Lecture 3: ANOVA Diagnostics and Remedial Measures

STA 106: Analysis of Variance

Suggested reading: ALSM Chapter 18

Xiner Zhou
Department of Statistics, University of California, Davis

ANOVA Diagnostics and Remedial Measures

- ANOVA Model Assumptions
- Check assumptions
- Remedies and concerns about assumption violations

Remedies

ANOVA Model Assumptions

- Independence of error terms
  - Independent random samples are obtained
  - would be violated, e.g. clustered within family or geographic region or time or other factors that give rise to similarity among units
- constancy or homogeneity of error variance
- Normality of error terms

often come together, detect violations using residuals

constancy or homogeneity of error variance

Normality of error variance

Solution for situation (1):

Regression with weighted least squares

Normality of error variance

Normality and constancy of error variance often come hand in hand

Solution for situation (2):

Transform Y's

Successful

If still not Successful (3):

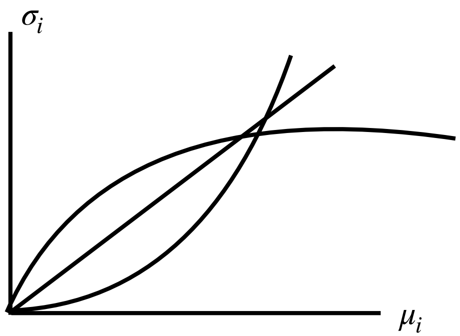
Nonparametric ANOVA

Remedies: Transform Y

Solution for (2): Transform Y's

- brings the distribution of error terms more close to normal
- Equalize the error variances across all factor levels

Solutions
- Practical Guide
- Box-Cox Transformation



Variance σ^2 increase when the mean μ increase

$\sigma_i \approx c\sqrt{\mu}_i$ $\quad Y' = \sqrt{Y}$ square root transformation

$\sigma_i \approx c\mu_i$ $\quad Y' = log_e Y$ log- transformation

$\sigma_i \approx c\mu_i^2$ $\quad Y' = \frac{1}{Y}$ reciprocal transformation

Plot these against factor levels:

$$\frac{s_i^2}{\bar{Y}_{i\cdot}} = c$$

$$\frac{s_i}{\bar{Y}_{i\cdot}} = c$$

$$\frac{s_i}{\bar{Y}_{i\cdot}^{\ 2}} = c$$

Approximate constancy of one of the three would suggest useful transformation, it will stabilize error variance and make the error distribution more close to normal.

Remedies: Transform Y

Solution for (2): Transform Y's

- brings the distribution of error terms more close to normal
- Equalize the error variances across all factor levels

Solutions
- Practical Guide
- Box-Cox Transformation

Box-Cox Transformation :

Uses power transformation of the data Y

Is a general data transformation technique (even outside ANOVA) used to stabilize variance and make the data more normally distributed

$$Y' = Y^{\lambda}$$

λ : a parameter to be determined by data using MLE for $Y_i^{\lambda} = \beta_1 + \beta_2 X_i + \epsilon_{ij}$

$$\lambda = 2 \quad Y' = Y^2$$

$$\lambda = .5 \quad Y' = \sqrt{Y}$$

$$\lambda = 0 \quad Y' = \log_e Y$$

$$\lambda = -.5 \quad Y' = \frac{1}{\sqrt{Y}}$$

$$\lambda = -1.0 \quad Y' = \frac{1}{Y}$$

Box-Cox is used only to provide a guide for selection a transformation, so overly precise results are not needed.

Choose a meaningful value of λ around the exact result provided by MLE.

Remedies: Transform Y

Solution for (2): Transform Y's

- brings the distribution of error terms more close to normal
- Equalize the error variances across all factor levels

Solutions
- Practical Guide
- Box-Cox Transformation

When such a transformation appropriately solve the non-constancy of error variance and normality, once can work with the transformed data for testing the equality of factor level means.

However, it is desirable present the results for estimating factor level effects in the original variable Y, for easier understanding of the meaning and significance of the results.

Effects of Departures form ANOVA Model Assumptions

What are the effects of any remaining departures from model assumptions?

