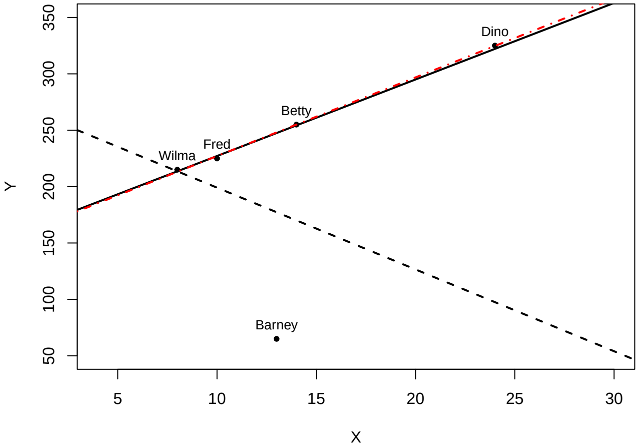
GSERM 2024
Regression for Publishing

June 19, 2024

Discrepancy, Leverage, and Influence



Note: Solid line is the regression fit for Wilma, Fred, and Betty only.
Long-dashed line is the regression for Wilma, Fred, Betty, and Barney.
Short-dashed (red) line is the regression for Wilma, Fred, Betty and Dino.

Discrepancy, Leverage, and Influence

$$\text{Influence} = \text{Leverage} \times \text{Discrepancy}$$

Leverage

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= \mathbf{H}\mathbf{Y} \end{aligned}$$

where

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

$$h_i = \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'$$

Variation:

$$\widehat{\text{Var}(\hat{u}_i)} = \hat{\sigma}^2[1 - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'] \tag{1}$$

$$\begin{aligned}\widehat{\text{s.e.}(\hat{u}_i)} &= \hat{\sigma}\sqrt{[1 - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i']} \\ &= \hat{\sigma}\sqrt{1 - h_i}\end{aligned} \tag{2}$$

"Standardized":

$$\tilde{u}_i = \frac{\hat{u}_i}{\hat{\sigma}\sqrt{1 - h_i}} \tag{3}$$

"Studentized": define

$$\begin{aligned}\hat{\sigma}^2_{-i} &= \text{Variance for the } N-1 \text{ observations} \neq i \\ &= \frac{\hat{\sigma}^2(N-K)}{N-K-1} - \frac{\hat{u}_i^2}{(N-K-1)(1-h_i)}.\end{aligned} \tag{4}$$

Then:

$$\hat{u}_i' = \frac{\hat{u}_i}{\hat{\sigma}_{-i}\sqrt{1-h_i}} \tag{5}$$

"DFBETA":

$$D_{ki} = \hat{\beta}_k - \hat{\beta}_{k(-i)} \tag{6}$$

"DFBETAS" (the "S" is for "standardized):

$$D^*_{ki} = \frac{D_{ki}}{\widehat{\text{s.e.}(\hat{\beta}_{k(-i)})}} \tag{7}$$

Cook's D:

$$\begin{aligned} D_i &= \frac{\tilde{u}_i^2}{K} \times \frac{h_i}{1-h_i} \\ &= \frac{h_i \hat{u}_i^2}{K \hat{\sigma}^2 (1-h_i)^2} \end{aligned} \tag{8}$$

```
> # No Barney OR Dino...
> summary(lm(Y~X,data=subset(flintstones,name!="Dino" & name!="Barney")))

Call:
lm(formula = Y ~ X, data = subset(flintstones, name != "Dino" &
    name != "Barney"))

Residuals:
      2       4       5
 0.7143 -2.1429  1.4286

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 159.2857     6.7763   23.51   0.0271 *
X             6.7857     0.6186   10.97   0.0579 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.673 on 1 degrees of freedom
Multiple R-squared:  0.9918, Adjusted R-squared:  0.9835
F-statistic: 120.3 on 1 and 1 DF,  p-value: 0.05787
```

Variance

```
> # No Barney (Dino included...)
> summary(lm(Y~X,data=subset(flintstones,name!="Barney")))

Call:
lm(formula = Y ~ X, data = subset(flintstones, name != "Barney"))

Residuals:
         2          3          4          5 
-8.771e-15  2.632e-01 -2.105e+00  1.842e+00 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 157.3684     2.4651   63.84 0.000245 ***
X             6.9737     0.1612   43.27 0.000534 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.987 on 2 degrees of freedom
Multiple R-squared:  0.9989, Adjusted R-squared:  0.9984 
F-statistic:  1873 on 1 and 2 DF,  p-value: 0.0005336
```

"COVRATIO":

$$\text{COVRATIO}_i = \left[(1 - h_i)\left(\frac{N - K - 1 + \hat{u}_i'^2}{N - K}\right)^K\right]^{-1} \tag{9}$$

Example: Federal Judicial Review, 1789-2021

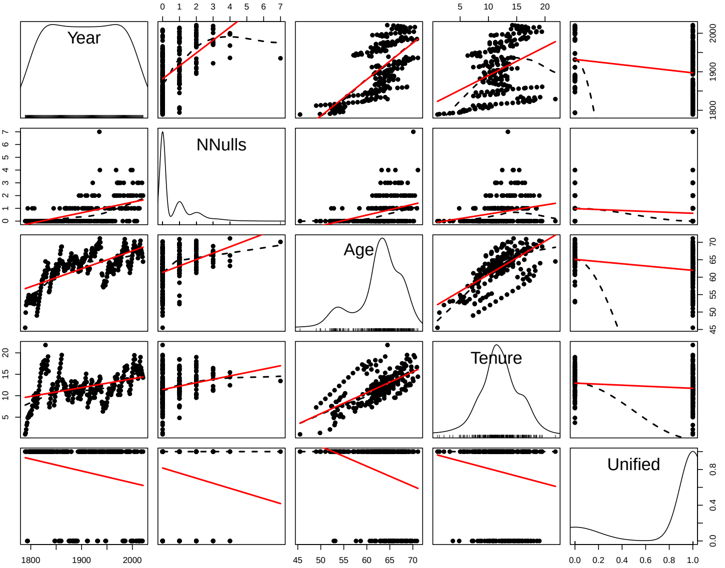
Dahl (1957):

- SCOTUS gets "out of step" with the other branches → judicial review
- Older / longer-serving justices will more likely to invalidate legislation

Data:

```
> psych::describe(NewDahl,fast=TRUE,skew=TRUE)
                  n    mean    sd  median    min     max  range  skew kurtosis   se
Year            233 1905.00 67.41 1905.00 1789.0 2021.00 232.00  0.00    -1.22 4.42
NConstDecisions 233   17.96 19.11   12.00    0.0   85.00  85.00  1.38     1.48 1.25
NNulls          233    0.70  1.06    0.00    0.0    7.00   7.00  1.96     5.39 0.07
Age             233   62.65  4.96   63.56   45.5   71.11  25.61 -0.84     0.29 0.32
Tenure          233   12.00  3.54   11.90    1.0   21.83  20.83 -0.19     0.25 0.23
Unified         233    0.78  0.42    1.00    0.0    1.00   1.00 -1.32    -0.26 0.03
```

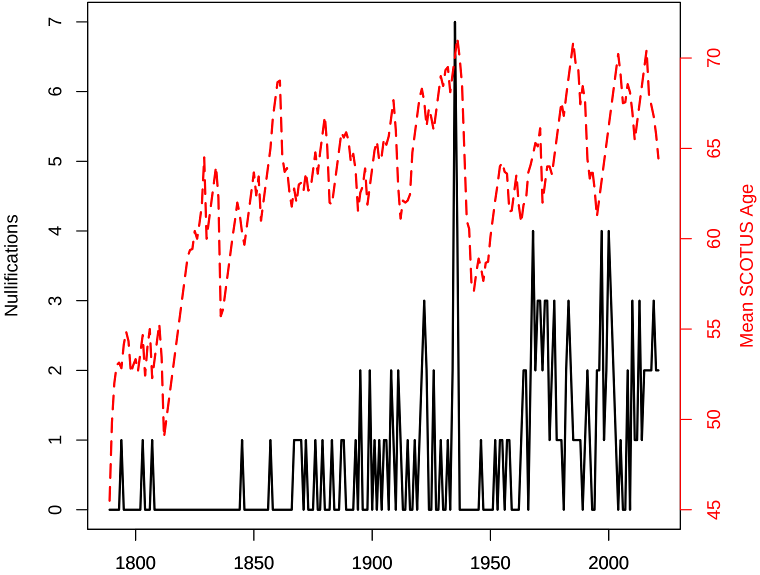
Example: Federal Judicial Review, 1789-2021



A Regression...

```
> Fit<-lm(NNulls~Age+Tenure+Unified,data=NewDahl)
> summary(Fit)

Call:
lm(formula = NNulls ~ Age + Tenure + Unified, data = NewDahl)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3632 -0.7014 -0.1433  0.3279  5.6837

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.63930    1.00986  -4.594 7.18e-06 ***
Age          0.08958    0.01829   4.899 1.82e-06 ***
Tenure      -0.01631    0.02494  -0.654    0.514
Unified     -0.10574    0.16025  -0.660    0.510
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9808 on 229 degrees of freedom
Multiple R-squared:  0.1572, Adjusted R-squared:  0.1461
F-statistic: 14.23 on 3 and 229 DF,  p-value: 1.545e-08
```

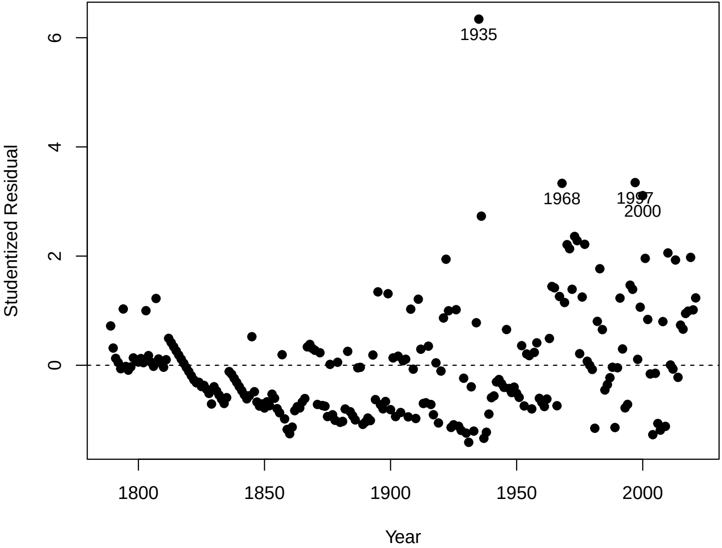
Federal Judicial Review and Mean SCOTUS Age



