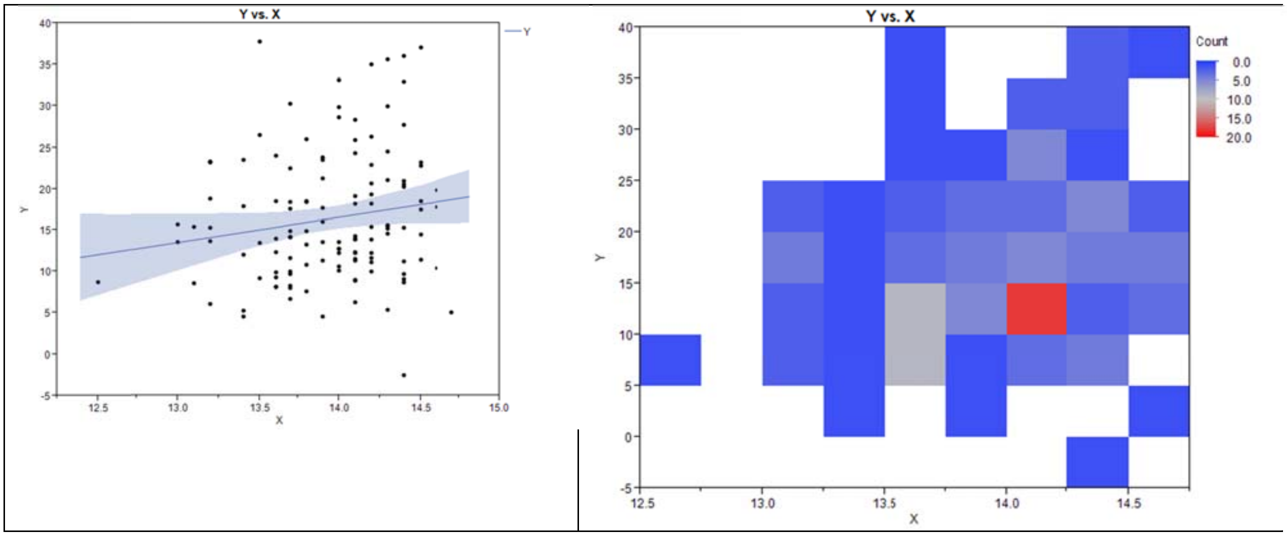
Using heatmap to interpret regression results in JMP

Chong Ho Yu, Ph.D.s
chonghoyu@gmail.com

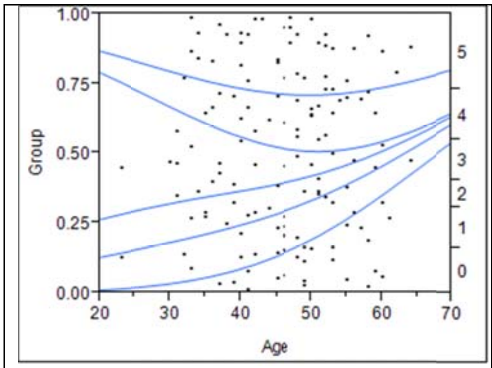
March 21, 2013

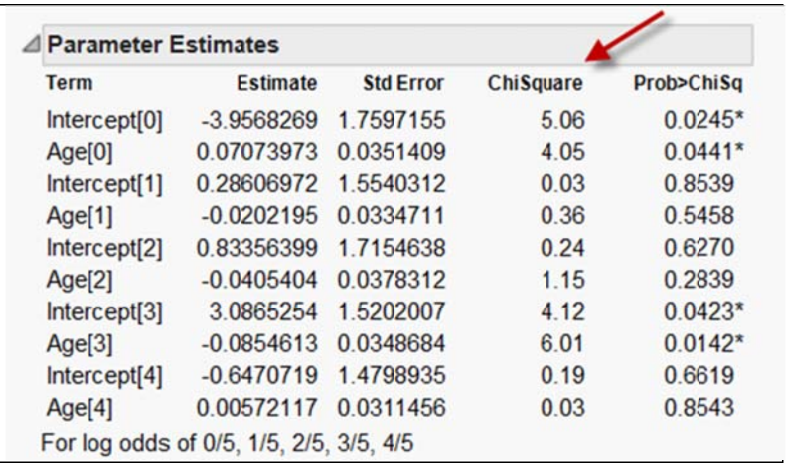
Scatterplot and regression line are very useful for investigating the relationship between the dependent and independent variables. However, when there are a lot of observations, the data points may overlap each other on the scatterplot, and as a result it hinders the analyst from seeing the hidden pattern. For example, the following scatterplot looks very normal. As X increases, Y increases.



