Chapter 8 Friday Practice

October 30, 2026 EN.553.211

Each original question is followed by its worked solution. The short checks also include their full prompts and answers.

Chapter 7 Review Question

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.

(a) 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.

(b) 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 $\Pr(\widehat{p} \leq 0.70)$. Check the expected counts, standardize, and use a direct normal approximation without continuity correction.

(c) 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?

(d) *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.

Chapter 7 Review Solution

(a) Estimate an unknown population proportion

The observed proportion is $\widehat{p} = 168/240 = 0.70$. The observed success and failure counts, 168 and 72, both exceed the course guideline of 5. The sampling assumptions are also needed; this count check alone does not repair dependence or selection bias.

$$\widehat{\mathrm{SE}}(\widehat{p}) = \sqrt{\frac{(0.70)(0.30)}{240}} \approx 0.0295804, \qquad ME = 1.96\widehat{\mathrm{SE}} \approx 0.0579776.$$

$$\boxed{CI_{95\%}(p) = 0.70 \pm 0.0579776 = [0.64202,\ 0.75798].}$$

We are approximately 95% confident that **64.20% to 75.80%** of this program's population of returned kits yield usable samples. The target is returned kits, not all kits distributed or all households. The margin is about 5.80 percentage points. About 95% of intervals from repeating this sampling procedure would contain the fixed p.

(b) A probability under a separate specified model

Here the model states $p = 0.75$. For a new sample, $X \sim \mathrm{Binomial}(240, 0.75)$ and $\widehat{p} = X/240$. The expected success and failure counts are 180 and 60, both greater than 5. Use the specified p in the sampling distribution:

$$\mathbb{E}(\widehat{p}) = 0.75, \qquad \mathrm{SE}(\widehat{p}) = \sqrt{\frac{(0.75)(0.25)}{240}} \approx 0.0279508.$$

With the requested direct normal approximation,

$$\Pr(\widehat{p} \le 0.70) \approx \Phi\left(\frac{0.70 - 0.75}{0.0279508}\right) = \Phi(-1.78885) \approx \boxed{0.0368}.$$

The underlying event is $X \le 168$. No continuity correction was used, as requested. The approximation describes a roughly 3.68% chance of this sample event under the stipulated model.

(c) Why the two standard errors differ

In part (a), p is fixed but unknown; $\widehat{p} = 0.70$ estimates it, so $\sqrt{\widehat{p}(1-\widehat{p})/n}$ estimates the SE for a CI. In part (b), $p = 0.75$ is given by the hypothetical model, so use $\sqrt{p(1-p)/n}$ to calculate a probability for a random sample proportion. One procedure estimates a parameter; the other predicts sample variation under a specified parameter value.

(d) Conservative sample-size planning

Four percentage points means $\varepsilon = 0.04$. Since $p(1-p) = \frac{1}{4} - (p - \frac{1}{2})^2 \le \frac{1}{4}$, the largest planned SE occurs at $p = 0.5$. Therefore require

$$1.96\sqrt{\frac{1/4}{n}} \le 0.04 \quad \Longleftrightarrow \quad n \ge \frac{1.96^2(1/4)}{0.04^2} = 600.25, \qquad \boxed{n = 601}.$$

Round *up*. This is a required **total** from the new program, not a number added to 240. It controls the planned normal-theory margin for every p; it does not guarantee exact 95% coverage where the normal approximation is poor, particularly near 0 or 1.

Problem 1 The GPA Example

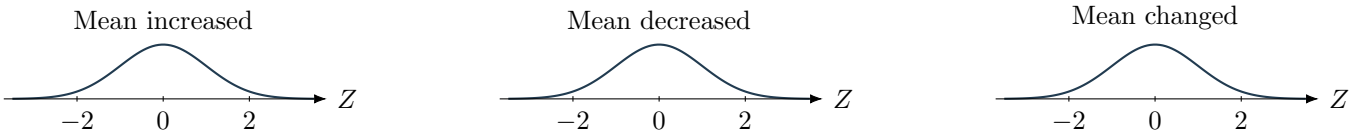
Use the example from the recorded Chapter 8 lecture. The historical reference mean GPA is 3.37. An independent random sample of 50 current students has $\overline{x} = 3.47$ and $s = .31$. Before collecting the sample, the research question was whether the population mean GPA had increased. The chosen significance level is $\alpha = .05$.