Residuals, etc.

```
> FitResid<-with(NewDahl,(Fit$model$NNulls-predict(Fit))) # residuals
> FitStandard<-rstandard(Fit) # standardized residuals
> FitStudent<-rstudent(Fit) # studentized residuals
> FitCooksD<-cooks.distance(Fit) # Cook's D
> FitDFBeta<-dfbeta(Fit) # DFBeta
> FitDFBetaS<-dfbetas(Fit) # DFBetaS
> FitCOVRATIO<-covratio(Fit) # COVRATIOs
```

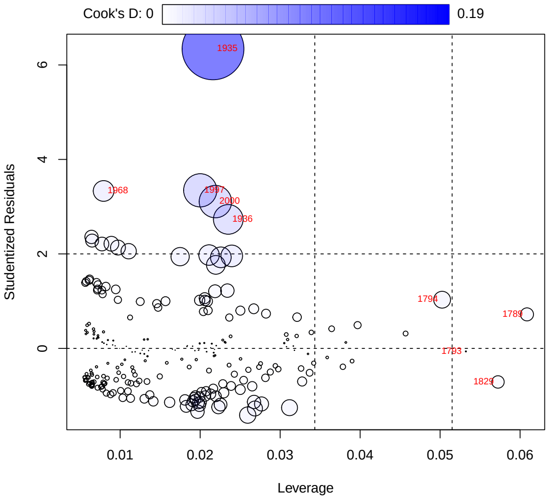
Studentized Residuals



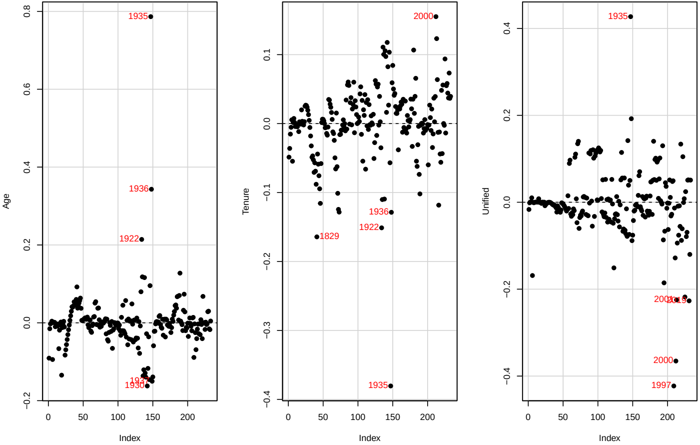
```
> max(FitStudent)
[1] 6.340266

> NewDahl$Year1935<-ifelse(NewDahl$Year==1935,1,0)
> summary(with(NewDahl, lm(NNulls~Age+Tenure+Unified+Year1935)))

Residuals:
    Min      1Q  Median      3Q     Max
-1.3054 -0.6574 -0.1317  0.3242  3.2313

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.887748   0.940647  -4.133 5.03e-05 ***
Age          0.076292   0.017026   4.481 1.18e-05 ***
Tenure      -0.007543   0.023086  -0.327    0.744
Unified     -0.168979   0.148415  -1.139    0.256
Year1935     5.809221   0.916242   6.340 1.22e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9063 on 228 degrees of freedom
Multiple R-squared:  0.2835, Adjusted R-squared:  0.2709
F-statistic: 22.55 on 4 and 228 DF,  p-value: 1.042e-15
```

“Bubble Plot”

```
> influencePlot(Fit,id=list(method="noteworthy",n=4,cex=0.7,
                  labels=NewDahl$Year,col="red"),xlab="Leverage")
```

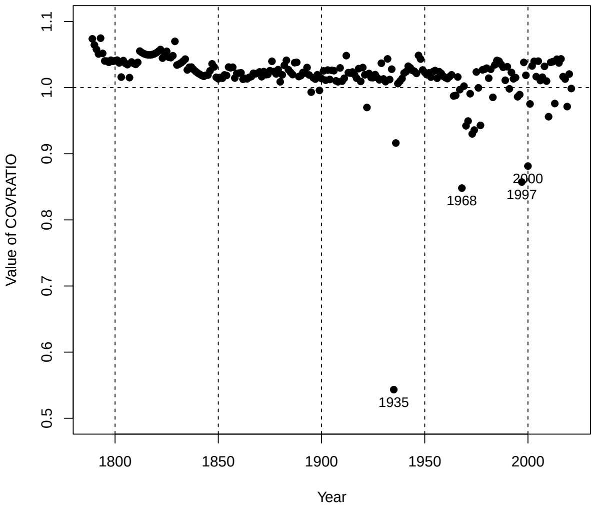


```
> dfbetasPlots(Fit,id.n=5,id.col="red",main="",pch=19,
               layout=c(1,3),labels=NewDahl$Year)
```



COVRATIO Plot

```
> plot(FitCOVRATIO~NewDahl$Year,pch=19,ylim=c(0.5,1.1),
    xlab="Year",ylab="Value of COVRATIO")
```



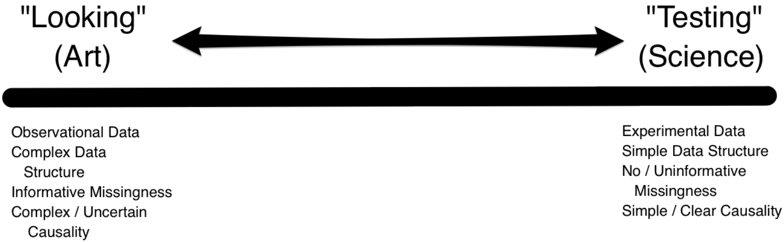
Sensitivity Analyses: Omitting Outliers

```
> out1<-c(1935) # one outlier
> LD2<-NewDahl[!(NewDahl$Year %in% out1),]
> out2<-c(1935,1968,1997,2000) # four outliers
> LD3<-NewDahl[!(NewDahl$Year %in% out2),]
> Fit2<-lm(NNulls~Age+Tenure+Unified,data=LD2)
> Fit3<-lm(NNulls~Age+Tenure+Unified,data=LD3)

> summary(Fit2)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.887748   0.940647  -4.133 5.03e-05 ***
Age          0.076292   0.017026   4.481 1.18e-05 ***
Tenure      -0.007543   0.023086  -0.327    0.744
Unified     -0.168979   0.148415  -1.139    0.256
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9063 on 228 degrees of freedom
Multiple R-squared:  0.1543, Adjusted R-squared:  0.1432
F-statistic: 13.87 on 3 and 228 DF,  p-value: 2.443e-08
```

Compare Models

	Dependent variable:		
	(1)	(2)	(3)
Age	0.090***	0.076***	0.079***
	(0.018)	(0.017)	(0.016)
Tenure	−0.016	−0.008	−0.015
	(0.025)	(0.023)	(0.021)
Unified	−0.106	−0.169	−0.066
	(0.160)	(0.148)	(0.139)
Constant	−4.639***	−3.888***	−4.077***
	(1.010)	(0.941)	(0.869)
Observations	233	232	229
R^2	0.157	0.154	0.162
Adjusted R^2	0.146	0.143	0.151
Residual Std. Error	0.981 (df = 229)	0.906 (df = 228)	0.836 (df = 225)
F Statistic	14.234*** (df = 3; 229)	13.869*** (df = 3; 228)	14.529*** (df = 3; 225)
Note:			*p<0.1; **p<0.05; ***p<0.01

Thinking About Diagnostics



One Approach

Pena, E.A. and E.H. Slate. 2006. "Global Validation of Linear Model Assumptions." *J. American Statistical Association* 101(473):341-354.

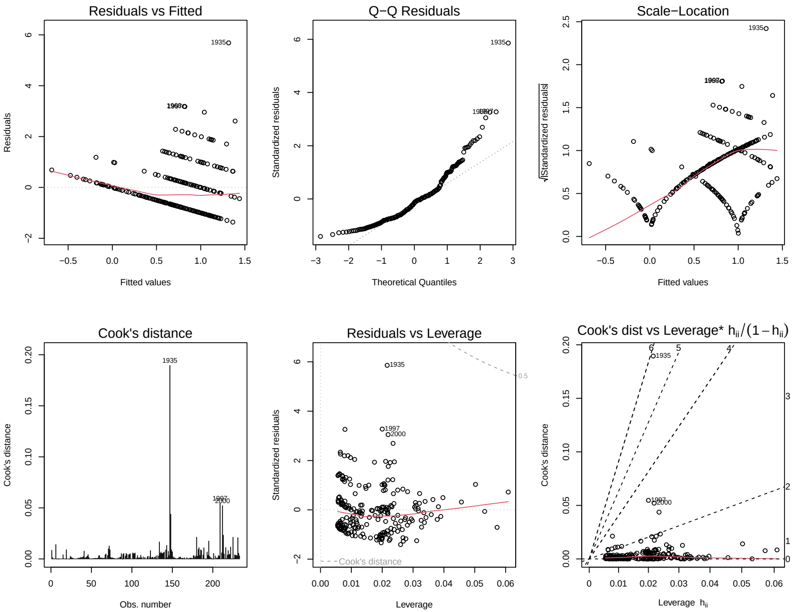
Tests for:

- Normality in $\hat{u}$s (via skewness & kurtosis tests)
- "Link function" (linearity / additivity)
- Constant variance and uncorrelatedness in $\hat{u}$s ("heteroskedasticity" test)

