

Logistic regression



Michael Friendly

Psych 6136

<http://friendly.github.io/psy6136>



2

Model-based methods: Overview

Structure

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

Advantages

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

3

Today's topics

- Model-based methods: Overview
- Logistic regression: one predictor, multiple predictors, fitting
- Visualizing logistic regression
- Effect plots
- Case study: Racial profiling
- Model diagnostics

loglm() vs. glm()

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

```
> 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 20.20  5 0.001144
Pearson         18.82  5 0.00207
```

What we can say:

Even the model with all pairwise associations fits poorly 😞

4

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     21.74   6
Model 2     20.20   5     1.531       1     0.21593
Saturated    0.00   0     20.204      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
```

Q: What can we say from this?

5

`loglm()` vs. `glm()`

With `glm()` you can test individual terms using `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
```

Q: Can someone help interpret the term for Gender:Admit ?

6

Dropping & adding terms

A useful strategy for model-building is to start with some model, and consider

- The effect of dropping high-order terms, one at a time
- The effect of adding terms w/in the scope of a larger model, one at a time
- MASS::dropterm() and MASS::addterm() do this for both `glm()` and `loglm()` models

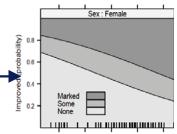
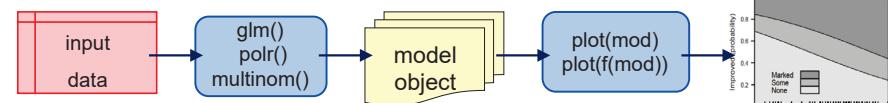
```
> MASS::dropterm(berk.glm2, test="Chisq")
Single term deletions

Model:
Freq ~ (Dept + Gender + Admit)^2
      Df Deviance     AIC      LRT Pr(Chi)
<none>    20.20  217.26
Dept:Gender  5  1148.90 1335.96 1128.70 <2e-16 ***
Dept:Admit   5   783.61  970.67  763.40 <2e-16 ***
Gender:Admit 1   21.74  216.80    1.53  0.2159
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7

Fitting & graphing models: 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

8

Modeling approaches: Overview

Association models

- Loglinear models
(contingency table form)
[Admit][Gender Dept]
[Admit Dept][Gender Dept]
[AdmitDept][AdmitGender][GenderDept]
- Poisson GLMs
(Frequency data frame)
 $\text{Freq} \sim \text{Admit} + \text{Gender} * \text{Dept}$
 $\text{Freq} \sim \text{Admit} * \text{Dept} + \text{Gender} * \text{Dept}$
 $\text{Freq} \sim \text{Admit} * (\text{Dept} + \text{Gender}) + \text{Gender} * \text{Dept}$
- Ordinal variables
 $\text{Freq} \sim \text{right} + \text{left} + \text{Diag}(\text{right}, \text{left})$
 $\text{Freq} \sim \text{right} + \text{left} + \text{Symm}(\text{right}, \text{left})$

Response models

- Binary response
- Categorical predictors: logit models
 $\text{logit}(\text{Admit}) \sim 1$
 $\text{logit}(\text{Admit}) \sim \text{Dept}$
 $\text{logit}(\text{Admit}) \sim \text{Dept} + \text{Gender}$
- Continuous/mixed predictors
- Logistic regression models
 $\text{Pr}(\text{Admit}) \sim \text{Dept} + \text{Gender} + \text{Age} + \text{GRE}$
- Polytomous response
- Ordinal: proportional odds model
 $\text{Improve} \sim \text{Age} + \text{Sex} + \text{Treatment}$
- General multinomial model
 $\text{WomenWork} \sim \text{Kids} + \text{HusbandIncome}$

12

Logistic regression

Response variable

- Binary response: success/failure, vote: yes/no
- Binomial data: x successes in n trials (grouped data)