Use the lecture's large-sample normal approximation, assuming its sampling conditions are adequate. The population standard deviation is unknown. Write Φ for the standard normal CDF.

(a) **Hypotheses and parameters.** Define the population parameter μ. Explain the roles of the *null hypothesis* and *alternative hypothesis*, and state both for this question. Identify the *null value* μ_0. Is a hypothesis a statement about μ or about the observed $\overline{x}$?

(b) **Statistic and reference distribution.** Write the test statistic in terms of the random quantities $\overline{X}$ and S, then calculate its observed value z using $\overline{x}$ and s. Explain what the numerator and denominator measure. Define the *null distribution*; state the distribution used here and the assumption under which it applies. Which is random before sampling: the statistic or the null value?

(c) **The p-value and its tail.** Define a p-value in words, including both what is assumed and what counts as "at least as favorable." For each separately specified research question below, state H_a, write a probability involving Z and the observed z, shade its region, and calculate the p-value. The data stay the same; these are three alternative questions, each chosen *before* seeing data.



Mark the observed boundary or boundaries and shade the relevant region. Use a normal table or calculator.

(d) **Significance and conclusion.** Define the significance level as used so far in this lecture. For the original *increase* question, state the decision at .05 and the conclusion about the population. Reassess the same evidence using a prechosen .01 level. Which of the data, z, p-value, threshold, and decision change?

(e) **Interpretation check.** Explain what is wrong with each statement and replace it with a correct one: (i) "The p-value is the probability that our null hypothesis is wrong." (ii) "If we fail to reject, we have proved that the population mean is 3.37."

Problem 1 Worked Solution

(a) A claim about a population parameter

μ is the mean GPA of the current student population from which the random sample is drawn. The null hypothesis supplies the reference claim $H_0 : \mu = 3.37$. The research question supplies $H_a : \mu > 3.37$. The null value is the fixed number $\mu_0 = 3.37$. Hypotheses concern μ; $\overline{x} = 3.47$ is an observed estimate, not the parameter.

(b) Random statistic and observed value

$$Z = \frac{\overline{X} - 3.37}{S/\sqrt{50}} \overset{\text{approx}}{\sim} N(0,1) \text{ under } H_0, \qquad z = \frac{3.47 - 3.37}{.31/\sqrt{50}} \approx 2.28099.$$

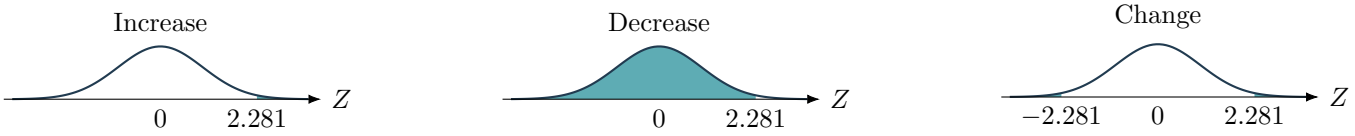
The numerator is the sample mean's discrepancy from the null value. The estimated sampling SE is $.31/\sqrt{50} \approx .0438406$ GPA units. The observed mean is about 2.281 SEs above the null value. Before sampling, $\overline{X}$, S, and Z vary; the stipulated null value does not. After sampling, $\overline{x}$, s, and z are fixed observations.

The *null distribution* is the reference distribution of the statistic when H_0 and the sampling assumptions hold. Here it is approximately standard normal under the lecture's large-sample method. Substituting an estimated SD does not make this an exact standard-normal result. The sample-size rule is a course convention, not a universal guarantee for arbitrary populations.

(c) Evidence depends on the alternative

A p-value is the probability, assuming H_0 and the model, of a statistic at least as favorable to the *specified* H_a as the observed statistic. With $z = 2.28099$:

Alternative	Tail probability	Approximate p-value
$\mu > 3.37$	$P(Z \geq z) = 1 - \Phi(z)$	.01127
$\mu < 3.37$	$P(Z \leq z) = \Phi(z)$	.98873
$\mu \neq 3.37$	$P(\lvert Z\rvert \geq \lvert z\rvert) = 2[1 - \Phi(\lvert z\rvert)]$	.02255



(d) Thresholds and conclusions

So far, α is the *prechosen threshold* for the p-value. Reject when p-value$\leq \alpha$; otherwise fail to reject. For the increase question, $.01127 < .05$: there is evidence of an increased population mean GPA. At .01, fail to reject: the evidence does not meet that stricter threshold. The data, observed z, and p-value stay the same. Only the threshold and decision change.

