures by the total number of intervals. Would averaging the two capture percentages give the same result here? Explain.

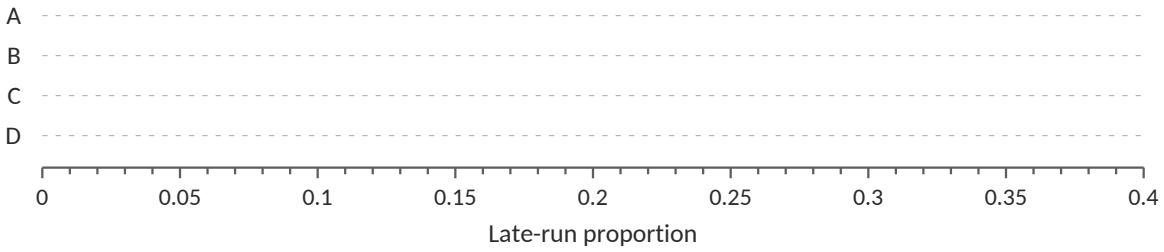
3. Would 95% confidence guarantee exactly 95 captures in the next 100 intervals? Use the displayed batch to explain what varies.

Name _____ Date _____

UNIT 3 / DAY 14 • INDEPENDENT WORK

Read an interval strip

A simulation uses a fixed late-run rate of 0.20 and n = 100 per sample. Four supplied 95% intervals, rounded, are A: 0.109–0.255, B: 0.167–0.332, C: 0.210–0.385, D: 0.085–0.221.



1. Draw the four labeled intervals on separate rows. Mark the known rate of 0.20 with a vertical line. Which intervals contain it, and what percentage of the four is that?

2. The next report concerns an unknown late-run rate and supplies one interval. Can its author mark it as a capture? Explain the distinction between the two reports.

Revise: Check your capture count against the endpoints. In your explanation, name the quantity that stayed fixed.

Exit: If 19 of 20 simulated intervals contain the known rate, what does the 19 count? What would remain unknown when measuring a real population?

Name _____ Date _____

UNIT 3 / DAY 14 HOMEWORK • HW14 • PRACTICE

Repair a coverage report

After Day 14: Homework and reading to prepare for Day 15. Turn in this two-sided HW14 sheet when requested. Keep PR15 on W089 in your binder.

The simulated population has a known positive-test rate of 40%. Each simulated sample produces one 95% interval. The table shows how many intervals contained 40% in each batch.

Batch	Intervals made	Intervals containing 40%	Percent containing 40%
1	10	9	
2	20	18	
3	30	29	

1. Compute coverage in each batch and across all 60 intervals. A classmate averages the three batch percentages. Does that give the combined coverage here? Explain using the batch sizes.

Name _____ Date _____

UNIT 3 / DAY 14 HOMEWORK • HW14 • PRACTICE CONTINUED

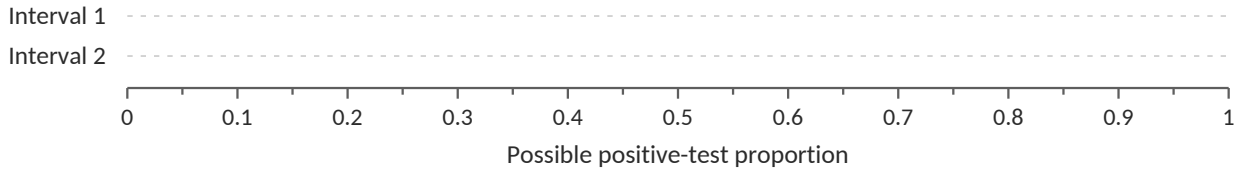
Repair a coverage report

3

Continue the same assignment. Use both sides of this sheet for the data and your work.

KEEP THE NEXT SHEET

Read PR15 on W089 and answer its two self-checks on W090 before Day 15. Keep that sheet in your binder.



2. Draw two example intervals on a common horizontal scale. Choose the endpoints so one interval includes 0.40 and the other misses it. Mark 0.40 with a vertical line. What varies between samples and what stays fixed?

3. Rewrite "95% of individual test results fall in the 95% interval." Then explain why 56 captures in 60 intervals is compatible with a 95% procedure.

UNIT 3 / AFTER DAY 14 / FOR DAY 15 • PR15 • PREPARATION READING

Use the interval, not just its center

Complete after Day 14 to prepare for Day 15.

Interpret a supplied confidence interval and apply a preselected decision threshold.

A bakery must decide whether to offer a smaller loaf. Before surveying anyone, management decides it will proceed only if the entire 95% confidence interval for customer interest is above 50%. It selects 50 customers at random from a large, current customer list. All respond, and 30 say they would order the loaf. The sample proportion, 30/50 = 60%, is a point estimate for everyone on the list.

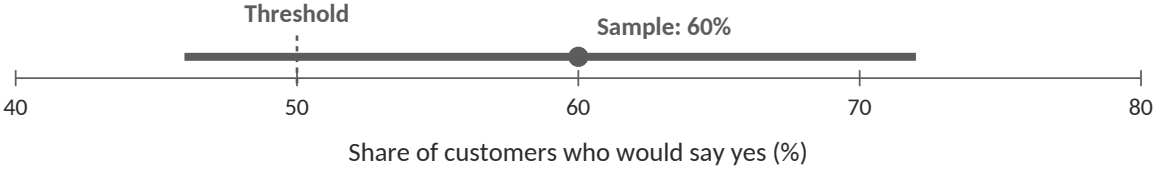
The supplied 95% interval runs from about 46% to 72%. In a report to the baker, we could write, "We are 95% confident that about 46% to 72% of customers on the list would answer yes when asked if they would order the smaller loaf." The interval allows for sampling variation. Another random group of 50 customers could give different endpoints. The question asked what people would do, so an actual count of future orders could tell a different story.

On the display, find the 50% mark. Some values in the interval exceed 50%, including the sample estimate of 60%. Other plausible values, such as 48%, fall below it. For the entire interval to lie above a cutoff, its lowest value must already exceed that cutoff. Here the lower endpoint is about 46%, so the bakery's survey does not meet its rule. The data leave both majority and minority interest plausible.

Setting the rule ahead of time matters. A baker who lowers the cutoff after seeing the interval has changed the decision being made. And an interval only addresses the population and responses the design supports. Even a very narrow interval from this list would leave interest among people outside the list unanswered.

Bakery survey	Value	Meaning
Sample share	60%	30 of the 50 respondents said yes
Supplied interval	46%–72%	Plausible shares of yes responses on the customer list
Trial rule	Above 50%	Compare the whole interval with the threshold

A decision about smaller loaves



UNIT 3 / AFTER DAY 14 / FOR DAY 15 • PR15 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

3

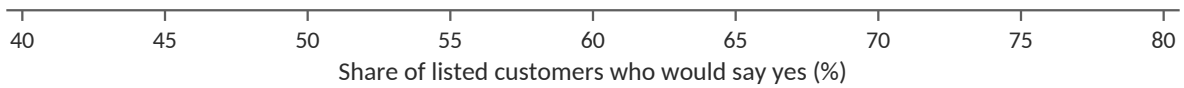
> **KEY IDEA**
>
> A point estimate above a target does not establish that the population value is above it.

1. Write one sentence to the baker interpreting the 95% interval. Include the people and response it describes.

2. Two bakeries with these same survey results had different rules before the survey. One requires the entire interval above 50%, the other above 45%. Which would proceed? Use the relevant endpoint and explain why 60% alone cannot settle either decision.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. OpenStax, Introductory Statistics 2e, §§1.2 and 8.3 (pp. 18–25, 420–424). The supplied interval uses the Wilson method in the NIST/SEMATECH e-Handbook, §7.2.4.1.

Name _____ Date _____

UNIT 3 / DAY 15 ● CLASSWORK ● T6

A trail map proposal

Interpret both confidence-interval endpoints for a population and apply a decision rule chosen in advance.

> **CHOOSE ONE GUIDED ACTIVITY**
>
> Your teacher will choose one activity. For the trail map, stay on this page. For the globe toss, use W093–W094. Everyone completes independent work on W092.

Suppose 80 users are selected in a simple random sample from a large list of trail-pass holders. All respond, and 52 say they would use a new map. The 95% interval is 54.1%–74.5%; the 90% interval is 55.9%–73.2%. The proposal needs a 95% lower endpoint above 55%.

1. Compute the point estimate and write the 95% interval interpretation with the population and measured response.

2. Does the 95% interval qualify the trail-map proposal for a trial? Cite the endpoint that decides it. Explain what would be wrong with switching to the 90% interval after seeing the results.

3. For each interval, subtract the lower endpoint from the upper. How much wider is the 95% interval in percentage points? Explain why the widths differ even though the sample is the same.

Name _____ Date _____

UNIT 3 / DAY 15 • INDEPENDENT WORK

A parcel locker

In this classroom case, a simple random sample of 120 customers from a large register yields 72 who would choose a parcel locker. Assume complete response. A supplied 95% interval is 51.1%–68.3%. A trial requires the entire interval to exceed 50%.

1. Calculate the estimate, compare both endpoints with the threshold, and give the decision.

2. Write the recommendation for a reader unfamiliar with the study. Address stated interest and actual use.

Revise: Underline the part of your recommendation that names the population. Check that your wording describes what customers said they would do.

Exit: A policy needs an entire 95% interval above 70%. The estimate is 74% and its interval is 65%–81%. What does the evidence establish about the rule?

Name _____ Date _____

UNIT 3 / DAY 15 ● OPTIONAL DATA LAB ● COLLECT

Globe toss | Collect your class data

Use these two pages if your teacher selects the globe activity. Then resume Day 15 Independent Work.

If your teacher chooses the globe activity, use these two pages in place of the trail-map activity. Estimate the proportion of valid catches classified as water. Then decide whether that estimate can tell you how much of Earth's surface is covered by water.

1. Everyone uses the right index finger. Start each gentle toss with the globe facing a different direction. At the first catch, hold still and check the point under the center of that fingertip. Take turns tossing, catching, or recording.

