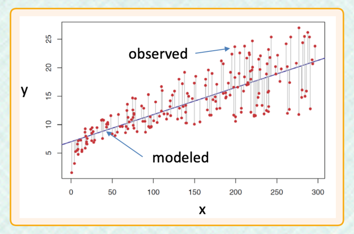
L2 Supervised Learning

Linear Regression

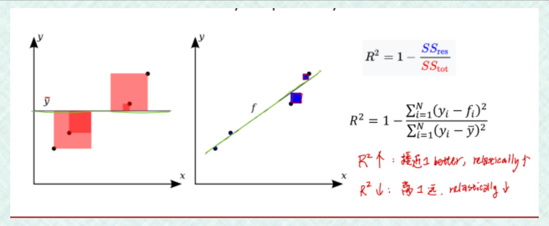
- Wy regression?
  - Increasing independent variable x increase/decreases the outcome y
  - Creates an **interpretable** model that explains the data
  - Powerful and flexible tool
- Linear Regression
  - Parametric Model
    - y=B0+B1x
    - x: observed features
    - y: observed response(outcome)
    - Model parameters B0 and B1 estimated from data
  - Prediction: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
  - Residual error: $e_i = y_i - \hat{y}_i$
  - Residual sum $\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$



- Mathematical intuition: OLS
  - RSS
    - $\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$
    - $\mathrm{RSS} = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \ldots$
  - Choose parameters that minimize RSS. Ordinary least squares coefficient estimates are sample means
    - $\bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$ are sample means
    - $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
    - $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$,

Model

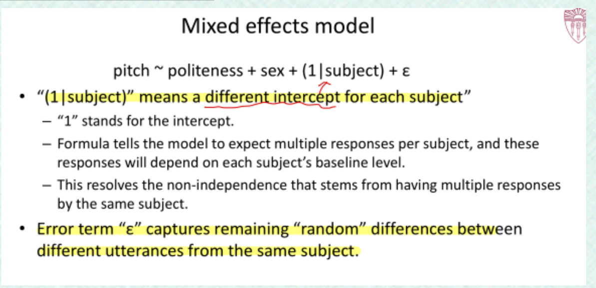
- Simple Model
  - Pitch (Y)
    - Dependent variable
    - Outcome
    - Response
  - Sex (X)
    - Independent variable
    - Explanatory variable
    - Predictor
    - Feature
    - fixed effect
  - Error(unmeasurable factor)
    - Error term
    - Random factors
  - Pitch = Sex + Error
- R-squared(R2)
  - Controls for the complexity of the model: Reflects how much variance in data is accounted for by the model considering how many variables it uses.
  - Coefficient of determination - R2
    - How much of the variation in y is explained by the variation in x?



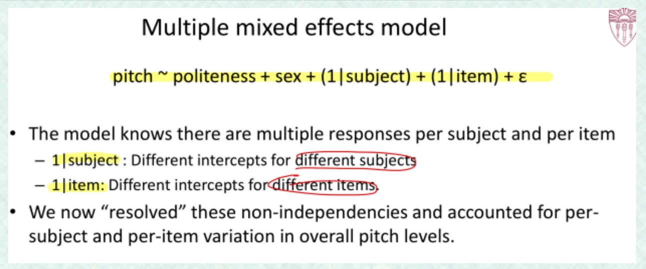
- P-value
  - Define: Probability our data could be observed given that the H0 is true
  - H0 Sex has no effect on Pitch
  - Ha Sex has effect on Pitch
  - P-value < 5%: statistically significant;

Mixed Effects Models

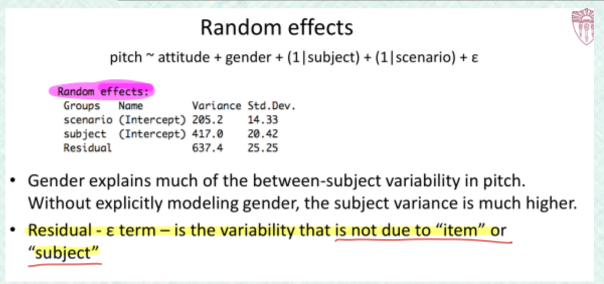
- Idependence assumption is often violated
  - Ex. Does politeness affect pitch?
    - Polite
    - Informal
    - Multiple responses
    - Multiple responses from the same subject cannot be regarded as independent from each other.
    - Pitch = Politeness + Sex + Error
  - Add a reandom effect
- Models
  - 1 Mixed effects Model



