Multiple Linear 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, 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 residual term.

Table 1: Achievement, Time Spent Studying, and Academic Ability (plus two categorical variables)

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

Note. Higher scores indicate higher levels of each variable.

1c. SPSS and JASP Results

Least-squares results for the sample data appear below.

Note: Show both SPSS and JASP analyses.

SPSS

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

JASP

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	84.500000000	2.802866498		30.147707733	$6.320083320 \times 10^{-12}$
H_1	(Intercept)	63.901747367	2.835634037		22.535258967	0.000000003
	Time	1.302288639	0.437042931	0.399660343	2.979772800	0.015455280
	Ability	2.524210155	0.499843330	0.677328755	5.050002674	0.000690342

Descriptives

	N	Mean	SD	SE
Achievement	12	84.500000000	9.709414363	2.802866498
Time	12	4.833333333	2.979729498	0.860173814
Ability	12	5.666666667	2.605355789	0.752101433

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 Z scores which have 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 and JASP 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 and JASP 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 by design and is expected for Z scores.

SPSS

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

JASP

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	0.000001000	0.288674910		0.000003464	0.999997298
H_1	(Intercept)	0.000005195	0.113188659		0.000045897	0.999964381
	Ztime	0.399660064	0.134124564	0.399660307	2.979767840	0.015455405
	Zability	0.677329125	0.134124759	0.677328556	5.049993244	0.000690351

Comment: Note that the unstandardized coefficients are equal to the standardized coefficients in the table above. SPSS and JASP 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: This is 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 = 0.399: A 1 standard deviation increase in Z_{Time} is predicted to result in a 0.399 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 which simply represent regression results with standard scores. By default, most statistical software automatically converts both dependent (DV) and independent variables (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. Note that converting to Z scores is just one of many ways researchers change the scale, or produce linear transformations, of variables to make results, hopefully, more interpretable.

As a rule, assume reported standardized results 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.

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 all IVs are standardized. (For examples when the intercept will not equal zero, see section 6 below.)

Once the regression equation is standardized, then the partial effect of a given X upon Y, or Z_x upon Z_y, becomes a focus on change in SD units. For the current example, as discussed above, the standardized solution is:

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

$= 0.399(Z_{X1}) + 0.677(Z_{X1})$

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 0.399 standard deviation increase in achievement, holding constant the effect of ability.

4b. Practice Interpretation and Example Publication

Many authors in psychology, sociology, education, political science, and the social sciences prefer to report standardized coefficients because it provides a common metric for reporting results and because the standardized coefficient can be viewed as an effect size to help judge variable contribution (i.e., the larger β_j or P_j in absolute value, the larger the predictive power of that variable in the regression equation). Objections to this practice are discussed later.

The following link provides example interpretation of coefficients presented by Thomas P. Vartanian of Bryn Mawr College.

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

Linked below is an example publication in which only standardized coefficients are reported. See Table 2, page 11, and Table 3, page 12.

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

Another example using path analysis to model student ratings of instruction. See Figure 1 page 401.

https://web.archive.org/web/20120314020220id_/http://www.charlesdarwinresearch.org/1985%20(Erdle,%20JEP).pdf

4c. Standardized Regression Equation: Only for Quantitative IVs (well, not really, see section 6)

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. Discussion of how to incorporate categorical variables via dummy variables is presented in section 6.

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." JASP only used the Standardized label.

4e. Cautions

Many statisticians argue that standardize coefficients offer no, or little, advantage over unstandardized coefficients, and often offer confusing information. For example, consider a regression model with salary in dollars as the dependent variable with the following predictors: number of hours worked, years of experience, years employed at current employer.

Which is easier to understand?

Standardized:

- For each 1 SD increase in hours worked, salary increases by 0.23 SDs.
- For each 1 SD increase in years of experience, salary increases by 0.11 SDs.
- For each 1 SD increase in years of employed at current employer, salary increases by 0.13 SDs.

Unstandardized:

- For each 1 additional hour worked, salary increases by $25.
- For each 1 additional year of experience, salary increases by $750.
- For each 1 additional year employed at current employer, salary increases $1500.

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.

These beliefs are not uniformly correct because 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 estimate of 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.

In some cases, however, standardized coefficients may be helpful in determining the relative contribution or predictive power of variables measured on widely different scales of measurement. For example, both the SAT and ACT are used to predict first-year college GPA. The SAT is scaled from 400 to 1600 while the ACT is scaled from 1 to 36. Even if both tests provided the same predictive power, the unstandardized coefficient for the SAT would be much smaller than the coefficient for the ACT due to the scale difference. However, once standardized, it is possible both tests would have the same, or a very close, standardized estimate.

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. If interested, 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, model fit statistics (R^2, adjusted R^2, MSE, SEE) remain the same because the linear transformation (i.e., converting to Z scores) don't change the underlying model. Similarly, model and coefficient inferential tests also provide the same results. 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.

