Linear Regression: Frequentist vs. Bayesian approaches

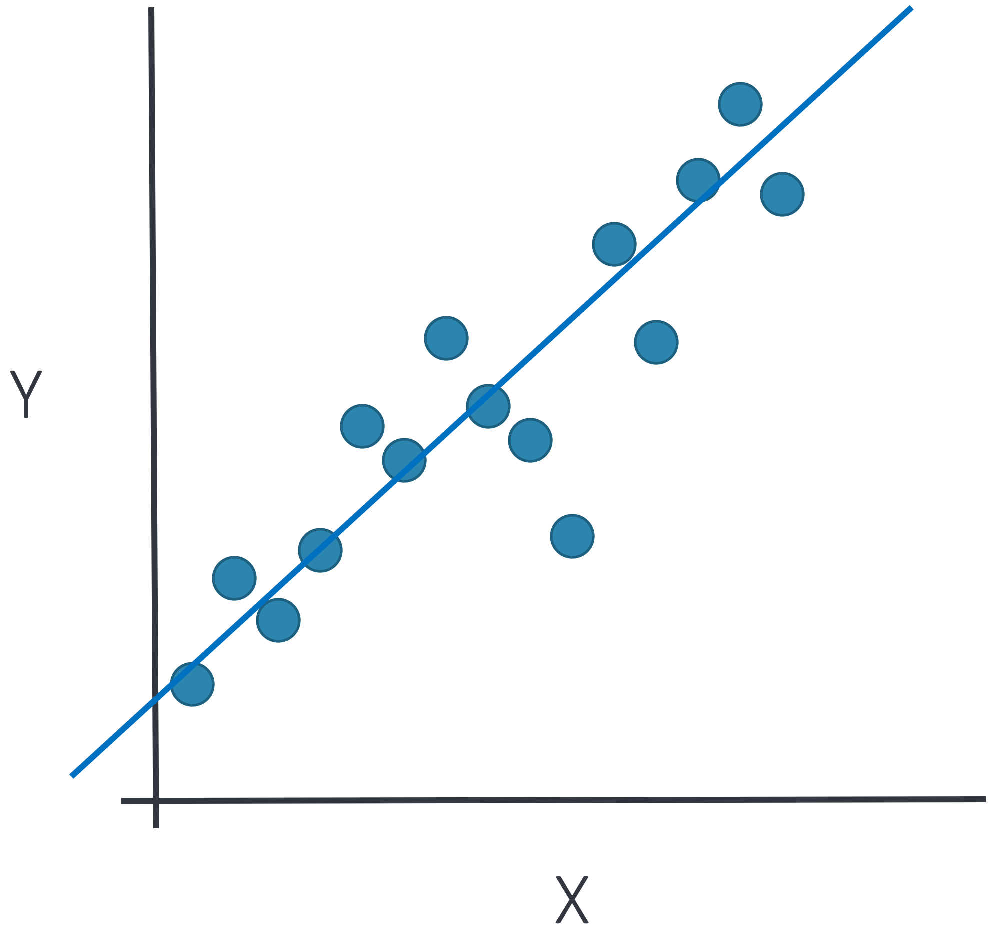
May 6, 2025

Prof. Gwendolyn Eadie

Terminology used in statistics and for regression

- X values can be called...
  - Covariates
  - Predictors
  - Explanatory variables
  - Independent variables
  - Features (computer science/machine learning)

- Y values can be called
  - Response variable
  - Dependent variable



Regression

- Estimate functional relationships
- Many types!
  - Linear regression, generalized linear models (GLMs), lasso regression, Poisson regression, logistic regression, …
- "standard" regression assumes X is fixed without error
- **Linear regression does not imply we are fitting a line**
  - E.g. "linear" regression means ***linear in the parameters***

