Statistical Control Techniques: Partial and Semipartial Correlation

Chapter 7

Today's Lecture

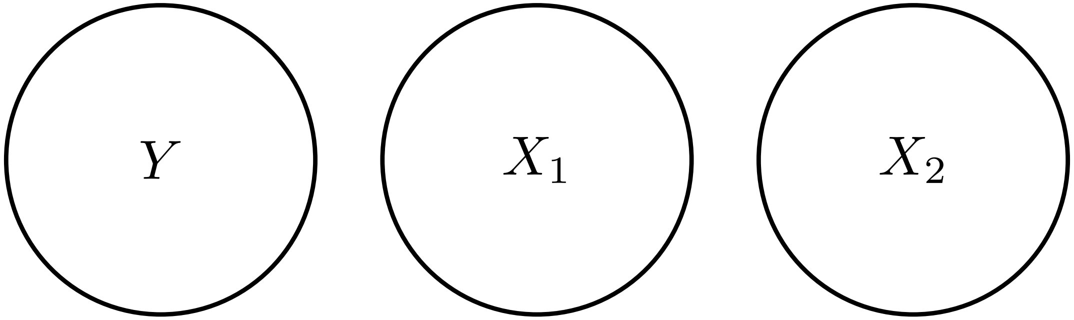
- Statistical control techniques.
- Partial correlation.
- Semipartial correlation.

Statistical Control

- Statistical control can be equated to controlling for variability.
- In experimental research (with human subjects), this is often accomplished by random assignment to a set of experimenter controlled conditions.
- In quasi-experimental research random assignment cannot be accomplished, although experimental conditions may exist.
- In nonexperimental or observational research, the “experimenter” does not control anything...but can observe many things.
- To obtain control when none is present, statistics can be useful.

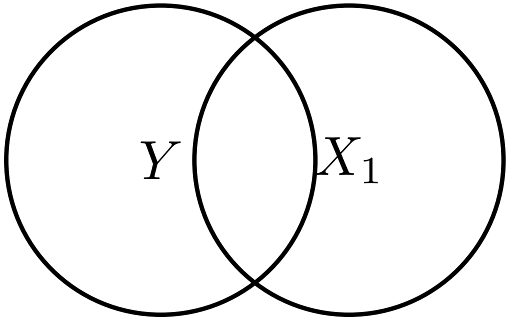
Venn Diagrams

One way to visualize the concepts described in this lecture is to imagine the variability of a variable as being represented by a circle.



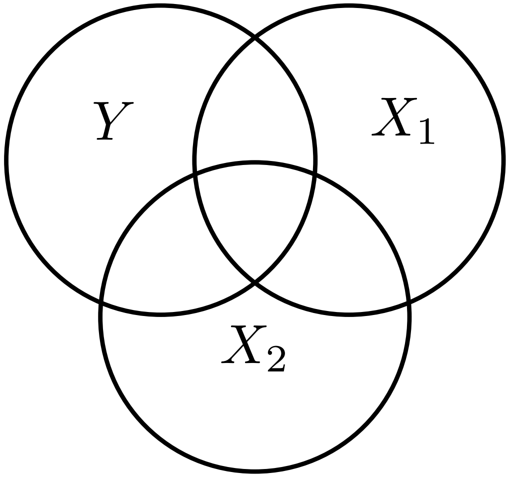
Venn Diagrams

Using X_1 to predict Y:



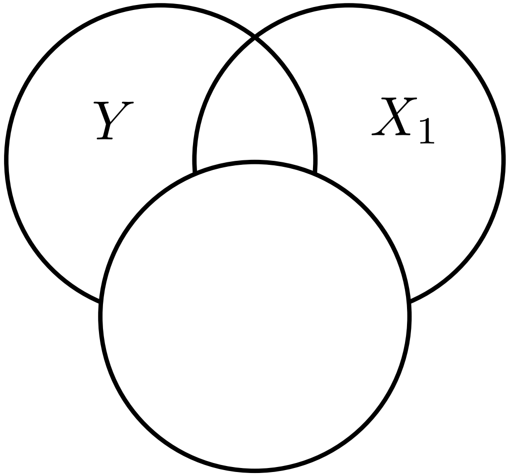
Venn Diagrams

Using X_1 and X_2 to predict Y:



Venn Diagrams

Using X_1 to predict Y, while controlling for the effects of X_2:



Today’s Example Data Set

From Weisberg (1985, p. 240).

