Chapter 11
Sampling Experiments and the Law of Averages

11.1 Put two cards numbers 0 and 1 into the box. Expect to get 0.5 heads, give or take 0.5 heads or so.

11.3 **a.** You expect to get n(π) = 400(0.50) = 200 heads, give or take $\sqrt{400(0.50)(1-0.50)}$ = 10 heads or so.

b. Approximately 68% of the time, you would get between 200 – 10 = 190 heads and 200 + 10 = 210 heads.

11.5 **a.** Expect n(π) = 80(.40) = 32 Democrats, give or take $\sqrt{80(0.40)(10.40)}$ = 4.38 Democrats or so.

b. Let K be the count of Democrats. Pr(Predict Republican Win) = Pr(K < 40) = (c.c.) Pr(K < 39.5) = (n.a.) Pr(z < (39.5 – 32)/4.38) = 0.9565.

11.7 Here μ_{SUM} = n(μ) = 19(180) = 3420 and σ_{SUM} = $\sqrt{n}$ (σ) = √[19](40) = 174.3560. Pr(Overload) = Pr(SUM > 3800) = Pr(z > (3800 – 3420)/174.3560] = 1 – Φ(2.1794) = 0.0146

11.9 Here μ_{SUM} = n(μ) = 80(190) = 15200 and σ_{SUM} = $\sqrt{n}$ (σ) = $\sqrt{80}$ (40) = 357.7709. Pr(Overload) = Pr(SUM > 16000) = Pr(z > (16000 – 15200)/357.7709] = 1 – Φ(2.2361) = 0.0127

11.11 **a.** Pr(X > 35) = Pr(z > (35 – 30)/5) = Pr(z > 1) = 0.1587

b. Pr(SUM > 140) = Pr(z > (140 – 4(30))/((√[4])(5))) = Pr(z > 2) = 0.0228

c. Pr(All weigh over 35gm) = $(0.1587)^4$ = 0.0006

11.13 **a.** $\mu_{SUM} = n(\mu) = 7(20) = 140$, $\sigma_{SUM} = \sqrt{n}\,(\sigma) = \sqrt{[7]}(5) = 13.2288$

b. Pr(SUM ≤ 150) = (c.c.) Pr(SUM < 150.5) = (n.a.) Pr(z < 0.7559) = 0.7863
Without continuity correction, the answer comes out to 0.7752

c. 99% of the time, the sales are less than 140 + $z_{0.99}$(13.2288) = 170.77. So an inventory of 171 containers would suffice 99% of the time.

11.15 Correct answer is (b), twice as accurate. Assuming no bias in the survey, accuracy is measured by the SE, which is inversely dependent on the square root of the sample size. So a sample of n = 4000 is √[4] =2 times as accurate as a sample of n = 1000.

11.17 $K - \mu_K$ is more likely to be smaller for Hospital A.
$P - \mu_P$ is more likely to be smaller for Hospital B.

11.19 **a.** $\mu_{\bar{X}} = E(\bar{X}) = \mu = 1$

b. $\sigma^2_{\bar{X}} = Var(\bar{X}) = \sigma^2/n = 4/9 = 0.4444$

c. Not shown. This is a normal curve centered at the mean $\mu_{\bar{X}} = 1$ and having inflection points at $\mu_{\bar{X}} \pm \sigma_{\bar{X}} = 1 \pm \frac{2}{3}$.

d. Approximately 95% of the population is within 1.96 σ's of μ. In other words, approximately 95% of the population lies between 1 – (1.96)(2) = –2.92 and 1 + (1.96)(2) = 4.92. This is a question that could have been answered in Chapter 8.

e. There is a 95% chance that the sample mean $\bar{X}$ will be within 1.96 $\sigma_{\bar{X}}$'s of $\mu_{\bar{X}} = 1$. In other words, there is a 95% chance that $\bar{X}$ will be between
1 – (1.96)(2/3) = –0.3067 and 1 + (1.96)(2/3) = 2.3067. Here we use the results of parts a and b to determine the variability of $\bar{X}$.

f. There is a 99% chance that the sample mean $\bar{X}$ will be within 2.576 $\sigma_{\bar{X}}$'s of $\mu_{\bar{X}}$. The number 2.576 is the 99.5th percentile; it is found from the z-table. In other words, there is a 95% chance that $\bar{X}$ will be between 1 – (2.576)(2/3) = –0.7173 and 1 + (2.576)(2/3) = 2.7173. Here we use the results of parts a and b to determine the variability of $\bar{X}$.

