

Doing a One-Way Analysis of Variance

Administrative Items

Getting help

- See me Monday 3-5:30 or Wednesday from 4-5:30.
- Send an e-mail to stine@wharton.
- Visit TAs, particularly for help using the computer.

Assignment #3

Review Questions (One-Way Anova)

What's it good for?

Comparing the averages among several groups.

Then why call it one-way analysis of variance?

It is an odd name since anova is a comparison of mean values, not variances. The name comes from the use of variances to determine if the differences among mean values are large.

It's called "one-way" because there is only one grouping of the observations into categories. That is, one column of a JMP spreadsheet defines the groups, not several.

What is important about an anova table?

The crucial term in the anova table is the F ratio and p-value in the far right column. These immediately tell you if the null hypothesis of equal means

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

can be rejected based on the size of the differences *between* the mean values relative to the variation *within* the groups.

What happens if you try anova with only two groups?

You will get the same p-value (or level of significance) as with a two-sample t-test, done in a two-sided manner. With anova, the tests we'll consider are always "two-sided" in that they reject H_0 if the means are far apart, regardless of the direction of the difference.

What's that Bonferroni thing all about?

Do a lot of tests and you tend to make many errors. Anova can lead to a lot of tests because you'll often make numerous comparisons (see next example). The Bonferroni method is a simple way to avoid excessive false positives (claiming a difference when it's not really there). To use Bonferroni, you simply conduct each test at a smaller α level than usual. If you are doing k tests, use $\alpha = 0.05/k$ rather than $\alpha = 0.05$. (What does this do to the power, though?)

Example Application

Gasoline mileage

Check out www.eia.doe.gov/price.html to find world crude price trends.

Context

An oil refiner believes that its new formulation of premium gasoline will produce higher mileage in cars than rival brands. To back up the claims it plans to use in advertising, it has conducted an *experiment*.

The refiner contacted a large rental car firm and rented 60 cars of the same model for a day. The cars were randomly divided into 4 groups of 15 each. The fuel was drained from the cars and replaced by fresh gasoline. Cars in each group received a different type of gasoline (B, E, S, X). The cars were then driven for a day along a standard course that combined interstate and city driving. At the end of the day, the mileage for each was measured and recorded.

Questions

Here are the questions that the refiner would like answered:

- (1) Is there evidence of a difference in mileage among the 4 types of gas?
- (2) Does the gas produced by the refiner (brand E) obtain the highest mileage?
- (3) Does the gas produced by the refiner obtain higher mileage than any rival brand?
- (4) Are there any other differences in mileage among the brands? For example, does brand X give higher mileage than any other?
- (5) Are the data suited to “the standard analysis”, or should alternatives be sought that would better accommodate the features of this data?

JMP

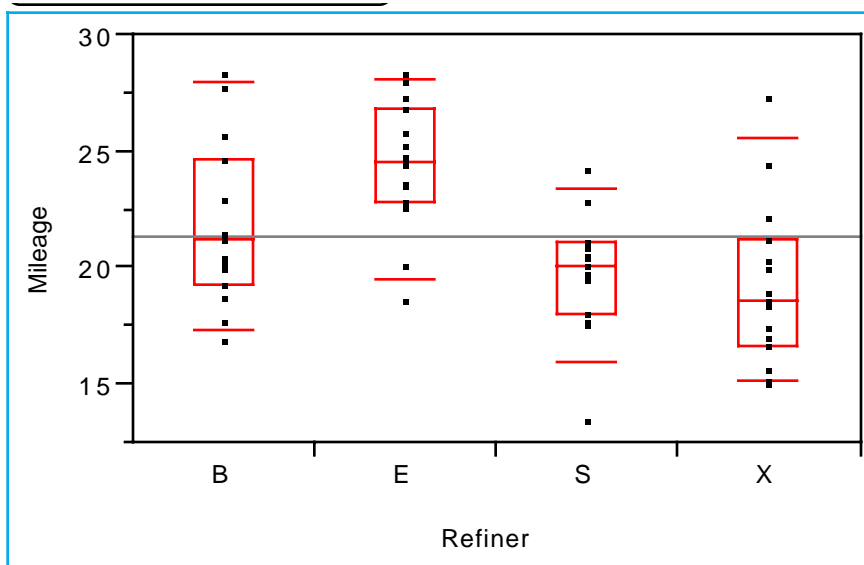
How do I build an analysis of variance with one factor?

