Scaling, Centering, and Collinearity

Read in the *Prestige.csv* data

Pineo-Porter occupational prestige score

Is the occupation a blue-collar profession?

```
               occupation prestige education blue_collar income
1 government administrators     68.8     13.11           0  12351
2          general managers     69.1     12.26           0  25879
3               accountants     63.4     12.77           0   9271
4       purchasing officers     56.8     11.42           0   8865
5                  chemists     73.5     14.62           0   8403
6                physicists     77.6     15.64           0  11030
```

Average education of occupational incumbents

Average income, in dollars,

These data are Canadian Census data from 1971, and are available as part of the **car** package.
Canada (1971) *Census of Canada*. Vol. 3, Part 6. Statistics Canada [pp. 19-1–19-21].

$$\text{Prestige} = \beta_0 + \beta_1(\text{Income}) + \beta_2(\text{Education}) + \beta_3(\text{Income})(\text{Education}) + \epsilon$$

```
Coefficients:
                     Estimate    Std. Error  t value   Pr(>|t|)
(Intercept)      -2.243231e+01  6.557889e+00 -3.42066 0.00092597 ***
income            3.739037e-03  1.003440e-03  3.72622 0.00033170 ***
education         5.475932e+00  5.682087e-01  9.63718 1.0855e-15 ***
income:education -1.876822e-04  7.369712e-05 -2.54667 0.01250027 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.243593 on 94 degrees of freedom
Multiple R-squared:  0.8260074,  Adjusted R-squared:  0.8204545
F-statistic:  148.751 on 3 and 94 DF,  p-value: < 2.2204e-16
```

$$\widehat{\text{Prestige}} = -22.4 + 0.004(\text{Income}) + 5.5(\text{Education}) - 0.0002(\text{Income})(\text{Education})$$

Scaling Predictors

We can scale any continuous predictors by multiplying or dividing by any value we want. This can alleviate really small or really large coefficients. The key is to get all of the predictors close to the same scale of magnitude.

```
                occupation prestige education type income
1 government administrators     68.8     13.11 prof  12351
2         general managers     69.1     12.26 prof  25879
3              accountants     63.4     12.77 prof   9271
4       purchasing officers     56.8     11.42 prof   8865
5                 chemists     73.5     14.62 prof   8403
6               physicists     77.6     15.64 prof  11030
```

Divide income by 1000

$$\text{Prestige} = \beta_0 + \beta_1(\text{Income}/1000) + \beta_2(\text{Education}) + \beta_3(\text{Income}/1000)(\text{Education}) + \epsilon$$

```
> lm.2 = lm(prestige ~ I(income/1000) + education +
     education:I(income/1000), data = Prestige)

Coefficients:
                             Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)               -22.43230970   6.55788913 -3.42066 0.00092597 ***
I(income/1000)              3.73903723   1.00343976  3.72622 0.00033170 ***
education                   5.47593167   0.56820872  9.63718 1.0855e-15 ***
I(income/1000):education  -0.18768224   0.07369712 -2.54667 0.01250027 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.243593 on 94 degrees of freedom
Multiple R-squared:  0.8260074,    Adjusted R-squared:  0.8204545
F-statistic:  148.751 on 3 and 94 DF,  p-value: < 2.2204e-16
```

$$\widehat{\text{Prestige}} = -22.4 + 3.7(\text{Income}/1000) + 5.5(\text{Education}) + -0.2(\text{Income}/1000)(\text{Education})$$

Scaling changes the size of the estimate and SE, but the t-values and p-values stay the same. Also, the model summaries are all identical.

Standardizing

One method of scaling that guarantees that all predictors will be on the same scale is standardizing. Standardizing a predictor means to divide it by its standard deviation.

```
> Prestige$income.std = Prestige$income / sd(Prestige$income)
> Prestige$education.std = Prestige$education / sd(Prestige$education)

> head(Prestige)
                 occupation prestige education type income  income.std education.std
1 government administrators     68.8     13.11 prof  12351 2.921147785   4.769117021
2          general managers     69.1     12.26 prof  25879 6.120669058   4.459906535
3               accountants     63.4     12.77 prof   9271 2.192693799   4.645432827
4       purchasing officers     56.8     11.42 prof   8865 2.096670319   4.154333820
5                  chemists     73.5     14.62 prof   8403 1.987402222   5.318420355
6                physicists     77.6     15.64 prof  11030 2.608716709   5.689472938
```

