



Linear model assumptions



Linear Model Assumptions

1. **Linearity:** The mean values of the outcome variable for each increment of the predictor(s) lie along a straight line. There is a linear relationship between predictors and target.
2. **Normally distributed errors:** the residuals (ε_i) are random, normally distributed with a mean of 0.
3. **Homoscedasticity:** At each level of the predictor variable(s), the variance of the residual terms should be constant.
4. **No perfect multicollinearity:** There should be no perfect linear relationship between two or more of the predictors.

Linear Model Assumptions

4. **Variable types:** Predictor variables should be quantitative or categorical with 2 categories. The target should be continuous and unbiased.
5. **Non-zero variance:** Predictors should not be constant.
3. **Predictors are uncorrelated to external variables.**
4. **Independence:** All values of the target are independent (each row in the dataset is independent).



When the assumptions are not met

- The coefficients cannot be trusted.



Multicollinearity

- If 2 variables are perfectly collinear (correlation=1), there is an infinite combination of coefficients that would work equally well.
- Perfect collinearity is rare.
- Partial collinearity is unavoidable.

THANK YOU

www.trainindata.com