

# Chapter 6 Model Assumption Checking and Remedies

Yibi Huang

# THREE ASSUMPTIONS We Need to Check

In the means model

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

we make 3 assumptions about the error term  $\varepsilon_{ij}$ 's:

1. the errors  $\varepsilon_{ij}$  are **independent**, randomly distributed
2. the errors  $\varepsilon_{ij}$  have **constant variance across treatments**
3. the errors  $\varepsilon_{ij}$  follow a **normal distribution**

As the 3 assumptions are all related to errors  $\varepsilon_{ij}$ , most of the model diagnostic methods are based on the *residuals*

$$\text{residual } e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i\bullet}$$

## Standardized and Studentized Residuals

Because the error  $\varepsilon_{ij}$  has mean 0 and SD  $\sigma$ , we sometimes standardize the residuals

$$\text{standardized residual } d_{ij} = \frac{e_{ij} - 0}{\hat{\sigma}} = \frac{e_{ij}}{\sqrt{\text{MSE}}}.$$

By the normality assumption,  $d_{ij}$  is approximately  $N(0, 1)$ . Observations with  $|d_{ij}| > 3$  are potential **outliers**.

A more accurate standardization is the

$$\text{studentized residual } s_{ij} = \frac{e_{ij}}{\sqrt{\text{MSE}(1 - \frac{1}{n_i})}}.$$

We divided by  $\sqrt{\text{MSE}(1 - \frac{1}{n_i})}$  because  $e_{ij} = y_{ij} - \bar{y}_{i\bullet}$  has SD  $\sigma\sqrt{1 - \frac{1}{n_i}}$ . Again, if the errors are normal,  $s_{ij}$  is approximately  $N(0, 1)$ . Observations with  $|s_{ij}| > 3$  are potential **outliers**.

## Example: Hodgkin's Disease

Hodgkin's disease is a type of lymphoma, which is a cancer originating from white blood cells called lymphocytes. The data set `Hodgkins.txt` contains plasma bradykininogen levels (in micrograms of bradykininogen per milliliter of plasma) in 3 types of subjects

- ▶ normal,
- ▶ in patients with active Hodgkin's disease, and
- ▶ in patients with inactive Hodgkin's disease.

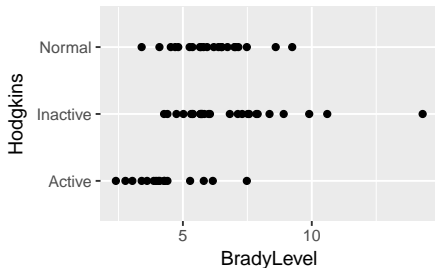
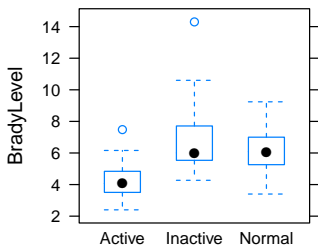
The globulin bradykininogen is the precursor substance for bradykinin, which is thought to be a chemical mediator of inflammation.

- ▶ Is this an experiment?
- ▶ We can use ANOVA to compare means of several samples in an observational study.

```

> hodgkins = read.table("Hodgkins.txt", header=T)
> library(hodgkins)
> bwplot(BradyLevel~Hodgkins, data=hodgkins)
> qqplot(BradyLevel, Hodgkins, data=hodgkins)

```



The distribution within each group looks right skewed. Let's fit the ANOVA model anyway and take a look at the residuals.

```

> brady1 = lm(BradyLevel ~ Hodgkins, data=hodgkins)
> anova(brady1)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hodgkins	2	65.893	32.946	10.67	0.0001042 ***
Residuals	62	191.449	3.088		

# Residuals

$$\text{Residual } e_{ij} = y_{ij} - \bar{y}_{i\bullet}$$

```
> round(brady1$res,2)
> round(brady1$res,2)
  1    2    3    4    5    6    7    8    9   10   11
-0.72 -0.29 -1.40 -0.40 -2.69  2.50  1.38  1.06  0.40 -2.01 -0.15
 12   13   14   15   16   17   18   19   20   21   22
 0.28  3.15 -0.44 -1.57  0.41  0.90  0.10  0.94 -1.28  0.64 -0.84
 23   24   25   26   27   28   29   30   31   32   33
-0.35 -1.27  0.97 -0.91 -0.21 -0.70  1.85  3.17 -0.44 -0.04 -0.26
 34   35   36   37   38   39   40   41   42   43   44
-1.91  1.50 -0.02 -1.54  0.09 -1.49  3.74 -1.84  7.44  3.04 -2.59
 45   46   47   48   49   50   51   52   53   54   55
-1.11 -1.12  0.99 -0.04  1.04  1.50 -1.14 -0.86 -2.11 -1.03  0.44
 56   57   58   59   60   61   62   63   64   65
 0.66 -1.54 -0.81 -1.18  0.71 -1.18  2.05 -1.47 -2.46  0.27
```

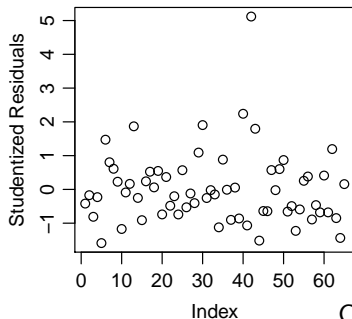
The 42nd observation has the largest studentized residual 7.44.  
Is it an outlier?

# Studentized Residuals

In R, `studres()` is the command to get the studentized residuals.

