Past-paper walk-through

Random Sampling and Data Collection ▪ 3.3

eXercise

Questions in this set

- 2025 Q2
- 2021 Q2
- 2019 Q6
- 2015 Q6

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

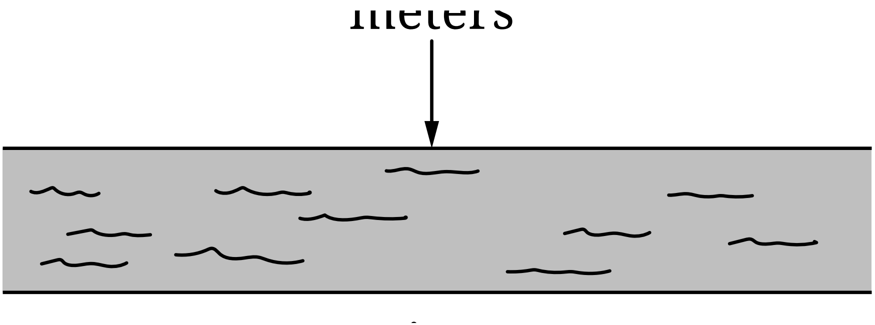
regions in the cabbage field closer to the river. To obtain the estimate, the farmer is considering three sampling methods.

- 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 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 6

2019 ▪ Q6

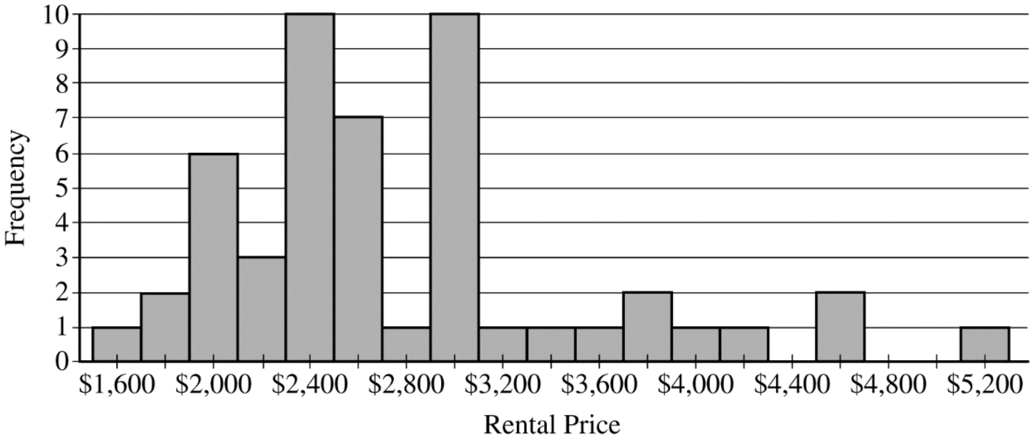
Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

Question 6

Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

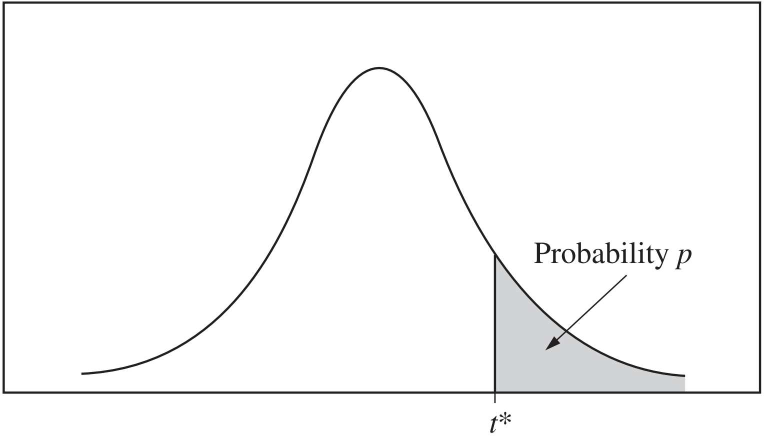
Question 6



Frequency
10
9
8
7
6
5
4
3
2
1
0
$1,600
$2,000
$2,400
$2,800
$3,200
$3,600
$4,000
$4,400
$4,800
$5,200
Rental Price

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

2019 ▪ Q6

Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(b) The distribution of the 50 rental prices of the available apartments is shown in the following histogram.

