Statistics Unit 7 worksheet: One sample interval for a proportion and sample size (answer key)

ANSWERS AT A GLANCE

1. (a) p is the proportion of all the city's rental apartments that have a working smoke alarm. (b) Random sample; 400 is at most 6,400; 296 successes and 104 failures, both at least 10. (c) $\hat{p} = 0.74$, SE ≈ 0.0219, ME ≈ 0.0430, interval 0.697 to 0.783. (d) We are 95% confident that between 69.7% and 78.3% of all the city's rental apartments have a working smoke alarm.
2. (a) 601 (b) 464 (c) 664 (d) 1,015
3. (a) p is the proportion of all the bay's adult horseshoe crabs that are female; conditions hold (360 is far below a tenth of the population; 126 and 234 are at least 10); $\hat{p} = 0.35$, SE ≈ 0.0251, ME ≈ 0.0414, interval 0.309 to 0.391. We are 90% confident that between 30.9% and 39.1% of the bay's adult horseshoe crabs are female. (b) 0.40 lies above the interval, so it is not plausible at this level: the sample is evidence that the share is now below 0.40. One third, 0.333, lies inside, so it is plausible, but so is every value from 0.309 to 0.391; the interval does not show that the share is one third.
4. (a) First cruise 764; children 463; excursion 714. (b) 764: every question must reach the margin, and $p^*(1-p^*)$ is largest for the guess nearest 0.5, so that question needs the most. (c) 784, which is 20 more; it guarantees a margin of at most 0.035 on every question whatever the shares turn out to be.
5. (a) Parameter: 0.38 is $\hat{p}$, and the 450 farms are the sample; p is the share of all the state's small farms. Conditions: the failures were not checked. Compute: 1.645 is the multiplier for 90%, not 95%. Conclude: it is about the 450 farms and calls 95% a probability about this one interval. (b) p is the proportion of all about 12,000 small farms in the state that sell part of their crop directly; Random, 450 ≤ 1,200, 171 and 279 both at least 10; ME ≈ 0.0448, interval 0.335 to 0.425; we are 95% confident that between 33.5% and 42.5% of all the state's small farms sell part of their crop directly to consumers.

WORKED SOLUTIONS

1. Housing inspectors in a city of about 64,000 rental apartments visit a simple random sample of 400 apartments and find that 296 have a working smoke alarm (made-up data). Use $z^* = 1.96$ for 95% confidence.

Carry out a 95% confidence interval for the share of all the city's rental apartments with a working smoke alarm, in four steps: (a) Parameter, (b) Conditions, with the numbers, (c) Compute (standard error and margin of error to four decimal places, the ends to three decimal places, from unrounded values), (d) Conclude in context.

ANSWER

(a) p is the proportion of all the city's rental apartments that have a working smoke alarm. (b) Random sample; 400 is at most 6,400; 296 successes and 104 failures, both at least 10. (c) $\hat{p} = 0.74$, SE ≈ 0.0219, ME ≈ 0.0430, interval 0.697 to 0.783. (d) We are 95% confident that between 69.7% and 78.3% of all the city's rental apartments have a working smoke alarm.

SOLUTION

Parameter. Let p be the proportion of all the city's about 64,000 rental apartments that have a working smoke alarm. The interval is about all of them, not the 400 visited.

Conditions. Random: the apartments were a simple random sample. 10%: a tenth of 64,000 is 6,400, and 400 is below it, so the draws are close to independent. Large Counts: there are 296 successes and 400 − 296 = 104 failures, both at least 10, so the sampling distribution of $\hat{p}$ is close to normal. All three hold, so a one-sample z interval may be used.

Compute. $\hat{p} = \dfrac{296}{400} = 0.74$ and

$$\mathrm{SE}_{\hat{p}} = \sqrt{\frac{0.74 \times 0.26}{400}} \approx 0.0219, \qquad z^* \, \mathrm{SE}_{\hat{p}} \approx 1.96 \times 0.021932 \approx 0.0430$$

The interval is 0.74 ± 0.0430, from about 0.697 to 0.783.

Free sample from math.wiki for classroom use. Do not post. math.wiki

Conclude. We are 95% confident that between 69.7% and 78.3% of all the city's rental apartments have a working smoke alarm.

2. Find the smallest sample size that gives a margin of error no larger than m at the stated level, using the planning guess p^* given ("none" means no prior estimate). Use $z^* = 1.645$, 1.96 and 2.576 for 90%, 95% and 99% confidence. Show the bound to two decimal places before you round.