```
glm(success ~ ..., family=binomial)
glm(cbind(Nsuccess, Nfail) ~ ..., family=binomial)
```

Explanatory variables

- Quantitative regressors: age, dose
- Transformed regressors: $\sqrt{\text{age}}$, $\log(\text{dose})$
- Polynomial regressors: age^2 , 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

13

Logistic regression: Extensions

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

Extensions of the framework for logistic regression allow us to handle more than two discrete outcomes. Explanatory variable remain the same

Explanatory variables

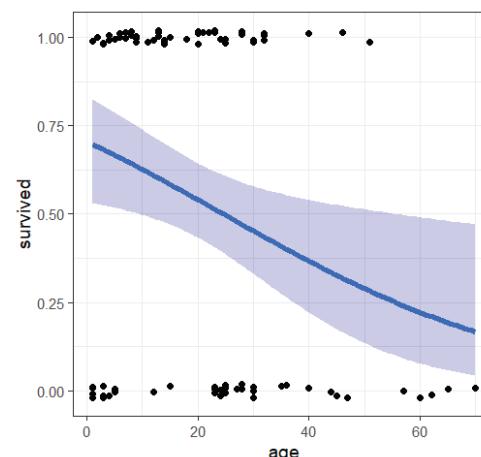
- Quantitative regressors: age, dose
- Transformed regressors: $\sqrt{\text{age}}$, $\log(\text{dose})$
- Polynomial regressors: age^2 , age^3 , ... (or better: splines)
- Categorical predictors: treatment, sex (dummy variables, contrasts)
- Interaction regressors: treatment \times age, sex \times age

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

14

Survival in the Donner Party

Data on the Donner Party records the fate of 90 people who set out to CA in 1846. They were trapped in an early winter storm near Reno, NV. Only 48 survived.



Who survived? Why?

Logistic regression can model the probability of the **binary (0/1)** outcome of survival

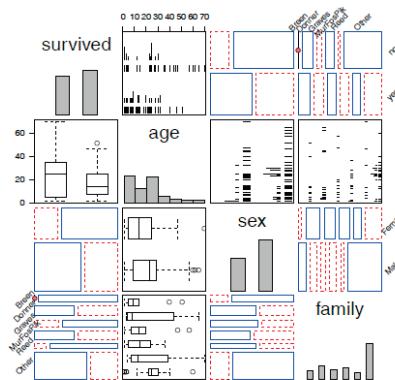
A quantitative predictor like age gives predicted probabilities (& CI)

Other predictors – sex, family, ... can give a more detailed understanding

15

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



Some possible models:

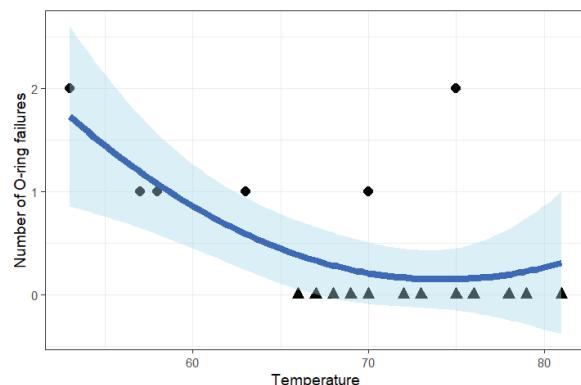
```
glm(survived ~ age, data=Donner, family=binomial)
glm(survived ~ age + sex + family, data=Donner, family=binomial)
glm(survived ~ age * sex, data=Donner, family=binomial)
```

16

Challenger: A better graph

This graph plots the number of failures out of 6 O-rings in all previous flights, including those with 0 failures

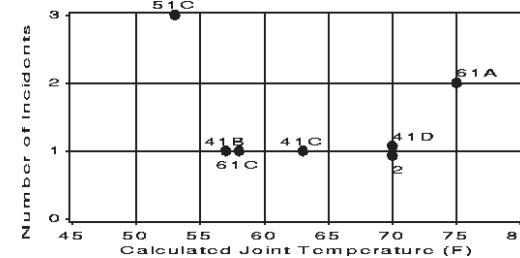
- It fits a simple quadratic regression, `nFailures ~ poly(Temperature, 2)`
- It should have been a warning that failures increase as temperature gets lower
- But it doesn't take into account that `nFailures ~ Bin(p, n=6)`



18

Challenger: A dataviz disaster

- The space shuttle *Challenger* exploded 73 sec. after takeoff on January 28, 1986, killing all 7 crew
 - Subsequent investigation revealed the [proximal cause](#): Low temperature → failures of the rubber O-rings joining rocket stages
 - The [anterior cause](#) was a failure of data analysis & visualization
- Data: 24 previous flights: temperature, # of “incidents”



Morton-Thiokol engineers prepared this bad graph

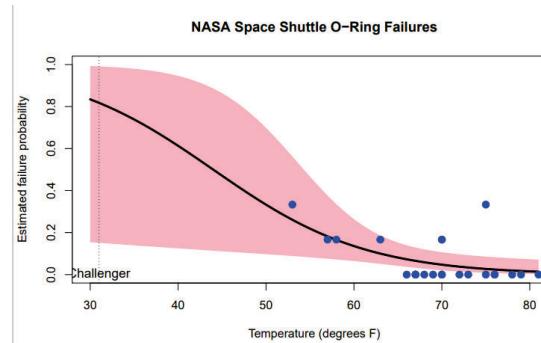
But, they also excluded all flights where there was no damage

17

Challenger: A better analysis

Logistic regression treats the # failures as a [binomial outcome with n = 6 trials](#)
The model provides

- Predicted probabilities outside the range of the data
- Confidence intervals, to judge model uncertainty



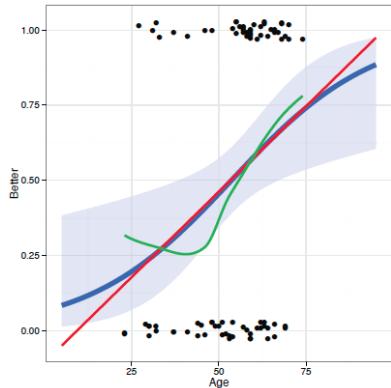
When the challenger was launched, the temp was 31° F

The CI band is very wide, but the predicted value is uncomfortably high

This analysis & graph might have saved lives!

19

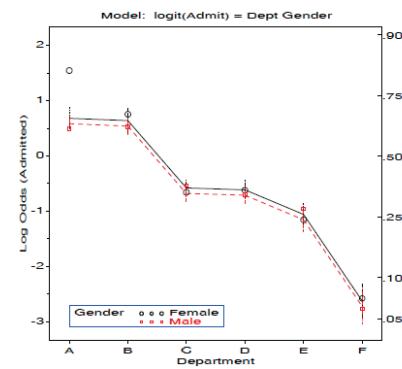
Example: Arthritis treatment



- 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

20

Example: Berkeley admissions



- 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

21

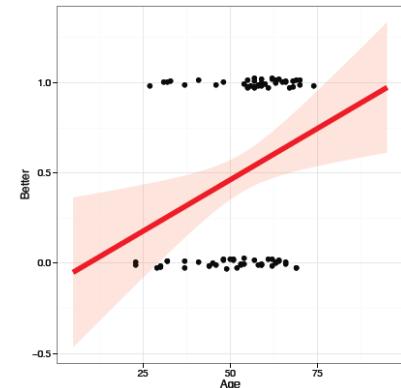
The Logistic Regression Model



22

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



23

Linear regression vs Logistic regression

OLS regression:

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

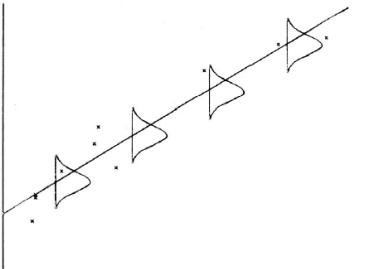


Fig. 2.1. Graphical representation of a simple linear normal regression.

y linear with x
constant residual variance

Logistic regression:

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

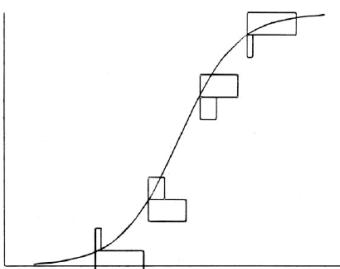


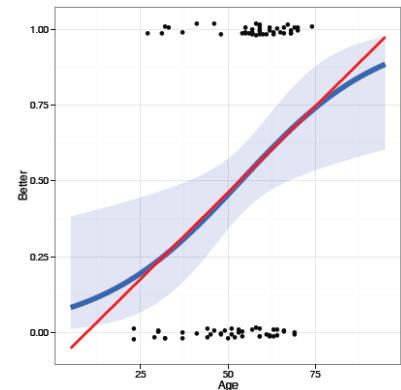
Fig. 2.2. Graphical representation of a simple linear logistic regression.

$y \sim \text{logit}(x)$
non-constant residual variance $\sim p(1-p)$

24

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.



25

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 $\pi(x)$

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

Thinking logically:

- Model is for the **log odds** of the marked response, $Y = 1$
- Can always back transform with logit^{-1} to get **probability** of $Y = 1$

26

Logistic regression: One predictor

The coefficients, α, β of this model have simple interpretations in terms of odds & log odds

$$\text{logit}[\pi(x)] \equiv \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x \quad \text{odds}(Y = 1) \equiv \frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) = e^\alpha (e^\beta)^x$$

β is the change in log odds for a **unit increase** in x
→ The odds of $Y=1$ are multiplied by e^β for each unit increase in x

α is the log odds **when $x=0$**

→ The odds of $Y=1$ when $x=0$ is e^α

In R, use **exp(coef(model))** to get these values

Another interpretation: In terms of probability, the **slope** of the logistic regression curve is $\beta/(1-\pi)$

This has the **maximum** value $\beta/4$ when $\pi = \gamma_2$

27

Logistic regression: Multiple predictors

- For a binary response, $Y \in \{0, 1\}$, let \mathbf{x} be a vector of p regressors, and π_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} ,

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

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

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

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

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

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

28

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))
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
```

