Article Talk

Read Edit View history

Search Wikipedia

Analysis of covariance

From Wikipedia, the free encyclopedia

"Ancova" redirects here. For the moth genus, see Ancova (moth).

Analysis of covariance (**ANCOVA**) is a general linear model which blends ANOVA and regression. ANCOVA evaluates whether the means of a dependent variable (DV) are equal across levels of a categorical independent variable (IV) often called a treatment, while statistically controlling for the effects of other continuous variables that are not of primary interest, known as covariates (CV) or nuisance variables. Mathematically, ANCOVA decomposes the variance in the DV into variance explained by the CV(s), variance explained by the categorical IV, and residual variance. Intuitively, ANCOVA can be thought of as 'adjusting' the DV by the group means of the CV(s).[1]

The ANCOVA model assumes a linear relationship between the response (DV) and covariate (CV):

$$y_{ij} = \mu + \tau_i + \mathrm{B}(x_{ij} - \overline{x}) + \epsilon_{ij}.$$

In this equation, the DV, y_{ij} is the jth observation under the ith categorical group; the CV, x_{ij} is the jth observation of the covariate under the ith group. Variables in the model that are derived from the observed data are μ (the grand mean) and $\overline{x}$ (the global mean for covariate x). The variables to be fitted are τ_i (the effect of the ith level of the IV), B (the slope of the line) and ϵ_{ij} (the associated unobserved error term for the jth observation in the ith group).

Under this specification, the *a* categorical treatment effects sum to zero $\left(\sum_i^a \tau_i = 0 \right)$. The standard assumptions of the linear regression model are also assumed to hold, as discussed below.[2]

Contents [hide]

Uses [edit]

Increase power [edit]

ANCOVA can be used to increase statistical power (the probability a significant difference is found between groups when one exists) by reducing the within-group error variance.[3] In order to understand this, it is necessary to understand the test used to evaluate differences between groups, the F-test. The F-test is computed by dividing the explained variance between groups

evaluate differences between groups, the F-test. The F-test is computed by dividing the explained variance between groups (e.g., medical recovery differences) by the unexplained variance within the groups. Thus,

$$F = \frac{MS_{between}}{MS_{within}}$$

If this value is larger than a critical value, we conclude that there is a significant difference between groups. Unexplained variance includes error variance (e.g., individual differences), as well as the influence of other factors. Therefore, the influence of CVs is grouped in the denominator. When we control for the effect of CVs on the DV, we remove it from the denominator making *F* larger, thereby increasing your power to find a significant effect if one exists at all.

Adjusting preexisting differences [edit]

Another use of ANCOVA is to adjust for preexisting differences in nonequivalent (intact) groups. This controversial application aims at correcting for initial group differences (prior to group assignment) that exists on DV among several intact groups. In this situation, participants cannot be made equal through random assignment, so CVs are used to adjust scores and make participants more similar than without the CV. However, even with the use of covariates, there are no statistical techniques that can equate unequal groups. Furthermore, the CV may be so intimately related to the IV that removing the variance on the DV associated with the CV would remove considerable variance on the DV, rendering the results meaningless.[4]
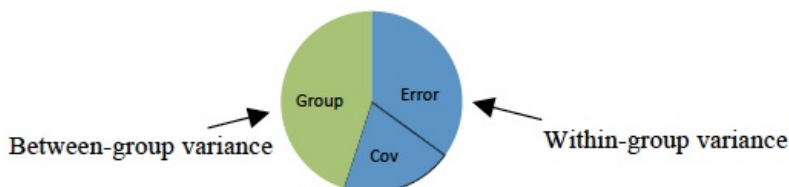
Assumptions [edit]

There are several key assumptions that underlie the use of ANCOVA and affect interpretation of the results.[2] The standard linear regression assumptions hold; further we assume that the slope of the covariate is equal across all treatment groups (homogeneity of regression slopes).

Assumption 1: linearity of regression [edit]

The regression relationship between the dependent variable and concomitant variables must be linear.

Assumption 2: homogeneity of error variances [edit]

The error is a random variable with conditional zero mean and equal variances for different treatment classes and observations.

Assumption 3: independence of error terms [edit]

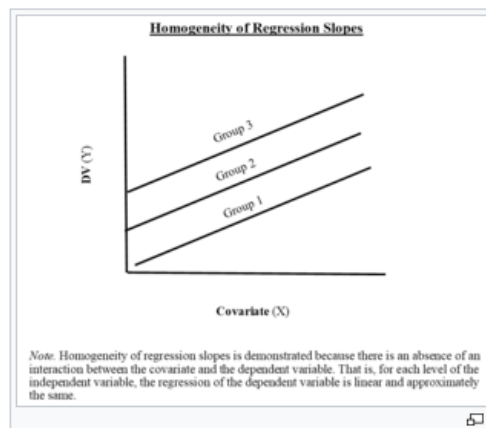The errors are uncorrelated. That is, the error covariance matrix is diagonal.

