REGRESSION MODEL WITH TWO EXPLANATORY VARIABLES

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

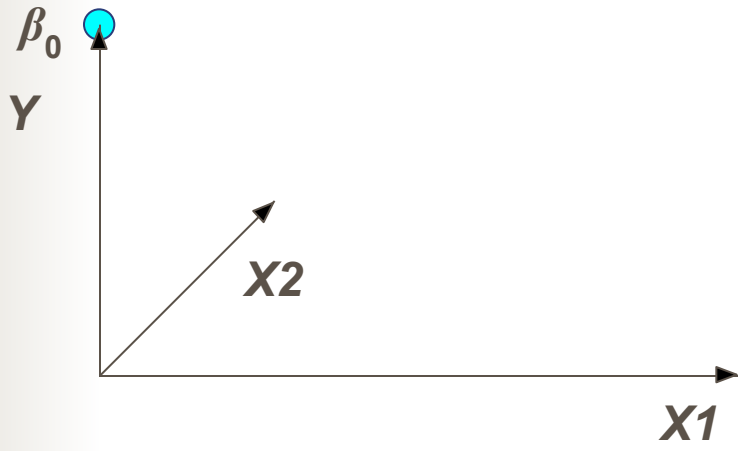
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

This sequence provides a geometrical interpretation of a multiple regression model with two explanatory variables.

Y – weekly salary ($)

X1 – length of employment (in months)

X2 – age (in years)



Specifically, we will look at weekly salary function model where weekly salary, *Y*, depend on length of employment X1, and age, X2.

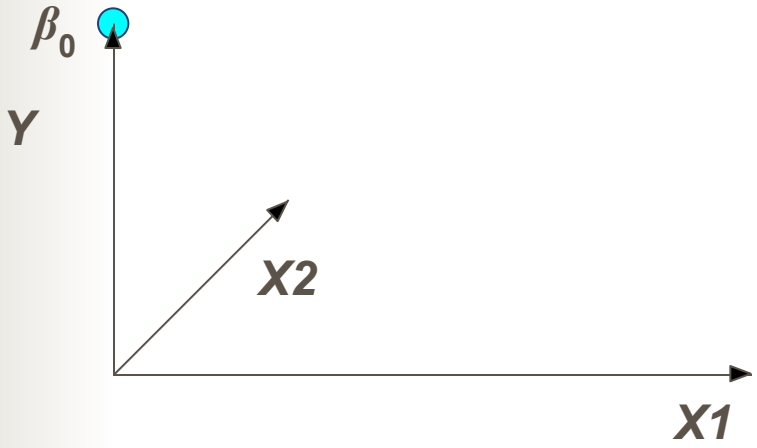
MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

Y – weekly salary ($)

X1 – length of employment (in months)

X2 – age (in years)



The model has three dimensions, one each for *Y, X1*, and X2. The starting point for investigating the determination of *Y* is the intercept, β_0.

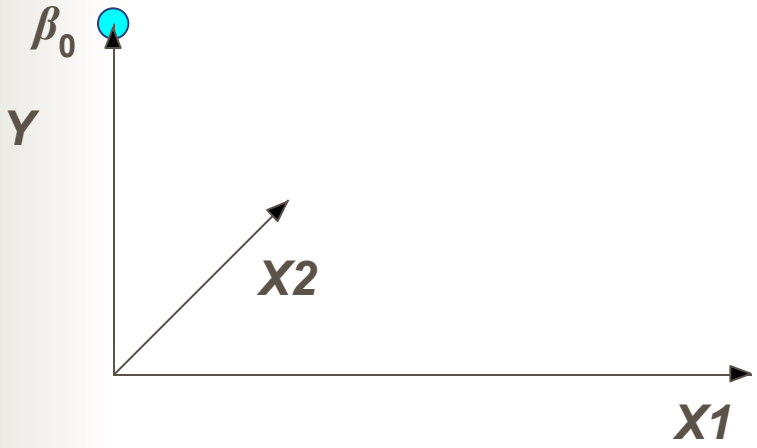
MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

Y – weekly salary ($)

X1 – length of employment (in months)

X2 – age (in years)



Literally the intercept gives *weekly salary* for those respondents who have no age (??) and no length of employment (??). Hence a literal interpretation of β_0 would be unwise.

