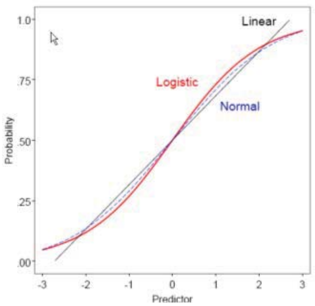
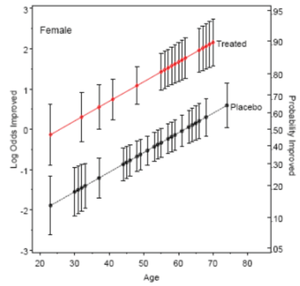
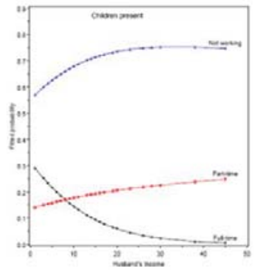
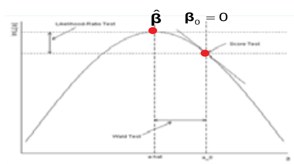


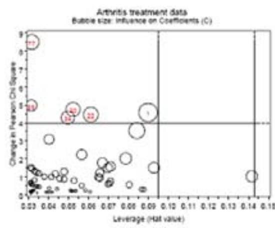




Logistic regression



Michael Friendly
Psychology 6140



Where to go from here?

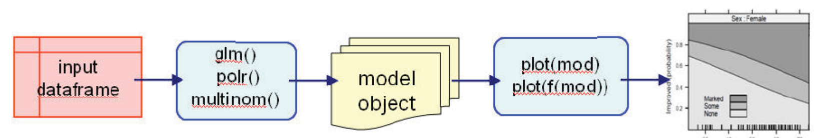
- What we've learned so far?
  - Tools for expressing, fitting, & understanding linear models, $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
  - Model selection methods
  - Model diagnostic methods
- So far this has been in the context of regression
- What we still have to learn?
  - Details for ANOVA
  - Multivariate extensions (MANOVA, MMReg)
  - Models for categorical responses
  - Other related methods
- Today, we'll consider one simple extension: categorical responses

Models for quantitative and categorical variables

	Dependent variables	
Independent variables	*Quantitative* $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta}$	*Categorical* $g(\mathbf{y}) = \mathbf{X}\,\boldsymbol{\beta}$
Quantitative	Regression	Logistic regression $\log\left(\frac{p}{1-p}\right) = \mathbf{X}\boldsymbol{\beta}$
Categorical (factors)	ANOVA	Loglinear models $\log(\mathbf{f}) = \mathbf{X}\,\boldsymbol{\beta}$
Both	Reg. w/ dummy vars ANCOVA Homogeneity of regression	General linear logistic model

Fitting & graphing in R

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

Logistic regression

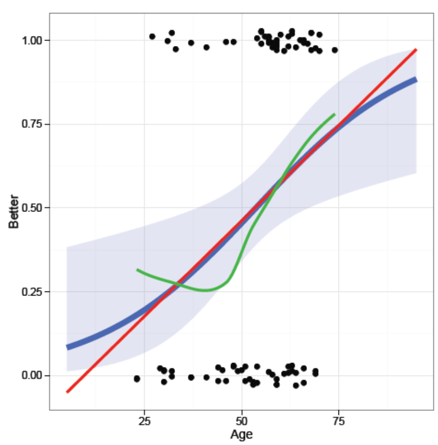
- The classical linear model assumes the response, Y, to be a quantitative variable
- In some cases, however, the response is categorical or dichotomous (binary outcome):
  - Improve vs. no improvement after treatment
  - Patient lives vs. dies
  - Applicant succeeds vs. fails
- Polytomous responses (later):
  - Improve: None, Some, Marked (ordered)
  - Women's paid work: none, part-time, full-time
  - Vote for: NDP, Liberal, Tories, Green (unordered)

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

For explanatory variables, this is the same as in ordinary linear models

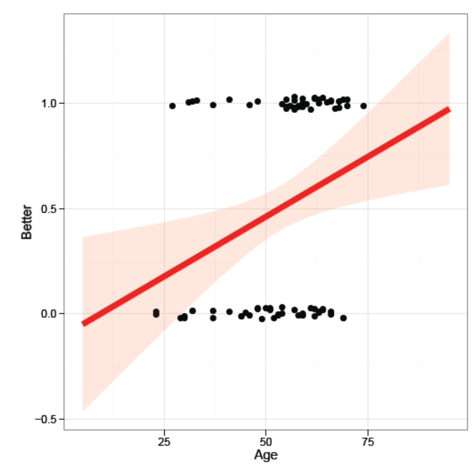
Arthritis treatment data



- 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

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: $\mathcal{V}(\hat{\pi}) = \hat{\pi}(1 - \hat{\pi}) \neq$ constant
  - Inferences, hypothesis tests are wrong!



