16.2: Logit Estimation

Logit is used when predicting limited dependent variables, specifically those in which YY is represented by 00's and 11's. By virtue of the binary dependent variable, these models do not meet the key assumptions of OLS. Logit uses maximum likelihood estimation (MLE), which is a counterpart to minimizing least squares. MLE identifies the probability of obtaining the sample as a function of the model parameters (i.e., the XX's). It answers the question, what are the values for BB's that make the sample most likely? In other words, the likelihood function expresses the probability of obtaining the observed data as a function of the model parameters. Estimates of AA and BB are based on maximizing a likelihood function of the observed YY values.
In logit estimation we seek P(Y=1)P(Y=1), the probability that Y=1Y=1. The odds that Y=1Y=1 are expressed as:

$$O(Y=1)=P(Y=1)1-P(Y=1)O(Y=1)=P(Y=1)1-P(Y=1) \tag{16.2.1}$$

Logits, LL, are the natural logarithm of the odds:

$$L=logeOL=logeP1-PL=logeOL=logeP1-P \tag{16.2.2}$$

They can range from $-\infty-\infty$, when P=0P=0, to $\infty\infty$, when P=1P=1. LL is the estimated systematic linear component:

$$L=A+B1Xi1+\ldots+BkXikL=A+B1Xi1+\ldots+BkXik \tag{16.2.3}$$

By reversing the logit we can obtain the predicted probability that Y=1Y=1 for each of the ii observations:

$$Pi=11-e-Li(16.2)(16.2)Pi=11-e-Li \tag{16.2.4}$$

where e=2.71828…e=2.71828…, the base number of natural logarithms. Note that LL is a linear function, but PP is a non-linear SS-shaped function as shown in Figure 16.2.2 Also note, that Equation 16.2 is the link function that relates the linear component to the non-linear response variable.

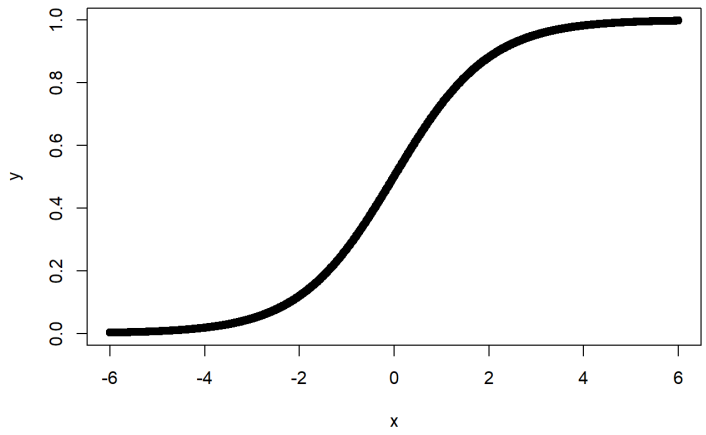

Figure 16.2.2: Predicted Probability as a Logit Function of XX

In more formal terms, each observation, ii, contributes to the likelihood function by PiPi if Yi=1Yi=1, and by 1−Pi1−Pi if Yi=0Yi=0. This is defined as:

$$PYii(1-Pi)1-YiPiYi(1-Pi)1-Yi \tag{16.2.5}$$

The likelihood function is the product (multiplication) of all these individual contributions:

$$\ell=\prod PYii(1-Pi)1-Yi\ell=\prod PiYi(1-Pi)1-Yi \tag{16.2.6}$$

The likelihood function is the largest for the model that best predicts Y=1Y=1 or Y=0Y=0; therefore when the predicted value of YY is correct and close to 11 or 00, the likelihood function is maximized.

To estimate the model parameters, we seek to maximize the log of the likelihood function. We use the log because it converts the multiplication into addition, and is therefore easier to calculate. The log likelihood is:

$$logel=n\sum i=1[YilogePi+(1-Yi)loge(1-Pi)]logel=\sum i=1n[YilogePi+(1-Yi)loge(1-Pi)] \tag{16.2.7}$$

The solution involves taking the first derivative of the log likelihood with respect to each of the BB's, setting them to zero, and solving the simultaneous equation. The solution of the equation isn't linear, so it can't be solved directly. Instead, it's solved through a sequential estimation process that looks for successively better fits'' of the model.