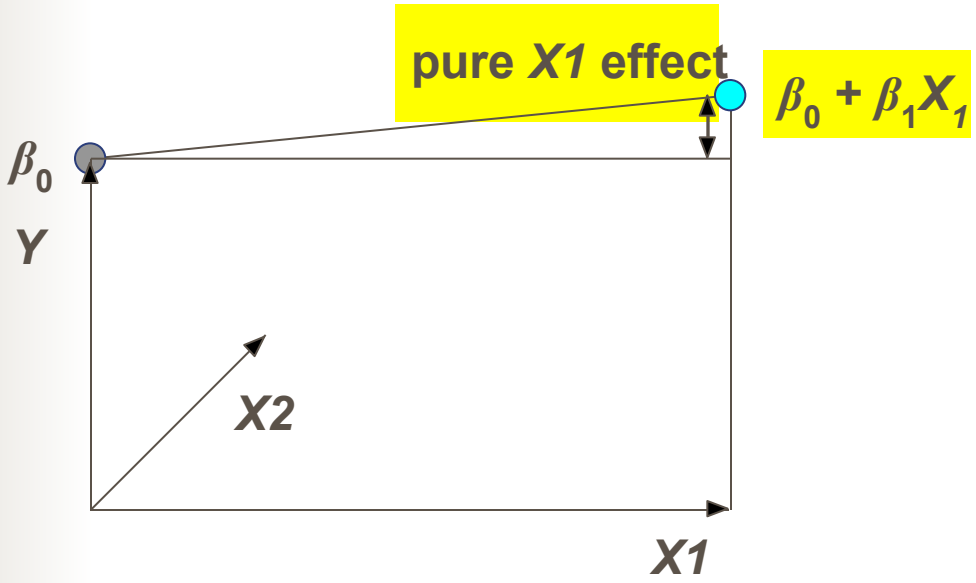
MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

Y – weekly salary ($)

X1 – length of employment (in months)

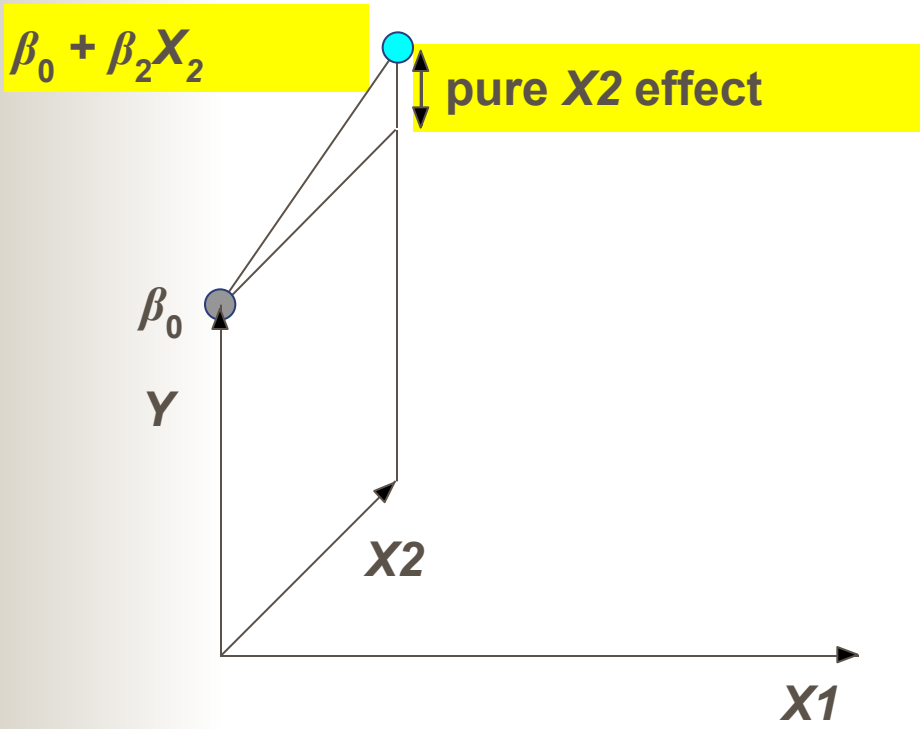
X2 – age (in years)



The next term on the right side of the equation gives the effect of *X1*. A one month of employment increase in *X1* causes weekly salary to increase by $\beta_1$dollars, holding *X2* constant.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

