

Two-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 the StatLab/TAs, particularly for help using the computer.

Plan for this week

Assignment #4

Review of One-Way Anova

Underlying “model” (see text, Defn 11.1)

Assume that the data have the form (I groups, with n_i in the i^{th})

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad i=1, \dots, I \quad j = 1, \dots, n_i$$

where

- (1) observations are independent,
- (2) observations have constant variance σ^2 , and
- (3) observations are normally distributed.

The error terms are thus also independent from one observation to another and are normally distributed with mean zero and variance σ^2 , written as

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Estimates of unknown parameters

Estimate the group means using the sample means

$$\hat{\mu}_i = \bar{y}_i$$

and estimate the error variance using the within sum of squares

$$\hat{\sigma}^2 = \frac{\text{Within SS}}{n - I} = \frac{\sum (y_{ij} - \bar{y}_i)^2}{n - I} = \frac{\sum (\hat{\varepsilon}_{ij})^2}{n - I}$$

Questions to answer in a one-way anova

- (a) Is there a difference among the underlying means μ_i ?
- (b) Is the group with the highest (lowest) average significantly better (worse)?
- (c) Are there significant differences between some mean values?
- (d) What should be done if the data are not normal?

Methods for answering these questions

- (a) F-test from the Anova table
- (b) Hsu's multiple comparisons
- (c) Tukey-Kramer multiple comparisons (confirm these “by hand”)
- (d) Outliers? Consider a nonparametric rank-sum comparison.

Example Application for One-Way Anova

Context from last time...

Claims of better mileage for a brand of gasoline, based on the following experiment:

A refiner rents 60 cars of the same model for a day and randomly divides the cars into 4 groups of 15 each. The fuel was drained from the cars and replaced by fresh gasoline of four types (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

- (a) Is there evidence of a difference in mileage among the 4 types of gas?
- (b) Is the formulation obtaining the highest mileage significantly the best?
- (c) Are there any other differences in mileage among the 4 brands?
- (d) 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.

Answering the questions

Analysis preliminaries

See notes from previous class regarding important preliminary steps (e.g., flaws in the data collection process, sources of dependence).

(a) Evidence of a difference among the 4 types of gas?

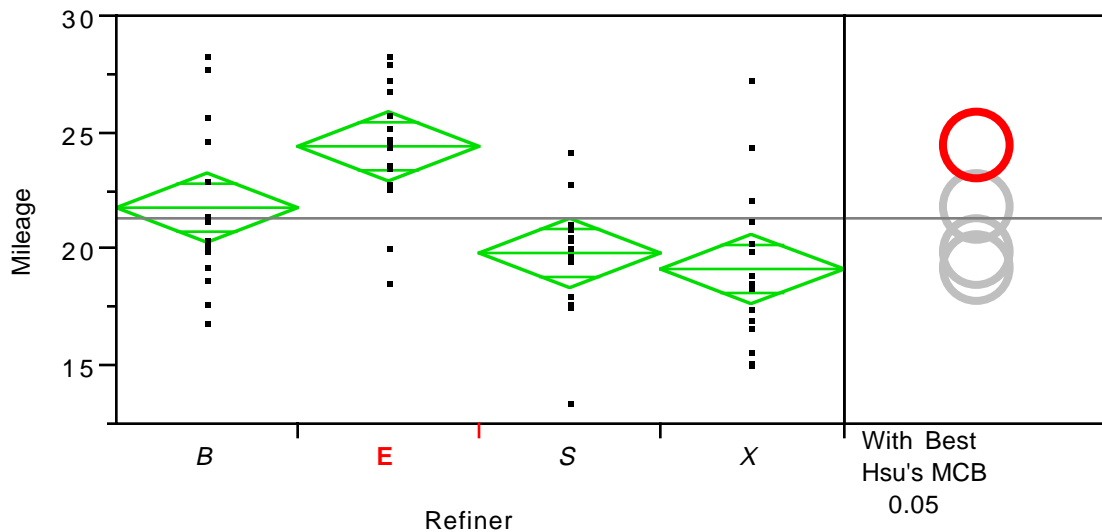
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob>F
Model	3	244.53	81.51	8.5227	
Error	56	535.58	9.56		Prob>F
C Total	59	780.11	13.22		<.0001

F-ratio indicates a significant difference exists, but does not indicate where the difference is. For that, we must go on to multiple comparison methods.