- *Fit Y by X* with continuous response (Y) and single categorical predictor (X)
- F-ratio in anova table handles the overall null hypothesis.
- Multiple comparison methods are used for other comparisons:
 - Graphically: Comparison circles show which are different.
 - Tabular summaries: Read the labels to interpret the output.

Doing a One-Way Anova (and answering those 5 questions)

- (1) Think about whether the data collection process might be flawed.
 - Randomized assignment to the groups?
 - Sample size adequate?
 - Confounded effects?
 - Lack of independence?
 - Other ways to reduce extraneous variation?

- (2) Plot the data and see what's going on. What do you see?
 - Outliers? Coding errors?
 - Skewness?
 - Differences in location (mean) or variation?



Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err	Mean	
B	15	21.80	3.46		0.894	
E	15	24.45	2.74		0.708	
S	15	19.92	2.59		0.668	
X	15	19.25	3.47		0.896	

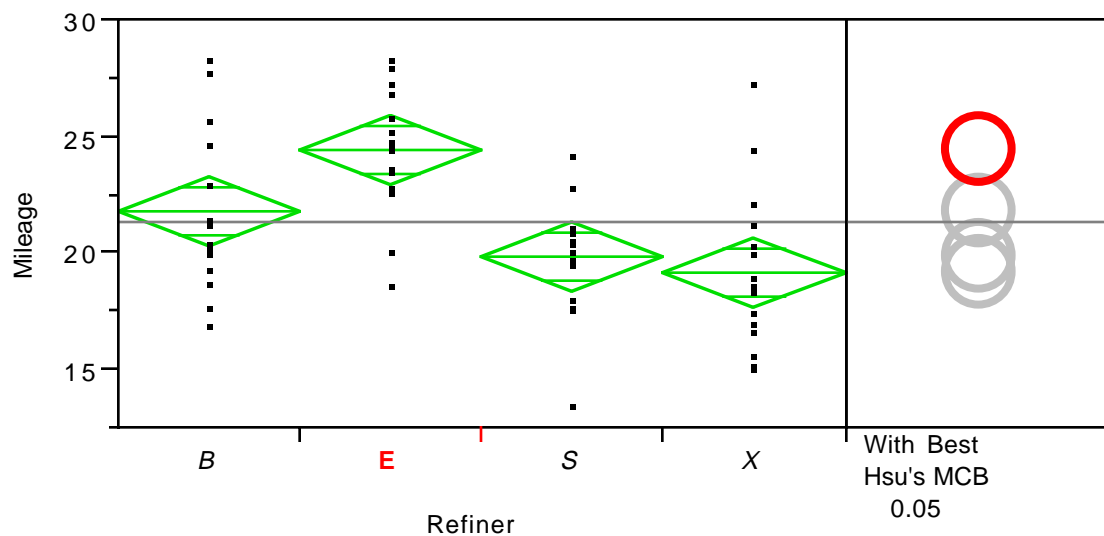
(3) Assuming all is well, is there a difference among the groups on average?

- Use the F-ratio from the Anova table.
- Large (i.e., significant) values of the F imply a difference exists, but do not identify for you *which* differences are meaningful.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob>F
Model	3	244.53	81.51	8.5227	
Error	56	535.58	9.56		
C Total	59	780.11	13.22		<.0001

(4a) To find whether the group with the largest mean is significantly higher than any other, use Hsu's multiple comparison. (analogous procedure for lowest)

- Comparison circles in JMP are easy to interpret. Those in difference colors are significantly different.
- So, in this case, the mileage obtained by brand E is highest and significant.



- Tabular summaries give one end of a confidence interval. When looking for the max, the intervals show are the lower endpoint of the max minus others. If this is negative, the difference is not significant. If positive, it is.

