

Chapter 21. Inference on Two Proportions

1. Reminder: We still need to do 6d (finding the sample size needed to estimate one proportion) from the Chapter 20 handout.
2. Comment about calculations and rounding.
3. Is there anything special to notice about using software to find confidence intervals and do hypothesis tests of two proportions?
4. What kinds of questions do we answer about two proportions? (And what conditions are necessary to use each of our techniques?)
 - a. Find a confidence interval for a difference of two proportions. (Separately: Large-sample and plus-four.)
 - b. Do a hypothesis test comparing two proportions.
5. How is this similar to two-sample mean problems?
6. How is this different from two-sample mean problems?
7. When we are setting up the problems, does it really matter which population proportion we put first and which we put second?
8. For one-proportion problems, why is there a difference in the formula for the SE for the confidence intervals and hypothesis tests for proportions?
9. For two-proportion problems, why is there another difference in the formula for the SE for the confidence intervals and hypothesis tests for proportions?

Ch. 22. Review

The main task for this chapter is to look at the statement of the problem itself and decide which technique is appropriate. We'll look over this chapter and the various review materials on the web for Test 4.

Discussion:

1. Reminder: We still need to do 6d (finding the sample size needed to estimate one proportion) from the Chapter 20 handout.
2. Comment about calculations and rounding.

The formulas in this chapter include several numbers very close to zero – some of which are in the denominator. The results of calculations like this are VERY dependent on how the rounding is done. If you calculate by hand or with a calculator, and write down rounded partial results, it is very likely that your final answer will look quite a bit different from the final answer in the back of the book, or from software. This is a good reason to focus on making sure you can write

the formula correctly, with the correct numbers, and then mostly find your answers for the z-scores using software. DO NOT spend much time practicing these calculations by hand. When you do calculations by hand, if you keep at least four digits after any leading zeroes in the numbers in the denominator, you have a better chance of getting an answer close to that given in the text.

3. Is there anything special to notice about using software to find confidence intervals and do hypothesis tests of two proportions?

Mostly not. We generalize from how we did one-proportion problems – to simply replace choosing the 1-sample procedure to choosing the 2-sample procedure.

And, just like the two-sample mean problems, we can put the data in in either order, so then we have to be careful to interpret our results correctly by paying attention to the order implicit in the question and noticing whether what we put into the software did the order we intended to use or the opposite order.

One change: Software gives us the opportunity to choose whether or not to pool the two samples for the SE. So we want to choose to “pool” when we do a hypothesis test and choose NOT to “pool” when we do a confidence interval. (See number 6 and 9 here for more discussion of why we do this.)

4. What kinds of questions do we answer about two proportions? (And what conditions are necessary to use each of our techniques?)
 - a. Find a confidence interval for a difference of two proportions. (Separately: Large-sample and plus-four.)

From the “Large-sample confidence interval” section (pages 517-520) write the sample size condition and read Example 21.2.

From the “Accurate confidence intervals” section (pages 520-522) write the sample size condition and skim over Example 21.3. For Exercise 21.5 at the end of this section

- Explain why the plus-four method is acceptable but the large-sample method would not be.
- Compute the estimates of both proportions, correct to at least 3 decimal places.
- Write the complete formula that needs to be evaluated, with all numbers plugged in, but don’t compute.
- Use software to find the interval.
- Write your conclusion in words.

- b. Do a hypothesis test comparing two proportions.

From the “Significance Tests for Comparing Proportions” section (pages 522-526) write the sample size condition and skim over Example 21.5. For exercise 21.7 at the end of this section

- Write the hypotheses and define the parameters.
- Compute the estimates of all three proportions, correct to at least 3 decimal places.
- Write the complete formula that needs to be evaluated, with all numbers plugged in, but don't compute.
- Write your hypotheses with the parameters in a different order, and then write the complete formula that needs to be evaluated, with all numbers plugged in, but don't compute.
- Use software to do the computations for each of these. Write the value for the z-score and the P-value for each beside your formulas. (See if the P-values agree with what you would find from the Normal table.)
- Choose either of these to use to write your conclusion in words.

5. How is this similar to two-sample mean problems?

We have to be careful to keep our chosen order of the parameters the same throughout the problem. We have to note whether we got the software to do the computation in the order we chose or not.

6. How is this different from two-sample mean problems?

In the denominator (SE) for the hypothesis test, instead of having two separate standard deviations to plug in separately, we combine all the data to give one combined estimator of the pop'n proportion when H_0 is true. Notice that the formula in our text and on our formula handout don't look the same, but are algebraically equivalent.

7. When we are setting up the problems, does it really matter which population proportion we put first and which we put second?

No. But we have to be careful all the way through to make sure we are staying clear about what order we plan to use and whether our software calculations are giving results in that order.

8. For one-proportion problems, why is there a difference in the formula for the SE for the confidence intervals and hypothesis tests for proportions?

In the hypothesis test problem, the denominator is the SE of the sampling dist'n of \hat{p} if H_0 is true, which means $p = p_o$, so we use p_o in the SE. In the confidence interval problem, we don't have that assumption, so we use the best estimator of p that we have, which is \hat{p} or \tilde{p} .

9. For two-proportion problems, why is there another difference in the formula for the SE for the confidence intervals and hypothesis tests for two proportions?

See the answer to number 6 in this list. This comes from the H_0 having the two population proportions equal, so all the data is estimating the same value of p , so that's why we "pool" the data.

Chapter 22. Look on the Course Calendar for the Test 4 announcement. Follow the link to "Summary information" which is where you expect the Test Review.

Hints about the statement of the problem:

Is the question about population mean(s) or population proportion(s)? Answer: Think carefully about the original data. Was it quantitative data or categorical data, where there are only two categories? If it was quantitative data, then the question is probably about mean(s). If it is categorical, with only two categories, then the question is probably about proportion(s).

Class exercise. Start with the problems in Ch. 22. For each, identify what type of method is needed.

Quiz 12: Due Wed. April 17.

For each four-step process, in the "Solve" part, calculate any sample proportions needed, give the formula in symbols and plug in numbers, but use software for calculations on these. Report the test statistic and the p-value.

21.32 (20 points)

21.34 (two questions, so two four-step processes. 40 points) .

The following problems are worth 5 points each: 22.41, 22.42, 22.43, 22.44, 22.45.

This problem is worth 15 points: 22.46