```
> library(MASS) # must load library "MASS" to studres()
> studres(brady1) # studentized residual
      1          2          3          4          5
-0.419472821 -0.170477001 -0.810287814 -0.228308941 -1.588932833
(... some output is omitted ...)
      61          62          63          64          65
-0.679398000  1.194860046 -0.848621213 -1.436837541  0.157258627

> plot(studres(brady1), ylab="Studentized Residuals")
```



There is a potential outlier with a studentized residual  $> 5$ .

## 3.4.1 Methods for Checking the Normality Assumption

- ▶ Histogram of the residuals, if normal, should be bell-shaped
  - ▶ Pros: simple, easy to understand
  - ▶ Cons: for a small sample, histogram may not be bell-shaped even though the sample is from a normal distribution
- ▶ Normal probability plot of the residuals
  - ▶ aka. normal QQ plot, in which QQ stands for “quantile-quantile”
  - ▶ the best method to assess normality
  - ▶ See next slide for details



## How to Make a Normal Probability Plot?

1. Given data  $(y_1, y_2, \dots, y_n)$ , standardize:  $x_i = (y_i - \bar{y})/s$ .
2. Arrange the data in increasing order:  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ .
3. Find quantiles of the  $N(0, 1)$  distribution:  $z_{(\frac{1}{n+1})}, z_{(\frac{2}{n+1})}, \dots, z_{(\frac{n}{n+1})}$ .

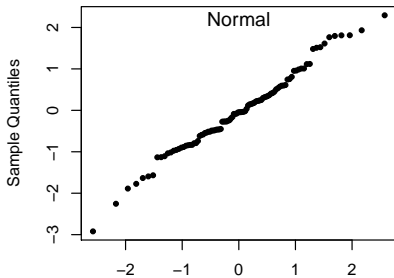
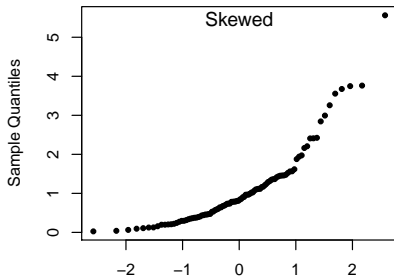
That is,  $z_{(\frac{i}{n+1})}$  is a value such that  $P(Z \leq z_{(\frac{i}{n+1})}) = \frac{i}{n+1}$  for  $Z \sim N(0, 1)$ .

4. Plot the  $x_{(i)}$  values against the  $z_{(\frac{i}{n+1})}$  values.

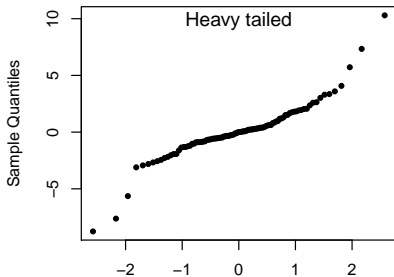
That is, plot the points  $(z_{(\frac{i}{n+1})}, x_{(i)})$  for  $i = 1, 2, \dots, n$

## Interpreting Normal Probability Plots

- ▶ If the data are approximately normal, the plot will be close to a straight line.
- ▶ Systematic deviations from a straight line indicate a non-normal distribution.
- ▶ Outliers appear as points that are far away from the overall pattern of the plot.
- ▶ R command: `qqnorm()`

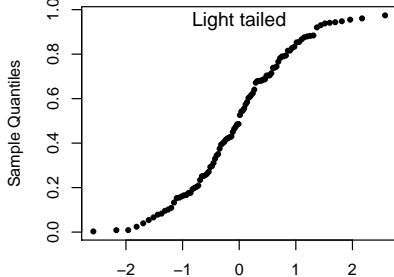
**Normal Q-Q Plot****Normal Q-Q Plot**

Theoretical Quantiles

**Normal Q-Q Plot**

Theoretical Quantiles

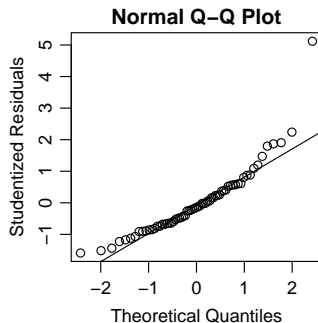
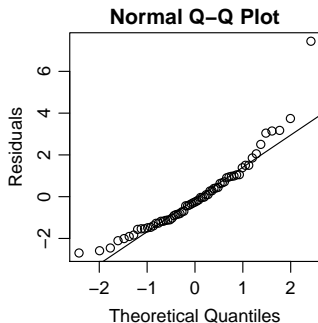
Theoretical Quantiles

**Normal Q-Q Plot**

Theoretical Quantiles

## Normal QQ Plot for The Hodgkin Data

```
> qqnorm(brady1$res, ylab="Residuals")  
> qqline(brady1$res)  
  
