Interaction between x_1 and x_2

Implies that the relationship between y and one x depends on the other x.

Parsimonious Model

A model with a small number of β parameters.

Recommendation for Assessing Model Adequacy

1. Conduct global F-test; if significant then:
2. Conduct t-tests on only the most important β's (*interaction* or *squared terms*)
3. Interpret value of $2s$
4. Interpret value of R_a^2

Recommendation for Testing Individual β's

1. If *curvature* (x^2) deemed important, do not conduct test for first-order (x) term in the model.
2. If *interaction* (x_1x_2) deemed important, do not conduct tests for first-order terms (x_1 and x_2) in the model.
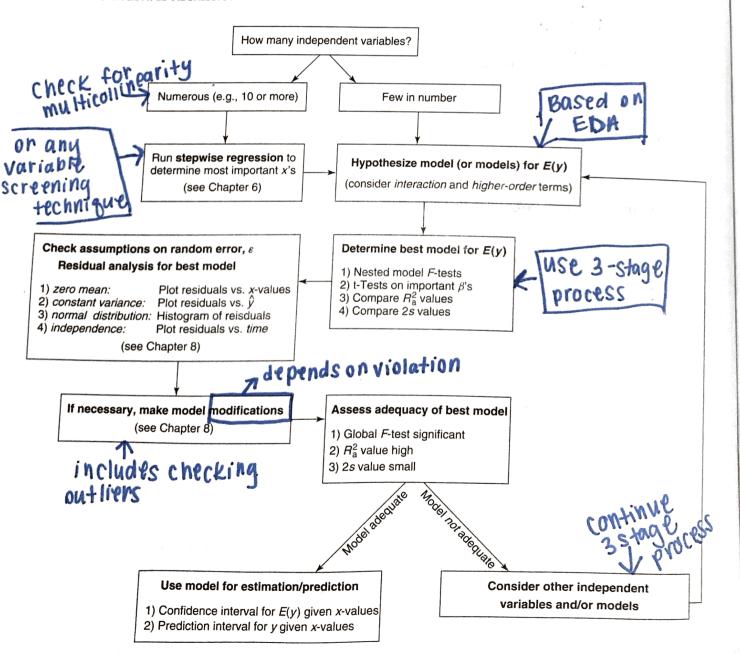
Extrapolation

Occurs when you predict y for values of x's that are outside of range of sample data.

Nested Models

Models where one model (the *complete model*) contains all the terms of another model (the *reduced model*) plus at least one additional term.

GUIDE TO MULTIPLE REGRESSION

Handwritten annotations: check for multicollinearity; or any variable screening technique; Based on EDA; use 3-stage process; depends on violation; includes checking outliers; continue 3 stage process

Flowchart:

How many independent variables?
→ Numerous (e.g., 10 or more) / Few in number

Numerous → Run **stepwise regression** to determine most important x's (see Chapter 6)

Few in number → **Hypothesize model (or models) for $E(y)$** (consider *interaction* and *higher-order* terms)

→ **Determine best model for $E(y)$**
1) Nested model F-tests
2) t-Tests on important β's
3) Compare R_a^2 values
4) Compare $2s$ values

→ **Check assumptions on random error, ε** — Residual analysis for best model
1) *zero mean:* Plot residuals vs. x-values
2) *constant variance:* Plot residuals vs. $\hat{y}$
3) *normal distribution:* Histogram of residuals
4) *independence:* Plot residuals vs. *time*
(see Chapter 8)

→ **If necessary, make model modifications** (see Chapter 8)

→ **Assess adequacy of best model**
1) Global F-test significant
2) R_a^2 value high
3) $2s$ value small

Model adequate → **Use model for estimation/prediction**
1) Confidence interval for $E(y)$ given x-values
2) Prediction interval for y given x-values

Model not adequate → **Consider other independent variables and/or models**