Lecture 3: ANOVA Diagnostics and Remedial Measures

STA 106: Analysis of Variance

Suggested reading: ALSM Chapter 18

Xiner Zhou

Department of Statistics, University of California, Davis

ANOVA Diagnostics and Remedial Measures

ANOVA Model Assumptions

Check assumptions
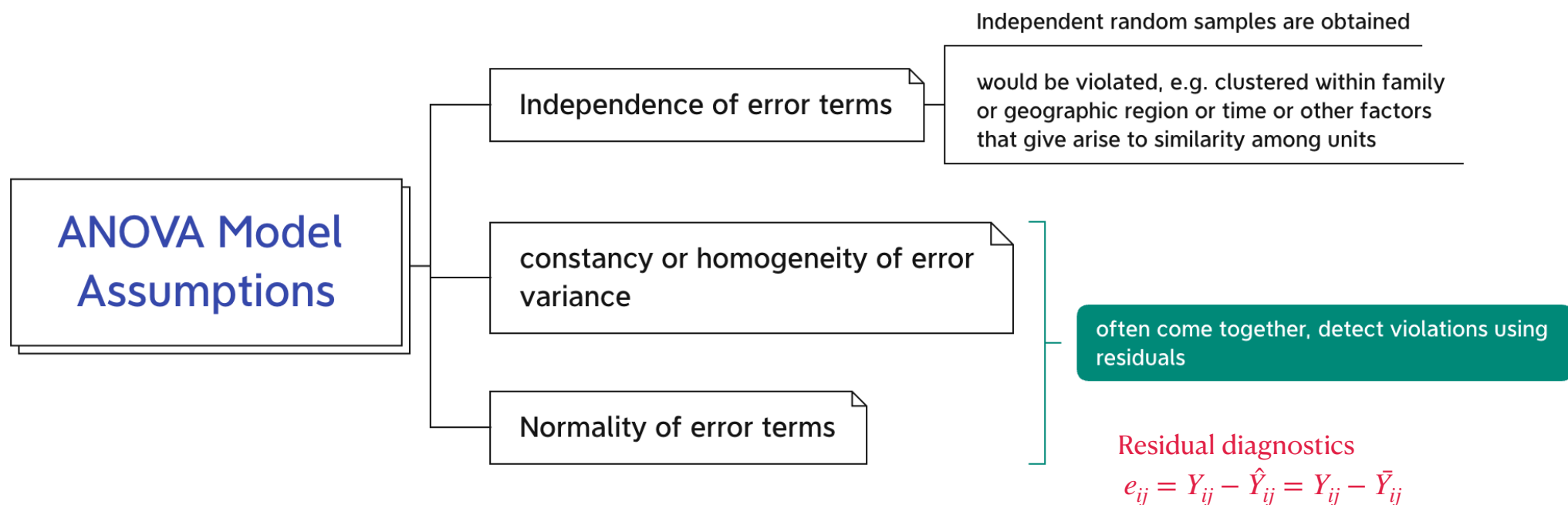
Remedies and concerns about assumption violations

ANOVA Model Assumptions

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

- ϵ_{ij} error term

We assume $\epsilon'_{ij}s$ are independently and identically distributed as $N(0,\sigma^2)$, for all i, j

$$\epsilon_{ij} \sim N(0,\sigma^2)$$

Independent random samples are obtained

would be violated, e.g. clustered within family or geographic region or time or other factors that give arise to similarity among units

ANOVA Model Assumptions

Independence of error terms

constancy or homogeneity of error variance

Normality of error terms

often come together, detect violations using residuals

Residual diagnostics
$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{ij}$$

What does "ANOVA model is appropriate" mean?

It is not necessary, nor possible, that an ANOVA model fit the data perfectly.

The major purpose of examination of the appropriateness of the model is to detect serious departure from the assumptions made by ANOVA Model

Examine whether the proposed ANOVA model is appropriate for the data at hand

Not appropriate

Yes appropriate

Remedial measures

Inferences based on the appropriate ANOVA model