PLSC 503 – Spring 2024

MLE: Testing and Inference + An Introduction to Models for Binary Responses

Christopher Zorn

March 25, 2024

Testing: The Plan

- “The Trinity”
- An example
- Practical advice

Inference, In General

1. Pick some $\mathbf{H}_A : \boldsymbol{\Theta} = \boldsymbol{\Theta}_A$
2. Estimate $\hat{\boldsymbol{\Theta}}$
3. Determine distribution of $\hat{\boldsymbol{\Theta}}$ under $\mathbf{H}_A$
4. Use (2) and (3) $\rightarrow \hat{\mathbf{S}} \sim h(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}})$ (*test statistic*)
5. Assess $\Pr(\hat{\mathbf{S}}|\mathbf{H}_A)$

Single Coefficients: Significance Testing

Consistency / Efficiency / Normality:

$$\hat{\boldsymbol{\Theta}}_{MLE} \overset{a}{\sim} \mathbf{N}[\boldsymbol{\Theta}, \mathbf{I}(\hat{\boldsymbol{\Theta}}_{MLE})]$$

Means that

$$\frac{\hat{\theta}_k - \theta_k}{\sqrt{\hat{\sigma}_k^2}} \sim N(0, 1)$$

So:

- Choose θ_A
- Estimate $\hat{\theta}_k$, $\hat{\sigma}_k^2$
- Compare $z_k = \frac{\hat{\theta}_k - \theta_A}{\sqrt{\hat{\sigma}_k^2}}$ to a z-table
- (Or, just look at your output...)

Single Coefficients: Confidence Intervals

- $\alpha \in (0, 1)$ = desired level of "significance"
- $(1 - \alpha) \times 100$-percent confidence intervals for $\hat{\theta}_k$ are:

$$\hat{\theta}_k \pm \left(z_\alpha \sqrt{\hat{\sigma}_k^2} \right)$$

- (Or just look at your output...)

More General Tests: "The Trinity"

- Likelihood-Ratio ("LR")
- Wald
- Lagrangian Multiplier (or "score")

Traits:

- Wald, LM $\xrightarrow{a}$ LR
- For the linear model, Wald $\geq$ LR $\geq$ LM

Generally:

$$\mathbf{R}\boldsymbol{\Theta} = \mathbf{r}$$

For one parameter:

$$\theta_2 = -2 \iff (0\ 1\ 0) \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = -2$$

For $>$ one parameter:

$$\boldsymbol{\Theta}_A : \theta_2 = 1, \theta_1 = 2\theta_3$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$r = \text{rows}(\mathbf{R}) \in [0, K]$$

We know that:

$$L(\hat{\mathbf{\Theta}}) \geq L(\mathbf{\Theta_A}), \text{ but}$$

By how much?

Odds of one thing vs. another:

$$\frac{\Pr(\text{Something})}{\Pr(\text{Something Else})}$$

Implies:

$$\frac{L(\mathbf{\Theta_A})}{L(\hat{\mathbf{\Theta}})} \ (\leq 1)$$

Suggests:

$$\ln L(\mathbf{\Theta_A}) - \ln L(\hat{\mathbf{\Theta}}) \ (\leq 0)$$

$$-2[\ln L(\mathbf{\Theta_A}) - \ln L(\hat{\mathbf{\Theta}})] \overset{a}{\sim} \chi^2_r$$

Traits:

- Intuition: Difference in $\ln L$ under constraint(s)
- Asymptotic
- Unreliable if $r > 100$ (or so)
- Easy to compute, but
- Requires that we have $\ln L(\boldsymbol{\Theta}_{\mathbf{A}})$ *and* $\ln L(\hat{\boldsymbol{\Theta}})$

Idea: If $\mathbf{\Theta_A}$, then

$$\mathbf{R\Theta} = \mathbf{r}$$

So:

$$\mathbf{R\Theta} - \mathbf{r} = \mathbf{0}$$

But...

- We have only $\hat{\mathbf{\Theta}}$ (from sample data)
- Possible that $\mathbf{R}\hat{\mathbf{\Theta}} - \mathbf{r} = \mathbf{0}$ *due to chance* (sampling variability).
- Solution: Account for *variability* in $\hat{\mathbf{\Theta}}$.

