Solutions

In-Class Activities

Activity 19-1: Christmas Shopping

a. The *amount expected to be spent on Christmas presents in 1999* is a quantitative variable.

b. The value \$857 is a statistic because it is a number that describes a sample. This statistic is represented by $\bar{x}$.

c. The parameter is the average (mean) amount expected to be spent by all American adults on Christmas presents in 1999. This parameter is represented by μ.

d. You do not *know* the value of the μ, but it is more likely to be close to \$857 than to be far from it.

e. The standard deviation of the sample mean $\bar{x}$ is $\sigma/\sqrt{n} = 250/\sqrt{922} = \$8.23$.

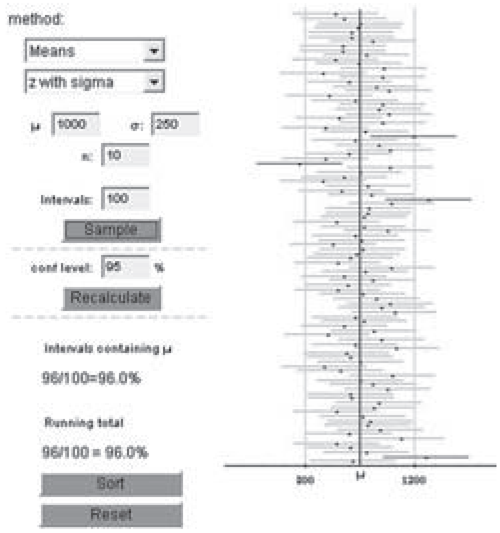
f. This interval estimate works out to be $\$857 \pm 2(\$8.23) = \$857 \pm \$16.47 =$ (\$840.53, \$873.47).

g. The sample standard deviation, s, is a reasonable substitute for σ.

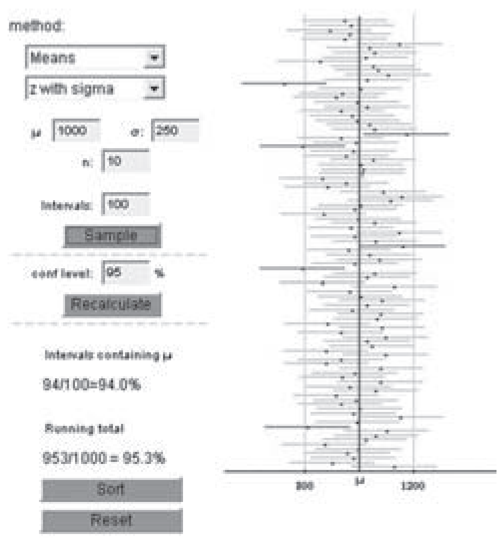
19

h. Answers will vary. The following is from one representative running of the applet:

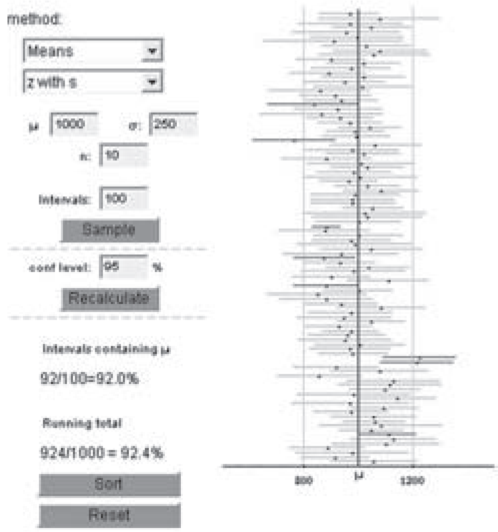
You find 96% of the intervals succeed in capturing the value of μ. Your total should be roughly 95%.



i. The running total percentage of intervals that succeed in capturing the population mean is 95.3%, which is very close to the expected 95%.

