Regression: Standardized Coefficients

1. The Regression Equation: Unstandardized Coefficients

Suppose a researcher is interested in determining whether academic achievement is related to students' time spent studying and their academic ability. Hypothetical data for these variables are presented in Table 1. In the corresponding regression equation for this model, achievement is denoted Y, time spent studying X_1, and academic ability X_2.

1a. Population Equation

The population regression model is

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i, \quad (1)$$

where

Y_i signifies the i^{th} student's achievement score;
β_1 is the population partial regression coefficient expressing the relationship between X_1 and Y, controlling for X_2;
β_2 is the population partial regression coefficient expressing the relationship between X_2 and Y, controlling for X_1;
β_0 is the population intercept for the equation; and
ε_i is, supposedly, a random error.

1b. Sample Equation

The sample regression equation for the hypothetical example of achievement is:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i, \quad (2)$$

where b_0 is the sample intercept; b_1 is the sample regression coefficient for X_1 controlling for the effect of X_2; b_2 is the sample regression coefficient for X_2 controlling for the effect of X_1; and e_i is the sample error term.

Table 1
Achievement, Time Spent Studying, and Academic Ability

Achievement	Time (in hours)	Ability
88	8	6
75	6	2
64	0	2
99	9	9
95	5	9
93	8	7
85	7	5
82	5	4
79	1	5
78	1	3
91	4	7
85	4	9

Note. Higher scores indicate higher levels of each variable.

1c. SPSS Results

Least squares results for the sample data appear below.

Descriptive Statistics

	Mean	Std. Deviation	N
achievement	84.5000	9.70941	12
time	4.8333	2.97973	12
ability	5.6667	2.60536	12

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	63.902	2.836		22.535	.000
	time	1.302	.437	.400	2.980	.015
	ability	2.524	.500	.677	5.050	.001

a Dependent Variable: achievement

1d. Unstandardized Coefficient Interpretation

The sample prediction model with estimates follows:

$Y' = b_0 + b_1X_{1i} + b_2X_{2i}$,

Achievement' = 63.90 + 1.30(time) + 2.52(ability)

Coefficient interpretation is the same as previously discussed in regression.

b_0 = 63.90: The predicted level of achievement for students with time = 0.00 and ability = 0.00.

b_1 = 1.30: A 1 hour increase in time is predicted to result in a 1.30 point increase in achievement holding constant ability.

b_2 = 2.52: A 1 point increase in ability is predicted to result in a 2.52 point increase in achievement holding constant time.

2. Z Scores

Recall that scores can be converted to a Z score which has a mean of 0.00 and a standard deviation of 1.00. One may use the following formula to calculate a Z score:

$$Z = \frac{X - M}{sd}$$

where X is the raw score, M is the mean, and sd is the standard deviation. Each of the three sets of scores in Table 1 is converted below to Z scores. The M and sd are provided above in the SPSS output.

Achievement converted to Z score: $Z_{Achievement}$

Achievement	Mean	X - M	Z = (X-M)/SD
88	84.5	3.5	0.360475
75	84.5	-9.5	-0.97843
64	84.5	-20.5	-2.11135
99	84.5	14.5	1.493397
95	84.5	10.5	1.081425
93	84.5	8.5	0.875439
85	84.5	0.5	0.051496
82	84.5	-2.5	-0.25748
79	84.5	-5.5	-0.56646
78	84.5	-6.5	-0.66945
91	84.5	6.5	0.669454
85	84.5	0.5	0.051496

Time converted to Z score: Z_{Time}

Time	Mean	X - M	Z = (X-M)/SD
8	4.8333	3.1667	1.062747296
6	4.8333	1.1667	0.391545543
0	4.8333	-4.8333	-1.622059717
9	4.8333	4.1667	1.398348172
5	4.8333	0.1667	0.055944666
8	4.8333	3.1667	1.062747296
7	4.8333	2.1667	0.727146419
5	4.8333	0.1667	0.055944666
1	4.8333	-3.8333	-1.28645884
1	4.8333	-3.8333	-1.28645884
4	4.8333	-0.8333	-0.27965621
4	4.8333	-0.8333	-0.27965621

