

CHAPTER 18:

NOMINAL AND ORDINAL DEPENDENT VARIABLES

18.1	Nominal Dependent Variable	466
18.2	Ordinal Dependent Variable	474
18.3	Conclusion	478
18.4	R Functions	479
18.5	Extensions	480
18.6	Applications	481
18.7	References and Further Readings	482

At first glance, this question appears to be a binary dependent variable problem. After all, there are only two parties, right? Well, even if you ignore third parties, there is a third option: abstention. In each US presidential election, a sizable number of registered voters decide not to vote. For instance, while Obama received 53% of the vote cast and McCain received 46%, a full 37.8% of the eligible voters did not vote. Thus, the distribution of votes in the 2008 US Presidential election is 33.0% Obama, 28.6% McCain, 0.6% other, and 37.8% none of the above. As such, conclusions based on those models that assume a binary outcome have definite issues with generalization to the voting public at large.

A better alternative is to specifically add in ‘abstention’ and model the three possible outcomes at once. Such a regression model is called a nominal regression model or a multinomial regression model.

There is a second type of dependent variable that is closely related to the nominal case—the ordinal dependent variable. The difference between the nominal and the ordinal is that the ordinal has more information contained in it. There is no ordering in the nominal case. There is an implicit ordering in the ordinal case.^a Examples of ordinal variables include ratings and indices.

If we model nominal dependent variables by using a series of binary regressions, we come up with odd results. If we force a nominal variable into just two categories, we lose information in the data. If we treat ordinal dependent variables simply as nominal, information is also lost. If we treat them as continuous, our conclusions may not match reality. This chapter examines how to model both the nominal dependent variable and the ordinal dependent variable more properly.



One of the most pervasive research questions in Political Science is to predict a person's vote based on demographic information. In other words, if you know a person's age, gender, income, education, and religion, how well can you predict how that individual will vote in the upcoming presidential election?

^aOrdinal is actually an abbreviation for ordered nominal.

18.1: Nominal Dependent Variable

A nominal variable is a categorical variable where there does not exist an inherent ordering in the categories. Examples may include job type, presidential vote (and non-vote), and beer brand choice. These variables are categorical — not continuous — and the categories have no inherent ordering. White Collar is not ‘greater than’ Professional. Voting Democratic is not ‘larger than’ voting Republican. Budweiser is not ‘more than’ Coors.¹ How do we model such dependent variables?

There are a couple ways of doing this. The first, easiest, and most understandable is to model the variable as a series of binary dependent variables. We already understand how this works, the testing of the model is already conceptually understood, and it works.² There are just a couple things to clarify.

18.1.1 MATHEMATICAL MODEL As with the simply binary dependent variable case, let us layout the mathematical background to the nominal dependent case. As in the binary dependent variables case, we are actually modeling the underlying probabilities of each of the outcomes. Also, as in the binary dependent variable case, the probabilities sum to one.

Thus, if we let p_j be the probability that category j is selected, then the following two conditions must hold:

$$0 < p_j < 1 \quad \text{for all } j \in \{1, 2, \dots, J\} \quad (18.1)$$

$$\sum_{j=1}^J p_j = 1 \quad (18.2)$$

Condition (18.1) must hold because we are dealing with probabilities bounded by 0 and 1, and Condition (18.2) holds because one of the J possible outcomes must happen. In the binary case, our two probabilities were p and $1-p$, which satisfies the second condition by default.

¹Of course, there may be a time when you are predicting Republican vote by examining an underlying level of conservatism. In such a case, Democratic–Republican would be ordered. Thus, it really depends on what you are predicting (as always).

²Usually. Nothing in statistics *always* is best. As you have seen by now, there are always methods that work better, but with trade-offs. The science here is to be aware of the strengths with the weaknesses and balance them to get closer to the true process you are trying to model.

When we generalize the binary case, we need to select an appropriate probability distribution—one that can model J possible outcomes with J different probabilities. That distribution is called the multinomial distribution.³ The probability density function for the multinomial distribution in the general case is

$$f_X(x) = \frac{n!}{x_1!x_2!\cdots x_J!} p_1^{x_1} p_2^{x_2} \cdots p_J^{x_J} \quad (18.3)$$

Here, x_i are non-negative integers and $\sum x_j = n$. The expected value of this distribution for a specified outcome is

$$\mathbb{E}[X] = np_j \quad (18.4)$$

Here, n is the total number of trials and p_j is the probability of outcome j .

