2.12 Tests for Homogeneity of Variance

- In an ANOVA, one assumption is the **homogeneity of variance (HOV)** assumption. That is, in an ANOVA we assume that treatment variances are equal:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2.$$

- Moderate deviations from the assumption of equal variances do not seriously affect the results in the ANOVA. Therefore, the ANOVA is robust to small deviations from the HOV assumption. We only need to be concerned about large deviations from the HOV assumption.
- Evidence of a large heterogeneity of variance problem is easy to detect in residual plots. Residual plots also provide information about patterns among the variance.
- Some researchers like to perform a hypothesis test to validate the HOV assumption. We will consider three common HOV tests: Bartlett's Test, Levene's Test, and the Brown-Forsythe Test.
- These tests are not powerful for detecting small or moderate differences in variances. This is okay because we are only concerned about large deviations from the HOV assumption.

2.12.1 Bartlett's Test

- To perform Bartlett's Test:
  1. Calculate $U = \dfrac{1}{C}\left[\nu \ln(s_p^2) - \sum_{i=1} a\nu_i \ln(s_i^2)\right]$ where

     $$s_p^2 = \frac{\sum_{i=1}^a \nu_i s_i^2}{\nu}, \qquad \nu_i = n_i - 1, \qquad \nu = \sum_{i=1}^a \nu_i, \qquad C = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^a \frac{1}{\nu_i} - \frac{1}{\nu}\right).$$

     Note: for a oneway ANOVA, $s_p^2 = MSE$ and $\nu = N - a$.
  2. Reject $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$ if $U > \chi^2(\alpha, a-1)$.
- Bartlett's Test is the uniformly most powerful (UMP) test for the homogeneity of variances problem *under the assumption that each treatment population is normally distributed.*
- Bartlett's Test has serious weaknesses if the normality assumption is not met.
  - The test's reliability is sensitive (not robust) to non-normality.
  - If the treatment populations are not approximately normal, the true significance level can be very different from the nominal significance level (say, $\alpha = .05$). This difference depends on the kurtosis (4th moment) of the distribution.
    - The true significance level will be smaller than the nominal level for a distribution with negative kurtosis (such as a uniform distribution).
    - The true significance level will be larger than the nominal level for a distribution with positive kurtosis (such as a double exponential distribution).
- Because of these problems, many statisticians do not recommend its use. They recommend Levene's Test (or the Brown-Forsythe Test) because these tests are not very sensitive to departures from normality.

2.12.2 Levene's Test

- To perform Levene's Test:
  1. Calculate each $z_{ij} = |y_{ij} - \overline{y}_{i\cdot}|$.
  2. Run an ANOVA on the set of z_{ij} values.
  3. If p-value $\leq \alpha$, reject H_o and conclude the variances are not all equal.
- Levene's Test is robust because the true significance level is very close to the nominal significance level for a large variety of distributions.
- It is not sensitive to symmetric heavy-tailed distributions (such as the double exponential and student's t distributions).

2.12.3 Brown-Forsythe Test

- To perform the Brown-Forsythe Test:
  1. Calculate each $z_{ij} = |y_{ij} - \widetilde{y}_i|$ where $\widetilde{y}_i$ is the median for the i^{th} treatment.
  2. Run an ANOVA on the set of z_{ij}'s.
  3. If p-value $\leq \alpha$, reject H_o and conclude the variances are not all equal.
- The Brown-Forsythe Test is relatively insensitive to departures from normality.
- It is not sensitive to skewed distributions (e.g., χ^2) and extremely heavy-tailed distributions (e.g., Cauchy). In these cases, it is more robust than Levene's Test.

2.12.4 Example of Bartlett's, Levene's, and Brown-Forsythe Tests

A textile company has five looms that weave cloth. The company is concerned that there may be significant variability in the strengths of the cloth produces by the looms. Five random samples of cloth are taken from the cloth produced by each loom. Each sample is tested and the strength is recorded. The data are:

		Loom		
1	2	3	4	5
14.0	13.9	14.1	13.6	13.8
14.1	13.8	14.2	13.8	13.6
14.2	13.9	14.1	14.0	13.9
14.0	14.0	14.0	13.9	13.8
14.1	14.0	13.9	13.7	14.0

SAS Output for HOV Tests

The GLM Procedure

Level of loom	N	cloth Mean	cloth Std Dev
1	5	14.0800000	0.08366600
2	5	13.9200000	0.08366600
3	5	14.0600000	0.11401754
4	5	13.8000000	0.15811388
5	5	13.8200000	0.14832397

