Statistics 191: Introduction to Applied Statistics

Weighted Least Squares, Transformations

Jonathan Taylor
Department of Statistics
Stanford University

February 22, 2010

Outline

Today's class

- Transformations to achieve linearity.
- Transformations to stabilize variance.
- Correcting for unequal variance: weighted least squares.

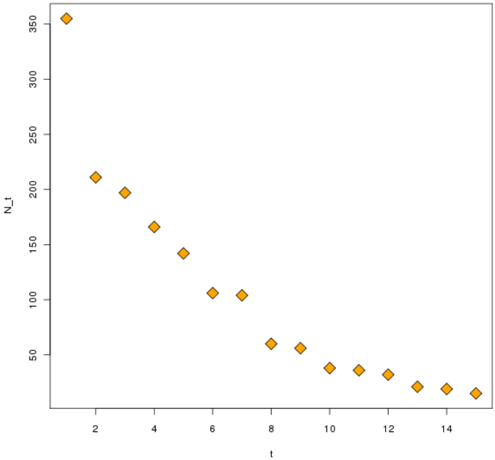
Transformations

Jonathan Taylor
Department of Statistics
Stanford University

Transformations to achieve linearity

- We have been working with *linear* regression models so far in the course.
- Many models are nonlinear, but can be *transformed* to a linear model.

Bacteria death



Transformations

Exponential growth model

- Suppose the expected number of cells grows like

$$E(n_t) = n_0 e^{\beta_1 t}, \qquad t = 1, 2, 3, \dots$$

- If we take logs of both sides

$$\log E(n_t) = \log n_0 + \beta_1 t.$$

- (Reasonable ?) model:

$$\log n_t = n_0 + \beta_1 t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \text{ independent}$$

Transformations

Logarithmic transformation

- This is slightly different than original model:

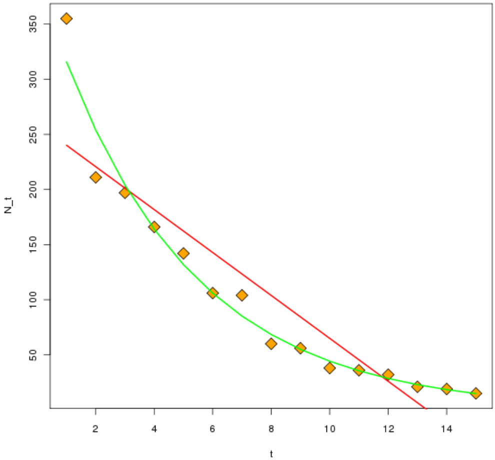
$$E(\log n_t) \leq \log E(n_t)$$

  but may be approximately true.
- If $\varepsilon_t \sim N(0, \sigma^2)$ then

$$n_t = n_0 \cdot \epsilon_t \cdot e^{\beta_1 t}.$$

- $\epsilon_t = e^{\varepsilon_t}$ is called a log-normal random $(0, \sigma^2)$ random variable.

Bacteria death, fitted values



Transformations

Linearizing regression function

Some models that can be linearized:

- $y = \alpha x^{\beta}$, use $\tilde{y} = \log(y), \tilde{x} = \log(x)$;
- $y = \alpha e^{\beta x}$, use $\tilde{y} = \log(y)$;
- $y = x/(\alpha x - \beta)$, use $\tilde{y} = 1/y, \tilde{x} = 1/x$.
- More in textbook.

