11.31 (a) Probability can be thought of as long-term relative frequency — over the next 10000 adults, approximately 10% will be left-handed; over the next 100000, the proportion could be even closer to 10%. But, it cannot be equated with short-term relative frequency.

(b) Outcomes cannot be assumed to be equally-likely. In this case the long-term relative frequency suggests they are not.

(c) The multiplication rule, not the addition rule, is in effect here.

(d) Probabilities must be between 0 and 1.

11.32 (a) $Pr[C] = 0.2$, $Pr[W] = 0.09$, $Pr[C \text{ and } W] = 0.02$

(b) $Pr[C \text{ or } W] = Pr[C] + Pr[W] - Pr[C \text{ and } W] = 0.27$

(c) $$Pr[C \mid W] = \frac{Pr[C \text{ and } W]}{Pr[W]} = \frac{0.02}{0.09} \doteq 0.222$$

(d) $$Pr[W \mid C] = \frac{Pr[C \text{ and } W]}{Pr[C]} = \frac{0.02}{0.2} = 0.1$$

(e) $Pr[C^c] = 1 - Pr[C] = 0.8$

(f) C and W disjoint would mean Warner Bros. has not produced any Hollywood comedy films. If it were true then $Pr[C \text{ and } W]$ would equal zero, and it does <u>not</u>.

(g) C and W independent would mean comedies are of the same relative frequency from other producers as from Warner Bros. Since $Pr[C] \neq Pr[C \mid W]$, they are <u>not</u> <u>independent</u>.

11.34 (a) $181/273$

(b) $232/273$

(c) $Pr[\text{has female} \mid \text{performer}] = 32/181$

(d) $Pr[\text{not performer} \mid \text{no female}] = \dfrac{83}{232}$

(e) $Pr[\text{performer and no female}] = \dfrac{149}{273}$

(f) $Pr[\text{not performer or female}] = \dfrac{122}{273}$

11.36 (a) $11/80$

(b) $1 - \dfrac{20}{80} = \dfrac{3}{4}$

(c) $Pr[\text{red or orange}] = \dfrac{11}{80} + \dfrac{12}{80} = \dfrac{23}{80}$

(d) $\left(\dfrac{20}{80}\right)^2 = 1/16$

(e) $Pr[\text{red}] \cdot Pr[\text{green} \mid \text{red}] = \dfrac{11}{80} \cdot \dfrac{11}{79} = \dfrac{121}{6320} \doteq 0.01915$.

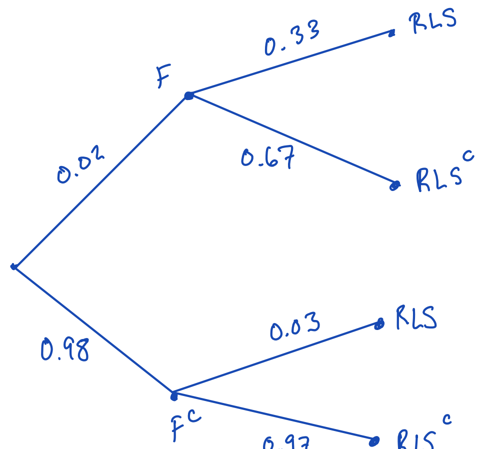
11.39 Answers in the back of the text.

11.57

$Pr(RLS \mid F) = 0.33$

$Pr(RLS \mid F^c) = 0.03$

$Pr(F) = 0.02$



$$Pr(F \mid RLS) = \frac{Pr(RLS \mid F)\, Pr(F)}{Pr(RLS)}$$

$$= \frac{Pr(RLS \mid F)\, Pr(F)}{Pr(RLS \text{ and } F) + Pr(RLS \text{ and } F^c)}$$

$$= \frac{Pr(RLS \mid F)\, Pr(F)}{Pr(RLS \mid F) Pr(F) + Pr(RLS \mid F^c)\, P(F^c)} = \frac{(0.02)(0.33)}{(0.02)(0.33) + (0.03)(0.98)}$$

$$= \boxed{0.183}$$

11.58 $P(D) = 1/38, \quad P(H) = P(D^c) = 37/38$

$P(+|H) = 0.0866, \quad P(-|H) = 0.9134$

$P(+|D) = 0.9989, \quad P(-|D) = 0.0011$

$$P(D|+) = \frac{P(+|D)\,P(D)}{P(+|D)\,P(D) + P(+|H)\,P(H)} = \frac{(0.9989)(1/38)}{(0.9989)(1/38) + (0.0866)(37/38)}$$

$$\doteq \boxed{0.2377}$$

11.83 (a) $0.217 + 0.363 + 0.165 + 0.145 + 0.067 + 0.026 + 0.018 = 1.0$

(b) - (d) See answers in the back of the text.

