Using JMP for quick IRT results

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

April 21, 2008

The objective of this document is to illustrate how to obtain quick results of Item Response Theory (IRT) modeling using JMP, which is available at http://www.jmp.com IRT is a proven assessment approach that can help you identify good/bad items, create better tests and give you better estimates of student's abilities. If students are getting better exams and teachers are more aware of the students' strengths and weaknesses, this can greatly improve AIMS preparation.

Although JMP has a nice graphical user face and its learning curve is not as steep as that of Winsteps, RUMM, and Bilog, please keep in mind that JMP is a general-purposed statistical package rather than a specialized assessment tool, thus its IRT information is limited. Nevertheless, for those who want immediate reports without going through scripting and programming, JMP is a good start. Also, as a general package, JMP allows you to do other things, such as compute descriptive and inferential statistics, data mining, six-sigma, experimental design, and much more. In this write-up, some seemingly advanced statistical concepts will be discussed, but using JMP makes this an easy thing to do. For clarity of interpretation, this illustration uses only 10 items and 500 students from the Bio data set.

Import data into JMP

JMP can read a wide variety of file formats, including SAS, Excel, text, Minitab, dbase, Access …etc. To read a dataset, simply open the file from the file menu. However, if you import the file from Excel, please make sure that the item responses are defined as numeric in Excel. If there is a "'" in front of the number, the value is left-aligned, or if the cell has a green triangle on

the upper left corner, then the so-called “number” is actually defined as a character (see Figure 1). To make the data set ready for JMP, please ask your IT personnel to output it properly.

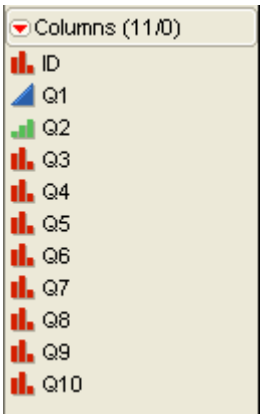
Figure 1. Numbers as characters in Excel

B2 '1

A	B	C	D	E	F	G	H	I	J	K	L
ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	
005090101	1	0	1	1	0	0	0	0	0	1	
005090719	1	0	1	1	0	1	0	0	1	1	
005090820	1	1	1	1	1	1	1	0	0	1	
005091440	1	1	1	1	1	1	0	1	1	1	
005091580	1	0	0	1	0	0	0	1	1	0	
005091880	1	1	1	1	1	1	0	0	1	1	

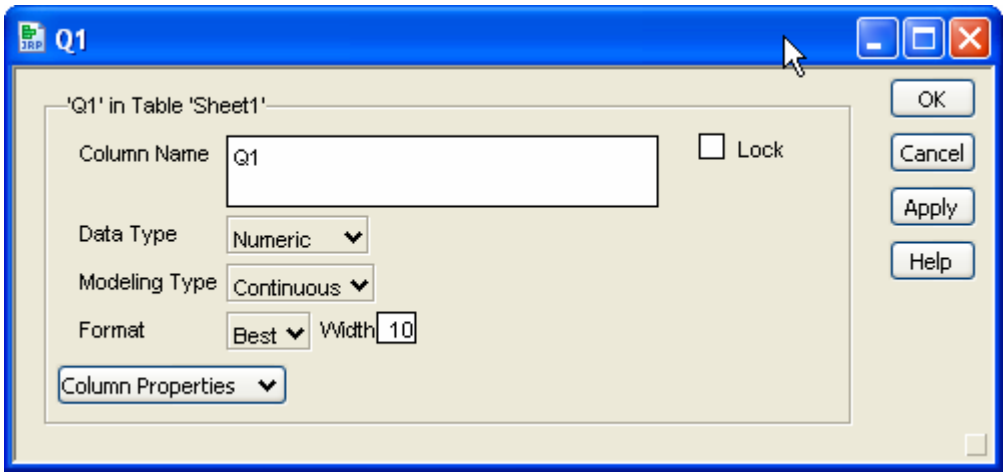
The second line of “defense” is inside JMP. If the data are not read properly, you can re-define the variable type in JMP. In JMP, there are three major variable types: nominal, ordinal, and continuous. Nominal data are depicted as a red bar chart, ordinal data are symbolized by a green histogram, and continuous-scaled data are represented as a blue curve (see Figure 2). Although binary item scores (1, 0) are discrete rather than continuous, they must be formatted as continuous in order to run IRT.

Figure 2. Variable types in JMP



To change the data type from “character” to “continuous numeric,” click on the variable header in the JMP spreadsheet, and make changes in the pop-up window as shown in Figure 3.

