Past-paper walk-through

Constructing a Confidence Interval for a Population Proportion ▪ 6.2

eXercise

Questions in this set

- 2022 Q4
- 2019 Q6
- 2018 Q2
- 2017 Q2
- 2016 Q5
- 2015 Q2

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 4

2022 ▪ Q4

A survey conducted by a national research center asked a random sample of 920 teenagers in the United States how often they use a video streaming service. From the sample, 59% answered that they use a video streaming service every day.

(a) Construct and interpret a 95% confidence interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.

(b) Based on the confidence interval in part (a), do the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would respond that they use a video streaming service every day is not 0.5? Justify your answer.

Solution 4

(a) Mark scheme

- Procedure: a one-sample z-interval for the proportion of all teenagers in the United States who would say they use a video streaming service every day; the parameter of interest is the population proportion. Conditions: a random sample of 920 teenagers was selected, which is enough for the interval to generalize to the population; 920 is less than 10% of all teenagers in the United States (independence, sampling without replacement from a finite population); there are more than 10 successes and 10 failures, since $n\hat{p} = (920)(0.59) \approx 542.8$ and $n(1-\hat{p}) = (920)(0.41) \approx 377.2$, so the sampling distribution of $\hat{p}$ is approximately normal. Calculation: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 0.59 \pm 1.96\sqrt{\dfrac{(0.59)(0.41)}{920}} = 0.59 \pm 0.032$, giving the interval

Solution 4

$(0.558,\ 0.622)$. Interpretation: we can be 95% confident that the true proportion of all teenagers in the United States who would say they use a video streaming service every day is between 0.558 and 0.622. This leaf bundles the three sections that together make up part (a) (worth 3 of the question's 4 raw points): Section 1 – identifies the appropriate procedure (one-sample *z*-interval, by name or by the endpoint calculations) AND states the parameter is the population proportion; Section 2 – verifies all three conditions (random sample; 10% condition; large-counts condition with the actual computed values 542.8 and 377.2) AND reports the correct interval endpoints (or values consistent with the stated procedure); Section 3 – states 95% confidence with a correct interpretation in context (the population of U.S. teenagers, not merely the sample of 920, and not implying the population proportion itself is random).

Solution 4

(b) Mark scheme

- The 95% confidence interval $(0.558,\ 0.622)$ indicates that any value between 0.558 and 0.622 is a plausible value for the proportion of all teenagers in the United States who use a video streaming service every day. Because 0.5 is NOT contained in the interval, the sample data provide convincing statistical evidence that the true proportion of all teenagers in the United States who would say they use a video streaming service every day is not 0.5. Full credit (Section 4 of the question) needs both: (1) a correct conclusion (convincing evidence, or not convincing evidence, that 0.5 is a plausible value) consistent with the interval calculated in part (a), (2) a correct justification based on whether 0.5 is contained in that interval. A hypothesis test may be used to reach the conclusion, but a

Solution 4

hypothesis test alone cannot satisfy the justification component (which must reference the confidence interval from part (a)).

Question 6

2019 ▪ Q6

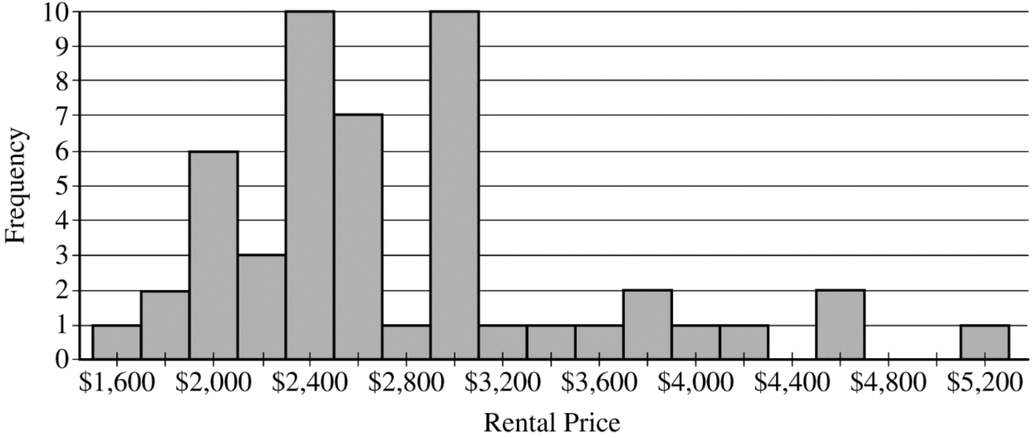
Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(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

