Chapter 9: Statistical Inference: Significance Tests About Hypotheses

Steps for Performing a Significance Test

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

Step 1 : Assumptions

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

Step 2 : Hypotheses

- 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

- 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 symbol H_0 denotes null hypothesis. This always has equality "=" sign.

    Ex: H_0: $p = 0.72$ H_0: $\mu = 42.3$

  - The ***alternative hypothesis*** states that the parameter falls in some alternative range of values. The symbol H_a denotes alternative hypothesis. The alternative hypothesis should express what the researcher hopes to show. This always has one of ">", "<", or "≠" signs.

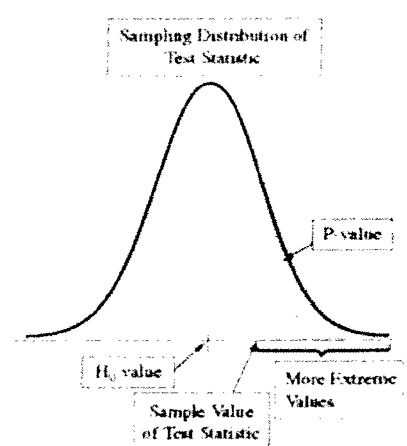
    Ex: H_a: $p < 0.47$ H_a: $\mu \neq 42$ H_a: $\mu > 3.45$

❖ The hypotheses should be formulated before viewing or analyzing the data!

Step 3: Test Statistic

- A ***test statistic*** describes how far the point estimate falls from the parameter value given in the null hypothesis
- We use the test statistic to assess the evidence against the null hypothesis by giving a probability, the P-Value.

Step 4: P-value



- To interpret a test statistic value, we use a probability summary of the evidence *against* the null hypothesis, H_0
  - First, we presume that H_0 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

- 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_0
  - The *P-value* is the probability that the test statistic equals the observed value or a value even more extreme

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

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

For example, we might decide that, we will reject H0 if the p-value ≤ 0.05 (significance level)

Significance Level

- The significance level is a nu
- mber 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*

Possible Decisions in a Hypothesis Test

P-value:	Decision about H_0:
$\leq \alpha$	Reject H_0
$> \alpha$	Fail to reject H_0

Steps of a Significance Tests about a Population Proportions

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: i.e. at least 15 successes and 15 failures $\{np \geq 15$ and $n(1-p) \geq 15\}$

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)

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:

$$z = \frac{\hat{p} - p_0}{se} \quad , where \quad se = \sqrt{\frac{p_0(1-p_0)}{n}} \qquad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Step 4: P-value

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

Step 5: Conclusion

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

Example:

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

- With random guessing, $p = 1/3$
- The astrologers' claim: $p > 1/3$

<u>Step 1: Assumptions</u>

- 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.
- $np=116(1/3) > 15$
- $n(1-p) = 116(2/3) > 15$

<u>Step 2: Hypotheses</u>

H_0: $p = 1/3$

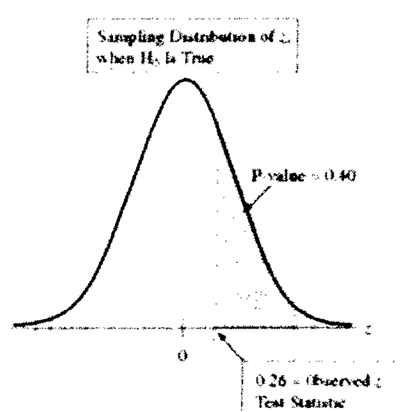
H_a: $p > 1/3$

<u>Step 3: Test Statistic:</u>

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

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

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



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

The P-value is 0.40

<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

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
- 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

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

"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*

Steps of a Significance Test about a Population Mean

- *Step 1: Assumptions*
  - The variable is quantitative
  - The data are obtained using randomization
  - The population distribution is approximately normal. This is most crucial when *n* is small and H_a is one-sided.

- *Step 2: Hypotheses:*
  - The null hypothesis has the form:
    - $H_0: \mu = \mu_0$
  - The alternative hypothesis has the form:
    - $H_a: \mu > \mu_0$ (one-sided test) or
    - $H_a: \mu < \mu_0$ (one-sided test) or
    - $H_a: \mu \neq \mu_0$ (two-sided test)

- *Step 3: Test Statistic*
  - The test statistic measures how far the sample mean falls from the null hypothesis value μ_0, as measured by the number of standard errors between them
  - The test statistic is:

$$t = \frac{\bar{x} - \mu_0}{se} \quad , where \quad se = s / \sqrt{n} \qquad\qquad t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

- *Step 4: P-value*
  - The P-value summarizes the evidence
  - It describes how unusual the data would be if H_0 were true
- *Step 5: Conclusion*
  - We summarize the test by reporting and interpreting the P-value

Decisions and Types of Errors in Significance Tests

Type I and Type II Errors

- When H_0 is true, a *Type I Error* occurs when H_0 is rejected
- When H_0 is false, a *Type II Error* occurs when H_0 is not rejected
- As P(Type I Error) goes *Down*, P(Type II Error) goes *Up*
  - The two probabilities are inversely related

Significance Test Results

TABLE 9.6: The Four Possible Results of a Decision in a Significance Test

Type I and Type II errors are the two possible incorrect decisions. We make a correct decision if we do not reject H_0 when it is true or if we reject it when it is false.