Note that the table also includes an estimate of σ^2 , namely

$$\hat{\sigma}^2 = \frac{\text{Within SS}}{n - I} = \text{Error Mean Square} = 9.56$$

(b) High mileage group significantly better than others?

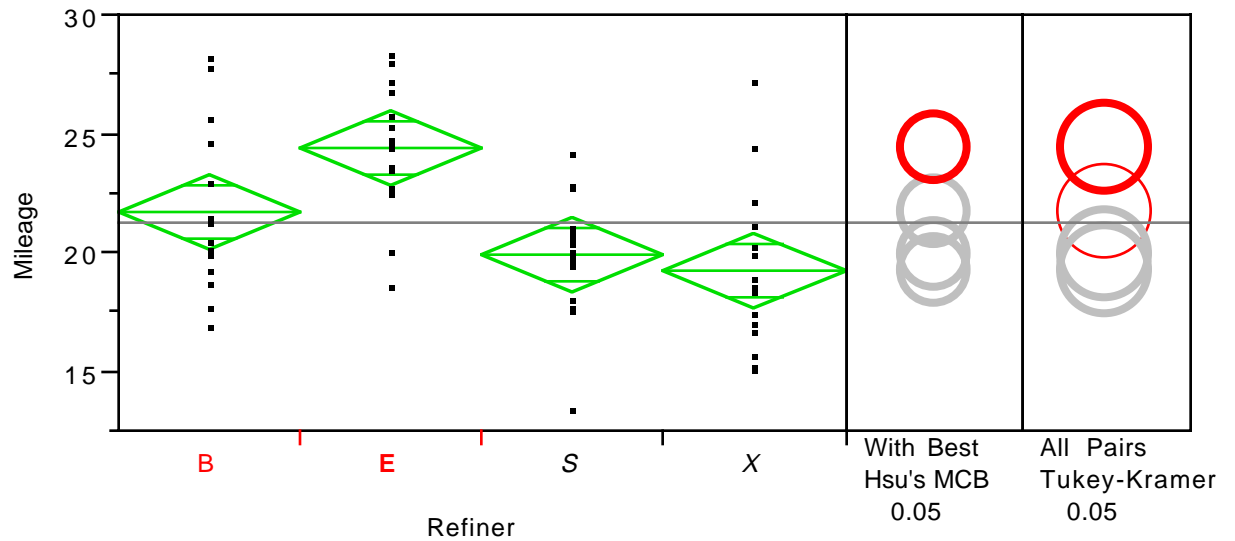


Mean[i]-Mean[j]-LSD	E	B	S	X
E	-2.38	0.27	2.15	2.83
B	-5.03	-2.38	-0.50	0.17
S	-6.91	-4.26	-2.38	-1.71
X	-7.59	-4.93	-3.05	-2.38

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

Yes, both graphically and in the table: E is better by a significant margin.

(c) Other differences in mileage among the 4 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.

It is useful to see how to calculate the values in the table (see text, Section 11.3) JMP shows the lower endpoints of confidence intervals for the absolute value of the difference in means. The confidence interval is computed as

$$(\text{difference in means}) \pm q_{\alpha}(\text{number of means, error df}) \sqrt{\hat{\sigma}^2 / \text{num per group}}$$

with q from Table 8. For example, the interval for E–B is ($\alpha=0.05$)

$$(24.45 - 21.8) \pm 3.74 \sqrt{9.56/15} = 2.65 \pm 3.74(.798) = 2.65 \pm 2.98$$

Notice that the term after the final “±” is the diagonal value in JMP’s table. The lower endpoint of this interval is –0.33, matching JMP’s value –0.34 (up to my rounding).

Thus, only the differences E–S and E–X are significant in this comparison. This does indeed obtain a different answer than using the previous Hsu’s comparison that showed a significant difference between E and B (WHY?).

(d) Use “the standard analysis”, or nonparametrically?

Start by checking assumptions. There’s no evident lack of constant variance, and the data appear close to normal.

Were the data contaminated by outliers, consider a nonparametric procedure. Below is JMP’s output for the Wilcoxon rank-sum test (also obtained with an option from the *Fit Y by X* output).

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

Since the data are close to normal, the two methods agree (compare chi-square to the Anova table).