2. Record W if the point is clearly water, L if it is clearly land, and U if you cannot tell. All three count as valid catches. A drop or a missed first contact is R: record the retry separately, toss again, and leave R out of the valid total.

3. Stop at the agreed time. Add the groups' counts in each column. If the class has no valid catches, ask your teacher for a practice tally and label it as supplied data.

Pool the recorded outcomes

Group	Clear water W	Clear land L	Uncertain U	Retries R	Valid n
1					
2					
3					
4					
5					
6					
Class total					

For each group and the class total, add W + L + U to find n. Leave retries in their own column.

First estimate: Use the class totals to estimate the percentage of valid catches that were clearly water. What does your denominator count?

Name _____ Date _____

UNIT 3 / DAY 15 ● OPTIONAL DATA LAB ● ANALYZE

Globe toss | Interpret the estimate

Use the same set of counts for the two ways of scoring uncertain catches.

> **CHOOSE YOUR DATA**
>
> Start with your class tally. If your teacher supplies practice counts instead, use that set for all three questions. Label which data you used.

1. Let n = W + L + U. Find the clear-water percentage, 100 × W/n. Recalculate it as 100 × (W + U)/n if all uncertain contacts count as water. How many percentage points apart are the two results?

2. Your teacher will supply a 95% interval from the clear-water count and valid-catch total, or use the interval supplied with the practice tally. Report the endpoints and whether both exceed 50%. This calculation assumes independent catches with a stable chance of clear-water contact. Explain one way the catching method could violate that assumption.

3. After your teacher gives the percentage of Earth's surface covered by water, compare it with your estimate. Would every equal-sized area of the globe have the same chance of touching the fingertip? Explain one reason the chances could differ.

Name _____ Date _____

UNIT 3 / DAY 15 HOMEWORK • HW15 • PRACTICE

An evening bus proposal

After Day 15: Homework and reading to prepare for Day 16. Turn in this two-sided HW15 sheet when requested. Keep PR16 on W097 in your binder.

Suppose 100 adult riders are selected in a random sample from a large, complete registered-rider list. All answer the survey, and 68 say they would try an evening bus. Supplied intervals are 58.3%–76.3% at 95% confidence and 59.9%–75.1% at 90%. The transit board requires a 95% lower endpoint above 60%.

1. Calculate the sample share and both interval widths. Compare the widths using percentage points; state what was held constant.

Name _____ Date _____

UNIT 3 / DAY 15 HOMEWORK • HW15 • PRACTICE CONTINUED

An evening bus proposal

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR16 on W097 and answer its two self-checks on W098 before Day 16. Keep that sheet in your binder.

2. Draft a standalone recommendation under the board's rule. Explain why the sample share alone does not settle it.

3. Now suppose the 100 responses instead came only from riders who clicked a link advertising longer hours. Which claim in your recommendation would you revise, and why?

UNIT 3 / AFTER DAY 15 / FOR DAY 16 • PR16 • PREPARATION READING

The sampling frame sets the reach

Complete after Day 15 to prepare for Day 16.

Identify the population actually represented by sampled records and the consequence of repeated visits.

A museum director wants to know whether people who might attend an exhibit would support it. A list of current adult members is available. That list is a sampling frame, the collection of cases from which someone can actually draw a sample. Random selection from the list gives those members a chance to answer. Nonmembers and children who might attend have no entry on this frame. A conclusion about all possible visitors would extend beyond the evidence.

Now consider a service office that draws a simple random sample of 50 visits from a much larger weekday log. Every selected visitor answers whether the wait on that visit was acceptable. Thirty say yes. The estimate is 30/50 = 60%. The supplied 95% interval, about 46%–72%, estimates the share of weekday visits in that log for which the visitor would report an acceptable wait.

Why say "visits" every time? Imagine a tiny log with six entries. One customer came five times and found the wait acceptable every time. A different customer came once and found it unacceptable. Five of the six visits had acceptable waits, but only one of the two people always found the wait acceptable. A frequent visitor contributes several opportunities to be sampled from a visit log.

Before writing a conclusion, identify one eligible case and the exact response recorded for it. If the office wants to know what distinct customers think, draw from a list with one entry per eligible customer and ask each selected person the same clearly defined question. Choosing more rows from the weekday visit log still leaves weekends out and still gives repeat visitors more entries.

Eligible cases	Outcome recorded	Who the register or log can represent
Current adult museum members	Favor the proposed exhibit?	Current adult members on the register
Weekday service visits	Was this visit's wait acceptable?	Weekday visits on the log

UNIT 3 / AFTER DAY 15 / FOR DAY 16 • PR16 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

3

> **KEY IDEA**
>
> The denominator and sampling frame must match the people or events named in the conclusion.

1. The director headlines the membership survey "Future visitors support the exhibit." Name a group the museum could defensibly make a claim about, and one group the headline adds without evidence.

2. In the six-entry example, calculate the acceptable share of visits and the share of distinct customers who always reported an acceptable wait. Why do the two percentages differ?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Service office (check 2)	Number meeting the definition	Total	Share
Acceptable visits			
Customers reporting acceptable waits on every visit			

After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. OpenStax, Introductory Statistics 2e, §§1.2 and 8.3 (pp. 18–25, 420–424). The supplied interval uses the Wilson method in the NIST/SEMATECH e-Handbook, §7.2.4.1.

Name _____ Date _____

UNIT 3 / DAY 16 • CLASSWORK • T6 AND T1

Feedback workshop after the checkpoint

Use the sampling design and confidence interval together to defend a population claim, then revise a specific error.

TODAY'S INDIVIDUAL TASK

Complete the assessment your teacher gives you first. The sample begins on A006 in Sample Assessments. After you turn it in, use this page for feedback.

After the individual checkpoint, examine this new classroom case. A random sample of 40 from a large register of current trail-pass holders yields 26 who support a refill station. All sampled holders respond. A supplied 95% interval is 49.5%–77.9%.

1. A draft says 'About two thirds of everyone in the county supports the station.' Identify one correct numerical detail and one unsupported extension.

2. Rewrite the claim to identify the group, outcome, estimate, and both endpoints. Could the office claim a majority at this confidence level?

Name _____ Date _____

UNIT 3 / DAY 16 • INDEPENDENT WORK

A second chance with a fresh case

After reviewing your checkpoint, consider a random sample in this classroom case of 50 from a large list of registered adult pool users. All respond and 30 favor a quieter evening. A supplied 95% interval is 46.2%–72.4%.

1. Write a standalone conclusion about support and evaluate whether the whole interval exceeds 50%.

2. A newsletter generalizes the result to all adults living near the pool. Who could have been left out? Describe a sampling plan for the newsletter's population.

Revise: Read your conclusion as if you lived near the pool. Would you know whose responses it describes? Revise any wording that broadens the claim.

Exit: A survey randomly selected only permit holders. Name one group outside its reach and one phrase that would fix an overly broad conclusion.

Name _____ **Date** _____

UNIT 3 / DAY 16 HOMEWORK • HW16 • PRACTICE

Revise and transfer

After Day 16: Homework and reading to prepare for Day 17. Turn in this two-sided HW16 sheet when requested. Keep PR17 on W103 in your binder.

Use the feedback on today's checkpoint, then consider a new classroom study. A workshop randomly samples 80 names from its large register. All respond; 49 say they would attend an evening session. The supplied 95% interval is 50.3%–71.2%.

1. Clarify one sentence from your checkpoint and explain the change. If it is unavailable or your sentence is already clear, correct this claim using today's workshop survey: "Most people will attend."

Name _____ Date _____

UNIT 3 / DAY 16 HOMEWORK • HW16 • PRACTICE CONTINUED

Revise and transfer

3

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR17 on W103 and answer its two self-checks on W104 before Day 17. Keep that sheet in your binder.

2. A trial requires the entire 95% interval to exceed 50%. Write a recommendation citing the observed share, both endpoints, and the population. What does the survey measure?

3. Explain 95% confidence across repeated random samples from the register. Now suppose the organizer asked only people who stayed after one evening event. Which group could those answers describe directly?

UNIT 4 / AFTER DAY 16 / FOR DAY 17 • PR17 • PREPARATION READING

Ask what the benchmark could produce

Complete after Day 16 to prepare for Day 17.

Explain a chance model, a simulated sample, and an upper-tail p-value.

Seven of 30 shuttle runs arrived late. Is that unusual for a service where each run has a 10% chance of being late? Before examining the records, the office chose to look for an increase. Its starting assumption, called the null hypothesis, is a 10% late rate. The alternative is that the rate is higher. Even when the rate is exactly 10%, different sets of 30 runs can produce different late counts.

Imagine drawing a digit at random from 0 through 9, with all ten digits equally likely. Let 0 represent a late run and the other nine digits represent on-time runs. Replace the digit and draw again independently. One draw represents one run. A batch of 30 draws represents one sample of 30 runs under the benchmark. For each batch, record just one number: its count of zeros.

Compare the simulated counts with the observed count. A batch with seven zeros matches the observation; a batch with eight exceeds it in the direction the office chose. Both count toward the upper tail. Six zeros falls short. Suppose 14 out of 500 simulated batches had seven or more zeros. The estimated p-value is 14/500 = 0.028. Under this model, 2.8% of the simulated batches produced a result at least as high as the one observed.

The denominator is 500 because each batch supplies one comparison with the observed sample. More simulated batches can sharpen that estimate. They cannot tell us whether the actual runs were independent. If several buses encountered the same road closure, one event could have produced several delays.

What one model result represents	What it records
One random digit	One shuttle run; a zero represents a late arrival
One batch of 30 digits	One sample of 30 runs; count its zeros
A batch with at least seven zeros	One qualifying sample in the upper-tail tally

UNIT 4 / AFTER DAY 16 / FOR DAY 17 • PR17 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> To estimate an upper-tail p-value, divide the number of benchmark batches that match or exceed the observed count by the total number of batches.

