

Logistic Regression

Multiple Predictors

Demetris Athienitis



Section 1

1 Model

2 Inference

Rather than just one predictor, multiple predictors x_1, x_2, \dots, x_k yield

$$\text{logit} [\pi(x)] = \alpha + \sum_{i=1}^k \beta_i x_i \quad \Leftrightarrow \quad \pi(x) = \frac{e^{\alpha + \sum_{i=1}^k \beta_i x_i}}{1 + e^{\alpha + \sum_{i=1}^k \beta_i x_i}}$$

Some predictors may be:

- quantitative, including polynomial and interaction terms
- qualitative (or binary)

Example (Horseshoe crab continued)

Add the color (4 levels) to the model by creating:

$$c_1 = \begin{cases} 1 & \text{medium light} \\ 0 & \text{o/w} \end{cases} \quad c_2 = \begin{cases} 1 & \text{medium} \\ 0 & \text{o/w} \end{cases} \quad c_3 = \begin{cases} 1 & \text{medium dark} \\ 0 & \text{o/w} \end{cases}$$

With $c_1 = c_2 = c_3 = 0$ indicating the remaining group, dark (i.e. base group).

$$\text{logit}[\pi(x)] = \alpha + \beta_1 x + \beta_2 c_1 + \beta_3 c_2 + \beta_4 c_3$$

Color	logit $[\pi(x)]$
medium light	$(\alpha + \beta_2) + \beta_1 x$
medium	$(\alpha + \beta_3) + \beta_1 x$
medium dark	$(\alpha + \beta_4) + \beta_1 x$
dark	$\alpha + \beta_1 x$

Example (continued)

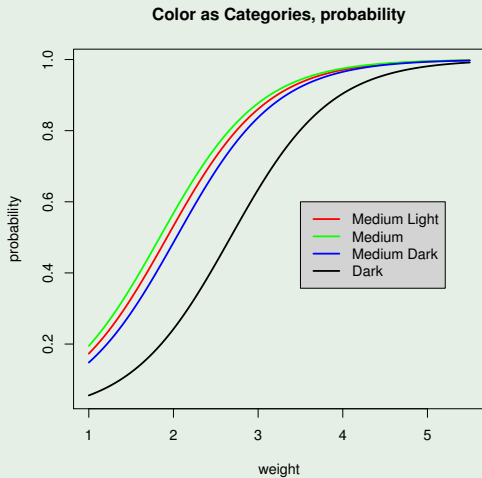
```
> crabs$color=factor(crabs$color,labels=c("ML","M","MD","D"))
> crabs$color=relevel(crabs$color,4)
> fit2=glm(y ~ weight + color, family=binomial,data=crabs)
> summary(fit2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.5266	1.0038	-4.510	6.50e-06	***
weight	1.6928	0.3888	4.354	1.34e-05	***
colorML	1.2694	0.8488	1.495	0.13479	
colorM	1.4143	0.5449	2.595	0.00945	**
colorMD	1.0833	0.5884	1.841	0.06561	.

Null deviance: 225.76 on 172 degrees of freedom
Residual deviance: 188.54 on 168 degrees of freedom
AIC: 198.54

Example (continued)



Section 2

1 Model

2 Inference

We already saw that for single parameter $H_0 : \beta = \beta_0$ we test

- using Wald test

$$\frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}} \underset{H_0}{\rightsquigarrow} N(0, 1)$$

- using the likelihood ratio test, G^2 , the difference in deviances between full and reduced (under the null) models

$$G^2 = D(y; \hat{\mu}_0) - D(y; \hat{\mu}_1)$$

Multiple parameters- LRT

In the Horseshoe crab, used parameters for fitting color (as a qualitative).
To test color, as a whole

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{vs} \quad H_1 : \text{at least one } \beta \neq 0$$

Null yields the *reduced model* (under null), μ_0

$$g(\mu_0) = \alpha + \beta_1 x$$

$$g(\mu_1) = \alpha + \beta_1 x + \beta_2 c_1 + \beta_3 c_2 + \beta_4 c_3$$

Fitting both models and obtaining *residual deviances* of each, create LRT

$$G^2 = D(y; \hat{\mu}_0) - D(y; \hat{\mu}_1) \xrightarrow[H_0]{d} \chi^2_{df}$$

where df is the dimension reduction of parameter vector, 3 in this case.

Example (Horseshoe crab continued)

$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$

- Residual deviance of null model 195.74 (when we fitted just weight)
- Residual deviance of full model 188.54

$$\begin{aligned} G^2 &= D(y; \hat{\mu}_0) - D(y; \hat{\mu}_1) \\ &= 195.74 - 188.54 = 7.2 \end{aligned}$$

with p-value of $1 - \text{pchisq}(7.2, 3) \approx 0.07$, might let us conclude that color is not significant.

However, looking at the individual test statistic values as well as the figure of probability curves we see that there is a more to this problem that we will be addressing.

Example (continued)

There is also an automatic way of performing this test once since we fitted the reduced model `fit` and the full model `fit2`