29

The `summary()` method gives details and tests of coefficients

```
> 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  G^2 for H_0: beta_Age = 0
Residual deviance: 109.16 on 82 degrees of freedom  G^2 for H_1: beta_Age != 0
```

How much better is this than the null model? $\Delta G^2_{(1)} = 116.45 - 109.16 = 7.29$

30

Interpreting coefficients

```
> coef(arth.logistic)
(Intercept)          Age
-2.64207  0.04925
```

```
> exp(coef(arth.logistic))
(Intercept)          Age
 0.07121  1.05048
> exp(10*coef(arth.logistic)[2])
Age
1.636
```

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

31

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.

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

- 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

32

Multiple predictors: Fitting

Fit the main effects model. Use $I(\text{Age} - 50)$ to center Age, making β_1 interpretable

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

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

```
> 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 **

broom::glance() gives model fit statistics

```
> broom::glance(arth.logistic2)
# A tibble: 1 x 8
  null.deviance df.null logLik  AIC   BIC deviance df.residual nobs
            <dbl>    <int>  <dbl> <dbl> <dbl>    <dbl>     <int> <int>
1          116.      83 -46.0  100.  110.    92.1       80     84
```

33

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

34

Hypothesis testing: Questions

- **Overall test:** How does my model, $\text{logit}(\pi) = \alpha + \mathbf{x}^T \boldsymbol{\beta}$ 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 χ^2 tests

35

Maximum likelihood estimation

In classical linear models using `lm()`, we fit using ordinary least squares.
All `glm()` models use maximum likelihood estimation— better properties

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

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

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

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

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

$$\mathbf{X}^\top \mathbf{y} = \mathbf{X}^\top \hat{\boldsymbol{\beta}}$$

36

37

Overall model tests

Likelihood ratio test (G^2)

- Compare **nested** models, similar to F tests in OLS
- Let L_1 = maximized value for **our model**
 $\text{logit}(\pi_i) = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}$ w/ k predictors
- Let L_0 = maximized likelihood for the **null model**
 $\text{logit}(\pi_i) = \beta_0$ under $H_0: \beta_1 = \beta_2 = \dots = \beta_k$
- Likelihood ratio test statistic:

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

Wald tests & confidence intervals

- Analogous to t -tests in OLS
- Test $H_0: \beta_i = 0$ $z = \frac{b_i}{s(b_i)} \sim \mathcal{N}(0,1)$ or $z^2 \sim \chi_1^2$
- Confidence interval $b_i \pm z_{1-\alpha/2} s(b_i)$

