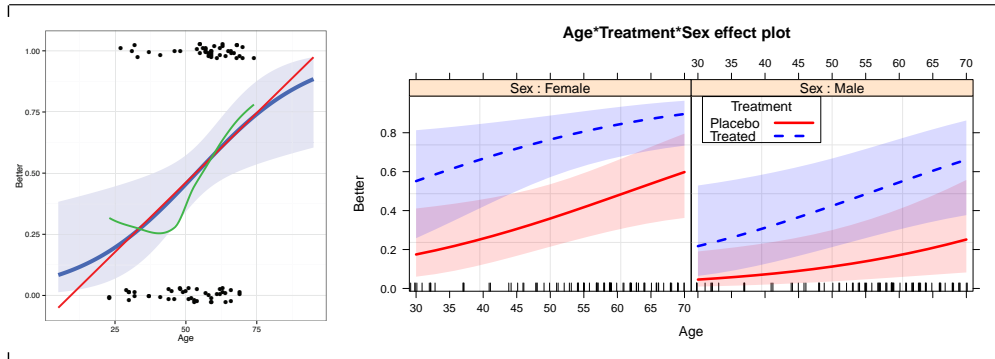


Michael Friendly

Psych 6136

November 1, 2017



### Structure

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

### Advantages

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

2/63

## loglm vs. glm

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

```
berk.mod1 <- loglm(~ Dept * (Gender + Admit), data=UCBAdmissions)
berk.mod2 <- loglm(~ (Admit + Dept + Gender)^2, data=UCBAdmissions)
anova(berk.mod2)

## Call:
## loglm(formula = ~(Admit + Dept + Gender)^2, data = UCBAdmissions)
##
## Statistics:
##              X^2 df  P(> X^2)
## Likelihood Ratio 20.204  5 0.0011441
## Pearson          18.823  5 0.0020740
```

Comparing models with **anova()** and **LRstats()**

```
anova(berk.mod1, berk.mod2, test="Chisq")

## LR tests for hierarchical log-linear models
##
## Model 1:
## ~Dept * (Gender + Admit)
## Model 2:
## ~(Admit + Dept + Gender)^2
##
##              Deviance df Delta(Dev) Delta(df) P(> Delta(Dev)
## Model 1          21.736  6
## Model 2          20.204  5      1.5312      1      0.21593
## Saturated         0.000  0      20.2043      5      0.00114

LRstats(berk.mod1, berk.mod2)

## Likelihood summary table:
##              AIC BIC LR Chisq Df Pr(>Chisq)
## berk.mod1  217 238    21.7  6    0.0014 **
## berk.mod2  217 240    20.2  5    0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## loglm vs. glm

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

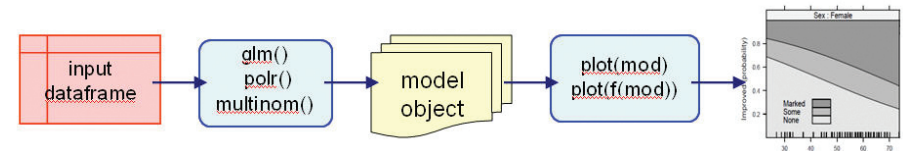
```
berkeley <- as.data.frame(UCBAdmissions)
berk.glm2 <- glm(Freq ~ (Dept+Gender+Admit)^2, data=berkeley,
                 family="poisson")
anova(berk.glm2, test="Chisq")

## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Freq
##
## Terms added sequentially (first to last)
##
##              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL              23      2650
## Dept              5       160      18      2491 <2e-16 ***
## Gender            1       163      17      2328 <2e-16 ***
## Admit             1       230      16      2098 <2e-16 ***
## Dept:Gender       5      1221      11       877 <2e-16 ***
## Dept:Admit        5       855       6       22  <2e-16 ***
## Gender:Admit      1        2       5       20  0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5/63

## Fitting and graphing: Overview

Object-oriented approach in R:



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

Plot methods:

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

6/63

## Objects and methods

How this works:

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

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

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

7/63

## Objects and methods

Methods for "glm" objects:

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

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

8/63

## Objects and methods

Some available `plot()` methods:

```
methods("plot")
```

```
## [1] plot,ANY-method      plot,color-method
## [3] plot.aareg*          plot.acf*
## [5] plot.ACF*            plot.augPred*
## [7] plot.bcnPowerTransform* plot.ca*
## [9] plot.coef.mer*       plot.compareFits*
## [11] plot.correspondence* plot.cox.zph*
## [13] plot.data.frame*     plot.decomposed.ts*
## [15] plot.default         plot.dendrogram*
## [17] plot.density*        plot.ecdf
## [19] plot.eff*            plot.efflist*
## [21] plot.effpoly*        plot.factor*
## [23] plot.formula*        plot.function
## [25] plot.gam*            plot.ggplot*
## [27] plot.gls*            plot.gnm*
## [29] plot.goodfit*        plot.gtable*
## [31] plot.hclust*         plot.histogram*
## [33] plot.HLtest*         plot.HoltWinters*
## [35] plot.intervals.lmList* plot.isoreg*
## [37] plot.jam*            plot.lda*
## [39] plot.lm*             plot.lme*
## [41] plot.lmList*         plot.lmList4*
## [43] plot.lmList4.confint* plot.loddsratio*
## [45] plot.loglm*          plot.mca*
## [47] plot.medpolish*      plot.merMod*
## [49] plot.mjca*           plot.mlm*
```

9/63

## Modeling approaches: Overview

### • Association models

#### • Loglinear models

(Contingency table form)

[Admit][GenderDept]

[AdmitDept][GenderDept]

[AdmitDept][AdmitGender][GenderDept]

