

Chapter 7 Problem 3

Fall-break problem 3. An unknown rate and a specified probability

A monitoring program calls a returned water sample *usable* if it meets a prespecified handling checklist. In a random sample of 240 returned kits, 168 samples are usable. Let p be the population proportion of usable samples among this program's returned kits. Assume independent Bernoulli outcomes with a common probability p , and a population much larger than the sample. These data are illustrative.

Use $z_{0.025} = 1.96$ and let Φ denote the standard normal CDF.

- Estimate p and check the observed success and failure counts. Construct an approximate 95% confidence interval for p , showing the estimated standard error and margin of error. Interpret the endpoints as percentages and identify the population they describe.
- Now consider a *separate hypothetical model* in which the population proportion is known to be $p = 0.75$. For a new independent sample of $n = 240$, find the approximate probability $\Pr(\hat{p} \leq 0.70)$. Check the expected counts, standardize, and use a direct normal approximation without continuity correction.
- Explain why the standard errors in parts (a) and (b) use different proportions. Which question estimates a fixed unknown parameter, and which assigns a probability to a random sample result?
- Think ahead using the sample-size planning section.* A new program wants a planned 95% normal-interval margin of error of at most 4 percentage points. No reliable advance estimate of its population proportion is available. Find the smallest integer sample size that meets this target under the worst-case choice of p . Justify that choice and state whether your answer is a required *total* or a number to add to 240.

Worked solution

(a) Estimate an unknown population proportion

$$\hat{p} = \frac{168}{240} = .70, \quad n\hat{p} = 168, \quad n(1 - \hat{p}) = 72.$$

Both observed counts exceed the course threshold of 5. With $z_{0.025} = 1.96$,

$$\widehat{\text{SE}} = \sqrt{\frac{.70(.30)}{240}} \approx .0295804,$$

$$\text{ME} = 1.96(.0295804) \approx .0579776,$$

$$CI_{95\%}(p) = [.70 - .0579776, .70 + .0579776] \approx [.64202, .75798].$$

The estimated population usable-sample rate among *this program's returned kits* is between about 64.20% and 75.80%, using an approximate 95% confidence procedure. Repeated intervals from this procedure would contain the fixed population p about 95% of the time under the assumptions. This does not describe all distributed kits or a probability distribution for the fixed parameter.

Worked Solution Continued

(b) Calculate under a separate specified model

Here $p = .75$ is stipulated, and the new sample is random. Expected counts are $240(.75) = 180$ and $240(.25) = 60$, both greater than 5. Thus

$$SE_{p=.75} = \sqrt{\frac{.75(.25)}{240}} \approx .02795085,$$

$$\Pr(\hat{p} \leq .70) \approx \Phi\left(\frac{.70 - .75}{\sqrt{.75(.25)/240}}\right) = \Phi(-1.788854) \approx .0368191 \approx 3.68\%.$$

Use the requested direct normal approximation with no continuity correction. Rounding to $z = -1.79$ in a normal table gives approximately .0367, consistent with the unrounded calculation.

(c) Why the two standard errors differ

In (a), p is unknown, so $\hat{p} = .70$ estimates it in the SE for an interval. In (b), the model supplies $p = .75$, so this determines the sampling variance. The first task estimates a fixed parameter; the second assigns a probability to a random sample result under a specified model.

(d) Plan for a new population

Since $p(1 - p) = \frac{1}{4} - (p - \frac{1}{2})^2 \leq \frac{1}{4}$, the worst case is $p = .5$. Require

$$1.96\sqrt{\frac{.25}{n}} \leq .04 \iff n \geq \frac{1.96^2(.25)}{.04^2} = 600.25.$$

The smallest integer is $n = 601$ **total observations in the new program**. It is not 601 more kits and not $601 - 240$. Check: $ME(600) \approx .0400083 > .04$, while $ME(601) \approx .0399750 < .04$. This controls the planned normal-form margin; exact interval coverage is still approximate.

Question reproduced verbatim from the course Chapter 7 chronological student packet, p. 17. The same question appears on p. 4 of the updated Chapter 8 student notes.