HW 6

Kaahan Motwani

Question 1

A From this data, the best estimate of the average weight of a mouse (answer in grams) will be the sample mean.

$$\frac{21 + 23 + 27 + 19 + 17 + 18 + 20 + 15 + 17 + 22}{10} = \mathbf{19.9}$$

B The formula for standard error, which is also can be understood as the standard deviation of the estimate of the mean, is

$$stderror(X) = \frac{stdunbiased(X)}{\sqrt{N}}$$

First, let us find the std unbiased estimate

$$\sqrt{\frac{\sum (x_i - mean(x))^2}{N-1}} = \sqrt{\frac{(21 - 19.9)^2 + (23 - 19.9)^2 + ... + (22 - 19.9)^2}{N-1}} = 3.51$$

Now, plugging this into the formula for standard error, we get:

$$\frac{3.51}{\sqrt{10}} = \mathbf{1.11}$$

C Using the same formula from part (b), we must find out the number of mice, N, to reduce standard error to 0.1. Using the formula and manipulating it to get N, we get:

$$stderror(X) = \frac{stdunbiased(X)}{\sqrt{N}}$$

$$\sqrt{N} = \frac{stdunbiased(X)}{stderror(X)}$$

$$N = (\frac{stdunbiased(X)}{stderror(X)})^2$$

$$N = (\frac{3.51}{0.1})^2 = \mathbf{1233}$$

Therefore, we must have at least 1233 mice

Question 2

A In order to find standard error, we must first find the stdunbiased(x), according to the same formula we used in question 1:

$$stderror(X) = \frac{stdunbiased(X)}{\sqrt{N}}$$

$$stdunbiased(X) = \sqrt{\frac{1}{10-1} * (3(1-.3)^2 + 7(0-.3)^2} = 0.48$$

Plugging this into the formula to get the standard error:

$$stderror(X) = \frac{0.48}{\sqrt{10}} = \mathbf{.15}$$

B

$$stderror(X) = \frac{0.48}{\sqrt{N}} = \mathbf{94}$$

Therefore, we would need 94 cards to get standard error of 0.05

Question 3

A a) Using the formula for standard error when we are given the standard deviation:

$$\frac{popsd(X)}{\sqrt{N}} = \frac{75}{\sqrt{40}} = 11.86$$

Since 68% confidence interval is within one standard deviation of the mean, we can estimate out confidence interval to be: [340 - 11.86, 340 + 11.86] = **[328.14, 351.86]**
b) Using the formula for standard error when we are given the standard deviation:

$$\frac{popsd(X)}{\sqrt{N}} = \frac{75}{\sqrt{40}} = 11.86$$

Since 68% confidence interval is within three standard deviations of the mean, we can estimate out confidence interval to be: [340 - 3 * 11.86, 340 + 3 * 11.86] = **[304.43, 375.58]**

B For this part, I would use a t-distribution, with degrees of freedom = N - 1 = 10 - 1 = 9 degrees of freedom

Question 4

A In order to get a 99% confidence interval for the probability of a female birth, we must find the standard error. In order to find the standard error, we must find the std unbiased estimate once again.

$$stdunbiased(X) = \sqrt{\frac{1}{2009-1} * (1026(1 - \frac{1026}{2009})^2 + 983(0 - \frac{1026}{2009})^2} = 0.50$$

Now, plugging this into the formula for standard error, we get:

$$\frac{.50}{\sqrt{2009}} = 0.0111$$

Thus, our confidence interval is

$$[\frac{1026}{2009} - 3 * 0.0111, \frac{1026}{2009} + 3 * 0.0111]$$

= **[.477, .544]**

B In order to get a 99% confidence interval for the probability of a female birth, we must find the standard error. In order to find the standard error, we must find the std unbiased estimate once again.

$$stdunbiased(X) = \sqrt{\frac{1}{2009-1} * (983(1 - \frac{983}{2009})^2 + 1026(0 - \frac{983}{2009})^2} = 0.50$$

Now, plugging this into the formula for standard error, we get:

$$\frac{.50}{\sqrt{2009}} = 0.0112$$

Thus, our confidence interval is

$$[\frac{983}{2009} - 3 * 0.0112, \frac{983}{2009} + 3 * 0.0012]$$

= **[0.456, 0.523]**

C These intervals do indeed overlap, suggesting that these means are not very different, statistically speaking, and they may be related.

Question 5

A For parts a, b, and c of this problem, I will be using the formulas:

$$popmean(x) = \frac{1}{N} \sum x_i * P(x_i)$$

$$stdunbiased(x) = \sqrt{\frac{\sum (x_i - mean(x))^2}{N-1}}$$

$$popmean(x) = \frac{stdunbiased(x)}{\sqrt{N}}$$

After performing calculations on the values in the dataset given in the textbook, I got the following values using the formulas above:

$$popmean(x) = 2.982$$

$$stdunbiased(x) = 0.3975$$

$$popmean(x) = 0.0517$$

Thus, the 99% confidence interval is [**2.827, 3.138**]

B After performing calculations on the values in the dataset given in the textbook, I got the following values using the formulas above:

$$popmean(x) = 0.7815$$

$$stdunbiased(x) = 0.2935$$

$$popmean(x) = 0.0424$$

Thus, the 99% confidence interval is [**0.654, 0.909**]

C These intervals do not overlap, suggesting that the difference between the means of the values from the two different regions is actually statistically significant.