

Single predictor - binary prediction model

If there is a binary predictor model, the regression coefficient is the difference between the averages of the two groups. Imagine a simple model:

$$y = \beta_0 + \beta_1 x, \quad (1)$$

where x is a binary variable. Suppose, we could plot two populations, and placed them next to each other, with one group on the left (i.e. $x = 0$) and another on the right ($x = 1$). Once we have calculated the regression line, it goes through the averages of both groups.

The coefficient on x is equal to:

$$\beta_1 = (\beta_0 + \beta_1 \cdot 1) - (\beta_0 + \beta_1 \cdot 0), \quad (2)$$

which is the difference in averages of two sub groups. In other words, depending on the sign, we could make the following statements

1. $\beta_1 > 0$ the subgroups for which $x = 1$ will on average have β_1 more, when it comes to the value of y
2. $\beta_1 < 0$ the subgroups for which $x = 1$ will on average have β_1 less, when it comes to the value of y

Multiple predictors

To say that we compare once coefficient difference, and hold other coefficient values constant is the following; say there is a model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \quad (3)$$

and β_1 is a binary variable. Therefore, the following holds:

1. β_0 - when analysing the intercept, we try to keep the other β 's at 0 value (same for the binary value). This is not always meaningful, as for some variables (e.g. age or IQ score) this might not be meaningful enough.
2. β_1 - we analyse the difference between two subgroups while holding the other values constant. Therefore, the average difference in the final score is equal to β_1 for the reference group *i.e.* $x_2 = 0$.
3. β_2 - the difference of a unit change is equal to β_2 , when holding the other predictors constant.
4. β_3 - the difference of a unit change is equal to β_3 , when holding the other predictors constant.

Least square estimate

For the model:

$$y = X\beta + \epsilon \quad (4)$$

we search for a $\hat{\beta}$ that minimizes the sum of squared errors:

$$\sum_{i=1}^n (y_i - X_i\hat{\beta})^2 \quad (5)$$

The estimate is also maximum likelihood estimate when i.i.d errors ϵ_i are normally distributed with equal variance.

Uncertainty of estimates

The standard estimation errors of $\hat{\beta}$ denote the uncertainty of estimation. In general, coefficient estimates within two standard errors are consistent with the data. The uncertainty for difference coefficient is correlated except for the studies with balanced designs.

Residuals

The differences between the actual and estimated values are called residuals and calculated in the following form:

$$r_i = y_i - X_i\hat{\beta} \quad (6)$$

The residuals are uncorrelated with all the predictors in the model. Say, for a model with a constant, the residuals are uncorrelated with it, and therefore have a mean of 0. This is not the assumption of regression theory.

Residual standard deviation

The residual standard deviation is calculated in the following way:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n r_i^2}{(n - k)}} \quad (7)$$

Therefore, $\hat{\sigma}$ denotes the accuracy of \hat{y}_i estimates. Also, here are the key points on the standard errors:

1. Standard errors for β are proportional to σ .
2. $\hat{\sigma}$ - the smaller the residual variance the better the fit.
3. $\hat{\sigma}^2$ (the residual standard deviation) denotes how much variance is not explained
4. $R^2 = 1 - \frac{\hat{\sigma}^2}{s_y^2}$ how much variance is explained by the model. s_y denotes standard deviation of the model.

5. The estimated variance $\hat{\sigma}^2$ has a sampling distribution centred at the true value of σ^2 and proportional to χ^2 distribution with $n - k$ degrees of freedom.

Things to be added

There is a section on page 42 discussing how model with less data generates higher value for R^2 . The reason being is that residual standard deviation is the same, but standard deviation of the data S_y^2 shrinks making R^2 larger. This could be supported with an example / image.