Reminder: To receive full credit for your homework, include as part of your solution a brief description of the problem that provides context.

This assignment includes problems related to determination of needed sample size, estimates of population proportions, and estimates of differences in two population means.

1. Exercise 6.21 (page 200).

   Solution: Suppose that $\bar{y} \pm$ SE is interpreted as a confidence interval for a large sample. What is the confidence level?

   For a large sample, the t distribution is essentially the same as the standard normal. The area between -1 and 1 under the standard normal curve is about 68%, so this is the confidence level.

2. Exercise 6.25 (page 203).

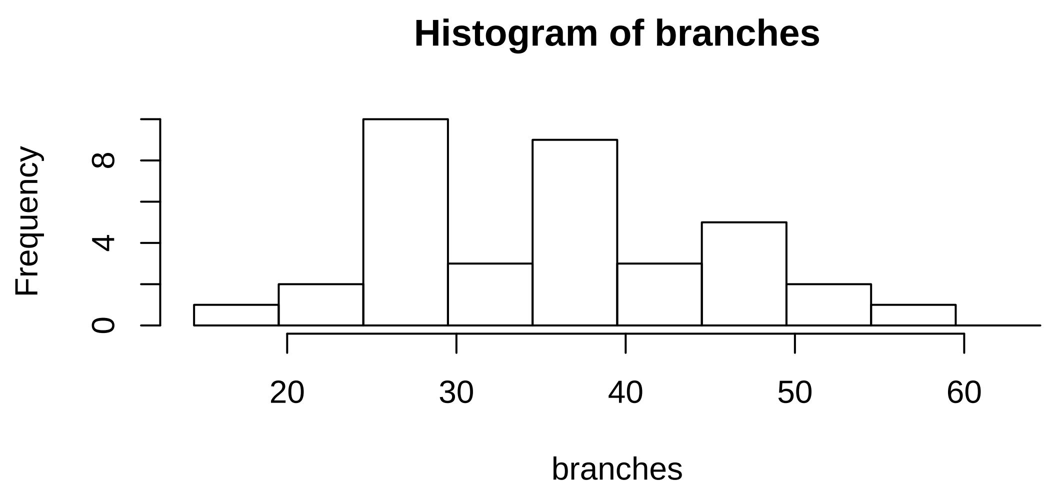
   Solution: Find n to plan a study on weight gains of young turkeys over three weeks so that the SE for the sample mean is less than (a) 20g and (b) 15g when you assume that the population standard deviation of weight gain is about 80g.

   (a) SE $= 80/\sqrt{n} \le 20$ implies that $n \ge (80/20)^2 = 16$.
   (b) SE $= 80/\sqrt{n} \le 15$ implies that $n \ge (80/15)^2 \doteq 28.4$, so let n be at least 29.

3. Exercise 6.27 (page 208).

   Solution: The number of dendritic branch segments from 36 nerve cells (taken from seven adult guinea pigs) has a sample mean of 35.67 and a sample standard deviation of 9.99. A 95% confidence interval is $32.3 < \mu < 39.1$.

   (a) On what grounds might the above analysis be criticized?
   The observations are not independent. There is a hierarchical structure to the sampling. We might expect counts from cells from the same guinea pig to be more alike than counts from different guinea pigs.



   (b) Construct a histogram.
   There is a lumpiness in the histogram which could be caused by clumps of observations close to one another. But, this could also be caused by chance variation in a small sample. To better assess the potential effect of dependence, we would need to know which counts came from which guinea pigs.

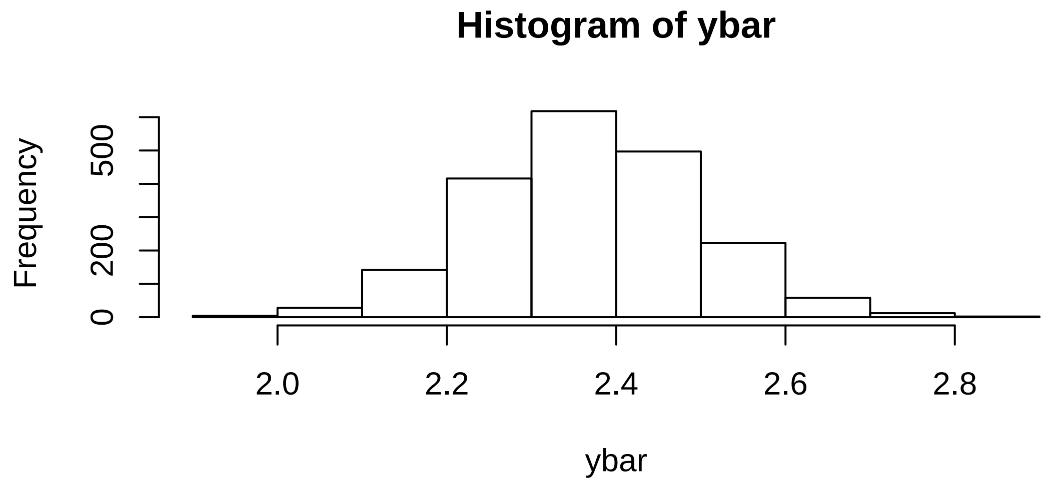
4. Exercise 6.29 (page 209).

