(c) No. The "95%" refers to the percentage (in a meta-experiment) of confidence intervals that would contain $\mu$. Since the width of a confidence interval depends on n, the percentage of observations contained in the confidence interval also depends on n, and would be very small if n were large.

**6.3.6** (a) This statement is false. The confidence interval allows us to make an inference concerning the mean of the entire population. We *know* that $59.77 < \bar{y} < 61.09$.

(b) This statement is true. (See part (a).)

**6.3.7** (a) In Table 4 we observe that $t_{0.10}$ is smaller than $t_{0.025}$ for all df values, thus mathematically the 80% confidence intervals are narrower than the 95% intervals.

(b) The larger sample size would yield a smaller standard error of the mean (and slightly smaller t-multiplier due to the increase in df). Thus, the interval using n=500 mice would be narrower than the interval with only n=86 mice.

**6.3.8** (a) This statement is false. The confidence interval concerns the *mean* of the population. It does not tell us where individual sample data points lie.

(b) This statement is false. The confidence interval concerns the *mean* of the population. It does not tell us where individual population values lie.

**6.3.9** The higher the confidence level, the greater the length of the interval. Thus, 90%, 85%, and 80% confidence intervals correspond to (0.822,0.858), (0.824,0.856), and (0.826,0.854), respectively.

**6.3.10** As the confidence level goes up the CI gets wider. Thus the 85% CI is (8.38, 10.66) – the most narrow – and the 95% CI is (7.75, 11.29) – the widest. This leaves (8.16, 10.88) as the 90% CI.

• **6.3.11** (a) $\bar{y} = 13.0$; $s = 12.4$; $n = 10$.
The degrees of freedom are $n - 1 = 10 - 1 = 9$. The critical value is $t_{0.025} = 2.262$. The 95% confidence interval for $\mu$ is

$$\bar{y} \pm t_{0.025}\frac{s}{\sqrt{n}}$$

$$13.0 \pm 2.262\left(\frac{12.4}{\sqrt{10}}\right)$$

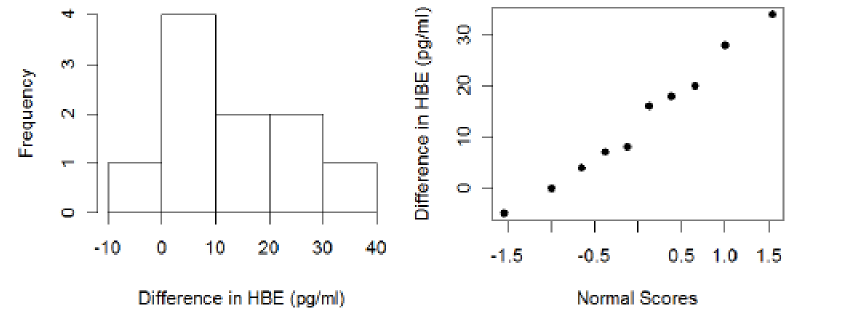
(4.1,21.9) or $4.1 < \mu < 21.9$ pg/ml.

(b) We are 95% confident that the average drop in HBE levels from January to May in the population of all participants in physical fitness programs like the one in the study is between 4.1 and 21.9 pg/ml.

(c) From the computed 95% confidence interval we have evidence that the mean drop in HBE is at least 4.1 pg/ml. So yes, there is compelling evidence that the HBE level tends to be lower in May than January, on average.



**6.3.12** A histogram and normal probability plot of the data, shown here, support the use of a normal curve model for these data.




**6.3.13** (a) $5{,}111 \pm (2.306)(818/\sqrt{9})$

(4482,5740) or $4{,}482 < \mu < 5{,}740$ units

(b) We are 95% confident that the average invertase activity of all fungal tissue incubated at 95% relative humidity for 24 hours is between 4,482 units and 5,740 units.

(c) To check that the data are from a normal population we can make a normal probability plot.

**6.3.14** $6.21 \pm (2.042)(1.84/\sqrt{36})$

(5.58,6.84) or $5.58 < \mu < 6.84$ µg/dl.

