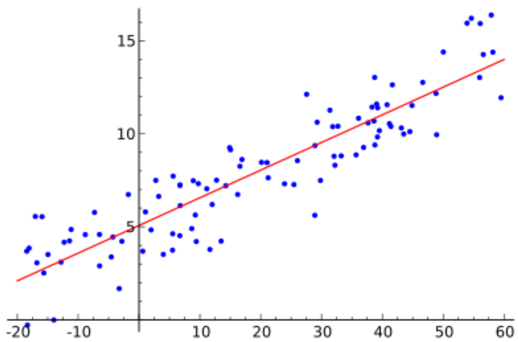
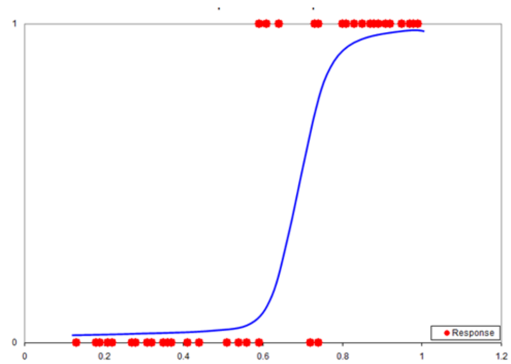
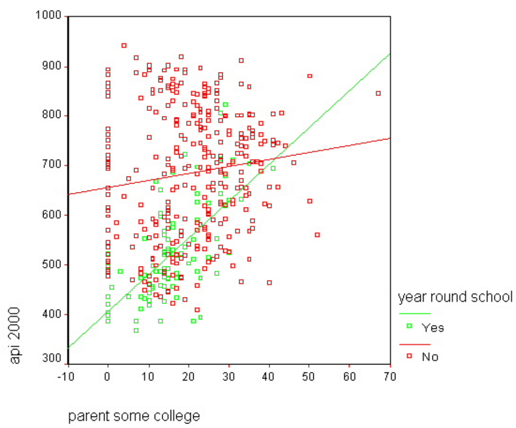
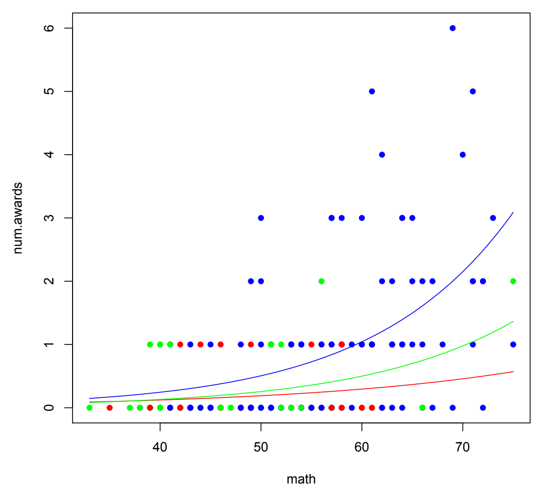
<table>
<tr><th>LINEAR MODELS</th><th>GENERALIZED LINEAR MODELS</th></tr>
<tr><td>


Simple linear regression with one continuous explanatory variable).
</td><td>


Logistic regression (binary response variable and one continuous explanatory variable)
</td></tr>
<tr><td>


Multiple linear regression model with one continuous variable and one categorical variable (2 levels).
</td><td>


Poisson regression with one continuous variable and one categorical variable (3 levels)
</td></tr>
</table>

<table>
<tr><th>LINEAR MODELS</th><th>GENERALIZED LINEAR MODELS</th></tr>
<tr><td>

Model statement:

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im} + \varepsilon_i$

Continuous response variable (normal distribution)

Predictor variables can be continuous and/or categorical

Can have transformations of x's and/or y to satisfy assumptions

Link function is identity link, with normal distribution (if you used a GLM to fit this)

</td><td>

Model statement:

$y_i = \mathbf{g^{-1}}\boldsymbol{(}\beta_1 x_{i1} + \ldots + \beta_m x_{im}\boldsymbol{)} + \varepsilon_i$