> library(MASS)  
> qqnorm(studres(brady1), ylab="Studentized Residuals")  
> qqline(studres(brady1))
```



Does the distribution of the residuals look normal?

It looks somewhat right-skewed, a potential outlier with  $s_{ij} > 5$ .

## Remedies for Non-Normality

Skewness can often be ameliorated by **transforming the response** — often a power transformation.

$$f_{\lambda}(y) = \begin{cases} y^{\lambda}, & \text{if } \lambda \neq 0 \\ \log(y), & \text{if } \lambda = 0. \end{cases}$$

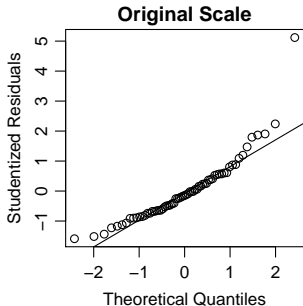
1. If **right-skewed**, try taking square root, logarithm, or other powers  $< 1$

$$y \longrightarrow \log(y), \sqrt{y}, \text{ or } y^{\lambda} \text{ with } \lambda < 1$$

2. If **left-skewed**, try squaring, cubing, or other powers  $> 1$

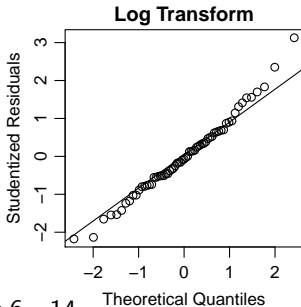
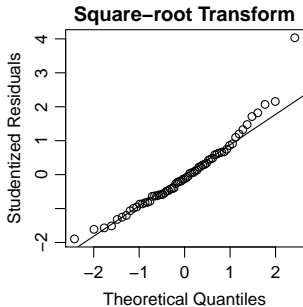
$$y \longrightarrow y^2, y^3, \text{ or } y^{\lambda} \text{ with } \lambda > 1$$

## Example: Hodgkin's Disease – QQ Plots



After a log transformation, the residuals are less skewed, and the outlier looks less extreme.

The square-root transformation also reduces skewness but not as much as the log transformation.



## R-Codes for Making the Plots on the Previous Slide

```
> brady1 = lm(BradyLevel ~ Hodgkins, data=hodgkins)
> brady2 = lm(sqrt(BradyLevel) ~ Hodgkins, data=hodgkins)
> brady3 = lm(log(BradyLevel) ~ Hodgkins, data=hodgkins)

> library(MASS)

> qqnorm(studres(brady1), main="Original Scale")
> qqline(studres(brady1))

> qqnorm(studres(brady2), main="Square-root Transform")
> qqline(studres(brady2))

> qqnorm(studres(brady3), main="Log Transform")
> qqline(studres(brady3))
```

## Box-Cox Method

Box-Cox method is an automatic procedure to select the “best” power  $\lambda$  that make the residuals of the model

$$y_{ij}^{\lambda} = \mu_i + \varepsilon_{ij}$$

closest to normal.

- ▶ We usually round the optimal  $\lambda$  to a convenient power like

$$-1, -\frac{1}{2}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{2}, 1, 2, \dots$$

since the practical difference of  $y^{0.5827}$  and  $y^{0.5}$  is usually small, but the square-root transformation is much easier to interpret.

- ▶ A confidence interval for the optimal  $\lambda$  can also be obtained. See Oehlert, p.129 for details.

We usually select a convenient power  $\lambda^*$  in this C.I.

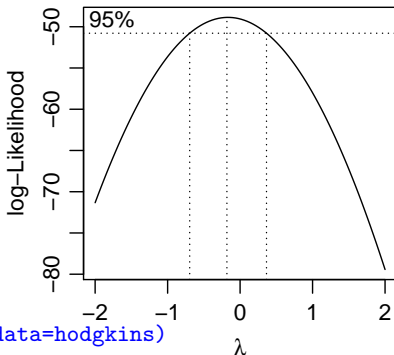


## Example: Hodgkin's Disease – Box-Cox

In R, one must first load the `MASS` library to use `boxcox()`.

The argument of the `boxcox()` function can be a model formula, a `lm` model.

