Correlation Theory

Now tri-variate

X ~ random variable

Y ~ random variable

Z ~ random variable

Covered Topics

- Partial Correlation (Pearson)

Partial Correlation (1)

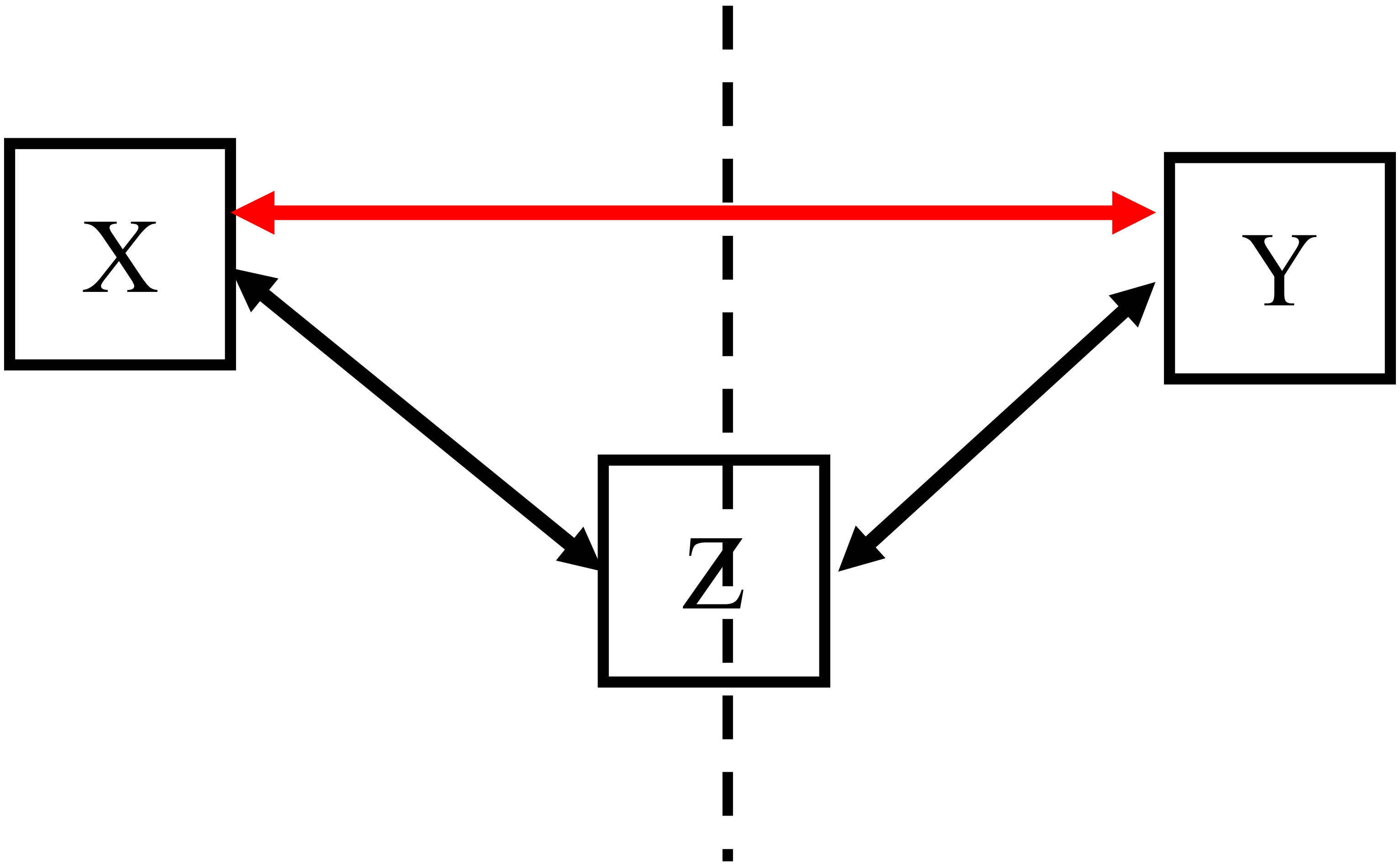
- X, Y and Z are three RV's
- They are assumed to be related.
- Ignoring Z, Corr (X,Y) = "direct" correlation (X,Y) + indirect effects from Z

Partial Correlation (2)

Partial Correlation

- Correlation when embedded effect of Z has been removed from both X and Y
- Parital Corr (X,Y) = “direct” corr(X,Y)

Partial Correlation (3)



Partial Correlation (4)

Assumptions

- X.Z are related as $X = \alpha_1 + \alpha_2 Z + \varepsilon$
- Y.Z are related as $Y = \beta_1 + \beta_2 Z + \xi$

Partial Correlation (5)

PCorr

- X.Z are related as $X = \alpha_1 + \alpha_2 Z + \varepsilon$
- Y.Z are related as $Y = \beta_1 + \beta_2 Z + \xi$