Ability converted to Z score: $Z_{Ability}$

Ability	Mean	X - M	Z = (X-M)/SD
6	5.6667	0.3333	0.127928578
2	5.6667	-3.6667	-1.407367888
2	5.6667	-3.6667	-1.407367888
9	5.6667	3.3333	1.279400927
9	5.6667	3.3333	1.279400927
7	5.6667	1.3333	0.511752694
5	5.6667	-0.6667	-0.255895538
4	5.6667	-1.6667	-0.639719655
5	5.6667	-0.6667	-0.255895538
3	5.6667	-2.6667	-1.023543771
7	5.6667	1.3333	0.511752694
9	5.6667	3.3333	1.279400927

3. Regression with Z Scores

One may use the Z scores calculated above in the regression model rather than the original raw scores. The Z scores are reproduced below, and SPSS results follow.

Table 2
Sample Data Converted to Z Scores.

$Z_{Achievement}$	Z_{Time}	$Z_{Ability}$
0.360475	1.062747296	0.127928578
-0.97843	0.391545543	-1.407367888
-2.11135	-1.622059717	-1.407367888
1.493397	1.398348172	1.279400927
1.081425	0.055944666	1.279400927
0.875439	1.062747296	0.511752694
0.051496	0.727146419	-0.255895538
-0.25748	0.055944666	-0.639719655
-0.56646	-1.28645884	-0.255895538
-0.66945	-1.28645884	-1.023543771
0.669454	-0.27965621	0.511752694
0.051496	-0.27965621	1.279400927

3a. SPSS Results

Descriptive Statistics

	Mean	Std. Deviation	N
z_ach	.0000	1.00000	12
z_time	.0000	1.00000	12
z_ability	.0000	1.00000	12

Comment: Note that the mean = 0.00 and sd = 1.00 for each of the three Z scores. This is be design and is expected for Z scores.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.195E-06	.113		.000	1.000
	z_time	.400	.134	.400	2.980	.015
	z_ability	.677	.134	.677	5.050	.001

a Dependent Variable: z_ach

Comment: Note that the unstandardized coefficients are equal to the standardized coefficients in the table above. SPSS automatically calculates Z score coefficients and reports them in the Standardized Coefficient column. Compare the Standardized Coefficients in the above table to the Standardized Coefficients in the regression results reported earlier.

3b. Interpretation of Coefficients with Z Scores

The coefficients for Z scores may be interested as follows:

b_0 = 5.195E-06 = 0.000005195 ≈ 0.000: The predicted value of Achievement (or more precisely $Z_{Achievement}$), in standard deviation units, when Z_{Time} and $Z_{Ability}$ both equal 0.00.

b_1 = 0.40: A 1 standard deviation increase in Z_{Time} is predicted to result in a 0.40 standard deviation increase in $Z_{Achievement}$ holding constant $Z_{Ability}$.

b_2 = 0.677: A 1 standard deviation increase in $Z_{Ability}$ is predicted to result in a 0.677 standard deviation increase in $Z_{Achievement}$ holding constant Z_{Time}.

As the above example shows, conversion of raw scores to Z scores simply changes the unit of measure for interpretation, the change from raw score units to standard deviation units.

4. The Regression Equation: Standardized Coefficients

The above analysis with Z scores produced Standardized Coefficients. Standardized coefficients simply represent regression results with standard scores. By default, most statistical software automatically converts both criterion (DV) and predictors (IVs) to Z scores and calculates the regression equation to produce standardized coefficients.

When most statisticians refer to standardized coefficients, they refer to the equation in which one converts both DV and IVs to Z scores. This, however, is not the only way to obtain standardized coefficients. One may opt, for example, to convert only the IVs to Z scores, or convert only the DV to Z scores. One may also opt to use a formula other than Z to obtain standardized scores. Gelman and Hill (2007, Data analysis using regression and multilevel/hierarchical models) argue that one should divide deviation scores not by one sd as done with Z scores, but instead by 2 sds. Note that converting to Z scores is just one of many ways researchers change the scale, or produce linear transformations, of variables in an attempt to make results more interpretable.

