Logit models

For a binary response, each loglinear model is equivalent to a logit model (logistic regression, with categorical predictors)

\[ \lambda_{jk} = \mu + \lambda_{1j}^A + \lambda_{jk}^D + \lambda_{jk}^G + \lambda_{jk}^{AD} + \lambda_{jk}^{DG} \]

So, for admitted \((i = 1)\) and rejected \((i = 2)\), we have:

\[ \log m_{1jk} = \mu + \lambda_{1j}^A + \lambda_{jk}^D + \lambda_{jk}^G + \lambda_{jk}^{AD} + \lambda_{jk}^{DG} \quad (7) \]

\[ \log m_{2jk} = \mu + \lambda_{2j}^A + \lambda_{jk}^D + \lambda_{jk}^G + \lambda_{jk}^{AD} + \lambda_{jk}^{DG} \quad (8) \]

Thus, subtracting \((7)-(8)\), terms not involving Admit will cancel:

\[ L_{jk} = \log m_{1jk} - \log m_{2jk} = \log \left( \frac{m_{1jk}}{m_{2jk}} \right) = \log \text{odds of admission} \]

\[ = (\lambda_{1j}^A - \lambda_{2j}^A) + (\lambda_{jk}^{AD} - \lambda_{jk}^{AD}) \]

\[ = \alpha + \beta_{jk}^{\text{Dept}} \quad \text{(renaming terms)} \]

where,

- \(\alpha\): overall log odds of admission
- \(\beta_{jk}^{\text{Dept}}\): effect on admissions of department,
- associations among predictors are assumed, but don’t appear in the logit model,
Logit models

Other loglinear models have similar, simpler forms as logit models, where only the relations of the response to the predictors appear in the equivalent logit model.

- Admit \perp Gender \perp Dept (mutual independence \equiv [A][D][G])

\[ \log m_{ijk} = \mu + \lambda_A^i + \lambda_D^j + \lambda_G^k \]
\[ \equiv L_{jk} = (\lambda_A^1 - \lambda_A^2) = \alpha \] (constant log odds)

- Admit \perp Gender \mid Dept, except for Dept. A

\[ \log m_{ijk} = \mu + \lambda_A^i + \lambda_D^j + \lambda_G^k + \lambda_{AD}^{ij} + \lambda_{DG}^{jk} + \delta_{j=1}\beta_{Gender} \]
\[ \equiv L_{jk} = \log(m_{1jk}/m_{2jk}) = \alpha + \beta_j^{Dept} + \delta_{j=1}\beta_{Gender} \]

where,
- \( \beta_j^{Dept} \): effect on admissions for department \( j \),
- \( \delta_{j=1}\beta_{Gender} \): 1 df term for effect of gender in Dept. A.

Fitting logit models

- Fitting procedures
  - PROC CATMOD, PROC LOGISTIC
  - PROC GENMOD / dist=poisson
  - SPSS: Logistic regression, Loglinear \rightarrow Logit, Generalized Linear Models
  - R: glm(), gnm()

- Visualization procedures
  - CATPLOT macro - plot predicted, observed log odds from CATMOD
  - INFLGLIM macro - influence plots for generalized linear models
  - HALFNORM macro - half-normal plot of residuals for generalized linear models

- SAS craft
  - All SAS procedures \rightarrow output dataset with obs., fitted values, residuals, diagnostics, etc.
  - New model \rightarrow new output dataset
  - Plotting steps remain the same
  - Similar ideas for SPSS, R

Plots for logit models

- Fit: PROC CATMOD; plot: CATPLOT macro
- Model: Admit \sim Gender + Dept \leftrightarrow loglinear [AD] [AG] [DG]

```sas
proc catmod order=data data=berkeley;
  weight freq;
  response / out=predict;
  model admit = dept gender / ml;
%catplot(data=predict, xc=dept, class=gender,
         type=FUNCTION, z=1.96, legend=legend1);
```

Plots observed and predicted on the logit scale (type=FUNCTION)

Main effects model—parallel profiles
Probabilities on a separate scale (added below)
Logit models: details

- **Model**: `Admit ~ Gender + Dept ↔ [AD] [AG] [DG]`

```sas
%include catdata(berkeley);
proc catmod order=data
   data=berkeley;
   weight freq;
   response / out=predict;
   model admit = dept gender / ml;
run;
```

PROC CATMOD output: Overall tests and goodness of fit

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>262.49</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dept</td>
<td>5</td>
<td>534.78</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>gender</td>
<td>1</td>
<td>1.53</td>
<td>0.2167</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>5</td>
<td>20.20</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

- No effect of Gender; big effect of Dept
- LR test (vs. saturated model): Model doesn't fit well— Why? How to modify?

