Chapter 8: Statistical Inference: Significance Tests About Hypotheses

Section 8.1: What Are the Steps for Performing a Significance Test?

Learning Objectives

1. 5 Steps of a Significance Test
2. Assumptions
3. Hypotheses
4. Calculate the test statistic
5. P-Value
6. Conclusion and Statistic Significance

Learning Objective 1: Significance Test

- **A *significance test* is a method of using data to summarize the evidence about a hypothesis**
- **A *significance test* about a hypothesis has *five steps***
  1. **Assumptions**
  2. **Hypotheses**
  3. **Test Statistic**
  4. **P-value**
  5. **Conclusion**

Learning Objective 2: Step 1: Assumptions

- **A (significance) test assumes that the data production used randomization**
- **Other assumptions may include:**
  - **Assumptions about the sample size**
  - **Assumptions about the shape of the population distribution**

Learning Objective 3: Step 2: Hypothesis

- **A *hypothesis* is a statement about a population, usually of the form that a certain parameter takes a particular numerical value or falls in a certain range of values**
- **The main goal in many research studies is to check whether the data support certain hypotheses**

Learning Objective 3: Step 2: Hypotheses

- **Each significance test has two hypotheses:**
  - **The *null hypothesis* is a statement that the parameter takes a particular value. It has a single parameter value.**
  - **The *alternative hypothesis* states that the parameter falls in some alternative range of values.**

Learning Objective 3:
Null and Alternative Hypotheses

- The value in the null hypothesis usually represents ***no effect***
  - The symbol H_o denotes null hypothesis
- The value in the alternative hypothesis usually represents ***an effect of some type***
  - The symbol H_a denotes alternative hypothesis
  - The alternative hypothesis should express what the researcher hopes to show.
- **The hypotheses should be formulated before viewing or analyzing the data!**

Learning Objective 4:
Step 3: Test Statistic

- **A *test statistic* describes how far the point estimate falls from the parameter value given in the null hypothesis (usually in terms of the number of standard errors between the two).**
- **If the test statistic falls far from the value suggested by the null hypothesis in the direction specified by the alternative hypothesis, it is good evidence against the null hypothesis and in favor of the alternative hypothesis.**
- **We use the test statistic to assesses the evidence against the null hypothesis by giving a probability , the P-Value.**

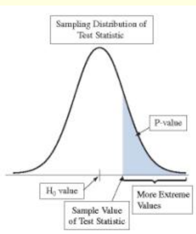
Learning Objective 5:
Step 4: P-value

- **To interpret a test statistic value, we use a probability summary of the evidence *against* the null hypothesis, H_o**
  - **First, we presume that H_o is true**
  - **Next, we consider the sampling distribution from which the test statistic comes**
  - **We summarize how far out in the tail of this sampling distribution the test statistic falls**

Learning Objective 5:
Step 4: P-value

- **We summarize how far out in the tail the test statistic falls by the tail probability of that value and values even more extreme**
  - **This probability is called a *P-value***
  - **The smaller the P-value, the stronger the evidence is against H_o**

Learning Objective 5:
Step 4: P-value



▲ **FIGURE 8.1: Suppose H_0 Were True. The P-value Is the Probability of a Test Statistic Value Like the Observed One or Even More Extreme.** This is the shaded area in the tail of the sampling distribution. **Question:** Which gives stronger evidence against the null hypothesis, a P-value of 0.20 or of 0.01? Why?

Learning Objective 5:
Step 4: P-value

- **The *P-value* is the probability that the test statistic equals the observed value or a value even more extreme**
- **It is calculated by presuming that the null hypothesis H is true**

The smaller the *P*-value, the stronger the evidence the data provide against the null hypothesis. That is, a small *P*-value indicates a small likelihood of observing the sampled results if the null hypothesis were true.