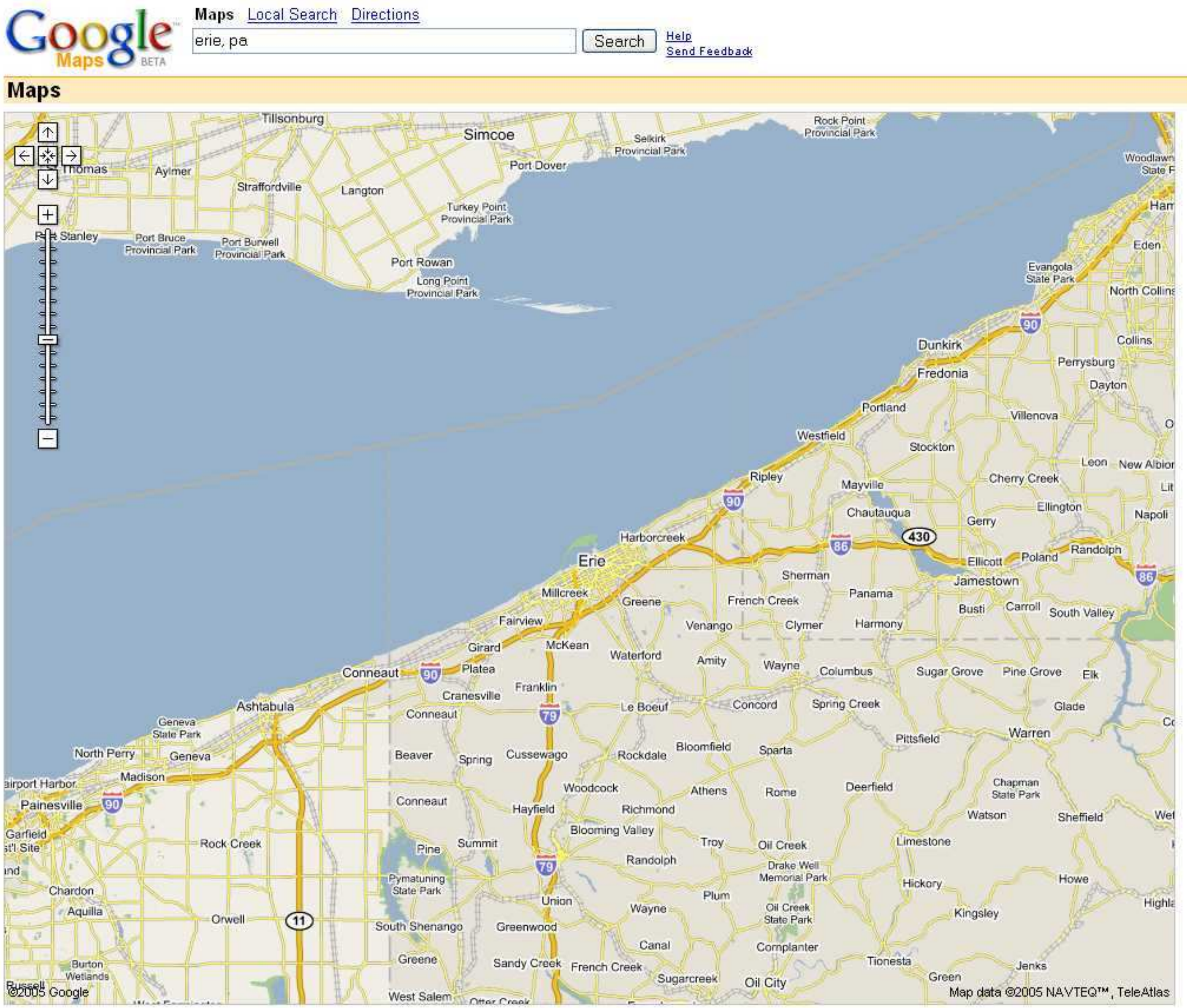
“Property taxes on a house are supposedly dependent on the current market value of the house. Since houses actually sell only rarely, the sale price of each house must be estimated every year when property taxes are set. Regression methods are sometimes used to make up a prediction function.”

Erie, Pennsylvania

We have data for 27 houses sold in the mid 1970's in Erie, Pennsylvania:

- X_1: Current taxes (local, school, and county) $\div$ 100 (dollars).
- X_2: Number of bathrooms.
- X_3: Lot size $\div$ 1000 (square feet).
- X_4: Living space $\div$ 1000 (square feet).
- X_5: Number of garage spaces.
- X_6: Number of rooms.
- X_7: Number of bedrooms.
- X_8: Age of house (years).
- X_9: Number of fireplaces.
- Y: Actual sale price $\div$ 1000 (dollars).