Introduction to Two-Way Anova

Two factors in the experiment

Often have more than one factor that is of interest.

For example, consider an experiment in web-page design. A company wants to learn about the effect of one-click checkouts. It also has two types of page layouts. One layout is more graphical, but loads more slowly and some consider it “complex”. The other layout is simpler, with less graphics. Rather than run two experiments, the company would like to run one. (This choice also has other advantages.)

The questions of interest:

Does the number of check-out clicks matter?

Does the layout matter?

And as a bonus

Is there an *interaction* between the two?

These can, of course, be followed up with multiple comparisons.

Interaction in two-way anova

Interaction occurs when certain combinations of factors have more or less effect than what you would expect based on their overall performance.

Learning about interaction is often the most important benefit of a two-way (and fancier) analysis of variance.

Underlying “model”

Assume that the data have the form

$$y_{ijk} = (\text{cell mean}) + \varepsilon_{ijk} \quad i=1, \dots, I \quad j = 1, \dots, J, \quad k = 1, \dots, n_{ij}$$

where we represent the cell mean as

cell mean = overall mean + row effect + column effect + interaction.

Again, the other assumptions are

(1) observations are independent,

(2) observations have constant variance σ^2 , and

(3) observations are normally distributed.

so that the error terms are $\varepsilon_{ijk} \sim N(0, \sigma^2)$

Estimate error variance

Again estimate the error variance using the error sum of squares

$$\hat{\sigma}^2 = \frac{\text{Error SS}}{n - IJ} = \frac{\sum (y_{ijk} - \bar{y}_{ij})^2}{n - IJ} = \frac{\sum (\hat{\varepsilon}_{ijk})^2}{n - IJ} = \text{Error MS}$$

JMP for Two-Way Anova

Fit model command

- Use the more complex “fit model” command.
- Watch with care the formation of interaction terms
(Add a factor twice, then with both it and the other term highlighted in the JMP dialog, use the “cross” button to build the interaction term)
- Profile plots are constructed by the button near the graphic for the interaction term in the output of the fit model command.

Concatenate function

- Use the concatenation of labels (formula) to look at individual cell means.
- The concatenate function is a character formula.

Example: Two-Way Analysis of Variance

Web page experiment

Customers in a “focus group” representing the target audience of the web site were randomized into six groups, with 10 in each group.

Each group was shown a page with either high or low graphics layout, and with either one, two, or three+ clickout options. After using the page, customers rated the “convenience and esthetics” of the design on a 0-100 scale.

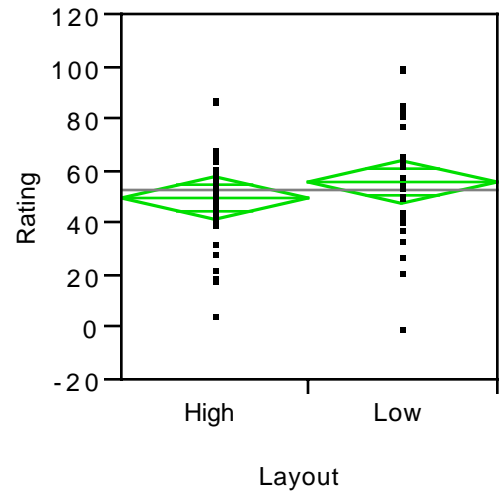
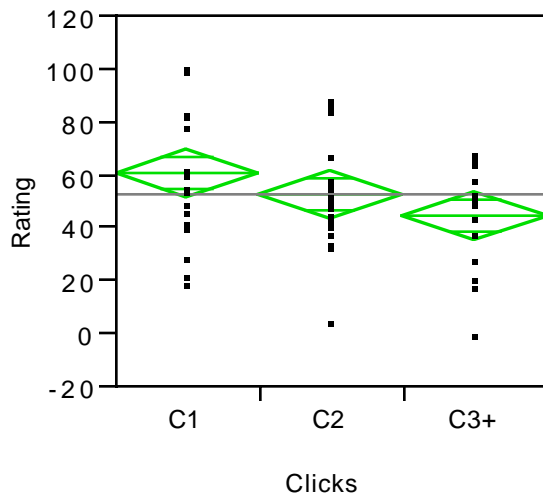
The design of the experiment thus has a “two-way” arrangement...