Learning Objective 6:
Step 5: Conclusion

- **The conclusion of a significance test reports the P-value and *interprets* what it says about the question that motivated the test**

Chapter 8: Statistical Inference: Significance Tests About Hypotheses

Section 8.2: Significance Tests About Proportions

Learning Objectives:

1. Steps of a Significance Test about a Population Proportion
2. Example: One-Sided Hypothesis Test
3. How Do We Interpret the P-value?
4. Two-Sided Hypothesis Test for a Population Proportion
5. Summary of P-values for Different Alternative Hypotheses
6. Significance Level
7. One-Sided vs Two-Sided Tests
8. The Binomial Test for Small Samples

Learning Objective 1:
Example: Are Astrologers' Predictions Better Than Guessing?

An astrologer prepares horoscopes for 116 adult volunteers. Each subject also filled out a California Personality Index (CPI) survey. For a given adult, his or her horoscope is shown to the astrologer along with their CPI survey as well as the CPI surveys for two other randomly selected adults. The astrologer is asked which survey is the correct one for that adult

- With random guessing, $p = 1/3$
- The astrologers' claim: $p > 1/3$
- The hypotheses for this test:
  - H_o: $p = 1/3$
  - H_a: $p > 1/3$

Learning Objective 1:
Steps of a Significance Test about a Population Proportion

Step 1: Assumptions

- **The variable is categorical**
- **The data are obtained using randomization**
- **The sample size is sufficiently large that the sampling distribution of the sample proportion is approximately normal:**
  - **$np \geq 15$ and $n(1-p) \geq 15$**

Learning Objective 1:
Steps of a Significance Test about a Population Proportion

Step 2: Hypotheses

- **The null hypothesis has the form**:
  - **H_0: $p = p_0$**
- **The alternative hypothesis has the form**:
  - **H_a: $p > p_0$ (one-sided test) or**
  - **H_a: $p < p_0$ (one-sided test) or**
  - **H_a: $p \neq p_0$ (two-sided test)**

Learning Objective 1:
Steps of a Significance Test about a Population Proportion

Step 3: Test Statistic

- The test statistic measures how far the sample proportion falls from the null hypothesis value, p_0, relative to what we'd expect if H_0 were true
- The test statistic is:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Learning Objective 1:
Steps of a Significance Test about a Population Proportion

Step 4: P-value

- The P-value summarizes the evidence
- It describes how unusual the observed data would be if H_0 were true

Learning Objective 1:
Steps of a Significance Test about a Population Proportion

Step 5: Conclusion

- We summarize the test by reporting and interpreting the P-value

Learning Objective 2:
Example 1

Step 1: Assumptions

- The data is categorical – each prediction falls in the category "correct" or "incorrect"
- Each subject was identified by a random number. Subjects were randomly selected for each experiment.
- np=116(1/3) > 15
- $n(1-p)$ = 116(2/3) > 15

Learning Objective 2:
Example 1

Step 2: Hypotheses

- H_0: p = 1/3
- H_a: p > 1/3

Learning Objective 2:
Example 1

Step 3: Test Statistic:

In the actual experiment, the astrologers were correct with 40 of their 116 predictions (a success rate of 0.345)

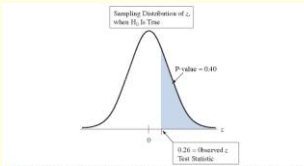
$$se = \sqrt{p_0(1-p_0)/n} = \sqrt{(1/3)(2/3)/116} = 0.0438$$

$$z = \frac{\hat{p} - p_0}{se} = \frac{0.345 - 1/3}{0.0438} = 0.26$$

Learning Objective 2: Example 1

<u>Step 4: P-value</u>

- **The P-value is 0.40**



▲ **FIGURE 8.2: Calculation of P-value, when $z = 0.26$, for Testing H_0: $p = 1/3$ against H_a: $p > 1/3$.** Presuming that H_0 is true, the P-value is the right-tail probability of a test statistic value even more extreme than observed. **Question:** Logically, why are the right-tail z-scores considered to be the more extreme values for testing H_0 against H_a: $p > 1/3$?

