

## STATISTICS 475: Applied Discrete Data Analysis

## Inference for Logistic Regression Parameters

(B&amp;L Section 2.2.2)

### 1 Problem to be solved

- We have formulated the logistic regression model for binary data as  $Y \sim \text{Bernoulli}(\pi)$ , where

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

- This leads to

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}$$

- We use ML to estimate parameters, leading to estimates  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ .
- We now need to use these parameter estimates to do inference on the corresponding parameters.

### 2 Individual Bernoulli Trials or Binomial?

First, a detail that was left out of the last lecture:

- In the placekick example, we had 1425 individual attempts that we modeled as independent Bernoulli trials
- In our analysis using `distance` as the explanatory variable, there were numerous placekick attempts at most distances.
- We fit our model on the 1425 attempts, but what if the data had been given to us already summarized by `distance`?

- That is, what if we had data as  $W_i \sim \text{Bin}(n_i, \pi_i)$ , where  $n_i$  is the number of trials at distance  $i$ ,  $\pi_i$  is the probability of success at distance  $i$ , and  $W_i$  is the RV for the count of successes at distance  $i$ ?
- The data would be pairs,  $(n_i, w_i)$ .
- How would we fit this model? *i.e. variables are weighted, each one represents a group of observations with some characteristic.*
- The answer is that it doesn't matter...mostly
  - You need to add a `weights=<n_i>` argument into the `glm()` function, where "n\_i" is whatever variable name contains the number of trials.
    - \* See the example at the very end of B&L Section 2.2.1, p. 74–76 for computational details
    - \* See the script for this lecture for some of the code
  - The resulting parameter estimates and standard errors are the same either way.
  - All inferences are the same either way
  - *Certain model-fit statistics (which we will learn about in a few minutes and again in Chapter 5) will be changed.*
    - \* We will worry about that when we get to it.
- So for all logistic regression models, you can fit the model to whatever format the data are in—binary or aggregated by explanatory variables—and not worry about having to rearrange your data every time you want to fit a model.
- There is a special version of this aggregation where data have been reduced to a single set of binomial trials for each unique value of explanatory variables (or combination of values when there is more than one  $x_j$ ).
  - This is called EXPLANATORY VARIABLE PATTERN (EVP) FORM.

Now on to inferences...

### 3 Inference on Regression Parameters

#### 3.1 Hypothesis tests for a single parameter

- Testing parameters is done in logistic regression for the same reasons it is done in linear regression
  - You may have a particular value for a parameter that you want to test,  $H_0 : \beta_j = c$
  - Most often, we are interested in  $H_0 : \beta_j = 0$  for any of the parameters  $j = 0, 1, \dots, p$ 
    - \* Represents the case there the variable can be dropped from the model.

- \* Not usually done for the intercept (Exercise: What does it mean when  $\beta_0 = 0$ ?)
- The technology is the same as what we have seen so far.
  - However, score-based inference is not usually done due to mathematical/computational difficulty.
  - Therefore, we focus on Wald and LR methods
- **Wald test of  $H_0 : \beta_j = c$** 
  - We have the parameter estimate  $\hat{\beta}_j$  and we get an estimate of its standard error directly from the likelihood computations,  $\sqrt{\widehat{Var}(\hat{\beta}_j)}$ .
  - Just put them together and run the test!
  - Usual properties associated with a Wald test
    - \* Often does not maintain type-1 error rate well unless sample size is large relative to the number of parameters being estimated
      - Also need enough successes and failures, but not sure how many and where among the explanatory variables they must occur.
    - \* No fixed recommendation for “large”
      - $n = 1425, p = 1$  in placekick example is pretty large.
      - $n = 30, p = 5$ , say, is not.
      - Otherwise...“I don’t know...fly casual.”
- **LR test of  $H_0 : \beta_j = c$** 
  - This is more complicated than before, because of the definition of the LR test statistic:  $-2 \log(\Lambda)$ , where
 