Plots for logit models: Output data set

PROC CATMOD output data set: observed & predicted, probabilities & logits

<table>
<thead>
<tr>
<th>dept</th>
<th>gender</th>
<th>admit</th>
<th><em>TYPE</em></th>
<th><em>OBS</em></th>
<th><em>PRED</em></th>
<th><em>SEPRED</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Male</td>
<td>Admit</td>
<td>FUNCTION</td>
<td>0.492</td>
<td>0.582</td>
<td>0.069</td>
</tr>
<tr>
<td>A</td>
<td>Male</td>
<td>Reject</td>
<td>PROB</td>
<td>0.621</td>
<td>0.642</td>
<td>0.016</td>
</tr>
<tr>
<td>A</td>
<td>Female</td>
<td>Admit</td>
<td>FUNCTION</td>
<td>1.544</td>
<td>0.682</td>
<td>0.099</td>
</tr>
<tr>
<td>A</td>
<td>Female</td>
<td>Reject</td>
<td>PROB</td>
<td>0.824</td>
<td>0.664</td>
<td>0.022</td>
</tr>
<tr>
<td>B</td>
<td>Male</td>
<td>Admit</td>
<td>PROB</td>
<td>0.534</td>
<td>0.539</td>
<td>0.086</td>
</tr>
<tr>
<td>B</td>
<td>Male</td>
<td>Reject</td>
<td>PROB</td>
<td>0.650</td>
<td>0.631</td>
<td>0.020</td>
</tr>
<tr>
<td>B</td>
<td>Female</td>
<td>Admit</td>
<td>PROB</td>
<td>0.754</td>
<td>0.639</td>
<td>0.116</td>
</tr>
<tr>
<td>B</td>
<td>Female</td>
<td>Reject</td>
<td>PROB</td>
<td>0.690</td>
<td>0.654</td>
<td>0.026</td>
</tr>
</tbody>
</table>

This contains both the observed and fitted logit values ( _TYPE_='FUNCTION' ) and probabilities ( _TYPE_='PROB' )

CATPLOT macro

- Plot logit values ( _TYPE_='FUNCTION' ) or probabilities ( _TYPE_='PROB' )
- With PScale macro, can plot on logit scale, with probability scale on right.

```sas
%pscale(lo=-4, hi=3, anno=pscale);
title 'Model: logit(Admit) = Dept Gender'
a=-90 'Probability (Admitted)';
axi1 order=(-3 to 2) offset=(4)
label=(a=90 'Log Odds (Admitted)');
axi2 label=('Department') offset=(4);
%catplot(data=predict, class=gender, xc=dept,
   type=FUNCTION, /* plot logit values */
z=1.96, /* show 1.96 x SE -> 95% CI */
anno=pscale); /* add probability scale */
```

→ no effect of Gender, except in Dept A (Females more likely admitted!)
Fitting and graphing other models

- Change MODEL statement → new fitted values
- Plotting step remains the same
- Admit ⊥ Gender | Dept, except for Dept. A ↔ Admit ∼ Dept + δ_j=1 Gender

```
proc catmod order=data data=berkeley;
response / out=predict;
model admit = dept dept1AG / ml;
%catplot(data=predict, xc=dept, class=gender, type=FUNCTION, z=1.96, legend=legend1);
```

Logit(Admit) = Dept DeptA*Gender
Gender Female Male
Log Odds (Admitted)
-3 -2 -1 0 1 2
Department A B C D E F

Fitting and graphing other models: details

```
%include catdata(berkeley);
data berkeley;
set berkeley;
*-- Dummy variable for Gender in Dept A;
department1AG = (gender='F') * (dept=1);
format dept dept .;
proc catmod order=data data=berkeley;
weight freq;
population dept gender;
direct dept1AG;
response / out=predict;
model admit = dept dept1AG / ml;
run;
```

PROC CATMOD output:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>291.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dept</td>
<td>5</td>
<td>571.45</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dept1AG</td>
<td>1</td>
<td>16.04</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.6685</td>
<td>0.0392</td>
<td>291.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dept</td>
<td>1.1606</td>
<td>0.0705</td>
<td>271.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>B</td>
<td>1.2113</td>
<td>0.0802</td>
<td>227.95</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>C</td>
<td>0.0528</td>
<td>0.0687</td>
<td>0.59</td>
<td>0.4426</td>
</tr>
<tr>
<td>D</td>
<td>0.00383</td>
<td>0.0727</td>
<td>0.00</td>
<td>0.9607</td>
</tr>
<tr>
<td>E</td>
<td>-0.4210</td>
<td>0.0871</td>
<td>23.34</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dept1AG</td>
<td>1.0521</td>
<td>0.2627</td>
<td>16.04</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Fits well! How to interpret?

```
proc print data=predict; var _obs_ _pred_ _sepred_; id dept gender; format numeric 6.3 dept dept.; where(_type_='FUNCTION');
```