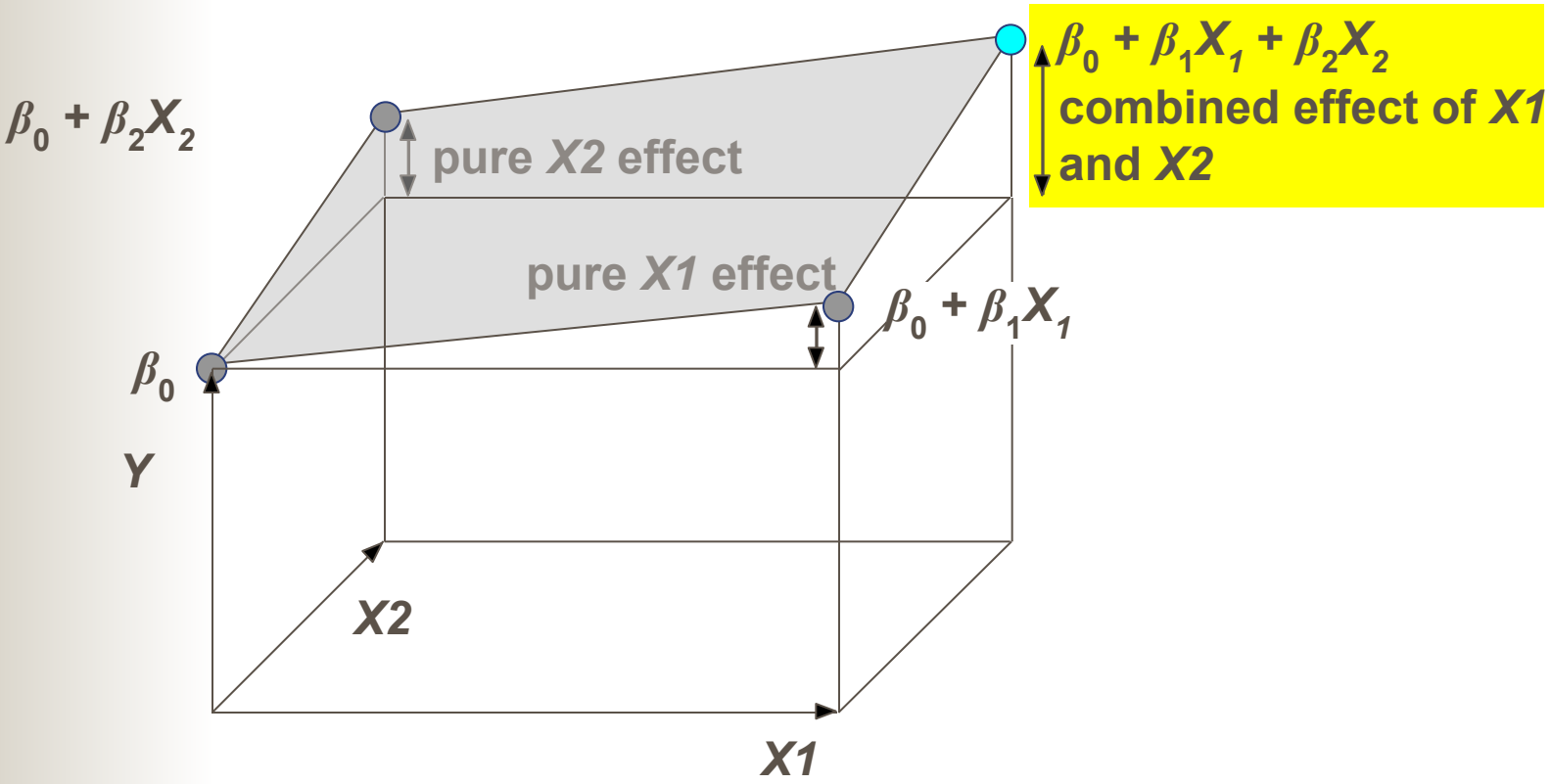
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$



Similarly, the third term gives the effect of variations in *X2*. A one year of age increase in *X2* causes weekly salary to increase by β_2 dollars, holding *X1* constant.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

