PLSC 503: "Multivariate Analysis for Political Research"

Bivariate Regression, IV: Stupid Regression Tricks
January 31, 2017

Introduction

The purpose of this class is to introduce you to and explain some "tricks" that one can (and nearly always should) use when doing regression analysis. By regression, I mean any sort of model in which a single response variable is some function of one or more covariates – so, while I'll be illustrating everything with bivariate regression, almost everything that follows is also applicable to multivariate OLS, GLM-type models (logit, probit, etc.), survival models, etc.

The day's discussion is organized around some general topics.

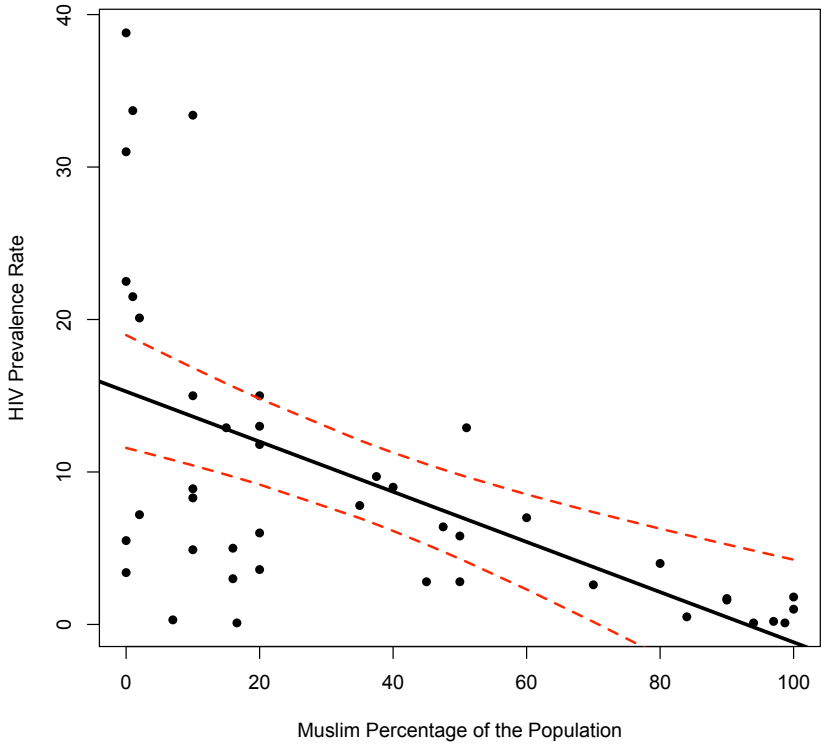
Topic #1: Scaling and Rescaling Variables