Example 18.1. Let us illustrate this with a typical “rolling a die example.” Assuming that the die is fair, then the probability of rolling a \square on the die is $\frac{1}{6}$. If we roll a fair die 100 times, what is the expected number of \square s?

This is a multinomial experiment. There are a fixed number of possible outcomes (six), the probabilities of each outcome are constant (they do not change as we roll the die), and the probabilities sum to one. As such, we know the expected value is

$$\mathbb{E}[X] = 100 \times \frac{1}{6} \approx 16.67$$

Thus, we expect (on average) to roll 16.67 \square s in 100 rolls of a fair die.

As we have a formula for our expected value, we have our mechanism for estimating the several p_j : in an experiment (or set of data), count the number of times outcome j occurred and divide by the total number of trials (or records). This is actually the Maximum Likelihood Estimate for p_j . Thus, our linear predictor is

$$\text{logit}(p_j) = \beta_{j,0} + \beta_{j,1}x_1 + \beta_{j,2}x_2 + \cdots + \beta_{j,k}x_k$$

Notice that this linear predictor has $k + 1$ parameters to estimate for each of the j categories. Thus, you will need more than $j(k + 1)$ pieces of data

³Recall that the distribution in the *binary* case was the *binomial* distribution.

to fit it. There are ways to reduce the dimensionality of the problem (reduce the number of parameters in need of estimation); however, these are beyond the scope of this book.

We need the logit link (or something just like it) to force our linear predictions to be in the range $p_j \in (0, 1)$. As any link that maps $L : \mathbb{R} \rightarrow (0, 1)$ is acceptable. We could use the log-log link, the complementary log-log link, or the probit link. As before, the choice of the link function is largely a matter of tradition. If you deviate from tradition, the burden of proof is on you to justify the selection. Furthermore, the differences are usually slight. If the differences are large, then there is something wrong with your research model. Because of this, it would behoove you to fit your research model using a couple different (appropriate) link functions to help determine the stability (robustness) of your results.

Note: Thus, there are two things that you need to take away from this discussion: First, we are able to fit the entire model at once because we have a distribution that can produce the necessary nominal results. Second, we model the underlying probabilities (like in the binary case), not the actual outcomes, as usual.

To see this in action, let us look at an extended example.

Example 18.2. The General Social Survey (GSS) at the University of Chicago conducts an extensive survey of adult Americans every year. The data is freely available.⁴ In this small subset of the data, `gssocc`, I would like to predict a person's occupation category (`occ`) based on race (`white`), years of education (`ed`), and years of experience (`exper`).

Before getting started, let us examine the variables involved. The race variable is binary, with a '1' representing the person identifying as 'white' and a '0' otherwise. As a side note, this is a race variable, not an ethnicity variable. Thus, Hispanics may self-identify as either white or non-white. Also note that this is a self-identification variable; that is, the individual being surveyed decided his or her reported race. Looking at a frequency count, a full 91.69% of the respondents stated they were white. This is significantly higher than the population at large, where approximately 80% of Americans are white. When we do the final analysis, we need to keep this in mind, as it is not necessarily representative of the nation as a whole.

⁴The data can be accessed from <http://www.norc.uchicago.edu/GSS+Website/>.

The median number of years of education in the sample is 12 years, which corresponds to graduating from high school. The mean number of years is 13.09, which indicates the sample is right skewed (mean is larger than the median). Furthermore, it is interesting to note that the first quartile is also 12 years. This indicates that at least 25% of the sample only graduated from high school. Digging a little, we find that 32.3% of the sample only graduated from high school. Additionally, 23.4% of the sample received a bachelor's degree or more, which is close to the population (27% have received a bachelor's degree or higher). Finally, 18.7% of the sample did not graduate from high school, which is close to the 15% estimate of the population. From this, it appears as though the sample is representative of the population in terms of educational attainment.