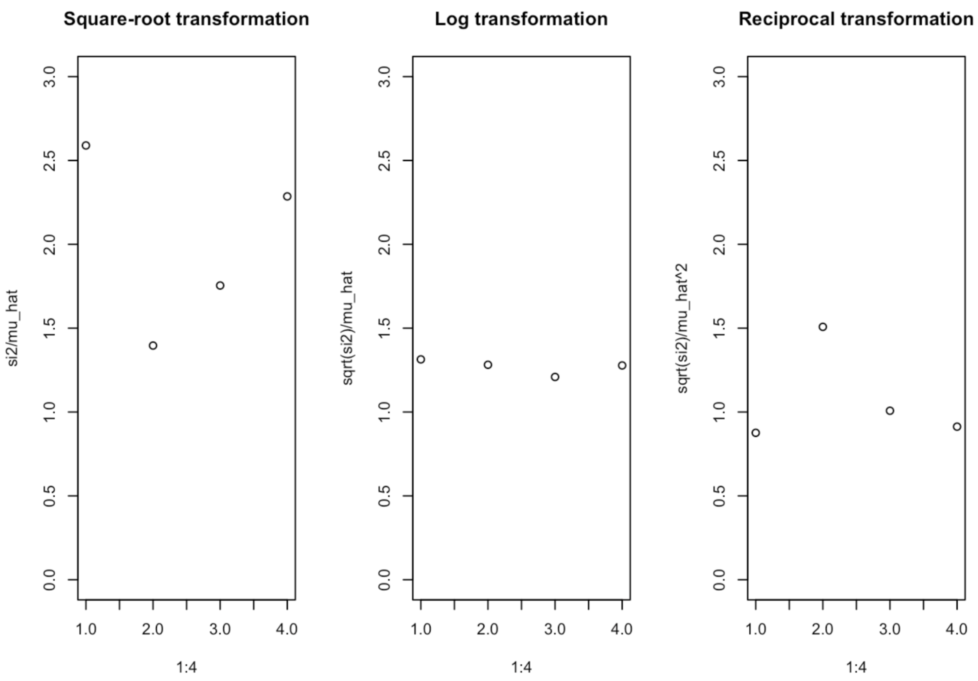
- **Nonnormality**
  - lack of normality is not a serious problem, as long as it's not extreme
- **Unequal Error Variances**
  - Inference based on F test (F test for equality of factor level means, Sheffe multiple comparision), is not affected substantially, but other inference procedures do
- **Non-independence of Error Terms**
  - Serious effects on inference
  - important to prevent in the design stage
  - more advanced modelling technique can rescue

Example

Regardless of your conclusions, practice with transformations.

For each shift. calculate $\bar{Y}_i$, and s_i.

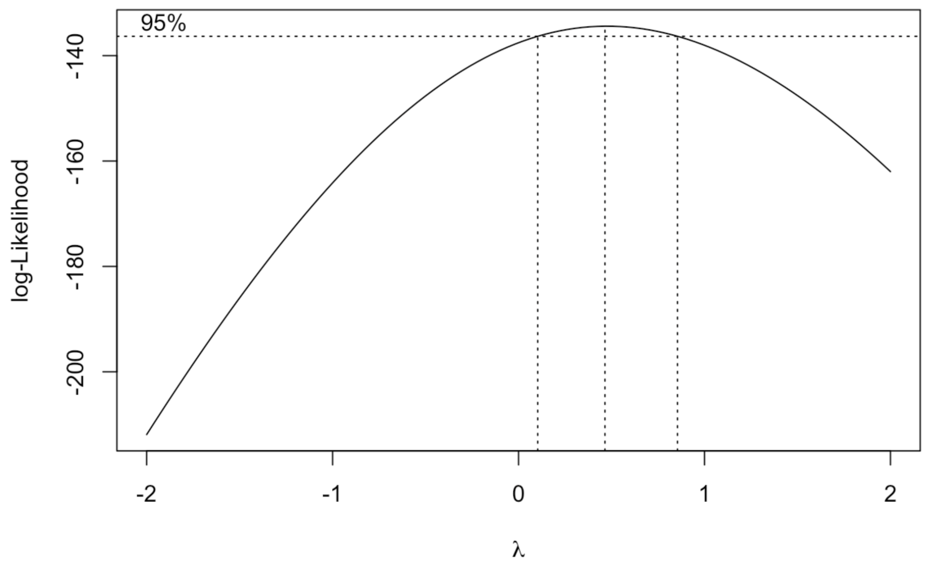
Examine the relation and determine the transformation that is most appropriate here. What do you conclude?



The plot suggests that log transformation of Y is appropriate.

Example

Use the Box-Cox procedure to find an appropriate power transformation of Y.
first adding the constant 1 to each Y observation. Does $\lambda = .5$a square-rocst transformation appear to be reasonable based on the Box-Cox procedure?

