Interpreting Multiple Regression

1. This is just one "Cliff Note" at best. There are a myriad of websites, textbooks, and research papers which give a much more comprehensive look at this, to wit:

http://www.statsoftinc.com/textbook/stmulreg.html#cresidual

As you read papers that employ OLS (Ordinary Least Squares) regression analysis, you will want to know three basic things:

1. Does the model explain a substantive proportion of the variance (around the variables in question) to a statistically significant degree?

The answer to the first part of this issue is found in the report of the regression coefficient, R Squared. The R^2 coefficient is some decimal proportion of 1, and, when multiplied by 100 expresses the percentage of the variance explained by the entire model. For example, if the R^2 for the model expressing, say, the relationship between religious practice and deviant behavior is .31, you can interpret that by saying, "The model explains 31% of the variance in deviance by these measures of religious practice." That also means that 69% of the variance in deviance is not explained by this model. Deciding whether this is substantive is a matter of subjective interpretation. However, if the coefficient is not statistically significant, it *cannot* be treated as substantively important.

The second issue is statistical significance. Statistical significance refers to a level of confidence that the percentage of the variance explained by the model did not happen by random chance. These levels of confidence are typically reported as 95% or 99% levels of certainty. If the R^2 noted above, .31, is statistically significant to a 99% confidence level, two asterisks typically follow the number. (R^2 = .31**). If the confidence level is 95%, the R^2 coefficient is followed by one asterisk (R^2 = .31*). If this level of confidence is not reached, the R^2 is typically not statistically significant and the meaning of the model is discounted. (Sometimes the confidence level is expressed in terms of error probability: 95% = 0.05; 99% = 0.01).

2. Which variables in the model contribute to the overall R^2, and are these contributions statistically significant?

Each variable in the model has an unstandardized *b* coefficient which is or is not a statistically significant contributor to the overall R^2. The independent variables in the model, as well as common control variables such as age, gender, race, or SES, contribute to the overall strength of R^2. Statistical significance is attached to each *b* coefficient in the same way as above. All this tells you is which variables in the model contribute to R^2 is a way that did not occur randomly. This is not a measure of the degree of contribution from each variable. Be very careful not to interpret a significant *b* if R^2 is not significant.

3. What is the relative proportion of contribution for each variable in the model?

Each variable in the model also has a Standardized Beta (β) coefficient attached to it. Standardization is a mathematical process which controls for differences in units of analysis (Z scores are calculated). The standardized β coefficients may then be compared to each other to glean the relative degree (vis-a-vis other variables) to which they support the strength of the overall model. In most cases, because of different units of analysis (income, age, education for example), you cannot compare *b's* so comparative contribution must be assessed using the standardized β. Note, however, that these are relative measures and not proportional expressions of percentages.