Concept of Regression

- *Estimating functional relationships*

$$E[Y|X] = f(X, \theta) + \epsilon$$

- Note the asymmetry in most regression analysis. This is not a fit to the joint distribution of (Y, X)
- *Homoscedastic errors:* ϵ is an n-vector with σ^2
- *Heteroscedastic errors*: ϵ is an n-vector with σ_i^2 *(known or unknown)*
- Errors-in-Variables models assume X has error as well

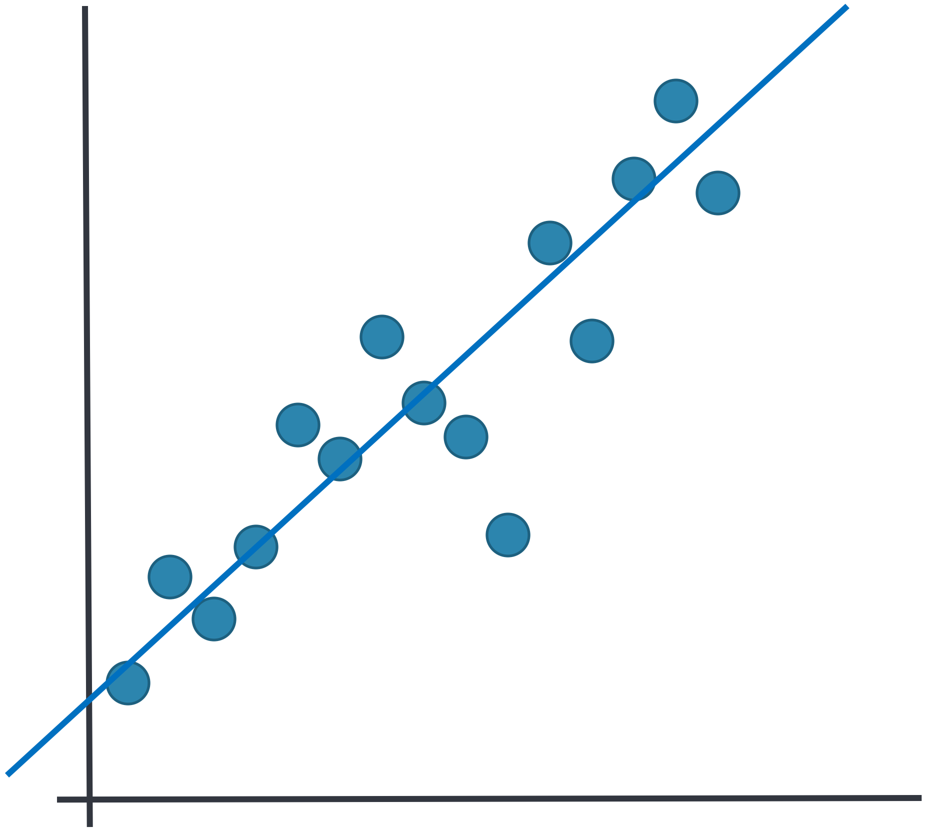
Are these models linear or non-linear?

Example	Linear or Non-Linear?
$Y = \beta_0 + \beta_1 X + \beta_2 e^{X} + \epsilon$	
$Y = \left(\frac{X}{\beta_0}\right)^{-\beta_1} + \epsilon$	
$Y = \beta_0 + \beta_1 \sin(X) + \beta_2 \cos(X) + \epsilon$	
$Y = \beta_0 e^{-\beta_1 X} + \epsilon$	