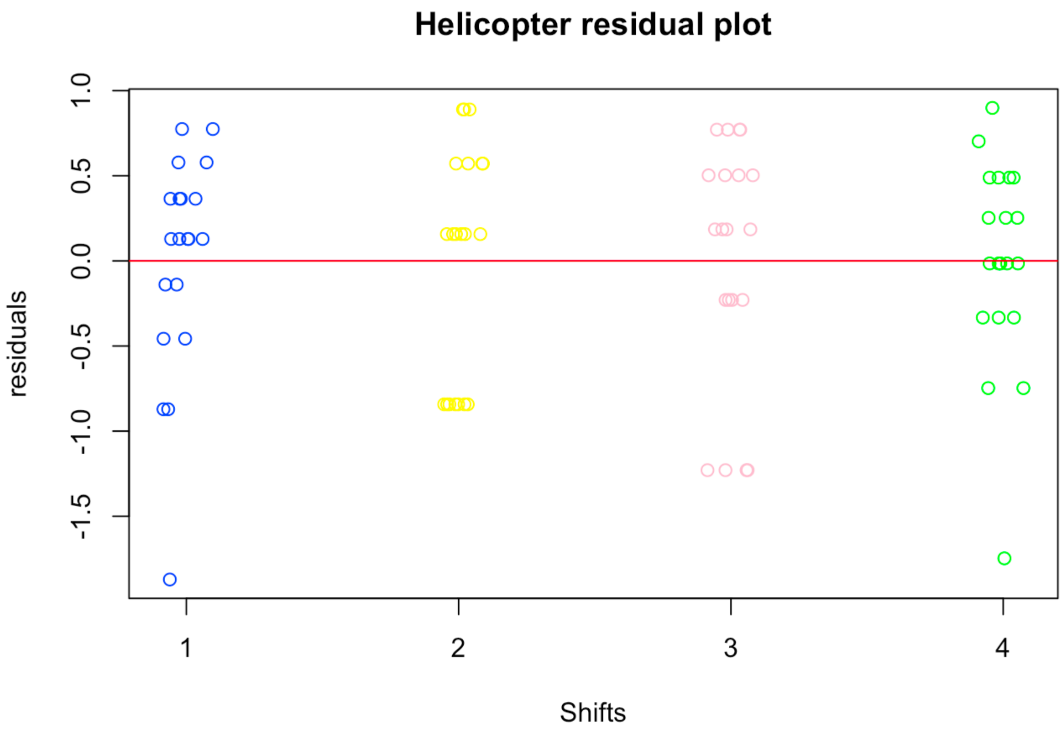


Box-Cox is used only to provide a guide for selection of transformation, so overly precise result is not needed.
Choose a meaningful value around the MLE.

In this case, the likelihood function is maximized around 0.5, it corresponds to a square-root transformation, although log transformation ($\lambda = 0$) is not bad choice either.

Example

The analyst decided to apply the square root transformation $Y' = \sqrt{Y}$ and examine its effectiveness. Obtain the transformed response data. fit ANOVA model. and obtain the residuals.



Example

Test whether the ANOVA model assumptions are met in the transformed data.
What are your findings about the effectiveness of the transformation?

```
##
##  Hartley's maximum F-ratio test of homogeneity of variances
##
## data:  y by i
## F Max = 1.4067, df = 19, k = 4, p-value = 0.8832
```