#### • Poisson GLMs

(Frequency data frame)

Freq ~ Admit + Gender\*Dept

Freq ~ Admit\*Dept + Gender\*Dept

Freq ~ Admit\*Dept + Admit\*Gender + Gender\*Dept

#### • Ordered variables

Freq ~ right+left+Diag(right:left)

Freq ~ right+left+Symm(right:left)

### • Response models

#### • Binary response

• Categorical predictors: Logit models

logit(Admit) ~ 1

logit(Admit) ~ Dept

logit(Admit) ~ Dept + Gender

• Continuous/mixed predictors:

Logistic regression models

Pr(Admit) ~ Dept + Age + GRE

#### • Polytomous response

• Ordinal: proportional odds model

Improve ~ Age + Sex + Treatment

• General: multinomial model

WomenWork ~ Kids + HusbandInc

Overview Modeling approaches

Examples

## Logistic regression models

### Response variable

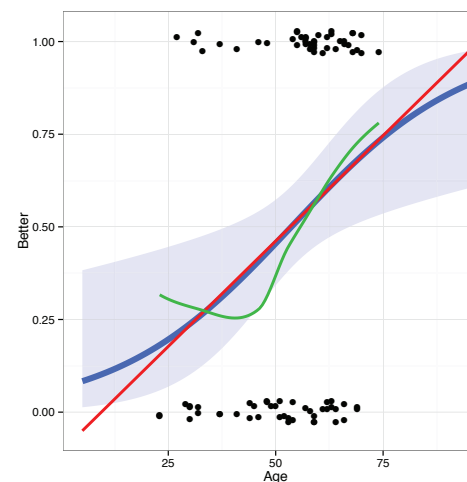
- Binary response: success/failure, vote: yes/no
- Binomial data:  $x$  successes in  $n$  trials (grouped data)
- Ordinal response: none < some < severe depression
- Polytomous response: vote Liberal, Tory, NDP, Green

### Explanatory variables

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

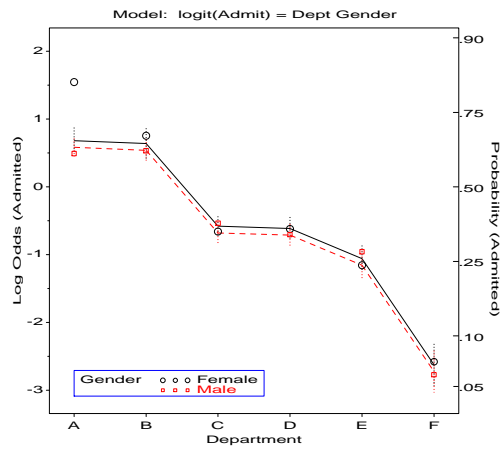
This is exactly the same as in classical ANOVA, regression models

## Arthritis treatment data



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

## Berkeley admissions data



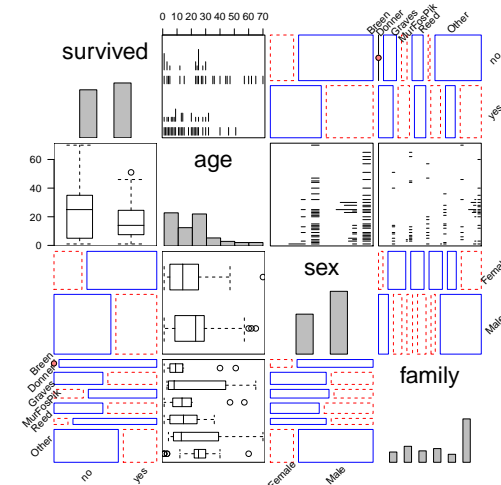
- Admit/Reject can be considered a **binomial response** for each Dept and Gender
- Logistic regression here is analogous to an ANOVA model, but for log odds(Admit)
- (With categorical predictors, these are often called **logit** models)
- Every such model has an equivalent **loglinear** model form.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

13/63

Binary response

## Survival in the Donner Party

- Binary response: **survived**
- Categorical predictors: **sex**, **family**
- Quantitative predictor: **age**
- Q: Is the effect of age linear?
- Q: Are there interactions among predictors?
- This is a **generalized pairs plot**, with different plots for each pair

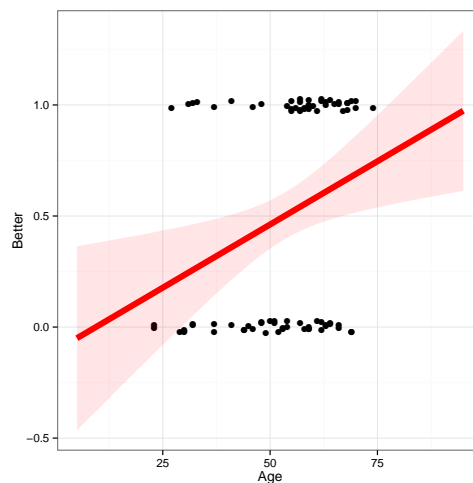


14/63

Binary response

## Binary response: What's wrong with OLS?

- For a binary response,  $Y \in (0, 1)$ , want to predict  $\pi = \Pr(Y = 1 | x)$
- 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:  $\mathcal{V}(\hat{\pi}) = \hat{\pi}(1 - \hat{\pi}) \neq \text{constant}$
  - Inferences, hypothesis tests are wrong!



