Linear Regression: via Ordinary Least Squares (OLS)

(1) parameters are linear

$$y = mx + b$$

dependent, slope, independent, intercept

(2) regression form:

$$y = \alpha + \beta x + \varepsilon$$

unexplained variance

error term → "residual"

$$y = \alpha + \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

(3) error term

$$\varepsilon \sim N(0, \sigma^2)$$

for unbiased + precise estimates, the error term must be normally distributed with a mean of zero

(4) residual testing concepts

(a) necessary to ensure condition in (3) is met

(b) residuals should be "independently + identically distributed"

↓

"iid"

independence of errors

AKA

zero conditional mean

↓

the ~~mean~~ mean error for all possible values of the covariates = 0

identically Distributed errors

AKA

homoskedasticity

↓

the variance of error for all possible values of the covariates is the same

normally distributed errors

(5) residual tests

(a) note: residuals are a sample of the error term

(b.) independence of errors

- most important in time series models
- formal tests: Durbin-Watson
  Breusch-Godrey
- plotting: residuals vs. observation # - should be ~0 throughout
  ACF (autocorrelation function)
  PACF (partial autocorrelation function)

(c.) homoskedastic errors

- when errors are not homoskedastic, they are said to be heteroskedastic
- formal tests: Breusch-Pagan
  White's Test
- plotting: residuals vs. predicted - should be no patterns
  residuals vs. observation # (time series) - should have same variance throughout

(d.) normal errors

- formal test: Jarque-Bera
- plotting: histogram of residuals