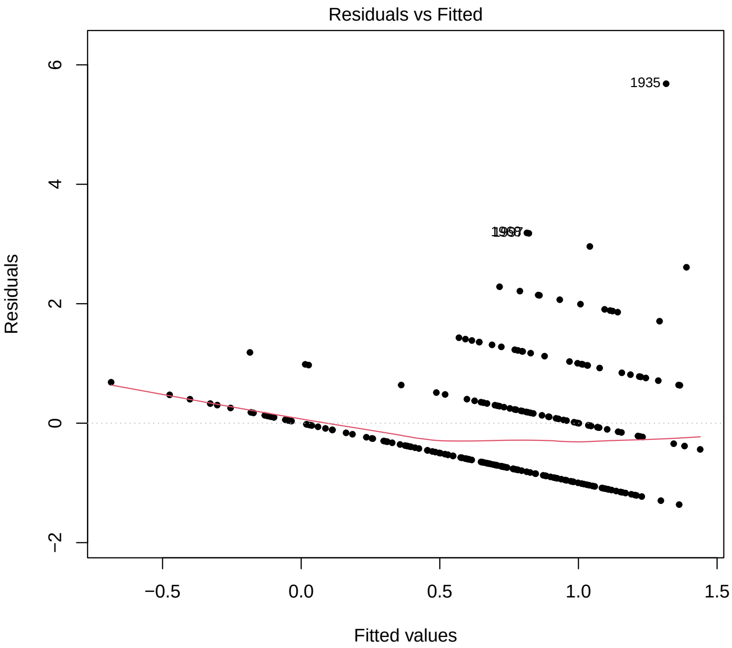
In Action

```
> Fit<-lm(NNulls~Age+Tenure+Unified,data=NewDahl)

> library(gvlma)
> Nope <- gvlma(Fit) # nope
> display.gvlmatests(Nope)

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance =  0.05

Call:
 gvlma(x = Fit)

                    Value   p-value                   Decision
Global Stat        402.68 0.000e+00 Assumptions NOT satisfied!
Skewness           111.88 0.000e+00 Assumptions NOT satisfied!
Kurtosis           243.82 0.000e+00 Assumptions NOT satisfied!
Link Function        5.07 2.434e-02 Assumptions NOT satisfied!
Heteroscedasticity  41.91 9.565e-11 Assumptions NOT satisfied!
```

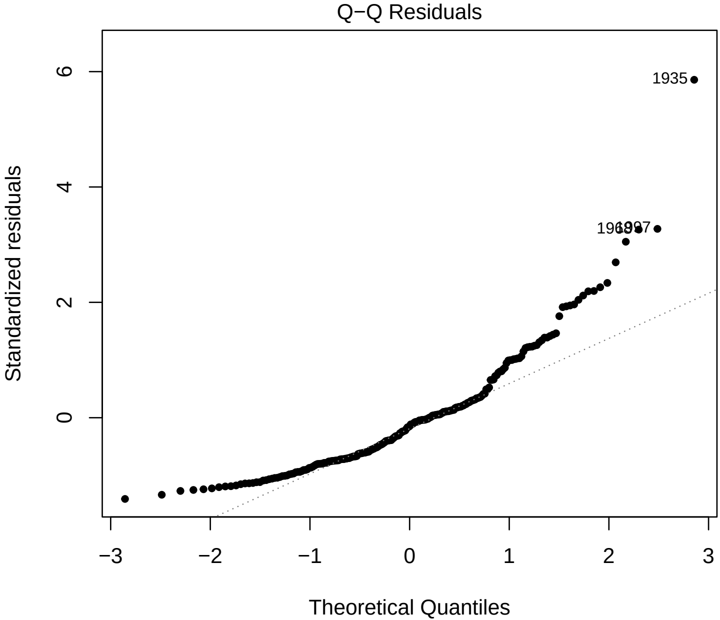
Another Approach: `plot(fit)`



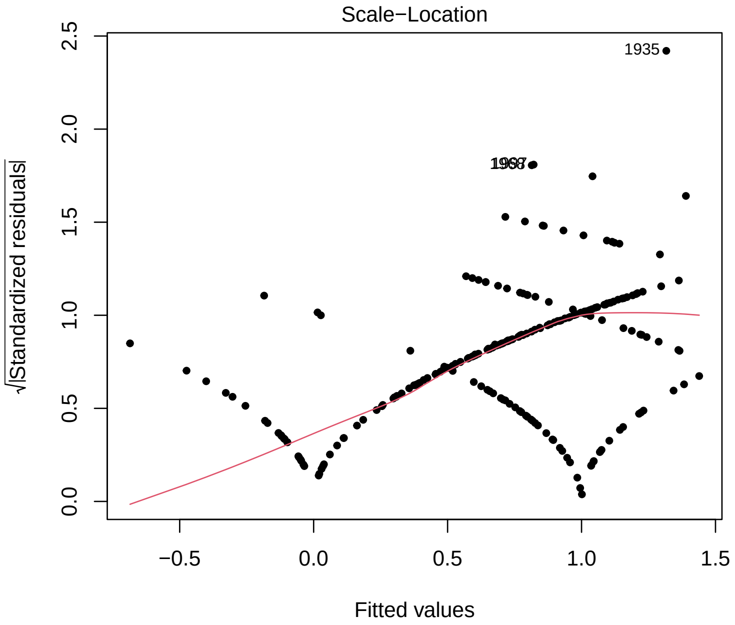
#1: Residuals vs. Fitted Values



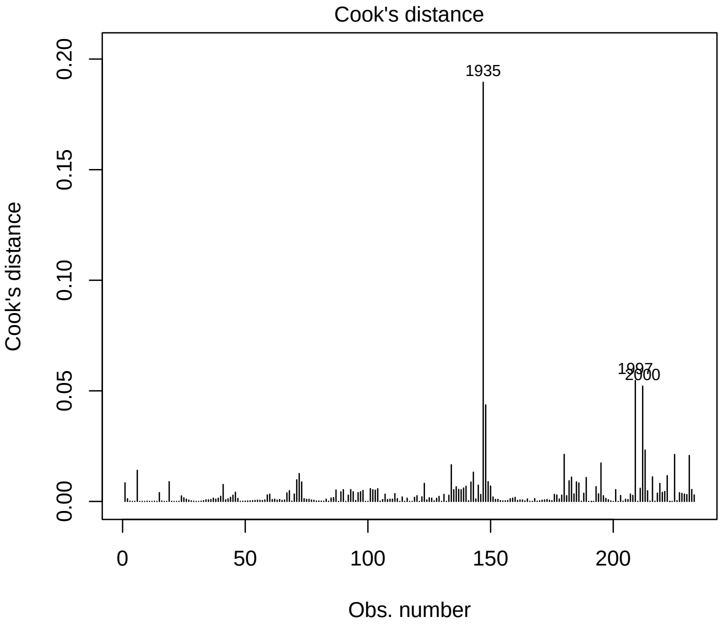
#2: Q-Q Plot of $\hat{u}$s



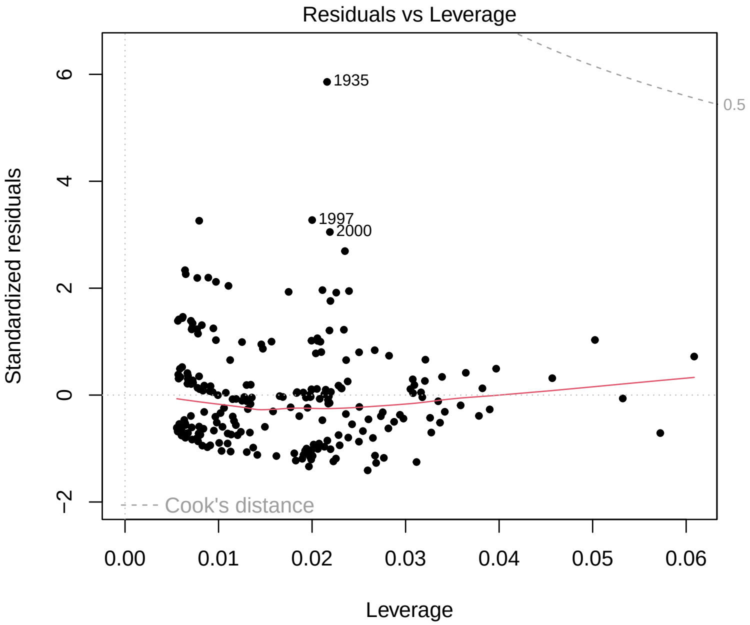
#3: "Scale-Location" Plot



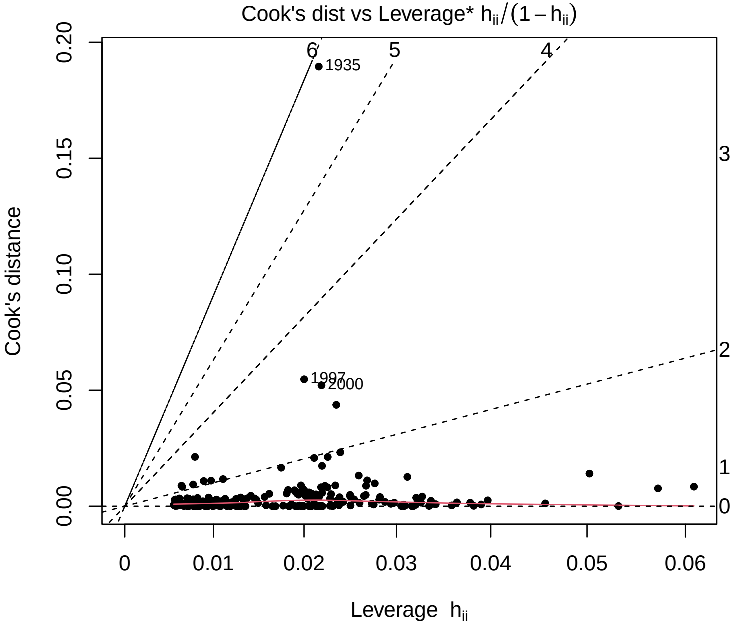
#4: Cook's D


Cook's distance
Cook's distance
0.00
0.05
0.10
0.15
0.20
1935
1997
2000
0
50
100
150
200
Obs. number

#5: Residuals vs. Leverage



#6: Cook's D vs. Leverage



“Variances”

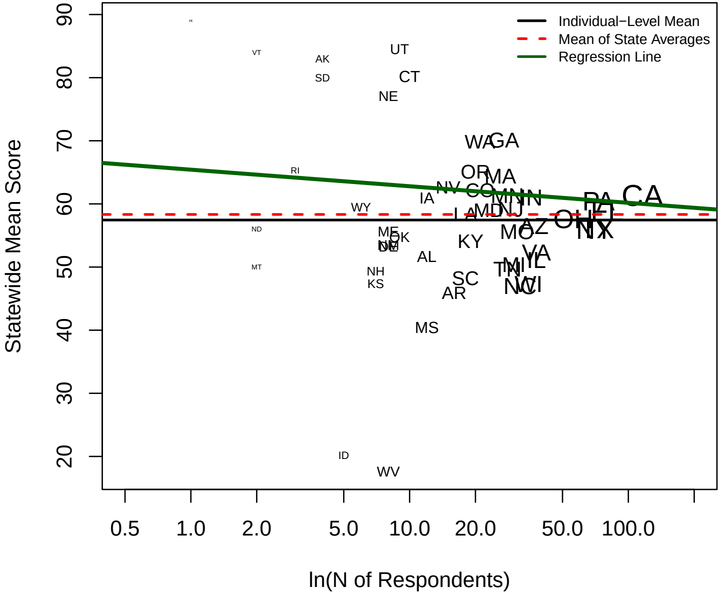
2016 ANES pilot study "feeling thermometer" toward gays and lesbians ($N = 1200$):

```
> summary(ANES$ftgay)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    0.0    40.5    54.0    57.4    88.5   100.0       1
> summary(ANES$presjob)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    2.00    4.00    4.19    7.00    7.00
```

Suppose we wanted to create aggregate measures, by state ($N = 51$). We would get:

```
> summary(StateFT)
    State               Nresp          meantherm      meanpresapp
 Length:50          Min.   :  1.00   Min.   :17.62   Min.   :2.000
 Class :character   1st Qu.:  8.00   1st Qu.:51.33   1st Qu.:3.755
 Mode  :character   Median : 18.00   Median :57.11   Median :4.236
                    Mean   : 24.00   Mean   :58.33   Mean   :4.146
                    3rd Qu.: 30.75   3rd Qu.:62.55   3rd Qu.:4.614
                    Max.   :116.00   Max.   :89.00   Max.   :5.800
```

Variances: Why We Care



Variances: A Generalization

Start with:

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i$$

with:

$$\text{Var}(u_i) = \sigma^2 / w_i$$

with w_i <u>known</u>.