The third independent variable is the years of experience in the job. There are no general statistics for the population, so we will have to make a large assumption that the sample represents the population.⁵ In the sample, the years of experience varies widely, from 2 to 66 years. The median is 17 years and the mean is 20.5 years. Thus, the sample is also right skewed. This makes sense as this is a count variable. Count variables tend to be right skewed as they cannot take on negative values. In fact, there is nothing in the distribution of the experience variable that looks wrong. With that said, however, one still needs to mention the caveat.

Looking at the correlations amongst the independent variables can help us avoid any unpleasant surprises due to collinearity and multicollinearity. The correlation matrix (Table 18.1) does not show any hint of multicollinearity. In fact, this correlation matrix suggests that these three variables are *effectively* independent of each other.⁶

Finally, let us note that there *may* be an inherent ordering in some of the jobs (White Collar greater than Blue Collar), but not for all five of the categories. As such, this is definitely a candidate for nominal regression.

NOMINAL REGRESSION: Now, let us model the outcome variable with the three independent variables. Actually, we need to step back and really think about what we mean by 'model the outcome'. Do I want to predict the probability

⁵This was a safe assumption with respect to the education variable, but not with respect to the white variable. As such, it needs to be mentioned that you are unable to check the representativeness of the experience variable.

⁶Pearson's product-moment correlation test indicates that the correlation between education and experience is statistically significant at the $\alpha = 0.05$ level ($t = -5.2152, df = 335, p \ll 0.0001$). However, the coefficient of -0.2740 is a low level of correlation.

	White	Education	Experience
White	1.0000	0.0243	-0.0794
Education	0.0243	1.0000	-0.2740
Experience	-0.0794	-0.2740	1.0000

Table 18.1: Correlation matrix for the three independent variables in the example, from *gssocc* data: `cor.x(gssocc)`.

	Estimate	Std. Error	z-value	p-value
Intercept	3.1036	1.0110	3.07	0.0021
White	0.7090	0.6213	1.14	0.2538
Years of education	-0.3721	0.0640	-5.81	0.0000
Years of experience	-0.0259	0.0113	-2.30	0.0215

Table 18.2: Results from the GLM (using the binomial family and the logit link) predicting whether or not a person is a blue collar worker. The AIC for this model is 304.75.

that a person will be Blue Collar given the x-variables? Or: Do I want to predict the job category given the input variables? These are different questions. They require slightly different methods.

The first question actually asks a binary question: What is the probability that a person will be Blue Collar (compared to *all* of the other job categories)? This is very much like the questions asked in Chapter 16. Here, the dependent variable takes on values 1 (Blue Collar) and 0 (not Blue Collar).

To answer *this* question, we need to create a variable called `bluecol` as an indicator variable for Blue Collared-ness. Thus, the model we fit will be

```
bluecol ~ white + ed + exper
```

We would fit it using a Generalized Linear Model, a binomial family, and a logit link. The results of the regression are in Table 18.2. From this model, we can perform all of the goodness of fit measures from Chapter 16.

Looking at the results from running the model, we see that greater levels of education and greater levels of experience are associated with a lower probability of being a blue collar worker. For Bob, an individual who responded that he was white, had 20 years of education, and 10 years of experience in their current job, the probability of being a blue collar worker is approximately 2% (as compared to not being a blue collar worker).

Note: This last part is subtle, but extremely important. Here is why: What is the probability that Bob is a white collar worker? If we do the same steps above, we get that the probability that Bob is a white collar worker (as compared to not being a white collar worker) is 13.1%. Similarly, if we continue performing separate logistic regressions, the probability that Bob is a professional is 96.9%; menial, 2.3%; and craft, 7.9%.

Note that all of these probabilities add up to more than 100%. There is something wrong here, since the probability that Bob holds one of these five job types cannot be greater than 100%.

The problem is that we kept changing the base category. In Chapter 16, we never mentioned the need to specify the base category since it always defaulted to the opposite of what we were modeling. In other words, we were actually measuring the probability of an event as compared to the probability of ‘not the event’. This ensured that the probabilities always added up to 100%. Within each of the above five regressions, if we added the probability of the event that Bob holds job type X with the probability that Bob holds job type not X, we always get 100%.

The lesson: comparing probabilities of events is not as easy as when we were only working in the binary realm. It is doable — easily so, with one small change. We need to select a base category that does not change throughout our analysis. The choice is up to you, as all choices are equally acceptable.

