

Chapter 4. Logistic Regression

1. Introduction to Logistic Regression

- Assume a single explanatory variable X , quantitative, and a binary response variable
- $\pi(x)$ is the probability of success at value x
- the logistic linear regression model is defined as $\text{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$.
- This implies that

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}.$$

a. Interpretation of β

- The logit increases by β for every 1 unit increase in x
- it determines the ratio of increase or decrease of the S-shaped curve for $\pi(x)$
- The median effective level (EL_{50}) is the level at which each outcome has a 50%
- It is obtained by solving $\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$ where $\pi(x) = 0.5$.
- this gives $x = -\alpha/\beta$

b. Example: Horseshoe Crabs data (p. 101)

- Let $\pi(x)$ be the probability that a female horseshoe crab of width x has a satellite,
- the simplest logistic model is $\text{logit}(\pi(x)) = \alpha + \beta x$
- The fit can be done easily in R as

```

> data<-read.table("crab.txt",header=T)
> attach(data)
> weight<-weight/1000; color=color-1;
> names(data)
[1] "color" "spine" "width" "satell" "weight"
> satell<-ifelse(satell>0,1,0)
> fit<-glm(satell~width,family=binomial(logit))
> summary(fit)

Call:
glm(formula = satell ~ width, family = binomial(logit))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0281  -1.0458   0.5480   0.9066   1.6941

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.3508     2.6287  -4.698 2.62e-06 ***
width         0.4972     0.1017   4.887 1.02e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 194.45  on 171  degrees of freedom
AIC: 198.45

Number of Fisher Scoring iterations: 4

> coef(fit)
(Intercept)      width
-12.3508177    0.4972306

```

– From R output, our logistic model is

- The estimate probability of a satellite is
- The median effective level is
- The interpretation of the logistic regression uses the odds and the odds ratio.
- The odds of response (1) (odds of success) are $\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = e^\alpha (e^\beta)^x$
- so, for every 1 unit increase in x , the odds of response 1 are multiplied by e^β
- For our example,

2. Inference for Logistic Regression

a. Confidence Interval

- A large sample Wald CI for β is $\hat{\beta} \pm z_{\alpha/2} SE(\hat{\beta})$
- Exponentiating the endpoints yields an interval for e^β
- In our example,

– The Wald CI can also be obtained from R as follows:

```
> ### Wald CI
> beta<-coef(fit)[2];beta
      width
0.4972306
> vb<-vcov(fit)
> seb<-sqrt(vb[2,2]);seb
[1] 0.1017355
> logwald.ci<-beta+c(-1,1)*qnorm(1-alpha/2)*seb;wald.ci
[1] 0.2978326 0.6966286
> wald.ci<-exp(logwald.ci);waldci
[1] 1.346936 2.006975
```

– The likelihood based CI for β is

– In the same way, the likelihood based CI for β is obtained from R by doing:

```
> ### Likelihood ratio CI
> library(MASS)
> lr<-confint(fit)
Waiting for profiling to be done...
> lr.ci<-lr[2,];lr.ci
      2.5 %      97.5 %
0.3083806 0.7090167
> elr.ci<-exp(lr.ci);elr.ci
      2.5 %      97.5 %
1.361219 2.031992
```

Interpretation of CI (Wald or Likelihood Ratio)

- For example, for Wald

b. Significance Testing

- For this example, want to test $H_0 : \beta = 0$ vs $H_a : \beta \neq 0$
- LRT is more powerful than Wald test.

- Wald test is obtained from R by

```
> ### Wald test
```

```

> waldt<-beta/seb;waldt
      width
4.887482
> pval=2*pnorm(waldt,lower.tail=F);pval
      width
1.02134e-06

```

and the likelihood ratio test (LRT) test is performed in R by doing

```

> ### Likelihood Test
> nul<-fit$null.deviance;nul
[1] 225.7585
> res<-fit$deviance;res
[1] 194.4527
> lrt<-nul-res;lrt
[1] 31.30586