Learning Objective 2: Example 1

<u>Step 5: Conclusion</u>

- **The P-value of 0.40 is not especially small**
- **It does not provide strong evidence against H_0: $p = 1/3$**
- **There is not strong evidence that astrologers have special predictive powers**

Learning Objective 3: How Do We Interpret the P-value?

- **A significance test analyzes the strength of the evidence *against* the null hypothesis**
- **We start by presuming that H_0 is true**
- **The *burden of proof* is on H_a**

Learning Objective 3: How Do We Interpret the P-value?

- **The approach used in hypothesis testing is called a *proof by contradiction***
- **To convince ourselves that H_a is true, we must show that data contradict H_0**
- ***If the P-value is small, the data contradict H_0 and support H_a***

Learning Objective 4: Two-Sided Significance Tests

- **A two-sided alternative hypothesis has the form H_a: $p \neq p_0$**
- **The P-value is the *two-tail* probability under the standard normal curve**
- **We calculate this by finding the tail probability in a single tail and then doubling it**

Learning Objective 4: Example 2

- **Study: investigate whether dogs can be trained to distinguish a patient with bladder cancer by smelling compounds released in the patient's urine**

Learning Objective 4:
Example 2

- **Experiment:**
  - **Each of 6 dogs was tested with 9 trials**
  - **In each trial, one urine sample from a bladder cancer patient was randomly place among 6 control urine samples**

31

Learning Objective 4:
Example 2

- **Results:**

In a total of 54 trials with the six dogs, the dogs made the correct selection 22 times (a success rate of 0.407)

32

Learning Objective 4:
Example 2

- **Does this study provide strong evidence that the dogs' predictions were better or worse than with random guessing?**

33

Learning Objective 4:
Example 2

Step 1: Check the sample size requirement:

- **Is the sample size sufficiently large to use the hypothesis test for a population proportion?**
  - **Is np_0>15 and $n(1-p_0)$>15?**
  - **54(1/7) = 7.7 and 54(6/7) = 46.3**
- **The first, np_0 is not large enough**
  - **We will see that the two-sided test is robust when this assumption is not satisfied**

34

Learning Objective 4:
Example 2

Step 2: Hypotheses

- **H_0: $p = 1/7$**
- **H_a: $p \neq 1/7$**

35

Learning Objective 4:
Example 2

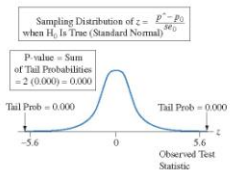
Step 3: Test Statistic

$$z = \frac{(0.407 - 1/7)}{\sqrt{\frac{(1/7)(6/7)}{54}}} = 5.55$$

36

Learning Objective 4: Example 2

Step 4: P-value



▲ **FIGURE 8.4: Calculation of P-value, when $z = 5.6$, for Testing H_0: p = 1/7 against H_a: $p \neq 1/7$.** Presuming H_0 is true, the P-value is the two-tail probability of a test statistic value even more extreme than observed. **Question:** Is the P-value of 0.000 strong evidence supporting H_0, or strong evidence against H_0?

Learning Objective 4: Example 2

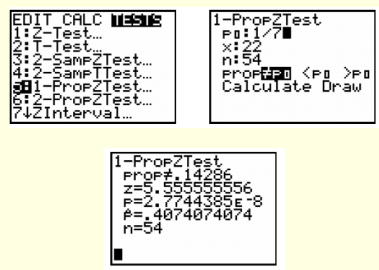
Step 5: Conclusion

- **Since the P-value is very small and the sample proportion is greater than 1/7, the evidence strongly suggests that the dogs' selections are *better* than random guessing**

