Linear model assumptions

Two more

1. Linear relation between y and x:
2. The sample $[y_i, x_i]$ is a random
3. $E(y \mid x) = \alpha + \beta x$
4. Homoskedasticity, residual variance is constant: $Var(y \mid x) = \sigma^2$
5. The response y has a normal distribution given the predictor: $y \sim N(\alpha + \beta x, \sigma)$

OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ are unbiased

OLS estimate of σ^2 is unbiased

$\hat{\alpha}$ and $\hat{\beta}$ have known distributions and we can calculate p-values

Linear regression is flexible!

Adding flexibility to our models

- The linear model we are using consists of making the parameters of probability distributions change according to some function

- The simplest function is a linear function

$$y_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta x_i$$

- Sometimes the relation between parameters and predictors is not linear

- We can use whatever functional shape we like, but it is useful to use transformations to linearize the relation