(a) $95\%,\ m = 0.04,$ none

(b) $90\%,\ m = 0.035,\ p^* = 0.30$

(c) $99\%,\ m = 0.05,$ none

(d) $95\%,\ m = 0.02,\ p^* = 0.12$

ANSWER

(a) 601 (b) 464 (c) 664 (d) 1,015

SOLUTION

Each uses $n \geq \left(\dfrac{z^*}{m}\right)^2 p^*(1-p^*)$, with $p^* = 0.5$ when there is no prior estimate, and rounds the bound up to a whole number.

(a) $\left(\dfrac{1.96}{0.04}\right)^2 (0.25) = 49^2(0.25) = 2{,}401 \times 0.25 = 600.25$, so $n = 601$.

(b) $\left(\dfrac{1.645}{0.035}\right)^2 (0.30)(0.70) = 47^2(0.21) = 2{,}209 \times 0.21 = 463.89$, so $n = 464$.

(c) $\left(\dfrac{2.576}{0.05}\right)^2 (0.25) = 51.52^2(0.25) = 2{,}654.3104 \times 0.25 \approx 663.58$, so $n = 664$.

(d) $\left(\dfrac{1.96}{0.02}\right)^2 (0.12)(0.88) = 98^2(0.1056) = 9{,}604 \times 0.1056 \approx 1{,}014.18$, so $n = 1{,}015$. A guess far from 0.5 cuts the sample to well under half of the 2,401 the safe guess would need at this level and margin.

3. Marine biologists capture a random sample of 360 adult horseshoe crabs from a large bay, whose adult population is in the hundreds of thousands, and find that 126 are female (made-up data). Use $z^* = 1.645$ for 90% confidence.

(a) Carry out a 90% confidence interval for the share of the bay's adult horseshoe crabs that are female: Parameter, Conditions with the numbers, Compute (four decimal places for SE and ME, three for the ends), Conclude in context.

(b) A survey from many years ago reported that 40% of the bay's adult crabs were female. A fishery council now says the share is about one third. For each value, say whether the interval makes it plausible, and what that does and does not show.

ANSWER

(a) p is the proportion of all the bay's adult horseshoe crabs that are female; conditions hold (360 is far below a tenth of the population; 126 and 234 are at least 10); $\hat{p} = 0.35$, SE ≈ 0.0251, ME ≈ 0.0414, interval 0.309 to 0.391. We are 90% confident that between 30.9% and 39.1% of the bay's adult horseshoe crabs are female. (b) 0.40 lies above the interval, so it is not plausible at this level: the sample is evidence that the share is now below 0.40. One third, 0.333, lies inside, so it is plausible, but so is every value from 0.309 to 0.391; the interval does not show that the share is one third.

SOLUTION

(a) *Parameter*. Let p be the proportion of all adult horseshoe crabs in the bay that are female.

Conditions. Random: the crabs were a random sample. 10%: the population is in the hundreds of thousands, so a tenth of it is well over 360. Large Counts: 126 females and 360 − 126 = 234 males, both at least 10. All three hold.

Free sample from math.wiki for classroom use. Do not post. math.wiki

Compute. $\hat{p} = \dfrac{126}{360} = 0.35$,

$$\mathrm{SE}_{\hat{p}} = \sqrt{\frac{0.35 \times 0.65}{360}} = \sqrt{0.00063194} \approx 0.0251, \qquad 1.645 \times 0.025139 \approx 0.0414$$

The interval is 0.35 ± 0.0414, from about 0.309 to 0.391.

Conclude. We are 90% confident that between 30.9% and 39.1% of all the bay's adult horseshoe crabs are female.

(b) The interval is the set of values of p the sample supports at 90% confidence. The old figure 0.40 lies beyond the upper end 0.391, so it is not plausible now; the sample is evidence, at this level, that fewer than 40% of adult crabs are female. (The old figure was itself an estimate with its own margin, so this does not measure a change precisely.) The council's one third, 0.333, lies inside the interval, so it is plausible. That does not confirm it: 0.31 and 0.38 are just as plausible.

4. A cruise line carried about 250,000 passengers last year. It plans one random sample of passengers that will answer three yes-or-no questions, and it wants a 95% interval with a margin of error of at most 0.035 for each question. Last year's smaller survey gave these estimates (made-up data): first cruise ever, 0.58; travelled with children, 0.18; booked a shore excursion, 0.35. Use $z^* = 1.96$.

(a) Find the sample size each question needs on its own, using last year's estimate as the planning guess.

(b) What sample size should the line use for the whole survey, and why is it the question whose guess is nearest 0.5 that decides?