Layout-Clicks	One-click	Two-click	Three+	
Low graphics	10	10	10	30
High graphics	10	10	10	30
	20	20	20	60

Initial conceptual analysis

Initial graphical analysis

- no severe outliers, similar variances.
- trends in mean values are weak: neither one-way analysis is significant.



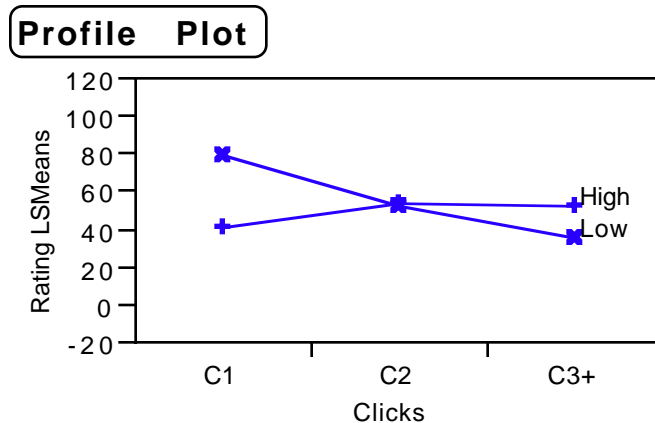
Anova table: overall differences in means?

- First check for interaction before considering marginal effects of row and column factors.
- In this example, we find a lot of interaction. ($F = 11.23$ with $p\text{-value} < .0001$)

Effect	Test					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F	
Clicks	2	2	2582.1	3.70	0.0312	
Layout	1	1	653.4	1.87	0.1768	
Layout*Clicks	2	2	7837.3	11.23	<.0001	

- In the presence of so much interaction, interpretation of the “marginal” effects is potentially misleading. Although differences in the cell means are present, you should not attempt to judge them from the marginal levels of “Clicks” or “Layout”

– Plot of cell means is very useful (aka, a profile plot)



The crossing lines in this figure (the points are the cell means) describe the interaction indicated in the anova table. The layout with low graphics is preferred when featured with one-click checkout, whereas users are ambivalent or prefer high graphics with the more elaborate checkouts.

Multiple Comparisons: which means are significantly different?

– Using JMP, you can simply reduce the problem to a one-way problem and use those tools. Join the labels of the two factors using the concatenation function, then do a one-way anova...

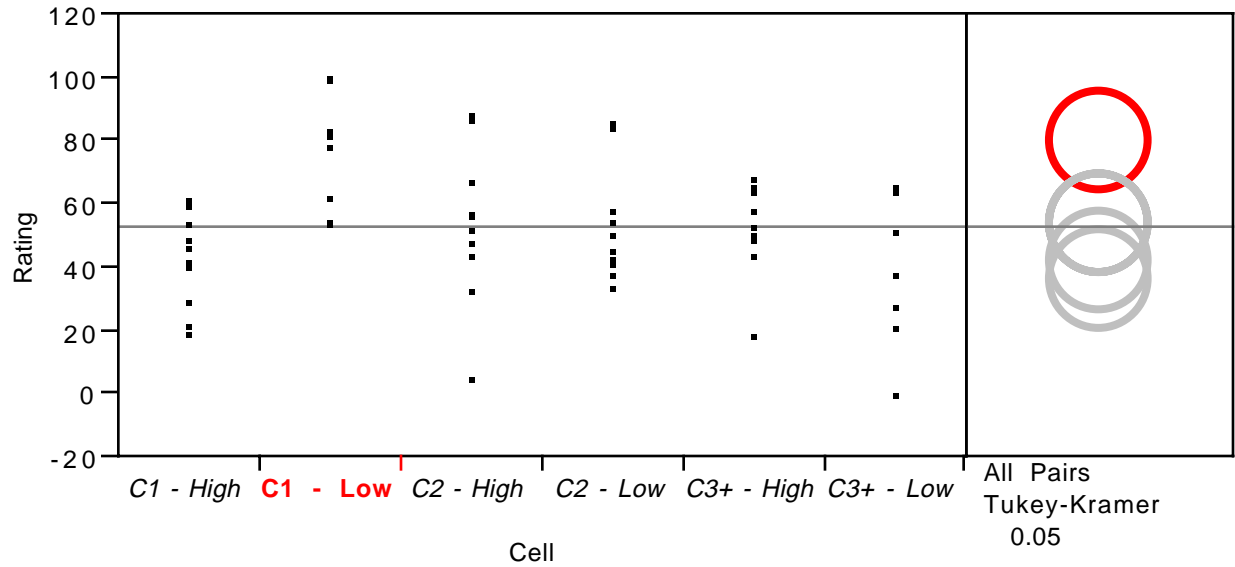
Join the labels of “Clicks” and “Layout” as a new column, then use this one column of categories as a one-way anova.