Erie, Pennsylvania



Lake Erie

Erie, Pennsylvania

Jack

Erie, Pennsylvania

Paula

Erie, Pennsylvania

For our example today, we will limit the size of the data to a smaller set of variables:

- X_1: Current taxes (local, school, and county) $\div$ 100 (dollars).
- X_2: Number of bathrooms.
- X_3: Living space $\div$ 1000 (square feet).
- X_4: Age of house (years).
- Y: Actual sale price $\div$ 1000 (dollars).

Correlation Matrix

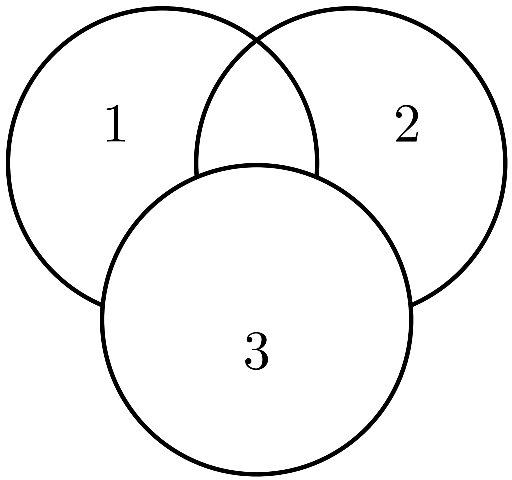
Variable	X_1	X_2	X_3	X_4	Y
X_1	1.000				
X_2	0.876	1.000			
X_3	0.832	0.901	1.000		
X_4	-0.371	-0.211	-0.178	1.000	
Y	0.915	0.924	0.929	-0.310	1.000

- X_1: Current taxes (local, school, and county) $\div$ 100 (dollars).
- X_2: Number of bathrooms.
- X_3: Living space $\div$ 1000 (square feet).
- X_4: Age of house (years).
- Y: Actual sale price $\div$ 1000 (dollars).

Partial Correlation

A partial correlation is a correlation between two variables from which the linear relations, or effects, of another variable(s) have been removed.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$



Partial Correlation in SPSS

- To perform a partial correlation in SPSS go to Analyze...Correlation...Partial.
- In the Variables box, put the variables you would like to get partial correlations for.
- In the Controlling For box, put the variables you would like to control for.

Partial correlation of $Y(1)$ (actual sale price $\div$ \$1000) with $X_3(2)$ (living space), controlling for $X_1(3)$ (current taxes):

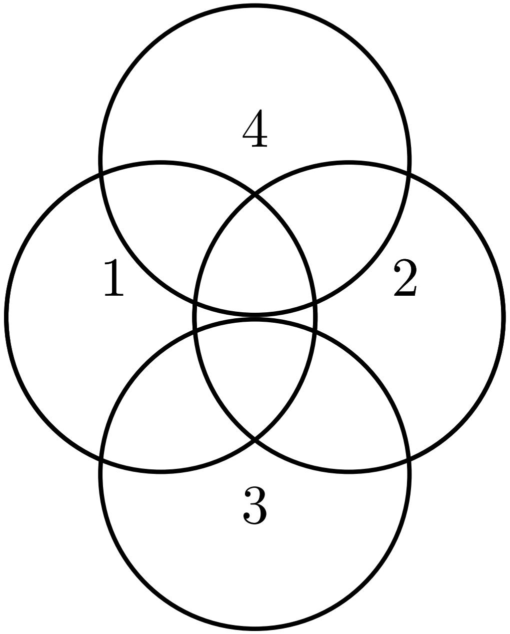
$$r_{yx_3.x_1} = \frac{r_{yx_3} - r_{yx_1} r_{x_3x_1}}{\sqrt{1 - r^2_{yx_1}}\sqrt{1 - r^2_{x_3x_1}}} =$$

$$\frac{0.929 - (0.915 \times 0.832)}{\sqrt{1 - 0.915^2}\sqrt{1 - 0.832^2}} = 0.748$$

Higher-Order Partial Correlation

The second-order partial correlation is the correlation between two variables with the effects of two other variables being removed:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{1 - r_{14.3}^2}\sqrt{1 - r_{24.3}^2}}$$



Second-order Partial Correlation

- To perform a partial correlation in SPSS go to Analyze...Correlation...Partial.
- In the Variables box, put the variables you would like to get partial correlations for.
- In the Controlling For box, put the variables you would like to control for - if more than one, a higher-order partial is given.

