Chapter 7 Practice 1

Question from the chronological student notes, page 5.

Practice 1. A mean, a margin, and a width

A laboratory takes a random sample of $n = 64$ independent water specimens from a large population. The sample mean concentration is $\overline{x} = 12.4$ mg/L and the population standard deviation is known to be $\sigma = 3.2$ mg/L. Assume the large-sample normal approximation for the sample mean is adequate. These data are illustrative.

Use $z_{0.025} = 1.96$ for 95% confidence and $z_{0.005} = 2.576$ for 99% confidence.

(a) Define the population parameter being estimated. Construct its approximate 95% confidence interval. Identify the standard error, margin of error, and *full width*, including their units.

(b) Interpret the interval and explain what the 95% confidence level means over repeated samples. A classmate says, "About 95% of individual water specimens have concentrations in this interval." Is that what a confidence interval for a mean says? Explain.

(c) For planning, suppose a future sample from the same population has $n = 256$, with $\sigma = 3.2$ mg/L still known. At 95% confidence, find its margin of error and full width. Compare both with part (a). Can you determine the future interval's endpoints from this information alone?

(d) Return to the *original* sample of 64 specimens. Construct the approximate 99% confidence interval. Explain why increasing the confidence level changes the width in this direction. Does changing the confidence level change the observed center?

Worked solutions

A mean, a margin, and a width

(a) Estimate the population mean

The parameter μ is the population mean concentration of the water specimens, in mg/L; its point estimate is $\overline{x} = 12.4$ mg/L. Since $\sigma = 3.2$ mg/L is known, use the standard error:

$$\mathrm{SE}(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{64}} = 0.4 \text{ mg/L}.$$

$$ME = 1.96(0.4) = 0.784 \text{ mg/L}, \qquad \text{full width} = 2ME = 1.568 \text{ mg/L}.$$

$$\boxed{CI_{95\%}(\mu) = 12.4 \pm 0.784 = [11.616,\ 13.184] \text{ mg/L}.}$$

(b) Interpretation and repeated sampling

We are approximately 95% confident that the population mean concentration is between 11.616 and 13.184 mg/L. Across repeated random samples, about 95% of intervals constructed by this method would contain the fixed population mean μ. Once this sample is observed, the displayed interval either contains μ or does not; the frequentist statement does not assign probability 0.95 to this fixed parameter.

The classmate's claim is incorrect. This interval estimates a *population mean*; it is not a range containing 95% of individual concentrations. Individual variation is measured by the population SD, whereas the CI uses the standard error of an average.

(c) Four times as many observations

For a sample of 256 from the same population, the known σ gives

$$\mathrm{SE}_{\text{new}} = \frac{3.2}{16} = 0.2, \qquad ME_{\text{new}} = 1.96(0.2) = 0.392, \qquad W_{\text{new}} = 0.784 \quad \text{(all in mg/L)}.$$

Quadrupling n divides the SE, margin, and full width by $\sqrt{4} = 2$. The future sample mean is unknown, so its endpoints cannot be determined. Their form is $\overline{x}_{\text{new}} \pm 0.392$; the known population SD keeps this width fixed at the specified confidence level and sample size.

(d) More confidence from the same observations

For the original sample, the center remains 12.4 mg/L and the SE remains 0.4 mg/L. Only the critical value changes:

$$ME_{99\%} = 2.576(0.4) = 1.0304 \text{ mg/L}, \qquad \boxed{CI_{99\%}(\mu) = [11.3696,\ 13.4304] \text{ mg/L}.}$$

The full width is 2.0608 mg/L, versus 1.568 mg/L at 95%. The larger central normal area requires a larger critical value, so higher confidence requires a wider interval for the same data.

Chapter 7 Problem 2

Question from the chronological student notes, page 16.

