

Multiple Regression

Conditional coefficients, model comparison, and shared predictor information

Ratiomera Statistics

Table of contents

Course-Page Summary	2
One Equation, Several Roles	2
Tests and Model Comparison	2
Interactions and Diagnostics	3
Expanded Reference	4
Purpose and foundations	4
Core ideas	4
Formula guide	5
Reading the explanatory figure	8
Interpretation checklist	8
How this topic connects	9

Course-Page Summary

Multiple regression models one quantitative outcome from several predictor terms at the same time. Each coefficient is conditional: it describes a fitted change associated with one predictor while the other terms in the equation are held fixed.

One Equation, Several Roles

For two quantitative predictors,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i.$$

Table 1: The model combines predictor contributions, but each term retains a specific conditional meaning.

Component	Interpretation
Intercept	Fitted mean when every quantitative predictor equals zero and every categorical predictor is at its reference level
Quantitative coefficient	Fitted outcome change for one additional predictor unit, conditional on the other terms
Dummy coefficient	Conditional difference from the named reference category at the quantitative reference values
Interaction coefficient	Change in one slope or group difference as another variable changes
Fitted value and residual	Combined model prediction and observed-minus-fitted difference for one case

Centering a quantitative predictor subtracts a meaningful reference such as its sample mean. This changes the intercept and any lower-order terms involved in an interaction, but does not change fitted values or overall fit. Centering improves interpretation; it does not remove predictor overlap.

Tests and Model Comparison

- An individual coefficient t test asks whether one population coefficient is zero conditional on the other included terms.
- The global F test asks whether all tested slopes are zero together.
- A nested-model F test asks whether a specified block of added terms improves fit beyond a smaller model.
- R^2 cannot decrease when predictors are added. Adjusted R^2 includes a complexity penalty and can decrease.
- AIC compares the relative fit-complexity balance among the stated candidate models. It does not identify a true or causal model.

Predictor overlap can make individual estimates less precise. A variance inflation factor summarizes how strongly one predictor is linearly explained by the other predictors. It is a diagnostic for coefficient uncertainty, not a universal pass-fail rule.

Interactions and Diagnostics

When an interaction is present, interpret conditional slopes or fitted values at meaningful combinations. Do not interpret a main-effect coefficient as one universal effect across all levels of the interacting variable.

Use residual-versus-fitted and normal Q-Q plots to assess modeled form, spread, and broad error shape. Leverage describes unusual predictor combinations. Cook's distance combines predictor position and residual information to prioritize sensitivity checks. Valid cases are not deleted merely because a diagnostic is large.

The progression from Topics 4 through 7 is one connected story: correlation describes paired movement, simple regression gives that movement a direction and units, partial correlation makes adjustment visible, and multiple regression places several conditional relationships in one fitted equation. Topic 8 shows that categorical group comparisons belong to the same general linear-model family.

Expanded Reference

Purpose and foundations

Multiple linear regression models the conditional mean of one numerical outcome using two or more predictors. It extends the fitted-line idea from simple regression. Instead of moving along one predictor axis, the fitted outcome can change across several predictor dimensions. Each coefficient describes the fitted change connected with one predictor while the other predictors in the model are held fixed.

That final phrase is a comparison rule, not a physical action. Consider a model predicting reasoning score from practice hours and prior score. The practice coefficient compares cases that differ by one practice hour but have the same modeled prior score. The prior-score coefficient compares cases that differ by one prior-score unit but have the same modeled practice hours. Whether those comparisons are well supported depends on the observed combinations of predictors and on the adequacy of the linear model.

Predictors may share information. Practice hours and prior preparation may both be connected with the same parts of the outcome and with one another. A simple-regression slope includes all outcome variation aligned with its one predictor. A multiple-regression coefficient isolates the predictor's conditional linear component given the others. A change between the simple and conditional coefficients is expected and informative, but it needs substantive interpretation.