Second-order Partial Correlation

Partial correlation of $Y(1)$ (actual sale price ÷ \$1000) with $X_3(2)$ (living space), controlling for $X_1(3)$ (current taxes) and $X_2(4)$ (number of bathrooms):

$$r_{12.3} = r_{yx_3.x_1} = 0.748$$

$$r_{14.3} = r_{yx_2.x_1} = 0.628$$

$$r_{24.3} = r_{x_3x_2.x_1} = 0.641$$

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{1 - r_{14.3}^2}\sqrt{1 - r_{24.3}^2}} =$$

$$\frac{0.748 - (0.628 \times 0.641)}{\sqrt{1 - 0.628^2}\sqrt{1 - 0.641^2}} = 0.578$$

Partial Correlations from Multiple Correlation

As you could probably guess, partial correlations can be found from regression analyses...

Squared partial correlation:

$$r^2_{12.3} = \frac{R^2_{1.23} - R^2_{1.3}}{1 - R^2_{1.3}}$$

- $R^2_{1.23}$ is the R^2 from a multiple regression with 1 being Y and 2 and 3 being the predictor variables.
- $R^2_{1.3}$ is the R^2 from a simple regression with 1 being Y and 2 being X - the single predictor variable.

Partial Correlations from Multiple Correlation

Partial correlation of $Y(1)$ (actual sale price $\div$ \$1000) with $X_3(2)$ (living space), controlling for $X_1(3)$ (current taxes):

$$R^2_{1.23} = 0.928$$

$$R^2_{1.3} = 0.838$$

$$r^2_{12.3} = \frac{R^2_{1.23} - R^2_{1.3}}{1 - R^2_{1.3}} = \frac{0.928 - 0.838}{1 - 0.838} = 0.865$$

$$r_{12.3} = \sqrt{r^2_{12.3}} = \sqrt{0.865} = 0.748$$

Partial Correlations from Multiple Correlation

Alternatively:

$$r^2_{12.3} = \frac{R^2_{2.13} - R^2_{2.3}}{1 - R^2_{2.3}}$$

Partial correlation of $Y(1)$ (actual sale price $\div$ \$1000) with $X_3(2)$ (living space), controlling for $X_1(3)$ (current taxes):

$$R^2_{2.13} = 0.865$$

$$R^2_{2.3} = 0.693$$

$$r^2_{12.3} = \frac{R^2_{2.13} - R^2_{2.3}}{1 - R^2_{2.3}} = \frac{0.865 - 0.693}{1 - 0.693} = 0.865$$

$$r_{12.3} = \sqrt{r^2_{12.3}} = \sqrt{0.865} = 0.748$$

Higher-Order Partial Correlations

Second-order squared partial correlation:

$$r^2_{12.34} = \frac{R^2_{1.234} - R^2_{1.34}}{1 - R^2_{1.34}}$$

$$r^2_{12.34} = \frac{R^2_{2.134} - R^2_{2.34}}{1 - R^2_{2.34}}$$

Third-order squared partial correlation:

$$r^2_{12.345} = \frac{R^2_{1.2345} - R^2_{1.345}}{1 - R^2_{1.345}}$$

$$r^2_{12.345} = \frac{R^2_{2.1345} - R^2_{2.345}}{1 - R^2_{2.345}}$$

Higher-Order Partial Correlations

Partial correlation of $Y(1)$ (actual sale price ÷ \$1000) with $X_3(2)$ (living space), controlling for $X_1(3)$ (current taxes) and $X_2(4)$ (number of bathrooms):

$$R^2_{1.234} = 0.935$$

$$R^2_{1.34} = 0.902$$

$$r^2_{12.34} = \frac{R^2_{1.234} - R^2_{1.34}}{1 - R^2_{1.34}} = \frac{0.935 - 0.901}{1 - 0.902} = 0.337$$

$$r_{12.34} = \sqrt{r^2_{12.34}} = \sqrt{0.337} = 0.578$$

Partial Correlations in Regression

You can obtain partial correlations (and semipartial correlations - called part correlations by SPSS) from the regression analysis subroutine:

- Go to Analyze...Regression...Linear.
- Click on Statistics box.
- Check Part and Partial Correlations.