```
> library(MASS)
> boxcox(brady1)
> boxcox(BradyLevel ~ Hodgkins, data=hodgkins)
```



The middle dash line marks the optimal  $\lambda$ , the right and left dash line mark the 95% C.I. for the optimal  $\lambda$ .

For the plot, we see the optimal  $\lambda$  is around  $-0.2$ , and the 95% C.I. contains 0. For simplicity, we use the log-transformed `BradyLevel` as our response.

## A Remark On the Log-Scale

In fact, for measurements of concentration, log scale more commonly used than the original scale. For example,

- ▶ A concentration of 10.1 and 10.001 are almost the same, but a concentration of 0.1 and 0.001 are very different since 0.1 is 100 times higher than 0.001.
- ▶ In the original scale, (10.1, 10.001) and (0.1, 0.001) differ by the same amount, 0.099.
- ▶ In log scale,

$$\log_{10} 0.1 - \log_{10} 0.001 = 2 \quad \text{much greater than} \\ \log_{10} 10.1 - \log_{10} 10.001 \approx 0.0043$$

## Example: Hodgkin's Disease

From the 2 ANOVA tables below, we see the differences of the 3 group of patients become more significant after a log transformation, because the outlier become less extreme and do not inflate the MSE as much.

Response: BradyLevel

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hodgkins	2	65.893	32.946	10.67	0.0001042 ***
Residuals	62	191.449	3.088		

Response: log(BradyLevel)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hodgkins	2	2.2526	1.12631	15.436	3.628e-06 ***
Residuals	62	4.5238	0.07297		

## Non-Parametric Tests

- ▶ Transformation does not always work.  
For example, it helps little for symmetric but heavy-tailed (many outliers) distributions.
- ▶ The ANOVA  $F$ -test is *robust* to non-normality, but it is not resistant to *outliers*.
- ▶ If outliers are unavoidable, and cannot be removed, try non-parametric tests, like permutation test in Chapter 2, and Kruskal-Wallis test in Section 3.11.1, that doesn't rely on normality assumption, which we will introduce soon after we finish Chapter 6.

## Part II: Constant Variance Assumption

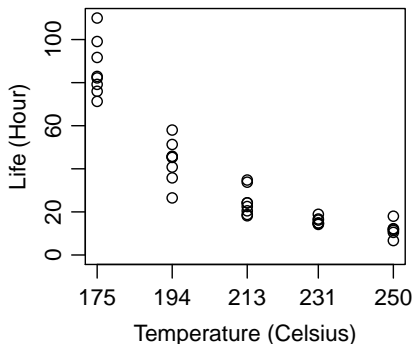
### Outline:

- ▶ Why Is Non-Constant Variance a Problem?
- ▶ Tools for checking nonconstant variance — Residual Plots
  - ▶ Residuals v.s. Fitted Values
  - ▶ Residuals v.s. Treatments
  - ▶ Residuals v.s. Other Variables
- ▶ Remedies
  - ▶ Transforming the Response — Variance-Stabilizing Transformation
  - ▶ Brown-Forsythe Modified  $F$ -test — an alternative to ANOVA  $F$ -test
  - ▶ Welch Test for Contrasts w/o Constant Variance Assumption

## Example: Resin Glue Failure Time

In previous lectures, the response of the resin glue experiment is  $\log_{10}(\text{lifetime})$ .

In the original data, the response is simply the failure time of the resin in hours without log transformation.

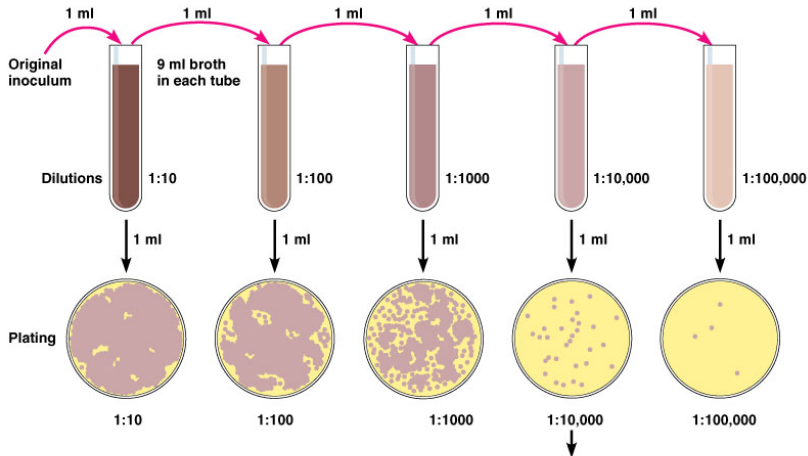


Temperature	Failure Time in Hours							
175°C	110.0	82.2	99.1	82.9	71.3	91.7	76.0	79.2
194°C	45.8	51.3	26.5	58.0	45.3	40.8	35.8	45.6
213°C	33.8	34.8	24.2	20.5	22.5	18.8	18.2	24.2
231°C	14.2	16.7	14.8	14.6	16.2	18.9	14.8	
250°C	18.0	6.7	12.0	10.5	12.2	11.4		

Data file: resinlife.txt

# Example — Serial Dilution Plating (Exercise 6.2 on p.143)

a method to count the number of bacteria in solution



**Calculation: Number of colonies on plate  $\times$  reciprocal of dilution of sample = number of bacteria/ml**  
(For example, if 32 colonies are on a plate of  $1/10,000$  dilution, then the count is  $32 \times 10,000 = 320,000/\text{ml}$  in sample.)

Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings.

## Example — Serial Dilution Plating (Cont'd)

- ▶ Goal: to compare 3 pasteurization methods for milk
- ▶ Design: 15 samples of milk randomly assigned to the 3 trts
- ▶ Response: the bacterial load in each sample after treatment, determined via serial dilution plating
- ▶ Data: <http://users.stat.umn.edu/~gary/book/fcdae.data/ex6.2>

	Method 1	Method 2	Method 3
	$26 \times 10^2$	$35 \times 10^3$	$29 \times 10^5$
	$29 \times 10^2$	$23 \times 10^3$	$23 \times 10^5$
	$20 \times 10^2$	$20 \times 10^3$	$17 \times 10^5$
	$22 \times 10^2$	$30 \times 10^3$	$29 \times 10^5$
	$32 \times 10^2$	$27 \times 10^3$	$20 \times 10^5$
Mean	$25.8 \times 10^2$	$27 \times 10^3$	$23.6 \times 10^5$
SD	492	5874	536656

$$\sqrt{\text{MSE}} = 309900$$



## Why Non-Constant Variability Causes Problems?

95%C.I. for the mean of Method 1:

$$\begin{aligned}\bar{y}_{1\bullet} \pm t_{0.025,15-3} \frac{\sqrt{\text{MSE}}}{\sqrt{n_1}} &= 2580 \pm 2.179 \frac{309900}{\sqrt{5}} \\ &= 2580 \pm 301965 \\ &= (-299385, 304545)\end{aligned}$$

which is way larger than the range of 5 observations for method 1 (2000-3200). What happened?

## Checking Constant Variance Assumption (1)

Tests for equality of variance are available, but not useful, because

1. the tests do not tell us how much nonconstant variance is present and how much it affects our inferences.
2. classical tests of constant variance are very sensitive to nonnormality in the errors. It is very difficult to tell *non-normality* from *normality with non-constant variance.*)

## Checking Constant Variance Assumption (2)

Better tool for checking constant variance — Residual Plots

- ▶ Residuals v.s. Fitted Values
- ▶ Residuals v.s. Treatments
- ▶ Residuals v.s. Other Variables

If the constant variance assumption is true, residuals will evenly spread around the zero line.

**Rule of thumb**: ANOVA  $F$ -tests for CRD can tolerate non-constant variance to some extent, so do tests for contrasts. It usually fine as long as

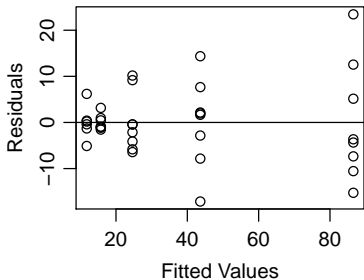
$$\frac{\max\{\hat{\sigma}_i\}}{\min\{\hat{\sigma}_i\}} \leq 2, 3 \text{ or even } 4,$$

especially when the group sizes  $n_i$  are (roughly) equal.

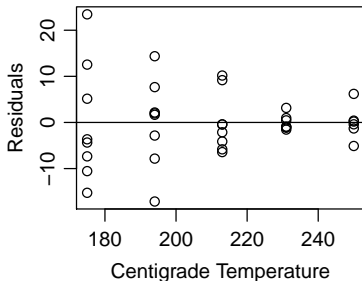
## Example: Resin Glue Data

