Lesson 8.3.1

8-43. a. The 200 shots each took represents a sample of shots, while the population is all possible shots for each. Jermaine's population proportion (his overall "skill") could still be higher than Aliah's despite this sample.

b. Jermaine: $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.8)(0.2)}{200}} = 0.028$,

Aliah: $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.84)(0.16)}{200}} = 0.026$

c. $p_{\text{Aliah}} > p_{\text{Jermaine}}$

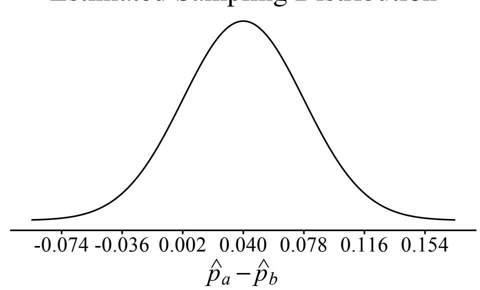
8-44. a. $\hat{p}_{\text{Aliah}} - \hat{p}_{\text{Jermaine}}$

b. $0.84 - 0.8 = 0.04$

c. Standard deviations, and therefore standard errors, follow the rule $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$. Therefore $SE^2_{\hat{p}_a - \hat{p}_j} = SE^2_{\hat{p}_a} + SE^2_{\hat{p}_j} = 0.028^2 + 0.026^2 = 0.00146$, so the standard error of $\hat{p}_{\text{Aliah}} - \hat{p}_{\text{Jermaine}}$ is about 0.038.

8-45. a. See diagram at right.

Estimated Sampling Distribution



b. 95% confidence interval = $0.04 \pm 1.96 \cdot 0.038 = (-0.035, 0.115)$. You can be 95% confident that Aliah is between –3.5% and 11.5% better than Jermaine: in this case a negative number would mean Jermaine is better than Aliah!

c. Since some values in the interval are positive but not every value in this interval is positive, it does not give evidence to accept or reject Aliah's claim; the best you can do is say her claim *could* still be true.

8-46. a. men: 46%, women: 39%

b. $p_m - p_w = 0$ or $p_w - p_m = 0$

c. The original problem stated the sample was random. The total population of men and women is very large compared to the sample sizes, so independent trials is met. And the counts of Republicans and non-Republicans is greater than 10 in both cases. A man who changes his party affiliation would not affect the proportion of women who are Republicans. Conditions met!

d. For men: estimate = 0.46, SE ≈ 0.0227. For women: estimate = 0.39, SE ≈ 0.0214. For the difference, estimate = 0.07, SE ≈ 0.0312.

e. Confidence interval = $0.07 \pm 1.96 \cdot 0.0312 = (0.01, 0.13)$. you can be 95% confident that the proportion of men who are Republicans is 0.01 to 0.13 higher than the proportion of women who are Republicans. Since 0 is not in the confidence interval, you can confidently reject the strategist's claim of independence and state that the variables of gender and Republican yes/no are associated variables.

8-47. Identify: Let p_1 = proportion of emperor penguins that are juveniles and p_2 = proportion of chinstrap penguins that are juveniles. $\hat{p}_1 \approx 0.2955$ and $\hat{p}_2 \approx 0.3309$.
Check conditions: both samples were random. The counts of successes and failures were all much greater than 10 (92 at lowest), satisfying large counts so sampling distribution ≈ normal. Assume the penguin populations are significantly larger than 4940 emperor penguins and 2780 chinstrap penguins, satisfying independent trials. Assume the samples of penguins are not associated.
Calculate: $SE_{\hat{p}_1} = \sqrt{\frac{0.2955(0.7045)}{494}} \approx 0.021$, $SE_{\hat{p}_2} = \sqrt{\frac{0.3309(0.6691)}{278}} \approx 0.028$, so the standard error of the difference is $\sqrt{0.021^2 + 0.028^2} \approx 0.035$. The confidence interval for the difference $p_1 - p_2$ is $-0.0354 \pm 1.96(0.035)$, or $(-0.104, 0.033)$.
Conclude: You can be 95% confident that there are between 10.4% fewer to 3.3% more juvenile emperor penguins than chinstrap penguins.

8-48. a. 6

b. $P(X = 7) = \frac{6}{36} \approx 0.167$

c. There are two ways to have a 5 when rolling 7: (2, 5) and (5, 2) so $P(5 \mid 7) = \frac{2}{6} \approx 0.333$.

8-49. a. $b = r\left(\frac{s_y}{s_x}\right) = 1.38$. Using the means of the explanatory and response variables as a point on the LSRL, $24.73 = a + 1.38(19.73)$. Solving for $a = -2.50$ mm, making the LSRL equation: $\hat{y} = -2.50 + 1.38x$, where x is leaf length, $\hat{y}$ is predicted seed cone length.

b. For every one mm increase in leaf length the predicted value of seed cone length will increase by 1.38 mm.

c. $\hat{y} = -2.50 + 1.38(16.33) = 20.03$ mm

d. The predicted value at $x = 21.6$ mm is $\hat{y} = -2.50 + 1.38(21.6) = 27.31$ mm. The residual is $y_{\text{observed}} - y_{\text{predicted}}$. Residual = 31.1 – 27.31 = 3.79 mm

e. There is no discernable pattern in the residual plot confirming that a liner model is most appropriate.