Figure 3. Change data type in JMP



Before any analysis is performed, one must verify that the data are clean. The easiest way to do so is simply plotting the distribution of each variable, including demographic variables and item responses. If there is any anomaly, you can spot it right away. For example, the gender plot should show male and female only. If there is a “third sex,” you have to go back to the original data to clean that up. By the same token, every item score is either “1” or “0.” If there is a “9,” chances are “9” denotes a missing value. To plot the data, go to **Analyze - Distribution**. Next, select the variables that you want to visualize to the Y box. You can select a block of variables by holding the shift key. Alternatively, you can select multiple variables by holding the control key (Figure 4a). Bar charts of gender and Question 1 are shown in Figure 4(b) and (c).

Figure 4 (a). Analyze the distributions

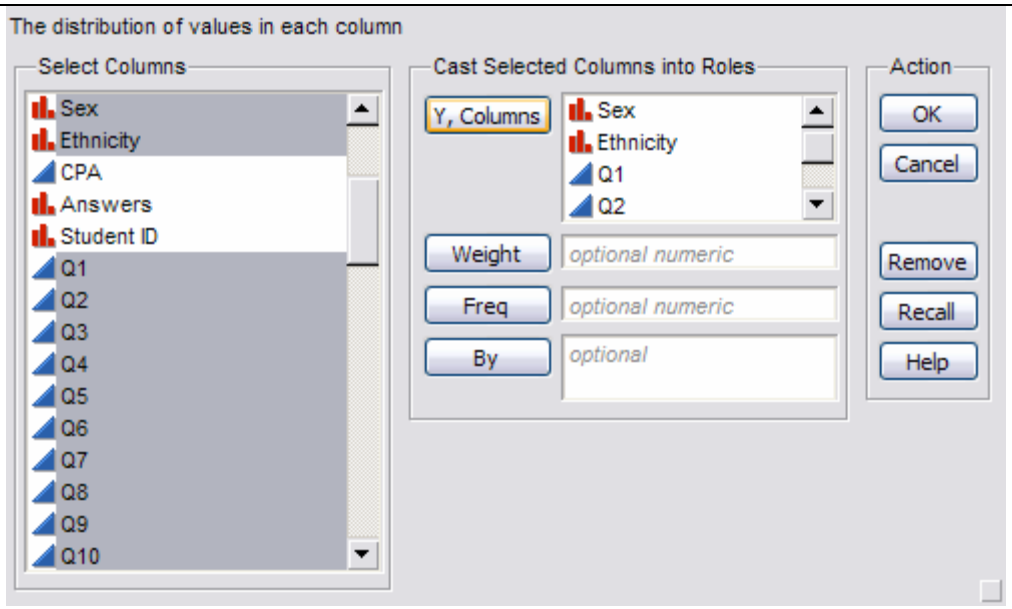


Figure 4(b) Barchart of gender

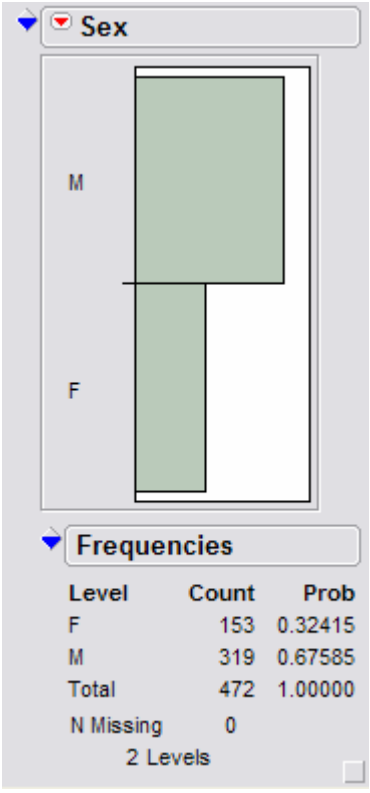
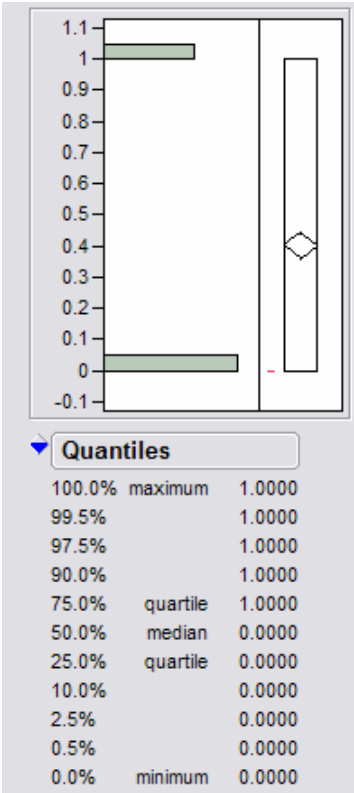
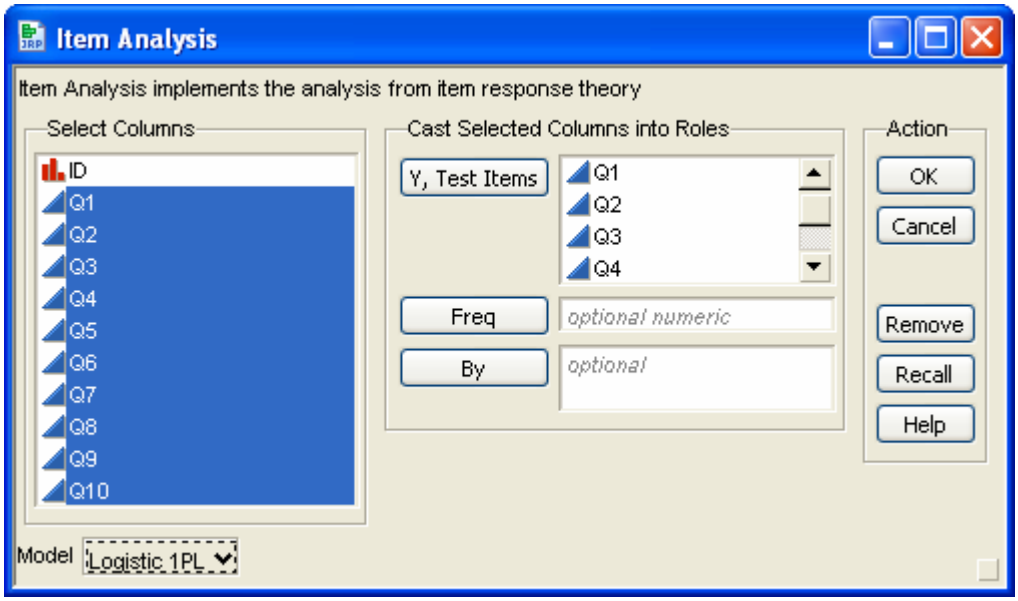


Figure 4(c) Barchart of Question 1



To run IRT, go to **Analyze – Multivariate methods – Item analysis**. Select all items, drag, and drop them into "**Y, Test Items**." In the model pop-up menu, select 1, 2, or 3 logistic model. As mentioned before, JMP is not as versatile as Winsteps, RUMM, and Bilog. You are confined to use logit and no option for probit is available. Next, click on OK (see Figure 5).

Figure 5. Item analysis pop-up window



In the resulting page, you can show or hide information by clicking on the blue triangle, and request more information by clicking on the red triangle. Figure 6a shows the item characteristic curves (ICC) and item information functions (IIF) for each item, Figure 6b shows each item difficulty parameter, and Figure 7 shows the test information function (TIF) for all items together. ICC tells you the probability of answering the item correctly at different levels of student ability, whereas IIF informs you how much reliable information about the student you can obtain at different levels of student ability. TIF is simply the sum of all IIFs in a test. For more information about ICC, IIF, and TIF, please consult http://www.creative-wisdom.com/multimedia/IRTTHA.htm or http://www.creative-wisdom.com/computer/sas/IRT.pdf Please note that in JMP each ICC has a vertical red line. The

red line shows the intersection of the probability and the ability when P = .5. In Question 1, students whose ability level is about -2 have 0.5 probability of answering the item correctly. In Question 2, students must have 0.5 ability in order to obtain 0.5 probability whereas in Question 3 it takes just -3.5 ability. In other words, if the item is easy, the red line leans toward left; if it is hard, it leans toward right. By looking at the location of the red line, the user can tell which item is a challenger and which one is a give-away.