Reality About H_0	Decision: Do not reject H_0	Decision: Reject H_0
H_0 true	Correct decision	Type I error
H_0 false	Type II error	Correct decision

Type I error occurs if we reject H_0 when it is actually true.

Decision Errors: Type I

- If we reject H_0 when in fact H_0 is true, this is a **Type I error**.
- If we *decide* there is a significant relationship in the population (reject the null hypothesis):
  - This is an incorrect decision only if H_0 is true.
  - The probability of this incorrect decision is equal to a.
- If we reject the null hypothesis when it *is true* and $a = 0.05$:
  - There really is no relationship and the extremity of the test statistic is due to chance.
  - About 5% of all samples from this population will lead us to incorrectly reject the null hypothesis and conclude significance.

P(Type I Error) = Significance Level α

- Suppose H_0 is true. The probability of rejecting H_0, thereby committing a Type I error, equals the significance level, α, for the test.
- We can control the probability of a Type I error by our choice of the significance level
- The more serious the consequences of a Type I error, the smaller α should be

Decision Errors: Type II

If we fail to reject H_0 when in fact H_a is true, this is a **Type II error**.

If we *decide* not to reject the null hypothesis and thus allow for the plausibility of the null hypothesis

We make an incorrect decision only if H_a is true. The probability of this incorrect decision is denoted by β .

Determine the null and alternative hypotheses.

1) The principal of a middle school claims that the seventh grade test scores at her school vary less than the test scores of seventh-graders at neighboring schools, which have variation described by $\sigma = 14.7$.

A) H_0: $\sigma = 14.7$
H_a: $\sigma > 14.7$

B) H_0: $\sigma = 14.7$
H_a: $\sigma \leq 14.7$

C) H_0: $\sigma = 14.7$
H_a: $\sigma \geq 14.7$

D) H_0: $\sigma \neq 14.7$
H_a: $\sigma \neq 14.7$

E) H_0: $\sigma = 14.7$
H_a: $\sigma < 14.7$

2) An automobile manufacturer claims that its new sedan will average better than 25 miles per gallon in the city. Let μ represent the true average mileage of the new sedan.

A) H_0: $\mu = 25$
H_a: $\mu \leq 25$

B) H_0: $\mu = 25$
H_a: $\mu > 25$

C) H_0: $\mu = 25$
H_a: $\mu \neq 25$

D) H_0: $\mu = 25$
H_a: $\mu < 25$

E) H_0: $\mu = 25$
H_a: $\mu \geq 25$

3) A cereal company claims that the mean weight of it's individual serving boxes is at least 14 oz.

A) H_0: $\mu = 14$
H_a: $\mu < 14$

B) H_0: $\mu = 14$
H_a: $\mu > 14$

C) H_0: $\mu = 14$
H_a: $\mu \geq 14$

D) H_0: $\mu = 14$
H_a: $\mu \neq 14$

E) H_0: $\mu = 14$
H_a: $\mu \leq 14$

Select the most appropriate answer.

4) Which P-value provides the strongest evidence against the null hypothesis?

A) 0.001
B) 1
C) -0.05
D) 0.99
E) 0.05

5) Which of the following would be an appropriate null hypothesis?

A) The population proportion is less than 0.41.
B) The population proportion is not equal to 0.41.
C) The sample proportion is less than 0.41.
D) The population proportion is equal to 0.41.
E) The sample proportion is equal to 0.41.

6) Which of the following would be an appropriate alternative hypothesis?

A) The population mean is equal to 3.4.
B) The sample mean is equal to 3.4.
C) The sample mean is not equal to 3.4.
D) The sample mean is greater than 3.4.
E) The population mean is greater than 3.4.

Find the P-value for the indicated hypothesis test.

7) A medical school claims that more than 28% of its students plan to go into general practice. It is found that among a random sample of 130 of the school's students, 39% of them plan to go into general practice. Find the P-Value for testing the school's claim.

A) 0.0280
B) 0.3078
C) 0.3461
D) 0.1635
E) 0.0026

For the given sample data and null hypothesis, compute the value of the test statistic, z

8) A claim is made that the proportion of children who play sports is less than 0.5, and the sample statistics include 1933 subjects with 30% saying that they play a sport.

A) 17.59
B) -17.59
C) -35.90
D) 0.50
E) 35.90

9) The claim is that the proportion of drowning deaths of children occurring at beaches is more than 0.25. The sample statistics include n = 615 drowning deaths of children with 30% of them occurring at beaches.

A) -2.86
B) 0.25
C) 2.71
D) 2.86
E) -2.71

10) A drug company claims that over 80% of all physicians recommend their drug. 1200 physicians were asked if they recommend the drug to their patients. 30% said yes.