The GLM Procedure

Dependent Variable: cloth

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.34160000	0.08540000	5.77	0.0030
Error	20	0.29600000	0.01480000		
Corrected Total	24	0.63760000			

R-Square	Coeff Var	Root MSE	cloth Mean
0.535759	0.872957	0.121655	13.93600

Source	DF	Type III SS	Mean Square	F Value	Pr > F
loom	4	0.34160000	0.08540000	5.77	0.0030

The GLM Procedure

Bartlett's Test for Homogeneity of cloth Variance			
Source	DF	Chi-Square	Pr > ChiSq
loom	4	2.5689	0.6323

The GLM Procedure

Levene's Test for Homogeneity of cloth Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
loom	4	0.0122	0.00304	0.67	0.6179
Error	20	0.0902	0.00451		

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of cloth Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
loom	4	0.0136	0.00340	0.57	0.6897
Error	20	0.1200	0.00600		

- From the following analysis in SAS, the p-values for Bartlett's Test, Levene's Test, and the Brown-Forsythe are .6323, .6179, and .6897, respectively.
- Therefore, we would **fail to reject** $\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$. Therefore, the HOV assumptions is reasonably met for the oneway ANOVA.
- And, assuming there are no serious violations of any other assumptions, we would reject H_0 : for the oneway ANOVA.

SAS Code for HOV Tests

```
DM 'LOG; CLEAR; OUT; CLEAR;';

ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\HOVTEST.PDF';
OPTIONS NODATE NONUMBER;

*************************************************;
*** 5 Looms,    Response = Cloth Output,   n=5 ***;
*** Bartlett's, Brown-Forsythe, Levene's Tests ***;
*************************************************;

DATA in; INPUT loom cloth @@; CARDS;
1 14.0 1 14.1 1 14.2 1 14.0 1 14.1
2 13.9 2 13.8 2 13.9 2 14.0 2 14.0
3 14.1 3 14.2 3 14.1 3 14.0 3 13.9
4 13.6 4 13.8 4 14.0 4 13.9 4 13.7
5 13.8 5 13.6 5 13.9 5 13.8 5 14.0
;
PROC GLM DATA=in;
     CLASS loom;
     MODEL cloth = loom / ss3 ;
     MEANS loom / HOVTEST=BARTLETT;
     MEANS loom / HOVTEST=BF;
     MEANS loom / HOVTEST=LEVENE(TYPE=ABS);

ODS GRAPHICS OFF;
RUN;
```

2.12.5 Data Analysis Options When the HOV Assumption is Not Valid

- If we reject $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$, then what options do we have to analyze the data? We will consider the following two options:

  1. Weighted least squares.
  2. Using a variance stabilizing transformation.

2.13 Weighted Least Squares

- Linear regression models (such as the models used in this course) that have a non-constant variance structure *(heterogeneity of variance)* can be fitted by the **weighted least squares (WLS)** method.
- With the WLS method, the squared deviation between the observed data value and the predicted value $(y_i - \widehat{y}_i)^2$ is multiplied by a weight w_i. This weight is inversely proportional to the variance of y_i.
- For simple linear regression, the WLS function is $W(\beta_0, \beta_1) \;=$

  To find the least squares normal equations, simultaneously solve $\partial W/\partial\beta_0 = 0$ and $\partial W/\partial\beta_1 = 0$.

  The WLS normal equations are:

$$\begin{aligned}\sum_{i=1}^{n} w_i y_i &= \widehat{\beta}_0 \sum_{i=1}^{n} w_i \,+\, \widehat{\beta}_1 \sum_{i=1}^{n} w_i x_i \\ \sum_{i=1}^{n} w_i x_i y_i &= \widehat{\beta}_0 \sum_{i=1}^{n} w_i x_i \,+\, \widehat{\beta}_1 \sum_{i=1}^{n} w_i x_i^2\end{aligned}$$

  The solution $\widehat{\beta}_0$ and $\widehat{\beta}_1$ to these equations are the WLS solutions.
- In some cases, the weights are known. For example, if an observed y_i is actually the mean on n_i observations and assuming the original observations comprising the mean have constant variance σ^2, then the variance of y_i is σ^2/n_i making the weights $w_i = n_i$.
- For a one factor CRD, the WLS function is

$$W(\mu, \tau_1, \ldots, \tau_a) \;=$$

  To find the least squares normal equations, you simultaneously solve

$$\partial W/\partial\mu = 0 \quad \text{and} \quad \partial W/\partial\tau_i = 0 \quad \text{for } i = 1, 2, \ldots, a.$$

