Overview

Model-based methods: Overview

Structure

- Explicitly assume some probability distribution for the data, e.g., binomial, Poisson, ...
- Distinguish between the systematic component—explained by the model—and a random component, which is not
- Allow a compact summary of the data in terms of a (hopefully) small number of parameters

Advantages

- Inferences: hypothesis tests and confidence intervals
- Can test individual model terms (anova())
- Methods for model selection: adjust balance between goodness-of-fit and parsimony
- Predicted values give model-smoothed summaries for plotting
- \( \implies \) Interpret the fitted model graphically

loglm vs. glm

With loglm() you can only test overall fit or difference between models

```r
berk.mod1 <- loglm(~ Dept * (Gender + Admit), data=UCBAdmissions)
berk.mod2 <- loglm(~(Admit + Dept + Gender)^2, data=UCBAdmissions)
anova(berk.mod2)
## Call:
## loglm(formula = ~(Admit + Dept + Gender)^2, data = UCBAdmissions)
## ## Statistics:
## X^2 df  P(> X^2)
## Likelihood Ratio 20.204 6 0.0011441
## Pearson 18.823 5 0.0020740
## LRstats(berk.mod1, berk.mod2)
## ## Likelihood summary table:
## AIC BIC LR Chisq Df Pr(>Chisq)
## berk.mod1 217.38 217.38 21.736 6 0.0011 **
## berk.mod2 217.240 217.240 20.2 5 0.0011 **
## ---
## Signif. codes: 0 '***' 0.001 '** ' 0.01 ' ' 0.05 ' . ' 1
```

Comparing models with anova() and LRstats()

```
anova(berk.mod1, berk.mod2, test="Chisq")
## LR tests for hierarchical log-linear models
## ## Model 1:
## ~Dept * (Gender + Admit)
## Model 2:
## ~(Admit + Dept + Gender)^2
## ## Deviance df Delta(Dev) Delta(df) P(> Delta(Dev)
## Model 1 21.736 6 1.5312 1 0.21593
## Model 2 20.204 5 0.0011441 0.0011441
## Saturated 0.000 0 20.2043 5 0.0011441
```
Overview

Model-based methods

With `glm()` you can test individual terms with `anova()` or `car::Anova()`

```r
berkeley <- as.data.frame(UCBAdmissions)
berk glm2 <- glm(Freq ~ (Dept+Gender+Admit)^2, data=berkeley,
  family="poisson")
anova(berk glm2, test="Chisq")
```

### Analysis of Deviance Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Deviance</th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>23</td>
<td>2650</td>
<td>18</td>
<td>2491</td>
</tr>
<tr>
<td>Dept</td>
<td>5</td>
<td>160</td>
<td>17</td>
<td>2328</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>813</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Admit</td>
<td>1</td>
<td>2098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dept:Gender</td>
<td>5</td>
<td>1221</td>
<td>11</td>
<td>877</td>
</tr>
<tr>
<td>Dept:Admit</td>
<td>5</td>
<td>855</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Gender:Admit</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>0.22</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fitting and graphing: Overview

Object-oriented approach in R:

- Fit model `obj <- glm(...)` → a `model object`
- `print(obj)` and `summary(obj)` → numerical results
- `anova(obj)` and `Anova(obj)` → tests for model terms
- `update(obj), add1(obj), drop1(obj)` for model selection

Plot methods:
- `plot(obj)` often gives diagnostic plots
- Other plot methods:
  - Mosaic plots: `mosaic(obj)` for "loglm" and "glm" objects
  - Effect plots: `plot(Effect(obj))` for nearly all linear models
  - Influence plots (car): `influencePlot(obj)` for "glm" objects

Objects and methods

How this works:

- Model objects have a "class" attribute:
  - `loglm()`: "loglm"
  - `glm()`: c("glm", "lm") — inherits also from `lm()`
- Class-specific methods have names like `method.class`, e.g.,
  - `plot.glm()`, `mosaic.loglm()`
- Generic functions (print(), summary(), plot()...) call the appropriate method for the class

```r
arth.mod <- glm(Better ~ Age + Sex + Treatment, data=Arthritis)
class(arth.mod)
```

```r
# [1] "glm" "lm"
```

