

Chapter 9 Significance Testing: Using Data to Test Hypotheses

In Chapter 8, we learned confidence intervals for an unknown parameter. Chapter 9 deals with significance test. These are different methods for coping with the uncertainty about the true value of a parameter caused by the sampling variation in estimates.

- **Confidence interval:** A fixed level of confidence is chosen. We determine a range of possible values for the parameter that are consistent with the data (at the chosen confidence level).
- **Significance test:** Only one possible value for the parameter, called the hypothesized value, is tested. We determine the strength of the evidence provided by the data against the proposition that the hypothesized value is the true value.

Note: This is a combination of Sections 9.1–9.5.

Hypothesis Testing:

Techniques for comparing two competing theories or hypotheses about the population.

Key Components:

- Population(s) of interest
- Research hypothesis
- Two competing hypotheses

H_0 : null hypothesis

H_1 : alternative hypothesis

- Data:
- Test Statistic:
- Decision Rule:

Setting up the Hypotheses

The null and alternative hypotheses are statements about a population parameter (not about the results in the sample). There will be a "**direction of extreme**" that is indicated by the alternative hypothesis. To see these ideas, let's try writing out some hypotheses to be put to the test.

Example 1. Mean area of apartments in a new development is advertised to be 1250 square feet. A tenant group thinks the apartments are smaller than advertised. They hire an engineer to measure a sample of apartments to test their suspicion.

$H_0 :$

$H_1 :$

Example 2. A sociologist asks a large sample of HS students which subject they like best. She suspects that a higher percent of males than of females will name math as favorite.

$H_0 :$

$H_1 :$

Example 3. Last year, service technicians took an average of 2.6 hours to respond to trouble calls. Manager wishes to assess if this year's data show a different average response time.

$H_0 :$

$H_1 :$

Let's Do a Test:

To learn how to do a test about the population mean (in the case when we assume the data are a random sample from a normal population with a known population standard deviation), we will just dive right in and use our knowledge from Section 7.2 to work it through.

Example: SSHA is a psychological test that measures motivation, attitude toward school, study habits. Scores: 0 to 200. The model for SSHA scores for US college students is normal with a mean of 115 and a standard deviation of 30.

A teacher thinks that older students have **better** attitudes (a higher score implies a better attitude). She will give the SSHA test to a random sample of 20 older students (at least 30 years old).

What is the population under study?

State the hypotheses to be tested regarding the mean of the population under study:

$H_0 :$

$H_1 :$

What will we base our decision on? We will look at the sample mean \bar{X} computed for our random sample of test scores.

Draw a picture of the sampling distribution for \bar{X} if the null hypothesis H_0 were true.

What kind of values of \bar{x} will lead us to reject H_0 ?

Values that are too small? too large? both directions?

The 20 test scores have been obtained and suppose the sample mean was 127.2.

Note that 127.2 is larger than 115, but is it large enough? Let's compute its standard score.

The standard score is called the **z-test statistic** and is given by:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} =$$

How likely is it to get a z-test statistic of 1.82 or even larger if the H_0 is true?

So if H_0 were true, we would expect to get a test statistic of 1.82 or larger only -----% of the time (in repeated samples). This probability you computed is called the **p-value**. It measures the compatibility of the data with the null hypothesis.

P-value = the probability of getting a test statistic as extreme or more extreme than the actual observed test statistic, assuming H_0 is true. (note that the direction of extreme matches H_1 .)

Our text (p. 381) says:

The P-value is the probability that, if the hypothesis was true, sampling variation would produce an estimate that is further away from the hypothesized value than the estimate we got from our data.

The P-value measures the strength of the evidence against H_0 .

The smaller the p-value, the stronger the evidence against H_0 .

In our example, p-value = ----- is somewhat small. How small is small enough?

We compare the p-value to the fixed amount of evidence required to reject H_0 . This fixed level, decided in advance, is our **significance level**, represented by α .

1. **Suppose the significance level $\alpha = 5\%$.**

What is the decision? Accept H_0 (i.e. fail to Reject H_0) or Reject H_0

What is the conclusion?

2. **Suppose the significance level $\alpha = 1\%$.**

What is the decision? Accept H_0 (i.e. fail to Reject H_0) or Reject H_0

What is the conclusion?