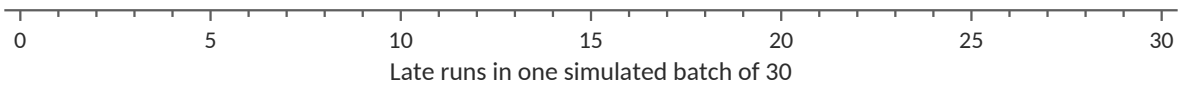
4

1. Which features of the shuttle model stay fixed each time you generate a fresh batch? What can change?

2. In a second simulation, 8 batches had exactly seven late runs, 3 had more than seven, and 189 had fewer than seven. Estimate the upper-tail p-value for seven observed late runs.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.



After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. OpenStax, Introductory Statistics 2e, §§4.3, 9.1 and 9.4 (pp. 237–239, 462–464, 467–469). The shuttle simulation was developed for this course.

Name _____ Date _____

UNIT 4 / DAY 17 • CLASSWORK • T7

Build a class shuttle simulation

I can estimate an upper-tail p-value and explain what one simulated batch represents.

In the shuttle example, six of 25 sampled runs were more than five minutes late. Simulate runs under the assumption that each run independently has a 10% chance of being late. Generate independent, equally likely digits 0–9 using a random-number tool. Alternatively, place ten same-size tokens in a bag, one labeled with each digit 0–9; mix and draw without looking, replacing the token after each draw. Count a zero as late. Count every other digit as on time. One batch of 25 digits represents one sample of 25 runs.

Record five digits in each numbered column. Count the zeros across all five columns for one batch.

Batch	1–5	6–10	11–15	16–20	21–25	Late count	At least six?
1							
2							

Class batches posted: _____ Batches with at least six late runs: _____

1. Make two batches of 25 digits. For tokens, return and remix after each draw. Record the digits and post the number of zeros in each batch. What does one posted number represent?

2. What fraction of the class batches had six or more late runs? In a separate simulation of 1,000 batches, 24 had exactly six late runs, 10 had seven, and one had eight or more. Find the fraction with six or more in the separate simulation. Compare it with your class fraction.

3. If no class batch reaches six, what should you report? Explain why you would still run more batches.

Name _____ Date _____

UNIT 4 / DAY 17 • INDEPENDENT WORK

Water sensors

A facility checks 20 independently selected sensors and finds four with a warning. The benchmark warning rate is 10%. An increase was the stated concern. Among 1,000 simulated batches under the benchmark, 100 have exactly four warnings, 23 have five, and 12 have six or more.

4

1. Compute the observed warning rate and the estimated upper-tail p-value. Label each denominator.

2. Write a sentence interpreting the simulated percentage so that someone who has not read the problem can understand it.

Revise: Circle each denominator in your first answer. Check that the sentence about the p-value describes whole simulated samples.

Exit: An upper-tail simulation produced 500 batches. Eight had exactly the observed count and 13 had a larger count. Estimate the upper-tail probability. What does its denominator count?

Name _____ **Date** _____

UNIT 4 / DAY 17 HOMEWORK • HW17 • PRACTICE

Inspect the orchard's sorting line

After Day 17: Homework and reading to prepare for Day 18. Turn in this two-sided HW17 sheet when requested. Keep PR18 on W109 in your binder.

An orchard expects 20% of crates to need a second inspection. It asks in advance whether that fraction has risen. In a sample of independent crate outcomes, 9 of 25 crates needed another inspection. Among 1,000 simulated batches at the benchmark, 29 had exactly 9 such crates, 15 had 10, and 6 had 11 or more.



Crates needing a second inspection in one batch of 25

1. Assign equally likely digits from 0 to 9 so that one digit represents one crate and 20% of crates need reinspection. State how many digits make one batch. On the number line, mark the observed count and shade all counts at least that large.

Name _____ Date _____

UNIT 4 / DAY 17 HOMEWORK • HW17 • PRACTICE CONTINUED

Inspect the orchard's sorting line

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR18 on W109 and answer its two self-checks on W110 before Day 18. Keep that sheet in your binder.

4

2. Calculate the observed crate fraction and the estimated p-value. Explain why counting only batches above nine changes the answer.

3. Write an interpretation for the inspection manager. State one check on the sampling method before applying the conclusion to all shipments.

UNIT 4 / AFTER DAY 17 / FOR DAY 18 ● PR18 ● PREPARATION READING

Size of a change and strength of evidence

Complete after Day 17 to prepare for Day 18.

Distinguish an observed effect from statistical significance and practical importance.

A workshop checks whether its defect rate has risen above 20%. Before inspecting any items, it sets a rule: open a review if the upper-tail p-value is at most 0.05. This cutoff is the significance level. Agreeing on the rule in advance gives everyone the same standard, whichever result the inspection produces.

Imagine two possible inspections of the same process. The first finds 7 defective items among 20, so 7/20 = 35% were defective. The second finds 14 among 40, also 35%. Each sample exceeds the 20% benchmark by 15 percentage points. That difference describes the size and direction of the observed change. Expressing it in ordinary counts may help a supervisor picture it: 35 defective items per hundred would be 15 more than the benchmark of 20 per hundred.

How unusual is each result? Suppose every item independently has a 20% chance of being defective. The chance of finding at least 7 defects in 20 items is about 0.087. The chance of finding at least 14 in 40 is about 0.019. Sample percentages from larger independent samples tend to vary less around the benchmark. The same observed excess therefore supplies stronger evidence in the second inspection. Only the second result is statistically significant at the chosen level.

Should that excess matter to this workshop? It depends on what a defect costs: perhaps a label can be reprinted immediately, or perhaps the item must be discarded. A useful recommendation gives the size of the observed excess, explains whether the evidence meets the cutoff, and names a practical consequence.

Inspection	Defective	Excess above 20%	Upper-tail p-value
7 of 20	35%	15 percentage points	0.087
14 of 40	35%	15 percentage points	0.019

UNIT 4 / AFTER DAY 17 / FOR DAY 18 • PR18 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> The observed difference measures size. The p-value measures how unusual the result would be under the benchmark. The consequences determine its practical importance.

1. In the two inspections, what stayed the same about the observed excess? Which result meets the workshop's investigation rule? Support both answers.

2. Another workshop observes 30% defective against the same 20% benchmark. The supplied p-value is 0.08. Give the excess and explain what the manager can conclude under the 0.05 rule.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Sample	Observed excess (points)	Upper-tail p-value	Meets 0.05 cutoff?
7 / 20			
14 / 40			
Check 2: another workshop			

After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. OpenStax, Introductory Statistics 2e, §§4.3 and 9.4–9.5 (pp. 237–239, 467–473). For effect size and statistical significance, see the American Statistical Association's 2016 statement on p-values, principle 5.

Name _____ Date _____

UNIT 4 / DAY 18 • CLASSWORK • T7

Compare two samples of late shuttle runs

I can compare a p-value with a preselected cutoff and report the size and practical importance of the observed change.

Two routes have the same 10% benchmark late rate. Use the two samples below. Before collecting the samples, the dispatcher chose to test for an increase on each route at a significance cutoff of 0.05.

Route	Late runs	Runs sampled	Upper-tail p-value	Observed late rate	Excess over 10% (points)
First	3	15	0.184		
Second	20	100	0.002		

1. For each route, find the observed late rate, subtract the benchmark, and report the difference in percentage points. Compare each p-value with 0.05. How can the same observed increase lead to different test decisions?

2. A reporter says the second route has a bigger lateness problem. Rewrite the claim using what these two samples establish.

Name _____ Date _____

UNIT 4 / DAY 18 • INDEPENDENT WORK

A trail counter misses walkers

In a random audit of 100 comparable crossings, the trail counter missed 15 crossings. The usual missed-crossing rate is 10%. The supplied upper-tail p-value is 0.073. The team set its cutoff at 0.05 before the audit. Treat the sampled crossing outcomes as independent under the benchmark model.

4

1. Find the observed excess and compare the p-value with the cutoff.

2. Write a conclusion for a ranger who thinks the audit proved that the old rate still holds.

Revise: Check the ranger's conclusion against the 0.05 rule. Revise any wording that treats a result above the cutoff as proof of the old rate.

Exit: A clinic sees a 3-percentage-point increase in late arrivals. The p-value is 0.011 and its preselected cutoff is 0.05. What else would the clinic need to know before changing staffing?

Name _____ Date _____

UNIT 4 / DAY 18 HOMEWORK • HW18 • PRACTICE

Two monitors, one decision standard

After Day 18: Homework and reading to prepare for Day 19. Turn in this two-sided HW18 sheet when requested. Keep PR19 on W115 in your binder.

Two independent audits each test for an increase, using a significance cutoff of 0.05 chosen in advance. A scanner misses 15 of 50 packages against a 20% benchmark (p-value 0.061). A labeler damages 8 of 30 labels against a 10% benchmark (p-value 0.008). Treat outcomes within each audit as independent with a constant rate under its benchmark model.

1. For each device, report the observed rate and its difference from its benchmark in percentage points. Show how you obtained each number.

Name _____ Date _____

UNIT 4 / DAY 18 HOMEWORK • HW18 • PRACTICE CONTINUED

Two monitors, one decision standard

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR19 on W115 and answer its two self-checks on W116 before Day 19. Keep that sheet in your binder.

4

2. Write a two-sentence report comparing each result with its own benchmark and the common significance cutoff.

3. Critique this recommendation: Stop both devices because both percentages rose. Identify one concrete piece of operational evidence needed to decide what to do.

UNIT 4 / AFTER DAY 18 / FOR DAY 19 • PR19 • PREPARATION READING

A decision can be wrong in two ways

Complete after Day 18 to prepare for Day 19.

Identify Type I and Type II errors from the actual condition and the decision made.

A greenhouse has to decide whether to send someone to check its cooling system. The sensor signals "check" when its readings suggest overheating. On any one occasion the greenhouse is actually safe or too hot, and the sensor either sounds the alarm or stays silent. To classify a decision, put a finger on the row for the actual temperature first, then move to the column for the sensor's action.