15/63

## OLS vs. Logistic regression

OLS regression:

- Assume  $y|x \sim N(0, \sigma^2)$

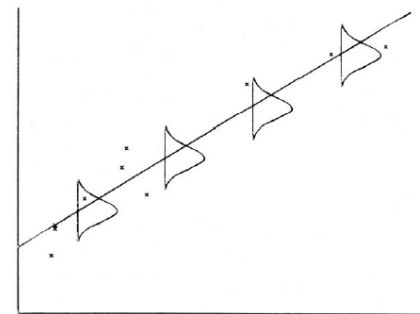


Fig. 2.1. Graphical representation of a simple linear normal regression.

Logistic regression:

- Assume  $\Pr(y=1|x) \sim \text{binomial}(p)$

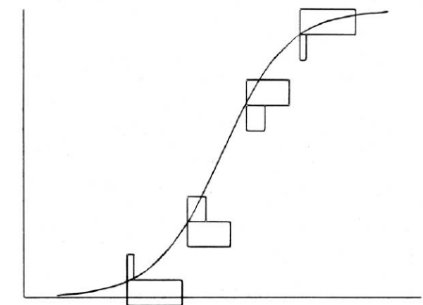
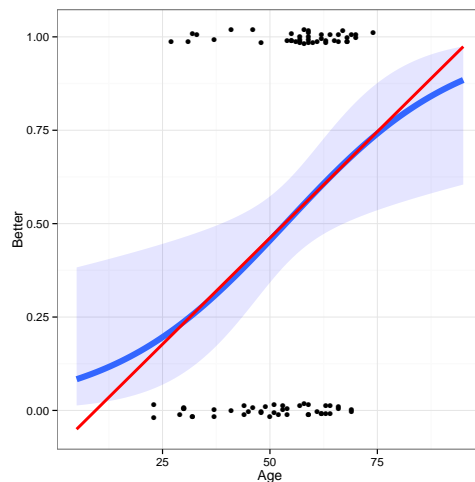


Fig. 2.2. Graphical representation of a simple linear logistic regression.

16/63

## Logistic regression

- 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.



17/63

## 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  $\pi(x)$

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

18/63

## 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

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

Thus:

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

19/63

## Logistic regression models: Multiple predictors

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

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

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

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

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

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

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

20/63

## Fitting the logistic regression model

Logistic regression models are the special case of [generalized linear models](#), fit in R using `glm(..., family=binomial)`. For this example, we define [Better](#) as any improvement at all:

```
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
```

21/63

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
```

## Interpreting coefficients

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

## (Intercept)      Age      (Intercept)      Age
##   -2.642071    0.049249      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.
- $\Pr(\text{Better})$  increases by  $\beta/4 = 0.0123$  for each year (near  $\pi = \frac{1}{2}$ )

23/63

## Multiple predictors

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

$$\text{logit}(\pi_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

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.

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

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

24/63



## Multiple predictors: Fitting

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

25/63

Hypothesis tests

## Hypothesis testing: Questions

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

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

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

27/63

## 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

26/63

Hypothesis tests

## Maximum likelihood estimation

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

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

- Under independence, joint likelihood is the product over all cases

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

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

$$\mathbf{x}^T \mathbf{y} = \mathbf{x}^T \hat{\mathbf{p}}$$

28/63

## Overall test

### ■ Likelihood ratio test ( $G^2$ )

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

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

29/63

## Wald tests and confidence intervals

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

$$z = \frac{b_k}{s(b_k)} \sim \mathcal{N}(0,1) \quad \text{or} \quad z^2 \sim \chi_1^2$$

(Wald chi-square)

- Confidence interval:

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

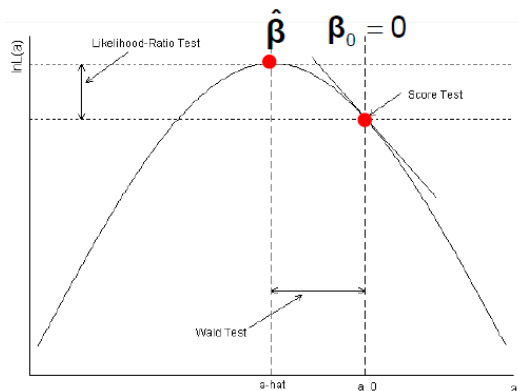
e.g.,

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

30/63

## LR, Wald and score tests

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



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

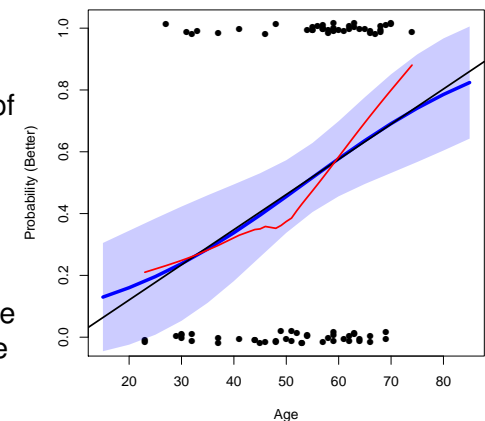
- LR test: diff in log L
- Wald test:  $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^2$
- Score test: slope at  $\boldsymbol{\beta} = 0$

31/63

## Plotting logistic regression data

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

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

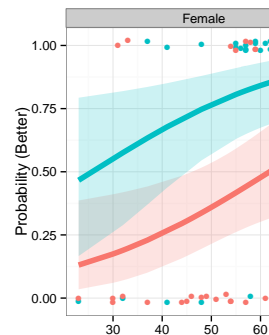
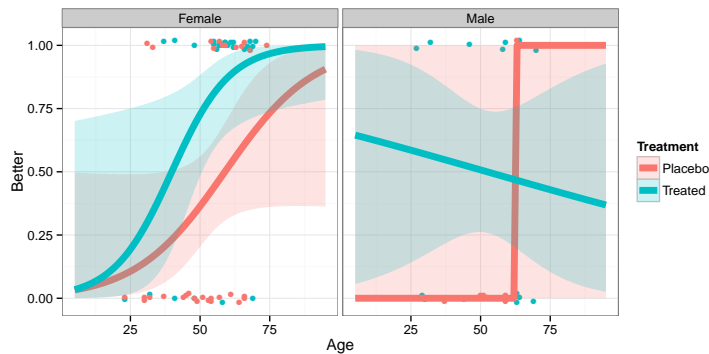


32/63



## Types of plots

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

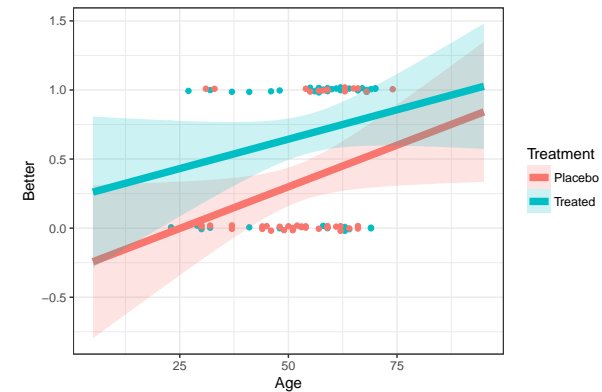


33/63

## Conditional plots with ggplot2

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

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

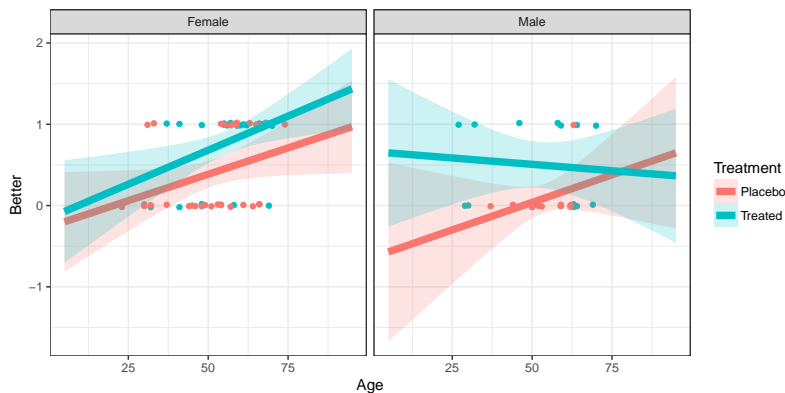


34/63

## Conditional plots with ggplot2

Conditional plot, faceted by Sex