OLS vs. Logistic

OLS regression:
- Assume y|x ~ N(0, σ^2)

Logistic regression:
- Assume Pr(y=1|x) ~ binomial(p)

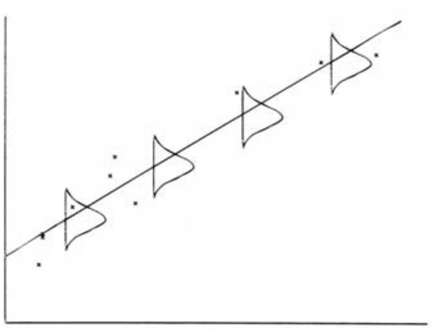


Fig. 2.1. Graphical representation of a simple linear normal regression.

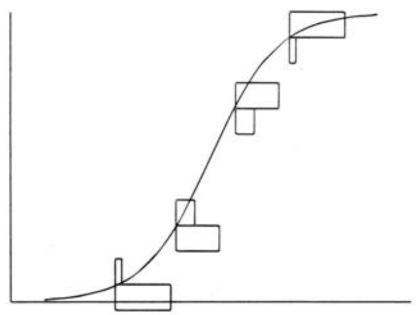
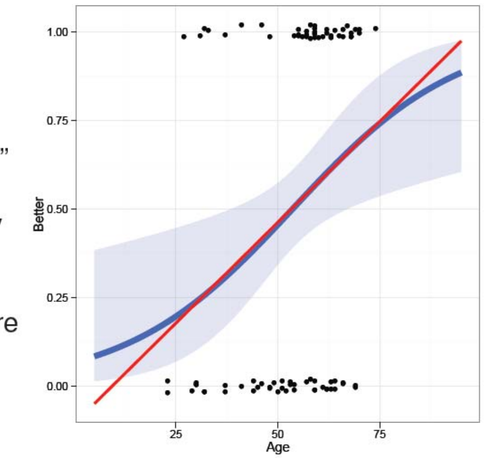


Fig. 2.2. Graphical representation of a simple linear logistic regression.

Logistic regression: binary response

- Logistic regression avoids these problems
- Models $\text{logit}(\pi_i) \equiv \log[\pi/(1-\pi)]$
- 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.



Probabilities, odds & logits

π=Prob(y=1)	Odds = π/(1- π)	Logit = log[π/(1- π)]
.05	5/95 = 0.0526	-2.94
.10	1/9 = 0.1111	-2.20
.30	3/7 = 0.4286	-0.85
.50	5/5 = 1	0.00
.70	7/3 = 2.333	0.85
.90	9/1 = 9	2.20
.95	95/5 = 19	2.94

- Prob: symmetric around p=0.5
- Logit: symmetric around logit(p) = 0

Logistic regression: One predictor

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,

$$\text{logit}[\pi(x)] \equiv \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \ .$$

- 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 $\pi(x)$

$$\pi(x) = \text{logit}^{-1}[\pi(x)] = \frac{1}{1+\exp[-(\alpha+\beta x)]}$$

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

$$\text{odds}(Y = 1) \equiv \frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x} \ . \qquad (1)$$

Thus:

- β is the change in the log odds associated with a unit increase in x.
- The odds are multiplied by e^{β} for each unit increase in x.
- α is log odds at $x = 0$; e^{α} is the odds of a favorable response at this x-value.
- In R, use `exp(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 $\pi = \frac{1}{2}$

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:

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

Fit and print:

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

##
## Call:  glm(formula = Better ~ Age, family = binomial, data = Arthritis)
##
## Coefficients:
## (Intercept)          Age
##     -2.6421       0.0492
##
## Degrees of Freedom: 83 Total (i.e. Null);  82 Residual
## Null Deviance:     116
## Residual Deviance: 109  AIC: 113
```


The `summary()` method gives details:

```
summary(arth.logistic)

##
## Call:
## glm(formula = Better ~ Age, family = binomial, data = Arthritis)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5106  -1.1277   0.0794   1.0677   1.7611
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.6421     1.0732   -2.46    0.014 *
## Age           0.0492     0.0194    2.54    0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 116.45  on 83  degrees of freedom
## Residual deviance: 109.16  on 82  degrees of freedom
## AIC: 113.2
##
## Number of Fisher Scoring iterations: 4
```

Why do I get z tests? Where is my R^2?

Interpreting coefficients

```
coef(arth.logistic)

## (Intercept)         Age
##   -2.642071    0.049249
```

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

## (Intercept)         Age
##    0.071214    1.050482
```

```
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 $\pi = \frac{1}{2}$)

Logistic regression: Multiple predictors

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

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

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

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

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

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

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

Arthritis data: Multiple predictors

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

$$\operatorname{logit}(\pi_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_2 x_{i2}$$

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.

$$x_2 = \begin{cases} 0 & \text{if Female} \\ 1 & \text{if Male} \end{cases} \qquad x_3 = \begin{cases} 0 & \text{if Placebo} \\ 1 & \text{if Treatment} \end{cases}$$

- α doesn't have a sensible interpretation here. Why?
- β_1: increment in log odds(Better) for each year of age.
- β_2: difference in log odds for male as compared to female.
- β_3: difference in log odds for treated vs. the placebo group

Fit the main effects model. Use `I(Age-50)` to center Age, making α interpretable.

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

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

