Please Read the Abstract

By

Abstract Writer

Submitted to the Department of People who read Abstracts and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members

MEMBER 1, Chairperson

MEMBER 2, Occasional Visitor

MEMBER 3

MEMBER 4, The One Who Never Answers Email

The One with an Extra Long Name

The Fifth Beatle

Date defended: October 02, 2016

The Thesis Committee for Abstract Writer certifies
that this is the approved version of the following thesis :

Please Read the Abstract

MEMBER 1, Chairperson

Date approved: October 06, 2016

Abstract

Hopefully, you are reading "ku-thesis.pdf". If you are reading a file with some other name, then there's been a goof up. You might have found this file by itself, but it should also be available inside a larger collection. The whole collection is available in various formats, but probably the most convenient will be "KU-thesis-20151102.zip". Versions are available at `http://pj.freefaculty.org/guides/Computing-HOWTO/KU-thesis` and also in the pages hosted by the office of the Graduate Dean.

This is the November 2015 edition of the KU dissertation template. It goes along with a new version of "kuthesis.cls" (version 2.2). We have made a few little changes that make it easier to adjust committee member names and roles.

The other difference is that we have turned on hyperlinks in the PDF output. We checked with the KU Graduate Dean's office and they said there is no policy prohibiting the use of hyperlinks, but there is also no guarantee from the dissertation hosting institution that they will either work or not cause problems down the line. We believe there is not too much danger here, this feature is easy to turn off if need be. Also, it is not entirely clear what color the links must be inside the document, so we take a conservative step here of making the links black, so they appear as simple text. The default settings for the hyperref package created ugly pink links with gross looking boxes around them, we can't have that. Users can change the color to blue. It is fairly easy to see how this can be done in the document preamble. Look for something like this "`urlcolor={black},citecolor={black},allcolors={black}`" and think of Henry Ford's famous comment, "You can have a car in any color you like, as long as its black."

Other than that, its the same thesis template.

Oh, wait. There's one more change. This Abstract is even more welcoming and helpful. You will love using LaTeX to prepare your thesis. The power to keep content structured, with a systematic framework for cross references among figures, tables, equations, and sections, will significantly improve your quality of life. This will be much better than using an ordinary "text editor."

Watch Out! Before you try to edit this and enhance it (and put your name on it), please **STOP**! Recompile this document exactly as it is before you make changes. Unzip the collection, consider installing the class file as described below, then try to compile `thesis-ku`. That zipped collection has everything in the right places, don't re-arrange directories yet. The document won't compile if the source code and the auxiliary material become inconsistent. Most importantly, the package includes "`kuthesis.cls`", the LaTeX class file that makes the whole thing work. If you downloaded the Zip file that includes the whole collection, you should be good to go.

Test your LaTeX setup by compiling "`thesis-ku.tex`". Until you reproduce the PDF you are reading now, you are not ready to start making changes. People often come forward with questions and our first question for them is "did you compile (without error) `thesis-ku.tex`?" Until we hear those people say "yes", we are generally unable to help with their revised documents. That's the only way we will know if your LaTeX setup is ready to go.

If you have used LaTeX before, you will find this more or less self explanatory. There's a LaTeX file "`thesis-ku.tex`". It is a "master" document, it links together the content of several files. Compile that with **`pdflatex`** and you will know soon enough if all is well. The class file that makes this go, "`kuthesis.cls`", is saved in the current working directory. It is not necessary to install it into your broader LaTeX distribution in order to build this test case. The only reason to install kuthesis.cls in your larger setup is to silence warnings from editors that don't remember to look in your current

directory. If you install `kuthesis.cls` in your system, you don't need to keep a copy of it with your dissertation. That may be convenient.

If you have never used LaTeX before, please don't make writing your dissertation your first LaTeX project. Start with some simpler example documents. Write some letters to your loved ones, experiment a little bit. In the KU Center for Research Methods and Data Analysis, there are occasional workshops for people who want to get started with LaTeX. There are many "how to get started with LaTeX" websites. The one offered by the author of this abstract is `http://pj.freefaculty.org/latex`.

