Key Algebraic Results in Linear Regression

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Key Algebraic Results in Linear Regression

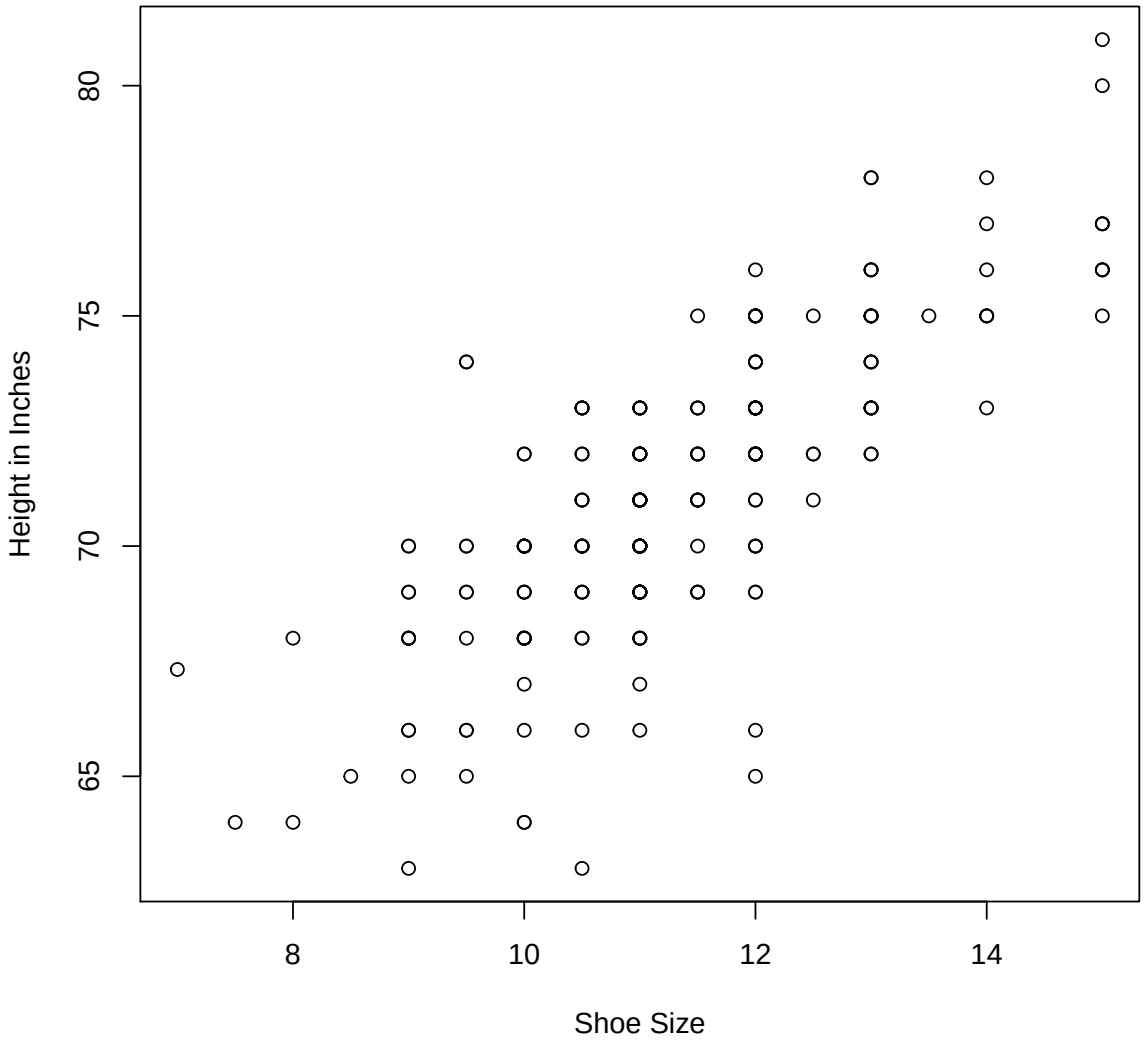
Introduction

- In this module, we explore the algebra of least squares linear regression systems with a special eye toward developing the properties useful for deriving factor analysis and structural equation modeling.
- A key insight is that important properties hold whether or not variables are observed.