Suppose the alarm sounds while an independent thermometer shows that the greenhouse is safely within range. The sensor has rejected a true benchmark of safe conditions. This is a false alarm, or Type I error. A staff member may interrupt other work to inspect equipment that is functioning normally. If the alarm sounds while the greenhouse really is too hot, the same action is a correct detection.

Now suppose the greenhouse is actually overheating. If the sensor stays silent, it has missed a real change. This is a Type II error. The delay can leave plants exposed to excessive heat. Silence when the greenhouse is safe is the other correct decision. Knowing the action alone never tells us which row of the table to use.

The same reasoning applies when a dispatcher flags a route for a possible increase in its late rate. A flag is a decision to investigate. The underlying rate may still be at the benchmark, or it may have risen. A statement about an error should identify what was actually true, what the dispatcher decided, and a consequence for someone who uses or operates the route.

Actual greenhouse condition	Alarm sounds	Alarm stays silent
Safe	False alarm (Type I)	Correct decision
Too hot	Correct detection	Missed event (Type II)

UNIT 4 / AFTER DAY 18 / FOR DAY 19 • PR19 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> An error label requires both what was true and what the rule decided.

4

1. An independent thermometer confirms overheating. The alarm stays silent. Name the error and give a possible consequence.

2. The sensor sounds twice this week. Can the manager count two Type I errors from that record alone? What additional evidence would settle the question?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Actual temperature	Alarm	Silence
Safe		
Too high		

After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §9.2, pp. 464–466. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 4 / DAY 19 • CLASSWORK • T8

Build a truth-by-action table

I can identify a false alarm and a missed change and explain a concrete consequence of each.

A lab uses a check to flag a purifier if its failure rate appears to exceed 5%. Compare a true rate of 5% with a true rate of 12%. Its decision is to inspect or continue using the purifier.

Actual condition	Inspect	Continue using
Rate = 5%		
Rate = 12%		

1. In each cell of the decision table, write "correct," "Type I error," or "Type II error" for that combination of true failure rate and action.

2. Write one consequence for each error and identify who bears it. Can the lab identify the error just by knowing that it inspected?

Name _____ Date _____

UNIT 4 / DAY 19 • INDEPENDENT WORK

Ordering replacement bulbs

A maintenance crew flags a batch if a sample suggests more than 8% of the bulbs fail early. Consider a true rate of 8% and a true rate of 18%. A flag triggers a supplier review.

1. Which combination of true rate and action is a Type I error? Describe a cost.

2. The crew did not flag a batch. Write two sentences explaining when that decision is correct and when it is a Type II error.

Revise: Read your Type II explanation again. Make sure it identifies both the true early-failure rate and the crew's decision.

Exit: Compare two possible true sign-damage rates: the benchmark and a higher rate. A trail office flags an increase. Under which rate is this a Type I error?

Name _____ Date _____

UNIT 4 / DAY 19 HOMEWORK • HW19 • PRACTICE

A school's meal-service review

After Day 19: Homework and reading to prepare for Day 20. Turn in this two-sided HW19 sheet when requested. Keep PR20 on W121 in your binder.

A cafeteria checks whether more than 10% of orders take over eight minutes. A flag triggers a workflow review. Compare two possible true rates, 10% and 22%.

True rate	Flag	No flag

1. Enter the two possible true slow-order rates in the first column. For each rate, label a flag and a no-flag decision as correct, Type I error, or Type II error.

Name _____ Date _____

UNIT 4 / DAY 19 HOMEWORK • HW19 • PRACTICE CONTINUED

A school's meal-service review

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR20 on W121 and answer its two self-checks on W122 before Day 20. Keep that sheet in your binder.

4

2. Write one standalone explanation of each error for the dining manager. Include the true rate, action, and person affected.

3. A student claims every flag shows there was a Type I error. Identify the missing fact and construct a specific counterexample.

UNIT 4 / AFTER DAY 19 / FOR DAY 20 ● PR20 ● PREPARATION READING

Power is detection under a stated alternative

Complete after Day 19 to prepare for Day 20.

Calculate power and miss rate, and distinguish both from the false-alarm rate.

How often would a review rule catch a real increase in shuttle delays? To answer, we need batches for which we know the actual late rate. A simulation lets us choose that rate. We apply the same review rule to each batch. A flag means that the batch triggered an investigation. Use the rows of the table as two separate trials of the rule.

In the first row, all 1,000 batches came from a model with the benchmark 10% late rate. The rule flagged 20 batches. Each of those 20 is a false alarm, giving an estimated false-alarm rate of 20/1,000 = 2%. The remaining 980 were correctly left unflagged. In the second row, all 1,000 batches came from a model with a higher true late rate of 25%. The rule flagged 700. These are correct detections, giving an estimated power of 700/1,000 = 70% at that particular alternative.

What happened in the other 300 batches in the second row? Delays really were more frequent, but the rule failed to flag those batches. Its estimated miss rate at 25% is 300/1,000 = 30%. Power and miss rate add to 100% because they divide the same 1,000 changed-condition batches into two groups. The 20 false alarms belong to a different row.

Before quoting a power value, name the change the rule was asked to detect. A rule that catches 70% of batches at a 25% late rate might catch a much smaller share at a 12% late rate. To estimate its power at 12%, run the same rule on new batches generated at 12%.

Assumed process	Batches flagged	Batches simulated
Benchmark: 10% late	20	1,000
Alternative: 25% late	700	1,000

UNIT 4 / AFTER DAY 19 / FOR DAY 20 • PR20 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> Power and miss rate use changed conditions; false-alarm rate uses the benchmark condition.

4

1. At the true 25% late rate, what fraction of batches did the rule leave unflagged? Explain what that fraction estimates.

2. Another 200 batches are generated at a true 15% late rate; 80 trigger the rule. Estimate its power at 15%. Which simulation row would you use to estimate false alarms?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Assumed process	Flagged batches	Unflagged batches	All batches
10% late (benchmark)			1,000
25% late			1,000
15% late (check 2)			200

After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §9.2, pp. 464–466. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 4 / DAY 20 • CLASSWORK • T8

Read the risk in natural counts

I can read a decision rule under two true rates and explain the tradeoff between false alarms and power.

Each shuttle batch contains 25 trips. The first rule flags a batch with at least five late trips. The second flags a batch with at least six. Use the table below. Dispatch will accept a rule only if its false-alarm rate at the benchmark is 5% or less.

Flag when ...	False-alarm rate at 10%	Power at 25%	False alarms per 1,000 at 10%	Misses per 1,000 at 25%
At least 5 late	9.8%	78.6%		
At least 6 late	3.3%	62.2%		

1. For each rule, estimate how many of 1,000 batches would be flagged at a true 10% late rate. In a separate 1,000 batches at a true 25% late rate, how many increases would that rule miss? Enter your counts in the last two columns of the table.

2. Which rule meets the 5% false-alarm limit? At a true 25% late rate, how many more increases per 1,000 batches would that rule miss? Name a true rate whose detection probability the table does not give.

Name _____ Date _____

UNIT 4 / DAY 20 • INDEPENDENT WORK

Inspecting orchard crates

An inspection rule flags a batch when enough of its 20 crates need re-inspection. Model the crates as independent with a common re-inspection probability. At a true 10% re-inspection rate, flagging at four or more gives 13.3% false alarms and at five or more gives 4.3%. At a true rate of 25%, the corresponding powers are 77.5% and 58.5%.

4

1. Find both miss rates at 25% and the increase in false alarms when the cutoff falls from five to four.

2. A crew can tolerate at most 8% false alarms. Write a recommendation describing one consequence of the eligible rule.

Revise: Mark the true rate attached to each percentage in your recommendation. Check which rate belongs with a miss.

Exit: At a true defect rate of 30%, a rule has 84% power. In repeated batches produced at that rate, what does the other 16% count?

Name _____ Date _____

UNIT 4 / DAY 20 HOMEWORK • HW20 • PRACTICE

A dispatcher for another service

After Day 20: Homework and reading to prepare for Day 21. Turn in this two-sided HW20 sheet when requested. Keep PR21 on W127 in your binder.

Dispatch compares two rules for batches of 30 independent trips. Each rule flags a batch when its number of late trips reaches the cutoff. The table gives the flag rate under two different assumptions about the true late rate.

Cutoff (late trips)	False-alarm rate at 10%	Power at 25%	Miss rate at 25%
At least 6	7.3%	79.7%	
At least 7	2.6%	65.2%	

1. Complete the miss-rate column. When the cutoff falls from seven to six, by how many percentage points do false alarms and power change? State the direction of each change.

Name _____ Date _____

UNIT 4 / DAY 20 HOMEWORK • HW20 • PRACTICE CONTINUED

A dispatcher for another service

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR21 on W127 and answer its two self-checks on W128 before Day 21. Keep that sheet in your binder.

4

2. Under a 3% false-alarm cap, choose the rule and write a consequence for riders. Then reconsider for an 8% cap with detection as the priority.

3. A supervisor reads 65.2% as the share of flagged batches that came from routes with a 25% late rate. What does 65.2% actually measure? What additional counts would answer the supervisor's question?

UNIT 4 / AFTER DAY 20 / FOR DAY 21 • PR21 • PREPARATION READING

Choose a rule with the costs in view

Complete after Day 20 to prepare for Day 21.

Apply a false-alarm constraint before comparing power and explain the resulting tradeoff.

A garden manager monitors how often soil readings fall below a moisture threshold. When the system flags a batch, someone goes out to inspect the irrigation lines. The regular process produces 10% low readings. The manager wants to detect a change to 25% low readings while keeping unnecessary inspections manageable. Before choosing a rule, she sets a maximum false-alarm rate of 4%.

Both proposed rules have been evaluated using the same batch size and conditions. At the 10% benchmark, Rule A falsely flags 6% of batches; Rule B falsely flags 3%. The first decision is about eligibility. Rule A exceeds the 4% cap; Rule B satisfies it. A high detection rate cannot make a rule eligible under a cap it exceeds.

