MATRIX NOTATION EXAMPLES & REGRESSION DECOMPOSITION

Geog 210B

Winter 2018

Review

DIGRESSION ON MATRIX ALGEBRA

Matrix Algebra Outline

- Identify vectors and special matrices
- Perform matrix operations: addition, subtraction, multiplication, inversion
- Matrix notation for regression models

Matrix

A matrix is any doubly subscripted array of elements arranged in rows and columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & ,\ldots, & a_{1n} \\ a_{21} & ,\ldots, & a_{2n} \\ \ldots & \ldots\ \ldots & \ldots \\ a_{m1} & ,\ldots, & a_{mn} \end{bmatrix} = \{A_{ij}\}$$

Row Vector

[1 x n] matrix

$$A\left[a_1, a_2, \ldots, a_n\right] = \left\{a_j\right\}$$

Column Vector

[m x 1] matrix

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix} = \{a_i\}$$

Square Matrix

Same number of rows and columns

$$\mathrm{B} = \begin{bmatrix} 5 & 4 & 7 \\ 3 & 6 & 1 \\ 2 & 1 & 3 \end{bmatrix}$$

Identity Matrix

Square matrix with ones on the diagonal and zeros elsewhere.

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Transpose Matrix

Rows become columns and columns become rows

$$A' = \begin{bmatrix} a_{11} & a_{21} & ,\ldots, & a_{m1} \\ a_{12} & a_{22} & ,\ldots, & a_{m2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{1n} & a_{2n} & ,\ldots, & a_{mn} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & ,\ldots, & a_{1n} \\ a_{21} & ,\ldots, & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & ,\ldots, & a_{mn} \end{bmatrix} = \{A_{ij}\}$$

Matrix Addition and Subtraction

A new matrix **C** may be defined as the additive combination of matrices **A** and **B** where: **C = A + B** is defined by:

$$\{C_{ij}\} = \{A_{ij}\} + \{B_{ij}\}$$

Note: all three matrices are of the same dimensions

Addition

If $$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and $$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

then $$C = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

Matrix Addition Example

$$A + B = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 8 & 10 \end{bmatrix} = C$$

Matrix Subtraction

C = A - B

Is defined by

$$\{C_{ij}\} = \{A_{ij}\} - \{B_{ij}\}$$

Matrix Multiplication

Matrices A and B have these dimensions:

[m x n] and [p x q]

Matrix Multiplication

Matrices A and B can be multiplied if:

[m x n] and [p x q]

n = p

Matrix Multiplication

The resulting matrix will have the dimensions:

[m x n] and [p x q]

m x q

Computation: A x B = C

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad [2 \text{ x } 2]$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \quad [2 \text{ x } 3]$$

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \end{bmatrix}$$

$$[2 \text{ x } 3]$$

Computation: A x B = C

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$$

[3 x 2] [2 x 3]

A and B can be multiplied

$$C = \begin{bmatrix} 2*1+3*1=5 & 2*1+3*0=2 & 2*1+3*2=8 \\ 1*1+1*1=2 & 1*1+1*0=1 & 1*1+1*2=3 \\ 1*1+0*1=1 & 1*1+0*0=1 & 1*1+0*2=1 \end{bmatrix} = \begin{bmatrix} 5\,2\,8 \\ 2\,1\,3 \\ 1\,1\,1 \end{bmatrix}$$

[3 x 3]

Computation: A x B = C

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$$

[3 x 2] [2 x 3]

Result is 3 x 3

$$C = \begin{bmatrix} 2*1+3*1=5 & 2*1+3*0=2 & 2*1+3*2=8 \\ 1*1+1*1=2 & 1*1+1*0=1 & 1*1+1*2=3 \\ 1*1+0*1=1 & 1*1+0*0=1 & 1*1+0*2=1 \end{bmatrix} = \begin{bmatrix} 5\,2\,8 \\ 2\,1\,3 \\ 1\,1\,1 \end{bmatrix}$$

[3 x 3]

Matrix Inversion

$$B^{-1}B = BB^{-1} = I$$

Like a reciprocal in scalar math

Like the number one in scalar math

See matrix algebra: http://www.catonmat.net/blog/mit-linear-algebra-part-three/

http://www.youtube.com/watch?v=S4n-tQZnU6o&feature=channel

REGRESSION IN MATRIX NOTATION

Matrix Notation:

$$y = X\beta + \epsilon$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The symbol α or β_0 is used here to indicate an intercept. All other βs are slopes.