Figure 6a. ICC and IIF

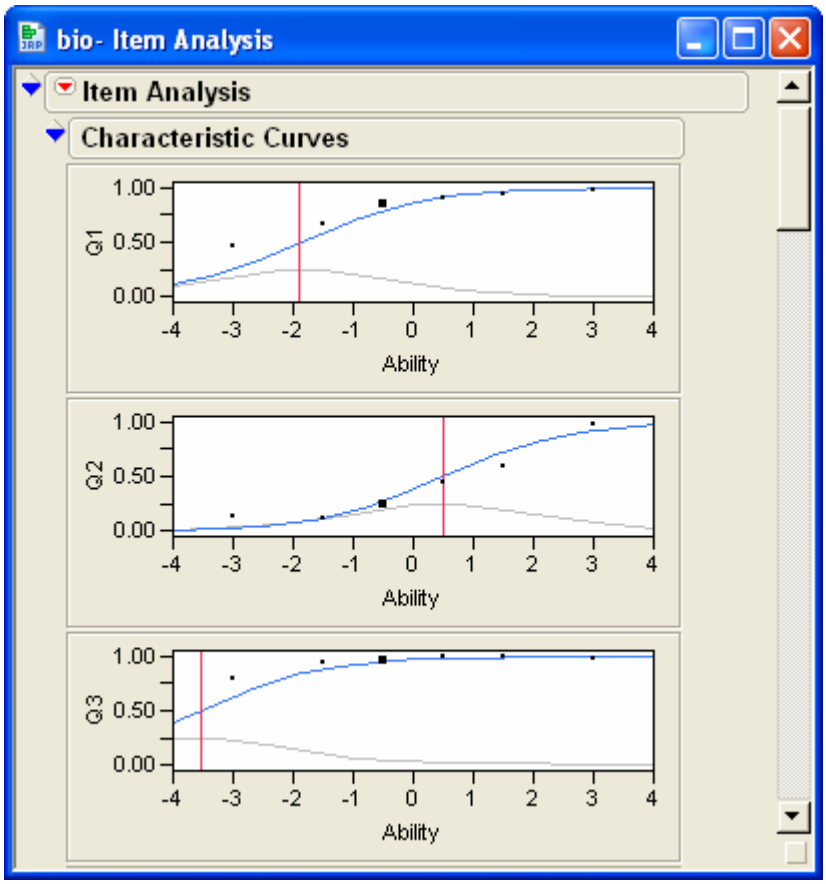


Figure 6b. Item difficulty parameters.

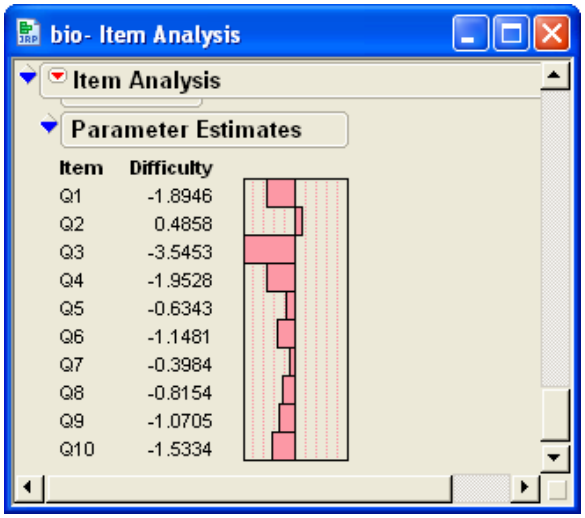


Figure 7. TIF.