PROC CATMOD: observed and predicted logits:

<table>
<thead>
<tr>
<th>dept</th>
<th>gender</th>
<th><em>OBS</em></th>
<th><em>PRED</em></th>
<th><em>SEPRED</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>M</td>
<td>0.492</td>
<td>0.492</td>
<td>0.072</td>
</tr>
<tr>
<td>A</td>
<td>F</td>
<td>1.544</td>
<td>1.544</td>
<td>0.025</td>
</tr>
<tr>
<td>B</td>
<td>M</td>
<td>0.534</td>
<td>0.543</td>
<td>0.086</td>
</tr>
<tr>
<td>B</td>
<td>F</td>
<td>0.754</td>
<td>0.754</td>
<td>0.086</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>-0.536</td>
<td>-0.616</td>
<td>0.069</td>
</tr>
<tr>
<td>C</td>
<td>F</td>
<td>-0.660</td>
<td>-0.618</td>
<td>0.069</td>
</tr>
<tr>
<td>D</td>
<td>M</td>
<td>-0.704</td>
<td>-0.665</td>
<td>0.075</td>
</tr>
<tr>
<td>D</td>
<td>F</td>
<td>-0.622</td>
<td>-0.665</td>
<td>0.075</td>
</tr>
<tr>
<td>E</td>
<td>M</td>
<td>-0.957</td>
<td>-1.090</td>
<td>0.095</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>-1.157</td>
<td>-1.090</td>
<td>0.095</td>
</tr>
<tr>
<td>F</td>
<td>M</td>
<td>-2.770</td>
<td>-2.676</td>
<td>0.152</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>-2.581</td>
<td>-2.676</td>
<td>0.152</td>
</tr>
</tbody>
</table>
**Logit models**

**Fitting and graphing other models: details**

```sas
proc catplot(data=predict, x=dept, class=gender,
           type=FUNCTION, / * plot the log odds */
           z=1.96); /* 95% error bars */
```

```
logit(Admit) = Dept DeptA*Gender
Gender Female
Male
Log Odds (Admitted)
-3 -2 -1 0 1 2
Department A B C D E F
```

**Diagnostic plots for Generalized Linear Models**

**INFLGLIM** macro: Influence plots for generalized linear models (Williams, 1987)
- Fit: PROC GENMOD; calculates additional diagnostic measures (Hat value, Cook’s D, etc.)
- Plot: measures of residual (GY=Δχ², χ² residual) vs. leverage (GX=hat value), bubble size (area, radius) \(\sim\) Cook’s D.
- \(\rightarrow\) which cells have undue impact on fitted model?

```
%inflglim(data=berkeley,
           class=dept gender admit,
           resp=freq,
           model=admit|dept gender|dept,
           dist=poisson,
           id=cell,
           gx=hat, gy=streschi);
```

- All cells which do not fit (|ri| > 2) are for department A.
- Males applying to dept A have large leverage \(\Rightarrow\) large influence (Cook’s D)
Influence plots in R

The `influencePlot()` function in the `car` package gives similar plots:

```r
berkeley <- as.data.frame(UCBAdmissions)
berk.mod <- glm(Freq ~ Dept * (Gender+Admit), data=berkeley, family="poisson")
influencePlot(berk.mod, id.n=3, id.col="red")
```

Diagnostic plots for Generalized Linear Models

- Plot ordered absolute residuals, $|r_i|$ vs. expected normal values, $|z_i|$
- Standard normal confidence envelope not suitable for GLMs
- Simulate reference ‘line’ and envelope with simulated confidence intervals

```sas
%halfnorm(data=berkeley, class=dept gender admit, resp=freq, model=dept|gender dept|admit, dist=poisson, id=cell);
```

Logistic regression models

**Response variable**
- Binary response: 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(dose)
- Polynomial regressors: $\text{age}^2$, $\text{age}^3$, ···
- Categorical predictors: treatment, sex
- Interaction regressors: treatment $\times$ age, sex $\times$ age

- Points with largest $|\text{residual}|$ labeled
- The model fits well, except in department A.
Logistic regression models: Binary response

- For a binary response, $Y \in (0,1)$, want to predict $\pi = \Pr(Y = 1 \mid x)$
- Linear regression will give predicted values outside $0 \leq \pi \leq 1$
- Logistic model:
  - $\text{logit}(\pi_i) \equiv \log[\pi_i/(1 - \pi_i)]$ avoids this problem
  - logit is interpretable as “log odds” that $Y = 1$
- Probit (normal transform) model → similar predictions, but is less interpretable