At the outset, understand that, because OLS is a linear estimator, any linear change in the scale of Y or X results in a linear change in the estimates that result. Let's consider some data from 43 African countries (during the year 2001), and think for the moment about the relationship between HIV/AIDS prevalence rates and the percentage of the population that is Muslim. That regression looks like this:

```
> africa<-read.dta("africa2001.dta")
> attach(africa)
> fit<-lm(adrate~muslperc)
> summary.lm(fit)

Call:
lm(formula = adrate ~ muslperc)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.2787     1.8322    8.34  2.3e-10 ***
muslperc     -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```

and the scatterplot looks like this:

Figure 1: Scatterplot of HIV/AIDS Rates on Muslim Population Percentage, 2001



This regression tells us that:

- We would expect the HIV/AIDS prevalence rate to be about 15.2 percent ($\pm$ 3.6 percent) when the Muslim population is zero.
- Each one-percent increase in the Muslim population would correspond to an expected decrease in the HIV/AIDS rate of about 0.16 percent ($\pm$ 0.07 percent).

Now, because the model is linear, any *linear* change to the scale of Y or X will yield a corresponding change in the estimated coefficients. In particular,

- Adding a *constant* to either Y or X will change the estimated *intercept* $\hat{\beta}_0$ of the resulting model. More specifically,
  - Adding a constant k to Y will change the estimated constant term by an amount equal to k (that is, to $\hat{\beta}_0 + k$), but will not affect its estimated standard error.
  - Adding a constant k to X will change the estimated constant by an amount equal to $-k\hat{\beta}_1$ (that is, to $\hat{\beta}_0 - k\hat{\beta}_1$), and will affect its standard error as well.
- Multiplying either Y or X by a constant will change the estimated slope $\hat{\beta}_1$ of the resulting model, and may also affect the intercept $\hat{\beta}_0$. Specifically,

- Multiplying Y by a constant k will result in an estimated intercept equal to $\hat{\beta}_0 \times k$, and an estimated slope equal to $\hat{\beta}_1 \times k$.
- Multiplying X by a constant k will result in an estimated slope equal to $\frac{\hat{\beta}_1}{k}$, but will not affect the estimated intercept (which remains $\hat{\beta}_0$).

- Any linear transformations of either Y or X will have no effect on the model's "fit," including the estimated R^2 or other summary statistics.

Intuitively:

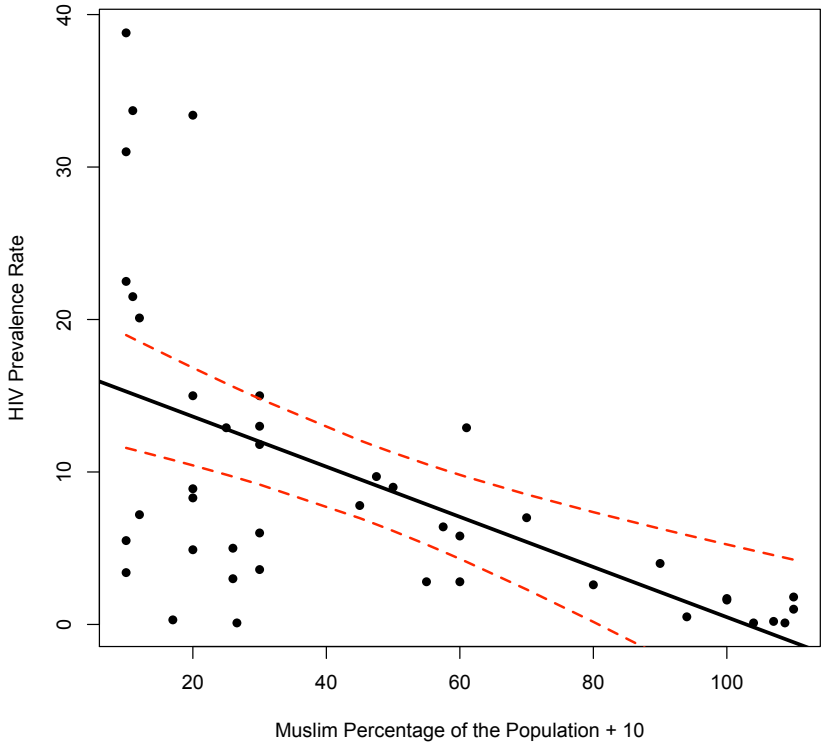
- Adding a constant to either Y or X changes the "location" of the relationship in the X-Y space, but not its "slope." Put differently, adding a constant "shifts" the relationship, but does not change its "shape."
- Multiplying Y or X by a constant "stretches," "shrinks," and/or "flips" the axis of the variable to which it is done, with the predictable effect on the coefficient estimates.

Consider two quick examples. First, what happens if we add ten to the `muslperc` variable:

```
> africa$muslplusten<-muslperc+10
> fit2<-lm(adrate~muslplusten,data=africa)
> summary(fit2)

Call:
lm(formula = adrate ~ muslplusten, data = africa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.9232     2.1152    8.00  6.6e-10 ***
muslplusten  -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```

The scatterplot looks like this:

Figure 2: Scatterplot of HIV/AIDS Rates on Rescaled Muslim Population Percentage



Notice that:

- The slope of the line remains the same, as does its estimated standard error.
- The estimate for the intercept changes; since we added ten to `muslperc`, the new intercept becomes 15.279 - (10 $\times$ -0.1644) = 16.923.
- The estimated standard error for the constant term changes as well. This is because the intercept is now "farther" from $\bar{X}$, and is also reflected in the broader range of the confidence interval for $\hat{Y}$ at $X = 0$.
- None of the summary statistics change (since we didn't muck around with Y).

