1. A class survey in a large class for first-year college students asked, "About how many minutes do you study on a typical weeknight?" 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. Regard these students as an SRS from the population of all first-year students at this university. Does the study give good evidence that students claim to study less than 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?

(c) What is the P-value of the test? Can you conclude that students do claim to study less than 2 hours per weeknight on the average?

2. Young men in North America and Europe (but not in Asia) tend to think they need more muscle to be attractive. One study presented 200 young American men with 100 images of men with various levels of muscle. Researchers measure level of muscle in kilograms per square meter (kg/m^2) of fat-free body mass. Typical young men have about 20 kg/m^2. Each subject chose two images, one that represented his own level of body muscle and one that he thought represented "what women prefer." The mean gap between self-image and "what women prefer" was 2.35 kg/m^2. Suppose that the "muscle gap" in the population of all young men has a Normal distribution with standard deviation 2.5 kg/m^2. We suspect (before seeing the data) that young men think women prefer more muscle than they themselves have.

(a) State null and alternative hypotheses for testing this suspicion.

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

(c) You can tell just from the value of z that the evidence in favor of the alternative is very strong (that is, the p-value is very small). Explain why this is true.

3. Successful hotel managers must have personality characteristics often thought of as feminine (such as "compassionate") as well as those often thought of as masculine (such as "forceful"). The Bem Sex-Role Inventory (BSRI) is a personality test that gives separate ratings for female and male stereotypes, both on a scale of 1 to 7. A sample of 148 male general managers of three-star and four-star hotels had mean BSRI femininity score x = 5.29. The mean score for the general male population is $\mu = 5.19$. Do hotel managers, on the average, differ significantly in femininity score from men in general? Assume that the standard deviation of scores in the population of all male hotel managers is the same as the $\sigma = 0.78$ for the adult male population.

(a) State null and alternative hypotheses in terms of the mean femininity score μ for male hotel managers.

(b) Find the z test statistic.

(c) What is the P-value for your z? What do you conclude about male hotel managers?

4. Every society has its own marks of wealth and prestige. In ancient China, it appears that owning pigs was such a mark. Evidence comes from examining burial sites. The skulls of sacrificed pigs tend to appear along with expensive ornaments, which suggests that the pigs, like the ornaments, signal the wealth and prestige of the person buried. A study of burials from around 3500 B.C. concluded that "there are striking differences in grave goods between burials with pig skulls and burials without them.... A test indicates that the two samples of total artifacts are significantly different at the 0.01 level." Explain clearly why "significantly different at the 0.01 level" gives good reason to think that there really is a systematic difference between burials that contain pig skulls and those that lack them.

5. How heavy a load (pounds) is needed to pull apart pieces of Douglas fir 4 inches long and 1.5 inches square? Here are data from students doing a laboratory exercise:

33190	32320	23040	24050	31860
33020	30930	30170	32590	32030
32720	31300	26520	30460	33650
28730	33280	32700	32340	31920

We are willing to regard the wood pieces prepared for the lab session as an SRS of all similar pieces of Douglas fir. Engineers also commonly assume that characteristics of materials vary Normally. Suppose that the strength of pieces of wood like these follows a Normal distribution with standard deviation 3000 pounds.

(a) Is there significant evidence at the $\alpha = 0.10$ level against the hypothesis that the mean is 32,500 pounds for the two-sided alternative?

(b) Is there significant evidence at the $\alpha = 0.10$ level against the hypothesis that the mean is 31,500 pounds for the two-sided alternative?

6. Breast-feeding mothers secrete calcium into their milk. Some of the calcium may come from their bones, so mothers may lose bone mineral. Researchers measured the percent change in mineral content of the spines of 47 mothers during three months of breast-feeding. Here are the data:

-4.7	2.2	-6.5	-4.0	0.3	-2.5	-4.9	-7.8	-3.1	-1.0
-3.0	-4.9	-4.7	-2.3	0.4	-2.7	-1.0	-3.6	-3.8	-5.3
-0.8	-6.5	-5.2	-5.9	0.2	-5.3	-1.8	-2.0	-2.5	-2.2
-8.3	-5.2	-2.1	-0.3	-5.1	-2.1	-6.8	-4.3	-5.7	-7.0
-2.2	-5.6	-4.4	-3.3	-6.2	-6.8	1.7			

The researchers are willing to consider these 47 women as an SRS from the population of all nursing mothers. Suppose that the percent change in this population has a Normal distribution with standard deviation $\sigma = 2.5\%$. Do these data give good evidence that, on the average, nursing mothers lose bone mineral?

7. Sulfur compounds cause "off-odors" in wine, so winemakers want to know the odor threshold, the lowest concentration of a compound that the human nose can detect. The odor threshold for dimethyl sulfide (DMS) in trained wine tasters is about 25 micrograms per liter of wine (μg/L). The untrained noses of consumers may be less sensitive, however. Here are the DMS odor thresholds for 10 untrained students:

30	30	42	35	22	33	31	29	19	23

(a) Assume that the standard deviation of the odor threshold for untrained noses is known to be \sigma = 7 μg/L.

Is there evidence that the mean threshold for untrained tasters is greater than 25 μg/L?

8. A confidence interval for the population mean μ tells us which values of μ are plausible (those inside the interval) and which values are not plausible (those outside the interval) at the chosen level of confidence. You can use this idea to carry out a test of any null hypothesis $H_0 : \mu = \mu_0$ starting with a confidence interval: reject H_0 if μ_0 is outside the interval and fail to reject if μ_0 is inside the interval. The alternative hypothesis is always two-sided, $H_a : \mu \neq \mu_0$, because the confidence interval extends in both directions from $\bar{x}$. A 95% confidence interval leads to a test at the 5% significance level because the interval is wrong 5% of the time. In general, confidence level C leads to a test at significance level $\alpha = 1 - C$.

A 95% confidence interval for a population mean is 30.7 ± 3.2. Use the method described above to answer these questions.

(a) With a two-sided alternative, can you reject the null hypothesis that $\mu = 33$ at the 5% ($\alpha = 0.05$) significance level? Why?

(b) With a two-sided alternative, can you reject the null hypothesis that $\mu = 34$ at the 5% significance level? Why?