Logistic regression models

Quantitative predictor: Linear and Logit regression on age

Except in extremes, linear and logistic models give similar predicted values

Arthritis treatment data

Linear and Logit Regressions on Age

Probability (Improved)

AGE

20 30 40 50 60 70 80

Logistic regression models

Fitting logistic models

Data:
- Frequency form (from PROC FREQ)— when all predictors are discrete
- Case form— when any predictors are quantitative

Models:
- CLASS statement (V7+)— no need for dummy variables
  - discrete predictors
  - can specify order and parameterization (effect, polynomial, reference cell)
- MODEL statement— allows GLM syntax, e.g.,

```plaintext
proc logistic;
  class Sex Treat;
  model Better = Sex | Treat | Age @2;
⇒ Better = Sex Treat Age Sex*Treat Sex*Age Treat*Age
```

Fitting
Logistic regression models: Binary response

Visualization

- **Goal:** see and understand the data and fitted model
- **LOGODDS** macro: Plot observed responses, fitted and smoothed probabilities
- **Model plots:**
  - OUTPUT statement →
    - fitted \( \hat{\pi}_i \), lower/upper \((1 - \alpha)\) CI, and/or
    - fitted logit, \((\alpha + \mathbf{x}_i^T \hat{\beta}) \pm z_{1-\alpha/2} \text{se}(\text{logit})\)
  - Plot with standard procedures (PROC GCHART, GPlot)
- Utility macros (BARS, LABEL, POINTS, PSSCALE, etc.) for custom displays
- Effect plots—plot hierarchical subset of effects, averaging over those not included.
- **INFLOGIS** macro: Influence plots for logistic regression models
- **ADDVAR** macro: Added variable plots for new predictors or transformations of old

**Example: Arthritis treatment data**

- **Predictors:** Sex, Treatment (treated, placebo), Age
- **Response:** improvement (none, some, marked)
  - Consider first as binary response: None vs. (Some or Marked)='Better'
- **Data in case form:**

```
data arthrit;
  length treat $7. sex $6. ;
  input id treat $ sex $ age improve @@ ;
  case = _n_;
  better = (improve > 0); /* -- Make binary response; */
datalines;
57 Treatment Male 27 1 9 Placebo Male 37 0
46 Treatment Male 29 0 14 Placebo Male 44 0
77 Treatment Male 30 0 73 Placebo Male 50 0
... (observations omitted)
56 Treatment Female 69 1 42 Placebo Female 66 0
43 Treatment Female 70 1 15 Placebo Female 66 1
71 Placebo Female 68 1
1 Placebo Female 74 2
;```

**LOGODDS macro: Empirical logit plots**

Problems with visualizing discrete outcomes:

- **Linearity:** Is a linear relation realistic?
- **Smoothing:** Discrete data often requires smoothing to see!

The **LOGODDS** macro:

- Show the data: Plot \((0/1)\) responses [stacked or jittered]
- Divide \(X\) into groups (e.g., deciles), empirical logit, \(\log \left( \frac{y_i + 1/2}{n_i - y_i + 1/2} \right)\), for each
- Linear logistic regression, plus smoothed curve (LOWESS macro)