• **6.3.15** $\bar{y} = 1.20$; $s = .14$; $n = 50$.
The degrees of freedom are $50 - 1 = 49$. From Table 4 with df = 50 (the df value closest to 49) we find that $t_{0.05} = 1.676$. The 90% confidence interval for $\mu$ is

$$\bar{y} \pm t_{0.05}\frac{s}{\sqrt{n}}$$

$$1.20 \pm 1.676\left(\frac{0.14}{\sqrt{50}}\right)$$

(1.17,1.23) or $1.17 < \mu < 1.23$ mm.

**6.3.16** (a) We are 95% confident that the mean Bayley Index of prematurely born infants who receive intravenous-feeding solutions is between 93.8 and 102.1.

(b) Although the center of the interval is 97.95, which is less than the general population average of 100, the interval extends above 100, so we cannot be sure that $\mu$ is less than 100.

**6.3.17** The 95% CI would be 97.95 ± t*SE where the t multiplier would be approximately 2 and the SE would be $20.04/\sqrt{n}$. Thus if n satisfied $2*20.04/\sqrt{n} = 100-97.95$ the CI would barely touch 100. Solving for n gives $\left(\frac{2*20.04}{100-97.95}\right)^2 = 382.25$. Using n = 383 would give a 95% CI of (95.90, 99.99) which excludes 100. [Note: We could use t=1.984 rather than t=2 since we know that n will be greater than 92 and thus the degrees of freedom will be large enough to make the t multiplier 1.984 or smaller. Using t=1.984 in the argument above gives n = 377. In fact, by trial and error we can find that n = 370 is large enough.]

**6.3.18** $\bar{y} = 10.3$; s = 0.9; n = 101. There are 100 degrees of freedom, so $t_{0.025} = 1.984$. The 95% confidence interval for μ is

$$10.3 \pm 1.984\left(\frac{0.9}{\sqrt{101}}\right)$$

(10.12,10.48) or $10.12 < \mu < 10.48$ g/dLi.

**6.3.19** 1 - 0.025 = 0.975. In Table 3, z = 1.96 corresponds to an area of 0.975. (A t distribution with df = ∞ is a normal distribution.)

• **6.3.20** 1 - 0.0025 = 0.9975. In Table 3, an area of 0.9975 corresponds to z = 2.81. A t distribution with df = ∞ is a normal distribution; thus, $t_{0.0025} = 2.81$ when df = ∞.

**6.3.21** The confidence coefficient is approximately 68%, because a t distribution with df = ∞ is a normal distribution, and the area under a normal curve between z = -1.00 and z = 1.00 is approximately 68%.

**6.3.22** **(a)** Smaller. The area included between t = -1.00 and t = 1.00 is smaller for a Student's t distribution than for a normal distribution.

**(b)** It is not affected, because of the Central Limit Theorem.

**6.4.1** **(a)** Guessed SD = 20 kg; n must satisfy the inequality

$$\frac{20}{\sqrt{n}} \le 5$$

so n = 16.

**(b)** n must satisfy the inequality

$$\frac{40}{\sqrt{n}} \le 5$$

so n = 64. The required sample size does not double, but rather is four times as large.

• **6.4.2** We use the inequality



$$\frac{\text{Guessed SD}}{\sqrt{n}} \le \text{Desired SE}.$$

In this case, the desired SE is 3 mg/dl and the guessed SD is 40 mg/dl. Thus, the inequality is

$\frac{40}{\sqrt{n}} \le 3$ or $\frac{40}{3} \le \sqrt{n}$ which means that $n \ge 177.8$, so a sample of n = 178 men is needed.

**6.4.3** Guessed SD = 1.2 cm. The desired SE is 0.2 cm, so n must satisfy

$$\frac{1.2}{\sqrt{n}} \le 0.2$$

which yields $n \ge 36$.

**6.4.4** Guessed SD = 80 g

**(a)** The desired SE is 20 g, so n must satisfy

$$\frac{80}{\sqrt{n}} \le 20$$

which yields $n \ge 16$.

**(b)** The desired SE is 15 g, so n must satisfy

$$\frac{80}{\sqrt{n}} \le 15$$

which yields $n \ge 28.4$, so n = 29.

