THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

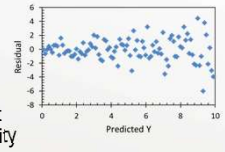
Lecture 26
Correcting for Heteroscedasticity: Stabilizing the Variance

Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Review of Heteroscedasticity



- Heteroscedasticity is often a by-product of other violations of assumptions
  - A misspecified model will almost always result in heteroscedasticity
  - Only if we are convinced this is not the case should we take remedial action against heteroscedasticity
- Result of heteroscedasticity will be an unbiased estimator that is inefficient (larger $SE(b_k)$)
  - Small amounts of heteroscedasticity (variance changing by less than a factor of ~4) don't matter much

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

What to Do if You Detect Heteroscedasticity?

- If you know how the variance changes with each y_i, use a weighted regression
  - More on this coming soon
- For a variance with unknown dependence on y_i, other regression methods can be used (such as GMM, generalized method of moments estimation)
  - We won't get into this
- If you know the general trend of variance change with the predictor variable, transform the data to make it homoscedastic

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Transforming the Data

- Our model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- Let's say we know that $var(\varepsilon_i) = kx_i$
- We can make the variance of the error term constant by diving our model equation by $\sqrt{x_i}$

Requires Multiple Regression

$$\frac{y_i}{\sqrt{x_i}} = \frac{\beta_0}{\sqrt{x_i}} + \beta_1\sqrt{x_i} + \frac{\varepsilon_i}{\sqrt{x_i}} \qquad y_i' = \frac{\beta_0}{\sqrt{x_i}} + \beta_1\sqrt{x_i} + \varepsilon_i'$$

(alternate approach: weighted regression with $w_i = 1/x_i$)

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Transforming the Data

- Logarithm transformation: If y grows exponentially or as a power of x, transform by taking the log
  - Let $y' = \ln(y)$, then use y' as the response
  - For a power law model, use $\ln(x)$ as the regressor too
- Square root transformation: applies to data exhibiting a Poisson distribution (variance = mean)
  - Fitting $\hat{y} \propto x^2$ is not the same as fitting $\hat{y}' = \sqrt{\hat{y}} \propto x$
- We can systematize such transformations using the Box-Cox transformation approach

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Box-Cox Transformation

- If we are willing to change the model, find the best power transformation that fits the data:

$$y_i^{(\lambda)} = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \ln(y_i) & \lambda = 0 \end{cases}$$

(requires that y > 0 for all data)

  - Use MLE to find the best fit λ, β_0, and β_1
  - Alternately, repeat the OLS for different values of λ and find the one with minimum s_ε
  - We often round the best-fit λ to the nearest multiple of 0.5 and then repeat the regression

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Box-Cox Transformation

- The Box-Cox transformation affects many things at once
  - It changes the model
  - It can change the error distribution (making it closer to or further from normal)
  - It can change heterscedasticity
  - Everything must be checked again after the transformation
- It is often used during exploratory work, at the beginning of model building

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Final Approach to Heteroscedasticity

- If the normal probability distribution with constant variance does not fit, try a different distribution
  - Ex: Gamma distribution for purely positive real numbers (positive support)
- We can do this using a Generalized Linear Model (glm)
  - More on this topic later in the semester

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Lecture 26: What have we learned?

- Under what circumstances is it appropriate to take remedial action against heteroscedasticity?
- What are the two main approaches to addressing heteroscedasticity?
- When should you choose weighted regression vs. data transformation?