Using the matrix operations we can move things around this equation and compute different entities of interest

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad \text{where} \quad \epsilon_i \sim^{iid} N(0, \sigma^2)$$

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\
\vdots &\;\;\vdots\;\;\vdots \\
Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n
\end{aligned}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- $\mathbf{X}$ is called the design matrix.
- β is the vector of parameters
- ϵ is the error vector
- $\mathbf{Y}$ is the response vector

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\beta + \epsilon \\ \mathbf{Y}_{n\times 1} &= \mathbf{X}_{n\times 2}\beta_{2\times 1} + \epsilon_{n\times 1} \end{aligned}$$

BACK TO OUR EXAMPLE

In Scalar Notation

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

$$y_1 = \alpha + \beta x_1 + \varepsilon_1$$
$$y_2 = \alpha + \beta x_2 + \varepsilon_2$$
$$y_3 = \alpha + \beta x_3 + \varepsilon_3$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$y_n = \alpha + \beta x_n + \varepsilon_n$$

In Matrix Notation

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} \alpha + \beta x_1 \\ \alpha + \beta x_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha + \beta x_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

n by 1 n by 2 2 by 1 n by 1

ID	X	Y	XY	X	Y hat	e=Y-Yhat
1	2	4	8	4	4.5	-0.5
2	3	7	21	9	6.25	0.75
3	1	3	3	1	2.75	0.25
4	5	9	45	25	9.75	-0.75
5	9	17	153	81	16.75	0.25
Sums	20	40	230	120	40	0
Average	4	8			8	0

$$X=\begin{bmatrix}1 & 2\\ 1 & 3\\ 1 & 1\\ 1 & 5\\ 1 & 9\end{bmatrix} \qquad Y=\begin{bmatrix}4\\ 7\\ 3\\ 9\\ 17\end{bmatrix}$$

SAME NUMERICAL EXAMPLE OF INTRO SLIDE

$$X'X=\begin{bmatrix}5 & 20\\ 20 & 120\end{bmatrix},\ (X'X)^{-1}=\begin{bmatrix}0.6 & -0.1\\ -0.1 & 0.025\end{bmatrix}$$

$$X'Y=\begin{bmatrix}40\\ 230\end{bmatrix},\ \hat{\boldsymbol{\beta}}=\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}=(X'X)^{-1}X'Y=\begin{bmatrix}1\\ 1.75\end{bmatrix}$$

$$\hat{Y}=1+1.75X,\ \hat{Y}=X\hat{\boldsymbol{\beta}}=[4.5\ \ 6.25\ \ 2.75\ \ 9.75\ \ 16.75]'$$

$$e=E=Y-\hat{Y}=[-0.5\ \ 0.75\ \ 0.25\ \ -0.75\ \ 0.25]'$$

$$\text{Sum of Squared Residuals } e'e=1.5$$

RECALL FROM EARLIER NOTES WE HAVE 3 UNKNOWN PARAMETERS IN THIS MODEL:

$$\alpha, \beta, \sigma_{\varepsilon}$$

The ε values underlying the sample data are never observed because we do not actually know α and β.

We could use the es (residuals) and it is possible to compute the variance of the residuals. Two alternative estimators are:

$$s^2 = \frac{\sum_{i=1}^{n} e^2}{(n-2)}$$

This estimator can be seen as recognizing the decrease in degrees of freedom in determining the value of the variance because when we use least squares as the estimation method we impose 2 constraints:

$$\sum_{i=1}^{n} e = \sum_{i=1}^{n} Xe = 0$$

$$s^2 = \frac{\sum_{i=1}^{n} e^2}{n}$$

Properties of Linear Regression

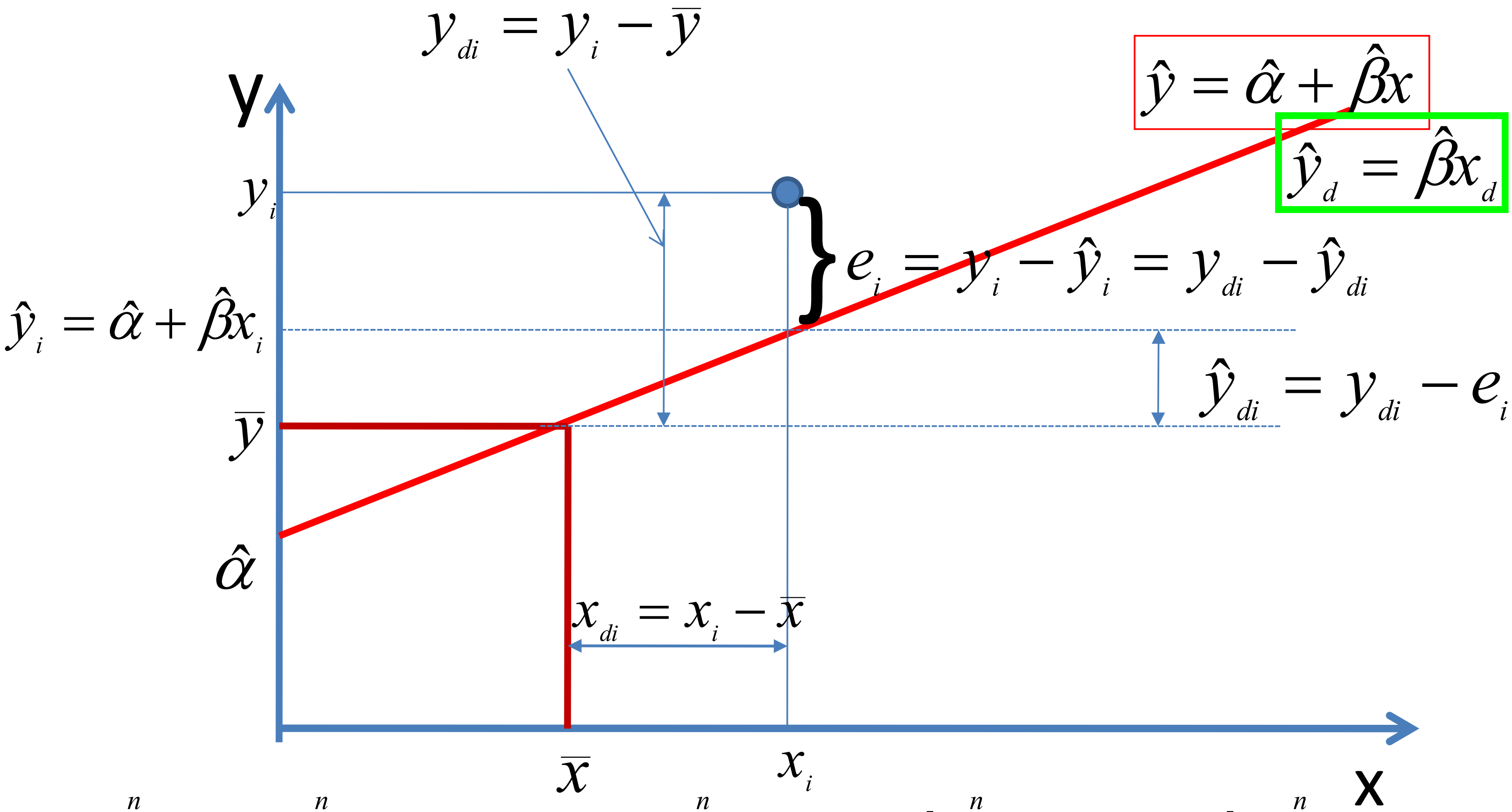
- The regression line passes through the point of x and y means
- The residuals have zero covariance with the sample x values

$$Cov(x\ e\ ) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(e_i - \overline{e}) = 0$$

- The residuals have zero covariance with the y hat values

$$Cov(\hat{y}\ e\ ) = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \overline{y})(e_i - \overline{e}) = 0$$

- The regression coefficients can be computed using deviations from the mean (next slide)
- The total variation of Y can be decomposed into explained and unexplained variation that can give us easy to understand "goodness of fit" statistics



$$\sum_{i=1}^{n} e^2 = \sum_{i=1}^{n} (y_{di} - \hat{y}_{di})^2 = \sum_{i=1}^{n} y_{di}^2 - 2\hat{\beta}\sum_{i=1}^{n} x_{di} y_{di} + \hat{\beta}^2 \sum_{i=1}^{n} x_{di}^2$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}} = 0 \quad => \hat{\beta} = \frac{\sum_{i=1}^{n} x_{di} y_{di}}{\sum_{i=1}^{n} x_{di}^2}$$

Another way to think about the meaning/nature of a slope

Decomposition of sum of squares

$$\hat{y}_{di} = y_{di} - e_i \quad => \quad y_{di} = \hat{y}_{di} + e_i = \hat{\beta} x_{di} + e_i$$

Take the square and sum over all observations n

$$(y_{di})^2 = (\hat{y}_{di} + e_i)^2 = (\hat{\beta} x_{di} + e_i)^2$$

$$\sum_{i=1}^{n} (y_{di})^2 = \sum_{i=1}^{n} (\hat{y}_{di} + e_i)^2 = \sum_{i=1}^{n} (\hat{\beta} x_{di} + e_i)^2$$

$$\sum_{i=1}^{n} (y_{di})^2 = \sum_{i=1}^{n} (\hat{y}_{di})^2 + \sum_{i=1}^{n} (e_i)^2 + 2\sum_{i=1}^{n} (\hat{y}_{di} e_i)$$

$$\sum_{i=1}^{n} (\hat{y}_{di} e_i) = ?$$

This looks like a covariance

Decomposition of sum of squares

$$Cov(\hat{y}\ e\ ) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(e_i - \bar{e}) = 0$$

$$= y_{di} \qquad = e_i \quad because\ \bar{e} = 0$$

$$Cov(\hat{y}\ e\ ) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{di})(e_i) = 0$$

$$\sum_{i=1}^{n}(y_{di})^2 = \sum_{i=1}^{n}(\hat{y}_{di})^2 + \sum_{i=1}^{n}(e_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_{di}e_i)$$

$$\sum_{i=1}^{n}(y_{di})^2 = \sum_{i=1}^{n}(\hat{y}_{di})^2 + \sum_{i=1}^{n}(e_i)^2$$