```
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
```

NB: Female is ref. category for Sex; Placebo for Treatment

Interpreting coefficients

```
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

Estimation & hypothesis tests

- Ordinary regression model is fit by least squares, because it has optimal properties
  - Unbiased: E(**b**) = β
  - Consistent: **b** → β as N → ∞
  - Minimum variance: Var(**b**) ≤ any other method
- These properties are attained for logistic regression when fit by maximum likelihood
  - Overall ***F*** tests → L.R. χ^2 tests
  - Partial ***t*** tests → Wald χ^2 or z tests
  - max. likelihood used almost everywhere else, other than classical regression/ANOVA

Maximum likelihood estimation

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

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

- Assuming independence, joint likelihood is product over all cases

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

- Find estimates that maximize $\mathcal{L}$ but simpler for log $\mathcal{L}$

$$\sum_i Y_i x_{ik} = \sum \hat{p}_i x_{ik} \quad \Rightarrow \quad \mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\mathbf{p}}$$

Analogous to linear model,

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \hat{\mathbf{y}}$$

Hypothesis testing: Questions

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

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

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

$$H_0 : \beta_k = 0 \quad \text{given other predictors retained}$$

- **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 → likelihood ratio G^2 tests
- *t*-tests → Wald *z* or χ^2 tests

Hypothesis tests

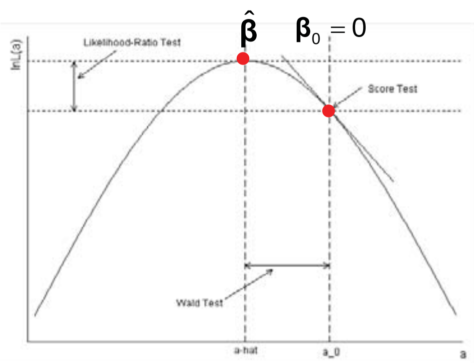
- Likelihood ratio test (G^2)
  - Compare *nested* models, similar to incremental F tests in OLS
  - Let $\mathcal{L}_1$ = maximized likelihood for **our** model $\text{logit}(\pi_i) = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}$ w/ *k* predictors
  - Let $\mathcal{L}_0$ = maximized likelihood for **null**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

$$G^2 = -2 \log\left(\frac{L_0}{L_1}\right) = 2(\log L_1 - \log L_0) \sim \chi_k^2$$

Other tests: Wald, score

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.3859	3	<.0001
Score	22.0051	3	<.0001
Wald	17.5147	3	0.0006

Testing Global Null Hypothesis: BETA=0



Different ways to measure departure from H_0 : $\beta = 0$

- LR test: diff in log L
- Wald test: $(\hat{\beta} - \beta_0)^2$
- Score test: slope at $\beta = 0$

Wald tests & confidence intervals

- Analogous to *t*-tests in OLS
- H_0: $\beta_i = 0$

$$z = \frac{b_k}{s(b_k)} \sim \mathcal{N}(0,1) \quad \text{or} \quad z^2 \sim \chi^2_1$$

(Wald chi-square)

- Confidence interval:

$$b_k \pm z_{1-\alpha/2} s(b_k)$$

e.g.,

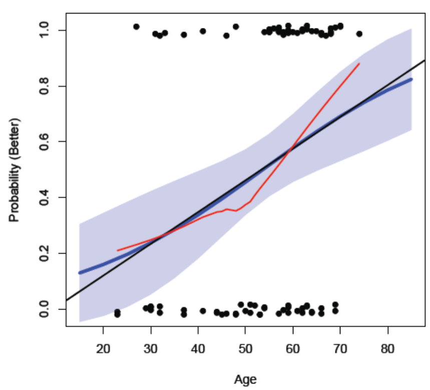
Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald 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

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 overplottting.

- 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



Full-model plots

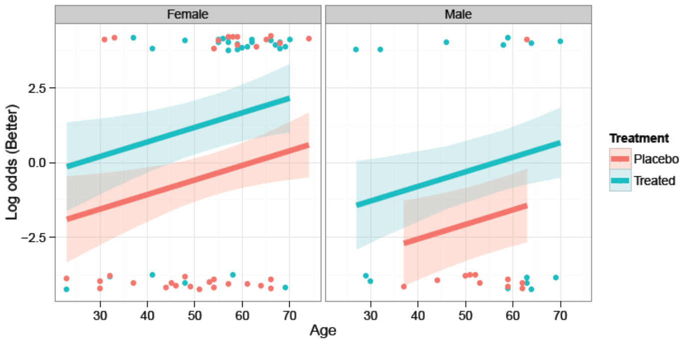
Full-model plots show the fitted values on the logit scale or on the response scale (probability), usually with confidence bands. This often requires a bit of custom programming.
Steps:

- 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()`, · · ·

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

##   ID Treatment  Sex Age Improved Better   fit se.fit
## 1 57   Treated Male  27     Some      1 -1.43  0.758
## 2 46   Treated Male  29     None      0 -1.33  0.728
## 3 77   Treated Male  30     None      0 -1.28  0.713
## 4 17   Treated Male  32   Marked      1 -1.18  0.684
```

Full-model plots

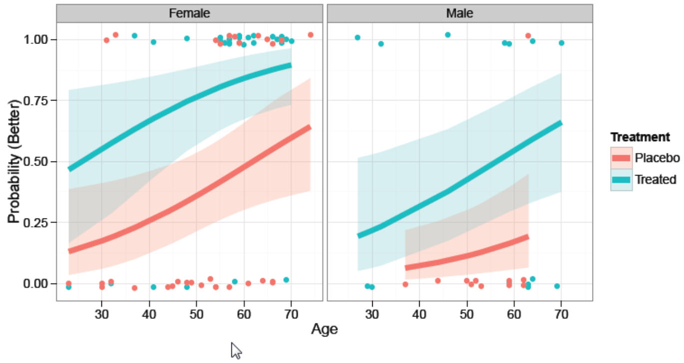
Ploting on the logit scale shows the additive effects of age, treatment and sex



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

Full-model plots

Ploting on the probability scale may be simpler to interpret



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

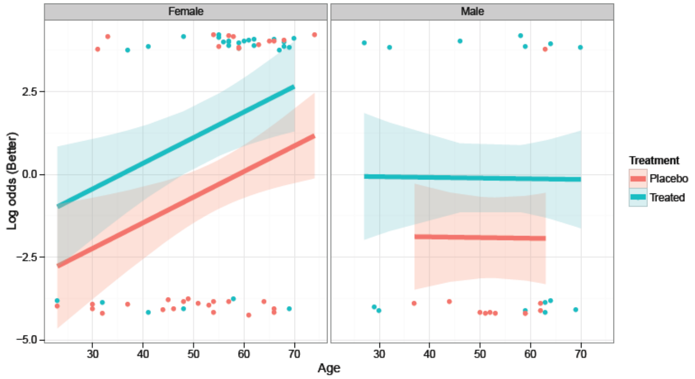
Models with interactions

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
## Sex           6.98  1    0.00823 **
## Treatment    11.90  1    0.00056 ***
## Sex:Age       3.42  1    0.06430 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

Models with interactions

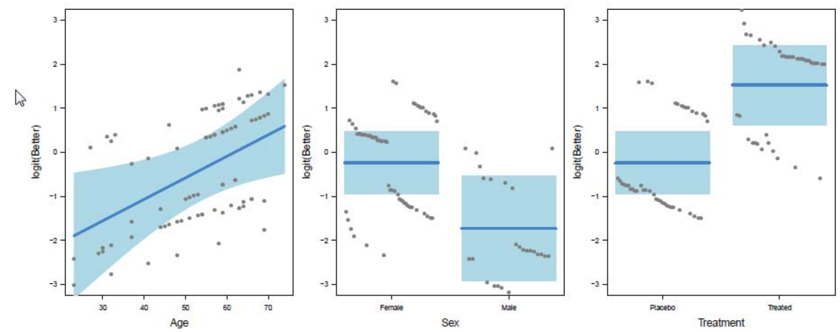


- Only the model changes
- `predict()` automatically incorporates the revised model terms
- Plotting steps remain the same
- This interpretation is quite different!

Complex models: `visreg` and `effects` packages

- 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


```
library(visreg)
visreg(arth.logistic2, ylab="logit(Better)", ...)
```



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

Effect plots: basic ideas

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

Data

	x1	x2	sex	x1x2	y	yhat
1	1	1	F	1	4.73	4.46
2	2	1	M	0	6.10	5.55
3	3	1	F	-1	4.32	4.34
4	1	1	F	1	4.84	4.46
5	2	1	F	0	4.73	4.40
...	...	...	...	...	...	...
29	2	2	M	0	6.10	6.15
30	3	2	F	1	6.71	7.14

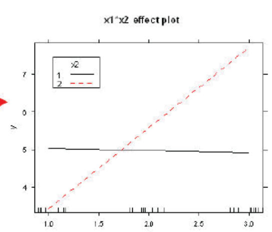
- Fit data: $\mathbf{X}\hat{\boldsymbol{\beta}} \Rightarrow \hat{\mathbf{y}}$
- Score data $\mathbf{X}^*\hat{\boldsymbol{\beta}} \Rightarrow \hat{\mathbf{y}}^*$
  - plot vars: vary over range
  - control vars: fix at means

Score data

	x1	x2	sex	x1:x2	y	yhat*
31	1	1	0.5	1	NA	5.030
32	2	1	0.5	2	NA	4.971
33	3	1	0.5	3	NA	4.912
34	1	2	0.5	2	NA	3.437
35	2	2	0.5	4	NA	5.574
36	3	2	0.5	6	NA	7.710

plot vars — control vars

plot

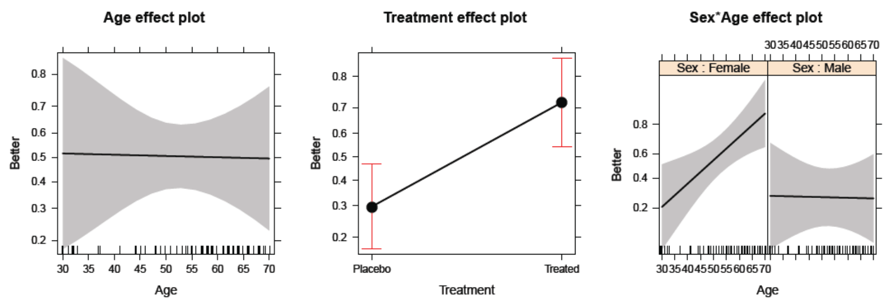


Plotting main effects:

```
library(effects)
arth.eff2 <- allEffects(arth.logistic2)
plot(arth.eff2, rows=1, cols=3)
```

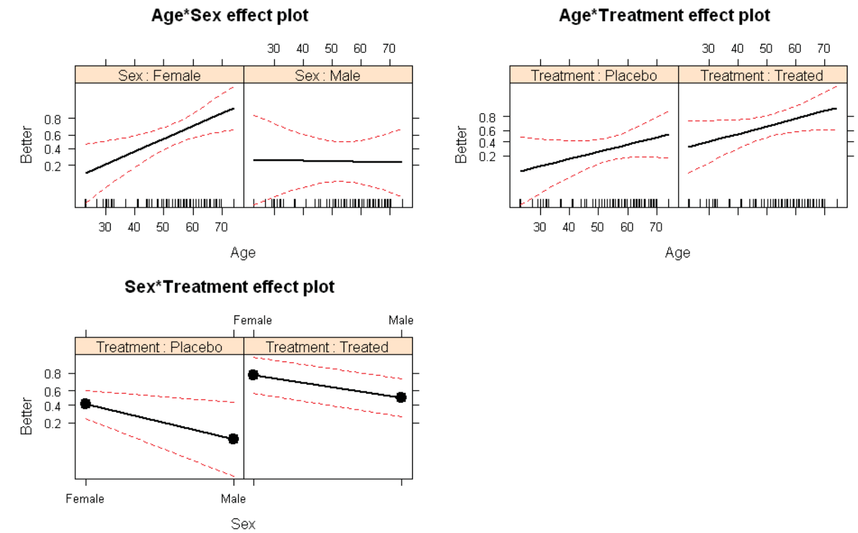
Model with interaction of Age x Sex

```
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


```
# fit model with all 2-way interactions
arth.logistic2 <- glm(Better ~ (Age + Sex + Treatment)^2, data=Arthritis,
                       family=binomial)
eff.logistic2<-allEffects(arth.logistic2)
plot(eff.logistic2)
```



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 $\Pr(Y = 1)$
- Important interactions have been omitted
- A few “wild” observations have a large impact on the fitted model or coefficients

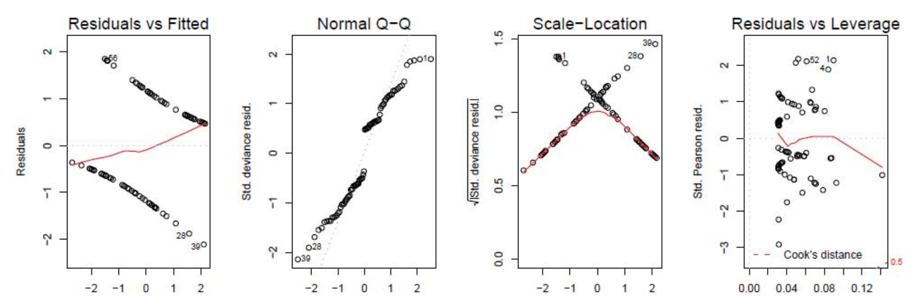
Model specification: Tools and techniques

- 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

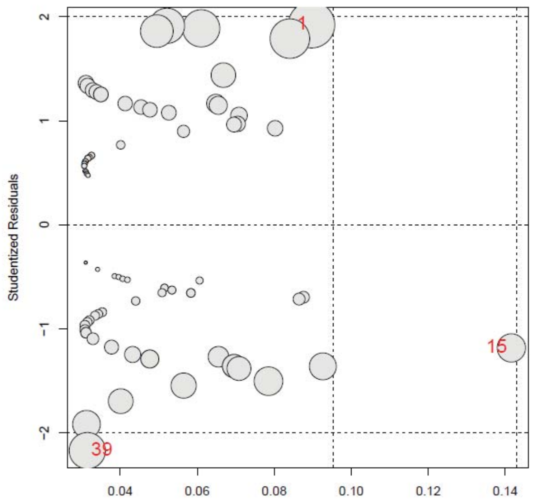
```
arth.mod1 <- glm(Better ~ Age + Sex + Treatment, data=Arthritis,
                 family='binomial')
plot(arth.mod1)
```