```

– To make a prediction for a new value of x , use the command `predict()` from R. Here is an example:

```

> ### Prediction of a new data
> new <- data.frame(width = 26.5)
> prednew<-predict(object =fit, newdata = new, type = "response", se = TRUE)
> prednew
$fit
      1
0.6954646

$se.fit
      1
0.03996454

$residual.scale
[1] 1
> alpha<- .05
> lower<-prednew$fit-qnorm(1-alpha/2)*prednew$se;lower
      1
0.6171356
> upper<-prednew$fit+qnorm(1-alpha/2)*prednew$se;upper
      1
0.7737937
> ci<-prednew$fit+c(-1,1)*qnorm(1-alpha/2)*prednew$se;ci
[1] 0.6171356 0.7737937

```

c. Model Checking

- Need to be able to check whether our model fits.
- If not, a more complicated model may be needed.
- For instance, we may need to include other term like a quadratic term (X^2).
- We may need other explanatory variable in addition to X .
- We use goodness-of-fit statistic to check model fit.

- Pearson χ^2
- Likelihood Ratio G^2

d. Residuals

- Let y_i be the number of successes for the n_i observation at the i th setting of the explanatory variable.
- Let $\hat{\pi}_i$ be the predicted probability of success using the fitted model.
- If the model is good, y_i should be close to $n_i\hat{\pi}_i$.
- The i th (Pearson) residual is

$$e_i = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1 - \hat{\pi}_i)}} \sim N(0, 1).$$

- When the model doesn't fit well, inspection of the residuals may help identify the problem or suggest alternative models to fit.
- Residuals behave similarly to Z scores.
- Lack of fit will be values very far from 0 ($\pm 2, 3$).

3. Logit Models for Qualitative Predictors

- Our explanatory variables have been continuous variables so far.
- When X is continuous, we look at models like

$$\log\left(\frac{\pi}{1 - \pi}\right) = \alpha + \beta x.$$

- When X is categorical, having levels $1, 2, \dots, I$:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \beta_i,$$

where X is at level i .

- Similar to one-way ANOVA model.
- Suppose X has I levels (this is a $I \times 2$ table),

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha + \beta_i, \quad \text{for } i = 1, 2, \dots, I.$$

Put restriction on β_i : (1) $\beta_1 = 0$, (2) $\beta_I = 0$, (3) $\sum_{i=1}^I \beta_i = 0$.

Assume restriction (2), then $\alpha = \log \left(\frac{\pi_I}{1 - \pi_I} \right)$. So

$$\begin{aligned} \beta_1 &= \log \left(\frac{\pi_1}{1 - \pi_1} \right) - \log \left(\frac{\pi_I}{1 - \pi_I} \right) \\ &= \log \left(\frac{\pi_1(1 - \pi_I)}{(1 - \pi_1)\pi_I} \right) = \log \left(\frac{\Omega_1}{\Omega_I} \right) \\ &\vdots \\ \beta_{I-1} &= \log \left(\frac{\Omega_{I-1}}{\Omega_I} \right) \end{aligned}$$

- An alternative approach uses dummy variables

$$X_i = \begin{cases} 1 & \text{if observation are at } i\text{th level of } X, \\ 0 & \text{otherwise.} \end{cases}$$

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha + \beta_1 x_1 + \dots + \beta_{I-1} x_{I-1}.$$

- As many parameters as categories for $X \Rightarrow$ model will fit the data perfectly. This model is called the saturated model.

Remark

$$\begin{aligned} \beta_i &= \log \left(\frac{\pi_i}{1 - \pi_i} \right) - \log \left(\frac{\pi_I}{1 - \pi_I} \right) \\ &= \log \left(\frac{\pi_i(1 - \pi_I)}{(1 - \pi_i)\pi_I} \right) = \log \left(\frac{\Omega_i}{\Omega_I} \right) \end{aligned}$$

To find

$$\log \left(\frac{\Omega_1}{\Omega_2} \right) = \beta_1 - \beta_2.$$

To find odds ratio of level 3 and level 5, we have

$$\frac{\Omega_3}{\Omega_5} = e^{\beta_3 - \beta_5}$$

– Now, assume that we have

$$\log \left(\frac{\pi}{1 - \pi} \right) = \alpha.$$

This means the levels of X has no effect on Y . We have the independent model.