Bivariate Linear Regression

- In bivariate linear regression performed on a sample of n observations, we seek to examine the extent of the linear relationship between two observed variables, X and Y.
- One variable (usually the one labeled Y) is the *dependent* or *criterion* variable, the other (usually labeled X) is the *independent* or *predictor* variable.
- Each data point represents a pair of scores, x_i, y_i that may be plotted as a point in the plane. Such a plot, called a *scatterplot*, is shown on the next slide.
- In these data, gathered on a group of male college students, the independent variable plotted on the horizontal (X) axis is shoe size, and the dependent variable plotted on the vertical (Y) axis is height in inches.

Bivariate Linear Regression



Bivariate Linear Regression

- It would be a rare event, indeed, if all the points fell on a straight line. However, if Y and X have an approximate linear relationship, then a straight line, properly placed, should fall close to many of the points.
- Choosing a straight line involves choosing the slope and intercept, since these two parameters define any straight line.
- The regression model in the sample is that

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \tag{1}$$

- Generally, the *least squares* criterion, minimizing $\sum_{i=1}^{n} e_i^2$ under choice of $\hat{\beta}_0$ and $\hat{\beta}_1$, is employed.
- Minimizing $\sum_{i=1}^{n} e_i^2$ is accomplished with the following well-known *least squares solution*.

$$\hat{\beta}_1 = \frac{r_{Y,X} S_Y}{S_X} = \frac{s_{Y,X}}{s_X^2} = s_{X,X}^{-1} s_{x,y} \tag{2}$$

$$\hat{\beta}_0 = \overline{Y}_\bullet - \beta_1 \overline{X}_\bullet \tag{3}$$

Bivariate Linear Regression

Deviation Score Formulas

- Suppose we were to convert X into deviation score form. This would have no effect on any variance, covariance or correlation involving X, but would change the mean of X to zero.
- What would be the effect on the least squares regression?
- Defining $x_i^* = x_i - \overline{X}_\bullet$, we have the new least squares setup

$$y_i = \hat{\beta}_0^* + \hat{\beta}_1^* x_i^* + e_i^* \tag{4}$$

- From the previous slide, we know that
  $\hat{\beta}_1^* = S_{Y,X^*}/S_{X^*,X^*} = S_{Y,X}/S_{X,X} = \hat{\beta}_1$, and that
  $\hat{\beta}_0^* = \overline{Y}_\bullet - \hat{\beta}_1^* \overline{X}_\bullet^* = \overline{Y}_\bullet$.
- Thus, if X is shifted to deviation score form, the slope of the regression line remains unchanged, but the intercept shifts to $\overline{Y}_\bullet$.
- It is easy to see that, should we also re-express the Y variable in deviation score form, the regression line intercept will shift to zero and the slope will *still* remain unchanged.

Bivariate Linear Regression

Variance of Predicted Scores

- Using linear transformation rules, one may derive expressions for the variance of the predicted ($\hat{y}_i$) scores, the residual (e_i) scores, and the covariance between them.
- For example consider the variance of the predicted scores. Remember that adding a constant (in this case $\hat{\beta}_0$) has no effect on a variance, and multiplying by a constant multiplies the variance by the square of the multiplier. So, since $\hat{y}_i = \hat{\beta}_1 x_i + \hat{\beta}_0$, it follows immediately that

$$\begin{aligned} s^2_{\hat{Y}} &= \hat{\beta}_1^2 S_X^2 \\ &= (r_{Y,X} S_Y / S_X)^2 S_X^2 \\ &= r^2_{Y,X} S_Y^2 \end{aligned} \tag{5}$$

Bivariate Linear Regression

Covariance of Predicted and Criterion Scores

- The covariance between the criterion scores (y_i) and predicted scores ($\hat{y}_i$) is obtained by the heuristic rule.
- Begin by re-expressing $\hat{y}_i$ as $\beta_1 x_i + \beta_0$, then recall that additive constant β_0 cannot affect a covariance.
- So the covariance between y_i and $\hat{y}_i$ is the same as the covariance between y_i and $\hat{\beta}_1 x_i$.
- Using the heuristic approach, we find that $S_{Y,\hat{Y}} = S_{Y,\hat{\beta}_1 X} = \hat{\beta}_1 S_{Y,X}$ Recalling that $S_{Y,X} = r_{Y,X} S_Y S_X$, and $\hat{\beta}_1 = r_{Y,X} S_Y / S_X$, one quickly arrives at

$$\begin{aligned} S_{Y,\hat{Y}} &= \hat{\beta}_1 S_{Y,X} \\ &= (r_{Y,X} S_Y S_X)(r_{Y,X} S_Y / S_X) \\ &= r^2_{Y,X} S^2_Y \\ &= S^2_{\hat{Y}} \end{aligned} \tag{6}$$

