**Section 2.2.2 – Hypothesis tests for regression parameters**

One way to assess the importance of an explanatory variable is through using a hypothesis test. For example, suppose we are interested in the rth explanatory variable xr in the model



If βr = 0, we see that xr would be excluded from the model. Thus, we are interested in hypothesis tests of the form:

H0: βr = 0

Ha: βr ≠ 0

Alternatively, we could state the hypotheses as:



Comments:

* The null hypothesis model terms are all included within the alternative hypothesis model. In other words, the null hypothesis model is a special case of the alternative hypothesis model.
* The null hypothesis model is often referred to as a reduced model and the alternative hypothesis model is often referred to as a full model.

The purpose of this section is to examine two ways that hypothesis tests of this form can be performed.

Wald test

The Wald statistic is



to test H0: βr = 0 vs. Ha: βr ≠ 0. For a large sample, the test statistic has an approximate standard normal distribution if the null hypothesis of βr = 0 is true. Thus, reject the null hypothesis if we observe a test statistic value that is “unusual” for a standard normal distribution. We define unusual to be . The p-value is  where Z has a standard normal distribution. Wald test statistics and p-values are automatically provided for individual βr parameters using code like summary(mod.fit).

The Wald test can also be performed for more than one regression parameter at the same time. However, because the Wald test has similar problems to those we saw in Chapter 1, we are not going to pursue how to perform these types of tests.

Likelihood ratio test (LRT)

Generally, a better test than the Wald is a LRT. The LRT statistic is again:



To perform a test of H0: βr = 0 vs. Ha: βr ≠ 0, we obtain the estimated probabilities of success from estimating



(suppose the estimates are , , … ) and the estimated probabilities of success from estimating



(suppose the estimates are , , … ). We can then find





 (1)

where I use  and  to be vectors of estimated regression parameters under the H0 and Ha models. If the null hypothesis was true, -2log(Λ) has an approximate  distribution for a large sample.

Questions:

* What happens to Λ and -2log(Λ) as evidence grows against the null hypothesis?
* When should the null hypothesis be rejected?
* How should we calculate the p-value?

The hypotheses can be generalized to allow for q different regression parameters to be set to 0 in H0. When this happens and the null hypothesis is true, -2log(Λ) has an approximate  distribution for a large sample.

Example: Placekicking (Placekick.R, Placekick.csv)

Consider the model with both the change and distance variables included in it:



To perform Wald tests on each of these variables, we can examine the summary(mod.fit2) output from earlier:

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

> summary(mod.fit2)

Call: glm(formula = good ~ change + distance, 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 \*\*\*

change -0.447783 0.193673 -2.312 0.0208 \*

distance -0.112889 0.008444 -13.370 <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: 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

Below is a summary of the tests using α = 0.05:

|  |  |
| --- | --- |
| **Change** | **Distance** |
| H0: β1 = 0 Ha: β1 ≠ 0  | H0: β2 = 0 Ha: β2 ≠ 0  |
| ZW = -2.31 | ZW = -13.37 |
| p-value = 0.0208 | p-value < 2×10-16 |
| Reject H0 because p-value is small | Reject H0 because p-value is small |
| There is sufficient evidence to indicate change has an effect on the probability of success given distance is in the model. | There is sufficient evidence to indicate distance has an effect on the probability of success given change is in the model. |

The change test has a p-value less than 0.05, but it is only a little less. In fact, if α = 0.01, there would not be a rejection of the null hypothesis. It is **preferable** to word a conclusion like

There is **marginal** evidence to indicate that a lead change placekick has an effect on the probability of success given distance is in the model.

There are a number of ways to perform LRTs in R. The easiest way to perform the tests of interest here is to use the Anova() function from the car package. This package is not automatically installed in R so you will need to install it prior to its use. The package corresponds to the book “An R Companion to Applied Regression” by Fox and Weisberg. Below is the R code and output:

> library(package = car)

> Anova(mod = mod.fit2, test = "LR")

Analysis of Deviance Table (Type II tests)

Response: good

 LR Chisq Df Pr(>Chisq)

change 5.246 1 0.02200 \*

distance 218.650 1 < 2e-16 \*\*\*

---

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

Notice the p-values are quite similar to the p-values for the Wald test. This is due to the large sample.

Within the stats package, the anova() function can perform LRTs. Below is what occurs with a somewhat naïve use of it:

> anova(object = 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 P(>|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 \*\*\*

---

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

The p-value for distance is the same as from using Anova(), but the p-value for change is not. The reason for the difference is due to the hypotheses being tested. The anova() function tests the model’s explanatory variables in an sequential manner. Thus, the change test p-value is actually for the test of



because it is listed first in the formula argument of glm(). The test for distance is listed second so anova() tests:



where change is assumed to be in both models.

