Math 154 – Fall 2017
Jo Hardin
warm-up # 4 (Hyp Test)

Name: ______________________________

Consider a two sample test of means (e.g., determining that treatment A extends life - on average - more than treatment B). Answer the questions below with respect to the logic of hypothesis testing.

1. Assume the null hypothesis is true.
   - *What does it mean for the null hypothesis to be true? Explain in words.*
2. Generate a sampling distribution for the relevant test statistic under the null hypothesis.
   - *What is a test statistic? What is THE test statistic? (provide either the technical formula or the general idea).*
   - *What is a sampling distribution? What is THE sampling distribution? (provide either the technical formula or the general idea).*
3. Compare the observed statistic to the sampling distribution to get the p-value.
   - *What is the p-value? (provide either the definition or the general idea).*

Solution:

1. Assume the null hypothesis is true.
   - *What does it mean for the null hypothesis to be true?*
     If the null hypothesis is true, the average increase in extended survival for treatment A is the same as that for treatment B. The notation for true treatment averages is often given by μ, leading to a null hypothesis of:
     $$H_0 : \mu_A = \mu_B$$
2. Generate a sampling distribution for the relevant test statistic under the null hypothesis.
   - *What is a test statistic? What is THE test statistic?*
     A test statistic is a number that comes from a sample and measures something about the data, usually in reference to the null hypothesis. Here, interest is in comparing the treatment averages, so their difference is paramount. The t-test statistic is given by:
     $$t^* = \frac{\overline{Y}_A - \overline{Y}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}}$$
   - *What is a sampling distribution? What is THE sampling distribution?*
     A sampling distribution represents the possible values and relative frequencies of the test statistic, typically under the assumption that the

null hypothesis is true. The sampling distribution gives a sense of how different $\overline{Y}_A$ and $\overline{Y}_B$ can be when the null hypothesis really is true. That is, the sampling distribution represents the natural variability of the test statistic when H_0 is true. Here:

$$t^* \sim t_{df} \quad df \approx \min(n_A - 1, n_B - 1)$$

3. Compare the observed statistic to the sampling distribution to get the p-value.

   - *What is the p-value?*
     The p-value is the probability of the observed data or more extreme if H_0 is true. That is, the p-value determines where on the sampling distribution our observed test statistic is. If the observed test statistic is not consistent with the sampling distribution (i.e., p-value is tiny), then H_0 can be discounted (rejected!) as a possibility. If the observed test statistic is consistent with the sampling distribution (i.e., p-value is big), then there is no reason to reject H_0.