Interpreting the Intercept in Linear Regression Models with Categorical Predictors

Kyle M. Lang

2021-03-24

In this document, I'll go through the interpretation of the estimated intercept in multiple linear regression models that contain categorical predictors. The interpretation of the intercept differs based on the nature of the variables on the right hand side (RHS) of the regression equation. For simple linear regression models, the following represent the intercept interpretations for some common types of predictor variables.

- Continuous X: The model-implied mean of Y (i.e., the expected value for Y, given our model) for cases with $X = 0$.
- Dummy-coded X: The model-implied mean of Y for cases in the reference group of X.
- Unweighted effects-coded X: The model-implied, unweighted mean of Y (see below for definition of the unweighted mean)
- Weighted effects-coded X: The model-implied, weighted mean of Y (see below for definition of the weighted mean)

When we combine different types of variables and/or codes on the RHS of a multiple linear regression model, the intercept is interpreted as a sort of "intersection" of the interpretations for the separate variable types/codes.

- Dummy-coded X_1 and continuous X_2: The model-implied mean of Y for cases in the reference group of X_1 with $X_2 = 0$.
- Unweighted effects-coded X_1 and continuous X_2: The model-implied, unweighted mean of Y for cases with $X_2 = 0$.
- Weighted effects-coded X_1 and continuous X_2: The model-implied, weighted mean of Y for cases with $X_2 = 0$.
- Dummy-coded X_1 and unweighted effects-coded X_2: The model-implied, unweighted mean of Y for cases in the reference group of X_1.
- Dummy-coded X_1 and weighted effects-coded X_2: The model-implied, weighted mean of Y for cases in the reference group of X_1.

- Weighted effects-coded X_1 and unweighted effects-coded X_2: The model-implied mean of Y calculated with proportional weighting across the groups of the X_1 factor and without weighting for the groups of the X_2 factor.

Note that the final combination (e.g., both weighted and unweighted effects codes) should not come up much, in practice. Usually, your data characteristics or your inferential goals will motivate the use of either all weighted effects codes or all unweighted effects codes. In the following section, I'll show a few examples of the combined interpretations described above.

1 Example Interpretations

We'll anchor our interpretations by refitting the same model with different coding schemes for the predictor variables. In our example model, we'll regress BMI onto a two-level factor representing biological sex and a three-level factor representing educational attainment level.

$$Y_{bmi} = \beta_0 + \beta_1 X_{female} + \beta_2 X_{ed.mid} + \beta_2 X_{ed.hi} + \varepsilon$$

We will use the *BMI* data from the **wec** package (Te Grotenhuis et al., 2016) to fit this model using different flavors of categorical coding for the *sex* and *education* factors.

1.1 Two Dummy Codes

When coding each factor using dummy codes, we get the following fitted regression equation.

$$\hat{Y}_{bmi} = 26.4 - 0.5X_{female} - 1.14X_{ed.mid} - 1.86X_{ed.hi}$$

The male group is the reference group for the *sex* factor, and the lowest educational attainment group is the reference group for the *education* factor. So, the intercept, $\hat{\beta}_0 = 26.4$, represents the model-implied mean of BMI (i.e., the expected BMI given our fitted model) for males in the lowest educational attainment group.

1.2 Two Unweighted Effects Codes

When coding each factor using unweighted effects codes, we get the following fitted regression equation.

$$\hat{Y}_{bmi} = 25.15 + 0.25X_{female} + 1X_{ed.mid} - 0.14X_{ed.hi}$$

In this case, we don't need to worry about keeping track of the reference groups because the intercept represents the (unweighted) grand mean of BMI. Since both factors are coded with unweighted effects codes, the model-implied, unweighted mean of BMI is $\hat{\beta}_0 = 25.15$.

1.2.1 What is the Unweighted Mean?

For this model, the unweighted mean is defined as the simple average of the group-specific means of BMI. Since the *sex* factor has two levels and the *education* factor has three levels, we have $2 \times 3 = 6$ group-specific means.

	highest	middle	lowest
female	26.07	25.29	24.71
male	26.22	24.74	23.89

Table 1: Group-Specific Means of BMI

Averaging these six group-specific means produces the unweighted mean of BMI ($\overline{BMI} = 25.15$). We call this quantity the *unweighted* mean because we compute it as the simple average of the group-specific means (i.e., we don't weight the means by their relative sample sizes).

1.3 Two Weighted Effects Codes

When coding each factor using weighted effects codes, we get the following fitted regression equation.

$$\hat{Y}_{bmi} = 24.98 - 0.23X_{female} + 0.02X_{ed.mid} - 0.7X_{ed.hi}$$

Again, we don't need to concern ourselves with the reference groups because the intercept represents the (weighted) grand mean of BMI. Since both factors are coded with weighted effects codes, the model-implied, weighted mean of BMI is $\hat{\beta}_0 = 24.98$.

1.3.1 What is the Weighted Mean?

In this case, the weighted mean is defined as the weighted average of the group-specific means of BMI. We weight the group-specific means shown in Table 1 by weights representing the proportion of the total sample size associated with each group.

	lowest	middle	highest
male	0.103	0.187	0.181
female	0.107	0.241	0.180

Table 2: Group-Specific Sample Weights

To compute the weighted mean, assume we have two matrices such as those shown in Tables 1 and 2:

- An $L \times K$ matrix of group-specific means of Y, M, with elements m_{lk}

- The corresponding matrix of group-specific sample weights, W, with elements w_{lk}

We can then compute the weighted mean of Y via the following formula:

$$\bar{Y} = \sum_{l=1}^{L} \sum_{k=1}^{K} w_{lk} m_{lk}$$

For our example data, the estimated weighted mean is $\overline{BMI} = 24.98$. When computing this quantity from the raw data, the weighted mean of Y will equal the arithmetic average of Y, $\bar{Y} = N^{-1} \sum_{n=1}^{N} y_n$.

