RATIOMERA

Complete Solutions

Probability

Ratiomera Statistics

These complete solutions use the same identifiers and order as the Exercise Sheet. Intermediate values are retained until the stated rounding step, so small differences caused by earlier rounding are acceptable where noted. All settings, values, data, and software outputs are constructed teaching material; they are not empirical findings.

Part I: Theory

A08: Probability Mass Functions and Densities

`T02-A08-V01`: From the number of exhibitions visited to the time spent in a museum

Identify the issue, part (a)

Because X records the number of exhibitions visited, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the time spent in a museum, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \leq b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \leq x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \leq y)$ accumulates area continuously under the density.

`T02-A08-V02`: From the number of messages received to the delay until the next message

Identify the issue, part (a)

Because X records the number of messages received, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the delay until the next message, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \leq b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \le x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \le y)$ accumulates area continuously under the density.

`T02-A08-V03`: From the count of transcription errors to the duration of an audio segment

Identify the issue, part (a)

Because X records the count of transcription errors, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the duration of an audio segment, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \le b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \le x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \le y)$ accumulates area continuously under the density.

`T02-A08-V04`: From the number of books borrowed to the mass of a returned parcel

Identify the issue, part (a)

Because X records the number of books borrowed, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the mass of a returned parcel, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \le b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \le x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \le y)$ accumulates area continuously under the density.

`T02-A08-V05`: From the count of survey reminders to a respondent's completion time

Identify the issue, part (a)

Because X records the count of survey reminders, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is a respondent's completion time, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \le b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \le x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \le y)$ accumulates area continuously under the density.

`T02-A08-V06`: From the number of route changes to the distance traveled
Identify the issue, part (a)

Because X records the number of route changes, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the distance traveled, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \le b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \le x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \le y)$ accumulates area continuously under the density.

`T02-A08-V07`: From the count of missing fields to a participant's age measured precisely
Identify the issue, part (a)

Because X records the count of missing fields, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is a participant's age measured precisely, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \le b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \le x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \le y)$ accumulates area continuously under the density.

`T02-A08-V08`: From the number of workshop sessions to the sound level in the room
Identify the issue, part (a)

Because X records the number of workshop sessions, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the sound level in the room, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \leq b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \leq x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \leq y)$ accumulates area continuously under the density.

`T02-A08-V09`: From the count of preserved photographs to the temperature of the archive
Identify the issue, part (a)

Because X records the count of preserved photographs, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the temperature of the archive, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \leq b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \leq x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \leq y)$ accumulates area continuously under the density.

`T02-A08-V10`: From the number of successful checks to the exact reaction time on a task
Identify the issue, part (a)

Because X records the number of successful checks, it has countable support and a PMF can assign mass $P(X = x)$ to each possible count. Because Y is the exact reaction time on a task, an ideal continuous model represents it with a density $f_Y(y)$.

Reason through the evidence, part (b)

$P(X = x)$ may be positive for one count, while $P(Y = y) = 0$ at every exact point even when nearby values are plausible.

Reason through the evidence, part (c)

For X, an interval probability is the sum of its included masses. For Y, it is density area, for example $P(a < Y \leq b) = \int_a^b f_Y(y)\,dy$.

State the conclusion and its limits, part (d)

In either case, a CDF records accumulated probability: $F_X(x) = P(X \leq x)$ jumps at supported counts, whereas $F_Y(y) = P(Y \leq y)$ accumulates area continuously under the density.

A14: Population, Sample, and Selection Bias

`T02-A14-V01`: Park-use QR survey
Identify the issue, part (a)

The target population is all city residents whose weekly use of any park is of interest.

Reason through the evidence, part (b)

The operational sampling frame is people entering the largest central park during the posting period who notice the code. The achieved sample is the 640 park visitors who scanned the code and submitted the survey. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the proportion of all city residents who use any park weekly.

State the conclusion and its limits, part (d)

The sample statistic is the proportion of the 640 respondents who report weekly park use. The main threats are specific to this design: Frequent park users are more likely to enter this park, and people who notice and scan a code may differ in interest or digital access from those who do not. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to draw a probability sample from a city-resident frame and follow up selected nonrespondents through more than one contact mode.

`T02-A14-V02`: Commuting survey of parking-permit holders

Identify the issue, part (a)

The target population is all enrolled students.

Reason through the evidence, part (b)

The operational sampling frame is the university's list of students holding parking permits. The achieved sample is the 820 parking-permit holders who replied. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the mean commuting time among all enrolled students.

State the conclusion and its limits, part (d)

The sample statistic is the mean commuting time among the 820 respondents. The main threats are specific to this design: The frame omits students who walk, cycle, use transit, or do not hold a permit, and reply behavior may depend on commuting burden. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to sample from the full enrollment register, stratify by likely travel mode if useful, and pursue responses from the selected students.

`T02-A14-V03`: Satisfaction after one sold-out exhibition

Identify the issue, part (a)

The target population is all museum visitors during the target season.

Reason through the evidence, part (b)

The operational sampling frame is visitors leaving the sold-out evening exhibition who were offered the exit survey. The achieved sample is the 510 attendees who completed that exit survey. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the mean satisfaction score among all visitors in the target season.

State the conclusion and its limits, part (d)

The sample statistic is the mean satisfaction score among the 510 respondents. The main threats are specific to this design: One unusually popular evening may not represent other dates or exhibitions, and survey completion may depend on whether an attendee had a notably good or bad experience. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to select visits across exhibitions, days, and times, then invite a probability sample of exiting visitors and document nonresponse.

`T02-A14-V04`: Digital-access survey inside an app

Identify the issue, part (a)

The target population is all library users.

Reason through the evidence, part (b)

The operational sampling frame is library users who use the smartphone app and were exposed to the survey notice. The achieved sample is the 430 app users who volunteered an answer. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the proportion of all library users who need better digital access.

State the conclusion and its limits, part (d)

The sample statistic is the proportion of the 430 respondents reporting that need. The main threats are specific to this design: Users without suitable devices or app access cannot enter the frame, and volunteers who answer an access survey may have unusually strong needs or engagement. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to sample from the full user register and offer accessible web, telephone, paper, and in-person response routes.

`T02-A14-V05`: Volunteer hours from large-charity lists

Identify the issue, part (a)

The target population is all volunteers in the region.

Reason through the evidence, part (b)

The operational sampling frame is membership lists supplied by large registered charities. The achieved sample is the 760 listed charity members whose records were used. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the mean weekly volunteer hours among all regional volunteers.

State the conclusion and its limits, part (d)

The sample statistic is the mean weekly hours recorded for those 760 listed members. The main threats are specific to this design: The lists exclude informal volunteers and members of small or unregistered groups, and formal membership records may overrepresent regular, long-term volunteers. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to build a broader frame from multiple organization types and community sources, then use probability selection within defined volunteer strata.

T02-A14-V06: Course-workload survey after grades
Identify the issue, part (a)

The target population is all students enrolled in the course.

Reason through the evidence, part (b)

The operational sampling frame is enrolled students whose accounts remained active on the platform after grades were released. The achieved sample is the 390 still-active students who supplied workload information. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the mean perceived workload among all enrolled students.

State the conclusion and its limits, part (d)

The sample statistic is the mean workload reported by the 390 respondents. The main threats are specific to this design: Students who disengaged, withdrew, or stopped using the platform are excluded, and willingness to answer after grading may be related to workload or course outcomes. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to select from the original course roster, contact students independently of later platform activity, and follow up nonrespondents.

T02-A14-V07: Transit delays inferred from hashtagged comments
Identify the issue, part (a)

The target population is all rider trips during the target period.

Reason through the evidence, part (b)

The operational sampling frame is publicly retrievable social-media comments that use the campaign hashtag. The achieved sample is the 1 240 retrieved hashtagged comments. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the proportion of all rider trips experienced as delayed.

State the conclusion and its limits, part (d)

The sample statistic is the proportion of the 1 240 comments that describe a delay. The main threats are specific to this design: People with extreme experiences are more likely to post, one person can contribute several comments, and comments rather than rider trips are the observed units. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to sample rider trips from operational records and obtain a response tied to each selected trip, while keeping the trip as the unit of analysis.

T02-A14-V08: Neighborhood-interest forms after performances
Identify the issue, part (a)

The target population is all residents in the surrounding neighborhood.

Reason through the evidence, part (b)

The operational sampling frame is ticket holders leaving the selected performances who were offered a form. The achieved sample is the 570 ticket holders who stayed and completed a form. Keeping these separate matters

because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the proportion of neighborhood residents interested in future programs.

State the conclusion and its limits, part (d)

The sample statistic is the proportion of the 570 respondents expressing interest. The main threats are specific to this design: Residents who do not already attend ticketed events are absent from the frame, and staying to complete a form may be related to enthusiasm for the center. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to use a neighborhood address frame, select residents independently of attendance, and offer several response modes.

`T02-A14-V09`: Sleep records from year-long wearable users

Identify the issue, part (a)

The target population is all users in the intended wearable-user population during the year.

Reason through the evidence, part (b)

The operational sampling frame is device users with accounts activated at the beginning of the observation period. The achieved sample is the 680 users retained for a full year with complete sleep records. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the population mean nightly sleep duration.

State the conclusion and its limits, part (d)

The sample statistic is the retained users' mean recorded nightly duration. The main threats are specific to this design: Year-long retention excludes intermittent or discontinued users, and retention or complete wear may depend on sleep habits, health, or satisfaction with the device. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to select users at enrollment, retain partial records under a prespecified missing-data plan, and compare retained with lost participants.

`T02-A14-V10`: Archive feedback shown only after a download

Identify the issue, part (a)

The target population is all archive search attempts during the target period.

Reason through the evidence, part (b)

The operational sampling frame is search attempts that reached at least one download and therefore received the feedback prompt. The achieved sample is the 450 submitted feedback forms from that restricted frame. Keeping these separate matters because the frame describes who or what had a route to selection, whereas the sample contains the units actually observed.

Reason through the evidence, part (c)

A matching population parameter is the proportion of all search attempts ending in successful retrieval.

State the conclusion and its limits, part (d)

The sample statistic is the proportion of the 450 sampled forms whose submitters report success. The main threats are specific to this design: The prompt appears only after a post-search success event, so failed searches have no route into the frame; response among downloaders can add further self-selection. A larger sample obtained through the same mechanism would reduce random sampling variability around that mechanism's frame-specific value, but it would not repair the systematic coverage or selection mechanisms just identified. A more defensible approach is to sample search attempts at initiation, invite feedback whether or not a download occurs, and link one response opportunity to each selected attempt.

A15: Coverage Error and the Population Behind a Percentage

`T02-A15-V01`: Education levels among football supporters

Identify the issue, part (a)

The broad claim names all people who support Northport FC.

Reason through the evidence, part (b)

The units with a route into the calculation are platform members who name Northport FC on a visible profile and report education information.

Reason through the evidence, part (c)

This creates coverage error because supporters who do not use the professional platform, do not name a club, or omit education have no route into the percentage. Platform membership is also related to education and employment. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among the analyzed platform profiles that named Northport FC and reported education, 64% reported a university degree." A more defensible study would define supporter status first, sample supporters through a frame that is not tied to professional-platform membership, and follow up selected nonrespondents. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V02`: Reading habits from an e-reader community

Identify the issue, part (a)

The broad claim names all adults in the country.