11.87 See answers in the back of the text.

11.89 See answers in the back of the text.

11.117 See answers in the back of the text.

11.120 Though this seems like it would be sampling w/out replacement, the sample size $(n=10)$, as compared to the U.S. population, is small enough to assume the count X of senior citizens is $\text{Binom}(10, 0.13)$.

$$\Pr[X=3] = \binom{10}{3}(0.13)^3(0.87)^7 = \text{dbinom}(3, 10, 0.13) \doteq 0.09946$$

$$\Pr[X=4] = \text{dbinom}(4, 10, 0.13) \doteq 0.02601.$$

11.127 See answers in the back of the text.

3.12 We expect the sampling distribution to be centered at the value of the population proportion, so we estimate that the population parameter is $p = 0.30$. The standard error is the standard deviation of the distribution of sample proportions. The middle of 95% of the distribution goes from about 0.16 to 0.44, about 0.14 on either side of $p = 0.30$. By the 95% rule, we estimate that $SE \approx 0.14/2 = 0.07$. (Answers may vary slightly.)

3.13 We expect the sampling distribution to be centered at the value of the population mean, so we estimate that the population parameter is $\mu = 85$. The standard error is the standard deviation of the distribution of sample means. The middle of 95% of the distribution goes from about 45 to 125, about 40 on either side of $\mu = 85$. By the 95% rule, we estimate that $SE \approx 40/2 = 20$. (Answers may vary slightly.)

3.23 (a) The value 30 is a population parameter and the notation is $\mu = 30$. The value 27.90 is a sample statistic and the notation is $\overline{x} = 27.90$.

(b) The distribution will be bell-shaped and the center will be at the population mean of 30. The sample mean 27.90 would represent one point on the dotplot.

(c) The dotplot will have 1000 dots and each dot will represent the mean for a sample of 75 co-payments.

3.24 (a) The two distributions centered at the population average are probably unbiased, distributions A and D. The two distributions not centered at the population average ($\mu = 2.61$) are biased, dotplots B and C. The sampling for Distribution B gives an average too high, and has large households over-represented. The sampling for Distribution C gives an average too low and may have been done in an area with many people living alone.

(b) The larger the sample size the lower the variability, so distribution A goes with samples of size 100, and distribution D goes with samples of size 500.

3.42 Using ME to represent the margin of error, an interval estimate for $\mu_1 - \mu_2$ is $\overline{x}_1 - \overline{x}_2 \pm ME = 5 \pm 8$ so an interval estimate of plausible values for the difference in population means is -3 to 13.

3.45 The 95% confidence interval estimate is $\hat{p} \pm 2 \cdot SE = 0.32 \pm 2(0.04) = 0.32 \pm 0.08$, so the interval is 0.24 to 0.40. We are 95% confident that the true value of the population proportion p is between 0.24 and 0.40.

3.46 The 95% confidence interval estimate is $\overline{x} \pm 2 \cdot SE = 55 \pm 2(1.5) = 55 \pm 3$, so the interval is 52 to 58. We are 95% confident that the true value of the population mean μ is between 52 and 58.

3.49 The 95% confidence interval estimate is $(\overline{x}_1 - \overline{x}_2) \pm$ margin of error $= 3.0 \pm 1.2$, so the interval is 1.8 to 4.2. We are 95% confident that the true difference in the population means $\mu_1 - \mu_2$ is between 1.8 and 4.2 (which means we believe that the mean of population 1 is between 1.8 and 4.2 units larger than the mean of population 2.)

3.50 The interval estimate is $(\hat{p}_1 - \hat{p}_2) \pm$ margin of error $= 0.08 \pm 0.03$, so the interval is 0.05 to 0.11. We are 95% confident that the true difference in population proportions $p_1 - p_2$ is between 0.05 and 0.11 (which

(b) The parameter we are estimating is the proportion, p, of *all* young people in the US who have been arrested by the age of 23. Using the information in the sample, we estimate that $p \approx 0.30$.

(c) If the margin of error is 0.01, the interval estimate is 0.30 ± 0.01 which gives 0.29 to 0.31. Plausible values for the proportion p range from 0.29 to 0.31.

(d) Since the plausible values for the true proportion are those between 0.29 and 0.31, it is very unlikely that the actual proportion is less than 0.25.

3.52 (a) The population is all people ages 18 and older living in the US. The sample is the 147,291 people who were actually contacted and asked whether or not they got health insurance from an employer. The parameter of interest is p, the proportion of the entire population of US adults who get health insurance from an employer. The relevant statistic is $\hat{p} = 0.45$, the proportion of people in the sample who get health insurance from an employer.

(b) An interval estimate is found by taking the best estimate ($\hat{p} = 0.45$) and adding and subtracting the margin of error (± 0.01). We are relatively confident that the population proportion is between 0.44 and 0.46, or that the percent of the entire population that receive health insurance from an employer is between 44% and 46%.

