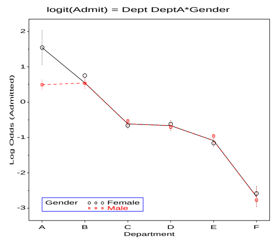
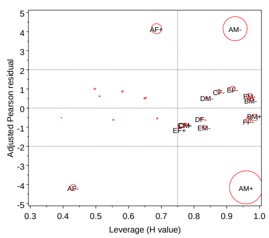
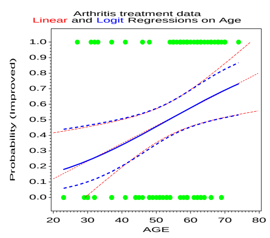
Part 3: Model-based methods for categorical data







Topics:

- Logit models
  - Plots for logit models
  - Diagnostic plots for generalized linear models
- Logistic regression models
  - Logistic regression: Binary response
  - Model plots
  - Effect plots for generalized linear models
  - Influence measures and diagnostic plots
  - Polytomous responses

Logit models

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

- Admit $\perp$ Gender $\mid$ Dept (conditional independence, [AD][DG])

$$\log\, m_{ijk} = \mu + \lambda_i^A + \lambda_j^D + \lambda_k^G + \lambda_{ij}^{AD} + \lambda_{jk}^{DG}$$

$$\leftrightarrow L_{jk} = \log(m_{1jk}/m_{2jk}) = (\lambda_1^A - \lambda_2^A) + (\lambda_{1j}^{AD} - \lambda_{2j}^{AD}) = \alpha + \beta_j^{\text{Dept}}$$

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

$$\log\, m_{ijk} = \mu + \lambda_i^A + \lambda_j^D + \lambda_k^G + \lambda_{ij}^{AD} + \lambda_{jk}^{DG} + \delta_{j=1}\lambda_{ik}^{AG}$$

$$\leftrightarrow L_{ij} = \log(m_{1jk}/m_{2jk}) = \alpha + \beta_j^{\text{Dept}} + \delta_{j=1}\beta^{\text{Gender}}$$

where,
  - $L_{ij} = \log(m_{ij1}/m_{ij2})$: log odds of admission,
  - β_j^{Dept}: effect on admissions of department,
  - $\delta_{j=1}\beta^{\text{Gender}}$: effect of gender in Dept. A.

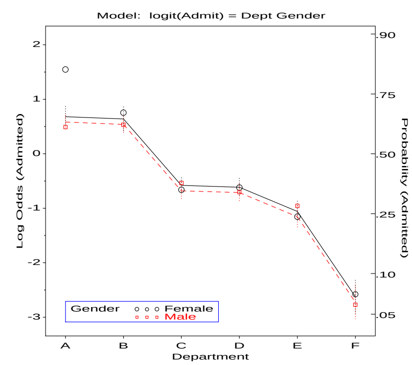
Logit models

- **Fitting procedures**
  - `PROC CATMOD`, `PROC LOGISTIC`
  - `PROC GENMOD / dist=poisson`
  - SAS/INSIGHT (Fit Y X) Options ⟶ Distribution poisson
  - SPSS: Logistic regression
  - R: glm()
- **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 ⟶ output dataset with obs., fitted values, residuals, diagnostics, etc.
  - New model ⟶ 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 ∼ Gender + Dept ↔ loglinear [AD] [AG] [DG]

```
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 for logit models

- **Model:** Admit $\sim$ Gender + Dept $\leftrightarrow$ [AD] [AG] [DG]

catberk2.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

Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	1	262.49	<.0001
dept	5	534.78	<.0001
gender	1	1.53	0.2167
Likelihood Ratio	5	20.20	0.0011

- No effect of Gender
- Model doesn't fit well— Why? How to modify?

Plots for logit models

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

dept	gender	admit	_TYPE_	_OBS_	_PRED_	_SEPRED_
A	Male		FUNCTION	0.492	0.582	0.069
A	Male	Admit	PROB	0.621	0.642	0.016
A	Male	Reject	PROB	0.379	0.358	0.016
A	Female		FUNCTION	1.544	0.682	0.099
A	Female	Admit	PROB	0.824	0.664	0.022
A	Female	Reject	PROB	0.176	0.336	0.022
B	Male		FUNCTION	0.534	0.539	0.086
B	Male	Admit	PROB	0.630	0.631	0.020
B	Male	Reject	PROB	0.370	0.369	0.020
B	Female		FUNCTION	0.754	0.639	0.116
B	Female	Admit	PROB	0.680	0.654	0.026
B	Female	Reject	PROB	0.320	0.346	0.026
...						
F	Male		FUNCTION	-2.770	-2.724	0.158
F	Male	Admit	PROB	0.059	0.062	0.009
F	Male	Reject	PROB	0.941	0.938	0.009
F	Female		FUNCTION	-2.581	-2.625	0.158
F	Female	Admit	PROB	0.070	0.068	0.010
F	Female	Reject	PROB	0.930	0.932	0.010

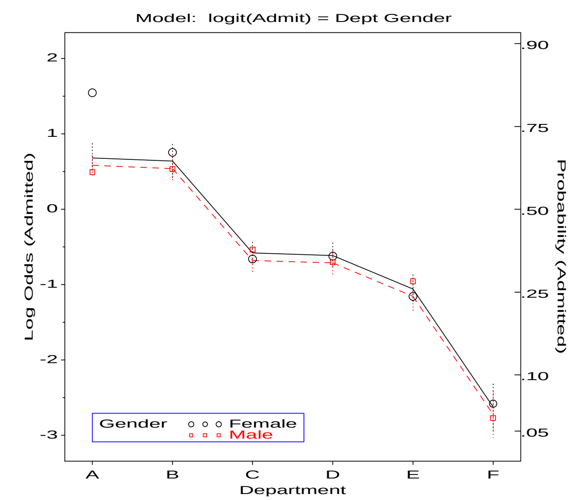
CATPLOT macro

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

··· catberk2.sas

```
%pscale(lo=-4, hi=3, anno=pscale);

title 'Model:  logit(Admit) = Dept Gender'
   a=-90 'Probability (Admitted)';
axis1 order=(-3 to 2) offset=(4)
      label=(a=90 'Log Odds (Admitted)');
axis2 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    */
```


CATPLOT macro

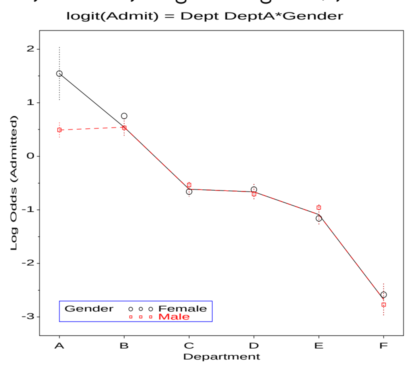


$\longrightarrow$ no effect of Gender, except in Dept A (Females more likely admitted!)

Fitting and graphing other models

- Change `MODEL` statement $\longrightarrow$ new fitted values
- Plotting step remains the same
- Admit $\perp$ Gender $|$ Dept, except for Dept. A $\leftrightarrow$ Admit $\sim$ Dept + $\delta_{j=1}$ 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);
```



Fitting and graphing other models

- Model: Admit $\perp$ Gender $|$ Dept, except for Dept. A
- Need to define a dummy variable for effect of Gender in Dept. A

catberk6.sas ···

```
%include catdata(berkeley);
data berkeley;
   set berkeley;
   *-- Dummy variable for Gender in Dept A;
   dept1AG = (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;
  ...
```


Fitting and graphing other models

`PROC CATMOD` output:

```
         Maximum Likelihood Analysis of Variance

     Source              DF   Chi-Square    Pr > ChiSq
     --------------------------------------------------
     Intercept            1       291.22        <.0001
     dept                 5       571.45        <.0001
     dept1AG              1        16.04        <.0001

     Likelihood Ratio     5         2.68        0.7489

         Analysis of Maximum Likelihood Estimates

                                 Standard      Chi-
Parameter       Estimate          Error     Square   Pr > ChiSq
---------------------------------------------------------------
Intercept        -0.6685         0.0392     291.22       <.0001
dept       A      1.1606         0.0705     271.21       <.0001
           B      1.2113         0.0802     227.95       <.0001
           C      0.0528         0.0687       0.59       0.4426
           D     0.00358         0.0727       0.00       0.9607
           E     -0.4210         0.0871      23.34       <.0001
dept1AG           1.0521         0.2627      16.04       <.0001
```

How to interpret?