WLS now minimizes:

$$\text{RSS} = \sum_{i=1}^{N} w_i (Y_i - \mathbf{X}_i \boldsymbol{\beta}).$$

which gives:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{WLS} &= [\mathbf{X}'(\sigma^2 \boldsymbol{\Omega})^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\sigma^2 \boldsymbol{\Omega})^{-1}\mathbf{Y} \\
&= [\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{Y}
\end{aligned}$$

where:

$$\mathbf{W} = \begin{bmatrix} \frac{\sigma^2}{w_1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma^2}{w_2} & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\sigma^2}{w_N} \end{bmatrix}$$

The variance-covariance matrix is:

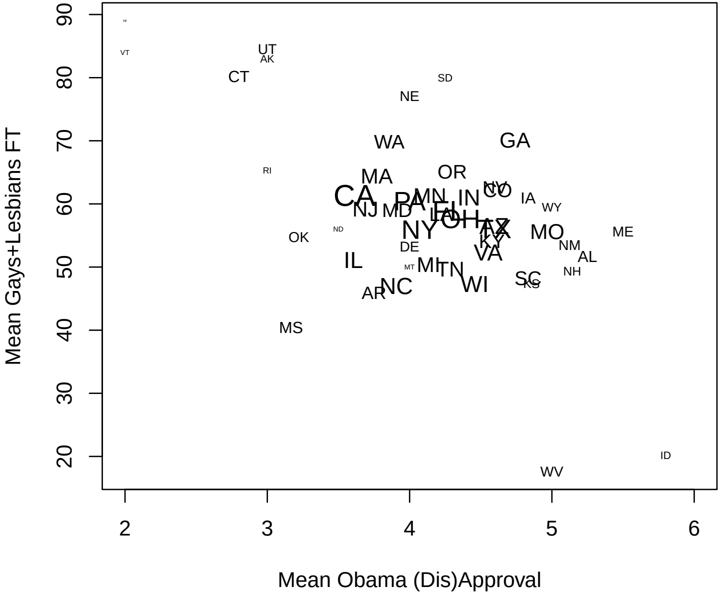
$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}_{WLS}) &= \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \\ &\equiv (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\end{aligned}$$

A common case is:

$$\mathrm{Var}(u_i) = \sigma^2 \frac{1}{N_i}$$

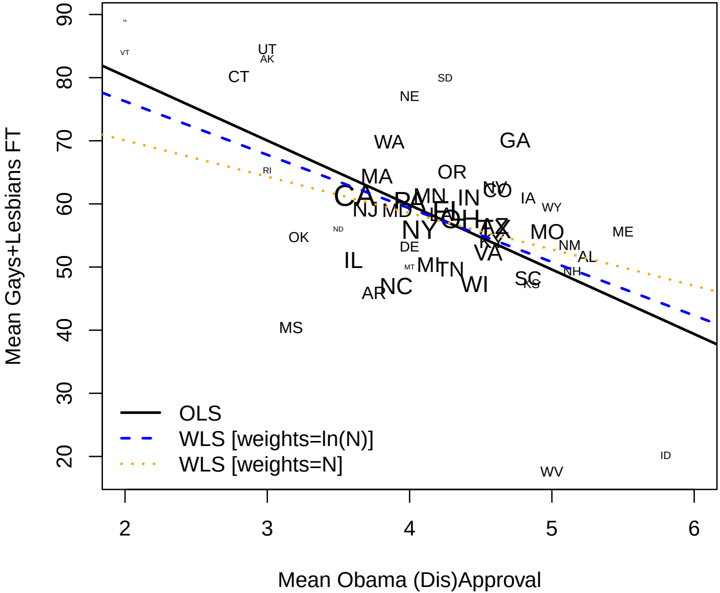
where N_i is the number of observations upon which (aggregate) observation i is based.

Feeling Thermometer Example



	Dependent variable:		
	Mean Gay/Lesbian FTs		
	OLS	WLS [1/ln(N)]	WLS [1/N]
Mean Presidential Approval	-10.216^{***}	-8.483^{***}	-5.756^{**}
	(1.976)	(2.200)	(2.187)
Constant	100.684^{***}	93.221^{***}	81.583^{***}
	(8.343)	(9.378)	(9.238)
Observations	50	50	50
R^2	0.358	0.237	0.126
Adjusted R^2	0.344	0.221	0.108
Residual Std. Error (df = 48)	11.130	17.072	37.914
F Statistic (df = 1; 48)	26.721^{***}	14.870^{***}	6.927^{**}
Note:			*p<0.1; **p<0.05; ***p<0.01

Regressions, Plotted



Recall that, if $\sigma_i^2 \neq \sigma_j^2 \; \forall \; i \neq j$,

$$\begin{aligned} \text{Var}(\beta_{\text{Het.}}) &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{Q}\,(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

where $\mathbf{Q} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})$ and $\mathbf{W} = \sigma^2\boldsymbol{\Omega}$.

We can rewrite $\mathbf{Q}$ as

$$\begin{aligned} \mathbf{Q} &= \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}) \\ &= \sum_{i=1}^{N} \sigma_i^2 \mathbf{X}_i \mathbf{X}_i' \end{aligned}$$

Estimate $\hat{\mathbf{Q}}$ as:

$$\widehat{\mathbf{Q}} = \sum_{i=1}^{N} \hat{u}_i^2 \mathbf{X}_i \mathbf{X}_i'$$

Yields:

$$\begin{aligned}\widehat{\text{Var}(\beta)}_{\text{Robust}} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\widehat{\mathbf{Q}}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\left[\mathbf{X}'\left(\sum_{i=1}^{N} \hat{u}_i^2 \mathbf{X}_i \mathbf{X}_i'\right)^{-1} \mathbf{X}\right](\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

"Robust" standard error estimates:

- are heteroscedasticity-consistent, but
- are biased in small samples, and
- are less efficient than "naive" estimates when $\text{Var}(u) = \sigma^2 \mathbf{I}$.
- Come in various "versions"
  - Called "HC0," "HC1," "HC2," "HC3," etc.
  - See the Long and Ervin (2000) paper for details...

“Clustering”

Huber / White

?????????

WLS / GLS

I know very little about my error variances...

I know a great deal about my error variances...

A common case:

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + u_{ij}$$

with

$$\sigma_{ij}^2 = \sigma_{ik}^2.$$

"Robust, clustered" estimator:

$$\widehat{\text{Var}(\boldsymbol{\beta})}_{\text{Clustered}} = (\mathbf{X}'\mathbf{X})^{-1}\left\{\mathbf{X}'\left[\sum_{i=1}^{N}\left(\sum_{j=1}^{n_j}\hat{u}_{ij}^2\mathbf{X}_{ij}\mathbf{X}_{ij}'\right)\right]^{-1}\mathbf{X}\right\}(\mathbf{X}'\mathbf{X})^{-1}$$

Regressions, Again

	Mean Gay/Lesbian FTs			
	OLS	OLS (robust)	WLS [1/ln(N)]	WLS [1/N]
Mean Presidential Approval	-10.216^{***}	-10.216^{***}	-8.483^{***}	-5.756^{**}
	(1.976)	(2.339)	(2.200)	(2.187)
Constant	100.684^{***}	100.684^{***}	93.221^{***}	81.583^{***}
	(8.343)	(9.722)	(9.378)	(9.238)
Observations	50		50	50
R^2	0.358		0.237	0.126
Adjusted R^2	0.344		0.221	0.108
Residual Std. Error (df = 48)	11.130		17.072	37.914
F Statistic (df = 1; 48)	26.721^{***}		14.870^{***}	6.927^{**}
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Expanded State-Level ANES Example

```
> psych::describe(StateData)
              n  mean    sd median trimmed   mad   min    max  range  skew kurtosis   se
State*       50 25.50 14.58  25.50   25.50 18.53  1.00  50.00  49.00  0.00    -1.27 2.06
NResp        50 24.00 23.74  18.00   19.48 16.31  1.00 116.00 115.00  1.79     3.34 3.36
LGBTTherm    50 58.33 13.74  57.11   58.11  8.51 17.62  89.00  71.38 -0.22     1.40 1.94
MeanCons     50  3.97  0.77   4.00    3.98  0.55  1.50   5.60   4.10 -0.47     1.28 0.11
MeanAge      50  4.74  0.64   4.78    4.74  0.43  3.10   6.50   3.40  0.11     1.10 0.09
MeanEducation 50  3.25  0.52   3.22    3.22  0.41  2.33   5.00   2.67  0.84     1.44 0.07
BornAgainProp 50  0.28  0.18   0.25    0.28  0.19  0.00   0.72   0.72  0.11    -0.62 0.02
```

Basic regression:

```
> OLS<-lm(LGBTTherm~MeanCons+MeanAge+MeanEducation+BornAgainProp,data=StateData)
> summary(OLS)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    71.6464    16.5774   4.322 8.44e-05 ***
MeanCons       -7.9256     2.5441  -3.115   0.0032 **
MeanAge        -2.6688     2.5572  -1.044   0.3022
MeanEducation   9.4772     3.3843   2.800   0.0075 **
BornAgainProp  -0.2273    10.5597  -0.022   0.9829
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.55 on 45 degrees of freedom
Multiple R-squared:  0.4589, Adjusted R-squared:  0.4108
F-statistic: 9.542 on 4 and 45 DF,  p-value: 1.128e-05
```

