Lecture Notes: Causal Inference Fall 2020

Claire Duquennois

7/28/2020

Contents

1 The Hodgepodge

1.1 Non-standard standard errors

A standard error is, of course, an estimate of the uncertainty around an estimated parameter. Formally we have

$$se = \sqrt{\widehat{Var(\hat{\beta})}}$$

in other words, the standard error is the square root of the estimated variance of the estimated parameter.

Just like calculating point estimates, it is incredibly important to get your standard errors right. You have to know what you don't know!

1.1.1 Robust standard errors

We are going to use the diamonds data set from `ggplot2` for this exercise so we don't need to load an external data set.

```
knitr::kable(head(diamonds))
```

carat	cut	color	clarity	depth	table	price	x	y	z
0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	2.43
0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	2.31
0.23	Good	E	VS1	56.9	65	327	4.05	4.07	2.31
0.29	Premium	I	VS2	62.4	58	334	4.20	4.23	2.63
0.31	Good	J	SI2	63.3	58	335	4.34	4.35	2.75
0.24	Very Good	J	VVS2	62.8	57	336	3.94	3.96	2.48

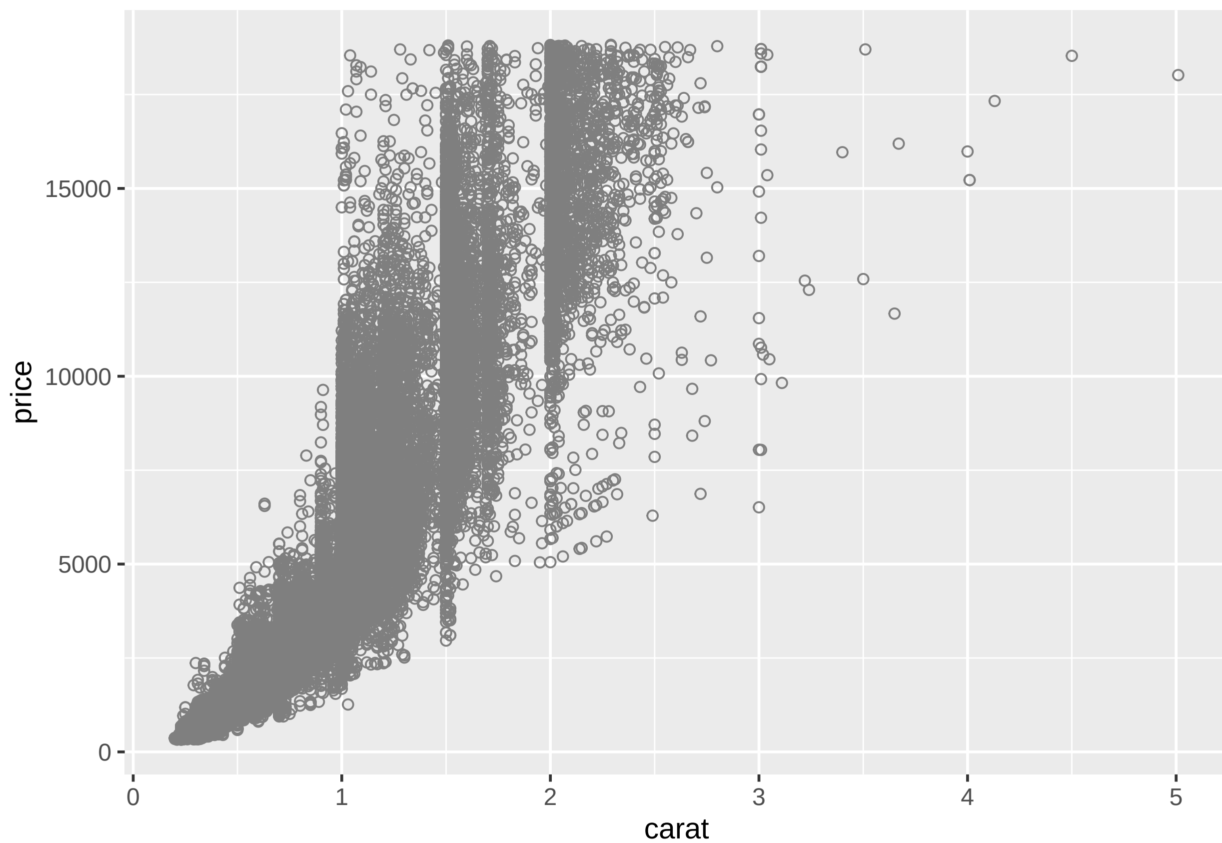
Lets regress price on carats and depth.

```
reg1<-felm(price~carat+depth, diamonds)

summary(reg1)
```

```
##
## Call:
##    felm(formula = price ~ carat + depth, data = diamonds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18238.9   -801.6    -19.6    546.3  12683.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4045.333    286.205   14.13   <2e-16 ***
## carat       7765.141     14.009  554.28   <2e-16 ***
## depth       -102.165      4.635  -22.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1542 on 53937 degrees of freedom
## Multiple R-squared(full model): 0.8507   Adjusted R-squared: 0.8507
## Multiple R-squared(proj model): 0.8507   Adjusted R-squared: 0.8507
## F-statistic(full model):1.536e+05 on 2 and 53937 DF, p-value: < 2.2e-16
## F-statistic(proj model): 1.536e+05 on 2 and 53937 DF, p-value: < 2.2e-16
```