$$\Lambda = \frac{\text{Maximum of likelihood function under } H_0}{\text{Overall maximum of likelihood function}}$$
    - \* The denominator of this is easy: it is just the value of the likelihood function at the  $p + 1$  MLEs  $\hat{\beta}_0, \dots, \hat{\beta}_p$ .
    - \* Numerator is tricky: In any logistic regression, there are at least two parameters.
    - \* We use a method called “profiling” described in the appendix at the end of the lecture to get a value
  - The test statistic  $-2 \log(\Lambda)$  is compared to  $\chi^2_\nu$ , where  $\nu = \frac{p+1}{1 \text{ for one parameter}}$  (Why?)
  - \* As usual, only 2-sided tests are available directly, but a 1-sided p-value can be computed using techniques previously described.

### 3.2 Confidence interval for a single parameter

- **Wald** confidence interval is easy,

$$\hat{\beta}_j \pm Z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_j)},$$

but does not maintain nominal confidence level well

- Inverted Wald test, so same properties.

- **LR** confidence interval is an inverted LR test

- Requires finding all values, say  $b$ , for which the LR test would not reject  $H_0 : \beta_j = b$ .
- Repeat the “profiling” process described in the Appendix to create likelihood profiles at different values of  $b$
- Find the endpoints of the CI using an iterative numerical algorithm that cleverly chooses values of  $b$  at which to evaluate the profile
- Result is sometimes referred to as a **PROFILE LR CONFIDENCE INTERVAL**

#### Example: Placekicking (Lecture 7 scripts.R, Placekick.csv)

Recall that we previously fit the probability of a successful kick as a function of distance using the Bernoulli trials:

```
> mod.fit <- glm(formula=good ~ distance,
                 family=binomial(link = logit), data=placekick)
```

Now we repeat the fit on the data aggregated by distance and show that it leads to the same parameter estimates:

```
> w<-aggregate(formula = good ~ distance, data = placekick,
FUN = sum)
> n<-aggregate(formula = good ~ distance, data = placekick,
FUN = length)
> w.n<-data.frame(distance = w$distance, success = w$good,
trials = n$good, proportion = round(w$good/n$good,4))
> head(w.n)
  distance success trials proportion
1       18         2     3      0.6667
2       19         7     7      1.0000
3       20       776   789      0.9835
4       21        19    20      0.9500
```

```

5      22      12      14      0.8571
6      23      26      27      0.9630

> # Using weights= to fit model to aggregated data.
> mod.fit.bin <- glm(formula=success/trials ~ distance,
weights=trials, family=binomial(link=logit), data=w.n)
> summary(mod.fit.bin)
Call: glm(formula = success/trials ~ distance,
family = binomial(link = logit), data = w.n, weights = trials)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0373  -0.6449  -0.1424   0.5004   2.2758

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.812080   0.326277   17.81  <2e-16 ***
distance    -0.115027   0.008339  -13.79  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 282.181 on 42 degrees of freedom
Residual deviance: 44.499 on 41 degrees of freedom
AIC: 148.46

Number of Fisher Scoring iterations: 5

```

Suppose we want to test whether the probability of success is related to the distance of the kick. The hypotheses are:

$$H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0$$

The Wald tests for testing the significance of a parameter (i.e., is it different from zero?) are given in the `summary()` output. For example,  $H_0: \beta_1 = 0$  has test statistic

$$Z_W = \frac{-0.115}{0.0083} = -13.79$$

and a p-value of “<2e-16”, which means that it is smaller than  $2 \times 10^{-16}$ , which is the smallest p-value that R reports. Thus, we strongly reject  $H_0$  and conclude that the distance of the kick is related to the probability that it is good.

The LR test is found using the `Anova()` function of the `car` package. Note that there is another function called `anova()` in the base `stats` package, but this function does things I don't like when  $p > 1$  (see “Sequential Tests” below), so I routinely use `Anova()`.