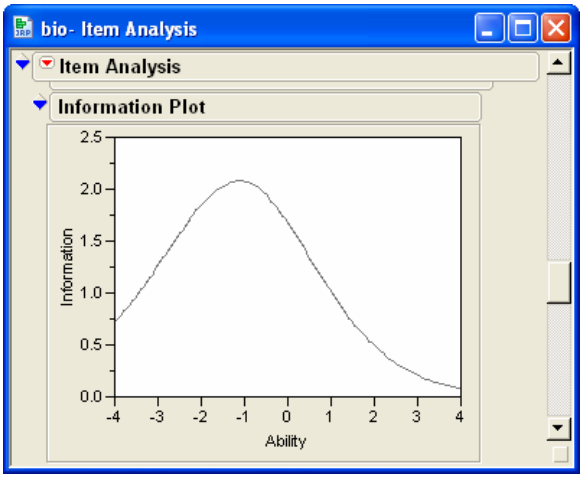
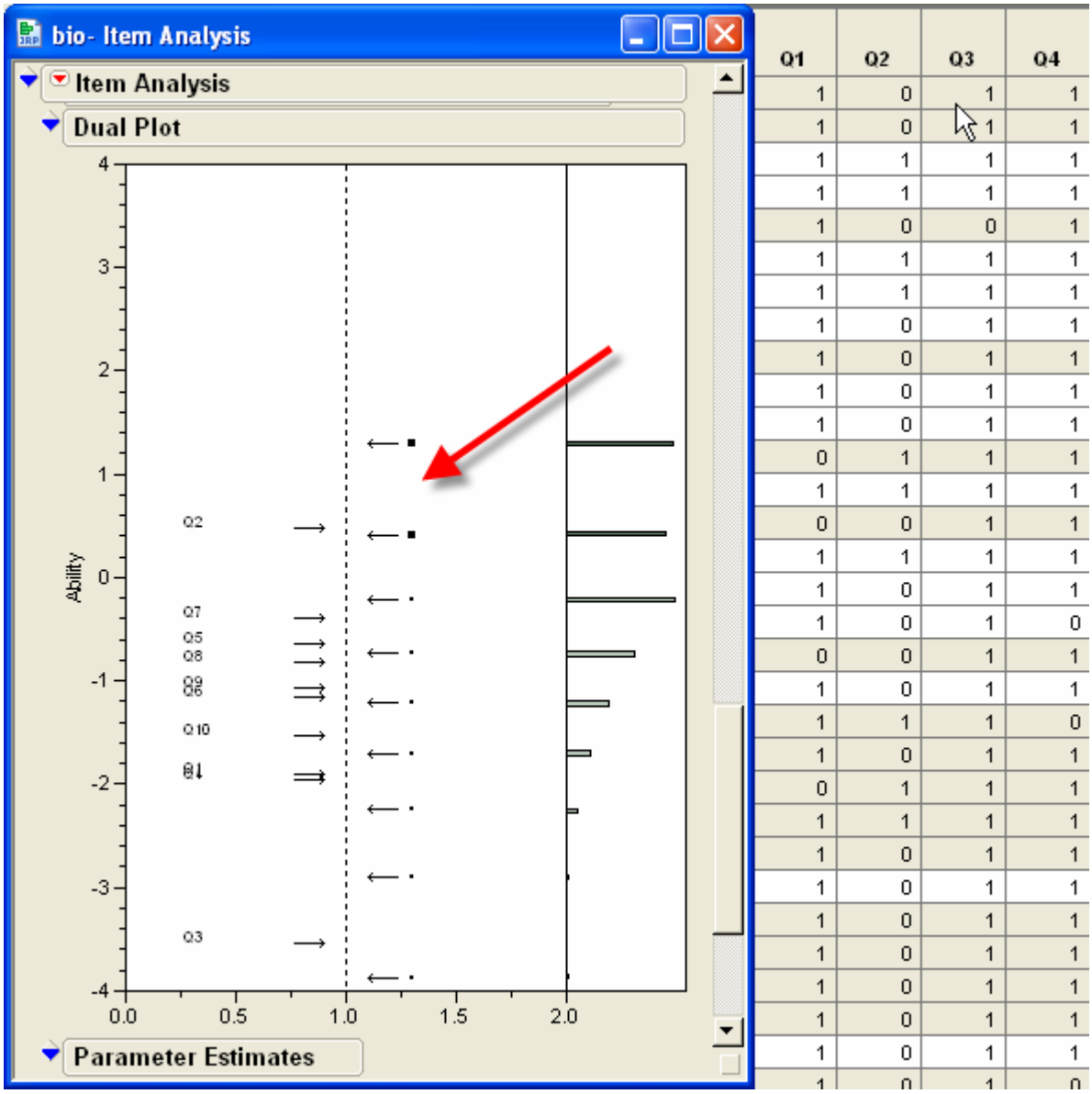


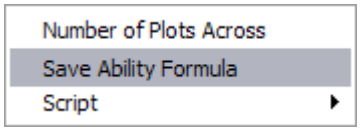
Figure 8 shows a dual plot, which is equivalent to the item-person map (IMP) in Rasch Unidimensional Measurement Modeling (RUMM). The attributes of all items and students are re-scaled in terms of logit, and therefore they can be compared side by side. JMP's graphs are dynamic and interactive. Logit is the natural log if the odds ratio. For more information, please consult the two websites mentioned in Page 5. If you want to identify the students who are above average (> 0), you can select the points and the corresponding rows in the spreadsheet are highlighted simultaneously.

Figure 8. Dual plot



Typically, the primary goal of IRT is to examine the quality of items. One should not make a firm judgment about student ability until items are validated. Nevertheless, one can conduct initial analysis using the ability estimates yielded from IRT modeling. To append the ability estimates to the original table, click the red triangle and choose **Save Ability Formula** (see Figure 9).

Figure 9. Save ability formula



After the ability estimates are saved, one can perform various exploratory analyses. For example, to detect whether there is a significant performance gap between different races, sexes, and school, one can use Fit Y and X by putting ability formula into Y and race, sex, school into X. To show the boxplots and the diamond plots, select Quantiles and Means/ANOVA from the red triangle (Figure 10).

Figure 10.