Response variable is typically **binary** (from binomial distribution) or **count** (from Poisson distribution)

Response probability distributions can be any of the exponential family of distributions (e.g.: *normal, inverse normal, gamma, binomial, negative binomial, Poisson, multinomial*)

Predictor variables can be continuous and/or categorical

Non-linear **link function** relates predictor variables and response variable (logit or probit link for binomial, log link for Poisson)

</td></tr>
<tr><td>

How model is fit

Least squares or maximum likelihood
Least squares: find parameter estimates such that the sum of squared differences between errors is the smallest (minimize SSE)

</td><td>

How model is fit

Maximum likelihood: find parameter estimates such that the likelihood of getting these y-values is the largest (maximize likelihood)

</td></tr>
<tr><td>

Features

Goes through the mean of all the x values and the y values
Sum of errors is zero; $\Sigma e_i = 0$

</td><td>

Features

1. Error structure
2. Link function
3. Linear predictor

</td></tr>
</table>

LINEAR MODELS	GENERALIZED LINEAR MODELS
Assumptions 1. Linearity between continuous predictor variables and response variable (*residual plot*) 2. errors are **independent** and identically distributed (**equal variance**/ homoskedastic, no correlation between errors due to space/time/etc.) and follow the **normal** distribution (*residual plot, heteroskedasticity tests, normality plot, normality tests, histogram of errors*) **Sampling assumptions** 1. Predictor variables are measured without error 2. Data represents a random sample from the population	**Assumptions** 1. Variances of y's is a function of the mean of y, given x (*residual plot* can be used to check for overdispersion) 2. Observations (errors) are independent **Sampling assumptions** 1. Predictor variables are measured without error 2. Data represents a random sample from the population
Hypothesis tests **F test** comparing full model to null model t tests to test contribution of each variable, given the other variables (equivalent to a partial **F test*** for that variable) *partial F tests can be used to test multiple variables at the same time by comparing nested models	**Hypothesis tests** **Likelihood ratio test** comparing full model to null model **Likelihood ratio test** comparing models with and without the variable of interest (can compare any set of nested models, therefore can test multiple variables at the same time)

LINEAR MODELS	GENERALIZED LINEAR MODELS
Measures of goodness of fit *Maximize* 1. $\mathbf{R^2}$ = SS reg / SSY = 1 – SSE/SSY 2. $\mathbf{R^2}$ **adjusted** $= 1 - \frac{(n-1)}{(n-m-1)} \frac{SSE}{SSY}$ If y was transformed: You will need to backtransform the predicted values into the original units. You can then calculate a Pseudo-R^2: Pseudo-R^2 = 1 – SSE/SSY	**Measures of goodness of fit** *Maximize* 1. $\text{Pseudo } R^2 = 1 - \left\{ \frac{L(0)}{L(\hat{\beta})} \right\}^{\frac{2}{n}}$ L(0) = likelihood of intercept only model 2. **Max-scaled** $\mathbf{R^2}$ $\tilde{R}^2 = \frac{R^2}{R^2 \max}$ where $R^2 \max = 1 - \{L(0)\}^{2/n}$
Minimize 1. **Root MSE** / Standard error of the estimate 2. **PRESS (Prediction sum of squares) statistic** – subset used to estimate parameters, then tested at prediction for remaining data 3. **Akaike's information criterion** (based on least squares) AIC = n ln(SSE) – nln(n) + 2p p = number of parameters n = sample size	*Minimize* 1. **Akaike's information criterion** $AIC = -2 \ln L + 2(k + s)$ Where k =levels for y – 1, and s is the number of predictor variables (including all dummy variables). AIC with correction for finite sample sizes $AICc = AIC + \frac{2k(k+1)}{n-k-1}$ n = sample size Penalty for extra parameters AICc converges to AIC when n is large 2. **Swartz's criterion (BIC)** $SC = -2 \ln L + (k + s) \ln(\sum_{j=1} f_j)$ f_j is the frequency for each proportion; sum of frequencies = n when single observations of y = 0 or 1 are used. Includes sample size in the form of frequencies
	Also see **classification table** (results for different probability cut-offs) for logistic regression