Special Case #1 : $h(\mu_i) = \mu_i$ i.e. $SD[y_i] \propto \mu_i$

standard deviation of the response is proportional to the mean.

$$g'(\mu_i) = \frac{1}{h(\mu_i)} = \frac{1}{\mu_i}$$

$$\therefore g(\mu_i) = \log(\mu_i)$$

natural

Thus $g(y_i) = \log(y_i)$ is the response transformation that stabilizes this sort of non-constant variance

Special Case #2 : $h(\mu_i) = \sqrt{\mu_i}$ i.e., $Var[y_i] \propto \mu_i$

variance of the response is proportional to the mean

$$g'(\mu_i) = \frac{1}{h(\mu_i)} = \frac{1}{\sqrt{\mu_i}}$$

$$\therefore g(\mu_i) = 2\sqrt{\mu_i}$$

Thus $g(y_i) = \sqrt{y_i}$ is the response transformation that stabilizes this sort of non-constant variance.

Alternatively we could the more general class of transformations called the Box-Cox Transformations (aka, "power transformations"):

$$g(y_i) = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(y_i) & \text{for } \lambda = 0 \end{cases}$$

natural

* all of the transformations we've discussed arise as special cases of this:
  - $\lambda = \frac{1}{2}$ (square root)
  - $\lambda = 0$ (natural-log)
  - $\lambda = -1$ (reciprocal)

* An optimal value of λ can be found algorithmically by choosing the value that maximizes the likelihood corresponding to the model with the transformed response.

The Box-Cox transformation is very effective, but if we care about interpretation then the log-transformation is more appropriate.

↳ The issue is that β_j is interpretted as the expected difference between

$$g(y_i \mid x_j = x+1) - g(y_i \mid x_j = x)$$

Interpeting this on the response scale is not possible in general. However when $g(\cdot) = \log(\cdot)$ then we have

$$\log(y_i \mid x_j = x+1) - \log(y_i \mid x_j = x)$$

$$\Rightarrow \beta_j = E\left[\log\left(\frac{y_i \mid x_j = x+1}{y_i \mid x_j = x}\right)\right]$$

$\therefore e^{\beta_j}$ is interpretted as the expected multiplicative change in y_i when x_j is incresed by 1 unit (all else equal)

Multicollinearity

A common issue in linear regression is when two or more explanatory variables are linearly related to one another (either exactly or approximately).

i.e. Given explanatory variables x_1, x_2, x_3

$$x_1 = a + b x_2 \quad \text{or} \quad x_1 \approx a + b x_2$$

$$x_1 = a + b x_2 + c x_3 \quad \text{or} \quad x_1 \approx a + b x_2 + c x_3$$

This is the problem of multicollinearity. If the linear relationship is exact, then the columns of X are linearly dependent and we cannot invert X^TX (and hence estimate the β's). If the linear relationship is approximate, then the columns of X are close to linearly dependent and the resulting problem is one of variance inflation.

↓

This is manifested by inflated $Var[\hat{\beta}_j]$ when x_j is in the model with other multicollear x's, vs. when these other x's are not in the model.