How To Customize This and Write a Dissertation.

This PDF document is produced from a source document. If you are using a LaTeX editor like TexShop, Tex Works, or TeX Studio, the source document you need to edit and compile is "`thesis-ku.tex`." If you plan to use LyX, the place to start is by opening "`thesis-ku.lyx`" in LyX's editor. This document was prepared with version 2.1 and, if you are using an older version, you should update.

The first user customization is to insert your name and the names of some committee members. The document has a top section called a "preamble." (If using LyX, click the Document menu, then Settings, and look down). In the preamble, you will find the following block:

```
% Set the title and author info
\title{Please Read the Abstract}
\author{Abstract Writer}

\dept{Department of People who read Abstracts}
\degreetitle{Doctor of Philosophy}
\papertype{Dissertation} %or Thesis
%% It is vital to have 7 entries, even
%%      if some are empty for committee and role
```

```
%% I mean, it is vital to leave the empty place
%%      holders, the empty braces {}
\committee{MEMBER 1}{MEMBER 2}{MEMBER 3}{MEMBER 4}{The
    One with an Extra Long Name}{The Fifth Beatle}{}
\role{Chairperson}{Occasional Visitor}{}{The One Who
    Never Answers Email}{}{}{}
```

Here you see the most important template updates for October, 2015. We have newly enhanced `\committee` and `\role` environments. Now, we allow users to specify up to 7 committee members and each one can have a different role. Why provide all of this luxury? There were unexpected requests from students. The original KU Thesis template had only "Chair" for the first member. Users asked for a secondary title "Co-Chair" for member 2. Then they wanted to change the title of the first member to "Lead Chair". Then they wanted to add a the third Co-Chair, and so forth.

Now we accommodate those requests by leaving blank spots for up to seven committee members and seven roles. It is vital for you to leave the blank place holders {} if you don't want to list additional members or roles. This example document shows some funny titles, we do that only to prove it is possible. Generally, there is only a Chair, possibly a Co-Chair. This example has more roles in this example for fun. And to make fun of a professor. You can change them to {}.

The other thing to change is the date section. Lower in the preamble, you should see a place to insert some dates. Originally, we told students "just write those dates in by hand, like we did back in the olden days." But they kept saying they wanted them type set, so here they are. If a date is currently unknown, just leave a blank inside the brackets

```
\datedefended{October 02, 2016}
\dateapproved{October 06, 2016}
```

After you make those changes in the preamble, compile the document.

Setup: Consider Installing kuthesis.cls This is not truly necessary, programs like pdflatex will (probably) find kuthesis.cls if you have it in the same directory as the document. On every LaTeX system I've ever seen, a latex processor will look in several places for "`kuthesis.cls`" while it processes thesis-ku.tex. I believe, with 97.5% confidence, it will find "`kuthesis.cls`" if you keep it in the same directory as your document. If you want to use kuthesis.cls for several documents, it is necessary to install "`kuthesis.cls`" into your LaTeX file structure. That is generally an easy process, but it is different on every type of computer. The CRMDA staff has recently prepared a guide for this, entitled "How Users can Install LaTeX packages without Help from System Administrators". We have included a pre-release snapshot in the current folder ("`32.latex_config.pdf`"). The gist of this is as follows:

1. Copy "`kuthesis.cls`" into a directory in your LaTeX distribution. Usually (almost always), you will find a directory component that ends in "tex/latex" and you can put kuthesis.cls in there. The only puzzle is finding the right place, and that's why you should review our guide.
2. Run "texhash" or (or whatever your LaTeX distribution provides) make the LaTeX distribution take notice of your new file. On Mac and Linux systems, we run "`sudo texhash`" to get this done as the administrator, but if you install "`kuthesis.cls`" in your home user directory, the `sudo` should not be necessary.

Does this work with LyX?