H_0: p = 0.8

A) -38.97
B) -86.60
C) 43.30
D) -43.30
E) -56.29

Perform a significance test for a population proportion using the critical value approach.

11) A poll of 1000 registered voters reveals that 48% of the voters surveyed prefer the Democratic candidate for the presidency. At the 0.05 level of significance, do the data provide sufficient evidence to conclude that the percentage of voters who prefer the Democratic candidate is less than 50%?

Assume that a simple random sample has been selected from a normally distributed population. Find the test statistic t.

12) Test the claim that for the population of female college students, the mean weight is given by $\mu = 132$ lb. Sample data are summarized as $n = 20$, $\bar{x} = 137$ lb, and $s = 14.2$ lb. Use a significance level of $\alpha = 0.1$.
Find the test statistic t.

A) -1.57
B) 7.04
C) 1.729
D) 1.57
E) 14.2

State conclusion to significance test in terms of the null hypothesis

13) In a random sample of 88 adults from a particular town, it is found that 6 of them have been exposed to a certain flu strain. At the 0.01 significance level, test the claim that the proportion of all adults in the town that have been exposed to this flu strain differs from the nationwide percentage of 8%.
H_0: p = 0.08 H_a: p ≠ 0.08.
$\alpha = 0.01$
Test statistic: z = -0.41. P-Value = 0.6828
State your conclusion in terms of H_0.

A) Do not reject H_a since the P-value is larger than α.
B) Reject H_a since the P-value is larger than α.
C) Since the P-value is larger than α, we conclude that the proportion of adults in this particular town that have been exposed to this flu strain differs from the nationwide percentage of 8%.
D) Reject H_0 since the P-value is larger than α.
E) Do not reject H_0 since the P-value is larger than α.

Assume that a simple random sample has been selected from a normally distributed population. Find the test statistic t.

14) Test the claim that the mean lifetime of a particular car engine is greater than 220,000 miles. Sample data are summarized as n = 23, $\bar{x}$ = 226,450 miles, and s = 11,500 miles. Use a significance level of $\alpha = 0.01$.
Find the test statistic t.

A) 2.69
B) -2.69
C) -2.24
D) 12.9
E) 2.24

15) Test the claim that the mean age of the prison population in one city is less than 26 years. Sample data are summarized as $n = 25$, $\overline{x} = 24.4$ years, and $s = 9.2$ years. Use a significance level of $\alpha = 0.05$.
Find the test statistic t.
A) 1.87
B) -4.35
C) 0.87
D) -1.87
E) -0.87

Classify the significance test as two-tailed, left-tailed, or right-tailed.

16) A manufacturer claims that the mean amount of juice in its 16 ounce bottles is 16.1 ounces. A consumer advocacy group wants to perform a significance test to determine whether the mean amount is actually less than this.
A) Left-tailed
B) Right-tailed
C) None of these
D) Middle-tailed
E) Two-tailed

17) In 1990, the average duration of long-distance telephone calls originating in one town was 15.3 minutes. A long-distance telephone company wants to perform a significance test to determine whether the average duration of long-distance phone calls has changed from the 1990 mean of 15.3 minutes.
A) Two-tailed
B) None of these
C) Middle-tailed
D) Left-tailed
E) Right-tailed

Assume that a simple random sample has been selected from a normally distributed population. State the final conclusion.

18) Test the claim that the mean lifetime of a particular car engine is greater than 220,000 miles. Sample data are summarized as $n = 23$, $\overline{x} = 226{,}450$ miles, and $s = 11{,}500$ miles. Use a significance level of $\alpha = 0.01$.
H_0: $\mu = 220{,}000$ H_a: $\mu > 220{,}000$
State your conclusion about H_0.
A) $t = 12.9$, reject H_0
B) $t = -2.69$, reject H_0do not reject H_0
C) $z = -2.69$, reject H_0
D) $t = 2.69$, reject H_0
E) $t = -2.69$, reject H_0

19) Test the claim that for the population of female college students at a particular university, the mean weight is given by $\mu = 132$ lb. Sample data are summarized as $n = 20$, $\overline{x} = 137$ lb, and $s = 14.2$ lb. Use a significance level of $\alpha = 0.1$.
$H_0 : \mu = 132$ $H_a : \mu \neq 132$
State your conclusion about H_0.
A) $t = 7.04$, reject H_0
B) $z = 1.57$, do not reject H_0
C) $t = 1.57$, do not reject H_0
D) $t = -1.57$, do not reject H_0
E) $t = 1.57$, reject H_0

Classify the conclusion of the significance test as a Type I error, a Type II error, or No error.

20) In the past, the mean lifetime for a certain type of flashlight battery has been 9.6 hours. The manufacturer has introduced a change in the production method and wants to perform a significance test to determine whether the mean lifetime has increased as a result. The hypotheses are:
H_0: $\mu = 9.6$ hours
H_a: $\mu > 9.6$ hours
Suppose that the results of the sample lead to nonrejection of the null hypothesis. Classify that conclusion as a Type I error, a Type II error, or a correct decision, if in fact the mean running time has increased.
A) No error
B) Type II error
C) Type I error