Cool. But of course, we should make sure that our OLS assumptions make sense. One easy way to do this is to plot the data:

```r
myPlot <- ggplot(data = diamonds, aes(y = price, x = carat)) +
geom_point(color = "gray50", shape = 21)

myPlot
```



There are a bunch of things about this plot that should give you the econometric heebie jeebies. From an OLS perspective, you should be very afraid that these data are definitely not homoskedastic. The higher the carat, the greater the variance in price. This means that our OLS standard errors are likely going to get things wrong.

Heteroskedasticity is scary- but thankfully all is not lost. All we have to do is tweak our original assumptions a little bit to relax the homoskedasticity assumption and allow for the variance to depend on the value of x_i.

We know that

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma^2}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$$

With heteroskedasticity σ^2 is no longer constant and becomes a function of the particular value of x_i an observation has, so

$$Var(u_i|x_i) = \sigma_i^2$$

Where are we going to find all these σ_i^2 for each individual observation?

Econometricians Eicker, Huber and White figured out a way to do this by basically using the square of the estimated residual of each observation, $\hat{u}_i^2$, as a stand-in for σ_i^2. With this trick, a valid estimator for $Var(be\hat{t}a_1)$, with heteroskedasticity of **any** form (including homoskedasticity), is

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\hat{u}_i^2}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$$

We commonly call the resulting standard errors “robust”, or “heteroskedasticity-robust”.

How can we find these in R?

```r
reg1<-felm(price~carat+depth, diamonds)

summary(reg1, robust=TRUE)
```

```
##
## Call:
##    felm(formula = price ~ carat + depth, data = diamonds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18238.9   -801.6    -19.6    546.3  12683.7
##
## Coefficients:
##             Estimate Robust s.e t value Pr(>|t|)
## (Intercept) 4045.333    369.176   10.96   <2e-16 ***
## carat       7765.141     25.105  309.31   <2e-16 ***
## depth       -102.165      5.946  -17.18   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1542 on 53937 degrees of freedom
## Multiple R-squared(full model): 0.8507   Adjusted R-squared: 0.8507
## Multiple R-squared(proj model): 0.8507   Adjusted R-squared: 0.8507
## F-statistic(full model, *iid*):1.536e+05 on 2 and 53937 DF, p-value: < 2.2e-16
## F-statistic(proj model): 4.878e+04 on 2 and 53937 DF, p-value: < 2.2e-16
```

Or if you want to put them in a stargazer table:

```r
stargazer(reg1, type = "latex" , se =  list(reg1$rse), header=FALSE)
```

Table 2:

	Dependent variable:
	price
carat	$7,765.141^{***}$
	(25.105)
depth	-102.165^{***}
	(5.946)
Constant	$4,045.333^{***}$
	(369.176)
Observations	53,940
R^2	0.851
Adjusted R^2	0.851
Residual Std. Error	1,541.649 (df = 53937)
Note:	*p<0.1; **p<0.05; ***p<0.01

It is worth noting that robust standard errors are larger than regular standard errors, and thus more conservative (which is the right thing to be... you want to know what you don't know).

