Past-paper walk-through

Potential Problems with Sampling ▪ 3.4

eXercise

Questions in this set

- 2025 Q2
- 2024 Q3
- 2021 Q2
- 2018 Q2
- 2016 Q3
- 2015 Q6
- 2014 Q4

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 2

2025 ▪ Q2

Aphids are tiny insects that feed on plants such as cabbage plants. A farmer wants to reduce the number of aphids in a cabbage field. A river is located 100 meters south of the cabbage field. The farmer divides the field into 25 regions of equal size, as shown in the diagram. Each region has approximately the same number of cabbage plants.

Farmer's House and Cabbage Field

[Diagram: a house icon sits above a 5×5 grid of numbered regions, labeled by row A-E (top to bottom) and organized in 5 columns. The grid cells, read row by row, are: Row A: 1, 2, 3, 4, 5 Row B: 6, 7, 8, 9, 10 Row C: 11, 12, 13, 14, 15 Row D: 16, 17, 18, 19, 20 Row E: 21, 22, 23, 24, 25 Below row E, an arrow labeled "100 meters" points down to a shaded band labeled "River".]

The farmer would like to estimate the proportion of cabbage plants in the field that are affected by aphids and believes that the extent of aphid damage is greater for the

Question 2

regions in the cabbage field closer to the river. To obtain the estimate, the farmer is considering three sampling methods.

- Sampling method I: Select region 3, which is closest to the farmer's house and farthest from the river. Examine every cabbage plant in the region for aphid damage.
- Sampling method II: Randomly select one row (A, B, C, D, or E). For every region in the selected row, examine every cabbage plant for aphid damage.
- Sampling method III: Randomly select one region from each of rows A, B, C, D, and E. For each selected region, examine every cabbage plant for aphid damage.

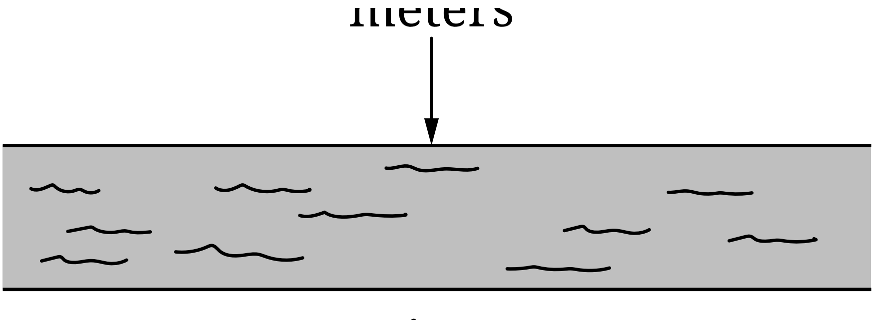
Question 2

(a) Explain whether sampling method I is an appropriate sampling method for the farmer to use to estimate the proportion of cabbage plants in the field that are damaged by aphids.

(b) Using sampling method II, the farmer randomly selected row E and examined every cabbage plant in row E. If the farmer's belief is correct, determine whether the selection of row E is likely to provide an overestimate or an underestimate of the proportion of cabbage plants in the field that are damaged by aphids. Justify your answer.

(c) Using the information provided in the diagram of the cabbage field, describe how to implement sampling method III, which requires a random selection of one region from each of rows A, B, C, D, and E.

Question 2



Solution 2

(a) Mark scheme

- No, sampling method I is not an appropriate sampling method. It is a convenience sample because region 3 is not selected randomly. If the farmer's belief is correct (aphid damage decreases farther from the river), region 3 — the row farthest from the river — likely has fewer aphid-damaged cabbage plants than most other regions in the field, so examining only region 3 would likely lead to an underestimate of the field-wide proportion of damaged plants. Full credit needs: (1) indicates method I is not appropriate, (2) an explanation referencing that the region is not representative / the sampling process is not random / it is a convenience sample, (3) sufficient context including any two of: sampling unit

Solution 2

(region), population (cabbage field or the 25 regions), statistic/parameter (proportion or count of damaged plants).