Methods for "glm" objects:

```r
library(MASS); library(vcdExtra)
methods(class="glm")
```

```r
# [1] add1 addterm anova
# [4] Anova asGnm assoc
# [7] avPlot Boot bootCase
# [10] cereplot coefci coeftest
# [13] coerce confidenceEllipse confint
# [16] cooks.distance deviance drop1
# [19] dropterm effects extractAIC
# [22] family formula gamma.shape
# [25] influence initialize leveragePlot
# [28] linearHypothesis logLik mmp
# [31] model.frame modFit mosaic
# [34] ncvTest nosbs predict
# [37] print profile qPlot
# [40] residualPlot residualPlots residuals
# [43] rstandard rstudent show
# [46] sieve sigmaHat slotsFromS3
# [49] summary vcov weights
# see '?methods' for accessing help and source code
```
**Object and methods**

Logistic regression models

- **Response variable**: success/failure, vote: yes/no
- **Binomial data**: x successes in n trials (grouped data)
- **Ordinal response**: none < some < severe depression
- **Polytomous response**: vote Liberal, Tory, NDP, Green

**Explanatory variables**

- **Quantitative regressors**: age, dose
- **Transformed regressors**: \( \sqrt{\text{age}}, \log(\text{dose}) \)
- **Polynomial regressors**: \( \text{age}^2, \text{age}^3 \)
- **Categorical predictors**: treatment, sex (dummy variables, contrasts)
- **Interaction regressors**: treatment \( \times \) age, sex \( \times \) age

This is exactly the same as in classical ANOVA, regression models

**Arthritis treatment data**

- The response variable, Improved is **ordinal**: "None" < "Some" < "Marked"
- A binary logistic model can consider just **Better = (Improved>"None")**
- **Other important predictors**: Sex, Treatment
- **Main Q**: how does treatment affect outcome?
- **How does this vary with Age and Sex?**
- **This plot shows the binary observations, with several model-based smoothings**

**Association models**

- **Loglinear models**
  - (Contingency table form)
  - [Admit][GenderDept]
  - [AdmitDept][GenderDept]
  - [AdmitDept][GenderDept][GenderDept]

- **Poisson GLMs**
  - (Frequency data frame)
  - Freq ~ Admit + Gender*Dept
  - Freq ~ Admit*Dept + Gender*Dept
  - Freq ~ right+left+Diag(right+left)
  - Freq ~ right+left+Symm(right+left)

**Response models**

- **Binary response**
  - Categorical predictors: Logit models
  - \( \logit(\text{Admit}) \sim 1 \)
  - \( \logit(\text{Admit}) \sim \text{Dept} \)
  - \( \logit(\text{Admit}) \sim \text{Dept} + \text{Gender} \)

- **Continuous/mixed predictors**
  - Logistic regression models
  - \( P(\text{Admit}) \sim \text{Dept} + \text{Age} + \text{GRE} \)

- **Polytomous response**
  - Ordinal: proportional odds model
  - \( \text{Improve} \sim \text{Age} \times \text{Sex} \times \text{Treatment} \)
  - General: multinomial model
  - \( \text{WomenWork} \sim \text{Kids} + \text{HusbandInc} \)
Berkeley admissions data

- Admit/Reject can be considered a **binomial response** for each Dept and Gender.
- Logistic regression here is analogous to an ANOVA model, but for log odds(Admit).
- (With categorical predictors, these are often called logit models)
- Every such model has an equivalent **loglinear** model form.
- This plot shows fitted logits for the main effects model, Dept + Gender.

Survival in the Donner Party

- Binary response: **survived**
- Categorical predictors: **sex**, **family**
- Quantitative predictor: **age**
- Q: Is the effect of age linear?
- Q: Are there interactions among predictors?

