Lecture 4. Checking Model Adequacy
Montgomery: 3-4, 15-1.1
Model Checking and Diagnostics

• Model Assumptions
  1 Model is correct
  2 Independent observations
  3 Errors normally distributed
  4 Constant variance

\[ y_{ij} = (\bar{y}_.. + (\bar{y}_i. - \bar{y}_..)) + (y_{ij} - \bar{y}_i.) \]

\[ y_{ij} = \hat{y}_{ij} + \hat{\varepsilon}_{ij} \]
observed = predicted + residual

• Note that the predicted response at treatment \( i \) is \( \hat{y}_{ij} = \bar{y}_i. \)

• Diagnostics use predicted responses and residuals.
Diagnostics

• Normality
  – Histogram of residuals
  – Normal probability plot / QQ plot
  – Shapiro-Wilk Test

• Constant Variance
  – Plot $\hat{e}_{i,j}$ vs $\hat{y}_{i,j}$ (residual plot)
  – Bartlett’s or Levene’s Test

• Independence
  – Plot $\hat{e}_{i,j}$ vs time/space
  – Plot $\hat{e}_{i,j}$ vs variable of interest

• Outliers
Constant Variance

- In some experiments, error variance \( (\sigma^2_i) \) depends on the mean response

\[
E(y_{ij}) = \mu_i = \mu + \tau_i.
\]

So the constant variance assumption is violated.

- Size of error (residual) depends on mean response (predicted value)

- Residual plot
  - Plot \( \hat{\epsilon}_{ij} \) vs \( \hat{y}_{ij} \)
  - Is the range constant for different levels of \( \hat{y}_{ij} \)

- More formal tests:
  - Bartlett’s Test
  - Modified Levene’s Test.
Bartlett’s Test

- $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2$

- Test statistic: $\chi^2_0 = 2.3026 \frac{q}{c}$

where

$$q = (N - a) \log_{10} S_p^2 - \sum_{i=1}^{a} (n_i - 1) \log_{10} S_i^2$$

$$c = 1 + \frac{1}{3(a-1)} \left( \sum_{i=1}^{a} (n_i - 1)^{-1} - (N - a)^{-1} \right)$$

and $S_i^2$ is the sample variance of the $i$th population and $S_p^2$ is the pooled sample variance.

- Decision Rule: reject $H_0$ when $\chi^2_0 > \chi^2_{\alpha, a-1}$.

Remark: sensitive to normality assumption.
Modified Levene’s Test

• For each fixed \( i \), calculate the median \( m_i \) of \( y_{i1}, y_{i2}, \ldots, y_{in_i} \).

• Compute the absolute deviation of observation from sample median:

\[
d_{ij} = |y_{ij} - m_i|
\]

for \( i = 1, 2, \ldots, a \) and \( j = 1, 2, \ldots, n_i \),

• Apply ANOVA to the deviations: \( d_{ij} \)

• Use the usual ANOVA \( F \)-statistic for testing \( H_0 : \sigma_1^2 = \ldots = \sigma_a^2 \).
options ls=80 ps=65;

title1 'Diagnostics Example';

data one;
  infile 'c:\saswork\data\tensile.dat';
  input percent strength time;

proc glm data=one;
  class percent;
  model strength=percent;
  means percent / hovtest=bartlett hovtest=levene;
  output out=diag p=pred r=res;

proc sort; by pred;
symbol1 v=circle i=sm50; title1 'Residual Plot';
proc gplot; plot res*pred/frame; run;

proc univariate data=diag normal noprint;
  var res; qqplot res / normal (L=1 mu=est sigma=est);
  histogram res / normal; run;
run;

proc sort; by time;
symbol1 v=circle i=sm75;
title1 'Plot of residuals vs time';
proc gplot; plot res*time / vref=0 vaxis=-6 to 6 by 1;
run;

symbol1 v=circle i=sm50;
title1 'Plot of residuals vs time';
proc gplot; plot res*time / vref=0 vaxis=-6 to 6 by 1;
run;
Diagnostics Example

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>475.760000</td>
<td>118.940000</td>
<td>14.76</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>161.200000</td>
<td>8.06000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>24</td>
<td>636.960000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Levene’s Test for Homogeneity of strength Variance
ANOVA of Squared Deviations from Group Means

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent</td>
<td>4</td>
<td>91.6224</td>
<td>22.9056</td>
<td>0.45</td>
<td>0.7704</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>1015.4</td>
<td>50.7720</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bartlett’s Test for Homogeneity of strength Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent</td>
<td>4</td>
<td>0.9331</td>
<td>0.9198</td>
</tr>
</tbody>
</table>
Non-constant Variance: Impact and Remedy

