One-Way Analysis of Variance

Administrative Items

Midterm

– Grades.
– Make-up exams, in general.

Getting help

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

Review Questions (Chi-square)

How do we calculate the chi-square statistic?

(a) Textbook formula, given that you can figure out the expected frequencies
- Page 399, 404
- Use the null hypothesis to obtain expected counts
  - Use the marginals if testing for independence
  - It's done on a case-by-case basis for goodness-of-fit.

(b) Use JMP software in cases of independence
- Use the column button to tell JMP that the column has frequencies

How do I use the value of the chi-square statistic?

Large values of the chi-square statistic indicate a deviation from the null hypothesis. The chi-square gets larger as the data depart further from the null hypothesis. (Look at the formula – deviations are squared and added.)

What assumptions should be checked?

- The observations used to build the table are independent.
- The *expected* counts are 5 or larger.

New Concepts and Terminology

Analysis of variance method for comparing many mean values

– Generalization of two-sample t-test
- Same assumptions.
- "Same" null H_0: $\mu_1 = \mu_2 = \ldots \mu_I$ vs some difference

– t-test is replaced by an overall test, the overall F test.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	1173.290	130.366	1.0361
Error	90	11324.500	125.828	**Prob>F**
C Total	99	12497.790	126.240	0.4180

Anova table summary

– Test H_0 using a "decomposition of the total variation"

– Anova table summarizes the "sources" of variation.

Model = differences *between* group averages
Error = differences *within* groups.
DF = degrees of freedom
Sum of squares = variation attributed to some source
Mean square = variation divided by degrees of freedom
F Ratio = Ratio of model mean squares to error mean square
Prob>F = p-value of the test of H_0.

Complicating issue

– It's too easy to find a significant effect when many comparisons are made.

– If you do 20 t-tests, you expect to find one error.

– Chance for at least one error among 20 independent α=0.05 tests is

$$P\{\text{error somewhere}\} = 1 - P\{\text{no error}\}$$
$$= 1 - 0.95^{20} \approx 0.64$$

Bonferroni method

– Simple means to control the chance for an error somewhere.

– Based on simple Bonferroni inequality

$$P\{\text{Error}_1 \text{ or Error}_2 \text{ or} \ldots \text{Error}_{20}\} \le \sum P\{\text{Error}_i\} = 20\ (0.05)$$

– Want to do 10 "honest" t-tests at one time? Use a p-value of 0.05/10 = 0.005 for each one rather than usual 0.05 cutoff.

– Moral: Being "punished" for not having better, sharper theory.

Other multiple comparison methods

– Goal: Locate important differences while avoiding false claims of significant differences.
– Problem: Suppose you have 10 groups $\Rightarrow$ 45 pairwise comparisons.
– Three methods of comparison, depending upon the question of interest:
  Least significant differences (LSD)... generally avoid.
  Hsu's comparisons (HSU)... Which is best/worst?
  Tukey-Kramer comparisons (HSD) ... Are any differences significant?

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)
– Graphically: Comparison circles show which are different.
– Tables summarize the differences. Read the labels to interpret the output.

Examples for Today

Selecting the best vendor Repairs.jmp, page 233

Question

"Does one vendor stand out as the best (lowest cost)?"

Data

10 service calls for each of 10 vendors, price of comparable repair.

– Initially model as a multiple regression using one categorical variable. t-ratios for the slopes suggest one is significant ($p = 0.0307$). Is this appropriate?

– Bonferroni says "no", none are significant. (all p-values are > 0.005).
– F-ratio says "no", the factor as a whole is not useful.
– Multiple comparisons via Hsu's procedure (Fit Y by X) also says "no".
Graphically (p 236) via linked comparison circles.
Tabular (p 237) The tables show *one* endpoint of confidence interval for the difference in mean values.

