Let y R to the power of n be the vector of the dependent variable (e. g. , wages), and X R to the power of n k be the matrix of regressors where the first column is typically a column of ones (intercept) and subsequent columns represent explanatory variables (e. g. , education, experience).

The vector R to the power of k contains the unknown coefficients, and R to the power of n is the vector of error terms. The linear regression model is given by y = X +. The classical Gauss-Markov assumptions ensure the OLS estimator has desirable properties: E[X] = 0 (errors average to zero), Var(X) = to the power of 2 I subscript n (constant variance), and rank(X) = k (no perfect multicollinearity).

The OLS estimator = (X^ X)^-1 X^ y minimizes the sum of squared residuals. It is unbiased (on average correct), meaning E[] =, with variance Var() = to the power of 2 (X^ X)^-1. The error variance to the power of 2 is estimated by to the power of 2 = 1n - k subscript i=1 to the power of n subscript i to the power of 2, where = y - X are the residuals.

For hypothesis testing, the t -statistic for H subscript 0: subscript j = 0 is t subscript j = subscript j to the power of 2 (X^ X) subscript jj^-1. Worked Example: Suppose we have n = 3 observations and k = 2 regressors (intercept and education). Let y = bmatrix 10 15 20 bmatrix (wages in $1000s) and X = bmatrix 1 & 12 1 & 16 1 & 18 bmatrix (years of education).

Compute X^ X = bmatrix 3 & 46 46 & 724 bmatrix, (X^ X)^-1 = 156 bmatrix 724 & -46 -46 & 3 bmatrix, and X^ y = bmatrix 45 710 bmatrix. Then = 156 bmatrix 724 45 - 46 710 -46 45 + 3 710 bmatrix = bmatrix -20 2 bmatrix. The estimated equation is wage = -20 + 2 education.

The intercept implies zero education corresponds to -$20, 000 wages (nonsensical here, illustrating small-sample limitations). The slope suggests each additional year of education increases wages by $2, 000 on average.