- Does not affect F-test dramatically when experiment is balanced

- Why concern?
  - Comparison of treatments depends on MSE
  - Incorrect intervals and comparison results

- Variance-Stabilizing Transformations
  - Common transformations
    \[ \sqrt{x}, \log(x), \frac{1}{x}, \arcsin(\sqrt{x}), \text{ and } \frac{1}{\sqrt{x}} \]
  - Box-Cox transformations
    1. approximate the relationship \( \sigma_i = \theta \mu_i^\beta \), then the transformation is \( X^{1-\beta} \)
    2. use maximum likelihood principle
  * Distribution often more “normal” after transformation
Ideas for Finding Proper Transformations

• Consider response $Y$ with mean $E(Y) = \mu$ and variance $Var(Y) = \sigma^2$.

• That $\sigma^2$ depends on $\mu$ leads to nonconstant variances for different $\mu$.

• Let $f$ be a transformation and $\tilde{Y} = f(Y)$; What is the mean and variance of $\tilde{Y}$?

• Approximate $f(Y)$ by a linear function (Delta Method):

\[
 f(Y) \approx f(\mu) + (Y - \mu)f'(\mu)
\]

Mean $\tilde{\mu} = E(\tilde{Y}) = E(f(Y)) \approx E(f(\mu)) + E((Y - \mu)f'(\mu)) = f(\mu)$

Variance $\tilde{\sigma}^2 = Var(\tilde{Y}) \approx [f'(\mu)]^2 Var(Y) = [f'(\mu)]^2 \sigma^2$

• $f$ is a good transformation if $\tilde{\sigma}^2$ does not depend on $\tilde{\mu}$ anymore. So, $\tilde{Y}$ has constant variance for different $f(\mu)$. 
Transformations

- Suppose $\sigma^2$ is a function of $\mu$, that is $\sigma^2 = g(\mu)$

- Want to find transformation $f$ such that $\tilde{Y} = f(Y)$ has constant variance: $\text{Var}(\tilde{Y})$ does not depend on $\mu$.

- Have shown $\text{Var}(\tilde{Y}) \approx [f'(\mu)]^2 \sigma^2 \approx [f'(\mu)]^2 g(\mu)$

- Want to choose $f$ such that $[f'(\mu)]^2 g(\mu) \approx c$

Examples

$g(\mu) = \mu$ (Poisson) $f(X) = \int \frac{1}{\sqrt{\mu}} d\mu \to f(X) = \sqrt{X}$

$g(\mu) = \mu (1 - \mu)$ (Binomial) $f(X) = \int \frac{1}{\sqrt{\mu(1-\mu)}} d\mu \to f(X) = \arcsin(\sqrt{X})$

$g(\mu) = \mu^{2\beta}$ (Box-Cox) $f(X) = \int \mu^{-\beta} d\mu \to f(X) = X^{1-\beta}$

$g(\mu) = \mu^2$ (Box-Cox) $f(X) = \int \frac{1}{\mu} d\mu \to f(X) = \log X$
Identify Box-Cox Transformation Using Data: Approximate Method

- From the previous slide, if $\sigma = \theta \mu^\beta$, the transformation is

$$f(Y) = \begin{cases} 
  Y^{1-\beta} & \beta \neq 1; \\
  \log Y & \beta = 1 
\end{cases}$$

So it is crucial to estimate $\beta$ based on data $y_{ij}$, $i = 1, \ldots, a$.

- We have $\log \sigma_i = \log \theta + \beta \log \mu_i$

- Let $s_i$ and $\bar{y}_i$ be the sample standard deviations and means. Because $\hat{\sigma}_i = s_i$ and $\hat{\mu}_i = \bar{y}_i$, approximately,

$$\log s_i = \text{constant} + \beta \log \bar{y}_i,$$

where $i = 1, \ldots, a$.

- We can plot $\log s_i$ against $\log \bar{y}_i$, fit a straight line and use the slope to estimate $\beta$. 
Identify Box-Cox Transformation: Formal Method

1. For a fixed \( \lambda \), perform analysis of variance on

\[
y_{ij}(\lambda) = \begin{cases} 
\frac{y_{ij}^\lambda - 1}{\lambda ar{y}_{\lambda} - 1} & \lambda \neq 0 \\
\hat{y} \log y_{ij} & \lambda = 0 
\end{cases}
\]

where \( \hat{y} = \left( \prod_{i=1}^{a} \prod_{j=1}^{n_i} y_{ij} \right)^{1/N} \).

