

## Regression with Matrix Algebra

Describe the solution for regression weights for raw scores using matrix algebra.

Describe the solution for standardized regression weights from a correlation matrix using matrix algebra.

Describe the sampling distributions of the  $b$  and beta weights.

What is the meaning of the covariance or correlation matrix of the  $b$  weights? How is it used?

In raw score form the regression equation is:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

This says that  $Y$ , our dependent variable, is composed of a linear part and error. The linear part is composed of an intercept,  $a$ , and  $k$  independent variables,  $X_1 \dots X_k$  along with their associated raw score regression weights  $b_1 \dots b_k$ .

In matrix terms, the same equation can be written:

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{e}$$

$\mathbf{y} =$	$\mathbf{X}$	$\mathbf{b} +$	$\mathbf{e}$
$N \times 1$	$N \times (1+k)$	$(1+k) \times 1$	$N \times 1$

This says to get  $Y$  for each person, multiply each  $X_i$  by the appropriate  $b_i$ , add them and then add error. The matrices look like this:

$Y_1$		1	$X_{11}$	$X_{12}$	.	.	.	$X_{1k}$		$a$		$e_1$
$Y_2$	=	1	$X_{21}$	$X_{22}$				$X_{2k}$		$b_1$	+	$e_2$
.		.	.	.				.		$b_2$		
.		.	.	.				.		.		

.		.	.	.				.		.		
$Y_N$		1	$X_{N1}$	$X_{N2}$	.	.	.	$X_{Nk}$		$b_k$		$e_N$

With raw scores, we create an augmented design matrix  $X$ , which has an extra column of 1s in it for the intercept. For each person, the 1 is used to add the intercept in the first row of the column vector  $\mathbf{b}$ .

If we solve for the  $b$  weights, we find that

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

To give you an idea why it looks like that, first remember the regression equation:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

Let's assume that error will equal zero on average and forget it to sketch a proof:

$$\mathbf{y} = \mathbf{X}\mathbf{b}$$

Now we want to solve for  $\mathbf{b}$ , so we need to get rid of  $\mathbf{X}$ . First we will make  $\mathbf{X}$  into a nice square, symmetric matrix by premultiplying both sides of the equation by  $\mathbf{X}'$ :

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$$

And now we have a square, symmetric matrix that with any luck has an inverse, which we will call  $(\mathbf{X}'\mathbf{X})^{-1}$ . Multiply both sides by this inverse, and we have

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

It turns out that a matrix multiplied by its inverse is the identity matrix ( $\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$ ):

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

and a matrix multiplied by the identity matrix is itself ( $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ ):

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

which is the desired result.

### A numerical example with one independent variable

Suppose our data look like this:

Y	X
1	6
2	7
3	8
3	9
4	7

We can represent the regression problem as:

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

										-1							
							1	6									1
	1	1	1	1	1		1	7			1	1	1	1	1		2
	6	7	8	9	7		1	8			6	7	8	9	7		3
							1	9									3
							1	7									4

	$\mathbf{X}'$	$\mathbf{X}$	$\mathbf{X}'$	$\mathbf{y}$
	5	37		
$\mathbf{X}'\mathbf{X} =$	N	$\Sigma X$		
	37	279		
	$\Sigma X$	$\Sigma X^2$		
	13			
$\mathbf{X}'\mathbf{y} =$	$\Sigma Y$			
	99			
	$\Sigma XY$			

Now to find the inverse of  $\mathbf{X}'\mathbf{X}$ .

If  $\mathbf{X}$  is a 2 by 2 matrix, then the inverse is a simple function of the elements each divided by the determinant.

Suppose  $\mathbf{A}$  has the following elements:

$\mathbf{A} =$	$a$	$b$	

	$c$	$d$	

Then  $|A|$  is  $ad-bc$ . The inverse of  $X$  is

$A^{-1} =$	$d/ X $	$-b/ X $	
	$-c/ X $	$a/ X $	

Or

$A^{-1} =$	$d/(ad-bc)$	$-b/(ad-bc)$
	$-c/(ad-bc)$	$a/(ad-bc)$

In our example

$X'X =$	5	37
	37	279

And

$(X'X)^{-1} =$	10.73	-1.42	(rounded)
	-1.42	.19	

To find  $b$ , we find

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

10.7308	-1.42308		13		-1.38	
-1.42308	.192308		99	=	.54	
$(\mathbf{X}'\mathbf{X})^{-1}$			$\mathbf{X}'\mathbf{y}$		$\mathbf{b}$	

Therefore  $a = -1.38$  and  $b = .54$ . The regression equation:

$$Y' = -1.38 + .54X.$$

### Deviation Scores and 2 IVs

The raw score computations shown above are what the statistical packages typically use to compute multiple regression.