Now consider only batches generated at the higher 25% rate. Rule A detects 82% and misses 18%. Rule B detects 71% and misses 29%. Choosing B means about 11 more missed batches per 100 batches at the higher rate. Under the benchmark, B produces about three fewer false alarms per 100 batches. Those are two different groups of 100; the overall number of inspections in a garden also depends on how often the harmful condition occurs.

A recommendation could read, "Use Rule B for irrigation inspections. Its 3% false-alarm rate meets the 4% limit. When the true share of low readings rises to 25%, the rule still misses about 29% of batches, so some irrigation problems may go uninvestigated." A change to the cap or to the size of change we want to detect deserves a fresh comparison.

Garden rule	False alarms at 10% benchmark	Power at 25% alternative
A	6%	82%
B	3%	71%

UNIT 4 / AFTER DAY 20 / FOR DAY 21 • PR21 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> A rule with greater power is not automatically the better choice when error costs differ.

4

1. Under the 4% cap, which rule may the manager use? At a true 25% share of low readings, what proportion would that rule miss?

2. Suppose the manager changes the cap to at most 6% and wants the highest power among eligible rules. Which rule should she choose? What extra cost does that involve at the benchmark?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Rule	Meets 4% cap?	Meets 6% cap?	Power at 25%	Miss rate at 25%
A				
B				

After you try both questions, check your reasoning on W157. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §9.2, pp. 464–466. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 4 / DAY 21 ● CLASSWORK ● T7 AND T8

Feedback after the decision memo

I can recommend a rule using its stated risk limit and explain what the actual sample contributes to the decision.

> **TODAY'S INDIVIDUAL TASK**
>
> Complete the assessment your teacher gives you first. The sample begins on A007 in Sample Assessments. After you turn it in, use this page for feedback.

After handing in your decision memo, consider a warehouse rule that flags batches for review. Its power is 75% when the true damage rate is 30%. A manager says, "Three quarters of the flagged batches really had that damage rate."

1. Which batches form the denominator of the 75% power figure? Rewrite the manager's statement so that it describes what the rule detects.

2. Clarify one risk statement in your memo. If it is unavailable, explain what counts you would need to determine the fraction of all flags that were false.

Name _____ Date _____

UNIT 4 / DAY 21 • INDEPENDENT WORK

A separate equipment review

A batch contains 40 independent device checks. The office allows at most 5% false alarms. Eight devices failed in this batch. The table gives the flag rate for each review rule under two different true failure rates.

4

Flag a batch with ...	False-alarm rate at a true 10%	Power at a true 20%
At least 7 failures	10.0%	71.4%
At least 8 failures	4.2%	56.3%

1. Which cutoff meets the office's limit? Does this batch trigger review? Give its observed excess above 10% in percentage points. Under a true 10% rate, what is the probability of at least the eight failures observed? Report that probability as the p-value.

2. Write a two-sentence recommendation including one cost of the eligible rule at the stated alternative.

Revise: Underline your statement about the observed batch. Check that your missed-batch percentage refers to repeated batches at the stated alternative.

Exit: A rule has 56% power at a stated alternative. The sample's p-value is 0.04. Which number concerns the observed sample, and how should the other be interpreted?

Name _____ Date _____

UNIT 4 / DAY 21 HOMEWORK • HW21 • PRACTICE

Repair a recommendation

After Day 21: Homework and reading to prepare for Day 22. Turn in this two-sided HW21 sheet when requested. Keep PR22 on W133 in your binder.

A booking office reviews changes among 20 independent reservations. At its 15% benchmark, flagging at 6 or more changes gives 6.7% false alarms; at 7 or more it gives 2.2%. At a true 30% change rate the respective powers are 58.4% and 39.2%. The office sets a 4% false-alarm cap and later observes 6 changes.

1. Choose a rule under the cap. At a true 30% rate, what share of batches would that rule miss, and who bears a cost?

Name _____ Date _____

UNIT 4 / DAY 21 HOMEWORK • HW21 • PRACTICE CONTINUED

Repair a recommendation

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR22 on W133 and answer its two self-checks on W134 before Day 22. Keep that sheet in your binder.

4

2. By how many percentage points does the observed change rate exceed 15%? Does this batch trigger the chosen review rule? Separately, what is the probability of at least the observed six changes under the benchmark?

3. Choose a sentence from your memo or feedback and clarify its risk or denominator. If your memo is unavailable or its conclusion is already clear, repair "There is a 39.2% chance the manager was right."

UNIT 5 / AFTER DAY 21 / FOR DAY 22 ● PR22 ● PREPARATION READING

Two percentages with the same numerator

Complete after Day 21 to prepare for Day 22.

Identify the conditioning group before calculating or interpreting a detection percentage.

A machine inspects 40 parts and flags 10 for review. A separate inspection of all 40 establishes which parts really are defective. There are eight defective parts, and the machine flagged six of them. The technician running the machine asks, "How many of the defects did we catch?" The employee opening a flagged box asks, "How often will I find a defect inside?" Both questions involve the same six parts.

For the technician, start with all eight defective parts. The machine found six and missed two. Its observed detection rate is 6/8 = 75%. For the employee checking flags, start with the ten flagged parts. Six were defective and four were sound. The share of flagged parts that were defective is 6/10 = 60%. Think of the denominator as the set of parts you put on the workbench before you start counting.

Here is a useful check on the wording. "The machine found 60% of the defects" would tell the technician that 60% of all eight defects were caught. The records say six out of eight. You can repair the sentence either by writing "The machine found 75% of the eight defective parts" or by keeping 60% and writing "Of the ten parts the machine flagged, 60% were defective."

An independent inspection of the unflagged parts matters. If the technician had opened only flagged boxes, four false alarms would still be visible, but the two missed defects would be hidden. In that case the technician could describe what turned up in the flagged boxes but could not calculate what share of all defects the machine detected.

Actual condition	Flagged	Unflagged	Total
Defective	6	2	8
Sound	4	28	32
All parts	10	30	40

UNIT 5 / AFTER DAY 21 / FOR DAY 22 • PR22 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
> "Of defective items" and "of flagged items" require different denominators.

1. In another batch, 12 parts were flagged and 10 parts were defective. Eight of the flagged parts were defective. What share of defects did the machine find, and what share of flags identified defects?

2. For the original batch of 40, suppose only the ten flagged boxes were opened. Which percentage could the technician calculate? What information would be missing?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

New batch (check 1)	Numerator counts	Denominator counts	Percentage
Share of defects found			
Share of flags identifying defects			

After you try both questions, check your reasoning on W158. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §3.4, pp. 183–186; §9.2, pp. 464–466. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 5 / DAY 22 • CLASSWORK • T1–T8

Check denominators, then repair a claim

Find the first unsupported step in a claim and show how the correction changes its meaning.

Everyone completes question 1 using the alert table. Then complete the additional model or interval question your teacher assigns. Explain the relevant group or rule to a partner before calculating. These are targeted review tasks for ideas already taught.

Alerts on 80 observed trips

Actual trip	Flagged	Unflagged	Total
Late	10	10	20
On time	6	54	60
Total	16	64	80

Read across a row or down a column according to the question. The cells describe observed trips.

1. Everyone: From the alert table, find the false-alarm share among trips that were truly on time. Name the denominator. Then explain what 6/16 measures and why it is a different rate.

2. Model station. A trip predicted to take 14 minutes took 17. A driver says its residual is −3 minutes because the trip was slow. Correct the calculation and interpret the sign.

3. Interval station. A 95% interval for the percentage of trips that arrive late is 8%–21%. Does this support the claim that at least 15% arrive late? Explain.

Name _____ Date _____

UNIT 5 / DAY 22 • INDEPENDENT WORK

Try your station in a new case

Complete the part of question 1 your teacher assigns for additional practice. Everyone answers question 2. In the scanner practice case, 12 of 50 deliveries were late. The scanner flagged 10 deliveries; seven flagged deliveries were late.

1. Model: A trip predicted to take 19 minutes took 24. Find and interpret its residual.
Interval: A 95% interval for the on-time delivery rate is 72%–86%. Does it establish that the rate exceeds 80%? Explain.
Denominator: What fraction of on-time deliveries did the scanner flag? Show how you find the numerator and denominator.

2. A colleague says, "The scanner caught 70% of the late deliveries." Correct the statement. Use the two relevant fractions to explain how the colleague got 70%.

Revise: Circle the two denominators in the scanner case. Rewrite any phrase that could refer to the wrong group of deliveries.

Exit: A report flagged 15 files. Nine of the 12 faulty files were flagged. What share of faulty files were flagged?

Name _____ **Date** _____

UNIT 5 / DAY 22 HOMEWORK • HW22 • PRACTICE

Correct and transfer

After Day 22: Homework and reading to prepare for Day 23. Turn in this two-sided HW22 sheet when requested. Keep PR23 on W139 in your binder.

For question 1, use a sentence from your marked checkpoint or this Day 22 scanner claim: "The scanner caught 70% of late deliveries." The scanner flagged 7 of 12 late deliveries; 7 of its 10 flags identified late deliveries. For questions 2 and 3, use the device-check counts in question 2.

1. Quote the sentence you chose, revise it, and explain which number, group, or conclusion needed to change.

Name _____ Date _____

UNIT 5 / DAY 22 HOMEWORK • HW22 • PRACTICE CONTINUED

Correct and transfer

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR23 on W139 and answer its two self-checks on W140 before Day 23. Keep that sheet in your binder.

5

True condition	Flag	No flag	Total
Faulty			
Not faulty			
Total			

2. A device flagged 20 of 100 checks. Of the 16 faulty cases, 12 were flagged. Complete the table. What fraction of faulty cases were flagged? What fraction of flagged cases were faulty? Label each denominator.

3. Write one question your table cannot answer about a different machine, and name the evidence needed.

UNIT 5 / AFTER DAY 22 / FOR DAY 23 ● PR23 ● PREPARATION READING

Plan the evidence before collecting it

Complete after Day 22 to prepare for Day 23.