```
> r1 <- lmtest::coeftest(arth.logistic2)
> r2 <- confint(arth.logistic2)
Waiting for profiling to be done...
> cbind(r1, r2)
            Estimate Std. Error z value Pr(>|z|) 2.5 % 97.5 %
(Intercept) -0.578     0.367   -1.6    0.116 -1.33  0.124
I(Age - 50)   0.049     0.021    2.4    0.018   0.01  0.092
SexMale      -1.488     0.595   -2.5    0.012  -2.73 -0.372
TreatmentTreated  1.760     0.536    3.3    0.001   0.75  2.875
```

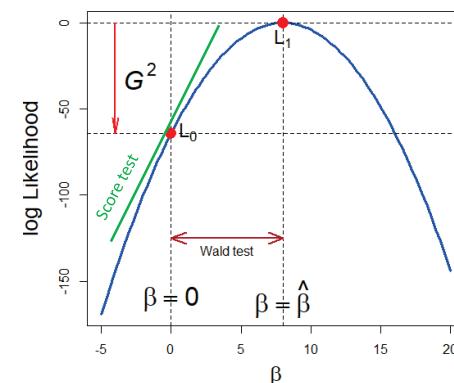
38

39

LR, Wald & Score tests

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.3859	3	<.0001
Score	22.0051	3	<.0001
Wald	17.5147	3	0.0006

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$



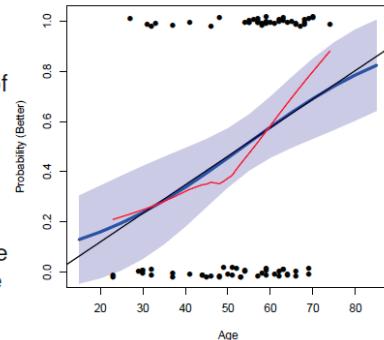
Different ways to measure departure from $H_0: \boldsymbol{\beta} = 0$

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

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.

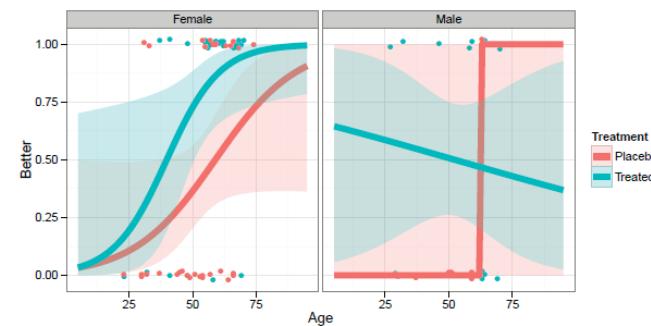
- Need to jitter the points
- Useful to show the fitted logistic curve
- Confidence band gives a sense of uncertainty
- Adding a non-parametric (loess) smooth shows possible nonlinearity
- NB: Can plot either on the response scale (probability) or the link scale (logit) where effects are linear



40

Types of plots

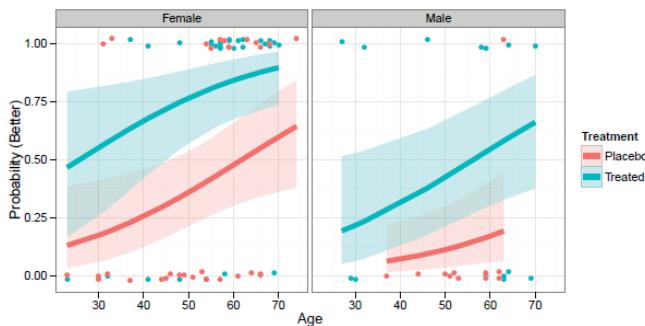
- **Conditional plots:** Stratified plot of Y or logit(Y) vs. one X, conditioned by other predictors--- only that `subset` is plotted for each panel



41

Types of plots

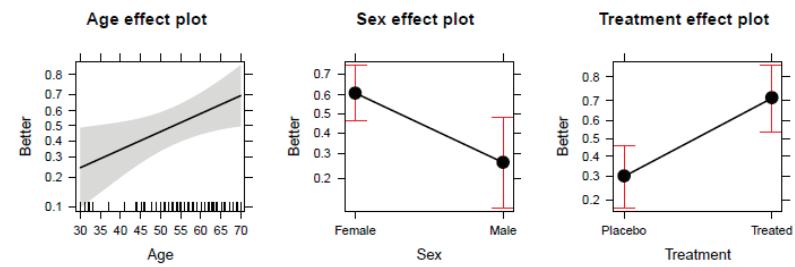
- **Full-model plots:** Plot of **fitted response surface**, showing all effects; usually shown in several panels



42

Types of plots

- **Effect plots:** plots of predicted effects for terms in the model, averaged over predictors not shown in a given plot

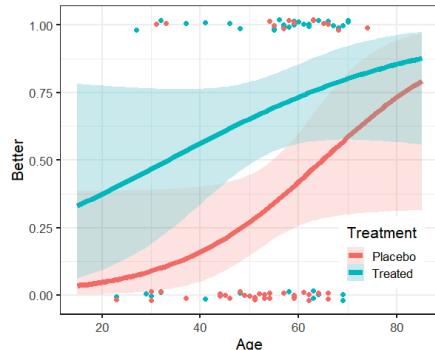


43

Conditional plots with ggplot2

Plot Arthritis data by Treatment, ignoring Sex; overlay fitted logistic reg. lines

```
gg <- ggplot(Arthritis, aes(Age, Better, color=Treatment)) +  
  xlim(15, 85) +  
  geom_jitter(height = 0.02, width = 0, size=2) +  
  stat_smooth(method = "glm", family = binomial, alpha = 0.2,  
    aes(fill=Treatment), size=2.5, fullrange=TRUE) +  
  theme_bw(base_size = 16) + theme(legend.position = c(.85, .2))  
gg # show the plot
```



geom_jitter() shows the observations more distinctly

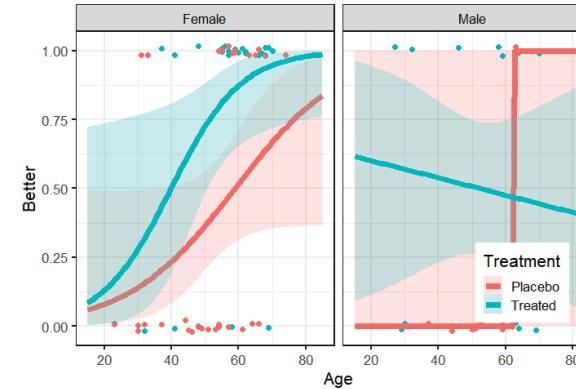
Fitted lines use method="glm", family=binomial

44

Conditional plots with ggplot2

Can show the conditional plots for M & F, simply by faceting by Sex

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



Only the data for each Sex is used in each plot

Plotting the data points shows that the data for males is too thin to give good estimates of separate regression

45

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
```

46

Plotting with ggplot2

Plot the fitted log odds, confidence band and observations

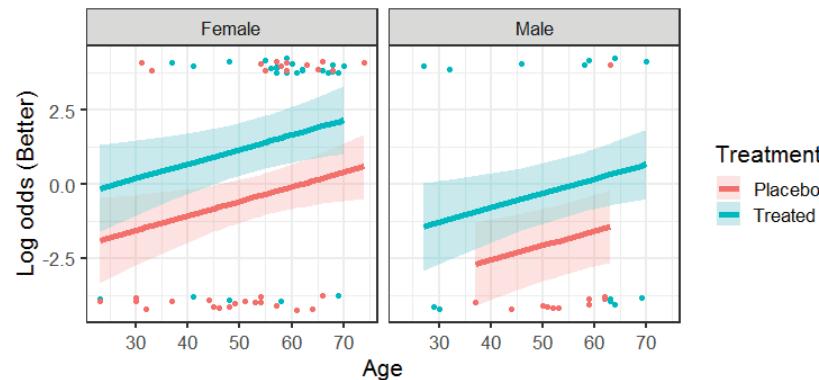
```
arth.fit2 <- arth.fit2 |>  
  mutate(obs = ifelse(Better==0, -4, 4)) # show obs at -4, 4  
  
ggplot( arth.fit2, aes(x=Age, y=fit, color=Treatment)) +  
  geom_line(size = 2) +  
  geom_ribbon(aes(ymin = fit - 1.96 * se.fit,  
                  ymax = fit + 1.96 * se.fit,  
                  fill = Treatment), alpha = 0.2,  
  color = "transparent") +  
  labs(x = "Age", y = "Log odds (Better)") +  
  geom_jitter(aes(y=obs), height=0.25, width=0) +  
  facet_wrap(~ Sex) +  
  theme_bw(base_size = 16)
```

Using `color=Treatment` gives separate points and lines for the two groups

47

Full-model plot

Plotting on the logit scale shows the **additive** effects of age, treatment and sex
NB: easier to compare the treatment groups within the **same** panel



