Eco 213: Basic Data Analysis and Econometrics
Lecture 3: Multiple Regression Analysis
February 23, 2019

Dr. Garima Malik
Department of Economics
Shiv Nadar University

Outline

- Multiple Regression Analysis
  - Design requirements
  - Multiple regression model
  - R^2
  - Testing R^2 and b's
  - Comparing models
  - Comparing partial regression coefficients

Multiple Regression Analysis (MRA)

- Method for studying the relationship between a dependent variable and two or more independent variables.
- Purposes:
  - Prediction
  - Explanation
  - Theory building

Design Requirements

- One dependent variable (criterion)
- Two or more independent variables (predictor variables).
- Sample size: >= 50 (at least 10 times as many cases as independent variables)

Assumptions

- **Independence**: the scores of any particular subject are independent of the scores of all other subjects
- **Normality**: in the population, the scores on the dependent variable are normally distributed for each of the possible combinations of the level of the X variables; each of the variables is normally distributed
- **Homoscedasticity**: in the population, the variances of the dependent variable for each of the possible combinations of the levels of the X variables are equal.
- **Linearity**: In the population, the relation between the dependent variable and the independent variable is linear when all the other independent variables are held constant.

Simple vs. Multiple Regression

- One dependent variable Y predicted from one independent variable X
- One regression coefficient
- **r^2**: proportion of variation in dependent variable Y predictable from X

- One dependent variable Y predicted from a set of independent variables (X1, X2 ….Xk)
- One regression coefficient for each independent variable
- **R^2**: proportion of variation in dependent variable Y predictable by set of independent variables (X's)

Example: Self Concept and Academic Achievement (N=103)

Statistic	Self-Concept: General (GSC)	Self-Concept: Academic (ASC)	Academic Achievement (AA)	Grade Point Average (GPA)
Correlation				
GSC	1.00			
ASC	.45	1.00		
AA	.15	.40	1.00	
GPA	.25	.50	.62	1.00
Mean	5.20	5.60	54.30	2.50
Standard Deviation	.92	1.26	10.65	.50