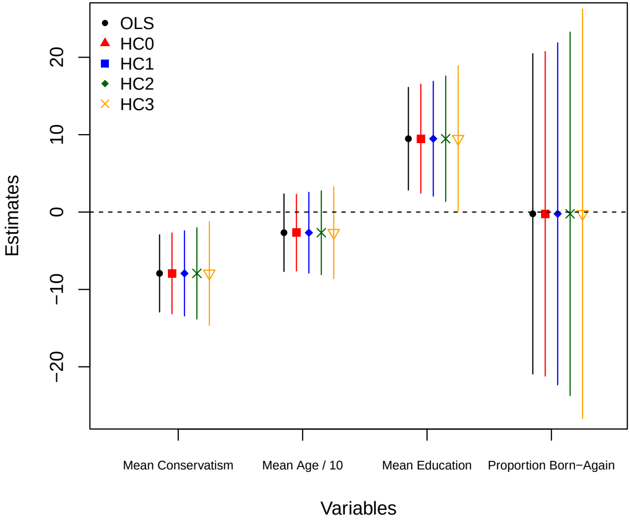
“Robust” SEs

```
> hccm(OLS,type="hc3") # "HC3" var-cov matrix
              (Intercept)   MeanCons      MeanAge MeanEducation BornAgainProp
(Intercept)     605.37713 -43.049519 -37.2508294   -89.9147335    122.746274
MeanCons        -43.04952  11.706690  -1.2344463     4.9693587    -38.744970
MeanAge         -37.25083  -1.234446   9.1700322    -0.6448676     -3.438703
MeanEducation   -89.91473   4.969359  -0.6448676    23.1479923     -4.406169
BornAgainProp   122.74627 -38.744970  -3.4387029    -4.4061691    182.301825
```

```
> sqrt(diag(hccm(OLS,type="hc3"))) # "HC3" robust SEs
  (Intercept)      MeanCons       MeanAge MeanEducation BornAgainProp
    24.604413      3.421504      3.028206      4.811236     13.501919
```

```
> coeftest(OLS,vcov.=vcovHC)

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)   71.64636   24.60441  2.9119 0.005572 **
MeanCons      -7.92559    3.42150 -2.3164 0.025147 *
MeanAge       -2.66876    3.02821 -0.8813 0.382838
MeanEducation  9.47715    4.81124  1.9698 0.055035 .
BornAgainProp -0.22727   13.50192 -0.0168 0.986644
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$\hat{\beta}$s and 95% CIs: Various Types of Robust SEs



Generalized Linear Models (GLMs)

The Exponential Family

$$f(z|\psi) = \Pr(Z = z|\psi)$$

Exponential if:

$$f(z|\psi) = r(z)s(\psi)\exp[q(z)h(\psi)]$$

provided that $r(z) > 0$ and $s(\psi) > 0$.

$$f(z|\psi) = \exp[\underbrace{\ln r(z) + \ln s(\psi)}_{\text{“additive”}} + \underbrace{q(z)h(\psi)}_{\text{“interactive”}}]$$

$$y = q(z)$$

$$\theta = h(\psi)$$

$$f[y|\theta] = \exp[y\theta - b(\theta) + c(y)].$$

- $b(\theta)$ is a "normalizing constant"
- $c(y)$ is a function solely of y
- $y\theta$ is a multiplicative term

A Familiar Family Member: Poisson

$$f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}.$$

$$\begin{aligned} f(y|\lambda) &= \exp\left\{\ln\left[\exp(-\lambda)\lambda^y/y!\right]\right\} \\ &= \exp[\underbrace{y\ln(\lambda)}_{y\theta} - \underbrace{\lambda}_{b(\theta)} - \underbrace{\ln(y!)}_{c(y)}] \end{aligned}$$

Family Nuisances

$$f(y|\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right]$$

Familiar Family Member II: Normal

$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(y-\mu)^2}{2\sigma^2}\right]$$

$$\begin{aligned}
f(y|\mu,\sigma^2) &= \exp\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right] \\
&= \exp\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}y^2 + \frac{1}{2\sigma^2}2y\mu - \frac{1}{2\sigma^2}\mu^2\right] \\
&= \exp\left[\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right] \\
&= \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} + \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\}
\end{aligned}$$

Normal, continued

$$f(y|\mu,\sigma^2) = \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} + \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\}$$

$\theta = \mu$, so:

- $y\theta = y\mu$
- $b(\theta) = \frac{\mu^2}{2}$
- $a(\phi) = \sigma^2$
- $c(y,\phi) = \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]$

Other Family Members

- Binomial ($\supset$ Bernoulli; also Multinomial)
- Exponential
- Gamma
- Logarithmic
- Inverse Gaussian
- Negative Binomial
- others...

Little Red Likelihood

$$\begin{aligned}
\ln L(\theta, \phi|y) &= \ln f(y|\theta, \phi) \\
&= \ln \left\{ \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \right\} \\
&= \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln L(\theta, \phi|y)}{\partial \theta} \equiv \mathbf{S} &= \frac{\partial}{\partial \theta}\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \\
&= \frac{y - \frac{\partial}{\partial \theta} b(\theta)}{a(\phi)}.
\end{aligned}$$

Among family members:

- $\mathbf{S}$ is a sufficient statistic for θ.
- $\mathrm{E}(\mathbf{S}) = 0$.
- $\mathrm{Var}(\mathbf{S}) \equiv \mathcal{I}(\theta) = \mathrm{E}[(\mathbf{S})^2|\theta]$

$$\mathsf{E}(Y) = \frac{\partial}{\partial\theta} b(\theta)$$

and

$$\mathsf{Var}(Y) = a(\phi)\frac{\partial^2}{\partial\theta^2} b(\theta)$$

Example: Poisson Again

$$
\begin{aligned}
\mathsf{E}(Y) &= \frac{\partial}{\partial\theta}\exp(\theta) \\
&= \exp(\theta)|_{\theta=\ln(\lambda)} \\
&= \lambda
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Var}(Y) &= 1 \times \frac{\partial^2}{\partial\theta^2}\exp(\theta)|_{\theta=\ln(\lambda)} \\
&= \exp[\ln(\lambda)] \\
&= \lambda
\end{aligned}
$$

Example: Normal Again

$$
\begin{aligned}
\mathsf{E}(Y) &= \frac{\partial}{\partial\theta}\left(\frac{\theta^2}{2}\right) \\
&= \theta|_{\theta=\mu} \\
&= \mu
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Var}(Y) &= \sigma^2 \times \frac{\partial^2}{\partial\theta^2}\left(\frac{\theta^2}{2}\right) \\
&= \sigma^2 \times \frac{\partial}{\partial\theta}\theta \\
&= \sigma^2
\end{aligned}
$$

Linear Model(s)

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i$$

$$\mathrm{E}(Y_i) \equiv \boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$$

The “Generalized” Part

$$g(\boldsymbol{\mu}_i) = \mathbf{X}_i\boldsymbol{\beta}.$$

$$\begin{aligned}\boldsymbol{\eta}_i &= \mathbf{X}_i\boldsymbol{\beta}\\ &= g(\boldsymbol{\mu}_i)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}_i &= g^{-1}(\boldsymbol{\eta}_i)\\ &= g^{-1}(\mathbf{X}_i\boldsymbol{\beta})\end{aligned}$$

Random component $\sim$ Exponential Family$(\cdot)$ with

$$\mathrm{E}(Y_i) = \boldsymbol{\mu}_i.$$

Systematic component:

$$g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i$$

or

$$g^{-1}(\boldsymbol{\eta}_i) = \boldsymbol{\mu}_i.$$

The Return of The Family

$$
\begin{aligned}
\boldsymbol{\theta}_i &= g(\boldsymbol{\mu}_i) \\
&= \boldsymbol{\eta}_i \\
&= \mathbf{X}_i \boldsymbol{\beta}
\end{aligned}
$$

$$g^{-1}(\boldsymbol{\theta}_i) = \boldsymbol{\mu}_i$$

GLM Example: Linear-Normal

$$f(y|\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2)$$

$$\boldsymbol{\mu}_i = \boldsymbol{\eta}_i$$

$$\begin{aligned} \boldsymbol{\mu}_i \equiv \boldsymbol{\theta}_i &= \boldsymbol{\eta}_i \\ Y_i &\sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma^2) \end{aligned}$$

GLM Example: Binary

$$f(y|\pi) = \pi^y(1-\pi)^{1-y}$$

$$\boldsymbol{\theta}_i = \ln\left(\frac{\boldsymbol{\mu}_i}{1-\boldsymbol{\mu}_i}\right)$$

$$\begin{aligned}
\boldsymbol{\mu}_i &= g^{-1}(\boldsymbol{\theta}_i) \\
&= \frac{\exp(\boldsymbol{\eta}_i)}{1+\exp(\boldsymbol{\eta}_i)} \\
Y_i &\sim \text{Bernoulli}(\boldsymbol{\mu}_i)
\end{aligned}$$

GLM Example: Counts (Independent Events)

$$f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$$

$$\ln(\lambda_i) = \boldsymbol{\eta}_i$$

$$\begin{aligned}
\boldsymbol{\mu}_i &= g^{-1}(\boldsymbol{\theta}_i) \\
&= \exp(\boldsymbol{\eta}_i) \\
Y_i &\sim \text{Poisson}(\lambda_i)
\end{aligned}$$

Common GLM Flavors

Distribution	Range of Y	Link(s) $g(\cdot)$	Inverse Link $g^{-1}(\cdot)$
Normal	$(-\infty, \infty)$	Identity: $\boldsymbol{\theta} = \boldsymbol{\mu}$ (Canonical)	$\boldsymbol{\theta}$
Binomial	$\{0,...n\}$	Logit: $\boldsymbol{\theta} = \ln\left(\frac{\boldsymbol{\mu}}{1-\boldsymbol{\mu}}\right)$ (Canonical)	$\frac{\exp(\boldsymbol{\theta})}{1+\exp(\boldsymbol{\theta})}$
		Probit: $\boldsymbol{\theta} = \Phi^{-1}(\boldsymbol{\mu})$	$\Phi(\boldsymbol{\theta})$
		C-Log-Log: $\boldsymbol{\theta} = \ln[-\ln(1-\boldsymbol{\mu})]$	$1 - \exp[-\exp(\boldsymbol{\theta})]$
Bernoulli	$\{0,1\}$	(same as Binomial)	(same as Binomial)
Multinomial	$\{0,...J\}$	(same as Binomial)	(same as Binomial)
Poisson	$[0, \infty]$ (integers)	Log: $\boldsymbol{\theta} = \ln(\boldsymbol{\mu})$ (Canonical)	$\exp(\boldsymbol{\theta})$
Gamma	$(0, \infty)$	Reciprocal: $\boldsymbol{\theta} = -\frac{1}{\boldsymbol{\mu}}$ (Canonical)	$-\frac{1}{\boldsymbol{\theta}}$

Note: The Bernoulli is a special case of the Binomial with $n = 1$. The multinomial is the J-outcome variant of the Binomial, and is also related to the Poisson (see, e.g., Agresti 2002).

GLMs: How-To

- Pick $f(Y)$
- Pick $g(\cdot)$
- Specify $\mathbf{X}$
- Estimate

- MLE
- IRLS ($\approx$ MLE):

$$\hat{\boldsymbol{\beta}}^{(t+1)} = [\mathbf{X}'\mathbf{W}^{(t)}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}^{(t)}\mathbf{z}^{(t)}$$

with

$$\underset{N\times N}{\mathbf{W}^{(t)}} = \mathrm{diag}\left[\frac{\left(\partial\mu_i^{(t)}/\partial\eta_i^{(t)}\right)^2}{\mathrm{Var}(Y_i)}\right]$$

and

$$\mathbf{z}^{(t)} \quad = \quad \boldsymbol{\eta}^{(t)} + (Y - \boldsymbol{\mu}^{(t)})\left(\frac{\partial\boldsymbol{\eta}^{(t)}}{\partial\boldsymbol{\mu}}\right).$$

IRLS, Intuitively

At iteration t:

1. Calculate $\mathbf{z}^{(t)}$, $\mathbf{W}^{(t)}$
2. Regress $\mathbf{z}^{(t)}$ on $\mathbf{X}$, using $\mathbf{W}^{(t)}$ as weights, to obtain $\hat{\boldsymbol{\beta}}^{(t+1)}$
3. Generate $\boldsymbol{\eta}^{(t+1)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(t+1)}$
4. Generate $\boldsymbol{\mu}^{(t+1)} = g^{-1}(\boldsymbol{\eta}^{(t+1)})$
5. Use $\boldsymbol{\eta}^{(t+1)}$ and $\boldsymbol{\mu}^{(t+1)}$ to calculate $\mathbf{z}^{(t+1)}$ and $\mathbf{W}^{(t+1)}$
6. Repeat until convergence.

"Response" Residuals:

$$\begin{aligned}\hat{u}_i &= Y_i - \hat{\mu}_i \\ &= Y_i - g^{-1}(\mathbf{X}_i\hat{\boldsymbol{\beta}})\end{aligned}$$

"Pearson" Residuals:

$$\hat{P}_i = \frac{\hat{u}_i}{[\text{Var}(\hat{u}_i)]^{1/2}}$$

"Deviance":

$$\begin{aligned}
\hat{d}_i &= -2[\ln L_i(\hat{\theta}) - \ln L_i(\theta_S)] \\
&= 2\left\{\left[\frac{Y_i\theta_S - b(\theta_S)}{a(\phi)} + c(Y_i,\phi)\right] - \left[\frac{Y_i\hat{\theta} - b(\hat{\theta})}{a(\phi)} + c(Y_i,\phi)\right]\right\} \\
&= 2\left[\frac{Y_i(\theta_S - \hat{\theta}) - b(\theta_S) + b(\hat{\theta})}{a(\phi)}\right]
\end{aligned}$$

"Deviance" Residuals:

$$\hat{r}_{Di} = \left(\frac{\hat{u}_i}{|\hat{u}_i|}\right)\sqrt{\hat{d}_i^2}$$

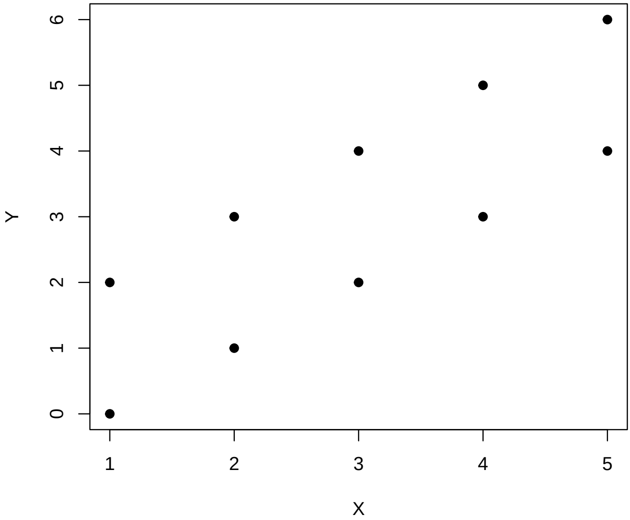
Toy Example: Linear-Normal

$$
\begin{aligned}
X &= \{1, 1, 2, 2, 3, 3, 4, 4, 5, 5\} \\
Y &= \{0, 2, 1, 3, 2, 4, 3, 5, 4, 6\}
\end{aligned}
$$

$$
\begin{aligned}
Y_i &= 0 + 1X_i + u_i \\
\hat{u}_i^2 &= 1 \,\forall\, i
\end{aligned}
$$

$$
\begin{aligned}
\text{“TSS”} \equiv \sum (Y_i - \bar{Y})^2 &= 30 \\
\text{“RSS”} \equiv \sum \hat{u}_i^2 &= 10 \\
\text{“MSS”} \;/\; \text{“ESS”} &= 20
\end{aligned}
$$

Toy Example: Plot



Toy Example: OLS