  After algebraic manipulation, this yields the following WLS normal equations:

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{n_i} w_{ij} y_{ij} &= \widehat{\mu} \sum_{i=1}^{a}\sum_{j=1}^{n_i} w_{ij} \,+\, \sum_{i=1}^{a}\left(\widehat{\tau}_i \sum_{j=1}^{n_i} w_{ij}\right) \\ \sum_{j=1}^{n_i} w_{ij} y_{ij} &= \widehat{\mu} \sum_{j=1}^{n_i} w_{ij} \,+\, \widehat{\tau}_i \sum_{j=1}^{n_i} w_{ij} \quad \text{for } i = 1, 2, \ldots, a\end{aligned}$$

  The solution to these $(a+1)$ equations subject to one constraint (such as $\sum_{i=1}^{a} \tau_i = 0$) are the WLS solutions.
- However, because the variance σ_i^2 of y_{ij} is typically unknown, we need to estimate the weight $1/\sigma_i^2$ from the data.
- For the one-factor CRD, we know the sample variance s_i^2 for treatment i is an unbiased estimate of σ_i^2 $(E(s_i^2) = \sigma_i^2)$. The estimated weight is $\widehat{w}_{ij} = 1/s_i^2$.
- SAS and Minitab will perform a WLS analysis. You just have to supply the weights.

2.13.1 Weighted Least Squares (WLS) Example

EXAMPLE: A company wants to test the effectiveness of a new chemical disinfectant. Six dosage levels were considered (1 through 5 grams per 100 ml). The experiment involved applying equal amounts of the disinfectant at each level to a surface that was covered with a common bacteria. The results are given below. The design was completely randomized.

Dose	%
1	5
1	1
1	3
1	5
1	2
1	6
1	1
1	3

Dose	%
2	13
2	13
2	6
2	7
2	11
2	4
2	14
2	12

Dose	%
3	12
3	16
3	9
3	18
3	16
3	7
3	14
3	13

Dose	%
4	17
4	13
4	16
4	19
4	26
4	15
4	23
4	27

Dose	%
5	22
5	30
5	27
5	32
5	32
5	43
5	29
5	26

The sample variances s_i^2 are

$s_1^2 =$ $\qquad$ $s_2^2 =$ $\qquad$ $s_3^2 =$ $\qquad$ $s_4^2 =$ $\qquad$ $s_5^2 =$

Thus, the weights $1/s_i^2$ are

$w_1 =$ $\qquad$ $w_2 =$ $\qquad$ $w_3 =$ $\qquad$ $w_4 =$ $\qquad$ $w_5 =$

SAS Output for WLS Example

SAMPLE VARIANCES AND WEIGHTS FOR EACH TREATMENT trt

Obs	trt	var_y	wgt
1	1	3.6429	0.27451
2	2	14.2857	0.07000
3	3	13.8393	0.07226
4	4	27.4286	0.03646
5	5	38.1250	0.02623

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	207.5551273	51.8887818	51.89	<.0001
Error	35	35.0000000	1.0000000		
Corrected Total	39	242.5551273			

R-Square	Coeff Var	Root MSE	y Mean
0.855703	11.86288	1.000000	8.429653

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	4	207.5551273	51.8887818	51.89	<.0001

Bonferroni (Dunn) t Tests for y

This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	35
Error Mean Square	1
Critical Value of t	2.99605