```
gg + facet_wrap(~ Sex)
```



The data is too thin for males to estimate each regression separately

## 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	Treated	Male	27	Some	1	-1.43	0.758
##	2	Treated	Male	29	None	0	-1.33	0.728
##	3	Treated	Male	30	None	0	-1.28	0.713
##	4	Treated	Male	32	Marked	1	-1.18	0.684

35/63

36/63

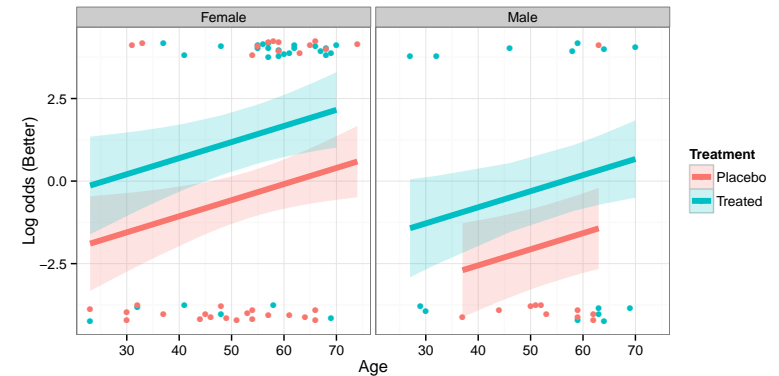
## Plotting with ggplot2 package

```
arth.fit2$obs <- c(-4, 4)[1+arth.fit2$Better]