```sas
%include catdata(arthrit);
%logodds(data=arthrit, x=age, y=better, /* vars to plot */
  smooth=0.5, /* LOWESS smoothing parameter */
  plot=logit); /* plot on logit scale */
```
Smoothing the binary observations

Can also use direct smoothing:

Arthritis data: linear logistic and lowess smooth

- SAS: PROC LOESS, lowess macro; R: lowess()
- There is a hint that the relation may be non-linear
- But data is thin at the extremes

PROC LOGISTIC: Model fitting and plotting

- Specify ordering of response levels (order= or descending options)
- Specify parameterizations for CLASS variables
- OUTPUT statement to get fitted logits and probabilities

The output includes:

Type III Analysis of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>1</td>
<td>6.2576</td>
<td>0.0124</td>
</tr>
<tr>
<td>treat</td>
<td>1</td>
<td>10.7596</td>
<td>0.0010</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>5.5655</td>
<td>0.0183</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex Female vs Male</td>
<td>4.427</td>
<td>1.380 - 14.204</td>
</tr>
<tr>
<td>treat Treated vs Placebo</td>
<td>5.811</td>
<td>2.031 - 16.632</td>
</tr>
<tr>
<td>age</td>
<td>1.050</td>
<td>1.008 - 1.093</td>
</tr>
</tbody>
</table>

Parameter estimates (reference cell coding):

- $\beta_1 = 1.49 \Rightarrow$ Females $e^{1.49} = 4.43 \times$ more likely better than Males
- $\beta_2 = 1.76 \Rightarrow$ Treated $e^{1.76} = 5.81 \times$ more likely better than Placebo
- $\beta_3 = 0.0487 \Rightarrow$ odds ratio=1.05 $\Rightarrow$ odds of improvement increase 5% each year. Over 10 years, odds of improvement $= e^{10 \times 0.0486} = 1.63$, a 63% increase.

PROC LOGISTIC: Full-model plots

Full-model plots display the fitted (predicted) values over all combinations of predictors:

- Plot fitted values from the dataset specified on the OUTPUT statement
- Plot either predicted probabilities or logits
- Confidence intervals or standard errors allow showing error bars

The first few observations from the results dataset:

<table>
<thead>
<tr>
<th>id</th>
<th>sex</th>
<th>treat</th>
<th>age</th>
<th>better</th>
<th>prob</th>
<th>lower</th>
<th>upper</th>
<th>logit</th>
<th>selogit</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>Male</td>
<td>Treated</td>
<td>27</td>
<td>1</td>
<td>0.194</td>
<td>0.103</td>
<td>0.334</td>
<td>-1.427</td>
<td>0.758</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>Placebo</td>
<td>37</td>
<td>0</td>
<td>0.063</td>
<td>0.032</td>
<td>0.120</td>
<td>-2.700</td>
<td>0.725</td>
</tr>
<tr>
<td>46</td>
<td>Male</td>
<td>Treated</td>
<td>29</td>
<td>0</td>
<td>0.209</td>
<td>0.115</td>
<td>0.350</td>
<td>-1.330</td>
<td>0.728</td>
</tr>
<tr>
<td>14</td>
<td>Male Placebo</td>
<td>44</td>
<td>0</td>
<td>0.086</td>
<td>0.047</td>
<td>0.152</td>
<td>-2.358</td>
<td>0.658</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>Male</td>
<td>Treated</td>
<td>30</td>
<td>0</td>
<td>0.217</td>
<td>0.122</td>
<td>0.357</td>
<td>-1.281</td>
<td>0.713</td>
</tr>
<tr>
<td>73</td>
<td>Male Placebo</td>
<td>50</td>
<td>0</td>
<td>0.112</td>
<td>0.065</td>
<td>0.188</td>
<td>-2.066</td>
<td>0.622</td>
<td></td>
</tr>
</tbody>
</table>

- prob, predicted probabilities, with CI (lower, upper)
- logit, predicted logit, with standard error selogit
PROC LOGISTIC: Full-model plots

Basic plots:
- Plot either logit or probability vs. one predictor (continuous or most levels)
- Separate curves for one factor (= factor)
- Separate panels for all others (BY statement)

```sas
proc gplot data=results;
  plot (logit prob) * age = treat; /* separate curves */
  by sex; /* separate panels */
  symbol1 v=circle i=join l=3 c=black; /* placebo */
  symbol2 v=dot i=join l=1 c=red; /* treated */
  barlen=selogit, out=bars;
by sex;

data bars;
Placebo
Treated
Female
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
Placebo
Treated
Male
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
Probability Improved
Log Odds Improved
-3
-2
-1
0
1
2
3
Age
20 30 40 50 60 70 80
```

- SYMBOL statement— define the point value (v=), interpolate option (i=), line style (l=), color (c=), etc.

Enhanced plots:

```sas
*-- Error bars, on logit scale;
%pscale(out=pscale,
Placebo
Treated
Female
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
Placebo
Treated
Male
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
Probability Improved
Log Odds Improved
-3
-2
-1
0
1
2
3
Age
20 30 40 50 60 70 80
```

- Custom legend and panel labels (LABEL macro)
- Probability scales at right (PScale macro)
- Show 67% error bars ±1 se (BARS macro)
- Separate curves for one factor (BY statement)
- Separate panels for all others (BY statement)
- Join ANNOTATE datasets
- Probability scales at right

```
*-- Error bars, on logit scale;
%pscale(out=pscale,
Placebo
Treated
Female
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
Placebo
Treated
Male
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
Probability Improved
Log Odds Improved
-3
-2
-1
0
1
2
3
Age
20 30 40 50 60 70 80
```

- Custom legend and panel labels (LABEL macro)
- Probability scales at right
- Show 67% error bars ±1 se (BARS macro)
- Separate curves for one factor (BY statement)
- Separate panels for all others (BY statement)
- Join ANNOTATE datasets
- Probability scales at right
Models with interactions

- Plotting fitted values
  - Only need to change the `MODEL` statement
  - Output dataset automatically incorporates all model terms
  - Plotting steps remain exactly the same