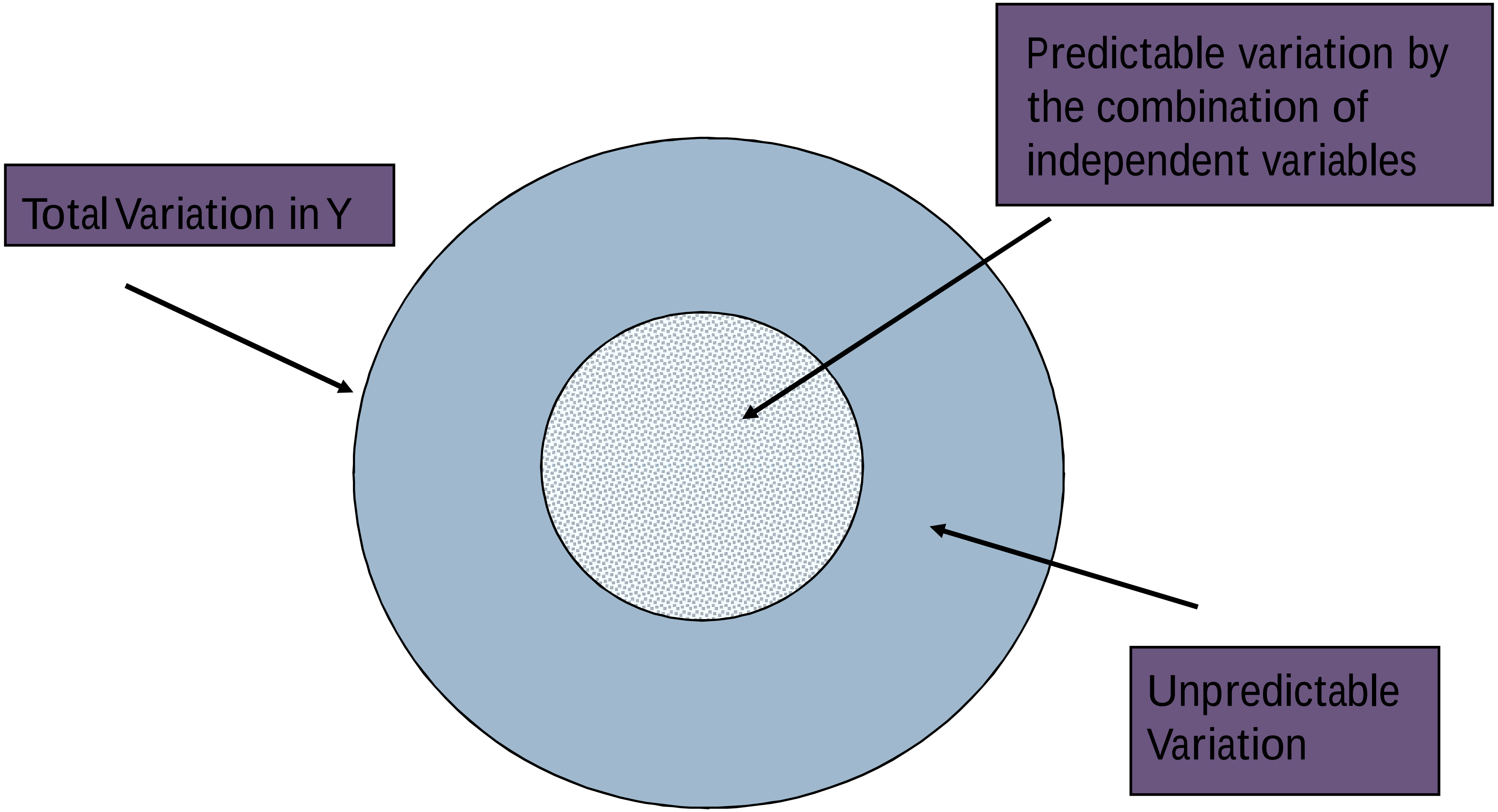
Example: The Model

- ^
- $Y = a + b_1X_1 + b_2X_2 + \ldots b_kX_k$
- The b's are called **partial regression coefficients**
- **Our example-Predicting AA:**
- **^**
  - **Y=** $36.83 + (3.52)X_{ASC} + (-.44)X_{GSC}$
- Predicted AA for person with GSC of 4 and ASC of 6
  - **^**
  - **Y=** $36.83 + (3.52)(6) + (-.44)(4) = 56.23$

Multiple Correlation Coefficient (R) and Coefficient of Multiple Determination (R^2)

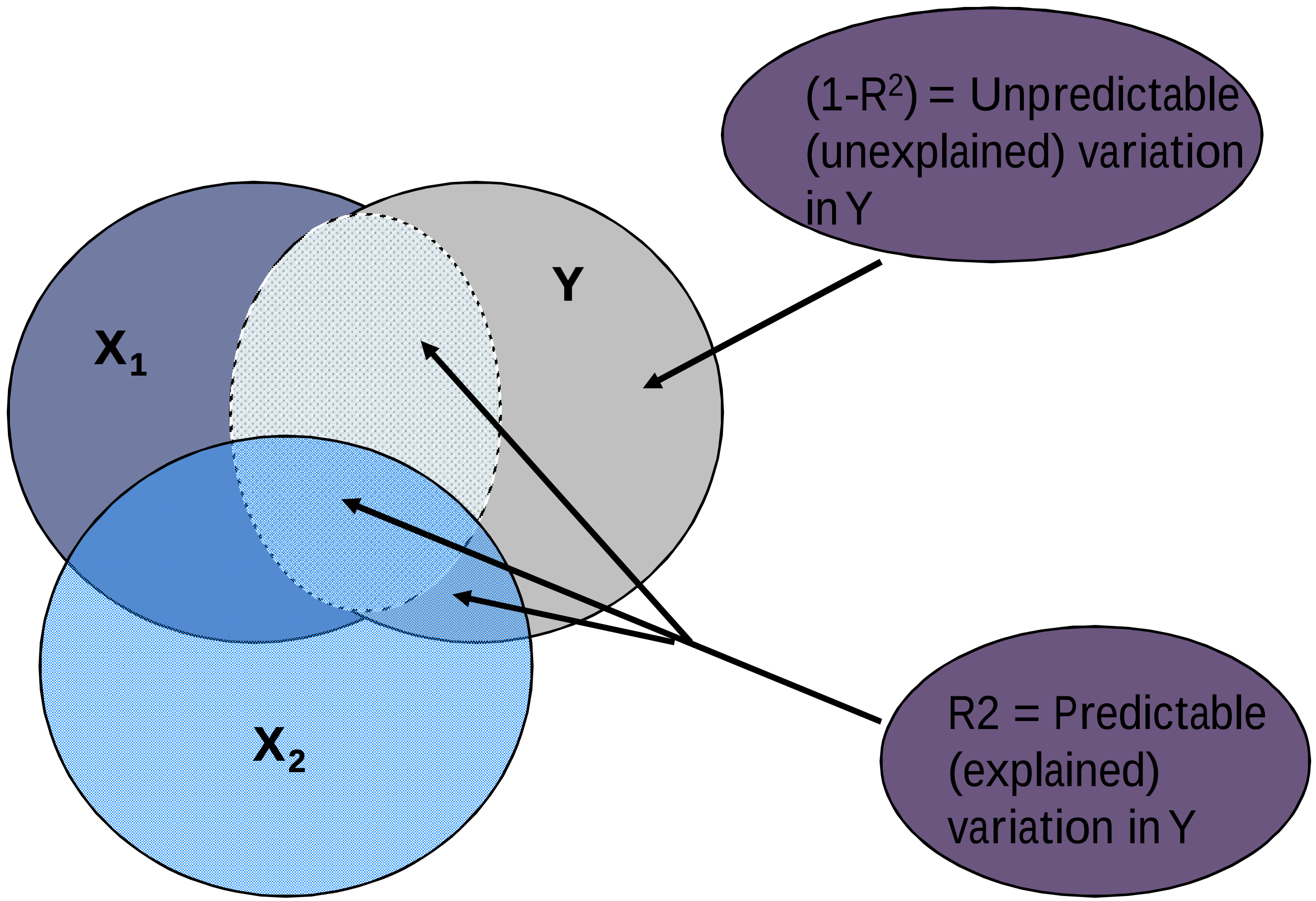
- R = the magnitude of the relationship between the dependent variable and the best linear combination of the predictor variables
- R^2 = the proportion of variation in Y accounted for by the set of independent variables (X's).

