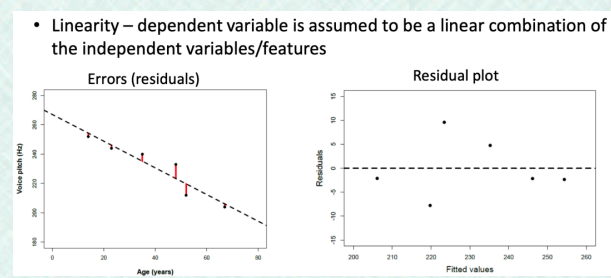
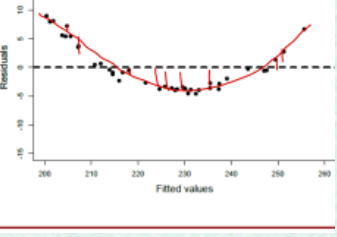


Dependent variable is assumed to be a linear combination of the independent variable/ features



Check the residuals!

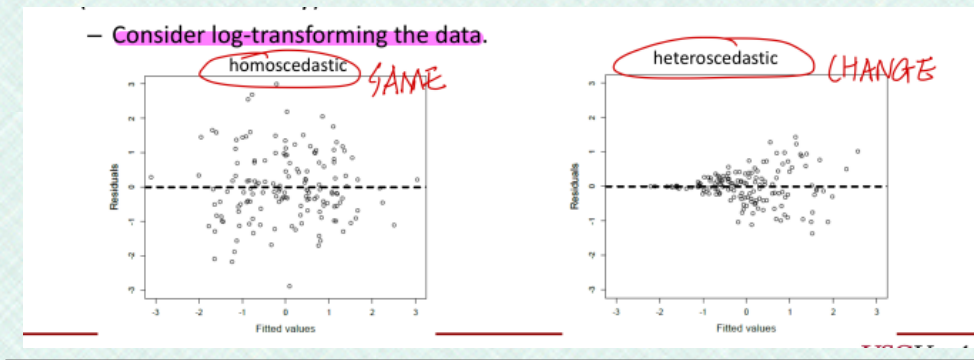
- Did you miss an important **fixed effect** that interacts with whatever fixed effects you already have in your model?
- Perform a **nonlinear transformation of your outcome**, e.g., by taking the log-transform.
- Perform a **nonlinear transformation of the fixed effects**. E.g., if age were related to pitch in a U-shaped way, then you could add age and age² (age-squared) as predictors.
- If you're seeing stripes in the residual plot, then you're most likely dealing with some kind of categorical data – use a different class of models, such as logistic models.



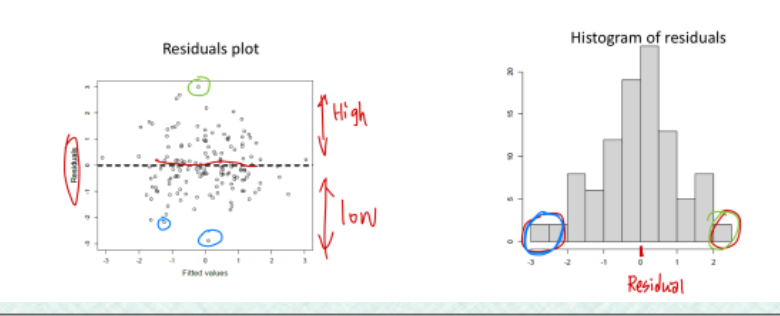
If two fixed effects are correlated with each other, they are collinear

- Fixed effects are correlated
- Intelligence syllables per second + words per second
- Cannot interpret coefficients
- Different linear combinations of fixed effects can produce the same response
- Use PCA, feature selection, etc. to choose a smaller set of explanatory fixed effects

Unequal variances: Does the variance of the data stay the same across the range of predicted values (homoscedasticity) or does it change (heteroscedasticity).



- Does the distribution of residuals look normal?
- Least important: linear models are robust to violations of normality



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 influential if including or excluding the point would considerably change the slope of the regression line.

DO NOT exclude them from analysis, unless there is an obvious error with the data.