Reason through the evidence, part (b)

The units with a route into the calculation are members of the e-reader forum who saw the invitation and chose to respond.

Reason through the evidence, part (c)

This creates coverage error because adults who do not use e-readers or the forum are absent, and highly engaged readers are especially likely to join and answer. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among responding members of this e-reader forum, 71% reported finishing at least two books per month." A more defensible study would draw a probability sample from a population-based adult frame and offer several response modes. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V03`: Cycling frequency from a route-planning app
Identify the issue, part (a)

The broad claim names all residents of the city.

Reason through the evidence, part (b)

The units with a route into the calculation are active users of the cycling app who allow ride recording.

Reason through the evidence, part (c)

This creates coverage error because residents who do not cycle, do not use the app, or disable recording are missing, while frequent cyclists are more likely to remain active users. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: “Among active app users with ride recording enabled, 58% recorded at least three rides per week.” A more defensible study would sample residents from a city register and measure cycling frequency whether or not they use an app. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V04`: Museum interest from newsletter subscribers
Identify the issue, part (a)

The broad claim names all residents of the region.

Reason through the evidence, part (b)

The units with a route into the calculation are museum newsletter subscribers who opened the message and completed the survey.

Reason through the evidence, part (c)

This creates coverage error because people already interested in the museum are more likely to subscribe, open the message, and answer the survey. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: “Among newsletter subscribers who answered, 82% said that they planned to visit the exhibition.” A more defensible study would sample regional residents independently of newsletter subscription and record nonresponse. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V05`: Remote-work preference from a coworking platform
Identify the issue, part (a)

The broad claim names all employed adults in the target region.

Reason through the evidence, part (b)

The units with a route into the calculation are account holders on the coworking platform who received and answered its poll.

Reason through the evidence, part (c)

This creates coverage error because the platform overrepresents people whose jobs can be done remotely, and volunteers with strong preferences may answer more often. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among responding account holders on this coworking platform, 76% preferred remote work on most weekdays." A more defensible study would sample employed adults across occupations and work arrangements from a suitable labor-force frame. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V06`: Language use from public profile fields

Identify the issue, part (a)

The broad claim names all residents of the country.

Reason through the evidence, part (b)

The units with a route into the calculation are platform members with public profiles who chose to list at least one language.

Reason through the evidence, part (c)

This creates coverage error because platform access and public-profile choices differ across residents, and listing a language does not establish daily use. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Of the public profiles with a language field in the analyzed platform data, 43% listed at least three languages." A more defensible study would sample residents through a population frame and ask a clearly defined question about everyday language use. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V07`: Student wellbeing from a study-planner app

Identify the issue, part (a)

The broad claim names all students enrolled at the universities of interest.

Reason through the evidence, part (b)

The units with a route into the calculation are users of the study-planner app who noticed and answered the wellbeing prompt.

Reason through the evidence, part (c)

This creates coverage error because students who use a planning app may differ in workload or organization, and willingness to answer may be connected with current stress. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among app users who answered the prompt, 61% reported high academic stress." A more defensible study would sample from complete enrollment lists and contact the selected students through more than one mode. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V08`: Concert attendance from ticket-account profiles

Identify the issue, part (a)

The broad claim names all residents in the target population.

Reason through the evidence, part (b)

The units with a route into the calculation are registered ticketing accounts with observable page-following activity.

Reason through the evidence, part (c)

This creates coverage error because residents without accounts are absent, one person can hold several accounts, and following a page is not the same outcome as attending a concert. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among the observed ticketing accounts, 67% followed at least one concert page last year." A more defensible study would sample people rather than accounts and ask or verify a clearly defined attendance outcome. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V09`: Public-transport satisfaction from a mobile-ticket sample

Identify the issue, part (a)

The broad claim names all passengers using the transport system in the target period.

Reason through the evidence, part (b)

The units with a route into the calculation are passengers who purchased a mobile ticket and received the in-app question.

Reason through the evidence, part (c)

This creates coverage error because cash, paper-ticket, pass, and accessibility-service users cannot enter the frame, and satisfaction may influence whether the prompt is answered. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among mobile-ticket users who answered the in-app question, 74% reported being satisfied." A more defensible study would sample trips across ticket types, routes, and times, then invite the selected passengers through accessible response routes. Increasing the number of records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

`T02-A15-V10`: Volunteer participation from organization websites

Identify the issue, part (a)

The broad claim names all formal and informal volunteers in the region.

Reason through the evidence, part (b)

The units with a route into the calculation are volunteers publicly listed by the large charities included in the web search.

Reason through the evidence, part (c)

This creates coverage error because informal volunteers, small organizations, and volunteers without public profiles are missing, while regular contributors are more likely to be featured. The percentage can be calculated correctly for the observed records and still fail to estimate the percentage in the broader population.

State the conclusion and its limits, part (d)

An honest descriptive statement is: "Among volunteers publicly listed by the included large charities, 69% were described as contributing every month." A more defensible study would construct a broader frame across organization sizes and informal community work, then sample volunteers within it. Increasing the number of

records drawn through the same restricted route would make the restricted-frame percentage more precise, but would not add the kinds of people who never entered that frame.

A16: Survivorship Bias and Missing Outcomes

`T02-A16-V01`: Damage patterns on returned delivery drones

Identify the issue, part (a)

The observed group contains drones that were damaged yet still returned and could be inspected.

Reason through the evidence, part (b)

Missing from that group are drones that failed to return, including any whose navigation units were critically damaged.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Damage to a navigation unit may prevent return, so the low observed mark count there can signal severe selection rather than safety. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to investigate failed-flight logs and recovered nonreturning drones before deciding where reinforcement is most valuable. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V02`: Study habits among course completers

Identify the issue, part (a)

The observed group contains enrollees who remained through completion and agreed to be interviewed.

Reason through the evidence, part (b)

Missing from that group are students who withdrew, stopped logging in, or did not agree to the interview.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Planning habits can be connected with persistence, so conditioning on completion can make the observed habit unusually common. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to follow the original enrollment cohort and collect comparable information from completers and noncompleters. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V03`: Reliability among devices still in service

Identify the issue, part (a)

The observed group contains sensors that survived in service for two years and were still available for inspection.

Reason through the evidence, part (b)

Missing from that group are sensors removed, discarded, or replaced earlier, possibly because corrosion caused failure.

Reason through the evidence, part (c)

The selection process is tied to the outcome: The outcome of interest can determine whether a sensor remains observable, leaving the least damaged units in the inspected group. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to use maintenance and replacement records for the full original sensor cohort, including failed units. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V04`: Satisfaction among returning museum visitors
Identify the issue, part (a)

The observed group contains visitors who were satisfied or motivated enough to return at least four times and attend again.

Reason through the evidence, part (b)

Missing from that group are one-time visitors and people who decided not to return.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Past satisfaction can influence return, so selecting on a later visit filters out many less satisfied experiences. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to sample first-time visits and follow those visitors whether or not they return. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V05`: Workload reports from retained employees
Identify the issue, part (a)

The observed group contains employees from the hiring cohort who remained for five years and responded.

Reason through the evidence, part (b)

Missing from that group are employees who resigned, were dismissed, or could not be contacted after leaving.

Reason through the evidence, part (c)

The selection process is tied to the outcome: First-year workload can affect leaving, so the retained employees may systematically report different experiences. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to collect workload information prospectively from the full hiring cohort and retain exit information. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V06`: Recovery among follow-up clinic attendees
Identify the issue, part (a)

The observed group contains treated patients who attended the final follow-up and supplied an outcome.

Reason through the evidence, part (b)

Missing from that group are patients who missed follow-up because they worsened, recovered elsewhere, moved, or disengaged.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Follow-up attendance may depend on recovery, so the observed percentage need not represent all treated patients. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to trace the full treated cohort and use several appropriate ways to obtain outcomes from missed appointments. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V07`: Durability among surviving archive files

Identify the issue, part (a)

The observed group contains files that survived, remained discoverable, and could still be opened.

Reason through the evidence, part (b)

Missing from that group are lost, corrupted, or undiscoverable files whose metadata may have contributed to their disappearance.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Requiring a file to be findable and open can remove precisely the failures needed to assess preservation. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to audit the original file inventory and count missing and corrupted files as outcomes rather than excluding them. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V08`: Confidence among competition finalists

Identify the issue, part (a)

The observed group contains entrants who advanced through every earlier round and reached the final.

Reason through the evidence, part (b)

Missing from that group are entrants eliminated in earlier rounds or who withdrew.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Initial confidence can influence performance and withdrawal, so finalists form a selected subset of entrants. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to measure confidence for all entrants

before the first round and retain their later competition status. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V09`: Travel times from completed app routes

Identify the issue, part (a)

The observed group contains recorded trips that remained active until the app registered completion.

Reason through the evidence, part (b)

Missing from that group are interrupted, abandoned, or exceptionally delayed trips whose app sessions ended early.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Long or problematic trips may be more likely to be closed early, which can make completed routes look faster. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to define every started trip as part of the cohort and investigate incomplete route records instead of silently dropping them. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

`T02-A16-V10`: Reading progress among active subscribers

Identify the issue, part (a)

The observed group contains subscribers who remained active for the full year and had readable progress data.

Reason through the evidence, part (b)

Missing from that group are people who canceled or whose accounts became inactive during the year.

Reason through the evidence, part (c)

The selection process is tied to the outcome: Reading engagement can affect cancellation, so active subscribers may display unusually high progress. This is survivorship bias, meaning that continued availability is required for observation even though failure to remain available can itself carry important information.

State the conclusion and its limits, part (d)

Looking only at the observed cases conditions the analysis on survival, completion, return, or retention. It can therefore hide failures and reverse the practical lesson. The next step is to retain the original subscriber cohort in the analysis and record progress up to cancellation or follow up former subscribers. The goal is not to guess the missing results, but to redesign collection so both continuing and noncontinuing cases contribute evidence.