Explaining Variation: How much?



Proportion of Predictable and Unpredictable Variation

Where:
Y= AA
X_1 = ASC
X_2 =GSC



Various Significance Tests

- Testing R^2
  - Test R^2 through an F test
  - Test of competing models (difference between R^2) through an F test of difference of R^2s
- Testing b
  - Test of each partial regression coefficient (b) by t-tests
  - Comparison of partial regression coefficients with each other - t-test of difference between partial regression coefficients (β)

Example: Testing R^2

- What proportion of variation in AA can be predicted from GSC and ASC?
  - Compute R^2: $R^2 = .16$ ($R = .41$) : 16% of the variance in AA can be accounted for by the composite of GSC and ASC
- Is R^2 statistically significant from 0?
  - F test: $F_{observed} = 9.52$, $F_{crit\ (05/2,100)} = 3.09$
  - Reject H_0: in the population there is a significant relationship between AA and the linear composite of GSC and ASC

Example: Comparing Models -Testing R^2

- Comparing models
  - $\hat{}$
  - **Model 1: $\hat{Y}$=** $35.37 + (3.38)X_{ASC}$
  - **Model 2: $\hat{Y}$=** $36.83 + (3.52)X_{ASC} + (-.44)X_{GSC}$
- Compute R^2 for each model
  - Model 1: $R^2 = r^2 = .160$
  - Model 2: $R^2 = .161$
- Test difference between R^2s
  - $F_{obs} = .119$, $F_{crit(.05/1,100)} = 3.94$
  - Conclude that GSC does not add significantly to ASC in predicting AA

Testing Significance of b's

- H0: $\beta = 0$
- $t_{observed} = \dfrac{b - \beta}{\text{standard error of } b}$
- with N-k-1 df

Example: t-test of b

- $t_{observed}$ = -.44 - 0/14.24
- $t_{observed}$ = -.03
- $t_{critical(.05,2,100)}$ = 1.97

- Decision: Cannot reject the null hypothesis.
- Conclusion: The population β for GSC is not significantly different from 0

Comparing Partial Regression Coefficients

- Which is the stronger predictor? Comparing b_{GSC} and b_{ASC}
- Convert to **standardized** partial regression coefficients (beta weights, β's)
  - $\beta_{GSC} = -.038$
  - $\beta_{ASC} = .417$
  - On same scale so can compare: ASC is stronger predictor than GSC
- Beta weights (β's) can also be tested for significance with t tests.

Different Ways of Building Regression Models

- **Simultaneous**: all independent variables entered together
- **Stepwise**: independent variables entered according to some order
  - By size or correlation with dependent variable
  - In order of significance
- **Hierarchical**: independent variables entered in stages

Practice:

- Grades reflect academic achievement, but also student's efforts, improvement, participation, etc. Thus hypothesize that best predictor of grades might be academic achievement and general self concept.
- Once AA and GSC have been used to predict grades, academic self-concept (ASC) is not expected to improve the prediction of grades (I.e. not expected to account for any additional variation in grades)

Objectives of Multiple Regression

- Establish the linear equation that best predicts values of a dependent variable Y using more than one explanatory variable from a large set of potential predictors $\{x_1, x_2, \ldots x_k\}$.
- Find that subset of all possible predictor variables that explains a significant and appreciable proportion of the variance of Y, trading off adequacy of prediction against the cost of measuring more predictor variables.

Expanding Simple Linear Regression

- Quadratic model.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon$$

Adding one or more polynomial terms to the model.

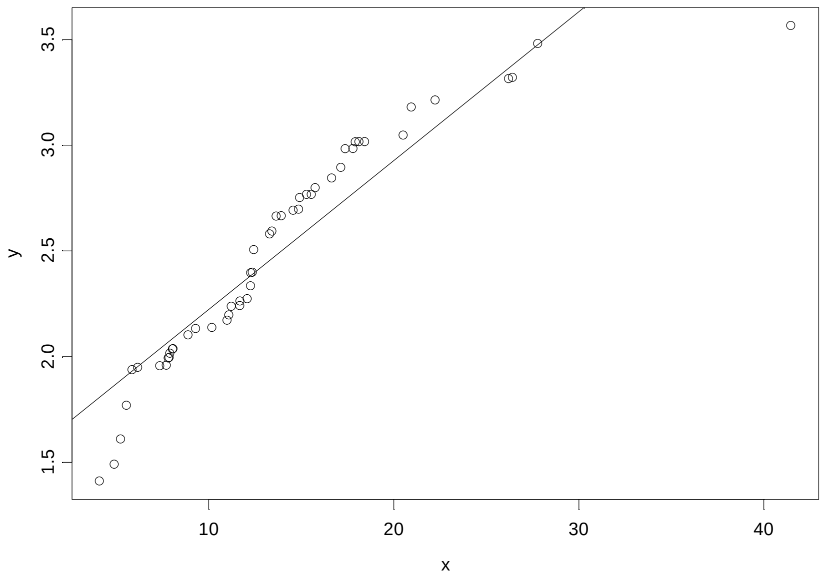
- General polynomial model.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \dots + \beta_k x_1^k + \varepsilon$$