– Decoding the table:
We want to know if the smallest value is really smaller than the rest. The difference

Avg(smallest) – Avg(other group)

will always be negative, so the lower endpoint of the confidence interval will also be negative. The upper endpoint might be negative or positive. If it's negative, then zero is *not* in the interval and the difference is significant. If it's positive (as all are in the table at the bottom of page 237), the differences are not significant.

– Use the hints that the JMP output provides
"If a column as any negative values, the mean [for that column] is significantly greater than the min [smallest mean value]."

– Checking assumptions: independence, constant variance, normality.

Conclude

Differences among the vendors are not significant. Get more data and have a more focused comparison next time. (Note the deceptive use of an inappropriate analysis to conclude a significant effect does exist, p 242.)

Headache pain relief

Headache.jmp, page 243

Question

"What claims can be made comparing these drugs?"

Data

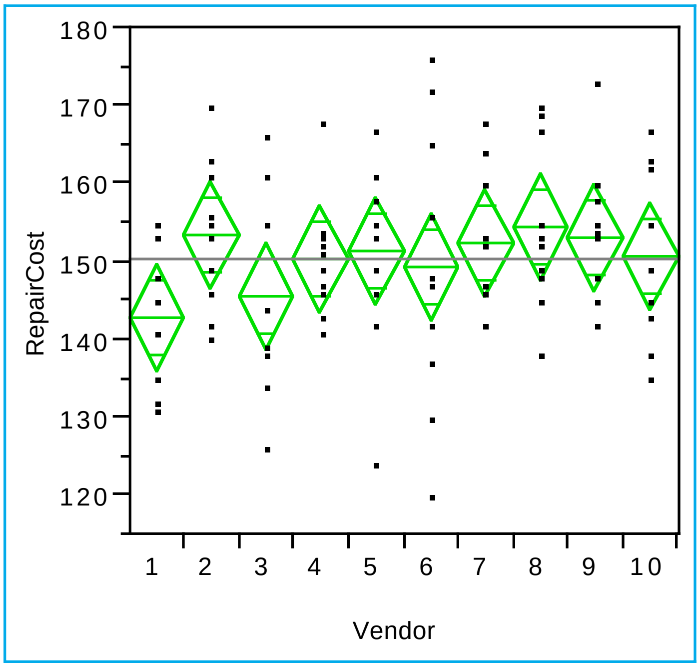
Tests of several "active compounds" and placebo.

– Similar in spirit to methods used in clinical trials.

– Outliers are still relevant (after all, it's still multiple regression).
– Use Tukey-Kramer since interested in *all possible comparisons* for marketing claims rather than just deciding which if any offers the most relief.