1.1.2 Clustered standard errors

Econometricians Haiku

T-stats looks too good

Try cluster standard errors

significance gone.

from Angrist and Pischke 2008

Suppose that every observation belongs to (only) one of G groups. The assumption we make when we cluster is that there is no correlation across groups- but we will allow for arbitrary within-group correlation. A great example: consider individuals within a village. In many cases it's pretty reasonable to think that individuals' error terms are correlated within a village, but that individuals' errors aren't correlated across villages.

I will spare you the matrix math needed to dive deeper into this. Suffice to say that "cluster-robust" estimates allow for a more complicated set of correlations to exist within observations within a cluster. One thing to be aware of though is that you will need to have a fairly large number of clusters (40+) for the estimate to be credible.

Clustering in R:

I use the `NOxEmissions` dataset from the `robustbase` package. This is a dataset of hourly NO_x readings, including NO_x concentration, auto emissions and windspeed. We are going to use the observation date as our cluster variable. This allows for arbitrary dependence between observations in the same day, and zero correlation across days. In this reasonable? ... Maybe. But we'll go with it for now:

```
nox <- as.data.frame(NOxEmissions) %>%
mutate(ones = 1)
noClusters <- felm(data = nox, LNOx ~ sqrtWS )

Clusters <- felm(data = nox, LNOx ~ sqrtWS |0|0| julday)

stargazer(noClusters,Clusters, type = "latex" , header=FALSE)
```

Table 3:

	Dependent variable:	
	LNOx	
	(1)	(2)
sqrtWS	-0.864^{***}	-0.864^{***}
	(0.020)	(0.048)
Constant	5.559^{***}	5.559^{***}
	(0.029)	(0.065)
Observations	8,088	8,088
R^2	0.185	0.185
Adjusted R^2	0.185	0.185
Residual Std. Error (df = 8086)	0.846	0.846
Note:	*p<0.1; **p<0.05; ***p<0.01	

In this case, the regular standard errors are smaller than the clustered standard errors. Be aware that this need not necessarily be the case - depending on the correlation between observations within a cluster, clustered standard errors can be smaller than regular standard errors.

1.1.3 Newey West Standard Errors

For time series data.

1.1.4 Conley Standard Errors

For spatial data.

1.2 Confidence intervals for predictions

You've already had to "predict" a value of the dependent variable, y, given certain values of the independent variables. But this prediction is just a guess, and we can construct a confidence interval to give a range of possible values for this prediction, to demonstrate this uncertainty.

There are two kinds of predictions we can make:

- A confidence interval for the **average** y given x_1, $x_2 \ldots$ x_k.
- A confidence interval for a **particular** y given x_1, $x_2 \ldots$ x_k.

We will use Wooldridge's birth weight data to construct both kinds of confidence intervals to demonstrate the (subtle) difference between them.

$$bweight = \beta_0 + \beta_1 lfaminc + \beta_2 meduc + \beta_3 parity + u$$

where bwght is birth weight in ounces, lfaminc is the log of family income in \$1000s, meduc is the education of the mother in years, and parity is the birth order of the child.

Estimating this equation in R, we get the following results:

```r
#using the bwght data from the wooldridge package
reg1<-lm(bwght~lfaminc+motheduc+parity, bwght)

summary(reg1)
```

```
##
## Call:
## lm(formula = bwght ~ lfaminc + motheduc + parity, data = bwght)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -94.533 -11.888   0.779  13.136 151.477
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.5652     3.3666  31.356  < 2e-16 ***
## lfaminc       2.1313     0.6506   3.276  0.00108 **
## motheduc      0.3172     0.2520   1.259  0.20829
## parity        1.5261     0.6119   2.494  0.01275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.21 on 1383 degrees of freedom
##   (1 observation deleted due to missingness)
```

```
## Multiple R-squared:  0.01633,    Adjusted R-squared:  0.0142
## F-statistic: 7.654 on 3 and 1383 DF,  p-value: 4.482e-05
```