However, we can also use matrix algebra to solve for regression weights using (a) deviation scores instead of raw scores, and (b) just a correlation matrix. We will, of course, now have to do both.

The deviation score formulation is nice because the matrices in this approach contain entities that are conceptually more intuitive. So let us suppose that we have subtracted the mean of Y from each Y score, and the mean of each X from its column. In that case, we have

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

which is the same equation as for raw scores except that the subscript  $d$  denotes deviation scores. The design matrix  $\mathbf{X}$  does not contain an initial column of 1s because the intercept for this model will be zero. (Why is the intercept zero?) Because the scores are deviation scores, the matrix  $\mathbf{X}'\mathbf{X}$  will be an SSP matrix:

	$\sum x_1^2$	$\sum x_1x_2$	...	$\sum x_1x_k$

$\mathbf{X}'\mathbf{X} =$	$\sum x_2 x_1$	$\sum x_2^2$	...	sym
	...		...	
	$\sum x_k x_1$	...		$\sum x_k^2$

And the matrix  $\mathbf{X}'\mathbf{y}$  will be a sums of cross products matrix:

	$\sum x_1 y$
$\mathbf{X}'\mathbf{y} =$	$\sum x_2 y$
	...
	$\sum x_k y$

For a two-variable problem, the solution for the  $b$  weights looks like this:

			-1	
	$\sum x_1^2$	$\sum x_1 x_2$		$\sum x_1 x_k$
$\mathbf{b} =$	$\sum x_2 x_1$	$\sum x_2^2$		$\sum x_2 y$
	$\mathbf{X}'\mathbf{X}$			$\mathbf{X}'\mathbf{y}$

This says multiply the matrix of cross products of  $x$  and  $y$  by the inverse of the SSCP matrix to find the  $b$  weights.

The determinant of  $\mathbf{X}'\mathbf{X}$  is found by multiplying the elements of the main diagonal and subtracting the product of the off-diagonal elements. The inverse of  $\mathbf{X}'\mathbf{X}$  is a simple function of the elements of  $\mathbf{X}'\mathbf{X}$  each divided by the determinant.

	$\sum x_1^2$	$\sum x_1 x_2$		
$ \mathbf{X}'\mathbf{X}  =$	$\sum x_2 x_1$	$\sum x_2^2$	$=$	$(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2$
		$\frac{\sum x_2^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	$\frac{-\sum x_1 x_2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	
$\mathbf{X}'\mathbf{X}^{-1} =$		$\frac{-\sum x_2 x_1}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	$\frac{\sum x_1^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	

The  $\mathbf{b}$  weights will be found by multiplying the above matrix by  $\mathbf{X}'\mathbf{y}$ :

	$\frac{\sum x_2^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	$\frac{-\sum x_1 x_2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	$\sum x_1 y$
$\mathbf{b} =$	$\frac{-\sum x_2 x_1}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	$\frac{\sum x_1^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$	$\sum x_2 y$
	$\mathbf{X}'\mathbf{X}^{-1}$		$\mathbf{X}'\mathbf{y}$
	$\frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$		



$$\mathbf{b} = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

Note that these formulas match those I gave you earlier without matrix algebra.

Recall our 2-variable problem in which we predicted

Chevy mechanics' job performance scores from mechanical aptitude and conscientiousness scores. One of the first things we did was to compute the following matrix:

	y	x1	x2
Y	29.75	139.5	90.25
x1	0.77	1091.8	515.5
x2	0.72	0.68	521.75

The table above has sums of squares, cross products, and correlations in it:

	y	x1	x2
Y	$\sum y^2$	$\sum x_1 y$	$\sum x_2 y$
x1	$r_{y1}$	$\sum x_1^2$	$\sum x_1 x_2$
x2	$r_{y2}$	$r_{12}$	$\sum x_2^2$

To solve for the  $b$  weights, we grab the following pieces:

--	--	--	--	--

			-1			
	$\sum x_1^2$	$\sum x_1x_2$		$\sum x_1y$		
<b>b =</b>	$\sum x_2x_1$	$\sum x_2^2$		$\sum x_2y$		
	<b>X'X</b>			<b>X'y</b>		
			-1			
	1091.8	515.5		139.5		
<b>b =</b>	515.5	521.75		90.25		
	<b>X'X</b>			<b>X'y</b>		
	.0017	-.0016		139.5	=	.086
<b>b =</b>	-.0016	.0036		90.25		.087
	<b>X'X<sup>-1</sup></b>			<b>X'y</b>	=	<b>b</b>

Note that this result agrees with our earlier estimates computed without matrix algebra.