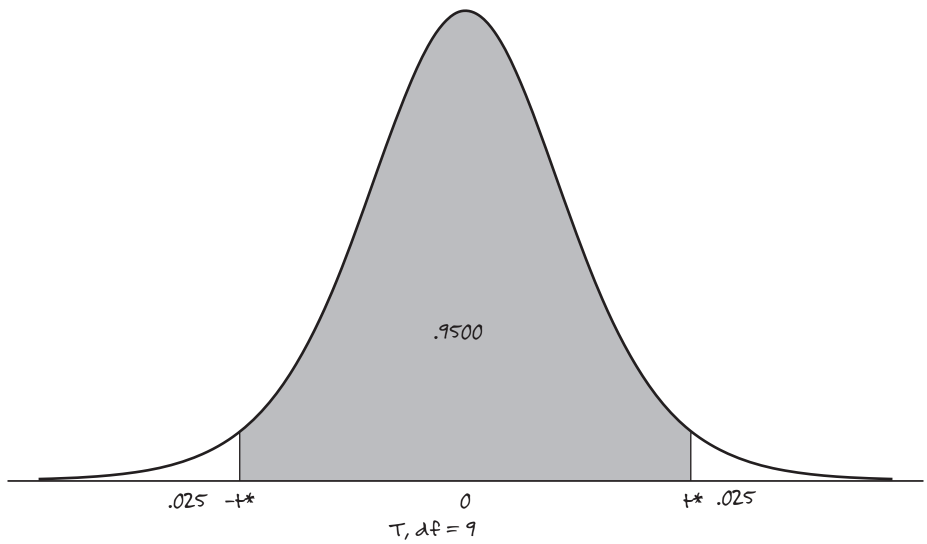


j. The running total percentage of intervals that succeed in capturing the population mean is 92.4%, which is noticeably less than 95%.



Activity 19-2: Exploring the *t*-Distribution

a. Here is a sketch of the $t(9)$ distribution:



b. See shading in plot above.

c. The area to the right of $t^* = 1 - .975 = .025$.

d. Using the *t*-table (Table III), $t^* = 2.262$.

e. This t^* critical value is greater than the z^* critical value for a 95% confidence interval because of the greater uncertainty introduced by estimating with s rather than σ. This will make the margins-of-error of your confidence intervals wider in order to achieve the stated confidence level in the long run.

f. For a 95% confidence interval, $t^* = 2.045$. This value is less than the previous t^* value, which makes sense because you have increased the sample size, decreasing the uncertainty in estimating σ by s.

g. For a 90% confidence interval $t^* = 1.699$; for a 99% confidence interval $t^* = 2.759$. The t^* for a 99% confidence interval is greater, which is appropriate because this interval claims more confidence (more certainty). In order to be more confident, the interval will need to be wider.

h. With 100 degrees of freedom, $t^* = 1.984$.

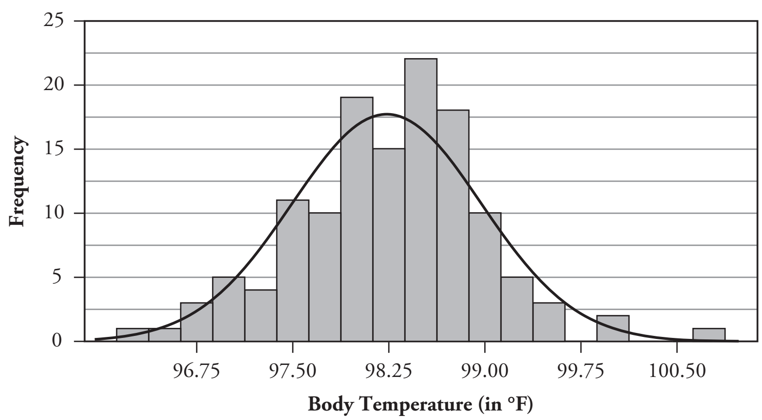
Activity 19-3: Body Temperatures

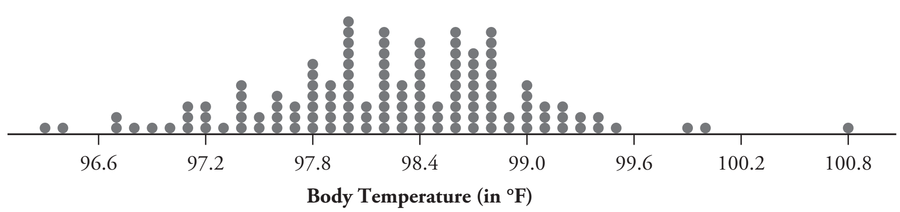
a. For a 95% CI with 129 degrees of freedom, you calculate $98.249 \pm (1.984)(.733)/\sqrt{130} = (98.1215, 98.3765)$.