Test statistic:

$$\mathbf{W} = (\mathbf{R}\hat{\boldsymbol{\Theta}} - \mathbf{r})' \left[\mathbf{R}\,\mathrm{Var}(\hat{\boldsymbol{\Theta}})\,\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\Theta}} - \mathbf{r})$$

Distribution:

$$\mathbf{W} \overset{a}{\sim} \chi^2_r$$

Traits:

(+) Easy, fast

(+) No need for $\ln L(\boldsymbol{\Theta}_\mathbf{A})$

(–) Uses $\mathrm{Var}(\hat{\boldsymbol{\Theta}})$, not $\mathrm{Var}(\boldsymbol{\Theta}_\mathbf{A})$ (potentially poor coverage)

(–) Can yield nonsensical results

Idea: If $\mathbf{\Theta_A}$, then

$$\left.\frac{\partial \ln L}{\partial \theta}\right|_{\mathbf{\Theta_A}} \approx \mathbf{0}$$

Consider a new problem:

$$\max_{\mathbf{\Theta}} \left[L(\mathbf{\Theta}) - \boldsymbol{\lambda}(\mathbf{\Theta} - \mathbf{\Theta_A})\right]$$

Yields:

$$\tilde{\mathbf{\Theta}} = \mathbf{\Theta_A}$$

$$\tilde{\boldsymbol{\lambda}} = \mathbf{g}(\tilde{\mathbf{\Theta}})$$

LM Tests (continued)

Suggests

$$LM = \mathbf{g}(\tilde{\boldsymbol{\Theta}})' \, \mathbf{I}(\tilde{\boldsymbol{\Theta}})^{-1} \mathbf{g}(\tilde{\boldsymbol{\Theta}})$$

$$LM \overset{a}{\sim} \chi^2_r$$

<u>Traits</u>:

$(+)$ No need for $\hat{\boldsymbol{\Theta}}$!

(–) No information on $\hat{\boldsymbol{\Theta}}$...

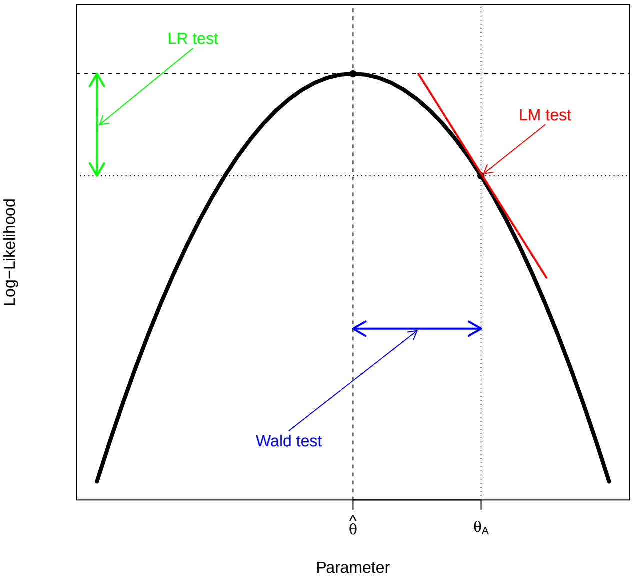
Tests, Conceptually (C. Franklin remix)

- The LR asks, "Did the likelihood change much under the null hypotheses versus the alternative?"
- The Wald test asks, "Are the estimated parameters very far away from what they would be under the null hypothesis?"
- The LM test asks, "If I had a less restrictive likelihood function, would its derivative be close to zero here at the restricted ML estimate?"