Comparisons significant at the 0.05 level are indicated by *.**				
trt Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
5 - 4	10.6250	2.0487	19.2013	***
5 - 3	17.0000	9.3642	24.6358	***
5 - 2	20.1250	12.4564	27.7936	***
5 - 1	26.8750	20.0292	33.7208	***
4 - 5	-10.6250	-19.2013	-2.0487	***
4 - 3	6.3750	-0.4297	13.1797	
4 - 2	9.5000	2.6586	16.3414	***
4 - 1	16.2500	10.3455	22.1545	***
3 - 5	-17.0000	-24.6358	-9.3642	***
3 - 4	-6.3750	-13.1797	0.4297	
3 - 2	3.1250	-2.4926	8.7426	
3 - 1	9.8750	5.4460	14.3040	***
2 - 5	-20.1250	-27.7936	-12.4564	***
2 - 4	-9.5000	-16.3414	-2.6586	***
2 - 3	-3.1250	-8.7426	2.4926	
2 - 1	6.7500	2.2649	11.2351	***
1 - 5	-26.8750	-33.7208	-20.0292	***
1 - 4	-16.2500	-22.1545	-10.3455	***
1 - 3	-9.8750	-14.3040	-5.4460	***
1 - 2	-6.7500	-11.2351	-2.2649	***

```
DM 'LOG; CLEAR; OUT; CLEAR;';
ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\WLS.PDF';
OPTIONS NODATE NONUMBER;

DATA in; INPUT trt y @@;  CARDS;
1  5   1  1   1  3   1  5   1  2   1  6   1  1   1  3
2 13   2 13   2  6   2  7   2 11   2  4   2 14   2 12
3 12   3 16   3  9   3 18   3 16   3  7   3 14   3 13
4 17   4 13   4 16   4 19   4 26   4 15   4 23   4 27
5 22   5 30   5 27   5 32   5 32   5 43   5 29   5 26
;
```

```
PROC SORT  DATA=in; BY trt;            <-- Sort the data by treatments.
PROC MEANS DATA=in noprint; BY trt;    <-- Calculate and save sample
     VAR y;                            <-- variances in `wset'.
     OUTPUT OUT=wset VAR=var_y;

DATA wset; SET wset;                   <-- Calculate the weights from
     wgt = 1/var_y;                    <-- the sample variance in wset.
DROP _FREQ_ _TYPE_;

PROC PRINT DATA=wset;
TITLE 'SAMPLE VARIANCES AND WEIGHTS FOR EACH TREATMENT trt';

DATA in; MERGE in wset; BY trt;        <-- Attach the weights by treatment.

PROC GLM DATA=in;
    WEIGHT wgt;                        <-- Include the WEIGHT statement.
    CLASS trt;
    MODEL y = trt / SS3;
     MEANS trt / BON;
TITLE 'WEIGHTED LEAST SQUARES EXAMPLE WITH BONFERRONI MCP';
RUN;
```

2.14 Variance Stabilizing Transformations

- If the homogeneity of variance assumption is only moderately violated, the F-test results are slightly affected when the design is balanced (equal n_i's). No transformation should be considered.
- If the homogeneity of variance assumption is either (i) seriously violated or (ii) moderately violated with very different n_i sample sizes (serious imbalance), then the effects on the F-test are more serious.
  - If the treatments having the larger variances have the smaller sample sizes, the true Type I error is larger than the nominal level.
  - If the treatments having the larger variances have the larger sample sizes, the true Type I error is smaller than the nominal level.
- A common approach to deal with nonconstant variance (heterogeneity of variance) is to apply a **variance-stabilizing transformation** of the response that will equalize the variances across treatments. We then perform the ANOVA on the transformed data.
- Sometimes the variance of the response increases or decreases as the mean of the response increases. If this is the case, the residuals vs predicted values plot would have a funnel shape. This is when a variance stabilizing transformation may be appropriate.
- The statistical problem is to use the data to determine the form of the required transformation.
- Let μ_i be the mean for treatment i. Suppose the standard deviation of y_{ij} is proportional to a power of μ_i. That is, $\sigma_i = \theta\mu_i^\alpha$ for some α and θ. θ is called the **constant of proportionality**. Notationally, we write $\sigma_i \propto \mu_i^\alpha$. The symbol $\propto$ means "is proportional to".
- The goal is to find a transformation $y^* = y^\lambda$ such that y^* has constant or near constant variance across all treatments.
- This implies that after transforming each y_{ij} to y^*_{ij}, we no longer have a HOV problem when the ANOVA is run with the y^*_{ij} values.
- It can be shown that the variance is constant if $\lambda = 1 - \alpha$. We will discuss how to estimate α or λ.

2.14.1 The Empirical Method

- If $\sigma_i = \theta\mu_i^\alpha$, then $\log(\sigma_i) = \log(\theta) + \alpha\log(\mu_i)$. A plot of $\log(\sigma_i)$ vs $\log(\mu_i)$ is linear with slope equal to α. Thus, a simple way to estimate α would be to

  1. Calculate s_i and $\overline{y}_{i\cdot}$ for treatment $i = 1, 2, \ldots, a$.
  2. Fit a regression line $\log(s_i) =$
  3. The least squares estimate of the slope $\widehat{\alpha}$ is the estimate of α.
  4. Transform each y_{ij} to $y_{ij}^* = y_{ij}^\lambda$ where $\lambda = 1 - \widehat{\alpha}$.
  5. Run the ANOVA on the y_{ij}^* values.

- Note that if $\alpha = 0$, then $\sigma_i = \theta$ for all i. Thus, the homogeneity of variance assumption is met without a transformation.