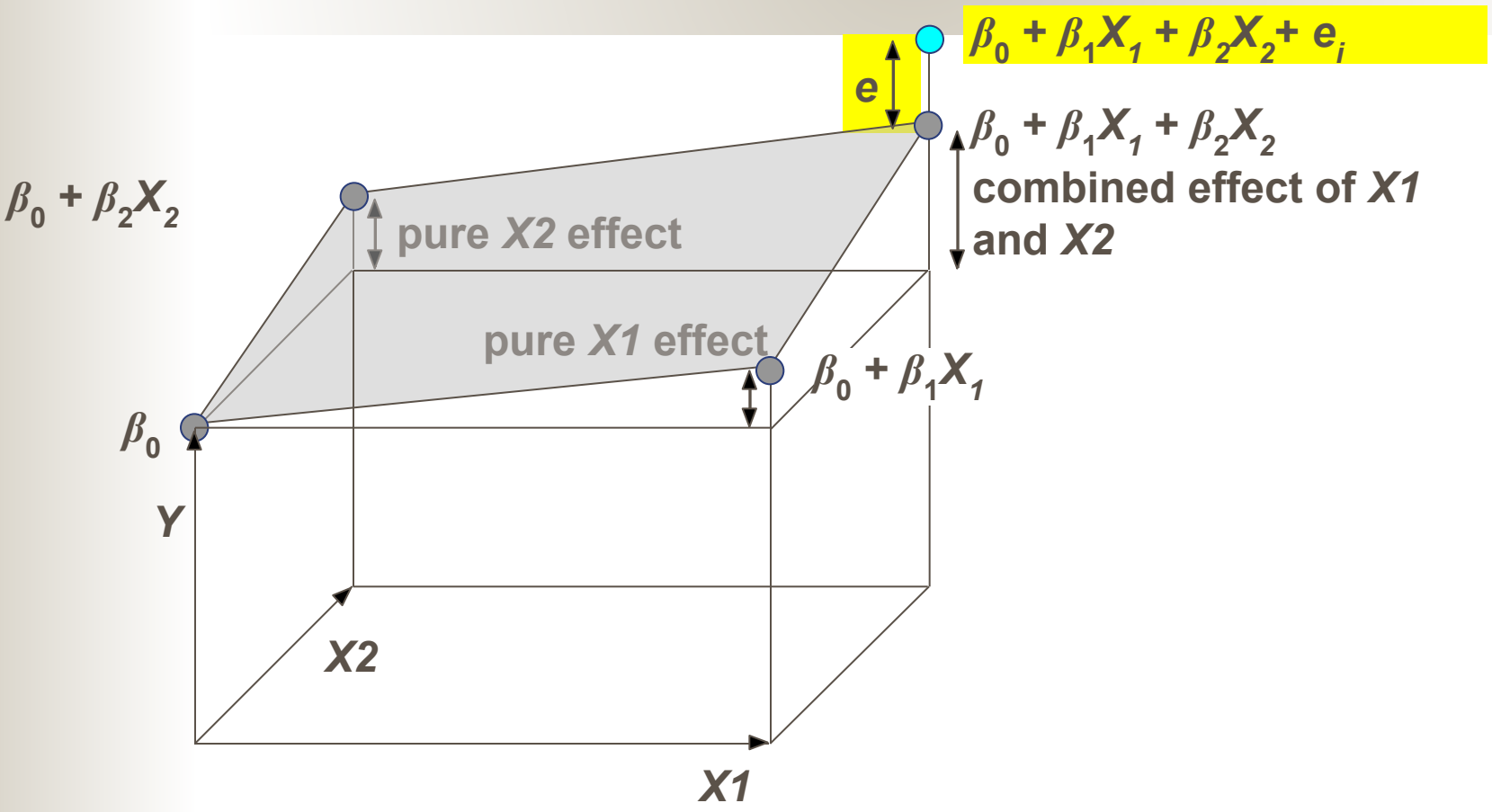
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$



Different combinations of *X1* and *X2* give rise to values of *weekly salary* which lie on the plane shown in the diagram, defined by the equation $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. This is the nonrandom component of the model.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