Tests, Conceptually (h.t.: Buse 1982)

- LR test $\approx$ manic mountaineer
- Wald test $\approx$ tired mountaineer
- LM test $\approx$ lazy mountaineer

Tests, Conceptually (adapted from Fox 1997, p. 570)



Tests, Practically

- All are asymptotically identical...
- Require different estimates, but similar information
- Generally, preference goes LR $>$ Wald $>$ LM

Tests in R:

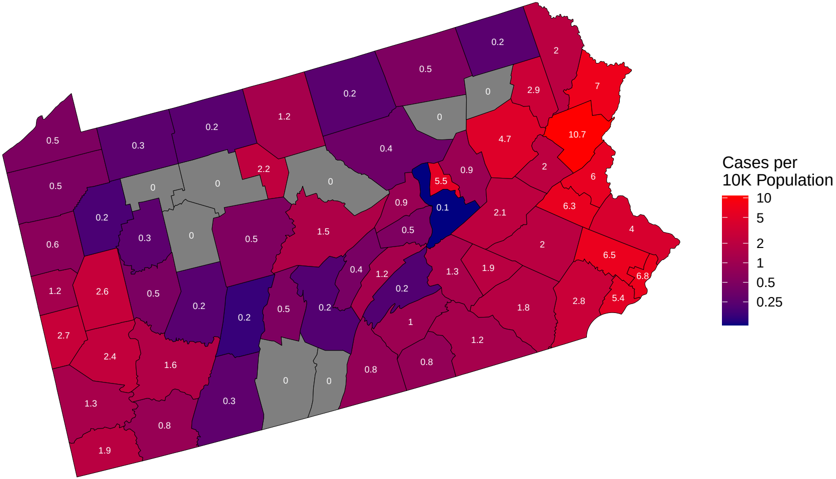
- Wald tests: `waldtest` (in `lmtest`), `wald.test` (in `aod`), etc.
- LR tests: `lrtest` (in `lmtest`), `RLRsim`, many others
- LMs "by-hand" (straightforward...)

Example: Historical COVID-19 in Pennsylvania

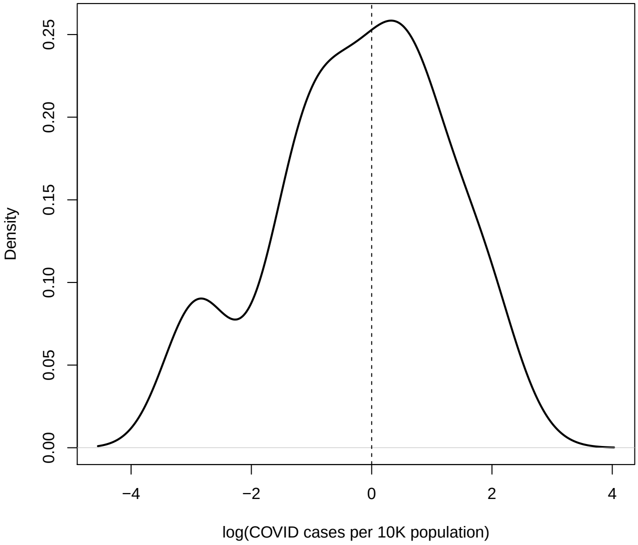
- COVID-19 cases, 67 counties, as of $3/30/2020$
- Source: `https://github.com/nytimes/covid-19-data`
- (Badly) Skewed $\rightarrow$ logged
- We're guessing $\sim N(\mu, \sigma^2)$...

PA COVID−19 Cases (per 10,000 population) by County, through March 30, 2020

Source: https://github.com/nytimes/covid−19−data



PA COVID-19 Cases per 10K Population, 3/30/2020 (logged)



```
> library(readr)
> library(maxLik)
> library(aod)
> library(lmtest)

# Get COVID data:

> COVID<-read_csv("https://raw.githubusercontent.com/PrisonRodeo/
   PLSC503-2024-git/master/Data/COVID-PA.csv")

# log-lik function:

> COVIDll <- function(param) {
+  mu <- param[1]
+  sigma <- param[2]
+  ll <- -0.5*log(sigma^2) - (0.5*((x-mu)^2/sigma^2))
+  ll
}

> x<-log(COVID$CasesPer10K+0.055)
```

```
> hats <- maxLik(COVIDll, start=c(0,1))
> summary(hats)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 5 iterations
Return code 2: successive function values
  within tolerance limit
Log-Likelihood: -56.4
2  free parameters
Estimates:
     Estimate Std. error t value Pr(> t)
[1,]   -0.217      0.172   -1.26    0.21
[2,]    1.407      0.122   11.58  <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
--------------------------------------------
```

```
> COVIDLM<-lm(x~1)
> summary(COVIDLM)

Call:
lm(formula = x ~ 1)

Residuals:
   Min     1Q Median     3Q    Max
-2.684 -0.972  0.163  0.969  2.595

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.217      0.173   -1.25     0.22

Residual standard error: 1.42 on 66 degrees of freedom
```

```
> hats$estimate
[1] -0.217  1.407
```

```
> hats$gradient
[1] 0.00000000422 0.00000223621
```

```
> hats$hessian
      [,1]  [,2]
[1,] -33.8   0.0
[2,]   0.0 -67.7
```