[Histogram titled "Rental Price"; y-axis "Frequency" from 0 to 10; x-axis with tick marks at \$1,600, \$2,000, \$2,400, \$2,800, \$3,200, \$3,600, \$4,000, \$4,400, \$4,800, \$5,200.]

Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?

Question 6

2019 ▪ Q6

Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(c) Instead of using the sample median as the point estimate for the population mean, Emma wants to explore using the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, that she can use to approximate the sampling distribution of the sample medians of size 50. Instead, Emma decides to use a theoretical sampling distribution to approximate the sampling distribution of the sample medians of size 50.

Question 6

Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

- Take a random sample, with replacement, of size 50 from the original sample.
- Calculate and record the median of the sample.
- Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table shows the results of generating 15,000 medians.

Bootstrap Distribution of Medians

Median	**Frequency**	**Median**	**Frequency**	**Median**	**Frequency**
2,345	3	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	2	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65

Question 6

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

Question 6

2019 ▪ Q6

Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(d)(i) Use the frequency table to find the following.
Value of the 5th percentile.

(d)(ii) Use the frequency table to find the following.
Value of the 95th percentile.

Question 6

2019 ▪ Q6

Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(e) Find the percentage of bootstrap medians in the table that are equal to or between those found in part (d).

Question 6

2019 ▪ Q6

Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

Solution 6

(a)

Mark scheme

- Because random sampling was used, the results of the sample may be generalized to the population of rental prices for one-bedroom apartments in the city that are listed on this particular website at the time the sample was taken.

Solution 6

(b) Mark scheme

- Because the distribution of the 50 rental prices in the sample is skewed to the right, the sample median provides a better indicator of typical rental prices than the sample mean. Some very large (high) rental prices pull the sample mean up, so the sample mean would overestimate the typical rental price, whereas the sample median would give a more accurate representation of a typical rental price. Must both (1) identify that using the mean overestimates the typical rental price, and (2) link this disadvantage to a feature of the distribution's shape (skewness) evident in the histogram – merely noting the mean is larger than the median is not by itself sufficient.

Solution 6

(c)

Mark scheme

- To determine the sampling distribution of median rental prices for random samples of size 50 from this population, Emma would need to obtain every possible sample of 50 one-bedroom apartments from this population (i.e., from the website's full listing population) and compute the median of each sample. The collection of all of these possible sample medians is the theoretical sampling distribution of the sample median.

Solution 6

(d)(i)

Mark scheme

- $(0.05)(15{,}000)=750$. The 5th percentile is a value $x_{0.05}$ such that at least 750 of the 15,000 bootstrap sample medians in the table are less than or equal to $x_{0.05}$. Cumulating frequencies from the smallest sample median in the table (going down the columns) until first reaching 750 values gives $x_{0.05}=\$2{,}500$.

Solution 6

(d)(ii)

Mark scheme

- $(0.95)(15{,}000)=14{,}250$. Similarly, cumulating frequencies from the smallest value until reaching 14,250 (i.e., leaving 750 values above) gives the 95th percentile $x_{0.95}=\$2{,}950$.

Solution 6

(e)

Mark scheme

- The percentage of the 15,000 bootstrap medians that are equal to or between the part-(d) values $(2,500 \mathit{and} 2{,}950)$ is $\frac{14{,}404}{15{,}000} \times 100\% \approx 96.03\%$. (Any answer using plausible part-(d) values between 2,345 and 3,062.5, computed consistently, is acceptable.)