Select quantiles and means/ANOVA

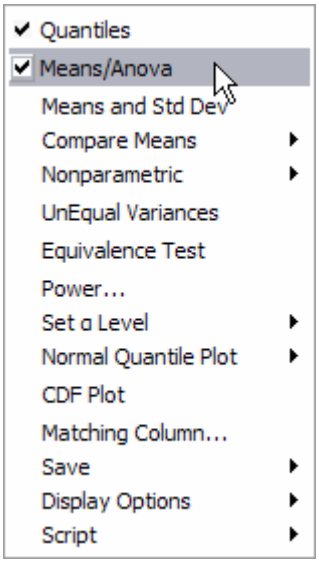
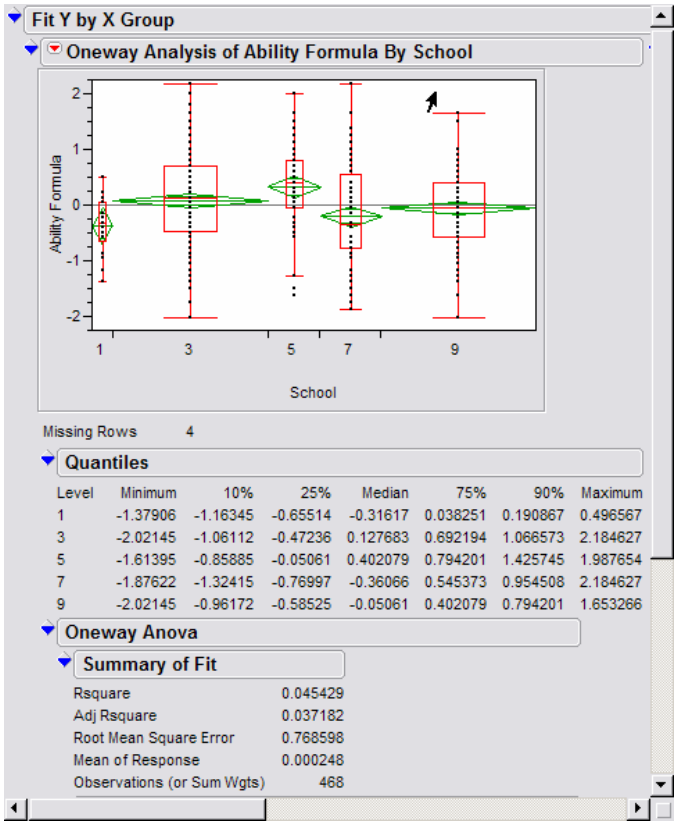


Figure 11 shows the results for school differences. Although parametric procedures such as t-tests and F-tests are widely used for between-group comparison, these procedures based upon centrality may mislead the researcher, especially in the case of heterogeneity of variance. As a remedy, more and more researchers endorse the use of confidence intervals (CI). By using CI, the researcher not only looks at the group differences by means, but also by variability. JMP provides a powerful tool named diamond plot to visualize variability, as demonstrated in Figure 11. It condenses a lot of important information:

- **Grand mean**: represented by a horizontal line. In IRT ability estimates, the mean is always zero.
- **Group means**: the horizontal line inside each diamond is the group means
- **Confidence intervals**: The diamond is the CI for each group
- **Quantile**: In addition to CI, JMP also provides the option of overlaying a boxplot showing quantile information

Figure 11. Boxplots and diamond plots for between-group comparison



The visualization of CI is straightforward--the flatter the diamond is, the tighter the CI is. If the two diamonds are not overlapped, which means the lower bound of one diamond is higher than the upper bound of another, then there *may be* a significant difference. It is important to point out that the probability of overlap is a function of the standard error. As the standard errors

become less homogeneous, the probability of overlap decreases. Thus, CI overlaps should be taken for exploratory purposes only; it could not be used as a replacement of hypothesis testing.