```sas
proc logistic data=arthrit descending;
class sex (ref=last) treat (ref=first) / param=ref;
model better = treat sex | age @2;;
output out=results p=prob l=lower u=upper
  xbeta=logit stdxbeta=selogit / alpha=.33;
```

Effect plots: basic ideas

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

Data

<table>
<thead>
<tr>
<th>xi</th>
<th>x2</th>
<th>y</th>
<th>yhat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>y</td>
<td>4.73</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4.46</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4.32</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4.94</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>4.73</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>6.10</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>6.71</td>
<td></td>
</tr>
</tbody>
</table>

* Fit data: $X\beta \Rightarrow \hat{y}$

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

Effect plots software

- General method
  - Create a grid of values for predictors in the effect (`EXPGRID` macro)
  - Fix other predictors at “typical” values (mean, median, proportion in the data)
  - Concatenate grid with data
  - Fit model → output data set → fitted values in the grid
  - Standard errors automatically calculated
  - Plot fitted values in the grid

- `EFFPLOT` macro
  - Works with `PROC REG`, `PROC GLM`, `PROC LOGISTIC`, `PROC GENMOD`
  - Uses `MEANPLOT` macro to do the plotting
  - Some limitations – can’t plot correct standard errors

- SAS 9.3 ODS Graphics
  - Several procedures now do effects-like plots: `LOGISTIC`, `GLM`, `GLIMMIX`
  - Easy; `PROC LOGISTIC` quite flexible

- R: effects package
  - Most general: Handles linear models (`lm()`), generalized linear models (`glm()`), multinomial (`multinom()`) and proportional-odds (`polr()`) models.
  - `allEffects(model)` calculates effects for all high-order terms in model
  - `plot(allEffects(model))` plots them

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 $g(\mu) = \eta \rightarrow \text{estimate } b \text{ of } \beta \text{ and covariance matrix } V(b) \text{ of } b.$
  - 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,” $X^*$
  - Calculate fitted effect values, $\tilde{\eta}^* = X^*b$.
  - Standard errors are square roots of $\text{diag}(X^*V(b)X^*)$.
  - Plot $\tilde{\eta}^*$, or values transformed back to scale of response, $g^{-1}(\tilde{\eta}^*)$.
  - Note: This provides a general means to visualize interactions in all linear and generalized linear models.
Effect plots: Example
- Cowles and Davis (1987)—Volunteering for a psychology experiment
  - Predictors: Sex, Neuroticism, Extraversion
  - → strong interaction, Neuroticism × Extraversion

Effect plots: SAS 9.3 ODS Graphics
```sas
proc logistic data=cowles outest=parm descending;
class Sex;
model Volunteer = Sex Extraver | Neurot / lackfit;
effectplot slicefit(x=Extraver sliceby=Neurot) / at(sex=1.5) noobs;
effectplot slicefit(x=Neurot sliceby=Extraver) / at(sex=1.5) noobs;
effectplot contour(x=Neurot y=Extraver) / at(sex=1.5) noobs;
run;
```

SAS 9.2: ODS Graphics
```sas
%include catdata(arthrit);
ods graphics on;
proc logistic data=arthrit descending
plots(only)=(effect(plotby=sex sliceby=treat showobs clband alpha=0.33));
class sex (ref=last) treat (ref=first) / param=ref;
model better = sex treat age / clodds=wald;
run;
ods graphics off;
```
Effect plots with the effects package in R

```r
> library(effects) ## load the effects package
> data(Cowles)
> mod.cowles <- glm(volunteer ~ sex + neuroticism*extraversion,
+              data=Cowles, family=binomial)
> summary(mod.cowles)
```

Coefficients:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | -2.358207 | 0.501320 | -4.704 | 2.55e-06 *** |
| sexmale | -0.247152 | 0.111631 | -2.214 | 0.02683 * |
| neuroticism | 0.110777 | 0.037648 | 2.942 | 0.00326 ** |
| extraversion | 0.166816 | 0.037719 | 4.423 | 9.75e-06 *** |
| neuroticism:extraversion | -0.008552 | 0.002934 | -2.915 | 0.00355 ** |

```r
> eff.cowles <- allEffects(mod.cowles, + xlevels=list(neuroticism=0:24, + extraversion=seq(0, 24, 8)))
> plot(eff.cowles, + 'neuroticism:extraversion', + ylab="Prob(Volunteer)", + ticks=list(at=c(.1,.25,.5,.75,.9)), layout=c(4,1), aspect=1)
```

Extended example: Arrests for Marihuana Possession

Context & background

- 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)

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)
> data(Arrests)
> some(Arrests)
```