– We can also have more than one explanatory variables.

– For example, Suppose X has levels $1, \dots, I$ and Z has levels $1, \dots, K$. More variables implies more models to consider.

Some potential models

$$\begin{aligned} \log \left(\frac{\pi(X = i, Z = k)}{(1 - \pi(X = i, Z = k))} \right) &= \alpha \\ &= \alpha + \beta_i^X + \beta_k^Z \\ &= \alpha + \beta_i^X + \beta_k^Z + \beta_{ik}^{XZ}. \end{aligned}$$

4. Multiple Logistic Regression

– As with ordinary regression, logistic regression generalizes to more than one explanatory variable.

Continuous

For example, with k explanatory variables X_1, \dots, X_k , we may fit the model

$$\log \left(\frac{\pi}{1 - \pi} \right) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k.$$

We need to first see the interpretation of the β 's:

e^{β_i} gives the multiplicative factor/effect on the odds of a unit increase in the i th explanatory variable (X_i) when the other variables are held fixed.

AZT and AIDS Example (p111)

Assume X and Z each takes values 0 and 1 to represent the 2 levels of each explanatory variable.

$$\text{logit}_1(P(Y = 1)) = \alpha + \beta_1 x + \beta_2 z.$$

The fit of this example using dummy variables is obtained from R.

```
> ### Example 1: AZT on AIDS
> y<-cbind(c(14,32,11,12),c(93,81,52,43))
> x<-c(1,0,1,0)
> z<-c(1,1,0,0)
> fitd<-glm(y~x+z,family=binomial)
> summary(fitd)
```

Call:

```
glm(formula = y ~ x + z, family = binomial)
```

Deviance Residuals:

1	2	3	4
-0.5547	0.4253	0.7035	-0.6326

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.07357	0.26294	-4.083	4.45e-05	***
x	-0.71946	0.27898	-2.579	0.00991	**
z	0.05548	0.28861	0.192	0.84755	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8.3499 on 3 degrees of freedom
Residual deviance: 1.3835 on 1 degrees of freedom
AIC: 24.86

Number of Fisher Scoring iterations: 4

– An alternative for this fit is to use ANOVA. In R, this is done easily by:

```
> ### Example 1 using ANOVA
> table.4.3<-expand.grid(azt=factor(c("yes","no"),levels=c("no","yes")),
+ race=factor(c("white","black"),levels=c("black","white")))
> table.4.3<-data.frame(table.4.3,yes=c(14,32,11,12),no=c(93,81,52,43))
> options(contrasts=c("contr.treatment","contr.poly"))
> fita<-glm(cbind(yes,no)~azt+race,family=binomial,data=table.4.3)
> summary(fita)
```

```
Call:
glm(formula = cbind(yes, no) ~ azt + race, family = binomial,
     data = table.4.3)
```

Deviance Residuals:

1	2	3	4
-0.5547	0.4253	0.7035	-0.6326

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.07357	0.26294	-4.083	4.45e-05	***
aztyes	-0.71946	0.27898	-2.579	0.00991	**
racewhite	0.05548	0.28861	0.192	0.84755	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8.3499 on 3 degrees of freedom
 Residual deviance: 1.3835 on 1 degrees of freedom
 AIC: 24.86

Number of Fisher Scoring iterations: 4

– Note that the deviance of the fit can be obtained in R by

```
> fita$deviance
[1] 1.38353
> sum(residuals(fita,type="pearson")^2)
[1] 1.391026
> nulc<-fita$null.deviance;nulc
[1] 8.349946
> resc<-fita$deviance;resc
[1] 1.38353
```

Conditional independence between X and Y controlling for Z if $\beta_1 = 0$:

$$\text{logit}_2(P(Y = 1)) = \alpha + \beta_2 z.$$

To test for conditional independence of AZT treatment and the development of AIDS symptoms, controlling for race. i.e.

$H_0 : \beta_1 = 0$ v.s. $H_a : \beta_1 \neq 0$. We want to compare model 2 (reduced model) to model 1 (full model).