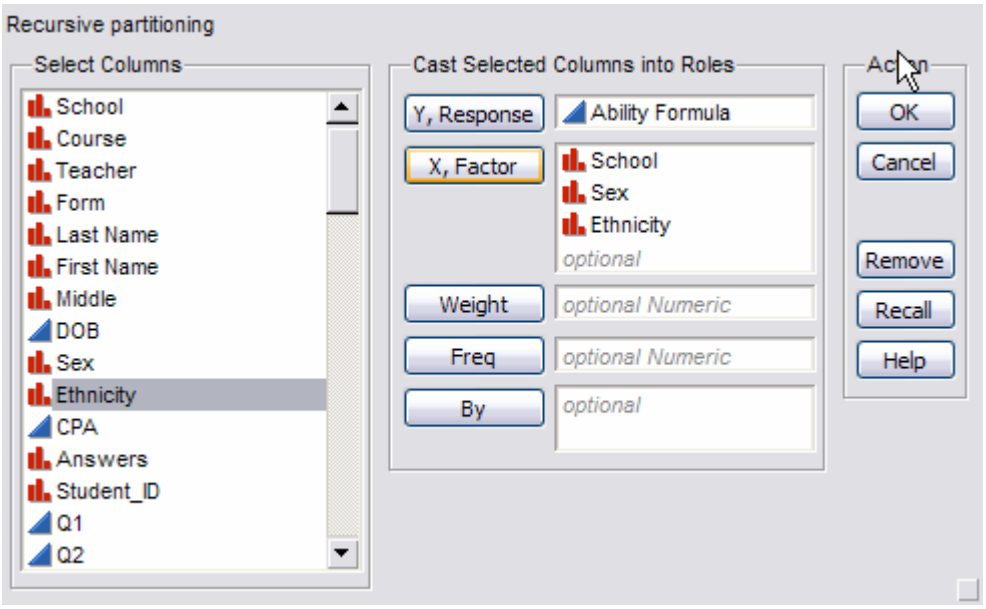
The above illustration shows the school effect only. There is also a gender effect and a race effect with respect to ability estimates. However, one may want to ask among the three variables which factor is considered most important. To answer this question, one can employ classification tree by selecting **Analyze-Modeling-Partition**.

Classification trees aim to find which independent variable(s) can make successively a decisive spilt of the data by dividing the original group of data into pairs of subgroups in the dependent variable. In programming terminology, a classification tree can be viewed as a set of "nested-if" logical statements. This example is commonly used for illustrating nested-if logic: When heart attack patients are admitted to a hospital, physiological measures, including heart rate, blood pressure, and background information are usually obtained. Subsequently, patients can be tracked to see if they survive the heart attack for a period of time. In order to determine which piece of information is most relevant to the survival of patients, a three-question decision tree was developed: What is the patient's minimum systolic blood pressure over the initial 24 hour period? What is his/her age? Does he/she display sinus tachycardia? The answers to these three questions can help the doctor to make a quick decision: "If the patient's minimum systolic blood pressure over the initial 24 hour period is greater than 91, if the patient's age is over 62.5 years, and if the patient displays sinus tachycardia, then and only then the patient is not predicted to survive for at least 30 days." These nested-if decisions can be translated into a leaf report showing the conditional probability of each combination of variables, and also into a graphical form as a tree structure. A classification tree can be somewhat complicated due to the size of the tree. But, graphical procedures can help to simplify interpretation for the final decision making.

Because classification trees can provide guidelines for decision-making, they are also known as decision trees. It is important to note that data mining focuses on pattern recognition, hence no probabilistic inferences and Type I error are involved.

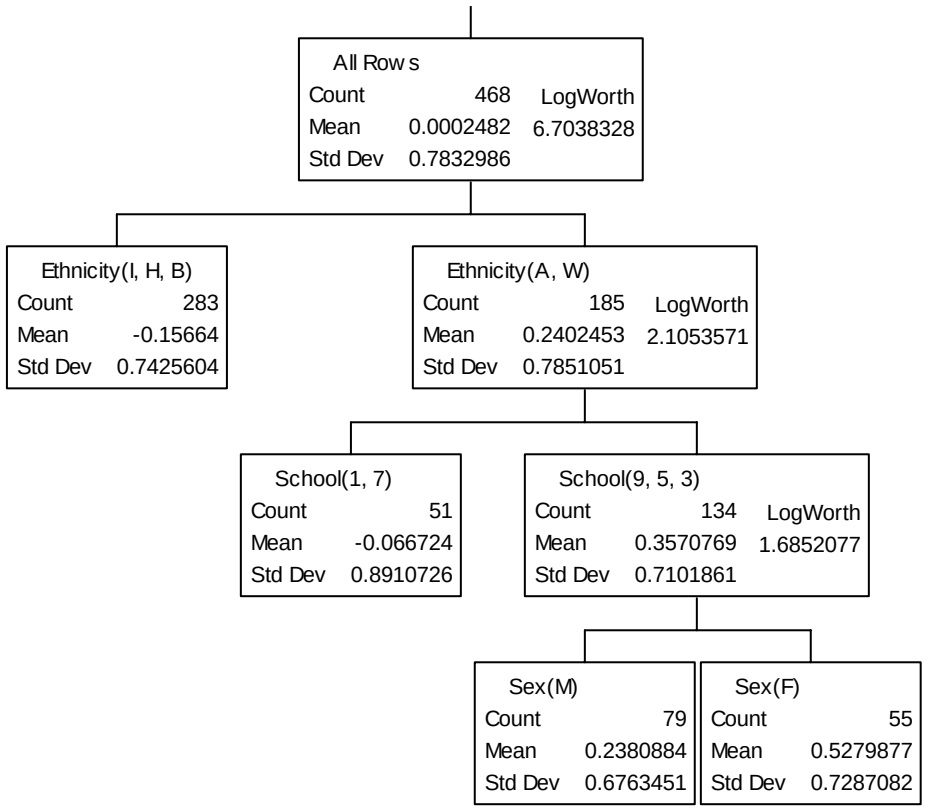
To construct a classification tree, in the dialog box ability formula should be put into Y while school, sex, and race should be placed inside X (Figure 12).