Since we can select any job type as our base, let us select Blue Collar, since it is first in our dataset, and since the level in the first record is the default base for most programs. (We will see shortly how to switch between the bases.)

To perform this modeling, you will have to load the `nnet` package, which comes with your base distribution of R. Once loaded with the `library(nnet)` command, to fit the better model, use the R command

```
multinom(occ ~ white + ed + exper)
```

Because of the large amount of output, the regression table is structured slightly different. The coefficients (in logit units) and the standard errors are still presented. The statistical significance is not. However, a quick rule of thumb is that the variable is statistically significant (at the $\alpha = 0.05$ level) if the parameter estimate is more than twice the standard error. Table 18.3 presents the output from modeling the data in the form given in the output.

Coefficients:				
	Constant Term	White	Education Level	Experience
Craft	-1.8328	-0.7642	0.1933	0.0230
Menial	-0.7412	-1.2365	0.0994	-0.0074
Prof	-12.2595	0.5376	0.8783	0.0309
WhiteCol	-6.9800	0.3349	0.4526	0.0299
Std. Errors:				
	Constant Term	White	Education Level	Experience
Craft	1.1861	0.6324	0.0775	0.0126
Menial	1.5195	0.1996	0.1023	0.0174
Prof	1.6681	0.7996	0.1005	0.0144
WhiteCol	1.7144	0.9340	0.1023	0.0153

Table 18.3: Results of the multinomial regression. Note that the p -values are not provided. To determine which independent variables are statistically significant for predicting the dependent variable levels, divide the coefficient estimate by the standard error. If that ratio is greater than 2, then the variable is statistically significant at the $\alpha = 0.05$ level.

Note that one of the five job types is missing: Blue Collar. This is because all of the probabilities are measured *with respect to* Blue Collar. Thus, these percentages are directly comparable (after transforming from logit units).

R is nice in that if you `predict` on a multinomial model, it will give you the category with the highest probability, by default. Thus, according to this model, Bob will most likely be a Professional (which was our conclusion above). If we want the probabilities for each of the possible job types for Bob, we need to add a `type="probs"` parameter to our function call:

```
predict(model.mnl, newdata=BOB, type="probs")
```

Such a call gives us the following probabilities (which sum to one, as they should):

BlueCol	Craft	Menial	Prof	WhiteCol
0.0020	0.0091	0.0020	0.9565	0.0304

Base switching: If you wish to switch your base category, there are two options. First, you can subtract the parameter estimates of the new base from all the other bases. Thus, if we want to change the base from Blue Collar to Professional, we would subtract the Professional parameter estimates from

the other parameter estimates. So, for example, the White Color estimates with Professional as the base will be $-6.9800 - -12.2595 = 5.2795$. Unfortunately, the standard errors are not so easily calculated — or at all reasonably calculable by hand.

Also unfortunately, most statistical programs require you to physically re-order the data to select a different base; most programs use the level of the first data point as the base category. R does allow you to switch among the bases without having to physically alter the data. Unfortunately, the method is rather arcane. Fortunately, the `RFS` package has a function, `set.base()` that allows you to change the bases much more easily.

Thus, to set `craft` as the base, we would use the command

```
occ <- set.base(occ, base="craft", data=gssocc)
```

I leave it as an exercise to rerun the analysis with `craft` as the base. Check that the parameter estimates follow the above observation.

INTERPRETATION: The interpretation of the coefficients (parameter estimates) is the same as for the binary dependent variable case. Just remember that the coefficients are in logit units. In R, however, this library does not require you to back-transform. To remember this, just look at the output — it is in proportions already (a quick check is that they sum to one).

GOODNESS OF FIT: The first check of the goodness of the model is the relative accuracy (see also Section 16.5). The accuracy is the number of correct predictions divided by the number of cases. The relative frequency divides this number by the accuracy of always selecting the modal category. For this dataset, the modal category is Professional, with 112 out of 337 cases belonging to Professionals. Thus, the relative accuracy is $\frac{169}{337} / \frac{112}{337} = 1.509$. Thus, this model improves accuracy by 50% over the null model. Is this good? It depends on your other models.

As Maximum Likelihood Estimation is used, the Akaike Information Criteria score is also reported. For this model, $AIC = 885$. Is this good? Again, it depends on your other models. In other words, model comparison needs another model. I leave it as an exercise to see that the null model has $AIC = 1027$. Thus, our model is much better than the null model.