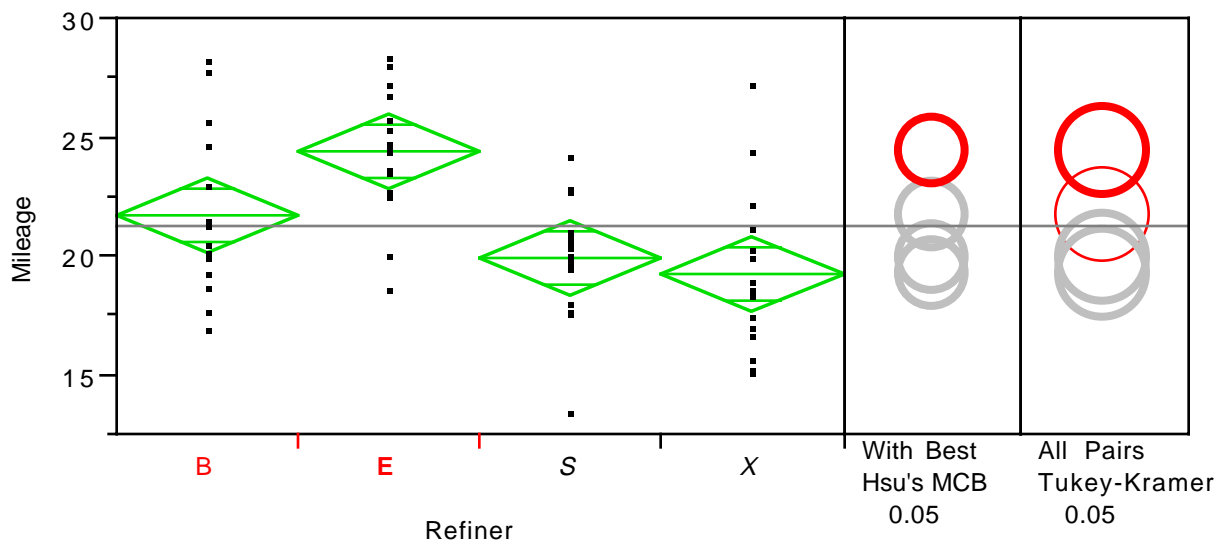
In this case, we see that all of B, S, and X are significantly lower than E.

Mean[i]-Mean[j]-LSD	E	B	S	X
E	-2.38	0.27	2.15	2.83
B	-5.03	-2.38	-0.50	0.17
S	-6.91	-4.26	-2.38	-1.71
X	-7.59	-4.93	-3.05	-2.38

If a column has any positive values, the mean is significantly less than the max.

- (4b)** To check for any pairwise difference in mean values, use Tukey's HSD procedure. These allow you to poke around and make any pairwise comparison at a cost of less power to identify an effect.

- In this example, the lower power of Tukey's method does not find a difference between brands E and B. We are free, however, to explore other comparisons among the brands.



Abs(Dif)-LSD	E	B	S	X
E	-2.99	-0.34	1.54	2.22
B	-0.34	-2.99	-1.11	-0.44
S	1.54	-1.11	-2.99	-2.32
X	2.22	-0.44	-2.32	-2.99

Positive values show pairs of means that are significantly different.

(5) Check assumptions

– Think again about whether independence is valid.

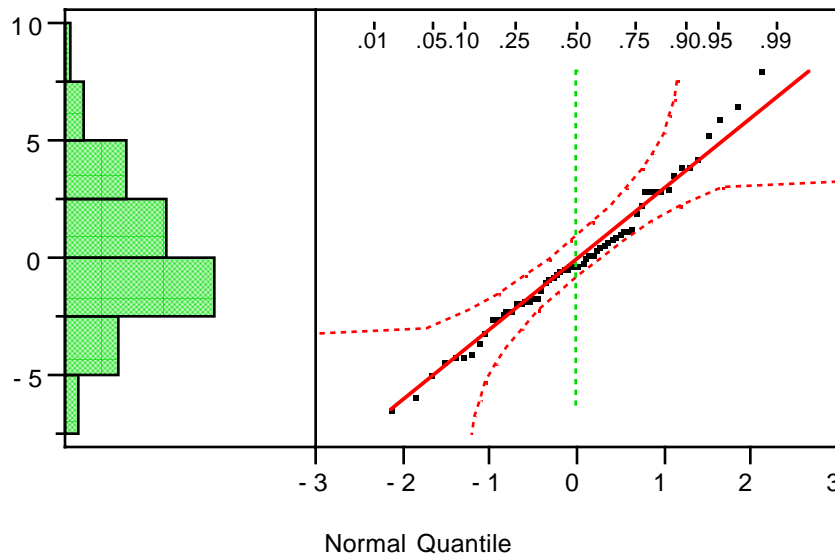
– Equal variance

This is usually ok if the ratio of the smallest to largest SD is only about 2 or smaller. (i.e., one boxplot is no more than twice as long as any other)

– Normality

Check this by considering the “residuals” of the analysis. Ideally, we’d do best by subtracting the group mean from each observation and considering the quantile plot of the residual values. This is tedious to do now, but will be easy once we get to regression.