*If distance is not related, its coefficient would be 0 (1 unit change in  $X_1$  leads to 0 change in  $Y$ , holding other explanatory variables constant).*

*→ model object*

```
> library(car)
> Anova(mod.fit.bin, test = "LR")
Analysis of Deviance Table (Type II tests)

Response: success/trials
      LR Chisq Df Pr(>Chisq)
distance  237.68  1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test statistic is  $-2\log(\Lambda) = 237.68$  which leads to a p-value of  $< 2 \times 10^{-16}$  on  $\chi^2_1$ . This leads to the same conclusion as the Wald test.

Confidence intervals for the regression parameters are found next. The `confint()` function gives LR intervals when applied to objects of the class `glm`, while `confint.default()` does Wald.

```
> # LR CI for all parameters
> round(confint(mod.fit.bin), digits=3)
Waiting for profiling to be done...
      2.5 % 97.5 %
(Intercept)  5.196  6.477
distance    -0.132 -0.099
>
> # Wald Intervals
> round(confint.default(mod.fit.bin), digits=3)
      2.5 % 97.5 %
(Intercept)  5.173  6.452
distance    -0.131 -0.099
```

Both results are very similar, owing to the large sample size. Using LR as an example, we are 95% confident that the interval  $-0.132 < \beta_1 < -0.099$  covers the true change in log odds for each 1-yard change in distance.

### 3.3 Hypothesis tests with multiple parameters (Model comparison tests)

- When  $p > 1$ , there are several different kinds of tests that can be considered.
- MANY of them can be expressed as a comparison of two models

- Alternative hypothesis is the “full model” you are working from
- Null model is the “reduced model” that you have after the constraints in the null hypothesis are applied to the full model
- First, recall that  $H_0 : \beta_j = 0$  vs.  $H_a : \beta_j \neq 0$  is equivalent to saying that the model does not need to have  $x_j$  in it.
  - If  $p = 1$  then the comparison is simple:
    - \*  $H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1$  (Full model)
    - $H_0 : \text{logit}(\pi) = \beta_0$  (Reduced model)
- However, tests of  $H_0 : \beta_j = 0$  can have different interpretations and different results depending on what other variables are assumed to be in the model when variable  $x_j$  is removed
  - For example, suppose that  $p = 2$ . Then a test of  $H_0 : \beta_2 = 0$  could be interpreted as a comparison of two different pairs of models:
    1.  $H_0 : \text{logit}(\pi) = \beta_0 + \beta_1 x_1$   
 $H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
    2.  $H_0 : \text{logit}(\pi) = \beta_0$   
 $H_a : \text{logit}(\pi) = \beta_0 + \beta_2 x_2$
  - Similar for tests of  $\beta_1$ .
  - With more than 2 variables, the number of possibilities become overwhelming
    - \* Each of these comparisons can give a different result (and usually does)
- There are several different strategies for how sets of tests for  $p \geq 2$  variables can be tested for significance. The two main ones are PARTIAL tests and SEQUENTIAL tests
  - Partial tests assume that *all other variables will remain in the model* when the test of  $\beta_j = 0$  is considered.
    - \* That is, the “full model” always includes all the variables in the problem.
    - \* For example, for  $p = 3$ , testing the significance of each variable's parameter results in these three pairs of hypotheses:
      1. Testing  $H_0 : \beta_1 = 0$  vs.  $H_a : \beta_1 \neq 0$  is equivalent to testing
 
$$\begin{aligned} H_0 : \text{logit}(\pi) &= \beta_0 + \beta_2 x_2 + \beta_3 x_3 \\ H_a : \text{logit}(\pi) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \end{aligned}$$
      2. Testing  $H_0 : \beta_2 = 0$  vs.  $H_a : \beta_2 \neq 0$  is equivalent to testing
 
$$\begin{aligned} H_0 : \text{logit}(\pi) &= \beta_0 + \beta_1 x_1 + \beta_3 x_3 \\ H_a : \text{logit}(\pi) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \end{aligned}$$
      3. Testing  $H_0 : \beta_3 = 0$  vs.  $H_a : \beta_3 \neq 0$  is equivalent to testing
 
$$\begin{aligned} H_0 : \text{logit}(\pi) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\ H_a : \text{logit}(\pi) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \end{aligned}$$

- Sequential tests assume that there is an ordering to the importance of the variables and assume that only more important variables are in the null model. For example, suppose we think the importance ordering is  $x_1 > x_2 > x_3$ . Then

1. Testing  $H_0 : \beta_1 = 0$  vs.  $H_a : \beta_1 \neq 0$  is equivalent to testing

$$H_0 : \text{logit}(\pi) = \beta_0$$

$$H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1$$