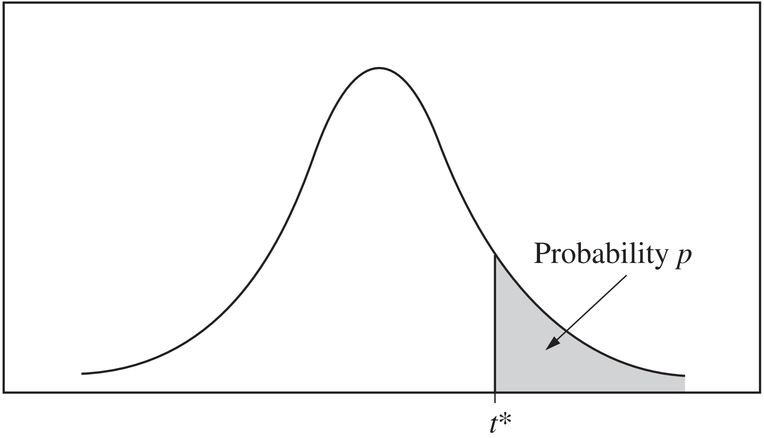
Question 6



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 2

2017 ▪ Q2

2. The manager of a local fast-food restaurant is concerned about customers who ask for a water cup when placing an order but fill the cup with a soft drink from the beverage fountain instead of filling the cup with water. The manager selected a random sample of 80 customers who asked for a water cup when placing an order and found that 23 of those customers filled the cup with a soft drink from the beverage fountain.

   (a) Construct and interpret a 95 percent confidence interval for the proportion of all customers who, having asked for a water cup when placing an order, will fill the cup with a soft drink from the beverage fountain.

   (b) The manager estimates that each customer who asks for a water cup but fills it with a soft drink costs the restaurant \$0.25. Suppose that in the month of June 3,000 customers ask for a water cup when placing an order. Use the confidence interval constructed in part (a) to give an interval estimate for the cost to the restaurant for the month of June from the customers who ask for a water cup but fill the cup with a soft drink.

Solution 2

(a) Mark scheme

- The appropriate procedure is a one-sample z-interval for a population proportion p (population $=$ all customers of the restaurant who ask for a water cup; $p =$ proportion of that population who fill the cup with a soft drink). Formula: $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$.
- Conditions: (1) Random sample – satisfied because the stem states a random sample of customers who asked for a water cup was used; (2) Large sample – number of successes $n\hat{p} = 23 \geq 10$ and number of failures $n(1-\hat{p}) = 57 \geq 10$, both satisfied (the response must use the actual numerical counts, not just assert the condition in words).

Solution 2

- Mechanics: $\hat{p} = 23/80 = 0.2875$. Confidence interval $= 0.2875 \pm 1.96\sqrt{0.2875(1-0.2875)/80} = 0.2875 \pm 1.96(0.0506) = 0.2875 \pm 0.0992$, i.e. (0.1883, 0.3867).
- Interpretation: "We can be 95 percent confident that, in the population of all customers of the restaurant who ask for a water cup, the proportion who will fill it with a soft drink is between 0.1883 and 0.3867." A full interpretation needs: (i) conveys inference about the population proportion (not just the sample of 80 customers); (ii) clearly identifies the parameter as the proportion of the water-cup population that fills the cup with a soft drink; (iii) mentions 95 percent confidence and interprets it correctly (e.g. "we can be 95 percent confident" – NOT as a 95 percent probability/chance that the interval captured the true proportion, which is incorrect).

Solution 2

- This leaf combines the three components College Board scores separately (procedure+conditions, correct interval, and interpretation), each worth 1 point in this scheme, so this leaf is worth 3 of the question's 4 points.

Solution 2

(b) Mark scheme

- Using the interval from part (a), a 95 percent interval estimate for the number of the 3,000 June customers who ask for a water cup but fill it with a soft drink is 3000×0.1883 to 3000×0.3867, i.e. 565 to 1,160 customers. At a cost of \$0.25 per such customer, the interval estimate for the cost to the restaurant in June is $565 \times \$0.25 = \$141.25$ to $1160 \times \$0.25 = \$290.00$, i.e. \$141.25 to \$290.00. Full credit needs the correct final dollar interval AND work shown for how it was obtained (multiplying the proportion-based customer-count interval by \$0.25). A response that correctly follows this method from an incorrect part (a) interval still earns full credit for this leaf.

Question 5

2016 ▪ Q5

A polling agency showed the following two statements to a random sample of 1,048 adults in the United States.
Environment statement: Protection of the environment should be given priority over economic growth.
Economy statement: Economic growth should be given priority over protection of the environment.
The order in which the statements were shown was randomly selected for each person in the sample.
After reading the statements, each person was asked to choose the statement that was most consistent with his or her opinion. The results are shown in the table.

	Environment Statement	**Economy Statement**	**No Preference**
Percent of sample	58%	37%	5%

Question 5

(a) Assume the conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the proportion of all adults in the United States who would have chosen the economy statement.

Question 5

Environment statement: Protection of the environment should be given priority over economic growth.

Economy statement: Economic growth should be given priority over protection of the environment.

Question 5

	Environment Statement	Economy Statement	No Preference
Percent of sample	58%	37%	5%