For these data, the “residuals” are pretty close to normal.



– If the data do appear contaminated by outliers, consider a nonparametric procedure (Wilcoxon rank-sum test). Here we get a significant result; since the data are close to normal, the two methods agree (compare to the Anova table).

Level	Count	Score Sum	Score Mean	(Mean-Mean0)/Std0
B	15	492	32.8000	0.580
E	15	679.5	45.3000	3.782
S	15	363.5	24.2333	-1.596
X	15	295	19.6667	-2.766

1-way Test, Chi-Square Approximation

ChiSquare	DF	Prob>ChiSq
18.7395	3	0.0003

Answers...

- (1) Is there evidence of a difference in mileage among the 4 types of gas?

Yes, the overall F-ratio ≈ 8.5 is significant ($p\text{-value} < 0.0001$).

- (2) Does the gas produced by the refiner (brand E) obtain the highest mileage?

Yes, by using Hsu's multiple comparison.

- (3) Does the gas produced by the refiner obtain higher mileage than any rival?

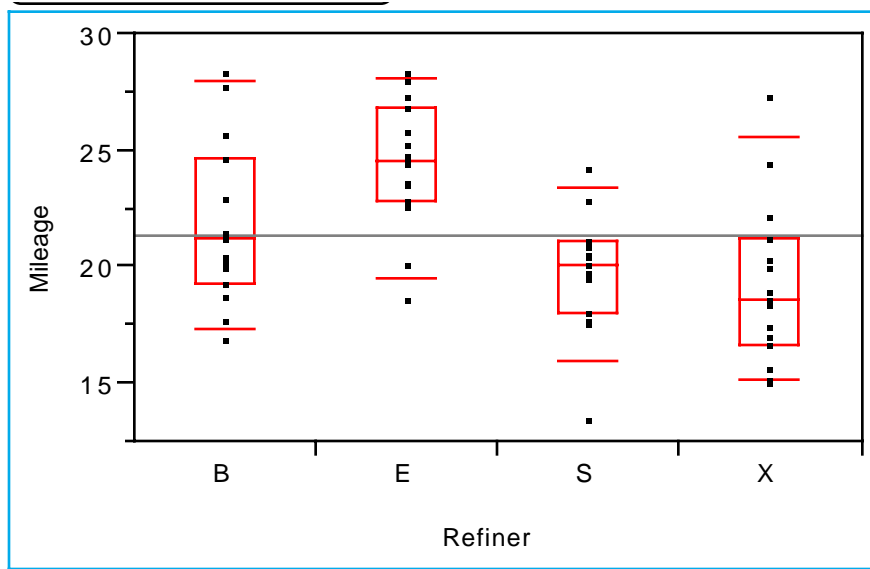
Yes, but use Hsu's comparison or you'll miss a significant comparison between Brand E and the second highest, Brand B.

- (4) Are there any other differences in mileage among the brands? For example, does brand X give higher mileage than any other?

Using Tukey's method, Brand X does worse than Brand E. It's the lowest among the three.

- (5) Are the data suited to "the standard analysis", or should alternatives be sought that would better accommodate the features of this data?

Yes, the assumptions seem reasonable.