1.2.0.1 Confidence interval for the average Birthweight Recall that our model gives us the expected value:

$$E[bweight|lfaminc, meduc, parity] = \beta_0 + \beta_1 log(faminc) + \beta_2 meduc + \beta_3 parity$$

and our regression gives us an estimate of this:

$$\hat{E}[bweight|faminc, meduc, parity] = \hat{y} = \hat{\beta_0} + \hat{\beta_1} log(faminc) + \hat{\beta_2} meduc + \hat{\beta_3} parity$$

In words: when we plug in particular values of the independent variables, we obtain a predication for y, which is an estimate for the expected value of y given the particular values for the explanatory variables.

Say we are interested in a confidence interval for the **average birthweight** for babies with a family income of \$14,500 (ln(14.5)=2.674), mothers with 12 years of education, and with 2 older siblings (parity=3). In other words, we are interested in:

$$\begin{aligned}
\hat{E}[bweight|faminc = 14.5, meduc = 12, parity = 3] &= 105.66 + 2.13 ln(faminc) + 0.317 meduc + 1.53 parity \\
\hat{y}_{faminc=14.5, meduc=12, parity=3} &= 105.66 + 2.13(2.674) + 0.317(12) + 1.53(3) \\
&= 119.75 ounces
\end{aligned}$$

How do we find a standard error for our estimate of the expected value of y for these particular values of the explanatory variables? Notice that this standard error is complicated because $\widehat{bweight}$ is a function of our $\hat{\beta}$'s which are all random variables. To avoid this computation, we want to transform our data. Before proceeding with the formal steps, recall that we have the following regression in mind

$$bweight = \beta_0 + \beta_1 lfaminc + \beta_2 meduc + \beta_3 parity + u$$

Then

$$\hat{\beta_0} = \hat{E}(bweight|lfaminc = 0, meduc = 0, parity = 0)$$

If we modify the regression by subtracting our particular values (specified above) for each of the independent variables, then we get the following regression

$$bweight = \beta_0 + \beta_1(lfaminc - 2.674) + \beta_2(meduc - 12) + \beta_3(parity - 3) + u$$

Then

$$\hat{\beta_0} = \hat{E}(bweight|lfaminc = 2.674, meduc = 12, parity = 3).$$

In other words, the new intercept is the predicted birthweight for babies with family incomes of \$14,500 (ln(14.5)=2.674), mothers with 12 years of education and with 2 older siblings. That's perfect! If we run this regression, `R` always outputs a standard error for the intercept coefficient. We can then grab the standard errors from there.

So step by step we need to:

1) Generate new variables: $\tilde{x}_j = x_j - \alpha_j$
2) Run the regression: $y = \tilde{\beta_0} + \tilde{\beta_1}\tilde{x_1} + ... + \tilde{\beta_k}\tilde{x_k} + \tilde{u}$
3) Then $\hat{E}[y|x_1 = \alpha_1, ..., x_k = \alpha_k] = \tilde{\beta_0}$

4) Plug these values into the formula for confidence intervals and interpret.

Below is the code for steps 1 and 2:

```r
#Step 1: generate new variables
bwght$lfaminc_0<-bwght$lfaminc-2.674
bwght$motheduc_0<-bwght$motheduc-12
bwght$parity_0<-bwght$parity-3

#step 2: run the new regression

reg2<-lm(bwght~lfaminc_0+motheduc_0+parity_0,bwght)

summary(reg2)
```

```
##
## Call:
## lm(formula = bwght ~ lfaminc_0 + motheduc_0 + parity_0, data = bwght)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -94.533 -11.888   0.779  13.136 151.477
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 119.6491     1.0066 118.864  < 2e-16 ***
## lfaminc_0     2.1313     0.6506   3.276  0.00108 **
## motheduc_0    0.3172     0.2520   1.259  0.20829
## parity_0      1.5261     0.6119   2.494  0.01275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.21 on 1383 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01633,    Adjusted R-squared:  0.0142
## F-statistic: 7.654 on 3 and 1383 DF,  p-value: 4.482e-05
```

Using this output, the 95% confidence interval for the average birthweight for babies given a family income of \$14,500, a mother with 12 years of education and with 2 older siblings is:

$$[119.64 - 1.96(1.007), 119.64 + 1.96(1.007)] = [117.6653, 121.6158]$$

1.2.0.2 Confidence Interval for a particular individual's birthweight

A confidence interval for the average of observations with specific characteristics is not the same as a confidence interval for a particular individual observation. In forming a confidence interval for a particular unit, we must account for another very important source of variation: the variance in the unobserved error, which measures our ignorance of the unobserved factors that affect y_i.