```
> linmod<-lm(Y~X)
> summary(linmod)

Call:
lm(formula = Y ~ X)

Residuals:
   Min     1Q Median     3Q    Max 
    -1     -1      0      1      1 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)   
(Intercept) 6.661e-16  8.292e-01       0  1.00000   
X           1.000e+00  2.500e-01       4  0.00395 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.118 on 8 degrees of freedom
Multiple R-squared:  0.6667, Adjusted R-squared:  0.625 
F-statistic:    16 on 1 and 8 DF,  p-value: 0.00395
```

Toy Example: Linear-Normal GLM

```
> linglm<-glm(Y~X,family="gaussian")
> summary(linglm)

Call:
glm(formula = Y ~ X, family = "gaussian")

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.661e-16  8.292e-01       0  1.00000
X           1.000e+00  2.500e-01       4  0.00395 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.25)

    Null deviance: 30  on 9  degrees of freedom
Residual deviance: 10  on 8  degrees of freedom
AIC: 34.379

Number of Fisher Scoring iterations: 2
```

Better GLM Example: Political Knowledge

- 2008 NES political knowledge
- Identify Speaker of the House, VP, British PM, and Chief Justice
- $Y_i =$ number of correct answers (out of four)

$$f(Y_i, p_i) = \binom{4}{Y_i} p_i^{Y_i} (1 - p_i)^{4 - Y_i}$$

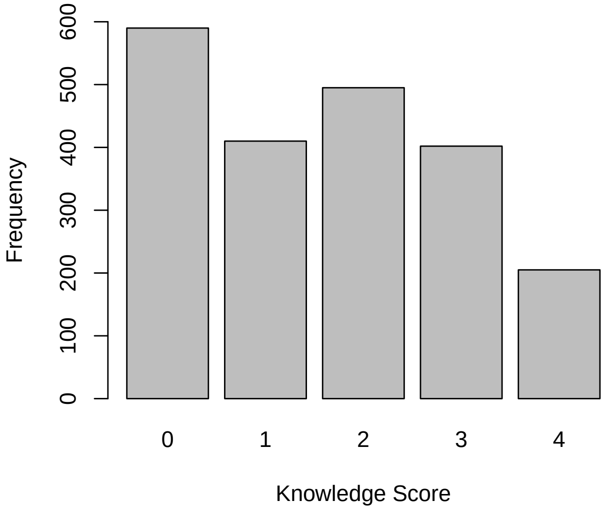
$$\begin{aligned} Y & \sim \text{Binomial}(4, p) \\ \text{E}(Y_i) & = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})} \end{aligned}$$

GLM Example Data (2008 NES)

```
> psych::describe(NES08[,4:16],fast=TRUE,skew=TRUE)
```

	vars	n	mean	sd	median	min	max	range	skew	kurtosis	se
knowledge	1	2102	1.63	1.33	2	0	4	4	0.23	-1.14	0.03
sex	2	2323	NaN	NA	NA	Inf	-Inf	-Inf	NA	NA	NA
race	3	2323	NaN	NA	NA	Inf	-Inf	-Inf	NA	NA	NA
age	4	2323	NaN	NA	NA	Inf	-Inf	-Inf	NA	NA	NA
female	5	2323	0.57	0.50	1	0	1	1	-0.28	-1.92	0.01
white	6	2323	0.62	0.49	1	0	1	1	-0.50	-1.75	0.01
oftenvote	7	2323	NaN	NA	NA	Inf	-Inf	-Inf	NA	NA	NA
conservative	8	1626	4.14	1.54	4	1	7	6	-0.13	-0.74	0.04
prayfreq	9	2323	NaN	NA	NA	Inf	-Inf	-Inf	NA	NA	NA
heterosexual	10	2274	0.96	0.20	1	0	1	1	-4.63	19.47	0.00
married	11	2308	0.42	0.49	0	0	1	1	0.31	-1.90	0.01
yrsofschool	12	2312	13.08	2.59	13	0	17	17	-0.75	1.81	0.05
income	13	2172	10.52	6.20	11	1	25	24	0.13	-0.82	0.13

Political Knowledge (2008 NES)



```
> nes08.binom<-glm(cbind(knowledge,4-knowledge)~age+female+white+
+                      conservative+heterosexual+married+yrsofschool+
+                      income,data=NES08,family=binomial)
> summary(nes08.binom)