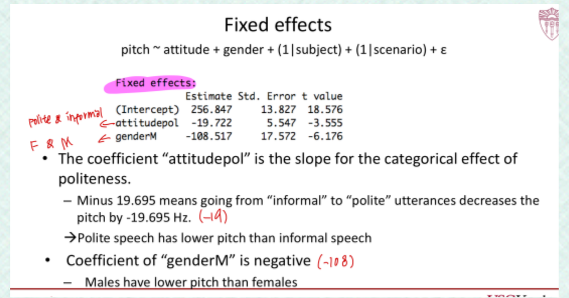
  - 2 Multiple mixed effects model



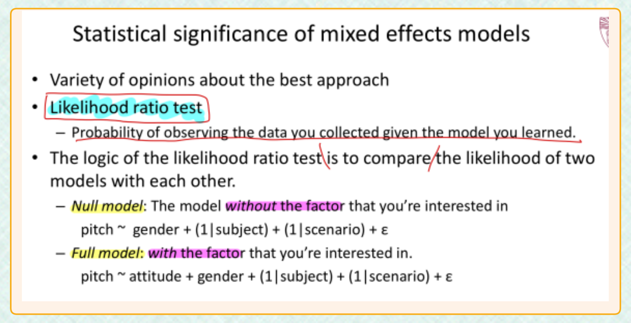
- Effect
  - 1 Random Effects



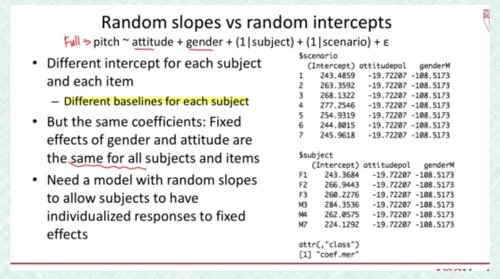
  - 2 Fixed Effects

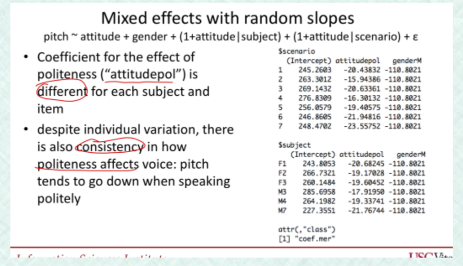


- Stat. Significance of Mixed Effects Models
  - Likelihood Ratio Test
    - 1 Null Model
    - 2 Full Model



- Random Slopes V.S random intercept

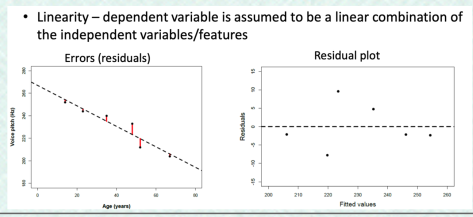




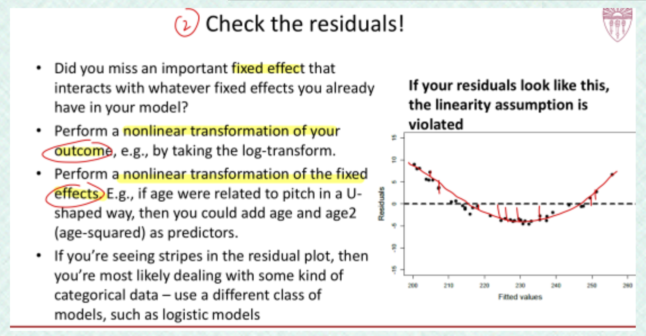
  - ????
  - ??????

Multiple regression Model

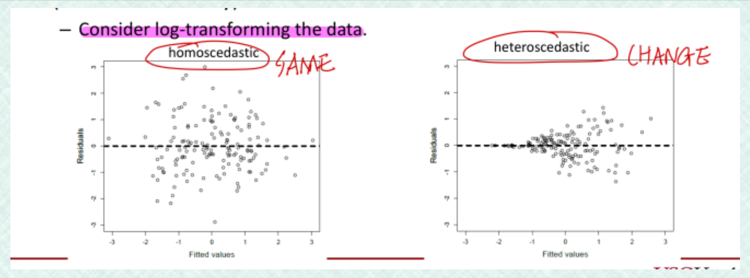
- Assumptions of linear models
  - 1 Linearity
    - Dependent variable is assumed to be a linear combination of the independent variable/ features



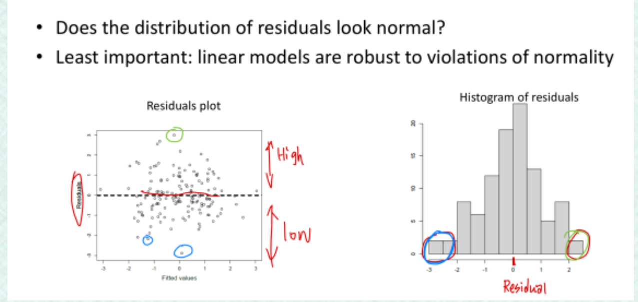
  - 2 Check residuals!