Match a decision, target population, sampling frame, and measured outcome.

A county program needs to decide whether to reserve a room for a Saturday orientation. It wants to know what share of currently registered families say they would attend. The family is the case: each registered family should have one chance to be selected on a complete list, even if it has more than one child in the program. This list of eligible cases is the sampling frame.

Last year's sign-in sheet looks convenient. Which current families could never appear on it? A family that just registered and a returning family that skipped last year's event would both be missing. A family that attended last year but left the program might still appear. Randomly choosing names from the old sheet would describe the people on that sheet. It cannot give newly registered families a chance to be selected.

Suppose the program randomly selects 40 families from its complete current list, and all respond. Of those families, 24 say they would attend, giving 24/40 = 60%. A supplied 95% interval runs from 45% to 74% for the share of listed families who would say yes. The program set a room-reservation rule in advance: reserve only if the entire interval is above 50%. The lower endpoint, 45%, leaves less than a majority plausible, so the evidence falls short of the rule.

Even a clear majority saying yes would establish interest in the orientation. How many families actually attend is a different outcome. After the event, a sign-in record can measure attendance. Until then, staffing recommendations should identify the survey result as stated intent.

Source	Who could appear?
Complete registration list	Any family registered for the program
Last year's attendance list	Families who attended last year, including any no longer registered

UNIT 5 / AFTER DAY 22 / FOR DAY 23 • PR23 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
>
> A well-collected answer to the wrong question cannot settle the intended decision.

1. Which list should the county sample to estimate interest among currently registered families? Give one specific family the other list could leave out.

2. Suppose a larger survey of registered families gives a 95% interval of 51%–67%. Would it meet the same reservation rule? What does a yes response measure?

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

List	Who could appear?	Who could be missing?
Registered families		
Last year's sign-ins		

After you try both questions, check your reasoning on W158. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §1.2, pp. 18–25; §8.3, pp. 420–424. Examples and data are constructed for this course. The first interval uses the Wilson method (NIST/SEMATECH e-Handbook, §7.2.4.1).

Name _____ Date _____

UNIT 5 / DAY 23 ● CLASSWORK ● T1–T8

Rehearse with a partner

Combine study design, numerical evidence, and a stated decision rule in a short recommendation.

In a supplied practice survey, all 80 randomly selected registered trail-program families respond, and 52 say they would attend a Saturday orientation. A supplied 95% interval for this stated-interest proportion is 54%–75%. The organizer requires the entire interval to exceed 55% before approving a trial.

1. Find the sample percentage and check the rule using both endpoints. Identify the target population.

2. Write a two-sentence recommendation. Give it to a partner who cannot see the question. Ask them to identify the group, the organizer's rule, and your decision from your words alone. Revise any detail they missed.

Name _____ Date _____

UNIT 5 / DAY 23 • INDEPENDENT WORK

Can timing data support a change?

A delivery service recorded waiting times at pickup sites 2–6 miles from its depot. Its model predicts the wait in minutes as 5 + 2d, where d is distance in miles.

1. Find the prediction at 4 miles and residual if the observed wait was 16 minutes. Interpret the sign.

5

2. A manager promises an exact wait for a 12-mile pickup. Use the observed range and residual to evaluate that promise.

Revise: Add one detail from the model's observed distance range or its residual to support your judgment about the 12-mile pickup.

Exit: A committee will act only if an entire confidence interval for support is above 60%. Its interval is 57%–72%. Which endpoint determines whether the committee's rule is met? Explain.

Name _____ **Date** _____

UNIT 5 / DAY 23 HOMEWORK ● HW23 ● PRACTICE

A different client

After Day 23: Homework and reading to prepare for Day 24. Turn in this two-sided HW23 sheet when requested. Keep PR24 on W145 in your binder.

In a supplied practice survey, all 120 randomly selected garden-club members respond, and 84 are willing to try a new pickup site. A supplied 95% interval for the stated-interest proportion is 61%–77%.

1. Write the estimate and apply a rule requiring the entire interval to exceed 60%. Use the lower endpoint explicitly.

Name _____ Date _____

UNIT 5 / DAY 23 HOMEWORK • HW23 • PRACTICE CONTINUED

A different client

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR24 on W145 and answer its two self-checks on W146 before Day 24. Keep that sheet in your binder.

5

2. Suppose the same 84 of 120 said yes, but they volunteered through an online post. Whose willingness can you now describe? Explain why that changes what you can recommend to the club.

3. The club wants to estimate what share of its members would actually use the pickup site during a trial week. What would you record, and which members belong in the denominator?

UNIT 5 / AFTER DAY 23 / FOR DAY 24 • PR24 • PREPARATION READING

Choose the output that answers the question

Complete after Day 23 to prepare for Day 24.

Select a descriptive summary, prediction, or interval to match the quantity being asked about.

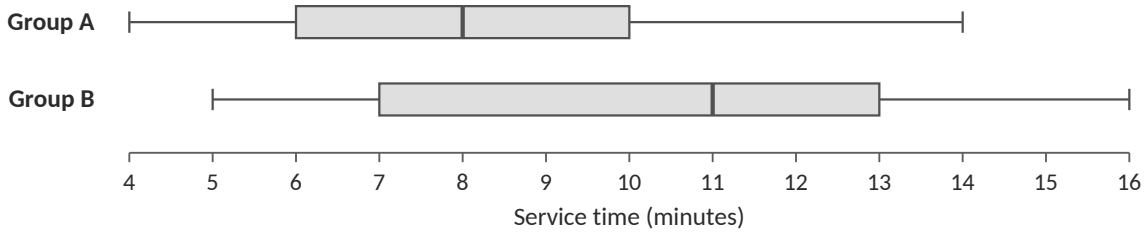
A service manager has two questions. At which counter did a typical visit take less time? And did more than 15% of visits across the service finish within five minutes? A visit lasting eight minutes and a visit lasting eighteen minutes provide different times for the first question. For the second, both have the same answer: neither finished within five minutes. Identify what you need to measure before choosing an output.

For the first question, use the boxplots. Find the line inside each box and read it against the minutes scale. The median was 8 minutes at A and 11 at B, so a typical visit in the recorded group at A was 3 minutes shorter. The ranges overlap. For example, the slowest recorded visit at A took 14 minutes and the fastest at B took 5 minutes. The median comparison describes the centers of the groups; individual visits can come out in the other order.

For the second question, use an interval for the share of visits that finished within five minutes. Suppose a random sample of visits across the service gives a 95% interval from 6% to 18%. The 15% target lies inside the interval. Values above and below the target are plausible, so the manager cannot establish from this sample that the overall share exceeds 15%.

Write the decision in the units of its question. Minutes describe a time; a percentage describes a share of visits. A model could answer a third question, such as how many minutes to allow for a specified workload. A prediction for one workload does not by itself establish what share of all visits meets a time target.

Service times for two groups



UNIT 5 / AFTER DAY 23 / FOR DAY 24 • PR24 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
> Identify the variable and target quantity before choosing which number to use.

1. Which counter had the shorter typical service time, and by how much? Identify the feature of the display that supports your conclusion.

2. After a staffing change, a new random sample of visit records gives a 95% interval of 17%–24% for the share of visits completed within five minutes. Has the service met the same 15% target? Write a sentence a manager could use.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Question	Outcome and unit	Relevant output
Typical service time		
Share finished within five minutes		

After you try both questions, check your reasoning on W158. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §2.4, pp. 94–98; §8.3, pp. 420–424; §12.3, pp. 623–626. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 5 / DAY 24 • CLASSWORK • T1–T8

Before the individual task

Use supplied statistical evidence to defend a conclusion independently in an unfamiliar setting.

> **TODAY'S INDIVIDUAL TASK**
>
> Complete the entry ticket and the brief tool-choice question below. Then begin the individual performance task on A009 in Sample Assessments. Self-check and revise afterward.

Complete the brief output-selection question below when your teacher directs. For the individual performance task, use the separate assessment your teacher distributes; its three sections are A, B, and C. Write and submit assessed answers on that assessment. The workbook is for the preparation and later self-check.

Section in the sample	Assessment pages
A. Compare scanning stations	A009
B. Plan a twenty-photo job	A010
C. Evaluate evidence and recommend a rule	A011–A012

1. A bus office wants to compare typical morning and evening trip times. It has boxplots of trip times and the percentage on time in each period. Which output would you use, and what would you compare?

Optional working space

Use this space for scratch calculations if permitted. Put each final answer on its assessment page.

Name _____ Date _____

UNIT 5 / DAY 24 • INDEPENDENT WORK

Self-check after the task

When the independent task ends, leave your first answers visible. At your teacher's signal, use your assessment to check one calculation and one conclusion. Mark revisions separately. Hand in the assessment after the self-check.

1. Find one answer with a signed quantity. Check its direction and units, and revise it if needed.

2. Read one conclusion without its question. Add any essential population, outcome, time window or comparison that is missing.

Exit: Which single answer would you revisit tomorrow, and what specific evidence would you recheck?

Name _____ **Date** _____

UNIT 5 / DAY 24 HOMEWORK • HW24 • PRACTICE

Rehearse a targeted revision

After Day 24: Homework and reading to prepare for Day 25. Turn in this two-sided HW24 sheet when requested. Keep PR25 on W151 in your binder.

A firm randomly sampled 100 orders from a large register of recent orders and found 18 returns. A supplied 95% interval for the register's return share is 12%–27%. A writer says, 'Returns are certainly above 15%.'

1. Name the first reasoning error. Rewrite the conclusion with an estimate, interval, and scope.

Name _____ Date _____

UNIT 5 / DAY 24 HOMEWORK • HW24 • PRACTICE CONTINUED

Rehearse a targeted revision

Continue the same assignment. Use both sides of this sheet for the data and your work.

> **KEEP THE NEXT SHEET**
>
> Read PR25 on W151 and answer its two self-checks on W152 before Day 25. Keep that sheet in your binder.

