Statistics and Hypothesis Testing
Due Date: Thursday 9/14 @ 11:59 PM

1. Define a p-value.

   The p-value is the probability of obtaining the same or more severe data, under the assumption that the null-hypothesis is correct

2. What is the p-value in the rigged coin example (from the stats slides), and what does this p-value indicate?

   p = 0.0059, This indicates that the situation with 16 heads would only happen by random chance 0.59% of the time

3. Suppose we are examining the mutation status of the gene *MSH2* between two groups: old and young colorectal cancer patients.
   a. Define the null hypothesis, H_0.

      Mutation status is the same between younger and older colorectal cancer patients

   b. Define the alternative hypothesis, H_a.

      Mutation status is different between younger and older CRC patients.

   c. Suppose we find that *MSH2* is mutated more frequently in younger patients than in old patients, with a p-value of 0.07. From this, what can we conclude? What if the p-value was 0.03? Be specific.

      There is a 7% chance that this link occurred due to random chance. With a p-value of $p < 0.05$, this is not significant. // There is a 3% chance that the link occurred due to random chance. With the same p-value, this is significant

   d. In a typical biological analysis, we look at many thousands of genes. Suppose we have 20,000 genes in our data set. Is there an issue if we set our p-value threshold at 0.05 as usual and examine each gene one at a time? (Hint: think about false positives.)

      5% of 20,000 genes is 1000. It is possible that not all 1000 of these genes is significant

4. Two students, Alex and Jamie, are looking at the incidence of seizures in a set of glioblastoma patients in a clinical study. They find that seizures concurrently occur with usage of a certain hypothetical drug, Placebomab, with a p-value of 0.04. <u>Determine whether the following statements are true or false. If false, explain why.</u>

   - Alex: “Because the p-value of 0.04 is less than 0.05, the drug Placebomab is most likely a cause of seizures in this group of patients.” (TRUE/FALSE)

     p-values cannot determine if something is the cause of another thing,

   - Jamie: “Even though the p-value of 0.04 is less than 0.05, we can only say that Placebomab and seizures are significantly correlated in glioblastoma patients.” (TRUE/FALSE)