The income of government administrators, \$12,351 Canadian Dollars, is 2.92 times larger than the standard deviation of \$4228.

```
> lm.3 = lm(prestige ~ income.std + education.std +
     education.std:income.std, data = Prestige)


Coefficients:
                            Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)              -22.4323097   6.5578891 -3.42066 0.00092597 ***
income.std                15.8091450   4.2426763  3.72622 0.00033170 ***
education.std             15.0529886   1.5619697  9.63718 1.0855e-15 ***
income.std:education.std  -2.1814059   0.8565719 -2.54667 0.01250027 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.243593 on 94 degrees of freedom
Multiple R-squared:  0.8260074,    Adjusted R-squared:  0.8204545
F-statistic:  148.751 on 3 and 94 DF,  p-value: < 2.2204e-16
```

Scaling changes the size of the estimate and SE, but the *t*-values and *p*-values stay the same. Also, the model summaries are all identical.

The coefficients now are interpreted as a one-standard deviation difference in X_1 is associated with a hat(beta)-unit difference in Y, controlling for X_2, X_3, …

Centering

We can *center* any continuous predictors by subtracting or adding any value we want.

Say poverty was defined by a \$5000 income or below, and a low-level of education was the 8th grade.

$$\text{Centered Income} = \text{Income} - 5000$$

$$\text{Centered Education} = \text{Education} - 8$$

A centered income value of zero would indicate an occupation at the poverty line (a raw income of \$5000).

A centered education value of zero would indicate an occupation with a low-level of education (a raw education of 8).

Centered model

```
> lm.4 = lm(prestige ~ I(income - 5000) + I(education - 8) +
     I(education - 8):I(income - 5000), data = Prestige)

Coefficients:
                                     Estimate    Std. Error  t value   Pr(>|t|)
(Intercept)                       3.256304e+01  1.061757e+00 30.66902 < 2.22e-16 ***
I(income - 5000)                  2.237579e-03  4.450897e-04  5.02726 2.3723e-06 ***
I(education - 8)                  4.537520e+00  3.406764e-01 13.31915 < 2.22e-16 ***
I(income - 5000):I(education - 8) -1.876822e-04  7.369712e-05 -2.54667     0.0125 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.243593 on 94 degrees of freedom
Multiple R-squared:  0.8260074,    Adjusted R-squared:  0.8204545
F-statistic:  148.751 on 3 and 94 DF,  p-value: < 2.2204e-16
```

Raw model

```
Coefficients:
                    Estimate    Std. Error  t value   Pr(>|t|)
(Intercept)      -2.243231e+01  6.557889e+00 -3.42066 0.00092597 ***
income            3.739037e-03  1.003440e-03  3.72622 0.00033170 ***
education         5.475932e+00  5.682087e-01  9.63718 1.0855e-15 ***
income:education -1.876822e-04  7.369712e-05 -2.54667 0.01250027 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.243593 on 94 degrees of freedom
Multiple R-squared:  0.8260074,    Adjusted R-squared:  0.8204545
F-statistic:  148.751 on 3 and 94 DF,  p-value: < 2.2204e-16
```

Centering with Only One Predictor

```
> lm.5 = lm(prestige ~ income, data = Prestige)
> lm.6 = lm(prestige ~ I(income - 5000), data = Prestige)


> summary(lm.5)$coefficients
                Estimate      Std. Error      t value         Pr(>|t|)
(Intercept) 27.596394991598 2.3803130069062 11.593599208 5.921240409e-20
income       0.002843574327 0.0002933466716  9.693562607 6.772967896e-16


> summary(lm.6)$coefficients
                     Estimate      Std. Error      t value         Pr(>|t|)
(Intercept)      41.814266625633 1.3587320151706 30.774476614 1.585609536e-51
I(income - 5000)  0.002843574327 0.0002933466716  9.693562607 6.772967896e-16
```

With only one predictor, only the intercept changes. The effect of income (raw or centered) remains the same…centering has no effect on the effect! A one-unit difference in the raw variable is the same as a one-unit difference on the centered variable.