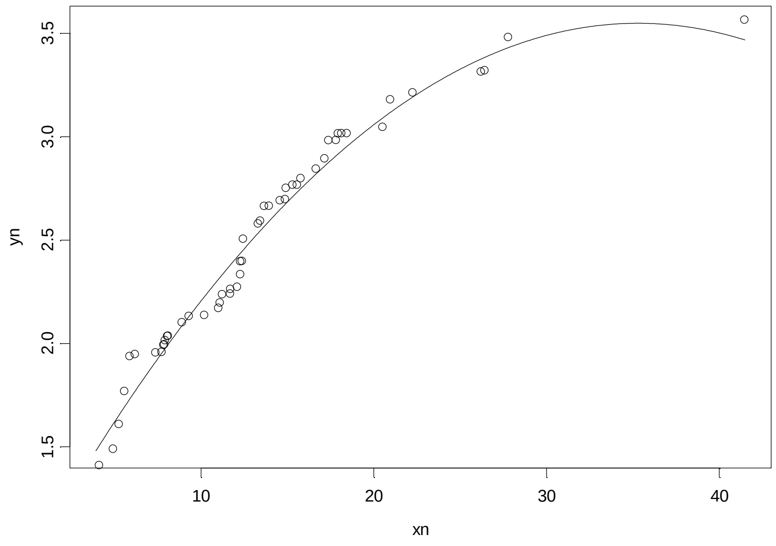
Any independent variable, x_i, which appears in the polynomial regression model as x_i^k is called a **k^{th}-degree term**.

Polynomial model shapes.

Linear



Adding one more terms to the model significantly improves the model fit.



Quadratic

Incorporating Additional Predictors

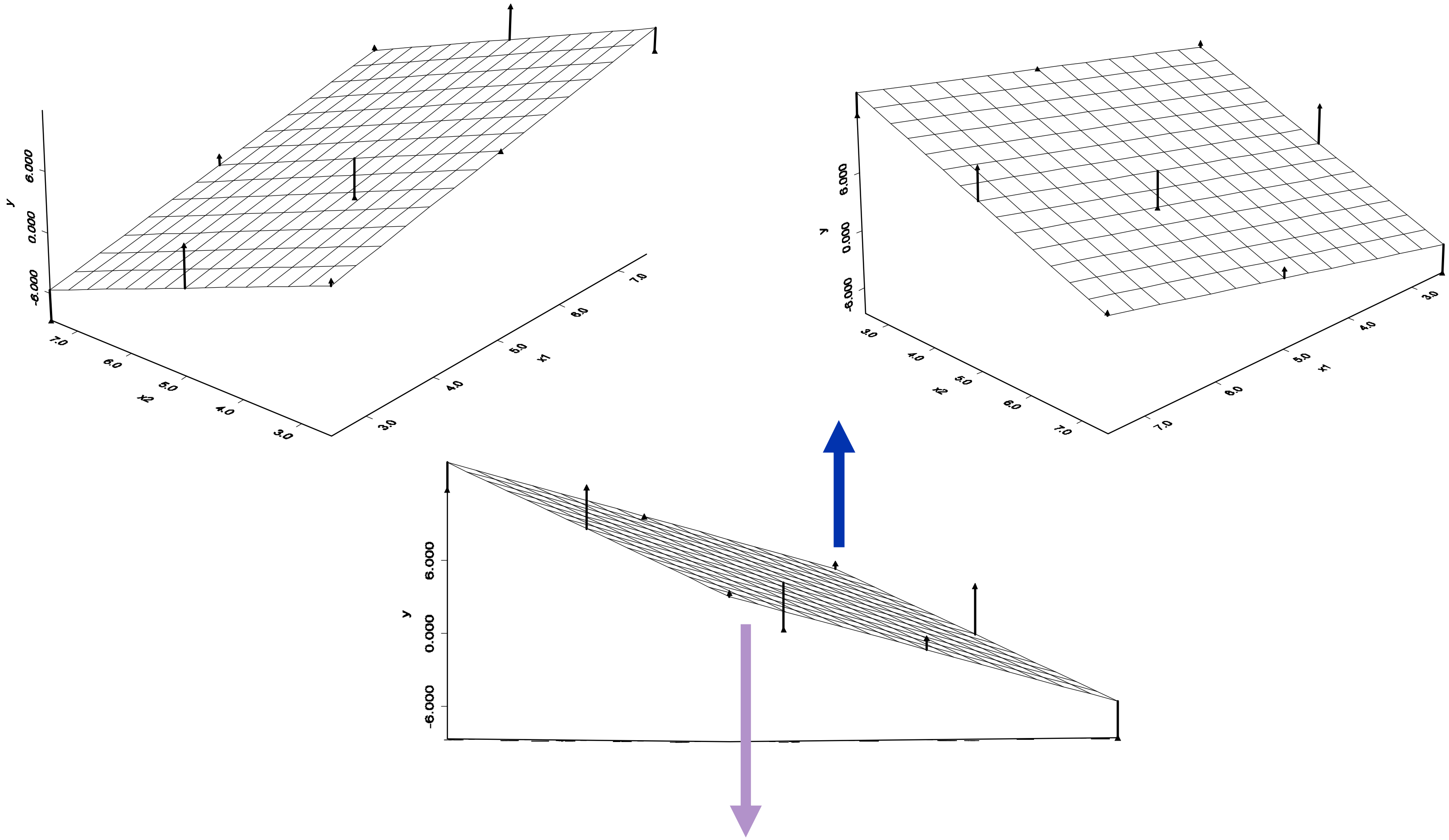
Simple additive multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + \varepsilon$$