For the second example, consider what happens if we multiply Y by -314:[1]

```
> africa$screwyrate<-adrate*(-314)
> fit3<-lm(screwyrate~muslperc,data=africa)
> summary(fit3)

Call:
lm(formula = screwyrate ~ muslperc, data = africa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4797.5      575.3   -8.34  2.3e-10 ***
muslperc        51.6       11.6    4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2600 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05

> anova(fit3)
Analysis of Variance Table

Response: screwyrate
          Df    Sum Sq   Mean Sq F value    Pr(>F)
muslperc   1 133908740 133908740  19.827 6.391e-05 ***
Residuals 41 276912962   6753975
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

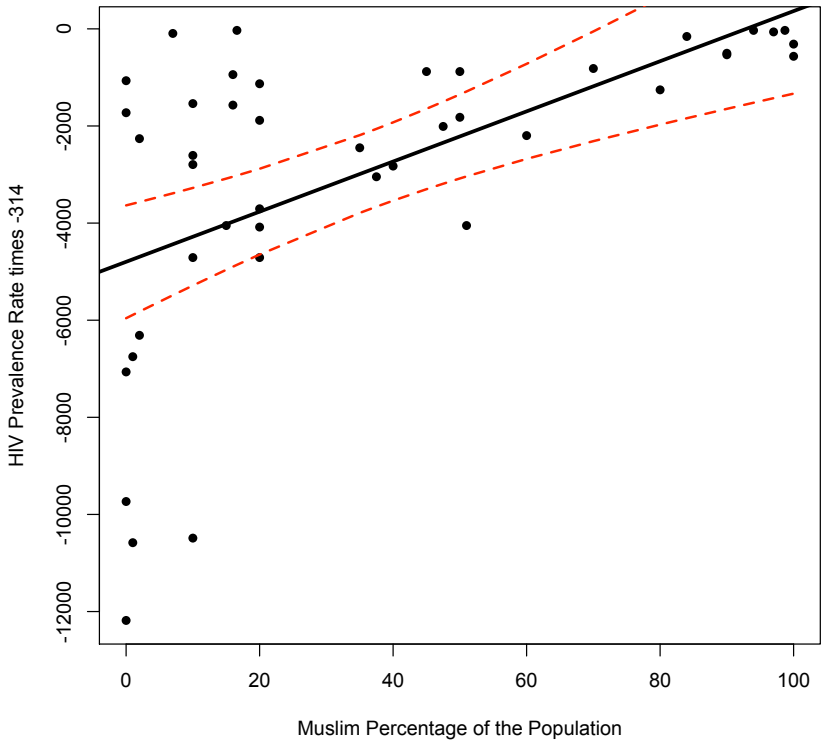
The resulting scatterplot appears on the next page. Notice a few things here:

- Because we have rescaled Y, the `Model`, `Residual`, and `Total` sums of squares have all changed, as has the `Root MSE` (but the R^2 and F statistics have not, indicating that the strength of the relationship / "fit" has not changed).
- The model now says that the expected value of $-314 \times$ *HIV/AIDS Rate* is $(-314 \times 15.279 =)$ -4797.5 percent ($\pm$ 1127.6 percent) when the Muslim population is zero, and
- that each one-percent increase in the Muslim population is associated with a $[-314 \times (-0.1644) =]$ 51.64 ($\pm$ 22.7 percent) increase in the expected value of ($-314 \times$ *HIV/AIDS Rate*).
- Relatedly, both the standard error estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ have changed, reflecting the fact that Y is now on a different "scale."

[1] I picked this day because my birthday – November 10 – is the 314th day of the year.

- Graphically, the scatterplot shows us that the Y-axis (and thus the relationship between X and Y) has been "flipped" (by virtue of the constant we multiplied being negative) and "stretched" (because the absolute value of that constant was greater than one).

Figure 3: Scatterplot of Rescaled HIV/AIDS Rates on Muslim Population Percentage



The point of this entire discussion is to point out how *linear transformations of X and Y do not affect the basic (linear) relationship between the two variables.* This means we can always "rescale" either or both of the variables linearly without changing our inferences, a fact that has some useful implications.

A Word About Percentages and Proportions

Percentages and proportions can always be thought of in terms of their inverses: The percentage of a population that has some characteristic is the same as 100 minus the percentage that does not have that characteristic. This sort of transformation is akin to an intercept shift; and, if done to both Y and X, gives us results that are identical to the original analysis in their findings (albeit somewhat different in their interpretation).

Consider again the HIV/AIDS – Muslim relationship:

```
> africa$nonmuslimpct<-100 - muslperc
> africa$noninfected<-100 - adrate
> fit4<-lm(noninfected~nonmuslimpct,data=africa)
> summary(fit4)

Call:
lm(formula = noninfected ~ nonmuslimpct, data = africa)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  101.1660     2.6808   37.74  < 2e-16 ***
nonmuslimpct  -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```