Model part	Meaning	Question to ask
Intercept b_0	Fitted outcome when every numerical predictor is zero and categorical predictors are at reference levels	Are those reference values meaningful and represented?
Numerical coefficient b_j	Fitted outcome difference for a one-unit predictor increase, holding others fixed	Which variables are being held fixed and in what units?
Indicator coefficient	Fitted difference from a stated reference category	Which category is the reference?
Interaction coefficient	Change in one predictor's slope across values or groups of another predictor	Which conditional slope is being modified?

Core ideas

Begin with a model chosen from the research question, not from a mechanical search across every available variable. A predictor may represent a focal exposure, a planned control, a group contrast, or a term needed to represent the functional form. Explain each role. Adding a variable changes the question answered by every conditional coefficient, so two models with different predictor sets are not interchangeable descriptions.

Categorical predictors enter through indicator variables. With three tutorial formats, one category becomes the reference and two indicators compare the other formats with it. Changing the reference category changes the printed intercept and contrasts but does not change fitted values. The reference should be stated in tables and prose.

An interaction means that the conditional association of one predictor differs across values of another. In a practice-by-format interaction, there is no single practice slope for all formats. The main practice coefficient is the slope within the reference format; each interaction coefficient tells

how another format's slope differs. Interpret the component coefficients together and show fitted lines or predicted values.

Evaluation level	Useful quantity or display	What it contributes
Individual coefficient	Estimate, standard error, interval, t test	Conditional direction, size, and uncertainty
Added predictor block	Nested-model F test and change in R^2	Whether the block adds modeled outcome variation
Whole model	R^2 , adjusted R^2 , overall F test	Sample fit and joint evidence for the predictor set
Model adequacy	Residual, quantile, leverage, and influence plots	Whether the fitted form and uncertainty assumptions are credible

R^2 cannot decrease when predictors are added, even if they contribute little useful information. Adjusted R^2 includes a penalty for the number of predictors and can decrease. Information criteria such as AIC also balance fit with model complexity, but comparisons are meaningful only among models fitted to the same outcome and observations. No single fit number replaces residual diagnostics or substantive judgment.

Strong predictor overlap means that predictors contain strongly overlapping linear information. It can enlarge coefficient standard errors and make individual estimates unstable while fitted values remain useful. It does not create bias by itself in an otherwise suitable model. Examine predictor relationships, coefficient uncertainty, and the design. Do not remove a conceptually necessary variable solely to make another coefficient significant. The supplied Statistics 1 materials do not define a universal numerical cutoff for deciding that predictor overlap is too strong.

The residual assumptions extend those of simple regression: an appropriate linear conditional mean, independent errors, suitable variance across fitted values, and a residual distribution adequate for the intended inference. Influential cases can alter several coefficients. Extrapolation can also occur in combinations of predictors even when each individual value lies within its observed range.

Formula guide

For p predictors, the population model is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i$$

The fitted sample value uses estimated coefficients, and the residual remains observed minus fitted:

$$\hat{y}_i = b_0 + \sum_{j=1}^p b_j x_{ji}, \quad e_i = y_i - \hat{y}_i$$

For two quantitative predictors, the conditional slopes can be expressed through the three pairwise correlations and the variables' standard deviations:

$$b_1 = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2} \frac{s_Y}{s_{X_1}}, \quad b_2 = \frac{r_{Y2} - r_{Y1}r_{12}}{1 - r_{12}^2} \frac{s_Y}{s_{X_2}}.$$

The subtraction represents correlation information shared with the other predictor, while the standard-deviation ratio returns the result to outcome units per predictor unit. If $|r_{12}| = 1$, the denominator is zero and the two separate slopes cannot be estimated from that model.

The residual standard error reports typical unexplained spread in outcome units:

$$s_e = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - p - 1}}.$$

Here p counts non-intercept predictor parameters. A categorical predictor may require more than one parameter.

For a quantitative predictor, the standardized coefficient is

$$\hat{\beta}_j = b_j \frac{s_{X_j}}{s_Y}.$$

It describes conditional fitted change in outcome standard deviations for a one-standard-deviation predictor difference. Unlike a bivariate correlation, it is conditional on the other model terms and is not restricted to the interval from -1 to $+1$.