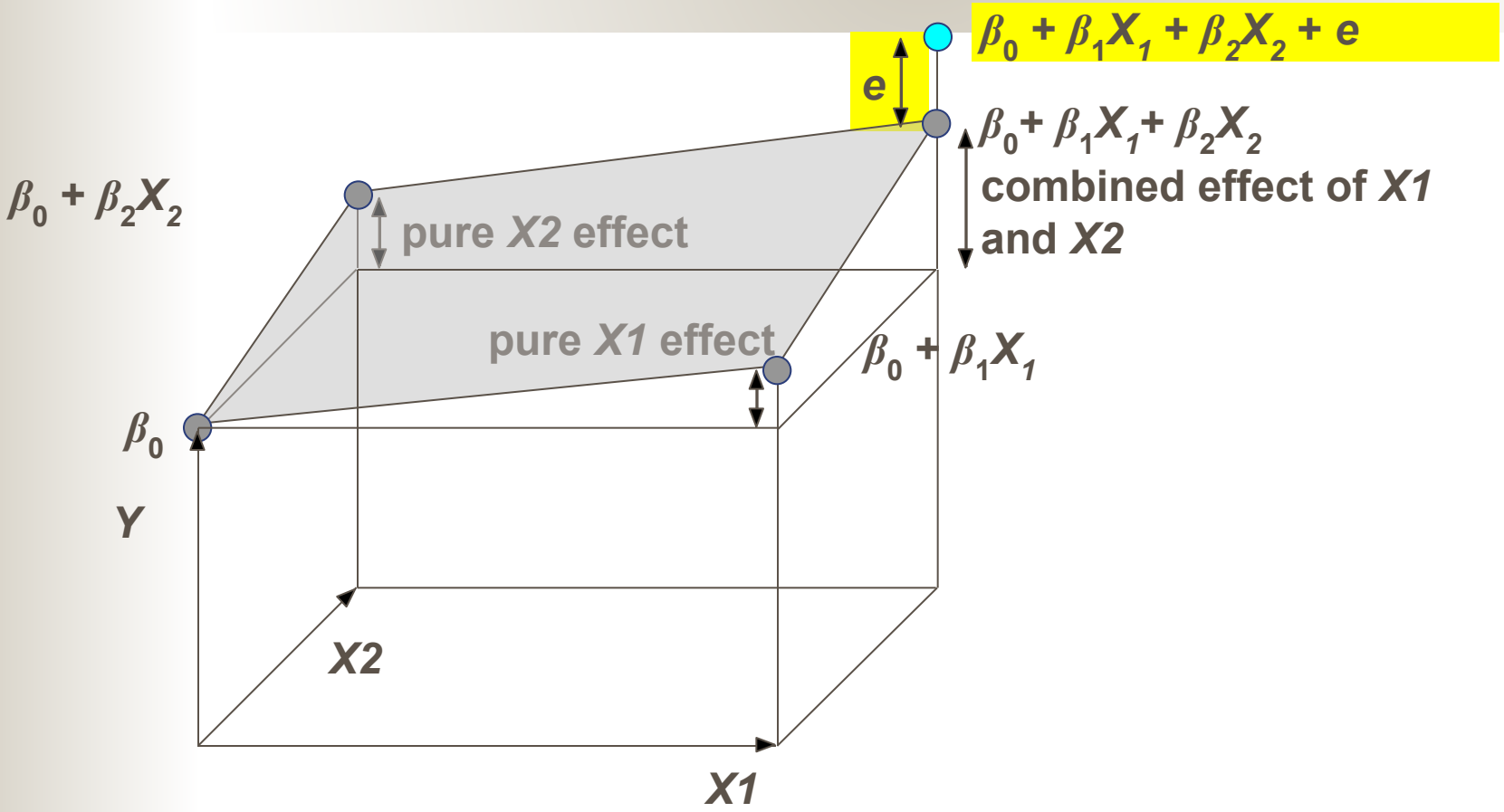
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$



The final element of the model is the error term, *e*. This causes the actual values of *Y* to deviate from the plane. In this observation, *e* happens to have a positive value.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

