The second common type of inference is **tests of significance**. It has the following goal: to assess the evidence provided by data about some claim concerning a population.

**Hypothesis Testing**

The 4 steps common to all tests of significance:

1. Identify the parameter, state the null hypothesis \( H_0 \) and the alternative hypothesis \( H_a \), and choose the type of test that fits your situation.
2. Calculate the value of the test statistic.
3. Find the P-value.
4. State your conclusion about the data in a sentence, using the P-value and/or comparing the P-value to a significance level for your evidence.

**Hypothesis Testing with respect to a population mean \( \mu \)**

In Chapter 17 we still use “simple conditions” as in the previous chapter.

**STEP 1:** State the null hypothesis \( H_0 \) and the alternative hypothesis \( H_a \) with respect to a population mean \( \mu \).

To do a significance test, you need 2 hypotheses:

- **Null Hypothesis** \( H_0 \): the statement being tested, usually phrased as “no effect” or “no difference”.
- **Alternative Hypothesis** \( H_a \): the statement we hope or suspect is true instead of \( H_0 \).

Hypotheses always refer to some population, not to a particular outcome / sample. We state hypothesis in terms of population parameters.

Alternative hypothesis can be one-sided or two-sided. In practice you should decide in advance whether a particular direction of effect is important or not before examining the data.

- **One-sided hypothesis:** covers just part of the range for your parameter. Alternative hypothesis is one-sided if it states that a parameter is larger than or smaller than the Null hypothesis value
  
  \[
  H_0 : \mu = 10 \quad \text{vs.} \quad H_a : \mu > 10
  \]
  
  or
  
  \[
  H_0 : \mu = 10 \quad \text{vs.} \quad H_a : \mu < 10
  \]

- **Two-sided hypothesis:** covers the whole possible range for your parameter. Alternative hypothesis is two-sided if a parameter could be either smaller or larger.
  
  \[
  H_0 : \mu = 10 \quad \text{vs.} \quad H_a : \mu \neq 10
  \]
Even though $H_a$ is what we hope or believe to be true, our test gives evidence for or against $H_0$ only. We never prove $H_0$ true, we can only state whether we have enough evidence to reject $H_0$ (which is evidence in favor of $H_a$) or that we don’t have enough evidence to reject $H_0$.

**STEP 2:** Calculate the value of the test statistic with respect to a population mean $\mu$.

A test statistic (calculated from the sample data) measures compatibility between the $H_0$ and the data.

The formula for the test statistic will vary between different types of problems. In problems like those we study in Chapter 16, the test statistic will be the $z$-score:

- compute the test statistic $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
- the P-value for a test of $H_0$ against:

$$H_a: \mu > \mu_0 \quad \text{is} \quad P(Z \geq z)$$

$$H_a: \mu < \mu_0 \quad \text{is} \quad P(Z \leq z)$$

$$H_a: \mu \neq \mu_0 \quad \text{is} \quad 2P(Z \geq |z|)$$

**STEP 3:** Find the P-value.

P-value: the probability, computed assuming that $H_0$ is true, that the test statistic would take a value as extreme or more extreme than that actually observed due to random fluctuation. It is a measure of how unusual your sample results are.

- Calculate the P-value by using the sampling distribution of the test statistic and appropriate distribution for testing.
- The smaller the P-value, the stronger the evidence against $H_0$ provided by the data.

STEP 4: Compare your P-value to a significance level. State your conclusion about the data in a sentence.

- Compare P-value to a significance level $\alpha$.

  If $P-value \leq 0.05$ then there is no more than 1 in 20 chance (because $0.05 = \frac{5}{100} = \frac{1}{20}$) of being wrong in rejecting $H_0$ - what is very small chance.

The most common values for significance level $\alpha$ are 0.05 and 0.01. In practice you should decide in advance how strong the evidence must be for you to reject the Null hypothesis, so significance level $\alpha$ should be specified prior to collecting data and any analysis.

- If the $P-value \leq \alpha$, we can reject $H_0$.

- If you can reject $H_0$, your results are significant.

- If you do not reject $H_0$, your results are not significant.

P-values can be more informative than a reject/do not reject $H_0$ based on $\alpha$. As P-value gets smaller the evidence for rejecting $H_0$ gets stronger.

Just because we use $\alpha = 0.05$ a lot doesn’t mean that’s the level you have to use - it’s just the most common.

If we fail to reject $H_0$, it may be because $H_0$ is true or because our sample size is insufficient to detect the alternative.

Problem 1.

Each of the following situations requires a significance test about a population mean \( \mu \). State the appropriate null hypothesis \( H_0 \) and alternative hypothesis \( H_a \) in each case:

a. Census Bureau data shows that the mean household income in the area served by a shopping mall is $72,500 per year. A market research firm questions shoppers at the mall to find out whether the mean household income of mall shoppers is higher than that of the general population.

b. Last year, your company’s service technicians took an average of 1.8 hours to respond to trouble calls from business customers who had purchased service contracts. Do this year’s data show a different average response time?

Problem 2.

An article in The Journal of the American Medical Association examined body temperatures of males and females. The mean body temperature of the 65 female subjects who participated in the study was 98.4 degrees Fahrenheit. Assume \( \sigma \) is known to be 0.72 degree.

a) Is there evidence that the population mean body temperature of females is different from 98.6? Perform a test of significance to answer this question. Make sure you state your hypotheses, calculate the value of the test statistic, compute the P-value, and state your conclusion in terms of the problem.

b) Is there evidence that the population mean body temperature of females is greater than 98.6? Perform a test of significance to answer this question. Make sure you state your hypotheses, calculate the value of the test statistic, compute the P-value, and state your conclusion in terms of the problem.

Problem 3.

A manufacturer produces tin cans with wall thickness having a Normal distribution with a standard deviation of \( \sigma = 0.07 \) mm. Tin cans that are too thick or too thin are undesirable. Optimal can thickness is 0.53 mm. A sample of 25 cans is randomly selected and the sample mean is found to be 0.50 mm. Is this evidence that the mean can width differs from the optimal width?

(a) State your hypotheses using mathematical notation (symbols).

(b) In words, state what the null hypothesis means.

(c) Calculate the value of the test statistic.

(d) Determine the p-value.

(e) State your conclusion in terms of the problem.

(f) Is the result above significant at the 1% level (\( \alpha = 0.01 \))? Why or why not?
**Problems for Chapter 17:** Tests of Significance: The Basics.

**Problem 4.**

You are going to proceed a survey among first-year college students about: “How many minutes do you study on a typical weeknight?”

Does the survey give good evidence that students claim to study less than 120 minutes (2 hours) per night on the average?

(a) State null and alternative hypotheses in terms of the mean study time in minutes for the population.

(b) What is the value of the test statistic $z$?

A class survey in a large class for first-year college students showed that the mean response of the 463 students was $\bar{x} = 118$ minutes. Suppose that we know that the study time follows a Normal distribution with standard deviation $\sigma = 65$ minutes in the population of all first-year students at this university.

(c) What is the P-value of the test? What can you conclude?

**Problem 5.**

The same survey, but now we are interested whether the survey gives good evidence that students claim to study more than 120 minutes (2 hours) per night on the average?

(a) State null and alternative hypotheses in terms of the mean study time in minutes for the population.

(b) What is the value of the test statistic $z$?

Now $\bar{x} = 137$ minutes, $n=269$, $\sigma = 65$.

(c) What is the P-value of the test? What can you conclude?