Fitting and graphing other models

`PROC CATMOD`: observed and predicted logits:

··· catberk6.sas ···

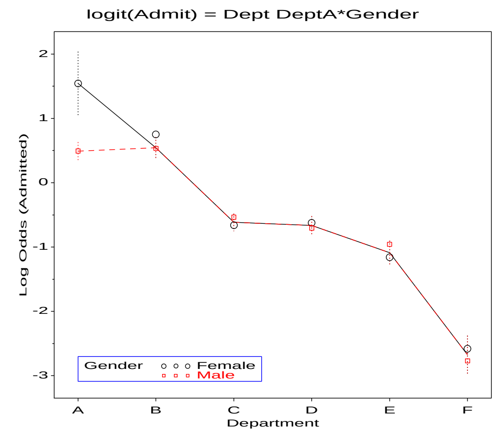
```
proc print data=predict;
  id dept gender;
  var _obs_ _pred_ _sepred_;
  format _numeric_  6.3 dept dept.;
  where(_type_='FUNCTION');
```

dept	gender	_OBS_	_PRED_	_SEPRED_
A	M	0.492	0.492	0.072
A	F	1.544	1.544	0.253
B	M	0.534	0.543	0.086
B	F	0.754	0.543	0.086
C	M	-0.536	-0.616	0.069
C	F	-0.660	-0.616	0.069
D	M	-0.704	-0.665	0.075
D	F	-0.622	-0.665	0.075
E	M	-0.957	-1.090	0.095
E	F	-1.157	-1.090	0.095
F	M	-2.770	-2.676	0.152
F	F	-2.581	-2.676	0.152

Fitting and graphing other models

··· catberk6.sas

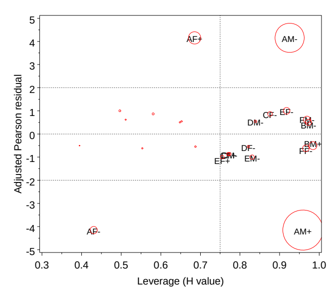
```
title 'logit(Admit) = Dept DeptA*Gender';
%catplot(data=predict, x=dept, class=gender,
   type=FUNCTION,       /* plot the log odds */
   z=1.96);             /* 95% error bars    */
```



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=$\Delta\chi^2$, χ^2 residual) vs. leverage (GX=hat value), bubble size (area, radius) $\sim$ Cook's D.
- $\longrightarrow$ which cells have undue impact on fitted model?



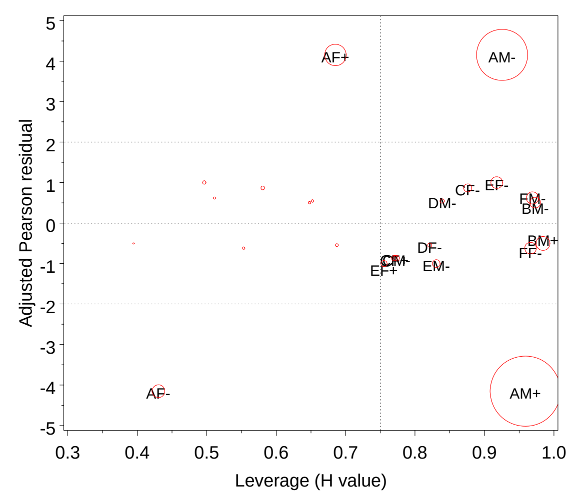
INFLGLIM macro: Example

- Berkeley data, model $[AD][GD] \leftrightarrow L_{ij} = \alpha + \beta_j^{\text{Dept}}$

genberk1.sas

```
%include catdata(berkeley);
*-- make a cell ID variable, joining factors;
data berkeley;
   set berkeley;
   cell = trim(put(dept,dept.)) ||
          gender ||
          trim(put(admit,yn.));

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


INFLGLIM macro: Example



- All cells which do not fit ($|r_i| > 2$) are for department A.
- Males applying to dept A have large leverage $\Rightarrow$ large influence (Cook's D)

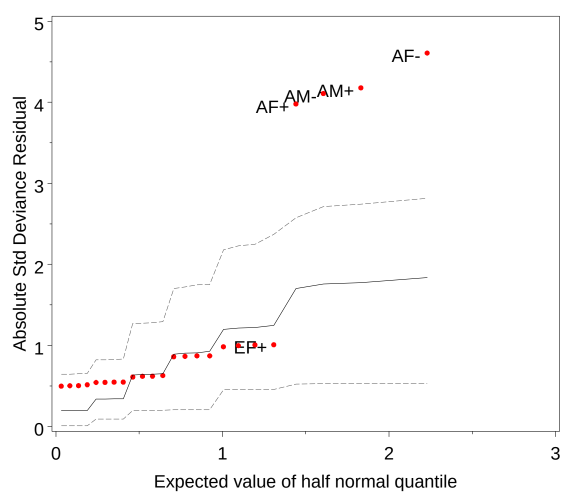
Diagnostic plots for Generalized Linear Models

HALFNORM macro: Half-normal plot of residuals (Atkinson, 1981)

- 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

··· genberk1.sas

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

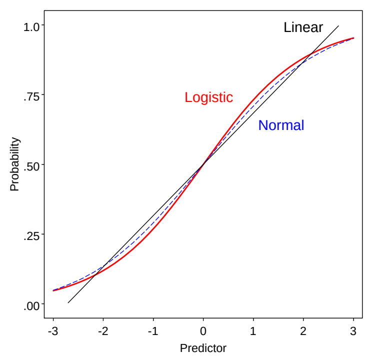
- Points with largest |residual| labeled
- The model fits well, except in department A.

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, Aliance, NDP
- **Explanatory variables**:
  - Quantitative regressors: age, dose
  - Transformed regressors: $\sqrt{\text{age}}$, log(dose)
  - Polynomial regressors: age^2, age^3, $\cdots$
  - Categorical predictors: treatment, sex
  - Interaction regessors: treatment $\times$ age, sex $\times$ age

Logistic regression models: Binary response

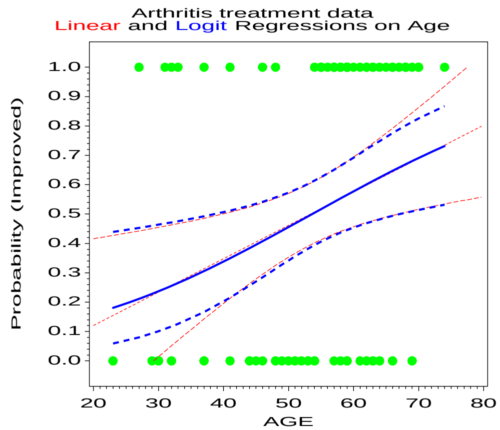
- For a binary response, $Y \in (0, 1)$, want to predict $\pi = \Pr(Y = 1 \mid x)$
- Linear regression will give predicted values outside $0 \leq \pi \leq 1$
- Logistic model:
  - $\text{logit}(\pi_i) \equiv \log[\pi/(1-\pi)]$ avoids this problem
  - logit is interpretable as "log odds" that $Y = 1$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



Logistic regression models: Binary response

Quantitative predictor: Linear and Logit regression on age

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



Logistic regression models: Binary response

- For a binary response, $Y \in (0,1)$, let $\boldsymbol{x}$ be a vector of p regressors, and π_i be the probability, $\Pr(Y = 1 \,|\, \boldsymbol{x})$.
- The logistic regression model is a linear model for the *log odds*, or *logit* that $Y = 1$, given the values in $\boldsymbol{x}$,

$$\begin{aligned} \text{logit}(\pi_i) \equiv \log\left(\frac{\pi_i}{1-\pi_i}\right) &= \alpha + \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\beta} \\ &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \end{aligned}$$

- An equivalent (non-linear) form of the model may be specified for the probability, π_i, itself,

$$\pi_i = \{1 + \exp(-[\alpha + \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\beta}])\}^{-1}$$

- The logistic model is a *linear model* for the log odds, but also a *multiplicative* model for the odds of "success,"

$$\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\beta}) = \exp(\alpha)\exp(\boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\beta})$$

  so, increasing x_{ij} by 1 increases $\text{logit}(\pi_i)$ by β_j, and multiplies the odds by e^{β_j}.

Logistic regression models: Binary response

- **Fitting**: `PROC LOGISTIC` (or `ROBUST` macro— M-estimation)
  - 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.,
      `model Better = Sex | Treat | Age @2;`

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, $(\alpha + \boldsymbol{x}_i^{\mathsf{T}} \hat{\boldsymbol{\beta}}) \pm z_{1-\alpha/2}\, se(\text{logit})$
    - Plot with standard procedures (`PROC GCHART`, `GPLOT`)
    - Utility macros (`BARS`, `LABEL`, `POINTS`, `PSCALE`, 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:

arthrit.sas

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

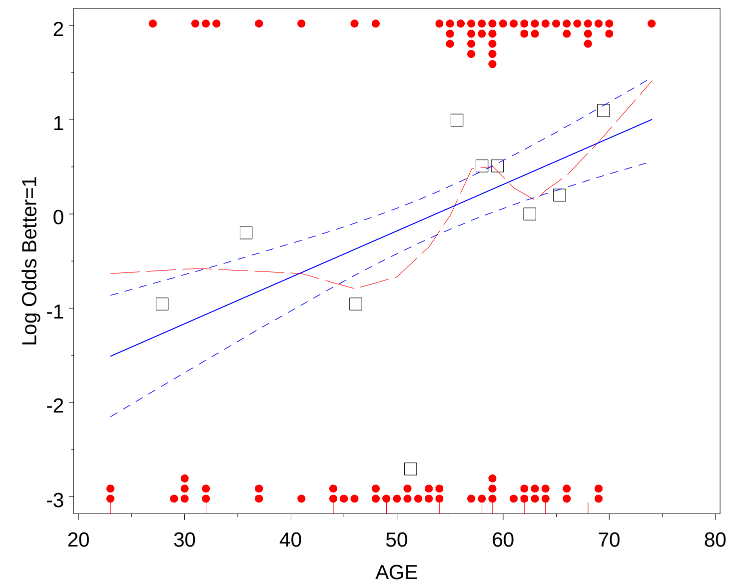

LOGODDS macro: Empirical logit plots

- **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), emprical 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)