Decomposition of sum of squares

$$\sum_{i=1}^{n}(y_{di})^2 = \sum_{i=1}^{n}(\hat{y}_{di})^2 + \sum_{i=1}^{n}(e_i)^2$$

$$\sum_{i=1}^{n}(y_{di})^2 = TSS$$

Total sum of squares in the dependent variable measured about its mean

$$\sum_{i=1}^{n}(\hat{y}_{di})^2 = ESS$$

EXPLAINED sum of squares (also called Regression sum of squares

$$\sum_{i=1}^{n}(e_i)^2 = RSS$$

RESIDUAL sum of squares also called unexplained sum of squares

We like to work with proportions: ESS/TSS

$$\mathrm{var}(x) = s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_{di})^2}{n}$$

Similar for y

The Pearson correlation coefficient:

$$r = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)}\sqrt{\mathrm{var}(y)}} = \frac{1/n\sum_{i=1}^{n} x_{di}y_{di}}{\sqrt{1/n\sum_{i=1}^{n} x_{di}^2}\sqrt{1/n\sum_{i=1}^{n} y_{di}^2}}$$

$$r = \frac{\sum_{i=1}^{n} x_{di}y_{di}}{n\sqrt{1/n\sum_{i=1}^{n} x_{di}^2}\sqrt{1/n\sum_{i=1}^{n} y_{di}^2}} = \left(\frac{\sum_{i=1}^{n} x_{di}y_{di}}{\sum_{i=1}^{n} x_{di}^2}\right)\frac{\sqrt{\sum_{i=1}^{n} x_{di}^2}}{\sqrt{\sum_{i=1}^{n} y_{di}^2}}$$

$$\mathrm{Re}\,member \quad that \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_{di}y_{di}}{\sum_{i=1}^{n} x_{di}^2}$$

Substitute for beta hat above

Correlation, β hat, and r square

$$r = \frac{\sum_{i=1}^{n} x_{di} y_{di}}{n\sqrt{1/n\sum_{i=1}^{n} x_{di}^2}\sqrt{1/n\sum_{i=1}^{n} y_{di}^2}} = \hat{\beta}\frac{\sqrt{\sum_{i=1}^{n} x_{di}^2}}{\sqrt{\sum_{i=1}^{n} y_{di}^2}} = \hat{\beta}\frac{s_x}{s_y}$$

$$r^2 = \frac{(\sum_{i=1}^{n} x_{di} y_{di})^2}{(\sqrt{\sum_{i=1}^{n} x_{di}^2}\sqrt{\sum_{i=1}^{n} y_{di}^2})^2} = \hat{\beta}\frac{\sum_{i=1}^{n} x_{di} y_{di}}{\sum_{i=1}^{n} y_{di}^2} = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