11.21 **a.** $\mu = 3.5$. The population box consists of 6 cards numbered 1, 2, 3, 4, 5, and 6. The mean of the box is 3.5.

b. $\bar{x}$

c. The standard error SE($\bar{X}$) = $\sigma/\sqrt{n}$ = 1.7078/10 = 0.1708.
σ can be found using to formulas of Section 4.7.

d. The sampling error.

11.23 **a.** $\mu_{\bar{X}} = \mu = 100$. $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 1.6$.
Pr(89 < $\bar{X}$ < 102) = Pr((89 – 100)/1.6 < Z < ((102 – 100)/1.6) = 0.7887

b. Pr($\bar{X}$ > 102) = 0.1056

11.25 **a.** $\mu_{\bar{X}} = \mu = 43{,}000$, $\sigma_{\bar{X}} = \sigma/\sqrt{36} = 4500/6 = 750$.

b. P-value = Pr($\bar{X}$ < 41,000) = Pr(Z < (41,000 – 43,000)/750) = 0.0038

c. Agree. This cannot be explained by sample variability. P-value is very small. If these brake pads were just as good as the old type, there would only be a 0.38% chance of getting results this poor.

11.27 **a.** Pr(X > 200) = Pr(Z > (200 – 176)/30) = Pr(Z > 0.8667) = 0.2119

b. Pr($\bar{X}$ > 200) = Pr(Z > (200 – 176)/(30/√ [10])) = 0.0057

11.29 **a.** The expected value of the new telephone sampling method can be approximated by the sample mean, which is $\bar{x}$ = 63.56%. Since the true percentage of Democrats is 72%, this method underestimated the true percentage by about 8.44%. Thus, the bias can be estimated by $\bar{x} - \mu$ = 63.56% – 72% = –8.44%.

b. The precision of the method is approximately s = 1.2%.

c. The accuracy of the method is approximately $\sqrt{[(1.2)^2 + (-8.44)^2]}$ = 8.52%.

11.31 $E(W) = E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = \{a_1\mu + a_2\mu + \ldots + a_n\mu\} = (a1 + a_2 + \ldots a_n)\mu = (1)\mu = \mu$

11.35 **a.** The largest number of aces is 100.

b. The smallest number of aces is 0.

c. The expected number of aces is $n(\pi) = 100/6 = 16.6667$. The standard error is $\sqrt{[(100)(1/6)(5/6)]} = 3.7268$

11.37 **a.** $\pi = .12$, $n = 1000$. $\mu_P = \pi = .12 = 12\%$. $\sigma = \sqrt{\dfrac{\pi(1-\pi)}{n}} = 0.0103 = 1.03\%$.

b. Pr(P < .10) = Pr(K < 100) =(c.c.) = Pr(K < 99.5) = 0.0230 = 2.30%.
The answer is 2.58% when done without the continuity correction.

11.39 Let K be the number in the exit poll who said they voted for the Democrat.

a. Pr(correct forecast) = Pr(K = 1) = .60

b. Pr(correct forecast) = Pr(K = 2 or 3) = $\binom{3}{2}(.60)^2(.40)^1 + \binom{3}{3}(.60)^3(.40)^0 = .648$

c. Pr(correct forecast) = Pr(K > 50) =(c.c.) Pr(K > 50.5)
= (n.a.) Pr(Z > (50.5 – (100)(.60)/$\sqrt{[(100)(.60)(.40)]}$ = 0.9738

11.41 $\mu_{SUM} = n(\mu) = 25(30) = 750$. $\sigma_{SUM} = \sqrt{n}\ (\sigma) = 10$.
Pr(SUM < 725) = Pr(Z < (725 – 750)/10) = Pr(Z < 2.5) = 0.0062.
Note that there is no continuity correction because SUM is continuous.