2.14.2 The Box-Cox Procedure

- Another approach is the **Box-Cox procedure** which will estimate the value of λ corresponding to the transformation y_{ij}^λ that maximizes the model R^2.

- To find the Box-Cox transformation,

  1. For a sequence of λ values, calculate $R^2(\lambda)$. $R^2(\lambda)$ is the model R^2 value from the ANOVA on the transformed $y^{(\lambda)}$ values.
  2. Select the λ that maximizes $R^2(\lambda)$ (which is equivalent to maximizing the likelihood function).
  3. Run the ANOVA on the y_{ij}^λ values.

- SAS can find the Box-Cox transformation using the TRANSREG procedure.

2.14.3 Transformation Example using the Empirical and Box-Cox Methods

EXAMPLE: We will use the same data used in the WLS example:

Dose	%	Dose	%	Dose	%	Dose	%	Dose	%
1	5	2	13	3	12	4	17	5	22
1	1	2	13	3	16	4	13	5	30
1	3	2	6	3	9	4	16	5	27
1	5	2	7	3	18	4	19	5	32
1	2	2	11	3	16	4	26	5	32
1	6	2	4	3	7	4	15	5	43
1	1	2	14	3	14	4	23	5	29
1	3	2	12	3	13	4	27	5	26

- We will see that the recommended transformation is a square root ($\lambda = .5$) transformation. The following SAS output contains

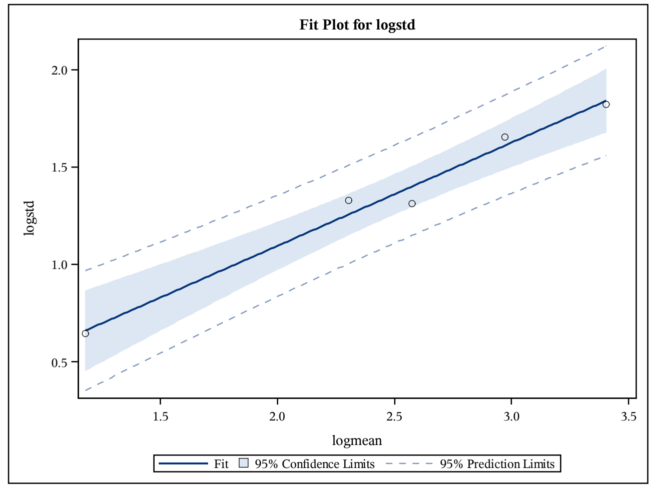
  – The empirical method results and the Box-Cox method results.

  – The analysis of the original data. Note that the variability increases with the treatment levels (from 1 to 5).

  – The analysis of the transformed (square root) data. Note that the variability is now nearly constant across the treatment levels (from 1 to 5).

EMPIRICAL SELECTION OF ALPHA

Obs	mean	std	logstd	logmean
1	3.250	1.90863	0.64638	1.17865
2	10.000	3.77964	1.32963	2.30259
3	13.125	3.72012	1.31376	2.57452
4	19.500	5.23723	1.65579	2.97041
5	30.125	6.17454	1.82044	3.40536



ANOVA TO FIND EMPIRICAL SELECTION OF ALPHA

The GLM Procedure

Variable: logstd

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.79599334	0.79599334	153.30	0.0011
Error	3	0.01557765	0.00519255		
Corrected Total	4	0.81157099			