   Solution: Data is the number of Ichneumon fly eggs on each of 242 larva of the moth *Ephestia*. The 95% confidence interval for the population mean is $2.12 < \mu < 2.61$ using the t distribution method.

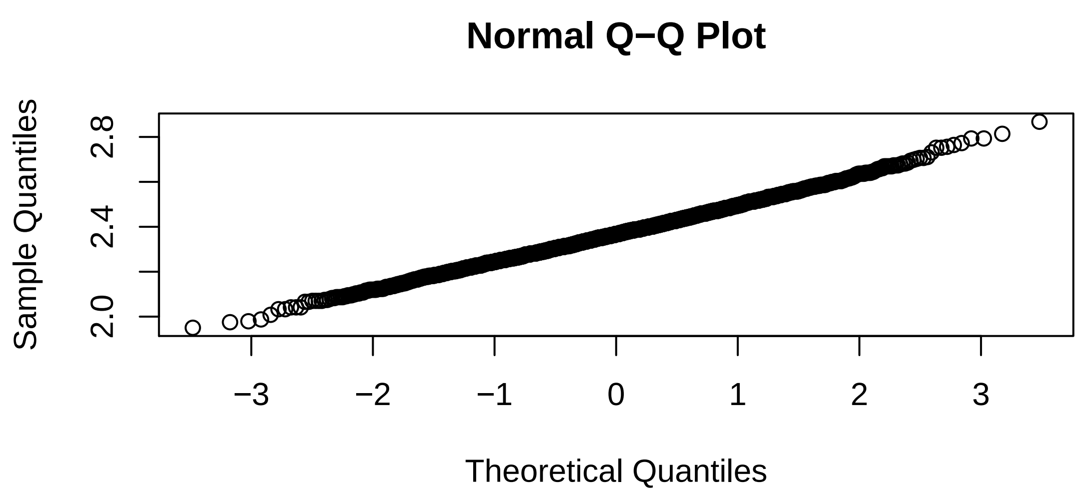
   (a) The distribution shows some right skew with one outlier (15 eggs) while the rest are distributed (with right skew) from 0 to 10. (There is a typo in the book. There should be a count of 2 for 10 eggs.) This strongly implies that the populaiton itself is skewed right.

(b) The central limit theorem implies that the sampling distribution of the sample mean will be approximately normal if the sample size is large enough. The sample size here is 242, which is large enough to overcome even fairly strong nonnormality in the population. The interval is valid.

As a check, here is an estimate of the true sampling distribution which is constructed by taking many samples (with replacement) of size 242 from the sample itself.



A normal probability plot of the sampled sample means also shows that the sampling distribution is approximately normal.



In the estimated sampling distribution, the mean is 2.369 and the standard deviation is 0.127, which are close to the theoretically determined values, 2.368 and $1.950/\sqrt{242} = 0.125$.

5. Exercise 6.34 (page 217).

Solution: Let p be the proportion of mice with white-spotted bellies in a natural population near Ann Arbor Michigan. In a sample of 580 mice, the number with white-spotted bellies is 28. An estimate of p is $\tilde{p} = (28+2)/(580+4) \doteq 0.051$. A 95% confidence interval for p is $\tilde{p} \pm 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$, or 0.051 ± 0.018. We are 95% confident that the proportion of white-bellied mice in the natural population near Ann Arbor, Michigan is between 0.033 and 0.069.

6. Exercises 6.40 and 6.41 (page 217).

   Solution: These exercises deal with determining necessary sample sizes so that the standard error for the sample proportion of people who can taste PTC is less than 0.01. Answers come from the expression for the standard error and solving for n, rounding up if necessary.

$$\text{SE}(\tilde{p}) = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

   If we assume that p will be close to 70%, the necessary size is 2096. If we assume that p is completely unknown, we can be safe by assuming that p will be close to 0.5 (see the graph on page 216). So, the necessary size is 2496.

7. Exercises 7.11 and 7.12 (page 238).

   Solution: A study examined whether relaxation training aided by biofeedback and meditation could help reduce high blood pressure. Subjects were randomly assigned to a biofeedback group or a control group. Data is reduction in systolic blood pressure after eight weeks.

   (a) Construct a 95% confidence interval.

   Here is an R function you may use to calculate the estimated number of degrees of freedom.

```
df2sample <- function(se1,n1,se2,n2) {
  return( (se1^2 + se2^2)^2 / (se1^4/(n1-1) + se2^4/(n2-1) ) )
}
```

   Plugging into this equation, the estimated degrees of freedom is 190 which we round up to 190.

   The confidence interval is 9.8 ± 3.7.

   Interpreted in the context of this setting, we are 95% confident that the difference in the mean reduction in systolic blood pressure due to the biofeedback treatment and the mean reduction due to participation in the study as a control in the population is between 6.1 and 13.5.

   If the populations are not normal, is the confidence interval valid?

   The Central Limit Theorem implies that the sampling distribution of the test statistic is approximately normal even if the population is not if the sample sizes are big enough. These samples sizes are big enough unless we had evidence of quite strong skewness, which we would not expect.

8. Exercise 7.15 (page 238).

   Solution: The data for this exercise is the head width (in mm) in two groups of female crickets. A 95% confidence interval for the difference in population means is

$$0.059 \pm 0.18$$

   We are 95% confident that the difference in mean head widths of all female crickets that successfully mate and those that do not is between -0.121 and 0.239 mm.