Past-paper walk-through

Justifying a Claim Based on a Confidence Interval for a Population Proportion ▪ 6.3

eXercise

Questions in this set

- 2022 Q4
- 2019 Q6
- 2017 Q2
- 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

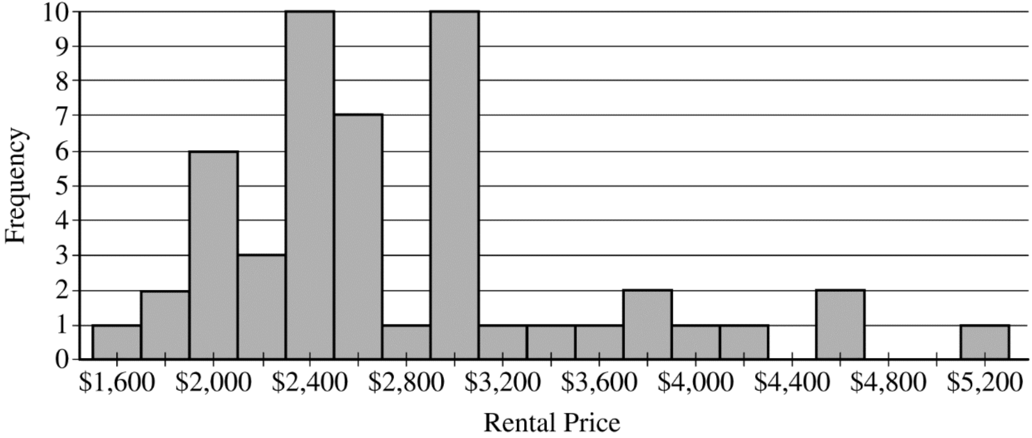
Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

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

Question 6

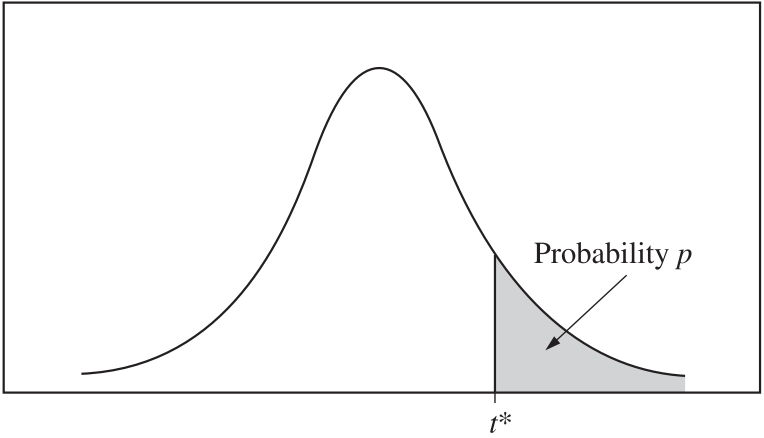
Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	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


Frequency
0
1
2
3
4
5
6
7
8
9
10
$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 ($2,500 and$ 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 and$2,950. Full credit requires: (1) using $2, 500 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

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 2

2015 ▪ Q2

To increase business, the owner of a restaurant is running a promotion in which a customer's bill can be randomly selected to receive a discount. When a customer's bill is printed, a program in the cash register randomly determines whether the customer will receive a discount on the bill. The program was written to generate a discount with a probability of 0.2, that is, giving 20 percent of the bills a discount in the long run. However, the owner is concerned that the program has a mistake that results in the program not generating the intended long-run proportion of 0.2.

The owner selected a random sample of bills and found that only 15 percent of them received discounts. A confidence interval for p, the proportion of bills that will receive a discount in the long run, is 0.15 ± 0.06. All conditions for inference were met.

Question 2

(a)(i) Consider the confidence interval 0.15 ± 0.06. Does the confidence interval provide convincing statistical evidence that the program is not working as intended? Justify your answer.

(a)(ii) Consider the confidence interval 0.15 ± 0.06. Does the confidence interval provide convincing statistical evidence that the program generates the discount with a probability of 0.2? Justify your answer.

Question 2

2015 ▪ Q2

To increase business, the owner of a restaurant is running a promotion in which a customer's bill can be randomly selected to receive a discount. When a customer's bill is printed, a program in the cash register randomly determines whether the customer will receive a discount on the bill. The program was written to generate a discount with a probability of 0.2, that is, giving 20 percent of the bills a discount in the long run. However, the owner is concerned that the program has a mistake that results in the program not generating the intended long-run proportion of 0.2.

The owner selected a random sample of bills and found that only 15 percent of them received discounts. A confidence interval for p, the proportion of bills that will receive a discount in the long run, is 0.15 ± 0.06. All conditions for inference were met.

(b) A second random sample of bills was taken that was four times the size of the original sample. In the second sample 15 percent of the bills received the discount.

Question 2

Determine the value of the margin of error based on the second sample of bills that would be used to compute an interval for p with the same confidence level as that of the original interval.

Question 2

2015 ▪ Q2

To increase business, the owner of a restaurant is running a promotion in which a customer's bill can be randomly selected to receive a discount. When a customer's bill is printed, a program in the cash register randomly determines whether the customer will receive a discount on the bill. The program was written to generate a discount with a probability of 0.2, that is, giving 20 percent of the bills a discount in the long run. However, the owner is concerned that the program has a mistake that results in the program not generating the intended long-run proportion of 0.2.

The owner selected a random sample of bills and found that only 15 percent of them received discounts. A confidence interval for p, the proportion of bills that will receive a discount in the long run, is 0.15 ± 0.06. All conditions for inference were met.

Question 2

(c) Based on the margin of error in part (b) that was obtained from the second sample, what do you conclude about whether the program is working as intended? Justify your answer.

Solution 2

(a)(i)

Mark scheme

- No. The confidence interval is $(0.09, 0.21)$, which includes the value 0.20. Because 0.20 is a plausible value for the true probability, the interval does NOT provide convincing statistical evidence that the program is not working as intended.

Solution 2

(a)(ii)

Mark scheme

- No. The confidence interval $(0.09, 0.21)$ contains values other than 0.20 (for example 0.10 or 0.15), so any value in the interval, not just 0.20, is a plausible value for the true probability. This does not provide convincing evidence that the probability is exactly 0.20.

Solution 2

(b)

Mark scheme

- The margin of error for a proportion is proportional to $1/\sqrt{n}$, so quadrupling the sample size divides the margin of error by $\sqrt{4} = 2$. New margin of error $= 0.06/2 = 0.03$.

Solution 2

(c)

Mark scheme

- Using the new margin of error 0.03, the confidence interval for p is $0.15 \pm 0.03 = (0.12, 0.18)$. This interval does not contain 0.20, so there is now convincing evidence that the program is not working as intended (not generating the discount with probability 0.20).