Week 7 Reading Questions

Mercy Melo (no collaborators)

Question 1: The population mean has no effect on the width of the confidence intervals. The population mean will only impact the location of the normal curve on the x-axis of the population distribution curve. For example, a population of peregrine falcons may have an average mass of 800 grams and a population of kestrels may have an average mass of 100 grams. This large difference in means would have no effect on each of the population's distribution curves nor the confidence intervals derived from said curves.

Question 2: The population standard deviation has a great effect on the width of the confidence intervals as standard deviation and standard error are factored into the calculation for confidence intervals. As standard deviation around the population parameter of interest increases, the confidence intervals will become wider in response. Alternatively, a small standard deviation in the population's parameter of interest will result in more narrow confidence intervals that are closer to the parameter's true value.

For example, I have two population of kestrels, A and B, each with 5 birds. Individuals in population A have the following masses: 80, 92, 135, 147, 73. Individuals in population B have the following masses: 95, 103, 107, 92, 101. Population A would have a much larger standard deviation as the spread between the points is much greater than what is seen in population B (where all of the points are relatively much closer together). In this case, we would have a wider confidence interval for the mean mass estimate of population A than we would for the same parameter in population B.

Question 3: Theoretically speaking, populations are infinite and therefore population size would have almost no effect on the width of confidence intervals as it cannot be altered.

If this question is referring to populations under the biological definition (a collection of organisms of the same species living within a defined geographical area), then population size may impact the size of the confidence intervals when compared to the sample size taken from that population. If the population size and sample size are relatively similar, the confidence intervals around the estimate of a parameter would be quite narrow. As sample size deviates further and further from the true population size, the confidence intervals would become wider as we increase the incidence rate of obtaining a non-representative sample.

Question 4: Sample size can have a great effect on the width of confidence intervals around a parameter estimate. As we increase the number of samples, we also increase the probability that our sample is representative of the true population, which will make our confidence intervals narrower. This also holds true when we think about how sample size can impact the sample distribution of repeated measures. As we increase sample size, it effectively normalizes our sample distribution and theoretically provides a more accurate estimate of the true value of the parameter.

Small sample sizes would therefore have the opposite effect. Collecting only a few samples increases our chances of having a non-representative sample. For example, if we look back to population A of our kestrels in question 2, we can see that there is a mean mass value of around 100 grams. If we have a sample size of 1 and happen to collect data on the bird with a mass of 73, we can see that this can lead us to drawing a false conclusion (and overall a bad estimate) about the mean mass value for the population. In this way, the confidence interval width for small sample sizes would be much higher than confidence intervals derived from large samplings.

Question 5: Let's use an example regarding the height of hockey players in the NHL. We can measure the height of every player on the Pittsburgh Penguins and find that they have an average height of 6'3". Then we measure the height of every player on the Toronto Maple Leafs and find that they have an average height of 6'1". We can see during our measurements that some players are much taller than average and some players are much shorter than average, but overall our mean for the 2 teams sits around 6'2". We can plot the heights of all of these players and find that they follow a "normal distribution", which means that most players are around 6'2", fewer players are 5'8" and even fewer players are 5'5" (and the same goes for tall and extra-tall players). Despite having these outliers, we can calculate a range of values (called a "confidence interval") that describes the average heights of these players. This range has a high-value and low-value that are equally-distant from the average of 6'2", but that encompasses most of the player's heights (let's say we calculated our confidence interval as 5'10" to 6'6").

Now let's say we measure the height of all players on 100 different teams (we'll have to go outside the NHL, but let's assume the AHL and international pro-league team players are similar heights to the NHL). Using our confidence intervals that we calculated using the Pittsburgh Penguins and Toronto ML's, we can predict that 95 out of these 100 teams will have an average height value that falls within our confidence interval- their team average height would be between 5'10" and 6'6". With that being said, it wouldn't be cause for concern if 5 of the teams had an average height outside of this range (maybe the Vegas Golden Knights are super tall and have an average height of 6'8").