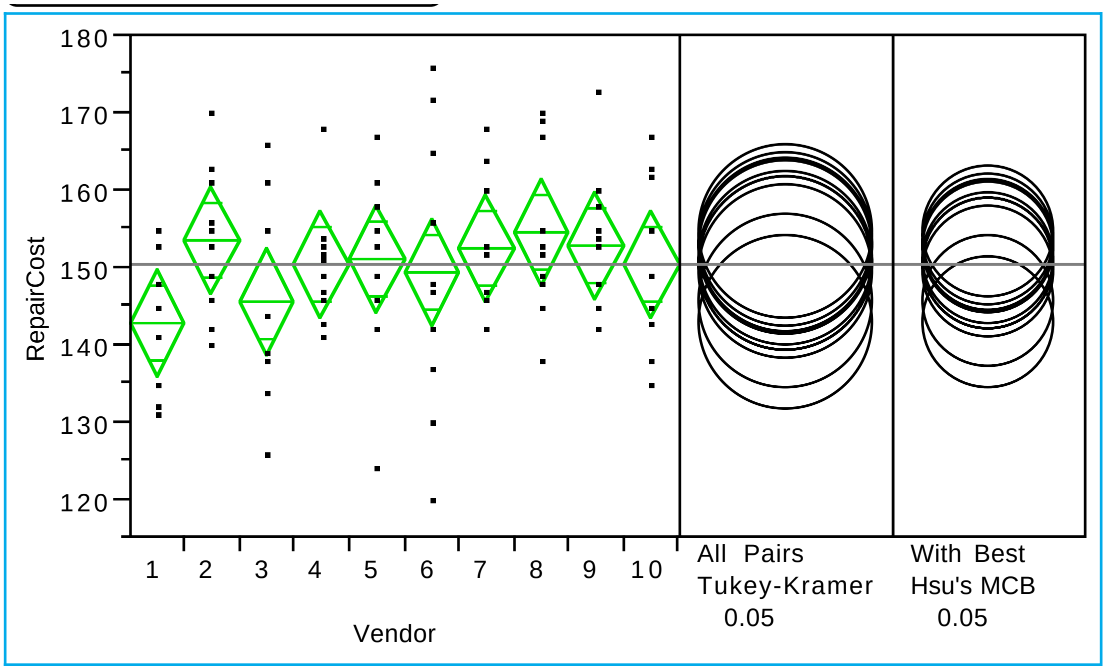
Conclude

Active compound #2 is clearly best, even allowing for all six possible pairwise comparisons. None of the others differs from each other in the amount of relief offered.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	1173.290	130.366	1.0361
Error	90	11324.500	125.828	**Prob>F**
C Total	99	12497.790	126.240	0.4180



Comparisons for all pairs using Tukey-Kramer HSD

q *
3.24444

Abs(Dif)-LSD	8	2	9	7	5	10	4	6	3	1
8	-16.28	-15.18	-14.68	-14.18	-12.98	-12.28	-12.08	-10.98	-7.38	-4.68
2	-15.18	-16.28	-15.78	-15.28	-14.08	-13.38	-13.18	-12.08	-8.48	-5.78
9	-14.68	-15.78	-16.28	-15.78	-14.58	-13.88	-13.68	-12.58	-8.98	-6.28
7	-14.18	-15.28	-15.78	-16.28	-15.08	-14.38	-14.18	-13.08	-9.48	-6.78
5	-12.98	-14.08	-14.58	-15.08	-16.28	-15.58	-15.38	-14.28	-10.68	-7.98
10	-12.28	-13.38	-13.88	-14.38	-15.58	-16.28	-16.08	-14.98	-11.38	-8.68
4	-12.08	-13.18	-13.68	-14.18	-15.38	-16.08	-16.28	-15.18	-11.58	-8.88
6	-10.98	-12.08	-12.58	-13.08	-14.28	-14.98	-15.18	-16.28	-12.68	-9.98
3	-7.38	-8.48	-8.98	-9.48	-10.68	-11.38	-11.58	-12.68	-16.28	-13.58
1	-4.68	-5.78	-6.28	-6.78	-7.98	-8.68	-8.88	-9.98	-13.58	-16.28

Positive values show pairs of means that are significantly different.

Mean[i]-Mean[j]+LSD	8	2	9	7	5	10	4	6	3	1
8	12.32	13.42	13.92	14.42	15.62	16.32	16.52	17.62	21.22	23.92
2	11.22	12.32	12.82	13.32	14.52	15.22	15.42	16.52	20.12	22.82
9	10.72	11.82	12.32	12.82	14.02	14.72	14.92	16.02	19.62	22.32
7	10.22	11.32	11.82	12.32	13.52	14.22	14.42	15.52	19.12	21.82
5	9.02	10.12	10.62	11.12	12.32	13.02	13.22	14.32	17.92	20.62
10	8.32	9.42	9.92	10.42	11.62	12.32	12.52	13.62	17.22	19.92
4	8.12	9.22	9.72	10.22	11.42	12.12	12.32	13.42	17.02	19.72
6	7.02	8.12	8.62	9.12	10.32	11.02	11.22	12.32	15.92	18.62
3	3.42	4.52	5.02	5.52	6.72	7.42	7.62	8.72	12.32	15.02
1	0.72	1.82	2.32	2.82	4.02	4.72	4.92	6.02	9.62	12.32

If a column has any negative values, the mean is significantly greater than the min.