2. Testing  $H_0 : \beta_2 = 0$  vs.  $H_a : \beta_2 \neq 0$  is equivalent to testing

$$H_0 : \text{logit}(\pi) = \beta_0 + \beta_1 x_1$$

$$H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

3. Testing  $H_0 : \beta_3 = 0$  vs.  $H_a : \beta_3 \neq 0$  is equivalent to testing

$$H_0 : \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

- Partial tests are by far the most common ones.

- \* Rarely have a prior ordering on the importance of variables

- \* An exception is polynomials:  $\text{logit}(\pi) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p$

- Then higher-order terms are less important than lower order terms.

• **FACT: The LR test statistic for comparing nested models can be conducted by comparing the RESIDUAL DEVIANCES of the two models**

- “Residual deviance” (often just called “deviance”) is a measure sort of like “Error sum of squares” in regression

- \* It measures the amount of “information” in  $Y$  that is not explained by the fitted model

- \* It is related to the log-likelihood of the model

*How the model fits the data  $\rightarrow$  less deviance.*

- \* The more variables you add to the model, the smaller the residual deviance becomes

- \* A model that explains the responses perfectly has residual deviance of 0  $\rightarrow$  Not necessarily a good thing!

- A SATURATED MODEL—one that has as many parameters as observations—does this

- Certain smaller models may do this as well, by luck

- \* The degrees of freedom for the residual deviance are  $n - p$ , just like in regression<sup>1</sup>

- “Nested model” means that the null model is a special case of the alternative model, where certain parameter restrictions have been applied (like setting some parameters to 0)

- \* Models  $\beta_0 + \beta_1 x_1$  and  $\beta_0 + \beta_2 x_2 + \beta_3 x_3$  are NOT nested, for example. There is something each model that is not in the other.

*by 0  
not 0  $\rightarrow$  Nested model*

<sup>1</sup>In the context of logistic regression to aggregated data, “n” refers to the total number of fitted points, not the total number of trials.

*## df for a test: ##*

*# of different parameters in  $H_0$  and  $H_a$*

- Denote the residual deviance for the null-hypotheses model by  $D_0$  and for the alternative-hypothesis model by  $D_a$ . Then the LR test stat can be written as

$$-2 \log(\Lambda) = D_0 - D_a$$

- \* This gets compared to  $\chi^2_\nu$ , where  $\nu$  is the difference in residual degrees of freedom between the two models.

- All of the tests in the examples above can be computed by fitting the two models in the corresponding null and alternative hypotheses, and taking the resulting differences in deviances.

- How many DF would they have? |

- All of this is important! Remember it!

**Simultaneous testing of multiple parameters**

- In addition to testing one parameter at a time, can test multiple parameters simultaneously

- For example “significance of the model” asks whether the *entire model* has any explanatory value

- \* Tests  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ .



- For example “significance of the model” asks whether the *entire model* has any explanatory value
  - \* Tests  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ .
  - \* For any  $p$  this is tested by comparing
 
$$H_0 : \text{logit}(\pi) = \beta_0$$

$$H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$
- Both models are fit and the LR test stat is found as the difference in deviances again.
  - \* How many DF would the model significance test have?  $p$
- In certain cases, there are other combinations of variables that you might want to consider removing at once and then determine whether this change has “significantly harmed” the model fit.
  - For example, in a large model, some variables may be related to measurements taken by one machine that is expensive to run. Do we need these measurements in our model, or would a simpler, much cheaper set of measurements do nearly as well?
  - Same technology applies to compare the nested models
  - $DF = \#$  variables removed.

**Example: Placekicking (Lecture 8 scripts.R, Placekick.csv)**

We now add the binary variable `change` to the model, to see whether the probability that the kick is successful,  $\pi$ , is affected by the pressure on the kicker to change the lead in the game, after accounting for the distance of the kick. To do this, we fit the model

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

So  $p = 2$  here,  $x_1$  is `distance`, and  $x_2$  is `change`. The model fit is given below:

```
> mod.fit2 <- glm(formula=good ~ distance + change,
family=binomial(link=logit), data=placekick)
> summary(mod.fit2)

Call:
glm(formula = good ~ distance + change,
family = binomial(link = logit), data = placekick)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7061   0.2282   0.2282   0.3750   1.5649

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.893181    0.333184   17.687  <2e-16 ***
distance    -0.112889    0.008444  -13.370  <2e-16 ***
change      -0.447783    0.193673   -2.312   0.0208 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1013.4  on 1424  degrees of freedom
Residual deviance:  770.5  on 1422  degrees of freedom
AIC: 776.5

Number of Fisher Scoring iterations: 6
```

*All Wald tests  
are passed*

- The estimated logistic regression model is

$$\text{logit}(\hat{\pi}) = 5.89 - 0.113\text{distance} - 0.448\text{change}$$

- When `change=1` (kick changes the lead) the log-odds of success decrease by 0.448 compared to when `change=0` (no lead change).

- \* At first glance, "pressure" reduces the probability of success.
- The Wald test for this effect ( $H_0 : \beta_2 = 0$ ) has  $Z_W = -2.31$  and  $p\text{-value}=0.021$ . This means that the change in probability (through the log-odds) is significant
- \* The sample size is large with lots of successes and failures, so I trust this result.

- Nonetheless, we do LR tests for each parameter next:

```
> # Partial tests (add each variable last)
> Anova(mod.fit2) # omitting test="LR" (default)
Analysis of Deviance Table (Type II tests)
```

*partial*

```
Response: good
      LR Chisq Df Pr(>Chisq)
distance 218.650 1 <2e-16 ***
change    5.246 1  0.022 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # Sequential tests
> anova(mod.fit2, test = "Chisq")
Analysis of Deviance Table
Model: binomial, link: logit
Response: good
Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                1424    1013.43
distance 1 237.681                1423     775.75 <2e-16 ***
change   1   5.246                1422     770.50 0.022 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Conclusions are the same: both p-values are  $< 0.05$  so we reject each respective  $H_0 : \beta_j = 0$  and conclude that both **distance** and **change** are important variables in this model, given that the other is in the model.
- In particular, we can test the significance of the entire model,  
 $H_0 : \text{logit}(\pi) = \beta_0$   
 $H_a : \text{logit}(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ 
  - To do this we fit a model with intercept only, **formula=Y~1**

```
> mod.fit0 <- glm(formula=good ~ 1, family=binomial(link=logit),
data=placekick)
> # Note 2 df
> anova(mod.fit0, mod.fit2, test = "Chisq")
Analysis of Deviance Table
```

```
Model 1: good ~ 1
Model 2: good ~ distance + change
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1424      1013.4
2      1422       770.5  2    242.93 < 2.2e-16 ***
```

–  $-2\log(\Lambda) = 242.9$  which gives a tiny p-value on  $\chi^2_2$  (2df!). Conclude that the overall model is a significantly better fit than assuming that all probabilities are constant.

\* This test does not tell you *which* variables might be useful. It need not be all of them!

## 4 What to learn from this

1. Everything. Seriously. This is fundamental to all we will do.
  - (a) Fitting a logistic regression model to binary, partially aggregated, or fully aggregated (EVP form) gives the same parameter estimates and inferences.
  - (b) Tests and CIs for a single parameter
    - i. Use LR where possible
    - ii. Sequential and partial tests
  - (c) Comparing nested models and residual deviances
2. Even some of the R hints in the scripts are helpful for future use.