Yes! I prepared this document with LyX! I like LyX. LyX is a pleasant-to-use editor, it can protect you from some of the details that ordinary LaTeX will present. LyX is free to use on all major computing platforms, including Linux, Macintosh, and Windows. LyX can be downloaded at `http://www.lyx.org`. Windows Users: We have done this many times. The best route is to take the large package they offer called the

"installer bundle", which includes the MikTEX distribution of LATEX. Unless you have a LATEX distribution already, in which case you don't need the whole bundle.

Why does LYX work? When the user asks LYX to preview the PDF output, a two step process happens behind the scenes. A file named "`thesis-ku.tex`" is created in a temporary directory, and then a program like **`pdflatex`** is asked to turn that into a PDF document. LYX will then send the PDF output to a PDF viewer, such as Yapp, Evince or Acrobat.

A warning: there will be warnings! The LYX user who opens thesis-ku.lyx will be slapped in the face with a series of warnings. These warnings will be

1. We can't find the class file "`kuthesis.cls`" and
2. We can't find the layout file "`kuthesis.layout`"

Users will generally receive those annoying warnings, they pop up over and over! To reassure you, however

1. Don't worry, those are just warnings.
2. Try to compile the document anyway, it may work
3. You can install "`kuthesis.cls`" and "`kuthesis.layout`" so as to silence those annoying warnings. (After installation, it is necessary in LYX to run Tools -> Reconfigure.)

We have already discussed the "kuthesis.cls" installation process. Next, the layout is considered.

What about the LYX layout file? This is truly superficial. LYX works, it will compile, but it is warning you that it might beautify its presentation for you if you give it a layout file. The warning can be silenced by copying our file "`kuthesis.layout`" (in a directory stylefiles) into the layouts directory within your LYX configuration folder. The only problem is finding that folder.

On Linux, copy into the user HOME folder, under ".lyx/layouts".

On Windows, the right place will be something in the neighborhood of the AppData directory inside the user's home folder. Here's a way to find out. Run LyX, in the top menu choose Tools/Preferences. Change something trivial, then close the menu. That will force LyX to create a user folder somewhere in your personal HOME directory. Then search inside your user directory for the directories and files that LyX created. It should work to search for a directory named "layouts" (or any of the other LyX related configuration files, like "lyxrc.defaults". Probably, searching for "*lyx*" will be sufficient). In MS Windows 7, for example, I found the LyX configuration folder is "C:\Users\pauljohn\AppData\Roaming\lyx212\layouts". Caution: The AppData folder is hidden/concealed and, in order to find it in Windows Explorer, it is necessary to change the folder options to NOT hide protected files and folders.

On Macintosh OS X, we recently followed same procedure. The place to put a layout file was "/Users/pauljohn/Library/Application Support/LyX-2.1/layouts". Documentation for OS X El Capitan indicates that Apple has followed the Microsoft policy of hiding this folder, so the person who uses the Finder will have to change some preferences to show the directory.

If this is an important project, please consider installing Git or some other version management program.

It is difficult to avoid clutter if you have to rename your file every day. I see user directories full of files named, "my_diss.lyx", "my_diss.2.lyx", "my_diss_final.lyx", "my_diss_rejected.lyx" and so forth. It is very difficult to keep old and new versions straight unless you use a version manager. A version manager will handle that for you. You will just have one file "my_diss.lyx" and you will have the ability to compare old and new snapshots. The CRMDA has a more-or-less comprehensive, yet easy to follow guide for Git, "Git it Together".

If you need help, ask, but please understand

I don't have as much time as I would like to help users. I want you to use LaTeX. I think doing anything else is finger painting. I've watched 15 MA and PhD theses written with this template and I know a bright student will succeed and be happy with the result. If you have a laptop computer and you want to show me some errors, I can give you 5 or 10 minutes. But not much more...

CRMDA offers some workshops and I teach graduate courses in which we show how to use LaTeX. If you find something wrong with kuthesis.cls, I will be glad to try to fix it.