```
> resinlife = read.table("resinlife.txt", header=T)
> lmorig = lm(life ~ as.factor(temp), data=resinlife)
> plot(lmorig$fit,lmorig$res,ylab="Residuals", xlab="Fitted Values")
> abline(h=0) # adding a zero line
> plot(temp,lmorig$res,ylab="Residuals",xlab="Centigrade Temperature")
> abline(h=0) # adding a zero line
```

Residuals v.s. Fitted Values



Residuals v.s. Treatments



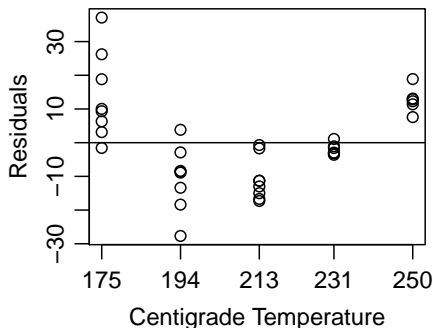
- ▶ Why do the points line up vertically?
- ▶ Variability of residuals increases with the fitted value, but decreases with the temperature

## More Residuals Plots

Residuals plots can be used to check many other things, like non-linearity.

E.g., for the resin glue data, one can check if the linear model  $y_{ij} = \beta_0 + \beta_1 T + \varepsilon_{ij}$  is appropriate by checking the residual plot

```
> lmorig1 = lm(life ~ temp, data=resinlife)
> plot(temp,lmorig1$res,ylab="Residuals",xlab="Centigrade Temperature")
> abline(h=0)
```



From the residual plot, we see

- ▶ non-constant variance across temperature
- ▶ lifetime is *curved*, not linear with temperature

## Plot of Residuals v.s. Variables Not in the Model

If there exist other variables that might affect the response, but are not included in the model, then one should check the plots of residuals versus these variables. For example,

- ▶ if experimental units come from different batches, then plot residuals v.s. batches
- ▶ if measurements are made by several operators, then plot residuals should v.s. operators

Patterns in such residual plots suggest these variables should

- ▶ either be included in the analysis  
(but note one CANNOT claim these variables have a *causal-effect* on the response, since they are not controlled in advance)
- ▶ or be controlled more carefully, e.g., by a **block design**, in future experiments

## Remedy 1: Variance-Stabilizing Transformation

If the SD  $\sigma$  (the spread of residuals) changes the mean  $\mu$  (the fitted values), you can try a *variance-stabilizing transformation* of the response to make the variance (closer to) constant.

- ▶ if the SD is proportional to the fitted value, then

$$y \rightarrow \log(y)$$

- ▶ if the SD is proportional to  $\sqrt{\text{the fitted value}}$ , i.e., the variance is proportional to the fitted value, then

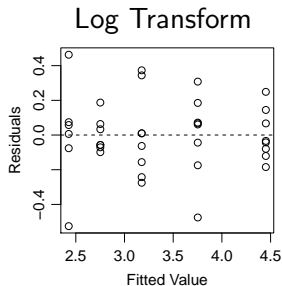
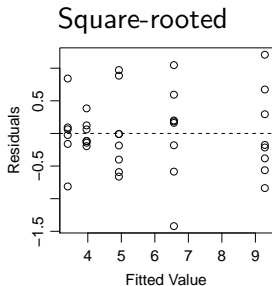
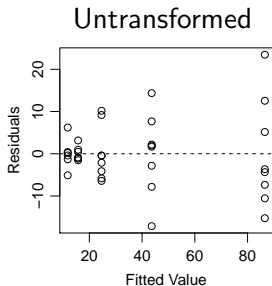
$$y \rightarrow \sqrt{y}$$

- ▶ In general, if the SD  $\sigma$  is proportional to (the fitted values) $^\alpha$ , then the variance-stabilizing transformation is

$$y \rightarrow \begin{cases} y^{1-\alpha} & \text{for } \alpha \neq 1 \\ \log(y) & \text{for } \alpha = 1 \end{cases}$$

# Example: Resin Glue