### Obtaining $\beta$ weights from a Correlation Matrix

With two standardized variables, our regression equation is

$$z_y' = \beta_1 z_1 + \beta_2 z_2$$

To solve for beta weights, we just find:

$$\beta = R^{-1}r$$

where **R** is the correlation matrix of the predictors (X variables) and **r** is a column vector of correlations between Y and each X.

Recall our earlier matrix:

	y	x1	x2
Y	29.75	139.5	90.25
x1	0.77	1091.8	515.5
x2	0.72	0.68	521.75

The matrix **R** is:

	x1	x2
x1	1	.68
x2	0.68	1

And the matrix **r** is:

y	
x1	0.77
x2	0.72

Therefore our beta weights are:

	1.86	-1.26		.77	=	.52
$\beta =$	-1.26	1.86		.72		.37

	$\mathbf{R}^{-1}$		$\mathbf{r}$	=	$\beta$
--	-------------------	--	--------------	---	---------

Note that this result agrees with our earlier estimates of beta weights calculated without matrix algebra.

If the predictors are all orthogonal, then the matrix  $\mathbf{R}$  is the identity matrix  $\mathbf{I}$ , and then  $\mathbf{R}^{-1}$  will equal  $\mathbf{R}$ . In such a case, the  $\beta$  weights will equal the simple correlations (we have noted before that  $r$  and  $\beta$  are the same when the independent variables are uncorrelated). The matrix inverse discounts the correlations in  $\mathbf{r}$  to make them weights that correspond to the unique parts of each predictor, that is,  $\beta$  weights. The inverse operation in a sense makes the predictors orthogonal.

### Sampling Covariance of Regression Weights

We can define a population in which a regression equation describes the relations between  $Y$  and some predictors, e.g.,

$$Y'_{JP} = \alpha + \beta_1 MC + \beta_2 C,$$

Where  $Y$  is job performance,  $\alpha$  and  $\beta$  are population parameters,  $MC$  is mechanical comprehension test scores, and  $C$  is conscientiousness test scores. If we take repeated samples from our population and estimate  $\beta_1$  and  $\beta_2$ , we will have two sampling distributions, one for each slope estimate. Each of the slope distributions will have a variance, known as the sampling variance (this variance is used to construct confidence intervals and significance tests). Because all the variables in this equation are measured on the same people, the measures are dependent. There will be a covariance between the two slope estimates. In other words, the two slope estimates are dependent and may covary (be correlated) across samples. Suppose our population parameters are as follows:

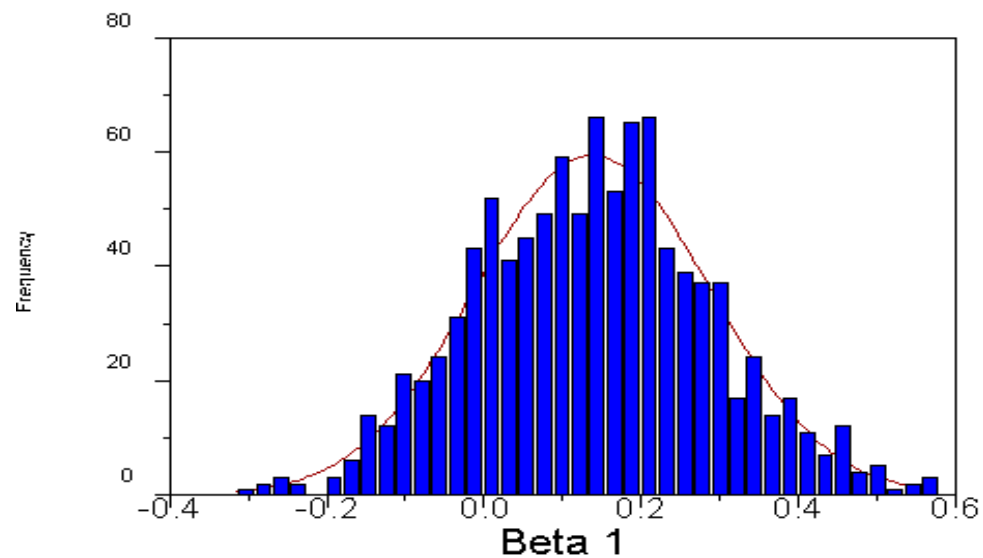
$\rho$ (population correlations)	$Y$	$X_1$	$X_2$
$Y$	1		
$X_1$	.45	1	
$X_2$	.50	.80	1