Correction For Attenuation

- Often in behavioral research, variables are not measured without error (non-perfect reliability).
- For these times, the correlation between two variables is dampened by extra noise associated with measurement.
- To correct for this noise, a so-called correction for attenuation accounts for the imperfections in measurement.
- Using these corrected values, one can then build corrected partial correlations.

Correction for attenuation:

$$r_{12}^{*} = \frac{r_{12}}{\sqrt{r_{11}}\sqrt{r_{22}}}$$

Corrected Partial Correlation

Recall the formula for the first-order partial correlation:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

First-order partial correlation correction for attenuation:

$$r_{12.3}^* = \frac{r_{33}r_{12} - r_{13}r_{23}}{\sqrt{r_{11}r_{33} - r_{13}^2}\sqrt{r_{22}r_{33} - r_{23}^2}}$$

Semipartial Correlation

- Semipartial correlation removes the effects of additional variables from one of the variables under study (typically X).

First-order semipartial correlation:

$$r_{1(2.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{23}^2}}$$

Or...

$$r_{2(1.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}}$$

Semipartial Correlation from Regression

First-order semipartial correlation:

$$r^2_{1(2.3)} = R^2_{1.23} - R^2_{1.3}$$

Second-order semipartial correlation:

$$r^2_{1(2.34)} = R^2_{1.234} - R^2_{1.34}$$

Third-order semipartial:

$$r^2_{1(2.345)} = R^2_{1.2345} - R^2_{1.345}$$

Semipartial Correlation from Regression

Semipartial correlation of $Y(1)$ (actual sale price $\div$ \$1000) with $X_3(2)$ (living space), controlling $X_3(2)$ (living space) for $X_1(3)$ (current taxes):

$$R^2_{1.23} = 0.928$$

$$R^2_{1.3} = 0.838$$

$$r^2_{1(2.3)} = R^2_{1.23} - R^2_{1.3} = 0.928 - 0.838 = 0.090$$

$$r_{1(2.3)} = \sqrt{r^2_{1(2.3)}} = \sqrt{0.090} = 0.300$$

Semipartial Correlation from Regression

Semipartial correlation of $Y(1)$ (actual sale price $\div$ \$1000) with $X_3(2)$ (living space), controlling $X_3(2)$ (living space) for $X_1(3)$ (current taxes) and $X_2(4)$ (number of bathrooms):

$$R^2_{1.234} = 0.935$$

$$R^2_{1.34} = 0.902$$

$$r^2_{1(2.34)} = R^2_{1.234} - R^2_{1.34} = 0.935 - 0.902 = 0.033$$

$$r_{1(2.3)} = \sqrt{r^2_{1(2.3)}} = \sqrt{0.090} = 0.182$$

Tests of Significance

Semipartial correlations are found using the formulas developed for the test of the increment in the proportion of variance accounted for in regression:

$$F = \frac{R^2_{y.12\ldots k_1} - R^2_{y.12\ldots k_2}/(k_1 - k_2)}{1 - R^2_{y.12\ldots k_1}/(N - k_1 - 1)},$$

where $k_1 > k_2$, with:

- $df_{num} = k_1 - k_2$
- $df_{denom} = N - k_1 - 1$

Multiple Regression

You can begin to see how multiple regression uses the statical concepts from today's to control the prediction of the Y variable:

$$R^2_{y.1234} = r^2_{y1} + r^2_{y(2.1)} + r^2_{y(3.12)} + r^2_{y(3.123)}$$

Multiple Partial Correlation

Multiple partial correlation:

$$R^2_{1.23(4)} = \frac{R^2_{1.234} - R^2_{1.4}}{\sqrt{1 - R^2_{1.4}}}$$

Two control variables:

$$R^2_{1.23(45)} = \frac{R^2_{1.2345} - R^2_{1.45}}{\sqrt{1 - R^2_{1.45}}}$$

Multiple Semipartial Correlation

Multiple semipartial correlation:

$$R^2_{1(23.4)} = R^2_{1.234} - R^2_{1.4}$$

Two control variables:

$$R^2_{1(23.45)} = R^2_{1.2345} - R^2_{1.45}$$

Final Thought

- Often times, experimental control is not possible.
- Using statistical techniques, one can get an idea of how variables are related to one another by removing the effects of spurious variables

- Partial and semipartial are statistical techniques to provide post-hoc control.

Next Time

- Midterm handed out.
- Midterm discussion.