```
%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       */
```

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

glogist1c.sas ···

```

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

The output includes:

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	1	6.2576	0.0124
treat	1	10.7596	0.0010
age	1	5.5655	0.0183

```
              Analysis of Maximum Likelihood Estimates

                                   Standard        Wald
Parameter           DF  Estimate      Error  Chi-Square  Pr > ChiSq

Intercept            1   -4.5033     1.3074     11.8649      0.0006
sex       Female     1    1.4878     0.5948      6.2576      0.0124
treat     Treated    1    1.7598     0.5365     10.7596      0.0010
age                  1    0.0487     0.0207      5.5655      0.0183

                        Odds Ratio Estimates

                                    Point          95% Wald
   Effect                        Estimate      Confidence Limits

   sex   Female vs Male             4.427       1.380      14.204
   treat Treated vs Placebo         5.811       2.031      16.632
   age                              1.050       1.008       1.093
```

Parameter estimates (reference cell coding):

- $\beta_1 = 1.49 \Rightarrow$ Females $e^{1.49}$=4.43 times more likely to be better than Males
- $\beta_2 = 1.76 \Rightarrow$ Treated $e^{1.76}$=5.81 times more likely to be 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: Model plots

- Plots of fitted values from the dataset specified on the OUTPUT statement
- Plot either predicted probabilities or logits

The first few observations from the `results` dataset:

```
 id sex   treat   age better   prob  lower  upper  logit selogit

 57 Male Treated  27    1     0.194  0.103  0.334 -1.427  0.758
  9 Male Placebo  37    0     0.063  0.032  0.120 -2.700  0.725
 46 Male Treated  29    0     0.209  0.115  0.350 -1.330  0.728
 14 Male Placebo  44    0     0.086  0.047  0.152 -2.358  0.658
 77 Male Treated  30    0     0.217  0.122  0.357 -1.281  0.713
 73 Male Placebo  50    0     0.112  0.065  0.188 -2.066  0.622
  ...
```


PROC LOGISTIC: Model plots

Basic plots:

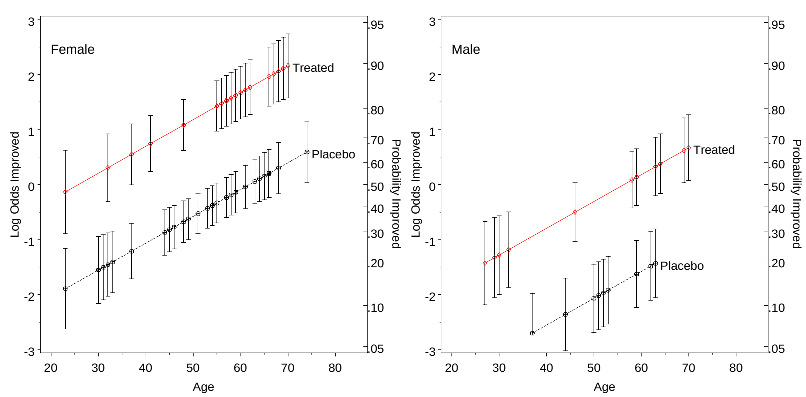
- Plot either logit or probability vs. one predictor (continuous or most levels)
- Separate curves for one factor
- Separate panels for all others (BY statement)

```
proc gplot data=results;
   plot (logit prob) * age = treat;
   by sex;
   symbol1 v=circle i=join l=3 c=black;
   symbol2 v=dot    i=join l=1 c=red;
```


PROC LOGISTIC: Model plots

Enhanced plots:

- Plot on logit scale, with probability scale at right (PSCALE macro)
- Show 67% error bars $\approx \pm 1$ se (BARS macro)
- Custom legend and panel labels (LABEL macro)



PROC LOGISTIC: Model plots

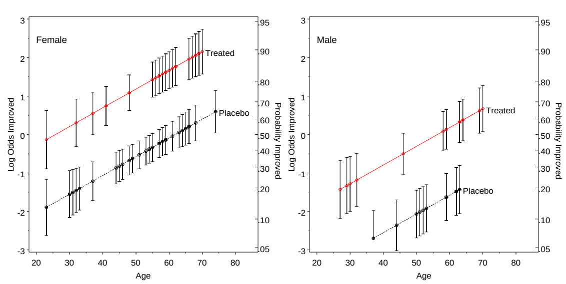
Enhanced plots:

··· glogist1c.sas ···

```
   *-- Error bars, on logit scale;
%bars(data=results, var=logit,
    class=age, cvar=treat, by=age,
    barlen=selogit, out=bars);

   *-- Custom legends and panel labels;
%label(data=results, y=logit, x=age, xoff=1, cvar=treat,
    by=sex, subset=last.treat, out=label1, pos=6, text=treat);
%label(data=results, y=2.5, x=20, size=2,
    by=sex, subset=first.sex, out=label2, pos=6, text=sex);

   *-- Probability scales at right;
%pscale(out=pscale,
    byvar=sex, byval=%str('Female','Male'));

   *-- Join ANNOTATE datasets;
data bars;
   set label1 label2 bars pscale;
proc sort;
   by sex;
```


··· glogist1c.sas

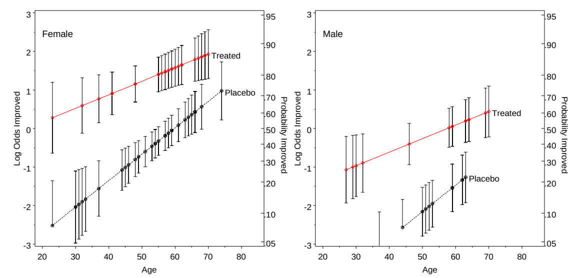
```
title ' '
    h=1.8 a=-90 'Probability Improved'    /* right axis label   */
    h=2.5 a=-90 ' ';                      /* extra space        */
goptions hby=0;                           /* suppress BY values */
proc gplot data=results;
    plot logit * age = treat /
        vaxis=axis1 haxis=axis2 hm=1 vm=1
        nolegend anno=bars frame;
    by sex;
    axis1 label=(a=90 'Log Odds Improved')
        order=(-3 to 3);
    axis2 order=(20 to 80 by 10) offset=(2,6);
    symbol1 v=+ i=join l=3 c=black;
    symbol2 v=- i=join l=1 c=red;
    label age='Age';
run;
```



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

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



Effect plots for generalized linear models

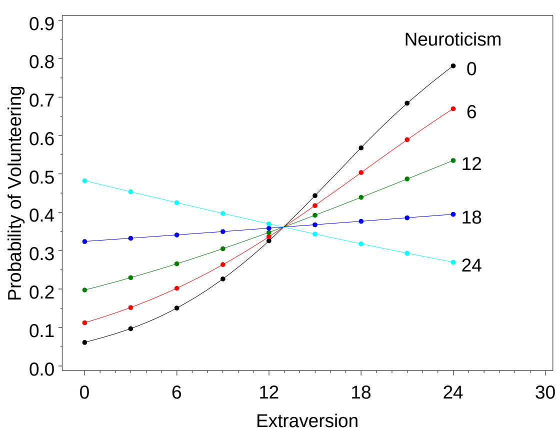
- For simple models, full model plots show the complete relation between response and all predictors.
- Fox (1987)— For complex models, often wish to plot a specific main effect or interaction (including lower-order relatives)— controlling for other effects
  - Fit full model to data with linear predictor (e.g., logit) $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ and link function $g(\boldsymbol{\mu}) = \boldsymbol{\eta} \rightarrow$ estimate $\boldsymbol{b}$ of $\boldsymbol{\beta}$ and covariance matrix $\widehat{V(\boldsymbol{b})}$ of $\boldsymbol{b}$.
  - 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," $\boldsymbol{X}^*$
  - Calculate fitted effect values, $\hat{\boldsymbol{\eta}}^* = \boldsymbol{X}^*\boldsymbol{b}$.
  - Standard errors are square roots of $\mathrm{diag}(\boldsymbol{X}^*\widehat{V(\boldsymbol{b})}\boldsymbol{X}^{*\mathsf{T}})$
  - Plot $\hat{\boldsymbol{\eta}}^*$, or values transformed back to scale of response, $g^{-1}(\hat{\boldsymbol{\eta}}^*)$.
- *Note*: This provides a general means to visualize interactions in *all* linear and generalized linear models.

Effect plots in SAS

- 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
- (Not yet a macro)

Effect plots: Example

- Cowles and Davis (1987)— Volunteering for a psychology experiment
  - Predictors: Sex, Neuroticism, Extraversion
  - ⟶ strong interaction, Neuroticism × Extraversion



Effect plots: Example

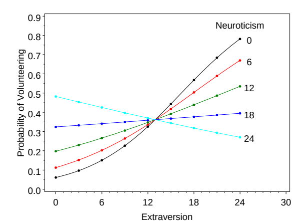
cowles3.sas

```
%include catdata(cowles);
%expgrid(Sex=0.5,            /* Fix Sex at mid value */
   Extraver=0 to 24 by 3, /* range of predictors  */
   Neurot=0 to 24 by 6,   /*    "          "      */
   _in_=1);               /* grid select value  */

*-- catenate grid values with data;
data cowles;
   set cowles _grid_;

*-- fit model, output fitted values;
proc logistic data=cowles outest=parm covout;
   model Volunter = Sex Extraver | Neurot  / covb;
   output out=predicted xbeta=logit stdxbeta=selogit
      p=prob u=upper l=lower / alpha=.33;

*-- select grid, replace labels;
data effect;
   set predicted (where =(_in_=1));
   label logit='log odds of Volunteering'
      prob = 'Probability of Volunteering';
```


Effect plots: Example

··· cowles3.sas

```
*-- Custom legend;
%label(data=effect,
   x=Extraver, y=prob,
   subset=Extraver=24,       /* at last.Extraver */
   text=put(Neurot,3.),      /* label text       */
   pos=6, xoff=.2, out=labels);

*-- Plot step;
proc gplot data=effect;
   plot prob * Extraver = Neurot /
      vaxis=axis1 haxis=axis2 vm=1
      anno=labels nolegend;
   symbol v=dot i=spline r=5;
   axis1 label=(a=90 r=0) order=(0 to .9 by .1);
   axis2 order=(0 to 30 by 6) offset=(3,1);
run; quit;
```



Effect plots with the `effects` package in R

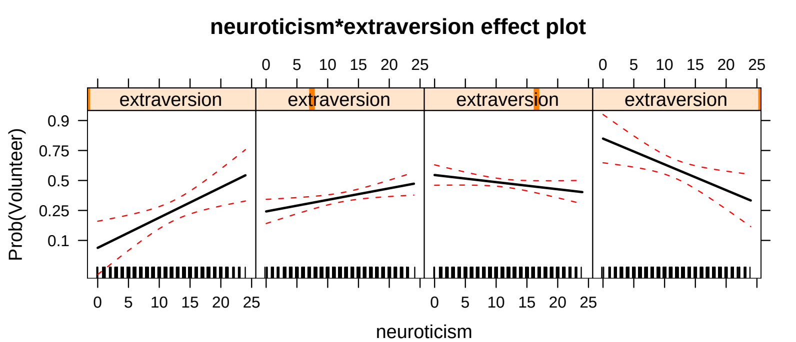
```
R> library(effects)  ## load the effects package
R> data(Cowles)
R> mod.cowles <- glm(volunteer ~ sex + neuroticism*extraversion,
+      data=Cowles, family=binomial)
R> summary(mod.cowles)
```

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

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1933.5  on 1420  degrees of freedom
Residual deviance: 1897.4  on 1416  degrees of freedom
AIC: 1907.4
```


Effect plots with the `effects` package in R

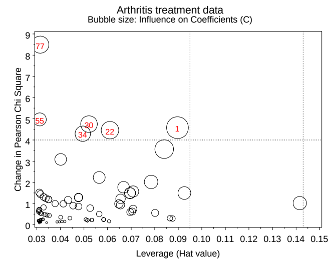
Calculate effects for all model terms, plot neuro:extra:

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



Influence measures and diagnostic plots

- **Leverage**: Potential impact of an individual case $\sim$ distance from the centroid in space of predictors
- **Residuals**: Which observations are poorly fitted?
- **Influence**: Actual impact of an individual case $\sim$ leverage $\times$ residual
  - ***C***, ***CBAR*** – analogs of Cook's D in OLS $\sim$ standardized change in regression coefficients when i-th case is deleted.
  - ***DIFCHISQ***, ***DIFDEV*** – $\Delta\chi^2$ when i-th case is deleted.



Influence measures and diagnostic plots

`PROC LOGISTIC` provides printed output with the `influence` and `iplots` options