Nevertheless, further insight can be unveiled when the analyst converts the scatterplot into a heatmap (press the heatmap icon on **Graphic Builder**). A heatmap uses the color spectrum to represent the frequency counts. The red color depicts a dense concentration whereas the blue color denotes fewer observations. In the preceding heatmap it is obvious that there is a high concentration of subjects when X is between 14 and 14.25, and Y is between 10 and 15.

It is important to point out that a regression line is programmed to pass through as many points as possible with the least square of residuals. If there is an outlier, the model will be driven by the outlier. In this model although there is no obvious outlier, some areas have fewer observations than the others. In this case, the analyst might choose a weighted regression to obtain a better estimation. Alternatively, the researcher could look at the heatmap and draw the conclusion that the prediction yielded from this model is more accurate given certain X and Y values. The premise is that the inference is stronger when there are more subjects.





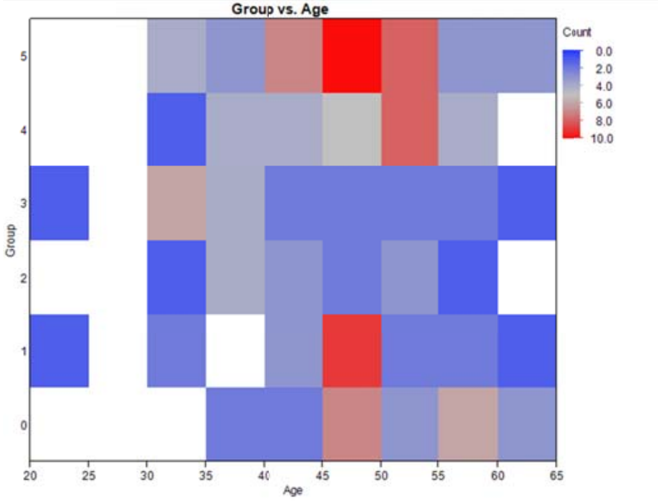
Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-3.9568269	1.7597155	5.06	0.0245*
Age[0]	0.07073973	0.0351409	4.05	0.0441*
Intercept[1]	0.28606972	1.5540312	0.03	0.8539
Age[1]	-0.0202195	0.0334711	0.36	0.5458
Intercept[2]	0.83356399	1.7154638	0.24	0.6270
Age[2]	-0.0405404	0.0378312	1.15	0.2839
Intercept[3]	3.0865254	1.5202007	4.12	0.0423*
Age[3]	-0.0854613	0.0348684	6.01	0.0142*
Intercept[4]	-0.6470719	1.4798935	0.19	0.6619
Age[4]	0.00572117	0.0311456	0.03	0.8543

For log odds of 0/5, 1/5, 2/5, 3/5, 4/5

A heatmap is also useful in multinominal logistic regression modeling. The preceding panels are typical results of such modeling. The variable Age is used to predict the group membership (0-5). For a single individual, one can either be in Group 0 or somewhere else. But when there are many people, it makes sense to ask about the probability of belonging to a specific group. For example, if 20 out of 100 people are classified into Group 0, then the probability is .2. The left panel above shows this type of logistic function: Given the age, what is the probability of belonging to a particular group. However, we can see that the data points are everywhere and thus the logistic function does not seem to be very helpful.

When we turn to the numeric output, as shown in the right panel, we can look at the significance of Age as a predictor of each group membership, based on the Chi-square statistics. Chi-square analysis is a test of goodness of fit between the actual and the predicted cell counts. The algorithm partitions age into several groups and then form an Age X Group crosstab table. Next, the discrepancy in each cell is computed and summed into the overall Chi-square statistics. Again, it is not easy to understand what is going on.



By converting the scatterplot into a heatmap, the analyst can see a visual equivalence to a crosstab table. There are five groups on the Y-axis and Age is partitioned into 9 groups. Hence, it becomes a 6X9 crosstab table.
If Age and a particular group has a perfect relationship, then we should see a pattern in the color spectrum (e.g. from blue to red or from red to blue). But a perfect relationship (like Snow White and Prince Charming) is hardly found. Nonetheless, based on the high concentration in the upper row the researcher can conclude that when Age is between 45 and 50, the subject tends to be in Group 5.
Happy JMPing!