Now that we have looked at our model, let us look at the parameter estimates. According to our model, Whites have a higher probability of being Professionals and White Collar workers than they are to be Craft or Menial

laborers. As for education, higher levels of education are associated with higher odds of being a Professional or a White Collar worker (both of these are statistically significant) than being a Blue Collar worker. Finally, years of experience are not a statistically significant predictor of job type, as none of the coefficients are statistically significant (coefficient / standard error > 2).⁷

So, we have a picture of Professionals and White Collar workers, when compared to Blue Collar workers: they are White and well educated. Not an earth-shattering conclusion, but it is encouraging to see that our conclusions do seem to reflect reality.

18.2: Ordinal Dependent Variable

Another variety of categorical dependent variables is ordinal. A variable is ordinal if it is categorical *and* the categories have an underlying order to them. Examples include movie ratings (number of stars), hurricane intensity, and so forth.

There are actually at least four ways of handling ordinal dependent variables:

1. Treat them as nominal. This allows us to fit ordinal data using previous techniques. Unfortunately, it is inefficient as it ignores important aspects of the data itself.
2. Treat their cumulative level as nominal. If the ordinal variable takes on values 1 – 5, then create nominal variables corresponding to Level 1, Levels 1 and 2, Levels 1–3, Levels 1–4, and Levels 1–5. This preserves much of the underlying information, but allows us to fit it using a previous method.
3. Assume that there is an underlying continuous process that you wish to fit. The ordinal nature is just several threshold values along the possible values. This reduces to a pseudo-OLS, where you also need to fit the threshold values, not just the slopes and intercepts. Using Maximum Likelihood methods, this is trivial to solve.

⁷This rule of thumb comes from the fact that in a Normal distribution, the ratio needs to exceed 1.96 to be statistically significant at the $\alpha = 0.05$ level. These parameter estimates are not guaranteed to be Normally distributed. As such, the rule of thumb is to be more conservative. Even with the rule of thumb, do not bet the farm.

4. Pretend that the ordinal values are continuous and fit it using ordinary least squares or one of its offsprings. This has the advantage of being easily fit.

Three of these ways have already been discussed, and you are quite adept at using them (Options 1, 2, and 4). Only the third option is completely new to you. This chapter focuses on how to fit Option Three.

18.2.1 OPTION THREE Let us assume that there is an underlying continuous process. We only experience this process through the ordinal variable. This is very similar to how we first looked at binary variables: underlying process exhibited only in the 0/1 outcomes. Here, there is more than just the one threshold (which traditionally defaulted to $\lambda = 0.500$). Thus, we have two sets of parameters to fit. The first is the parameters which describe the process. The second is the position of those threshold values. Without going into the details, we will use Maximum Likelihood Estimation as our fitting method because it has many nice properties. Thus, our underlying process is

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Our thresholding process is shown in Figure 18.1. The line represents the underlying continuous process that you are trying to model. The A, B, C, and D represent the observed ordinal values. The threshold values, τ_1 , τ_2 , and τ_3 are the values that separate the observed ordinal values.

This model is very straight forward and understandable. The fitting is also straight forward. The results are also straight forward.

Example 18.3. Let us use some more data from the GSS. This data explores the ‘warmth of feeling’ the respondent has for the president. The demographic information is the gender (male), the race (white), the age, and the number of years of education (ed). The response variable has four ordered levels: Strongly Disagree (SD), Disagree (D), Agree (A), and Strongly Agree (SA). Our goal is to explain a person’s feelings toward the president based solely on demographic information.