**Binary response: What’s wrong with OLS?**

- For a binary response, \( Y \in (0, 1) \), want to predict \( \pi = \Pr(Y = 1 \mid x) \)
- A **linear probability model** uses classical linear regression (OLS)
- Problems:
  - Gives predicted values and CIs outside \( 0 \leq \pi \leq 1 \)
  - Homogeneity of variance is violated: \( \forall(\hat{\pi}) = \hat{\pi}(1 - \hat{\pi}) \neq \text{constant} \)
  - Inferences, hypothesis tests are wrong!

**OLS vs. Logistic regression**

- OLS regression:
  - Assume \( y(x) \sim N(0, \sigma^2) \)
- Logistic regression:
  - Assume \( \Pr(y=1 \mid x) \sim \text{binomial}(p) \)
Logistic regression

- Logistic regression avoids these problems
- Models \( \logit(\pi_i) \equiv \log[\pi_i/(1 - \pi_i)] \)
- \( \logit \) is interpretable as “log odds” that \( Y = 1 \)
- A related probit model gives very similar results, but is less interpretable
- For \( 0.2 \leq \pi \leq 0.8 \) fitted values are close to those from linear regression.

Logistic regression models: Multiple predictors

- For a binary response, \( Y \in (0, 1) \), let \( x \) be a vector of \( p \) regressors, and \( \pi_i \) be the probability, \( \Pr(Y = 1 \mid x) \).
- The logistic regression model is a linear model for the log odds, or \( \logit \) that \( Y = 1 \), given the values in \( x \),
  \[
  \logit(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + x_i^T \beta
  \]
- An equivalent (non-linear) form of the model may be specified for the probability, \( \pi_i \), itself,
  \[
  \pi_i = \{1 + \exp[-(\alpha + x_i^T \beta)]\}^{-1}
  \]
- The logistic model is also a multiplicative model for the odds of “success,”
  \[
  \frac{\pi_i}{1 - \pi_i} = \exp(\alpha + x_i^T \beta) = \exp(\alpha) \exp(x_i^T \beta)
  \]
- Increasing \( x_{ij} \) by 1 increases \( \logit(\pi_i) \) by \( \beta_j \), and multiplies the odds by \( e^{\beta_j} \).

Logistic regression: One predictor

The coefficients of this model have simple interpretations in terms of odds and log odds:
- The odds can be expressed as a multiplicative model
  \[
  \text{odds}(Y = 1) = \frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) = e^\alpha (e^\beta)^x .
  \]
- \( \beta \) is the change in the log odds associated with a unit increase in \( x \).
- The odds are multiplied by \( e^\beta \) for each unit increase in \( x \).
- \( \alpha \) is log odds at \( x = 0 \); \( e^\alpha \) is the odds of a favorable response at this \( x \)-value.
- In R, use \( \exp(\text{coef(obj)}) \) to get these values.
- Another interpretation: In terms of probability, the slope of the logistic regression curve is \( \beta \pi(1 - \pi) \)
- This has the maximum value \( \beta/4 \) at \( \pi = \frac{1}{2} \)

For a single quantitative predictor, \( x \), the simple linear logistic regression model posits a linear relation between the log odds (or \( \logit \)) of \( \Pr(Y = 1) \) and \( x \),
  \[
  \logit[\pi(x)] = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x .
  \]
- When \( \beta > 0 \), \( \pi(x) \) and the log odds increase as \( x \) increases; when \( \beta < 0 \) they decrease with \( x \).
- This model can also be expressed as a model for the probabilities \( \pi(x) \)
  \[
  \pi(x) = \log^{-1}[\pi(x)] = \frac{1}{1 + \exp[-(\alpha + \beta x)]}
  \]
Fitting the logistic regression model

Logistic regression models are the special case of **generalized linear models**, fit in R using `glm(..., family=binomial)`

For this example, we define Better as any improvement at all:

```r
data("Arthritis", package="vcd")
Arthritis$Better <- as.numeric(Arthritis$Improved > "None")
```

Fit and print:

```r
arth.logistic <- glm(Better ~ Age, data=Arthritis, family=binomial)
arth.logistic
```