Types of outliers: influential points

- How do outliers affect the regression line in this plot?
 - Without the outlier, there is no observable relationship between x and y.
- Influential points:** does leaving one data point out substantially change regression coefficients?
- Do not exclude them from analysis, unless there is an obvious error with the data

Types of outliers

- How do outliers affect the regression line in this plot?
 - To answer, consider what the regression line would be without the outliers.
 - (top) **Outliers** pull the regression line away from the observations in the larger group of data.
 - (bottom) This **outlier** does not influence the regression line.

Overfitting

- Avoid overfitting
 - It may be tempting to create an "optimal" model
 - But, a complex model that performs well on training data, may not generalize
 - It may have learned to specialize to existing data
 - = reduce parameters (simplify the model)
 - Signs of overfitting: very high R2 on training data

Information leakage

- Avoid information leakage
 - Never test on the same data used for training
 - 5-fold cross validation
 - Train on a random 80% of data, test on 20%
 - Average results over 5 random splits
 - Leave one out (for small data)
 - Train on N-1 data points, test on 1 point
 - For hyperparameter tuning
 - Train on 3 folds, validate on 1 fold, test on 1 fold

Bias-variance tradeoff

- Bias
 - difference between average prediction of the model and true value
 - Model **underfits** the data, **oversimplifies** the model
 - Could also be due to **systematic errors**
- Variance
 - variability of model prediction for a given feature value
 - Repeated sampling of population results in different data
 - Model **overfits**, does not generalize to test data

Noise, bias, and variance

If there is lots of noise in the data, repeatedly sampling the data may yield different models.

The best case is if we always hit the bull's eye. There are two sources of error: Bias and Variance, which we both want to minimize.

- 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.
- 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.

- NEVER test on the SAME data used for training
- N-fold Cross validation
- Hyperparameter Tunning

Bias-Variance Tradeoff

- Feature Transformation
- Forward Selection
- Backward Elimination
- Feature Selection

L2 Supervised Learning

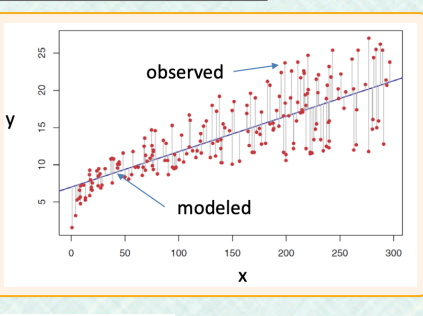
Linear Regression

Why regression?

- Increasing independent variable x increase/decreases the outcome y
- Creates an **interpretable** model that explains the data

Linear Regression

- Parametric Model
 - $y = B_0 + B_1x$
 - x: observed features
 - y: observed response(outcome)
 - Model parameters B0 and B1 estimated from data
- Prediction: $\hat{y} = \beta_0 + \beta_1x$
- Residual error: $e_i = y_i - \hat{y}_i$
- Residual sum: $RSS = e_1^2 + e_2^2 + \dots + e_n^2$



Mathematical intuition: OLS

- RSS: $RSS = e_1^2 + e_2^2 + \dots + e_n^2$
 $RSS = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots$
- Choose parameters that minimize RSS. Ordinary least squares coefficient estimates 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
 - 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?
 - $R^2 = 1 - \frac{RSS}{TSS}$
 - $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
 - $R^2 \uparrow$: **More** variance, **accuracy**?
 - $R^2 \downarrow$: **Less** variance, **accuracy**?

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;

Independence 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

Mixed effects model

pitch ~ politeness + sex + (1|subject) + ε

- "(1|subject)" means a **different intercept for each subject**
- "1" stands for the intercept.
- Formula tells the model to expect multiple responses per subject, and these responses will depend on each subject's baseline level.
- This resolves the non-independence that stems from having multiple responses by the same subject.
- Error term "ε" captures remaining "random" differences between different utterances from the same subject.

Multiple mixed effects model

pitch ~ politeness + sex + (1|subject) + (1|item) + ε

- The model knows there are multiple responses per subject and per item
- 1|subject: Different intercepts for **different subjects**
- 1|item: Different intercepts for **different items**.
- We now "resolved" these non-independencies and accounted for per-subject and per-item variation in overall pitch levels.