Note: $$ESS = \sum_{i=1}^{n} (\hat{y}_{di})^2 = \sum_{i=1}^{n} (\hat{\beta} x_{di})^2 = \hat{\beta}^2 \sum_{i=1}^{n} (x_{di})^2 = \hat{\beta}\frac{\sum_{i=1}^{n} x_{di} y_{di}}{\sum_{i=1}^{n} x_{di}^2}\sum_{i=1}^{n} x_{di}^2 = \hat{\beta}\sum_{i=1}^{n} x_{di} y_{di}$$

```
> summary(HHPMT.lm)

Call:
lm(formula = TotDist ~ HHSIZ, data = SmallHHfile)

Residuals:
   Min      1Q Median      3Q     Max
-186.1  -49.1  -26.5   11.4 5717.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.2118     1.1817   10.33   <2e-16 ***
HHSIZ        21.7305     0.4053   53.62   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 114.7 on 42429 degrees of freedom
Multiple R-squared:  0.06345,   Adjusted R-squared:  0.06343
F-statistic:  2875 on 1 and 42429 DF,  p-value: < 2.2e-16

> anova(HHPMT.lm) # anova table
Analysis of Variance Table

Response: TotDist
             Df    Sum Sq  Mean Sq F value    Pr(>F)
ESS HHSIZ         1  37815166 37815166  2874.6 < 2.2e-16 ***
RSS Residuals 42429 558154359    13155
TSS = ESS+RSS
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Some Notes on R output

- Mean Squares = the Sum of Squares divided by their respective Degrees of Freedom

- Adjusted R square. As Xs are added to the model, each X will explain some of the variance in the dependent variable Y. You can add Xs to the model and continue to improve the ability of the model to explain the dependent variable, but this increase in R-square would be simply due to chance variation (the coefficients will not be significantly different than zero but you get the illusion of better fit). The adjusted R-square attempts to yield a more honest value to estimate the R-squared for the population by penalizing models with many coefficients.

- Adjusted R-squared = 1 - ((1-R-sq)(N-1 / N - k - 1)). From this formula, you can see that when the number of observations is small and the number of Xs is large, there will be a much greater difference between R-square and adjusted R-square (because the ratio of (N-1 / N - k - 1) will be much less than 1.

- When the number of observations is very large compared to the number of Xs, the value of R-square and adjusted R-square will be much closer because the ratio of (N-1)/(N-k-1) is closer to 1.

Hypotheses Testing in CLR

Testing for significance of regression coefficients

$$\mathrm{Re}\,member \quad that \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_{di} y_{di}}{\sum_{i=1}^{n} x_{di}^2} \qquad let's\ write\ w_i = \frac{x_{di}}{\sum_{i=1}^{n} x_{di}^2}$$

It can be shown that:

$$\hat{\beta} = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i (\alpha + \beta x_i + \varepsilon_i)$$

Simplifying and taking the expectation:

$$E(\hat{\beta}) = \beta$$

Least squares = unbiased estimate of beta

$$\mathrm{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)^2] = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n} x_{di}^2}$$

Hypotheses Testing in CLR

Similar considerations for the intercept

$$E(\hat{\alpha}) = \alpha,\ \mathrm{var}(\hat{\alpha}) = \sigma_{\varepsilon}^{2}\left[\frac{1}{n} + \frac{\overline{x}^{2}}{\sum_{i=1}^{n} x_{di}}\right]$$

$$\mathrm{cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\sigma_{\varepsilon}^{2}\overline{x}}{\sum_{i=1}^{n} x_{di}^{2}}$$

Think of circumstances when the covariance can be zero

Considering that ε is assumed normally distributed

$$\hat{\beta} \sim N(\beta, \sigma_{\varepsilon}^{2} / \sum_{i=1}^{n} x_{di}^{2})$$

We do not have the population value of sigma (**σ**) but we will use its estimate s

$$\frac{\hat{\beta} - \beta}{\dfrac{s}{\sqrt{\sum_{i=1}^{n} x_{di}^{2}}}} \sim t(n-2)$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} e_{i}^{2}}{(n-2)}}$$

$$H_0 : \beta = \beta_0$$
$$H_a : \beta \neq \beta_0$$

Reject the null at the 95% confidence if

=0

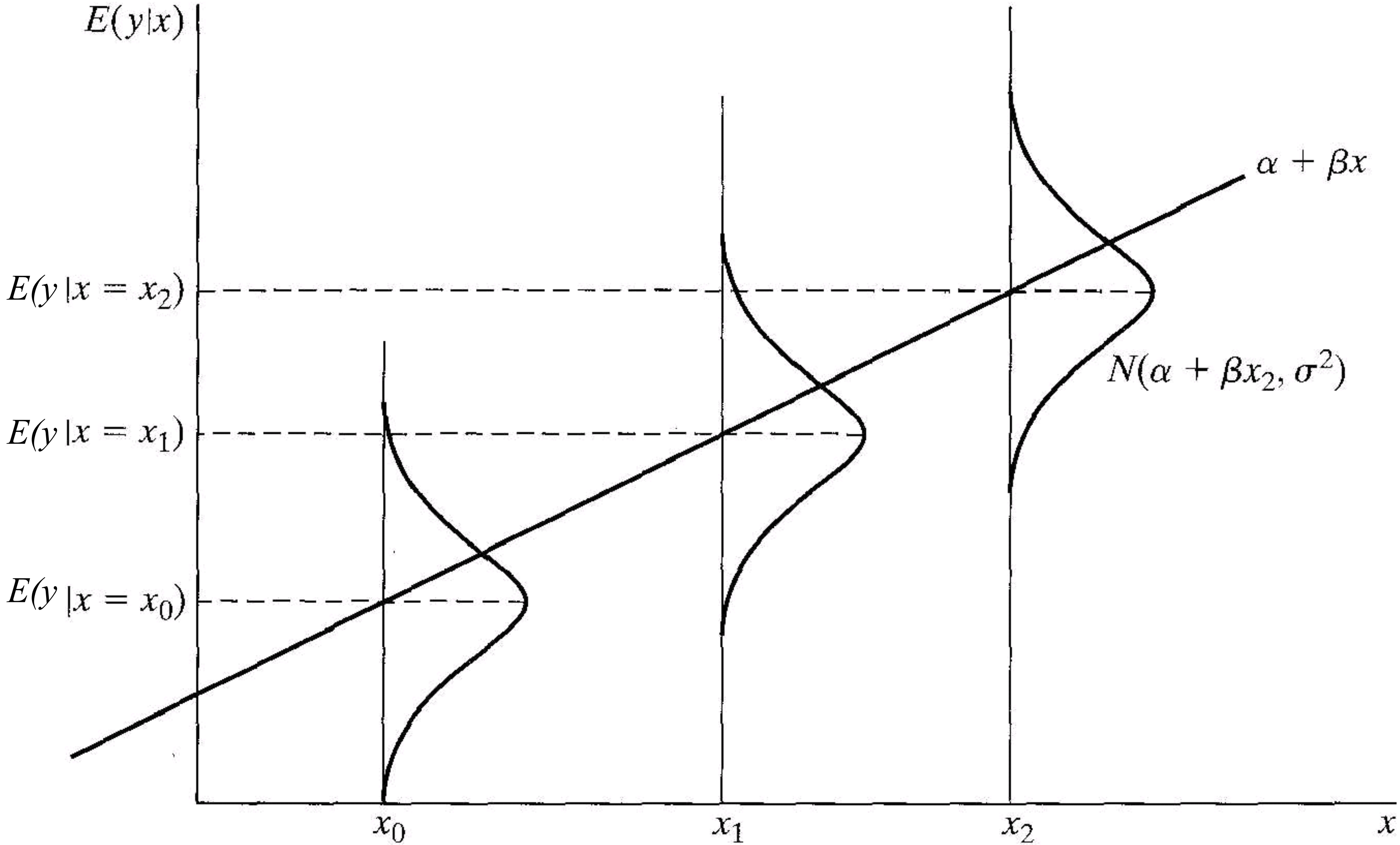
$$\left| \frac{\hat{\beta} - \beta_0}{s \Big/ \sqrt{\sum_{i=1}^{n} x_{di}^2}} \right| > t_{0.025}(n-2)$$

We can also compute the following:

F statistic

$$F = \frac{ESS/1}{RSS/(n-2)} \sim F(1, n-2)$$

$$F = \frac{ESS/1}{RSS/(n-2)} > F_{0.95}(1, n-2)$$



Keep in mind always this representation of a linear model

MORE EXAMPLES OF MATRIX NOTATION FOR THE LINEAR REGRESSION MODEL

$$\sum \epsilon_i^2 = [\epsilon_1 \, \epsilon_2 \, \cdots \, \epsilon_n] \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \epsilon' \epsilon$$

Normal Equations