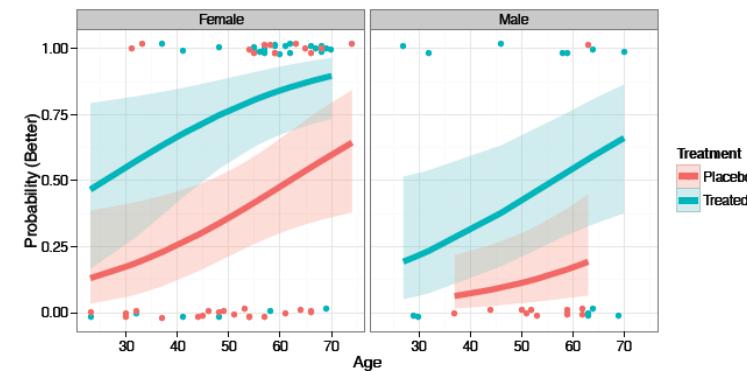
These plots show model uncertainty (confidence bands)
Jittered points show the data

48

Full-model plot

Plotting on the probability scale may be simpler to interpret
Use `predict(... type = "response")` to get fitted probabilities

```
arth.fit2r <- cbind(Arthritis,
                      predict(arth.logistic2, se.fit = TRUE, type="response"))
```



49

Models with interactions

Is the linear effect of age the same for females, males?

- We can test this by adding an **interaction** of Sex \times Age
- **`update()`** makes it easy to add/subtract terms from a model
- **`car:::Anova()`** gives partial tests of each term after all others

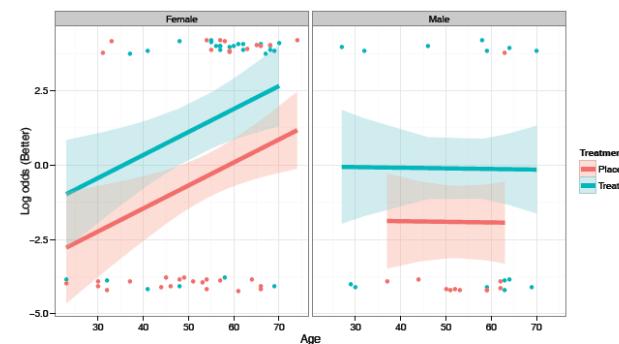
```
> arth.logistic4 <- update(arth.logistic2, . ~ . + I(Age-50):Sex)
> car:::Anova(arth.logistic4)
Analysis of Deviance Table (Type II tests)

Response: Better
          LR Chisq Df Pr(>Chisq)
I(Age - 50)      6.16  1   0.01308 *
Sex             6.98  1   0.00823 **
Treatment       11.90  1   0.00056 ***
I(Age - 50):Sex  3.42  1   0.06430 .
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1
```

The interaction term Age:Sex is not quite significant, but plot the fitted model anyway

50

Models with interactions



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

51

Effect plots: Basic ideas

Show a given **marginal effect**, **controlling** / adjusting for other model effects

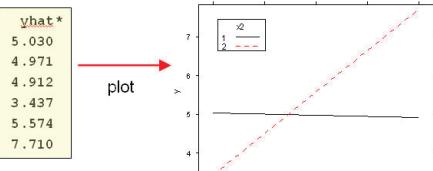
Data

	x1	x2	sex	x1*x2	y
1	1	1	F	1	4.73
2	2	1	M	0	6.10
3	3	1	F	-1	4.32
4	1	2	F	1	4.84
5	2	1	F	0	4.73
...
29	2	2	M	0	6.10
30	3	2	F	1	6.71

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

Score data

	x1	x2	sex	x1*x2	y	vthat*
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



52

Effect plots: Details

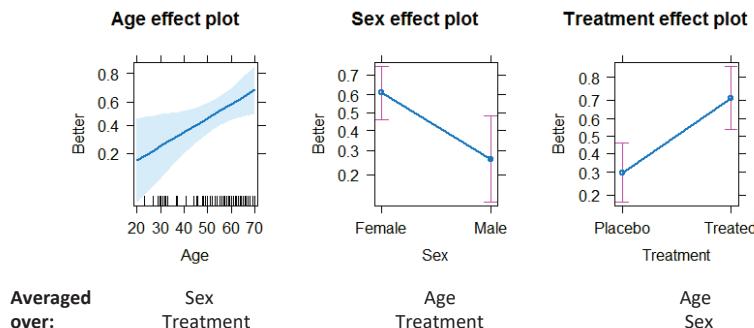
- For simple models, full model plots show the complete relation between response and *all predictors*.
- Fox(1987)— For complex models, often wish to plot a specific main effect or interaction (including lower-order relatives)— *controlling for other effects*
 - Fit full model to data with linear predictor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

53

Plotting main effects

allEffects() calculates effects for all high-order terms in the model
The response is plotted on the logit scale, but labeled with probabilities

```
library(effects)
arth.eff2 <- allEffects(arth.logistic2)
plot(arth.eff2, rows=1, cols=3, lwd=2)
```

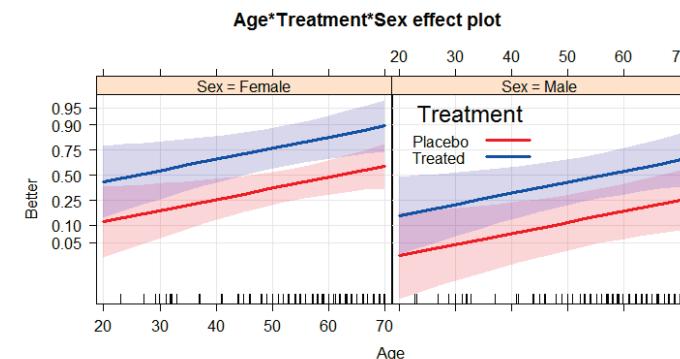


54

Full-model plot

The full-model plot is simply the **Effect()** of the highest-order interaction of factors