More moving parts...

```
> -(solve(hats$hessian))
       [,1]   [,2]
[1,] 0.0296 0.0000
[2,] 0.0000 0.0148
```

```
> sqrt(-(solve(hats$hessian)))
      [,1]  [,2]
[1,] 0.172 0.000
[2,] 0.000 0.122
```

Restricted model: fix $\mu = 0$:

```
> COVIDllAlt <- function(param) {
+      sigma <- param[1]
+    ll <- -0.5*log(sigma^2) - (0.5*((x-0)^2/sigma^2))
+    ll
+ }
```

```
> hatsF <- maxLik(COVIDll, start=c(0,1))
> hatsR <- maxLik(COVIDllAlt, start=c(1))
```

LR tests

Log-likelihoods:

```
> hatsF$maximum
[1] -56.4

> hatsR$maximum
[1] -57.2
```

Testing:

```
> -2*(hatsR$maximum-hatsF$maximum)
[1] 1.57

> pchisq(-2*(hatsR$maximum-hatsF$maximum),df=1,lower.tail=FALSE)
[1] 0.21
```

```
> library(lmtest) # install as necessary

> lrtest(hatsF,hatsR)
Likelihood ratio test

Model 1: hatsF
Model 2: hatsR
  #Df LogLik Df Chisq Pr(>Chisq)
1   2  -56.4
2   1  -57.2 -1  1.57       0.21


> # Compare to Wald:
>
> wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:1,H0=0)
Wald test:
----------

Chi-squared test:
X2 = 1.6, df = 1, P(> X2) = 0.21
```

Test $\mu = \sigma = 2$:

```
> aod::wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:2,H0=c(2,2),verbose=TRUE)
Wald test:
----------

Coefficients:
[1] -0.22  1.41

Var-cov matrix of the coefficients:
      [,1]  [,2]
[1,] 0.030 0.000
[2,] 0.000 0.015

Test-design matrix:
   [,1] [,2]
L1    1    0
L2    0    1

Positions of tested coefficients in the vector of coefficients: 1, 2

H0:  -0.2166 = 2; 1.4073 = 2

Chi-squared test:
X2 = 190.0, df = 2, P(> X2) = 0.0
```

More Wald tests

Test $\mu = 0$, $\sigma = 2$:

```
> aod::wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:2,H0=c(0,2))
Wald test:
----------

Chi-squared test:
X2 = 25.4, df = 2, P(> X2) = 0.0000031
```

Test $\mu = -0.2$, $\sigma = 1.5$:

```
> aod::wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:2,H0=c(-0.2,1.5))
Wald test:
----------

Chi-squared test:
X2 = 0.59, df = 2, P(> X2) = 0.74
```

Nonsensical Wald Test

Test : $\sigma = 0$:

```
> aod::wald.test(Sigma=vcov(hats),b=coef(hats),Terms=2:2,H0=0)
Wald test:
----------

Chi-squared test:
X2 = 134.0, df = 1, P(> X2) = 0.0
```

Attempt the Same Nonsensical LR Test...