– The reduced model is obtained in R as

```

> ### Fitting the reduced model
> fit1<-glm(cbind(yes,no)~race,family=binomial,data=table.4.3)
> summary(fit1)

Call:
glm(formula = cbind(yes, no) ~ race, family = binomial, data = table.4.3)

Deviance Residuals:
    1      2      3      4 
-2.1028  1.8649 -0.4126  0.4294 

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.41838    0.23239  -6.103 1.04e-09 ***
racewhite    0.08797    0.28547   0.308  0.758
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.3499  on 3  degrees of freedom
Residual deviance: 8.2544  on 2  degrees of freedom
AIC: 29.731

Number of Fisher Scoring iterations: 4

```

The LR statistic

$LRT_{2/1}$ = residual deviance (2) – residual deviance (1) = 8.2544 – 1.3825 = 6.87
and the $df_{2/1} = 2 - 1 = 1$. There is evidence of association. In R, we can do

```

> ### Testing the impact of AZT
> nulc1<-fit1$null.deviance;nulc1
[1] 8.349946
> resc1<-fit1$deviance;resc1
[1] 8.254436
> lrtm<-resc1-nulc1;lrtm
[1] 6.870906

```

– Note that in this specific case, we can also look at the Wald statistic P-value = 0.0099 .

Horseshoe Crabs with Color and Width Example (p116)

- Use the horseshoe crabs data with color and width as predictors.
- Color has 5 categories (light, medium light, medium, medium dark, dark).
- Sample contains no light crabs, so we used the other 4 categories.
- Treat color in the regression model using 3 dummy variables. (0, 0, 0) gives the 4th one.

$$\text{logit}(P(Y = 1)) = \alpha + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 x.$$

- Assuming a model with no interaction, the fit is done using R.

```
> ### Example 2: Horseshoes crabs with colors and width predictors
> colfac<-factor(data$color,levels=c("5","4","3","2"),labels=c("dark","med-dark",
+ , "med","med-light"))
> satell<-ifelse(satell>0,1,0)
> crabfit<-glm(satell~colfac+width,family=binomial)
> summary(crabfit)
```