Partial Correlation (6)

By OLS

$$\hat{\alpha}_2 = \frac{\sum x_i z_i}{\sum z_i^2}, \hat{\alpha}_1 = \bar{X} - \hat{\alpha}_2 \bar{Z}$$

where $x_i = X_i - \bar{X},$

$$z_i = Z_i - \bar{Z}$$

Partial Correlation (7)

By OLS

$$\hat{\beta}_2 = \frac{\sum y_i z_i}{\sum z_i^2}, \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{Z}$$

where $y_i = y_i - \bar{Y},$

$$z_i = Z_i - \bar{Z}$$

Partial Correlation (8)

X without Z $$X - \alpha_2 Z = \alpha_1 + \varepsilon$$

Y without Z $$Y - \beta_2 Z = \beta_1 + \xi$$

Partial Corr between X and Y $(\rho_{XY.Z})$

$$= corr(\alpha_2 + \varepsilon, \beta_1 + \xi) = corr(\varepsilon, \xi)$$

Partial Correlation (9)

Sample partial correlation coefficient

$$r_{XY.Z} = \frac{\sum\left(X_i - \hat{X}_i\right)\left(Y_i - \hat{Y}_i\right)}{\sqrt{\sum\left(X_i - \hat{X}_i\right)^2}\sqrt{\sum\left(Y_i - \hat{Y}_i\right)^2}}$$

Partial Correlation (9)

Sample partial correlation coefficient

$$r_{XY.Z} = \frac{\sum \left(x_i - \hat{\alpha}_2 z_i\right)\left(y_i - \hat{\beta}_2 z_i\right)}{\sqrt{\sum \left(x_i - \hat{\alpha}_2 z_i\right)^2} \sqrt{\sum \left(y_i - \hat{\beta}_2 z_i\right)^2}}$$

Partial Correlation (9)

Sample partial correlation coefficient

$$\sum\left(x_i-\hat{\alpha}_2 z_i\right)^2=\sum x_i^2-2\hat{\alpha}_2\sum x_i z_i+\hat{\alpha}_2^2\sum z_i^2$$

$$=\sum x_i^2-2\frac{\sum x_i z_i}{\sum z_i^2}\sum x_i z_i+\left(\frac{\sum x_i z_i}{\sum z_i^2}\right)^2\sum z_i^2$$

$$=\sum x_i^2-\frac{\left(\sum x_i z_i\right)^2}{\sum z_i^2}$$

Partial Correlation (10)

Sample partial correlation coefficient

$$= \sum x_i^2 - \frac{\left(\sum x_i z_i\right)^2}{\sum x_i^2 \sum z_i^2} \sum x_i^2$$

$$= \left(1 - r_{XZ}^2\right) \sum x_i^2$$

Partial Correlation (11)

Sample partial correlation coefficient

$$\sum \left(y_i - \hat{\alpha}_2 z_i \right)^2 = \left(1 - r_{YZ}^2 \right) \sum y_i^2$$

Partial Correlation (12)

Sample partial correlation coefficient

$$\sum\left(x_i-\hat{\alpha}_2 z_i\right)\left(y_i-\hat{\beta}_2 z_i\right)$$

$$=\sum x_i y_i-\hat{\alpha}_2\sum y_i z_i$$

$$-\hat{\beta}_2\sum x_i z_i+\hat{\alpha}_2\hat{\beta}_2\sum z_i^2$$

Partial Correlation (13)

Sample partial correlation coefficient

$$= \sum x_i y_i - \frac{\sum x_i z_i}{\sum z_i^2} \sum y_i z_i$$

$$- \frac{\sum y_i z_i}{\sum z_i^2} \sum x_i z_i + \frac{\sum x_i z_i}{\sum z_i^2} \frac{\sum y_i z_i}{\sum z_i^2} \sum z_i^2$$

$$= \sum x_i y_i - \frac{\sum y_i z_i}{\sum z_i^2} \sum x_i z_i$$

Partial Correlation (14)

Sample partial correlation coefficient

$$= r_{YX}\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}$$

$$-\frac{r_{YZ}\sqrt{\sum y_i^2}\sqrt{\sum z_i^2}}{\sum z_i^2} r_{XZ}\sqrt{\sum x_i^2}\sqrt{\sum z_i^2}$$

$$= \left(r_{YX} - r_{YZ}r_{XZ}\right)\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}$$

Partial Correlation (15)

Sample partial correlation coefficient

$$r_{XY.Z} = \frac{\left(r_{YX} - r_{YZ}r_{XZ}\right)\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}{\sqrt{\left(1-r_{XZ}^2\right)\sum x_i^2}\sqrt{\left(1-r_{YZ}^2\right)\sum y_i^2}}$$

$$= \frac{r_{YX} - r_{YZ}r_{XZ}}{\sqrt{1-r_{XZ}^2}\sqrt{1-r_{YZ}^2}}$$