2. Step 1 generates a transformed data \( y_{ij}(\lambda) \). Apply ANOVA to the new data and obtain \( SS_E \). Because \( SS_E \) depends on \( \lambda \), it is denoted by \( SS_E(\lambda) \).

- Repeat 1 and 2 for various \( \lambda \) in an interval, e.g., \([-2,2]\), and record \( SS_E(\lambda) \).

3. Find \( \lambda_0 \) which minimizes \( SS_E(\lambda) \) and pick up a meaningful \( \lambda \) in the neighborhood of \( \lambda_0 \). Denote it again by \( \lambda \).

4. The transformation is:

\[
\tilde{y}_{ij} = y_{ij}^{\lambda_0} \text{ if } \lambda_0 \neq 0; \\
\tilde{y}_{ij} = \log y_{ij} \text{ if } \lambda_0 = 0.
\]
### An Example: boxcox.dat

```
trt  response
1    0.948916
1    0.431494
1    3.486359
   ...  
2    3.469623
2    0.840701
2    3.816014
2    1.234756
   ...  
3    10.680733
3    19.453816
3    3.810572
3    10.832754
3    3.814586
```
Approximate Method: trans.sas

options nocenter ps=65 ls=80;
title1 'Increasing Variance Example';
data one;
  infile 'c:\saswork\data\boxcox.dat'; input trt resp;
  proc glm data=one; class trt;
    model resp=trt; output out=diag p=pred r=res;

  title1 'Residual Plot'; symbol1 v=circle i=none;
  proc gplot data=diag; plot res*pred /frame;

  proc univariate data=one noprint;
    var resp; by trt; output out=two mean=mu std=sigma;
  data three;
    set two; logmu = log(mu); logsig = log(sigma);

  proc reg; model logsig = logmu;

  title1 'Mean vs Std Dev'; symbol1 v=circle i=rl;
  proc gplot; plot logsig*logmu / regeqn; run;
Residual Plot

Residual Plot
**Plot of** $\log s_i$ **vs** $\log \mu_i$

**Mean vs Std Dev**

Regression Equation:

$\log s_i = -0.293928 + 1.212067 \times \log \mu_i$
formal method: trans1.sas

options ls=80 ps=65 nocenter;
title1 'Box-Cox Example';

data one;
  infile 'c:\saswork\data\boxcox.dat';
  input trt resp;
  logresp = log(resp);

proc univariate data=one noprint;
  var logresp; output out=two mean=mlogresp;

data three;
  set one; if _n_ eq 1 then set two;
  ydot = exp(mlogresp);
  do l=-1.0 to 1.0 by .25;
    den = l*ydot**(l-1); if abs(l) eq 0 then den = 1;
    yl=(resp**l -1)/den; if abs(l) < 0.0001 then yl=ydot*log(resp);
    output;
  end;
keep trt yl l;

proc sort data=three out=three; by l;
proc glm data=three noprint outstat=four;
   class trt; model yl=trt; by l;

data five; set four;
   if _SOURCE_ eq 'ERROR'; keep l SS;

proc print data=five;
run;

symbol1 v=circle i=sm50;
proc gplot;
   plot SS*l;
run;
### $SS_E(\lambda)$ and $\lambda$

<table>
<thead>
<tr>
<th>OBS</th>
<th>L</th>
<th>SS</th>
<th>OBS</th>
<th>L</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.00</td>
<td>2150.06</td>
<td>10</td>
<td>0.25</td>
<td>112.37</td>
</tr>
<tr>
<td>2</td>
<td>-1.75</td>
<td>1134.83</td>
<td>11</td>
<td>0.50</td>
<td>154.23</td>
</tr>
<tr>
<td>3</td>
<td>-1.50</td>
<td>628.94</td>
<td>12</td>
<td>0.75</td>
<td>253.63</td>
</tr>
<tr>
<td>4</td>
<td>-1.25</td>
<td>369.35</td>
<td>13</td>
<td>1.00</td>
<td>490.36</td>
</tr>
<tr>
<td>5</td>
<td>-1.00</td>
<td>232.32</td>
<td>14</td>
<td>1.25</td>
<td>1081.29</td>
</tr>
<tr>
<td>6</td>
<td>-0.75</td>
<td>158.56</td>
<td>15</td>
<td>1.50</td>
<td>2636.06</td>
</tr>
<tr>
<td>7</td>
<td>-0.50</td>
<td>119.28</td>
<td>16</td>
<td>1.75</td>
<td>6924.95</td>
</tr>
<tr>
<td>8</td>
<td>-0.25</td>
<td>100.86</td>
<td>17</td>
<td>2.00</td>
<td>19233.39</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>98.09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Plot of $SS_E(\lambda)$ vs $\lambda$

