Let $X$ be a measurement with mean $\mu$ and standard deviation $\sigma$. Suppose someone has made the claim that $\mu = M$. If true, then for large random samples of $n$ measurements, the distribution of all possible sample means $\bar{x}$ is approximately $N(M, \sigma / \sqrt{n})$. (The distribution is exactly $N(M, \sigma / \sqrt{n})$ for any sample size when $X$ is normally distributed.) Thus, the bounds that contain the various $\bar{x}$ with probability $r = 1 - \alpha$ are given by

\[
M - \frac{z_{\alpha/2} \sigma}{\sqrt{n}} \leq \bar{x} \leq M + \frac{z_{\alpha/2} \sigma}{\sqrt{n}},
\]

or approximately

\[
M - \frac{z_{\alpha/2} S}{\sqrt{n}} \leq \bar{x} \leq M + \frac{z_{\alpha/2} S}{\sqrt{n}}.
\]

**Example 1.** Suppose the packaged weights of a product are to have a mean of 50 oz. with a standard deviation of 0.7 oz. But a random check 49 packages yields $\bar{x} = 49.7$ oz. Is that unusual? What is the probability of obtaining an $\bar{x}$ as low as 49.7 (i.e. 49.7 or lower) with a sample of size 49?

**Solution.** If $\mu = 50$ oz. and $\sigma = 0.7$, were true, then 95% of the $\bar{x}$ from random samples of size $n = 49$ should be from $50 - \frac{1.96 \times 0.7}{\sqrt{49}}$ to $50 + \frac{1.96 \times 0.7}{\sqrt{49}}$ which is from 49.804 oz. to 50.196 oz. We see that $\bar{x} = 49.7$ is well below the left bound of the confidence interval; thus it is a rare event to obtain an $\bar{x}$ this low.

To find the probability of having $\bar{x} \leq 49.7$, we note that $\bar{x} = N(\mu, \sigma / \sqrt{n}) = N(50, \frac{0.7}{\sqrt{49}}) = N(50, 0.1)$. We then can compute $P(\bar{x} \leq 49.7)$ with the command `normalcdf(-1E99, 49.7, 50, 0.1)`:

We see that $P(\bar{x} \leq 49.7) = 0.00135$.

So it is very rare to obtain an $\bar{x}$ as low as 49.7 if $\mu$ really equals 50 with $\sigma = 0.7$.

If $\mu = 50$ oz. and $\sigma = 0.7$ were true, then there would be only a 0.00135 probability of obtaining an $\bar{x}$ of 49.7 or lower with a sample of size 49. This small probability, called the $P$-value, gives evidence to reject the claim that $\mu = 50$ oz. in favor of the alternative that $\mu < 50$ oz.
If $\mu = M$, then most often we should have $M - \frac{z_{\alpha/2} \sigma}{\sqrt{n}} \leq \bar{x} \leq M + \frac{z_{\alpha/2} \sigma}{\sqrt{n}}$.

Equivalently we would have $-\frac{z_{\alpha/2} \sigma}{\sqrt{n}} \leq \bar{x} - M \leq \frac{z_{\alpha/2} \sigma}{\sqrt{n}}$, or

$$-\frac{z_{\alpha/2} \sigma}{\sqrt{n}} \leq \frac{\bar{x} - M}{\sigma} \leq \frac{z_{\alpha/2} \sigma}{\sqrt{n}}.$$

The value $\frac{\bar{x} - M}{\sigma}$ is called the $z$-test statistic, denoted by $z$. So the test statistic should be within $\pm z_{\alpha/2}$. When the test statistic is too small or too large, that is beyond $\pm z_{\alpha/2}$, then $\bar{x}$ is also too small or too large, and we have evidence to reject the claim that $\mu = M$. In the previous example, the test statistic is $z = \frac{49.7 - 50}{0.7} = -3$. And this value is beyond even the 99% $z$-scores of $\pm 2.576$ which leave only 0.005 probability at each tail. Thus, there is less than 0.005 probability of having a $z$-statistic as low as $-3$. (As we have seen, the probability is actually about 0.00135.)

The Z-Test Feature

Item 1 in the STAT TESTS menu is the Z-Test screen that computes the $z$-test statistic and $P$-value $p$ for such an hypothesis test about the mean. The Inpt can be set to Stats (if we already have $n$ and $\bar{x}$) or to Data if we have entered measurements into a list. The previous example can be written formally as

A null hypothesis of $H_0: \mu = 50$ versus an alternative hypothesis of $H_a: \mu < 50$.

We have chosen this left-sided alternative because the value of $\bar{x} = 49.7$ is less than the initial claim of $\mu = 50$.

Level of Significance

The level of significance, denoted by $\alpha$, is the pre-assigned allowable bound for the $P$-value $p$. If $p < \alpha$, then we believe we have enough evidence to reject the null hypothesis. Generally, we take $\alpha = 0.05$. In the above example, the $P$-value is 0.00135 which is less than $\alpha = 0.05$; so we have strong evidence to reject $H_0$. 
Example 2. IQ scores are to have a mean of \( \mu = 100 \) with \( \sigma = 15 \). But a random sample of 36 varsity athletes at a school yielded an average score of \( \bar{x} = 104 \). At the 0.05 level of significance, is there evidence to reject that \( \mu = 100 \) among varsity athletes at that school? Use the \( P \)-value to state the conclusion. Then explain in terms of the test stat.

Solution. We shall test \( H_0: \mu = 100 \) versus \( H_a: \mu > 100 \) using the Z-Test screen. We obtain a \( P \)-value of \( p = 0.0548 \). (You can enter either Calculate or Draw).