Part II: Calculator Practice

A01: Sequential Conditional Probability

`T02-A01-V01`: Completing an archive search

Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.62 \times 0.81 = 0.5022$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.62 \times 0.81 \times 0.74 = 0.3716$. Thus the model says that a proportion 0.3716 will complete the entire sequence ending with "identify the relevant letter."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.74 = 0.26$. Therefore $P(A \cap B \cap C') = 0.62 \times 0.81 \times 0.26 = 0.1306$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

`T02-A01-V02`: Passing a three-stage language assessment

Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.68 \times 0.77 = 0.5236$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.68 \times 0.77 \times 0.84 = 0.4398$. Thus the model says that a proportion 0.4398 will complete the entire sequence ending with "pass the interview."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.84 = 0.16$. Therefore $P(A \cap B \cap C') = 0.68 \times 0.77 \times 0.16 = 0.0838$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

`T02-A01-V03`: Finishing a digital registration

Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.73 \times 0.86 = 0.6278$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.73 \times 0.86 \times 0.79 = 0.4960$. Thus the model says that a proportion 0.4960 will complete the entire sequence ending with "submit the consent form."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.79 = 0.21$. Therefore $P(A \cap B \cap C') = 0.73 \times 0.86 \times 0.21 = 0.1318$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

`T02-A01-V04`: Solving a fieldwork sequence

Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.57 \times 0.83 = 0.4731$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.57 \times 0.83 \times 0.91 = 0.4305$. Thus the model says that a proportion 0.4305 will complete the entire sequence ending with "upload the record correctly."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.91 = 0.09$. Therefore $P(A \cap B \cap C') = 0.57 \times 0.83 \times 0.09 = 0.0426$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

`T02-A01-V05`: Completing a library research task
Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.66 \times 0.72 = 0.4752$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.66 \times 0.72 \times 0.88 = 0.4182$. Thus the model says that a proportion 0.4182 will complete the entire sequence ending with "evaluate its methods accurately."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.88 = 0.12$. Therefore $P(A \cap B \cap C') = 0.66 \times 0.72 \times 0.12 = 0.0570$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

`T02-A01-V06`: Advancing through a music audition
Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.71 \times 0.69 = 0.4899$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.71 \times 0.69 \times 0.82 = 0.4017$. Thus the model says that a proportion 0.4017 will complete the entire sequence ending with "pass the sight-reading task."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.82 = 0.18$. Therefore $P(A \cap B \cap C') = 0.71 \times 0.69 \times 0.18 = 0.0882$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

`T02-A01-V07`: Completing a laboratory protocol
Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.64 \times 0.87 = 0.5568$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.64 \times 0.87 \times 0.76 = 0.4232$. Thus the model says that a proportion 0.4232 will complete the entire sequence ending with "label the result correctly."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.76 = 0.24$. Therefore $P(A \cap B \cap C') = 0.64 \times 0.87 \times 0.24 = 0.1336$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

T02-A01-V08: Finishing an online safety course
Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.78 \times 0.75 = 0.5850$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.78 \times 0.75 \times 0.89 = 0.5206$. Thus the model says that a proportion 0.5206 will complete the entire sequence ending with "submit the final reflection."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.89 = 0.11$. Therefore $P(A \cap B \cap C') = 0.78 \times 0.75 \times 0.11 = 0.0643$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

T02-A01-V09: Completing a map-reading challenge
Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.59 \times 0.82 = 0.4838$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.59 \times 0.82 \times 0.85 = 0.4112$. Thus the model says that a proportion 0.4112 will complete the entire sequence ending with "identify the final landmark."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.85 = 0.15$. Therefore $P(A \cap B \cap C') = 0.59 \times 0.82 \times 0.15 = 0.0726$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

T02-A01-V10: Passing a data-entry check
Set up the calculation, part (a)

Let A, B, and C denote success at the three stages in order.

By the chain rule, $P(A \cap B) = P(A)P(B \mid A) = 0.69 \times 0.84 = 0.5796$.

Work through the calculation, part (b)

$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) = 0.69 \times 0.84 \times 0.73 = 0.4231$. Thus the model says that a proportion 0.4231 will complete the entire sequence ending with "submit the clean record."

Interpret and check the result, part (c)

Conditional on the first two successes, failure at the third stage has probability $1 - 0.73 = 0.27$. Therefore $P(A \cap B \cap C') = 0.69 \times 0.84 \times 0.27 = 0.1565$, the modeled proportion who reach the third stage but do not complete it. The later-stage probabilities refer to groups already restricted by earlier success, so replacing them with unrelated marginal probabilities would discard the conditions in the scenario.

A02: Independent Joint Events

T02-A02-V01: Two independent quality checks
Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.78 \times 0.64 = 0.4992$. This is the modeled probability that a scanned page passes the image check and the same page passes the metadata check. For

, the general addition rule gives $P(A \cup B) = 0.78 + 0.64 - 0.4992 = 0.9208$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.78(1 - 0.64) = 0.2808$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V02`: Independent workshop attendance
Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.55 \times 0.72 = 0.3960$. This is the modeled probability that a resident attends the morning session and the resident attends the evening session. For

, the general addition rule gives $P(A \cup B) = 0.55 + 0.72 - 0.3960 = 0.8740$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.55(1 - 0.72) = 0.1540$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V03`: Two independent sensor alerts
Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.18 \times 0.27 = 0.0486$. This is the modeled probability that the temperature sensor triggers and the vibration sensor triggers. For

, the general addition rule gives $P(A \cup B) = 0.18 + 0.27 - 0.0486 = 0.4014$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.18(1 - 0.27) = 0.1314$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V04`: Independent item selections
Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.36 \times 0.41 = 0.1476$. This is the modeled probability that a selected book is a translation and it has a hard cover. For

, the general addition rule gives $P(A \cup B) = 0.36 + 0.41 - 0.1476 = 0.6224$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.36(1 - 0.41) = 0.2124$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V05`: Independent survey events
Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.22 \times 0.63 = 0.1386$. This is the modeled probability that a response arrives on Monday and it is submitted from a mobile device. For

, the general addition rule gives $P(A \cup B) = 0.22 + 0.63 - 0.1386 = 0.7114$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.22(1 - 0.63) = 0.0814$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V06`: Two independent route events

Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.74 \times 0.58 = 0.4292$. This is the modeled probability that a bus arrives within five minutes and a connecting train has an available seat. For

, the general addition rule gives $P(A \cup B) = 0.74 + 0.58 - 0.4292 = 0.8908$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.74(1 - 0.58) = 0.3108$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V07`: Independent coding checks

Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.83 \times 0.69 = 0.5727$. This is the modeled probability that a record has a valid date and it has a valid category code. For

, the general addition rule gives $P(A \cup B) = 0.83 + 0.69 - 0.5727 = 0.9473$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.83(1 - 0.69) = 0.2573$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V08`: Two independent draws

Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.45 \times 0.32 = 0.1440$. This is the modeled probability that a selected token is blue and a second, replaced draw is triangular. For

, the general addition rule gives $P(A \cup B) = 0.45 + 0.32 - 0.1440 = 0.6260$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.45(1 - 0.32) = 0.3060$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V09`: Independent study events

Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.67 \times 0.88 = 0.5896$. This is the modeled probability that a participant completes the diary and the laboratory file uploads successfully. For

, the general addition rule gives $P(A \cup B) = 0.67 + 0.88 - 0.5896 = 0.9604$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.67(1 - 0.88) = 0.0804$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

`T02-A02-V10`: Independent catalog features
Set up the calculation, part (b)

Independence permits $P(A \cap B) = P(A)P(B) = 0.39 \times 0.76 = 0.2964$. This is the modeled probability that an item is digitized and its creator field is complete. For

, the general addition rule gives $P(A \cup B) = 0.39 + 0.76 - 0.2964 = 0.8536$, the probability that at least one of those events occurs. For

Interpret and check the result, part (c)

, independence of A and B also gives independence of A and B', so $P(A \cap B') = 0.39(1 - 0.76) = 0.0936$. Independence is used to replace joint probabilities with products; the addition rule itself is valid whether or not the events are independent.

A03: Contingency Tables and Event Relationships

`T02-A03-V01`: Reading format and course completion
Reason before calculating, part (a)

$P(Y \mid G) = 12/30 = 0.4000$ for the Audio row, while $P(Y \mid G^c) = 28/70 = 0.4000$ for the Text row. These conditional proportions are equal, so the variables are independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Audio and Completed contains 12 of the 100 observations, so $P(G \cap Y) = 12/100 = 0.1200$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

`T02-A03-V02`: Museum membership and event attendance
Reason before calculating, part (a)

$P(Y \mid G) = 24/40 = 0.6000$ for the Member row, while $P(Y \mid G^c) = 18/60 = 0.3000$ for the Nonmember row. These conditional proportions are different, so the variables are not independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Member and Attended contains 24 of the 100 observations, so $P(G \cap Y) = 24/100 = 0.2400$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

T02-A03-V03: Study location and deadline success

Reason before calculating, part (a)

$P(Y \mid G) = 21/35 = 0.6000$ for the Library row, while $P(Y \mid G^c) = 27/45 = 0.6000$ for the Home row. These conditional proportions are equal, so the variables are independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Library and On time contains 21 of the 80 observations, so $P(G \cap Y) = 21/80 = 0.2625$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

T02-A03-V04: Caption use and quiz completion

Reason before calculating, part (a)

$P(Y \mid G) = 30/50 = 0.6000$ for the Captions row, while $P(Y \mid G^c) = 18/30 = 0.6000$ for the No captions row. These conditional proportions are equal, so the variables are independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Captions and Completed contains 30 of the 80 observations, so $P(G \cap Y) = 30/80 = 0.3750$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

T02-A03-V05: Transit pass and campus visits

Reason before calculating, part (a)

$P(Y \mid G) = 16/40 = 0.4000$ for the Pass row, while $P(Y \mid G^c) = 14/60 = 0.2333$ for the No pass row. These conditional proportions are different, so the variables are not independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Pass and Frequent contains 16 of the 100 observations, so $P(G \cap Y) = 16/100 = 0.1600$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

T02-A03-V06: Reminder choice and response

Reason before calculating, part (a)

$P(Y \mid G) = 27/40 = 0.6750$ for the Reminder row, while $P(Y \mid G^c) = 33/60 = 0.5500$ for the No reminder row. These conditional proportions are different, so the variables are not independent in this empirical table.

This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Reminder and Responded contains 27 of the 100 observations, so $P(G \cap Y) = 27/100 = 0.2700$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

`T02-A03-V07`: Workshop track and certification
Reason before calculating, part (a)

$P(Y \mid G) = 18/30 = 0.6000$ for the Methods row, while $P(Y \mid G^c) = 24/60 = 0.4000$ for the Writing row. These conditional proportions are different, so the variables are not independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Methods and Certified contains 18 of the 90 observations, so $P(G \cap Y) = 18/90 = 0.2000$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

`T02-A03-V08`: Device type and form completion
Reason before calculating, part (a)

$P(Y \mid G) = 14/35 = 0.4000$ for the Tablet row, while $P(Y \mid G^c) = 30/45 = 0.6667$ for the Laptop row. These conditional proportions are different, so the variables are not independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Tablet and Complete contains 14 of the 80 observations, so $P(G \cap Y) = 14/80 = 0.1750$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

`T02-A03-V09`: Volunteer role and return visit
Reason before calculating, part (a)

$P(Y \mid G) = 22/40 = 0.5500$ for the Guide row, while $P(Y \mid G^c) = 11/40 = 0.2750$ for the Archive row. These conditional proportions are different, so the variables are not independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Guide and Returned contains 22 of the 80 observations, so $P(G \cap Y) = 22/80 = 0.2750$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

T02-A03-V10: Tutorial format and practice submission
Reason before calculating, part (a)

$P(Y \mid G) = 26/40 = 0.6500$ for the Live row, while $P(Y \mid G^c) = 39/60 = 0.6500$ for the Recorded row. These conditional proportions are equal, so the variables are independent in this empirical table. This is a descriptive statement about the displayed empirical distribution, not proof that the same relationship holds in a wider population.

Work through the calculation, part (b)

The intersection of Live and Submitted contains 26 of the 100 observations, so $P(G \cap Y) = 26/100 = 0.2600$.

Interpret and check the result, part (c)

G and Y are not disjoint because that intersection is nonempty. Disjoint events would have an intersection count of zero.

A04: Bayes' Theorem and Base Rates

T02-A04-V01: Accessibility-needs screening
Reason before calculating, part (a)

The false-positive probability is $1 - 0.91 = 0.09$.