5

2. Suppose the firm had examined only calls from customers with complaints. Could those calls support an estimate for all recent orders? Explain how the records differ from a random sample of orders.

UNIT 5 / AFTER DAY 24 / FOR DAY 25 ● PR25 ● PREPARATION READING

Revise the reasoning, then transfer it

Complete after Day 24 to prepare for Day 25.

Repair an unsupported conclusion and identify what new evidence would justify a wider claim.

A trail group asks 30 weekend visitors whether they favor a new sign. Eighteen say yes. A student reports, "Sixty percent of town residents favor the sign." The arithmetic is right: 18/30 = 60%. The first unsupported step is changing the people in the denominator. Weekday visitors and residents who never use the trail had no chance to answer.

A revision that can stand alone is, "Of the 30 weekend trail visitors who answered the survey, 18 (60%) favored the proposed sign." That sentence tells a reader who was observed, what they were asked, and where the percentage came from. The selection method has not been specified, so a claim about every weekend visitor would require further evidence.

What would justify a claim about visitors across the season? Suppose the trail requires a free registration and keeps a complete list of individual visitors. A random sample from that season's list would give weekday and weekend visitors a chance to be selected. With one entry per visitor and appropriate follow-up for nonresponse, it could support an estimate for the season's visitors. A list of visits can give a frequent visitor several entries; a study about visitors needs to account for that difference.

The town might instead want to know whether its residents favor the sign. That requires a frame that includes residents who never visit the trail. This is what to look for in a revision: identify the first step the evidence does not support, repair the claim using what you have, and explain what additional evidence would support the wider conclusion.

Sample design	Scope to consider
Only weekend visitors were sampled	The surveyed weekend visitors who responded
Random sample across the whole season	The visitors on that season's sampling frame

UNIT 5 / AFTER DAY 24 / FOR DAY 25 • PR25 • KEEP YOUR RESPONSES

Check your understanding

Answer using the reading. These responses are for your preparation and are not collected.

> **KEY IDEA**
> Correct arithmetic is only one part of a valid statistical argument.

1. At a different trailhead, 15 of 24 weekend visitors surveyed wanted more picnic tables. Rewrite "Most residents want more tables" as one sentence the records support.

2. If the county wants the views of all town residents, would randomly selecting visitors from an entire season of trail registrations suffice? Describe a more suitable frame.

Optional working space

Use this organizer for the checks or to work through the example if helpful. No extra response is required.

Sample	Group the evidence describes
Weekend sample	
Season-long visitor list	
Proposed resident list	

After you try both questions, check your reasoning on W158. Bring any question you still have to class.

Source and adaptation. Concepts adapted from OpenStax, Introductory Statistics 2e, §§1.1–1.2, pp. 5–25; §1.4, pp. 34–37. Examples and data are constructed for this course.

Name _____ Date _____

UNIT 5 / DAY 25 ● CLASSWORK ● T1–T8

Revise from feedback

Repair an earlier answer and use the same reasoning in a fresh problem.

> **TODAY'S INDIVIDUAL TASK**
>
> Complete the entry ticket and identify what needs revising. The individual transfer task begins on A013 in Sample Assessments.

Use a response from your marked performance task. If it is unavailable, use this draft: "The model predicted a 14-minute wait and the actual wait was 19 minutes. The residual is +5 minutes, so the wait was five minutes shorter than predicted." If your own response was sound, change one study condition and reconsider your conclusion.

1. Identify the first step you would change and the rubric row it affects. If your work was already correct, identify the study condition you changed.

2. Write the corrected response or the conclusion under your changed condition. Underline the wording or number that explains why your conclusion changed.

Name _____ Date _____

UNIT 5 / DAY 25 • INDEPENDENT WORK

Transfer to a new setting

Complete exactly one teacher-assigned route in the Day 25 sample transfer assessment, then complete its common task. The assigned route is worth 8 points and the common task 20 points, for 28 points total. Submit the assessment before completing the reflection below.

1. Describe one decision you made on the transfer task. Explain which fact in the new problem guided that decision. If it uses the same reasoning as your earlier revision, identify the connection.

2. Write one question you would ask before accepting a statistical claim outside class.

Exit: Write one rule for judging a statistical claim and one example of when you would use it.

RESPONSE GUIDE

Check after your first attempt

Equivalent reasoning is welcome. Correct your own work; these are self-checks, not collected assignments.

PR01 ● Start in class: what does a record describe?

1. Across both players' practice sessions, 10 of 18 serves landed in bounds, about 56%. The counts are 6 + 4 in bounds and 12 + 6 attempts.

2. The percentage describes attempts by two players. To study players, define a rule such as "landed at least one serve in bounds" and count the players who met it.

PR02 ● From individual waits to a distribution

1. The corrected total is 23 − 2 + 3 = 24 minutes for six visits, so the mean library wait is 24 ÷ 6 = 4 minutes.

2. No. Both 2 and 3 minutes belong to the bin from 2 up to 5 minutes. Its bar still counts five visitors, even though one visitor's time changed.

PR03 ● Why one long wait changes the mean

1. The total rises by 13 − 8 = 5 minutes across five waits. The mean increases by 1 minute, from 4 to 5 minutes.

2. The reordered waits are 2, 3, 4, 8, and 9 minutes. The corrected value moves past the old middle, so the median rises from 3 to 4 minutes.

PR04 ● A fair comparison needs a scale

1. For the second event, (12 − 18) ÷ 3 = −2. The 12-minute and 22-minute runners both finished two SDs below the means of their own events.

2. The time was 3 minutes above the second event's 18-minute mean. An SD there is 3 minutes, so 21 minutes is one SD above the mean.

PR05 ● Two different uses of chance

1. The manager randomly sampled 30 shifts from the term roster. Their average waits can be used to estimate the mean of the shift averages for all shifts on that roster. No sign was assigned or compared.

2. For example, people who know the task may choose the familiar sign and finish more quickly because of their experience. The time difference would mix sign choice with prior experience.

PR06 ● A comparison that can stand on its own

1. Among the recorded pickups, the median wait was 8 minutes in the morning and 10 minutes in the afternoon, so the typical morning wait by this measure was 2 minutes shorter.

2. In the second log, morning pickups had the shorter median wait (8 versus 10 minutes), while afternoon pickups had the more consistent middle half (IQR 3 versus 6 minutes).

PR07 ● Two measurements, one case

1. The new four-mile shuttle trip took 14 minutes. At the same distance its point lies halfway between the 12-minute and 16-minute trips, two minutes from each.

2. The two-mile trip took 9 minutes. Doubling 9 gives 18 minutes, but the four-mile trips took 12 and 16 minutes. Recording stops or traffic on each trip could help explain the times.

RESPONSE GUIDE • CONTINUED

Check after your first attempt

PR08 • A prediction and its residual

1. The model predicts 4 + 2(6) = 16 seconds for six labels. This is 2 seconds longer than the five-label prediction.

2. For the six-label job, the residual is 13 − 16 = −3 seconds. The job finished 3 seconds sooner than the model predicted.

PR09 • What does the model leave out?

1. Both trips covered four miles, but they took 12 and 17 minutes. A model using distance alone gives both the same prediction. The log does not establish what caused the five-minute difference.

2. For the ordinary four-mile trip, 14 minutes exceeds the 13-minute slot, while 12.5 minutes fits. The scheduling decision depends on which model the dispatcher uses.

PR10 • An equation can go farther than the evidence

1. Eight envelopes gives 2 + 0.5(8) = 6 minutes and lies within the recorded range of 4–10. Twenty exceeds that range. Time comparable twenty-envelope jobs to check whether 12 minutes is reasonable.

2. Ten exceeds the recorded hand-sealed range of 4–8 envelopes, so that prediction requires extrapolation. Ten is inside the machine-sealed range of 6–12 envelopes.

PR11 • From a prediction to a recommendation

1. For the four-mile trip, the model predicts 13 minutes within the recorded 2–5-mile range, leaving 1 minute before the gate locks. A comparable trip ran 3 minutes over its prediction, so I would request a longer slot for this deadline.

2. The log ends at five miles. Record comparable trips around seven miles under similar conditions, then check their actual times against the model's predictions before scheduling the longer trip.

PR12 • From one sample to a distribution of estimates

1. Three of the five draws gave a pretzel because 0.60 × 5 = 3. Packets were replaced, so the same pretzel packet could have appeared more than once.

2. The plot has one dot at 0.40 and two stacked dots at 0.60, one for each sample of five draws. The population proportion stays at 8/20 = 0.40 because each packet is returned.

PR13 • Why sample size changes precision

1. For 10 deliveries the width is 90% − 70% = 20 points. For 40 it is 85% − 75% = 10 points. The middle half of the larger-sample estimates is 10 percentage points narrower.

2. The ten-delivery estimate is 80%, exactly the model rate. The forty-delivery estimate is 70%, ten points away. One pair can behave this way even though larger samples usually show less spread across many repetitions.

PR14 • Confidence belongs to the procedure

1. The 12%–27% interval missed the known 30% share. Its upper endpoint was 3 percentage points below 30%. In this simulation the true share was set in advance.

2. The observed capture rate is 23/25 = 92%. The population share stayed at 30%, while new random samples produced different intervals. A 95% procedure can have a batch capture rate different from 95%.

RESPONSE GUIDE • CONTINUED

Check after your first attempt

PR15 • Use the interval, not just its center

1. We are 95% confident that about 46% to 72% of customers on the bakery's current list would answer yes when asked if they would order a smaller loaf.

2. The 45% bakery would proceed because even the lower endpoint, about 46%, exceeds 45%. The 50% bakery would wait because 46% falls below 50%. Both rules use the full interval, while 60% is a single sample estimate.

PR16 • The sampling frame sets the reach

1. The survey can describe support among current adult members on the register, under suitable response conditions. Children or nonmembers who might visit could not be selected and are outside that group.