gg2 <- ggplot(arth.fit2, aes(x=Age, y=fit, color=Treatment)) +
  geom_line(size = 2) +
  geom_ribbon(aes(ymin = fit - 1.96 * se.fit,
                ymax = fit + 1.96 * se.fit,
                fill = Treatment), alpha = 0.2,
            color = "transparent") +
  labs(x = "Age", y = "Log odds (Better)") +
  geom_point(aes(y=obs), position=position_jitter(height=0.25, width=0))
gg2 + facet_wrap(~ Sex)
```

## Full-model plots

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



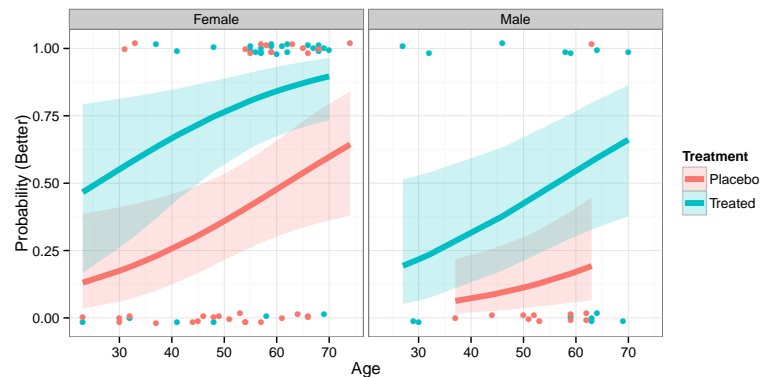
These plots show the data (jittered) as well as model uncertainty (confidence bands)

37/63

38/63

## Full-model plots

Plotting 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)

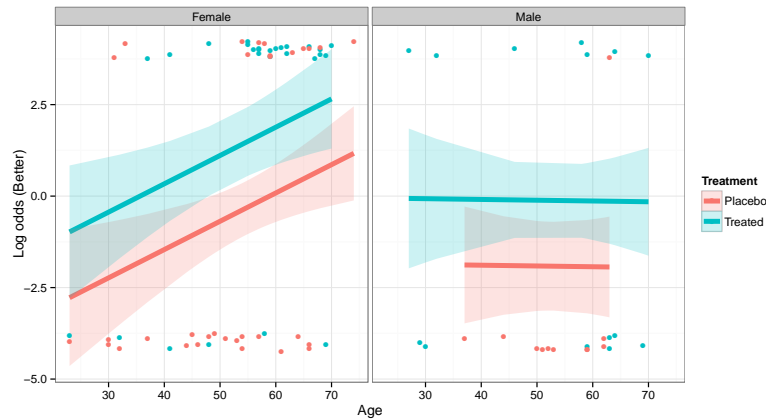
## Analysis of Deviance Table (Type II tests)
##
## Response: Better
##          LR   Chisq Df Pr(>Chisq)
## I(Age - 50)          0
## Sex              6.98  1   0.00823 **
## Treatment       11.90  1   0.00056 ***
## Sex:Age          3.42  1   0.06430 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

39/63

40/63

# Models with interactions



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

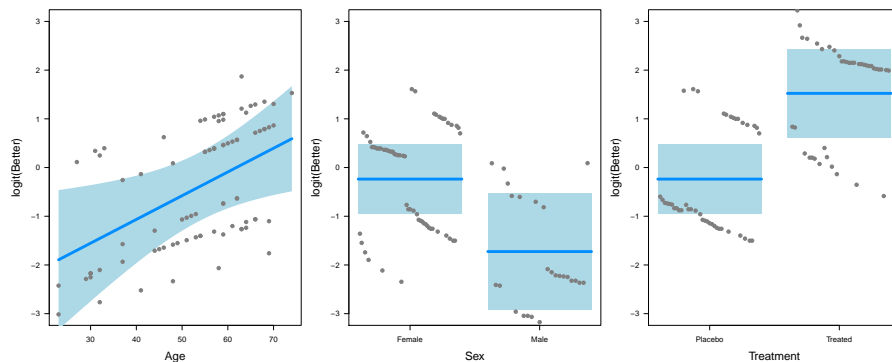
41/63

# The visreg package

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

42/63

```
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** ( $r_j$ ): the coefficient  $\hat{\beta}_j$  in the full model is the slope of the simple fit of  $r_j$  on  $x_j$ .

43/63

# Effect plots: basic ideas

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

## Data

	x1	x2	sex	x1:x2	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:  $\mathbf{X}\hat{\beta} \Rightarrow \hat{\mathbf{y}}$

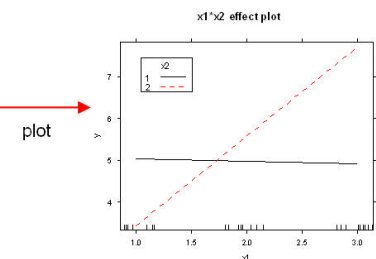
• Score data  $\mathbf{X}^* \hat{\beta} \Rightarrow \hat{\mathbf{y}}^*$

- 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



44/63

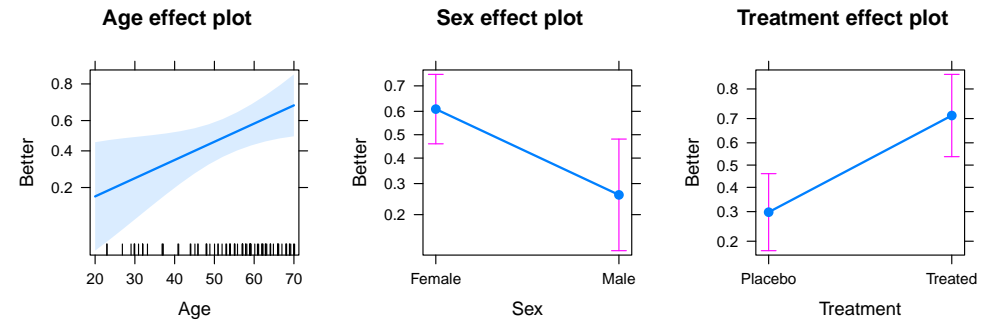
# Effect plots for generalized linear models: Details

- For simple models, full model plots show the complete relation between response and *all predictors*.
- Fox(1987)— For complex models, often wish to plot a specific main effect or interaction (including lower-order relatives)— *controlling for other effects*
  - Fit full model to data with linear predictor (e.g., logit)  $\eta = \mathbf{X}\beta$  and link function  $g(\mu) = \eta \rightarrow$  estimate  $\mathbf{b}$  of  $\beta$  and covariance matrix  $\widehat{V}(\mathbf{b})$  of  $\mathbf{b}$ .
  - Construct “score data”
    - Vary each predictor in the term over its’ range
    - Fix other predictors at “typical” values (mean, median, proportion in the data)
    - $\rightarrow$  “effect model matrix,”  $\mathbf{X}^*$
  - Use `predict()` on  $\mathbf{X}^*$ 
    - Calculate fitted effect values,  $\hat{\eta}^* = \mathbf{X}^* \mathbf{b}$ .
    - Standard errors are square roots of  $\text{diag } \mathbf{X}^* \widehat{V}(\mathbf{b}) \mathbf{X}^{*\top}$
  - Plot  $\hat{\eta}^*$ , or values transformed back to scale of response,  $g^{-1}(\hat{\eta}^*)$ .
- Note:** This provides a general means to visualize interactions in *all* linear and generalized linear models.