Additive (Effect) Assumption - The expected change in **y** per unit increment in $\mathbf{x_j}$ is constant and does not depend on the value of any other predictor. This change in y is equal to $\boldsymbol{\beta_j}$.

Additive regression models:

For two independent variables, the response is modeled as a surface.



Interpreting Parameter Values (Model Coefficients)

- "Intercept" - value of y when all predictors are 0.

  β_0

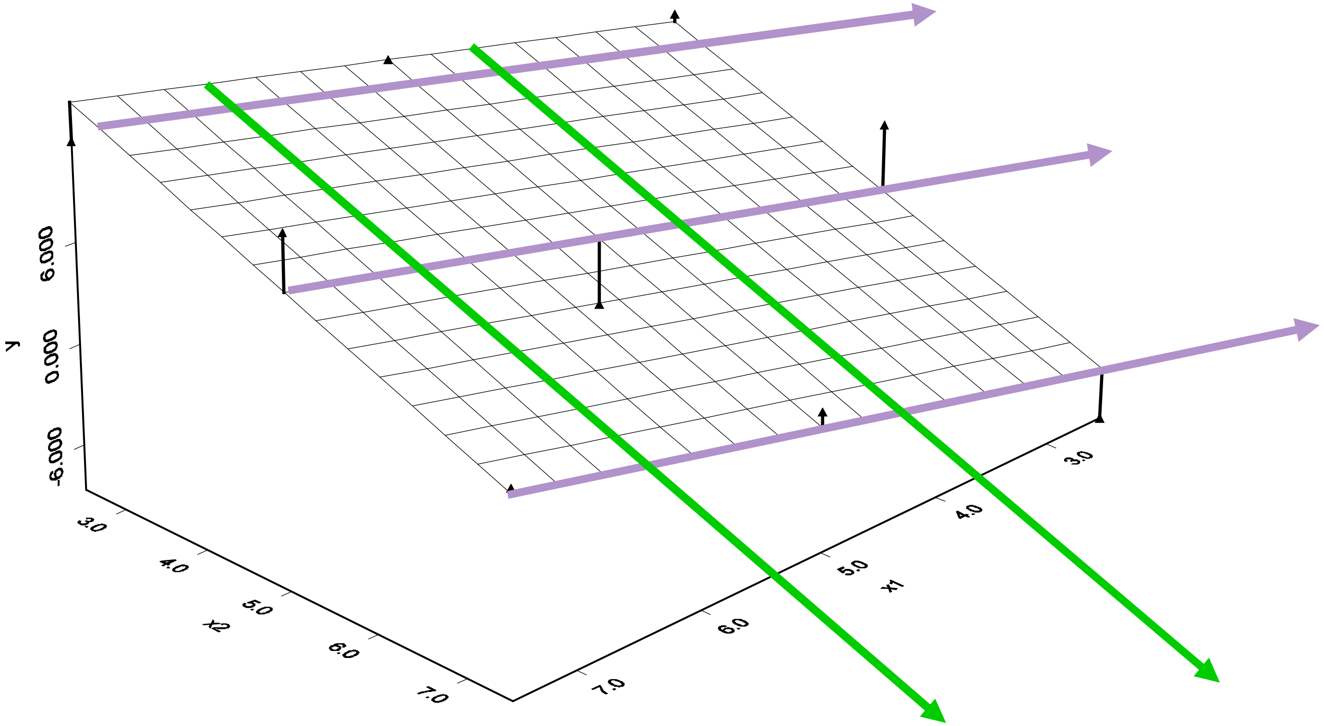
- "Partial slopes"

  $\beta_1, \beta_2, \beta_3, \ldots \beta_k$

β_j - describes the **expected change** in **y** per unit increment in $\mathbf{x_j}$ when all other predictors in the model are held at a constant value.