(e) Repair the interpretations

(i) Under the null model, about 1.13% of samples would yield a standardized increase at least this large. The p-value is not the probability that H_0 is wrong (or true). The null model is assumed when this tail probability is calculated.

(ii) Failure to reject means insufficient evidence for the specified alternative at that level; it does not prove equality.

Problem 2 The Chapter 7 Connection

Return to Chapter 7 Problem 3. In a random sample of 240 returned water-sample kits, 168 are usable. Let p be the population proportion of usable samples *among this program's returned kits*. Assume independent Bernoulli outcomes with a common p and a population much larger than the sample. These data are illustrative.

New question. Before sampling, the program asks whether the usable-sample proportion *differs from* .75. The decline comparison below is a second, separately preplanned question. Use $\alpha = .05$.

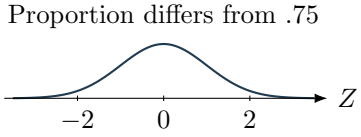
(a) **Retrieve the confidence interval.** Use your Chapter 7 work, or recompute the estimate $\widehat{p}$, estimated standard error, and approximate 95% confidence interval. Use $z_{.025} = 1.96$. Does the interval include .75? State what population quantity the interval estimates.

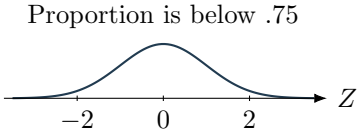
(b) **Build a null reference model.** State H_0 and H_a for the new question. Check the expected success and failure counts under H_0. The model from Chapter 7 Problem 3(b) supplies

$$Z = \frac{\widehat{p} - .75}{\sqrt{.75(.25)/240}} \overset{\text{approx}}{\sim} N(0,1) \quad \text{when } p = .75.$$

Use this supplied model to calculate the observed z. Explain why the interval in (a) uses an estimated standard error but this reference calculation uses .75. In each calculation, identify what is fixed and what varies across repeated samples.

(c) **Turn the probability into evidence.** For the *differs* question, mark the observed boundaries and shade the appropriate region on the first curve. Calculate the p-value using the direct normal approximation, without continuity correction. Give a decision and a contextual conclusion. Does this decision agree with the fact you observed about the 95% interval?





(d) **Compare a planned decline question.** A second analyst had specified $H_a : p < .75$ *before* seeing the same data. Shade that analyst's region on the second curve. Use your earlier calculation to obtain the p-value and decision at .05. Explain how the two analysts can reach different decisions without changing the data or the null distribution.

(e) **Reconcile the interval and the tests.** A classmate says, "A two-sided 95% confidence interval decides every 5% test of its null value." Use your results to assess the claim. Identify both the difference in tail questions and the difference between the two standard-error formulas. Should an analyst choose the direction of H_a after seeing which tail gives the smaller p-value? Explain.

Problem 2 Worked Solution

(a) Retrieve the original estimation calculation

$$\widehat{p} = \frac{168}{240} = .70, \qquad \widehat{\mathrm{SE}} = \sqrt{\frac{.70(.30)}{240}} \approx .0295804.$$

The observed success/failure counts are 168 and 72. The 95% normal interval is

$$.70 \pm 1.96(.0295804) = [.6420224, .7579776].$$

It includes .75. It estimates the population usable-sample proportion among this program's *returned kits*, about 64.20% to 75.80%. It is not an interval for the probability that the fixed parameter lies inside the realized endpoints.

(b) Condition on the null claim

For the new question, $H_0 : p = .75$ and $H_a : p \neq .75$. The expected counts under H_0 are $240(.75) = 180$ and $240(.25) = 60$, meeting the course's normal-approximation check. Use the supplied reference model:

$$\mathrm{SE}_0 = \sqrt{\frac{.75(.25)}{240}} \approx .02795085, \qquad z = \frac{.70 - .75}{.02795085} \approx -1.788854.$$

For estimation, the unknown fixed p is estimated by the random $\widehat{p}$, giving a plug-in SE. Across repeated samples both interval center and estimated SE vary. For the null reference calculation, $p = .75$ and its SE are fixed; the sample count and $\widehat{p}$ vary. The observed .70 is fixed after collection. The symbol p for the proportion is distinct from the phrase p-value.