```
proc logistic data=arthrit ;
   model better = sex treat age / influence iplots;
```

```
                     The LOGISTIC Procedure
                     Regression Diagnostics
          Deviance Residual               Hat Matrix Diagonal

Case             (1 unit = 0.26)                (1 unit = 0.01)
Number  Value   -8  -4   0 2 4 6 8     Value   0 2 4 6 8  12  16

   1    1.812  | *      |        |    0.089  |          *      |
   2    0.360  |        |*       |    0.031  |    *            |
   3    0.685  |        |  *     |    0.087  |          *      |
   4    0.425  |        | *      |    0.034  |    *            |
   5    0.700  |        |  *     |    0.086  |          *      |
   6    0.488  |        | *      |    0.038  |    *            |
   7    1.703  |  *     |        |    0.084  |         *       |
   8    0.499  |        | *      |    0.039  |    *            |
   9    1.396  |   *    |        |    0.066  |        *        |
  10    0.511  |        | *      |    0.040  |     *           |
  11    1.142  |    *   |        |    0.064  |       *         |
  12    0.523  |        | *      |    0.041  |     *           |
  13    1.234  |        |    *   |    0.065  |       *         |
  14    0.599  |        | *      |    0.051  |      *          |
  15    1.121  |    *   |        |    0.065  |       *         |
  16    0.599  |        | *      |    0.051  |      *          |
```

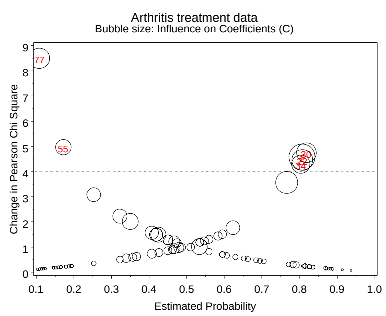
```
17    1.319  |          |    *    |    0.069  |        *         |
18    0.640  |          | *       |    0.058  |       *          |
19    1.319  |          |    *    |    0.069  |        *         |
20    0.640  |          | *       |    0.058  |       *          |
21    1.340  |          |    *    |    0.070  |        *         |
22    1.814  | *        |         |    0.061  |       *          |
23    1.022  |    *     |         |    0.070  |        *         |
24    0.529  |          | *       |    0.060  |       *          |
25    1.449  |          |    *    |    0.078  |         *        |
26    0.619  |          | *       |    0.053  |      *           |
27    0.909  |     *    |         |    0.080  |         *        |
...
```

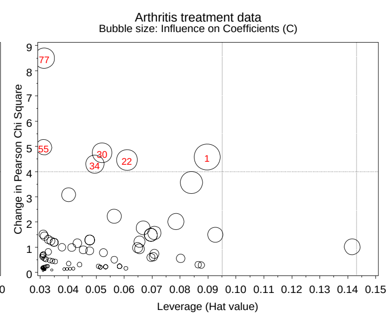
Problems:

- Way too much output
- Doesn't highlight unusual cases well
- Index plots don't consider combinations of measures

INFLOGIS macro

- Specialized version of INFLGLIM macro for logistic regression
- Plots a measure of change in χ^2 (DIFCHISQ or DIFDEV) vs. predicted probability or leverage.
- Bubble symbols show actual influence (C or CBAR)
- Shows standard cutoffs for "large" values
- Labels outlying cases





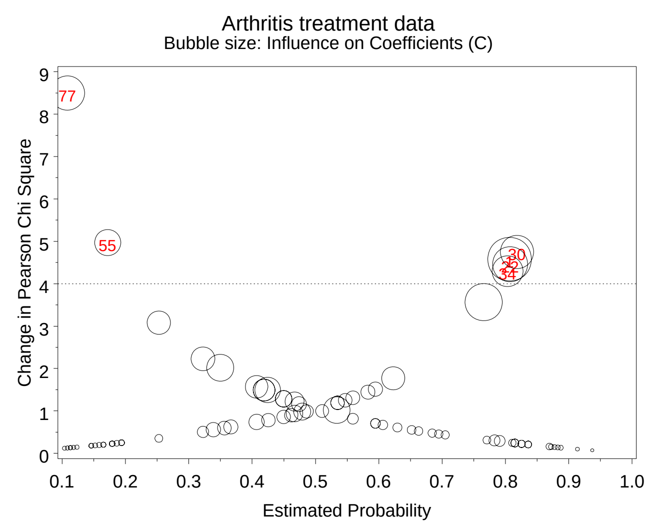
INFLOGIS macro: Example

```
%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 */
```

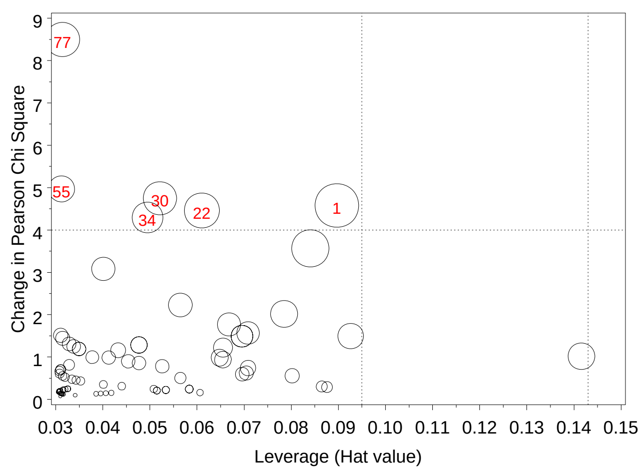
Printed output lists cases with "large" leverage, residual or influence:

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

INFLOGIS macro: Example



INFLOGIS macro: Example



Polytomous responses

- m categories $\longrightarrow$ $(m-1)$ comparisons (logits)
- **Response categories *ordered***, e.g., None, Some, Marked improvement
  - Proportional odds model
    - Uses adjacent-category logits

None	Some or Marked
None or Some	Marked

    - Assumes slopes are the same for all $m-1$ logits
  - Nested dichotomies

None	Some or Marked
Some	Marked

    - Model each logit separately
    - G^2 s are additive $\longrightarrow$ combined model
- **Response categories *unordered***, e.g., vote NDP, Liberal, Tory, Aliance
  - Multinomial logistic regression
    - Uses generalized logits (`LINK=GLOGIT`) in `PROC LOGISTIC` (V8.2+)
  - Nested dichotomies

Ordinal response: Proportional odds model

Arthritis treatment data:

Sex	Treatment	Improvement None	Some	Marked	Total
F	Active	6	5	16	27
F	Placebo	19	7	6	32
M	Active	7	2	5	14
M	Placebo	10	0	1	11

- Model logits for adjacent category cutpoints:

$$\text{logit}\,(\theta_{ij1}) = \log \frac{\pi_{ij1}}{\pi_{ij2} + \pi_{ij3}} = \text{logit (None vs. [Some or Marked])}$$

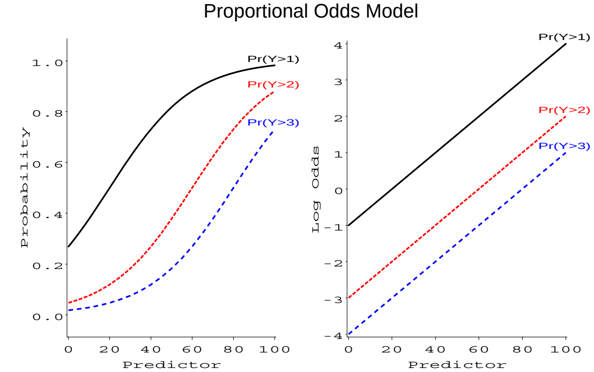
$$\text{logit}\,(\theta_{ij2}) = \log \frac{\pi_{ij1} + \pi_{ij2}}{\pi_{ij3}} = \text{logit ([None or Some] vs. Marked)}$$

- Consider a logistic regression model for each logit:

$$\text{logit}(\theta_{ij1}) = \alpha_1 + \boldsymbol{x}'_{ij}\,\boldsymbol{\beta}_1$$

$$\text{logit}(\theta_{ij2}) = \alpha_2 + \boldsymbol{x}'_{ij}\,\boldsymbol{\beta}_2$$

- Proportional odds assumption: ***regression functions are parallel*** on the logit scale i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.



Proportional odds model: Fitting and plotting

Similar to binary response models, except:

- Response variable has $m > 2$ levels; output dataset has `_LEVEL_` variable
- Must ensure that response levels are ordered as you want— use `order=data` or `descending` options.
- Validity of analysis depends on proportional odds assumption. Test of this assumption appears in `PROC LOGISTIC` output.

Example, using dependent variable `improve`, with values 0, 1, and 2:

glogist2a.sas ···

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

proc print data=results(obs=6);
   id id treat sex;
   var improve _level_ prob lower upper logit;
   format prob lower upper logit selogit 6.3;
run;
```