4g. Regression with One Predictor: β_1 = Pearson r

The standardized coefficient, β_1, equals the Pearson correlation, r, for a regression equation with only one predictor. For example, the correlation between Achievement and Time is 0.7195 as shown below.

Pearson's Correlations

Variable		Achievement	Time	Ability
1. Achievement	Pearson's r	—		
	p-value	—		
2. Time	Pearson's r	0.719569789	—	
	p-value	0.008333839	—	
3. Ability	Pearson's r	0.866092510	0.472310446	—
	p-value	0.000269651	0.121035259	—

The simple regression of Achievement on Time is shown below. Note the unstandardized coefficient for Time is 2.344, but the standardized coefficient is 0.7195 which is the same value as the correlation between Achievement and Time.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	84.500000000	2.802866498		30.147707733	$6.320083320\times10^{-12}$
H_1	(Intercept)	73.167235495	4.016000654		18.218930176	0.000000005
	Time	2.344709898	0.715547548	0.719569789	3.276805160	0.008333839

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 squared semi-partial correlation.
*$p < .05$.

6. Supplemental Topic: Computing Standardized Values for Dummy Variables

Most authors write that categorical variables are difficult to use, or should not be used, when standardized regression equations are the focus. Some have also argued for changes to the standardization practice to make the use of categorical variables, specifically dummy variables, more interpretable. For example, Gelman and Hill (2007) argue that one should divide deviation scores not by one SD as done with Z scores, but instead by 2 SDs. Gelman (2008) provides a more detailed discussion of this suggestion in the article linked below.

http://www.stat.columbia.edu/~gelman/research/published/standardizing7.pdf

Correct standardized estimates can be obtained, however, with dummy variables, but to obtain these estimates requires that one first convert all quantitative variables to Z scores and leave the dummy variables unchanged (i.e., maintain the 0 and 1 coding). Next, one then enters the standardized variables and the original dummy variables into the regression and uses to **unstandardized** estimates as the corrected **standardized** coefficients for the dummy variables.

Below is an illustrated example showing how to obtain a correct standardized solution with dummy variables. The description below is long due to extra steps to help explain the results obtained, but the process is simple:

- convert all IVs and DVs that are quantitative/scale variables to Z scores, but leave all dummy/indicator variables with 0,1 coding – do not convert dummy variables to Z scores;
- enter the Z scored variables (the IV and DV except for dummy variables) and the dummy variables (with 0,1 coding) into the regression analysis; and
- use the Unstandardized coefficients column of software results for the correct standardized values.

(1) Initial Regression Analysis

Run the regression analysis to obtain the unstandardized estimates so those can be examined and reported in the resulting output (such as in an APA styled table with both unstandardized and standardized estimates). Below is the regression of Achievement on Time, Ability, and group status with dummy variables for Groups A and B included in the equation and Group C used as the reference category.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	84.500000000	2.802866498		30.147707733	$6.320083320\times10^{-12}$
H_1	(Intercept)	66.310750820	3.402640418		19.488027728	0.000000234
	Time	1.756128161	0.619159070	0.538939496	2.836311773	0.025177335
	Ability	2.091488130	0.596471496	0.561215178	3.506434329	0.009906818
	Group A	−4.843056360	3.701241168	−0.245592167	−1.308495216	0.232027777
	Group B	−1.608352635	3.479631730	−0.081559821	−0.462219212	0.657947044

Descriptives

	N	Mean	SD	SE
Achievement	12	84.500000000	9.709414363	2.802866498
Time	12	4.833333333	2.979729498	0.860173814
Ability	12	5.666666667	2.605355789	0.752101433
Group A	12	0.333333333	0.492365964	0.142133811
Group B	12	0.333333333	0.492365964	0.142133811

Incorrect values for dummy variables

The JASP output above provides the correct estimates for the unstandardized coefficients, but the standardized coefficients for the dummy variables are incorrect. They are incorrect because the dummy variables were converted to Z scores by the software and that should not occur; indicator (dummy) variables should not be standardized. These are nominal variables and the SD of these dummy codes, 0 and 1, is nonsense and should not be used to convert unstandardized coefficients to standardized coefficients.

(2) Initial Estimates of Standardized Differences

(2a) Cohen d for Group A vs. C and Group B vs. C Comparisons in Achievement

While this step is not necessary, it is helpful in this illustration to calculate the correct standardized estimates to use as a confirmation value in this process. To calculate the standardized mean difference, effect formula for standardized mean differences, or d, or Cohen's d, will be used. (Example source for more information: https://en.wikipedia.org/wiki/Effect_size.)

The general formula for effect size d is a mean difference divided by the SD of the dependent variable. The SD for Achievement is shown in the JASP output above and is 9.70941.

The unstandardized coefficients for the dummy variables provide the mean difference for Group A vs. Group C, and Group B vs. Group C. These are reported below in the second column. The conversion to standardized estimates is shown in column three, and column four shows the result.