## 5 Appendix: Profile likelihood

- If  $H_0$  pertains to only one parameter, then there are  $p$  additional parameters that still need to be estimated
  - When  $p = 1$ , the likelihood function has a dome-like shape when plotted against  $\beta_0$  and  $\beta_1$
  - Suppose we want to test  $H_0 : \beta_1 = c$  vs.  $H_a : \beta_1 \neq c$

- \* So need to refit the model, fixing  $\beta_1$  at the hypothesized value, and find the MLEs of the remaining parameters under that constraint
  - This gives us the maximized likelihood under  $H_0$
- \* Fixing  $\beta_j = c$  is like taking a cross-section of the dome, resulting in a concave curve called a **LIKELIHOOD PROFILE**.
- \* Then find the value of  $\beta_0$  that maximizes this profile
  - The new estimate is likely to be different from the original MLE!
- \* Get the value of the likelihood function at the top of the profile
  - Plug that maximized value into the equation above.
- \* The evaluated likelihood at the MLEs with one parameter fixed is smaller than the unrestricted version unless, by sheer luck, the original MLE  $\hat{\beta}_j$  is exactly equal to the hypothesized value.
- \* The  $\chi^2_{1,1-\alpha}$  critical value tells us how much smaller is too much smaller on the log-likelihood
- When  $p > 1$ , it is a higher-dimensional dome-like shape that we can no longer picture, but the concept and process is the same.
  - \* Fix any parameters specified under  $H_0$
  - \* Create a profile likelihood curve in the remaining dimensions
  - \* Maximize that likelihood to obtain the needed numerator for  $\Lambda$

## 6 Appendix: Additional details about sequential and partial tests

- Note that the LR test statistics are close to the squares of the Wald test stats from the partial tests, but not quite the same.
  - $2.312^2 = 5.345$ ,  $-13.370^2 = 178.757$  vs.  $5.246$ ,  $218.65$
  - In large samples the squared Wald and LR statistic grow closer together because the distribution of the square of a  $N(0, 1)$  random variable is exactly  $\chi^2_1$ .
- Note that the last sequential test from any model matches the partial test for the corresponding variable (why??)
  - The remaining partial tests are almost always different from their respective sequential tests
  - Also note that you get different sequential test results for both variables if we reverse the order in the **formula** specification

```

> mod.fit2a <- glm(formula=good ~ change + distance,
family=binomial(link=logit), data=placekick)
> anova(mod.fit2a, test = "Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: good

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                1424    1013.43
change    1    24.277      1423     989.15 8.343e-07 ***
distance  1   218.650      1422     770.50 < 2.2e-16 ***

```

– The partial tests don't change

- We can perform any of these tests by naming the models in `anova(<H0 model>, <Ha model>)`

– For example, the test for `change` given that `distance` is in the model:

```

> anova(mod.fit, mod.fit2, test = "Chisq")
Analysis of Deviance Table

Model 1: good ~ distance
Model 2: good ~ distance + change
      Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          1423      775.75
2          1422      770.50  1    5.2455    0.022 *

```

– Residual deviance for null model (Model 1) is  $D_0 = 775.75$

– Residual deviance for alternative model (Model 2) is  $D_a = 770.50$

\*  $-2 \log(\Lambda) = D_0 - D_a = 5.25$

## 7 Exercises (due when announced)

Complete the following exercises from B&L, Chapter 2. As always, add proper interpretations on all confidence intervals and tests.

1. Using the data from Exercises 4 and 5, compute 95% LR and Wald confidence intervals for the effect of temperature on the probability of o-ring failure.
2. For our placekick example,
  - (a) Fit a model containing all explanatory variables
  - (b) We know that **distance** is important. Using a single model-comparison LR test, determine whether *all of the rest* of the variables can be dropped. Use  $\alpha = 0.05$ .
  - (c) Use partial LR tests with  $\alpha = 0.05$  to determine which variables seem to contribute significantly to the model.
    - i. Given the results of this test, could you use a model with **Distance** and **PAT** only? Explain.

Practice exercises, not marked:

1. What does  $\beta_0 = 0$  imply about the probability of success?