Now, let us fit this data with ordinal regression. The command in R is `polr()`, which requires the `MASS` package. The actual command is

```
polr(warm ~ male + white + age + ed )
```

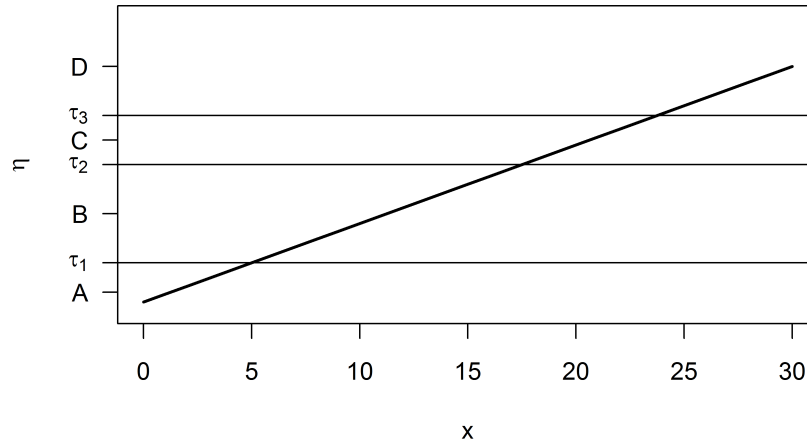


Figure 18.1: Schematic diagram of the thresholding process. The line represents the linear continuous process. The τ s represent the threshold values. A, B, C, and D represent the ordinal outcomes.

Variables:	Value	Std. Error	t-value
Woman	0.743	0.078	9.50
White	-0.400	0.118	-3.39
Age	-0.020	0.0024	-8.17
Years of Education	0.098	0.013	7.52
Thresholds:			
SD — D	-1.700	0.237	-7.18
D — A	0.111	0.233	0.48
A — SA	1.979	0.236	8.37

Table 18.4: Result of ordinal regression in R.

This command will give the coefficients of the underlying linear regression and the threshold values separating the four categories. From Table 18.4, we see that the equation for the underlying linear process is

$$\eta = 0.743 \times \text{Woman} + -0.400 \times \text{white} + -0.020 \times \text{age} + 0.098 \times \text{ed}$$

The thresholds are also listed. The threshold between Strongly Disagree and Disagree is at $\tau_1 = -1.700$. The threshold between Disagree and Agree is $\tau_2 = 0.111$. The threshold between Agree and Strongly Agree is $\tau_3 = 1.979$. Thus, to calculate our prediction, we calculate the prediction based on the linear model, η , and compare that value to the intervals described by the thresholds. Thus, for Bob, who is Male, White, 40 years old and has 20 years of education, we have

$$\eta = 0.740 \times 0 + -0.400 \times 1 + -0.020 \times 40 + 0.098 \times 20 = 0.76$$

As $\eta = 0.76$, we have our prediction that Bob agrees with the president. If we actually want probabilities that Bob Strongly Disagrees, Disagrees, Agrees, or Strongly Agrees, we would have to back-transform using the inverse of the logit function. Or, we could just ask the computer to do it for us:

```
BOB = data.frame(male='Men', white='White', age=40, ed=20)
predict(model.oll, newdata=BOB, type="probs")
```

This gives the probabilities as

SD	D	A	SA
0.0785	0.263	0.429	0.229

Thus, it is far from certain that Bob agrees (or strongly agrees) with the president (although the probability he does is $0.429 + 0.229 = 0.658$).

ACCURACY: Finally, let us look at the accuracy of the model. I leave it as an exercise to show that the relative accuracy is 1.105, which indicates that the model is about 10.5% better than the null model (modal category is Agree). This is not a fantastic increase in accuracy, but we do know how certain demographics feel about the president: Whites tend to disagree, Males tend to disagree, older people tend to disagree, and lesser educated people tend to disagree.

Of course, we could have added in a quadratic education term to the model to see if both the highly educated and the lesser educated both support the president. I also leave this as an exercise to show that there is no evidence of this. Thus, we can conclude that the relationship between education and presidential support is linear.

18.3: Conclusion

In this chapter, we examined the special issues behind fitting dependent variables that are either nominal or ordinal. Nominal dependent variables are still basically fit with a series of logistic (or other link) regressions. The alteration comes about because we need to keep the same base category throughout in order to make our results comparable.

The ordinal dependent variable can be fit using a technique similar to the previous chapter: fit an underlying linear function, then create thresholds to divide a constant function into an ordinal response.

In both cases, predictions in \mathbb{R} follow the typical structure, with the addition of being able to just predict the outcome category or being able to predict the probabilities associated with the case fitting in each bin.

18.4: R Functions

In this chapter, we were introduced to several R functions that will be useful in the future. These are listed here.

