**Quiz 3, Version 1**

**Figure 1: This graph from Morris et al. (2020), published in the journal Proceedings of the National Academy of Sciences, illustrates how plant growth rates are affected by various drivers of plant growth. Use it to answer Questions 1 and 2.**



Q1. In Figure 1, the points show the mean growth rate from a sample, and the vertical lines show an interval that represents plus or minus (+/-) one standard error (SE) from the mean estimate. Which of these statements about this plot is **true**?

1. Disturbance has the largest amount of natural variability in the sample, because the vertical lines are the longest (the SE is the largest).
2. The vertical lines are 95% confidence intervals.
3. The vertical lines for Disturbance are the longest, creating the widest interval, so the true population mean is less likely to be inside that interval than in the intervals for any other driver.
4. The vertical lines for Disturbance are the longest, creating the widest interval, so the true population mean is more likely to be inside that interval than in the intervals for any other driver.

Q2. If they had used R for their analysis, which of these formulas would Morris et al. (2020) have used to calculate the standard error for the Disturbance sample data, if the sample size (n) was 100?

1. SE<- sd(DisturbanceSampleData)/sqrt(100)
2. SE<- sd(DisturbanceSampleData)
3. SE<-mean(DisturbanceSampleData)-2\*sd(DisturbanceSampleData)/sqrt(100)
4. SE<-mean(DisturbanceSampleData)+2\*sd(DisturbanceSampleData)

**Figure 2. The following figure from the NY Times shows the percent of surveyed United States students who use e-cigarettes and other tobacco products. Use it to answer Questions 3 and 4.**

NYTimes



Q3. In 2018, 28% of surveyed high school students reported that they used any tobacco product (Fig. 2). Which of these statements about the accuracy of this estimate is **incorrect**?

1. The percent of surveyed high school students who used tobacco products in 2018 (28%) is exactly the same as the percent of **all** United States high school students who used tobacco products in 2018, because there is no confidence interval on the graph.
2. The percent of surveyed high school students who used tobacco products in 2018 (28%) would be more likely to be close to the percent of **all** United States high school students who used tobacco products in 2018 if the number of surveyed students was very large (e.g., 10,000 students) than if the sample size was smaller (e.g., 100 students).
3. The percent of surveyed high school students who used tobacco products in 2018 (28%) would be more likely to be close to the percent of **all** United States high school students who used tobacco products in 2018 if the survey was of a random sample of all United States high school students, rather than a convenient or biased sample of United States high school students.
4. The CDC could have accurately estimated the percent of **all** United States high school students who used tobacco products in 2018 by doing many surveys of randomly sampled high school students with n>30 for each sample, and then taking the average of the sampling distribution (the average of all percent estimates from all surveys).

Q4. If the CDC was very interested in the true percent of **all** United States high school students who used tobacco products in 2018 and wanted to construct a confidence interval that was most likely to include the true percentage, which of these options should they pick?

1. Construct a 99.9% confidence interval
2. Construct a 0.1% confidence interval
3. Construct a 95% confidence interval
4. Construct a 99% confidence interval

**Fig. 3. Let’s say that this figure from the New York Times represents a census of the whole population of first-time mothers in the United States in 2016. Use this figure to answer Questions 5 and 6.**



Q5. If we were to sample this population of first-time mothers in the United States in 2016, which of these **would** you expect from the sample?

1. The sample mean would be around 26 years old.
2. The sample mean would be around 34 years old.
3. The sample standard deviation would be around 40 years.
4. The sample distribution would be heavily skewed to the right.

Q6. If we were to make a sampling distribution of this population of first-time mothers in the United States in 2016, which of these would you **NOT** expect from the sampling distribution?

1. The sampling distribution would be wider than the population distribution.
2. The sampling distribution would be approximately normally distributed, if your sample size was large enough.
3. The sampling distribution would be narrower than the sample distribution.
4. The sampling distribution would have roughly the same mean as the population distribution.

**Quiz 3, Version 2**

**Dataset:**

*Imagine that you randomly sample fish from two different lakes, so that you have two samples from two different populations. You calculate some sample statistics and here is what you find:*

*Sample 1 from Population 1:*

*Sample size(n)=100 fish*

*Sample mean (x̄)=25 cm*

*Sample standard deviation (s) = 10 cm*

*Sample 2 from Population 2:*

*Sample size(n)=100 fish*

*Sample mean (x̄)=25 cm*

*Sample standard deviation (s) = 5 cm*

Q1. Which of these is the 95% confidence interval for the mean population fish length for Population 1?

1. 25∓(2\*10/sqrt(100)) = 23 cm - 27 cm
2. 25∓(2\*10) = 5 cm - 45 cm
3. 25∓sqrt(100) = 15 cm - 35 cm
4. 25∓(2/sqrt(100)) = 24.8 cm - 25.2 cm

Q2. Which of these statements is the **best** interpretation of the confidence interval you calculated in Question 1?

1. If we took 20 samples from Population 1 and calculated 95% confidence intervals based on each sample, we would expect 19/20 of those confidence intervals to contain the true mean population fish length for Population 1.
2. We can be 95% certain that the true mean population fish length for Population 1 is within the 95% confidence interval.
3. We can be 95% confident that if we took another sample from Population 1 (Sample 2), the 95% confidence interval for that sample would contain the true mean population fish length for Population 1.
4. We can be 95% sure that the true mean population fish length for Population 1 is approximately 25 cm.

Q3. If you calculate a 95% confidence interval for the mean population fish length for Population 2, which of these statements will be true?

1. The 95% confidence interval for the mean population fish length for Population 2 is narrower (less wide) than the 95% confidence interval for the mean population fish length for Population 1.
2. The 95% confidence interval for the mean population fish length for Population 2 is wider than the 95% confidence interval for the mean population fish length for Population 1.
3. The 95% confidence interval for the mean population fish length for Population 2 is exactly the same as the 95% confidence interval for the mean population fish length for Population 1.
4. The 95% confidence interval for the mean population fish length for Population 2 is 24.8 cm - 25.2 cm.

Q4. Which of these interpretations regarding the true population mean for Population 1 is **incorrect**?

1. According to the Law of Large Numbers, our sample size is too big to accurately estimate the true population mean.
2. According to the Law of Large Numbers, since our random sample size is “large enough” (n>30), our sample mean (25 cm) is likely to be close to the true population mean, because sampling error decreases with increasing sample size.
3. According to the Central Limit Theorem, since our random sample size is “large enough” (n>30), a **sampling** distribution based on repeated samples of Population 1 would be normally distributed, and the sampling distribution mean would be close to the true population mean.
4. We do not know the true population mean, but we can use our random sample and statistics to make inferences about which values for the true population mean are likely or unlikely.

Q5. If you calculated a 99% confidence interval for Population 1 based on Sample 1, how would it compare to the 95% confidence interval that you calculated in Question 1?

1. The 99% confidence interval would be wider than the 95% confidence interval and more likely to include the true population mean.
2. The 99% confidence interval would be narrower (less wide) than the 95% confidence interval and less likely to include the true population mean.
3. The 99% confidence interval would be the same as the 95% confidence interval and equally likely to include the true population mean.
4. The 99% confidence interval would be 24.8 cm - 25.2 cm.

Q6. Match the R functions to their uses.

*Functions:*

ggplot()

rnorm()

sd()

geom\_vline()

*Uses:*

Make a plot in R.

Randomly draw values from a normal distribution.

Calculate a standard deviation.

Add a vertical line to a plot.