Call:
glm(formula = cbind(knowledge, 4 - knowledge) ~ age + female +
    white + conservative + heterosexual + married + yrsofschool +
    income, family = binomial, data = NES08)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.155103   0.247251  -8.716  < 2e-16 ***
age           0.012301   0.001806   6.811 9.68e-12 ***
female       -0.229704   0.058547  -3.923 8.73e-05 ***
white         0.185427   0.063213   2.933  0.00335 **
conservative -0.030712   0.018824  -1.632  0.10277
heterosexual -0.073058   0.138717  -0.527  0.59842
married       0.170145   0.057919   2.938  0.00331 **
yrsofschool   0.099083   0.012940   7.657 1.90e-14 ***
income        0.010540   0.005197   2.028  0.04256 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3167.1  on 1350  degrees of freedom
Residual deviance: 2962.6  on 1342  degrees of freedom
  (972 observations deleted due to missingness)
AIC: 4558.4

Number of Fisher Scoring iterations: 4
```

GLMs: Other Topics + Extensions

Other Topics:

- Generalizations for Overdispersion (binomial)
- Diagnostics (leverage, etc.)
- Joint Mean-Dispersion Models

Extensions:

- Bias-reduced models (a la Firth 1993)
- "Generalized additive models" (GAMs)
- "Generalized estimating equations" (GEEs)
- "Vector" GLMs (Yee and Wild 1996; Yee and Hastie 2003)

GLMs: References

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd Ed. London: Chapman & Hall.

Dobson, Annette J., and and Adrian G. Barnett. 2008. *An Introduction to Generalized Linear Models*, 3rd Ed. London: Chapman & Hall.

Faraway, Julian J. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models.* London: Chapman & Hall / CRC.

Dunn, Peter K., and Gordon K. Smyth. 2018. *Generalized Linear Models With Examples in R*. New York: Springer.

Hardin, James W., and Joseph W. Hilbe. 2012. *Generalized Linear Models and Extensions*, 3rd Ed. College Station, TX: Stata Press.

Binary Response Models

Linear Probability Model (LPM)

$$E(Y) = \mathbf{X}\boldsymbol{\beta}$$

$$Y \in \{0, 1\}$$

$$\begin{aligned} E(Y) &= 1[\Pr(Y = 1)] + 0[\Pr(Y = 0)] \\ &= \Pr(Y = 1) \end{aligned}$$

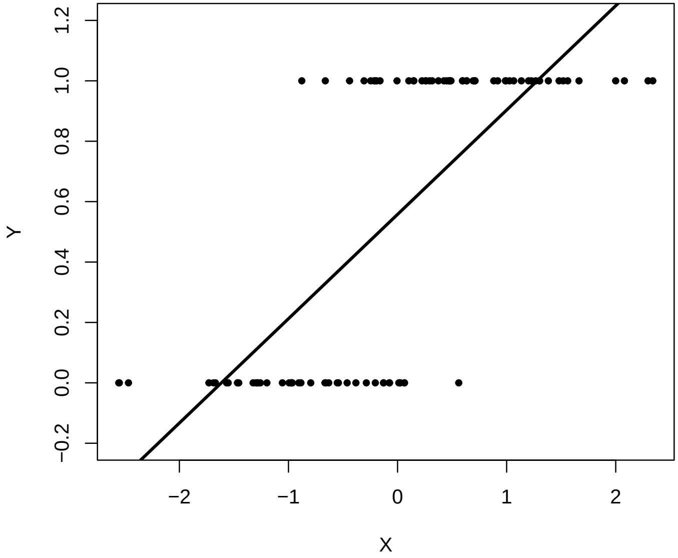
So:

$$\Pr(Y_i = 1) = \mathbf{X}_i\boldsymbol{\beta}$$

or:

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + u_i$$

LPM Illustrated



LPM Issues

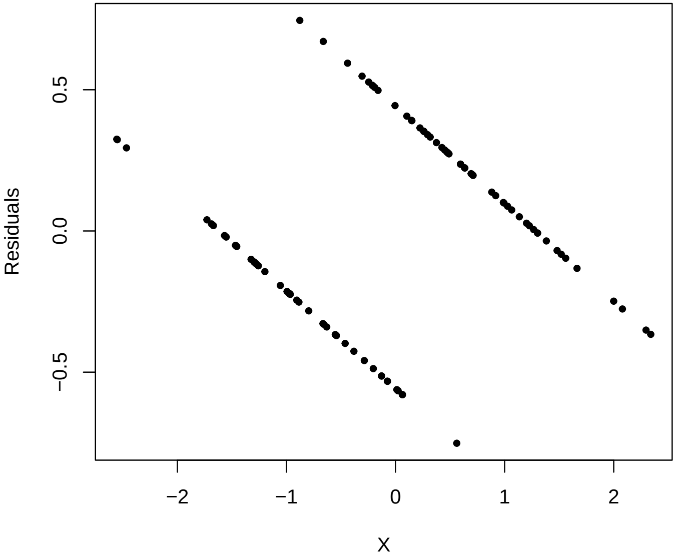
Variance:

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{E}(Y)[1 - \mathrm{E}(Y)] \\ &= \mathbf{X}_i\boldsymbol{\beta}(1 - \mathbf{X}_i\boldsymbol{\beta}) \end{aligned}$$

Residuals:

$$\hat{u}_i \in \{1 - \mathbf{X}_i\hat{\boldsymbol{\beta}}, -\mathbf{X}_i\hat{\boldsymbol{\beta}}\}$$

LPM Residuals



Whither The LPM?

Various thoughts:

- Issues:
  - Model misspecification → bias, inconsistency
  - Creates heteroscedasticity
  - Can yield predicted values outside $(0, 1)$
  - See, e.g., See: Chen, Kaicheng, Robert S. Martin, and Jeffrey M. Wooldridge. 2023. "Another Look at the Linear Probability Model and Nonlinear Index Models." Working paper: Michigan State University.
- The rehabilitation of the LPM:
  - "Logit is hard" / "OLS is awesome" / "It doesn't matter anyway"
  - More-or-less entirely due to (famous) economists
  - Examples: here, here, etc.
- Takeaway: **Pay attention to what people in your discipline / field are doing.**

Start with:

$$Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + u_i$$

And:

$$\begin{aligned} Y_i &= 0 \text{ if } Y_i^* < 0 \\ Y_i &= 1 \text{ if } Y_i^* \geq 0 \end{aligned}$$

So:

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(Y_i^* \geq 0) \\ &= \Pr(\mathbf{X}_i\boldsymbol{\beta} + u_i \geq 0) \\ &= \Pr(u_i \geq -\mathbf{X}_i\beta) \\ &= \Pr(u_i \leq \mathbf{X}_i\beta) \\ &= \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}} f(u)du \end{aligned}$$

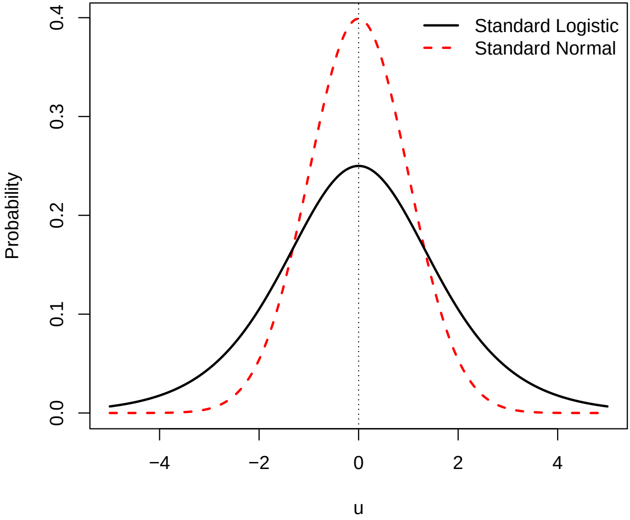
"Standard logistic" PDF:

$$\Pr(u) \equiv \lambda(u) = \frac{\exp(u)}{[1+\exp(u)]^2}$$

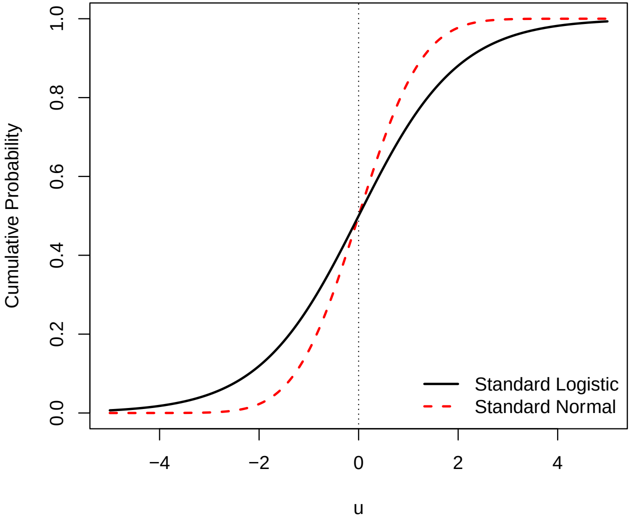
CDF:

$$\begin{aligned}\Lambda(u) &= \int \lambda(u) du \\ &= \frac{\exp(u)}{1+\exp(u)} \\ &= \frac{1}{1+\exp(-u)}\end{aligned}$$

Standard Normal and Logistic PDFs



Standard Normal and Logistic CDFs



For the standard logistic:

- $\lambda(u) = 1 - \lambda(-u)$
- $\Lambda(u) = 1 - \Lambda(-u)$
- $\mathrm{Var}(u) = \frac{\pi^2}{3} \approx 3.29$

Logistic → "Logit"

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\
&= \Pr(u_i \leq \mathbf{X}_i\boldsymbol{\beta}) \\
&= \Lambda(\mathbf{X}_i\boldsymbol{\beta}) \\
&= \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}
\end{aligned}
$$

$$
\text{(equivalently)} = \frac{1}{1 + \exp(-\mathbf{X}_i\boldsymbol{\beta})}
$$

Likelihoods

$$L_i = \left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]^{1-Y_i}$$

$$L = \prod_{i=1}^{N}\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]^{1-Y_i}$$

$$\begin{aligned}\ln L \quad = \quad & \sum_{i=1}^{N} Y_i \ln\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right) + \\ & (1-Y_i)\ln\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]\end{aligned}$$

Digression I: Logit as an Odds Model

$$\text{Odds}(Z) \equiv \Omega(Z) = \frac{\Pr(Z)}{1 - \Pr(Z)}.$$

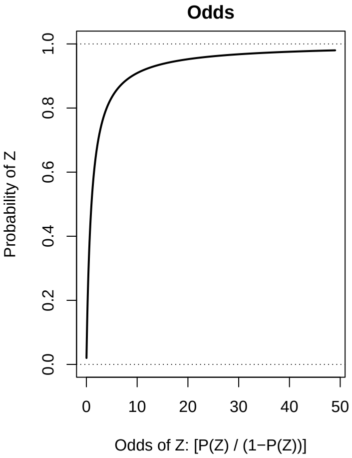
$$\ln[\Omega(Z)] = \ln\left[\frac{\Pr(Z)}{1 - \Pr(Z)}\right]$$

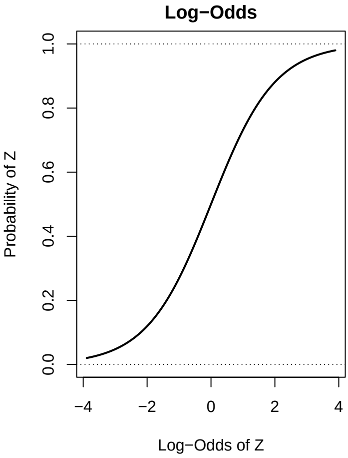
$$\ln[\Omega(Z_i)] = \mathbf{X}_i\boldsymbol{\beta}$$

$$\begin{aligned} \Omega(Z_i) &= \frac{\Pr(Z)}{1 - \Pr(Z)} \\ &= \exp(\mathbf{X}_i\boldsymbol{\beta}) \end{aligned}$$

$$\Pr(Z_i) = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}$$

Visualizing Log-Odds





Y Be Normal?

Standard Normal PDF:

$$\Pr(u) \equiv \phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

Standard Normal CDF:

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

Normal → "Probit"

$$\begin{aligned}\Pr(Y_i = 1) &= \Phi(\mathbf{X}_i\boldsymbol{\beta}) \\ &= \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{X}_i\boldsymbol{\beta})^2}{2}\right) d\mathbf{X}_i\boldsymbol{\beta}\end{aligned}$$

$$L = \prod_{i=1}^{N} \left[\Phi(\mathbf{X}_i\boldsymbol{\beta})\right]^{Y_i} \left[1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})\right]^{(1-Y_i)}$$

$$\ln L = \sum_{i=1}^{N} Y_i \ln\Phi(\mathbf{X}_i\boldsymbol{\beta}) + (1 - Y_i) \ln[1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})]$$

Digression II: The Random Utility Model

$$Y \in \{SQ, A\}$$

$$\begin{aligned} Y_i &= A \quad \text{if} \quad \mathrm{E}[\mathrm{U}_i(A)] \geq \mathrm{E}[\mathrm{U}_i(SQ)] \\ &= SQ \quad \text{if} \quad \mathrm{E}[\mathrm{U}_i(A)] < \mathrm{E}[\mathrm{U}_i(SQ)] \end{aligned}$$

$$\mathrm{E}[\mathrm{U}_i(A)] = \mathbf{X}_{iA}\boldsymbol{\beta} + u_{iA}$$

So:

$$\begin{aligned} \Pr(Y = A) &= \Pr\{\mathrm{E}[\mathrm{U}_i(A)] \geq \mathrm{E}[\mathrm{U}_i(SQ)]\} \\ &= \Pr\{(\mathbf{X}_{iA}\boldsymbol{\beta} + u_{iA}) \geq \mathrm{E}[\mathrm{U}_i(SQ)]\} \end{aligned}$$

Digression II: The Random Utility Model

Normalize:

$$\mathrm{E}[\mathrm{U}_i(SQ)] = 0$$

Then:

$$\begin{aligned}
\Pr(Y = A) &= \Pr\{(\mathbf{X}_{iA}\boldsymbol{\beta} + u_{iA}) \geq 0\} \\
&= \Pr\{u_{iA} \geq -\mathbf{X}_{iA}\boldsymbol{\beta}\} \\
&= F(\mathbf{X}_{iA}\boldsymbol{\beta})
\end{aligned}$$

Another Model: Complementary Log-Log

Uses:

$$\Pr(Y_i = 1) = 1 - \exp[-\exp(\mathbf{X}_i\beta)]$$

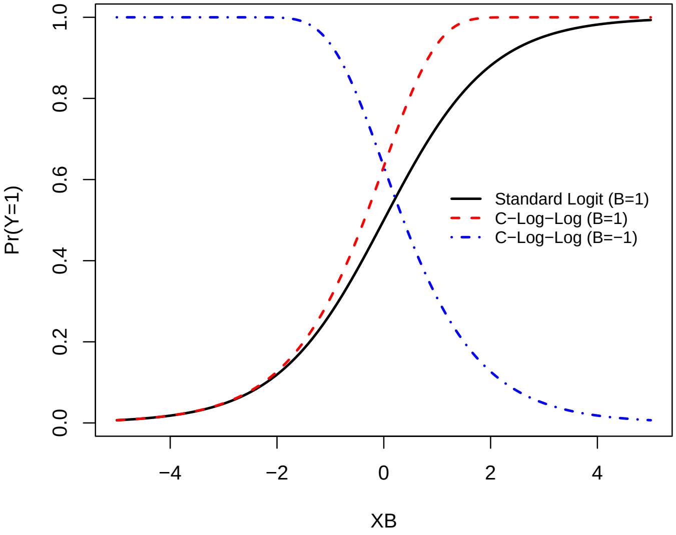
or

$$\ln\{-\ln[1 - \Pr(Y_i = 1)]\} = \mathbf{X}_i\beta$$

Likelihood is:

$$\begin{aligned} \ln L &= \sum_{i=1}^{N} Y_i \ln\{1 - \exp[-\exp(\mathbf{X}_i\beta)]\} + \\ &\quad (1 - Y_i) \ln\{1 - \{1 - \exp[-\exp(\mathbf{X}_i\beta)]\}\} \end{aligned}$$

Logit and C-log-log CDFs



All require that:

- "Threshold" $= Y^* > 0$
- $\mathrm{E}(u_i|\mathbf{X}, \boldsymbol{\beta}) = 0$
- $\mathrm{Var}(u_i) = \frac{\pi^2}{3}$ or 1.0.

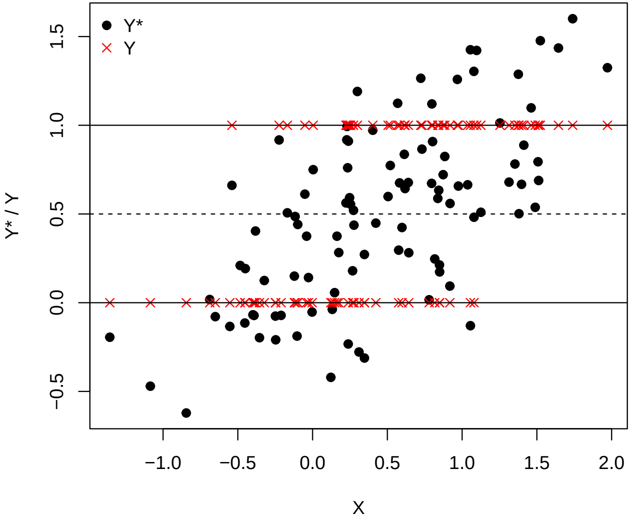
In general:

- The Universe: Logit $>$ Probit
- The (Social Science) Universe: Meh...
- $\hat{\boldsymbol{\beta}}_{\text{Logit}} \approx 1.8 \times \hat{\boldsymbol{\beta}}_{\text{Probit}}$
- Four reasons to prefer / use logit

A Toy Example