2. Five of six visits had acceptable waits, about 83%. One of two customers always reported an acceptable wait, or 50%. The first calculation counts five visits by the same person. The second counts each person once.

PR17 • Ask what the benchmark could produce

1. The batch size stays at 30, draws are independent, and each run has a 10% chance of being late. The number of late runs in a batch can change.

2. $(8 + 3)/(8 + 3 + 189) = 11/200 = 0.055$. Count matches and larger counts; each of the 200 batches contributes one result.

PR18 • Size of a change and strength of evidence

1. Both samples are 35% defective, 15 percentage points above 20%. Only 14 of 40 meets the rule: 0.019 is below 0.05; 0.087 for 7 of 20 exceeds it.

2. The observed excess is 10 percentage points. Since 0.08 exceeds 0.05, the sample gives insufficient evidence of an increase under this rule. The true rate could still have risen.

PR19 • A decision can be wrong in two ways

1. This is a Type II error. The greenhouse really was too hot, and the sensor missed it. Plants could remain in damaging heat while staff receive no warning.

2. No. A Type I error requires a safe temperature when an alarm sounded. Independent temperature records at the two alarm times would identify which, if any, were false alarms.

PR20 • Power is detection under a stated alternative

1. 300 of the 1,000 batches generated at 25% went unflagged, or 30%. This estimates how often the rule misses that particular increase.

2. Power at 15% is about 80/200 = 40%. False alarms must be estimated from batches generated at the 10% benchmark, where a flag would be mistaken.

PR21 • Choose a rule with the costs in view

1. Only B qualifies: its 3% false-alarm rate is below 4%, whereas A's is 6%. At the 25% alternative, B misses about 100% − 71% = 29% of batches.

2. Both qualify under an inclusive 6% cap. A has higher power at the stated alternative, 82% versus 71%. At the benchmark, A would produce about 6 false alarms per 100 batches versus 3 for B.

RESPONSE GUIDE • CONTINUED

Check after your first attempt

PR22 • Two percentages with the same numerator

1. The machine found 8/10 = 80% of the defective parts. Among the 12 flags, 8/12 (about 67%) identified defects.

2. Of the ten flagged parts, six were defective, so the share defective among flags is 60%. The number of defective unflagged parts would be unknown, leaving the detection rate undetermined.

PR23 • Plan the evidence before collecting it

1. Use the complete current registration list, with one entry per family. Last year's attendance sheet could omit a family that registered this year or one that did not attend last year.

2. Yes. The full 51%–67% interval exceeds 50%, so the rule supports reserving the room. The outcome is a family's stated intention to attend. Actual turnout must be observed.

PR24 • Choose the output that answers the question

1. The lines inside the boxes mark medians of 8 minutes for A and 11 for B. Counter A's median for the recorded visits was 3 minutes shorter.

2. The 95% interval puts the share of visits finished within five minutes at 17%–24%. Even the lower endpoint exceeds 15%, so the new sample supports meeting the target.

PR25 • Revise the reasoning, then transfer it

1. Of the 24 weekend visitors surveyed at that trailhead, 15 (62.5%) said they wanted more picnic tables.

2. No. Residents who never use the trail could not appear in the visitor frame. A random sample from a complete list of town residents, with appropriate follow-up, could support an estimate of support among residents.

FORMULA APPENDIX / DESCRIPTIVE STATISTICS

Summarize a group or compare positions

Predictions and proportions W160 • Samples and intervals W161 • Tests and errors W162

Find the mean or median

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Each x_i is one observation and n is the number of observations. The summation sign means to add them. To find the median, order the values. Use the middle value for an odd count or the mean of the two middle values for an even count.

For waits of 2, 3, 3, 5, and 12 minutes, the mean is 25/5 = 5 minutes and the median is 3 minutes.

Compare spread

$$\text{range} = \text{maximum} - \text{minimum} \qquad \text{IQR} = Q_3 - Q_1$$

The range spans all the data. The interquartile range spans the middle half. For this course, split ordered data into halves, omitting the middle observation if n is odd. Q_1 and Q_3 are the medians of the lower and upper halves.

For the same waits, $Q_1 = 2.5$ and $Q_3 = 8.5$ minutes. The IQR is 6 minutes and the range is 10 minutes.

Compare one observation with its own group

$$z = \frac{x - \mu}{\sigma} \qquad \text{or} \qquad z = \frac{x - \bar{x}}{s}$$

Use the reference group's mean and standard deviation (SD). The population values are μ and σ. For a sample use its mean and sample SD, s. The SD must be positive. A z-score gives a signed distance in SD units. To recover the raw value, add z times the reference SD to the reference mean.

A 16-minute trip compared with a mean of 20 and an SD of 2 has z = −2. Its time is two SDs below that group's mean.

> **WHICH SUMMARY FITS?**
>
> If one very long wait stretches the right tail, report the median and IQR to describe a typical wait and the middle half. Also report the long wait if it matters to the decision. For a roughly symmetric distribution without extremes, the mean and SD can be useful.

Use with HW03 for line times and HW04 for comparing relay legs.

FORMULA APPENDIX / PREDICTION AND PROBABILITY

Predict a response or choose a denominator

Predict a response or a change

$$\hat{y} = a + bx \qquad \hat{y}_2 - \hat{y}_1 = b(x_2 - x_1)$$

Here x is the input, ŷ is the predicted response, a is the intercept, and b is the change in predicted response per unit of x. Check the input range and conditions used to fit the line.

At 2 minutes per mile, increasing a route from 3 to 5 miles adds 2(5 − 3) = 4 predicted minutes.

Check a prediction against an observation

$$e = y - \hat{y}$$

The residual e compares the observed response y with its prediction at the same input. It has the response units. A positive residual means the observation was larger. Use its absolute value to compare error sizes.

If a model predicts 12 minutes and the trip takes 10, then e = −2 minutes. The model overestimated the time by 2 minutes.

Calculate a sample share or pool counts

$$\hat{p} = \frac{x}{n} \qquad \hat{p}_{\text{pooled}} = \frac{x_1 + x_2}{n_1 + n_2}$$

Here x is the number in the category being counted and n is the number of eligible cases. To pool nonoverlapping groups measuring the same outcome, add the counts and add the denominators. Multiply a proportion by 100 to report a percentage.

If 1 of 2 jobs and 2 of 8 jobs finish early, the pooled result is 3 of 10 jobs, or 30%. A simple average of 50% and 25% gives 37.5% and weights the jobs unequally.

Read "among" or "given" in a two-way table

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

B is the condition and A is the outcome. With counts, divide the number meeting both descriptions by the number meeting B. This requires a nonzero denominator.

Eight of 12 damaged items were flagged. If there were 20 flags, detection was 8/12, while the share of flags that identified damage was 8/20.

Use with HW08 for cycling-time predictions, HW12 to pool counts, and HW22 to interpret scanner results.

FORMULA APPENDIX / SAMPLING AND ESTIMATION

Describe repeated samples or read an interval

Compare the spread of sample proportions

$$\mu_{\hat{p}} = p \qquad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Here p is the fixed model probability of a success, and n is the number of independent trials in each sample. The standard deviation here describes the sample proportions over repeated samples. Use a model with the same p on every trial.

At the same p, increasing the number of trials from 10 to 40 halves this SD. One larger sample can still land farther from p.

Check whether intervals capture a known value

$$L \le p \le U \qquad \text{observed coverage} = \frac{\text{intervals capturing } p}{\text{intervals made}}$$

L and U are the endpoints of one interval. Include equality. In a simulation the true p is known, so you can check captures. Confidence describes the long-run performance of the procedure under its assumptions.

If 18 of 20 intervals contain the simulation's true proportion, that batch has 90% coverage. A fresh batch may give a different percentage.

Interpret supplied interval endpoints

$$\text{interval width} = U - L$$

Use the supplied Wilson interval for the population proportion. Describe its population, outcome, confidence level, and both endpoints. When using percentages, report widths in percentage points. Compare the relevant endpoint with the stated decision rule.

For endpoints 54% and 66%, the width is 12 percentage points. A requirement that the lower endpoint exceed 60% is not met.

In a survey, check the sampling frame, selection and response before extending a conclusion to a population. An interval describes sampling uncertainty under its assumptions.

Use with HW13 for sampling variation, HW14 for repeated intervals, and HW15 for a policy threshold.

FORMULA APPENDIX / INFERENCE AND DECISIONS

Judge evidence and the cost of a decision

Estimate a p-value from a simulation

$$\text{estimated tail probability} = \frac{\text{as or more extreme batches}}{\text{benchmark batches simulated}}$$

Use the alternative chosen before collecting data to decide which simulated results count as extreme. Include ties. For a proposed increase, count batches at or above the observed result. At the chosen significance level α, the course rule is significant when the p-value is at most α.

With 9 qualifying batches out of 300, the simulated p-value is about 0.030. The denominator counts batches, not the individual observations inside them.

Report the size of an observed difference

$$\text{observed excess in percentage points} = 100(\hat{p} - p_0)$$

Here $\hat{p}$ is the observed sample proportion and p_0 is the benchmark, both written as decimals. The sign gives the direction. Use the practical threshold or consequences in the question to judge importance.

A sample rate of 0.26 against a benchmark of 0.20 gives an excess of 6 percentage points. This calculation alone does not give a p-value.

Count false alarms, detections, and misses

$$\text{power} = 1 - \beta \qquad \text{expected count} = N \times \text{probability}$$

At the benchmark, the false-alarm rate is the fraction of batches flagged (Type I errors). At a stated alternative, power is the fraction flagged and β is the fraction missed (Type II errors). N denotes the number of comparable batches under the stated condition.

At a specified alternative, 80% power implies about 200 misses per 1,000 batches. The false-alarm rate requires separate batches generated at the benchmark.

Use with HW17 to estimate a tail probability, HW18 to compare significance and importance, and HW20–21 to evaluate a flagging rule.

Organization and notation follow the College Board AP Statistics Exam Reference Information. Formulas, explanations, and examples were selected for St. Mary's Applied Math.