The response profile displays the ordering of the outcome variable.

Response Profile

Ordered Value	improve	Total Frequency
1	2	28
2	1	14
3	0	42

Test of Proportional Odds Assumption:

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
2.4916	3	0.4768

Parameter estimates:

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	2	1	-4.6826	1.1949	15.3566	<.0001
Intercept	1	1	-3.7836	1.1530	10.7680	0.0010
sex	Female	1	1.2515	0.5321	5.5330	0.0187
treat	Treated	1	1.7453	0.4772	13.3774	0.0003
age		1	0.0382	0.0185	4.2361	0.0396

Odds ratios:

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
sex	Female vs Male	3.496	1.232	9.918
treat	Treated vs Placebo	5.728	2.248	14.594
age		1.039	1.002	1.077

Output data set (RESULTS) for plotting:

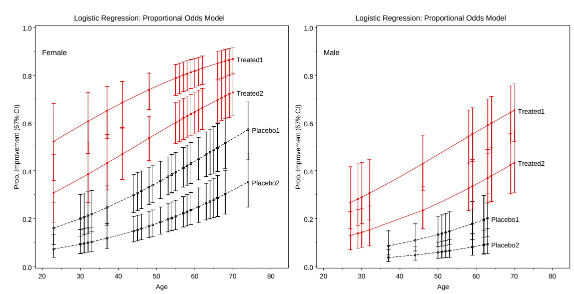
id	treat	sex	improve	_LEVEL_	prob	lower	upper	logit
57	Treated	Male	1	2	0.129	0.069	0.229	-1.907
57	Treated	Male	1	1	0.267	0.157	0.417	-1.008
9	Placebo	Male	0	2	0.037	0.019	0.069	-3.271
9	Placebo	Male	0	1	0.085	0.048	0.149	-2.372
46	Treated	Male	0	2	0.138	0.076	0.238	-1.830
46	Treated	Male	0	1	0.283	0.171	0.429	-0.931
...								

To plot predicted probabilities in a single graph, combine values of `TREAT` and `_LEVEL_`

··· glogist2a.sas ···

```
   *-- combine treatment and _level_, set error bar color;
data results;
   set results;
   treatl = trim(treat)||put(_level_,1.0);
   if treat='Placebo' then col='BLACK';
                      else col='RED';
proc sort data=results;
   by sex treatl age;
```

`··· plot prob * age = treatl; by sex;`



Add error bars and legends:

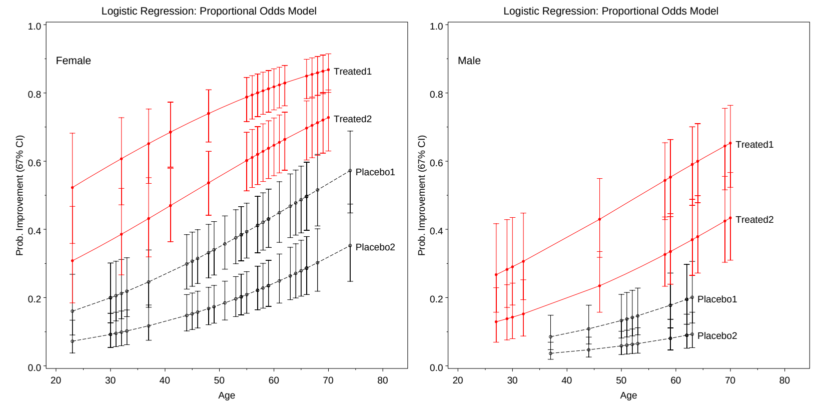
··· glogist2a.sas ···

```
   *-- Error bars, on prob scale;
%bars(data=results, var=prob,
   class=age, cvar=treatl, by=age,
   lower=lower, upper=upper,
   color=col, out=bars);
proc sort data=bars;
   by sex treatl age;

   *-- Custom legends, for treat-level and sex;
%label(data=results, y=prob, x=age, xoff=1, cvar=treatl,
   by=sex, subset=last.treatl, out=label1, pos=6, text=treatl);
%label(data=results, y=0.9, x=20, size=2,
   by=sex, subset=first.sex, out=label2, pos=6, text=sex);

   *-- Combine the annotate data sets;
data bars;
   set label1 label2 bars;
   by sex;
```


Plot step:

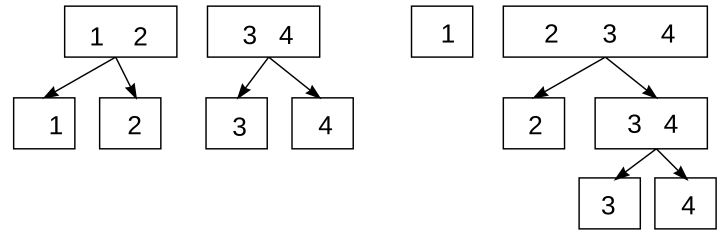
··· glogist2a.sas

```
goptions hby=0;
proc gplot data=results;
   plot prob * age = treatl /
        vaxis=axis1 haxis=axis2 hminor=1 vminor=1
        nolegend anno=bars name=glogist2a';
   by sex;
   axis1 label=(a=90 'Prob. Improvement (67% CI)')
         order=(0 to 1 by .2);
   axis2 order=(20 to 80 by 10)
         offset=(2,5);
   symbol1 v=circle  i=join line=3 c=black;
   symbol2 v=circle  i=join line=3 c=black;
   symbol3 v=dot     i=join line=1 c=red;
   symbol4 v=dot     i=join line=1 c=red;
run;
```

Polytomous response: Nested dichotomies

- m categories $\rightarrow (m-1)$ comparisons (logits)
- If these are formulated as $(m-1)$ nested dichotomies:
  - Each dichotomy can be fit using the familiar binary-response logistic model,
  - the $m-1$ models will be statistically independent (G^2 statistics will be additive)



Example: Women's Labour-Force Participation

Data: *Social Change in Canada Project* , York ISR (Fox, 1997)

- **Response:** not working outside the home (n=155), working part-time (n=42) or working full-time (n=66)
- Model as two nested dichotomies:
  - Working (n=106) vs. NotWorking (n=155)
  - Working full-time (n=66) vs. working part-time (n=42).
- **Predictors:**
  - Children? — 1 or more minor-aged children
  - Husband's Income — in $1000s
  - Region of Canada (not considered here)

