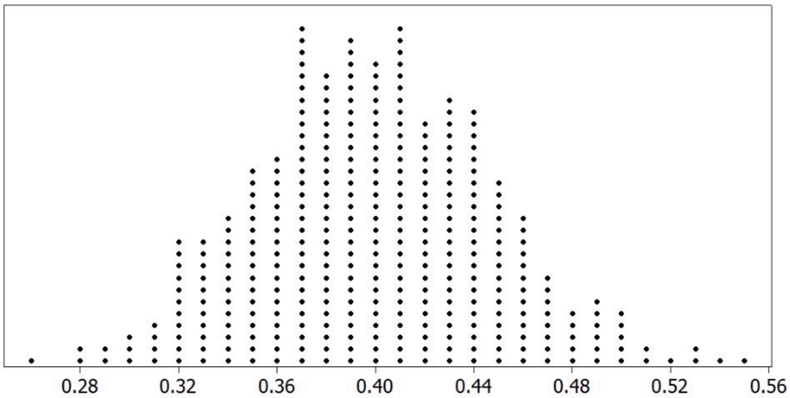
6.16 (a) The graph below shows the sample proportions for 1000 samples of size 100 which were simulated from a population where $p = 0.4$. The distribution is relatively symmetric and bell-shaped so a normal distribution is appropriate. The mean of these simulated sample proportions is 0.398 and the standard error is 0.047. Answers will vary with different simulations but will always be approximately 0.4 and 0.049, respectively.

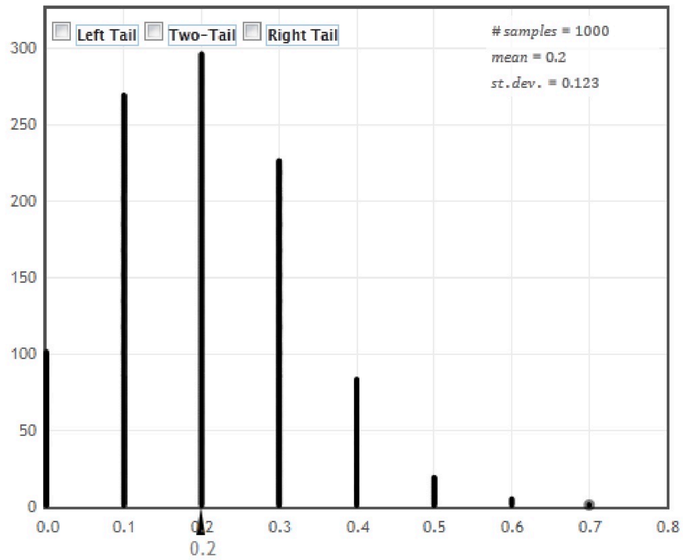


(b) The Central Limit Theorem for proportions says the expected mean of the sample proportions is p, which is 0.4, and the expected standard error is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(1-0.4)}{100}} = 0.049$$

These are both very close to those obtained with simulations.

6.18 (a) The graph below shows the sample proportions for 1000 samples of size 10 which were simulated from a population where $p = 0.2$. The distribution is not symmetric since the left tail is truncated at zero. This is not surprising since $np = 10 \cdot 0.2 = 2$ which is well less than the desired value of 10 that is needed in the Central Limit Theorem. The mean of these simulated sample proportions is 0.20 and the standard error is 0.123. Answers will vary with different simulations but should be approximately 0.20 and 0.126, respectively.



(b) The Central Limit Theorem for proportions says the expected mean of the sample proportions is p, which is 0.2, and the expected standard error is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2(1-0.2)}{10}} = 0.126$$

These are both very close to those obtained with simulations. Note that even though the sample size is not large enough to give a normal distribution, the mean and standard error of the sample proportions still agree with the Central Limit Theorem.

6.30 The desired margin of error is $ME = 0.05$ and we have $z^* = 1.96$ for 95% confidence. Since we are given no information about the population parameter, we use the conservative estimate $\tilde{p} = 0.5$. We use the formula to find sample size:

$$n = \left(\frac{z^*}{ME}\right)^2 \tilde{p}(1-\tilde{p}) = \left(\frac{1.96}{0.05}\right)^2 (0.5 \cdot 0.5) = 384.2.$$

We round up to $n = 385$. In order to ensure that the margin of error is within the desired ±5%, we should use a sample size of 385 or higher.

6.32 The desired margin of error is $ME = 0.03$ and we have $z^* = 1.645$ for 90% confidence. We estimate that p is about 0.3, so we use $\tilde{p} = 0.3$. We use the formula to find sample size:

$$n = \left(\frac{z^*}{ME}\right)^2 \tilde{p}(1-\tilde{p}) = \left(\frac{1.645}{0.03}\right)^2 (0.3 \cdot 0.7) = 631.4.$$

We round up to $n = 632$. In order to ensure that the margin of error is within the desired ±3%, we should use a sample size of 632 or higher.