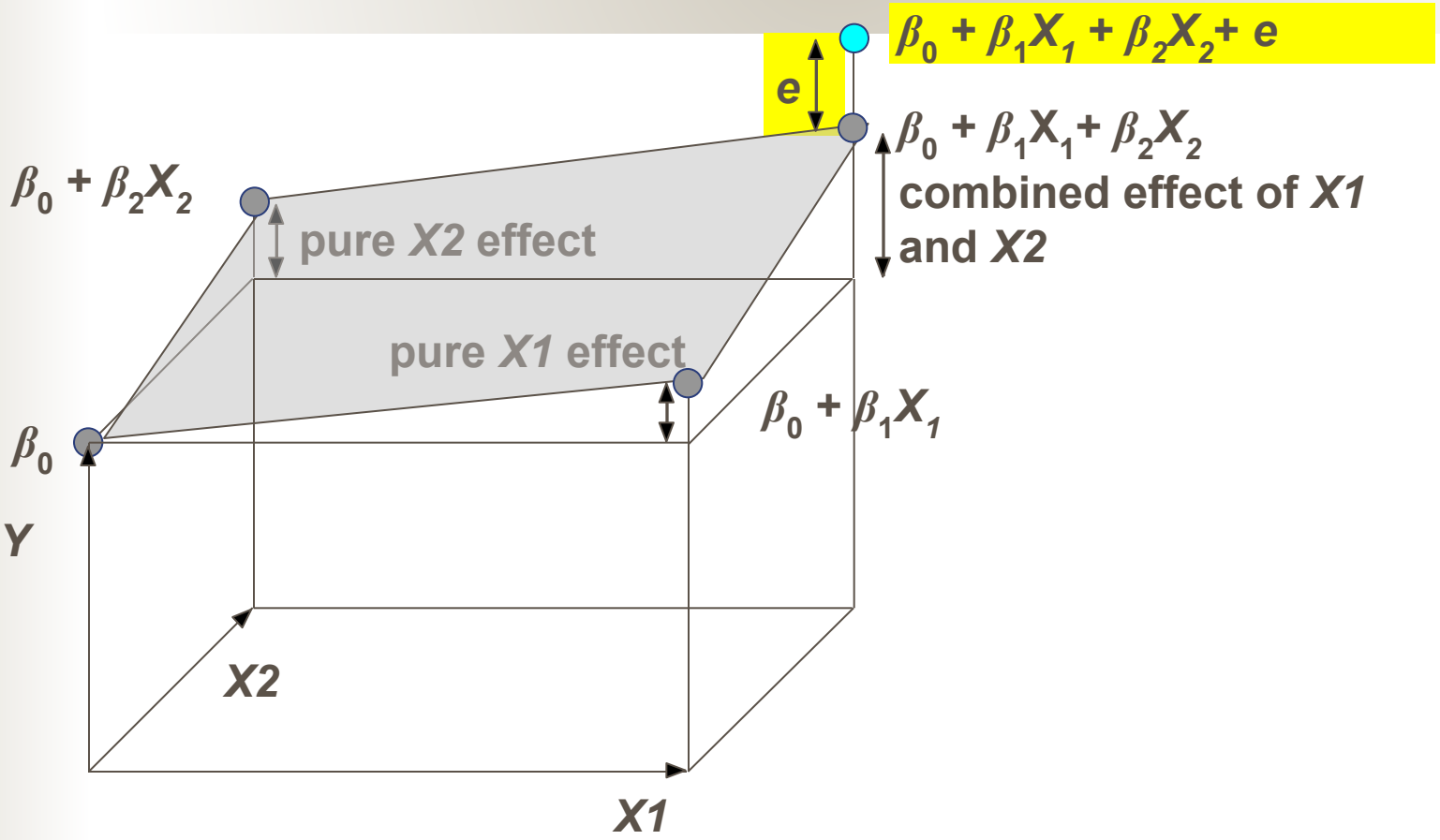
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$



A sample consists of a number of observations generated in this way. Note that the interpretation of the model does not depend on whether *X1* and *X2* are correlated or not.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$



However we do assume that the effects of *X1* and *X2* on *salary* are additive. The impact of a difference in *X1* on *salary* is not affected by the value of *X2*, or vice versa.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

Slope coefficients are interpreted as partial slope/partial regression coefficients:

- **b_1 = average change in Y associated with a unit change in X_1, with the other independent variables held constant (all else equal);**

- **b_2 = average change in Y associated with a unit change in X_2, with the other independent variables held constant (all else equal).**

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

The regression coefficients are derived using the same least squares principle used in simple regression analysis. The fitted value of *Y* in observation *i* depends on our choice of b_0, b_1, and b_2.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}$$

The residual e_i in observation *i* is the difference between the actual and fitted values of *Y*.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$SSE = \sum e_i^2 = \sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2$$

We define *SSE*, the sum of the squares of the residuals, and choose b_0, b_1, and b_2 so as to minimize it.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$SSE = \sum e_i^2 = \sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2$$

$$= \sum (Y_i^2 + b_0^2 + b_1^2 X_{1i}^2 + b_2^2 X_{2i}^2 - 2b_0 Y_i - 2b_1 X_{1i} Y_i$$
$$- 2b_2 X_{2i} Y_i + 2b_0 b_1 X_{1i} + 2b_0 b_2 X_{1i} + 2b_1 b_2 X_{1i} X_{2i})$$
$$= \sum Y_i^2 + nb_0^2 + b_1^2 \sum X_{1i}^2 + b_2^2 \sum X_{2i}^2 - 2b_0 \sum Y_i$$
$$- 2b_1 \sum X_{1i} Y_i - 2b_2 \sum X_{2i} Y_i + 2b_0 b_1 \sum X_{1i}$$
$$+ 2b_0 b_2 \sum X_{2i} + 2b_1 b_2 \sum X_{1i} X_{2i}$$

$$\frac{\partial SSE}{\partial b_0} = 0 \qquad \frac{\partial SSE}{\partial b_1} = 0 \qquad \frac{\partial SSE}{\partial b_2} = 0$$

First we expand *SSE* as shown, and then we use the first order conditions for minimizing it.