Example: Women's Labour-Force Participation

wlfpart.sas

```
proc format;
   value labour    /* labour-force participation */
      1 ='working full-time'  2 ='working part-time'
      3 ='not working';
   value kids      /* children in the household */
      0 ='Children absent'  1 ='Children present';
data wlfpart;
   input case labour husinc children region;
   working = labour < 3;
   if working then
      fulltime = (labour = 1);
datalines;
  1  3  15  1  3
  2  3  13  1  3
  3  3  45  1  3
  4  3  23  1  3
  5  3  19  1  3
  6  3   7  1  3
  7  3  15  1  3
  8  1   7  1  3
  9  3  15  1  3
  ... more data lines ...
```


Example: Women's Labour-Force Participation

First, try proportional odds model for `labour`

```
proc logistic data=wlfpart;
   model labour  = husinc children;
   title2 'Proportional Odds Model: Fulltime/Parttime/NotWorking';
```

The score test *rejects* the Proportional Odds Assumption

```
          Score Test for the Proportional Odds Assumption

                Chi-Square       DF      Pr > ChiSq
                   18.5638        2          <.0001
```


Fit separate models for each of `working` and `fulltime`:

```
proc logistic data=wlfpart nosimple descending;
   model working = husinc children ;
   output out=resultw p=predict xbeta=logit;
   title2 'Nested Dichotomies';

proc logistic data=wlfpart nosimple descending;
   model fulltime = husinc children ;
   output out=resultf p=predict xbeta=logit;
```

- `descending` option used to model the $\Pr(Y = 1)$
- `output` statement $\longrightarrow$ datasets for plotting

Output for WORKING dichotomy:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Odds Ratio
INTERCPT	1	1.3358	0.3838	12.1165	0.0005	.
HUSINC	1	-0.0423	0.0198	4.5751	0.0324	0.959
CHILDREN	1	-1.5756	0.2923	29.0651	0.0001	0.207

Output for FULLTIME dichotomy:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Odds Ratio
INTERCPT	1	3.4778	0.7671	20.5537	0.0001	.
HUSINC	1	-0.1073	0.0392	7.5063	0.0061	0.898
CHILDREN	1	-2.6515	0.5411	24.0135	0.0001	0.071

$$\log\left(\frac{\Pr(\text{working})}{\Pr(\text{not working})}\right) = 1.336 - 0.042\ \text{H\$} - 1.576\ \text{kids}$$

$$\log\left(\frac{\Pr(\text{fulltime})}{\Pr(\text{parttime})}\right) = 3.478 - 0.107\ \text{H\$} - 2.652\ \text{kids}$$

Combined tests for Nested Dichotomoies

- Nested dichotomies $\longrightarrow$ χ^2 tests and df for the separate logits are independent
- $\longrightarrow$ add, to give tests for the full m-level response

Global tests of BETA=0

Test	Response	ChiSq	DF	Prob ChiSq
Likelihood Ratio	working	36.4184	2	<.0001
	fulltime	39.8468	2	<.0001
	ALL	76.2652	4	<.0001

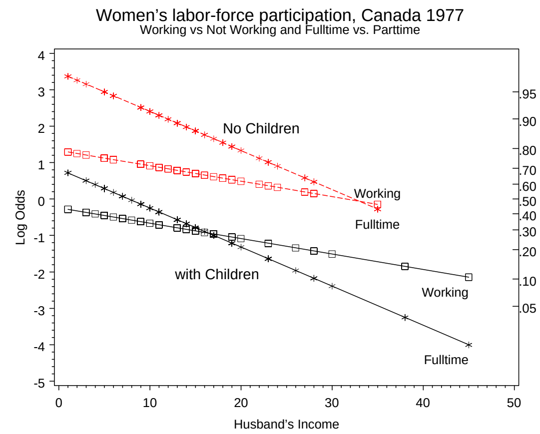
... (Score & Wald tests deleted) ...

Wald tests of maximum likelihood estimates

Variable	Response	WaldChiSq	DF	Prob ChiSq
Intercept	working	12.1164	1	0.0005
	fulltime	20.5536	1	<.0001
	ALL	32.6700	2	<.0001
children	working	29.0650	1	<.0001
	fulltime	24.0134	1	<.0001
	ALL	53.0784	2	<.0001
husinc	working	4.5750	1	0.0324
	fulltime	7.5062	1	0.0061
	ALL	12.0813	2	0.0024

Model visualization

- Join output datasets (`resultsw` and `resultsf`)
- Combine Response & Children $\longrightarrow$ `event`
- `plot logit * husinc = event;` $\longrightarrow$ separate lines



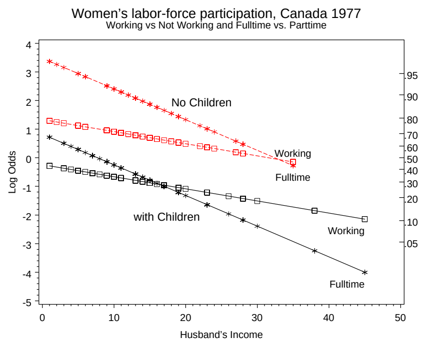
Model visualization

- Join output datasets (`resultsw` and `resultsf`)
- Combine Response & Children $\longrightarrow$ `event`

```
*-- Join the results datasets to create one plot;
data both;
   set resultw(in=inw)      /* working  */
       resultf(in=inf);     /* fulltime */
   if inw then do;
      if children=1 then event='Working, with Children ';
      else event='Working, no Children ';
   end;
   else do;
      if children=1 then event='Fulltime, with Children ';
      else event='Fulltime, no Children ';
   end;
```

Model visualization

```
proc gplot data=both;
   plot logit * husinc = event /
        anno=lbl nolegend frame vaxis=axis1;
   axis1 label=(a=90 'Log Odds') order=(-5 to 4);
   title2 'Working vs Not Working and Fulltime vs. Parttime';
   symbol1 v=dot    h=1.5 i=join l=3 c=red;
   symbol2 v=dot    h=1.5 i=join l=1 c=black;
   symbol3 v=circle h=1.5 i=join l=3 c=red;
   symbol4 v=circle h=1.5 i=join l=1 c=black;
```



Polytomous response: Generalized Logits

- Models the probabilities of the m response categories as $m-1$ logits comparing each of the first $m-1$ categories to the last (reference) category.
- Logits for any pair of categories can be calculated from the $m-1$ fitted ones.
- With k predictors, $x_1, x_2, \ldots, x_k$, for $j = 1, 2, \ldots, m-1$,

$$\begin{aligned} L_{jm} \equiv \log\left(\frac{\pi_{ij}}{\pi_{im}}\right) &= \beta_{0j} + \beta_{1j}\, x_{i1} + \beta_{2j}\, x_{i2} + \cdots + \beta_{kj}\, x_{ik} \\ &= \boldsymbol{\beta}_j^{\mathsf{T}} \boldsymbol{x}_i \end{aligned}$$

  - One set of fitted coefficients, $\boldsymbol{\beta}_j$ for each response category except the last.
  - Each coefficient, β_{hj}, gives the effect on the log odds of a unit change in the predictor x_h that an observation belongs to category j vs. category m.
- Probabilities are calculated as:

$$\pi_{ij} = \frac{\exp(\boldsymbol{\beta}_j^{\mathsf{T}} \boldsymbol{x}_i)}{\sum_{i=1}^{m} \exp(\boldsymbol{\beta}_j^{\mathsf{T}} \boldsymbol{x}_i)}$$

Polytomous response: Generalized Logits

- SAS:
  - In V8.2+, can use `PROC LOGISTIC` with `LINK=GLOGIT` option.
    - output dataset $\longrightarrow$ fitted probabilities, $\widehat{\pi}_{ij}$ for all m categories
    - Overall tests and specific tests for each predictor, for all m categories