$$\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})\beta$$

Solution to normal equations

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$
\begin{aligned}
\mathbf{X}'\mathbf{X} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \\
(\mathbf{X}'\mathbf{X})^{-1} &= \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} = \frac{1}{nSS_X} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} \\
\mathbf{X}'\mathbf{Y} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}
\end{aligned}
$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} = \frac{1}{nSS_X} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Compare X’X to the $(X'X)^{-1}$

Inverse, determinant, and adjoint defined in
http://www.sosmath.com/matrix/inverse/inverse.html

$$
\begin{aligned}
\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{nSS_X}\left[\begin{array}{cc} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{array}\right]\left[\begin{array}{c} \sum Y_i \\ \sum X_iY_i \end{array}\right] \\
&= \frac{1}{nSS_X}\left[\begin{array}{c} (\sum X_i^2)(\sum Y_i) - (\sum X_i)(\sum X_iY_i) \\ -(\sum X_i)(\sum Y_i) + n\sum X_iY_i \end{array}\right] \\
&= \frac{1}{SS_X}\left[\begin{array}{c} \bar{Y}(\sum X_i^2) - \bar{X}\sum X_iY_i \\ \sum X_iY_i - n\bar{X}\bar{Y} \end{array}\right] \\
&= \frac{1}{SS_X}\left[\begin{array}{c} \bar{Y}(\sum X_i^2) - \bar{Y}(n\bar{X}^2) + \bar{X}(n\bar{X}\bar{Y}) - \bar{X}\sum X_iY_i \\ SP_{XY} \end{array}\right] \\
&= \frac{1}{SS_X}\left[\begin{array}{c} \bar{Y}SS_X - SP_{XY}\bar{X} \\ SP_{XY} \end{array}\right] = \left[\begin{array}{c} \bar{Y} - \frac{SP_{XY}}{SS_X}\bar{X} \\ \frac{SP_{XY}}{SS_X} \end{array}\right] = \left[\begin{array}{c} b_0 \\ b_1 \end{array}\right]
\end{aligned}
$$

$$
\begin{aligned}
SS_X &= \sum X_i^2 - n\bar{X}^2 = \sum (X_i - \bar{X})^2 \\
SP_{XY} &= \sum X_iY_i - n\bar{X}\bar{Y} = \sum (X_i - \bar{X})(Y_i - \bar{Y})
\end{aligned}
$$

Fitted Values

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \mathbf{Xb}$$

Hat matrix

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{Xb} \\ \hat{\mathbf{Y}} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \hat{\mathbf{Y}} &= \mathbf{HY} \end{aligned}$$

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_{p-1} X_{i,p-1} + \epsilon_i \text{ for } i = 1, 2, \ldots, n$$

where

- Y_i is the value of the response variable for the ith case.
- $\epsilon_i \sim^{iid} N(0, \sigma^2)$ (exactly as before!)
- β_0 is the intercept (think multidimensionally).
- $\beta_1, \beta_2, \ldots, \beta_{p-1}$ are the regression coefficients for the explanatory variables.
- $X_{i,k}$ is the value of the kth explanatory variable for the ith case.
- Parameters as usual include all of the β's as well as σ^2. These need to be estimated from the data.

Model in Matrix Form

$$
\begin{aligned}
\mathbf{Y}_{n\times 1} &= \mathbf{X}_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1} \\
\epsilon &\sim N(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n}) \\
\mathbf{Y} &\sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})
\end{aligned}
$$

Design Matrix $\mathbf{X}$:

$$
\mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{bmatrix}
$$

Coefficient matrix β:

$$
\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}
$$

Parameter Estimation

Least Squares

Find $\mathbf{b}$ to minimize $SSE = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$
Obtain normal equations as before: $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$

Least Squares Solution

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Fitted (predicted) values for the mean of Y are

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y},$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

MANY X VARIABLES LINEAR REGRESSION MODEL

The dependent variable we try to explain

Random variable = error term

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

Intercept = constant to be estimated

Many "Slopes" = "sensitivity" of y to the values of each x_k

k = number of explanatory variables

Using an estimation method we try to find estimates of the parameter values for the intercept, slope, and variance of the error term using the sample data

$$\begin{aligned} E(y_i) &= E(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i) \\ &= E(\alpha) + E(\beta_1 x_{1i}) + E(\beta_2 x_{2i}) + \ldots + E(\beta_k x_{ki}) + E(\varepsilon_i) \\ &= \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} \end{aligned}$$