11.43 **a.** E(SUM) = (100)(50) = 5000, SE(SUM) = $\sqrt{[100]}(\sigma) = 250$.

b. Pr(SUM < 4500) = Pr(Z < –2) = 0.0228.

11.45 **a.** $Pr(X > 200) = e^{\{-200/150\}} = 0.2636$

b. $Pr(\overline{X} > 200) = Pr(Z > (200 - \mu)/(\sigma/\sqrt{[50]}) = Pr(Z > (200 - 150)/(150/\sqrt{[50]}) = Pr(Z > 2.357)$ $= 0.0092$

11.47 Solve $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 40/\sqrt{n} < 5$. Solution is $\sqrt{n} > 8$ or $n > 64$.
The sample should be of 65 or more.

11.49 **a.** Solve $SE_P = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.5/\sqrt{n} \leq 0.03$.

Solution is $\sqrt{n} \geq 16.6667$ or $n \geq 277.8$.
Select a sample of 278 or more.

b. Solve $SE_P = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.5/\sqrt{n} \leq 0.02$.

Solution is $\sqrt{n} \geq 25$ or $n \geq 625$.
Select a sample of 625 or more.

c. Solve $SE_P = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.5/\sqrt{n} \leq 0.01$.

Solution is $\sqrt{n} \geq 50$ or $n \geq 2500$.
Select a sample of 2500 or more.

11.51 Exercise 11.50 applies to SE_P; that is, for a fixed sample size, the largest possible SE_P occurs when $\pi = 1 - \pi = 0.5$. As in Exercise 11.49, solve $SE_P = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.5/\sqrt{n} \leq 0.01$.

Solution is $\sqrt{n} \geq 50$ or $n \geq 2500$.
Select a sample of 2500 or more.

11.53 $E(SUM) = n(\mu) = 16(2) = 32$ sec., $SE(SUM) = \sqrt{n}(\sigma) = 4(2) = 8$ sec.

11.55 **a.** X is a Bernoulli random variable with $\mu = \pi = 0.30$ and $\sigma = \sqrt{\pi(1-\pi)} = \sqrt{(0.30)(0.70)}$ = 0.4583

b. Y is a Bernoulli random variable with $\mu = \pi = 0.60$ and $\sigma = \sqrt{\pi(1-\pi)} = \sqrt{(0.60)(0.40)}$ = 0.4899

c. The random variable Z is 1 whenever both X and Y are 1; that is, the response to both questions Q1 and Q2 are Yes. Otherwise Z is 0.
Then Z is a Bernoulli random variable with $\mu = \pi = 0.18$ and
$\sigma = \sqrt{\pi(1-\pi)} = \sqrt{(0.18)(1-0.18)} = 0.3842$

d. All of X, Y and Z are Bernoulli random variables.

e. E(D) = E(X) – E(Y) = 0.30 – 0.60 = –0.30.
D^2 is a Bernoulli random variable which is 1 whenever X and Y differ in sign.
When X and Y are both 0 or both 1, then D^2 is 0.
Thus $E(D^2) = .54$.
Then $\sigma_D^2 = E(D^2) - E^2(D) = 0.54 - (0.30)^2 = 0.45$ and $\sigma_D = 0.6708$.

11.57 Let S denote the sample standard deviation based on a random sample of size n from a population with mean μ and standard deviation σ.
We know that variances of nonconstant random variables are positive and, from Theorem 7.2 , that $Var(S) = E(S^2) - E^2(S)$.
We know from Theorem 11.1 that sample variances are unbiased estimators of population variances; that is $E(S^2) = \sigma^2$.
Putting it all together, $Var(S) = E(S^2) - E^2(S) = \sigma^2 - E^2(S) > 0$.
So $E^2(S) < \sigma^2$ and $E(S) < \sigma$.
Thus, on average, the sample standard deviation will underestimate the population standard deviation.