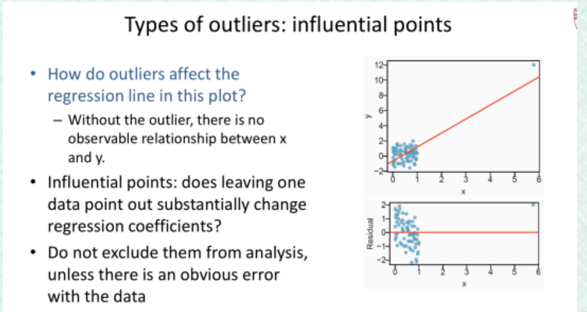
  - 3 Absence of Collinearity
    - If two fixed effects are correlated with each other, they are collinear
    - Intelligence syllables persecond + words per second — Fixed effects are correlated
    - Different linear combinations of fixed effects can produce the same response — Cannot interpret coefficients
    - Use PCA, feature selection, etc. to choose a smaller set of explanatory fixed effects
- Homoskedasticity
  - Unequal variances: Does the variance of the data stage the same across the range of predicted values (homoscedasticity) or does it change (heteroskedasticity).

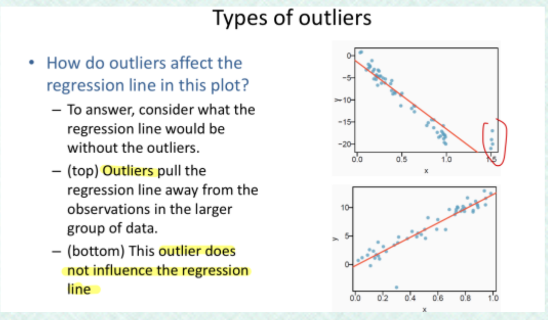


- Normality of residuals



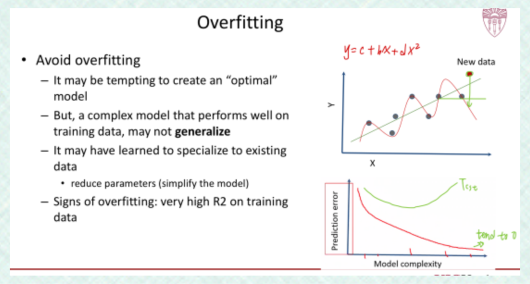
- Outliers
  - Influential points: does leaving one data point out substantially change regression coefficients?
  - Outliers that lie away from the center of the cloud in the x-direction are called high leverage points.
  - A point is infuential if including or excluding the point would considerably change the slope of the regression line.
  - DO NOT exclude them from analysis, unless there is a obvious error with the data.



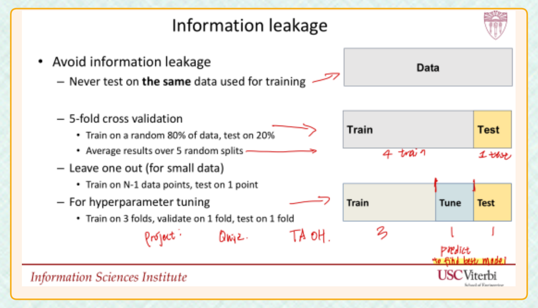


Pitfall and Best Practices

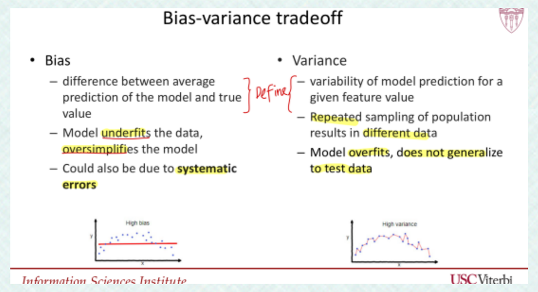
- Overfitting

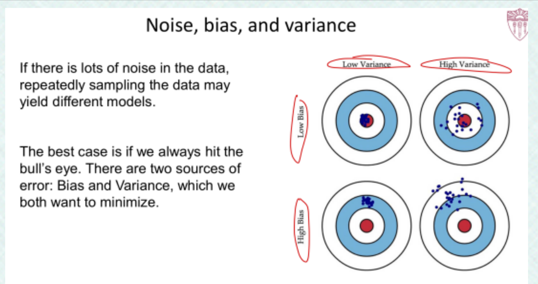


- Information Leakage
  - 1 NEVER test on the SAME data used for training
  - 2 N-fold Cross validation
  - 3 Hyperparameter Tunning



- Noise
  - Bias
  - Variance
  - Bias-Variance Tradeoff

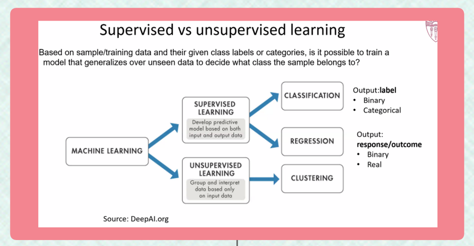




1.Unfortunately, it is not always possible to minimize both variance and bias at the same time. In general, bias is reduced if we add more and more parameters to a model and make it more complex.
2.However, the more complex the model becomes the more variance we introduce in the model. In its core the problem alludes to over- and under-fitting.

- Feature engineering
  - 1 Feature Transformation
  - 2 Feature Selection
    - 1 Forward Selection
    - 2 Backward Elimination

Machine Learning



- Supervised Learning
  - Classification
  - Regression
  - LABEL
- Unsupervised Learning
  - Clustering
  - UN-LABEL