```
> lm1 = lm(life ~ as.factor(temp))
> lm2 = lm(sqrt(life) ~ as.factor(temp))
> lm3 = lm(log(life) ~ as.factor(temp))
> plot(lm1$fit, lm1$res, xlab="Fitted Value", ylab="Residuals")
> abline(h=0,lty=2)
> plot(lm2$fit, lm2$res, xlab="Fitted Value", ylab="Residuals")
> abline(h=0,lty=2)
> plot(lm3$fit, lm3$res, xlab="Fitted Value", ylab="Residuals")
> abline(h=0,lty=2)
```





## Box-Cox Again

In many cases, we don't have a good idea what is the value of  $\alpha$ . We can still try power transformation of the response.

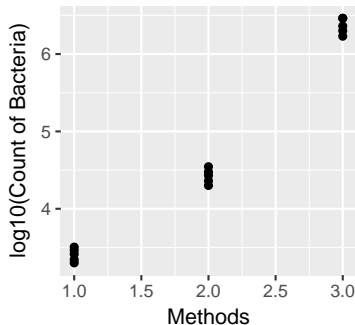
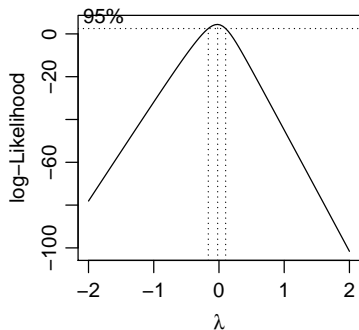
$$f_{\lambda}(y) = \begin{cases} y^{\lambda}, & \text{if } \lambda \neq 0 \\ \log(y), & \text{if } \lambda = 0. \end{cases}$$

How to select  $\lambda$ ?

- ▶ Trial and error: try convenient power like  $-1, -1/2, -1/3, 0, 1/3, 1/2, 2, \dots$  and then check residual plots for each of them for the constant variance.
- ▶ **Box-Cox method:**  
Though Box-Cox is developed to select a power transformation making the residuals as normal as possible, it's been shown that the optimal  $\lambda$  is often close to the variance-stabilizing  $\lambda$ .

## Example – Count of Bacteria Revisit

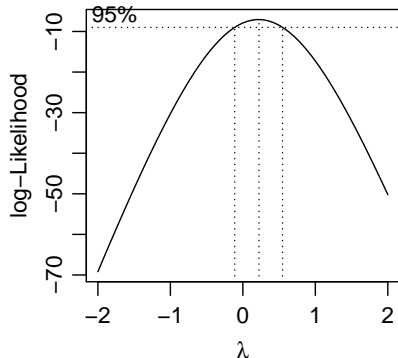
```
ex6.2 = read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/ex6  
library(MASS)  
boxcox(count ~ as.factor(method), data=ex6.2)  
qplot(method, log10(count), data=ex6.2,  
        xlab="Methods", ylab="log10(Count of Bacteria)")
```



After log transformation, the 3 groups look even in variability.

## Example: Resin Glue – Box-Cox

```
> library(MASS)
> boxcox(life ~ as.factor(temp), data=resinlife)
```



The 95% C.I. for  $\lambda$  contains both 0 and 1/2. As  $\lambda = 1/2$  is very close to the boundary of the C.I,  $\lambda = 0$  seems to be a better choice, which is consistent with the Arrhenius Law.

## Drawbacks of Transformation

- ▶ Except for a few special transformation (log,  $\sqrt{\quad}$ , inverse), the transformed response usually lacks natural interpretation (How to interpret  $y^{0.1}$ ?)
- ▶ Unless having a good interpretation on the transformed response, think again before making transformations

Remember that ANOVA tests have some tolerance for non-constant variance. If

$$\frac{\max\{\hat{\sigma}_i\}}{\min\{\hat{\sigma}_i\}} \leq 2, 3, \text{ or even } 4,$$

don't worry too much about non-constant variance.  
In that case, it is fine to leave the response untransformed.

```
> library(mosaic)
> sd(life ~ temp, data=resinlife)
      175      194      213      231      250
12.895514  9.557785  6.378703  1.661181  3.646917

> sd(sqrt(life) ~ temp, data=resinlife)
      175      194      213      231      250
0.6802037 0.7505140 0.6223047 0.2048698 0.5305554

