4.2

(a) Proportion. Each respondent reports whether or not they worry a great deal about federal spending and the budget deficit, so this is a categorical variable and we use a proportion.

(b) Mean. Each TV news program and newspaper report a number: revenue.

(c) Proportion. Each student reports whether or not they use geolocation services on their smart phones, so this is a categorical variable and we use a proportion.

(d) Proportion. Each user reports whether or not they purchased any Groupon coupons, so this is a categorical variable and we use a proportion.

(e) Mean. Each user reports a number: how many Groupon coupons they purchased over the last year.

4.4

(a) Use the sample mean to estimate the population mean: 171.1. Likewise, use the sample median to estimate the population median: 170.3.

(b) Use the sample standard deviation (9.4) and sample IQR ($177.8 - 163.8 = 14$).

(c) In order to determine if 180 cm or 155 cm are considered unusual observations we need to calculate how many standard deviations away from the mean this observation is, i.e. calculate the Z score.

$$Z = \frac{180 - 171.1}{9.4} = 0.95 \qquad Z = \frac{155 - 171.1}{9.4} = -1.71$$

Neither of these observations is more than two standard deviations away from the mean, so neither would be considered unusual.

(d) No, sample point estimates only estimate the population parameter, and they vary from one sample to another. Therefore we cannot expect to get the same mean and standard deviation with each random sample.

(e) We use the standard error of the mean to measure the variability in means of random samples of same size taken from a population. The variability in the means of random samples is quantified by the standard error. Based on this sample, $SE_{\bar{x}} = \frac{9.4}{\sqrt{507}} = 0.417$.

4.8

(a) We are 90% confident that US residents experience poor mental health 3.40 to 4.24 days per month.

(b) 90% of random samples of size 1,151 will yield a confidence interval that contains the true average number of bad mental health days that US residents experience per month.

(c) To be more sure they capture the actual mean, they require a wider interval, unless they collect more data.

(d) Less data means less precision. The estimate will probably be less accurate with less data, so the interval will be larger.

4.12

(a) False, inference is made on the population parameter, not the point estimate. The point estimate is always in the confidence interval.

(b) False, the strongly right skewed distribution in the sample indicates that the population distribution is most likely right skewed as well, but with a large enough sample size we can assume that the sampling distribution is nearly normal and calculate a confidence interval.

(c) False, the confidence interval is not about a sample mean. The true interpretation of the confidence level would be that 95% of random samples produce confidence intervals that include the true population mean.

(d) True, this is the correct interpretation of the confidence interval.

(e) False. If we only want to be 90% confident rather than 95% confident that we capture the parameter, a smaller interval is required.

(f) False, since in calculation of the standard error we divide the standard deviation by square root of the sample size, in order to cut the standard error to a third of what it is now (and hence the margin of error) we would need to sample $3^2 = 9$ times the number of people in the initial sample.

$$ME_{1/3\text{rd as big}} = z^\star \frac{s}{\sqrt{n}} \rightarrow \frac{1}{3} ME_{original} = z^\star \frac{s}{\sqrt{9n}} = z^\star \frac{s}{3\sqrt{n}} = \frac{1}{3} z^\star \frac{s}{\sqrt{n}} = \frac{1}{3} ME_{original}$$

(g) True, margin of error can be calculated as half the width of the interval:

$$ME = \frac{89.11 - 80.31}{2} = \frac{8.8}{2} = 4.4$$

4.16

(a) $H_0 : \mu = 1100$ (The current average calorie intake is 1100 calories)
$H_A : \mu \neq 1100$ (The current average calorie intake is different than 1100 calories.)

Note: we use a two-sided test because we would be very interested in identifying an increase, not just a decrease, even if a decrease is what we hoped to find. If we used a one-sided test with $\mu < 1100$, we would not be statistically justified in pointing out an increase, even if it was an extremely obvious increase.

(b) $H_0 : \mu = 462$ (The current average Verbal Reasoning score is 462)
$H_A : \mu \neq 462$ (The current average Verbal Reasoning score is not 462)

4.18 First, the hypotheses should be about the population mean (μ), not the sample mean. Second, if she would also have any interest in if the data actually showed a decrease, then she should run a two-sided test. The correct way to set up these hypotheses is shown below:

$$H_0 : \mu = 23.44 \text{ } \textit{years old}$$
$$H_A : \mu \neq 23.44 \text{ } \textit{years old}$$

4.20

(a) This claim does not seem plausible since \$100 is in the interval.

(b) A 90% confidence interval will be narrower than a 95% confidence interval. Even without calculating the interval, we can tell that 100 minutes would not be in it, so we would not reject her claim based on a 90% confidence interval either.

4.38

(a) The distribution of scores on this final exam is likely left skewed since the median is greater than the mean.

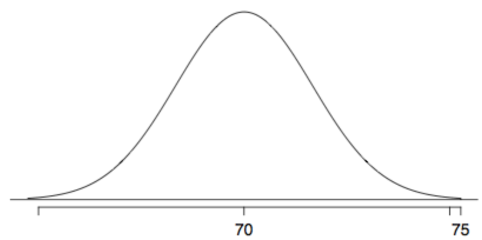
(b) Since the distribution is probably left skewed, the median would be more than the mean, and a majority of observations would be be above the mean.

(c) The shape of the precise distribution is unknown, so we cannot.

(d) Even though the population distribution is not normal, if the conditions for inference are satisfied we may be able to use the Central Limit Theorem to estimate this probability.

1. Independence: Students are sampled randomly and 40 students are likely less than 10% of all students who take this introductory statistics course, so the observations are independent.
2. Sample size: The sample size is large enough since $40 \geq 30$.
3. Skew: The data are skewed but a sample size of 40 might be large enough.

$$
\begin{aligned}
\bar{X} &\sim N\left(\mu_{\bar{x}} = 70, SE_{\bar{x}} = \frac{10}{\sqrt{40}}\right) \\
P(\bar{X} > 75) &= P\left(z > \frac{75 - 70}{\frac{10}{\sqrt{40}}}\right) \\
&= P(z > 3.16) \\
&= 1 - 0.9992 = 0.0008
\end{aligned}
$$



(e) Halving the sample size would increase the standard of error by a factor of $\sqrt{2}$, i.e. the new standard error will be $\sqrt{2}$ times the old standard error:

$$SE_{2n} = \frac{s}{\sqrt{n}} \rightarrow SE_n = \frac{s}{\sqrt{\frac{1}{2}n}} = \frac{s}{\sqrt{\frac{1}{2}}\sqrt{n}} = \sqrt{2}\frac{s}{\sqrt{n}} = \sqrt{2}SE_{2n}$$