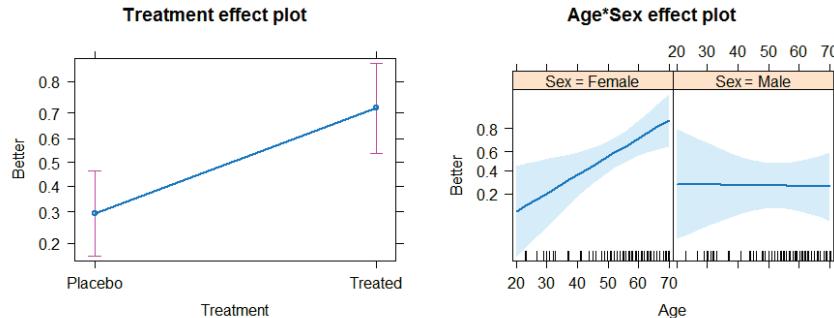
```
arth.full <- Effect(c("Age", "Treatment", "Sex"), arth.logistic2)
plot(arth.full, multiline=TRUE, ci.style="bands",
     colors = c("red", "blue"), lwd=3, . . .)
```



55

Model with interaction of Age × Sex

```
arth.eff4 <- allEffects(arth.logistic4)
plot(arth.eff4, lwd=2)
```



Only the high-order terms: Treatment & Age * Sex are shown & need to be interpreted
Q: How would you describe this?

56

Case study: Arrests for marijuana

- In Dec. 2002, the *Toronto Star* examined the issue of **racial profiling**, by analyzing a data base of 600,000+ arrest records from 1997-2002.
- 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, ... -- harsher treatment
- Response variable:** released: "Yes", "No"
 - Main predictor of interest: skin-colour of arrestee (black, white)
 - Other predictors: year, age, sex, ...

58

Race & Crime

Toronto Star investigation of racial disparities in treatment by Toronto Police Services

FOI request → > ½ M arrests, 1997–2002

Evidence for **racial profiling**?

Only look at discretionary charges:

Simple marijuana possession
Non-moving auto infractions

THE SATURDAY STAR

The photo that never was

GARTH WOOLLEY, CS
• Waterfront: Dress rehearsal of what could be, \$1, B45
• Underwater: Divers in a hole, E1
• Wheely: The Bug goes mobile, G1
• Paul Martin: The man who would be king, H1
• Carol Shields: Her last now? Unseen, J1

Also inside...

October 19, 2002

thestar.com ONTARIO EDITION

AN INVESTIGATION INTO RACE AND CRIME

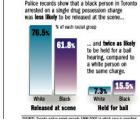


SUNG POLICE: Jason Burke, falsely accused of dealing drugs during Carnivale two years ago, says he was a victim of racial profiling.

Singled out

Star analysis of police crime data shows justice is different for blacks and whites

Telling numbers



Blacks arrested by Toronto police are treated more harshly than whites, according to an analysis of police data. This is despite the fact that blacks are less likely to be charged with simple drug possession, we take note. Despite this, they are twice as likely to be held for bail, compared to a white person arrested for the same charge.

At the station, arrested black people are more likely to be held over night than whites, even though they have been charged with the same offence.

Once at the station, arrested black people are more likely to be held over night than whites, even though they have been charged with the same offence.

The Toronto crime data also shows that black people are more likely to be charged with more serious offences than white people, such as assault or breaking and entering.

"It's clear that there is a racial disparity in the way police treat black people," says Michael Friendy, York University professor.

"We don't treat people differently because of their race," he says.

Chrétien expected to keep cabinet minister

Ethics report has 'wiggle room' to save MacAulay

BY TIM ROBERTS
AND LES WESTFORD
OTTAWA — Jean Chrétien may be able to keep his cabinet minister today if an ethics report reveals that he exceeded his authority when he accepted a gift.

Ethics committee Howard MacAulay, 68, was charged with accepting a \$1,000 gift from a police officer and delivered it to the Prime Minister's office in December 1999, where it was received by Chrétien's chief of staff, Tony Acciari.

It was due to be referred to the Ethics Committee of the House of Commons, but the committee said last night that it would not do so. Instead, Wilson's report, the Prince Edward Islander said, will be referred to the Standing Senate Committee on Ethics.

Chretien's office said MacAulay has been cleared of any wrongdoing by mid-November.

Community members said last night that they were disappointed by the report.

Wilson's report, the Prince Edward Islander said, will be referred to the Standing Senate Committee on Ethics.

Chretien's office said MacAulay has been cleared of any wrongdoing by mid-November.

MacAulay's spokesman, Alvin Morris, said last night that the committee had recommended the awarding of a \$100,000 fine to MacAulay.

MacAulay's office said last night that the committee had recommended the awarding of a \$100,000 fine to MacAulay.

Toronto Star meets mosaic displays

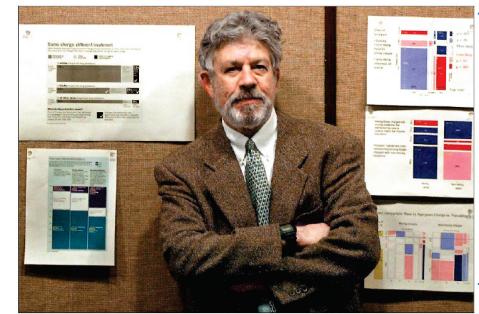
How to communicate these results most effectively?

- What is the message? What features are directly comprehensible to the audience?

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

Race and Crime

Graphic designer's early attempts



My early attempts

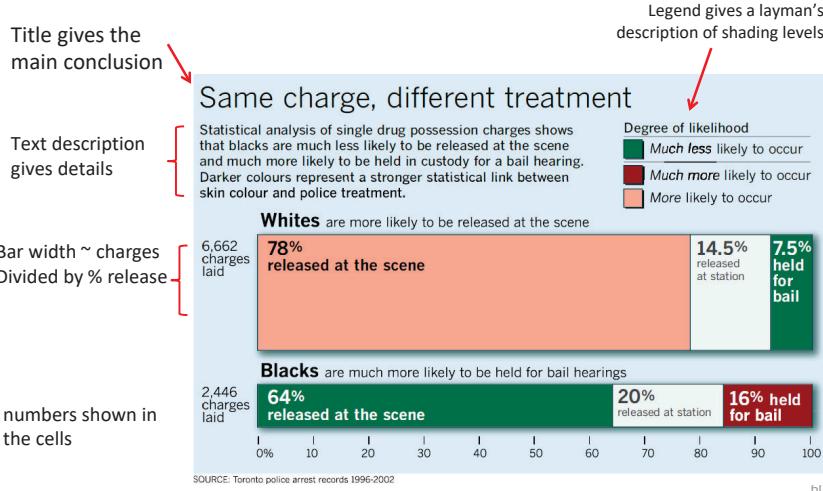
York University professor Michael Friendly's expert statistical analysis provided confirmation for the *Toronto Star's* series on racial profiling by a star writer.