To produce tests like Anova(), we need to estimate the H0 and Ha models separately and then use their model fit objects in a different way with anova():

> mod.fit.Ho <- glm(formula = good ~ distance, family =

 binomial(link = logit), data = placekick)

> anova(mod.fit.Ho, mod.fit2, test = “Chisq”)

Analysis of Deviance Table

Model 1: good ~ distance

Model 2: good ~ change + distance

 Resid. Df Resid. Dev Df Deviance P(>|Chi|)

1 1423 775.75

2 1422 770.50 1 5.2455 0.02200 \*

Signif. Codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ‘ 1

This use of anova() helps to emphasize a reduced and full model approach to obtaining the -2log(Λ) statistic. Also, this approach is helpful to know when Anova() may not be available (with more complex models than logistic regression) or when more than one variable is being tested at a time.

Question: Why do you think these function names involve the word “anova” when analysis of variance is not being performed?

Of course, you could program into R the -2log(Λ) statistic yourself and not even use Anova() or anova(). Please see the example in the book for it. Again, knowing how to perform a test in this manner is useful when there are no Anova() or anova() functions available, like for a statistical research problem! Also, going through calculations like this is helpful to understand the test statistic being used.

Deviance

In the previous example's output, the word “deviance” and its abbreviation “dev” appeared a number of times. Deviance refers to the amount that a particular model deviates from another model as measured by -2log(Λ). For example, the -2log(Λ) = 5.246 value used for testing change in the last example (given distance is in the model) is a measure of how much the estimated probabilities of success for the null hypothesis model deviate from the estimated probabilities of success for the alternative hypothesis model.

Residual deviance – This denotes how much the null hypothesis model of interest deviates from using the observed proportion of successes for each “observation” (yi = 0 or 1 for Bernoulli response data or wj/nj for binomial response data) to estimate π (πi or πj).

Using the observed proportion of successes is equivalent to using a model of the form  (i = 1, …, n) for Bernoulli response data or of the form  (j = 1, …, J) for binomial response data. Because the number of parameters is equal to the number of observations, this model is frequently referred to as the saturated model because no more additional parameters can be estimated.

Null deviance – This denotes how much the model  (or ) deviates from using the observed proportion of successes for each observation to estimate πi or (πj). Because  contains only the intercept term (just one parameter), every πi is estimated to be the same value for this particular model.

Deviance statistics are calculated often as an intermediary step for performing a LRT to compare two models. For example, consider testing:



The residual deviance for the model  tests



Ha:Saturated model

with

 (2)

Note: If written in terms of binomial response data, we would have:



The residual deviance for the model  tests



Ha:Saturated model

with

 (3)

By subtracting (3) from (2) we obtain Equation (1):



Thus, we obtain the desired -2log(Λ) by subtracting the residual deviances of the models in our original hypotheses!

The degrees of freedom for the test can be found from subtracting the corresponding degrees of freedom for (3) from (2).

Example: Placekicking (Placekick.R, Placekick.csv)

Below is the R code demonstrating the above discussion with respect to



> mod.fit.Ho <- glm(formula = good ~ distance, family =

 binomial(link = logit), data = placekick)

> names(mod.fit.Ho)

 [1] "coefficients" "residuals"

 [3] "fitted.values" "effects"

 [5] "R" "rank"

 [7] "qr" "family"

 [9] "linear.predictors" "deviance"

[11] "aic" "null.deviance"

[13] "iter" "weights"

[15] "prior.weights" "df.residual"

[17] "df.null" "y"

[19] "converged" "boundary"

[21] "model" "call"

[23] "formula" "terms"

[25] "data" "offset"

[27] "control" "method"

[29] "contrasts" "xlevels"

> df <- mod.fit.Ho$df.residual - mod.fit2$df.residual

> stat <- mod.fit.Ho$deviance - mod.fit2$deviance

> pvalue <- 1 - pchisq(q = stat, df = df)

> data.frame(Ho.resid.dev = mod.fit.Ho$deviance,

 Ha.resid.dev = mod.fit2$deviance, df = df, stat =

 round(stat,4), pvalue = round(pvalue,4))

 Ho.resid.dev Ha.resid.dev df stat pvalue

1 775.745 770.4995 1 5.2455 0.022

Final note: Use the easiest ways to perform the LRTs unless asked to otherwise. Some of the longer ways to perform the LRT were shown to:

1. Reinforce how the statistic was found,
2. Prepare you for material later in the course, and
3. Show common ways that others may perform tests.