Conditions for linear regression

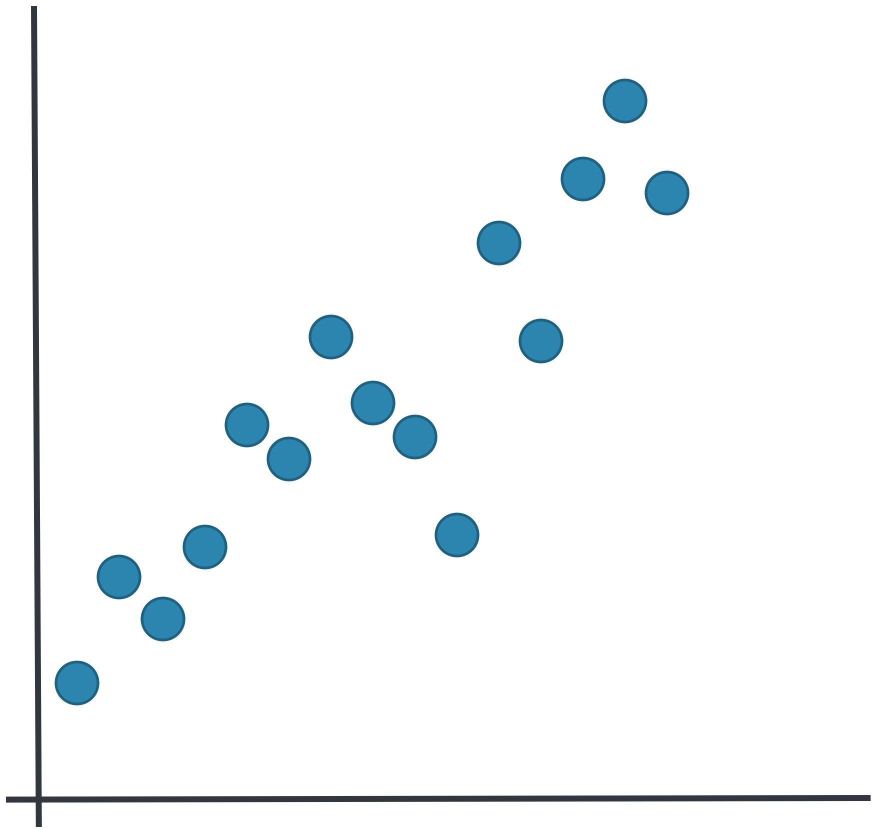
- Nearly normal residuals
  - Residuals should be normally distributed about 0
  - No trends in residuals
  - No major outlier(s) or "influential points" (we'll get to this!)
- Constant variability
  - Variability above/below least squares line shouldn't change as x changes
- Independent observations
  - e.g., typically don't apply to time series data

Frequentist: Fitting a line using Ordinary Least Squares (OLS) Regression



- Useful when the relationship between two quantities can be summarized by a straight line
- Correlation describes the *strength* of the correlation between x and y, whereas the regression line is used to describe the relationship between x and y
- The regression line is a model which follows the equation:

Frequentist:
Fitting a line using Ordinary Least Squares (OLS) Regression



- A regression line can tell you something about the effect of the predictor or independent variable on the response variable
- Slope of a regression line is related to the correlation of the points:
- Intercept of a regression line is:

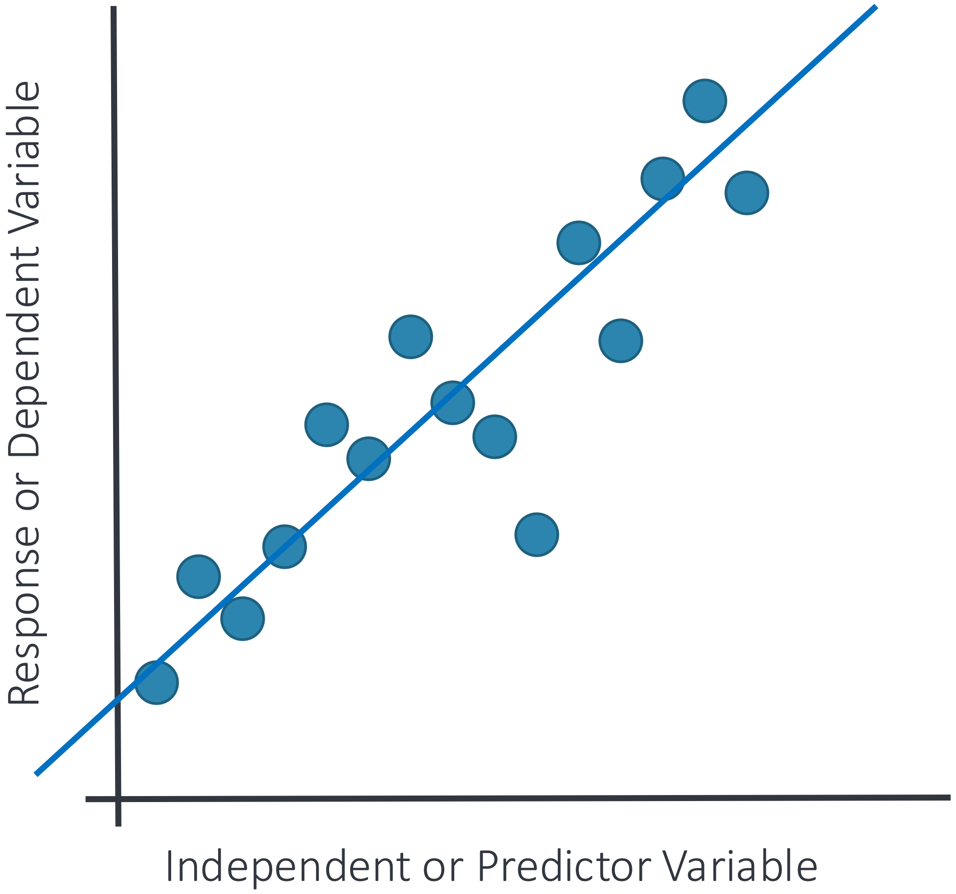
Relationship between correlation and slope (for simple L.R.)

Correlation r between x and y :

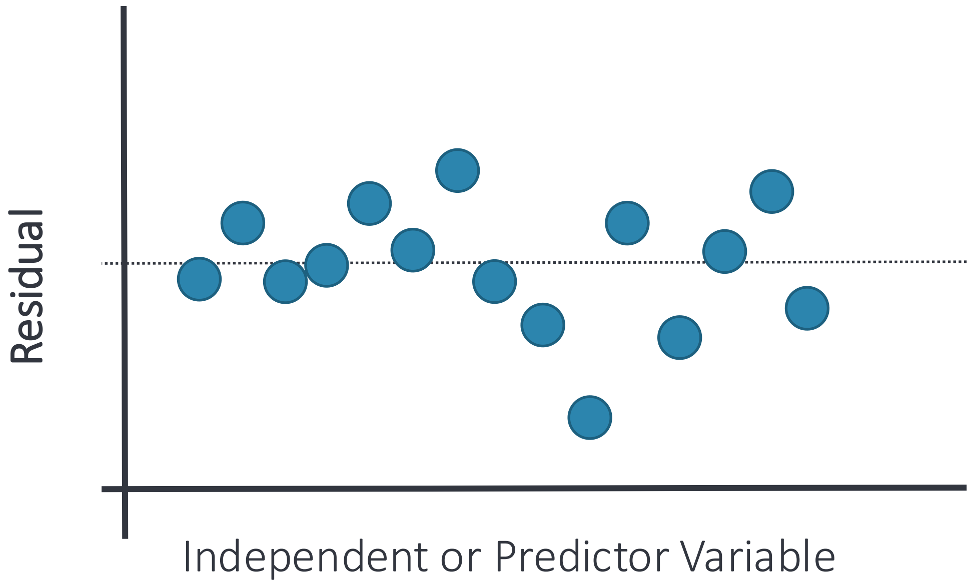
(always between -1 and 1)

- r tells you about the *strength* of the relationship
- Slope tells you change in Y per unit change in X

Frequentist:
How we find the least squares line



- The least squares line **minimizes the sum of squared residuals:**



Linear Regression more generally

$$Y = X\beta + \epsilon$$

Dimensions of each:

It helps to visualize:

Equivalently:

More concisely:

