Hypothesis Testing

Example: Building specifications in a city require that the mean breaking strength of residential sewer pipe be more that 2,400 lbs/ft of length. Each manufacturer that wants to sell pipe to the city must show that its pipe meets the specification.

For each manufacturer, we want to decide whether the mean breaking strength of its pipe exceeds 2400 lbs/ft.

Our approach is to define a pair of hypotheses as follows:

- The manufacturer's pipe does not meet specifications.
  Null Hypothesis (H_0): $\mu \leq 2{,}400$
- The manufacturer's pipe meets specifications.
  Alternative (Research) Hypothesis (H_a): $\mu > 2{,}400$

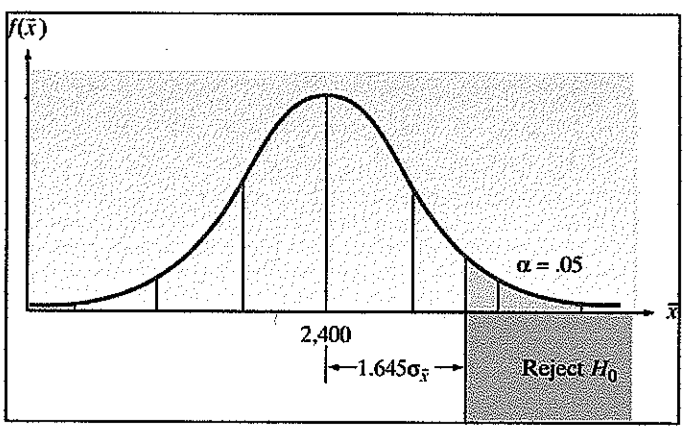
Suppose we test 50 sections of a manufacturer's sewer pipe and find the mean of the 50 measurements to be $\bar{x} = 2{,}460$ and standard deviation $s = 200$ lbs/ft.

Do those results provide strongly "convincing" evidence that the population mean exceeds 2,400 lbs/ft? Strongly "convincing" evidence will exist when the value of $\bar{x}$ exceeds 2,400 by an amount that cannot readily be attributed to sampling variability.

To decide we compute a ***test statistic***, which is the z-value that measures the distance between the value of $\bar{x}$ and the value of $\mu = 2{,}400$. If we can be strongly convinced that $\mu = 2{,}400$ can be rejected in favor of H_a: $\mu > 2{,}400$, then we can certainly reject the null hypothesis H_0: $\mu \leq 2{,}400$.

$$z = \frac{\bar{x} - 2400}{\sigma_{\bar{x}}} = \frac{\bar{x} - 2400}{\sigma / \sqrt{n}} \approx \frac{\bar{x} - 2400}{s / \sqrt{n}} \approx \frac{2460 - 2400}{200 / \sqrt{50}} \approx 2.12$$

So, the sample mean of 2460 lies $2.12\sigma_{\bar{x}}$ above the hypothesized value of $\mu = 2{,}400$. As we examine the figure below we note that if the hypothesis $\mu = 2{,}400$ is true, then the chance of observing $\bar{x}$ more than 1.645 standard deviations above 2,400 is only 0.05. Hence since our sample mean of 2460 is more than 1.645 standard deviations above 2,400, either H_0 and a rare event has occurred or H_a is true and the population mean exceeds 2,400. In this case we will probably reject the null hypothesis and conclude that the alternative hypothesis is probably true. (Of course we could be wrong; so how much faith do we have in our conclusion?)



What kinds of errors might we make in hypothesis testing?