Bivariate Linear Regression

Covariance of Predicted and Residual Scores

- Calculation of the covariance between the predicted scores and residual scores proceeds in much the same way. Re-express e_i as $y_i - \hat{y}_i$, then use the heuristic rule. One obtains

$$\begin{aligned} S_{\hat{Y},E} &= S_{\hat{Y},Y-\hat{Y}} \\ &= S_{\hat{Y},Y} - S^2_{\hat{Y}} \\ &= S^2_{\hat{Y}} - S^2_{\hat{Y}} \quad \text{(from Equation 6)} \\ &= 0 \end{aligned} \tag{7}$$

Bivariate Linear Regression

Covariance of Predicted and Residual Scores

- Calculation of the covariance between the predicted scores and residual scores proceeds in much the same way.
- Re-express e_i as $y_i - \hat{Y}_i$, then use the heuristic rule. One obtains

$$\begin{aligned} S_{\hat{Y},E} &= S_{\hat{Y},y-\hat{Y}} \\ &= S_{\hat{Y},y} - S^2_{\hat{Y}} \\ &= S^2_{\hat{Y}} - S^2_{\hat{Y}} \quad \text{(from Equation 6)} \\ &= 0 \end{aligned} \tag{8}$$

- Predicted and error scores always have exactly zero covariance, and zero correlation, in linear regression.

Bivariate Linear Regression

Additivity of Variances

- Linear regression partitions the variance of Y into non-overlapping portions.
- Using a similar approach to the previous proofs, we may show easily that

$$S_Y^2 = S_{\hat{Y}}^2 + S_E^2 \tag{9}$$

Multiple Linear Regression

- Multiple linear regression with a single criterion variable and several predictors is a straightforward generalization of bivariatelinear regression.
- To make the notation simpler, assume that the criterion variable Y and the p predictor variables X_j, $j = 1, \ldots, p$ are in deviation score form.
- Let $\mathbf{y}$ be an $n \times 1$ vector of criterion scores, and $\mathbf{X}$ be the $n \times p$ matrix with the predictor variables in columns. Then the multiple regression prediction equation in the sample is

$$\begin{aligned} \mathbf{y} &= \hat{\mathbf{y}} + \mathbf{e} \\ &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \end{aligned} \tag{10}$$

Multiple Linear Regression

- The least squares criterion remains essentially as before, i.e., minimize $\sum e_i^2 = \mathbf{e}'\mathbf{e}$ under choice of $\hat{\boldsymbol{\beta}}$. The unique solution is

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \tag{11}$$

which may also be written as

$$\hat{\boldsymbol{\beta}} = \mathbf{S}_{XX}^{-1}\mathbf{S}_{XY} \tag{12}$$

Multivariate Linear Regression

- The notation for multiple linear regression with a single criterion generalizes immediately to situations where more than one criterion is being predicted simultaneously.
- Specifically, let $n \times q$ matrix $\mathbf{Y}$ contain q criterion variables, and let $\hat{\boldsymbol{\beta}}$ be a $p \times q$ matrix of regression weights. The least squares criterion is satisfied when the sum of squared errors across all variables (i.e. $\text{Tr}(\mathbf{E}'\mathbf{E})$) is minimized.
- The unique solution is the obvious generalization of Equation 11, i.e.,

$$\hat{\mathbf{B}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \tag{13}$$

Multivariate Linear Regression

- We will now prove some multivariate generalizations of the properties we developed earlier for bivariate linear regression systems.
- First, we prove that $\hat{\mathbf{Y}} = \mathbf{XB}$ and $\mathbf{E} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}$ are uncorrelated. To do this, we examine the covariance matrix between them, and prove that it is a null matrix. Recall from the definition of the sample covariance matrix that, when scores in $\mathbf{Y}$ and $\mathbf{X}$ are in deviation score form, that $\mathbf{S_{YX}} = 1/(n-1)\mathbf{Y}'\mathbf{X}$. Hence, (moving the $n-1$ to the left of the formula for simplicity),

Multivariate Linear Regression

$$
\begin{aligned}
(n-1)\mathbf{S_{YE}} &= \hat{\mathbf{Y}}'\mathbf{E} \\
&= \left(\mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right) \\
&= \hat{\mathbf{B}}'\mathbf{X}'\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right) \\
&= \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} \\
&= \mathbf{Y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \\
&= 0 \qquad (14)
\end{aligned}
$$