Solution 2

(b) Mark scheme

- The selection of row E is likely to provide an overestimate of the proportion of all cabbage plants in the field that are damaged by aphids. If the farmer's belief that aphid damage is greater for regions closer to the river is correct, then row E — the row of regions located closest to the river — is likely to have a greater proportion of cabbage plants damaged by aphids than regions farther from the river, so examining only row E would likely overestimate the field-wide proportion. Full credit needs: (1) indicates the selection of row E is likely to produce an overestimate, (2) justifies this based on row E's location as closest to the river, (3) links that location to the farmer's belief to explain why it produces an overestimate.

Solution 2

(c) Mark scheme

- Write the region numbers for row A (1 through 5) onto same-size slips of paper, put them in a hat, mix well, and select one — this is the region chosen from row A. Repeat this process independently for each other row: row B (6 through 10), row C (11 through 15), row D (16 through 20), and row E (21 through 25), selecting one region number from each row. This results in the selection of exactly one region from each of the five rows (a stratified random sample using the rows as strata). The farmer then examines every cabbage plant in each of the 5 selected regions for aphid damage to determine the proportion damaged. (Equivalent alternative: use a random number generator to produce one two-digit integer from each range 01–05, 06–10, 11–15, 16–20, and 21–25, and select the

Solution 2

corresponding regions.) Full credit (2 marks) requires all 4 components: (1) describes a random selection process using the row groupings as strata, (2) selection of a region within each stratum is equally likely, (3) selections across the five strata are independent of one another, (4) the process results in exactly one region selected from each row/stratum.

Question 3

2024 ▪ Q3

A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(a) Is this an observational study or an experiment? Justify your answer in context.

Question 3

2024 ▪ Q3

A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(b) Model D has an autopilot feature, in which the car controls its own motion with human supervision. James owns a Model D car and will investigate whether using the autopilot feature results in higher mileage than not using the autopilot. James will drive his car on 70 different days to and from work, using the same route at the same time each day. James will record the mileage each day.

James will use a completely randomized design to conduct his investigation. Describe an appropriate method James could use to randomly assign the two treatments, driving

Question 3

using the autopilot feature and driving without using the autopilot feature, to 35 days each.

Question 3

2024 ▪ Q3

A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(c) After the investigation was completed, James verified that the conditions for inference were met and conducted a hypothesis test. He discovered the mean mileage when using the autopilot feature was significantly higher than the mean mileage when not using the autopilot feature.

James is a member of a Model D club with thousands of members who all drive Model D cars. He will give a presentation at a Model D club members' meeting later this year and would like to state that the results of his hypothesis test apply to all Model D cars in his club. Another member of the club who is a statistician tells James his findings

Question 3

do not apply to all Model D cars in the club. What change would James need to make to his original study to be able to generalize to all Model D cars in the club?

Solution 3

(a)

Mark scheme

- This is an observational study. The researchers had the car owners estimate their own mileage; the car owners were not randomly assigned a car model, so no treatment was imposed by the researchers. Full credit needs: (1) identifies "observational study" (stating it is an experiment scores Incorrect), (2) justification based on no treatment being imposed (e.g. "no random assignment of treatment"), (3) context – car models, drivers, cars, owners, or members (mileage alone is not sufficient context).

Solution 3

(b) Mark scheme

- Number the 70 days in the experiment from 1 to 70. Using a random number generator, generate 35 unique integers from 1 to 70 inclusive (no repeats). Assign the days with those 35 unique integers for James to drive the car WITH autopilot, and assign the remaining 35 days for James to drive WITHOUT autopilot. (Equivalent alternative: label 70 equally-sized slips of paper, 35 ”with autopilot” and 35 ”without autopilot,” mix them in a bag, and each day select one slip without replacement to determine that day’s driving method.) Full credit needs: (1) appropriate labels created for the units/treatments (the 70 days, and the two treatments), (2) a correctly described random process so every possible random assignment is equally likely (a coin-flip-with-stopping-rule method does NOT

Solution 3

satisfy this, since it biases the last few days), (3) the process results in exactly 35 days assigned to each treatment, using selection without replacement/without repeats.