Mean Centering

Applied researchers often use mean centering.

$$X_{\text{Mean Centered}} = X - \bar{X}$$

A value of zero on a mean centered variable would indicate that observation is at the mean. A negative value would indicate that observation was below the mean. And, a positive value would indicate the observation was above the mean.

```
> lm.7 = lm(prestige ~ I(income - mean(income)), data = Prestige)
> summary(lm.7)


Coefficients:
                              Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)               4.732755e+01 1.233964e+00 38.35407 < 2.22e-16 ***
I(income - mean(income)) 2.843574e-03 2.933467e-04  9.69356  6.773e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.21562 on 96 degrees of freedom
Multiple R-squared:  0.4946442,    Adjusted R-squared:   0.48938
F-statistic: 93.96516 on 1 and 96 DF,  p-value: 6.772968e-16
```

The average estimated prestige *for an occupation of average income* is 47.3.

A one-dollar difference in income is positively associated with a 0.003-unit difference in prestige, on average.

Centering with Multiple Predictors

```
> lm.8 = lm(prestige ~ income + education, data = Prestige)
> lm.9 = lm(prestige ~ I(income - mean(income)) + I(education - mean(education)),
     data = Prestige)


> summary(lm.8)$coefficients
                 Estimate      Std. Error       t value        Pr(>|t|)
(Intercept) -7.621035238457 3.1162308846861 -2.445593899 1.630283433e-02
income       0.001241536839 0.0002184936174  5.682256779 1.451954490e-07
education    4.292107598661 0.3360644972585 12.771678156 2.453044661e-22


> summary(lm.9)$coefficients
                                    Estimate      Std. Error       t value        Pr(>|t|)
(Intercept)                      47.327551020408 0.7525423402781 62.890216918 3.190258860e-79
I(income - mean(income))          0.001241536839 0.0002184936174  5.682256779 1.451954490e-07
I(education - mean(education))   4.292107598661 0.3360644972585 12.771678156 2.453044661e-22
```

With main-effects models, centering only affects the intercept—just like in a single predictor model. All other effects in the model (partial effects) are the same as they would be if we used the non-centered variables.

Now consider the fitted interaction model…

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(X_1) + \hat{\beta}_2(X_2) + \hat{\beta}_3(X_1)(X_2)$$

$\hat{\beta}_1$ is the expected difference in Y for a one-unit difference on X_1, holding all of the other terms constant.

This is a problem for the interaction term, since if X_1 varies (differs by one) so will $\hat{\beta}_3(X_1)(X_2)$

The only solution to this (holding the interaction constant) is to set X_2 (or conversely X_1 if you want to study the partial effect of X_2) to 0. Thus the interpretation of $\hat{\beta}_1$ is conditional on X_2 being 0.

This means the values and interpretation of the effects of $\hat{\beta}_1$ will change depending on where $X_2 = 0$ is located.

```
> Prestige$income.z = (Prestige$income - mean(Prestige$income)) / sd(Prestige$income)

> Prestige$education.z = (Prestige$education - mean(Prestige$education)) /
    sd(Prestige$education)



> head(Prestige)
                occupation prestige education type income  income.std education.std
1 government administrators     68.8     13.11 prof  12351 2.921147785   4.769117021
2         general managers     69.1     12.26 prof  25879 6.120669058   4.459906535
3              accountants     63.4     12.77 prof   9271 2.192693799   4.645432827
4      purchasing officers     56.8     11.42 prof   8865 2.096670319   4.154333820
5                 chemists     73.5     14.62 prof   8403 1.987402222   5.318420355
6               physicists     77.6     15.64 prof  11030 2.608716709   5.689472938
       incomeM     income.z  education.z
1  5412.142857 1.2800315051 0.8421067322
2 18940.142857 4.4795527777 0.5328962465
3  2332.142857 0.5515775194 0.7184225379
4  1926.142857 0.4555540394 0.2273235312
5  1464.142857 0.3462859416 1.3914100657
6  4091.142857 0.9676004288 1.7624626486
```

```
> lm.10 = lm(prestige ~ income.z + education.z + income.z:education.z, data = Prestige)


Coefficients:
                       Estimate Std. Error  t value   Pr(>|t|)
(Intercept)          48.5671126  0.8788165 55.26422 < 2.22e-16 ***
income.z              7.2427416  1.1914372  6.07900 2.5734e-08 ***
education.z          11.4730479  0.9073084 12.64515 < 2.22e-16 ***
income.z:education.z -2.1814059  0.8565719 -2.54667     0.0125 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.243593 on 94 degrees of freedom
Multiple R-squared:  0.8260074, Adjusted R-squared:  0.8204545
F-statistic:  148.751 on 3 and 94 DF,  p-value: < 2.2204e-16
```

The average estimated prestige *for an occupation of average income* and *average education-level* is 48.6.

