Nonlinear Least Squares: briefly...

Generalized Linear Models: analogs of regression, analysis of variance and analysis of covariance with non-normal response variables: Binomial, Poisson, Multinomial, Gamma, etc.

Hierarchical and Mixed Models: Coming soon...

Time Series:

\[ Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \epsilon \]

Highly technical, applications in Economics etc. Time permitting...

ETC! Sorry, we won't have time...
Binomial GLM: Logistic Regression

Given $X,$

$$Y \sim \text{Binomial}(n, p)$$

and

$$\log(\text{odds}) = \log \left( \frac{p}{1 - p} \right) = \beta_0 + \beta_1 X$$

or equivalently:

$$p = \frac{\text{odds}}{1 + \text{odds}} = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$
Logistic Curves with Different Slopes

\[ \log(\text{odds}) = \beta x \]
Logistic Curves with Different intercepts

\[
\log(\text{odds}) = B_0 + x
\]

- \(B_0 = 2\)
- \(B_0 = 0\)
- \(B_0 = -2\)
Odds
The odds for an event of probability $p$ are

$$odds = \frac{p}{1 - p}$$

The probability of an event with odds $a/b$ (often written $a$ to $b$ in favor, or $b$ to $a$ against) is

$$p = \frac{odds}{1 + odds} = \frac{a}{a + b}$$

Example: if $odds = 2$, or 2 to 1 in favor, then $p = \frac{2}{3}$, and

$$\frac{2/3}{1 - 2/3} = \frac{2/3}{1/3} = 2$$
Interpretation of Logistic Models

The Model:

\[ \log(\text{odds}) = \beta_0 + \beta_1 X \]

or

\[ \text{odds} = e^{\beta_0 + \beta_1 X} \]

or

\[ p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \]
Given two sets of conditional probabilities $\mathbb{P}(A|B)$ and $\mathbb{P}(A|B^c)$, the odds ratio is

$$OR = \frac{\text{odds}(A|B)}{\text{odds}(A|B^c)}$$

Thus the log of the odds ratio is

$$\log(\text{odds}(A|B)) - \log(\text{odds}(A|B^c))$$
The logistic model is

\[ \log(\text{odds}) = \beta_0 + \beta_1 X \]

Thus to compute the log of the odds ratio for two different values of \( X \), we subtract the linear model for the log-odds.

\[ \log(\text{odds}(X_1)) - \log(\text{odds}(X_2)) = \beta_1 (X_1 - X_2) \]

For a one unit change in \( X \), that’s just

\[ \log(\text{OR}) = \beta_1. \]
### Example: Orings

Coefficients:

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 6.89699  | 2.94427    | 2.343   | 0.01915  |
| temp           | -0.14212 | 0.04588    | -3.097  | 0.00195  |

# odds of failure at 32F
> exp(6.89699 -0.14212*32)
[1] 10.47666

# log odds ratio for failure at 32F vs 75F
> -0.14212*(32-75)
[1] 6.11116

# odds ratio for failure at 32F vs 75F
> exp(-0.14212*(32-75))
[1] 450.8614
\[ OR = e^\hat{\beta} \]

Simple Idea: exponentiate the CI for the log of the odds ratio:

\[ \hat{\beta} \pm SE(\hat{\beta}) \]

yielding

\[ e^{\hat{\beta} \pm SE(\hat{\beta})} \]
a CI for the Odds Ratio in R

\[ SE(\hat{\beta} \Delta X) = \Delta X SE(\hat{\beta}) \]

Coefficients:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 6.89699 | 2.94427 | 2.343 | 0.01915 |
| temp | -0.14212 | 0.04588 | -3.097 | 0.00195 |

\[
> 0.045 \times (75-32)
\]

[1] 1.935

# CI for odds ratio for failure at 32F vs 75F

\[
> \exp((-0.14212 + c(-1,1) \times 1.96 \times 0.04588) \times (32-75))
\]

[1] 21545.589263 9.434693

(If we use 75 − 32 we get the inverse odds ratio.)
Maximum Likelihood Estimation:
Choose the parameters ($\beta$’s) to maximize the probability of the observed data.
Binomial(n,p) Log-Likelihood: MLE for p

X = 10, N = 25, X/N = .4
Simple Example: Log-Likelihood

> Darwin.glm <- glm(Crossed > Self ~ 1, data = Darwin, family = binomial)
> summary(Darwin.glm)

Call:
glm(formula = Crossed > Self ~ 1, family = binomial)

Deviance Residuals:

            Min       1Q   Median       3Q      Max
-2.007 0.535 0.535 0.535 0.535

Coefficients:

                    Estimate Std. Error z value  Pr(>|z|)
(Intercept) 1.8718    0.7596 2.4647   0.0137
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11.78 on 14 df
Residual deviance: 11.78 on 14 df

Number of Fisher Scoring iterations: 4

Iterative searches can fail! Fortunately that is rare for logistic regression and other glm’s. Example?
# compare residual deviance to \(-2\times\text{logliklihood}\)
# Residual deviance was 11.78

```r
> a <- ifelse(with(Darwin, Crossed > Self), 1, 0)

> -2*sum(dbinom(a, 1, 13/15, log=TRUE))
[1] 11.78023
```
Definition:
residual deviance is $-2$ times the log of the probability of observing the data we got, assuming the estimated parameter values:

$$-2 \log(P(data|\hat{\beta})) = -2 \log(Likelihood(\hat{\beta}))$$

For the normal distribution, the deviance is almost the negative of residual sum of squares (RSS). We have been working with deviance already when we do least squares!
Analogy: Reductions in residual deviance are like reductions in RSS.

The `anova` function is designed to handle model comparisons for GLM’S too!

    anova(m1, m2, test = "Chisq")

The *Chisquared* test (aka $\chi^2$ test, *Likelihood Ratio Test*) is the natural analog of the F-test for linear models.
Anova Example

```r
> Anorexia.glm <- glm(Y > 0 ~ Therapy,
                      data=Anorexia, family=binomial)
> summary(Anorexia.glm)

Coefficients:

                  Estimate  Std. Error   z value  Pr(>|z|)
(Intercept)     -0.3102     0.3970    -0.781  0.4346
TherapyFamily   1.4888     0.6961     2.139  0.0324
TherapyCog/Behav 0.7455     0.5544     1.345  0.1787
```

Albyn Jones  Math 141
> anova(Anorexia.glm, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: Y > 0

Terms added sequentially (first to last)

<table>
<thead>
<tr>
<th>Df</th>
<th>Deviance</th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>70</td>
<td>96.716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapy</td>
<td>2</td>
<td>5.2193</td>
<td>68</td>
<td>91.497</td>
</tr>
</tbody>
</table>
Warnings!

All tests are approximate now!

Z tests are less reliable than the chisquared test done by the anova function!
Logistic regression is just regression with some minor changes!

Use the anova function to test hypotheses, don’t rely on the Z-tests unless you have a large sample size!