Solution 3

(c) Mark scheme

- In order to generalize his findings to all Model D cars in his club, James would need to randomly select Model D cars from the club, then carry out a similar study using the Model D cars that were randomly sampled from the club. Full credit needs: (1) states that more cars must be sampled, (2) states that random sampling is required for generalization, (3) is in context, including sampling from the population of interest (Model D cars/owners/members in his club). Any discussion of experimental design beyond sample selection (e.g. conditions, two-sample design) is ignored in scoring.

Question 2

2021 ▪ Q2

Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(a) One aspect of the study is to record the number of miles each subject walks per day. The researchers are deciding whether to have subjects wear an activity tracker to record the data or to have subjects keep a daily journal of the miles they walk each day. Describe what bias could be introduced by keeping the daily journal instead of wearing the activity tracker.

Question 2

2021 ▪ Q2

Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(b) During the course of the study, the subjects will have their cholesterol levels measured each month by a doctor. The researchers will perform a significance test at the end of the study to determine whether the average cholesterol level for subjects who walk fewer miles each day is greater than for those who walk more miles each day. Selecting a random sample creates a reasonable representative sample of the target population. Explain the benefit of using a representative sample from the population.

Question 2

2021 ▪ Q2

Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(c) Suppose the researchers conduct the test and find a statistically significant result. Would it be valid to claim that increased walking causes a decrease in average cholesterol levels for adults in the target population? Explain your reasoning.

Solution 2

(a) Mark scheme

- Keeping a daily journal relies on subjects self-reporting, and subjects may not accurately recall or may forget to log all their walking, leading to systematic (consistent) underreporting – or overreporting – of daily miles walked, unlike an activity tracker which records automatically. This introduces response bias: if most subjects underreport, the estimate of mean daily miles walked for the target population would be biased too low (or biased high if overreporting). Credit requires (1) stating that journaling could cause a bias avoided by trackers with a reasonable explanation (self-reporting/forgetting), AND (2) describing the direction of the bias (systematic under- or over-reporting, or a biased – too low/too high – estimate of the population parameter).

Solution 2

(b)

Mark scheme

- A representative sample is necessary to make a valid, unbiased inference/generalization about the target population – e.g., to estimate cholesterol levels in the population or the association between cholesterol level and amount of walking for all adults in the target population, not just the people sampled. Credit requires (1) explaining that a representative sample is needed for valid generalization to the population, AND (2) referring to estimation/inference about a population cholesterol parameter or the walking-cholesterol association.

Solution 2

(c) Mark scheme

- No. Because amounts of walking were not randomly assigned to subjects (this is an observational study, not a controlled experiment), a statistically significant result does not allow a causal claim. The researchers can only conclude that cholesterol level is associated with (not caused by) daily miles walked. There may be a confounding variable – e.g., diet – associated with both amount of walking and cholesterol level (people who walk more may also eat healthier, lower-cholesterol diets), so the observed association could actually be driven by that confounder rather than by walking itself. Credit requires (1) stating a causal claim cannot be made, AND (2) a valid explanation based on lack of random

Solution 2

assignment / observational-study status / a plausible confounding variable linked to both walking and cholesterol.

Question 2

2018 ▪ Q2

An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.
[Boxed question text: "Do you regularly recycle plastic bottles?"]
Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as $(0.584,\ 0.816)$.
(a) How many students were in the sample selected by the environmental science teacher?

Question 2

2018 ▪ Q2

An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.

[Boxed question text: "Do you regularly recycle plastic bottles?"]

Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as $(0.584,\ 0.816)$.

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

Question 2

2018 ▪ Q2

An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.

[Boxed question text: "Do you regularly recycle plastic bottles?"]

Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as $(0.584,\ 0.816)$.

(c) The statistics teacher at the high school was concerned about the potential bias in the survey. To obtain a potentially less biased estimate of the proportion, the statistics teacher used an alternate method for collecting student responses. A random sample

Question 2

of 300 students was selected, and each student was given the following instructions on how to respond to the question.

- In private, flip a fair coin.
- If heads, you must respond no, regardless of whether you regularly recycle.
- If tails, please truthfully respond yes or no.

(c)(i) What is the expected number of students from the sample of 300 who would be required to respond no because the coin flip resulted in heads?
(c)(ii) The results of the sample showed that 213 of the 300 selected students responded no. Based on the results of the sample, give a point estimate for the proportion of all students at the high school who would respond yes to the question.

Solution 2

(a) Mark scheme

- Using $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$ with the given interval (0.584, 0.816): $\hat{p} = (0.584 + 0.816)/2 = 0.7$, margin of error $= 0.816 - 0.7 = 0.116$, and $z^* = 1.96$ (95

Solution 2

(b) Mark scheme

- Bias may be introduced because students responded directly, in person, to the environmental science teacher — an authority figure known to care about the environment — rather than answering anonymously. Students might say 'yes' (that they recycle) even when they don't, because they don't want to admit to the teacher that they don't recycle. This would make the point estimate of the proportion who would respond yes GREATER than the true proportion of all students who would respond yes. Full credit requires three things: (1) a reason for bias tied specifically to the teacher personally asking/collecting responses (not question wording, voluntary response, or ordinary sampling variability), (2) an explicit contrast between what students would say versus what they actually do

Solution 2

(e.g., 'students might say yes when they actually don't recycle'), and (3) a description of the effect on the point estimate that does not contradict the bias described (here: the estimate is too high).

Solution 2

(c)

Mark scheme

- This is the stem/setup text introducing the statistics teacher's randomized-response method (private coin flip; heads forces a 'no' answer, tails means answer truthfully) — it has no scorable content of its own beyond what is captured and graded in parts (c-i) and (c-ii), which the SG scores together as a single part (c). No independent marks are allocated to this stem; see 2ci and 2cii for the full method and scoring.

Solution 2

(c)(i)

Mark scheme

- Each of the 300 students privately flips a fair coin; a head (probability 1/2) forces the student to respond 'no' regardless of their true recycling habit. Expected number of students forced to respond no because of heads $= 300 \times \frac{1}{2} = 150$. Full credit requires a single number, 150 (an approximate phrasing like 'about 150' or '2248150' is acceptable; a range such as '147-153' is not).

Solution 2

(c)(ii) Mark scheme

- 150 students are expected to be forced to say no by a head, and the other 150 are expected to answer truthfully. Of the 213 students who actually answered no, $213 - 150 = 63$ are attributed to truthful 'no' responses. That leaves $150 - 63 = 87$ of the truthful-tails students who must have truthfully answered yes. Point estimate for the proportion of ALL students at the school who would respond yes: $87/150 = 0.58$. Full credit requires this part-c pair to be scored together: an answer of 150 in (c-i) together with an answer of 0.58 (or an equivalent verbal form, e.g. '87 out of 150') in (c-ii), with supporting work. The response must clearly present 0.58 as a point estimate of the POPULATION

Solution 2

proportion (not label it as itself the exact population proportion, and not give a confidence interval as the final answer instead of the point estimate).

Question 3

2016 ▪ Q3

Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.
"Study Finds Strong Association Between Smoking and Alzheimer's"
The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

(a) Identify the explanatory and response variables in the study.
Explanatory variable:
Response variable:

(b) Is the study described in the article an observational study or an experiment? Explain.

Question 3

(c) Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.

Solution 3

(a)

Mark scheme

- Explanatory variable: the person's degree of cigarette smoking. Response variable: whether the person develops Alzheimer's disease during the course of the study. (Full credit requires both variables described with some indication of status/levels – e.g. smoking vs. not smoking, developing vs. not developing Alzheimer's – not just the bare topic names.)