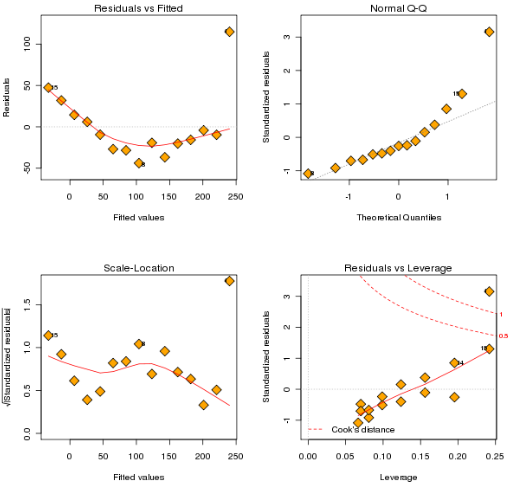
Transformations

Jonathan Taylor
Department of Statistics
Stanford University

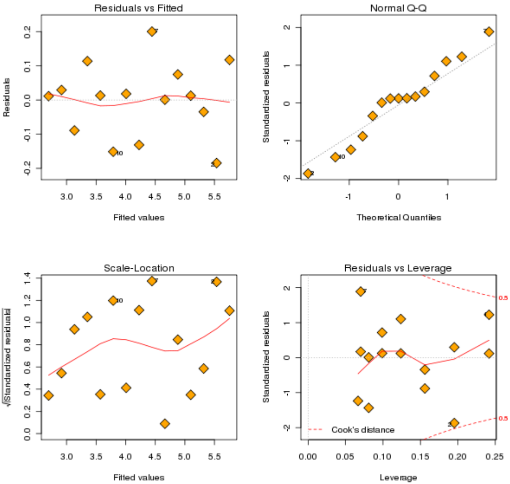
Caveats

- Just because expected value linearizes, doesn't mean that the errors behave correctly.
- In some cases, this can be corrected using weighted least squares (more later).
- Constant variance, normality assumptions should still be checked.

Bacteria death, untransformed



Bacteria death, transformed



Stabilizing variance

- Sometimes, a transformation can turn non-constant variance errors to "close to" constant variance.
- Example: by the "delta rule" (see next slide), if

$$Var(Y) = \sigma^2 E(Y)$$

then

$$\mathrm{Var}(\sqrt{Y}) \simeq \frac{\sigma^2}{4}.$$

Transformations

Delta rule

- Very important tool in statistics.
- Taylor series expansion:

$$f(Y) = f(E(Y)) + \dot{f}(E(Y))(Y - E(Y)) + \dots$$

- $$\mathrm{Var}(f(Y)) \simeq \dot{f}(E(Y))^2 \mathrm{Var}(Y)$$

- Previous example:

$$\mathrm{Var}(\sqrt{Y}) \simeq \frac{\mathrm{Var}(Y)}{4E(Y)}$$

Transformations

Caveats

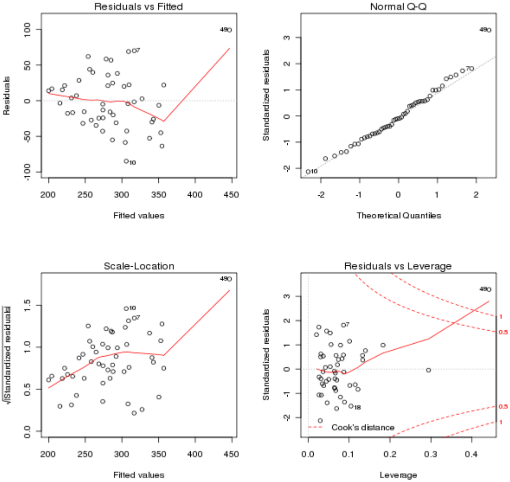
- Just because a transformation makes variance constant doesn't mean regression function is still linear (or even that it was linear)!
- The models are approximations, and once a model is selected our standard "diagnostics" should be used to assess adequacy of fit.
- It is possible to have non-constant variance but have the variance stabilizing transformation may destroy linearity of the regression function. *Solution:* try weighted least squares (WLS).

