Chapter 20 Summary
Testing Hypotheses about Proportions

What have we learned?

- We can use what we see in a random sample to test a particular hypothesis about the world.
    - Hypothesis testing complements our use of confidence intervals.
- Testing a hypothesis involves proposing a model, and seeing whether the data we observe are consistent with that model or so unusual that we must reject it.
    - We do this by finding a P-value—the probability that data like ours could have occurred if the model is correct.
- We've learned the process of hypothesis testing, from developing the hypotheses to stating our conclusion in the context of the original question.
- We know that confidence intervals and hypothesis tests go hand in had in helping us think about models.
    - A hypothesis test makes a yes/no decision about the plausibility of a parameter value.
    - A confidence interval shows us the range of plausible values for the parameter.

Hypotheses

- In Statistics, a hypothesis proposes a model for the world. Then we look at the data.
- If the data are consistent with that model, we have no reason to disbelieve the hypothesis.
    - Data consistent with the model *lend support* to the hypothesis, but do not *prove* it.
- But if the facts are inconsistent with the model, we need to make a choice as to whether they are inconsistent enough to disbelieve the model.
    - If they are inconsistent enough, we can reject the model.
- Think about the logic of jury trials:
    - To prove someone is guilty, we start by *assuming* they are innocent.
    - We retain that hypothesis until the facts make it unlikely beyond a reasonable doubt.
    - Then, and only then, we reject the hypothesis of innocence and declare the person guilty.
- The same logic used in jury trials is used in statistical tests of hypotheses:
    - We begin by assuming that a hypothesis is true.
    - Next we consider whether the data are consistent with the hypothesis.
    - If they are, all we can do is retain the hypothesis we started with. If they are not, then like a jury, we ask whether they are unlikely beyond a reasonable doubt.
- The statistical twist is that we can quantify our level of doubt.
    - We can use the model proposed by our hypothesis to calculate the probability that the event we've witnessed could happen.
    - That's just the probability we're looking for—it quantifies exactly how surprised we are to see our results.
    - This probability is called a P-value.

Hypotheses (cont.)

- When the data are consistent with the model from the null hypothesis, the P-value is high and we are unable to reject the null hypothesis.
    - In that case, we have to "retain" the null hypothesis we started with.
    - We can't claim to have proved it; instead we "*fail to reject the null hypothesis*" when the data are consistent with the null hypothesis model and in line with what we would expect from natural sampling variability.
- If the P-value is low enough, we'll "*reject the null hypothesis*," since what we observed would be very unlikely were the null model true.

Testing Hypotheses

- The null hypothesis, which we denote H_0, specifies a population model parameter of interest and proposes a value for that parameter.
    - We might have, for example, H_0: $p = 0.20$, as in the chapter example.
- We want to compare our data to what we would expect given that H_0 is true.
    - We can do this by finding out how many standard deviations away from the proposed value we are.
- We then ask how likely it is to get results like we did if the null hypothesis were true.

A Trial as a Hypothesis Test

- Hypothesis testing is very much like a court trial.
- The null hypothesis is that the defendant is innocent.
- We then present the evidence—collect data.
- Then we judge the evidence—"Could these data plausibly have happened by chance if the null hypothesis were true?"
    - If they were very unlikely to have occurred, then the evidence raises more than a reasonable doubt in our minds about the null hypothesis.
- Ultimately, we must make a decision. How unlikely is unlikely?
- Some people advocate setting rigid standards—1 time out of 20 (0.05) or 1 time out of 100 (0.01).
- But if *you* have to make the decision, you must judge for yourself in any particular situation whether the probability is small enough to constitute "reasonable doubt."

What to Do with an "Innocent" Defendant

- If the evidence is not strong enough to reject the presumption of innocent, the jury returns with a verdict of "not guilty."
    - The jury does not say that the defendant is innocent.
    - All it says is that there is not enough evidence to convict, to reject innocence.
    - The defendant may, in fact, be innocent, but the jury has no way to be sure.
- Said statistically, we will *fail to reject* the null hypothesis.
    - We never declare the null hypothesis to be true, because we simply do not know whether it's true or not.
    - Sometimes in this case we say that the *null hypothesis has been retained*.
- In a trial, the burden of proof id on the prosecution.
- In a hypothesis test, the burden of proof is on the unusual claim.
- The null hypothesis is the ordinary state of affairs, so it's the alternative to the null hypothesis that we consider unusual (and for which we must marshal evidence).

