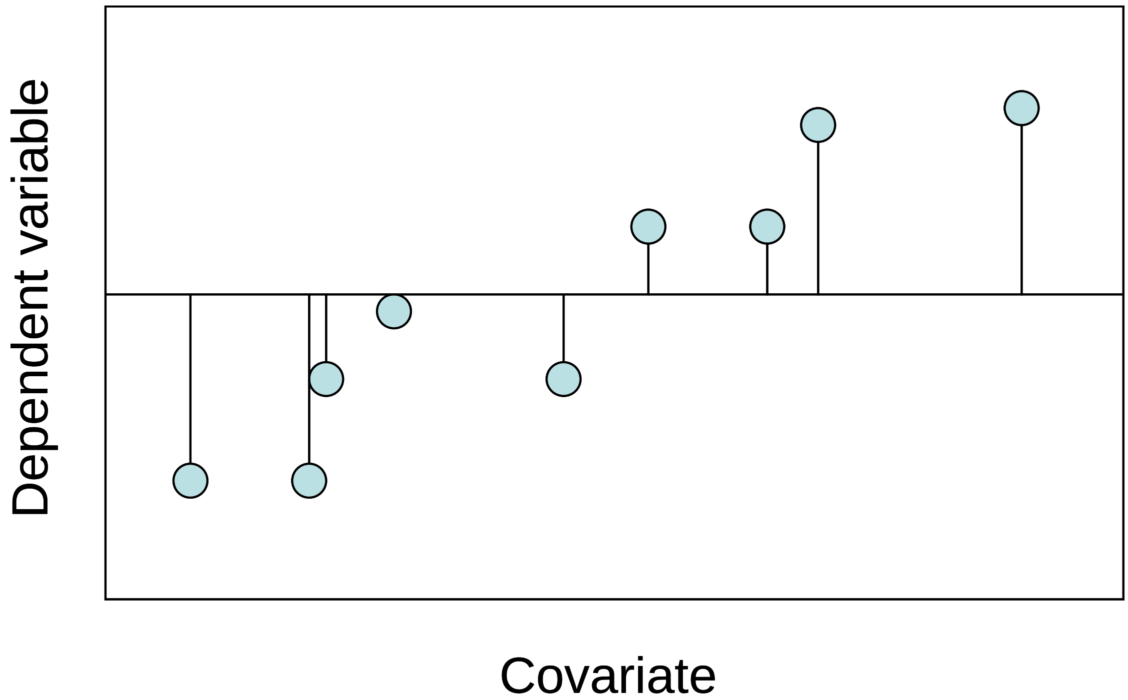
Analysis of Covariance

The analysis of covariance (ANCOVA) is a combination of regression analysis and analysis of variance. Covariance is used when the response variable *Y*, in addition to being affected by the treatments, is linearly related to another variable *X*. Hence, ANCOVA's usefulness depends on a strong correlation between the covariate and the dependent variable.

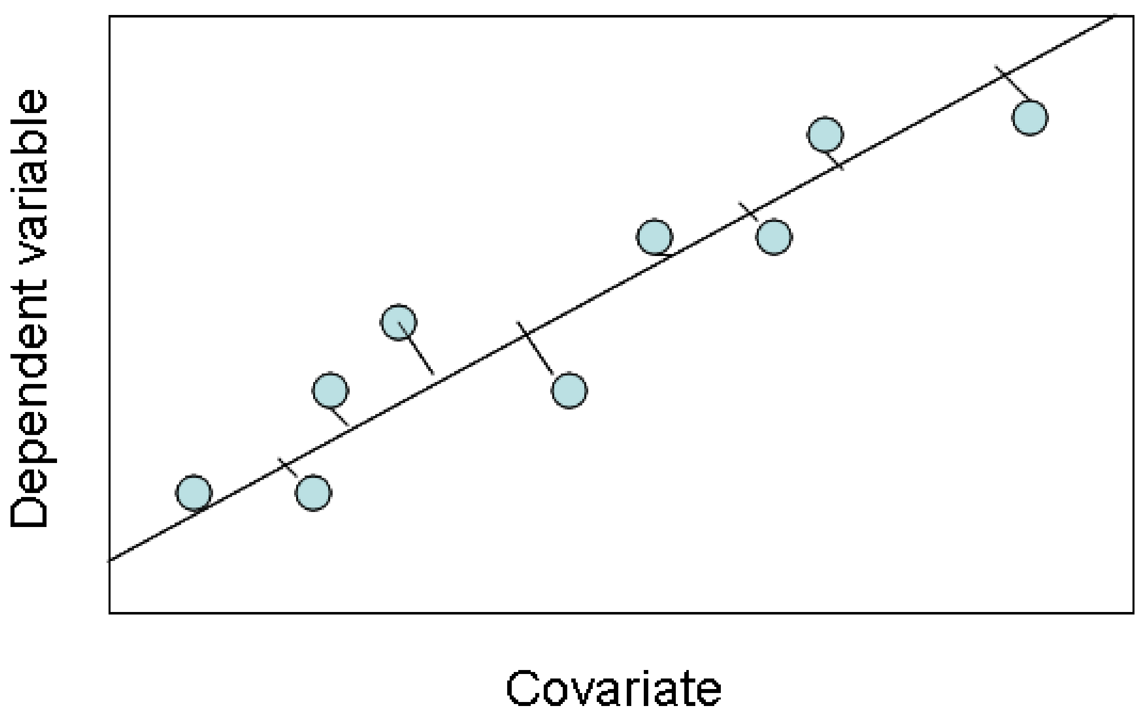
I. Partitioning variability
  a. ANOVA
    i. The variability is divided into two components
      1. Treatment effect
      2. Error – experimental and individual differences
  b. ANCOVA
    i. The variability is divided into three basic components
      1. Treatment effect
      2. Error
      3. Covariate