Multivariate Linear Regression

- The preceding result makes it easy to show that the variance-covariance matrix of $\mathbf{Y}$ is the sum of the variance-covariance matrices for $\hat{\mathbf{Y}}$ and $\mathbf{E}$. Specifically,

$$
\begin{aligned}
(n-1)\mathbf{S}_{\mathbf{YY}} &= \mathbf{Y}'\mathbf{Y} \\
&= \left(\hat{\mathbf{Y}} + \mathbf{E}\right)' \left(\hat{\mathbf{Y}} + \mathbf{E}\right) \\
&= \left(\hat{\mathbf{Y}}' + \mathbf{E}'\right) \left(\hat{\mathbf{Y}} + \mathbf{E}\right) \\
&= \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \mathbf{E}'\hat{\mathbf{Y}} + \hat{\mathbf{Y}}'\mathbf{E} + \mathbf{E}'\mathbf{E} \\
&= \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \mathbf{0} + \mathbf{0} + \mathbf{E}'\mathbf{E} \\
&= \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \mathbf{E}'\mathbf{E}
\end{aligned}
$$

Multivariate Linear Regression

- Consequently

$$\mathbf{S_{YY}} = \mathbf{S_{\hat{Y}\hat{Y}}} + \mathbf{S_{EE}} \tag{15}$$

- Notice also that

$$\mathbf{S_{EE}} = \mathbf{S_{YY}} - \mathbf{B'S_{XX}B} \tag{16}$$

Extensions to Random Variables and Random Vectors

- In the previous section, we developed results for sample bivariate regression, multiple regression and multivariate regression.
- We saw that, in the sample, a least squares linear regression system is characterized by several key propertiesSimilar relationships hold when systems of random variables are related in a linear least-squares regression system.
- In this section, we extend these results to least-squares linear regression systems relating random variables or random vectors.
- We will develop the results for the multivariate regression case, as these results include the bivariate and multiple regression systems as special cases.

Extensions to Random Variables and Random Vectors

- Suppose there are p criterion variables in the *random vector* $\mathbf{y}$, and q predictor variables in the random vector $\mathbf{x}$. For simplicity, assume all variables have means of zero, so no intercept is necessary. The prediction equation is

$$\begin{aligned} \mathbf{y} &= \mathbf{B}'\mathbf{x} + \mathbf{e} && (17) \\ &= \hat{\mathbf{y}} + \mathbf{e} && (18) \end{aligned}$$

- In the population, the least-squares solution also minimizes the average squared error, but in the long run sense of minimizing the expected value of the sum of squared errors, i.e., $\operatorname{Tr} \mathrm{E}(\mathbf{e}\mathbf{e}')$.
- The solution for $\mathbf{B}$ is

$$\mathbf{B} = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}} \qquad (19)$$

with $\boldsymbol{\Sigma}_{\mathbf{xx}} = \mathrm{E}(\mathbf{x}\mathbf{x}')$ the variance-covariance matrix of the random variables in $\mathbf{x}$, and $\boldsymbol{\Sigma}_{\mathbf{xy}} = \mathrm{E}(\mathbf{x}\mathbf{y}')$ the covariance matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$.

Extensions to Random Variables and Random Vectors

- The covariance matrix between predicted and error variables is null, just as in the sample case. The proof is structurally similar to its sample counterpart, but we include it here to demonstrate several frequently used techniques in the matrix algebra of expected values.

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\hat{\mathbf{y}}\mathbf{e}} &= \mathrm{E}\left(\hat{\mathbf{y}}\mathbf{e}'\right) \\
&= \mathrm{E}\left(\mathbf{B}'\mathbf{x}(\mathbf{y}-\mathbf{B}'\mathbf{x})'\right) \\
&= \mathrm{E}\left(\boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}\mathbf{y}' - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{x}\mathbf{x}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}}\right) \\
&= \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\,\mathrm{E}(\mathbf{x}\mathbf{y}') - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\,\mathrm{E}(\mathbf{x}\mathbf{x}')\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}} \\
&= \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}} - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}} \\
&= \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}} - \boldsymbol{\Sigma}_{\mathbf{yx}}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}} \\
&= \mathbf{0} \qquad (20)
\end{aligned}
$$

Extensions to Random Variables and Random Vectors

- We also find that

$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \boldsymbol{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} + \boldsymbol{\Sigma}_{\mathbf{ee}} \tag{21}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{ee}} = \boldsymbol{\Sigma}_{\mathbf{yy}} - \mathbf{B}'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{B} \tag{22}$$

- Consider an individual random variable y_i in $\mathbf{y}$. The correlation between y_i and its respective $\hat{y}_i$ is called "the multiple correlation of y_i with the predictor variables in $\mathbf{x}$."
- Suppose that the variables in $\mathbf{x}$ were uncorrelated, and that they and the variables in $\mathbf{y}$ have unit variances, so that $\boldsymbol{\Sigma}_{\mathbf{xx}} = \mathbf{I}$, an identity matrix, and, as a consequence, $\mathbf{B} = \boldsymbol{\Sigma}_{\mathbf{xy}}$.