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_1) - \mathrm{Cov}(X_1, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

We thus obtain three equations in three unknowns. Solving for b_0, b_1, and b_2, we obtain the expressions shown above.

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_1) - \mathrm{Cov}(X_1, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

The expression for b_0 is a straightforward extension of the expression for it in simple regression analysis.

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_1) - \mathrm{Cov}(X_1, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

However, the expressions for the slope coefficients are considerably more complex than that for the slope coefficient in simple regression analysis.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$b_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_1) - \mathrm{Cov}(X_1, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

For the general case when there are many explanatory variables, ordinary algebra is inadequate. It is necessary to switch to matrix algebra.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES

In matrix notation OLS may be written as:

$$Y = Xb + e$$

The normal equations in matrix form are now

$$X^T Y = X^T X b$$

And when we solve it for *b* we get:

$$b = (X^T X)^{-1} X^T Y$$

where Y is a column vector of the Y values and X is a matrix containing a column of ones (to pick up the intercept) followed by a column of the X variables containing the observations on them and *b* is a vector containing the estimators of regression parameters.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} \end{bmatrix} \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

MATRIX ALGEBRA: SUMMARY

A **vector** is a collection of n numbers or elements, collected either in a column (a **column vec**tor) or in a row (a **row vector**).
A **matrix** is a collection, or array, of numbers of elements in which the elements are laid out in columns and rows. The dimension of matrix is $n \times m$ where n is the number of rows and m is the number of columns.

Types of matrices

A matrix is said to be **square** if the number of rows equals the number of columns. A square matrix is said to be **symmetric** if its (i, j) element equals its (j, i) element. A **diagonal** matrix is a square matrix in which all the off-diagonal elements equal zero, that is, if the square matrix A is diagonal, then $a_{ij} = 0$ for $i \neq j$.
The **transpose** of a matrix switches the rows and the columns. That is, the transpose of a matrix turns the $n \times m$ matrix A into the $m \times n$ matrix denoted by A^T, where the (i, j) element of A becomes the (j, i) element of A^T; said differently, the transpose of a matrix A turns the rows of A into the columns of A^T. The **inverse** of the matrix A is defined as the matrix for which $A^{-1}A=1$. If in fact the inverse matrix A^{-1} exists, then A is said to be **invertible** or **nonsingular**.

Vector and matrix multiplication

The matrices A and B can be multiplied together if they are conformable, that is, if the number of columns of A equals the number of rows of B. In general, matrix multiplication does not commute, that is, in general $AB \neq BA$.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE

Data for weekly salary based upon the length of employment and age of employees of a large industrial corporation are shown in the table.

Employee	Weekly salary ($)	Length of employment (X1, months)	Age (X2, years)
1	639	330	46
2	746	569	65
3	670	375	57
4	518	113	47
5	602	215	41
6	612	343	59
7	548	252	45
8	591	348	57
9	552	352	55
10	529	256	61
11	456	87	28
12	674	337	51
13	406	42	28
14	529	129	37
15	528	216	46
16	592	327	56

Calculate the OLS estimates for regression coefficients for the available sample. Comment on your results.

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE

Y-weekly salary ($) X1 –length of employment (months) X2-age (years)

i	Y	X_1	X_2
1	639	330	46
2	746	569	65
3	670	375	57
4	518	113	47
5	602	215	41
6	612	343	59
7	548	252	45
8	591	348	57
9	552	352	55
10	529	256	61
11	456	87	28
12	674	337	51
13	406	42	28
14	529	129	37
15	528	216	46
16	592	327	56

X

1	330	46
1	569	65
1	375	57
1	113	47
1	215	41
1	343	59
1	252	45
1	348	57
1	352	55
1	256	61
1	87	28
1	337	51
1	42	28
1	129	37
1	216	46
1	327	56