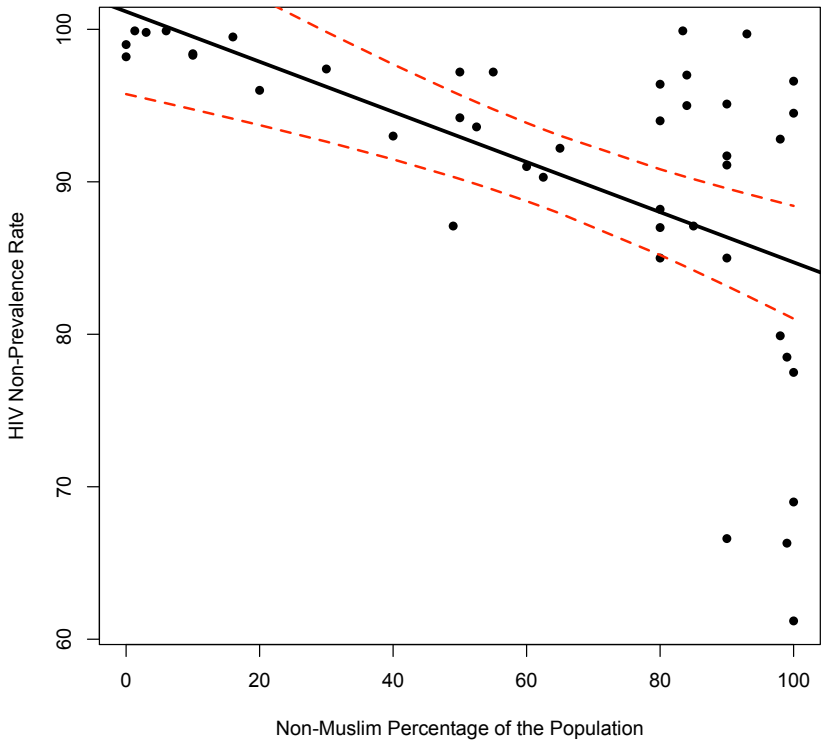
This regression tells us that:

- We would expect a country with no non-Muslims to have a non-infection rate of 101 percent ($\pm$ 5.25 percent).
- A one-percent increase in *non-Muslim* corresponds to an expected decrease in the non-infection rate of about 0.16 percent ($\pm$ 0.07 percent).

In other words, the relationship is exactly the same, but the interpretation is "reversed," as we would expect it to be. The scatterplot of this relationship also looks like a "mirror image" of the original one (see below).

Naturally, we'd be unlikely to use a transformation like this in this case, since it obscures (rather than clarifies) the nature of the relationship between X and Y. The importance of all this lies in the substantive interpretation of the covariates' associations with Y. Depending on the point you're trying to make, discussing (e.g.) the "percentage of Muslims" or "the percentage of non-Muslims" might make more sense.

Figure 4: Scatterplot of HIV/AIDS Non-Infection Rates on Non-Muslim Population Percentage



Interpreting the Constant Term

The foregoing discussion has obvious implications for if – and how – one interprets the estimate of $\hat{\beta}_0$ and its standard error estimates. The important thing to remember is that the constant term is the expected value of Y when $X = 0$. If $X = 0$ is not a "useful" value – because X can never equal zero, perhaps, or because it does not do so in the data – you may want to consider rescaling X so that $X = 0$ is meaningful.

One possibility that can be useful is to *center* X by subtracting its mean from it; call this $\tilde{X}_i = X_i - \bar{X}$. We can then regress Y on $\tilde{X}$. This has two potentially useful effects:

1. It means that the estimate of the intercept is now the expected value of Y when $\tilde{X} = 0$; that is, when X is at its mean. It thus prevents the estimated intercept from being nonsensical.