(c) The two-sided question

$$p\text{-value} \approx P(|Z| \geq 1.788854) = 2\Phi(-1.788854) \approx .0736383.$$

Shade both tails beyond -1.789 and $+1.789$. Since $.07364 > .05$, fail to reject. There is insufficient evidence at 5% that the population usable proportion differs from .75. This agrees with the observed inclusion of .75 in the 95% interval for these data. It does not prove exact equivalence of the two procedures for every sample.

Problem 2 Solution Continued

(d) The planned lower-tail question

For $H_a : p < .75$, shade only $Z \leq -1.788854$. Original Problem 3(b)'s probability becomes the one-sided p-value:

$$p\text{-value} \approx \Phi(-1.788854) \approx .0368191 < .05.$$

Reject: there is evidence of a decline below .75 for this preplanned directional question. The null model, observed z, and data are identical; the tail event differs. The two-sided event includes equally large deviations in *both* directions and has twice the area here. Rounding z to -1.79 gives table values .0367 and .0734.

(e) Why the interval does not decide every test

A two-sided 95% interval addresses both directions; it cannot automatically determine a one-sided 5% test. Also, this Chapter 7 Wald interval uses $\widehat{p}$ in its SE, whereas this null test uses $p_0 = .75$. Thus even the two-sided comparison is not an exact inversion of *this* interval method. Agreement in (c) is a result for these data. A 95% proportion interval and a 5% test need not agree when their constructions differ.

The alternative represents the research question chosen before inspecting the sample. Choosing the smaller tail after seeing the result changes the procedure and does not justify the planned one-sided interpretation.

Problem 2(c)–(d): the tail regions



Short Checks and Solutions

1. Recovery time hypotheses

Question. A treatment is intended to *reduce* mean recovery time from 12 days. State H_0 and H_a. Would a sample mean above 12 favor this alternative?

Solution. The population parameter is the mean recovery time μ under the treatment. Use $H_0 : \mu = 12$ days and $H_a : \mu < 12$ days. A sample mean above 12 points away from this alternative and does not support a reduction. Choose the alternative before observing the sample.

2. The same discrepancy with twice the standard error

Question. Suppose the discrepancy points toward H_a. If $\overline{x} - \mu_0$ stays fixed but the standard error doubles, what happens to z? Explain why the same raw difference now provides weaker evidence.

Solution. Since $z = (\overline{x} - \mu_0)/SE$, doubling the denominator halves z when the numerator is fixed. Its magnitude moves toward zero. A discrepancy in the direction of H_a now represents less unusual sampling variation under the null, so its corresponding p-value increases. The raw difference is unchanged; its size relative to sampling uncertainty is smaller.

3. A p value above the significance level

Question. A test yields a p-value of 0.08 with $\alpha = 0.05$. State the decision and explain why it does not prove H_0.

Solution. Because $.08 > .05$, fail to reject H_0. The data do not provide enough evidence for the specified alternative at this significance level. This does not prove the null; a sample can fail to detect an effect that is present.

Short Checks Continued

4. The GPA rejection region

Question. Return to the GPA example in Problem 1. For $\alpha = .05$, use $z_{.05} \approx 1.645$ to specify the rejection region. Compare your observed statistic with it. Does this agree with your p-value decision?

Solution. For the prechosen increase question, $H_a : \mu > 3.37$ and $\alpha = .05$. The upper-tail convention gives

$$R_0 = [z_{.05}, \infty) \approx [1.645, \infty).$$

The observed $z \approx 2.281$ belongs to this region. Reject H_0, agreeing with the upper-tail p-value $.01127 < .05$. Both methods express the same decision rule under the same reference model.

5. The two sided critical value

Question. Why does a two-sided 5% test use 1.96 rather than 1.645?

Solution. A two-sided test divides the total null rejection probability .05 between two tails, .025 each. Thus $P(Z > 1.96) \approx .025$, and the rejection region is

$$(-\infty, -1.96] \cup [1.96, \infty).$$

Using 1.645 in both tails would allocate about .05 to each tail, for total size about .10, rather than .05.