

# STA441s18 Assignment Four: Quiz on Tuesday February 6th in tutorial

$$F = \left( \frac{n-p}{s} \right) \left( \frac{a}{1-a} \right) \quad a = \frac{sF}{n-p+sF}$$

Wisconsin Power and Light studied the effectiveness of two devices for improving the efficiency of gas home-heating systems. The electric vent damper (EVD) reduces heat loss through the chimney when the furnace is in the off cycle by closing off the vent. It is controlled electrically. The thermally activated vent damper (TVD) is the same as the EVD except it is controlled by the thermal properties of a set of bimetal fins set in the vent. Ninety test houses were used. New vent dampers were installed in each house. Forty houses were randomly chosen to have an electric vent damper, and the remaining 50 were given a thermal vent damper. For each house, energy consumption was measured for a period of several weeks with the vent damper active ("vent damper in") and for an equal period with the vent damper not active ("vent damper out"). Here are the variables:

```
House Identification Number
Type of furnace (1=Forced air 2=Gravity 3=Forced water)
Chimney area
Chimney shape (1=Round 2=Square 3=Rectangular)
Chimney height in feet
Type of Chimney liner (0=Unlined 1=Tile 2=Metal)
Type of house (1=Ranch 2=Two-story 3=tri-level
               4=Bi-level 5=One and a half stories)
House age in yrs (99=99+)
Type of damper (1=EVD 2=TVD)
Energy consumpt with damper active (in)
Energy consumpt with damper inactive (out)
```

The raw data are available in [furnace.data.txt](#). There is a lesson here. Never trust what you are told about a data file. This is *nothing* compared to what you will encounter in practice. When in doubt, use common sense. Feel free to edit the data file manually if necessary. It's clumsy to fix the main problem with code.

Write a SAS program that reads and labels the data, including `proc format` where appropriate. Create the following new variables:

- Average energy consumption with vent damper in and vent damper out.
- The difference between energy consumption with vent damper in and vent damper out.
- A categorical variable with three values: Ranch, Two-story and Other.

Then,

1. Run `proc means` to obtain sample sizes, means, medians and standard deviations of the quantitative variables. Search for `proc means` online to see how to get medians; this is the quickest way to access documentation. Run `proc freq` to get frequency distributions of the categorical variables (a variable may occur in both sets). Be able to answer basic questions like "What is the median chimney height?" (there are roughly 3 feet to a meter. Does this make sense?), or "What percentage of houses have a forced water furnace?" (my answer is 7.78%)
2. Ignoring all other variables, test whether there is more energy consumption with the damper active or the damper inactive. Be able to answer questions like the following:
  - a. What is the value of the test statistic? The answer is a single number from the printout.
  - b. What is the p-value? The answer is a single number from the printout.
  - c. Do you reject the null hypothesis at the 0.05 level? Answer Yes or No.