> sd(log(life) ~ temp, data=resinlife)
      175      194      213      231      250
0.1440872 0.2393055 0.2451210 0.1012540 0.3177677
```

After the log transformation, the ratio of the largest and smallest SD is 2.2, which is acceptable

## Brown-Forsythe Modified $F$ -test

If transformation doesn't work well, or you don't want to transform for lacking a good interpretation, try tests that don't rely on the constant variance assumption:

- ▶ **Brown-Forsythe modified  $F$ -test** — an alternative of ANOVA  $F$ -test:

$$BF = \frac{\sum_{i=1}^g n_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}{\sum_{i=1}^g s_i^2 (1 - n_i/N)} = \frac{SS_{Trt}}{\sum_{i=1}^g s_i^2 (1 - n_i/N)}$$

in which  $s_i^2$  is the sample variance in treatment  $i$ . Under the null hypothesis of equal treatment means, BF is approximately distributed as an  $F$ -distribution with  $g - 1$  and  $\nu$  degrees of freedom, where

$$\nu = \frac{(\sum_{i=1}^g d_i)^2}{\sum_{i=1}^g d_i^2 / (n_i - 1)} \quad \text{in which} \quad d_i = s_i^2 (1 - n_i/N).$$

## Welch Tests for Contrasts w/o Constant Variance Assumption

If transformation doesn't work well, try tests that doesn't rely on the constant variance assumption:

- ▶ **Welch test for a contrast**  $\sum_{i=1}^g \omega_i \mu_i$ : The test statistic is

$$t = \frac{\sum_{i=1}^g \omega_i \bar{y}_{i\bullet}}{\sqrt{\sum_{i=1}^g \omega_i^2 s_i^2 / n_i}}$$

which is approximately a  $t$ -distribution with  $\nu$  degrees of freedom, where

$$\nu = \frac{(\sum_{i=1}^g \omega_i^2 s_i^2 / n_i)^2}{\sum_{i=1}^g \frac{1}{n_i - 1} \frac{\omega_i^4 s_i^4}{n_i^2}}$$

Observe that this is a generalization of the two-sample test without equal variance assumption.

## Part III — Checking for Dependent Errors

- ▶ Among the 3 assumptions, violation of the independence assumption causes severest problem. Most of our the analysis (ANOVA, test of contrasts, multiple comparisons, etc.) have little tolerance on dependence of errors
- ▶ There are various forms of dependence, *serial dependence* and *spatial dependence* are two common ones
- ▶ **Remedies for Dependence**
  - ▶ There isn't much we can do about dependence using our current machinery, since no simple transformation can remove dependence.
  - ▶ Analysis of dependent data requires tools like **time series** or **spatial statistics**, which is beyond the scope of this class



## Example: Standard Gravity

The National Bureau of Standards performed 8 series of experiments in 1924-1935 to determine  $g$ , the standard gravity.

The data are given in the table below (in deviations from  $9.8m/s^2 \times 10^5$ , e.g., the first measurement of  $g$  is  $9.80076 m/s^2$ ), with series 1 representing the earliest set of experiments and series 8 the last.

Series	Measurements												
1	76	82	83	54	35	46	87	68					
2	87	95	98	100	109	109	100	81	75	68	67		
3	105	83	76	75	51	76	93	75	62				
4	95	90	76	76	87	79	77	71					
5	76	76	78	79	72	68	75	78					
6	78	78	78	86	87	81	73	67	75	82	83		
7	82	79	81	79	77	79	79	78	79	82	76	73	64
8	84	86	85	82	77	76	77	80	83	81	78	78	78

## Weird ANOVA $F$ -Test

```
> g = c(76,82,83,54,35,46,87,68,87,95,98,100,109,109,100,81,75,68,67,
105,83,76,75,51,76,93,75,62,95,90,76,76,87,79,77,71,
76,76,78,79,72,68,75,78,78,78,78,86,87,81,73,67,75,82,83,
82,79,81,79,77,79,79,78,79,82,76,73,64,
84,86,85,82,77,76,77,80,83,81,78,78,78)
> series = c(rep(1,8),rep(2,11),rep(3,9),rep(4,8),rep(5,8),rep(6,11),
rep(7,13),rep(8,13))
> lmg = lm(g ~ as.factor(series))
> anova(lmg)
Response: g
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(series)	7	2818.6	402.66	3.5675	0.002357 **
Residuals	73	8239.4	112.87		

ANOVA rejects the  $H_0$  of the 8 series having equal means. What does this mean? Will you conclude that

- (a)  $g$  had changed in the 8 series of measurements, or
- (b) the ANOVA  $F$ -test failed?

If your answer is (a), how do you explain the change of  $g$ ?

If your answer is (b), why the ANOVA  $F$ -test failed?

## Example: Standard Gravity

The National Bureau of Standards<sup>1</sup> is the government agency that measures things. The following statement is taken from the NIST website:

*Founded in 1901, NIST is a non-regulatory federal agency within the U.S. Department of Commerce. NIST mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life.*

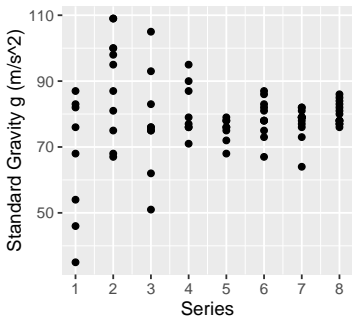
Thus, it is safe to assume that the NBS scientists were trying hard to measure the same quantity  $g$  (e.g., all experiments were done in the same location) throughout all eight series of experiments.

---

<sup>1</sup>now called the National Institute of Standards and Technology

```
> qplot(series, g, xlab="Series",  
  ylab="Standard Gravity g (m/s^2)") +  
  scale_x_continuous(breaks=1:8)
```

Variance decreases with series,  
which makes sense since the  
accuracy of measurement  
improves as time goes by.