Extensions to Random Variables and Random Vectors

- Then the correlation between a particular y_i and its respective $\hat{y}_i$ is

$$\begin{aligned} r_{y_i,\hat{y}_i} &= \frac{\sigma_{y_i\hat{y}_i}}{\sqrt{\sigma^2_{y_i}\sigma^2_{\hat{y}_i}}} \\ &= \frac{\mathrm{E}\left(y_i(\mathbf{b}_i'\mathbf{x})'\right)}{\sqrt{(1)(\mathbf{b}_i'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{b}_i)}} \\ &= \frac{\mathrm{E}\left(y_i\mathbf{x}'\mathbf{b}_i\right)}{\sqrt{(\mathbf{b}_i'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{b}_i)}} \\ &= \frac{\mathrm{E}\left(y_i\mathbf{x}'\right)\mathbf{b}_i}{\sqrt{(\mathbf{b}_i'\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{b}_i)}} \\ &= \frac{\sigma_{y_i\mathbf{x}}\mathbf{b}_i}{\sqrt{(\mathbf{b}_i'\mathbf{b}_i)}} \\ &= \frac{\mathbf{b}_i'\mathbf{b}_i}{\sqrt{(\mathbf{b}_i'\mathbf{b}_i)}} \end{aligned} \tag{23}$$

Extensions to Random Variables and Random Vectors

- It follows immediately that, when the predictor variables in $\mathbf{x}$ are orthogonal with unit variance, squared multiple correlations may be obtained directly as a sum of squared, standardized regression weights.
- In subsequent chapters, we will be concerned with two linear regression prediction systems known (loosely) as "factor analysis models," but referred to more precisely as "common factor analysis" and "principal component analysis."
- In each system, we will be attempting to reproduce an observed (or "manifest") set of p random variables in as (least squares) linear functions of a smaller set of m hypothetical (or "latent") random variables.

Partial Correlation

- In many situations, the correlation between two variables may be substantially different from zero without implying any causal connection between them.
- A classic example is the high positive correlation between number of fire engines sent to a fire and the damage done by the fire.
- Clearly, sending fire engines to a fire does not usually cause damage, and it is equally clear that one would be ill-advised to recommend reducing the number of trucks sent to a fire as a means of reducing damage.

Partial Correlation

- In situations like the house fire example, one looks for (indeed often hypothesizes on theoretical grounds) a "third variable" which is causally connected with the first two variables, and "explains" the correlation between them.
- In the house fire example, such a third variable might be "size of fire."
- One would expect that, if size of fire were held constant, there would be, if anything, a negative correlation between damage done by a fire and the number of fire engines sent to the fire.

Partial Correlation

- One way of statistically holding the third variable "constant" is through partial correlation analysis.
- In this analysis, we "partial out" the third variable from the first two by linear regression, leaving two linear regression error, or *residual* variables. We then compute the "partial correlation" between the first two variables as the correlation between the two regression residuals.
- A basic notion connected with partial correlation analysis is that, if, by partialling out one or more variables, you cause the partial correlations among some (other) variables to go to zero, then you have "explained" the correlations among the (latter) variables as being "due to" the variables which were partialled out.

Partial Correlation

- If, in terms of Equation 18 above, we "explain" the correlations in the variables in $\mathbf{y}$ by the variables in $\mathbf{x}$, then $\mathbf{e}$ should have a correlation (and covariance) matrix which is diagonal, i.e., the variables in $\mathbf{e}$ should be uncorrelated once we "partial out" the variables in $\mathbf{x}$ by linear regression.
- Recalling Equation 22 we see that this implies that $\mathbf{\Sigma_{yy} - B'\Sigma_{xx}B}$ is a diagonal matrix.

Partial Correlation

- This seemingly simple result has some rather surprisingly powerful ramifications, once one drops certain restrictive mental sets.
- In subsequent lectures, we shall see how, at the turn of the 20th century, this result led Charles Spearman to a revolutionary linear regression model for human intelligence, and an important new statistical technique for testing the model with data. What was surprising about the model was that it could be tested, even though the predictor variables (in $\mathbf{x}$) are never directly observed!