3.58 (a) This is a matched pairs design since all participants participated in both treatments (canned soup for five days and fresh soup for five days). There might be a great deal of variability in people's BPA concentrations and a matched pairs experiment reduces that variability.

(b) The population is all people, and we are estimating $\mu_C - \mu_F$, where μ_C is mean urinary BPA concentration after eating canned soup for five days and μ_F is mean urinary BPA concentration after eating fresh soup for five days. Since this is a matched pairs design, we could also use μ_D where μ_D is the mean difference in urinary BPA concentration between the two treatments.

(c) We are 95% confident that BPA concentration is, on average, between 19.6 and 25.5 μg/L higher in people who have eaten canned soup for five days than it is in people who have eaten fresh soup for five days.

(d) A larger sample size increases the accuracy, so we would expect the confidence interval to be narrower.

3.64 (a) Interval is for the mean, not all students.

(b) Interval is for the population mean, not the sample mean.

(c) The interval is not uncertain, only whether or not it captures the population mean.

(d) Interval is trying to capture the mean, not 95% of individual student pulse rates.

(e) Scope of inference could apply to the mean pulse rate for all students at this college, but sample was not taken from all U.S. college students.

(f) The population mean pulse rate is a single fixed value.

(g) Interval is for the population mean, not other sample means.

3.66 (a) Yes.

(b) Yes.

(c) No. A bootstrap sample has the same sample size as the original sample.

(d) No. The value 78 is not in the original sample.

(e) Yes.

(f) Yes

3.76 (a) We find for the 8 values in the table that $\overline{x} = 34.0$ and $s = 14.63$.

(b) We put the 8 values on the 8 slips of paper and mix them up. Draw one and write down the value and put it back. Mix them up, draw another, and do this 8 times. The resulting 8 numbers form a bootstrap sample, and the mean of those 8 numbers form one bootstrap statistic.

(c) We expect that the bootstrap distribution will be bell-shaped and centered at approximately 34.

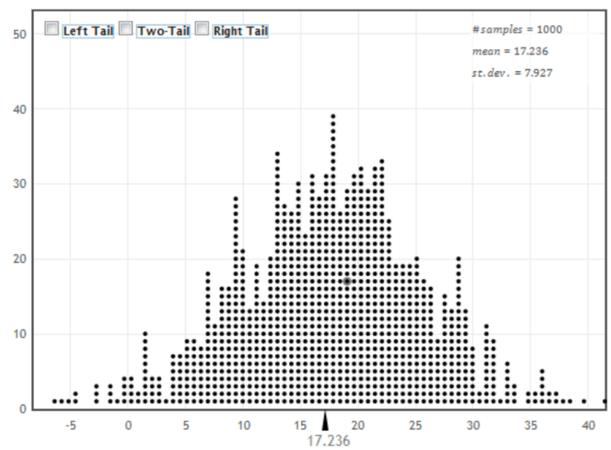
(d) The population parameter of interest is the mean, μ, number of ants on all possible peanut butter sandwich bits set near this ant hill. There are other possible answers for the population; for example, you might decide to limit it to the time of day at which the student conducted the study. The best point estimate is the sample mean $\overline{x} = 34$.

(e) We have

$$\begin{array}{rcl} \overline{x} & \pm & 2 \cdot SE \\ 34.0 & \pm & 2(4.85) \\ 34.0 & \pm & 9.7 \\ 24.3 & \text{to} & 43.7. \end{array}$$

We are 95% confident that the mean number of ants to climb on a bit of peanut butter sandwich left near an ant hill is between 24.3 ants and 43.7 ants.

3.82 Using *StatKey* or other technology, we create a bootstrap distribution to estimate the difference in means $\mu_t - \mu_c$ where μ_t represents the mean immune response for tea drinkers and μ_c represents the mean immune response for coffee drinkers. In the original sample the means are $\overline{x}_t = 34.82$ and $\overline{x}_c = 17.70$, respectively, so the point estimate for the difference is $\overline{x}_t - \overline{x}_c = 34.82 - 17.70 = 17.12$. We see from the bootstrap distribution that the standard error for the differences in bootstrap means is about $SE = 7.9$. This will vary for other sets of bootstrap differences.



For a 95% confidence interval, we have

$$\begin{array}{rcl} (\overline{x}_t - \overline{x}_c) & \pm & 2 \cdot SE \\ (34.82 - 17.70) & \pm & 2(7.9) \\ 17.12 & \pm & 15.8 \\ 1.32 & \text{to} & 32.92. \end{array}$$

We are 95% sure that the mean immune response is between 1.32 and 32.92 units higher in tea drinkers than it is in coffee drinkers.