**6.4.5** The sample mean for the females must have been (197.2 + 235.1)/2 = 216.15, the mid-point of the CI. Thus the margin of error was 235.1 – 216.15 = 18.95, which is $t^*\frac{SD}{\sqrt{29}}$ where t = 2.048. Thus the SD must have been $\frac{18.95*\sqrt{29}}{2.048} = 49.8$. For the male CI we want the margin of error to be 10, but this is (approximately) 2xSE so we want the SE to be 5 cm. Thus $5 = \frac{49.8}{\sqrt{n}}$. Solving for n and rounding up gives n = 100.

• **6.4.6** Guessed SD = 1.5 in

The SE should be no more than .25 in, so n must satisfy

$$\frac{1.5}{\sqrt{n}} \le .25$$

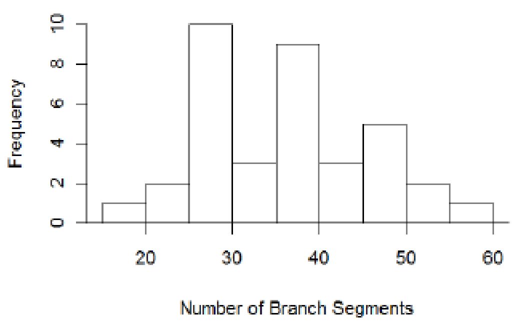
which yields $n \ge 36$.

• **6.5.1** The fact that the mean is less than the SD casts doubt on the condition that the population is normal, for the following reason. In a normal population, about 15% of the observations fall more than one SD below the mean, whereas this sample cannot have any observations that far below the mean because $\bar{y} - s$ is negative and the observed variable (serum SGOT) cannot be negative.

**6.5.2 (a)** There were 36 cells, but only seven guinea pigs, so there is a hierarchical structure in the data, which suggests that the observations are not independent.

**(b)**

| Number of branch segments | Frequency |
|---|---|
| 15-19 | 1 |
| 20-24 | 2 |
| 25-29 | 10 |
| 30-34 | 3 |
| 35-39 | 9 |
| 40-44 | 3 |
| 45-49 | 5 |
| 50-54 | 2 |
| 55-59 | 1 |
| Total | 36 |




The distribution has two or perhaps three modes, which may reflect the hierarchical structure in the data (that is, different modes may represent different animals or groups of animals.)

**6.5.3** The outlier (1,060) suggests that the population distribution is not normal but rather is skewed to the right or long-tailed. Because the sample size is small, Student's t method is not appropriate if the population is not normal.

**6.5.4 (a)** No. The observed frequency distribution is highly skewed, which suggests that the population distribution is skewed.

**(b)** Even if the population distribution is not normal, Student's t method is approximately valid if the sample size is large. In this case, the sample size is n = 242, which is quite large.



**6.5.5 (a)** $\bar{y} = 5.68$; $s = 1.54$; $n = 9$.

The 90% confidence interval for μ is

$$5.68 \pm 1.860\left(\frac{1.54}{\sqrt{9}}\right)$$

$(4.73, 6.63)$ or $4.73 < \mu < 6.63\sqrt{\text{cm}}$.

**(b)** We are 90% confident that the average square root of the diameters of all American Sycamore trees in the population is between 4.73 and $6.63\sqrt{\text{cm}}$.

• **6.5.6** In this experiment the treatments have been assigned to the flasks, so the entire flasks are the independent units. Though randomly drawn from within each flask, the three aliquots are not independent from one another in the larger context of this study: the aliquots from the same flask will be more similar to each other than aliquots from different flasks. Due to this hierarchical sampling, the SE as described would be an underestimate of the true SE. The SE as computed violates Condition 1(b). (One correct analysis would be to compute the mean density of the three aliquots from each flask, and then use these values as the independent measurements for analysis)

**6.5.7** In this experiment the independently observed units are the entire plants. The oil measurements from the seeds within each plant will likely be more similar than measurements across plants. Due to this hierarchical sampling (and lack of independence in the observations), the SE as described would be an underestimate of the true SE. The SE as computed violates Condition 1(b). (One correct analysis would be to compute the mean oil measurements for the five seeds from each plant, and then use these values as the independent measurements for analysis)

**6.5.8** Nearly 12% [ = 100 × (30/255)%] of the entire population has been sampled. For our SE calculations to be valid, the population must be large relative to the sample size. For our SE formula to be valid (without making special adjustments), we should sample no more than 5% of the population.