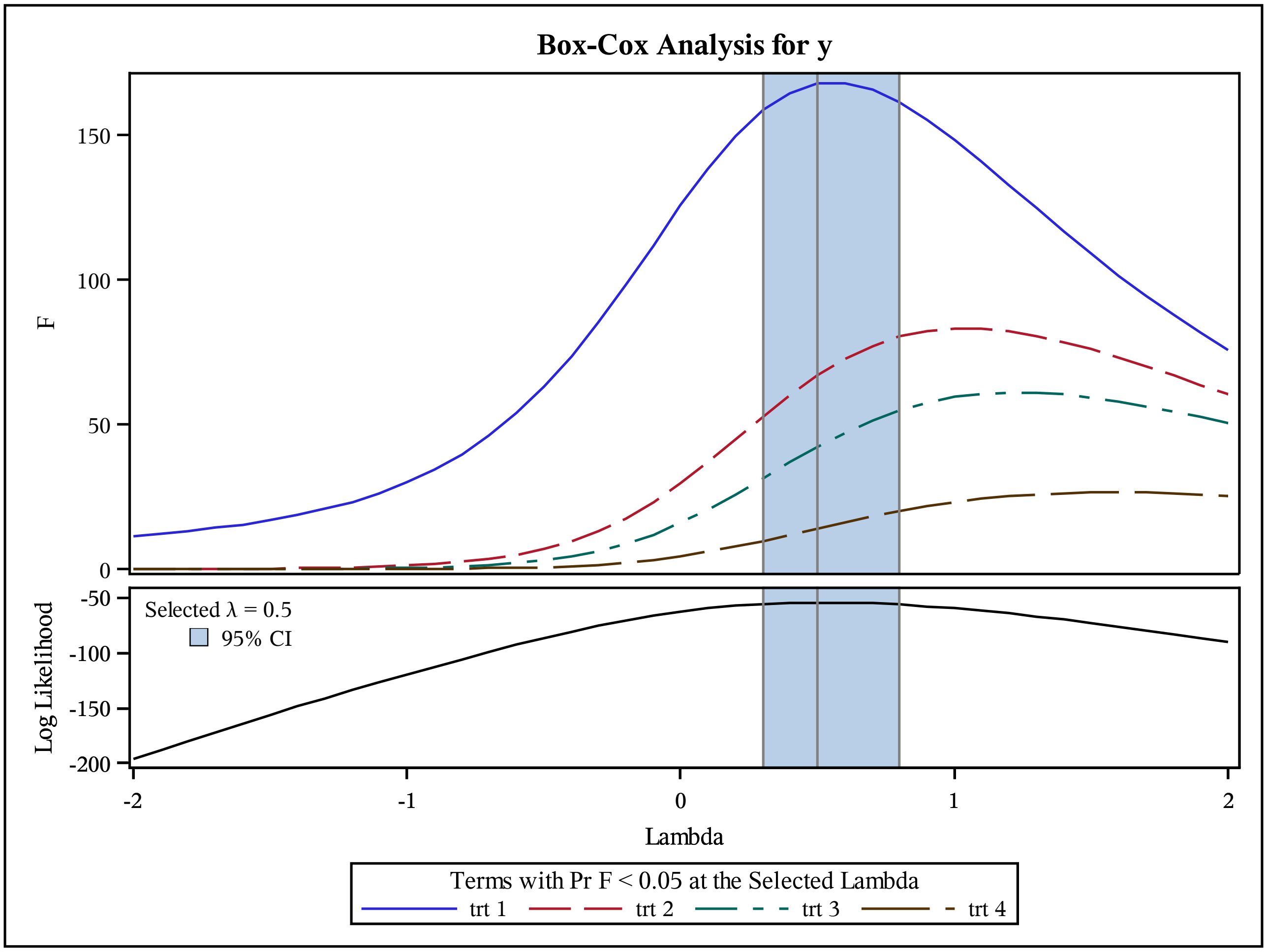
R-Square	Coeff Var	Root MSE	logstd Mean
0.980806	5.325109	0.072059	1.353200

Source	DF	Type III SS	Mean Square	F Value	Pr > F
logmean	1	0.79599334	0.79599334	153.30	0.0011

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.0347067133	0.11126036	0.31	0.7755
logmean	0.5303019549	0.04283106	12.38	0.0011

Find the Box-Cox Transformation using PROC TRANSREG

The TRANSREG Procedure



Level of trt	N	y Mean	y Std Dev
1	8	3.2500000	1.90862703
2	8	10.0000000	3.77964473
3	8	13.1250000	3.72011905
4	8	19.5000000	5.23722937
5	8	30.1250000	6.17454452

Level of trt	N	sqrty Mean	sqrty Std Dev
1	8	1.72499261	0.56000051
2	8	3.10359092	0.64827099
3	8	3.58746278	0.53995849
4	8	4.38144283	0.58842037
5	8	5.46489064	0.54507700