<table>
<thead>
<tr>
<th>released</th>
<th>colour</th>
<th>year</th>
<th>age</th>
<th>sex</th>
<th>employed</th>
<th>citizen</th>
<th>checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>915</td>
<td>No</td>
<td>2001</td>
<td>35</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>1568</td>
<td>Yes</td>
<td>2002</td>
<td>21</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>2981</td>
<td>Yes</td>
<td>2000</td>
<td>23</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>3381</td>
<td>Yes</td>
<td>1998</td>
<td>23</td>
<td>Male</td>
<td>No</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>3516</td>
<td>Yes</td>
<td>2002</td>
<td>22</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>4128</td>
<td>No</td>
<td>2001</td>
<td>29</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>4142</td>
<td>Yes</td>
<td>1998</td>
<td>23</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>4634</td>
<td>Yes</td>
<td>2001</td>
<td>18</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>4732</td>
<td>Yes</td>
<td>1999</td>
<td>21</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>5183</td>
<td>Yes</td>
<td>1999</td>
<td>19</td>
<td>Male</td>
<td>Yes</td>
<td>Yes</td>
<td>0</td>
</tr>
</tbody>
</table>
Extended example: Arrests for Marihuana 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)
> Anova(arrests.mod)
```

Analysis of Deviance Table (Type II tests)

<table>
<thead>
<tr>
<th>Response: released</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>colour</td>
<td>19.572</td>
<td>1</td>
<td>9.687e-06  ***</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

Evidence for different treatment of blacks and whites ("racial profiling"), controlling (adjusting) for other factors

```r
> plot(effect("colour", arrests.mod), multiline = FALSE, ylab = "Probability(released)")
```

- 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

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

- Opposite age effects for blacks and whites:
  - Young blacks treated more harshly than young whites
  - Older blacks treated less harshly than older whites

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)", ask = FALSE)
```

Influence measures and diagnostic plots

- **Leverage**: Potential impact of an individual case ∼ distance from the centroid in space of predictors
- **Residuals**: Which observations are poorly fitted?
- **Influence**: Actual impact of an individual case ∼ leverage × residual
  - C, CBAR – analogs of Cook’s D in OLS ∼ standardized change in regression coefficients when i-th case is deleted.
  - DIFCHISQ, DIFDEV ∼ Δχ² when i-th case is deleted.

```
Influence measures and diagnostic plots
```

PROC LOGISTIC: printed output with the influence option

```r
1 proc logistic data=arrests descending;
2 class colour year sex citizen employed;
3 model released = colour/year colour/age sex employed citizen checks;
4 model better = sex treat age / influence;
5 effectplot interaction (x=year sliceby=colour) / clm alpha=0.33 noobs;
6 effectplot slicefit (x=age sliceby=colour) / clm alpha=0.33 obs(fringe jitter);
7 run;
```
Influence measures and diagnostic plots

PROC LOGISTIC: plotting diagnostic measures with the `plots` option

```sas
proc logistic data=arthrit descending
   plots(only label)=(leverage dpc);
   class sex (ref=last) treat (ref=first) / param=ref;
   model better = sex treat age ;
run;
```

The option `plots(label)=dpc` gives plots of $\Delta \chi^2$ (DIFCHISQ, DIFDEV) vs. $\hat{p}$. Points are colored according to the influence measure C.

The two bands of points correspond to better = \{0, 1\}

INFLOGIS macro

- Specialized version of `INFLGLIM` macro for logistic regression
- Plots a measure of change in $\chi^2$ (DIFCHISQ or DIFDEV) vs. predicted probability or leverage.
- Bubble symbols show actual influence (C or CBAR)
- Shows standard cutoffs for “large” values
- Flexible labeling of unusual cases

INFLOGIS macro: Example

```sas
%include data(arthrit);
%inflogis(data=arthrit,
   class=sex treat,
   /* CLASS variables */
   y=better,  /* response */
   x=sex treat age,  /* predictors */
   id=case,  /* case ID */
   gy=DIFCHISQ,  /* graph ordinate */
   gx=PRED HAT,  /* graph abscissas */
   loptions=descending);
```

Printed output lists cases with “large” leverage, residual or influence:

```plaintext
case better sex treat age pred hat difchisq difdev c
  1 1 Male Treated 27 .806 .09 4.578 3.695 0.451
  22 1 Male Placebo 63 .807 .06 4.460 3.565 0.290
  30 1 Female Placebo 31 .818 .05 4.749 3.657 0.261
  34 1 Female Placebo 33 .803 .05 4.296 3.464 0.224
  55 0 Female Treated 58 .172 .03 4.970 3.676 0.160
  77 0 Female Treated 69 .108 .03 8.498 4.712 0.276
```
Diagnostic plots in R

In R, plotting a glm object gives the “regression quartet”

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

```r
cook.levels <- c(.5,Inf)
plot(arth.mod1, cook.levels = cook.levels)
```
Influence measures and diagnostic plots

The Donner Party: A graphic tale of survival & influence

History:
- Apr–May, 1846: Donner/Reed families set out from Springfield, IL to CA
- Jul: Bridger’s Fort, WY, 87 people, 23 wagons

TRAIL OF THE DONNER PARTY

Donner Party: Who lived and died?