Solution 6

(f) Mark scheme

- Using the results from parts (d) and (e), an approximate 96 percent confidence interval for the median rental price of all one-bedroom apartments listed on this website for this city is $(2,500$,2,950). Interpretation: we are approximately 96 percent confident that the median rental price of all one-bedroom apartments listed on this website for this city is between $2,500\mathit{and}$2,950. Full credit requires: (1) using $2,500\mathit{and}$2,950 (the part-(d) percentile values) as the interval endpoints, (2) stating an approximate 90% or 96% confidence level (both accepted, consistent with part (e)), (3) a correct statement in context that the interval is specifically for the median (not the mean) rental price.

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.

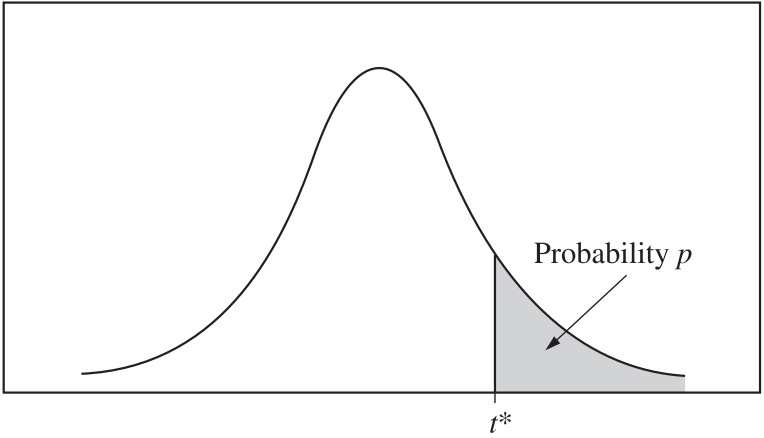
Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

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.

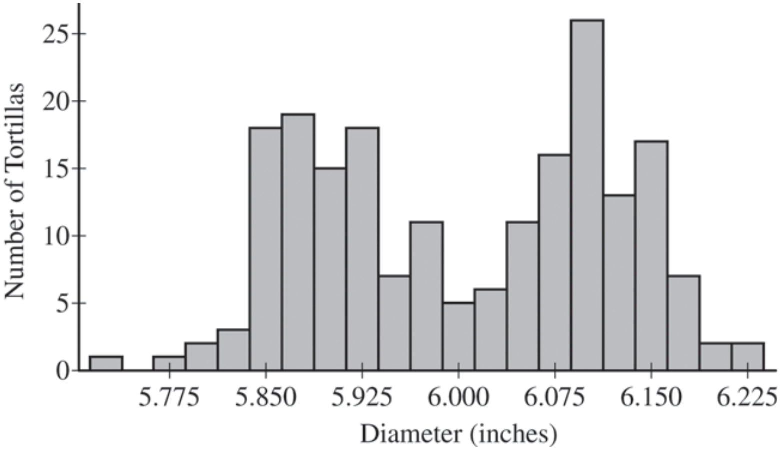
(a) Will a sample obtained by using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Why or why not.

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

Question 6

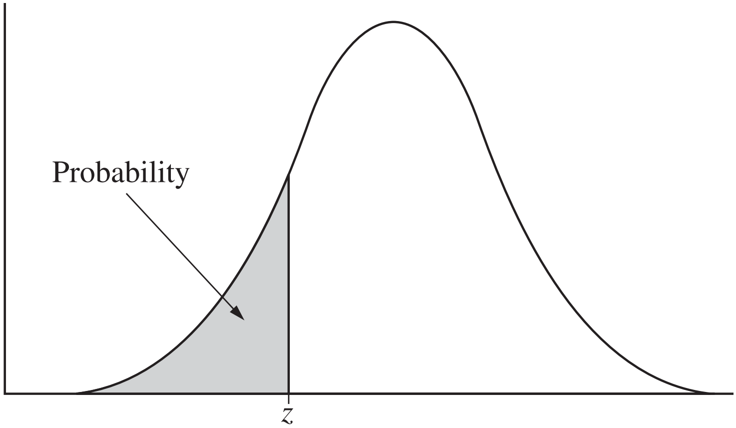


Table entry for z is the probability lying below z.

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.