Man behind the numbers

59

Racial profiling: Presentation graphic

Together, we created this (nearly) **self-explaining** infographic



Arrests for marijuana: Data

Response variable: released

Control variables:

- year, age, sex
- employed, citizen: Yes, No
- checks: # of police databases (previous arrests, convictions, parole status) where the arrestee's name was found

```
> library(car)      # for Anova()
> data(Arrests, package = "carData")
> some(Arrests)
```

	released	colour	year	age	sex	employed	citizen	checks
218	Yes	White	2000	24	Male	Yes	Yes	0
1301	No	Black	1999	17	Male	Yes	No	1
1495	Yes	White	1998	23	Male	Yes	Yes	0
1732	Yes	Black	2000	18	Male	Yes	Yes	2
1838	Yes	Black	1997	27	Male	No	Yes	5
2257	No	White	2001	19	Male	No	Yes	2
3100	No	Black	2000	19	Male	No	Yes	4
3843	Yes	White	1999	20	Male	Yes	Yes	0
4580	Yes	Black	1999	26	Male	Yes	Yes	1
4833	Yes	Black	1998	38	Male	Yes	Yes	0

61

Arrests for marijuana: Model

year is numerical. But may be non-linear. Convert to a **factor**

Fit model with all main effects, but allow **interactions** of colour:year and colour:age

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

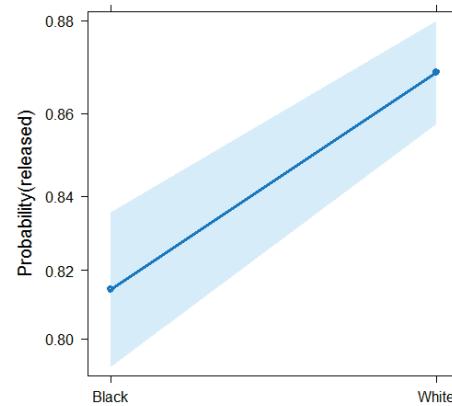
Analysis of Deviance Table (Type II tests)

Response: released
          LR Chisq Df Pr(>Chisq)
employed    72.7   1    < 2e-16 ***
citizen     25.8   1    3.8e-07 ***
checks      205.2  1    < 2e-16 ***
colour      19.6   1    9.7e-06 ***
year        6.1    5    0.29785
age         0.5    1    0.49827
colour:year 21.7   5    0.00059 ***
colour:age   13.9   1    0.00019 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

62

Effect plot: Skin colour

```
plot(Effect("colour", arrests.mod), lwd=3, ci.style="bands", ...)
```



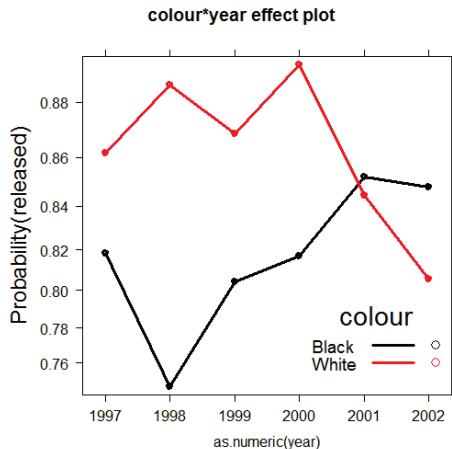
- Effect plot for colour shows average effect **controlling** (adjusting) for **all** other factors simultaneously
- (The Star analysis controlled for these one at a time.)
→ Evidence for different treatment of blacks & whites
- Even Francis Nunziata could understand this.
- However, effect smaller than reported by the Star

63

Effect plots: Interactions

The story turned out to be more nuanced than reported by the *Toronto Star*

```
plot(Effect(c("colour", "year"), arrests.mod), multiline=TRUE, ...)
```



Up to 2000, strong evidence for differential treatment of blacks & whites

Also: evidence to support Police claim of effect of training to reduce racial effects in treatment

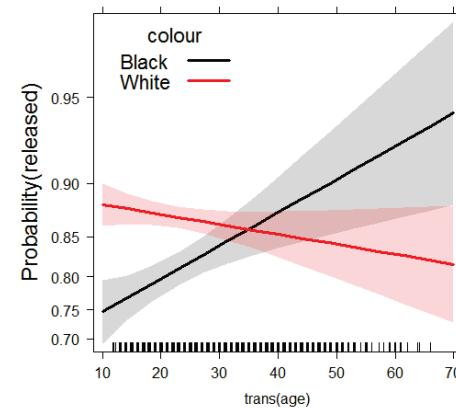
64

Effect plots: Interactions

A more surprising finding ...

```
plot(Effect(c("colour", "year"), arrests.mod), multiline=TRUE, ...)
```

Effects of skin colour and age on release



Opposite age effects for blacks & whites:

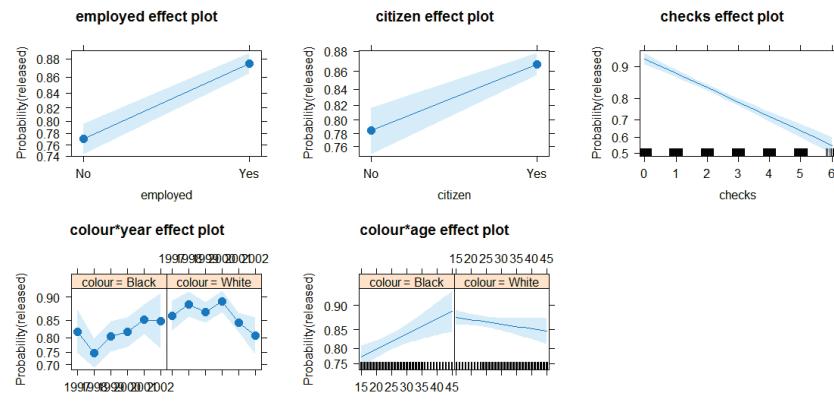
- Young blacks treated **more harshly** than young whites
- Older blacks treated **less harshly** than older whites