Graphical depiction of β_j.

β_1 - slope in direction of x_1.



β_2 - slope in direction of x_2.

Multiple Regression with Interaction Terms

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 +$$

$$\beta_3 x_3 + \ldots + \beta_k x_k +$$

$$\beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 +$$

$$\ldots + \beta_{1k} x_1 x_k + \ldots$$

$$+ \beta_{k-1,k} x_{k-1} x_k + \varepsilon$$

cross-product terms quantify the interaction among predictors.

Interactive (Effect) Assumption: The effect of one predictor, $\mathbf{x_i}$, on the response, **y**, will depend on the value of one or more of the other predictors.

Interpreting Interaction

Interaction Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \underline{\beta_{12} x_{i1} x_{i2}} + \varepsilon_i$$

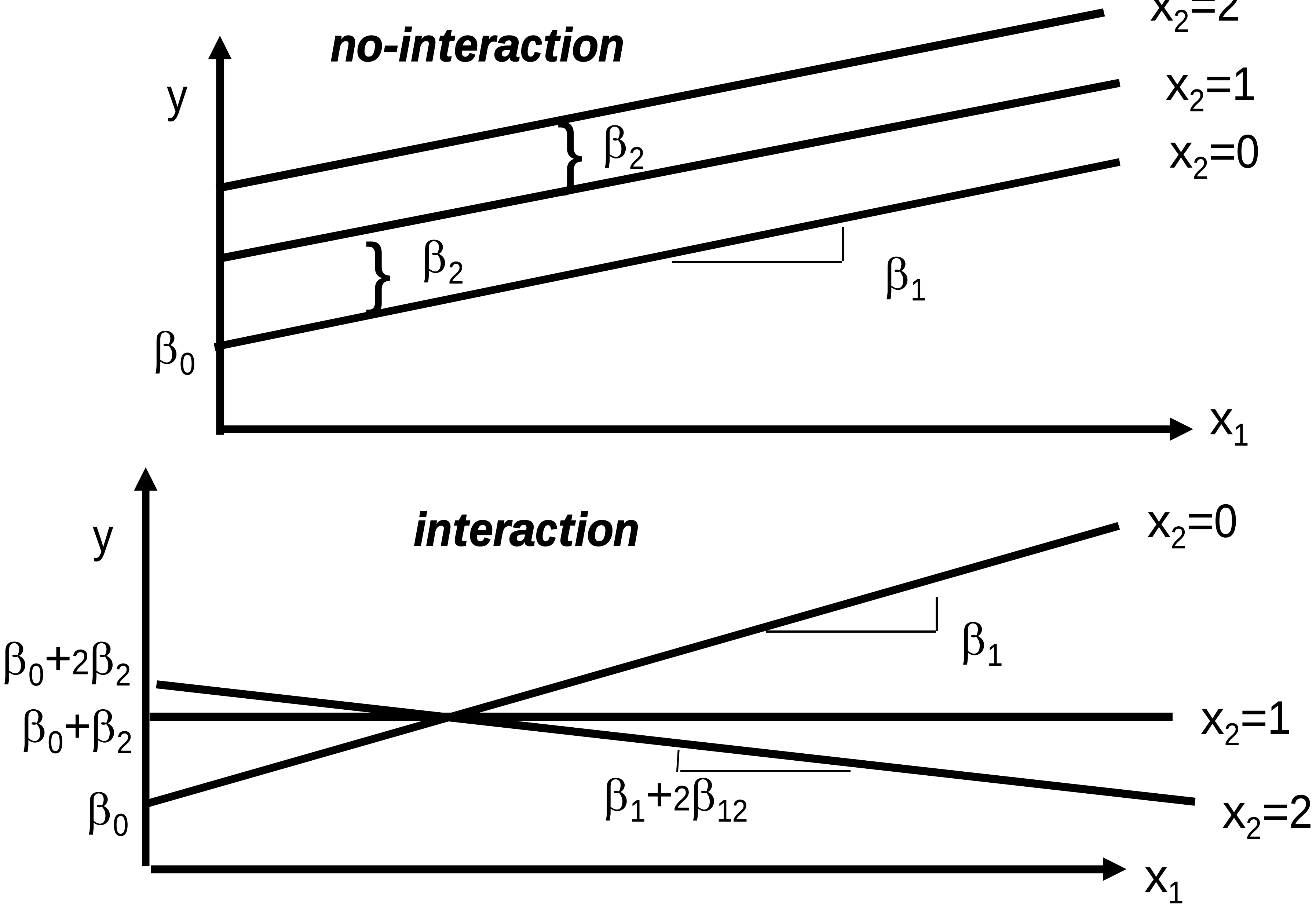
or Define: $$x_{i3} = x_{i1} x_{i2}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i3} + \varepsilon_i$$

No difference

$\boldsymbol{\beta_1}$ – No longer the expected change in Y per unit increment in X_1!

