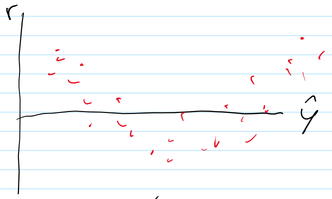
18. Box-Cox Transform

October 20, 2025 9:41 AM

Recall from last time:

Sometimes our data is not linear!

- looking at the data: Fitted vs Residual plot



← Hints that the data is nonlinear

- Fix the data / Model by transforming y and/or x_i
  - example: $y \leftarrow \log(y)$, $x_i \leftarrow x_i^2$
  - Problem: How to pick a good transformation?
- One Solution is to use outside knowledge
  → Physics (cars) → Geometry (tree)
- Another solution is to let the data decide!

Box-Cox transformation

We are going to replace y_i with $y_i^{(\lambda)} = \begin{cases} \frac{1}{\lambda}(y_i^{\lambda} - 1) & , \lambda \neq 0 \\ \log(y_i) & , \lambda = 0 \end{cases}$

(take y_i to the λ-power)

λ is treated as an unknown parameter
→ Estimate it from the data!

We proceed as before with unknowns β, σ^2, λ
and Maximize the likelihood.

Step 1: Joint Normal, $y \sim N(X\beta, \sigma^2 I)$

$$\text{PDF:} \quad f(y) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - X_{i\cdot}\beta)^2\right)$$

Step 2: Change of Variables, $y_i \leftarrow y_i^{(\lambda)}$

- $\dfrac{dy_i^{(\lambda)}}{dy_i} = y_i^{\lambda - 1}$ (Note $\dfrac{dy_i^{(\lambda)}}{dy_j} = 0$ if $i \neq j$)
- Jacobian $= \prod_{i=1}^{n} y_i^{\lambda - 1}$

of y_i:

- Jacobian $= \prod_{i=1}^{n} y_i^{\lambda-1}$

New PDF: $f(y_i^{(\lambda)}) = (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i^{(\lambda)} - X_{i\cdot}\beta)^2\right)\prod_{i=1}^{n} y_i^{\lambda-1}$

Step 3: Maximize the likelihood

$$\log(L(\beta,\sigma^2,\lambda)) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\underbrace{(y_i^{(\lambda)} - X_{i\cdot}\beta)^2}_{\text{residual Sum of Squares}} + \underbrace{(\lambda-1)\sum \log(y_i)}_{\text{only depends on }\lambda}$$

if we optimize in σ^2 and β

we set
- $\hat\beta = (X^TX)^{-1}X^T Y^{(\lambda)}$
- $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i^{(\lambda)} - X_{i\cdot}\beta)^2 = \frac{SS_{res}^{(\lambda)}}{n}$ ← MLE for σ^2. Not the unbiased estimator, $\frac{SS_{res}}{n-p-1}$

Step 4: plug in $\hat\beta$, $\hat\sigma^2$ and optimize for λ

$$\log L = -\frac{n}{2}\log(2\pi\hat\sigma^2) - \frac{n}{2} + (\lambda-1)\sum\log y_i$$

$$= C - \frac{n}{2}\log\hat\sigma^2 + \log\left(\left(\prod y_i\right)^{\lambda-1}\right)$$

Geometric Mean: $\gamma = \left(\prod_{i=1}^{n} y_i\right)^{1/n}$

$$= C - \frac{n}{2}\log\hat\sigma^2 + \frac{n}{2}\log\left(\gamma^{2(\lambda-1)}\right)$$

$$= C - \frac{n}{2}\log\left(\frac{\hat\sigma^2}{\gamma^{2(\lambda-1)}}\right)$$ ← log likelihood optimize over λ

Step 5: Optimize over λ

- pick the λ that, Maximizes the log Likelihood, which is the same as Minimizing the SS_{res}

$$\frac{Y^{(\lambda)}}{\gamma^{\lambda-1}} = X\theta + \varepsilon$$

- Use a computer to get $\hat\lambda$
  - $\hat\lambda$ is an estimator for λ, so we can do things like get a confidence int for λ.

e.g. we could test the Hypothesis

$H_0: \lambda = 1$ ← Don't transform / My data looks linear already

$H_1: \lambda \neq 1$ ← Do transform.