

# Transformations and Weighting

Dr. Austin Brown

Kennesaw State University

Some of the materials in today's lecture were adapted from those created by Dr. Taasoobshirazi as well as my former professor, Dr. Khalil Shafie (Thank you Drs. S & T!).

- We've learned over the past several weeks that linear regression, while a powerful tool, relies upon assumptions in order for us to reasonably rely upon the inferences we make.
- In last week's class, we talked about visual and testing methods in order to assess the linearity, normality, and constant variance assumptions.
- But we didn't talk about what to do in cases where the assumptions are violated!

- One common and traditional method to attempt to correct these issues is through variable transformation. We have two primary types of transformations:
  - ① **Linearizing Transformations:** Typically transform the explanatory variables to achieve linearity.
  - ② **Variance Stabilizing Transformations:** Typically transform the response variable to, well, stabilize the variance to achieve constant variance or homoskedacity. This assumption also typically goes hand in hand with the normality assumption.

# Variance Stabilizing Transformations

- When our variance/normality assumption is violated, one way to potentially correct it is through a variance stabilizing transformation.
- We do this because on occasion when the response variable actually follows a different distribution than the normal distribution, the variance and the mean are interrelated.
- For example, if  $Y \sim \text{Poisson}(\mu)$ ,  $E[Y] = \text{Var}[Y] = \mu$ . Or if  $Y \sim \text{BIN}(n = 1, p)$ ,  $E[Y] = p$  and  $\text{Var}[Y] = p(1 - p)$ .

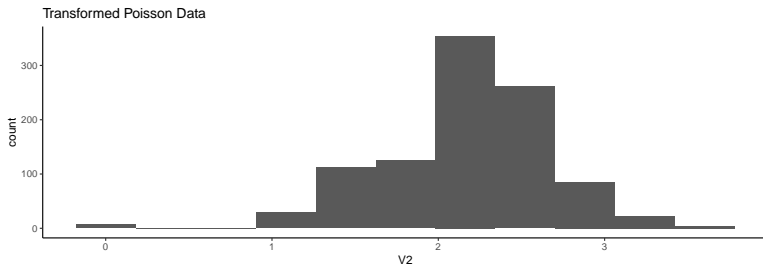
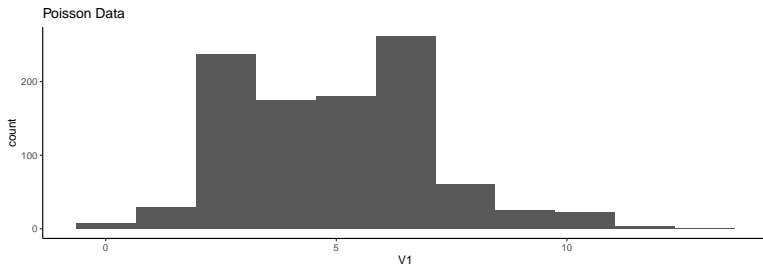
# Variance Stabilizing Transformations

- For regression, this is obviously problematic.
- For us, we assume the mean vector of our response variable,  $\mathbf{X}\beta$ , is independent of the variance,  $\sigma^2\mathbf{I}$ .
- Thus, the logic behind a variance stabilizing transformation is to brute force the response variable into possibly having a mean and variance that are independent of each other, and then running the regression on the transformed response.

# Variance Stabilizing Transformations

- The Montgomery Regression text, Chapter 5, illustrates several common transformations to use on the response if your data follow a specific pattern/distribution. In my experience, only two of these have been particularly useful:
- If your data follow a Poisson distribution, use the square root transformation. If your data are proportions (i.e.,  $0 \leq y_i \leq 1$ ), it may be useful to use the arcsine of the square root of the response transformation ( $y'_i = \sin^{-1}(\sqrt{y_i})$ ).

# Variance Stabilizing Transformations





# Variance Stabilizing Transformations

- Let's do an example of a VST using the Energy dataset on D2L.

# Variance Stabilizing Transformations

- Generally speaking, choosing the right VST is not so cut and dry. Unless our data very clearly follows a particular pattern, it can be hard for us to know which VST to use.
- Fortunately, there is a method we can use called Box-Cox Transformations in which we obtain a value,  $\lambda$ , which we raise our response variable,  $y$ , to.
- The value of  $\lambda$  is obtained through maximum likelihood estimation.

# Variance Stabilizing Transformations

- Let's use the Box-Cox transformation on the Energy data.

- While I would say that VST's are much more common in practice, it may also be necessary sometimes to perform a linearizing transformation, which solves the problem it sounds like it solves.
- Sometimes when we have data that follow a particular functional relationship (e.g., exponential), we can transform the regression equation to linearize it (note, this is actually the same rationale behind general linear models, such as logistic regression or count regression).

# Linearizing Transformations

- For example, suppose we had data which followed this formula:

$$y = \beta_0 e^{\beta_1 x} \varepsilon$$

- We can linearize it by using the natural log:

$$\ln y = \ln \beta_0 + \beta_1 x + \ln \varepsilon$$

- Note, now the assumption of normality and constant variance is on the transformed residuals (which assumes the untransformed residuals must follow the log-normal distribution. . . as you can see, this starts getting hairy quickly).

- Let's go through an example using the Windmill data in D2L.

- Another method for helping stabilize the variance is called “weighted least squares (WLS).”
- Basically, in the estimation process, we apply weights to our predictor variables to help those with smaller error variances have a “louder voice” in the final result, so to speak.
- And conversely, we put less weight on those with greater error variances so that aren’t overly detrimental.

- We'll define an  $n \times n$  weighting matrix,  $W$  (which has to be positive definite and invertible). Our normal equation is now:

$$(X^T W X) \hat{\beta} = X^T W Y$$

- Now, if  $X$  is invertible, our Weighted Least Squares estimators are:

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y$$



# Weighted Least Squares

- One issue with WLS is that we have to know  $W$ . Often, we rely on past experience or contextual information in order to help us.
- If we don't, it can sort of be a game of guess and check.
- Analyzing the residuals (such as plotting each regressor against the residuals) can be a good place to start.
- I would say most of the time, we're looking for a discernible relationship between the residuals and the explanatory variables. Then our weights are just the inverse of the identified relationship.
  - This is for sure not an exact science.