Dummy Variable Comparison	Unstandardized Mean Difference	Cohen's d (standardized mean difference) = mean difference / SD of Achievement	d (standardized mean difference)
Group A vs. Group C	$b_3 = -4.84305$	$\beta_3 = -4.84305 / 9.70941$	-0.4988
Group B vs. Group C	$b_4 = -1.60835$	$\beta_4 = -1.60835 / 9.70941$	-0.1656

The table below shows the calculated standardized estimate obtained using the d formula and the result from JASP (and SPSS) using the default method for obtaining standardized estimates.

Dummy Variable Comparison	d (standardized mean difference)	Incorrect standardized estimate from software
Group A vs. Group C	-0.4988	-0.2455
Group B vs. Group C	-0.1656	-0.0815

Note that that standardized estimates given by JASP (and other software) are incorrect with values of -0.2455 (vs. correct estimate -0.4988) for Group A and for Group B the value is -0.0815 (vs. correct estimate -0.1656). Again, these estimates are incorrect because, by default, statistical software converts the dummy variable into a Z score which means the values of the dummy are no longer 0 and 1. Dummy variables should not be converted to Z scores because they are not quantitative variables with meaningful standard deviations.

(2b) Cohen's d for Group C vs Overall Mean in Achievement

Before discussing how to incorporate dummy variables into a standardized regression equation, the predicted mean for Group C is calculated because it will help with interpretation of the standardized equation shown below. Using the regression equation Group C's adjusted mean can be calculated as follows:

Group C Mean = 66.3107 + 1.7561*(Time Mean) + 2.0914*(Ability Mean) + -4.843(A dummy) + -1.6083(B dummy)
Group C Mean = 66.3107 + 1.7561*(4.8333) + 2.0914*(5.6666) + -4.843(0) + -1.6083(0)
Group C Mean = 66.3107 + 1.7561*(4.8333) + 2.0914*(5.6666)
Group C Mean = 86.6496

The overall mean for Achievement is 84.50 with a SD = 9.70941 (see JASP output above), so the Group C is 0.221 standard deviations above the overall mean as shown by the calculation below.

d effect size for Group C = (86.6496 – 84.50) / 9.70941 = 0.221

(3) Create Z scores for all Scale Variables (all non-categorical variables)

To obtain correct standardized estimates with dummy variables, one must first standardize all quantitative variables by converting them to Z scores, then enter those into a regression equation with dummy variables which are NOT standardized (i.e., convert all quantitative, scale variables to Z scores, but leave dummy variables as is with 0 and 1 coding).

With this approach one can obtain the correct standardized regression estimates. See the output below in which Achievement, Time, and Ability were converted to Z scores and entered into the regression analysis with the dummy variables for Groups A and B. The Z scores provided for Achievement, Time, and Ability provided in Table 2 above were used for this regression analysis.

The correct standardized results are now shown in the **Unstandardized** column, not the **Standardized** column. For the scale variables, Time and Ability, the estimates are the same for both unstandardized and standardized, but only the unstandardized column provides the correct estimates for the dummy variables because these dummy variables have not been converted to Z scores. Note that the estimates for the dummy variables for Group A (-0.49) and B (-0.16) now match the calculated effect size d provided above in step 2a.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	0.000001000	0.288674910		0.000003464	0.999997298
H_1	(Intercept)	0.221484751	0.253791489		0.872703618	0.411757794
	Ztime	0.538939023	0.190014518	0.538939351	2.836304446	0.025177599
	Zability	0.561215545	0.160053460	0.561215074	3.506425568	0.009906935
	Group A	−0.498799251	0.381201865	−0.245591965	−1.308491111	0.232029094
	Group B	−0.165648547	0.358377648	−0.081559770	−0.462217853	0.657947971

Corrected Standardized Estimates

Notice that the intercept in the Unstandardized column, which is now the correct standardized output, is no longer zero in this regression equation. The intercept value 0.2214 represents the standardized difference between Group C, the reference category, and the overall mean for Achievement. This same value was calculated above in step 2b.

Instructor's note.
It seems differences in sample sizes for subgroups affect accuracy of the Standardized solution. In this case each effect size is half the size it should be, i.e., -.24559*2 = -.4911 (~-.4987) and -0.08155*2 = -0.1631 (~.1656). Experiment with different group sizes to identify relation, if any, that exists between standardized estimate and correct value in SD terms.
Update – seems group size does not affect estimates. Using the sex variable the standardized estimate is incorrect whether sex has equal or unequal category sizes, and whether Group is included or excluded from the analysis.

Sex = 2.0906 / 9.7094 = 0.2153 (without other dummy variables), JASP β = .1108
Sex = -6.6616 / 9.7094 = -0.6861 (with other dummy variables), JASP β = -.3532

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

7. Conversions and 8. 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).

References

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine*, *27*(15), 2865-2873.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, 666-687.