Conclusions and Consequences for a Test of Hypothesis		
	True State of Nature	
Conclusion	H_o True	H_a True
Accept H_0 (Assume H_0 True)	Correct decision	Type II error (probability β)
Reject H_0 (Assume H_a True)	Type I error (probability α)	Correct decision

In our example the probability of making a ***Type I error*** is

$$\alpha = P(z > 1.645 \text{ when } \mu = 2{,}400) = 0.05.$$

Now suppose that it had turned out that the sample mean breaking strength for the 50 sections of pipe tested was $\bar{x} = 2{,}430$ lbs/ft with a standard deviation of $s = 200$. In this case our test statistic is

$$z = \frac{2{,}430 - 2{,}400}{200/\sqrt{50}} = 1.06.$$

The mean of this sample is only 1.06 standard deviations above the hypothesized mean of $\mu = 2400$, and that value does not fall in the region ($z > 1.645$) where we would reject the null hypothesis if we use $\alpha = 0.05$. So, even though the sample mean exceeds the city's specification of 2,400 by 30 lbs/ft, it does not exceed the specification to provide strongly "convincing" evidence that the population mean exceeds 2,400.

If we proceed to accept the null hypothesis H_0: $\mu \leq 2{,}400$ and conclude the manufacturer's pipe does not meet specifications, we risk making a ***Type II error***: concluding that the null hypothesis is true (the pipe does not meet specifications) when in fact it is false (the pipe does meet specifications). We denote the probability of making a Type II error by β. In practice it is usually difficult to determine β. So we will avoid the potential of a Type II error by not concluding the null hypothesis is true. Instead we will simply say that ***the sample evidence is insufficient to reject H_0 at a =0.05***.

We can consider the null hypothesis the "status quo" hypothesis, because the effect of not rejecting H_o is to maintain the status quo.

Elements of a Test of Hypothesis

- *The Null hypothesis (H_0)* – the status quo. What we will accept unless proven otherwise. Stated as *H_0: parameter = value*
- *The Alternative (research) hypothesis (H_a)* – theory that contradicts H_0. Will be accepted if there is evidence to establish its truth
- *Test Statistic* – sample statistic used to determine whether or not to reject H_0 and accept H_a
- *The rejection region* – the region that will lead to H_0 being rejected and *Ha* accepted. Set to minimize the likelihood of a Type I error
- *The assumptions* – clear statements about the population being sampled
- *The Experiment and test statistic calculation* – performance of sampling and calculation of value of test statistic
- *The Conclusion* – decision to (not) reject H_0, based on a comparison of test statistic to rejection region

A Few Questions:

- If you test a hypothesis and reject the null hypothesis in favor of the alternative hypothesis, does your test prove that the alternative hypothesis is correct?

- According to a University of Florida wildlife ecology and conservation researcher, the average level of mercury uptake in wading birds in the Everglades has declined over the past several years. Five years ago, the average level was 15 parts per million (ppm).

  a. Give the null and alternative hypotheses for testing whether the average level today is less than 15 ppm.

  b. Describe a Type I error for this test.

  c. Describe a Type II error for this test.

Large Sample Test of Hypothesis about a Population Mean

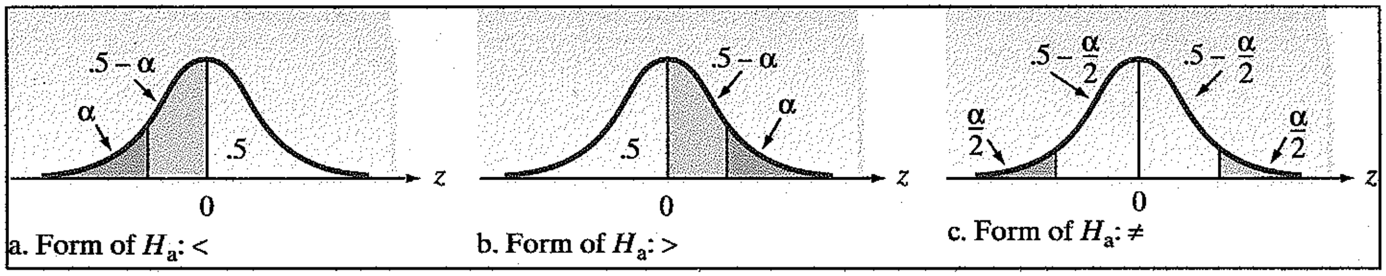
The null hypothesis is the status quo, expressed in one of three forms:

H_0: $\mu = 2400$
H_0: $\mu \leq 2400$
H_0: $\mu \geq 2400$

It represents what must be accepted if the alternative hypothesis is not accepted as a result of the hypothesis test.