Code

```
## [1] 0.8832147
```

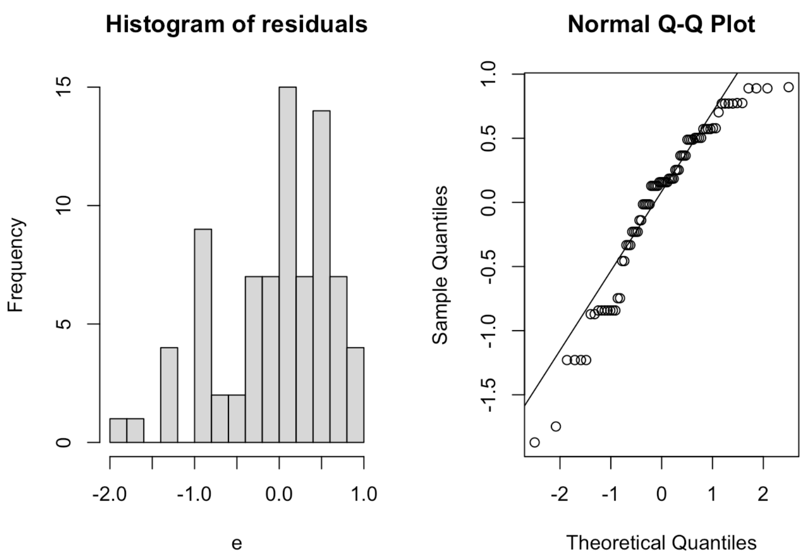
Code

```
## [1] 0.3862421
```

Code

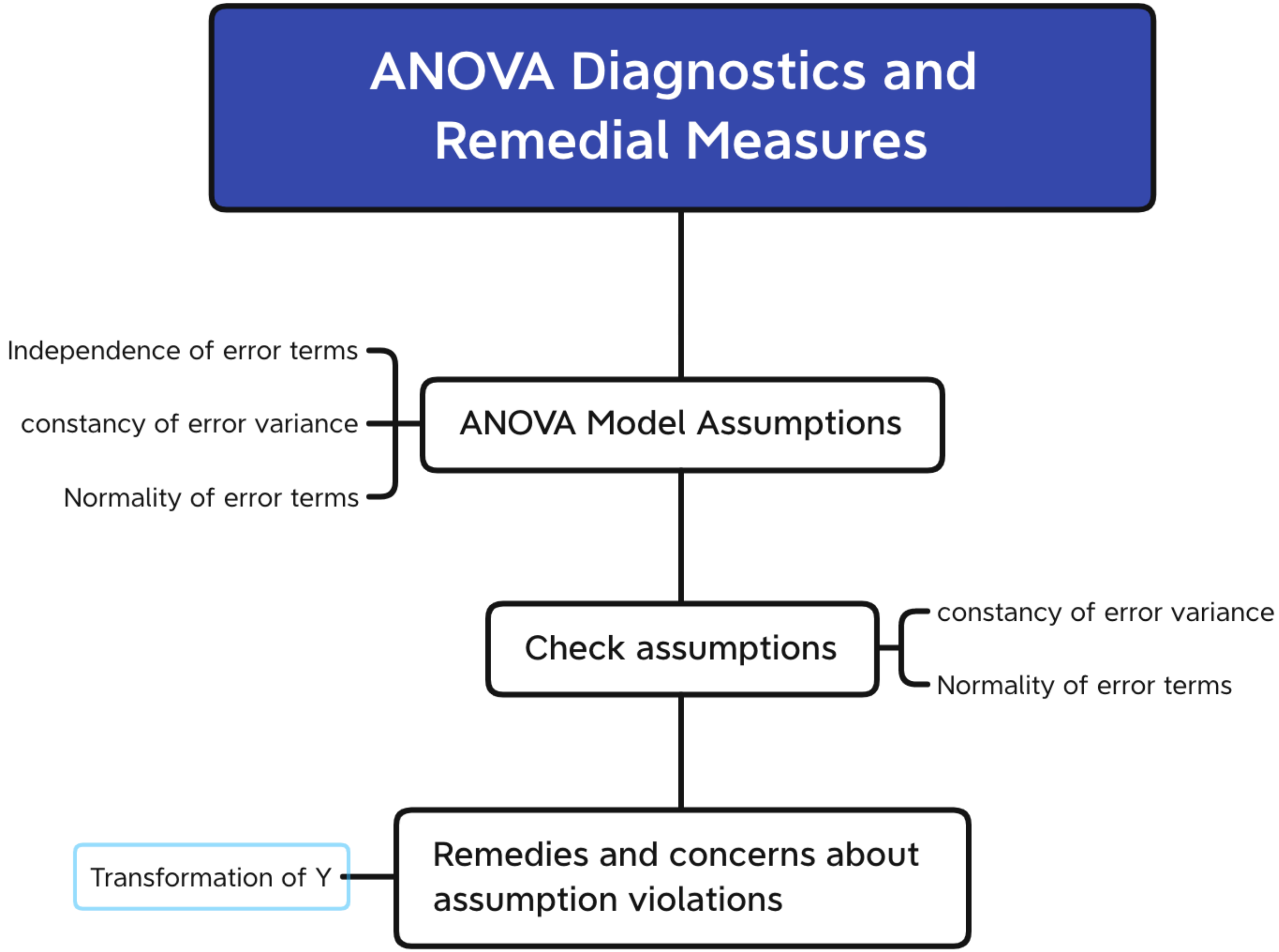
```
## [1] 0.7632153
```

Both the Hartley and Brown-Forsythe test do not reject the null with large p-values, therefore we can conclude that the equal variance assumption is not violated.



The QQ-plot suggests that the transformed data might be a little left-skewed. But since normality is not a serious concern for ANOVA model, we can proceed analysis with the square-root transformed data.

Summary


ANOVA Diagnostics and Remedial Measures
Independence of error terms
constancy of error variance
Normality of error terms
ANOVA Model Assumptions
Check assumptions
constancy of error variance
Normality of error terms
Transformation of Y
Remedies and concerns about assumption violations