The `summary()` method gives details:

```r
summary(arth.logistic)
```

Interpreting coefficients

```r
coef(arth.logistic)
```

- **Intercept**: -2.6421
- **Age**: 0.0492

```r
exp(coef(arth.logistic))
```

- **Intercept**: 0.0712
- **Age**: 1.0505

```r
exp(10*coef(arth.logistic)[2])
```

- **Age**: 1.6364

Interpretations:

- **log odds(Better)** increase by $\beta = 0.0492$ for each year of age
- **odds(Better)** multiplied by $e^\beta = 1.05$ for each year of age— a 5% increase
- over 10 years, odds(Better) are multiplied by $exp(10 \times 0.0492) = 1.64$, a 64% increase.
- **Pr(Better)** increases by $\beta/4 = 0.0123$ for each year (near $\pi = \frac{1}{2}$)

Multiple predictors

The main interest here is the effect of Treatment. Sex and Age are **control variables**. Fit the main effects model (no interactions):

$$
\text{logit}(\pi_i) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3
$$

where $x_1$ is Age and $x_2$ and $x_3$ are the factors representing Sex and Treatment, respectively. R uses dummy (0/1) variables for factors.

- $\alpha$: doesn’t have a sensible interpretation here. Why?
- $\beta_1$: increment in log odds(Better) for each year of age.
- $\beta_2$: difference in log odds for male as compared to female.
- $\beta_3$: difference in log odds for treated vs. the placebo group.
Multiple predictors: Fitting

Fit the main effects model. Use $I(Age-50)$ to center Age, making $\alpha$ interpretable.

```
arth.logistic2 <- glm(Better ~ I(Age-50) + Sex + Treatment,
data=Arthritis, family=binomial)
```

```
library(lmtest)
coeftest(arth.logistic2)
##
## z test of coefficients:
##
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -0.5781    0.3674  -1.57  0.116
## I(Age - 50)         0.0487     0.0207   2.36  0.018 *
## SexMale             -1.4878    0.5948  -2.50  0.012 *
## TreatmentTreated    1.7598     0.5365   3.28  0.001 **
## ---
## Signif. codes: 0 ' ***' 0.001 ' **' 0.01 ' *' 0.05 '.' 0.1 ' ' 1
```

```
cbind(coef=coef(arth.logistic2),
      OddsRatio=exp(coef(arth.logistic2)), exp(confint(arth.logistic2)))
##
##          coef OddsRatio 2.5 % 97.5 %
## (Intercept)   -0.5781  0.561  0.2647  1.132
## I(Age - 50)   0.0487  1.050  1.0100  1.096
## SexMale      -1.4878  0.226  0.0652  0.689
## TreatmentTreated 1.7598  5.811 2.1187 17.727
```

- $\alpha = -0.578$: At age 50, females given placebo have odds(Better) of $e^{-0.578} = 0.56$.
- $\beta_1 = 0.0487$: Each year of age multiplies odds(Better) by $e^{0.0487} = 1.05$, a 5% increase.
- $\beta_2 = -1.49$: Males $e^{-1.49} = 0.26 \times$ less likely to show improvement as females. (Or, females $e^{1.49} = 4.437 \times$ more likely than males.)
- $\beta_3 = 1.76$: Treated $e^{1.76} = 5.81 \times$ more likely Better than Placebo

Interpreting coefficients

```
coeftest() in lmtest gives just the tests of coefficients provided by
summary():
```

Hypothesis testing: Questions

- **Overall test**: How does my model, $	ext{logit}(\pi) = \alpha + \mathbf{x}^T \beta$ compare with the null model, $	ext{logit}(\pi) = \alpha$?

  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

- **One predictor**: Does $x_k$ significantly improve my model? Can it be dropped?

  $H_0 : \beta_k = 0$ given other predictors retained

- **Lack of fit**: How does my model compare with a perfect model (saturated model)?

