Homework 4 – Solution Key

Exercise 5.5

Success="infected" $p = 0.25$, $1 - p = 0.75$

The number of trials is $n = 4$,

a) $\Pr\{j\_successes\} = \Pr\{Y = j\} = {}_nC_j p^j (1-p)^{n-j}$

i) $\Pr\{\hat{p} = 0\} = \Pr\{Y = 0\} = {}_4C_0(0.25^0)(0.75^4) = 1*(1)(0.75^4) = .3164$

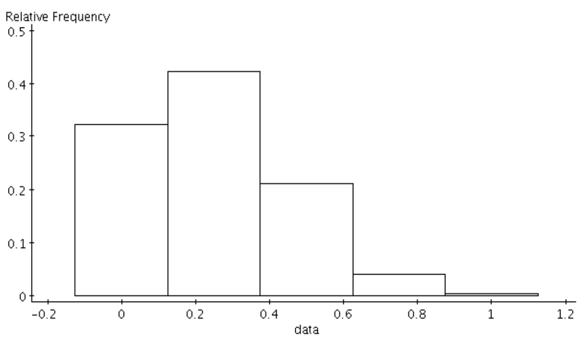
ii) $\Pr\{\hat{p} = 0.25\} = \Pr\{Y = 1\} = {}_4C_1(0.25^1)(0.75^3) = 4*(0.25^1)(0.75^3) = .4219$

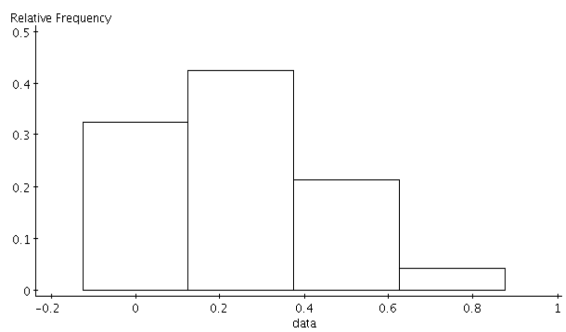
iii) $\Pr\{\hat{p} = 0.50\} = \Pr\{Y = 2\} = {}_4C_2(0.25^2)(0.75^2) = 6*(0.25^2)(0.75^2) = .2109$

iv) $\Pr\{\hat{p} = 0.75\} = \Pr\{Y = 3\} = {}_4C_3(0.25^3)(0.75^1) = 4*(0.25^3)(0.75^1) = .0469$

v) $\Pr\{\hat{p} = 1\} = \Pr\{Y = 4\} = {}_4C_4(0.25^4)(0.75^0) = 1*(0.25^4)(0.75^0) = .0039$

b)





X-axis represents $\hat{p}$. Since $\Pr\{\hat{p} = 1\}$ is almost zero, both graphs are correct as the last bar on some graphs may not be visible.

Exercise 5.15

a) In this population, $\mu = 176$, $\sigma = 30$, normal distribution. $\Pr\{166 < Y < 186\}$

$$z = \frac{y-\mu}{\sigma} = \frac{186-176}{30} = 0.33$$

From Table 3, the area below 0.33 is 0.6293

$$z = \frac{y-\mu}{\sigma} = \frac{166-176}{30} = -0.33$$

From Table 3, the area below -0.33 is 0.3707

What we want is the area between: $0.6293 - 0.3707 = .2586$

Thus, the percentage of 17-year olds that have serum cholesterol values between 166 and 186 is 25.86%.

b) In this sampling distribution of $\bar{Y}$, many groups of $n = 9$

The mean of the sampling distribution of $\bar{Y}$ is $\mu_{\bar{Y}} = \mu = 176$

The standard deviation is $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{9}} = 10$, and the shape of the distribution is normal because the population distribution is normal (from part 3a of Theorem 5.1)

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{186 - 176}{10} = 1.00$$

From Table 3, the area below 1.00 is $.8413$

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{166 - 176}{10} = -1.00$$

From Table 3, the area below -1.00 is $.1587$
What we want is the area between: $.8413 - .1587 = .6826$

Thus, percentage of groups that would have mean (group) cholesterol values between 166 and 186 is 68.26%.

c) The probability of an event can be interpreted as the long-run relative frequency of occurrence of the event (Chapter 3.3). The question in part c) is just a rewording of the question in part b).

Thus, $\Pr\{166 \le \bar{Y} \le 186\} = .6826$

Exercise 5.35

a) Intuitively, I would think that the smaller hospital has a greater likelihood of recording more days of having at least 60% boys. (There is no officially "correct" answer for part a).

b) $p = 0.5$
For $n = 45$, $mean = np = 22.5$, $SD = \sqrt{np(1-p)} = \sqrt{(45)(.5)(.5)} = 3.354$

Question asks for $\Pr\{\hat{p} \ge 0.60\} = \Pr\{Y \ge 27\}$ since 27 out of 45 is 0.60.

$$z = \frac{27 - 22.5}{3.354} = 1.34$$

From Table 3, the area below 1.34 is 0.9099
What we want is the area above: $1 - 0.9099 = .0901$

For $n = 15$, $mean = np = 7.5$, $SD = \sqrt{np(1-p)} = \sqrt{(15)(.5)(.5)} = 1.936$

Question asks for $\Pr\{\hat{p} \ge 0.60\} = \Pr\{Y \ge 9\}$ since 9 out of 15 is 0.60.

$$z = \frac{9 - 7.5}{1.936} = .77$$

From Table 3, the area below 0.77 is 0.7794
What we want is the area above: $1-0.7794=.2206$
Thus, in the larger hospital, 9.01% of days have 60% or more boys. In the smaller hospital, 22.06% of days have 60% or more boys. The smaller hospital records more days.

Exercise 5.50

For $\Pr\{1175 \le Y \le 1225\}$ where Y represent a 10-second count
In this population, $\mu = 1200$, $\sigma = 35$, normal distribution

$$z = \frac{y-\mu}{\sigma} = \frac{1225-1200}{35} = 0.71$$

From Table 3, the area below 0.71 is 0.7611

$$z = \frac{y-\mu}{\sigma} = \frac{1175-1200}{35} = -0.71$$

From Table 3, the area below -0.71 is 0.2389

What we want is the area between: $0.7611 - .2389 = .5222$

Thus, the probability that a 10-second count is between 1175 and 1225 is .5222.

For $\Pr\{1175 \le \bar{Y} \le 1225\}$

In this sampling distribution of $\bar{Y}$, where $\bar{Y}$ represent the mean of six 10-second counts, $n = 6$

The mean of the sampling distribution of $\bar{Y}$ is $\mu_{\bar{Y}} = \mu = 1200$

The standard deviation is $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{6}} = 14.29$, and the shape of the distribution is normal because the population distribution is normal (from part 3a of Theorem 5.1)

$$z = \frac{\bar{y}-\mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{1225-1200}{14.29} = 1.75$$

From Table 3, the area below -1.75 is 0.9599

$$z = \frac{\bar{y}-\mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{1175-1200}{14.29} = -1.75$$

From Table 3, the area below -1.75 is 0.0401
What we want is the area between: $0.9599 - 0.0401 = .9198$
Thus, the probability that the mean of six 10-second counts is between 1175 and 1225 is .9198

Since the .9198>.5222, this shows that the mean of 6 counts is more precise, in that it is more likely to be near the correct value (1200) than a single count.