Call:

```
glm(formula = satell ~ colfac + width, family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1124	-0.9848	0.5243	0.8513	2.1413

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-12.7151	2.7617	-4.604	4.14e-06	***
colfacmed-dark	1.1061	0.5921	1.868	0.0617	.
colfacmed	1.4023	0.5484	2.557	0.0106	*
colfacmed-light	1.3299	0.8525	1.560	0.1188	
width	0.4680	0.1055	4.434	9.26e-06	***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 225.76 on 172 degrees of freedom
Residual deviance: 187.46 on 168 degrees of freedom
AIC: 197.46

Number of Fisher Scoring iterations: 4

– The model has a different intercept parameter for crabs of different colors.

For example, the logit model for dark crab is

For example, the logit model for medium crab is

– However, the slope on width is always the same. So Regardless of the color, a 1 cm increase in width has a multiplicative effect of $e^{.468} = 1.60$ on the odds of having a satellite.

– The exponentiated difference between two-color parameter estimates is an

odds ratio comparing those colors.

– So, at any given width, the estimated odds that a medium crab has a satellite are $e^{1.4023-1.1061} = 1.34$ times the estimated odds for a medium dark crab.

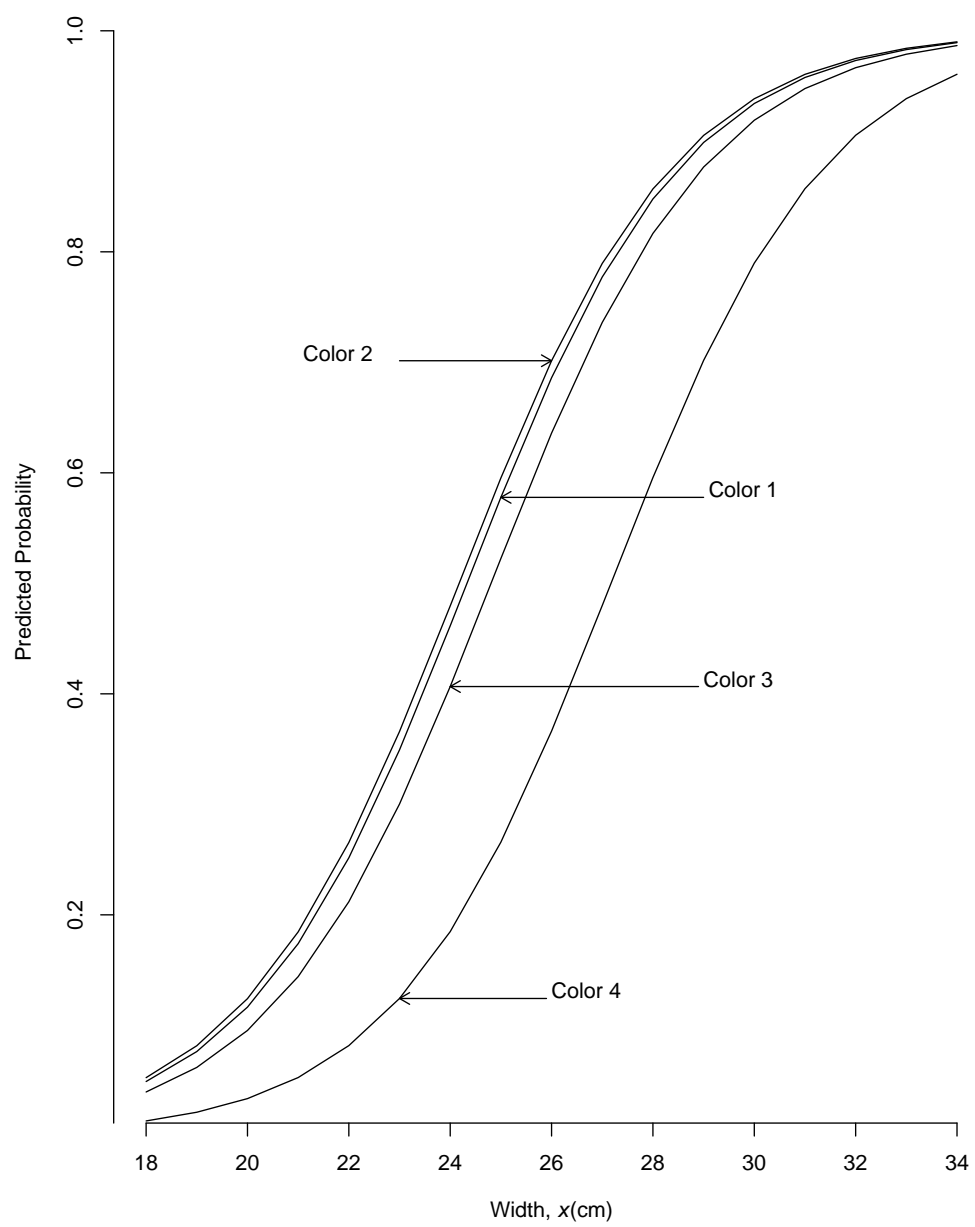
– The confidence intervals for the parameters are obtained from R

```
> library(MASS)
> confint(crabfit)
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)   -18.45674069 -7.5788795
colfacmed-dark -0.02792233  2.3138635
colfacmed      0.35269965  2.5260703
colfacmed-light -0.27377584  3.1356611
width          0.27128167  0.6870436
```

– For the graph, predict the probability at widths from 18 to 34cm each of the colors.