ANOVA -- ORIGINAL DATA

The GLM Procedure

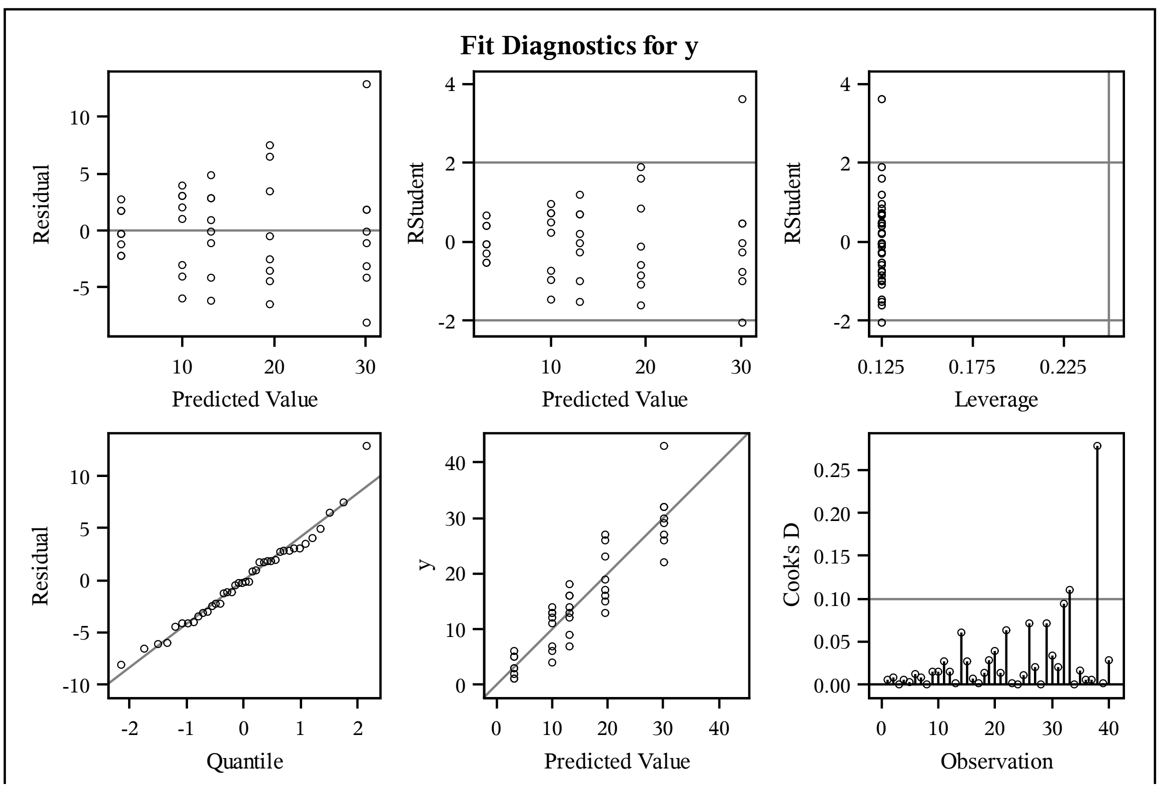
Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	3323.150000	830.787500	42.68	<.0001
Error	35	681.250000	19.464286		
Corrected Total	39	4004.400000			

R-Square	Coeff Var	Root MSE	y Mean
0.829875	29.02523	4.411835	15.20000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	4	3323.150000	830.787500	42.68	<.0001