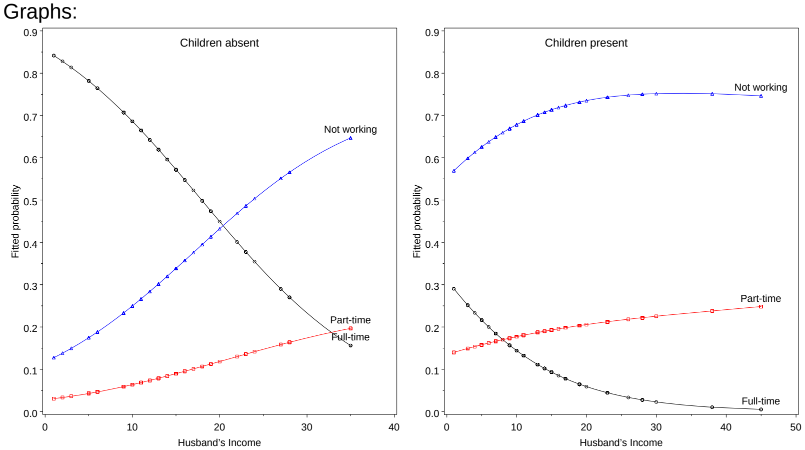
```
proc logistic data=wlfpart;
   model labor = husinc children / link=glogit;
   output out=results p=predict xbeta=logit;
```

  - `PROC CATMOD` with `RESPONSE=LOGITS` statement.
    - Same model, same predicted probabilities
    - Different syntax, output dataset format, plotting steps

```
proc catmod data=wlfpart;
   direct husinc;
   model labor = husinc children;
   response logits / out=results;
```


Example: Women's Labour Force Participation

Graphs:



Example: Women's Labour Force Participation

wlfpart5.sas ···

```
title 'Generalized logit model';
proc logistic data=wlfpart;
   model labor = husinc children / link=glogit;
   output out=results p=predict xbeta=logit;
```

Response profile:

Ordered Value	labor	Total Frequency
1	1	66
2	2	42
3	3	155

Logits modeled use labor=3 as the reference category.

Overall and Type III tests:

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	77.6106	4	<.0001
Score	76.4850	4	<.0001
Wald	58.4351	4	<.0001

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
husinc	2	12.8159	0.0016
children	2	53.9806	<.0001

These are comparable to the combined tests for the nested dichotomies models.

output dataset `results` (for plots):

case	labor	husinc	children	_LEVEL_	logit	predict
1	3	15	1	1	-2.03423	0.09333
1	3	15	1	2	-1.30743	0.19305
1	3	15	1	3	.	0.71363
2	3	13	1	1	-1.83977	0.11142
2	3	13	1	2	-1.32122	0.18715
2	3	13	1	3	.	0.70143
3	3	45	1	1	-4.95114	0.00528
3	3	45	1	2	-1.10067	0.24830
3	3	45	1	3	.	0.74642
4	3	23	1	1	-2.81207	0.04464
4	3	23	1	2	-1.25230	0.21238
4	3	23	1	3	.	0.74298
5	3	19	1	1	-2.42315	0.06486
5	3	19	1	2	-1.27987	0.20346
5	3	19	1	3	.	0.73168
6	3	7	1	1	-1.25639	0.18478
6	3	7	1	2	-1.36257	0.16616
...						

Example: Women's Labour Force Participation

··· wlfpart5.sas

```
proc sort data=results;   by children husinc _level_;

   *-- Curve labels;
%label(data=results, x=husinc, y=predict, cvar=_level_,
   by=children, subset=last._level_, text=put(_level_, labor.),
   pos=2, out=labels1);

   *-- Panel labels;
%label(data=results, x=20, y=0.85,
   by=children, subset=last.children, text=put(children, kids.),
   pos=2, size=2, out=labels2);
data labels;
   set labels1 labels2;
   by children;

goptions hby=0;
proc gplot data=results;   plot predict * husinc = _level_ /
      vaxis=axis1 hm=1 vm=1 anno=labels nolegend;
   by children;
   axis1 order=(0 to .9 by .1) label=(a=90);
   symbol1 i=join v=circle   c=black;
   symbol2 i=join v=square   c=red;
   symbol3 i=join v=triangle c=blue;
   run;
```

Conclusions

- **Summarization & exposure**
  - Effective data analysis requires *summarization*— hypothesis tests, model fits (& comparisons!), parameter estimates (& precision!)
  - Also requires *exposure*— displays to help the viewer see (& understand!) patterns, trends, and anomalies.
- **Graphical methods for categorical data**
  - Many new methods developed over the last 10–15 years
  - Some novel, others extend familiar methods for quantitative data
  - Described and illustrated in *VCD*
- **Theory into practice**
  - To be useful, statistical methods must be:
    - available— implemented in standard software
    - accessible— easy to use (or at least easier)
  - *VCD* provides ∼ 40 general macros and SAS/IML programs
  - The `vcd` package for R does the same for R users.

References

Atkinson, A. C. Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68:13–20, 1981.

Bangdiwala, K. Using SAS software graphical procedures for the observer agreement chart. *Proceedings of the SAS User's Group International Conference*, 12:1083–1088, 1987.

Bickel, P. J., Hammel, J. W., and O'Connell, J. W. Sex bias in graduate admissions: Data from Berkeley. *Science*, 187:398–403, 1975.

Bowker, A. H. Bowker's test for symmetry. *Journal of the American Statistical Association*, 43: 572–574, 1948.

Cowles, M. and Davis, C. The subject matter of psychology: Volunteers. *British Journal of Social Psychology*, 26:97–102, 1987.

Fox, J. Effect displays for generalized linear models. In Clogg, C. C., editor, *Sociological Methodology, 1987*, pp. 347–361. Jossey-Bass, San Francisco, 1987.

Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications, Thousand Oaks, CA, 1997.

Friendly, M. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89:190–200, 1994.

Friendly, M. Conceptual and visual models for categorical data. *The American Statistician*, 49: 153–160, 1995.

Friendly, M. Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics*, 8(3):373–395, 1999.

Friendly, M. Multidimensional arrays in SAS/IML. In *Proceedings of the SAS User's Group International Conference*, volume 25, pp. 1420–1427. SAS Institute, 2000.

Friendly, M. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56(4):316–324, 2002.

Friendly, M. and Kwan, E. Effect ordering for data displays. *Computational Statistics and Data Analysis*, 37, 2002. In press.

Hartigan, J. A. and Kleiner, B. Mosaics for contingency tables. In Eddy, W. F., editor, *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, pp. 268–273. Springer-Verlag, New York, NY, 1981.

Hoaglin, D. C. and Tukey, J. W. Checking the shape of discrete distributions. In Hoaglin, D. C., Mosteller, F., and Tukey, J. W., editors, *Exploring Data Tables, Trends and Shapes*, chapter 9. John Wiley and Sons, New York, 1985.

Koch, G. and Edwards, S. Clinical efficiency trials with categorical data. In Peace, K. E., editor, *Biopharmaceutical Statistics for Drug Development*, pp. 403–451. Marcel Dekker, New York, 1988.

Landis, J. R. and Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174., 1977.

Ord, J. K. Graphical methods for a class of discrete distributions. *Journal of the Royal Statistical Society, Series A*, 130:232–238, 1967.

Tufte, E. R. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

Tukey, J. W. Some graphic and semigraphic displays. In Bancroft, T. A., editor, *Statistical Papers in Honor of George W. Snedecor*. Iowa State University Press, Ames, IA, 1972.

Tukey, J. W. *Exploratory Data Analysis*. Addison Wesley, Reading, MA, 1977.

van der Heijden, P. G. M. and de Leeuw, J. Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, 50:429–447, 1985.

Williams, D. A. Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, 36:181–191, 1987.