At prevalence 0.02, the positive-result paths are $P(+ \cap D) = 0.72 \times 0.02 = 0.0144$ and $P(+ \cap D') = 0.09 \times 0.98 = 0.0882$. Thus $P(D \mid +) = 0.0144/(0.0144 + 0.0882) = 0.1404$.

Interpret and check the result, part (b)

At prevalence 0.11, the corresponding paths are 0.0792 and 0.0801, giving $P(D \mid +) = 0.0792/(0.0792 + 0.0801) = 0.4972$. This posterior is the chance that the condition described as "an accessibility support need" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

T02-A04-V02: Rare transcription-error detection
Reason before calculating, part (a)

The false-positive probability is $1 - 0.93 = 0.07$.

At prevalence 0.01, the positive-result paths are $P(+ \cap D) = 0.84 \times 0.01 = 0.0084$ and $P(+ \cap D') = 0.07 \times 0.99 = 0.0693$. Thus $P(D \mid +) = 0.0084/(0.0084 + 0.0693) = 0.1081$.

Interpret and check the result, part (b)

At prevalence 0.08, the corresponding paths are 0.0672 and 0.0644, giving $P(D \mid +) = 0.0672/(0.0672 + 0.0644) = 0.5106$. This posterior is the chance that the condition described as "a transcription error" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

T02-A04-V03: Conservation-risk screening
Reason before calculating, part (a)

The false-positive probability is $1 - 0.88 = 0.12$.

At prevalence 0.04, the positive-result paths are $P(+ \cap D) = 0.79 \times 0.04 = 0.0316$ and $P(+ \cap D') = 0.12 \times 0.96 = 0.1152$. Thus $P(D \mid +) = 0.0316/(0.0316 + 0.1152) = 0.2153$.

Interpret and check the result, part (b)

At prevalence 0.16, the corresponding paths are 0.1264 and 0.1008, giving $P(D \mid +) = 0.1264/(0.1264 + 0.1008) = 0.5563$. This posterior is the chance that the condition described as "an object that is at conservation

risk" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

`T02-A04-V04`: Duplicate-record detection
Reason before calculating, part (a)

The false-positive probability is $1 - 0.86 = 0.14$.

At prevalence 0.03, the positive-result paths are $P(+ \cap D) = 0.90 \times 0.03 = 0.0270$ and $P(+ \cap D') = 0.14 \times 0.97 = 0.1358$. Thus $P(D \mid +) = 0.0270/(0.0270 + 0.1358) = 0.1658$.

Interpret and check the result, part (b)

At prevalence 0.14, the corresponding paths are 0.1260 and 0.1204, giving $P(D \mid +) = 0.1260/(0.1260 + 0.1204) = 0.5114$. This posterior is the chance that the condition described as "a record that is a duplicate" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

`T02-A04-V05`: Language-support screening
Reason before calculating, part (a)

The false-positive probability is $1 - 0.94 = 0.06$.

At prevalence 0.05, the positive-result paths are $P(+ \cap D) = 0.76 \times 0.05 = 0.0380$ and $P(+ \cap D') = 0.06 \times 0.95 = 0.0570$. Thus $P(D \mid +) = 0.0380/(0.0380 + 0.0570) = 0.4000$.

Interpret and check the result, part (b)

At prevalence 0.19, the corresponding paths are 0.1444 and 0.0486, giving $P(D \mid +) = 0.1444/(0.1444 + 0.0486) = 0.7482$. This posterior is the chance that the condition described as "a need for language support" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

`T02-A04-V06`: Damaged-image detection
Reason before calculating, part (a)

The false-positive probability is $1 - 0.90 = 0.10$.

At prevalence 0.02, the positive-result paths are $P(+ \cap D) = 0.88 \times 0.02 = 0.0176$ and $P(+ \cap D') = 0.10 \times 0.98 = 0.0980$. Thus $P(D \mid +) = 0.0176/(0.0176 + 0.0980) = 0.1522$.

Interpret and check the result, part (b)

At prevalence 0.12, the corresponding paths are 0.1056 and 0.0880, giving $P(D \mid +) = 0.1056/(0.1056 + 0.0880) = 0.5455$. This posterior is the chance that the condition described as "an image that is damaged" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

`T02-A04-V07`: Research-integrity screening
Reason before calculating, part (a)

The false-positive probability is $1 - 0.96 = 0.04$.

At prevalence 0.01, the positive-result paths are $P(+ \cap D) = 0.81 \times 0.01 = 0.0081$ and $P(+ \cap D') = 0.04 \times 0.99 = 0.0396$. Thus $P(D \mid +) = 0.0081/(0.0081 + 0.0396) = 0.1698$.

Interpret and check the result, part (b)

At prevalence 0.07, the corresponding paths are 0.0567 and 0.0372, giving $P(D \mid +) = 0.0567/(0.0567 + 0.0372) = 0.6038$. This posterior is the chance that the condition described as "a submission that warrants

integrity review" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

T02-A04-V08: Equipment-failure warning
Reason before calculating, part (a)

The false-positive probability is $1 - 0.89 = 0.11$.

At prevalence 0.06, the positive-result paths are $P(+ \cap D) = 0.85 \times 0.06 = 0.0510$ and $P(+ \cap D') = 0.11 \times 0.94 = 0.1034$. Thus $P(D \mid +) = 0.0510/(0.0510 + 0.1034) = 0.3303$.

Interpret and check the result, part (b)

At prevalence 0.21, the corresponding paths are 0.1785 and 0.0869, giving $P(D \mid +) = 0.1785/(0.1785 + 0.0869) = 0.6726$. This posterior is the chance that the condition described as "an impending equipment failure" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

T02-A04-V09: Cataloging-anomaly detection
Reason before calculating, part (a)

The false-positive probability is $1 - 0.92 = 0.08$.

At prevalence 0.03, the positive-result paths are $P(+ \cap D) = 0.74 \times 0.03 = 0.0222$ and $P(+ \cap D') = 0.08 \times 0.97 = 0.0776$. Thus $P(D \mid +) = 0.0222/(0.0222 + 0.0776) = 0.2224$.

Interpret and check the result, part (b)

At prevalence 0.15, the corresponding paths are 0.1110 and 0.0680, giving $P(D \mid +) = 0.1110/(0.1110 + 0.0680) = 0.6201$. This posterior is the chance that the condition described as "a record that contains a cataloging anomaly" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

T02-A04-V10: Support-priority classification
Reason before calculating, part (a)

The false-positive probability is $1 - 0.95 = 0.05$.

At prevalence 0.04, the positive-result paths are $P(+ \cap D) = 0.87 \times 0.04 = 0.0348$ and $P(+ \cap D') = 0.05 \times 0.96 = 0.0480$. Thus $P(D \mid +) = 0.0348/(0.0348 + 0.0480) = 0.4203$.

Interpret and check the result, part (b)

At prevalence 0.18, the corresponding paths are 0.1566 and 0.0410, giving $P(D \mid +) = 0.1566/(0.1566 + 0.0410) = 0.7925$. This posterior is the chance that the condition described as "a case that genuinely has high support priority" is truly present after a positive result. Sensitivity instead conditions on the condition already being present. The higher base rate raises the share of positive results that are true positives.

A05: Discrete Expectation, Variance, PMF, and CDF

T02-A05-V01: Number of follow-up questions
Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.20+0.35+0.30+0.15=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 0(0.20) + 1(0.35) + 3(0.30) + 5(0.15) = 2.0000$. Across many comparable observations, the long-run average number of follow-up questions would approach 2.0000. Next, $E(X^2) = 0^2(0.20) + 1^2(0.35) + 3^2(0.30) + 5^2(0.15) = 6.8000$, so $\text{Var}\,(X) = 6.8000 - 2.0000^2 = 2.8000$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(0) = 0.20$, $F(1) = 0.55$, $F(3) = 0.85$, $F(5) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

`T02-A05-V02`: Daily archive requests

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.25+0.40+0.20+0.15=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 1(0.25) + 2(0.40) + 4(0.20) + 6(0.15) = 2.7500$. Across many comparable observations, the long-run average daily archive requests would approach 2.7500. Next, $E(X^2) = 1^2(0.25) + 2^2(0.40) + 4^2(0.20) + 6^2(0.15) = 10.4500$, so $\text{Var}\ (X) = 10.4500 - 2.7500^2 = 2.8875$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(1) = 0.25$, $F(2) = 0.65$, $F(4) = 0.85$, $F(6) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

`T02-A05-V03`: Completed practice sets

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.10+0.30+0.45+0.15=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 0(0.10) + 2(0.30) + 3(0.45) + 7(0.15) = 3.0000$. Across many comparable observations, the long-run average completed practice sets would approach 3.0000. Next, $E(X^2) = 0^2(0.10) + 2^2(0.30) + 3^2(0.45) + 7^2(0.15) = 12.6000$, so $\text{Var}\ (X) = 12.6000 - 3.0000^2 = 3.6000$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(0) = 0.10$, $F(2) = 0.40$, $F(3) = 0.85$, $F(7) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

`T02-A05-V04`: Reported route changes

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.45+0.25+0.20+0.10=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 0(0.45) + 1(0.25) + 2(0.20) + 4(0.10) = 1.0500$. Across many comparable observations, the long-run average reported route changes would approach 1.0500. Next, $E(X^2) = 0^2(0.45) + 1^2(0.25) + 2^2(0.20) + 4^2(0.10) = 2.6500$, so $\text{Var}\ (X) = 2.6500 - 1.0500^2 = 1.5475$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(0) = 0.45$, $F(1) = 0.70$, $F(2) = 0.90$, $F(4) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

T02-A05-V05: Weekly community meetings

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.30+0.25+0.35+0.10=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 1(0.30) + 3(0.25) + 4(0.35) + 8(0.10) = 3.2500$. Across many comparable observations, the long-run average weekly community meetings would approach 3.2500. Next, $E(X^2) = 1^2(0.30) + 3^2(0.25) + 4^2(0.35) + 8^2(0.10) = 14.5500$, so $\text{Var}\ (X) = 14.5500 - 3.2500^2 = 3.9875$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(1) = 0.30$, $F(3) = 0.55$, $F(4) = 0.90$, $F(8) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

T02-A05-V06: Successful file recoveries

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.15+0.25+0.40+0.20=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 0(0.15) + 2(0.25) + 5(0.40) + 6(0.20) = 3.7000$. Across many comparable observations, the long-run average successful file recoveries would approach 3.7000. Next, $E(X^2) = 0^2(0.15) + 2^2(0.25) + 5^2(0.40) + 6^2(0.20) = 18.2000$, so $\text{Var}\ (X) = 18.2000 - 3.7000^2 = 4.5100$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(0) = 0.15$, $F(2) = 0.40$, $F(5) = 0.80$, $F(6) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

T02-A05-V07: Museum rooms visited

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.20+0.30+0.35+0.15=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 2(0.20) + 4(0.30) + 5(0.35) + 9(0.15) = 4.7000$. Across many comparable observations, the long-run average museum rooms visited would approach 4.7000. Next, $E(X^2) = 2^2(0.20) + 4^2(0.30) + 5^2(0.35) + 9^2(0.15) = 26.5000$, so $\text{Var}\ (X) = 26.5000 - 4.7000^2 = 4.4100$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(2) = 0.20$, $F(4) = 0.50$, $F(5) = 0.85$, $F(9) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

T02-A05-V08: Optional readings completed

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.25+0.30+0.25+0.20=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 0(0.25) + 1(0.30) + 4(0.25) + 6(0.20) = 2.5000$. Across many comparable observations, the long-run average optional readings completed would approach 2.5000. Next, $E(X^2) = 0^2(0.25) + 1^2(0.30) + 4^2(0.25) + 6^2(0.20) = 11.5000$, so $\text{Var}\ (X) = 11.5000 - 2.5000^2 = 5.2500$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(0) = 0.25$, $F(1) = 0.55$, $F(4) = 0.80$, $F(6) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

`T02-A05-V09`: Verified oral-history segments

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.15+0.35+0.30+0.20=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 1(0.15) + 2(0.35) + 3(0.30) + 5(0.20) = 2.7500$. Across many comparable observations, the long-run average verified oral-history segments would approach 2.7500. Next, $E(X^2) = 1^2(0.15) + 2^2(0.35) + 3^2(0.30) + 5^2(0.20) = 9.2500$, so $\text{Var}\ (X) = 9.2500 - 2.7500^2 = 1.6875$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(1) = 0.15$, $F(2) = 0.50$, $F(3) = 0.80$, $F(5) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

`T02-A05-V10`: Data-quality warnings

Set up the calculation, part (a)

All masses are nonnegative and their sum is 0.40+0.25+0.20+0.15=1.00, so the table is a valid PMF.