Assumption 4: normality of error terms [edit]

The residuals (error terms) should be normally distributed $\epsilon_{ij} \sim N(0, \sigma^2)$.

Assumption 5: homogeneity of regression slopes [edit]

The slopes of the different regression lines should be equivalent, i.e., regression lines should be parallel among groups.

The fifth issue, concerning the homogeneity of different treatment regression slopes is particularly important in evaluating the appropriateness of ANCOVA model. Also note that we only need the error terms to be normally distributed. In fact both the independent variable and the concomitant variables will not be normally distributed in most cases.

Note. Homogeneity of regression slopes is demonstrated because there is an absence of an interaction between the covariate and the dependent variable. That is, for each level of the independent variable, the regression of the dependent variable is linear and approximately the same.

Conducting an ANCOVA [edit]

Test multicollinearity [edit]

If a CV is highly related to another CV (at a correlation of 0.5 or more), then it will not adjust the DV over and above the other CV. One or the other should be removed since they are statistically redundant.

Test the homogeneity of variance assumption [edit]

Tested by Levene's test of equality of error variances. This is most important after adjustments have been made, but if you have it before adjustment you are likely to have it afterwards.

Test the homogeneity of regression slopes assumption [edit]

To see if the CV significantly interacts with the IV, run an ANCOVA model including both the IV and the CVxIV interaction term. If the CVxIV interaction is significant, ANCOVA should not be performed. Instead, Green & Salkind[5] suggest assessing group differences on the DV at particular levels of the CV. Also consider using a moderated regression analysis, treating the CV and its interaction as another IV. Alternatively, one could use mediation analyses to determine if the CV accounts for the IV's effect on the DV.

Run ANCOVA analysis [edit]

If the CV×IV interaction is not significant, rerun the ANCOVA without the CV×IV interaction term. In this analysis, you need to use the adjusted means and adjusted MSerror. The adjusted means (also referred to as least squares means, LS means, estimated marginal means, or EMM) refer to the group means after controlling for the influence of the CV on the DV.

Follow-up analyses [edit]

If there was a significant main effect, it means that there is a significant difference between the levels of one IV, ignoring all other factors.[6] To find exactly which levels are significantly different from one another, one can use the same follow-up tests as for the ANOVA. If there are two or more IVs, there may be a significant interaction, which means that the effect of one IV on the DV changes depending on the level of another factor. One can investigate the simple main effects using the same methods as in a factorial ANOVA.
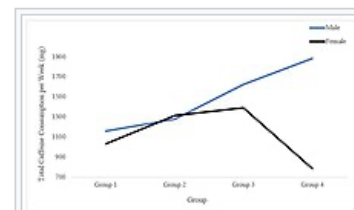
Simple main effects plot showing a small Interaction between the two levels of the independent variable.

Power considerations [edit]

While the inclusion of a covariate into an ANOVA generally increases statistical power by accounting for some of the variance in the dependent variable and thus increasing the ratio of variance explained by the independent variables, adding a covariate into ANOVA also reduces the degrees of freedom. Accordingly, adding a covariate which accounts for very little variance in the dependent variable might actually reduce power.

See also [edit]

- MANCOVA (Multivariate analysis of covariance)

References [edit]

1. ^ Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs: Prentice-Hall, Inc.
2. ^ *a* *b* Montgomery, Douglas C. "Design and analysis of experiments" (8th Ed.). John Wiley & Sons, 2012.
3. ^ Tabachnick, B. G.; Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Pearson Education.
4. ^ Miller, G. A.; Chapman, J. P. (2001). "Misunderstanding Analysis of Covariance". *Journal of Abnormal Psychology*. **110** (1): 40–48. doi:10.1037/0021-843X.110.1.40. PMID 11261398.
5. ^ Green, S. B., & Salkind, N. J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data* (6th ed.). Upper Saddle River, NJ: Prentice Hall.
6. ^ Howell, D. C. (2009) *Statistical methods for psychology* (7th ed.). Belmont: Cengage Wadsworth.

External links [edit]

- Examples of all ANOVA and ANCOVA models with up to three treatment factors, including randomized block, split plot, repeated measures, and Latin squares, and

Wikiversity has learning resources about *ANCOVA*

- including randomized block, split plot, repeated measures, and Latin squares, and their analysis in R ⊠ (University of Southampton)
- One-Way Analysis of Covariance for Independent Samples ⊠
- Use of covariates in randomized controlled trials by G.J.P. Van Breukelen and K.R.A. Van Dijk (2007) ⊠

V · T · E	**Statistics**	[show]
V · T · E	**Design of experiments**	[show]
V · T · E	**Least squares** and **regression analysis**	[show]

Categories: Analysis of variance | Covariance and correlation

This page was last edited on 7 May 2021, at 16:00 (UTC).

WIKIMEDIA project   Powered by MediaWiki