**Homework: Standard Error and 95% Confidence Intervals**

1. A student measures the running speed of 30 classmates. The mean running speed equals 7.12 mph, with a standard deviation equal to 0.6 mph. How would the student report the mean and standard error for these data?

ANSWER: standard error (SE) is calculated as the standard deviation (sd) divided by the square-root of the sample size. Therefore, SE = 0.6 / sqrt(30) = 0.109. The student could report the mean and SE as mean ± SE: 7.1 ± 0.11

2. A neuroscience researcher requires a measure of the standard deviation for the rate of neuron firing in mice. However, the researcher cannot find the desired standard deviation in published research papers. Instead, the researcher finds a published article that reports the *standard error* and sample size for a measurement of firing rate. Does this information help the researcher? If so, how?

ANSWER: Given that SE = sd / sqrt(sample size), the researcher could use the values of SE and sample size to calculate sd: sd = SE \* sqrt(sample size). (This kind of information can often be obtained from figures that display SE.)

3. A researcher measured 15 lab mice at birth; the average mass equaled 1.27g with a standard deviation of 0.08g. Calculate a 95% confidence interval (CI) for the mass of mice at birth in this sample.

ANSWER: Given the relatively small sample size, we would calculate 95% CI as: 95% CI = t \* SE, where t is the appropriate value from a t-distribution. We can start by calculating SE: SE = 0.08 / sqrt(15) = 0.0206. We can obtain the t-value from a t-distribution: as our sample size equals 15, the degrees of freedom (df) equals 14 (i.e., n-1). Therefore, we obtain the t-value for df=14, a type-1 error rate of 0.05 and a 2-tailed test. By reading a t-table from the internet, I found that this value equals 2.145. Therefore, the mean and 95% CI’s equal:

1.27g ± 2.145\*0.0206 = 1.27 ± 0.044. Alternatively, the 95% CI’s range from approximately 1.226 to 1.314.

4. The following three figures each display a mean and 95% CI’s for one Control and one Manipulated treatment. Based on the presented means and 95% CI’s, determine whether the Control and Manipulated treatments differ significantly (or whether it is impossible to determine from the information presented). *(This question is modified from Whitlock & Schluter’s “The analysis of biological data”.)*

a) The 95% confidence intervals do not overlap between the treatments.



ANSWER: In (a), the Control and Manipulated treatments will differ significantly. To understand why, remember that we can (very loosely) think of 95% as a range of plausible values for a measurement (a mean value, in this case). If the 95% CI’s do not overlap between the two treatments then, very roughly speaking, we can say that the range of plausible values for each treatment do not overlap, suggesting that the means of each treatment are different. Or, in other words, it is not plausible that the means of the two treatments would be the same.

b) The 95% confidence interval for one treatment includes the mean of the other treatment.



ANSWER: In (b), the Control and Manipulated treatments will not differ significantly (i.e., a statistically non-significant result). Recall our reasoning for (a), above: very roughly speaking, the 95% CI’s indicate a range of plausible values for the mean of each treatment. In (b) notice that the 95% CI’s of the Manipulated treatment include the mean of the Control treatment. This implies that the mean of the Control treatment is a ‘plausible’ value for the mean of the Manipulated treatment; i.e., it is plausible that the means do not differ between the treatments.

c) The 95% confidence intervals overlap between the two treatments, but the 95% confidence interval for one treatment does not overlap the mean of the other.



ANSWER: In (c) it is difficult to know, based on the means and 95% CI’s, whether the Control and Manipulated groups would differ significantly. In this case we would need to perform a statistical test to determine whether the treatments differ significantly.

5) The figure, below, comes from the paper in a major journal.



This figure conveys the number of mutations that are found at nucleotide sites along the X-chromosome of a fish species; the left end of the figure (Start position 0 Mega-bases (Mb)) represents one end of the X-chromosome and the right end (Start position n ~25 Mb) represents the opposite end; hence the X chromosome is approximately 25Mb long.

The horizontal grey bar centered at zero on the y-axis conveys the 95% confidence interval for the number of mutations found, on average, on autosomes. You will notice two squiggly lines (coloured black and orange) that often occur within the grey region (i.e., the 95% confidence interval for autosomes) and sometimes do not. Roughly speaking, these squiggly black and orange lines convey the numbers of mutations for positions along the X-chromosome for two different populations (one colour for each population).

1. The authors aimed to use this figure to compare the number of mutations along the X-chromosome between the two populations (i.e. between the black and orange squiggly lines). What is missing from this plot that would allow this comparison?

ANSWER: If the authors aim to illustrate a comparison between populations (i.e., between the black and orange squiggly lines) then the figure should include 95% CI’s for each of these coloured lines. Note that, if the 95% CI’s for the black and orange lines were as large as for the autosomes (i.e., the grey horizontal bar) then we’d see extensive overlap between the 95% CI’s for the black and orange lines, which would be informative (see answers to question 4).

1. By including the horizontal grey region (representing the 95% confidence intervals for mutation numbers on autosomes) the authors invite the readers to compare the number of mutations on the X-chromosome to what’s found on the autosomes. Is this comparison valid, given the available information?

ANSWER: There are at least two reasons why this comparison would be inappropriate. First, as noted above, the black and orange lines do not include 95% CI’s, so we lack appropriate information to compare the means represented by the black and orange lines to the autosomes. Second, even if the black and orange lines represented the *true* values of mutation number on the X-chromosome (i.e., we estimated mutation numbers on X-chromosomes perfectly for these two populations) a comparison between the squiggly lines and autosomes would remain problematic. Recall that, *by definition*, a 95% confidence interval will include the true value of a population 95% of the time (i.e., 95% of experiments), meaning that 95% CI’s will *not include* the true value 5% of the time. Now, we could consider different locations along the X-chromosome as different ‘experiments’. In this case, we expect the squiggly lines to fall outside the 95% CI’s for the autosomes 5% of the time, even if there is no difference between the X-chromosomes and autosomes. This fact complicates matters.