Compile Early, Compile Often

The truly bad part of using LaTeX is that the errors and warnings are, generally, not understandable. They are not understandable to me and I've been using LaTeX for about 20 years. I mean, well, they are simply too abstract and don't generally have accurate recommendations about what is needed to fix something. If you insert an illegal character, such as an "_" in a LaTeX box, the document will fail with this unhelpful message

```
Missing $ inserted.
Missing $ inserted.
Missing } inserted.
Extra }, or forgotten \endgroup.
...
I've inserted a begin-math/end-math symbol since I think
    you left one out. Proceed, with fingers crossed.
```

At best, the message will point to the line you inserted to cause the problem, but often the LaTeX system cannot guess where your mistake is, it can only say something is wrong. The best way to defend yourself: ***Compile Early, Compile Often.*** Don't make a lot of changes without trying to compile. It is easy to insert a feature that breaks the

document. Compile often so you know when things have gone wrong.

I urge you to keep your LaTeX document in Git so you can make frequent commits and it will be easy to stop back in time. This is one reasy why you should work on separate chapters, so if you break something, then you are more likely to find your mistakes.

I don't have time/ability to answer a million email questions about LaTeX or LyX. If you are stuck on a tough problem, I may be able to help, but if I have to exert myself, you should know I expect you to give me a t-shirt and/or add me to the acknowledgements in your dissertation.

Paul Johnson <pauljohn@ku.edu>

Prof. Political Science & Director, Center for Research Methods and Data Analysis

2016-03-28

Version History

Version 2.3, March 28, 2016. Section fonts were set as "Large", exceeding Graduate Office recommendation. Now they are "large"

Version 2.2, November 02, 2015. More flexible Committee member and role titles. Hyperlinks

Version 2.1, May 20, 2013. First, the Graduate School no longer exists, so we rephrase the front matter to refer to Graduate Faculty. Second, the page numbering after the front matter begins at arabic number 1.

Version 2, October 2, 2012. This is a minor update to allow insertion of dates in the document preamble.

And that's the end of the abstract.

Acknowledgements

I would like to thank all of the little people who made this thesis possible. Sleepy, Dopey, Grumpy, you know who you are.

Contents

List of Figures

List of Tables

Chapter 1

Ordinal Outcomes Regression

1.1 Introduction

This is my best effort to succinctly explain the theory behind the ordinal logistic regression model (with apologies to the probit model).

The main takeaway point is supposed to be this:

> The same data leads to different estimates from different programs. That happens because the ordinal model can be written down in several different ways. None of them are wrong, but they are different, and as a result the user must be cautious.

Estimates obtained from four different programs are offered in Table 1.1. If we line these up side by side, we see that estimates from one of the routines for R matches Stata (after chopping off the small differences in the decimals), while SAS appears to provide the "wrong sign" for the first row and the second procedure for R seems to provide the "wrong signs" for the second and third rows.

Table 1.1: Ordinal Regression Results

	R: polr	R: lrm	SAS	Stata
$\hat{b}_1$	-0.28	-0.28	0.28	-0.28
$\hat{\zeta}_1$	-4.24	4.24	-4.24	-4.24
$\hat{\zeta}_2$	-2.32	2.32	-2.32	-2.32

None of these are actually wrong, they are all correct *given the model they specified*. This the point at which the student may be tempted to give up. Please don't. I've worked very hard to clear this up in the following sections.

1.2 Extending the Logit Model to deal with Ordinal Dependent Variables

The easiest way to understand regression with ordinal dependent variables is to extend the "cumulative probability interpretation" of the two category model (Pinheiro & Bates, 2000).

In the two category model, y_i is 1 with probability

$$F(b_0 + b_1 X_i) = \int_{-\infty}^{b_0+b_1X_i} f(e_i)de_i \tag{1.1}$$

And, of course, the probability that y_i is 0 will be $1 - F(b_0 + b_1 X_i)$. The formula F is a "cumulative distribution function" (CDF), it represents the probability that a random variable e_i will be as small or smaller than $b_0 + b_1 X_i$. The function f is a "probability density function" (PDF), which represents the probability that e_i is equal to some particular value. This is illustrated in Figure 1.1. The "probability density function" f is defined from left to right and the possible outcomes are divided into two sets by the line drawn at $e_i = b_0 + b_1 X_i$. The area under the curve on the left side is the probability of getting a "yes" (or 1). The area on the right is the chance of a "no" (0).