II. General Linear Model
  a. One-way ANOVA
    i. $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where μ is the grand mean of the population, α_i is the fixed treatment effect for group *i*, and ε_{ij} is the error term of the j^{th} item in group *i*.
  b. Two-way ANOVA
    i. $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, where μ equals the parametric mean of the population, α_i is the fixed treatment effect for the i^{th} group of treatment *A*, β_j is the fixed treatment effect of the j^{th} group of treatment *B*, $(\alpha\beta)_{ij}$ is the interaction effect in the subgroup representing the i^{th} group of factor *A* and the j^{th} group of factor *B*, and ε_{ijk} is the error term of the k^{th} item in subgroup *ij*.
  c. ANCOVA
    i. $Y_{ij} = \mu + \alpha_i + \beta_{within}(X_{ij} - \overline{X}_i) + \varepsilon_{ij}$, where observation Y_{ij} is the j^{th} item in the i^{th} group, μ is the grand mean of the population, α_i is the fixed treatment effect for group *i*, $\beta_{within}(X_{ij} - \overline{X}_i)$ is the effect explained by the difference of the variate X_{ij} from its mean for *X*, and ε_{ij} is, as usual, the random deviation

III. Error variability
  a. ANOVA
    i. The error variability comes from the subject within group deviation from the mean of the group



b. ANCOVA
  i. The error variability is based on the deviation of the score from the regression line, which will be smaller than that in the ANOVA.



IV. Hypotheses
   a. The slopes of the regression lines are equal
   b. The Y-intercept of the regression lines are equal

V. We first have to test the assumption of parallelism (equal slopes). If the regression lines for each group are parallel, then a test for equality of the Y-intercept is equivalent to testing the differences of the means in the simulated experiment with constant X

VI. Assumptions
   a. All the assumptions that underlie ANOVA
      i. The error terms, and the observations, within each group come from a normally distributed population
      ii. The variances of the error term should be approximately equal in each group
      iii. The error terms and the observations should be independent
   b. Plus assumptions specific to ANCOVA
      i. A linear relationship between covariate and dependent variable
      ii. The covariate is independent of the treatment groups
      iii. The covariate is assumed to be a fixed variable with no error associated with it

VII. Choice of a good covariate
   a. Reduce error variance
   b. Yield a more precise and accurate estimate of group effects
   c. Takes into account the relationship between the covariate and the dependent variable

VIII. Data table

Data table for one-way ANCOVA

Sample 1			Sample J			Sample K	
X	Y		X	Y		X	Y
X_{11}	Y_{11}	…	X_{ij}	Y_{1j}	…	X_{1k}	Y_{1k}
X_{21}	Y_{21}	…	X_{2j}	Y_{2j}	…	X_{2k}	Y_{2k}
…	…	…	…	…	…	…	…
X_{i1}	Y_{i1}	…	X_{ij}	Y_{ij}	…	X_{ik}	Y_{ik}
…	…	…	…	…	…	…	…
X_{n1}	Y_{n1}	…	X_{nj}	Y_{nj}	…	X_{nk}	Y_{nk}

mean	$\bar{X}_1$	$\bar{Y}_1$	$\bar{X}_j$	$\bar{Y}_j$	$\bar{X}_k$	$\bar{Y}_k$

IX. Formulas for calculating Sum of Squares

a. For X

i. $$SS_{A(X)} = n\sum_{i=1}^{K}(\bar{X}_i - \bar{\bar{X}})^2$$

ii. $$SS_{W(X)} = \sum_{i=1}^{K}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$$

iii. $$SS_{T(X)} = \sum_{i=1}^{K}\sum_{j=1}^{n}(X_{ij} - \bar{\bar{X}})^2$$

b. For Y

i. $$SS_{A(Y)} = n\sum_{i=1}^{K}(\bar{Y}_i - \bar{\bar{Y}})^2$$

ii. $$SS_{W(Y)} = \sum_{i=1}^{K}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2$$

iii. $$SS_{T(Y)} = \sum_{i=1}^{K}\sum_{j=1}^{n}(Y_{ij} - \bar{\bar{Y}})^2$$

c. For XY

i. $$SP_A = n\sum_{i=1}^{K}(\bar{Y}_i - \bar{\bar{Y}})(\bar{X}_i - \bar{\bar{X}})$$

ii. $$SP_W = \sum_{i=1}^{K}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)(X_{ij} - \bar{X}_i)$$

iii. $$SP_T = \sum_{i=1}^{K}\sum_{j=1}^{n}(Y_{ij} - \bar{\bar{Y}})(X_{ij} - \bar{\bar{X}})$$

X. Summary table

Summary table for ANCOVA

Source	Sum of Squares (adjusted)	df	Mean Squares (adjusted)	F
Among (adjusted)	$SS_{A(adj)}$	$df_{A(adj)}$	$MS_{A(adj)}$	F
Within (adjusted)	$SS_{W(adj)}$	$df_{W(adj)}$	$MS_{W(adj)}$	
Total	$SS_{T(adj)}$	$df_{T(adj)}$	$MS_{T(adj)}$	

XI. Formulas for summary table

a. $SS_{T(adj)} = SS_{T(Y)} - \frac{(SP_T)^2}{SS_{T(X)}}$

b. $SS_{W(adj)} = SS_{W(Y)} - \frac{(SP_W)^2}{SS_{W(X)}}$

c. $SS_{A(adj)} = SS_{T(adj)} - SS_{W(adj)}$