(c) How many passengers would the safe guess $p^* = 0.5$ require, and what would the extra passengers buy?

ANSWER

(a) First cruise 764; children 463; excursion 714. (b) 764: every question must reach the margin, and $p^*(1-p^*)$ is largest for the guess nearest 0.5, so that question needs the most. (c) 784, which is 20 more; it guarantees a margin of at most 0.035 on every question whatever the shares turn out to be.

SOLUTION

(a) The factor in front is $\left(\dfrac{1.96}{0.035}\right)^2 = 56^2 = 3{,}136$.

First cruise: 3,136 × 0.58 × 0.42 = 3,136 × 0.2436 ≈ 763.93, so 764.

Children: 3,136 × 0.18 × 0.82 = 3,136 × 0.1476 ≈ 462.87, so 463.

Excursion: 3,136 × 0.35 × 0.65 = 3,136 × 0.2275 ≈ 713.44, so 714.

(b) One sample answers all three questions, so it must be large enough for the most demanding one, 764 passengers. The margin is $z^*\sqrt{p(1-p)/n}$, and $p(1-p)$ grows as p moves toward 0.5, so the guess 0.58 gives the largest factor and the largest n. The plan stays well under a tenth of 250,000.

(c) 3,136 × 0.25 = 784, a whole number, so 784 passengers, 20 more than 764. The safe guess makes $p^*(1-p^*)$ as large as it can be, so the margin is at most 0.035 on every question even if, say, the first-cruise share comes out at 0.50 rather than near 0.58.

5. A state has about 12,000 small farms. A simple random sample of 450 finds 171 that sell part of their crop directly to consumers (made-up data). A student hands in this 95% interval.

Parameter: $p = 0.38$, the proportion of the 450 farms that sell directly.

Conditions: Random, yes. 10%: 450 is less than 1,200. Large Counts: 171 ≥ 10.

Compute: SE $= \sqrt{0.38 \times 0.62/450} \approx 0.0229$, and with $z^* = 1.645$ the interval is 0.38 ± 0.0376, from 0.342 to 0.418.

Free sample from math.wiki for classroom use. Do not post. math.wiki

Conclude: There is a 95% probability that between 34.2% and 41.8% of the 450 farms sell directly.

(a) Find the mistake in each of the four steps and say why it is a mistake.

(b) Write each step correctly (use $z^* = 1.96$; four decimal places for the margin, three for the ends).

ANSWER

(a) Parameter: 0.38 is $\hat{p}$, and the 450 farms are the sample; p is the share of all the state's small farms. Conditions: the failures were not checked. Compute: 1.645 is the multiplier for 90%, not 95%. Conclude: it is about the 450 farms and calls 95% a probability about this one interval. (b) p is the proportion of all about 12,000 small farms in the state that sell part of their crop directly; Random, 450 ≤ 1,200, 171 and 279 both at least 10; ME ≈ 0.0448, interval 0.335 to 0.425; we are 95% confident that between 33.5% and 42.5% of all the state's small farms sell part of their crop directly to consumers.

SOLUTION

(a) *Parameter*. The parameter is a population share and is unknown; 0.38 is the sample share $\hat{p}$, and "the 450 farms" is the sample, whose share is already known exactly. *Conditions*. Large Counts needs at least 10 successes AND at least 10 failures; only the successes were checked. *Compute*. For 95%, $z^* = 1.96$; 1.645 leaves 0.05 in each tail and belongs to 90%, so the student's margin 1.645 × 0.022881 ≈ 0.0376 is too small and the interval 0.342 to 0.418 is too narrow for the level claimed. *Conclude*. Two errors: the sentence is about the sample instead of all the state's small farms, and "95% probability" treats the level as a chance for this one interval. The 95% describes the method: about 95% of samples give an interval that captures p.

(b) *Parameter*. Let p be the proportion of all the state's about 12,000 small farms that sell part of their crop directly to consumers.

Conditions. Random sample. A tenth of 12,000 is 1,200, and 450 ≤ 1,200. There are 171 successes and 450 − 171 = 279 failures, both at least 10.

Compute. $\hat{p} = \dfrac{171}{450} = 0.38$, SE $= \sqrt{0.38 \times 0.62/450} \approx 0.022881$, and ME = 1.96 × 0.022881 ≈ 0.0448. The interval is 0.38 ± 0.0448, from about 0.335 to 0.425.

Conclude. We are 95% confident that between 33.5% and 42.5% of all the state's small farms sell part of their crop directly to consumers.

Free sample from math.wiki for classroom use. Do not post. math.wiki