Suppose y_i can have 3 values, 0, 1, and 2. (Keep in mind that this model can be written down in several ways. We tackle my favorite first, and then consider the others.) Leave the predictive part of the model ($b_0 + b_1 X_i$) the same, but we now introduce two new positive constants (Π_0 and Π_1) that divide the space. Considering Figure 1.2, it should be easy to see why some people call these new parameters "thresholds".

To summarize the effect of these new thresholds, we write down 1 equation for each possible outcome. My tendency is to write the thresholds as positive values like so:

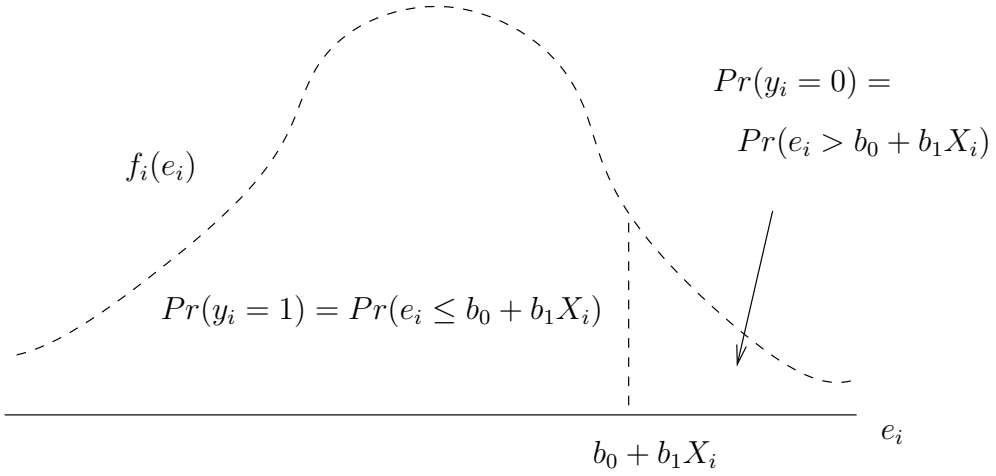


Figure 1.1: Dichotomous Outcome Variable

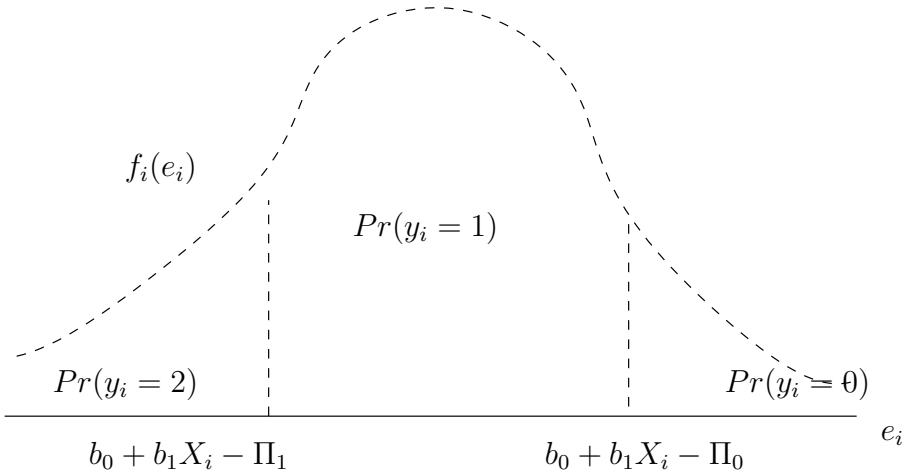


Figure 1.2: Ordinal Logit

$$y_i = \begin{cases} 2 & if\, b_0 + b_1 X_i - e_i \geq \Pi_1 \\ 1 & if\, \Pi_0 \leq b_0 + b_1 X_i - e_i < \Pi_1 \\ 0 & if\, b_0 + b_1 X_i - e_i < \Pi_0 \end{cases} \tag{1.2}$$

Note we don't really need 3 equations. If we have two, say $Pr(y_i = 0)$ and $Pr(y_i = 1)$, then the chance of ending up in the other category is $1 - Pr(y_i = 0) - Pr(y_i = 1)$.

