Tina Gao & David Lee

3/12/12

Period 4B

Statistics: Green School Project

In today's society, saving energy and resources is not only good for the environment, but also important economically. With the implementation of new energy saving methods, the advantages can be tremendous. We studied and created hypothesis tests for electric and gas energy used by the school from 2009-2010 and from 2010-2011. This essentially compares the energy usage of both years and reveals if there is actually a statistically significant decrease. A decrease would mean the new energy saving equipment is truly beneficial to reducing the amount of energy used.

The reason we chose the school years of 09-10 and 10-11 is that these were the only complete sets of data available. Although the 11-12 school year data is more current and therefore better justified, the school year is not over yet, which can potentially lead to problems when trying to compare with other school years. In addition, other events can contribute to how much energy is used in the last months of school; this often cannot be foreseen and is not modeled off of earlier years. The best way to decrease variability and achieve an accurate comparison is to compare two years with paralleled data, as we have done. However, even with a carefully carried out hypothesis test, issues can still come up. For example, the natural variability of different sets of data can affect how each turn out. There is a possibility that this particular sample of data is really rare and is actually the opposite of our conclusion. However, this is very unlikely because our P-values (the probability that something different happens instead of the norm) are either very high or very low. The risk of Type I Error (probability that we fail to find a

difference between energy usage when it exists) is 5% because that is our "cutoff level" when it comes to determining whether or not energy was saved. Of course, because of individual rounding and calculations, the numbers may not be exact. This is not a huge problem in this case because our numbers were all fairly far from the standard cutoff.

After performing a hypothesis test on the electricity data, we found that there is about a 1.6% chance that more energy was used in the 10-11 school year if, in reality, the amount of energy used in both years was the same. This means it is highly likely that energy usage was conserved during the 10-11 school year, indicating that the new green equipment works well for electricity. After performing a hypothesis test on the gas data, we found that there is about a 63.2% chance that more energy was used in the 10-11 school year if, in reality, the amount of energy used in both years was the same. This means it is pretty likely that energy usage was not conserved during the 10-11 school year, indicating that the new green equipment does not really work for gas. Indeed, it seems to have made things worse by increasing energy consumption, but then again, that may simply be due to variation or other external factors.

Through two successful hypothesis tests using the data from the 09-10 and 10-11 school years, it was determined that the new energy saving methods work well for electric, but do not seem to work from gas. In the future, we can use more electric energy saving methods to help make our school more green, but we can stick with old methods for gas, or simply use the less expensive method.

David L
Tina G.

Electricity

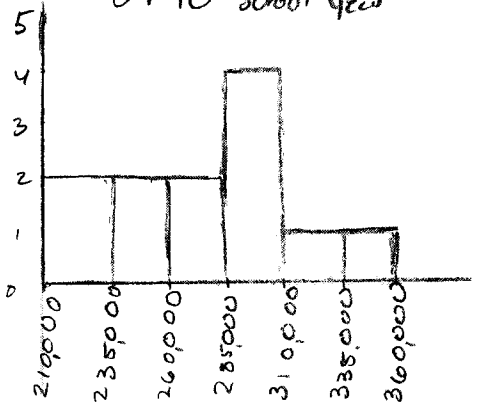
$\mu = \bar{y}_1 - \bar{y}_2$ $\bar{y}_1$ = the mean Kilowatts of electricity that Mammoths Ridge High school used in a single month in the 09-10 school year

$\bar{y}_2$ = the mean Kilowatts of electricity that MRHS used in a single month in the 10-11 school year

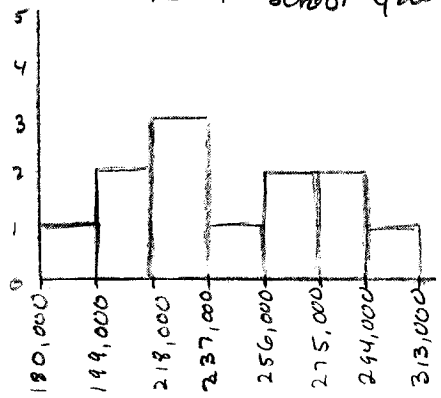
$H_0 - \mu = 0$ $H_A - \mu > 0$

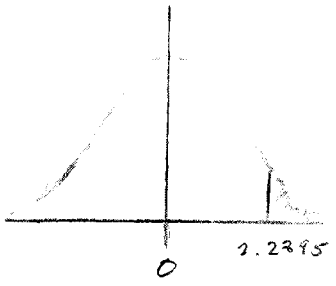
Assume data are independent of each other. Assume sample months are representative of all months. Assume sample size is less than 10% of the population of all months.