For ANOVA, regression, these tests are carried out using $F$-tests and $t$-tests. In logistic regression (fit by maximum likelihood) we use

- $F$-tests $\rightarrow$ likelihood ratio $\chi^2$ tests
- $t$-tests $\rightarrow$ Wald $z$ or $\chi^2$ tests

Maximum likelihood estimation

- Likelihood, $\mathcal{L} = \Pr(\text{data} | \text{model})$, as function of model parameters

  For case $i$,

  $$
  \mathcal{L}_i = \begin{cases} 
  p_i & \text{if } Y_i = 1 \\
  1-p_i & \text{if } Y_i = 0 
  \end{cases} = p_i^{Y_i}(1-p_i)^{1-Y_i} \quad \text{where } \quad p_i = \frac{Y_i}{1+(1+\exp(\mathbf{x}_i \beta))}
  $$

  Under independence, joint likelihood is the product over all cases

  $$
  \mathcal{L} = \prod_i^n p_i^{Y_i}(1-p_i)^{1-Y_i}
  $$

  $\implies$ Find estimates $\hat{\beta}$ that maximize log $\mathcal{L}$. Iterative, but this solves the “estimating equations”

  $$
  \mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\beta}
  $$
Overall test

- Likelihood ratio test (G^2)
  - Compare *nested* models, similar to incremental F tests in OLS
  - Let \( \mathcal{L}_1 \) = maximized likelihood for our model
    \( \logit(\pi_i) = \beta_0 + \mathbf{x}_i^T \mathbf{\beta} \) \( \text{w/ k} \) predictors
  - Let \( \mathcal{L}_0 \) = maximized likelihood for null model
    \( \logit(\pi_i) = \beta_0 \) \( \text{under} \ H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \)
  - Likelihood-ratio test statistic:
    \[
    G^2 = -2 \log \left( \frac{\mathcal{L}_0}{\mathcal{L}_1} \right) = 2(\log \mathcal{L}_1 - \log \mathcal{L}_0) \sim \chi^2_k
    \]

Wald tests and confidence intervals

- Analogous to t-tests in OLS
- \( H_0 : \beta_i = 0 \)
  \[
  z = \frac{b_k}{s(b_k)} \sim \mathcal{N}(0,1) \quad \text{or} \quad z^2 \sim \chi^2
  \]
- Confidence interval:
  \[
  b_k \pm z_{1-\alpha/2} s(b_k)
  \]

Visualizing

Plotting logistic regression data

Plotting a binary response together with a fitted logistic model can be difficult because the 0/1 response leads to much overplotting.

- Need to jitter the points
- Useful to show the fitted logistic curve
- Confidence band gives a sense of uncertainty
- Adding a non-parametric (loess) smooth shows possible nonlinearity
- NB: Can plot either on the response scale (probability) or the link scale (logit) where effects are linear
Types of plots

- **Conditional plots**: Stratified plot of Y or logit(Y) vs. one X, conditioned by other predictors—only that subset is plotted for each
- **Full-model plots**: plots of fitted response surface, showing all effects; usually shown in separate panels
- **Effect plots**: plots of predicted effects for terms in the model, averaged over predictors not involved in a given term.

Conditional plots with ggplot2

Plot of Arthritis treatment data, by Treatment (ignoring Sex)

```r
library(ggplot2)
gg <- ggplot(Arthritis, aes(Age, Better, color=Treatment)) +
  xlim(5, 95) + theme_bw() +
geom_point(position = position_jitter(height = 0.02, width = 0)) +
stat_smooth(method = "glm", family = binomial, alpha = 0.2,
aes(fill=Treatment), size=2.5, fullrange=TRUE)
gg # show the plot
```

Full-model plots

Full-model plots show the fitted values on the logit scale or on the response scale (probability), usually with confidence bands. This often requires a bit of custom programming.

Steps:

- Obtain fitted values with `predict(model, se.fit=TRUE)`—`type="link"` (logit) is the default
- Can use `type="response"` for probability scale
- Join this to your data (```cbind()```)
- Plot as you like: ```plot()```, ```ggplot()```, ...

```r
arth.fit2 <- cbind(Arthritis, predict(arth.logistic2, se.fit = TRUE))
head(arth.fit2[-9], 4)
```

The data is too thin for males to estimate each regression separately
**Full-model plots**

Plotting on the logit scale shows the additive effects of age, treatment and sex.