As a rule assume standardized results reported used full standardization (both DV and IVs were converted to standard scores), and that the Z formula was used for standardization. This means the interpretations discussed in these notes will apply. If researchers opted for other forms of standardized, normally this practice will be made explicit.

(Note: For those interested in standardization by dividing by 2 sd, Gelman has a separate article here: http://www.stat.columbia.edu/~gelman/research/published/standardizing7.pdf)

4a. Standardized Regression Equation

The standardized regression equation is:

$Z'_y = \beta_1 Z_{X1} + \beta_2 Z_{X2}$

or

$Z'_y = P_1 Z_{X1} + P_1 Z_{X1}$

where

Z'_y is the predicted value of Y in Z scores;
β_1 and P_1 represent the standardized partial regression coefficient for X_1;
β_2 and P_2 represent the standardized partial regression coefficient for X_2;

and Z_{X1} and Z_{X2} are the Z score values for the variables X_1 and X_2, respectively.

Note the absence of the intercept – the intercept will always equal 0.00 when standardization is based upon Z scores and both DV and IVs are standardized.

Once the regression equation is standardized, then the partial effect of a given X upon Y, or Z_x upon Z_y, becomes somewhat easier to interpret because interpretation is in sd units for all predictors. For the current example, as discussed above, the standardized solution is:

$Z'_y = P_1 Z_{X1} + P_1 Z_{X1}$

$= 0.400(Z_{X1}) + 0.677(Z_{X1})$

The standardized partial coefficient represents the amount of change in Z_y for a *standard deviation* change in Z_x. So, if X_1, time spent studying, were increased by one standard deviation, then one would anticipate a 0.40 standard deviation increase in achievement, holding constant the effect of ability.

4b. Practice Interpretation

Many authors in psychology, sociology, education, political science, and the social sciences in general prefer to report standardized coefficients. Below is linked an example publication in which only standardized coefficients are reported. Open the following article and find Table 2, page 11. Interpret the coefficients presented in that table.

http://www.bwgriffin.com/gsu/courses/edur8132/notes/StandardizedRegression_Sample2.pdf

As a second example, the following link provides interpretation of coefficients presented by Thomas P. Vartanian of Bryn Mawr College:

http://www.bwgriffin.com/gsu/courses/edur8132/notes/StandardizedRegression_Vartanian_Examples.pdf

4c. Standardized Regression Equation—Only for Quantitative IVs, No Qualitative IVs

In most cases statisticians argue that the standardized equation is only appropriate when quantitative, continuous predictors are present. Categorical predictors, such as the use of dummy variables, should not be present in a standardized regression equation. There are exceptions to this convention. Gelman and Hill (2007), for example, offer ways of incorporating and interpreting standardized categorical variables.

4d. Labels

Standardize coefficients are often called beta, beta weights, beta coefficients, or path coefficients in path analysis. As the SPSS results tables above show, SPSS uses two labels: "Standardized Coefficients" and "Beta."

4e. Cautions

Many statisticians argue that standardize coefficients offer no, or little, advantage over unstandardized coefficients, and often offer confusing information. In some disciplines researchers routinely prefer standardize coefficient over unstandardized because they believe standardize coefficients are more interpretable, provide an assessment of predictor importance (i.e., the larger the standardized coefficient in absolute value, the more important the predictor), and are better for comparing across groups and studies. In general these beliefs are incorrect. Standardized coefficients are dependent upon the sample sd, and if that value is inflated or deflated relative to the population sd, then standardized coefficients will provide an incorrect inference for the population value. It is possible, for example, for two groups to have the same unstandardized slope coefficient, yet have different standardized values due to differences in group sds.

However, standardized coefficients may be helpful in learning whether two predictors that have very different scales of measurement appear to have similar statistical effects or predictive power.