09-10 School year



10-11 School year





Both histograms roughly symmetrical and unimodal. No obvious outliers or skewness. Both are nearly normal.

2 sample T-test

$$t_{df} = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{SE(\bar{y}_1 - \bar{y}_2)} \qquad SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}$$

$$= \frac{(277225 - 242150) - 0}{15320} \qquad = \sqrt{\frac{36764^2}{12} + \frac{38267^2}{12}}$$

$t_{df} = 2.2895$ ✓ $p = 0.016$ ✓ $= 15320$

$df = 21.97$ ✓ $\alpha = 0.05$ Reject null hypothesis

There is sufficient evidence at the 0.05 level of significance to suggest that the mean difference of the Kilowatts of energy used by MRHS during the 09-10 school year is greater than the ones used in the 10-11 school year. The p value of 0.016 is less than the alpha level of 0.05. The probability of error is 0.016.

David L.
Tina G.

Gas

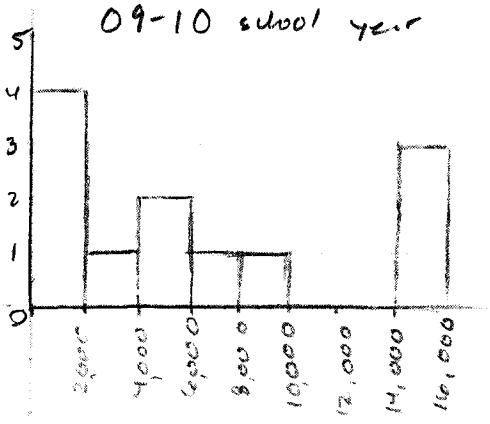
$\mu = \bar{Y}_3 - \bar{Y}_4$ $\bar{Y}_3$ = the mean therms of gas used by MRHS in a single month in the 09-10 school year

$\bar{Y}_4$ = the mean therms of gas used by MRHS in a single month in the 10-11 school year

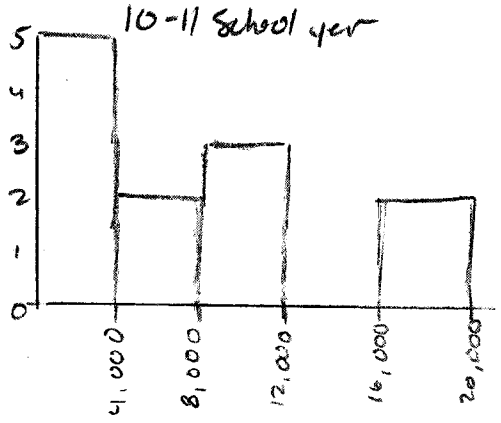
$H_0 - \mu = 0$ $H_A - \mu > 0$

Assume data are independent of each other. Assume sample months are representative of all months. Assum sample size is less than 10% of the population of all months.

09-10 school year



10-11 School year



2 sample T-test

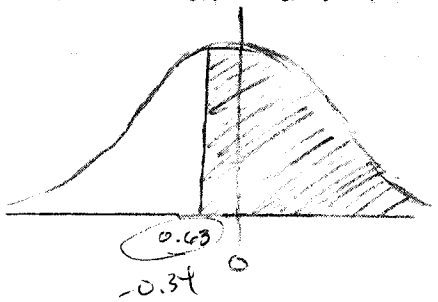
$$t_{df} = \frac{(\bar{Y}_1 - \bar{Y}_2) + (\mu_1 - \mu_2)}{SE(\bar{Y}_1 - \bar{Y}_2)}$$

$$SE(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{5917^2}{12} + \frac{6799^2}{12}} = 2601$$

$$t_{df} = \frac{(6308 - 7193.5) + 0}{2601}$$

$t_{df} = -0.340$ $p = 0.632$ $\alpha = 0.05$ ✓

$df = 21.59$



Fail to reject null hypothesis ✓

There is not sufficient evidence to suggest that the mean difference of the therms of gas used by MRHS during the 09-10 school year is greater than the ones used in the 10-11 school year. This is at the 0.05 level of significance.