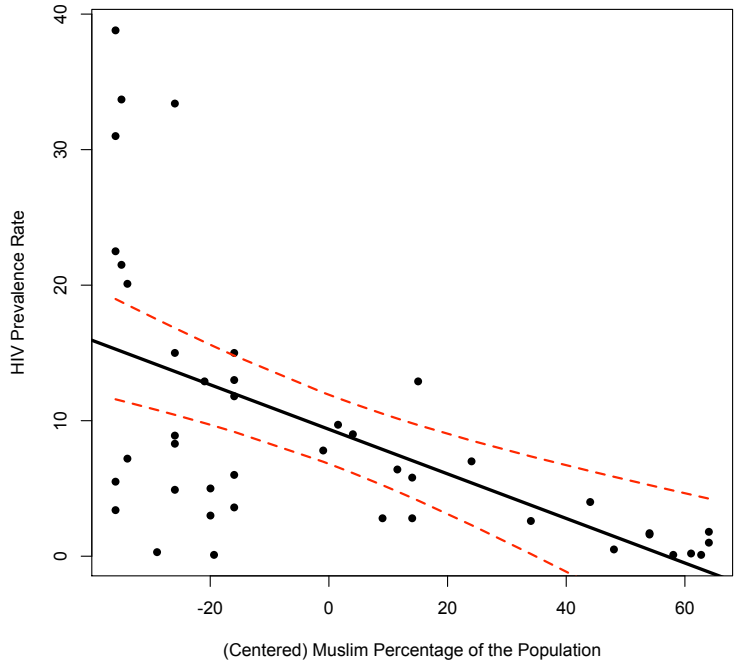
2. Relatedly, it will yield the smallest estimated standard error for the intercept, since $\bar{\tilde{X}} = 0$ by construction.

For example:

```
> africa$muslcenter<-muslperc - 35.96047
> fit5<-lm(adrate~muslcenter,data=africa)
> summary(fit5)

Call:
lm(formula = adrate ~ muslcenter, data = africa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.3651     1.2622    7.42  4.2e-09 ***
muslcenter   -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```

In this regression, the constant term tells us that, at the mean value of *Muslim Population*, the expected HIV infection rate is roughly 9.4 percent.

Figure 5: Scatterplot of HIV/AIDS Infection Rates on (Centered) Muslim Population Percentage



Note as well that if you "center" both X and Y, then the intercept will be meaningful, but it will also (by construction) be zero, since that is the mean of both the transformed X and Y variables (recall that, in the simple bivariate case, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$).

Rescaling Variables: Rules of Thumb

There are many rules-of-thumb that one needs to keep in mind when estimating and interpreting statistical models. The issue of rescaling variables raises two of them which, in some cases, have the potential to be in conflict with each other:

1. *Whenever possible, retain covariates in a metric that is the most "natural" and easily interpretable.*
2. *Whenever possible, rescale covariates so that coefficient estimates are easily reported and understood, and to eliminate large differences between the sizes of estimates.*

Both of these are good ideas in and of themselves, and as long as they don't conflict, both should be followed. It is easy to find examples of where this might be difficult to do, however. Consider a regression of HIV/AIDS rates on population:

```
> fit6<-lm(adrate~population,data=africa)
> summary(fit6)

Call:
lm(formula = adrate ~ population, data = africa)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.06e+01   1.91e+00    5.53    2e-06 ***
population  -7.05e-05   6.71e-05   -1.05      0.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.95 on 41 degrees of freedom
Multiple R-Squared: 0.0262,Adjusted R-squared: 0.00241
F-statistic:  1.1 on 1 and 41 DF,  p-value: 0.3
```

Here, population is just that – the number of people in the country – and ranges from 470,000 for Equitorial Guinea to about 117 million for Nigeria. As a result, the estimate of $\hat{\beta}_1$ reflects the expected change in HIV/AIDS rates associated with a *one person* increase in population. Not surprisingly, this effect is very small – the estimate is $\hat{\beta}_1 = -0.0000000705$. This estimate has two immediately apparent qualities:

1. It is hard to understand in and of itself, and
2. It is somewhat nonsensical (in that talking about a one-person change in a country's population is such a small number as to be negligible in real terms).

In addition, variables like these can tax the precision of the software/hardware used to estimate them, and in extreme cases even lead to rounding errors that can be misleading.

This is a case where the most natural unit of measure for population – the person – ought to give way to ease of interpretability and reporting. The best option is to simply divide `population` by a constant such that a one-unit change is meaningful. Since a natural way of thinking of countries' populations is in terms of millions of individuals, we might consider that:

```
> africa$popmil<-africa$population / 1000000
> fit7<-lm(adrate~popmil,data=africa)
> summary(fit7)

Call:
lm(formula = adrate ~ popmil, data = africa)