```
> set.seed(7222009)
> ystar<-rnorm(100,0.5,0.5)
> y<-ifelse(ystar>0.5,1,0)
> x<-ystar+(0.5*rnorm(100))
> data<-data.frame(ystar,y,x)

> head(data)
     ystar y       x
1  0.17977 0  0.2677
2  0.79428 1  1.5079
3  0.82408 1  0.8842
4  0.24658 0  0.8172
5  0.50966 1  1.1255
6 -0.07852 0 -0.6506
```

A Toy Example



```
> mylogit<-glm(y~x,family=binomial(link="logit"), data=data)
> myprobit<-glm(y~x,family=binomial(link="probit"), data=data)
> mycloglog<-glm(y~x,family=binomial(link="cloglog"), data=data)
```

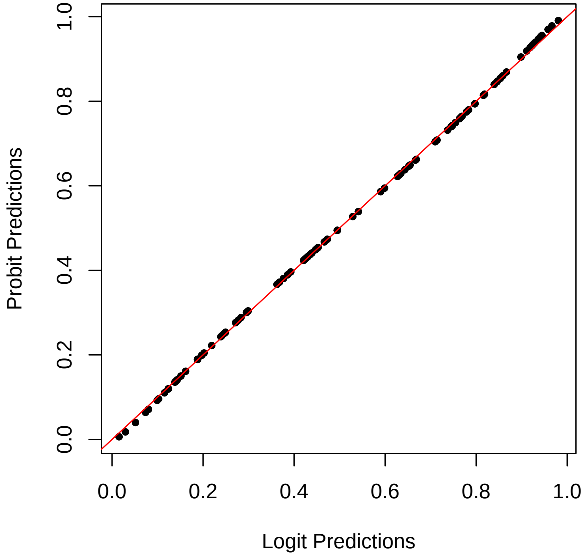
Logit, Probit, and C-Log-Log Models (Simulated Data)

	Logit	*Probit*	*C-Log-Log*
X	2.428^{***}	1.458^{***}	1.613^{***}
	(0.500)	(0.272)	(0.309)
Constant	-0.861^{***}	-0.519^{***}	-1.048^{***}
	(0.318)	(0.183)	(0.250)
Observations	100	100	100
Log Likelihood	-49.690	-49.490	-49.522
Akaike Inf. Crit.	103.380	102.979	103.044
Note:			*p<0.1; **p<0.05; ***p<0.01

Note:

- zs, Ps, $\ln L$s, AICs nearly identical
- Residuals, too
- $\hat{\beta}_{\text{Logit}}$ is $\frac{2.428}{1.458} = 1.54 \times \hat{\beta}_{\text{Probit}}$

Toy Example: Predicted Probabilities



Note: C-Log-Log Isn't "Reversible"

Suppose we generate a new dependent variable:

$$Y_{i\text{New}} = 1 - Y_i$$

What happens to our estimates?

	$\hat{\beta}_0$			$\hat{\beta}_1$		
	Y		Y_{New}	Y		Y_{New}
Probit	-0.52	$\leftrightarrow$	0.52	1.46	$\leftrightarrow$	-1.46
Logit	-0.86	$\leftrightarrow$	0.86	2.43	$\leftrightarrow$	-2.43
C-Log-Log	-1.05	$\leftrightarrow$	0.11	1.61	$\leftrightarrow$	-1.66