These plots show the data (jittered) as well as model uncertainty (confidence bands).

**Models with interactions**

Allow an interaction of Age x Sex

```
arth.logistic4 <- update(arth.logistic2, . ~ . + Age:Sex)
library(car)
Anova(arth.logistic4)
```

```
## Analysis of Deviance Table (Type II tests)

## Response: Better
##             LR Chisq Df Pr(>Chisq)
## I(Age - 50)     0.0        0
## Sex          6.981 1 0.00823 **
## Treatment   11.901 1 0.00056 ***
## Sex:Age    3.421 1 0.06430 .
## ---
## Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 .’ 1
```

Interaction is NS, but we can plot it the model anyway.

```
```

These plots show the data (jittered) as well as model uncertainty (confidence bands).
Models with interactions

The \texttt{visreg} package

- Provides a more convenient way to plot model results from the model object
- A consistent interface for linear models, generalized linear models, robust regression, etc.
- Shows the data as partial residuals or rug plots
- Can plot on the response or logit scale
- Can produce plots with separate panels for conditioning variables

- Only the model changes
- \texttt{predict()} automatically incorporates the revised model terms
- Plotting steps remain the same
- This interpretation is quite different!

\begin{verbatim}
library(visreg)
visreg(arth.logistic2, ylab="logit(Better)", ...)
\end{verbatim}

Effect plots: basic ideas

Show a given effect (and low-order relatives) controlling for other model effects.

Data

\begin{verbatim}
  x1  x2  sex  x3\times2  y  yhat
  1   1   F   1  1.1  4.73  4.66
  2   1   M  0  2  5.10  5.55
  3   1   F  0  4  6.10  5.34
  4   1   F  1  4.84  4.46
  5   2   F  0  4.73  4.40
  ... 
  29  2   H  0  6.10  6.15
  30  3   F  1  6.71  7.14
\end{verbatim}

- Fit data: $x\beta \Rightarrow \hat{y}$
- Score data $x \cdot \hat{\beta} \Rightarrow \hat{y}^*$
  - plot vars: vary over range
  - control vars: fix at means

Score data

\begin{verbatim}
  x1  x2  sex  x3\times2  y  yhat*
  31  1   1   0.5  1  5.03  5.03
  32  2   1   0.5  2  4.97  4.97
  33  3   1   0.5  3  4.91  4.91
  34  1   2   0.5  2  5.43  5.43
  35  2   2   0.5  4  5.97  5.97
  36  3   2   0.5  6  7.31  7.31
\end{verbatim}

- One plot for each variable in the model
- Other variables: continuous— held fixed at median; factors— held fixed at most frequent value
- Partial residuals ($r_j$): the coefficient $\hat{\beta}_j$ in the full model is the slope of the simple fit of $r_j$ on $x_j$. 

{41/63}{42/63}
Effect plots for generalized linear models: Details

- For simple models, full model plots show the complete relation between response and all predictors.
- Fox(1987)— For complex models, often wish to plot a specific main effect or interaction (including lower-order relatives)— controlling for other effects

  - Fit full model to data with linear predictor (e.g., logit) \( \eta = X\beta \) and link function \( g(\mu) = \eta \rightarrow \) estimate \( \hat{b} \) of \( \beta \) and covariance matrix \( \hat{V}(\hat{b}) \) of \( \hat{b} \).
  - Construct “score data”
    - Vary each predictor in the term over its range
    - Fix other predictors at “typical” values (mean, median, proportion in the data)
    - \( \rightarrow \) “effect model matrix,” \( X^* \)
  - Use \texttt{predict()} on \( X^* \)
    - Calculate fitted effect values, \( \hat{\eta}^* = X^* \hat{b} \).
    - Standard errors are square roots of \( \text{diag} X^* \hat{V}(\hat{b}) X^*^T \)
  - Plot \( \hat{\eta}^* \), or values transformed back to scale of response, \( g^{-1}(\hat{\eta}^*) \).