Residuals:
   Min     1Q Median     3Q    Max
-10.41  -7.34  -3.17   3.17  28.32

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.59       1.91    5.53    2e-06 ***
popmil        -70.46      67.14   -1.05      0.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.95 on 41 degrees of freedom
Multiple R-Squared: 0.0262,Adjusted R-squared: 0.00241
F-statistic:  1.1 on 1 and 41 DF,  p-value: 0.3
```

This makes a lot more sense, in substantive terms, but does nothing to the nature of the relationship between the variables.

Of course, as a practical matter, the balance between the value of retaining a variable's "natural" units and rescaling it for the same of clarity and interpretability is a judgement call. While I can't say I've ever seen it in any textbook, I have my own relatively simple rule for knowing when to rescale a covariate:

> **If an estimated coefficient has more than three digits to the left of the decimal place, or more than two zeros to the right of the decimal place, the variable associated with that estimate should be rescaled.**

Of course, this is just my rule. But, used consistently, it works pretty well. We'll return to this topic a bit later as well, in the context of multivariate regression models.

Topic #2: Proper Use and Interpretation of Dichotomous ("Dummy") Covariates

Dichotomous covariates present some challenges in interpretation in the context of regression models. Consider a model like this:

$$Y_i = \beta_0 + \beta_1 X_{Di} + u_i \tag{1}$$

where X_D indicates that X is a dichotomous variable, coded either zero or one.

Dummy Variables and t-tests

The first thing to note is that there are some striking similarities between the regression in (1) and a t-test for the difference of means in $Y|X_D = 0$ versus $Y|X_D = 1$. Consider our Africa data again:

```
> fit8<-lm(adrate~subsaharan,data=africa)
> summary(fit8)

Call:
lm(formula = adrate ~ subsaharan, data = africa)

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.27       3.88    0.33     0.75
subsaharanSub-Saharan    9.41       4.19    2.25     0.03 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.51 on 41 degrees of freedom
Multiple R-Squared: 0.11,Adjusted R-squared: 0.088
F-statistic: 5.05 on 1 and 41 DF,  p-value: 0.03

> t.test(adrate~subsaharan,var.equal=TRUE)