Gary King provides a useful discussion of the problem with standardized coefficients in his report "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science" which is linked below. Read the section entitled "The Race of the Variables" beginning on page 669.

http://www.bwgriffin.com/gsu/courses/edur8132/notes/King_Standardized_Coefficients.pdf

4f. Model Fit and Inference, Coefficient Inference

Since standardized coefficients are just linear transformations of the model variables, these don't change the underlying model. As a result, all model fit and inference procedures, and coefficient inference procedures, previously discussed still apply. For example, to perform hypothesis testing upon $B_1(Z_{Time})$, just perform the normal hypothesis test on the unstandardized coefficient—the same t-ratio applies.

5. APA Style

To include standardized coefficients, simply add a column in the regression results table for these coefficients. See the column labeled "β" below.

Table 2. Regression of Achievement on Time Spent Studying and Academic Ability

Variable	b	se	β	ΔR^2	95%CI	t
Time	1.30	0.437	0.400	.124	0.31, 2.29	2.98*
Ability	2.52	0.500	0.677	.356	1.39, 3.65	5.05*
Intercept	63.90	2.836	na	na	57.49, 70.32	22.54*

Note. R^2 = .874, adj. R^2 = .846, F = 31.27*, df = 1,9, MSE = 14.49, n = 12. The symbol ΔR^2 represents the semi-partial correlation squared.
*p < .05.

Material below this point not developed; will not be on Tests in EDUR 8132 until further development.

6. Conversions and 7. Exercises for Conversions

Exercise for standardized and unstandardized change in regression

1. IV is years experience on job (M = 12.3, SD = 5) and DV is salary (M = $40,000, SD = $8,000). Regression results are b_0 = 25,000 and b_1 = 1,000.

(a) What is the predicted salary difference, in dollars, between people with 25 years of experience difference? In SD units, what is the predicted salary difference for these two people?

(b) A three SD difference in years of experience results in how much change in salary in raw units (dollars)? Results in how much change in salary in SD units?

(c) If years of experience declines by 8 years, what change results in salary in both raw units (dollars) and standardized units?

(d) Note that the standardized regression coefficient is not reported. However, it can be calculated using the information reported. Find the value of P_1 using the data above. (Hint --- it is not as difficult as it first appears; in fact, you have already calculated information needed to determine P_1).

2. IVs are number of publications (M = 10, SD = 3), overall evaluation rating of work performance (M = 4, SD = .8), and count of number of committees served (M = 3, SD = 1). The DV is recommendation for merit pay increase, in dollars, for the year (M = $1,500, SD = $250). Regression results, in standardized coefficients, are P_1(publications) = .6, P_2 (evaluation) = 2.2, and P_3(number of committees) = .1.

(a) We wish to compare the difference in merit pay recommendation between two individuals. The first has 7 publications, an evaluation rating of 3.0, and served on 3 committees. The second individual has 10 publications, an evaluation rating of 3.8, and served on 4 committees. In both dollars and SD units, what is the predicted difference in merit pay recommendation between these two?

(b) Decreasing the work performance evaluation for an individual by 3 SDs results in what change in merit pay recommendation (provide change in both dollars and SD units)?

(c) Again, we wish to compare two individuals in terms of merit pay differences. The first individual has 2 SD more publications than the second, has a work evaluation rating that is one SD below the second individual, and has served on the same number of committees. What is predicted difference in merit pay recommendation for the two individuals in both dollars and SD units?

(d) Note that the unstandardized regression coefficients for b_1, b_2, and b_3 are not reported. Using the data provided, calculate the values for these three. (Hint --- this problem is similar to (d) in #1 above, but requires working from standardized to unstandardized. Remember, the definition for a slope, whether it is unstandardized or standardized, is rise/run [recall the scatterplot presented and discussed the first couple of weeks of class]. So, for example, the standardized coefficient for publications is .6, this means for a 1 SD run across the X axis [SD change in publications], we get an increase or rise of .6 SD in merit pay [a .6 rise on the Y axis]. Thus, the formula for rise/run is .6/1.0 = standardized slope of .6 --- use this to solve for the unstandardized coefficient).