Our population beta weights will be:

	1	.8	-1	.45
	.8	1		.50
	<b>R</b>			<b>r</b>
	2.78	-2.22		.45
=	-2.22	2.78		.50
	<b>R<sup>-1</sup></b>			<b>r</b>
				$\beta$

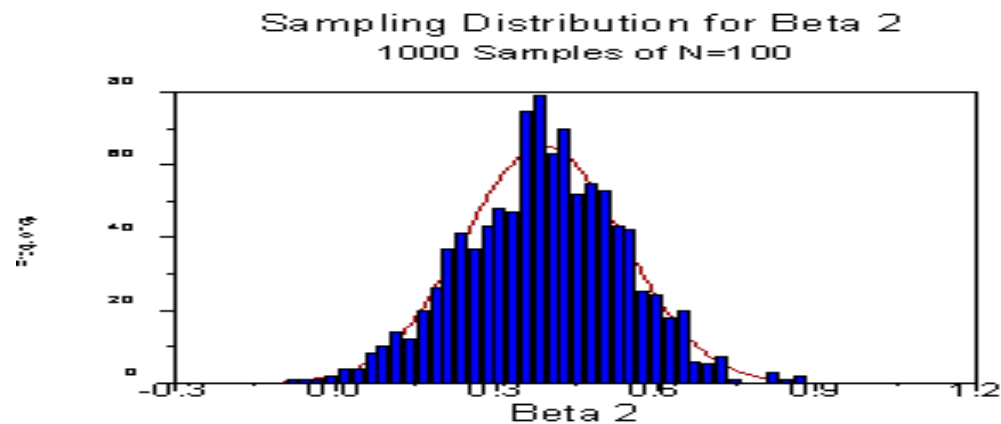
(Note: These  $\beta$  weights are betas both in the sense of being standardized and being population values.) I took 1,000 samples of size 100 from this population. The sampling distribution for beta1 looks like this:

### Sampling Distribution beta 1



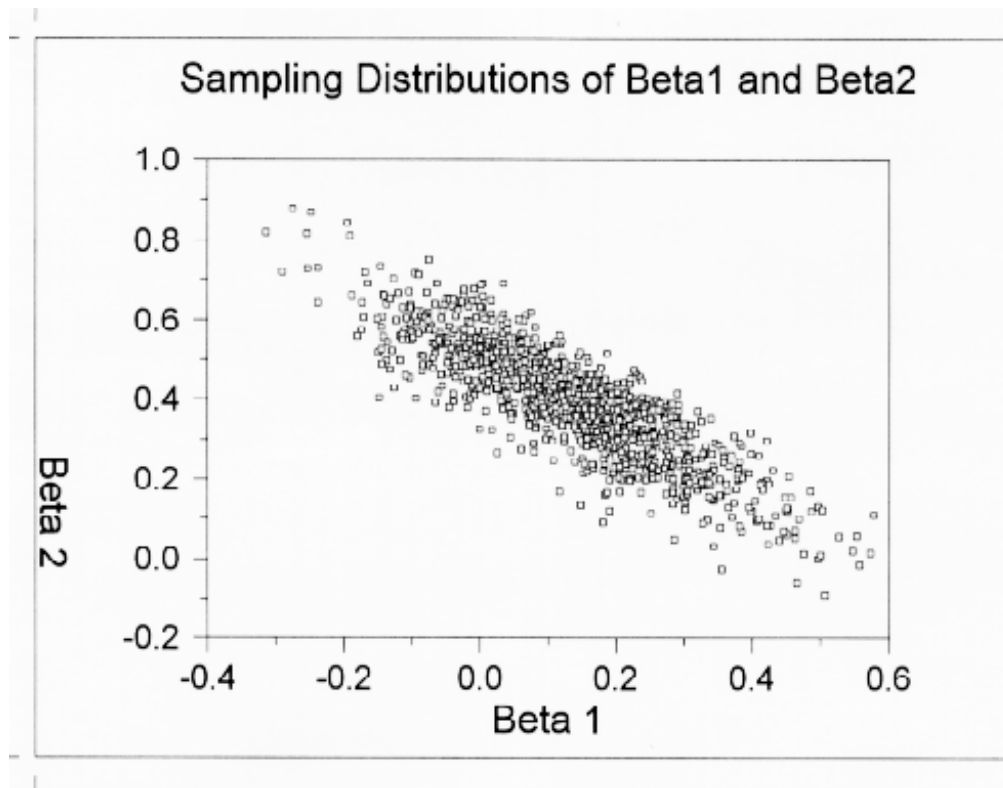
Its mean is .1376, which is close to its expected value of .1388, and its standard deviation is .1496.

