Vartanian: Data I
Partial Correlation and Regression

Partial Correlation Coefficient:

\[ r_{ij,k} = \frac{r_{ij} - (r_{ik})(r_{jk})}{\sqrt{1 - r_{ik}^2} \sqrt{1 - r_{jk}^2}} \]

Example:
Say we are trying to determine the effect of education on income, controlling for years of experience.

i=income
j=education
k=experience.

\( r_i = 0.50 \)
\( r_k = 0.40 \)
\( r_{jk} = -0.33 \)

\[ r_{ij,k} = \frac{0.50 - (0.40)(-0.33)}{\sqrt{1 - 0.40^2} \sqrt{1 - (-0.33)^2}} \]

\[ r_{ij,k} = \frac{0.632}{0.917 \times 0.944} = \frac{0.632}{0.866} \approx 0.73 \]

To get higher order partials, we would use the following formulas:

\[ r_{ij,kl} = \frac{r_{ij,k} - (r_{ik,l})(r_{jk,l})}{\sqrt{1 - r_{ij,k}^2} \sqrt{1 - r_{ij,l}^2}} \]
**Partial b coefficient:**
How do we determine the effect of each independent variable while controlling for all other variables?

\[
a_{ijk} = \bar{X}_i - b_{ijk} \bar{X}_j - b_{ikj} \bar{X}_k
\]

\[
b_{ijk} = \frac{b_{ij} - (b_{ik})(b_{kj})}{1 - b_{jk}b_{ij}}
\]

\[
b_{ikj} = \frac{b_{ik} - (b_{ij})(b_{jk})}{1 - b_{kj}b_{ik}}
\]
Example:

Let's say we have 3 variables:

Y: Income
X1: education
X2: experience.

We are going to determine the effect of education on income, controlling for experience.

run bivariate regressions, where the left hand side variable is the dependent variable and the right hand side variable is the independent variable:

\[ Y = X1 \quad \rightarrow \quad b_{ij} \]
\[ Y = X2 \quad \rightarrow \quad b_{ik} \]
\[ X1 = X2 \quad \rightarrow \quad b_{jk} \]
\[ X2 = X1 \quad \rightarrow \quad b_{kj} \]

Here, \( i \) refers to the dependent variable income.
\( j \) refers to education
\( k \) refers to experience.

We get our beta coefficient estimates for each of the regressions.

Let's say that

\[ Y \text{ bar} = 20,000 \]
\[ X1 \text{ bar} = 12 \]
\[ X2 \text{ bar} = 15 \]

\[ b_{ij} = 0.67 \]
\[ b_{ik} = 0.82 \]
\[ b_{jk} = -0.06 \]
\[ b_{kj} = -0.13 \]

\[ b_{yx1,x2} = b_{ij,k} \]

\[ 0.67 - (0.82)(-0.13) \over [1 - (-0.06)(-0.13)] \]

\[ .67 + .1066 / 1-.0078 = .7766/.9922 = .7688 \]

This indicates the partial b coefficient estimate for \( x_1 \) as the independent variable and \( y \) as the dependent variable, controlling for \( x_2 \), is .7688. In other words, as \( x_1 \) increases by
one unit, y will increase by .7688 units.

\[ b_{yx2,x1} = b_{ik,j} = \]

\[ [0.82 - (0.67)(-0.06)] \text{ over } [1 - (-0.13)(-0.06)] = \]

\[ [0.8602 \text{ over } 0.99] = 0.86696 \]

This gives us the b coefficient estimate for \( x_2 \) as the independent variable and \( y \) as the dependent variable, while we control for \( x_1 \). As \( x_2 \) increases by one unit, \( y \) will increase by .86696 units.

\[ a_{y,x1x2} = 20.00 - .7688(12) - 0.86696(15) = 1997.77 \]

This is the intercept for the regression equation.

****************************************************************

**Standardized Beta Coefficients**

Example:

<table>
<thead>
<tr>
<th>Spell</th>
<th>Std Dev</th>
<th>B</th>
<th>Beta</th>
<th>Sign t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kids</td>
<td>1.43</td>
<td>0.027133</td>
<td>0.010463</td>
<td>.7782</td>
</tr>
<tr>
<td>Edlevel</td>
<td>0.69</td>
<td>-0.316962</td>
<td>-0.201974</td>
<td>.1170</td>
</tr>
<tr>
<td>Age</td>
<td>9.63</td>
<td>-0.010321</td>
<td>-0.026814</td>
<td>.4715</td>
</tr>
<tr>
<td>Workb</td>
<td>686.80</td>
<td>-0.000897</td>
<td>-0.166245</td>
<td>.0000</td>
</tr>
</tbody>
</table>

\[ \beta_{ij,k} = b_{ij,k} \frac{s_j}{s_i} \]

Where \( i \) is the dependent variable and \( j \) is the independent variable. We multiply the partial b coefficient estimate by the standard deviation of the variable we’re examining (the independent variable in question) divided by the standard deviation of the dependent variable.

By standardizing the coefficient, we are better able to determine the relative effect of each of the independent variables at a glance. Remember that the standardized beta's are in standard deviation units. For example, a one standard deviation increase in age will decrease spell length by .1443 SD units. A one standard deviation increase in the number of kids will decrease spell length by .0982 standard deviation units. From the unstandardized b's we know, for example, that a one unit increase in age will decrease
spell length by .0643 years.

**Multiple Correlation:**

\[ R_{1,23}^2 = r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23} (1 - r_{12}^2) \]

<table>
<thead>
<tr>
<th>Prop explain by 2 and</th>
<th>Prop expl by 2</th>
<th>Additional explained by 3</th>
<th>Proportion unexplained by 2</th>
</tr>
</thead>
</table>

How much of the dependent variable, which is variable number 1, is explained by variables 2 and 3.

\[ R_{1,23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{12}^2} \]

**Example:**

We know the correlation coefficients for 3 variables:

- Income and education: 0.62
- Income and experience: 0.42
- Education and experience: -0.12

How do we then determine an \( R^2 \) value? How much of the variance do the two variables explain?

\[ R_{1,23}^2 = \frac{0.62^2 + 0.42^2 - 2(0.62 \times 0.42 \times -0.12)}{1 - (-0.12)^2} \]

\[ R_{1,23}^2 = \frac{0.3844 + 0.1764 + 0.03125}{0.9856} = \frac{0.6233}{0.9856} = 0.6324 \]
**Determining the Corrected Multiple Coefficient:**

\[
\overline{R}^2 = R^2 - \frac{k}{N-k-1} (1 - R^2)
\]

Where \( k \) = number of independent variables. Basically, \( \overline{R}^2 \) will decrease when the variables we add only contribute a small amount to the explained variance. We’re correcting \( R^2 \) by the number of degrees of freedom. Why is this measure used? Because as you add more and more independent variables into your regression model, the uncorrected \( R^2 \) value will either stay the same or go up. The corrected value will stop us from simply adding more and more variables into our models since the value of the corrected \( R^2 \) will decrease when these additional variables adding little or nothing to the explained variance of the dependent variable.