6.34 The sample size is definitely large enough to use the normal distribution. For a confidence interval using the normal distribution, we use

$$\text{Sample statistic} \pm z^* \cdot SE.$$

The relevant sample statistic for a confidence interval for a proportion is $\hat{p} = 0.20$. For a 99% confidence interval, we have $z^* = 2.576$, and the standard error is $SE = \sqrt{\hat{p}(1-\hat{p})/n}$. The confidence interval is

$$\begin{array}{rcl} \hat{p} & \pm & z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \\ 0.20 & \pm & 2.576 \cdot \sqrt{\dfrac{0.20(0.80)}{1000}} \\ 0.20 & \pm & 0.033 \\ 0.167 & \text{to} & 0.233 \end{array}$$

We are 99% confident that the proportion of US adults who say they never exercise is between 0.167 and 0.233. The margin of error is ±3.3%.

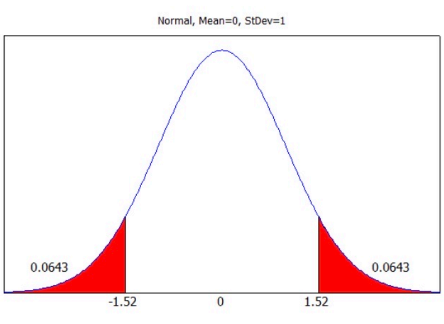
6.60 Since $np_0 = 120(0.75) = 90$ and $n(1-p_0) = 120(0.25) = 30$, the sample size is large enough to use the normal distribution. In general, the standardized test statistic is

$$z = \frac{\text{Sample Statistic} - \text{Null Parameter}}{SE}.$$

In this test for a proportion, the sample statistic is $\hat{p} = 0.69$ and the parameter from the null hypothesis is $p_0 = 0.75$. The standard error is $SE = \sqrt{p_0(1-p_0)/n}$. The standardized test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.69 - 0.75}{\sqrt{\frac{0.75(0.25)}{120}}} = -1.52.$$

This is a two-tail test, so the p-value is two times the area below −1.52 in a standard normal distribution. Using technology or a table, we see that the p-value is $2(0.0643) = 0.1286$. This p-value is not very small and is not significant at any reasonable significance level. We do not find evidence to support the alternative hypothesis that $p \neq 0.75$.



6.66 The sample proportion of questions having B as the correct answer is $\hat{p} = 90/400 = 0.225$. If all the choices were equally likely, we would expect each to be correct about 1/5, or 20%, of the time. If p represents the proportion of time B is the correct choice on all AP multiple choice questions, the hypotheses are:

$$\begin{aligned} H_0 &: \quad p = 0.20 \\ H_a &: \quad p > 0.20 \end{aligned}$$

The test statistic is:

$$z = \frac{\text{Sample statistic} - \text{Null parameter}}{SE} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.225 - 0.20}{\sqrt{\frac{0.2(0.8)}{400}}} = 1.25.$$

This is an upper-tail test, so the p-value is the area above 1.25 in a standard normal distribution. We find the p-value is 0.106. Even at a 10% level, we do not reject H_0. We do not find evidence that B is more likely to be the correct choice.

6.108 For a confidence interval for μ using the t-distribution, we use

$$\overline{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

We use a t-distribution with df $= 49$, so for a 90% confidence interval, we have $t^* = 1.68$. The confidence interval is

$$\begin{array}{ccc} 137.0 & \pm & 1.68 \cdot \dfrac{53.9}{\sqrt{50}} \\ 137.0 & \pm & 12.8 \\ 124.2 & \text{to} & 149.8 \end{array}$$

The best estimate for μ is $\overline{x} = 137$, the margin of error is ± 12.8, and the 90% confidence interval for μ is 124.2 to 149.8. We are 90% confident that the mean of the entire population is between 124.2 and 149.8.

6.118 We use a t-distribution with df $= 98$, so for a 95% confidence interval, we have $t^* = 1.98$. The confidence interval is

$$\begin{array}{ccc} \overline{x} & \pm & t^* \cdot \dfrac{s}{\sqrt{n}} \\ 564 & \pm & 1.98 \cdot \dfrac{122}{\sqrt{99}} \\ 564 & \pm & 24.3 \\ 539.7 & \text{to} & 588.3 \end{array}$$

We are 95% sure that the mean number of unique genes in the gut bacteria of European individuals is between 539.7 and 588.3 million.

6.128 (a) For a confidence interval for μ using the t-distribution, we use

$$\overline{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

Since the sample size is so large, we can use either a t-distribution or a standard normal distribution. Using a t-distribution with df $= 119$, for a 99% confidence interval, we have $t^* = 2.62$. The confidence interval is

$$\begin{array}{ccc} 290 & \pm & 2.62 \cdot \dfrac{87.6}{\sqrt{120}} \\ 290 & \pm & 21.0 \\ 269.0 & \text{to} & 311.0 \end{array}$$

We are 99% confident that the mean number of polyester microfibers entering wastewater when washing a fleece garment is between 269 and 311 per liter.

(b) The margin of error is ± 21.0 microfibers per liter.