Solution 3

(b)

Mark scheme

- Observational study – because the people in the study were not assigned a degree of cigarette smoking by the researchers. Instead, each person's smoking status was passively observed and recorded, not manipulated or assigned as a treatment.

Solution 3

(c) Mark scheme

- A confounding variable is related to the explanatory variable and possibly influences the response variable. Here, it is plausible that people who exercise more regularly are more health-conscious and therefore less likely to smoke cigarettes than people who do not exercise regularly (exercise status related to smoking status). It is also plausible that people who exercise more regularly are less likely to develop Alzheimer's disease than those who do not (exercise status may influence the response). If both relationships hold, smoking would appear associated with developing Alzheimer's simply because both are associated with exercise – even if there is no actual cause-and-effect relationship between smoking and Alzheimer's. Full credit requires both components: (1) a reasonable

Solution 3

explanation that exercise status is related to smoking status, and (2) a statement that exercise status might influence whether the person develops Alzheimer's disease.

Question 6

2015 ▪ Q6

Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.
[Two overlapping bell-shaped normal distribution curves labeled “Production Line A” (solid curve, centered at 5.9) and “Production Line B” (dashed curve, centered at 6.1), sharing a horizontal axis “Diameter (inches)” marked 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, and a vertical axis labeled “Probability”.]
The tortillas produced at the two production lines are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected per

Question 6

day at each corporation, and the diameters are measured. From the sample of 200 tortillas, the manager of the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

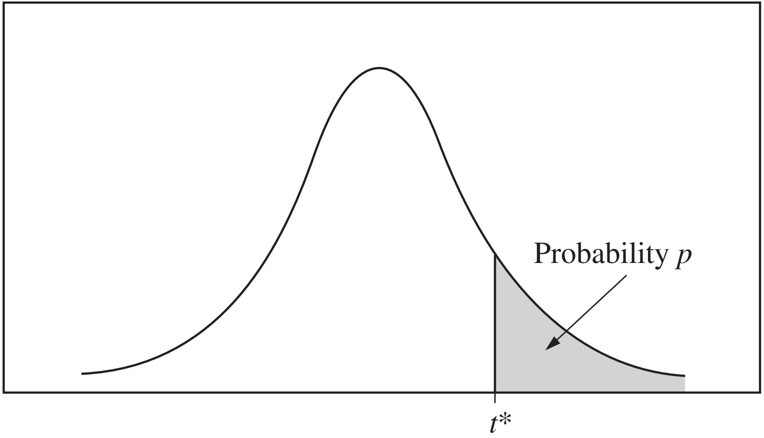
Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

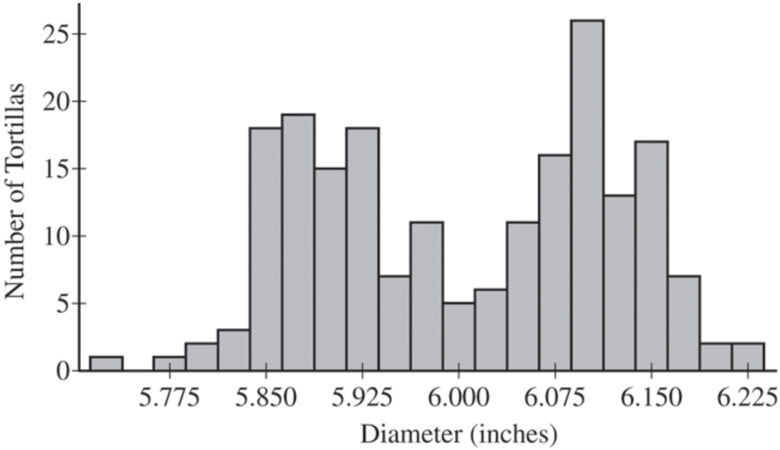
(a) Will a sample obtained by using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Why or why not.