The coefficient of determination compares residual and total sums of squares:

$$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}}$$

Adjusted R^2 accounts for sample size n and the number of predictors p :

$$R^2_{\text{adjusted}} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

The overall F statistic compares model mean square with residual mean square:

$$F = \frac{SS_{\text{model}}/p}{SS_{\text{residual}}/(n - p - 1)}$$

The global null is $H_0 : \beta_1 = \dots = \beta_p = 0$. A significant result says that at least one non-intercept population coefficient differs from zero under the model, but it does not identify which coefficient. An individual coefficient uses

$$t = \frac{b_j}{SE(b_j)}, \quad df = n - p - 1.$$

This test concerns coefficient j conditional on the exact other terms in the model. Its standard error is not the residual standard error.

For two nested models, the added contribution of q new predictors can be tested by comparing their reduction in residual sum of squares with the larger model's residual mean square:

$$F = \frac{(SS_{\text{residual, reduced}} - SS_{\text{residual, full}})/q}{SS_{\text{residual, full}}/(n - p - 1)}$$

The reduced model must be obtainable by setting the added full-model coefficients to zero, and both models must use the same outcome and analyzed cases. In this formula, p is the number of non-intercept predictor parameters in the full model, so the denominator uses the full model's residual degrees of freedom. For one added predictor, the semipartial correlation gives the same fit increment:

$$sr_j^2 = R^2_{\text{larger}} - R^2_{\text{smaller}} = \Delta R^2.$$

Only the candidate predictor is residualized for a semipartial correlation. Topic 6's partial correlation residualizes both focal variables.

Categorical predictors require indicators. With an intercept and k categories, use $k - 1$ indicators. For one quantitative predictor X and one binary indicator D , an additive model is

$$\hat{Y} = b_0 + b_1X + b_2D.$$

When $D = 0$, the fitted line is $b_0 + b_1X$. When $D = 1$, it is $(b_0 + b_2) + b_1X$. The lines are parallel, and b_2 is the fitted group difference at the same value of X .

An interaction allows the slopes to differ:

$$\hat{Y} = b_0 + b_1X + b_2D + b_3XD.$$

The reference-group slope is b_1 , the comparison-group slope is $b_1 + b_3$, and b_3 is the difference between slopes. The coefficient b_2 is the group difference at $X = 0$, so centering X may give that comparison a more useful reference point.

The Akaike information criterion used for candidate-model comparison is

$$AIC = -2\log(L) + 2k,$$

where L is the fitted likelihood and k is the number of estimated likelihood parameters. Smaller AIC indicates a better relative fit-complexity balance only among models fitted to the same outcome and cases. It does not prove that the selected model is true, causal, or accurate on new data.

Quantity	Question	Essential limit
R^2	How much sample outcome variation does this fitted model represent?	Cannot decrease when terms are added to the same OLS model
Adjusted R^2	Does the added sample fit outweigh its in-sample parameter penalty?	Is not validation on new cases
Nested F	Do the added coefficients jointly improve fit?	Requires truly nested models and the same cases
AIC	Which stated candidate has the best relative fit-complexity balance?	Has no universal pass mark

These formulas quantify sample fit and uncertainty. They do not determine which predictors make scientific sense or whether a conditional coefficient has a causal interpretation.