For the most part, the key assumptions required for logit models are analogous to those required for OLS. The key differences are that (a) we do not assume a linear relationship between the XXs and YY, and (b) we do not assume normally distributed, homoscedastistic residuals. The key assumptions that are retained are shown below.

Logit Assumptions and Qualifiers - The model is correctly specified - True conditional probabilities are logistic function of the XX's - No important XX's omitted; no extraneous XX's included - No significant measurement error - The cases are independent - No XX is a linear function of other XX's - Increased multicollinearity leads to greater imprecision - Influential cases can bias estimates - Sample size: n−k−1n−k−1 should exceed 100100 - Independent covariation between the XXs and YY is critical

The following example uses demographic information to predict beliefs about anthropogenic climate change.

```
ds.temp <- ds %>%
  dplyr::select(glbcc, age, education, income, ideol, gender) %>%
  na.omit()

logit1 <- glm(glbcc ~ age + gender + education + income, data = ds.temp, family = binomi
summary(logit1)
```

```
##
## Call:
## glm(formula = glbcc ~ age + gender + education + income, family = binomial(),
##     data = ds.temp)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.707  -1.250   0.880   1.053   1.578
##
## Coefficients:
##                   Estimate    Std. Error z value       Pr(>|z|)
## (Intercept)  0.4431552007  0.2344093710   1.891       0.058689 .
## age         -0.0107882966  0.0031157929  -3.462       0.000535 ***
## gender      -0.3131329979  0.0880376089  -3.557       0.000375 ***
## education    0.1580178789  0.0251302944   6.288 0.000000000322 ***
## income      -0.0000023799  0.0000008013  -2.970       0.002977 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3114.5  on 2281  degrees of freedom
## Residual deviance: 3047.4  on 2277  degrees of freedom
## AIC: 3057.4
##
## Number of Fisher Scoring iterations: 4
```

As we can see, age and gender are both negative and statistically significant predictors of climate change opinion. Below we discuss logit hypothesis tests, goodness of fit, and how to interpret the logit coefficients.

16.2.1 Logit Hypothesis Tests

In some ways, hypothesis testing with logit is quite similar to that using OLS. The same use of pp-values is employed; however, they differ in how they are derived. The logit analysis makes use of the Wald zz-statistic, which is similar to the tt-stat in OLS. The Wald zz

score compares the estimated coefficient to the asymptotic standard error, (aka the normal distribution). The pp-value is derived from the asymptotic standard-normal distribution. Each estimated coefficient has a Wald zz-score and a pp-value that shows the probability that the null hypothesis is correct, given the data.

z=BjSE(Bj)(16.3)(16.3)z=BjSE(Bj)

16.2.2 Goodness of Fit

Given that logit regression is estimated using MLE, the goodness-of-fit statistics differ from those of OLS. Here we examine three measures of fit: log-likelihood, the pseudo R2R2, and the Akaike information criteria (AIC).

Log-Likelihood

To test for the overall null hypothesis that all BB's are equal to zero (similar to an overall FF-test in OLS), we can compare the log-likelihood of the demographic model with 4 IVs to the initial null model," which includes only the intercept term. In general, a smaller log-likelihood indicates a better fit. Using the deviance statistic G2G2 (aka the likelihood-ratio test statistic), we can determine whether the difference is statistically significant. G2G2 is expressed as:

G2=2(logeL1−logeL0)(16.4)(16.4)G2=2(logeL1−logeL0)

where L1L1 is the demographic model and L0L0 is the null model. The G2G2 test statistic takes the difference between the log likelihoods of the two models and compares that to a χ2χ2 distribution with qq degrees of freedom, where qq is the difference in the number of IVs. We can calculate this in `R` . First, we run a null model predicting belief that greenhouse gases are causing the climate to change, using only the intercept:

```
logit0 <- glm(glbcc ~ 1, data = ds.temp)
summary(logit0)
```

```
##
## Call:
## glm(formula = glbcc ~ 1, data = ds.temp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5732  -0.5732   0.4268   0.4268   0.4268
##
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept)  0.57318    0.01036   55.35 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2447517)
##
##     Null deviance: 558.28  on 2281  degrees of freedom
## Residual deviance: 558.28  on 2281  degrees of freedom
## AIC: 3267.1
##
## Number of Fisher Scoring iterations: 2
```