18.4.0 PACKAGES

RFS

MASS

nnet

18.4.0 STATISTICS

multinom() This modeling function allows you to fit nominal dependent variables. Its structure is standard in that its main argument is the formula. In order to use the `multinom()` function, you must load the `nnet` library.

polr() This modeling function allows you to fit ordinal dependent variables when there is an underlying linear function that drives the process. In order to use the `polr()` function, you must load the `MASS` package.

predict(model, newdata) As with almost all statistical packages, R has a `predict` function. It takes two parameters, the model, and a dataframe of the independent values from which you want to predict. If you omit `newdata`, then it will predict based on the independent variables of the data itself, which can be used to calculate residuals. The dataframe must list all independent variables with their associate new values. You can specify multiple new values for a single independent variable.

set.base()

18.5: Extensions

This section offers suggestions on things you can practice from this chapter. Save the scripts in your Chapter 18 folder. For each of the following problems, please save the associated R script in the chapter folder as `ext0x.R`, where `x` is the problem number.

1. In Section 18.1.1, we fit a multinomial model to the `gssocc` data. The base used was 'Blue Collar.' Refit the model using 'Craft' as the base category.
2. Determine the AIC of the null model in Section 18.1.1.
3. As mentioned in Section 18.2.1, calculate the relative accuracy of the model of Example 18.3.
4. As mentioned in Section 18.2.1, add a quadratic education term to the model of Example 18.3 to see if both the highly educated and the lesser educated both support the president.

18.6: Applications

- Paul D. Allison and Nicholas A. Christakis. (1994) “Logit Models for Sets of Ranked Items.” *Sociological Methodology* 24: 199–228.
- John Fox and Robert Andersen. (2006) “Effect Displays for Multinomial and Proportional-Odds Logit Models.” *Sociological Methodology* 36: 225–55.
- Daniel Carson Johnson. (1997) “Formal Education vs. Religious Belief: Soliciting New Evidence with Multinomial Logit Modeling.” *Journal for the Scientific Study of Religion* 36(2): 231–46.
- Mark R. Killingsworth and Cordelia W. Reimers. (1983) “Race, Ranking, Promotions, and Pay at a Federal Facility: A Logit Analysis.” *Industrial and Labor Relations Review* 37(1): 92–107.
- Alan B. Lowther and John R. Skalski. (1998) “A Multinomial Likelihood Model for Estimating Survival Probabilities and Overwintering for Fall Chinook Salmon Using Release: Recapture Methods.” *Journal of Agricultural, Biological, and Environmental Statistics* 3(2): 223–36.
- Christopher Winship and Robert D. Mare. (1984) “Regression Models with Ordinal Variables.” *American Sociological Review* 49(4): 512–25.
- Judith E. Zeh, Daijin Ko, Bruce D. Krogman and Ronald Sonntag. (1986) “A Multinomial Model for Estimating the Size of a Whale Population from Incomplete Census Data.” *Biometrics* 42(1): 1–14.

18.7: References and Further Readings

- B. R. Bhat and N. V. Kulkarni. (1966) "On Efficient Multinomial Estimation." *Journal of the Royal Statistical Society. Series B (Methodological)* 28(1): 45–52.
- Zhen Chen and Lynn Kuo. (2001) "A Note on the Estimation of the Multinomial Logit Model with Random Effects." *The American Statistician* 55(2): 89–95.
- Jean-Yves Dauxois and Syed N. U. A. Kirmani. (2003) "Testing the Proportional Odds Model under Random Censoring." *Biometrika* 90(4): 913–22.
- Byung Soo Kim and Barry H. Margolin. (1992) "Testing Goodness of Fit of a Multinomial Model Against Overdispersed Alternatives." *Biometrics* 48(3): 711–19.
- Bercedis Peterson and Frank E. Harrell, Jr. (1990) "Partial Proportional Odds Models for Ordinal Response Variables." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 39(2): 205–17.
- G. A. F. Seber and S. O. Nyangoma. (1974) "Residuals for Multinomial Models." *Biometrika* 87(1): 183–91.
- M. Stone. (1974) "Cross-Validation and Multinomial Prediction." *Biometrika* 61(3): 509–15.
- Y. K. Tse. (1987) "A Diagnostic Test for the Multinomial Logit Model." *Journal of Business & Economic Statistics* 5(2): 283–86.