§2.3.2 Box-Cox family of transformations [ACD p. 389]

Suppose a normal linear model applies not to $y$, but to

$$
y^{(\gamma)} = \begin{cases} 
\frac{y^{\gamma-1}}{\gamma} & \gamma \neq 0 \\
\log y & \gamma = 0
\end{cases}
$$

As $\gamma$ varies from $-2$ to $2$ say, this encompasses

- reciprocal trans. \hspace{1cm} $\gamma = -1$
- log \hspace{1cm} $\gamma = 0$
- square root \hspace{1cm} $\gamma = \frac{1}{2}$
- original scale \hspace{1cm} $\gamma = 1$
- square \hspace{1cm} $\gamma = 2$

Assume $y_j > 0$; otherwise, apply trans. to $y_j + 5$ with 5 chosen large enough to make all $y_j + 5$ pos.

Assume a normal linear model:

$$y^{(\gamma)} = X\beta + \varepsilon \hspace{1cm} \text{Var}(\varepsilon) = \sigma^2$$

The Jacobian of the trans. from $y_j^{(\gamma)}$ to $y_j$ is

$$\frac{dy_j}{dy_j^{(\gamma)}} = y_j^{\gamma-1}$$

$$\Rightarrow f(y_j; \beta, \sigma^2, \gamma) = \frac{y_j^{\gamma-1}}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_j^{(\gamma)} - x_j^T\beta)^2 \right\}$$
and log lik for $y_1, \ldots, y_n$ is
\[
 l(\beta, \sigma^2, \eta) \propto -\frac{1}{2} \left\{ n \log \sigma^2 + \frac{1}{\sigma^2} \sum_{j=1}^{n} (y_j \eta - x_j^T \beta)^2 \right\} \\
+ (\eta - 1) \sum_{j=1}^{n} \log y_j .
\]

If $\eta$ is regarded as fixed, then
\[
 \hat{\beta}_\eta = (X^T X)^{-1} X^T y(\eta), \quad \sigma^2_\eta = \frac{\text{RSS}(\hat{\beta}_\eta)}{n}.
\]

$\Rightarrow$ profile log lik for $\eta$ is
\[
 l_p(\eta) \propto -\frac{n}{2} \left\{ \log \text{RSS}(\hat{\beta}_{\eta}) - \log g^2(\eta - 1) \right\}
\]
where $g = \left( \prod_{j=1}^{n} y_j \right)^{1/n}$ is the geometric mean of $y_1, \ldots, y_n$.

(\text{check details!})

Obtain an approx. 95\% confidence interval for $\eta$.

Recall LR test:
\[
 2 \left\{ l(\hat{\eta}) - l(\eta_0) \right\} \sim \chi^2_1 \quad \text{under H}_0 .
\]

Retain all $\eta_0$ within critical region
\[
 2 \left\{ l(\hat{\eta}) - l(\eta_0) \right\} \leq C_1 .
\]
$c_1 = 3.84$ for 5% critical point of $X^2$.

$\Rightarrow \ell(\hat{\lambda}) - \ell(\lambda_0) \leq \frac{c_1}{2}$ would be retained by LR test, approx.

\[ \ell(\lambda) \]

\[ l_{max} \]

\[ L \rightarrow \hat{\lambda} \rightarrow \lambda \rightarrow U \]

\[ \{1.92 \text{ drop} \]

\[ \text{approx. 95\% c.i. for } \lambda \]

Remarks:

1. Box & Cox suggest using the prof. lik. fact. for the largest linear model to be considered as guide in choosing $\lambda$, which is then fixed at that value.

2. Exact value of $\lambda$ rarely used - use nearby easily interpreted value.

3. Usually choose transformation for convenience, or based on experience, or on external information.
Case Study 2: Quine study

See handout from Venables & Ripley, Ch. 6.

- An unbalanced factorial design: 4 factors,
- 3 x 2 levels, 10 x 4 levels => $2^3 \times 4 = 32$ cells.
- "Full" model is: \( \text{eth} \times \text{sex} \times \text{age} \times \text{lnr} \)
- Note empty cells - no slow learners in Form 3.
- Diagnosic plots \( \Rightarrow \) variance increases with mean
- \( \Rightarrow \) multiplicative relationship, e.g. \( \sigma_i^2 = \beta \mu_i \sigma_o^2 \)

where \( \sigma_o^2 = \) baseline variation, \( \beta \) captures systematic dependency present \( \Rightarrow \) log trans. of data may be appropriate.

Note that variance of st. dev. also increases with increasing mean ("fan" shape).

Transformation: do either formally (using Box-Cox; see library (MASS) in R) or informally.

Initial analysis fixing \( \alpha = 1 \) doesn't work very well; c.f. for \( x \) does not contain 0.

\( \Rightarrow \) V&R then fit "displaced log" transformation to model
\log(y+\alpha) \text{ and estimate } \alpha \text{ via profile lik}

\Rightarrow \hat{\alpha} = 2.5 \text{; } 95\% \text{ c.c. excludes } \alpha = 1.

\Rightarrow \text{add } 2.5 \text{ to each response in days, then log response, then fit model.}


(1.) There are marginality restrictions: if interaction \text{a*b retained in model, so must main effects a,b be.}

(2.) Because design is unbalanced, order of fitting matters, and analysis is done sequentially.

(Without balance, do not get orthogonal decomposition of SS and the least squares estimates of parameters differ according to which terms are already included in the model.)

(3.) p 174. Best to use backward selection or stepwise selection. V&R start with full model, then drop 4\text{th}-order interaction term. Then proceed by sequentially dropping non-marginal terms using