Work through the calculation, part (b)

$E(X) = \sum xP(X = x) = 0(0.40) + 2(0.25) + 4(0.20) + 7(0.15) = 2.3500$. Across many comparable observations, the long-run average data-quality warnings would approach 2.3500. Next, $E(X^2) = 0^2(0.40) + 2^2(0.25) + 4^2(0.20) + 7^2(0.15) = 11.5500$, so $\text{Var}\ (X) = 11.5500 - 2.3500^2 = 6.0275$; this variance is expressed in squared count units.

Interpret and check the result, part (c)

The cumulative values are $F(0) = 0.40$, $F(2) = 0.65$, $F(4) = 0.85$, $F(7) = 1.00$. A PMF plot places a separate bar or point mass at each support value. A CDF is instead a nondecreasing, right-continuous step function that accumulates those masses and ends at 1.

A06: Exact Binomial Probabilities

`T02-A06-V01`: Completed consent checks

Set up the calculation, part (a)

The model is $X \sim B(8, 0.62)$.

$P(X = 5) = \binom{8}{5}0.62^5(1 - 0.62)^3 = 0.2815$. This is the modeled probability that exactly 5 of the 8 consent checks meet the success definition.

Work through the calculation, part (b)

$P(X = 7) = \binom{8}{7}0.62^7(1 - 0.62)^1 = 0.1071$, the corresponding probability for exactly 7.

Work through the calculation, part (c)

$E(X) = n\pi = 8(0.62) = 4.9600$ and Var $(X) = n\pi(1-\pi) = 8(0.62)(0.38) = 1.8848$. Across repeated groups of 8, the mean count would approach 4.9600.

Interpret and check the result, part (d)

The setting must keep the number of consent checks fixed at 8; classify each one only as success or failure according to whether it is completed; keep the success probability at 0.62; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V02`: Correctly classified images

Set up the calculation, part (a)

The model is $X \sim B(9, 0.74)$.

$P(X = 6) = \binom{9}{6}0.74^6(1-0.74)^3 = 0.2424$. This is the modeled probability that exactly 6 of the 9 images meet the success definition.

Work through the calculation, part (b)

$P(X = 8) = \binom{9}{8}0.74^8(1-0.74)^1 = 0.2104$, the corresponding probability for exactly 8.

Work through the calculation, part (c)

$E(X) = n\pi = 9(0.74) = 6.6600$ and Var $(X) = n\pi(1-\pi) = 9(0.74)(0.26) = 1.7316$. Across repeated groups of 9, the mean count would approach 6.6600.

Interpret and check the result, part (d)

The setting must keep the number of images fixed at 9; classify each one only as success or failure according to whether it is classified correctly; keep the success probability at 0.74; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V03`: Returned diary prompts

Set up the calculation, part (a)

The model is $X \sim B(7, 0.58)$.

$P(X = 3) = \binom{7}{3}0.58^3(1-0.58)^4 = 0.2125$. This is the modeled probability that exactly 3 of the 7 diary prompts meet the success definition.

Work through the calculation, part (b)

$P(X = 5) = \binom{7}{5}0.58^5(1-0.58)^2 = 0.2431$, the corresponding probability for exactly 5.

Work through the calculation, part (c)

$E(X) = n\pi = 7(0.58) = 4.0600$ and Var $(X) = n\pi(1-\pi) = 7(0.58)(0.42) = 1.7052$. Across repeated groups of 7, the mean count would approach 4.0600.

Interpret and check the result, part (d)

The setting must keep the number of diary prompts fixed at 7; classify each one only as success or failure according to whether it is returned; keep the success probability at 0.58; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V04`: Successful archive searches

Set up the calculation, part (a)

The model is $X \sim B(10, 0.43)$.

$P(X = 4) = \binom{10}{4}0.43^4(1-0.43)^6 = 0.2462$. This is the modeled probability that exactly 4 of the 10 archive searches meet the success definition.

Work through the calculation, part (b)

$P(X = 6) = \binom{10}{6}0.43^6(1 - 0.43)^4 = 0.1401$, the corresponding probability for exactly 6.

Work through the calculation, part (c)

$E(X) = n\pi = 10(0.43) = 4.3000$ and $\text{Var}\ (X) = n\pi(1 - \pi) = 10(0.43)(0.57) = 2.4510$. Across repeated groups of 10, the mean count would approach 4.3000.

Interpret and check the result, part (d)

The setting must keep the number of archive searches fixed at 10; classify each one only as success or failure according to whether it succeeds; keep the success probability at 0.43; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V05`: Usable sensor readings

Set up the calculation, part (a)

The model is $X \sim B(6, 0.81)$.

$P(X = 4) = \binom{6}{4}0.81^4(1 - 0.81)^2 = 0.2331$. This is the modeled probability that exactly 4 of the 6 sensor readings meet the success definition.

Work through the calculation, part (b)

$P(X = 6) = \binom{6}{6}0.81^6(1 - 0.81)^0 = 0.2824$, the corresponding probability for exactly 6.

Work through the calculation, part (c)

$E(X) = n\pi = 6(0.81) = 4.8600$ and $\text{Var}\ (X) = n\pi(1 - \pi) = 6(0.81)(0.19) = 0.9234$. Across repeated groups of 6, the mean count would approach 4.8600.

Interpret and check the result, part (d)

The setting must keep the number of sensor readings fixed at 6; classify each one only as success or failure according to whether it is usable; keep the success probability at 0.81; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V06`: On-time tutorial submissions

Set up the calculation, part (a)

The model is $X \sim B(12, 0.67)$.

$P(X = 8) = \binom{12}{8}0.67^8(1 - 0.67)^4 = 0.2384$. This is the modeled probability that exactly 8 of the 12 tutorial submissions meet the success definition.

Work through the calculation, part (b)

$P(X = 10) = \binom{12}{10}0.67^{10}(1 - 0.67)^2 = 0.1310$, the corresponding probability for exactly 10.

Work through the calculation, part (c)

$E(X) = n\pi = 12(0.67) = 8.0400$ and $\text{Var}\ (X) = n\pi(1 - \pi) = 12(0.67)(0.33) = 2.6532$. Across repeated groups of 12, the mean count would approach 8.0400.

Interpret and check the result, part (d)

The setting must keep the number of tutorial submissions fixed at 12; classify each one only as success or failure according to whether it arrives on time; keep the success probability at 0.67; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V07`: Verified catalog entries

Set up the calculation, part (a)

The model is $X \sim B(9, 0.52)$.

$P(X = 4) = \binom{9}{4}0.52^4(1 - 0.52)^5 = 0.2347$. This is the modeled probability that exactly 4 of the 9 catalog entries meet the success definition.

Work through the calculation, part (b)

$P(X = 7) = \binom{9}{7}0.52^7(1 - 0.52)^2 = 0.0853$, the corresponding probability for exactly 7.

Work through the calculation, part (c)

$E(X) = n\pi = 9(0.52) = 4.6800$ and $\text{Var}\,(X) = n\pi(1 - \pi) = 9(0.52)(0.48) = 2.2464$. Across repeated groups of 9, the mean count would approach 4.6800.

Interpret and check the result, part (d)

The setting must keep the number of catalog entries fixed at 9; classify each one only as success or failure according to whether it is verified; keep the success probability at 0.52; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V08`: Completed interview appointments

Set up the calculation, part (a)

The model is $X \sim B(11, 0.76)$.

$P(X = 8) = \binom{11}{8}0.76^8(1 - 0.76)^3 = 0.2539$. This is the modeled probability that exactly 8 of the 11 interview appointments meet the success definition.

Work through the calculation, part (b)

$P(X = 9) = \binom{11}{9}0.76^9(1 - 0.76)^2 = 0.2680$, the corresponding probability for exactly 9.

Work through the calculation, part (c)

$E(X) = n\pi = 11(0.76) = 8.3600$ and $\text{Var}\,(X) = n\pi(1 - \pi) = 11(0.76)(0.24) = 2.0064$. Across repeated groups of 11, the mean count would approach 8.3600.

Interpret and check the result, part (d)

The setting must keep the number of interview appointments fixed at 11; classify each one only as success or failure according to whether it is completed; keep the success probability at 0.76; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

`T02-A06-V09`: Correct route choices

Set up the calculation, part (a)

The model is $X \sim B(8, 0.35)$.

$P(X = 2) = \binom{8}{2}0.35^2(1 - 0.35)^6 = 0.2587$. This is the modeled probability that exactly 2 of the 8 route choices meet the success definition.

Work through the calculation, part (b)

$P(X = 4) = \binom{8}{4}0.35^4(1 - 0.35)^4 = 0.1875$, the corresponding probability for exactly 4.

Work through the calculation, part (c)

$E(X) = n\pi = 8(0.35) = 2.8000$ and $\text{Var}\,(X) = n\pi(1 - \pi) = 8(0.35)(0.65) = 1.8200$. Across repeated groups of 8, the mean count would approach 2.8000.

Interpret and check the result, part (d)

The setting must keep the number of route choices fixed at 8; classify each one only as success or failure according to whether it is correct; keep the success probability at 0.35; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

T02-A06-V10: Successful audio transcriptions
Set up the calculation, part (a)

The model is $X \sim B(10, 0.69)$.

$P(X = 6) = \binom{10}{6}0.69^6(1 - 0.69)^4 = 0.2093$. This is the modeled probability that exactly 6 of the 10 audio transcriptions meet the success definition.

Work through the calculation, part (b)

$P(X = 8) = \binom{10}{8}0.69^8(1 - 0.69)^2 = 0.2222$, the corresponding probability for exactly 8.

Work through the calculation, part (c)

$E(X) = n\pi = 10(0.69) = 6.9000$ and $\text{Var}\ (X) = n\pi(1 - \pi) = 10(0.69)(0.31) = 2.1390$. Across repeated groups of 10, the mean count would approach 6.9000.