- Note: This provides a general means to visualize interactions in all linear and generalized linear models.

Full model plots:

\texttt{arth.full <- Effect(c("Age", "Treatment", "Sex"), arth.logistic2)}
\texttt{plot(arth.full, multiline=TRUE, ci.style="bands", colors = c("red", "blue"), lwd=3, ...)}

Model with interaction of Age x Sex

\texttt{plot(allEffects(arth.logistic4), rows=1, cols=3)}

- Only the high-order terms for Treatment and Sex*Age need to be interpreted
- (How would you describe this?)
- The main effect of Age looks very different, averaged over Treatment and Sex
Case study: Arreests for Marijuana Possession

In Dec. 2002, the *Toronto Star* examined the issue of racial profiling, by analyzing a data base of 600,000+ arrest records from 1996-2002.

They focused on a subset of arrests for which police action was discretionary, e.g., simple possession of small quantities of marijuana, where the police could:
- Release the arrestee with a summons— like a parking ticket
- Bring to police station, hold for bail, etc.— harsher treatment

**Response variable:** released – Yes, No

**Main predictor** of interest: skin-colour of arrestee (black, white)

---

### Arrests for Marijuana Possession: Data

#### Data

**Control variables:**
- year, age, sex
- employed, citizen – Yes, No
- checks — Number of police data bases (previous arrests, previous convictions, parole status, etc.) in which the arrestee’s name was found.

```r
library(effects) # for Arrests data
data(Arrrests) # for Anova()
some(Arrests)
```

```r
## 243  
## 514  
## 628  
## 671  
## 1278 
## 2015 
## 2162 
## 2632 
## 3240 
## 4324
```

---

![Graphical Statistics](image)

**Man behind the numbers**

... Which got turned into this infographic:

**Same charge, different treatment**

Statistical analysis of single drug possession charges shows that blacks are much less likely to be released at the scene and much more likely to be held in custody for a bail hearing. Darker colours represent a stronger statistical link between skin colour and police treatment.

- **Whites** are more likely to be released at the scene
- **Blacks** are much more likely to be held for bail hearings

**Degree of likelihood**
- Much less likely to occur
- Much more likely to occur
- More likely to occur

<table>
<thead>
<tr>
<th></th>
<th>Released at Scene</th>
<th>Released at Station</th>
<th>Held for Bail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td>78%</td>
<td>14.5%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Blacks</td>
<td>64%</td>
<td>20%</td>
<td>16%</td>
</tr>
</tbody>
</table>

**SOURCE:** Toronto police arrest records 1996-2000

**Hey, they even spelled likelihood correctly!**
Arrests for Marijuana Possession: Model

To allow possibly non-linear effects of year, we treat it as a factor:

```r
> Arrests$year <- as.factor(Arrests$year)
```

Logistic regression model with all main effects, plus interactions of colour:year and colour:age

```r
> arrests.mod <- glm(released ˜ employed + citizen + checks + colour * year + colour * age, family = binomial, data = Arrests)
```

```r
> Anova(arrests.mod)
```

Analysis of Deviance Table (Type II tests)

<table>
<thead>
<tr>
<th></th>
<th>LR Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>employed</td>
<td>72.673</td>
<td>1</td>
<td>&lt; 2.2e-16  ***</td>
</tr>
<tr>
<td>citizen</td>
<td>25.783</td>
<td>1</td>
<td>3.820e-07   ***</td>
</tr>
<tr>
<td>checks</td>
<td>205.211</td>
<td>1</td>
<td>&lt; 2.2e-16   ***</td>
</tr>
<tr>
<td>year</td>
<td>6.087</td>
<td>5</td>
<td>0.2978477</td>
</tr>
<tr>
<td>age</td>
<td>0.459</td>
<td>1</td>
<td>0.4982736</td>
</tr>
<tr>
<td>colour:year</td>
<td>21.720</td>
<td>5</td>
<td>0.0005917  ***</td>
</tr>
<tr>
<td>colour:age</td>
<td>13.886</td>
<td>1</td>
<td>0.0001942  ***</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Effect plots: colour**

```r
> plot(Effect("colour", arrests.mod), ci.style="bands", ...)
```