Here’s a table of the cell means and SDs. These are the mean values plotted in the profile plot shown above.

Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err	Mean	
C1 - High	10	42.2	14.89		4.71	
C1 - Low	10	79.6	17.67		5.59	
C2 - High	10	53.7	24.51		7.75	
C2 - Low	10	53.4	18.20		5.75	
C3+ - High	10	53.5	14.95		4.73	
C3+ - Low	10	36.2	20.08		6.35	

Shown graphically with comparison circles...

The one-click option with low graphics layout is significantly different from the others using Tukey-Kramer.



Here are details for the Tukey-Kramer figure...
What other differences are significant?

Comparisons for all pairs using Tukey-Kramer HSD

	q *					
	2.95448					
Abs(Dif)-LSD	C1 - Low	C2 - High	C3+ - High	C2 - Low	C1 - High	C3+ - Low
C1 - Low	-24.68	1.22	1.42	1.52	12.72	18.72
C2 - High	1.22	-24.68	-24.48	-24.38	-13.18	-7.18
C3+ - High	1.42	-24.48	-24.68	-24.58	-13.38	-7.38
C2 - Low	1.52	-24.38	-24.58	-24.68	-13.48	-7.48
C1 - High	12.72	-13.18	-13.38	-13.48	-24.68	-18.68
C3+ - Low	18.72	-7.18	-7.38	-7.48	-18.68	-24.68

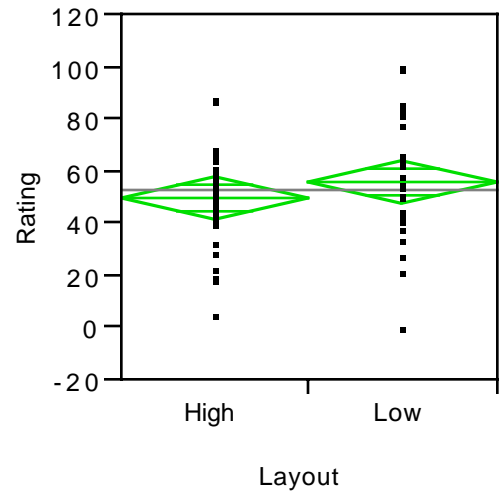
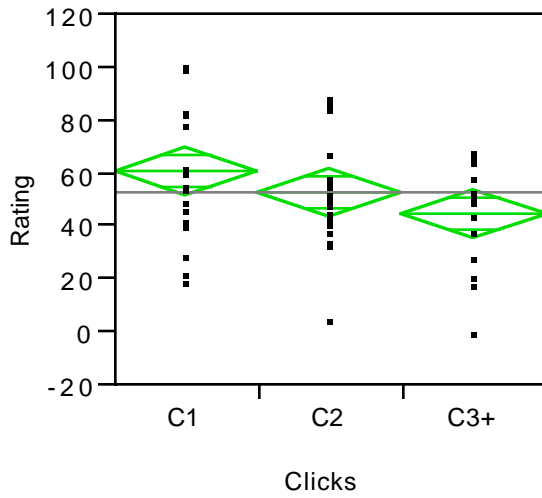
Positive values show pairs of means that are significantly different.

Conclusions

Significant interaction, so that marginal effects cannot be interpreted directly.

Tukey-Kramer intervals find significant differences among cell means.

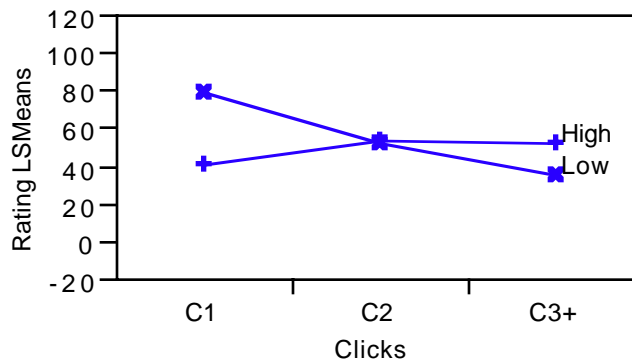
Business implications.