If \( \mu = 100 \) and \( \sigma = 15 \) were true, then there would still be a 5.48\% chance of obtaining an \( \bar{x} \) of 104 or higher with a sample of size 36. This \( P \)-value is larger than our level of significance \( \alpha = 0.05 \); thus, we do not have significant evidence to reject \( H_0 \).

For an alternative of \( H_a: \mu > 100 \) at \( \alpha = 0.05 \) level of significance, the \( z \)-score that creates 0.05 probability at the right-tail is 1.645. The test stat is \( z = (\bar{x} - M) / (\sigma / \sqrt{n}) = 4 / (15 / \sqrt{36}) = 1.6 \), which is within 1.645. So \( \bar{x} = 104 \) is not too high and we do not have enough evidence to reject \( \mu = 100 \).

Two-Sided Alternative

At times we test \( H_0: \mu = M \) versus the alternative \( H_a: \mu \neq M \). In this case, the \( P \)-value gives the probability of \( \bar{x} \) being as far away in either direction from \( M \) as the obtained \( \bar{x} \).

In the previous example, if we test \( H_0: \mu = 100 \) versus \( H_a: \mu \neq 100 \), then we obtain a \( P \)-value of \( p = 0.1096 \).

If \( \mu = 100 \) and \( \sigma = 15 \) were true, then there would be a 10.96\% chance of obtaining an \( \bar{x} \) as far away from 100 (in either direction) as 104 with a sample of size 36. (That is, the probability of being at least 4 away from \( M = 100 \): \( P(\bar{x} \leq 96) + P(104 \leq \bar{x}) = 0.1096 \).) This large \( P \)-value of 0.1096 means that we do not have enough evidence to reject \( H_0 \).
Practice Homework

In each case, compute the $P$-value and explain your conclusion in terms of the $P$-value (see the boxed explanations in Examples 1 and 2.)

1. A student group claims that, on average, first-year students at a university study 2.5 hours per night during the school week. A skeptic suspects that they study less than that on average. A class survey finds that the average study time claimed by 269 students is $\bar{x} = 137$ minutes. Assume that the study time follows a normal distribution with standard deviation $\sigma = 65$ minutes. Carry out a test of $H_0: \mu = 150$ minutes against the one-sided alternative $H_a: \mu < 150$ minutes.

2. The systolic blood pressure for males 35 to 44 years of age varies normally with mean 128 and standard deviation 15. But in a large company, a sample of 72 executives in this age group have a mean systolic blood pressure of $\bar{x} = 126.07$. Is this evidence that the company’s executives in this age group have a different mean systolic blood pressure from the general population? Carry out a two-sided test.

3. The DMS odor thresholds (in µg/liter) for 10 untrained wine tasters are

   31  31  43  36  23  34  32  30  20  24

Assume that the odor threshold is normally distributed with $\sigma = 7$. Is there significant evidence that the mean threshold for untrained tasters is greater than 25?

4. Explain the conclusions of Exercises 1 and 3 in terms of the test-statistic, $z$-score, and $\alpha = 0.05$ level of significance.
1. We test $H_0: \mu = 150$ versus $H_a: \mu < 150$ with the Z-Test.

The $z$-statistic is $-3.28$ which gives a $P$-value of about 0.00052.

Proper Explanation: If $\mu = 150$ and $\sigma = 65$ were true, then there would be only a 0.00052 probability of obtaining an $\bar{x}$ of 137 or lower with a sample of size 269. This small $P$-value gives strong evidence to reject the claim that $\mu = 150$ in favor of the alternative that $\mu < 150$.

2. We test $H_0: \mu = 128$ versus $H_a: \mu \neq 128$ with the Z-Test. The $z$-statistic is $-1.09177$ which gives a two-sided $P$-value of 0.274933.

Proper Explanation: If $\mu = 128$ and $\sigma = 15$ were true, then there would still be a 27.49% chance of obtaining a sample mean as far away from 128 as $\bar{x} = 126.07$ with a sample of size 72. The large $P$-value means we do not have evidence to reject $H_0$.

3. We test $H_0: \mu = 25$ versus $H_a: \mu > 25$. Enter the data into a list, then apply the Z-Test with the Inpt on Data. The $z$-statistic is 2.44 which gives a $P$-value of about 0.00735.

If $\mu = 25$ and $\sigma = 7$ were true, then there would be only a 0.00735 probability of obtaining an $\bar{x}$ of 30.4 or higher with a sample of size 10. This small $P$-value gives strong evidence to reject the claim that $\mu = 25$ in favor of the alternative that $\mu > 25$.

4. Ex. 1: The test stat is $\frac{\bar{x} - M}{\frac{\sigma}{\sqrt{n}}} = \frac{137 - 150}{\frac{65}{\sqrt{269}}} = -3.28$. For an alternative of $H_a: \mu < 150$ at $\alpha = 0.05$ level of significance, the $z$-score that creates 0.05 probability at the left-tail is $-1.645$. The test stat is beyond $-1.645$; so $\bar{x} = 137$ is too low and gives us significant evidence to reject $\mu = 150$.

Ex. 3: For an alternative of $H_a: \mu > 25$ at $\alpha = 0.05$ level of significance, the $z$-score that creates 0.05 probability at the right-tail is $+1.645$. The test stat of 2.44 is beyond 1.645, so $\bar{x} = 30.4$ is too high and gives us enough evidence to reject $\mu = 25$. 