Question 6

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.



Question 6


Number of Tortillas
0
5
10
15
20
25
5.775
5.850
5.925
6.000
6.075
6.150
6.225
Diameter (inches)

Question 6

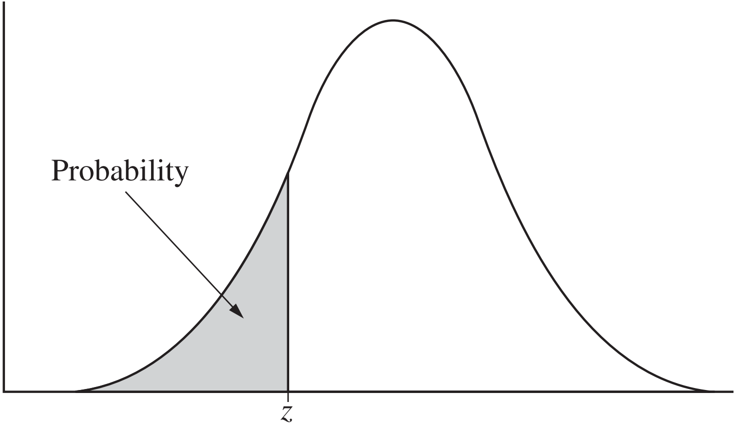


Table entry for z is the probability lying below z.

Question 6

2015 ▪ Q6

Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.
[Two overlapping bell-shaped normal distribution curves labeled “Production Line A” (solid curve, centered at 5.9) and “Production Line B” (dashed curve, centered at 6.1), sharing a horizontal axis “Diameter (inches)” marked 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, and a vertical axis labeled “Probability”.]
The tortillas produced at the two production lines are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected per day at each corporation, and the diameters are measured. From the sample of 200 tortillas, the manager of

Question 6

the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Choosing the shape of the histogram as a guide, explain which method, Method 1 or Method 2, was most likely used to obtain a such a sample. [Histogram titled with y-axis "Number of Tortillas" from 0 to 25 and x-axis "Diameter (inches)" with bins at approximately 5.775, 5.850, 5.925, 6.000, 6.075, 6.150, 6.225. The histogram is bimodal, with one cluster of taller bars centered around 5.850-5.925

Question 6

(peak near 20-22) and a second cluster of bars centered around 6.075-6.150 (peak near 20-22), with a smaller gap/dip around 6.000.]

Question 6

2015 ▪ Q6

Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.
[Two overlapping bell-shaped normal distribution curves labeled “Production Line A” (solid curve, centered at 5.9) and “Production Line B” (dashed curve, centered at 6.1), sharing a horizontal axis “Diameter (inches)” marked 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, and a vertical axis labeled “Probability”.]
The tortillas produced at the two production lines are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected per day at each corporation, and the diameters are measured. From the sample of 200 tortillas, the manager of

Question 6

the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas on the sample on a given day? Explain why.

Question 6

2015 ▪ Q6

Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.
[Two overlapping bell-shaped normal distribution curves labeled “Production Line A” (solid curve, centered at 5.9) and “Production Line B” (dashed curve, centered at 6.1), sharing a horizontal axis “Diameter (inches)” marked 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, and a vertical axis labeled “Probability”.]
The tortillas produced at the two production lines are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected per day at each corporation, and the diameters are measured. From the sample of 200 tortillas, the manager of

Question 6

the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(d)(i) Each day, the distribution of the 200,000 tortillas made that day has mean diameter 6 inches with standard deviation 0.11 inch.

For samples of size 200 taken from one day's production, describe the sampling distribution of the sample mean diameter for samples that are obtained using Method 1.

Question 6

(d)(ii) Suppose that one of the two sampling methods will be selected and used every day for one year (365 days). The sample mean of the 200 diameters will be recorded each day. Which of the two methods will result in less variability in the distribution of the 365 sample means? Explain.