Y

639
746
670
518
602
612
548
591
552
529
456
674
406
529
528
592

X^T

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
330	569	375	113	215	343	252	348	352	256	87	337	42	129	216	327
46	65	57	47	41	59	45	57	55	61	28	51	28	37	46	56

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE

	16	4291	779
X^TX	4291	1417105	227875
	779	227875	39771

	9192
X^TY	2617701
	457709

det	2105037674

min11	1417105	227875
	227875	39771

det	4432667330

min12	4291	227875
	779	39771

det	-6857264

min13	4291	1417105
	779	227875

det	-126113170

min21	4291	779
	227875	39771

det	-6857264

min22	16	779
	779	39771

det	29495

min23	16	4291
	779	227875

det	303311

min31	4291	779
	1417105	227875

det	-126113170

min32	16	779
	4291	227875

det	303311

min33	16	4291
	4291	1417105

det	4260999

matrix of minors

cofactor matrix

minors XTX	4432667330	-6857264	-126113170
	-6857264	29495	303311
	-126113170	303311	4260999

	4432667330	6857264	-126113170
$(X^TX)D$	6857264	29495	-303311
	-126113170	-303311	4260999

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE

$(X^TX)^{-1}$	2,1057	0,0033	-0,0599
	0,0033	0,00001	-0,0001
	-0,0599	-0,0001	0,002

	9192
X^TY	2617701
	457709

vector of parameters' estimates

$b=(X^TX)^{-1}X^TY$	461,85	=b0
	0,671	=b1
	-1,383	=b2

Y-weekly salary ($) **X1 –length of employment (months)** **X2-age (years)**

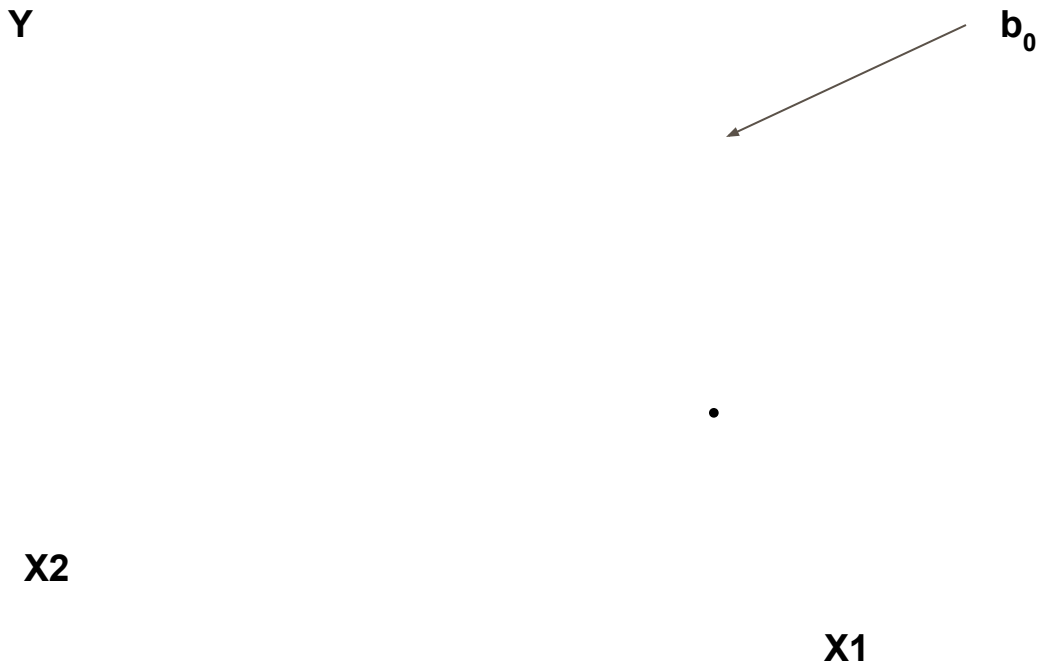
Our regression equation with two predictors (X1, X2):

$$\hat{y}_i = 461{,}85 + 0{,}671 \cdot X_1 - 1{,}383 \cdot X_2$$

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE

These are our data points in 3dimensional space (graph drawn using *Statistica 6.0*)

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE



Data points with the regression surface (*Statistica 6.0*)

MULTIPLE REGRESSION WITH TWO EXPLANATORY VARIABLES: EXAMPLE

Data points with the regression surface (*Statistica 6.0*) after rotation.