The sampling distribution for beta2 is:



The mean of the distribution is .3893, which is close to the expected value of .3889, and the standard deviation is .1482.

The scatter plot of the pairs of beta weights for the 1000 samples is:



As you can see, there is a negative correlation between the beta weights. When one is relatively large, the other is relatively small. This happens whenever the predictors are correlated. When the predictors are correlated, one predictor tends to get more than its share in a given sample. Correlated predictors are pigs -- they hog the variance in  $Y$ .

The variance covariance matrix of the  $b$  weights is:

$$\mathbf{C} = \frac{\mathbf{e}'\mathbf{e}}{N - k - 1} (\mathbf{X}_d'\mathbf{X}_d)^{-1} = s_{y|2 \dots k}^2 (\mathbf{X}_d'\mathbf{X}_d)^{-1}$$

which is the variance of estimate (mean square residual) times the inverse of the SSCP matrix (the inverse of the deviation scores premultiplied by the transpose of the deviation scores). The diagonal elements of this matrix are the sampling variances of the  $b$  weights. If we take the square roots of these elements, we have the standard errors of the  $b$  weights, which are used in testing them. The off-diagonal elements of  $\mathbf{C}$  are the covariances of the  $b$  weights.

In our example, the sum of squares of error was 9.88. The df was  $20 - 2 - 1 = 17$ . Therefore, the variance of estimate is  $9.88/17 = .58$ . The inverse of our SSCP matrix is

.0017	-.0016
-.0016	.0036
$(\mathbf{X}'\mathbf{X})^{-1}$	

Therefore our variance covariance matrix  $\mathbf{C}$  is

.0001	-.0001
-.0001	.0002
$\mathbf{C}$	

The standard error of  $b_1$  is  $\text{sqrt}(c_{11}) = .031$ .

The standard error of  $b_2$  is  $\text{sqrt}(c_{22}) = .046$ .

The test statistics for  $b_1$  and  $b_2$  are

$$t = b_1/s_1 = .086/.031 = 2.81.$$

$$t = b_2/s_2 = .087/.046 = 1.89.$$

These  $t$  values agree with our earlier calculations made without matrix algebra within rounding error.

To test for the difference between slopes (e.g., is  $b_1$  equal to  $b_2$ ) includes terms for both the variances of the  $b$  weights and also the covariance of the  $b$  weights. Specifically,



$$s_{b_1-b_2} = \sqrt{c_{11} + c_{22} - 2c_{12}}$$

This test is analogous to a two-sample t-test where we have the standard error of the difference defined as

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2}$$

The difference is that the error variances for the two means are independent, and so the covariance between the two is zero. The  $b$  weights are dependent, so we need to include the covariance to be accurate.

In our example

$$s_{b_1-b_2} = \sqrt{.0001 + .0002 - 2(-.0001)} = \sqrt{.0005} = .022$$

$$t = \frac{b_1 - b_2}{s_{b_1-b_2}} = \frac{.086 - .087}{.022} = -.045$$

Note: This test is only meaningful when both  $b$  weights are of the same kind and measured on the same scale. For example, if one IV were points scored by the home team and the other was points scored by the visiting team for the same type of game this would be okay. But if one were for dollars and one were for days, this would not be okay, because the size of the unit influences the size of the  $b$  weight.