

# Regression in SPSS

## Background

Regression measures the extent to which one or more predictors (X variables) relate to an outcome variable (the Y variable).

## Preparing Your Data

Before running correlations, you'll need to form composites for your continuous predictors (Xs) and outcome (Y). For more details, see the "Composites, Descriptive Statistics, and Histograms in SPSS" handout. You'll also need to dummy code any categorical/nominal predictors. For more details, see the "Dummy Coding" handout.

## Unstandardized Regression (*b*)

The syntax for running regression is

```
mixed y with x1 [x2 ...]
/fix ed=x1 [x2 ...]
/print=solution.
```

For example, imagine that we want to predict life satisfaction (called "ls" in our dataset) from extraversion (called "e" in our dataset). The regression would be:

```
mixed ls with e
/fix ed=e
/print=solution.
```

## Reading the Output

When running a regression, you'll see several tables output. One of them is called "Estimates of Fixed Effects:"

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.989903	.107652	1485	18.485	.000	1.778737	2.201069
e	.422894	.032946	1485	12.836	.000	.358269	.487519

a. Dependent Variable: ls.

In this table, the "Estimate" column contains the regression coefficients. The "95% Confidence Interval" column contains the 95% CI for the coefficient. Remember that we always report regression coefficients and CIs to two decimal places (and include leading zeros):

Extraversion predicted life satisfaction,  $b = 0.42$ , 95% CI [0.36, 0.49].

## Interpreting Unstandardized Coefficients

The interpretation of a regression coefficient is the “increase in Y per unit increase in X.” So, in the example data above, both extraversion and life satisfaction were rated on a scale from 1 (low) to 5 (high). The coefficient ( $b = 0.42$ ) means that, for each one unit increase in extraversion, a person would likely report 0.42 units higher life satisfaction. For example, a person who is 3/5 in extraversion would be expected to rate their life satisfaction as 3.26/5. In contrast, a person who is 4/5 in extraversion would be expected to rate their life satisfaction 3.68/5 (0.42 points higher than 3.26).

## Standardized Regression ( $\beta$ )

If you standardize all variables before including them in a regression, your regression coefficients are called “standardized regression coefficients” and you report them as a  $\beta$  (beta) instead of a  $b$ .

For example, we can standardize both life satisfaction and extraversion using the following syntax:

```
descriptives ls e /save.
```

This will save people’s standardized life satisfaction and extraversion scores as “zls” and “ze” in our dataset. We can then run a regression using the standardized variables:

```
mixed zls with ze  
/fixed=ze  
/print=solution.
```

Here is sample output:

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.000000	.024612	1485	.000	1.000	-.048278	.048278
Ze	.316027	.024620	1485.000	12.836	.000	.267733	.364321

a. Dependent Variable: Zscore(ls).

Because **both** X (extraversion) and Y (life satisfaction) are standardized, we report our regression coefficients using  $\beta$ s:

Extraversion predicted life satisfaction,  $\beta = 0.32$ , 95% CI [0.27, 0.36].

If and only if there is exactly one predictor (X variable) in the regression,  $\beta$  is the same as the correlation (i.e.,  $\beta = r$ ). This fact means that you can use regression to get confidence intervals for correlations if you’d like!

Extraversion predicted life satisfaction,  $r = .32$ , 95% CI [.27, .36].

## Interpreting standardized coefficients

Even if both of your variables are standardized, the interpretation of a regression coefficient is the same: the increase in Y per unit increase in X. The only difference is that when both variables are standardized, a “unit increase” refers to the standard deviations. Thus,  $\beta = 0.32$  could be interpreted as, “for each SD increase in extraversion, life satisfaction tends to increase 0.32 SDs.” For example, a person who is exactly average in extraversion ( $z = 0$ ) would be expected to be exactly average in life satisfaction, as well ( $z = 0$ ). In contrast, a person 1 SD above the mean in extraversion would be expected to report 0.32 SDs above the mean in life satisfaction.

## How do I know whether to standardize my variables or not?

In general, you should *always* standardize variables that have an arbitrary Likert-type scale (e.g., “rate your extraversion on a scale from 1-5”). In such measures, the numbers are arbitrary, and thus a “unit increase” on the original scale is less meaningful than “a standard deviation increase.”

In general, you should *not* standardize variables that have a meaningful numeric scale (e.g., income in dollars, number of sick days last month, time in minutes). The only exception is that it’s oftentimes wise to report  $\beta$ s (betas) in addition to  $b$ s if you don’t standardize all variables, e.g.:

People with more Facebook friends reported higher happiness,  $b = 0.03$ , 95% CI [0.01, 0.05],  $\beta = 0.23$ .

## Categorical Predictors

If you have dummy-coded categorical predictors (i.e., variables that are either 0 or 1 and represent groups; for more information, see the handout on Dummy Coding), you can run three slightly different types of regression.

## Unstandardized Regression ( $b$ )

If your outcome and dummy-coded predictor are both unstandardized, your regression will compute the raw mean difference between the groups on the original scale.

```
mixed avoidance with male
/fixed=male
/print=solution.
```

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.661881	.021273	1804	125.132	.000	2.620160	2.703603
male	-.194662	.042428	1804.000	-4.588	.000	-.277875	-.111449

a. Dependent Variable: avoidance.

On the original 1 to 5 scale, men were 0.19 points lower in attachment avoidance than were women,  $b = -0.19$ , 95% CI [-0.28, -0.11].

### Standardized Outcome Only ( $d$ )

If you standardize your outcome variable, but keep your predictor dummy-coded, your regression will compute the standardized mean difference between groups (i.e., how many standard deviations the groups differ by).

```
mixed zavoidance with male  
/fixed=male  
/print=solution.
```

**Estimates of Fixed Effects<sup>a</sup>**

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.048502	.026677	1804	1.818	.069	-.003819	.100824
male	-.244117	.053207	1804	-4.588	.000	-.348470	-.139763

a. Dependent Variable: Zscore(avoidance).

Men were 0.24 *SDs* lower in attachment avoidance than were women,  $d = -0.24$ , 95% CI [-0.35, -0.13].

### Standardized Regression ( $\beta$ )

If you standardize all variables, you will obtain the standardized regression coefficients:

```
mixed zavoidance with zmale  
/fixed=zmale  
/print=solution.
```

**Estimates of Fixed Effects<sup>a</sup>**

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.012966	.023082	1804	-.562	.574	-.058235	.032304
Zmale	-.105987	.023101	1804	-4.588	.000	-.151294	-.060680

a. Dependent Variable: Zscore(avoidance).

Men were lower than women in attachment avoidance,  $\beta = -0.11$ , 95% CI [-0.15, -0.06].

If *and only if* there is exactly one predictor (X variable) in the regression,  $\beta$  is the same as the correlation (i.e.,  $\beta = r$ ).

Men were lower than women in attachment avoidance,  $r = -.11$ , 95% CI [-.15, -.06].

# What Symbol Should I Use When Reporting My Regression Coefficients?