$\boldsymbol{\beta_{12}}$ – No easy interpretation! The effect on y of a unit increment in X_1, now depends on X_2.



no-interaction
y
x_2=2
x_2=1
x_2=0
β_2
β_2
β_1
β_0
x_1
interaction
y
x_2=0
β_1
$\beta_0+2\beta_2$
$\beta_0+\beta_2$
x_2=1
$\beta_1+2\beta_{12}$
β_0
x_2=2
x_1

A Protocol for Multiple Regression

Identify all possible predictors.

Establish a method for estimating model parameters and their standard errors.

Develop tests to determine if a parameter is equal to zero (i.e. no evidence of association).

Reduce number of predictors appropriately.

Develop predictions and associated standard error.

Estimating Model Parameters
Least Squares Estimation

Assuming a random sample of n observations $(y_i, x_{i1}, x_{i2}, \ldots, x_{ik})$, i=1,2,...,n. The estimates of the parameters for the best predicting equation:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}$$

Is found by choosing the values:

$$\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$$

which minimize the expression:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2$$

Normal Equations

Take the partial derivatives of the SSE function with respect to $\beta_0, \beta_1, \ldots, \beta_k$, and equate each equation to 0. Solve this system of k+1 equations in k+1 unknowns to obtain the equations for the parameter estimates.

$$
\begin{array}{lll}
1 & \sum_{i=1}^{n} y_i & = n\hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}\hat{\beta}_1 + \cdots + \sum_{i=1}^{n} x_{ik}\hat{\beta}_k \\
x_{i1} & \sum_{i=1}^{n} x_{i1} y_i & = \sum_{i=1}^{n} x_{i1}\hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}^2\hat{\beta}_1 + \cdots + \sum_{i=1}^{n} x_{i1}x_{ik}\hat{\beta}_k \\
\vdots & \vdots & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
x_{ik} & \sum_{i=1}^{n} x_{ik} y_i & = \sum_{i=1}^{n} x_{ik}\hat{\beta}_0 + \sum_{i=1}^{n} x_{ik}x_{i1}\hat{\beta}_1 + \cdots + \sum_{i=1}^{n} x_{ik}^2\hat{\beta}_k
\end{array}
$$

An Overall Measure of How Well the Full Model Performs

Coefficient of Multiple Determination

- Denoted as R^2.
- Defined as the **proportion of the variability** in the dependent variable y that is accounted for by the independent variables, $x_1, x_2, \ldots, x_k$, through the regression model.
- With only one independent variable (k=1), $R^2 = r^2$, the square of the **simple correlation coefficient**.

Computing the Coefficient of Determination

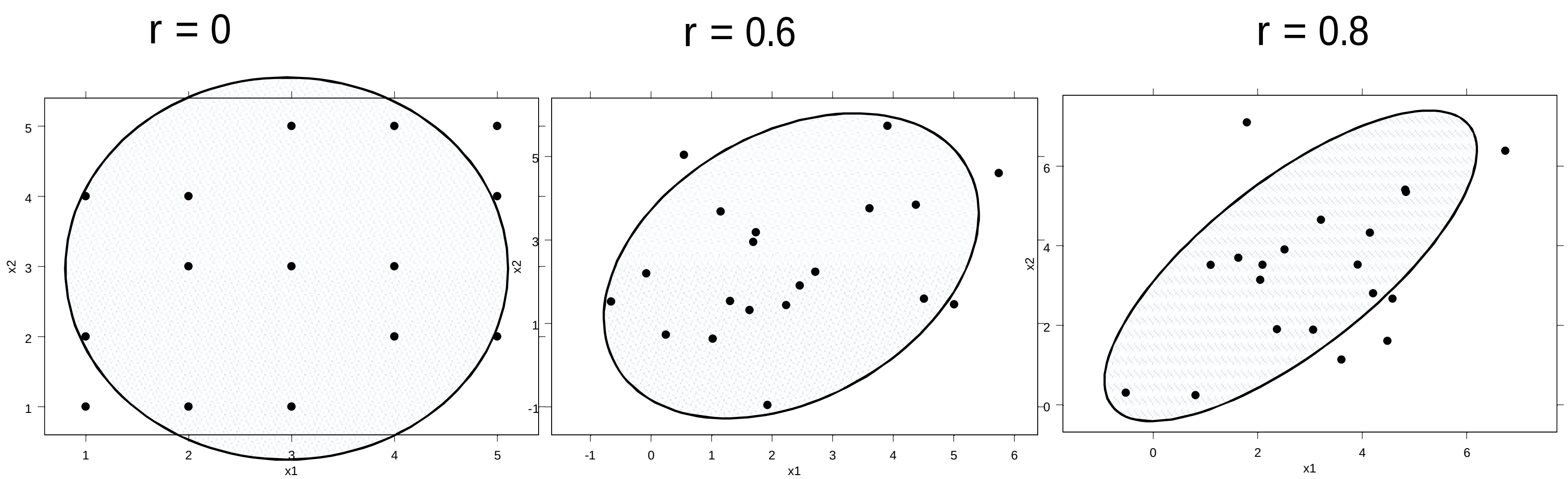
$$R^2_{y\cdot x_1 x_2 \cdots x_k} = \frac{SSR}{S_{yy}} = \frac{S_{yy} - SSE}{S_{yy}}, \qquad 0 \le R^2 \le 1$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{TSS}$$

$$\begin{aligned} SSE &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} y_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_{1i} y_i - \cdots - \hat{\beta}_k \sum_{i=1}^{n} x_{ik} y_i \end{aligned}$$