Review of Linear Regression

- Model

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}$$

- X often called the "design matrix"
- X is an $n \times p$ matrix of covariates/explanatory variables. These could be
  - Measurements
  - Fixed by design
  - Introduced to increase model flexibility

- In practice, the intercept β_0 may be encoded in X:

Least Squares estimators for β

- If the linear regression model is $Y = \beta_0 + \beta_1 X + \epsilon$, then the OLS estimator minimizes the *residual sums of squares (RSS):*

- Least squares estimator of β:

$$\hat{\beta}_{LS} = \mathrm{argmin}_{\beta} \sum_{i}^{n} \left(y_i - x_i^T \beta\right)^2$$

Maximum Likelihood Estimator (MLE) for β

Assuming normally-distributed, independent errors

If we assume that $y_i \sim N(x_i^T \beta, \sigma^2)$, then we can write out the likelihood function

Likelihood function

$y_i|X \sim N(X\beta, \sigma^2 I)$

The likelihood function is

$$L(\beta, \sigma^2; y) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left\{-\frac{1}{2}(y - X\beta)^T(y - X\beta)\right\}$$

The log-likelihood is then

And now we can do maximum likelihood estimation …

Maximum likelihood estimate of β

$$\ell(\beta,\sigma^2;y) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)$$

Take derivative w.r.t β and set equal to zero:

M.L.E. of β: $\hat{\beta}_{ML} = (X^TX)^{-1}X^TY = \hat{\beta}_{LS}$

Slope and Intercept Estimators

- Under the assumption that ϵ are i.i.d., then the slope and intercept estimators are unbiased and asymptotically distributed as:

assuming σ is known.

- When the true variance is unknown, the distributions of the estimators are not known (because we don't know β_1 nor σ^2)

Uncertainties in the line

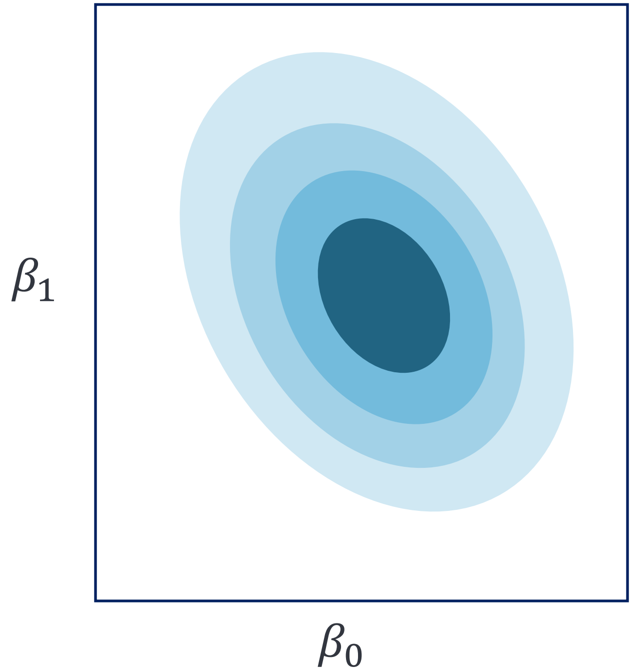
- Because we know that $\widehat{\beta_1} \sim N(\beta_1, \frac{\sigma^2}{s})$, then

And we can use this to construct confidence intervals for Y at each X using

→ do this for a range of x-values, and you will get confidence bands about the best-fit (OLS) line that are hyperbolas.

Point Estimates vs. Bayesian Inference

- The estimators for β_0, β_1, etc. In previous slides are *point estimates*
- A maximum likelihood estimate is also a point estimate of a parameter
- In Bayesian inference…
  - We try to infer the *distribution* for a parameter
  - We get a *posterior distribution*
  - The posterior distribution encodes all the information from
    - Prior assumptions
    - Model assumptions
    - Data
  - We report the whole posterior distribution in our results
  - We can report credible intervals to express uncertainty



Linear Regression in Bayesian context

- The model for y is

- The likelihood under this model is

- We must set priors on the parameters