65

Effect plots: allEffects

All high-order terms 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)", ...)
```



66

Model diagnostics

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

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

Model specification: Tools and techniques

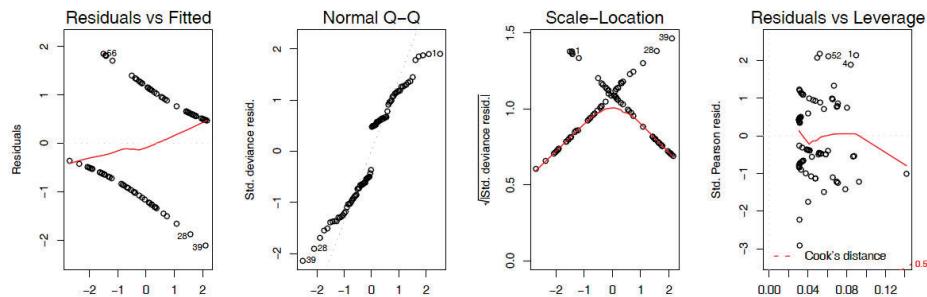
- Use non-parametric smoothed curves to detect non-linearity
- Consider using polynomial terms (X^2, X^3, \dots) or **regression splines** (e.g., `ns(X, 3)`)
- Use `update(model, ...)` to test for interactions—formula: $. \sim ?$

67

Diagnostic plots in R

In R, plotting a `glm` object gives the “regression quartet” – 4 basic diagnostic plots

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



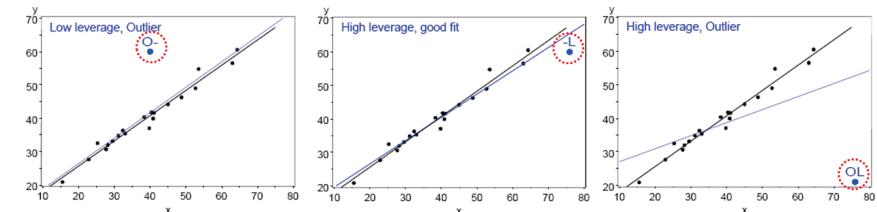
These plots often look peculiar for logistic regression models
Better versions are available in the car package

68

Unusual data: Leverage & Influence

- “Unusual” observations can have dramatic effects on least-squares estimates in linear models
- 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 & Y
- Heuristic formula:

$$\text{Influence} = \text{X leverage} \times \text{Y residual}$$



69

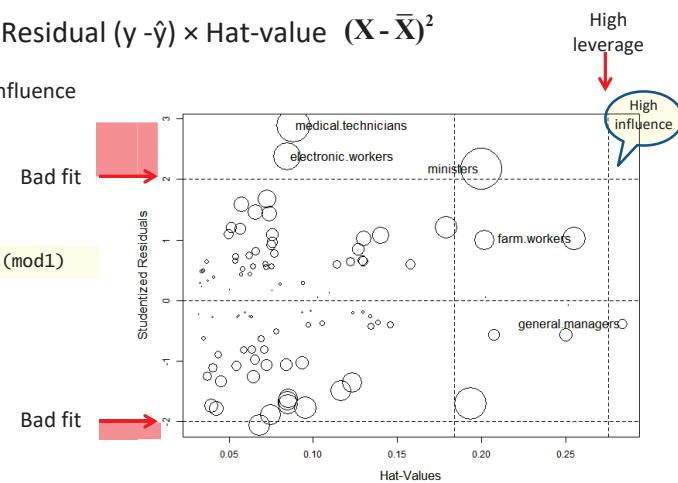
Influence plots

Influence (Cook's D) measures impact of individual obs. on coefficients, fitted values

Influence \sim Residual $(y - \hat{y}) \times \text{Hat-value } (X - \bar{X})^2$

Bubble size \sim influence

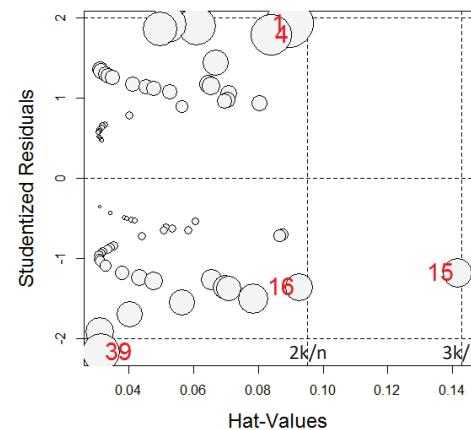
```
influencePlot(mod1)
```



70

Influence plots in R

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



X axis: Leverage (“hat values”)
notable values: $> 2k/n, 3k/n$

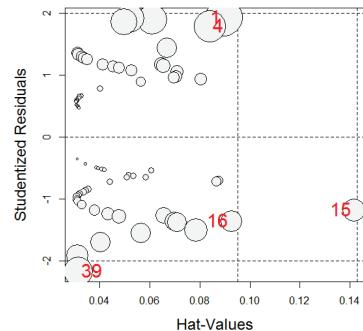
Y axis: Studentized residuals

Bubble size \sim Cook's D
(influence on coefficients)

71

Which cases are influential?

	Treatment	Sex	Age	Better	StudRes	Hat	CookD
1	Treated	Male	27	1	1.92	0.0897	0.1128
4	Treated	Male	32	1	1.79	0.0840	0.0818
15	Treated	Female	23	0	-1.18	0.1416	0.0420
16	Treated	Female	32	0	-1.36	0.0926	0.0381
39	Treated	Female	69	0	-2.17	0.0314	0.0690



case 1: younger male: moderate Hat,
better than predicted → large Cook D

case 39: older female: small Hat, but did
not improve with treatment

72

Looking ahead

- Logistic regression models need not always have linear effects—models **nonlinear** in Xs sometimes useful
- **Polytomous** outcomes can be handled as well
 - e.g., Improved = {"None", "Some", "Marked"}
- If ordinal,
 - the **proportional odds** model is a simple extension
 - **nested dichotomies** provides an alternative approach
- Otherwise, **multinomial logistic regression** is the way

73

Summary

- `loglm()` provides only overall tests of model fit
- Model-based methods, `glm()`, provide hypothesis tests, CIs & tests for individual terms
- Logistic regression: A `glm()` for a binary response
 - linear model for the log odds $\text{Pr}(Y=1)$
 - All similar to classical ANOVA, regression models
- Plotting
 - Conditional, full-model plots show data and fits
 - Effect plots show predicted effects averaged over others
- Model diagnostics
 - Influence plots are often informative

74