Fall-break problem 2. Comparing means and allocating a sample

A laboratory measures protein concentration in cultures produced by two processes. It takes an independent random sample from each process, with a different culture for every observation. The illustrative results are:

	Process 1	Process 2
Sample size	$n_1 = 64$	$n_2 = 100$
Sample mean (mg/mL)	$\overline{x}_1 = 5.9$	$\overline{x}_2 = 5.5$
Sample SD (mg/mL)	$s_1 = 1.6$	$s_2 = 1.5$

Both population means and SDs are unknown. Assume observations are independent within and between samples, both populations are much larger than the samples, and the normal approximations for the sample means are adequate. Use $z_{0.025} = 1.96$.

(a) Define $\Delta = \mu_1 - \mu_2$ in context and give its point estimate. Starting from independence, write the variance of $\overline{X}_1 - \overline{X}_2$ and calculate its estimated standard error. Why do the variance contributions add although the means are subtracted?

(b) Construct an approximate 95% confidence interval for Δ. Show its margin of error and endpoints, with units.

(c) A researcher says, "Process 1 has the larger sample mean, so its population mean must be larger." Assess the claim using your interval. Does including zero establish equality? Explain what 95% confidence describes.

(d) The laboratory can measure 36 additional independent cultures, all from either Process 1 or Process 2. For planning, hold the SD estimates at 1.6 and 1.5. Calculate the anticipated standard error under each option and choose the option with the smaller anticipated margin of error. Explain why this does not determine the future interval's center.

Worked solutions

Comparing means and allocating a sample

(a) Parameter, point estimate, and standard error

$\Delta = \mu_1 - \mu_2$ is the difference in population mean protein concentration, Process 1 minus Process 2. Its estimate is $5.9 - 5.5 = 0.4$ mg/mL. Independence gives

$$\begin{aligned}\operatorname{Var}(\overline{X}_1 - \overline{X}_2) &= \operatorname{Var}(\overline{X}_1) + \operatorname{Var}(\overline{X}_2) - 2\operatorname{Cov}(\overline{X}_1, \overline{X}_2) \\ &= \frac{\sigma_1^2}{64} + \frac{\sigma_2^2}{100}.\end{aligned}$$

The covariance is zero, and the coefficient -1 is squared in a variance. Thus the contributions add even though the means are subtracted. Replacing the unknown population SDs by the sample SDs gives

$$\widehat{\mathrm{SE}}(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{1.6^2}{64} + \frac{1.5^2}{100}} = \sqrt{0.0400 + 0.0225} = 0.25 \text{ mg/mL}.$$

(b) Approximate 95% confidence interval

$$ME = 1.96(0.25) = 0.49 \text{ mg/mL}, \qquad \boxed{CI_{95\%}(\Delta) = 0.4 \pm 0.49 = [-0.09,\ 0.89] \text{ mg/mL}.}$$

(c) What including zero does and does not establish

The positive point estimate does not establish a positive population difference. The interval includes zero and small negative differences, so these data do not establish which population mean is larger at this confidence level. Including zero does not establish equality; equality remains compatible with this interval method, alongside other values.

We are approximately 95% confident that the population mean difference is between -0.09 and 0.89 mg/mL. Over repeated independent samples, about 95% of intervals constructed this way would contain the fixed Δ. The confidence level does not describe the proportion of individual cultures or assign a post-observation probability to Δ.

(d) Where the additional observations help more

Holding the SD estimates fixed, the anticipated standard errors are

$$\text{Add 36 to Process 1:} \quad \sqrt{\frac{2.56}{100} + \frac{2.25}{100}} \approx 0.2193 \text{ mg/mL},$$

$$\text{Add 36 to Process 2:} \quad \sqrt{\frac{2.56}{64} + \frac{2.25}{136}} \approx 0.2378 \text{ mg/mL}.$$

The corresponding anticipated 95% margins of error are 0.4299 and 0.4661 mg/mL. Choose **Process 1**. Adding observations there reduces the total variance contribution by more:

$$0.0400 - 0.0256 = 0.0144 \quad > \quad 0.0225 - \frac{2.25}{136} \approx 0.005956.$$

The new observations can change both sample means and SD estimates. These calculations plan precision; they neither fix the future center nor guarantee the realized width.