Multicollinearity

A further **assumption** in multiple regression (absent in SLR), is that the predictors (x_1, x_2, ... x_k) are statistically uncorrelated. That is, the predictors do not co-vary. When the predictors are significantly correlated (correlation greater than about 0.6) then the multiple regression model is said to suffer from problems of multicollinearity.



Multicollinearity leads to

- Numerical instability in the estimates of the regression parameters – wild fluctuations in these estimates if a few observations are added or removed.
- No longer have simple interpretations for the regression coefficients in the additive model.

Ways to detect multicollinearity

- Scatterplots of the predictor variables.
- Correlation matrix for the predictor variables – the higher these correlations the worse the problem.
- Variance Inflation Factors (VIFs) reported by software packages.Values larger than 10 usually signal a substantial amount of collinearity.

What can be done about multicollinearity

- Regression estimates are still OK, but the resulting confidence/prediction intervals are very wide.
- Choose explanatory variables wisely! (E.g. consider omitting one of two highly correlated variables.)
- More advanced solutions: *principal components analysis*; *ridge regression*.

Testing in Multiple Regression

- Testing individual parameters in the model.
- Computing predicted values and associated standard errors.

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \sigma^2)$$

Overall AOV F-test

H_0: None of the explanatory variables is a significant predictor of Y

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

Reject if: $F > F_{k,n-k-1,\alpha}$

Standard Error for Partial Slope Estimate

The estimated standard error for: $\hat{\beta}_j$

$$s_{\hat{\beta}_j} = \hat{\sigma}_{\varepsilon} \sqrt{\frac{1}{S_{x_j x_j}(1 - R^2_{x_j \bullet x_1 x_2 \cdots x_{j-1} x_{j+1} \cdots x_k})}}$$

where

$$\hat{\sigma}_{\varepsilon} = \sqrt{\frac{SSE}{n-(k+1)}}$$

$$S_{x_j x_j} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

and $R^2_{x_j \bullet x_1 x_2 \cdots x_{j-1} x_{j+1} \cdots x_k}$ **is the coefficient of determination for the model with x_j as the dependent variable and all other x variables as predictors.**

What happens if all the predictors are truly independent of each other?

$$R^2_{x_j \bullet x_1 x_2 \cdots x_{j-1} x_{j+1} \cdots x_k} \to 0 \quad s_{\hat{\beta}_j} \to \frac{\hat{\sigma}_{\varepsilon}}{\sqrt{S_{x_j x_j}}}$$

If there is high dependency?

$$R^2_{x_j \bullet x_1 x_2 \cdots x_{j-1} x_{j+1} \cdots x_k} \to 1 \quad s_{\hat{\beta}_j} \to \infty$$

Confidence Interval

100(1-α)% Confidence Interval for $\hat{\beta}_j$

$$\boxed{\hat{\beta}_j \pm t_{n-(k+1),\alpha/2} s_{\hat{\beta}_j}}$$

df for SSE

Reflects the number of data points minus the number of parameters that have to be estimated.

Testing whether a partial slope coefficient is equal to zero.

$$H_0 \quad \beta_j = 0$$

Alternatives:

Rejection Region:

$$H_a \quad \beta_j > 0 \longrightarrow t > t_{n-(k+1),\alpha}$$

$$\beta_j < 0 \longrightarrow t < -t_{n-(k+1),\alpha}$$

$$\beta_j \neq 0 \longrightarrow |t| > t_{n-(k+1),\alpha/2}$$

Test Statistic:

$$t = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}$$

Predicting Y

- We use the least squares fitted value, $\hat{y}$, as our predictor of a single value of y at a particular value of the explanatory variables ($x_1, x_2, \ldots, x_k$).
- The corresponding interval about the predicted value of y is called a prediction interval.
- The least squares fitted value also provides the best predictor of E(y), the mean value of y, at a particular value of ($x_1, x_2, \ldots, x_k$). The corresponding interval for the mean prediction is called a confidence interval.
- Formulas for these intervals are much more complicated than in the case of SLR; they cannot be calculated by hand

Minimum R^2 for a "Significant" Regression

Since we have formulas for R^2 and F, in terms of n, k, SSE and TSS, we can relate these two quantities.

We can then ask the question: what is the min R^2 which will ensure the regression model will be declared significant, as measured by the appropriate quantile from the F distribution?

The answer (below), shows that this depends on n, k, and SSE/TSS.

$$\boxed{R^2{}_{\min} = \frac{k}{n-k-1}\frac{SSE}{TSS}F_{k,n-k-1,\alpha}}$$