1. (6.2, p. 116)

   X is a binomial random variable because it possesses the following characteristics:

   a) There are $n = 5102$ total number of fixed independent trials.
   b) The probability of success $p = 0.1$ remains constant between trials (ie, each trial has same probability of success), where we define a success as developing breast cancer.
   c) X counts the number of successes (or women developing breast cancer) out of n total trials.

2. (6.4, p. 122) We have $n = 2$ trials. Defining success as being a carrier of Tay-Sachs disease, we have probability of success $p = \frac{1}{28} \approx 0.0357 \Rightarrow q = 0.9643$, so that the # of successes $X \sim b\left(2, \frac{1}{28}\right)$. Therefore,

   $$\Pr(\text{neither}) = \Pr(X=0) = {}_n\mathrm{C}_x p^x q^{n-x} = {}_2\mathrm{C}_0 p^0 q^2 = \left(\frac{2!}{0!2!}\right)(1)\left(\frac{27}{28}\right)^2 = 0.9298469$$

   $$\Pr(\text{both}) = \Pr(X=2) = {}_n\mathrm{C}_x p^x q^{n-x} = {}_2\mathrm{C}_2 p^2 q^0 = \left(\frac{2!}{2!0!}\right)\left(\frac{1}{28}\right)^2(1) = 0.0012755$$

3. (6.6, p. 123) We have $n = 8$ trials. Defining success as reaching a live person, we have probability of success $p = 0.15 \Rightarrow q = 0.85$, so that the # of successes $X \sim b(8, 0.15)$. Therefore,

   $$\Pr(X=0) = {}_8\mathrm{C}_0 p^0 q^8 = (1)(1)(0.85)^8 = \left(\frac{8!}{0!8!}\right)(1)(0.85)^8 = 0.2724905$$

   $$\Pr(X=1) = {}_8\mathrm{C}_1 p^1 q^7 = \left(\frac{8!}{1!7!}\right)(0.15)^1(0.85)^7 = (8)(0.15)(0.85)^7 = 0.3846925$$

   $$\Pr(X=2) = {}_8\mathrm{C}_2 p^2 q^6 = \left(\frac{8!}{2!6!}\right)(0.15)^2(0.85)^6 = (28)(0.15)^2(0.85)^6 = 0.23760419$$

   $$\begin{aligned} \Longrightarrow \Pr(X \le 2) &= \Pr(X=0) + \Pr(X=1) + \Pr(X=2) \\ &= 0.2725 + 0.3847 + 0.2376 \\ &= 0.8948 \end{aligned}$$

4. (6.8, p. 124)

   (a) From 6.6 we know that $n = 15$ and $p = 0.15$ so
   $E(X) = \mu = np = 8(0.15) = 1.2$
   $V(X) = \sigma^2 = npq = 8(0.15)(0.85) = 1.02$

   (b) Now we have $n = 50$ so
   $E(X) = \mu = np = 50(0.15) = 7.5$
   $V(X) = \sigma^2 = npq = 50(0.15)(0.85) = 6.375$

5. (6.10, p. 128) We have $n = 2$ trials. Defining success as the student being a smoker, we have probability of success $p = 0.2 \Rightarrow q = 0.8$, so that the # of successes $X \sim b(2, 0.2)$. Therefore,

$$\Pr(\text{both}) = \Pr(X = 2) = {}_n\mathrm{C}_x p^x q^{n-x} = {}_2\mathrm{C}_2 p^2 q^0 = \left(\frac{2!}{2!0!}\right)(0.2)^2(1) = 0.04$$

6. (16.10, p. 366) For large-sample confidence intervals about the population proportion, we use the plus-4 method. We have

   - $\hat{p} = \dfrac{170}{2673} = 0.0636$
   - $\tilde{p} = \dfrac{172}{2677} = 0.0643 \Rightarrow \tilde{q} = 0.9357,\ \tilde{n} = 2677$
   - $SE_{\tilde{p}} = \sqrt{\dfrac{\tilde{p}\tilde{q}}{\tilde{n}}} = \sqrt{\dfrac{0.0643(0.9357)}{2677}} = 0.00474$
   - $\alpha = 0.05 \Rightarrow z_{1-\alpha/2} = 1.96$
   - $95\%\ CI:\ \tilde{p} \pm z_{1-\alpha/2} \cdot SE_{\tilde{p}} = 0.0643 \pm 1.96(0.00474) = (0.0550, 0.0736)$

7. (16.18, p. 371)

   - We will test the null hypothesis $H_0 : p = 0.5$ against the two-sided alternative $H_A : p \neq 0.5$.
   - We set our significance level at $\alpha = 0.05$ and calculate the following:
   - $SE_{\hat{p}} = \sqrt{\dfrac{p_0 q_0}{n}} = \sqrt{\dfrac{0.5(0.5)}{262}} = 0.03089$
   - $z_{\text{stat}} = \dfrac{\hat{p} - p_0}{SE_p} = \dfrac{0.5725 - 0.5}{0.03089} = \dfrac{0.0725}{0.03089} = 2.347$
   - Exact p-values from software: $p_{\text{1-sided}} = 0.00946 \Longrightarrow p_{\text{2-sided}} = 0.01892$
   - Our 2-sided p-value is less than α, therefore we reject the null hypothesis that the proportion of males with leukemia is 0.5.
   - We have sufficient evidence to conclude that the proportion of males with leukemia is significantly different (higher) than 50%. That is, there is evidence of a gender preference in favor of males for this disease.

8. (16.20, p. 371) To be conservative, we use $p^* = q^* = 0.5$. Given $m = 0.03$, we use our formula to calculate

$$n = \frac{z_{1-\alpha/2}^2 p^* q^*}{m^2} = \frac{(1.96)^2 0.5(0.5)}{0.03^2} = \frac{0.9604}{0.0009} = 1067.11$$

$$\Rightarrow \ n = 1068$$

We do not need to do any computation for the previous question in order to answer the second part. We know conceptually that n must increase in order to reduce the margin of error, so since the previous problem required $m = 0.06$, it must be that our computed sample size for $m = 0.03$ is greater than that of problem 16.19.

9. C

$\hat{p} = \dfrac{88}{120}$

$$\Rightarrow \ SE_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{\frac{88}{120}\left(\frac{32}{120}\right)}{120}} = \sqrt{0.001629} = 0.04037 \approx 0.04$$

10. E

- $\hat{p} = \dfrac{32}{120}$
- $\tilde{p} = \dfrac{34}{124} \approx 0.2742 \ \Rightarrow \ \tilde{q} = 0.7358, \tilde{n} = 124$
- $SE_{\tilde{p}} = \sqrt{\dfrac{\tilde{p}\tilde{q}}{\tilde{n}}} = \sqrt{\dfrac{\frac{34}{124}\left(\frac{90}{124}\right)}{124}} = \sqrt{0.0016} = 0.04$
- $\alpha = 0.05 \ \Rightarrow \ z_{1-\alpha/2} = 1.96$
- $95\% \ CI: \ \tilde{p} \pm z_{1-\alpha/2} \cdot SE_{\tilde{p}} = 0.2742 \pm 1.645(0.04) = (0.2084, 0.3400)$

11. C

As confidence level increases, width increases.