In order to translate this into a model involving the cumulative probability distribution, rearrange so that the random variable e_i is by itself.

$$y_i = \begin{cases} 2 & if\, e_i \leq b_0 + b_1 X_i - \Pi_1 \\ 1 & if\, b_0 + b_1 X_i - \Pi_1 < e_i \leq b_0 + b_1 X_i - \Pi_0 \\ 0 & if\, b_0 + b_1 X_i - \Pi_0 < e_i \end{cases} \tag{1.3}$$

As in the dichotomous case, the probabilities of the various outcomes are calculated by use of cumulative probability. Rearrange 1.2 to convert these into probabilities of the individual outcomes.

$$\begin{aligned} Pr(y_i = 2) &= Pr(e_i \leq b_0 + b_1 X_i - \Pi_1) && = F(b_0 + b_1 X_i - \Pi_1) \\ Pr(y_i = 1) &= Pr(b_0 + b_1 X_i - \Pi_1 \leq e_i < b_0 + b_1 X_i - \Pi_0) \\ &= 1 - F(b_0 + b_1 X_i - \Pi_0) - F(b_0 + b_1 X_i - \Pi_1) \\ Pr(y_i = 0) &= Pr(b_0 + b_1 X_i - \Pi_0 < e_i) && = 1 - F(b_0 + b_1 X_i - \Pi_0) \end{aligned} \tag{1.4}$$

Note that any one category can be thought of as a "residual" category after the others have been assigned their shares. The middle category, $y_i = 1$, is left over if we "chop off" the outcomes on the left ($y_i = 2$) and the right ($y_i = 0$). We are left with the chance of ending up in the middle. In that sense, the probability of landing in the middle is equal to 1.0 minus the chance of a very small amount of random noise ($e_i \leq b_0 + b_1 X_i - \Pi_1$) and minus the chance of having a very large random noise ($b_0 + b_1 X_i - \Pi_0 < e_i$). Similarly, the chances of being in the top category equal 1 minus the

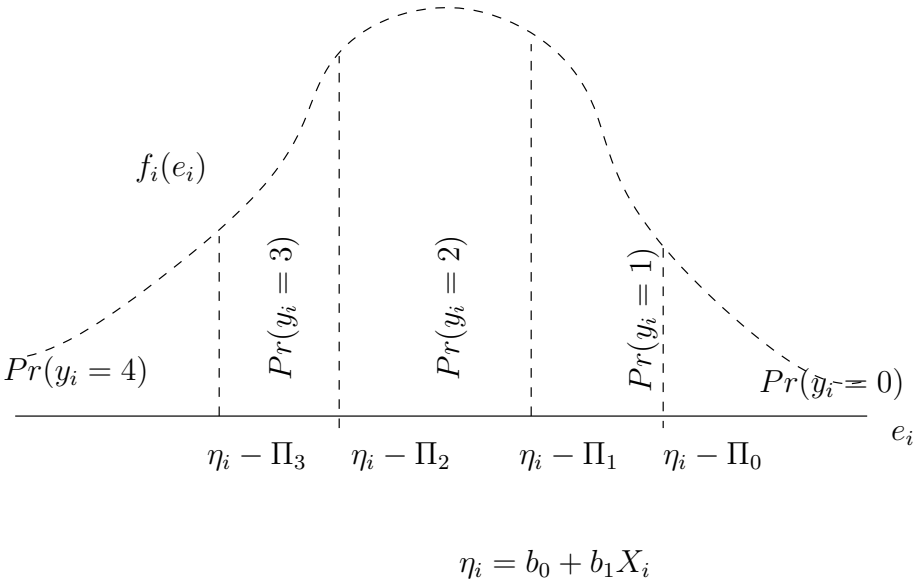


Figure 1.3: Ordinal Model with More Categories

chance of ending up in the lower categories.