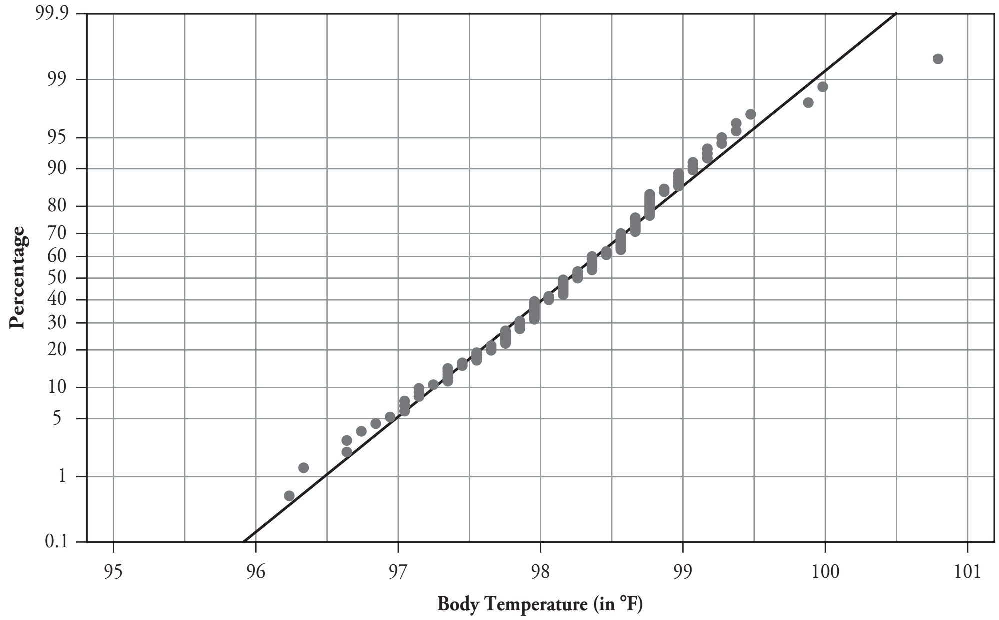
b. You are 95% confident the average body temperature of a sample of 130 healthy adults is between 98.12°F and 98.38°F.

When you say 95% confident, you mean that if you repeated this procedure of creating confidence intervals in this same manner (using random samples of 130 healthy adults), in the long run 95% of the intervals would contain the population mean body temperature of all healthy adults and 5% of the intervals would not contain this parameter.

c. Here are graphs of the sample data:







Yes, the body temperatures appear to be roughly normally distributed.

d. No, the normality of the population of body temperatures is not required for this *t*-procedure to be valid with these data because the sample size is large ($n = 130$).

e. You do not know whether this was a simple random sample of healthy adults. If you assume that it was, then the other technical condition required for the validity of this *t*-interval is satisfied.

f. Using Minitab's One-Sample T procedure,

```
Variable       N      Mean    StDev    SE Mean         95% CI
body temp    130    98.2492    0.7332     0.0643   (98.1220, 98.3765)
```

90% CI: (98.1427, 98.3558)   99% CI: (98.0811, 98.4174)

g. The midpoints of all three intervals are the same: 98.249 (the sample mean). The 90% confidence interval is the most narrow (width = .2131), followed by the 95% confidence interval (width = .2545), whereas the 99% confidence interval is the widest (width = .3363).

h. It does not appear that 98.6 is a plausible value for the mean body temperature for the population of all healthy adults because this value is not contained in any of the confidence intervals.

i. If the sample size had been only 13, but the sample mean and standard deviation had been the same, the 95% CI would be much wider (though it would have the same center) because (i) the t^* value used to create the interval would be much greater and (ii) the standard error would be greater (the square root of 13 is much smaller than the square root of 130).

j. For a 95% confidence interval with 12 degrees of freedom, you calculate $98.249 \pm (2.179)\,(.733)/\sqrt{13} = (97.807, 98.692)$°F.

As predicted, the midpoint is still 98.249, but the width is much greater (.88597). In fact, with this interval, 98.6 would be a plausible value for the mean body temperature for the population of all healthy adults.

19

Activity 19-4: Sleeping Times

a. Here is the completed table:

Sample Number	Sample Size	Sample Mean	Sample SD
3	30	6.6	0.825
1	10	6.6	0.825
2	10	6.6	1.597
4	30	6.6	1.597

b. They all have a sample mean of 6.6 hours.

c. The most important difference between samples 1 and 2 are the spreads (sample SDs). Sample 1 has a much smaller SD than does sample 2.

d. The most important difference between samples 1 and 3 is the sample size. Sample 3 uses a sample of size 30, whereas sample 1 uses a sample of size 10.

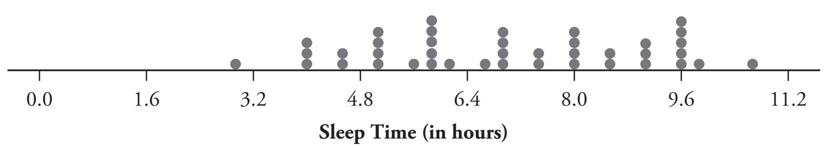
e. Sample 1 produced a more precise estimate of μ. This result makes sense because sample 1 has a smaller SD than sample 2, and so sample 1 will have a smaller margin-of-error.

f. Sample 3 produced a more precise estimate of μ. This result makes sense because sample 3 has a larger sample size, and so sample 3 will have a smaller margin-of-error.