We estimate the parameters using the least (minimum) squares concept. The squared residuals:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_{2i} x_{2i} - \ldots - \hat{\beta}_k x_{ki})^2$$

Minimization is done by differentiation (Derivatives set to zero).

The solution in Matrix notation is:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

What dimensions does this have?

Linear Model Assumptions

1. **The model is linear in the variables X and the disturbance ε**
2. **The random disturbance is centered at zero:**

$$E[\varepsilon] = \begin{bmatrix} E[\varepsilon_1] \\ E[\varepsilon_2] \\ \dots \\ E[\varepsilon_n] \end{bmatrix} = 0$$

implies

$$E[Y] = X\beta$$

Linear Model Assumptions

3. Homoskedastic disturbances:

$$E[\,\varepsilon\varepsilon'\,] = \sigma^2 I$$

This is the variance covariance matrix of the disturbances. It is an n by n matrix with n the number of observations.

- **The disturbances are independently (covariance is zero) identically (variance is always the same) distributed – written as:**

$$E[\varepsilon\varepsilon'] = \begin{bmatrix} E[\varepsilon_1\varepsilon_1]\ E[\varepsilon_1\varepsilon_2]\ \ldots E[\varepsilon_1\varepsilon_n] \\ E[\varepsilon_2\varepsilon_1]\ E[\varepsilon_2\varepsilon_2]\ \ldots E[\varepsilon_2\varepsilon_n] \\ \ldots \\ E[\varepsilon_n\varepsilon_1]\ E[\varepsilon_n\varepsilon_2]\ \ldots E[\varepsilon_n\varepsilon_n] \end{bmatrix} = \begin{bmatrix} \sigma^2\ 0 \ldots 0 \\ 0\ \sigma^2 \ldots 0 \\ \ldots \\ 0\ 0 \ldots \sigma^2 \end{bmatrix}$$

This is called **homoskedasticity** of the error terms

Linear Model Assumptions

4. **A reasonable assumption, but not necessary, is that the disturbances are also Normally distributed**

$$\varepsilon \sim N[0, \sigma^2 I]$$

Linear Model Assumptions

5. There are no exact linear relationships among the variables

a) (full)Rank (X) = K = number of independent variables. That is the number of independent columns in X is K

b) Have at least K observations

In econometrics literature these five properties are the Gauss-Markov theorem assumptions

The β estimates using Least Squares are:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

$$Y = X\beta + \varepsilon$$

Combine these equations to get:

$$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon$$

When X is nonstochastic and E(X'ε) = 0 then the estimate of β is unbiased:

$$E(\widehat{\beta}) = \beta$$

The variance of the parameter estimates is:

$$Var[\hat{\beta}] = \sigma^2 (X'X)^{-1}$$

The parameter estimates are a linear function of the disturbances ε and by virtue of assumption 4 we can say that:

$$\hat{\beta} \sim N[\beta, \sigma^2 (X'X)^{-1}]$$

multivariate normally distributed.

Note 1: Ideally we could use the standard normal distribution to test hypotheses regarding the estimates.
Note 2: Usually σ is not known and we need to estimate it.

An unbiased estimate of σ can be obtained from the residuals ei:

$$s^2 = \frac{e'e}{n - K}$$

Compare this with the 2 variable model

then the var-covariance matrix can be computed as (called the square of the standard error of the regression):

$$\hat{V}ar[\hat{\beta}] = s^2 (X'X)^{-1}$$

since we are using s instead of σ we cannot use the standard normal test and we use the t-test: T-statistic = estimate of β/(standard error of β).

Goodness of fit (how well our model replicates the data we use) is checked using indicators.

The most popular is called coefficient of determination or R-squared

Total Variation in the y variable is:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Total Sum of Squares (SST) = Regression Sum of Squares (SSR*) + Error Sum of Squares (SSE)

This is the same as:

Total Sum of Squares (TSS) = Explained Sum of Squares (ESS) + Residual Sum of Squares (RSS)

Regression Sum of Squares = variation we capture with xs and bs

*Note I am using different ss etc her because in literature you will find them both ways

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Percent of variation we explain by exercising our model

Takes values between 0 and 1

Usually with large samples = lower values (harder to explain variation)

Small sample = higher values

Terminology

- Dependent variable
- Independent variables
- Explanatory variable
- Disturbance
- Random error term
- Coefficients
- Parameters to estimate
- Significance
- Goodness-of-fit measure

Some more terminology

- Autocorrelation
- Covariate (the x variables in regression)