45/63

## Plotting main effects:

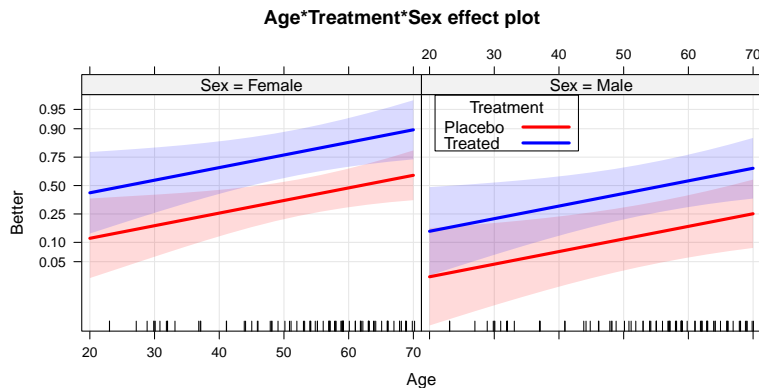
```
library(effects)
arth.eff2 <- allEffects(arth.logistic2)
plot(arth.eff2, rows=1, cols=3)
```



46/63

## Full model plots:

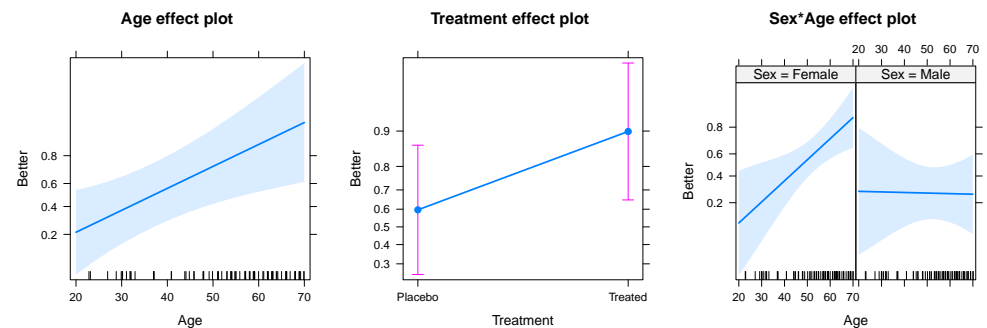
```
arth.full <- Effect(c("Age", "Treatment", "Sex"), arth.logistic2)
plot(arth.full, multiline=TRUE, ci.style="bands", colors = c("red", "blue"), lwd=3, ...)
```



47/63

## 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

48/63

# Case study: Arrests for Marijuana Possession

## Context & background

- In Dec. 2002, the *Toronto Star* examined the issue of **racial profiling**, by analyzing a data base of 600,000+ arrest records from 1996-2002.
- They focused on a subset of arrests for which police action was **discretionary**, e.g., simple possession of small quantities of marijuana, where the police could:
  - Release the arrestee with a summons— like a parking ticket
  - Bring to police station, hold for bail, etc.— harsher treatment
- **Response** variable: released – Yes, No
- Main **predictor** of interest: skin-colour of arrestee (black, white)

The *Toronto Star* meets mosaic displays...

B SECTION » TORONTO STAR » WEDNESDAY, DECEMBER 11, 2002 ★ thestar.com

## Race and Crime

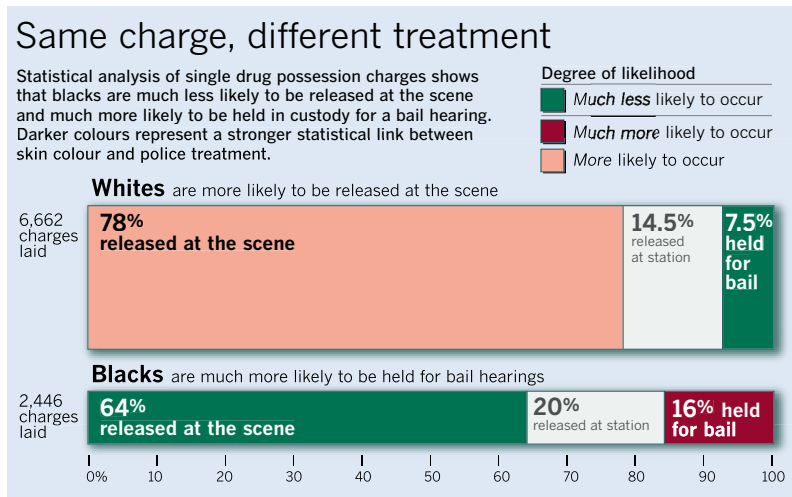


York University professor Michael Friendly's expert statistical analysis provided confirmation for the *Toronto Star*'s series on racial profiling by city police.

## Man behind the numbers

49 / 63

... Which got turned into this infographic:



... Hey, they even spelled likelihood correctly!