The four basic parts of a hypothesis test

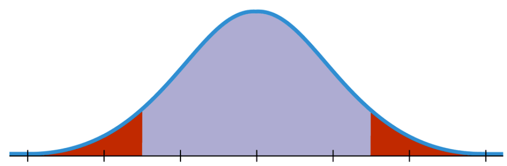
1. Hypotheses
   - The null hypothesis: To perform a hypothesis test, we must first translate our question of interest into a statement about model parameters.
     - In general, we have H_0: *parameter = hypothesized value*.
   - The alternative hypothesis: The alternative hypothesis, H_A, contains the values of the parameter we accept if we reject the null.
2. Model
   - To plan a statistical hypothesis test, specify the *model* you will use to test the null hypothesis and the parameter of interest.
   - All models require assumptions, so state the assumptions and check any corresponding conditions.
   - Your plan should end with a statement like
     - *Because the conditions are satisfied, I can model the sampling distribution of the proportion with a Normal model.*
     - Watch out, though. It might be the case that your model step ends with "*Because the conditions are not satisfied, I can't proceed with the text.*" If that's the case, stop and reconsider.
   - Each test we discuss in the book has a name that you should include in your report.
   - The test about proportions is called a one-proportion *z*-test.
     - One-Proportion *z*-Test
       - The conditions for the one-proportion *z*-test are the same as for the one proportion *z*-interval. We test the hypothesis H_0: $p = p_0$ using the statistic $z = \frac{(\hat{p} - p_0)}{SD(\hat{p})}$ where $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}}$
     - When the conditions are met and the null hypothesis is true, this statistic follows the standard Normal model, so we can use that model to obtain a P-value.
3. Mechanics
   - Under "mechanics" we place the actual calculation of our test statistic from the data.
   - Different tests will have different formulas and different test statistics.
   - Usually, the mechanics are handled by a statistics program or calculator, but it's good to know the formulas.
   - The ultimate goal of the calculation is to obtain a P-value.
     - The P-value is the probability that the observed statistic value (or an even more extreme value) could occur if the null model were correct.
     - If the P-value is small enough, we'll reject the null hypothesis.
     - Note: The P-value is a conditional probability—it's the probability that the observed results could have happened *if the null hypothesis is true*.
4. Conclusion
   - The conclusion in a hypothesis test is always a statement about the null hypothesis.
   - The conclusion must state either that we reject or that we fail to reject the null hypothesis.
   - And, as always, the conclusion should be stated in context.

The four basic parts of a hypothesis test (cont.)

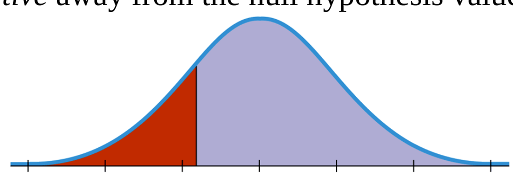
4. Conclusion (cont.)
    - Your conclusion about the null hypothesis should never be the end of a testing procedure.
    - Often there are actions to take or policies to change.

Alternative Alternatives

- There are three possible alternative hypotheses:
    - H_A: *parameter* < *hypothesized value*
    - H_A: *parameter* ≠ *hypothesized value*
    - H_A: *parameter* > *hypothesized value*
- H_A: *parameter* ≠ *value* is known as a two-sided alternative because we are equally interested in deviations on either side of the null hypothesis value.
- For two-sided alternatives, the P-value is the probability of deviating in *either* direction from the null hypothesis value.



- The other two alternative hypotheses are called one-sided alternatives.
- A one-sided alternative focuses on deviations from the null hypothesis value in only one direction.
- Thus, the P-value for one-sided alternatives is the probability of deviating *only in the direction of the alternative* away from the null hypothesis value.



P-Values and Decisions: What to Tell About a Hypothesis Test

- How small should the P-value be in order for you to reject the null hypothesis?
- It turns out that our decision criterion is context-dependent.
    - When we're screening for a disease and want to be sure we treat all those who are sick, we may be willing to reject the null hypothesis of no disease with a fairly large P-value.
    - A longstanding hypothesis, believed by many to be true, needs stronger evidence (and a correspondingly small P-value) to reject it.
- Another factor in choosing a P-value is the importance of the issue being tested.
- Your conclusion about any null hypothesis should be accompanied by the P-value of the test.
    - If possible, it should also include a confidence interval for the parameter of interest.
- Don't just declare the null hypothesis rejected or not rejected.
    - Report the P-value to show the strength of the evidence against the hypothesis.
    - This will let each reader decide whether or not to reject the null hypothesis.

*A Better Confidence Interval for Proportions

- Confidence intervals and hypothesis tests for proportions use different standard deviation formulas because of the special property that links a proportion's value with its standard deviation.
- We can improve on the confidence interval for a proportion with a simple adjustment.
  - The adjustment brings the standard deviations for the test and confidence interval into closer agreement.
- The improved method is a little strange:
  - It takes the original counts and adds four *phony* observations, two to the successes and two to the failures.
- We use the adjusted proportion $\tilde{p} = \frac{y+2}{n+4}$
- For convenience, we write $\tilde{n} = n+4$
- We modify the interval by using these adjusted values for both the center of the interval *and* the margin of error.
- The adjusted interval is $\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$
- This adjusted form gives better performance overall and works much better for proportions near 0 or 1.
- It has the additional advantage that we no longer need to check the Success/Failure Condition.

What Can Go Wrong?

- Hypothesis tests are so widely used—and so widely misused—that the issues involved are addressed in their own chapter (Chapter 21).
- There are a few issues that we can talk about already, though:
- Don't base your null hypothesis on what you see in the data.
  - *Think* about the situation you are investigating and develop your null hypothesis appropriately.
- Don't base your alternative hypothesis on the data, either.
  - Again, you need to *Think* about the situation.
- Don't make your null hypothesis what you want to show to be true.
  - You can reject the null hypothesis, but you can never "accept" or "prove" the null.
- Don't forget to check the conditions.
  - We need randomization, independence, and a sample that is large enough to justify the use of the Normal model.