Test : $\sigma = 0$:

```
> hatsDumb <- maxLik(COVIDll, start=c(mu=0,sigma=0),
+                              fixed="sigma")

> summary(hatsDumb)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 0 iterations
Return code 100: Initial value out of range.
--------------------------------------------

> lrtest(hatsF,hatsDumb)
Error in attr(ll, "df") <- sum(activePar(object)) :
  attempt to set an attribute on NULL
```

Binary Response Models

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

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

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

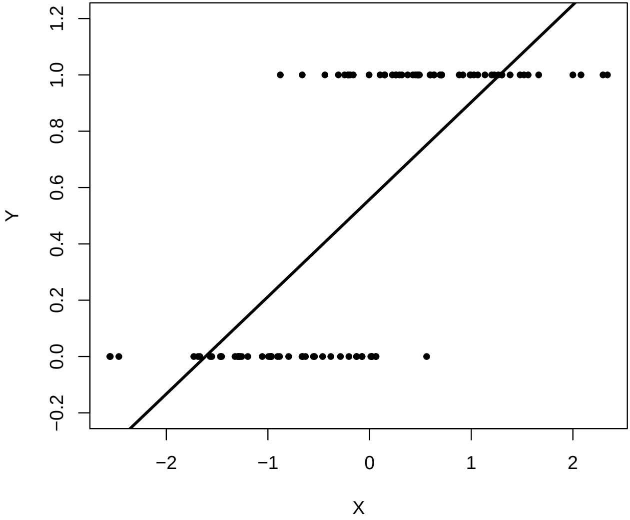
So:

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

or:

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

LPM Illustrated



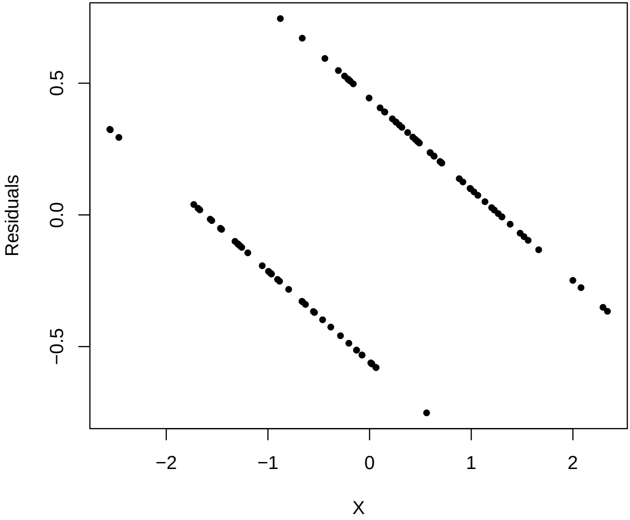
Variance:

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

Residuals:

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

LPM Residuals



Concerns:

- Predictions $\notin [0, 1]$
- Functional form $\rightarrow \frac{\partial \mathrm{E}(Y)}{\partial X} = \beta$ (reasonable?)

When *can* you use an LPM?

- When $\overline{Pr(Y_i = 1)} \approx 0.5$, and
- linearity seems reasonable, and
- you're a lazy economist at a fancy place.

See: Chen, Kaicheng, Robert S. Martin, and Jeffrey M. Wooldridge. 2023. Working paper: Michigan State University.

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:

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

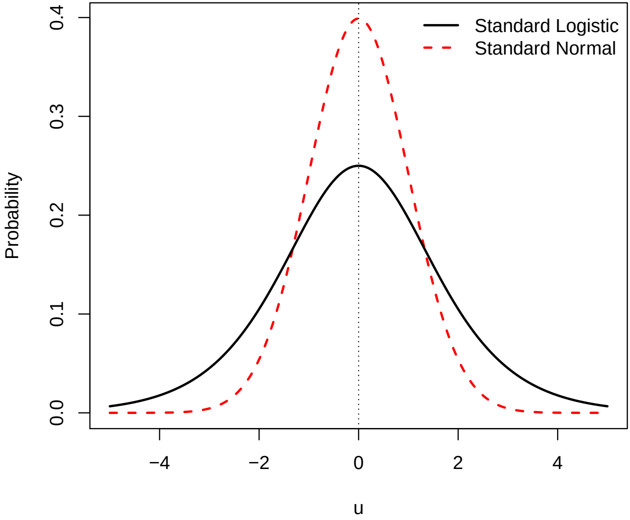
"Standard logistic" PDF:

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

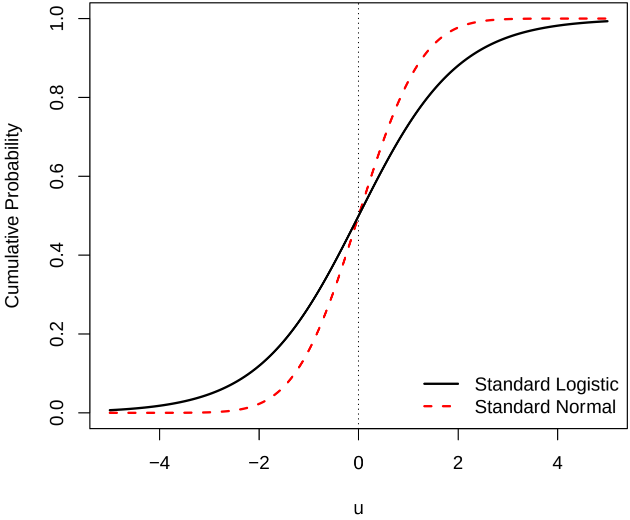
CDF:

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

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 $\rightarrow$ "Logit"

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

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

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

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

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

Digression I: Logit as an Odds Model

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

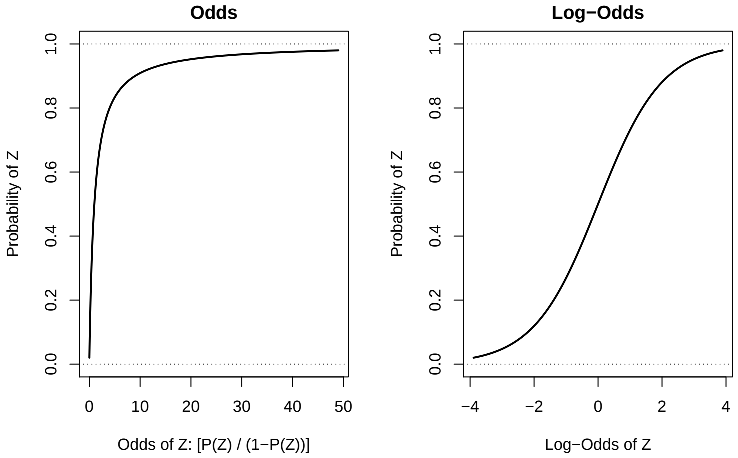
$$\ln[\Omega(Z)] = \ln\left[\frac{\Pr(Z)}{1 - \Pr(Z)}\right]$$

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

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

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

Visualizing Log-Odds



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

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

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

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

So:

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

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:

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

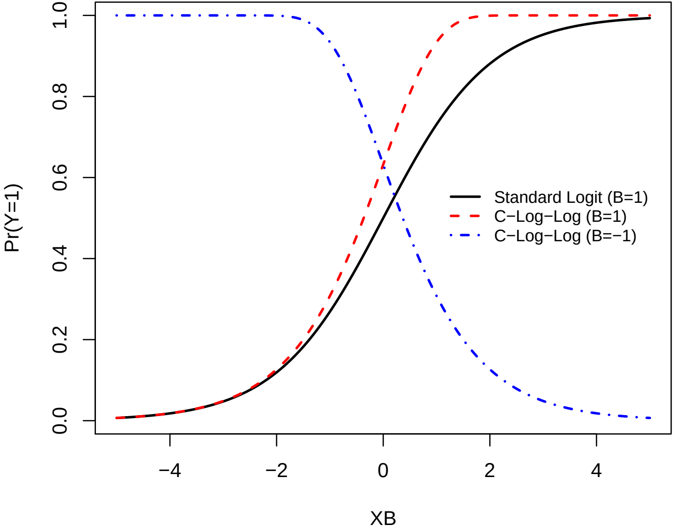
or

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

Likelihood is:

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

Logit and C-log-log CDFs



All require that:

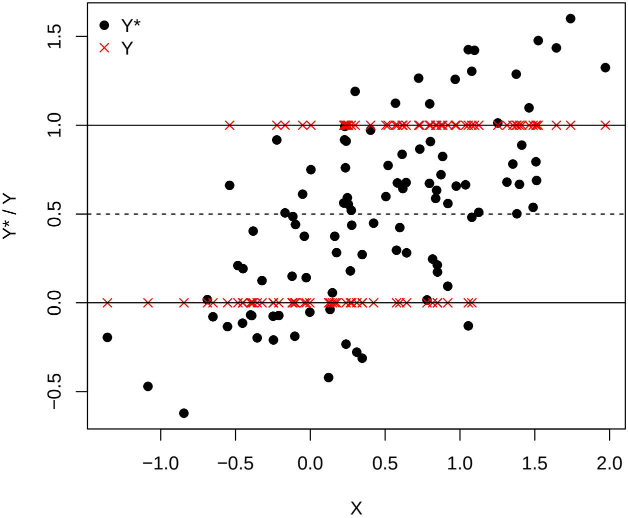
- "Threshold" $= Y^* > 0$
- $\mathrm{E}(u_i|\mathbf{X}, \boldsymbol{\beta}) = 0$
- $\mathrm{Var}(u_i) = \frac{\pi^2}{3}$ or 1.0.

In general:

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

```
> 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



```
> myprobit<-glm(y~x,family=binomial(link="probit"), data=data)
> summary(myprobit)

Call:
glm(formula = y ~ x, family = binomial(link = "probit"), data = data)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.963  -0.777   0.304   0.776   2.165

Coefficients:
            Estimate Std. Error z value    Pr(>|z|)
(Intercept)   -0.519      0.183   -2.83      0.0047 **
x              1.458      0.272    5.35 0.000000087 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 137.989  on 99  degrees of freedom
Residual deviance:  98.979  on 98  degrees of freedom
AIC: 103

Number of Fisher Scoring iterations: 5
```

```
> mylogit<-glm(y~x,family=binomial(link="logit"), data=data)
> summary(mylogit)

Call:
glm(formula = y ~ x, family = binomial(link = "logit"), data = data)

Deviance Residuals:
   Min       1Q  Median       3Q      Max
-1.959  -0.769   0.340   0.766   2.135

Coefficients:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept)   -0.861      0.318   -2.71    0.0067 **
x              2.428      0.500    4.86 0.0000012 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 137.99  on 99  degrees of freedom
Residual deviance:  99.38  on 98  degrees of freedom
AIC: 103.4

Number of Fisher Scoring iterations: 4
```

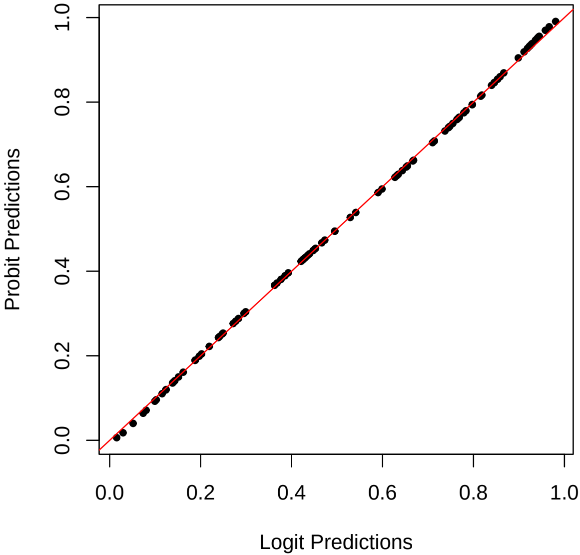
Toy Example: C-Log-Log

```
> mycloglog<-glm(y~x,family=binomial(link="cloglog"), data=data)
> summary(mycloglog)

Call:
glm(formula = y ~ x, family = binomial(link = "cloglog"), data = data)

Deviance Residuals:
   Min       1Q  Median       3Q      Max
-2.001  -0.783   0.187   0.812   1.995

Coefficients:
            Estimate Std. Error z value   Pr(>|z|)
(Intercept)   -1.048      0.250   -4.20 0.00002710 ***
x              1.613      0.309    5.22 0.00000018 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 137.989  on 99  degrees of freedom
Residual deviance:  99.044  on 98  degrees of freedom
AIC: 103

Number of Fisher Scoring iterations: 6
```

Note:

- zs, Ps, $\ln L$s (via "residual deviance") 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