The effect of income on prestige *depends on* the effect of education (or vice-versa).

Collinearity

Multicollinearity, or collinearity, occurs when a regression model includes two or more *highly related* predictors. Sometimes this is obvious. For example, including predictors of peer smoking *and* perceptions of school smoking norms, is likely to produce collinearity because these two predictors are highly correlated.

Other times, it isn't as obvious. For example, a weighted combination of several predictors might be related to another predictor. This would also produce collinearity issues, but isn't as detectable as just looking at the correlations matrix.

Collinearity produces unstable parameter estimates and inflated standard errors. What often happens is that predictors which are otherwise important predictors fail to be statistically reliable when included with other collinear predictors in the same model.

Go back to the initial model using the raw variables of income and education

```
                   Estimate Std. Error   t value Pr(>|t|)
(Intercept)      -22.432310   6.557889 -3.420660 0.000926
income             0.003739   0.001003  3.726220 0.000332
education          5.475932   0.568209  9.637183 0.000000
income:education  -0.000188   0.000074 -2.546670 0.012500
```

```
> cor(Prestige[c("prestige", "income", "education")])
              prestige       income    education
prestige  1.0000000000 0.7033094378 0.8664797685
income    0.7033094378 1.0000000000 0.5740978939
education 0.8664797685 0.5740978939 1.0000000000
```

Income and education are fairly highly correlated with each other. This might suggest that including both might produce misleading regression results…

Variance Inflation Factors

To examine how much of a problem collinearity might be, we examine the variance inflation factor for each predictor. This informs us how much larger the variance (SE^2) is in the particular model relative to a model in which all of the predictors are independent.

```
> library(car)
> vif(lm.1)
          income        education income:education
     33.276947976      4.510336048     47.051136899
```

The variance for the interaction effect is 47 times larger than would be expected in a model where income, education and the interaction were all unrelated.

Another way to consider this is to convert the variance inflation to the standard error inflation by computing the square root of each VIF

```
> sqrt(vif(lm.1))
          income        education income:education
     5.768617510      2.123755176      6.859383128
```

The SE for the interaction effect is 7 times larger than would be expected in a model where income, education and the interaction were all unrelated.

$$t = \frac{\hat{\beta}}{\mathrm{SE}_{\hat{\beta}}}$$

A larger SE, all things being equal, would result in a lower *t*-value….which means a larger *p*-value.

$$\hat{\beta} \pm 2(\mathrm{SE}_{\hat{\beta}})$$

A larger SE, all things being equal, would also result in a larger margin of error (wider CIs) for the estimates.

Possible Solution for Interaction Models

For interaction models, the easiest potential solution is two center any predictors that make up an interaction term…and to create the interaction term using those centered predictors.

```
Prestige$income.c = Prestige$income - mean(Prestige$income)
Prestige$education.c = Prestige$education - mean(Prestige$education)

lm.11 = lm(prestige ~ income.c + education.c + income.c:education.c, data = Prestige)

vif(lm.11)

           income.c          education.c income.c:education.c
        2.624254796          1.521855942          1.950390018
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-22.432310	6.557889	-3.420660	0.000926
income	0.003739	0.001003	3.726220	0.000332
education	5.475932	0.568209	9.637183	0.000000
income:education	-0.000188	0.000074	-2.546670	0.012500

Raw model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	48.567113	0.878817	55.264222	0.0000
income.c	0.001713	0.000282	6.078996	0.0000
education.c	4.173631	0.330058	12.645147	0.0000
income.c:education.c	-0.000188	0.000074	-2.546670	0.0125

Centered model

The information related to the interaction term does not change, but that for the main-effects does, alleviating our collinearity problems.