Random effects

pitch ~ attitude + gender + (1|subject) + (1|scenario) + ε

Group	Intercept	Estimate	Std. Dev.	t value
scenario (Intercept)	285.2	18.807	18.936	
subject (Intercept)	417.0	15.547	13.555	
subject (attitude)	437.0	28.42	23.25	

- Gender explains much of the between-subject variability in pitch. Without explicitly modeling gender, the subject variance is much higher.
- Residual - ε term** is the variability that is **not due to "item" or "subject"**

Fixed effects

pitch ~ attitude + gender + (1|subject) + (1|scenario) + ε

	Intercept	Estimate	Std. Dev.	t value
gender (Intercept)	256.867	18.807	18.936	
gender (attitude)	-19.722	15.547	13.555	
gender (gender)	-108.527	17.572	14.176	

- The coefficient "attitudepol" is the slope for the categorical effect of politeness.
- Minus 19.695 means going from "informal" to "polite" utterances decreases the pitch by 19.695 Hz.
- Polite speech has lower pitch than informal speech
- Coefficient of "genderM" is negative (-19.5)
- Males have lower pitch than females.

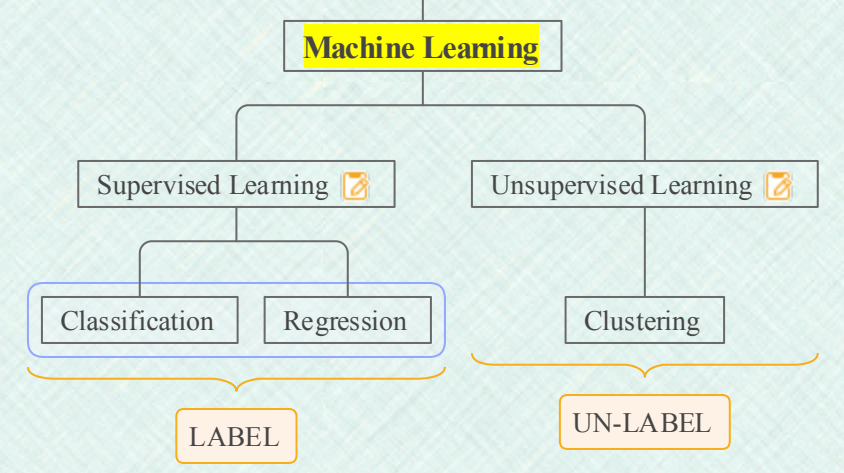
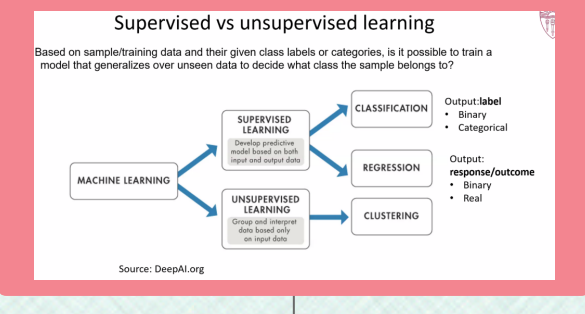
Stat. Significance of Mixed Effects Models

Likelihood Ratio Test

- Null Model
- Full Model

Statistical significance of mixed effects models

- Variety of opinions about the best approach
- Likelihood ratio test**
 - Probability of observing the data you collected given the model you learned.
- The logic of the likelihood ratio test is to compare the likelihood of two models with each other.
 - Null model:** The model **without the factor** that you're interested in
pitch ~ gender + (1|subject) + (1|scenario) + ε
 - Full model:** **with the factor** that you're interested in
pitch ~ attitude + gender + (1|subject) + (1|scenario) + ε



Random slopes vs random intercepts

Full = pitch ~ attitude + gender + (1|attitude|subject) + (1|scenario) + ε

- Different intercept for each subject and each item
- Different baselines for each subject and item
- But the same coefficients: Fixed effects of gender and attitude are the same for all subjects and items
- Need a model with random slopes to allow subjects to have individualized responses to fixed effects

Mixed effects with random slopes

pitch ~ attitude + gender + (1|attitude|subject) + (1|scenario) + ε

	Intercept	Estimate	Std. Dev.	t value
gender (Intercept)	256.867	18.807	18.936	
gender (attitude)	-19.722	15.547	13.555	
gender (gender)	-108.527	17.572	14.176	

- despite individual variation, there is also consistency in how politeness affects voice pitch tends to go down when speaking politely