We would like to construct a confidence interval for the birthweight of baby $i = 1$. Let $bweight_{i=1}$ denote that particular baby's birthweight with

$$bweight_{i=1} = \beta_0 + \beta_1 lfaminc_{i=1} + \beta_2 meduc_{i=1} + \beta_3 parity_{i=1} + u_{i=1}$$

Our best prediction of $bweight_{i=1}$ is $\widehat{bwieght}_{i=1}$ where

$$\widehat{bweight}_{i=1} = \hat{\beta}_0 + \hat{\beta}_1 lfaminc_{i=1} + \hat{\beta}_2 meduc_{i=1} + \hat{\beta}_3 parity_{i=1}$$

Now, there is some error, $\hat{u}_{i=1}$ associated with using $\widehat{bweight}_{i=1}$ to predict $bweight_{i=1}$ where

$$\begin{aligned}
\hat{u}_{i=1} &= bweight_{i=1} - \widehat{bweight}_{i=1} \\
&= (\beta_0 + \beta_1 lfaminc_{i=1} + \beta_2 meduc_{i=1} + \beta_3 parity_{i=1} + u_{i=1}) - (\hat{\beta}_0 + \hat{\beta}_1 lfaminc_{i=1} + \hat{\beta}_2 meduc_{i=1} + \hat{\beta}_3 parity_{i=1})
\end{aligned}$$

Finding the expected value, we get:

$$\begin{aligned}
E[\hat{u}_{i=1}] &= E[bweight_{i=1} - \widehat{bweight}_{i=1}] \\
&= (\beta_0 + \beta_1 lfaminc_{i=1} + \beta_2 meduc_{i=1} + \beta_3 parity_{i=1} + E[u_{i=1}]) - (E[\hat{\beta}_0] + E[\hat{\beta}_1] lfaminc_{i=1} + E[\hat{\beta}_2] meduc_{i=1} + E[\hat{\beta}_3] parity_{i=1}) \\
&= 0
\end{aligned}$$

Finding the variance we get

$$\begin{aligned}
Var(\hat{u}_{i=1}) &= Var(bweight_{i=1} - \widehat{bweight}_{i=1}) \\
&= Var(\beta_0 + \beta_1 lfaminc_{i=1} + \beta_2 meduc_{i=1} + \beta_3 parity_{i=1} + u_{i=1} - \widehat{bweight}_{i=1}) \\
&= Var(\widehat{bweight}_{i=1}) + Var(u_{i=1}) \\
&= Var(\widehat{bweight}_{i=1}) + \sigma^2 \\
\widehat{Var}(\hat{u}_{i=1}) &= Var(\widehat{bweight}_{i=1}) + \hat{\sigma}^2 \\
&= Var(\widehat{bweight}_{i=1}) + \frac{\sum \hat{u}_i^2}{n-k-1} \\
&= Var(\widehat{bweight}_{i=1}) + \frac{SSR}{n-k-1}
\end{aligned}$$

So you should see that there are two sources of variation in $\hat{u}_{i=1}$. First we have the sampling error in $\widehat{bweight}_{i=1}$ which arises because we have estimated the population parameters β. Second we have the variance of the error in the population ($u_{i=1}$).