Two Sample t-test

data:  adrate by subsaharan
t = -2.248, df = 41, p-value = 0.03
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -17.8659  -0.9576
sample estimates:
mean in group Not Sub-Saharan     mean in group Sub-Saharan
                        1.267                        10.678
```

Notice a number of interesting similarities between these two things:

- The estimated intercept $\hat{\beta}_0$ is the same as the mean of Y when $X_D = 0$.
- The slope estimate $\hat{\beta}_1$ is the difference between $\bar{Y}|X_D = 1$ and $\bar{Y}|X_D = 0$; that is, $(\bar{Y}|X_D = 1) - (\bar{Y}|X_D = 0)$, which implies that
- the mean of Y when $X_D = 1$ (which is the same as the expected value of Y given $X_D = 1$) is equal to $\hat{\beta}_0 + \hat{\beta}_1$, and
- the estimated standard error of $\hat{\beta}_1$ is equal to the standard error in the difference of means. That, in turn, means that
- the t-score for $\hat{\beta}_1 = 0$ is exactly the same as the t-statistic for the difference of means, as is the level of significance, 95-percent confidence interval, etc.

All of this suggests that there is little reason to do a regression of Y on X_D alone, since a t-test is in many respects simpler and more easily understood. (That said, there's nothing "wrong" with including dichotomous Xs in a multivariate regression).

Interpreting Dummy Covariates

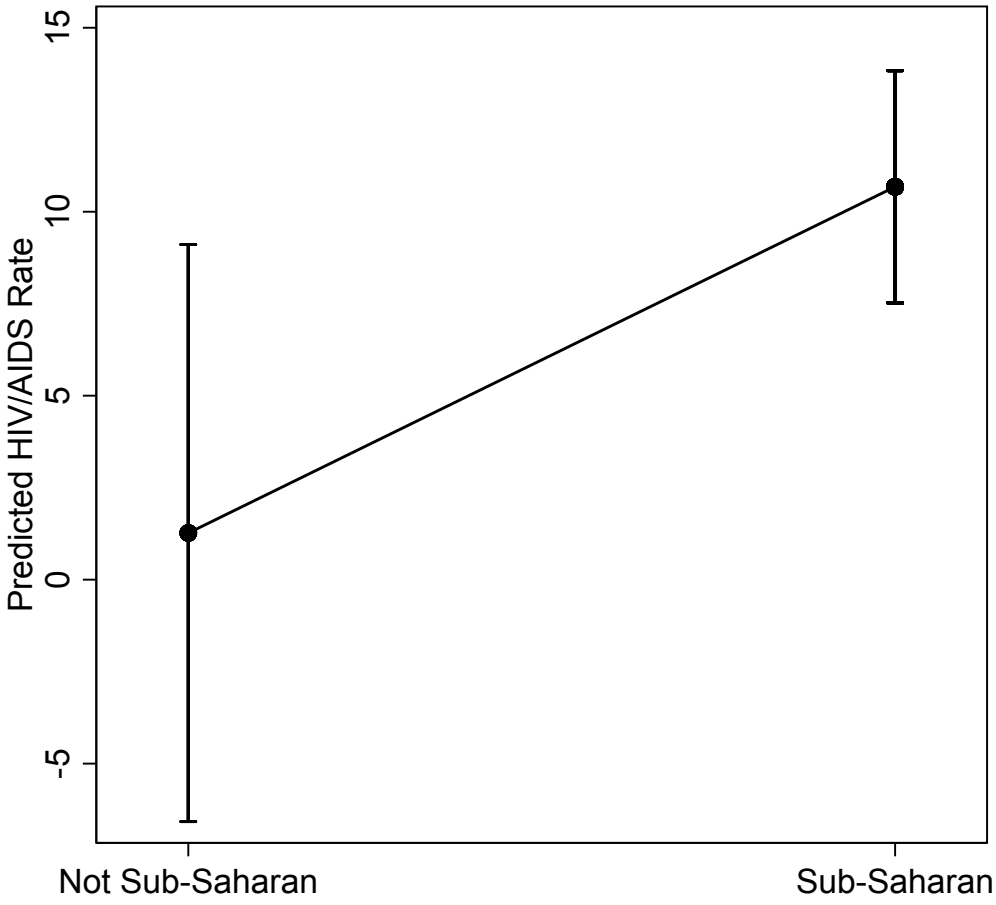
"Standard" methods for interpreting the influence of continuous covariates can generally be used with dichotomous variables as well; the key point is that some modification of presentation is usually warranted.

- Verbally, we can interpret the estimated effects in the standard fashion: $\hat{\beta}_1$, for example, is still the influence of a one-unit change in X_D on the expected value of Y. A key difference, however, is that we occasionally have dichotomous covariates which reflect qualitative (rather than quantitative) differences. In those instances, we rarely talk about "a one-unit change in X_D," but instead refer to the categories they represent. So, we'd typically say
    - "Our results indicate that countries in sub-Saharan Africa average HIV/AIDS infection rates which are nearly eight times greater, on average, than in Saharan African nations."

  or

    - "We find a statistically significant difference in HIV/AIDS infection rates between Saharan and sub-Saharan Africa, with average infection rates in the latter more than seven times greater than those in the former (10.7 versus 1.3 percent, respectively)."
- Graphically, it is usually better to use range-plots rather than scatterplots for things like predicted values of Y and their associated confidence intervals. So, to illustrate the effect of `subsaharan` on `adrate` in the model we estimated above, we could use:

Figure 6: Expected Values of HIV/AIDS Infection Rates in Saharan and Sub-Saharan Africa



This illustrates clearly that:

- The expected HIV/AIDS rate in Saharan Africa is about 1.3 percent, but with a relatively broad 95 percent confidence interval.
- The expected rate in sub-Saharan Africa is higher – 10.7 percent or so – and has a smaller associated 95 percent confidence interval (in large part because we have much more data on sub-Saharan Africa).

We'll spend a lot more time on issues related to dichotomous covariates in a couple weeks.

Topic #3: General Reporting of Regression Results

Finally, some general rules for reporting regression results. Some of these are from my "Ten Rules for Writing Up Quantitative Analyses" bit; others are more specific to this course, and/or aren't there for some other reason.

General Rules

1. *Every table, and every figure, should be able to "stand on its own."*

    - This means that, if someone were given the table / figure, along with the title of the paper/book to which it belonged, they'd be able to understand what was happening in it.
    - At a minimum, that means that each table / figure should have:
        (a) an informative *title*,
        (b) *labels*, wherever necessary,
        (c) a descriptive *caption*, which should include whatever information necessary to let the reader know what's going on.
    - Don't worry about being too verbose, or repeating things that are in the text.
    - Look at any paper Gary King has written in the past couple decades for good examples – he's one of the few people that are truly meticulous about this.

2. *Always report the N.*

    - Every table, and every figure, should tell the reader the number of observations on which it is based.
    - If the N varies within the table / figure, tell the reader that in the caption (*pace* Rule #1), or in the columns of the table.

3. *Use variable descriptions, not (software-based) variable names.*