Other analyses, e.g., (Ramsay and Schafer, 1997):
- Log Odds (survive) ∼ linear with Age
- Odds (survive | Women / survive | Men) = 4.9
- Ignored children

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>MALE</th>
<th>SURVIVED</th>
<th>DEATH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antoine</td>
<td>23</td>
<td>1</td>
<td>0</td>
<td>29DEC46</td>
</tr>
<tr>
<td>Breen, Edward</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, Margaret I.</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, James</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, John</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, Mary</td>
<td>40</td>
<td>0</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, Patrick</td>
<td>51</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, Patrick Jr.</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, Peter</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Breen, Simon</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Burger, Charles</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>27DEC46</td>
</tr>
<tr>
<td>Denton, John</td>
<td>28</td>
<td>1</td>
<td>0</td>
<td>26FEB47</td>
</tr>
<tr>
<td>Dolan, Patrick</td>
<td>40</td>
<td>0</td>
<td>1</td>
<td>27DEC46</td>
</tr>
<tr>
<td>Donner, Elitha Cumi</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Donner, Eliza Poor</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Donner, Elizabeth</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>14MAR47</td>
</tr>
<tr>
<td>Donner, Francis E.</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>Donner, George</td>
<td>62</td>
<td>1</td>
<td>0</td>
<td>18MAR47</td>
</tr>
<tr>
<td>Donner, George Jr.</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
</tbody>
</table>

Empirical logit plots

Is a linear logistic model satisfactory for these data?
- Discrete data often requires smoothing to see!

%logodds(data=donner, y=Died, x=Age, smooth=0.5);

⇒ relation with Age is quadratic: youngest and oldest most likely to perish.
Quadratic model?
- Fit: \( \Pr(\text{Death}) \sim \text{Age} + \text{Age}^2 + \text{Male} \)
- Statistical evidence for \( \text{Age}^2 \) equivocal:
  - Wald \( \chi^2_{(1)} = 2.84, p = 0.09 \); but
  - LR \( G^2_{(1)} = 4.40, p = 0.03 \)

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Standard Estimate</th>
<th>Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCPT</td>
<td>-1.7721</td>
<td>0.5673</td>
<td>9.7588</td>
<td>0.0018</td>
</tr>
<tr>
<td>AGE</td>
<td>0.0168</td>
<td>0.0184</td>
<td>0.8355</td>
<td>0.3607</td>
</tr>
<tr>
<td>AGE2</td>
<td>0.00208</td>
<td>0.00123</td>
<td>2.8439</td>
<td>0.0917</td>
</tr>
<tr>
<td>MALE</td>
<td>1.3745</td>
<td>0.5066</td>
<td>7.3617</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

- Males: \( \exp(1.3745) = 3.95 \) times as likely to die, controlling for Age, \( \text{Age}^2 \)

Visual evidence is persuasive (but the data are thin at older ages)

### Who was influential?

- Patrick Breen, James Reed: Older men who survived
- Elizabeth & Tamsen Donner, Elizabeth Graves: Older women who survived

### Why are they influential?

- Patrick Breen, James Reed: Older men who survived
- Elizabeth & Tamsen Donner, Elizabeth Graves: Older women who survived
- Moral lessons of this story:
  - Don’t try to cross the Donner Pass in late October; if you do, bring food
  - Plots of fitted models show only what is included in the model
  - Discrete data often need smoothing (or non-linear terms) to see the pattern
  - Always examine model diagnostics — preferably graphic
Logit models
- Analogous to ANOVA models for a binary response
- Equivalent to loglinear model, including interaction of all predictors
- Fitting: SAS: PROC CATMOD, PROC LOGISTIC; R: glm()
- Visualization: plot fitted logits (or probabilities) vs. factors (CATPLOT macro)

Logistic regression
- Analogous to regression models for a binary response
- Coefficients: increment to log odds / $\Delta X$; $\exp \beta \sim$ multiplier of odds per $\Delta X$
- Discrete responses: smoothing often useful
- Visualization: plot fitted logits (or probabilities) vs. predictors

Effect plots
- Plot a main effect or interaction in the context of a more complex model
- Shows that effect controlling for (averaged over) all other model effects
- SAS: EFFPLOT macro; R: effects package

Influence & diagnostics
- Influence plots highlight unusual cases/cells — large impact on fitted model
- Probability plots of residuals help to check model assumptions
- SAS: INFLGLIM macro, HALFNORM macro; R: plot(my.glm), influencePlot(my.glm)