Question 6

2015 ▪ Q6

Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.
[Two overlapping bell-shaped normal distribution curves labeled “Production Line A” (solid curve, centered at 5.9) and “Production Line B” (dashed curve, centered at 6.1), sharing a horizontal axis “Diameter (inches)” marked 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, and a vertical axis labeled “Probability”.]
The tortillas produced at the two production lines are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected per day at each corporation, and the diameters are measured. From the sample of 200 tortillas, the manager of

Question 6

the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(e) A government inspector will visit the factory on June 22 to observe the sampling and to determine if the factory is in compliance with the advertised mean diameter of 6 inches. The manager knows that, with both sampling methods, the sample mean is an unbiased estimator of the population mean. However, the manager is unsure which method is more likely to produce a sample mean that is close to 6 inches on the day of sampling. Based on your previous answers, which of the two sampling methods,

Question 6

Method 1 or Method 2, is more likely to produce a sample mean close to 6 inches? Explain.

Solution 6

(a)

Mark scheme

- No. A sample obtained using Method 2 will not be representative of the population of all tortillas made that day - it represents only the tortillas from one production line, not the whole population, because the diameter distributions differ between the two production lines.

Solution 6

(b)

Mark scheme

- Method 1 was most likely used. The bimodal shape of the histogram shows tortillas were selected from both production lines, consistent with Method 1; Method 2 would tend to produce a unimodal distribution centered at either 5.9 or 6.1 inches, not the observed bimodal shape.

Solution 6

(c)

Mark scheme

- Method 2 would give less variability in the sample of 200 tortillas on a given day, because that sample comes from only one production line. Since the two lines' diameter distributions differ, sampling from both lines (Method 1) produces more variable sample data.

Solution 6

(d)(i)

Mark scheme

- The sampling distribution of the sample mean diameter for samples of size 200 obtained using Method 1 is approximately Normal, with mean 6 inches and standard deviation $\dfrac{0.11}{\sqrt{200}} \approx 0.0078$ inch.

Solution 6

(d)(ii)

Mark scheme

- Method 1 would give less variability in the 365 sample means. With Method 2, roughly half of the 365 daily sample means will be close to 5.9 inches (production line A) and the other half close to 6.1 inches (production line B) - a bimodal spread. With Method 1, the sample means will all be close to 6 inches, tightly clustered as shown by the small standard deviation (≈ 0.0078 inch) from part (d).

Solution 6

(e)

Mark scheme

- Method 1. On a single day, Method 1 is more likely to produce a sample mean close to the true mean of 6 inches. Since the sample mean is an unbiased estimator of the population mean under both methods, the manager should pick the method with less variability in the distribution of the sample mean - which, based on the previous part, is Method 1.

Question 4

2014 ▪ Q4

As part of its twenty-fifth reunion celebration, the class of 1988 (students who graduated in 1988) at a state university held a reception on campus. In an informal survey, the director of alumni development asked 50 of the attendees about their incomes. The director computed the mean income of the 50 attendees to be $189,952. In a news release, the director announced, "The members of our class of 1988 enjoyed resounding success. Last year's mean income of its members was $189,952!"
(a) What would be a statistical advantage of using the median of the reported incomes, rather than the mean, as the estimate of the typical income?

Question 4

2014 ▪ Q4

As part of its twenty-fifth reunion celebration, the class of 1988 (students who graduated in 1988) at a state university held a reception on campus. In an informal survey, the director of alumni development asked 50 of the attendees about their incomes. The director computed the mean income of the 50 attendees to be $189,952. In a news release, the director announced, "The members of our class of 1988 enjoyed resounding success. Last year's mean income of its members was $189,952!"

(b) The director felt the members who attended the reception may be different from the class as a whole. A more detailed survey of the class was planned to find a better estimate of the income as well as other facts about the alumni. The staff developed two methods based on the available funds to carry out the survey.

Method 1: Send out an e-mail to all 6,826 members of the class asking them to complete an online form. The staff estimates that at least 600 members will respond.