    - In the analyses above, we used a covariate called `muslperc`, that was the Muslim percentage of the population. "Muslperc" is not a word, and should never appear in any table, figure, or other presentation of results. Instead, one should use "Muslim population percentage," or "Percent Muslim," or – if you're really tight on space – "Muslim."
    - Also, be consistent: If the same variable(s) appear in more than one table or figure, use *exactly* the same descriptive phrase for that variable in every table. Conversely, if you have three different measures of (e.g.) economic openness, each should be given a different, distinctive name to distinguish it from the other two.

Tables

Tables are a natural way to present the results of many (most?) statistical analyses. Some things to remember:

1. *Use column headings descriptively.*

- Don't say "Model I," "Model II," etc.
- Instead, say "POLITY Model," "Vanhanen Model," and so forth.

2. *Use multiple rows / columns rather than multiple tables.*

   - Whenever possible, combine shorter / smaller tables into larger ones.
   - Tables that are "wider" than they are "long" can be set using a landscape orientation.
   - *But*: don't do so at the expense of clarity. (This is a judgement call, and something one gets better at with practice).

3. *Learn about significant digits, and don't report more than 4-5 of them.*

   - That is, don't report a coefficient estimate as 0.0435099012, with a standard error of 0.016658321.
   - See the discussion of rescaling, above, if you have problems with "too many digits" on either side of the decimal place.

4. *Use a figure to replace a table when you can.*

   - It is often the case that you can display much more information much more clearly in a graph than in a table.
   - *Ceteris paribus*, a figure is better than a table.
   - For examples, see (e.g.) Gelman et al. (2002 *American Statistician*), Kastellec and Leoni (2007 *Perspectives on Politics*).

A Word About *s

If you ask people in political science about *s, you will get a host of responses. These can generally be grouped into two camps:

1. "*s are stupid. They are intellectually bankrupt, and indicative of a weak mind. They smell bad, chew with their mouths open, and leave the toilet seat up."

2. "*s are an integral part of every table. They are worthy substitutes for standard errors, make presentations more parsimonious and elegant, and generally make papers look shiny and clean. No work should be published unless it contains lots of *s."

As in nearly all such matters, the truth lies somewhere in between (though, for my money, it's closer to #1 than #2).

That said, if you're going to use *s, here are three rules of thumb:

1. **s go next to the coefficient estimates, not the standard errors.*
2. *Never use more than two.* The most widely-used convention has * indicating $p < 0.05$ and ** indicating $p < 0.01$.
3. *Tell the reader whether the significance levels indicated by the *s are one- or two-tailed.*

Figures

Note that all the rules that are relevant to tables also pertain to figures: naming variables, consolidation, etc. There are a few others that are specific to figures, though:

1. *Report the scale of axes, and label them.*
    - As before, labels should be informative, not variable names.
    - Scales should be clearly marked in units that are relevant to the variable in question.
2. *Use as much "space" as you need, but no more.*
    - E.g., if your Y variable is a percentage (and so theoretically can be anything between zero and 100), but it only ranges between 2% and 15% (as does, say, unemployment in the OECD), *don't* feel compelled to use a 100-percent scale for the Y-axis.
    - Both R and Stata (and most other good programs) do a pretty nice job of handling this for you.
    - As long as you abide by Rule #1 as well, your reader(s) have only themselves to blame if they misinterpret what they see.
3. *Use color sparingly.*
    - Color graphs are great, but many people can't print them, and many journals won't publish them.
    - Instead, use different symbols (circles, triangles, squares, plusses, etc.) and/or types of lines (solid, dashed, dotted, etc.) to indicate different things.

What All This Means

The results:

```
> fit<-lm(adrate~muslperc)
> summary.lm(fit)

Call:
lm(formula = adrate ~ muslperc)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.2787     1.8322    8.34  2.3e-10 ***
muslperc     -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```

The table:

Table 1: OLS Regression Model of HIV/AIDS Rates in Africa, 2001.

Variables	Model I
(Constant)	15.28
	(1.83)
Muslim Percentage of the Population	-0.164*
	(0.037)
Adjusted R^2	0.31

Note: $N = 43$. Cell entries are coefficient estimates; numbers in parentheses are estimated standard errors. Asterisks indicate $p < .05$ (one-tailed). See text for details.