

Homework 8 — Inference for Proportions and for a Variance

STAT 515 · Ott & Longnecker §10.2, §7.2 · Exam 2 review

(Due: Monday, Nov 2, 2026, before class)

Please submit a hard copy of your homework before class and email your R code, if applicable, to Alex Goebel at AJGOEBEL@email.sc.edu.

This assignment completes the material for **Exam 2 (Wednesday, Nov 4, in class)**. Show your work throughout.

Useful values: $z^* = 1.645$ (90%), 1.96 (95%). For $df = 19$: $\chi^2_{0.025} = 32.852$ and $\chi^2_{0.975} = 8.907$.

Problem 1: Confidence Interval for a Proportion (15 points)

In a survey of $n = 980$ randomly selected university students, 637 reported sleeping fewer than 7 hours on a typical weeknight.

- (a) Compute the sample proportion $\hat{p}$.
 - (b) Verify that the large-sample conditions are met, stating the rule you are using.
 - (c) Compute the standard error $\sqrt{\hat{p}(1 - \hat{p})/n}$ and construct a 95% confidence interval for the population proportion p .
 - (d) Interpret your interval in one sentence, in context.
 - (e) A campus health office claims that “more than two-thirds of students are sleep-deprived.” Does your interval support that claim? Explain.
 - (f) Without recomputing, state whether a 90% interval would be wider or narrower than your 95% interval, and why.
-

Problem 2: Wilson’s Adjusted (WAC) Interval (12 points)

In a small pilot survey, only $n = 20$ students were asked the same sleep question, and 3 reported sleeping fewer than 7 hours.

- (a) Compute $\hat{p}$ and check the large-sample conditions. Are they satisfied?
- (b) Compute the ordinary (Wald) 95% interval $\hat{p} \pm 1.96\sqrt{\hat{p}(1 - \hat{p})/n}$ anyway. What is wrong with the interval you obtain?

(c) Compute Wilson's adjusted estimate

$$\tilde{p} = \frac{x + 2}{n + 4}$$

and the Wilson adjusted 95% confidence interval

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}.$$

(d) Compare the two intervals. Explain in one or two sentences why the adjustment “adds two successes and two failures,” and what problem this fixes.

(e) For which of the two surveys — Problem 1 ($n = 980$) or this one ($n = 20$) — does the choice between the Wald and Wilson intervals matter more? Why?

Problem 3: Determining the Sample Size for a Proportion (10 points)

A polling organization wants to estimate the proportion of registered voters who support a ballot measure, with a 90% confidence interval and a margin of error of at most ± 0.04 .

(a) With no prior information about p , what value of p should be used in the sample size formula, and why does that choice give the largest required n ?

(b) Compute the required sample size using

$$n = \frac{(z^*)^2 p(1 - p)}{E^2}.$$

(c) Suppose an earlier poll suggests $p \approx 0.25$. Recompute the required sample size.

(d) Explain in one sentence why using prior information reduces the required sample size, and what risk the organization takes by relying on it.

Problem 4: Confidence Interval for a Variance (15 points)

A bottling plant records the fill volume (in mL) of $n = 20$ randomly selected bottles. The sample standard deviation is $s = 4.6$ mL. Fill volumes are assumed to be normally distributed.

(a) State the degrees of freedom and write down the formula for a $100(1 - \alpha)\%$ confidence interval for σ^2 :

$$\left(\frac{(n - 1)s^2}{\chi_{\alpha/2}^2}, \frac{(n - 1)s^2}{\chi_{1-\alpha/2}^2} \right).$$

(b) Compute $(n - 1)s^2$ and construct the 95% confidence interval for σ^2 .

(c) Convert your answer to a 95% confidence interval for the standard deviation σ .

- (d) Explain why this interval is **not** symmetric about s^2 , unlike a t interval for a mean.
- (e) The plant's specification requires $\sigma < 5$ mL. Based on your interval in (c), is the plant meeting the specification? Explain.
-

Problem 5: Normality, Robustness, and Choosing a Procedure (10 points)

- (a) Define **robustness** as the term is used in this course.
- (b) The t interval for a mean is robust to moderate departures from normality, but the χ^2 interval for a variance is **not**. Explain what this difference means for a data analyst, and why it makes the normality check more important in Problem 4 than in Problem 1 of Homework 6.
- (c) Suppose the normal Q-Q plot of the 20 fill volumes shows the points curving upward sharply at both ends, well away from the reference line. Name the feature of the distribution this indicates, and state whether you would still report the interval from Problem 4(b). Justify your answer.
- (d) Suppose instead the Q-Q plot shows points scattered randomly but closely about a straight line. What do you conclude, and which procedures from this assignment are then safe to use?
-

Problem 6: Proportions and Variances in R (8 points)

- (a) Explain what each line computes and report the output:

```
x <- 637; n <- 980
phat <- x / n
se <- sqrt(phat * (1 - phat) / n)
phat + c(-1, 1) * qnorm(0.975) * se
```

- (b) Write the R code that reproduces the Wilson adjusted interval in Problem 2(c).
- (c) Complete the following to reproduce the variance interval in Problem 4(b):

```
n <- 20; s <- 4.6
(n - 1) * s^2 / qchisq(c(_____, _____), df = n - 1)
```