- Deterministic relationship
- Disturbance (the error term epsilon)
- Exogeneity
- Explained variable
- Homoscedasticity
- Loglinear model
- Nonstochastic regressors
- Regressand (the dependent variable y)
- Regressor (the independent variables xs)
- Spherical disturbances (only same variance - no covariance)

Limitations due to assumptions

- **(1) <u>Linearity in x and ε</u>**
  - **Dummy variables and transformation of xs may not be enough**

- **(2) <u>Do we know all relevant variables?</u> (specification)**
  - **Do experiments with exclusion-inclusion of variables and use the theory**
- **(3) <u>Prior notions regarding β</u>**
  - **Incorporation of "prior" information can be done. Example: constrained least squares.**
- **(4) <u>Observational errors</u>**
  - **Use of a proxy for Y or X**
  - **Measurement errors in Y or X**
  - **The problem is more severe when the errors are for the Xs.**

Limitations due to assumptions

- **(5) <u>Aggregation</u>**
  - **Need to have information regarding aggregation used to obtain Xs. Spatial issues are particularly thorny.**
- **(6) <u>X-non stochastic</u>**
  - **Lagged dependent variable?**
  - **Errors in the Xs? Instrumental relationships?**
- **(7) <u>Simultaneous equations</u>**
  - **Interdependence among many ys**

- **(8) <u>Heteroskedastic disturbances</u>**
  - **Unequal variance**
- **(9) <u>Correlated disturbances</u>**
  - **Covariance different than zero**

Terminology: the multiple regression model that violates the classical assumptions of homoskedasticity and lack of correlation is called the generalized regression model

In general instead of an identify matrix for the variance covariance matrix of the disturbances we have Ω.

$$E[\varepsilon\varepsilon'] = \sigma^2 \Omega$$

Heteroskedasticity: When the scale of the dependent variable and the model's explanatory power varies from observation to observation.

Example: High variability in trip making at high levels of income (stereotype: older people are richer but they make less trips and high income people tend to make more trips)

Heteroskedastic Disturbances

$$E[\varepsilon\varepsilon'] = \begin{bmatrix} E[\varepsilon_1\varepsilon_1] \; E[\varepsilon_1\varepsilon_2] \; \ldots E[\varepsilon_1\varepsilon_n] \\ E[\varepsilon_2\varepsilon_1] \; E[\varepsilon_2\varepsilon_2] \; \ldots E[\varepsilon_2\varepsilon_n] \\ \ldots \\ E[\varepsilon_n\varepsilon_1] \; E[\varepsilon_n\varepsilon_2] \; \ldots E[\varepsilon_n\varepsilon_n] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 \; 0 \ldots 0 \\ 0 \; \sigma_2^2 \ldots 0 \\ \ldots \\ 0 \; 0 \ldots \sigma_n^2 \end{bmatrix}$$

Autocorrelation: When the data contain "memory" effects.

Example: The y values of one row depend on the y values of another row (maybe they are households that live by or they belong to the same social network)

In spatial analysis we will see how we build neighborhoods

$$E[\varepsilon\varepsilon'] = \begin{bmatrix} E[\varepsilon_1\varepsilon_1]\ E[\varepsilon_1\varepsilon_2]\ \ldots E[\varepsilon_1\varepsilon_n] \\ E[\varepsilon_2\varepsilon_1]\ E[\varepsilon_2\varepsilon_2]\ \ldots E[\varepsilon_2\varepsilon_n] \\ \ldots \\ E[\varepsilon_n\varepsilon_1]\ E[\varepsilon_n\varepsilon_2]\ \ldots E[\varepsilon_n\varepsilon_n] \end{bmatrix} = \begin{bmatrix} 1\ \rho_1\ 0 \ldots \rho_{n-1} \\ \rho_1\ 1\ \ \ldots \rho_{n-2} \\ \ldots \\ \rho_{n-1}\ \rho_{n-2}\ \ldots 1 \end{bmatrix}$$

The distribution of the estimator for beta is different now:

$$\hat{\beta} \sim N[\beta, \sigma^2 (X'X)^{-1}(X'\Omega X)(X'X)^{-1}]$$

If instead of the more complex equation for the variance covariance matrix we use the usual least squares estimate $\sigma 2$ (X'X)-1 we would get biased standard errors for the coefficient estimates and then the t-tests would be misleading.

Other problems may arise with the ordinary least squares depending on the particular form of Ω.

In general: Ordinary least squares, when $\Omega \neq I$, produces unbiased, consistent, and asymptotically normally distributed estimates.

The estimates are not efficient (i.e., minimum variance) and the usual inference procedures (t-tests etc) are not valid.

Usually we don't know the elements in Ω.

One way is to get an estimate of the Var-covariance matrix from Ordinary least squares and then produce weights for each observation.

Then, estimate a regression model on the weighted observations.