Decision Rule: Reject H_0 if _____

and conclude that the data are _____

Approach to hypothesis testing:

1. State the null and alternative hypotheses.
2. Calculate the test statistic.
3. Find the p-value (assuming _____ is true).
4. State the decision (by comparing to a significance level) and conclusion in terms of the scenario.

Summary: We have just conducted our first One-Sample Z-test for the Population Mean μ . It is a One-Sided to the Right (Upper level) Z-test.

In the above example, we assume we know the population s.d. σ , so the statistic $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ has a standard normal $N(0, 1)$ distribution.

If we do NOT know σ , we would estimate σ by sample s.d. S and use statistic $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$. What is the distribution of T under the null hypothesis H_0 ?

Note that $N(0, 1)$ is a special T distribution with d.f.= ∞ .

Next we will summarize the T-test for population mean μ in general and do examples of One-Sided to the Left, and Two-sided testing problems.

One-sample T test for $H_0 : \mu = \mu_0$

Assume:

Test Statistic: _____ has a _____ distribution under H_0

The observed value (computed from data) is

If $H_1 : \mu > \mu_0$ then the p-value is:

If $H_1 : \mu < \mu_0$ then the p-value is:

If $H_1 : \mu \neq \mu_0$ then the p-value is:

The test requires we have a random sample from a normal population. (If the sample size is large, the assumption of normality is not so crucial and the result is approximate.)

Example – Using the t table to find a p-value

$H_0 : \mu = 0$ versus $H_1 : \mu > 0$ with $n = 15$ observations and the observed t statistic is $t = 1.97$

- (a) What are the degrees of freedom?
- (b) Give the 2 critical values t^* from the t table which bound $t = 1.97$.
- (c) What are the right-tail probabilities p for these two entries?
- (d) Between what two values does the p-value fall?
- (e) Is the value of $t = 1.97$ significant at the 5% level? At the 1% level?
- (f) If have the software, find the exact p-value.

Example – Using the t table to find a p-value

$H_0 : \mu = 64$ versus $H_1 : \mu \neq 64$ with $n = 30$ observations and the observed t statistic is $t = 1.12$

- (a) What are the degrees of freedom?
- (b) Give the 2 critical values t^* from the t table which bound $t = 1.12$.
What are the right-tail probabilities p for these two entries?

(c) How would you report the p-value for this test?

(d) Is the value of $t = 1.12$ significant at the 10% level? At the 5% level?

(e) If have the software, find the exact p-value.

Example: Researcher wants to test if HS students complete a maze more quickly while listening to classical music. For the general HS population, the time to complete the maze follows a normal distribution with a mean of 40 seconds. A random sample of 100 HS students are timed while listening to classical music. The mean time was 39.1 seconds and the standard deviation was 4.12. Conduct a test.

(a) State the appropriate hypotheses:

(b) Give the observed test statistic:

(c) Give the degrees of freedom:

(d) Draw the picture of the p-value and find the (bounds for the) p-value.

(e) Are the results statistically significant at the 5% level?

(f) Are the results statistically significant at the 1% level?

The p-value shows us the amount of evidence against H_0 .

Example: Daily intake of calcium (mg) for 38 women in bone health study. We are concerned that the average calcium intake the population of such women is not meeting the RDA of 1200 mg, that is, the mean is less than the 1200 mg level. Here is the summary from Stata.

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
calcium	38	926.0263	427.2295	374	2433

(a) State the appropriate hypotheses:

(b) Give the observed test statistic:

(c) Give the degrees of freedom:

(d) Draw the picture of the p-value and find the (bounds for the) p-value.

(e) What is your conclusion at a 5% significance level?

Example: 12 Radon Detectors exposed to room with 105 pCi/l of radon (over 3 days). The 12 detectors in the sample gave a mean of 104.13 and standard deviation of 9.32.

(a) Based on these data, is there evidence at the 5% level that the mean reading for these detectors differ from the true value of 105?

(b) Give a 95% confidence interval for the mean reading μ for this type of detector.

(c) Did the 95% CI contain the hypothesized value of 105?

So the value of 105 was a plausible value for μ at the 95% confidence level, and thus we accepted the hypothesis that $\mu = 105$ at the 5% significance level.