```
> ### Plot of the predicted values
> res1<-predict(crabfit, type="response", newdata=data.frame(width=seq(18,34,1),
+ colfac="med-light"))
> res2<-predict(crabfit, type="response", newdata=data.frame(width=seq(18,34,1),
+ colfac="med"))
> res3<-predict(crabfit, type="response", newdata=data.frame(width=seq(18,34,1),
+ colfac="med-dark"))
> res4<-predict(crabfit, type="response", newdata=data.frame(width=seq(18,34,1),
> plot(seq(18,34,1),res1,type="l",bty="L",ylab="Predicted Probability", axes=F,
+ xlab=expression(paste("Width, ", italic(x), "(cm)")))
> axis(2, at=seq(0,1,.2))
> axis(1, at=seq(18,34,2))
> lines(seq(18,34,1),res2) # add colors 2-4
> lines(seq(18,34,1),res3)
> lines(seq(18,34,1),res4)
> # add arrows and text
> arrows(x0=29, res1[25-17],x1=25, y1=res1[25-17], length=.09)
> text(x=29.1, y=res1[25-17], "Color 1", adj=c(0,0))
> arrows(x0=23, res2[26-17],x1=26, y1=res2[26-17], length=.09)
> text(x=21.1, y=res2[26-17], "Color 2", adj=c(0,0))
> arrows(x0=28.9, res3[24-17],x1=24, y1=res3[24-17], length=.09)
> text(x=29, y=res3[24-17], "Color 3", adj=c(0,0))
> arrows(x0=25.9, res4[23-17],x1=23, y1=res4[23-17], length=.09)
```

```
> text(x=26, y=res4[23-17], "Color 4", adj=c(0,0))
```



5. Model Checking

a. Definition

- There are typically two test statistics for goodness of fit.

$$X^2 = \sum \frac{(\text{observed} - \text{fitted})^2}{\text{fitted}}$$

$$G^2 = 2 \sum (\log \text{observed} / \text{fitted}).$$

- There is a lack of fit when X^2 or G^2 is too big.

b. Likelihood Ratio Model Comparison Tests

- The log likelihood-ratio statistic equals

$$l = -2(L_0 - L_1)$$

- L_1 = the maximized log of the likelihood function from a complex model, say M_1 .
- L_0 = the maximized log of the likelihood function from a simpler (nested) model, say M_0 .
- The goodness of model fit statistic G^2 is a special case of the likelihood ratio test statistic where
 - $M_0 = M$, the model we're testing.
 - $M_1 = M_S$, the most complex model possible or the “saturated” model.
- For Generalized Linear Models such as Poisson and logistic regression, G^2 is equal to “deviance” of the model.
 - L_S = maximized log of the likelihood function from the saturated model M_S .
 - L_0 = maximized log of the likelihood function from the simpler model M_0 .
 - L_1 = maximized log of the likelihood function from the complex model M_1 .

- We want to compare the fit of the model M_0 and M_1 .

$$\text{deviance for } M_0 = G^2(M_0) = 2(L_0 - L_S)$$

$$\text{deviance for } M_1 = G^2(M_1) = 2(L_1 - L_S)$$

- The likelihood ratio statistic is

$$G^2(M_0|M_1) = G^2(M_0) - G^2(M_1)$$

- and

$$df = df_0 - df_1.$$

- Assuming that M_1 holds, this statistic tests
 - Whether the lack of fit of M_0 is significantly larger than that of M_1 .
 - Whether the parameters in M_1 that are not in M_0 equal zero.
- In R, Null Deviance = $2(L_N - L_S)$ where L_N is the “Null Model” and
- Residual Deviance = $2(L_P - L_S)$ where L_P is the “Proposed Model”
- The Null Model assumes one parameter for all of the data points, which means you only estimate 1 parameter, the intercept.
- The null deviance shows how well the response is predicted by the model with nothing but an intercept.
- The Proposed Model assumes you can explain your data points with p parameters + an intercept term, so you have $p + 1$ parameters.
- The residual deviance shows how well the response is predicted by the model when the predictors are included.

Example: Horseshoe crab on width

- Using the Horseshoe crab on width example, we can test $H_0 : \beta = 0$ vs $H_a : \beta \neq 0$ by doing the following.
- M_0 = Model with only an intercept, i.e. $\text{logit}(\pi(x)) = \alpha$
- $G^2(M_0) = 225.76$ with $df_0 = 172$.
- M_1 = Model with an intercept and width, i.e. $\text{logit}(\pi(x)) = \alpha + \beta x$
- $G^2(M_1) = 194.45$ with $df_1 = 171$.
- $G^2(M_0|M_1) = 225.76 - 194.45 = 31.31$ with $df = 172 - 171 = 1$