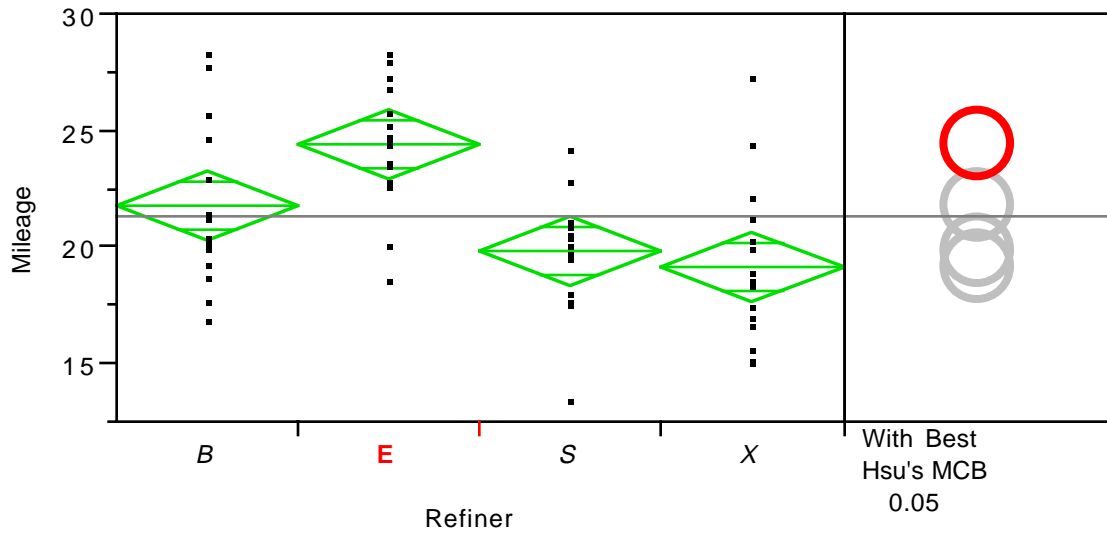


Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err	Mean
B	15	21.80	3.46		0.894
E	15	24.45	2.74		0.708
S	15	19.92	2.59		0.668
X	15	19.25	3.47		0.896

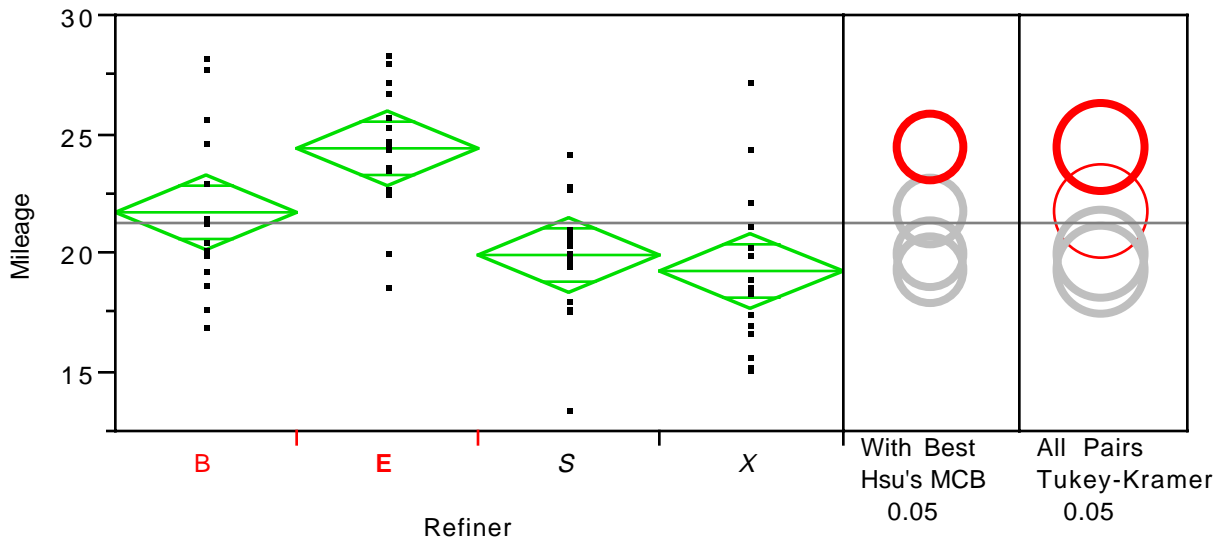
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	244.53	81.51	8.5227
Error	56	535.58	9.56	Prob>F
C Total	59	780.11	13.22	<.0001



Mean[i]-Mean[j]-LSD	E	B	S	X
E	-2.38	0.27	2.15	2.83
B	-5.03	-2.38	-0.50	0.17
S	-6.91	-4.26	-2.38	-1.71
X	-7.59	-4.93	-3.05	-2.38

If a column has any positive values, the mean is significantly less than the max.



Abs(Dif)-LSD	E	B	S	X
E	-2.99	-0.34	1.54	2.22
B	-0.34	-2.99	-1.11	-0.44
S	1.54	-1.11	-2.99	-2.32
X	2.22	-0.44	-2.32	-2.99

Positive values show pairs of means that are significantly different.