Connection between Confidence Intervals and two-sided tests.

Reject $H_0 : \mu = \mu_0$ in favor of $H_1 : \mu \neq \mu_0$ at the significance level α if the $1 - \alpha$ confidence interval DOES NOT contain the hypothesized value μ_0 .

Example:

Cockroach Biology study of sugar levels absorbed.

95% CI for mean amount sugar (mg) in hindguts of cockroaches: 4.2 ± 2.3 .

Test $H_0 : \mu = 7$ vs. $H_1 : \mu \neq 7$ at $\alpha = 0.05$. Reject H_0 Accept H_0 Can't tell

Test $H_0 : \mu = 6$ vs. $H_1 : \mu \neq 6$ at $\alpha = 0.05$. Reject H_0 Accept H_0 Can't tell

Test $H_0 : \mu = 6$ vs. $H_1 : \mu \neq 6$ at $\alpha = 0.01$. Reject H_0 Accept H_0 Can't tell

Test $H_0 : \mu = 6$ vs. $H_1 : \mu \neq 6$ at $\alpha = 0.10$. Reject H_0 Accept H_0 Can't tell

Testing Hypotheses about a Population Proportion p

With a large sample size, we will do a Z-test – but be careful on the form of the standard deviation in the denominator.

Large-Sample Z Test for the population proportion p

To test the hypothesis $H_0 : p = p_0$ we compute the Z statistic

which has approximately the $N(0, 1)$ distribution when H_0 is true.

The p-value is computed as follows:

If $H_1 : p > p_0$ then the p-value is:

If $H_1 : p < p_0$ then the p-value is:

If $H_1 : p \neq p_0$ then the p-value is:

Example:

Of the 500 household respondents, 44% had no children at home. The US Census reports that 48% of households have no children. Do our results contradict census results? Test the hypotheses to assess if the population proportion has changed from the Census value of 0.48 using a 5% significance level.

Example:

Joshua, a starting basketball player, made only 36.2% of his free throws last season. During the summer he worked on developing a softer shot in the hope of improving his free-throw accuracy. In the first eight games of this season Joshua made 22 free throws in 42 attempts.

(a) Did his work in the summer result in a higher probability of success? Answer this by doing a test at the 5% significant level.

(b) Give a 90% CI for Joshua's free-throw success probability for the new season.

(c) What assumptions are needed for the validity of the test and CI calculations that you performed?

Summary and Comments on Significance Tests**Approach to hypothesis testing:**

1. State the null and alternative hypotheses.
2. Calculate the test statistic.
3. Find the p-value (assuming _____ is true).
4. State the decision (by comparing to a significance level) and conclusion in terms of the scenario.

P-Values

- The P-value measures the strength of the evidence against H_0 .
- The smaller the P-value, the stronger the evidence against H_0 .
- A large P-value provides no evidence against the null hypothesis.
- A large P-value does not imply that the null hypothesis is true.
- Suggestions for verbal translation of P-values are given in Table 9.3.2 (page 379).

Significance

- If the P-value $\leq \alpha$, the effect is said to be significant at the α -level.
- If you always test at the 5% level, you will reject one true null hypothesis in 20 **over the long run**.
- A two-sided test of $H_0 : \mu = \mu_0$ is significant at the 5% level if and only if μ_0 lies outside a 95% confidence interval for μ .
- In reports on research, the word “significant” used alone often means “significant at the 5% level” (i.e. P-value < 0.05). “Nonsignificant”, “does not differ significantly” and even “is no different” often mean P-value > 0.05 .
- No real distinction between P-value of 0.049 and 0.051.
- A nonsignificant result does **not** imply that H_0 is true.

What statistical significance doesn't mean

A statistically significant result implies that we rejected H_0 . That is, there was good evidence of an “effect” but the effect can be small, not of any practical importance — with a large enough sample size even a small effect can be significant.

Idea:

Likewise, there could be an effect – one that is of practical importance, but with a too small sample size, we may failing to reject H_0 , we may not have a statistically significant result.

Statistical Significance \neq Practical Significance

Two Types of Error

We may make two types of error when testing a hypothesis.

- A type I error is
- A type II error is

The significance level = the probability of making a Type I error.

The power of a test = 1 – probability of making a Type II error.