51 / 63

## Arrests for Marijuana Possession: Data

### Data

### Control variables:

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

```
library(effects) # for Arrests data
library(car)     # for Anova()
data(Arrests)
some(Arrests)
```

##	released	colour	year	age	sex	employed	citizen	checks
## 243	Yes	White	2000	18	Male	Yes	Yes	2
## 514	Yes	White	1999	24	Male	Yes	Yes	1
## 628	Yes	White	1997	19	Male	Yes	Yes	3
## 671	Yes	White	2000	30	Male	No	Yes	1
## 1278	Yes	Black	2002	21	Male	No	Yes	3
## 2015	Yes	Black	2000	19	Male	No	Yes	4
## 2162	Yes	Black	2000	27	Male	Yes	No	1
## 2632	No	White	1997	28	Male	No	No	5
## 3240	Yes	White	2000	28	Male	Yes	Yes	3
## 4324	Yes	Black	1999	22	Male	Yes	No	0

52 / 63

## Arrests for Marijuana Possession: Model

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

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

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

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

Analysis of Deviance Table (Type II tests)

Response: released

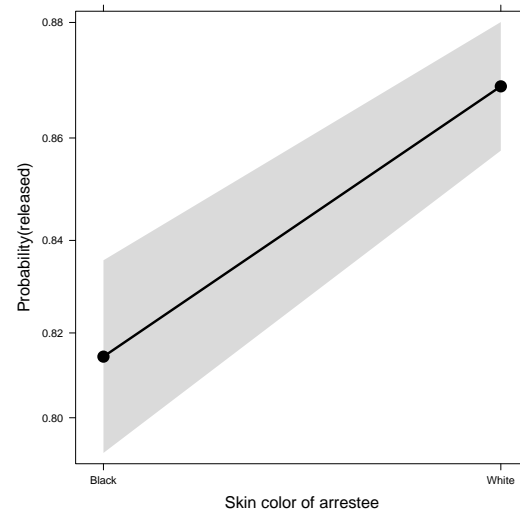
	LR	Chisq	Df	Pr(>Chisq)
employed	72.673	1	< 2.2e-16	***
citizen	25.783	1	3.820e-07	***
checks	205.211	1	< 2.2e-16	***
colour	19.572	1	9.687e-06	***
year	6.087	5	0.2978477	
age	0.459	1	0.4982736	
colour:year	21.720	5	0.0005917	***
colour:age	13.886	1	0.0001942	***

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

53/63

## Effect plots: colour

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



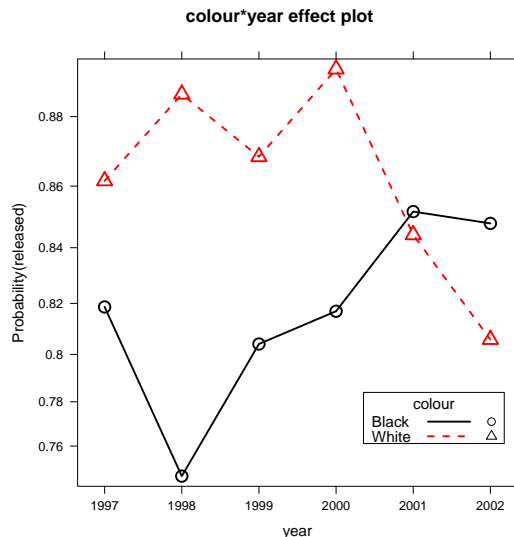
- Effect plot for colour shows average effect **controlling** (adjusting) for *all* other factors simultaneously
- (The *Star* analysis, controlled for these one at a time.)
- $\Rightarrow$  Evidence for different treatment of blacks and whites ("racial profiling")
- (Even Frances Nunziata could understand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

54/63

## Effect plots: Interactions

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

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



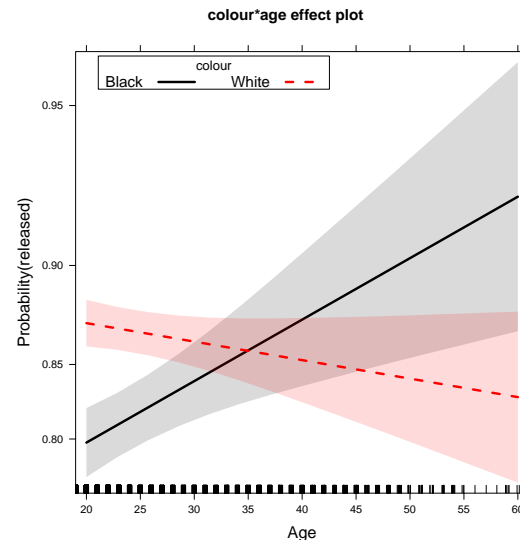
- Up to 2000, strong evidence for differential treatment of blacks and whites
- Also evidence to support Police claim of effect of training to reduce racial effects in treatment

55/63

## Effect plots: Interactions

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

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



A more surprising finding:

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

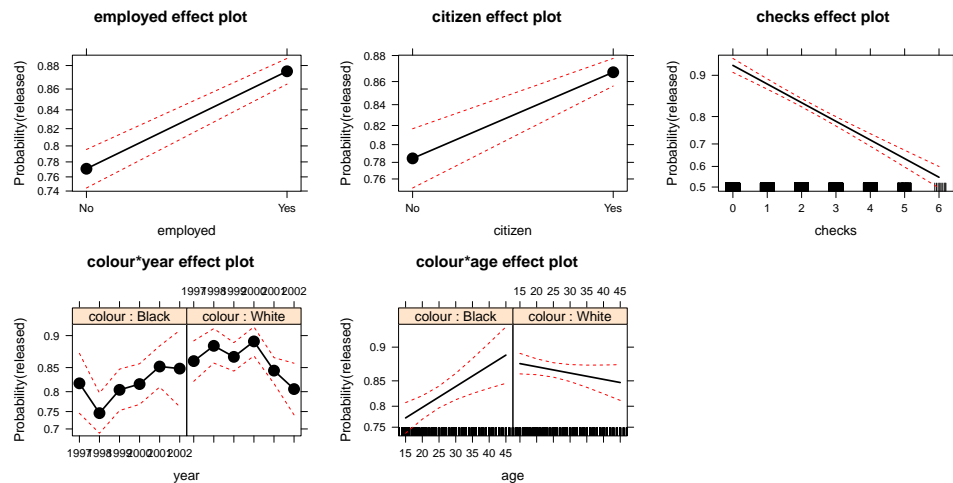
56/63



## Effect plots: allEffects

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

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



57/63

## 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

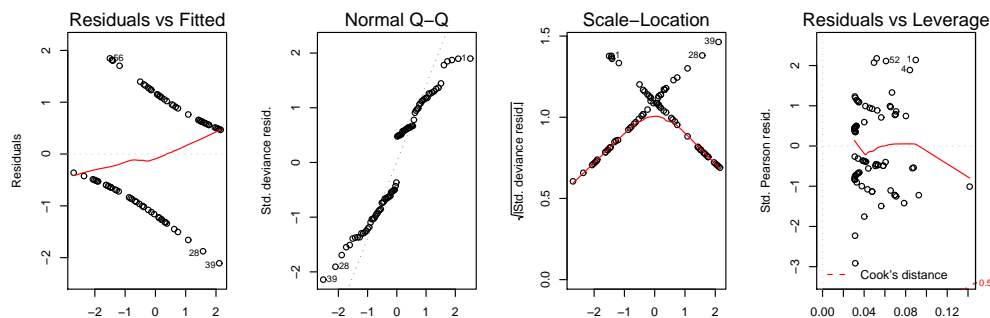
- Use non-parametric smoothed curves to detect non-linearity
- Consider using polynomial terms ( $X^2, X^3, \dots$ ) or **regression splines** (e.g., `ns(X, 3)`)
- Use `update(model, ...)` to test for interactions—formula: `. ~ .^2`

58/63

## 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

59/63

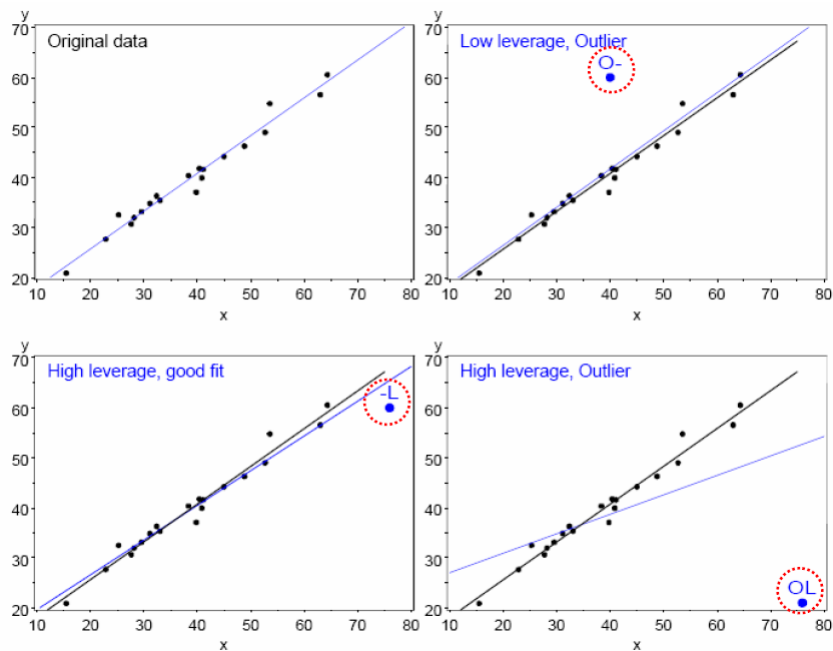
## Unusual data: Leverage and Influence

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

$$\text{Influence} = \text{Leverage}_X \times \text{Residual}_Y$$

60/63

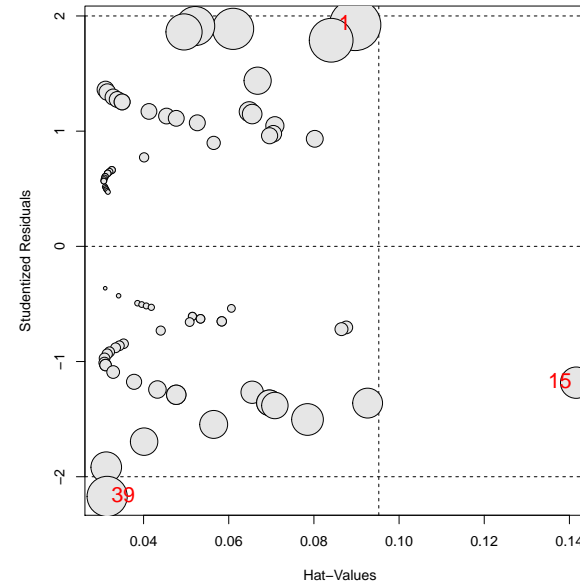
## Effect of adding one more point in simple linear regression (new point in blue)



61/63

## Influence plots in R

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

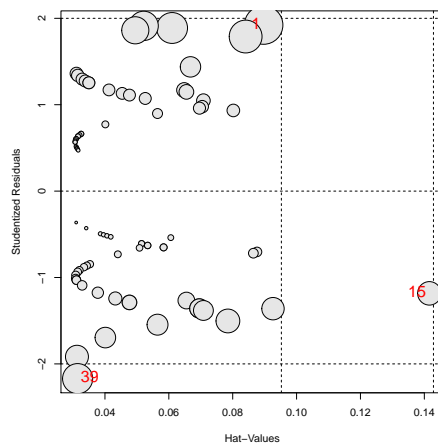


- X axis: Leverage ("hat values")
- Y axis: Studentized residuals
- Bubble size  $\sim$  Cook D (influence on coefficients)

62/63

## 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



63/63