Correcting for unequal variance

Jonathan Taylor
Department of Statistics
Stanford University

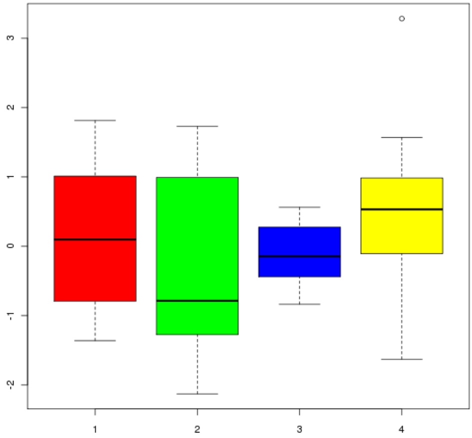
Example: education expenditure in 1975

- Variables:
  1. Y – per capita education expenditure by state
  2. X_1 – per capita income in 1973 by state
  3. X_2 – proportion of population under 18
  4. X_3 – proportion in urban areas
  5. `Region` – which region of the country are the states located in ?
- Hypothesis: weights vary by Region: i.e. variability of expenditure varies by rough geographic region.


Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

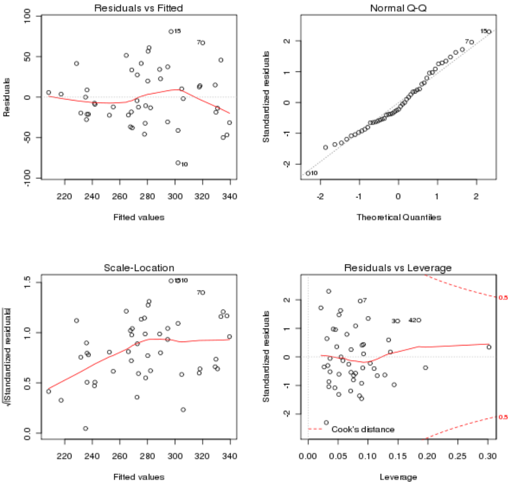
Boxplot of residuals

Jonathan Taylor
Department of Statistics
Stanford University

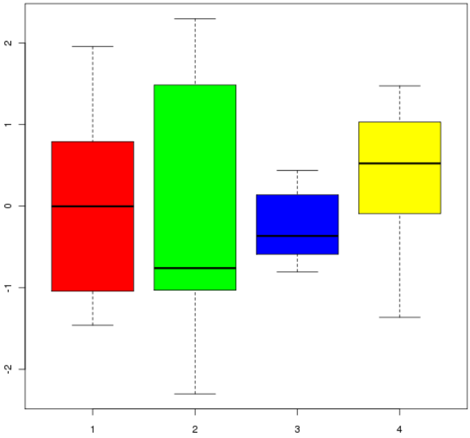


Education expenditure, AK removed

Jonathan Taylor
Department of Statistics
Stanford University



Boxplot of residuals, AK removed



Correcting for unequal variance

Reweighting

- If you have a reasonable guess of variance as a function of the predictors, you can use this to "reweight" the data.
- Hypothetical example

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2 X_i^2).$$

- Setting $\tilde{Y}_i = Y_i / X_i$, $\tilde{X}_i = 1/X_i$, model becomes

$$\tilde{Y}_i = \beta_0 \tilde{X}_i + \beta_1 + \epsilon_i, \epsilon_i \sim N(0, \sigma^2).$$

Correcting for unequal variance

Jonathan Taylor
Department of Statistics
Stanford University

Weighted Least Squares

- Fitting this model is equivalent to minimizing

$$\sum_{i=1}^{n} \frac{1}{X_i}^2 \left(Y_i - \beta_0 - \beta_1 X_i\right)^2$$

- Weighted Least Squares

$$SSE(\beta, w) = \sum_{i=1}^{n} w_i \left(Y_i - \beta_0 - \beta_1 X_i\right)^2, \qquad w_i = \frac{1}{X_i^2}.$$

- In general, weights should be like:

$$w_i = \frac{1}{\mathrm{Var}(\varepsilon_i)}.$$

Correcting for unequal variance

Jonathan Taylor
Department of Statistics
Stanford University

Common weighting "schemes"

- If you have a qualitative variable, then it is easy to estimate weight within groups (our example today).
- "Often"
$$\text{Var}(\varepsilon_i) = \text{Var}(Y_i) = V(E(Y_i))$$
- Many non-Gaussian models behave like this: logistic, Poisson regression – upcoming lectures.