Increasing Variance Example
Using Proc Transreg

```plaintext
proc transreg data=one;
model boxcox(y/lambda=-2.0 to 2.0 by 0.1)=class(trt); run;
```

The TRANSREG Procedure
Transformation Information
for BoxCox(y)

<table>
<thead>
<tr>
<th>Lambda</th>
<th>R-Square</th>
<th>Log Like</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.0</td>
<td>0.10</td>
<td>-108.906</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>0.18</td>
<td>-22.154</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.19</td>
<td>-19.683</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.20</td>
<td>-17.814*</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.20</td>
<td>-16.593*</td>
</tr>
<tr>
<td>-0.1</td>
<td>0.21</td>
<td>-16.067&lt;</td>
</tr>
<tr>
<td>0.0 +</td>
<td>0.21</td>
<td>-16.284*</td>
</tr>
<tr>
<td>0.1</td>
<td>0.22</td>
<td>-17.289*</td>
</tr>
<tr>
<td>0.2</td>
<td>0.22</td>
<td>-19.124</td>
</tr>
<tr>
<td>0.3</td>
<td>0.22</td>
<td>-21.820&lt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>0.10</td>
<td>-174.641+</td>
</tr>
</tbody>
</table>

* Confidence Interval
< Best Lambda
+ Convenient Lambda
Nonnormality

<table>
<thead>
<tr>
<th>trt</th>
<th>nitrogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.80</td>
</tr>
<tr>
<td>2</td>
<td>0.60</td>
</tr>
<tr>
<td>3</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>1.20</td>
</tr>
<tr>
<td>5</td>
<td>0.74</td>
</tr>
<tr>
<td>6</td>
<td>1.26</td>
</tr>
</tbody>
</table>
Test: Shapiro-Wilk
Statistic: W = 0.910027
p Value: Pr < W = 0.0149
Kruskal-Wallis Test: a Nonparametric alternative

*a* treatments, $H_0$: *a* treatments are not different.

- Rank the observations $y_{ij}$ in ascending order
- Replace each observation by its rank $R_{ij}$ (assign average for tied observations)
- Test statistic
  - $H = \frac{1}{S^2} \left[ \sum_{i=1}^{a} \frac{R_{i}^2}{n_i} - \frac{N(N+1)^2}{4} \right] \approx \chi^2_{a-1}$
  - where $S^2 = \frac{1}{N-1} \left[ \sum_{i=1}^{a} \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right]$
- Decision Rule: reject $H_0$ if $H > \chi^2_{\alpha, a-1}$.
- Let $F_0$ be the $F$-test statistic in ANOVA based on $R_{ij}$. Then
  $$F_0 = \frac{H/(a - 1)}{(N - 1 - H)/(N - a)}$$
options nocenter ps=65 ls=80;

data new;
  input strain nitrogen @@;
cards;
  1 2.80 1 7.04 1 0.41 1 1.73 1 0.18
  2 0.60 2 1.14 2 0.14 2 0.16 2 1.40
  3 0.05 3 1.07 3 1.68 3 0.46 3 4.87
  4 1.20 4 0.89 4 3.22 4 0.77 4 1.24
  5 0.74 5 0.20 5 1.62 5 0.09 5 2.27
  6 1.26 6 0.26 6 0.47 6 0.46 6 3.26
;
proc npar1way;
  class strain;
  var nitrogen;
run;
The NPAR1WAY Procedure
Analysis of Variance for Variable nitrogen
    Classified by Variable strain

<table>
<thead>
<tr>
<th>strain</th>
<th>N</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2.4320</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0.6880</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>1.6260</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1.4640</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.9840</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1.1420</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among</td>
<td>5</td>
<td>9.330387</td>
<td>1.866077</td>
<td>0.7373</td>
<td>0.6028</td>
</tr>
<tr>
<td>Within</td>
<td>24</td>
<td>60.739600</td>
<td>2.530817</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable nitrogen
Classified by Variable strain

<table>
<thead>
<tr>
<th>strain</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>93.00</td>
<td>77.50</td>
<td>17.967883</td>
<td>18.60</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>57.00</td>
<td>77.50</td>
<td>17.967883</td>
<td>11.40</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>78.50</td>
<td>77.50</td>
<td>17.967883</td>
<td>15.70</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>93.00</td>
<td>77.50</td>
<td>17.967883</td>
<td>18.60</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>68.00</td>
<td>77.50</td>
<td>17.967883</td>
<td>13.60</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>75.50</td>
<td>77.50</td>
<td>17.967883</td>
<td>15.10</td>
</tr>
</tbody>
</table>

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square     2.5709
DF             5
Pr > Chi-Square 0.7658