Any probability distribution can be used for the random error e_i, the two most common being Logistic and Normal. If the Normal is chosen, it is customary to call this a "probit" model and the symbol for the cumulative distribution is usually $\Phi()$.

What if your dependent variable have more categories? Add more thresholds! See the example in Figure 1.3.

Chapter 2

Elementary Regression

2.1 Example

Here's what I would do to begin a regression exercise. I'm using the dataset "Prestige" that is in the car package for R.

```
library(car)

pdf(file="car.inc.ed.pdf", height=5,width=5, onefile=F,
    paper="special")
plot(income~education, xlab="Education", ylab="Income", main="",
    data=Prestige)
dev.off()
```

That plot is presented in the top panel in Figure 2.1.

A regression model Gelman et al. (2003)can be fitted to that scatterplot with R's lm function.

```
mymod <- lm(income~education, data=Prestige)
```

mymod is an "object", a complicated structure that contains a great deal of information. Observe:

```
attributes(mymod)
```

```
$names
[1] "coefficients"  "residuals"     "effects"       "rank"
[5] "fitted.values" "assign"        "qr"            "df.residual"
[9] "xlevels"       "call"          "terms"         "model"
$class
[1] "lm"
```

We can access values from mymod by 2 methods.

1. Direct access

```
> mymod$coefficients
> mymod$coefficients
(Intercept)    education
  -2853.5856     898.8128
```

2. Or by "accessor" methods.

```
> coef(mymod)
(Intercept)    education
  -2853.5856     898.8128
```

Some values are needed so regularly that a "summary" method is written to gather and summarize them.

```
> summary(mymod)
Call:
lm(formula = income ~ education, data = Prestige)
Residuals:
     Min       1Q   Median       3Q      Max
-5493.20 -2433.80   -41.92  1491.50 17713.14
Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2853.6      1407.0  -2.028   0.0452 *
education      898.8       127.0   7.075 2.08e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3483 on 100 degrees of freedom
Multiple R-squared: 0.3336,     Adjusted R-squared: 0.3269
F-statistic: 50.06 on 1 and 100 DF,  p-value: 2.079e-10
```

There are some functions to make more beautiful tables that are closer to the needs of journals. Years ago, I wrote a function called "outreg" and it would make a nice looking table. You can use that if you want, but if I were you, I would use the regression tables from either of these packages for R: "memisc" or "apsrtable". Since I already have memisc installed, I will demonstrate that:

```
> library(memisc)
> toLatex(mtable(mymod))
```

The LaTeX code can be pasted into a document in the same way that we handled cross tabulation tables. Please see Table 2.1

Observe:

The function "predict" can be used to obtain predicted values for example values of the input. One specifies a "newdata" option which must be a data frame, but we can create the new dataframe "on the fly" without too much trouble.

```
> predict(mymod, newdata=data.frame(education=c(8,9,10,11,12)))
       1        2        3        4        5
4336.917 5235.730 6134.543 7033.356 7932.169
```

Sometimes we want to plot the predictive line in the model, and that can be done in a variety of ways. One way is to find some points and then use them with the lines function.

```
> range(Prestige$education)
```

Table 2.1: My Regression Table

(Intercept)	−2853.586*
	(1407.039)
education	898.813***
	(127.035)
R-squared	0.334
adj. R-squared	0.327
sigma	3483.378
F	50.060
p	0.000
Log-likelihood	−975.609
Deviance	1213392025.001
AIC	1957.218
BIC	1965.093
N	102

```
[1]   6.38  15.97
> predict(mymod,
    newdata=data.frame(education=range(Prestige$education)))
        1         2
 2880.840 11500.456
> X1 <- range(Prestige$education)
> Y1 <- predict(mymod,
    newdata=data.frame(education=range(Prestige$education)))
> lines(X1,Y1)
```