Reading the explanatory figure

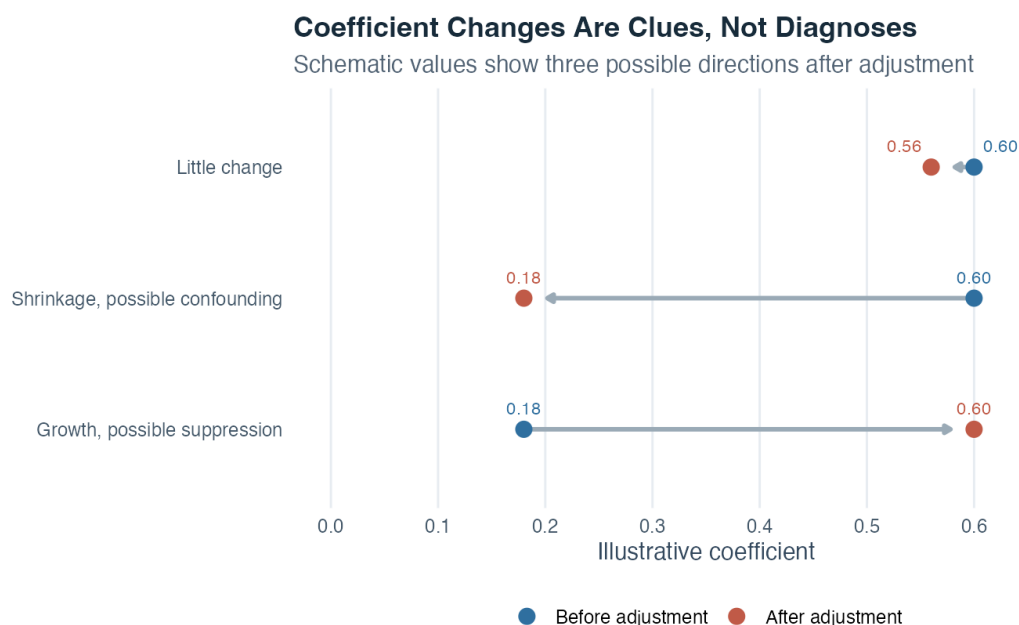


Figure 1: Three horizontal coefficient comparisons show little change, shrinkage, and growth between values before and after adjustment in a multiple-regression setting.

Each row compares a blue coefficient before adjustment with an orange coefficient after other predictors enter the model. In the upper row, 0.60 changes to 0.56. The conditional result is similar to the unadjusted result, so the added predictors changed this particular coefficient little. That does not prove that the predictors are irrelevant; they may improve prediction or matter for other coefficients.

In the middle row, 0.60 shrinks to 0.18. The focal predictor shared substantial outcome-related information with the added variables. Confounding is one possible substantive explanation, but the graph labels it as possible because coefficient movement alone cannot identify a causal role. Measurement overlap, selection, functional form, or sampling variation may also matter.

In the lower row, 0.18 grows to 0.60. Adjustment has revealed a stronger conditional association, a pattern often described as possible suppression. Again, the label is a clue rather than a conclusion. Inspect the predictor relations, coefficient intervals, design, and model diagnostics. The horizontal distance visualizes numerical movement; it does not display uncertainty, so a full analysis also needs confidence intervals.

Interpretation checklist

State the outcome, every predictor, units, coding, and reference categories. Explain why each predictor is included and whether interactions were planned. Inspect distributions, predictor relationships, missingness, and supported combinations. Translate each coefficient as a conditional fitted difference while naming what is held fixed. For interactions, report conditional slopes or predicted values rather than interpreting one term alone.

Compare nested models only when they use the same observations and outcome. Report coefficient estimates and intervals, R^2 , adjusted R^2 , relevant model comparisons, and diagnostics. Check residual shape, changing variance, leverage, influence, and predictor overlap. Keep prediction, association, and causation separate. If coefficients change across models, describe the change and investigate its source instead of assigning an automatic causal label.

How this topic connects

Multiple regression gathers the preceding association tools into one framework. Covariance and correlation introduced shared linear variation. Simple regression turned that pattern into a directional fitted equation with residuals. Partial correlation showed that “holding a variable constant” can be understood by residualizing focal variables. A multiple-regression coefficient applies the same conditional logic while keeping the outcome unit and allowing several predictors to be assessed together.

Analysis of variance is the next expression of this framework. Group membership can be represented by indicator predictors, so comparing group means becomes a regression model with categorical information. The ANOVA F test asks whether the group terms collectively account for outcome variation beyond residual variation. What first appears to be a separate method is therefore another view of the same general linear-model reasoning.

Document ID: topic-07-multiple-regression-summary-en

Notes 1 of 3

Multiple Regression | Topic 7

Notes 2 of 3

Multiple Regression | Topic 7

Notes 3 of 3

Multiple Regression | Topic 7