

RELATIONSHIP BETWEEN ANOVA AND REGRESSION

-Pooja Radhakrishnan

- ANOVA model is a special case of regression model in which all the predictors are categorical.
- ANOVA is a statistical model that is used to predict the continuous outcome of a response variable based on one or more *categorical* independent/predictor variables.
- Linear Regression is a statistical model that is used to predict the continuous outcome of a response variable based on one or more *continuous* independent/predictor variables.
- Both the models can be used only if the outcome/response variable is a continuous response variable and not for categorical response variables.
- Linear Regression is based on the concept of Least Squares, whereas ANOVA has the same in addition to many other ways to produce it. If there are categorical predictor variables, then the regression technique converts the categorical predictor variable to continuous predictor variable using indicator variables (0/1 for binary or n indicator variables for n levels.).
- Both the models can be used to check and plot the residuals.
- Mathematically, both ANOVA and Linear Regression are used as models to test against the same hypothesis.
- In ANOVA, each category's mean of the categorical variable is compared to the grand mean.
- In Regression, the intercept of each category of the categorical variable is compared with the reference group's intercept.
- ANOVA is concerned to present the differences between the categories' means in the data.
- Linear Regression is mostly concerned to estimate the sample mean response and an associated variance.
- The F-statistic in the case of the ANOVA and in the case of linear regression will be the same. The reason is that because, both the procedures test the same hypothesis, but with different structure.
- ANOVA will check qualitatively if the ratio obtained is high enough to interpret that no grouping is unlikely.
- Linear Regression will check qualitatively if the ratio obtained is high enough to interpret and suggest that an intercept model is sparse.
- Example:
Response variable: To determine how many hours of sleep a person has in a night
Predictor variable: Age in years
Model to be used: **Regression**
Response variable: To determine how many hours of sleep a person has in a night
Predictor variable: Parent/Child, Activities performed before sleep
Model to be used: **ANOVA**
Response variable: To determine how many hours of sleep a person has in a night
Predictor variable: Age in years, Parent/Child, Activities performed before sleep
Model to be used: **Regression/ANOVA**