Better versions of these plots are available in the car package

Influence plots in R

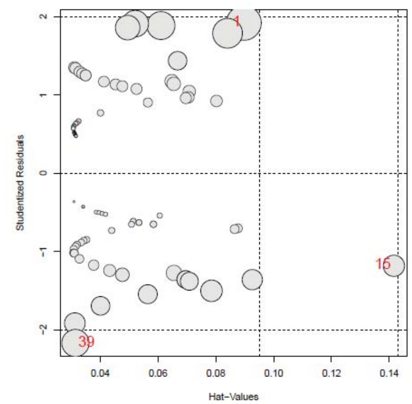
```
library(car)
influencePlot(arth.logistic2)
```



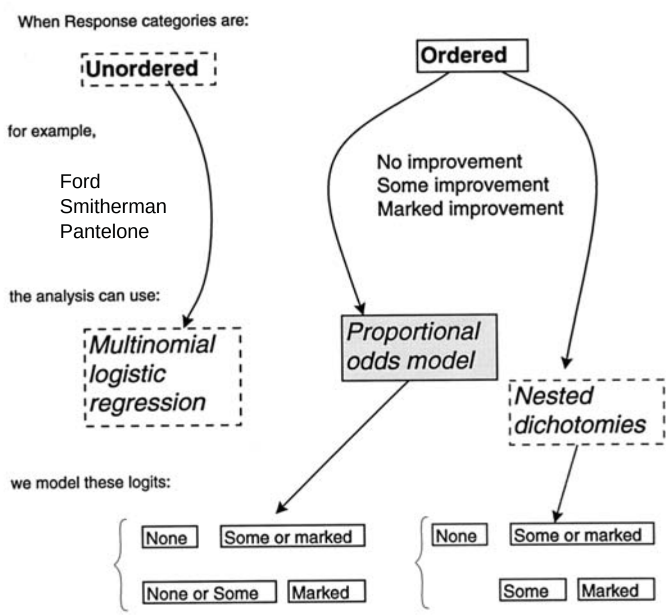
- X axis: Leverage ("hat values")
- Y axis: Studentized residuals
- Bubble size ∼ Cook D (influence on coefficients)

Which cases are influential?

	ID	Treatment	Sex	Age	Better	StudRes	Hat	CookD
1	57	Treated	Male	27	1	1.922	0.08968	0.3358
15	66	Treated	Female	23	0	-1.183	0.14158	0.2049
39	11	Treated	Female	69	0	-2.171	0.03144	0.2626



Polytomous responses: Overview



Polytomous responses: Overview

- m categories $\rightarrow$ $(m-1)$ 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; only intercepts vary
  - Nested dichotomies [None | Some or Marked], [Some | Marked]
    - Model each logit separately
    - G^2 s are additive $\rightarrow$ 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:

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

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

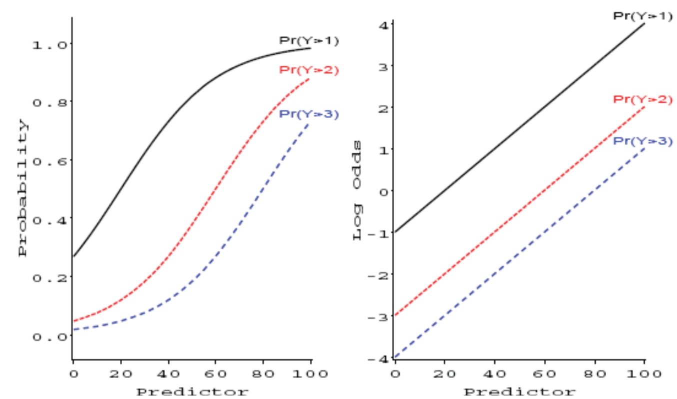
- Consider a logistic regression model for each logit:

$$\text{logit}(\theta_{ij1}) = \alpha_1 + \boldsymbol{x}'_{ij}\, \beta_1$$

$$\text{logit}(\theta_{ij2}) = \alpha_2 + \boldsymbol{x}'_{ij}\, \beta_2$$

- Proportional odds assumption: ***regression functions are parallel*** on the logit scale i.e., $\beta_1 = \beta_2$.

Proportional Odds Model



i.e., logits for adjacent response comparisons differ only in intercept

→ if true, simplifies interpretation

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

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

Proportional odds model: fitting & 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