Correcting for unequal variance

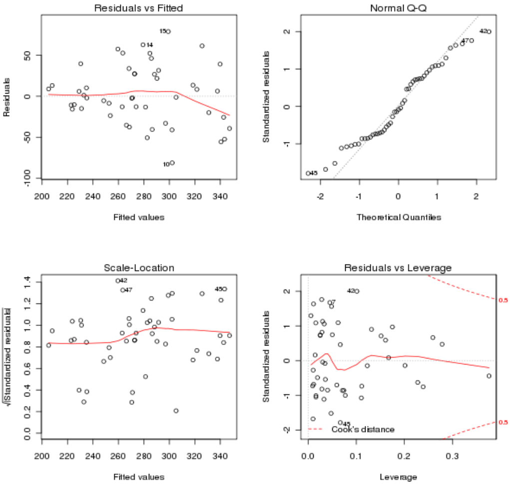
Two stage procedure

- Suppose we have a hypothesis about the weights, i.e. they are constant within `Region`, or they are something like

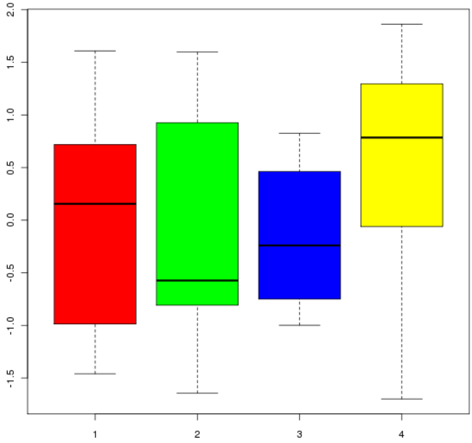
$$w_i \beta_0 + \beta_1 X_{i1}^2.$$

- As in models with autocorrelation, we pre-whiten:
  1. Fit model using OLS (Ordinary Least Squares) to get initial estimate $\widehat{\beta}_{OLS}$
  2. Use predicted values from this model to estimate w_i.
  3. Refit model using WLS (Weighted Least Squares).
  4. If needed, iterate previous two steps.

Education expenditure – weighted



Boxplot of residuals – weighted



Correcting for unequal variance

Jonathan Taylor
Department of Statistics
Stanford University

Unequal variance: effects on inference

- So far, we have just mentioned that things *may* have unequal variance, but not thought about how it affects inference.
- In general, if we ignore unequal variance, our estimates of variance are no longer unbiased. The covariance has the "sandwich form"

$$\text{Cov}(\widehat{\beta}) = (X'X)^{-1}(XW^{-1}X)(X'X)^{-1}.$$

with $W = (\sigma_i^2)$.

Correcting for unequal variance

Efficiency

- The efficiency of an unbiased estimator of β is 1 / variance.
- Estimators can be compared by their efficiency: the more efficient, the better.
- The other reason to correct for unequal variance (besides so that we get valid inference) is for efficiency.

Correcting for unequal variance

Jonathan Taylor
Department of Statistics
Stanford University

Efficiency – example

- Suppose

$$Z_i = \mu + \varepsilon_i, \qquad \varepsilon_i \sim N(0, i^2 \cdot \sigma^2), 1 \leq i \leq n.$$

- Two unbiased estimators of μ:

$$\widehat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n Z_i$$

$$\widehat{\mu}_2 = \frac{1}{\sum_{i=1}^n i^{-2}} \sum_{i=1}^n i^{-2} Z_i$$

- The estimator $\widehat{\mu}_2$ will always have lower variance, hence tighter confidence intervals.

Efficiency of estimators

Statistics 191: Introduction to Applied Statistics

Jonathan Taylor
Department of Statistics
Stanford University