# Past-paper walk-through

## Biased and Unbiased Point Estimates ▪ 5.4

eXercise

# Questions in this set

- 2019 Q6
- 2018 Q2
- 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 6

2019 ▪ Q6

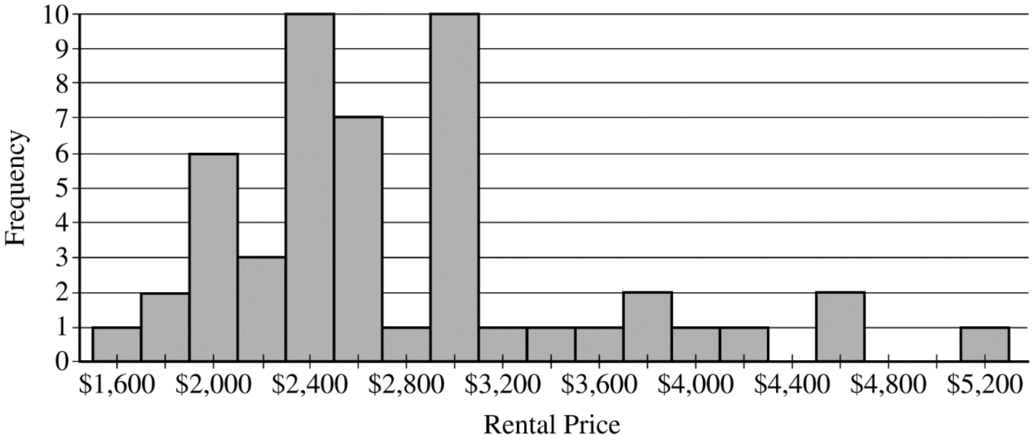
Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

**(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 |

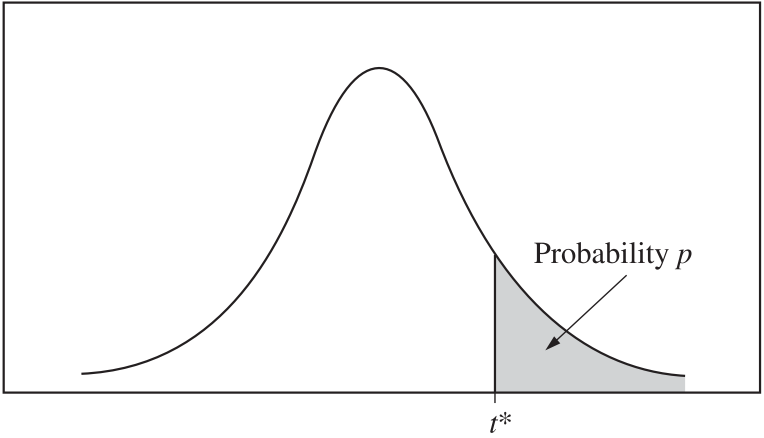
# Question 6



10
9
8
7
6
5
4
3
2
1
0
Frequency
$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 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |

# 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 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 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.

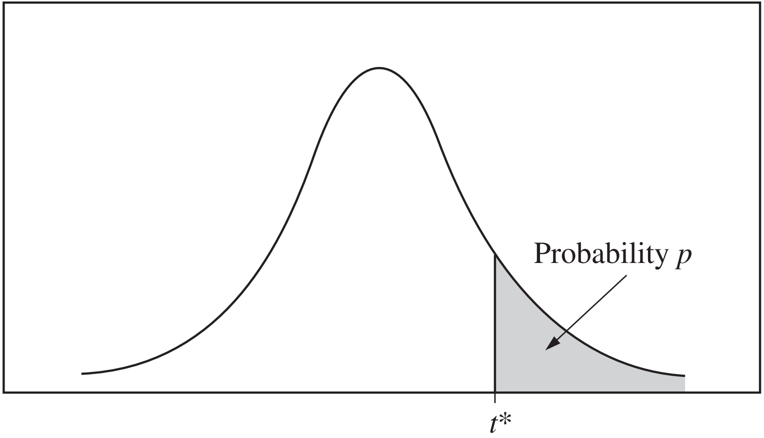
Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

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.

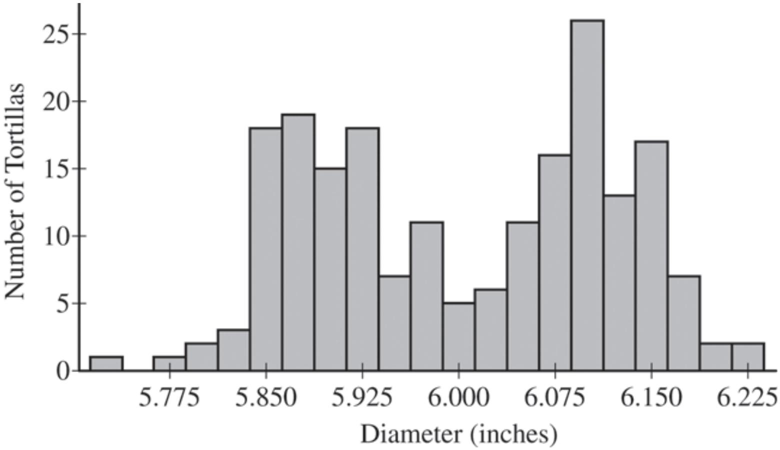
**(a)** Will a sample obtained by using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Why or why not.

# 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

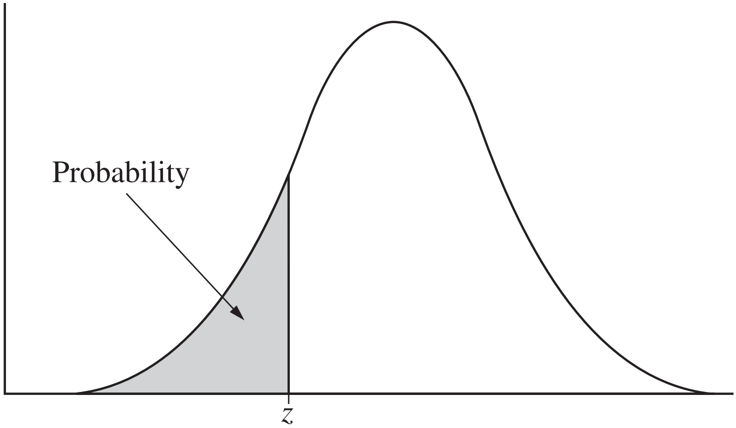



Table entry for $z$ is the probability lying below $z$.

# 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 ($\approx 0.0078$ 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.