                                        Point          95% Wald
     Effect                          Estimate      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
```

e^{β_i}

Output data set (RESULTS) for plotting:

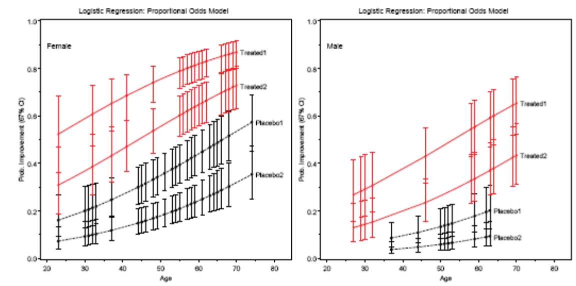
id	treat	sex	improve	_LEVEL_	prob	lower	upper	logit
57	Treated	Male	1	2	0.129	0.069	0.229	-1.907
57	Treated	Male	1	1	0.267	0.157	0.417	-1.008
9	Placebo	Male	0	2	0.037	0.019	0.069	-3.271
9	Placebo	Male	0	1	0.085	0.048	0.149	-2.372
46	Treated	Male	0	2	0.138	0.076	0.238	-1.830
46	Treated	Male	0	1	0.283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

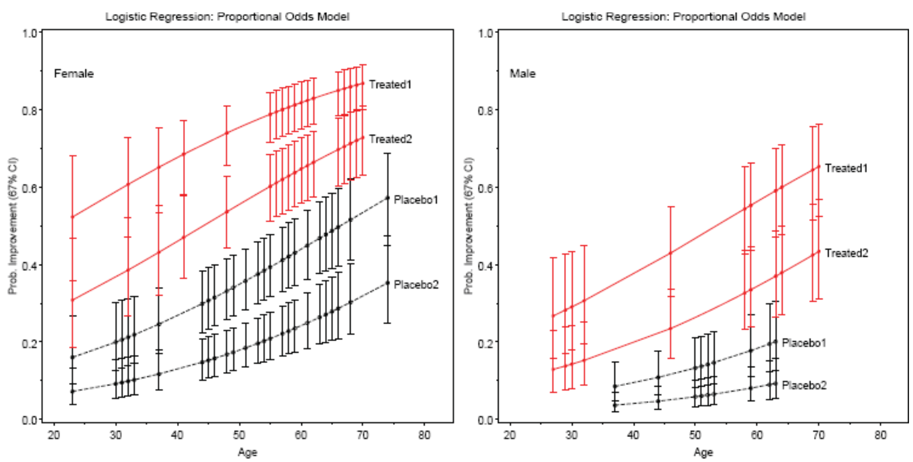
To plot predicted probabilities in a single graph, combine values of TREAT and _LEVEL_

··· 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;
```





Interpretation:

- Effects of age, treatment and sex are similar to what we saw before
- There are substantial differences among the 3 response categories. Intercept 2 here is for the distinction between none vs. (some, marked)

Proportional odds models in R

- **Fitting:** `polr()` in MASS package

The response, `Improved` has been defined as an *ordered* factor

```
data(Arthritis, package="vcd")
head(Arthritis$Improved)

## [1] Some   None   None   Marked Marked Marked
## Levels: None < Some < Marked
```

Fitting:

```
library(MASS)          # for polr()
library(car)           # for Anova()

arth.polr <- polr(Improved ~ Sex + Treatment + Age,
                  data=Arthritis)
summary(arth.polr)
Anova(arth.polr)       # Type II tests
```

Proportional odds models in R

The `summary()` function gives standard statistical results:

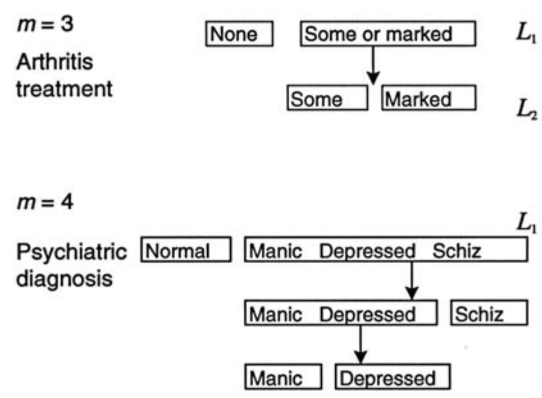
```
> summary(arth.polr)
```

```
Call:
polr(formula = Improved ~ Sex + Treatment + Age, data = Arthritis)

Coefficients:
                    Value   Std. Error   t value
SexMale          -1.25168     0.54636   -2.2909
TreatmentTreated  1.74529     0.47589    3.6674
Age               0.03816     0.01842    2.0722

Intercepts:
            Value   Std. Error t value
None|Some   2.5319  1.0571     2.3952
Some|Marked 3.4309  1.0912     3.1442

Residual Deviance: 145.4579
AIC: 155.4579
```

- Results are similar, but less convenient than proc logistic (no p-values)
- Test of proportional odds assumption requires the VGLM package

Nested dichotomies

For an *m* category response, the *m-1* comparisons can be represented by *m-1* logit models for a set of *m-1* nested dichotomies among the response categories.