Now we can compute the $Var(\widehat{bweight}_{i=1})$ exactly the way we did before (by subtracting the specific values we are interested in and re-running the regression and looking at the intercept term's standard errors). Second we can compute $\frac{SSR}{n-k-1}$ from our regression output. The 95% confidence interval for $bweight_{i=1}$ is then

$$\hat{y} \pm 1.96 * se(\hat{u}_{i=1})$$

Steps in computing a confidence interval for a particular y when $x_j = \alpha_j$:

1) Generate new variables: $\tilde{x}_j = x_j - \alpha_j$
2) Run the regression: $y = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_1 + ... + \tilde{\beta}_k \tilde{x}_k + \tilde{u}$
3) Then $\hat{E}[y|x_1 = \alpha_1, ..., x_k = \alpha_k] = \tilde{\beta}_0$ and the standard error of the estimate is $se(\tilde{\beta}_0)$
4) Get an estimate for the variance of $\hat{u} = \hat{\sigma}^2$ from the R output
5) compute the standard error: $\sqrt{se(\tilde{\beta}_0)^2 + \hat{\sigma}^2}$
6) Plug these values into the formula for confidence intervals and interpret.

Below is the code:

```r
#Step 1: generate new variables
bwght$lfaminc_0<-bwght$lfaminc-2.674
bwght$motheduc_0<-bwght$motheduc-12
bwght$parity_0<-bwght$parity-3

#step 2: run the new regression

reg2<-lm(bwght~lfaminc_0+motheduc_0+parity_0,bwght)

summary(reg2)
```

```
##
## Call:
## lm(formula = bwght ~ lfaminc_0 + motheduc_0 + parity_0, data = bwght)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -94.533 -11.888   0.779  13.136 151.477
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 119.6491     1.0066 118.864  < 2e-16 ***
## lfaminc_0     2.1313     0.6506   3.276  0.00108 **
## motheduc_0    0.3172     0.2520   1.259  0.20829
## parity_0      1.5261     0.6119   2.494  0.01275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.21 on 1383 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01633,    Adjusted R-squared:  0.0142
## F-statistic: 7.654 on 3 and 1383 DF,  p-value: 4.482e-05
```

```r
#step 4: get the estimate of the variance

summary(lm(bwght~lfaminc_0+motheduc_0+parity_0,bwght))$sigma^2
```

```
## [1] 408.5987
```

The confidence interval for a particular baby's birthweight with a family income of $14,500 (ln(14.5=2.674)), a mother with 12 years of education and with 2 older siblings we have:

$$\begin{aligned}
SE &= \sqrt{se(\tilde{\beta}_0)^2 + \hat{\sigma}^2} = \sqrt{(1.007^2) + 408.59} = 20.239 \\
CI &= [119.64 - 1.96 * (20.239); 119.64 + 1.96 * (20.239)] \\
&= [79.972; 159.308]
\end{aligned}$$

1.3 Variable Transformations

1.3.1 Scaling

1.3.2 Standardizing

Standardizing variables eliminates the units in order to be able to compare the magnitude of estimates across independent variables. This can make interpretation easier if you have variables with weird arbitrary units that are unfamiliar to people. We can solve this issue by standardizing the variables.

Suppose we have a regression with two variables, x_1 and x_2:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{u}$$

We know that our regression must go through the point of averages, or think if we average the previous equation, and use the fact that the $\hat{u}_i$'s have a zero sample average, or if we plugged in $\bar{x}_1$ and $\bar{x}_2$, we would predict $\bar{y}$:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2$$

We can subtract the second equation from the first to get:

$$\begin{aligned}\hat{y} - \bar{y} &= (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{u}) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2) \\ &= \hat{\beta}_1 (x_1 - \bar{x}_1) + \hat{\beta}_2 (x_2 - \bar{x}_2) + \hat{u}\end{aligned}$$

With a little bit of additional algebra, dividing both sides of this equation by the standard deviation of y, σ_y and multiplying each independent variable by $1 = \frac{\sigma_x}{\sigma_x}$. we can get the entire regression into standard units:

$$(\frac{y - \bar{y}}{\hat{\sigma}_y}) = \frac{\hat{\sigma}_{x_1}}{\hat{\sigma}_y} \hat{\beta}_1 (\frac{x_1 - \bar{x}_1}{\hat{\sigma}_{x_1}}) + \frac{\hat{\sigma}_{x_2}}{\hat{\sigma}_y} \hat{\beta}_2 (\frac{x_2 - \bar{x}_2}{\hat{\sigma}_{x_2}}) + \frac{\hat{u}}{\hat{\sigma}_y}$$

Now we can say that controlling for x_2 a one standard deviation increase in x_1 leads to a $\frac{\hat{\sigma}_{x_1}}{\hat{\sigma}_y} \hat{\beta}_1$ standard deviation increase in the predicted y. We call this new term the standardized coefficient or "beta" coefficient. In R, we can get these coefficients by using the `lm.beta` command following a regression using `lm`.