We then calculate the log likelihood for the null model,

logeL0(16.5)(16.5)logeL0

```
logLik(logit0)
```

```
## 'log Lik.' -1631.548 (df=2)
```

Next, we calculate the log likelihood for the demographic model,

logeL0(16.6)(16.6)logeL0

Recall that we generated this model (dubbed “logit1”) earlier:

```
logLik(logit1)
```

```
## 'log Lik.' -1523.724 (df=5)
```

Finally, we calculate the GG statistic and perform the chi-square test for statistical significance:

```
G <- 2*(-1523 - (-1631))
G
```

```
## [1] 216
```

```
pchisq(G, df = 3, lower.tail = FALSE)
```

```
## [1] 0.000000000000000000000000000000000000000000000001470144
```

We can see by the very low p-value that the demographic model offers a significant improvement in fit.

The same approach can be used to compare nested models, similar to nested FF-tests in OLS. For example, we can include ideology in the model and use the `anova` function to see if the ideology variable improves model fit. Note that we specify the χ2χ2 test.

```
logit2 <- glm(glbcc ~ age + gender + education + income + ideol,
              family = binomial(), data = ds.temp)
summary(logit2)
```

```
##
## Call:
## glm(formula = glbcc ~ age + gender + education + income + ideol,
##     family = binomial(), data = ds.temp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6661  -0.8939   0.3427   0.8324   2.0212
##
## Coefficients:
##                  Estimate    Std. Error z value             Pr(>|z|)
## (Intercept)  4.0545788430  0.3210639034  12.629 < 0.0000000000000002 ***
## age         -0.0042866683  0.0036304540  -1.181             0.237701
## gender      -0.2044012213  0.1022959122  -1.998             0.045702 *
## education    0.1009422741  0.0293429371   3.440             0.000582 ***
## income      -0.0000010425  0.0000008939  -1.166             0.243485
## ideol       -0.7900118618  0.0376321895 -20.993 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3114.5  on 2281  degrees of freedom
## Residual deviance: 2404.0  on 2276  degrees of freedom
## AIC: 2416
##
## Number of Fisher Scoring iterations: 4
```

```
anova(logit1, logit2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: glbcc ~ age + gender + education + income
## Model 2: glbcc ~ age + gender + education + income + ideol
##   Resid. Df Resid. Dev Df Deviance              Pr(>Chi)
## 1      2277     3047.4
## 2      2276     2404.0  1   643.45 < 0.00000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we can see, adding ideology significantly improves the model.

Pseudo R2R2

A measure that is equivalent to the R2R2 in OLS does not exist for logit. Remember that explaining variance in YY is not the goal of MLE. However, a pseudo'' R2R2 measure exists that compares the residual deviance of the null model with that of the full model. Like the R2R2 measure, pseudo R2R2 ranges from 00 to 11 with values closer to 11 indicating improved model fit.

Deviance is analogous to the residual sum of squares for a linear model. It is expressed as:

deviance=−2(logeL)(16.7)(16.7)deviance=−2(logeL)

It is simply the log-likelihood of the model multiplied by a −2−2. The pseudo R2R2 is 11 minus the ratio of the deviance of the full model L1L1 to the deviance of the null model L0L0:

pseudoR2=1−−2(logeL1)−2(logeL0)(16.8)(16.8)pseudoR2=1−−2(logeL1)−2(logeL0)

This can be calculated in 'R' using the full model with ideology.