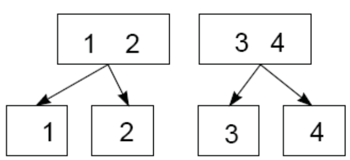


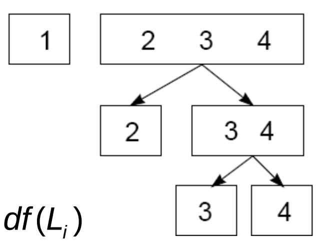
$m=3$: $L_1 = log \frac{\pi_1}{\pi_2 + \pi_3}$, $L_2 = log \frac{\pi_2}{\pi_3}$

$m=4$: $L_1 = log \frac{\pi_1}{\pi_2 + \pi_3 + \pi_4}$, $L_2 = log \frac{\pi_4}{\pi_2 + \pi_3}$, $L_3 = log \frac{\pi_2}{\pi_3}$

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)





$$G^2_{all} = \sum_{i}^{m-1} G^2(L_i) \qquad df_{all} = \sum df(L_i)$$

- This allows the slopes to differ for each logit
- Some hand-calculation is required for overall tests

Ex: 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)

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 ...
```

labour	working	fulltime
1	1	1
2	1	0
3	0	.

Ex: 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

NB: The score test is anti-conservative: *p*-values often too small. Use with caution.

Fitting nested dichotomies

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

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

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

→H$ and kids have **greater** impact on full vs. parttime choice than on working vs. not working

Combined tests for nested dichotomies

- Nested dichotomies → χ^2 tests and df for the separate logits are independent
- → 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

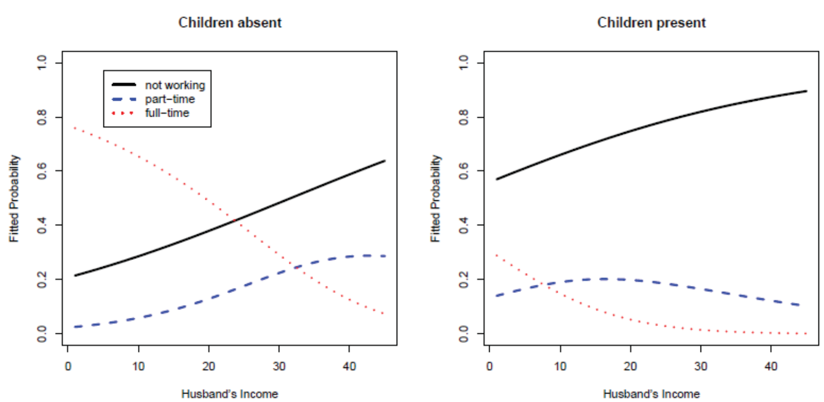
Model visualization

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

Model visualization

Alternatively, you can find the predicted probabilities for each response category and plot these in relation to the predictors.

This is often easier to interpret.



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 $j = 1, 2, \ldots, m-1$,

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

  - 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:

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

- 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

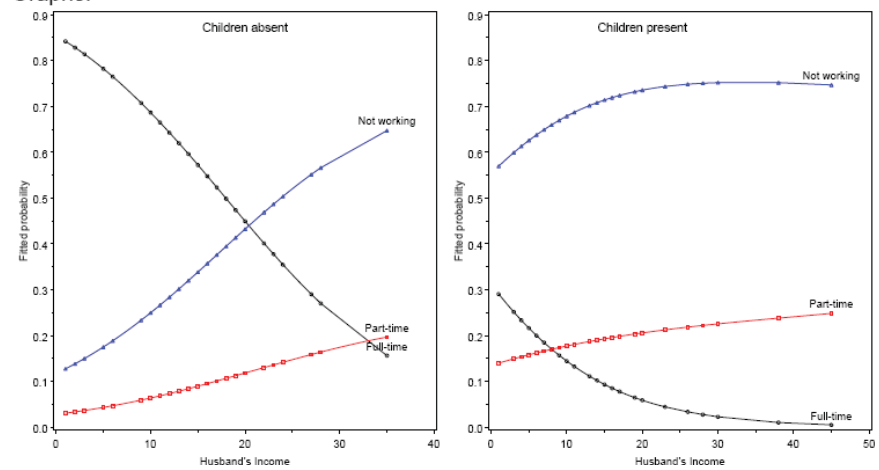
```
proc logistic data=wlfpart;
   model labor = husinc children / link=glogit;
   output out=results p=predict xbeta=logit;
```

  - `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;
```


Ex: Women's labour force participation

Graphs:



Fitting & plotting generalized logits

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

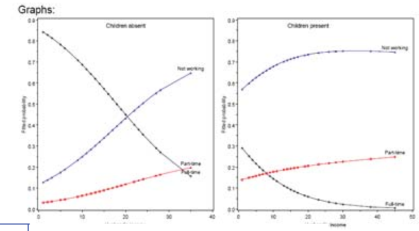
Effect	DF	Wald Chi-Square	Pr > ChiSq
husinc	2	12.8159	0.0016
children	2	53.9806	<.0001

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
...						

NB: for each case, predicted probabilities of labor category sum to 1



... 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;
```


Summary

- Logistic regression
  - Extends regression to case of binary response
  - Fit by max. likelihood, not OLS
    - *F* tests → L.R. χ^2 tests;
    - ***t*** tests → Wald χ^2
  - Interpretation of coefficients:
    - β_i = increment to log odds Y=1 for $\Delta x_i = 1$
    - $\exp(\beta_i)$ = multiplier of odds ratio
- Polytomous response
  - Ordered response categories:
    - Proportional odds model,
    - nested dichotomies
  - Unordered: Multinomial logistic regression