Figure 12. Recursive partitioning



Click the button “Spilt” three times and the importance of the variables are rank-ordered. According to the classification tree (Figure 13), the most crucial factor to ability is school, the second is race, and the third is gender. Based on the leaf report, one can conjecture that if a student is a white or Asian, and if the school is 9, 5, and 3, and the gender is female, it is most likely that the ability estimate is above average (0.5279).

Figure 13. Classification tree and leaf report



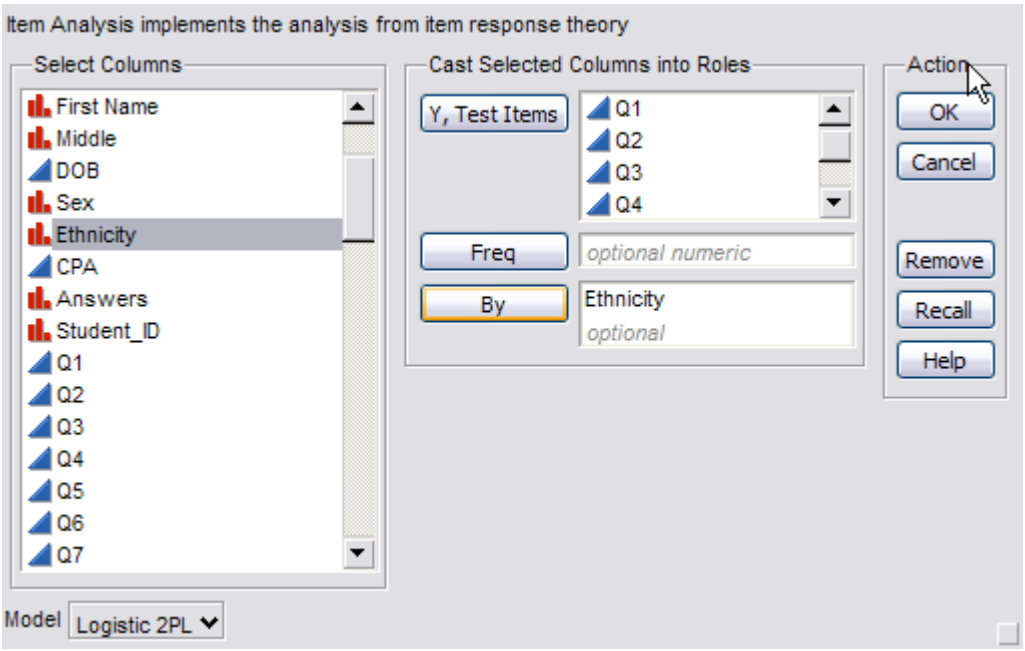
Leaf Report

Leaf Label	Mean	Count
Ethnicity(I, H, B)	-0.1566404	283
Ethnicity(A, W)&School(1, 7)	-0.0667239	51
Ethnicity(A, W)&School(9, 5, 3)&Sex(M)	0.23808836	79
Ethnicity(A, W)&School(9, 5, 3)&Sex(F)	0.52798772	55

Re-analysis by demographic variables

In this example, after the initial analysis, you may want to take different demographic variables into account while computing item attributes. For example, you may want to partition the data analysis by race to examine whether item parameters vary across different ethnic groups. To do so, you can simply re-run the analysis by dragging “Ethnicity” into the cell “By”, as shown in Figure 14.

Figure 14. IRT by race



Summary

Running IRT is no longer a technical work for psychometricians only. Unlike syntax-based IRT software packages, such as Bilog and Winsteps, JMP allows users to obtain quick results by pointing and clicking, dragging and dropping. More importantly, every report in JMP is presented in a graphical fashion; you can make accurate and meaningful interpretations without invoking numeric-based statistical reasoning. Happy JMPing!