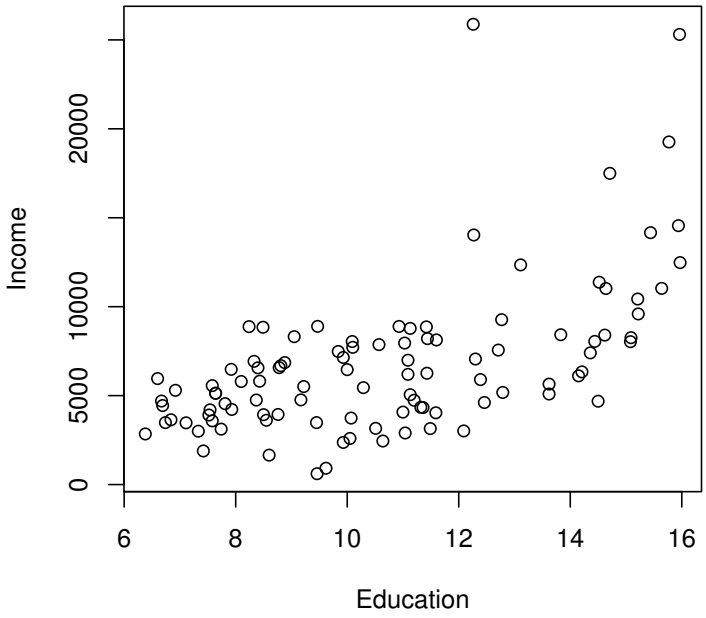
That adds a line to the figure, as seen in the bottom panel.

Because it is very common to add a regression line to a plot, the R function abline has been customized to do this for us automatically if you give it a regression object. The following commands were the ones I actually used to produce the figure.

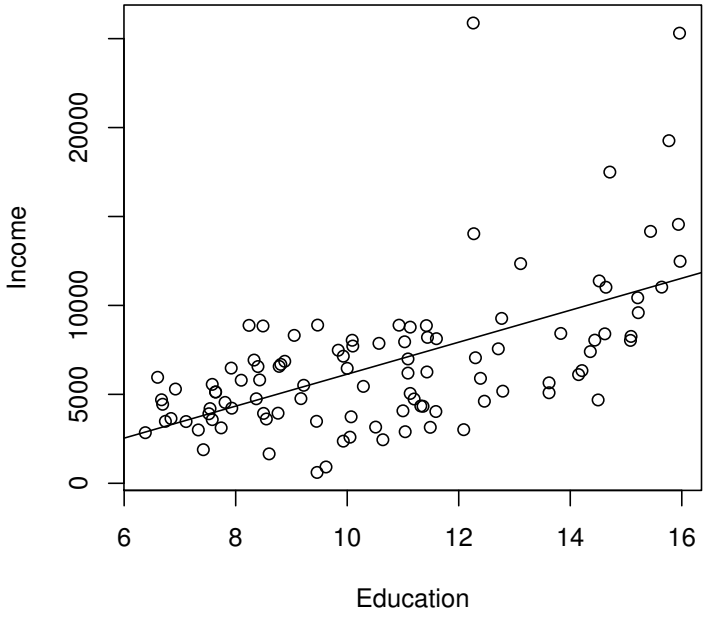
```
pdf(file="car.inc.ed-fit.pdf" height=5, width=5, onefile=F,
    paper="special")
plot(income~education, xlab="Education", ylab="Income",
```

```
    main="", data=Prestige)
abline(mymod)
dev.off()
```

I went to the trouble of illustrating the “old fashioned” way because I think it is possible to become too dependent on simplifying functions, especially when they don’t do exactly what you want.



(a) A Scatterplot



(b) Add the "regression line"

Figure 2.1: Income Depends on Education

References

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis, Second Edition*. Chapman & Hall, 2 edition.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer.

Appendix A

My Appendix, Next to my Spleen

There could be lots of stuff here