

14.3: Standardized Regression Coefficients

In most cases, the various IVs in a model are represented on different measurement scales. For example, ideology ranges from 1 to 7, while age ranges from 18 to over 90 years old. These different scales make comparing the effects of the various IVs difficult. If we want to directly compare the magnitudes of the effects of ideology and age on levels of environmental concern, we would need to **standardize** the variables.

One way to standardized variables is to create a ZZ-score based on each variable. Variables are standardized in this way as follows:

$$Z_i = \frac{X_i - \bar{X}}{s_x} \quad (14.1)$$

where s_x is the s.d. of X . Standardizing the variables by creating ZZ-scores re-scales them so that each variables has a mean of 00 and a s.d. of 11. Therefore, all variables have the same mean and s.d. It is important to realize (and it is somewhat counter-intuitive) that the standardized variables retain all of the variation that was in the original measure.

A second way to standardize variables converts the unstandardized B , into a standardized B' .

$$B'_k = \frac{B_k s_y}{s_x} \quad (14.2)$$

where B_k is the unstandardized coefficient of X_k , s_x is the s.d. of X_k , and s_y is the s.d. of Y . Standardized regression coefficients, also known as beta weights or “betas”, are those we would get if we regress a standardized Y onto standardized X 's.

Interpreting Standardized Betas

- The standard deviation change in Y for a one-standard deviation change in X
- All X 's on an equal footing, so one can compare the strength of the effects of the X 's
- *Cannot be used for comparisons across samples*
- Variances will differ across different samples

We can use the `scale` function in `R` to calculate a ZZ score for each of our variables, and then re-run our model.

```
stan.ds <- ds.temp %>%
  dplyr::select(glbcc_risk, age, education, income, ideol, gender) %>%
  scale %>%
  data.frame()

ols3 <- lm(glbcc_risk ~ age + education + income + ideol + gender, data = stan.ds)
summary(ols3)
```

```
##
## Call:
## lm(formula = glbcc_risk ~ age + education + income + ideol +
##     gender, data = stan.ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.92180 -0.54357  0.06509  0.48646  2.20164
##
## Coefficients:
##              Estimate      Std. Error t value
## (Intercept)  0.0000000000000001685  0.0167531785616065292    0.000
## age         -0.0187675384877126518  0.0169621356203379960   -1.106
## education    0.0395657731919867237  0.0178239180606745221    2.220
## income      -0.0466922668201090602  0.0178816880127353542   -2.611
## ideol       -0.5882792369403809785  0.0170882328807871603  -34.426
## gender      -0.0359158695199312886  0.0170016561132237121   -2.112
##
##              Pr(>|t|)
## (Intercept)    1.00000
## age            0.26865
## education      0.02653 *
## income         0.00908 **
## ideol          < 0.0000000000000002 ***
## gender         0.03475 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7984 on 2265 degrees of freedom
## Multiple R-squared:  0.364, Adjusted R-squared:  0.3626
## F-statistic: 259.3 on 5 and 2265 DF, p-value: < 0.00000000000000022
```

In addition, we can convert the original unstandardized coefficient for ideology, to a standardized coefficient.

```
sdX <- sd(ds.temp$ideol, na.rm=TRUE)
sdY <- sd(ds.temp$glbcc_risk, na.rm=TRUE)
ideology.prime <- ols1$coef[5]*(sdX/sdY)
ideology.prime
```

```
##      ideol
## -0.5882792
```

Using either approach, standardized coefficients allow us to compare the magnitudes of the effects of each of the IVs on YY.

This page titled [14.3: Standardized Regression Coefficients](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Jenkins-Smith et al. \(University of Oklahoma Libraries\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.