The alternative hypothesis can take one of three forms:

One-tailed, upper tail	Ha: $\mu < 2400$
One-tailed, upper tail	Ha: $\mu > 2400$
Two-tailed	Ha: $\mu \neq 2400$



The null hypothesis will be specified as that parameter value closest to the alternative in one-tailed tests and as the complementary (or only unspecified) value in two-tailed tests. (*Example*: H_0: $\mu = 2{,}400$)

The rejection regions corresponding to typical values selected for α are shown in the table below. Note that the smaller α we select, the more evidence (larger z) we will need to reject H_0.

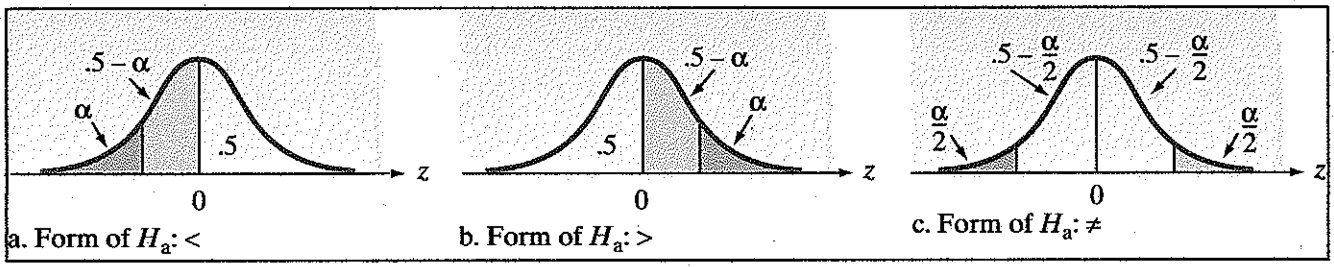


TABLE 8.2 Rejection Regions for Common Values of α

	Alternative Hypotheses		
	Lower-Tailed	Upper-Tailed	Two-Tailed
$\alpha = .10$	$z < -1.28$	$z > 1.28$	$z < -1.645$ or $z > 1.645$
$\alpha = .05$	$z < -1.645$	$z > 1.645$	$z < -1.96$ or $z > 1.96$
$\alpha = .01$	$z < -2.33$	$z > 2.33$	$z < -2.575$ or $z > 2.575$

Example: Suppose we have: n = 100, $\bar{x}$ = 11.85, s = 0.5, and we want to test if $\mu \neq 12$ with a 99% confidence level, our setup would be as follows:

H_0: $\mu = 12$

Ha: $\mu \neq 12$

Test statistic $$z = \frac{\bar{x}-12}{\sigma_{\bar{x}}} = \frac{\bar{x}-12}{\sigma/\sqrt{n}} = \frac{11.85-12}{\sigma/\sqrt{100}} \approx \frac{11.85-12}{s/10} = \frac{-.15}{.5/10} = -3$$

Rejection region z < -2.575 or z > 2.575 (two-tailed)

Since z falls in the rejection region, we conclude that at the 0.01 level of significance the observed mean differs significantly from 12.

p-Values for Statistical Tests

The ***p-value***, or observed significance level, is the smallest α that can be set that will result in the research hypothesis being accepted. Here are the steps in calculation the p-value:

- Determine value of test statistic z
- The p-value is the area to the right of z if Ha is one-tailed, upper tailed
- The p-value is the area to the left of z if Ha is one-tailed, lower tailed.
- The p-valued is twice the tail area beyond z if Ha is two-tailed.

In the above example, p-value = $P(z < -3$ or $z > 3) = 2P(z < -3) \approx 0.0026$ (If H_0 is true 0.0026 is the probability of observing a value of z that is as contradictory to H_0, and supportive of the alternative hypothesis.) We can therefore interpret this p-value as a strong indication that $\mu \neq 12$.