Interpret and check the result, part (d)

The setting must keep the number of audio transcriptions fixed at 10; classify each one only as success or failure according to whether it is successful; keep the success probability at 0.69; and make the trial outcomes independent. If any condition fails, this binomial calculation is not justified.

A07: Binomial Tail Probabilities by Complement

T02-A07-V01: More than 3 records requiring manual review
Set up the calculation, part (a)

Here $X \sim B(40, 0.04)$, and the complement of $X > 3$ is $X \leq 3$.

Work through the calculation, part (b)

Therefore $P(X > 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = 1 - [0.1954 + 0.3256 + 0.2646 + 0.1396] \approx 1 - 0.9252 = 0.0748$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9252 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0748 to seeing more than 3 successes among the 40 records, where one success means that one unit requires manual review. The complement uses 4 lower-tail terms, whereas a direct sum would require the values 4 through 40.

T02-A07-V02: More than 5 visitors requesting audio guides
Set up the calculation, part (a)

Here $X \sim B(25, 0.12)$, and the complement of $X > 5$ is $X \leq 5$.

Work through the calculation, part (b)

Therefore $P(X > 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)] = 1 - [0.0409 + 0.1395 + 0.2283 + 0.2387 + 0.1790 + 0.1025] \approx 1 - 0.9291 = 0.0709$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9291 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0709 to seeing more than 5 successes among the 25 museum visitors, where one success means that one unit requests an audio guide. The complement uses 6 lower-tail terms, whereas a direct sum would require the values 6 through 25.

T02-A07-V03: More than 4 invalid survey links
Set up the calculation, part (a)

Here $X \sim B(30, 0.06)$, and the complement of $X > 4$ is $X \leq 4$.

Work through the calculation, part (b)

Therefore $P(X > 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] = 1 - [0.1563 + 0.2992 + 0.2769 + 0.1650 + 0.0711] \approx 1 - 0.9685 = 0.0315$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9685 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0315 to seeing more than 4 successes among the 30 survey links, where one success means that one unit is returned as invalid. The complement uses 5 lower-tail terms, whereas a direct sum would require the values 5 through 30.

`T02-A07-V04`: More than 6 items needing preservation work
Set up the calculation, part (a)

Here $X \sim B(35, 0.09)$, and the complement of $X > 6$ is $X \leq 6$.

Work through the calculation, part (b)

Therefore $P(X > 6) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)] = 1 - [0.0369 + 0.1276 + 0.2145 + 0.2333 + 0.1846 + 0.1132 + 0.0560] \approx 1 - 0.9660 = 0.0340$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9660 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0340 to seeing more than 6 successes among the 35 items, where one success means that one unit needs preservation work. The complement uses 7 lower-tail terms, whereas a direct sum would require the values 7 through 35.

`T02-A07-V05`: More than 4 participants missing a reminder
Set up the calculation, part (a)

Here $X \sim B(28, 0.08)$, and the complement of $X > 4$ is $X \leq 4$.

Work through the calculation, part (b)

Therefore $P(X > 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] = 1 - [0.0968 + 0.2358 + 0.2768 + 0.2086 + 0.1134] \approx 1 - 0.9314 = 0.0686$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9314 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0686 to seeing more than 4 successes among the 28 participants, where one success means that one unit misses a reminder. The complement uses 5 lower-tail terms, whereas a direct sum would require the values 5 through 28.

`T02-A07-V06`: More than 5 uploads needing a second attempt
Set up the calculation, part (a)

Here $X \sim B(32, 0.07)$, and the complement of $X > 5$ is $X \leq 5$.

Work through the calculation, part (b)

Therefore $P(X > 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)] = 1 - [0.0981 + 0.2362 + 0.2755 + 0.2074 + 0.1132 + 0.0477] \approx 1 - 0.9780 = 0.0220$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9780 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0220 to seeing more than 5 successes among the 32 uploads, where one success means that one unit needs a second attempt. The complement uses 6 lower-tail terms, whereas a direct sum would require the values 6 through 32.

T02-A07-V07: More than 5 sampled pages containing annotations
Set up the calculation, part (a)

Here $X \sim B(20, 0.15)$, and the complement of $X > 5$ is $X \leq 5$.

Work through the calculation, part (b)

Therefore $P(X > 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)] = 1 - [0.0388 + 0.1368 + 0.2293 + 0.2428 + 0.1821 + 0.1028] \approx 1 - 0.9327 = 0.0673$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9327 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0673 to seeing more than 5 successes among the 20 sampled pages, where one success means that one unit contains annotations. The complement uses 6 lower-tail terms, whereas a direct sum would require the values 6 through 20.

T02-A07-V08: More than 4 interviews requiring rescheduling
Set up the calculation, part (a)

Here $X \sim B(24, 0.11)$, and the complement of $X > 4$ is $X \leq 4$.

Work through the calculation, part (b)

Therefore $P(X > 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] = 1 - [0.0610 + 0.1810 + 0.2572 + 0.2331 + 0.1513] \approx 1 - 0.8835 = 0.1165$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.8835 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.1165 to seeing more than 4 successes among the 24 interviews, where one success means that one unit requires rescheduling. The complement uses 5 lower-tail terms, whereas a direct sum would require the values 5 through 24.

T02-A07-V09: More than 3 route observations showing delay
Set up the calculation, part (a)

Here $X \sim B(36, 0.05)$, and the complement of $X > 3$ is $X \leq 3$.

Work through the calculation, part (b)

Therefore $P(X > 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = 1 - [0.1578 + 0.2990 + 0.2753 + 0.1642] \approx 1 - 0.8963 = 0.1037$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.8963 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.1037 to seeing more than 3 successes among the 36 route observations, where one success means that one unit shows a delay. The complement uses 4 lower-tail terms, whereas a direct sum would require the values 4 through 36.

T02-A07-V10: More than 5 forms containing an optional comment
Set up the calculation, part (a)

Here $X \sim B(18, 0.18)$, and the complement of $X > 5$ is $X \leq 5$.

Work through the calculation, part (b)

Therefore $P(X > 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)] = 1 - [0.0281 + 0.1110 + 0.2071 + 0.2425 + 0.1996 + 0.1227] \approx 1 - 0.9111 = 0.0889$. The approximation sign is necessary because the displayed component terms are rounded; the value 0.9111 was calculated from unrounded terms.

Interpret and check the result, part (c)

The model assigns probability 0.0889 to seeing more than 5 successes among the 18 forms, where one success means that one unit contains an optional comment. The complement uses 6 lower-tail terms, whereas a direct sum would require the values 6 through 18.

A09: Standard-Normal Probabilities

`T02-A09-V01`: Standard-normal regions, set 1
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le -0.45) = \Phi(-0.45) = 0.3264$. Shade left of −0.45.

Work through the calculation, part (b)

$P(Z > 1.36) = 1 - \Phi(1.36) = 0.0869$. Shade the right tail.

Interpret and check the result, part (c)

$P(-0.80 < Z \le 0.95) = \Phi(0.95) - \Phi(-0.80) = 0.6171$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V02`: Standard-normal regions, set 2
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le -1.12) = \Phi(-1.12) = 0.1314$. Shade left of −1.12.

Work through the calculation, part (b)

$P(Z > 0.84) = 1 - \Phi(0.84) = 0.2005$. Shade the right tail.

Interpret and check the result, part (c)

$P(-0.35 < Z \le 1.42) = \Phi(1.42) - \Phi(-0.35) = 0.5590$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V03`: Standard-normal regions, set 3
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le 0.28) = \Phi(0.28) = 0.6103$. Shade left of 0.28.

Work through the calculation, part (b)

$P(Z > 1.74) = 1 - \Phi(1.74) = 0.0409$. Shade the right tail.

Interpret and check the result, part (c)

$P(-1.05 < Z \le 0.62) = \Phi(0.62) - \Phi(-1.05) = 0.5855$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V04`: Standard-normal regions, set 4
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le -0.93) = \Phi(-0.93) = 0.1762$. Shade left of −0.93.

Work through the calculation, part (b)

$P(Z > 1.18) = 1 - \Phi(1.18) = 0.1190$. Shade the right tail.

Interpret and check the result, part (c)

$P(-0.44 < Z \le 1.27) = \Phi(1.27) - \Phi(-0.44) = 0.5680$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V05`: Standard-normal regions, set 5
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le 0.67) = \Phi(0.67) = 0.7486$. Shade left of 0.67.

Work through the calculation, part (b)

$P(Z > 2.05) = 1 - \Phi(2.05) = 0.0202$. Shade the right tail.

Interpret and check the result, part (c)

$P(-1.33 < Z \le 0.71) = \Phi(0.71) - \Phi(-1.33) = 0.6694$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V06`: Standard-normal regions, set 6
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le -1.48) = \Phi(-1.48) = 0.0694$. Shade left of −1.48.

Work through the calculation, part (b)

$P(Z > 0.56) = 1 - \Phi(0.56) = 0.2877$. Shade the right tail.

Interpret and check the result, part (c)

$P(-0.92 < Z \le 1.08) = \Phi(1.08) - \Phi(-0.92) = 0.6811$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V07`: Standard-normal regions, set 7
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le 0.14) = \Phi(0.14) = 0.5557$. Shade left of 0.14.

Work through the calculation, part (b)

$P(Z > 1.51) = 1 - \Phi(1.51) = 0.0655$. Shade the right tail.

Interpret and check the result, part (c)

$P(-0.68 < Z \le 1.19) = \Phi(1.19) - \Phi(-0.68) = 0.6347$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V08`: Standard-normal regions, set 8
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le -0.76) = \Phi(-0.76) = 0.2236$. Shade left of −0.76.

Work through the calculation, part (b)

$P(Z > 1.89) = 1 - \Phi(1.89) = 0.0294$. Shade the right tail.

Interpret and check the result, part (c)

$P(-1.21 < Z \le 0.37) = \Phi(0.37) - \Phi(-1.21) = 0.5312$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V09`: Standard-normal regions, set 9
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le 0.91) = \Phi(0.91) = 0.8186$. Shade left of 0.91.

Work through the calculation, part (b)

$P(Z > 1.24) = 1 - \Phi(1.24) = 0.1075$. Shade the right tail.

Interpret and check the result, part (c)

$P(-0.57 < Z \le 1.63) = \Phi(1.63) - \Phi(-0.57) = 0.6641$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

`T02-A09-V10`: Standard-normal regions, set 10
Set up the calculation, part (a)

Write $\Phi(z) = P(Z \le z)$.

$P(Z \le -0.22) = \Phi(-0.22) = 0.4129$. Shade left of −0.22.

Work through the calculation, part (b)

$P(Z > 2.17) = 1 - \Phi(2.17) = 0.0150$. Shade the right tail.

Interpret and check the result, part (c)

$P(-1.46 < Z \le 0.88) = \Phi(0.88) - \Phi(-1.46) = 0.7384$. Shade between the two cutoffs. Boundary inclusions do not change probabilities for a continuous distribution.

A10: Probabilities for a General Normal Distribution

`T02-A10-V01`: Modeling reading fluency score
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{100} = 10.00$.

