A.2 Worksheet

Which Gum Lasts Longer?

Background: (Background taken from: Brandt (2001) "FORMULATION CHALLENGE: CONFECTIONERY - A STICKY Situation," found online at: http://www.preparedfoods.com/CDA/ArticleInformation/features/BNP__Features__Item/0,1231,114008,00.html

"Gum chewing dates back to ancient civilizations. Ancient Greeks chewed mastic tree resin, ancient Central American Mayans chewed chicle, and American Indians chewed gum made from spruce tree resin. This gum was eventually replaced by paraffin wax gum. Today's chewing gums are made mostly of synthetic materials."

"Long-lasting flavor is one of the 'holy grails' of the chewing gum industry. For most chewing gums today, flavor lasts about 12 to 13 minutes as a standard…"

Problem:

We want to determine if there is a significant difference in the mean number of minutes it takes for two different brands of chewing gum to lose their flavor.

Instructions:

The Data Sheet contains the gum data that was collected in class (the length of time, in minutes, that the flavor lasted for Brand 1 and Brand 2 gums).

1.

(a) Explain why our two samples (the number of minutes that the flavor lasted for chewing gum Brand 1 and chewing gum Brand 2) are **not** independent.

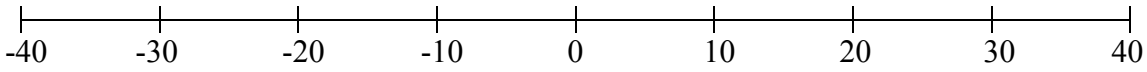
(b) Explain how our data collection scheme would have to be changed for this experiment in order to get two samples of flavor duration values that are independent. Be sure to clearly explain why the samples would be independent, under your modified data collection scheme.

2. Calculate the differences in the number of minutes the flavor lasted for the two gums (Brand 1 minus Brand 2). Enter the differences into the appropriate column on the Data Sheet.

3. Construct a boxplot of the differences.

min = _____ quartile 1 = _____ median = _____ quartile 3 = _____ max = _____

Boxplot:



Based on the boxplot, would you conclude that there is a difference in the number of minutes that flavor is retained for Brand 1 and Brand 2 gums? Explain.

4. Explain why we cannot make a formal statistical conclusion that compares the flavor durations of the two brands of gum based solely on an examination of the boxplot constructed in Question 3.

5. Recall, that in order to test $H_0 : \mu = \mu_0$, we use the test statistic: $t = \dfrac{\overline{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$, where μ_0 is the null hypothesized value for μ, $\overline{x}$ is the mean of the sample, s is the standard deviation of the sample, n is the sample size, and $n-1$ degrees of freedom are used for the test.

Let μ_D = the mean difference in the number of minutes that flavor is retained (Brand 1 minus Brand 2).

To test: $H_0 : \mu_D = \mu_{D_0}$ based on a simple random sample of n_D differences from the population, we use the test statistic: $t = \dfrac{\overline{x}_D - \mu_{D_0}}{\dfrac{s_D}{\sqrt{n_D}}}$, where μ_{D_0} is the null hypothesized difference, $\overline{x}_D$ is the mean of the sample differences, s_D is the standard deviation of the sample differences, and $n_D - 1$ degrees of freedom are used for the test.

Perform a hypothesis test to determine if there is a significant difference in the mean number of minutes that the flavor lasts for Brand 1 and Brand 2 gums.

(1) H_0 : H_A :

(2) test statistic = $t = \dfrac{\overline{x}_D - \mu_{D_0}}{\dfrac{s_D}{\sqrt{n_D}}}$ =

(3) *P*-value =

(4) conclusion =

6. Give a practical interpretation of the *P*-value that you calculated in Question 5.

7. Construct a 95% confidence interval for μ_D (use the formula: $\overline{x}_D \pm t_{n-1,\frac{\alpha}{2}} \dfrac{s_D}{\sqrt{n_D}}$).

Explain how the confidence interval gives the same conclusion as the hypothesis test performed in Question 5.

8.

(a) Would you recommend that this activity be used in future introductory statistics classes? Yes_____ No _____

(b) Do you think that the instructions for this activity are clear? Why or why not?

(c) Do you think that participating in this activity helped you to think about independent samples versus dependent samples? Why or why not?

(d) Why can we not ignore pairing and analyze paired samples data as if we had two independent samples?