- d. Are the results statistically significant at the 0.05 level? Answer Yes or No.
  - e. In plain, non-technical language, what do you conclude, if anything? Say something about furnaces and vent dampers. Statistical terminology is absolutely not allowed here. Pretend you are writing a quick email to your boss, who failed the only Statistics course he ever took, and is touchy about it. Remember that she may forward the email to other "statistical experts" he knows, so beware of accepting  $H_0$  if the test is not significant.
3. We are told that houses were randomly assigned to electrical versus thermal vent dampers. Though we are always polite, we never fully believe anything the client says. If houses were randomly assigned, would you expect to see relationships between type of vent damper and other characteristics of the house? Answer Yes or No and briefly say why. Following up on this point,
    - a. Test the relationship between type of vent damper and the other categorical variables, one at a time and ignoring all other variables in each analysis. For type of house, use the 3-category variable you created. In every case, the null hypothesis is independence of vent damper type and the other variable.
    - b. To test for a relationship between vent damper type and quantitative characteristics of the house, just test for mean differences in age, chimney height and chimney area. For now, ignore all other variables. Be able to state your conclusions if any.
    - c. Does it seem like houses could have been randomly assigned to type of vent damper?
  4. Ignoring all other variables, is there a tendency for houses that consume lots of energy with the vent damper inactive to also consume a lot of energy with the vent damper active?
    - a. Answer the question Yes or No. If the answer is No, add a sentence that protects you against accusations that you are accepting the null hypothesis.
    - b. What is the value of the test statistic? The answer is a single number from the printout.
    - c. What is the p-value? The answer is a single number from the printout.
    - d. Do you reject the null hypothesis at the 0.05 level? Answer Yes or No.
    - e. Are the results statistically significant at the 0.05 level? Answer Yes or No.
    - f. What proportion of the variation in energy consumption with vent damper active is explained by energy consumption with vent damper inactive? Your answer is a single number that may or may not be on the printout.
    - g. The equation for predicting energy consumption with vent damper active from energy consumption with vent damper inactive is  $\text{Predicted } y = b_0 + b_1 x$ . What are  $b_0$  and  $b_1$ ? The answer is a pair of numbers from your printout. (One of the residuals may be an outlier, but don't worry about that for now.)
    - h. For a house that consumed 10 BTU (British Thermal Units) with vent damper out, what is the predicted energy expenditure with vent damper in? A calculator may be helpful here.
  5. The main purpose of this study is to find out whether type of vent damper has an effect on energy consumption. The most obvious way to do the analysis is to make type of vent damper the explanatory variable and energy consumption with vent damper in (that is, vent damper active) the response variable. You would not expect type of vent damper to affect energy consumption when the damper is not active. Carry out the suggested test. There is only one explanatory variable.
    - a. What is the value of the test statistic? The answer is a single number from the printout.
    - b. What is the p-value? The answer is a single number from the printout.
    - c. Do you reject the null hypothesis at the 0.05 level? Answer Yes or No.
    - d. Are the results statistically significant at the 0.05 level? Answer Yes or No.
    - e. What proportion of the variation in energy consumption with vent damper active is explained by type of vent damper? Your answer is a single number on the printout.
    - f. State the results in plain, non-statistical language.
  6. Here's another natural way to do the analysis. Do the two types of vent damper differ in the mean amount of energy they *save*? Your response variable should be the *difference* of two variables in the raw data file. Again, there is only one explanatory variable.
    - a. Answer the question Yes or No. If the answer is Yes, say which type saves more energy. If the answer is No, add a sentence that protects you against accusations that you are accepting the null hypothesis.

- b. What is the value of the test statistic? The answer is a single number from the printout.
  - c. What is the p-value? The answer is a single number from the printout.
  - d. Do you reject the null hypothesis at the 0.05 level? Answer Yes or No.
  - e. Are the results statistically significant at the 0.05 level? Answer Yes or No.
  - f. What is the sample mean saving in energy consumption by using an electrical vent damper (compared to not using it)? The answer is a single number from your printout.
  - g. What is the sample mean saving in energy consumption by using a thermal vent damper (compared to not using it)? The answer is a single number from your printout.
7. One more try. One may view analyzing the differences as a way to take the baseline measurement (energy consumption with vent damper inactive) into consideration. If you think about it, it's a regression with the regression coefficient for energy consumption with vent damper in forced to equal one. What if we allowed that estimated regression coefficient to vary, and just controlled for energy consumption with vent damper inactive, using an ordinary regression. Give it a try. There are two explanatory variables, and no product terms.
- a. What is the value of the test statistic? The answer is a single number from the printout.
  - b. What is the p-value? The answer is a single number from the printout.
  - c. Do you reject the null hypothesis at the 0.05 level? Answer Yes or No.
  - d. Are the results statistically significant at the 0.05 level? Answer Yes or No.
  - e. What proportion of the *remaining* variation in energy consumption with vent damper active is explained by type of vent damper? Your answer is a single number on the printout.
  - f. State the results in plain, non-statistical language.

Bring *both* your log file and your results file to the quiz. Do not write anything on the printouts except your name and student number. You may be asked to hand one or both of them in. These two files *must* be generated by the same SAS program or you may lose a lot of marks. There must be no errors, no warnings and **no notes about invalid data** in your log file. Bring a calculator.



This assignment is licensed under a [Creative Commons Attribution-ShareAlike 3.0 \(or later\) Unported License](https://creativecommons.org/licenses/by-sa/3.0/). Use and share it freely.