$z = (79 - 72)/10.00 \approx 0.7000$. Retaining the unrounded quotient gives $P(X \le 79) = \Phi((79 - 72)/10.00) = 0.7580$. Thus the model places proportion 0.7580 of reading fluency score values at or below 79 score points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (68 - 72)/10.00 \approx -0.4000$, and $P(X > 68) = 1 - \Phi((68 - 72)/10.00) = 0.6554$. This is the modeled proportion above 68 score points, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V02`: Modeling archive processing time
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{64} = 8.00$.

$z = (51 - 45)/8.00 \approx 0.7500$. Retaining the unrounded quotient gives $P(X \leq 51) = \Phi((51 - 45)/8.00) = 0.7734$. Thus the model places proportion 0.7734 of archive processing time values at or below 51 minutes; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (39 - 45)/8.00 \approx -0.7500$, and $P(X > 39) = 1 - \Phi((39 - 45)/8.00) = 0.7734$. This is the modeled proportion above 39 minutes, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V03`: Modeling wellbeing score

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{81} = 9.00$.

$z = (64 - 58)/9.00 \approx 0.6667$. Retaining the unrounded quotient gives $P(X \leq 64) = \Phi((64 - 58)/9.00) = 0.7475$. Thus the model places proportion 0.7475 of wellbeing score values at or below 64 score points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (52 - 58)/9.00 \approx -0.6667$, and $P(X > 52) = 1 - \Phi((52 - 58)/9.00) = 0.7475$. This is the modeled proportion above 52 score points, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V04`: Modeling museum visit duration

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{225} = 15.00$.

$z = (105 - 90)/15.00 \approx 1.0000$. Retaining the unrounded quotient gives $P(X \leq 105) = \Phi((105 - 90)/15.00) = 0.8413$. Thus the model places proportion 0.8413 of museum visit duration values at or below 105 minutes; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (78 - 90)/15.00 \approx -0.8000$, and $P(X > 78) = 1 - \Phi((78 - 90)/15.00) = 0.7881$. This is the modeled proportion above 78 minutes, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V05`: Modeling memory score

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{144} = 12.00$.

$z = (124 - 110)/12.00 \approx 1.1667$. Retaining the unrounded quotient gives $P(X \leq 124) = \Phi((124 - 110)/12.00) = 0.8783$. Thus the model places proportion 0.8783 of memory score values at or below 124 score points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (103 - 110)/12.00 \approx -0.5833$, and $P(X > 103) = 1 - \Phi((103 - 110)/12.00) = 0.7202$. This is the modeled proportion above 103 score points, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V06`: Modeling sound-level index

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{49} = 7.00$.

$z = (42 - 38)/7.00 \approx 0.5714$. Retaining the unrounded quotient gives $P(X \leq 42) = \Phi((42 - 38)/7.00) = 0.7161$. Thus the model places proportion 0.7161 of sound-level index values at or below 42 index points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (33 - 38)/7.00 \approx -0.7143$, and $P(X > 33) = 1 - \Phi((33 - 38)/7.00) = 0.7625$. This is the modeled proportion above 33 index points, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V07`: Modeling course confidence score

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{121} = 11.00$.

$z = (75 - 66)/11.00 \approx 0.8182$. Retaining the unrounded quotient gives $P(X \leq 75) = \Phi((75 - 66)/11.00) = 0.7934$. Thus the model places proportion 0.7934 of course confidence score values at or below 75 score points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (59 - 66)/11.00 \approx -0.6364$, and $P(X > 59) = 1 - \Phi((59 - 66)/11.00) = 0.7377$. This is the modeled proportion above 59 score points, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V08`: Modeling response time

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{3600} = 60.00$.

$z = (575 - 520)/60.00 \approx 0.9167$. Retaining the unrounded quotient gives $P(X \leq 575) = \Phi((575 - 520)/60.00) = 0.8203$. Thus the model places proportion 0.8203 of response time values at or below 575 milliseconds; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (485 - 520)/60.00 \approx -0.5833$, and $P(X > 485) = 1 - \Phi((485 - 520)/60.00) = 0.7202$. This is the modeled proportion above 485 milliseconds, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V09`: Modeling community trust score

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{64} = 8.00$.

$z = (54 - 48)/8.00 \approx 0.7500$. Retaining the unrounded quotient gives $P(X \leq 54) = \Phi((54 - 48)/8.00) = 0.7734$. Thus the model places proportion 0.7734 of community trust score values at or below 54 score points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (43 - 48)/8.00 \approx -0.6250$, and $P(X > 43) = 1 - \Phi((43 - 48)/8.00) = 0.7340$. This is the modeled proportion above 43 score points, represented by the right tail. Both interpretations depend on the stated normal model.

`T02-A10-V10`: Modeling cataloging accuracy score

Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{36} = 6.00$.

$z = (88 - 84)/6.00 \approx 0.6667$. Retaining the unrounded quotient gives $P(X \leq 88) = \Phi((88 - 84)/6.00) = 0.7475$. Thus the model places proportion 0.7475 of cataloging accuracy score values at or below 88 score points; shade the left side of that cutoff.

Interpret and check the result, part (b)

$z = (79 - 84)/6.00 \approx -0.8333$, and $P(X > 79) = 1 - \Phi((79 - 84)/6.00) = 0.7977$. This is the modeled proportion above 79 score points, represented by the right tail. Both interpretations depend on the stated normal model.

A11: Inverse Standard-Normal Quantiles

`T02-A11-V01`: Finding the 70% and 92% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.70$ gives $z_{70\%} = 0.5244$. Because 0.70 is above 0.50, this cutoff is positive.

Interpret and check the result, part (b)

$\Phi(z) = 0.92$ gives $z_{92\%} = 1.4051$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

`T02-A11-V02`: Finding the 15% and 88% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.15$ gives $z_{15\%} = -1.0364$. Because 0.15 is below 0.50, this cutoff is negative.

Interpret and check the result, part (b)

$\Phi(z) = 0.88$ gives $z_{88\%} = 1.1750$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

`T02-A11-V03`: Finding the 80% and 96% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.80$ gives $z_{80\%} = 0.8416$. Because 0.80 is above 0.50, this cutoff is positive.

Interpret and check the result, part (b)

$\Phi(z) = 0.96$ gives $z_{96\%} = 1.7507$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

`T02-A11-V04`: Finding the 28% and 90% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.28$ gives $z_{28\%} = -0.5828$. Because 0.28 is below 0.50, this cutoff is negative.

Interpret and check the result, part (b)

$\Phi(z) = 0.90$ gives $z_{90\%} = 1.2816$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

`T02-A11-V05`: Finding the 50% and 94% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.50$ gives $z_{50\%} = 0.0000$. Because 0.50 is equal to 0.50, this cutoff is zero.

Interpret and check the result, part (b)

$\Phi(z) = 0.94$ gives $z_{94\%} = 1.5548$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

T02-A11-V06: Finding the 75% and 97% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.75$ gives $z_{75\%} = 0.6745$. Because 0.75 is above 0.50, this cutoff is positive.

Interpret and check the result, part (b)

$\Phi(z) = 0.97$ gives $z_{97\%} = 1.8808$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

T02-A11-V07: Finding the 32% and 68% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.32$ gives $z_{32\%} = -0.4677$. Because 0.32 is below 0.50, this cutoff is negative.

Interpret and check the result, part (b)

$\Phi(z) = 0.68$ gives $z_{68\%} = 0.4677$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

T02-A11-V08: Finding the 82% and 95% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.82$ gives $z_{82\%} = 0.9154$. Because 0.82 is above 0.50, this cutoff is positive.

Interpret and check the result, part (b)

$\Phi(z) = 0.95$ gives $z_{95\%} = 1.6449$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

T02-A11-V09: Finding the 11% and 62% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.11$ gives $z_{11\%} = -1.2265$. Because 0.11 is below 0.50, this cutoff is negative.

Interpret and check the result, part (b)

$\Phi(z) = 0.62$ gives $z_{62\%} = 0.3055$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

T02-A11-V10: Finding the 78% and 93% z-quantiles
Set up the calculation, part (a)

A quantile begins with cumulative probability q and solves $\Phi(z) = q$ for a location.

$\Phi(z) = 0.78$ gives $z_{78\%} = 0.7722$. Because 0.78 is above 0.50, this cutoff is positive.

Interpret and check the result, part (b)

$\Phi(z) = 0.93$ gives $z_{93\%} = 1.4758$; its expected sign is positive. The inputs are areas and the outputs are positions on the z-axis, which reverses an ordinary forward CDF calculation.

A12: Sampling Distribution of the Mean

`T02-A12-V01`: Precision of a sample mean for reading score
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 64$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{196} = 14.00$, so SE $= 14.00/\sqrt{49} = 2.0000$ score points. Across repeated samples, their means are centered at 64 score points with standard deviation 2.0000 score points.

Work through the calculation, part (b)

With variance 100, SE $= \sqrt{100}/\sqrt{49} = 1.4286$ score points. The smaller population variance decreases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 121$, SE $= \sqrt{196}/\sqrt{121} = 1.2727$ score points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

`T02-A12-V02`: Precision of a sample mean for processing time
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 52$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{225} = 15.00$, so SE $= 15.00/\sqrt{36} = 2.5000$ minutes. Across repeated samples, their means are centered at 52 minutes with standard deviation 2.5000 minutes.

Work through the calculation, part (b)

With variance 144, SE $= \sqrt{144}/\sqrt{36} = 2.0000$ minutes. The smaller population variance decreases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 100$, SE $= \sqrt{225}/\sqrt{100} = 1.5000$ minutes. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

`T02-A12-V03`: Precision of a sample mean for wellbeing index
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 71$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{144} = 12.00$, so SE $= 12.00/\sqrt{64} = 1.5000$ index points. Across repeated samples, their means are centered at 71 index points with standard deviation 1.5000 index points.

Work through the calculation, part (b)

With variance 256, SE $= \sqrt{256}/\sqrt{64} = 2.0000$ index points. The larger population variance increases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 81$, SE $= \sqrt{144}/\sqrt{81} = 1.3333$ index points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

T02-A12-V04: Precision of a sample mean for memory score
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 105$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{324} = 18.00$, so SE $= 18.00/\sqrt{81} = 2.0000$ score points. Across repeated samples, their means are centered at 105 score points with standard deviation 2.0000 score points.

Work through the calculation, part (b)

With variance 225, SE $= \sqrt{225}/\sqrt{81} = 1.6667$ score points. The smaller population variance decreases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 144$, SE $= \sqrt{324}/\sqrt{144} = 1.5000$ score points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

T02-A12-V05: Precision of a sample mean for trust rating
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 48$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{100} = 10.00$, so SE $= 10.00/\sqrt{25} = 2.0000$ rating points. Across repeated samples, their means are centered at 48 rating points with standard deviation 2.0000 rating points.

Work through the calculation, part (b)

With variance 169, SE $= \sqrt{169}/\sqrt{25} = 2.6000$ rating points. The larger population variance increases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 64$, SE $= \sqrt{100}/\sqrt{64} = 1.2500$ rating points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

T02-A12-V06: Precision of a sample mean for reaction time
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 480$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{2500} = 50.00$, so SE $= 50.00/\sqrt{100} = 5.0000$ milliseconds. Across repeated samples, their means are centered at 480 milliseconds with standard deviation 5.0000 milliseconds.