We use the bwght2 dataset to look at how parent ages correlate with birth weights. (Note: birth weights here will be measured in grams). I estimate four different regressions of the type

$$birthweight_i = \beta_0 + \beta_1 motherage_i + \beta_2 fatherage_i + \epsilon_i$$

scaling either the dependent and/or independent variables.

```
bwght<-bwght2

reg1<-lm(bwght~mage+fage, bwght)
reg2<-lm(scale(bwght)~scale(mage)+scale(fage), bwght)
reg3<-lm(scale(bwght)~mage+fage, bwght)
reg4<-lm(bwght~scale(mage)+scale(fage), bwght)

meandep1<-round(mean(bwght$bwght),2)
meandep2<-round(mean(scale(bwght$bwght)),2)
sddep1<-round(sd(bwght$bwght),2)
sddep2<-round(sd(scale(bwght$bwght)),2)

stargazer(reg1,reg2, reg3, reg4, type = "latex" , header=FALSE,
          add.lines=list(c("Mean",meandep1,meandep2, meandep2, meandep1 ),
                         c("SD",sddep1,sddep2, sddep2, sddep1 )))
```

Interpreting column 1:

- Having a mother that is **a year** older at birth predicts a birthweight that is 3.992 **grams** less (not statistically significant).
- Having a father that is **a year** older at birth predicts a birthweight that is 9.313 **grams** more (highly statistically significant).

Table 4:

	Dependent variable:			
	bwght	scale(bwght)		bwght
	(1)	(2)	(3)	(4)
mage	−3.992		−0.007	
	(3.943)		(0.007)	
fage	9.313***		0.016***	
	(3.291)		(0.006)	
scale(mage)		−0.033		−19.044
		(0.033)		(18.812)
scale(fage)		0.092***		53.205***
		(0.033)		(18.803)
Constant	3,221.030***	−0.001	−0.312**	3,400.304***
	(87.703)	(0.023)	(0.152)	(13.448)
Mean	3401.12	0	0	3401.12
SD	576.54	1	1	576.54
Observations	1,826	1,826	1,826	1,826
R^2	0.005	0.005	0.005	0.005
Adjusted R^2	0.004	0.004	0.004	0.004
Residual Std. Error (df = 1823)	574.677	0.997	0.997	574.677
F Statistic (df = 2; 1823)	4.912***	4.912***	4.912***	4.912***
Note:				*p<0.1; **p<0.05; ***p<0.01

Interpreting column 2:

- Having a mother whose age is **one standard deviation higher** at birth predicts a birthweight that is 0.033 **standard deviations** lower (not statistically significant).
- Having a father whose age is **one standard deviation higher** at birth predicts a birthweight that is 0.092 **standard deviations** higher (highly statistically significant).

Interpreting column 3:

- Having a mother that is **a year** older at birth predicts a birthweight that is 0.007 **standard deviations** lower (not statistically significant).
- Having a father that is **a year** older at birth predict a birthweight that is 0.016 **standard deviations** higher (highly statistically significant).

Interpreting column 4:

- Having a mother whose age is **one standard deviation higher** at birth predicts a birthweight that is 19.044 **grams** lower (not statistically significant).
- Having a father whose age is **one standard deviation higher** at birth predicts a birthweight that is 53.205 **grams** higher (highly statistically significant).

Notes:

- You can also standardize y, x_1 and x_2 directly in the data set and then run your regression on the standardized variables though this involves more coding.
- You do not need to standardize all the variables. You could just standardize x_1 and adjust your interpretation accordingly (you will need an intercept in this case).

1.3.3 Logs

1.3.4 Quadratics

1.3.5 Interactions with continuous variables

1.4 Non-linear estimation strategies

1.4.1 Logits