Dependent Variable: y



ANOVA -- SQUARE ROOT TRANSFORMATION

The GLM Procedure

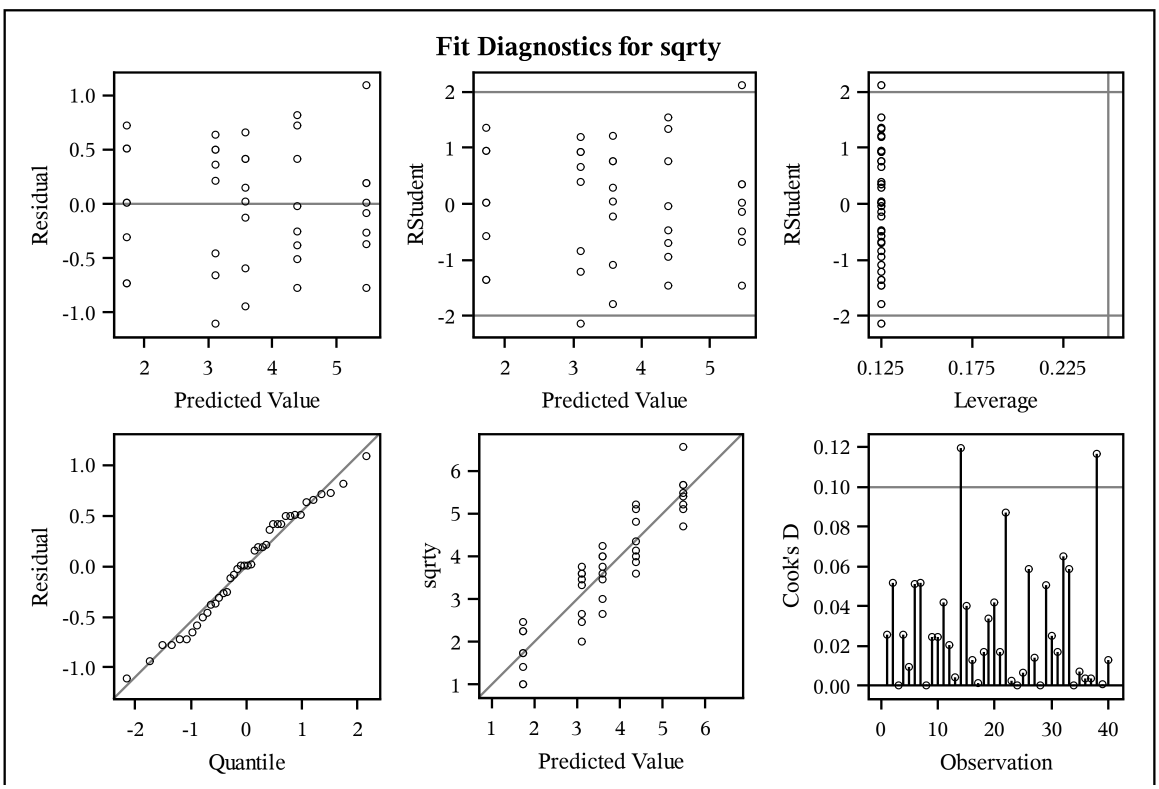
Variable: sqrty

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	62.69546580	15.67386645	46.96	<.0001
Error	35	11.68130939	0.33375170		
Corrected Total	39	74.37677520			

R-Square	Coeff Var	Root MSE	sqrty Mean
0.842944	15.81701	0.577712	3.652476

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	4	62.69546580	15.67386645	46.96	<.0001

Dependent Variable: sqrty



```
DM 'LOG; CLEAR; OUT; CLEAR;';

ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\BOXCOX.PDF';
OPTIONS NODATE NONUMBER;

*******************************************;
*** Variance Stabilizing Transformations ***;
*******************************************;
DATA in; INPUT trt y @@;  CARDS;
1  5   1  1   1  3   1  5   1  2   1  6   1  1   1  3
2 13   2 13   2  6   2  7   2 11   2  4   2 14   2 12
3 12   3 16   3  9   3 18   3 16   3  7   3 14   3 13
4 17   4 13   4 16   4 19   4 26   4 15   4 23   4 27
5 22   5 30   5 27   5 32   5 32   5 43   5 29   5 26

*******************************************************;
*** Find the transformation using the empirical method ***;
*******************************************************;

PROC SORT DATA=in; BY trt;
PROC MEANS DATA=in NOPRINT; BY trt;
     VAR y; OUTPUT OUT=yset MEAN=mean STD=std;
DATA yset; SET yset;
     logstd =LOG(std);   logmean=LOG(mean);
PROC PRINT DATA=yset;
     VAR mean std logstd logmean;
TITLE 'EMPIRICAL SELECTION OF ALPHA';

PROC GLM DATA=yset;
     MODEL logstd=logmean / SS3 solution;
TITLE 'ANOVA TO FIND EMPIRICAL SELECTION OF ALPHA';

*** Use the output of the GLM procedure regressing the  ***;
*** log standard deviations on the log means. Apply the ***;
*** Apply the transformation to the response and        ***;
*** rerun the analysis with the transformed response.   ***;

*****************************************************;
*** Find the transformation using the Box-Cox method ***;
*****************************************************;

PROC TRANSREG DATA=in;
     MODEL BOXCOX(y / LAMBDA=-2 to 2 by .1) = CLASS(trt);
TITLE 'Find the Box-Cox Transformation using PROC TRANSREG';

*** Use the output of the TRANSREG procedure to find the ***;
*** the Box-Cox transformation. Apply the transformation ***;
*** to the response and rerun the analysis with the      ***;
*** transformed response                                 ***;

*************************************;
*** ANOVA BEFORE A TRANSFORMATION ***;
*************************************;
PROC GLM DATA=in PLOTS=(DIAGNOSTICS);
     CLASS trt;
     MODEL y = trt / SS3;
     MEANS trt ;
TITLE 'ANOVA -- ORIGINAL DATA';

*************************************;
*** ANOVA AFTER A TRANSFORMATION ***;
*************************************;

DATA in; SET in;;  sqrty = SQRT(y);

PROC GLM DATA=in PLOTS=(DIAGNOSTICS);
     CLASS trt;
     MODEL sqrty = trt / SS3;
     MEANS trt;
TITLE 'ANOVA -- SQUARE ROOT TRANSFORMATION';
RUN;
```