Work through the calculation, part (b)

With variance 1600, SE $= \sqrt{1600}/\sqrt{100} = 4.0000$ milliseconds. The smaller population variance decreases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 400$, SE $= \sqrt{2500}/\sqrt{400} = 2.5000$ milliseconds. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

`T02-A12-V07`: Precision of a sample mean for confidence score
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 59$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{121} = 11.00$, so SE $= 11.00/\sqrt{49} = 1.5714$ score points. Across repeated samples, their means are centered at 59 score points with standard deviation 1.5714 score points.

Work through the calculation, part (b)

With variance 196, SE $= \sqrt{196}/\sqrt{49} = 2.0000$ score points. The larger population variance increases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 100$, SE $= \sqrt{121}/\sqrt{100} = 1.1000$ score points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

`T02-A12-V08`: Precision of a sample mean for visit duration
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 82$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{400} = 20.00$, so SE $= 20.00/\sqrt{64} = 2.5000$ minutes. Across repeated samples, their means are centered at 82 minutes with standard deviation 2.5000 minutes.

Work through the calculation, part (b)

With variance 256, SE $= \sqrt{256}/\sqrt{64} = 2.0000$ minutes. The smaller population variance decreases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 144$, SE $= \sqrt{400}/\sqrt{144} = 1.6667$ minutes. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

`T02-A12-V09`: Precision of a sample mean for accuracy score
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 88$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{81} = 9.00$, so SE $= 9.00/\sqrt{36} = 1.5000$ score points. Across repeated samples, their means are centered at 88 score points with standard deviation 1.5000 score points.

Work through the calculation, part (b)

With variance 144, SE $= \sqrt{144}/\sqrt{36} = 2.0000$ score points. The larger population variance increases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 49$, SE $= \sqrt{81}/\sqrt{49} = 1.2857$ score points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

`T02-A12-V10`: Precision of a sample mean for sound index
Reason before calculating, part (a)

For an unbiased sample mean, $E(\bar{X}) = \mu = 42$. Independence gives SD $(\bar{X}) = \sigma/\sqrt{n}$.

$\sigma = \sqrt{169} = 13.00$, so SE $= 13.00/\sqrt{25} = 2.6000$ index points. Across repeated samples, their means are centered at 42 index points with standard deviation 2.6000 index points.

Work through the calculation, part (b)

With variance 100, SE $= \sqrt{100}/\sqrt{25} = 2.0000$ index points. The smaller population variance decreases the SE relative to part

Work through the calculation, part (a)

.

Interpret and check the result, part (c)

With $n = 64$, SE $= \sqrt{169}/\sqrt{64} = 1.6250$ index points. The larger sample size decreases the SE through the square root of n. A smaller SE means that repeated sample means cluster more tightly around the population mean.

A13: Intervals under a Normal Model

`T02-A13-V01`: Interval probabilities for focus score
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{81} = 9.00$.

The endpoints are $z_a = (50 - 50)/9.00 \approx 0.0000$ and $z_b = (59 - 50)/9.00 \approx 1.0000$. Using the unrounded z-scores, $P(50 < X \leq 59) = \Phi((59 - 50)/9.00) - \Phi((50 - 50)/9.00) = 0.3413$. The model therefore places proportion 0.3413 of focus score values between 50 and 59 score points.

Interpret and check the result, part (b)

$z_c = (43 - 50)/9.00 \approx -0.7778$ and $z_d = (61 - 50)/9.00 \approx 1.2222$, giving $P(43 < X \leq 61) = \Phi((61 - 50)/9.00) - \Phi((43 - 50)/9.00) = 0.6708$. This is the modeled proportion between 43 and 61 score points. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V02`: Interval probabilities for reading score
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{100} = 10.00$.

The endpoints are $z_a = (65 - 70)/10.00 \approx -0.5000$ and $z_b = (82 - 70)/10.00 \approx 1.2000$. Using the unrounded z-scores, $P(65 < X \leq 82) = \Phi((82 - 70)/10.00) - \Phi((65 - 70)/10.00) = 0.5764$. The model therefore places proportion 0.5764 of reading score values between 65 and 82 score points.

Interpret and check the result, part (b)

$z_c = (58 - 70)/10.00 \approx -1.2000$ and $z_d = (76 - 70)/10.00 \approx 0.6000$, giving $P(58 < X \leq 76) = \Phi((76 - 70)/10.00) - \Phi((58 - 70)/10.00) = 0.6107$. This is the modeled proportion between 58 and 76 score points. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V03`: Interval probabilities for visit duration
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{225} = 15.00$.

The endpoints are $z_a = (80 - 80)/15.00 \approx 0.0000$ and $z_b = (95 - 80)/15.00 \approx 1.0000$. Using the unrounded z-scores, $P(80 < X \leq 95) = \Phi((95 - 80)/15.00) - \Phi((80 - 80)/15.00) = 0.3413$. The model therefore places proportion 0.3413 of visit duration values between 80 and 95 minutes.

Interpret and check the result, part (b)

$z_c = (62 - 80)/15.00 \approx -1.2000$ and $z_d = (101 - 80)/15.00 \approx 1.4000$, giving $P(62 < X \leq 101) = \Phi((101 - 80)/15.00) - \Phi((62 - 80)/15.00) = 0.8042$. This is the modeled proportion between 62 and 101 minutes. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V04`: Interval probabilities for memory score
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{144} = 12.00$.

The endpoints are $z_a = (96 - 105)/12.00 \approx -0.7500$ and $z_b = (117 - 105)/12.00 \approx 1.0000$. Using the unrounded z-scores, $P(96 < X \leq 117) = \Phi((117 - 105)/12.00) - \Phi((96 - 105)/12.00) = 0.6147$. The model therefore places proportion 0.6147 of memory score values between 96 and 117 score points.

Interpret and check the result, part (b)

$z_c = (88 - 105)/12.00 \approx -1.4167$ and $z_d = (122 - 105)/12.00 \approx 1.4167$, giving $P(88 < X \leq 122) = \Phi((122 - 105)/12.00) - \Phi((88 - 105)/12.00) = 0.8434$. This is the modeled proportion between 88 and 122 score points. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V05`: Interval probabilities for trust index
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{64} = 8.00$.

The endpoints are $z_a = (40 - 44)/8.00 \approx -0.5000$ and $z_b = (52 - 44)/8.00 \approx 1.0000$. Using the unrounded z-scores, $P(40 < X \leq 52) = \Phi((52 - 44)/8.00) - \Phi((40 - 44)/8.00) = 0.5328$. The model therefore places proportion 0.5328 of trust index values between 40 and 52 index points.

Interpret and check the result, part (b)

$z_c = (33 - 44)/8.00 \approx -1.3750$ and $z_d = (49 - 44)/8.00 \approx 0.6250$, giving $P(33 < X \leq 49) = \Phi((49 - 44)/8.00) - \Phi((33 - 44)/8.00) = 0.6494$. This is the modeled proportion between 33 and 49 index points. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V06`: Interval probabilities for response time
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{2500} = 50.00$.

The endpoints are $z_a = (475 - 500)/50.00 \approx -0.5000$ and $z_b = (560 - 500)/50.00 \approx 1.2000$. Using the unrounded z-scores, $P(475 < X \leq 560) = \Phi((560 - 500)/50.00) - \Phi((475 - 500)/50.00) = 0.5764$. The model therefore places proportion 0.5764 of response time values between 475 and 560 milliseconds.

Interpret and check the result, part (b)

$z_c = (410 - 500)/50.00 \approx -1.8000$ and $z_d = (535 - 500)/50.00 \approx 0.7000$, giving $P(410 < X \leq 535) = \Phi((535 - 500)/50.00) - \Phi((410 - 500)/50.00) = 0.7221$. This is the modeled proportion between 410 and 535 milliseconds. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V07`: Interval probabilities for wellbeing score
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{121} = 11.00$.

The endpoints are $z_a = (62 - 62)/11.00 \approx 0.0000$ and $z_b = (74 - 62)/11.00 \approx 1.0909$. Using the unrounded z-scores, $P(62 < X \leq 74) = \Phi((74 - 62)/11.00) - \Phi((62 - 62)/11.00) = 0.3623$. The model therefore places proportion 0.3623 of wellbeing score values between 62 and 74 score points.

Interpret and check the result, part (b)

$z_c = (48 - 62)/11.00 \approx -1.2727$ and $z_d = (69 - 62)/11.00 \approx 0.6364$, giving $P(48 < X \leq 69) = \Phi((69 - 62)/11.00) - \Phi((48 - 62)/11.00) = 0.6362$. This is the modeled proportion between 48 and 69 score points. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V08`: Interval probabilities for cataloging score
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{49} = 7.00$.

The endpoints are $z_a = (81 - 86)/7.00 \approx -0.7143$ and $z_b = (93 - 86)/7.00 \approx 1.0000$. Using the unrounded z-scores, $P(81 < X \leq 93) = \Phi((93 - 86)/7.00) - \Phi((81 - 86)/7.00) = 0.6038$. The model therefore places proportion 0.6038 of cataloging score values between 81 and 93 score points.

Interpret and check the result, part (b)

$z_c = (74 - 86)/7.00 \approx -1.7143$ and $z_d = (90 - 86)/7.00 \approx 0.5714$, giving $P(74 < X \leq 90) = \Phi((90 - 86)/7.00) - \Phi((74 - 86)/7.00) = 0.6729$. This is the modeled proportion between 74 and 90 score points. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V09`: Interval probabilities for sound level
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{36} = 6.00$.

The endpoints are $z_a = (36 - 36)/6.00 \approx 0.0000$ and $z_b = (42 - 36)/6.00 \approx 1.0000$. Using the unrounded z-scores, $P(36 < X \leq 42) = \Phi((42 - 36)/6.00) - \Phi((36 - 36)/6.00) = 0.3413$. The model therefore places proportion 0.3413 of sound level values between 36 and 42 decibels.

Interpret and check the result, part (b)

$z_c = (27 - 36)/6.00 \approx -1.5000$ and $z_d = (39 - 36)/6.00 \approx 0.5000$, giving $P(27 < X \leq 39) = \Phi((39 - 36)/6.00) - \Phi((27 - 36)/6.00) = 0.6247$. This is the modeled proportion between 27 and 39 decibels. Endpoint inclusion does not affect a continuous-model probability.

`T02-A13-V10`: Interval probabilities for confidence rating
Set up the calculation, part (a)

The standard deviation is $\sigma = \sqrt{64} = 8.00$.

The endpoints are $z_a = (51 - 55)/8.00 \approx -0.5000$ and $z_b = (63 - 55)/8.00 \approx 1.0000$. Using the unrounded z-scores, $P(51 < X \leq 63) = \Phi((63 - 55)/8.00) - \Phi((51 - 55)/8.00) = 0.5328$. The model therefore places proportion 0.5328 of confidence rating values between 51 and 63 rating points.

Interpret and check the result, part (b)

$z_c = (43 - 55)/8.00 \approx -1.5000$ and $z_d = (59 - 55)/8.00 \approx 0.5000$, giving $P(43 < X \leq 59) = \Phi((59 - 55)/8.00) - \Phi((43 - 55)/8.00) = 0.6247$. This is the modeled proportion between 43 and 59 rating points. Endpoint inclusion does not affect a continuous-model probability.