Question 4

Method 2: Select a simple random sample of members of the class and contact the selected members directly by phone. Follow up to ensure that all responses are obtained. Because method 2 will require more time than method 1, the staff estimates that only 100 members of the class could be contacted using method 2.

Which of the two methods would you select for estimating the average yearly income of all 6,826 members of the class of 1988? Explain your reasoning by comparing the two methods and the effect of each method on the estimate.

Solution 4

(a) Mark scheme

- (Q4 is graded holistically across three official sections — section 1 = part (a), section 2 = part (b) component 1, section 3 = part (b) component 2 — via a combination table into an overall 0-4 score; this leaf covers section 1 only, so it carries about 1 of the 4 total marks.) The median is less affected by skewness and outliers than the mean. With a variable such as income, a small number of very large incomes could dramatically increase the mean but not the median, so the median would provide a better estimate of a typical income value. Full credit needs two components: (1) describes how skewness or outliers affect the mean but not (or less than) the median (e.g. 'the mean is affected by skewness/outliers,' 'the median is not affected by skewness/outliers,' or 'the mean

Solution 4

is greater/less than the median when there is right/left skewness or outliers'), (2) makes a conjecture about a relevant characteristic of the distribution of incomes, such as skewness or an outlier. Partial credit for only one of the two components. Component (1) is NOT satisfied by unexplained claims such as 'don't use the mean for skewed distributions,' 'use the median for skewed distributions,' or an incorrect statement about means/medians (e.g. claiming the median is higher than the mean for right-skewed distributions). A single sentence can satisfy both components (e.g. 'If there was a billionaire in the sample, the mean would be higher than the median'). If a response instead argues that the mean is the more appropriate estimate of typical income, this leaf is scored one level down from what it otherwise would earn.

Solution 4

(b) Mark scheme

- (This leaf covers sections 2 and 3 of Q4's three-section holistic scoring — about 3 of the question's 4 total marks.) Method 2 is better than Method 1. A sample obtained via Method 1 (e-mail everyone, ~600 respond) could be biased because of the voluntary nature of the response — it is plausible that class members with larger incomes would be more likely to respond, so the Method-1 sample mean would tend to overestimate the mean income of all class members. With Method 2 (a smaller random sample with follow-up), despite the smaller sample size, the random selection is likely to produce a sample that is more representative of the entire class, giving an unbiased estimate of the class's mean income — but if some of the members randomly selected for Method 2 do not respond, the incomes of

Solution 4

responders may differ from the incomes of non-responders (e.g. if members with larger incomes are more or less willing to respond), and this nonresponse bias may produce a misleading estimate/conclusion about the mean income, including a direction consistent with whichever bias is identified (e.g. 'the sample mean is likely to be higher than the mean of the population'). Full credit needs, for section 2: chooses Method 2 AND identifies a relevant characteristic of Method 1 (its voluntary/self-selected nature is likely biased) AND a relevant characteristic of Method 2 (uses random selection, so is more representative) — responses that compare the two methods, or that only discuss negative characteristics of Method 1 and positive characteristics of Method 2, can satisfy both components even without explicitly stating 'Method 2'; and for section 3: indicates the incomes of responders may differ from the incomes of non-responders AND indicates the biased sampling method may produce a misleading estimate/conclusion about the

Solution 4

mean income, including direction. A single sentence can satisfy both the first component of section 2 and the first component of section 3 (e.g. 'In Method 1, rich people are more likely to respond.'). Responses that focus on the larger sample size of Method 1 can still satisfy the section-3 direction component if they describe the effect as reducing the variability of the estimate; responses focused on untruthful survey answers can satisfy it if the effect on the estimate is stated appropriately. Discussions of conditions for inference are extraneous and should be ignored. Partial credit is earned section-by-section for providing only one of a section's two components (or, for section 2, for choosing Method 1 outright, which cannot earn credit for that section); this leaf's overall credit should reflect the AP holistic combination of sections 2 and 3.