1.4 Dummy Code and Unweighted Effects Code

When we code *sex* with a dummy code and *education* with unweighted effects codes, we get the following fitted regression equation.

$$\hat{Y}_{bmi} = 25.4 - 0.5X_{female} + 1X_{ed.mid} - 0.14X_{ed.hi}$$

The reference groups for *sex* is still males. So, the intercept, $\hat{\beta}_0 = 25.4$, represents the model-implied, unweighted mean of BMI for males. To compute the analogous marginal mean from the raw data, we would simply average the three group-specific means in the first row of Table 1.

1.5 Dummy Code and Weighted Effects Code

When coding *sex* with a dummy code and *education* with weighted effects codes, we get the following fitted regression equation.

$$\hat{Y}_{bmi} = 25.24 - 0.5X_{female} + 0.02X_{ed.mid} - 0.7X_{ed.hi}$$

Here, the intercept, $\hat{\beta}_0 = 25.24$, represents the model-implied, weighted mean of BMI for males. To compute the analogous marginal mean from the raw data, we would weight the three group-specific means in the first row of Table 1 by the corresponding sample weights in the first row of Table 2 and sum these products.

2 Estimated Intercepts $\neq$ Data-Derived Means?

Often, the estimated intercept from a given model does not match the corresponding marginal mean that we would calculate from the raw data. We can see an example of this discordance if we review the example with two dummy-coded predictors from above. When we use dummy codes for both the *sex* and *education* factors, we get an estimated intercept of $\hat{\beta}_0 = 26.402$, but the corresponding marginal mean from Table 1 is $\overline{BMI}_{male,ed.low} = 26.074 \neq \hat{\beta}_0$.

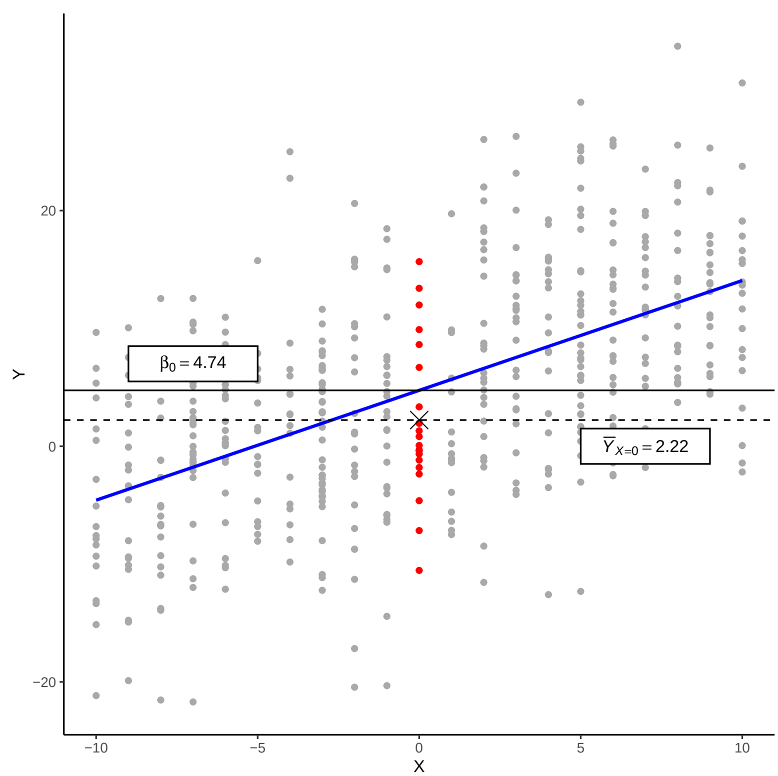
Although this discordance may be disturbing, we don't really need to worry about it. Remember that, for this example, the intercept is the *model-implied* mean of males in the lowest educational group. Our model does not perfectly reproduce our data, though, so there can be discrepancies. If we estimate the appropriate marginal means using our fitted model (e.g., by using the **emmeans** package, Lenth, 2021), we see that the relevant marginal mean estimated from our fitted model matches the estimated intercept.

	lowest	middle	highest
male	26.40	25.26	24.55
female	25.90	24.76	24.05

Table 3: Marginal Means Estimated from the Model

2.1 How is this Possible?

To get some intuition for how the discordance between the model-based estimates and the data-derived quantities can occur, consider the following figure.



The above figure shows a scatterplot of Y against X with the fitted regression line overlaid. Notice that the estimated intercept is $\hat{\beta}_0 = 4.74$. If, however, we average all of the Y values for which $X = 0$ (i.e., the red points in the figure), we get an estimated mean of $Y_{X=0} = 2.22 \neq \hat{\beta}_0$. Although these two quantities share the same interpretation, they need not match in practice. Regardless of what types of predictors are included, the model from which the intercept is derived is only approximating the ideal relationship between X and Y. So, the estimated intercept may not match the analogous quantity estimated directly from the raw data.

References

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=emmeans` (R package version 1.5.5-1)

Te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2016). A novel method for modelling interaction between categorical variables. *International Journal of Public Health*, 1-5. Retrieved from `http://doi.org/10.1007/s00038-016-0902-0`