```
pseudoR2 <- 1 - (logit2$deviance/logit2$null.deviance)
pseudoR2
```

```
## [1] 0.2281165
```

The pseudo R2R2 of the model is 0.2281165. Note that the psuedo R2R2 is only an approximation of explained variance, and should be used in combination with other measures of fit such as AIC.

Akaike Information Criteria

Another way to examine goodness-of-fit is the Akaike information criteria (AIC). Like the adjusted R2R2 for OLS, the AIC takes into account the parsimony of the model by penalizing for the number of parameters. But AIC is useful only in a comparative manner – either with the null model or an alternative model. It does not purport to describe the percent of variance in YY accounted for, as does the pseudo R2R2.

AIC is defined as -2 times the residual deviance of the model plus two times the number of parameters, or kk IVs plus the intercept:

AIC=−2(logeL)+2(k+1)(16.9)(16.9)AIC=−2(logeL)+2(k+1)

Note that smaller values are indicative of a better fit. The AIC is most useful when comparing the fit of alternative (not necessarily nested) models. In `R`, AIC is given as part of the `summary` output for a `glm` object, but we can also calculate it and verify.

```
aic.logit2 <- logit2$deviance + 2*6
aic.logit2
```

```
## [1] 2416.002
```

```
logit2$aic
```

```
## [1] 2416.002
```

16.2.3 Interpreting Logits

The logits, LL, are logged odds, and therefore the coefficients that are produced must be interpreted as logged odds. This means that for each unit change in ideology, the predicted logged odds of believing climate change has an anthropogenic cause decrease by -0.7900119. This interpretation, though mathematically straightforward, is not terribly informative. Below we discuss two ways to make the interpretation of logit analysis more intuitive.

Calculate Odds

Logits can be used to directly calculate odds by taking the antilog of any of the coefficients:

antilog=eBantilog=eB

For example, the following retuns odds for all the IVs.

```
logit2 %>% coef() %>% exp()
```

```
## (Intercept)         age      gender   education      income       ideol
##  57.6608736   0.9957225   0.8151353   1.1062128   0.9999990   0.4538394
```

Therefore, for each 1-unit increase in the ideology scale (i.e., becoming more conservative), the odds of believing that climate change is human caused decrease by 0.4538394.

Predicted Probabilities

The most straightforward way to interpret logits is to transform them into predicted probabilities. To calculate the effect of a particular independent variable, XiXi, on the probability of Y=1Y=1, set all XjXj's at their means, then calculate:

^P=11+e−^LP^=11+e−L^

We can then evaluate the change in predicted probabilities that YY=1 across the range of values in XiXi.

This procedure can be demonstrated in two steps. First, create a data frame holding all the variables except ideology at their mean. Second, use the `augment` function to calculate the predicted probabilities for each level of ideology. Indicate `type.predict = "response"` .

```
library(broom)
log.data <- data.frame(age = mean(ds.temp$age),
                       gender = mean(ds.temp$gender),
                       education = mean(ds.temp$education),
                       income = mean(ds.temp$income),
                       ideol = 1:7)
log.data <- logit2 %>%
  augment(newdata = log.data, type.predict = "response")
log.data
```

```
## # A tibble: 7 x 7
##     age gender education income ideol .fitted .se.fit
## * <dbl>  <dbl>     <dbl>  <dbl> <int>   <dbl>   <dbl>
## 1  60.1  0.412      5.09 70627.     1   0.967 0.00523
## 2  60.1  0.412      5.09 70627.     2   0.929 0.00833
## 3  60.1  0.412      5.09 70627.     3   0.856 0.0115
## 4  60.1  0.412      5.09 70627.     4   0.730 0.0127
## 5  60.1  0.412      5.09 70627.     5   0.551 0.0124
## 6  60.1  0.412      5.09 70627.     6   0.357 0.0139
## 7  60.1  0.412      5.09 70627.     7   0.202 0.0141
```

The output shows, for each case, the ideology measure for the respondent followed by the estimated probability (pp) that the individual believes man-made greenhouse gasses are causing climate change. We can also graph the results with 95%95% confidence intervals. This is shown in Figure 16.2.3