Effect Test

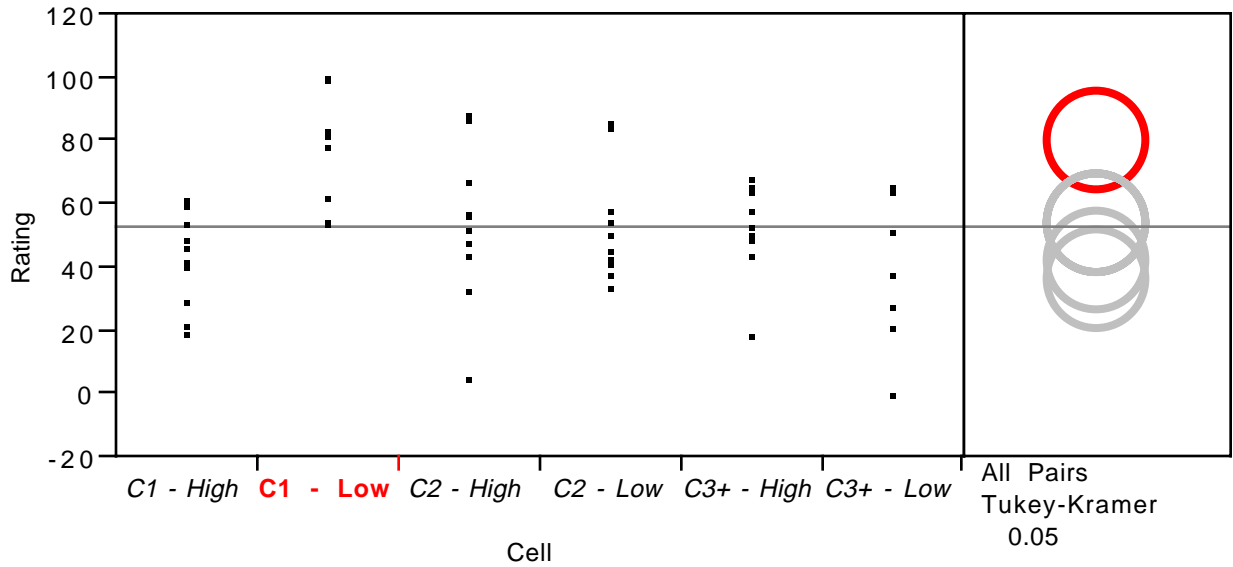
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Clicks	2	2	2582.1	3.70	0.0312
Layout	1	1	653.4	1.87	0.1768
Layout*Clicks	2	2	7837.3	11.23	<.0001

Profile Plot



Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err	Mean
C1 - High	10	42.2	14.89		4.71
C1 - Low	10	79.6	17.67		5.59
C2 - High	10	53.7	24.51		7.75
C2 - Low	10	53.4	18.20		5.75
C3+ - High	10	53.5	14.95		4.73
C3+ - Low	10	36.2	20.08		6.35



$$(difference\ in\ means) \pm q_{\alpha}(number\ of\ means, error\ df) \sqrt{\hat{\sigma}^2 / num\ per\ group}$$

Comparisons for all pairs using Tukey-Kramer HSD

	q *					
	2.95448					
Abs(Dif)-LSD	C1 - Low	C2 - High	C3+ - High	C2 - Low	C1 - High	C3+ - Low
C1 - Low	-24.68	1.22	1.42	1.52	12.72	18.72
C2 - High	1.22	-24.68	-24.48	-24.38	-13.18	-7.18
C3+ - High	1.42	-24.48	-24.68	-24.58	-13.38	-7.38
C2 - Low	1.52	-24.38	-24.58	-24.68	-13.48	-7.48
C1 - High	12.72	-13.18	-13.38	-13.48	-24.68	-18.68
C3+ - Low	18.72	-7.18	-7.38	-7.48	-18.68	-24.68

Positive values show pairs of means that are significantly different.