Transformations

- **The need for transformation**

1) Linearity of the regression. Diagnosis: a plot of e_i against y_i reveals a trend (linear, logarithmic, exponential, etc.)

2) Normality of the residuals - outliers or skewness / heavy tails of y_i. Transformation on y may do the trick.

3) Variance stabilization

* Our hope is to settle all of these issues at once, but we may not be able to. Either way, we'll try.

* **Variance stabilization transformations**

* Suppose we have a sequence of random variables $Y_1, \dots, Y_n$ w/ $E[Y_i] = \mu_i$ and $Var(Y_i) = \sigma_i^2$

* Suppose in particular that σ_i^2 is a function of μ_i:

$$\sigma_i^2 = g(\mu_i)$$

ex: $Y_i \sim Po(\lambda) \Rightarrow \sigma^2 = \mu \Rightarrow g(\mu) = \mu$

- Aim: find a function $h(Y_i)$ whose variance is constant.

- To do this, we use a Taylor approximation for $h(Y_i)$:

$$h(Y_i) - h(\mu_i) \approx (y_i - \mu_i)\, h'(\mu_i)$$

$$\Rightarrow \quad Var(h(Y_i)) \approx (h'(\mu_i))^2\, \sigma_i^2 \qquad (*)$$

Note that the RHS(*) will be constant if

$$(h'(\mu))^2\, g(\mu) = \text{constant} \qquad (**)$$

- Point: If we know $g(\mu)$, we can exactly determine the function / transformation h of Y_i that gives a constant variance by solving (**).

- For many distributions, we can solve (**) and hence come up w/ a variance stabilizing transformation.

- Example: $Y_i \sim Po(\lambda_i)$

Then, $\sigma_i^2 = \lambda_i \Rightarrow g(\lambda_i) = \lambda_i$

Solving (**) gives $h(\lambda_i) = \sqrt{\lambda_i}$ and in fact

$$(h'(\lambda_i))^2 \lambda_i = (\tfrac{1}{2}\lambda_i^{-1/2})^2 \lambda_i$$
$$= \tfrac{1}{4}$$

So, the $\sqrt{Y_i}$ is a variance stabilizing transform for Poisson (count) data, and in so doing we expect $Var(Y_i) \equiv \frac{1}{4}$.

* <u>Box - Cox Families of Transformations</u>

* Many transformations we can take can be embedded into a power law family. This class, <u>known</u> as <u>Box - Cox transformations</u> can be <u>written</u> as:

$$h(y) = \frac{y^\lambda - 1}{\lambda} \qquad \text{(Box - Cox)}$$

where λ can take on any real number. For $\lambda = 0$, we have $h(y) = \ln(y)$.

* <u>Choosing λ</u>

* The general approach for choosing from a possible Box - Cox transformation is to calculate λ through maximum likelihood estimation.

* I forego the details here but show the example of how to do this from page 111 of Faraway.

* The boxcox( ) function will return a confidence interval for λ for which you choose the most interpretable value.

* Notably: if $\lambda = 1$ is in the CI, we have no good reason to perform a transformation.

* Important point about Inference after a Transformation: Any inference that you make will now be about the linear effect of x on $h(y)$. You have to think carefully about how to bring CIs on β for this model back to CIs for β for x on y.

Example: $h(y) = x\beta + \varepsilon$

$$\Rightarrow \quad y = h^{-1}(x\beta + \varepsilon)$$