```
log.df <- log.data %>%
  mutate(upper = .fitted + 1.96 * .se.fit,
         lower = .fitted - 1.96 * .se.fit)

ggplot(log.df, aes(ideol, .fitted)) +
  geom_point() +
  geom_errorbar(aes(ymin = lower, ymax = upper, width = .2))
```

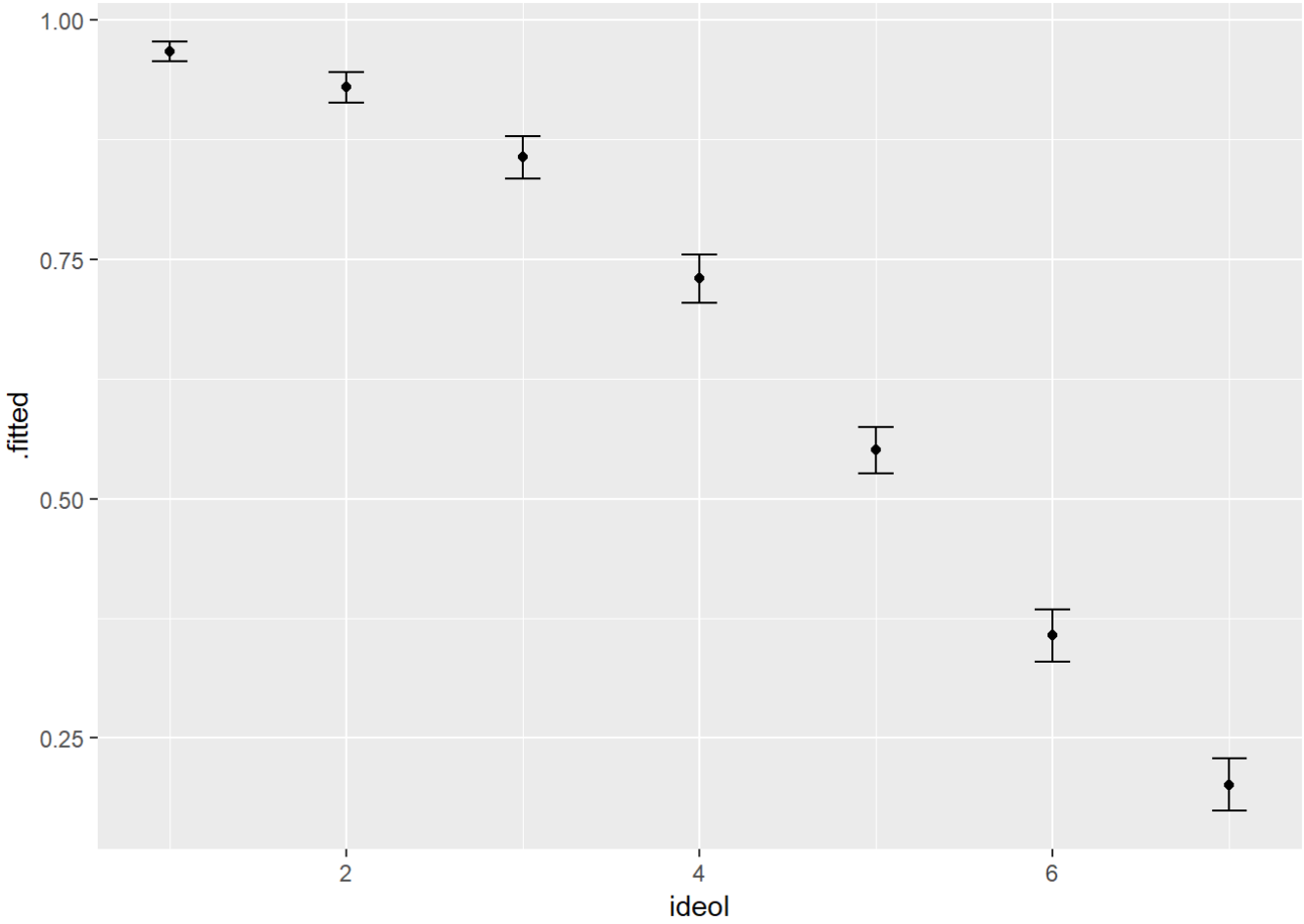

Figure 16.2.3: Predicted Probability of believing that Greenhouse Gases cause Climate Change by Ideology

We can see that as respondents become more conservative, the probability of believing that climate change is man-made decreases at what appears to be an increasing rate.