Learning Objective 4: Example 2

- **Insight:**
  - **In this study, the subjects were a *convenience sample* rather than a random sample from some population**
  - **Also, the dogs were not randomly selected**
  - **Any inferential predictions are highly tentative. They are valid only to the extent that the patients and the dogs are representative of their populations**
  - **The predictions become more conclusive if similar results occur in other studies**

Learning Objective 4: Example 2



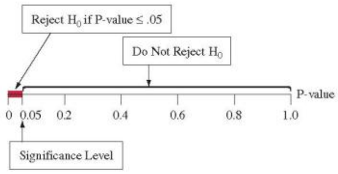
Learning Objective 5: Summary of P-values for Different Alternative Hypotheses

Alternative Hypothesis	P-value
H_a: $p > p_0$	Right-tail probability
H_a: $p < p_0$	Left-tail probability
H_a: $p \neq p_0$	Two-tail probability

Learning Objective 6: The Significance Level Tells Us How Strong the Evidence Must Be

- **Sometimes we need to make a decision about whether the data provide sufficient evidence to reject H_0**
- **Before seeing the data, we decide how small the P-value would need to be to reject H_0**
- **This cutoff point is called the *significance level***

Learning Objective 6: The Significance Level Tells Us How Strong the Evidence Must Be



▲ **FIGURE 8.6: The Decision in a Significance Test.** Reject H_0 if the P-value is less than or equal to a chosen **significance level,** usually 0.05.

43

Learning Objective 6: Significance Level

- **The significance level is a number such that we reject H_0 if the P-value is less than or equal to that number**
- **In practice, the most common significance level is 0.05**
- **When we reject H_0 we say the results are *statistically significant***

44

Learning Objective 6: Possible Decisions in a Hypothesis Test

P-value:	Decision about H_0:
≤ α	Reject H_0
> α	Fail to reject H_0

45

Learning Objective 6: Report the P-value

- **Learning the actual P-value is more informative than learning only whether the test is "statistically significant at the 0.05 level"**
- **The P-values of 0.01 and 0.049 are both statistically significant in this sense, but the first P-value provides much stronger evidence against H_0 than the second**

46

Learning Objective 6: "Do Not Reject H_0" Is Not the Same as Saying "Accept H_0"

- **Analogy: Legal trial**
  - **Null Hypothesis: Defendant is Innocent**
  - **Alternative Hypothesis: Defendant is Guilty**
  - **If the jury acquits the defendant, this does not mean that it accepts the defendant's claim of innocence**
  - **Innocence is plausible, because guilt has not been established *beyond a reasonable doubt***

47

Learning Objective 7: One-Sided vs Two-Sided Tests

- **Things to consider in deciding on the alternative hypothesis:**
  - **The context of the real problem**
  - **In most research articles, significance tests use two-sided P-values**
  - **Confidence intervals are two-sided**

48

Learning Objective 8:
The Binomial Test for Small Samples

- **The test about a proportion assumes normal sampling distributions for $\hat{p}$ and the z-test statistic.**
  - **It is a large-sample test because the CLT requires that the expected numbers of successes and failures be at least 15. In practice, the large-sample z test still performs quite well in two-sided alternatives even for small samples.**
  - **Warning: For one-sided tests, when p_0 differs from 0.50, the large-sample test does not work well for small samples**

Learning Objective 9:
Class Exercise 1

- In a survey by Media General and the Associated Press, 813 of the 1084 respondents indicated support for a ban on household aerosols. At the 1% significance level, test the claim that more than 70% of the population supports the ban.

Learning Objective 9:
Class Exercise 2

- In a Roper Organization poll of 2000 adults, 1280 have money in regular savings accounts. Use this sample data to test the claim that less than 65% of all adults have money in regular savings accounts. Use a 5% level of significance.

Learning Objective 9:
Class Exercise 3

- According to a Harris Poll, 71% of Americans believe that the overall cost of lawsuits is too high. If a random sample of 500 people results in 74% who hold that belief, test the claim that the actual percentage is 71%. Use a 10% significance level.