```
> anova(fit2,fit,test="LRT")
```

Analysis of Deviance Table

Model 1: $y \sim \text{weight} + \text{color}$

Model 2: $y \sim \text{weight}$

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	168	188.54			
2	171	195.74	-3	-7.1949	0.06594 .

Example (Horseshoe crab continued)

$$\text{logit}[\pi(x)] = \alpha + \beta_1 x + \beta_2 c_1 + \beta_3 c_2 + \beta_4 c_3$$

What about testing $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, i.e. are all predictors not significant. Recall,

Null deviance: 225.76 on 172 degrees of freedom

Residual deviance: 188.54 on 168 degrees of freedom

$$\begin{aligned} G^2 &= D(y; \hat{\mu}_0) - D(y; \hat{\mu}_1) \\ &= 225.76 - 188.54 = 37.22 \end{aligned}$$

with p-value $1 - \text{pchisq}(37.22, 4) = 1.622771\text{e-}07$

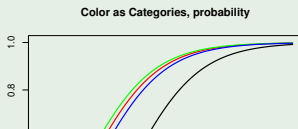
Example (Horseshoe crab continued)

- Testing color via LRT yielded a p-value of 0.07, so is color significant?
- Some coefficients corresponding to color that compare each color level vs color base level were significant

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.5266	1.0038	-4.510	6.50e-06	***
weight	1.6928	0.3888	4.354	1.34e-05	***
colorML	1.2694	0.8488	1.495	0.13479	
colorM	1.4143	0.5449	2.595	0.00945	**
colorMD	1.0833	0.5884	1.841	0.06561	.

- From the plot we notice that there may be in fact be only two groups: dark and not dark



Do **NOT** put too much emphasis on any one tool.

Remember the (somewhat funny) Oral-B commercials? 4 out of 5 dentists recommend Oral-B.

Example (continued)

Create a new binary variable, dark - yes/no.

```
> dark=ifelse(unclass(color)==4,1,0)
> fit2.2=glm(y ~ weight + dark, family=binomial,data=crabs)
> summary(fit2.2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.3134	0.8984	-3.688	0.000226	***
weight	1.7292	0.3825	4.520	6.18e-06	***
dark	-1.2954	0.5222	-2.481	0.013110	*

Null deviance: 225.76 on 172 degrees of freedom
Residual deviance: 189.17 on 170 degrees of freedom
AIC: 195.17

Example (continued)

Testing dark vs non dark via $H_0 : \beta_2 = 0$ for this model

- Via Wald test, p-value = 0.013110
- Via LRT, $G^2 = 195.74 - 189.17$ with 1 degree of freedom yields, p-value = 0.01039651

Why did the p-value drop from 0.07 to about 0.01?

Because we tested using a method that uses less degrees of freedom (1 instead of 3) and hence has more power in detecting significance.

We learned

- Model with multiple predictors
- Reviewed individual tests
- Simultaneous tests on multiple parameter coefficients