Transformations

- A **variance stabilizing transformation** may be useful when the variance of y appears to depend on the value of the regressor variables, or on the mean of y. Table 5.1 lists some commonly used variance stabilizing transformations. For example, if the variance of y is proportional to the mean of y, it is useful to consider transforming to $\sqrt{y}$.
- Why does this work? In general, if y has mean μ_y and variance σ_y^2, then a function $h(y)$ has approximate mean $h(\mu_y)$ and variance approximately equal to $(h'(\mu_y))^2\sigma_y^2$. (Proof of this is given in Stat 4066 and/or 5067, using a Taylor series approximation.)

- Example: Suppose Y has a Poisson distribution with mean $\mu_Y = \mu$ and variance $\sigma_Y^2 = \mu$. let $Z = h(Y) = \sqrt{Y}$. Then $h'(Y) = .5Y^{-1/2}$, so that $h'(\mu) = .5\mu^{-1/2}$, and the variance of Z is approximately $(h'(\mu_y))^2\sigma_y^2 = (.5\mu^{-1/2})^2\mu = .25$. The variance stabilizing transformation for the Poisson distribution is the $\sqrt{}$ transform.
- Regression type models where Y has a Poisson distribution are a subset of generalized linear models. GLM's are the topic of chapter 13 in Montgomery, Peck and Vining (which we will not cover) and are the main topic of Stat 4620. Some of the theory of GLM's is discussed in Stat 4066. The GLM family also includes binomial response variables.

Intrinsically linear models

- Some models are intrinsically linear, and can be appropriately transformed to give a linear relation.
- For example, if $V = kHW$,
- then $log(V) = log(k) + log(H) + log(W)$,
- In Economics, the Cobb-Douglas production function is $P = kL^{\alpha}C^{\gamma}\epsilon$,
- the model $log(P) = \beta_0 + \beta_1 log(L) + \beta_2 log(C) + log(\epsilon)$ may be a useful regression model.
- This will typically assume that $log(\epsilon)$ has a normal distribution.

- In biochemistry, where y is reaction rate and x is substrate concentration, the Michaelis-Menten equation states that

$$y = \frac{V_{max} x}{K_m + x}$$

- V_{max} and K_m are parameters to be estimated.
- Note that as $x \to \infty$, $y \to V_{max}$.
- The Lineweaver-Burk plot, or double reciprocal plot, is a plot of $1/y$ vs $1/x$, provides a convenient means of estimating the two model parameters.

$$\frac{1}{y} = \frac{K_m}{V_{max}} \frac{1}{x} + \frac{1}{V_{max}}$$

- or

$$\frac{1}{y} = \beta_0 + \beta_1 \frac{1}{x}$$

- with $\beta_0 = \frac{1}{V_{max}}$ and $\beta_1 = \frac{K_m}{V_{max}}$
- Find the least squares estimators of β_0 and β_1 and transform to get estimators of K_m and V_{max}.

- Nowadays biochemists fit a nonlinear regression (model is nonlinear in x)

$$y_i = \frac{V_{max} x_i}{K_m + x_i} + \epsilon_i$$

assuming the ϵ_i are a sample from a $N(0, \sigma^2)$ population,

- Nonlinear regression is discussed in Chapter 12 of Montgomery, Peck and Vining. We will not discuss this topic.
- but the Lineweaver-Burk plot is still often included in research papers in order to visualize the linearized relationship, and allow quick ballpark estimates. (ie the x intercept is $-1/K_m$ and the y intercept is $1/V_{max}$)

- If the relationship between y and a regressor appears to be nonlinear, buthe assumptions of i.i.d. $N(0, \sigma^2)$ errors appears to be approximately satisfied, then it may be more appropriate to transform the regressor variable x rather than y.
- The Box-Cox transformation gives a class of transformations which are meant to simultaneously correct for nonnormality and/or nonconstant variance, and are described in section 5.4.1. The response variable of the Box-Cox transformation is

$$\frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}}$$

where $\dot{y}$ is the geometric mean of the observations. You are not responsible for the Box-Cox transformation.

Weighted least squares

- Weighted least squares can be used to estimate the parameters of regression models with nonconstant error variance. This is the topic of section 5.5.1.
- If $\mathbf{y} = \mathbf{X}\beta + \epsilon$, with $E(\epsilon) = \mathbf{0}$ and $Cov(\epsilon) = \sigma^2\mathbf{V}$,
  - the generalized least squares estimator is given by

$$\hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$$

  - with covariance matrix

$$Cov(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

- This works because:
  - the covariance matrix of $V^{-1/2}\epsilon$ is $\sigma^2\mathbf{I}$, where $V^{-1/2}V^{-1/2} = V^{-1}$
  - so we transform the model to

$$(V^{-1/2}\mathbf{y}) = (V^{-1/2}\mathbf{X})\beta + V^{-1/2}\epsilon$$

  - and then carry out least squares using the data $(V^{-1/2}\mathbf{y})$ and $(V^{-1/2}\mathbf{X})$.

Gauss-Markov theorem

The **Gauss-Markov theorem** (appendix C.11) states that $\hat{\boldsymbol{\beta}}$ is the minimum variance unbiased estimator of $\boldsymbol{\beta}$, also known as the best linear unbiased estimator (BLUE).