Activity 19-5: Sleeping Times

Results will vary by class, but here are the results from one college class.

a. The following graph displays the results:



These 40 sleep times are roughly normally distributed, centered around 7 hours, with a minimum of 3 hours and a maximum of 10.5 hours. There are no noticeable outliers.

b. The sample size is $n = 40$; the sample mean is $\bar{x} = 6.981$ hours; and the sample standard deviation is $s = 1.981$ hours.

c. The sample size is large ($40 > 30$), but the sample was not randomly selected, because it consisted of the students in this one class. It might not be representative of students at the entire school with regard to sleep hours, as students in a statistics class may tend to be mostly from one type of major who may tend to study and sleep more or less than the typical student.

d. For a 90% CI, with 39 degrees of freedom, you calculate $6.981 \pm (1.685)(.31322) = (6.45322, 7.50878)$ hours.

You are 90% confident the average amount of sleep per night obtained by all students at the school is between 6.45 and 7.51 hours.

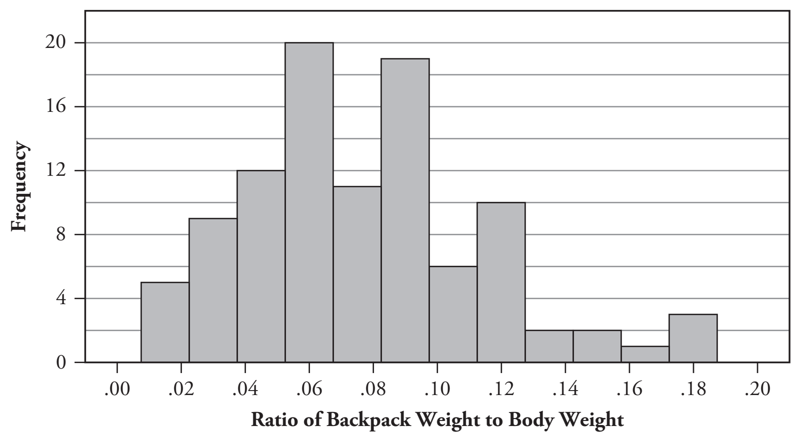
e. Seven of the 40 sleep times fall within this interval. This is 17.5%.

f. No, this percentage is not close to 90%, but there is no reason that it should be. Your interval is designed to estimate the *average* sleep time. It is not telling you anything about the individual sleep times. They may or may not fall within this interval.

Activity 19-6: Backpack Weights

a. The observational units are the students. The variable is the *ratio of backpack weight to body weight,* which is quantitative. The sample is the 100 Cal Poly students whose weights were recorded by the student researchers. The population is all Cal Poly students at the time the study was conducted.

b. The following histogram reveals that the distribution of these weight ratios is a bit skewed to the right. The center is around .07 or .08 (mean $\bar{x} = .077$, median = .071). The five-number summary is (.016, .050, .071, .096, .181), so students in the sample carried as little as 1.6% of their weight in their backpacks and as much as 18.1% of their weight in their backpacks. The standard deviation of these ratios is $s = .037$.



c. Calculating a 99% CI for the population mean by hand using the formula

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

with $100 - 1 = 99$ rounding down to 80 degrees of freedom gives $.0771 \pm 2.639 \times .0366/\sqrt{100}$, which is $.0771 \pm .0097$, which corresponds to the interval from .0674 through .0868. (This t^* value should be based on 99 degrees of freedom, but we used 80 degrees of freedom here, the closest value less than 99 that appears in Table III.) Using technology gives a slightly more accurate 99% CI for μ of .0675 through .0867.

d. You are 99% confident that the mean weight ratio of backpack-to-body weights among all Cal Poly students at the time of this study is between .0674 and .0868. In other words, you are 99% confident that the average Cal Poly student carries between 6.74% and 8.68% of his/her body weight in his/her backpack. By "99% confidence," you mean that 99% of all intervals constructed with this method would succeed in capturing the actual value of the population mean weight ratio.

e. The first condition is that the sample be randomly selected from the population. This is not literally true in this case because the student researchers did not obtain a list of all students at the university and select randomly from that list, but they did try to obtain a representative sample. The second condition is either that the population of weight ratios is normal or that the sample size is large. In this case, the sample size is large ($n = 100$, which is greater than 30), so this condition is satisfied even though the distribution of ratios in the sample is somewhat skewed (and so presumably is the population).

f. You do not expect 99% of the sample, nor 99% of the population, to have a weight ratio between .0673 and .0869. You are 99% confident that the population *mean* weight ratio is between these two endpoints. In fact, only 18 of the 100 students in the sample have a weight ratio in this interval.