- Effect plot for colour shows average effect controlling (adjusting) for all other factors simultaneously.
- (The *Star* analysis, controlled for these one at a time.)
- Evidence for different treatment of blacks and whites ("racial profiling").
- (Even Frances Nunziata could understand this.)
- NB: Effects smaller than claimed by the *Star*

**Effect plots: Interactions**

The story turned out to be more nuanced than reported by the *Toronto Star*, as shown in effect plots for interactions with colour.

```r
> plot(effect("colour:year", arrests.mod), multiline = TRUE, ...)
```

- Up to 2000, strong evidence for differential treatment of blacks and whites.
- Also evidence to support Police claim of effect of training to reduce racial effects in treatment.

```r
> plot(effect("colour:age", arrests.mod), multiline = TRUE, ...)
```

A more surprising finding:
- Opposite age effects for blacks and whites—
- Young blacks treated *more* harshly than young whites.
- Older blacks treated *less* harshly than older whites.
Effect plots: allEffects

All model effects can be viewed together using `plot(allEffects(mod))`

```r
> arrests.effects <- allEffects(arrests.mod, xlevels = list(age = seq(15, +
   45, 5)))
> plot(arrests.effects, ylab = "Probability(released)"
```

Model diagnostics

As in regression and ANOVA, the validity of a logistic regression model is threatened when:

- Important predictors have been omitted from the model
- Predictors assumed to be linear have non-linear effects on \( \text{Pr}(Y = 1) \)
- Important interactions have been omitted
- A few “wild” observations have a large impact on the fitted model or coefficients

Model specification: Tools and techniques

- Use non-parametric smoothed curves to detect non-linearity
- Consider using polynomial terms \((X^2, X^3, \ldots)\) or regression splines (e.g., `ns(X, 3)`)
- Use `update(model, ...)` to test for interactions— formula: \( . \sim .^2 \)

Diagnostic plots in R

In R, plotting a `glm` object gives the “regression quartet” — basic diagnostic plots

```r
arth.mod1 <- glm(Better ˜ Age + Sex + Treatment, data=Arthritis,
   family='binomial')
plot(arth.mod1)
```

Unusual data: Leverage and Influence

- “Unusual” observations can have dramatic effects on estimates in linear models
  - Can change the coefficients for the predictors
  - Can change the predicted values for all observations
- Three archetypal cases:
  - Typical X (low leverage), bad fit — Not much harm
  - Unusual X (high leverage), good fit — Not much harm
  - Unusual X (high leverage), bad fit — BAD, BAD, BAD
- Influential observations: unusual in both X and Y
- Heuristic formula:

\[
\text{Influence} = \text{Leverage}_X \times \text{Residual}_Y
\]

Better versions of these plots are available in the `car` package
Effect of adding one more point in simple linear regression (new point in blue)

Influence plots in R

library(car)
influencePlot(arth.logistic2)

Which cases are influential?

<table>
<thead>
<tr>
<th>ID</th>
<th>Treatment</th>
<th>Sex</th>
<th>Age</th>
<th>Better</th>
<th>StudRes</th>
<th>Hat</th>
<th>CookD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treated</td>
<td>Male</td>
<td>27</td>
<td>1</td>
<td>1.922</td>
<td>0.08968</td>
<td>0.3358</td>
</tr>
<tr>
<td>15</td>
<td>Treated</td>
<td>Female</td>
<td>23</td>
<td>0</td>
<td>-1.183</td>
<td>0.14158</td>
<td>0.2049</td>
</tr>
<tr>
<td>39</td>
<td>Treated</td>
<td>Female</td>
<td>69</td>
<td>0</td>
<td>-2.171</td>
<td>0.03144</td>
<td>0.2626</td>
</tr>
</tbody>
</table>

- X axis: Leverage (“hat values”)
- Y axis: Studentized residuals
- Bubble size ~ Cook D (influence on coefficients)