The ANOVA  $F$ -test here may not be reliable because at least the constant variance assumption is not met

```
> round(sd(g ~ series), 2)  
  1      2      3      4      5      6      7      8  
19.25 15.29 15.76  8.30  3.65  5.84  4.74  3.36
```

Will a transformation work here? Box-Cox?

## Brown-Forsythe Modified $F$ -test

In view of the nonconstant variability, let's try the Brown-Forsythe modified  $F$ -test

$$BF = \frac{SS_{Trt}}{\sum_{i=1}^g s_i^2(1 - n_i/N)}$$

The numerator  $SS_{Trt} = 2819$  can be found in the ANOVA table above. The denominator is found using  $R$  (see the codes below) to be 888.5747

```
> sds = sd(g ~ series); sds
      1      2      3      4      5      6      7      8
19.24977 15.29349 15.75595  8.29694  3.65474  5.83874  4.73665  3.35506
> ni = c(8, 11, 9, 8, 8, 11, 13, 13)
> di = sds^2*(1-ni/sum(ni))
> BFbottom = sum(di)
> BFbottom
[1] 888.5747
> BF = 2819/BFbottom
```

The  $BF$ -statistic is thus  $BF = \frac{2819}{888.5747} = 3.1725$ .

Under the null hypothesis of equal treatment means, BF is approximately distributed as an  $F$ -distribution with  $g - 1$  and  $\nu$  degrees of freedom, where the degrees of freedom  $\nu$  is calculated as 29.46 in the R code below.

$$\nu = \frac{(\sum_{i=1}^g d_i)^2}{\sum_{i=1}^g d_i^2 / (n_i - 1)} \quad \text{in which} \quad d_i = s_i^2(1 - n_i/N)$$

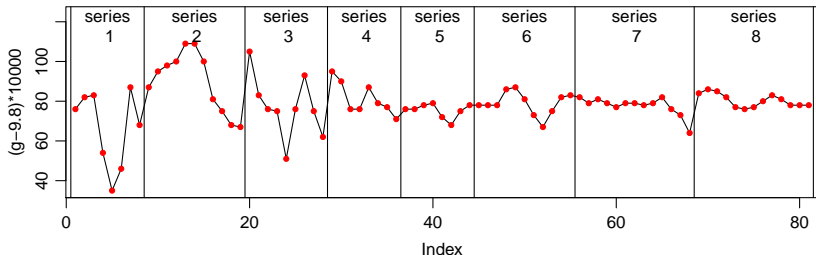
```
> nu = (BFbottom)^2/sum(di^2/(ni-1))
> nu
[1] 29.46249
> 1 - pf(BF, df1 = 7, df2 = nu)           # P-value of the BF-test
[1] 0.01263341
```

However, the BF-test, not relying on the constant variance assumption, also rejects the null hypothesis of equal mean at a  $P$ -value 0.0126. Why the BF-test also failed?

# Tools for Checking Serial Dependence

1. **Time plot:** a plot of residuals v.s. the order they are measured)
  - ▶ It's better to keep track of the order units are measured.
  - ▶ A smooth time-plot is a sign of positive serial dependence, since a smooth time plot means successive residuals are too close together
2. **Autocorrelation Plots**
  - ▶ Lag 1 autocorrelation plot: plotting  $(e_1, \dots, e_{n-1})$  v.s.  $(e_2, \dots, e_n)$
  - ▶ Lag  $k$  autocorrelation plot: plotting  $(e_1, \dots, e_{n-k})$  v.s.  $(e_{1+k}, \dots, e_n)$
  - ▶ Any trend in the autocorrelation plot is a sign of serial dependence.
3. **Autocorrelation**
  - ▶ the Lag- $k$  *autocorrelation coefficient* is the correlation coefficient of  $(e_1, \dots, e_{n-k})$  v.s.  $(e_{1+k}, \dots, e_n)$ ,  $k = 1, 2, 3, \dots$

The observations in each series are in fact given in time order taken. We can thus make a time plot.



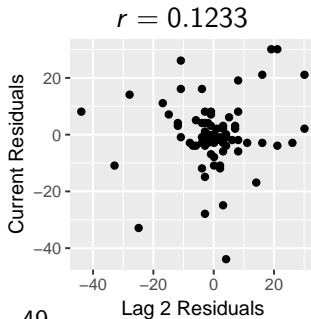
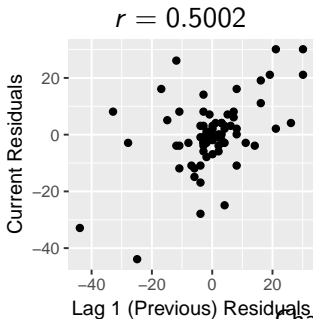
We can see a lot of measurements are close to the previous measurements, which indicates a positive serial correlation.

It's not surprising that scientists in NIST might unconsciously match their results with the previous measurement, which was often regarded as the most accurate one till then.



# Autocorrelation Plots

```
> lm0 = lm(g ~ 1) # Null model: All series had the same mean
> res = lm0$res # Residual of null model
> qqplot(res[2:81],res[1:80], xlab="Current Residual",
  ylab="Lag 1 (Previous) Residual")
> cor(res[2:81],res[1:80])
[1] 0.5002454
> qqplot(res[1:79], res[3:81], ylab="Current Residuals",
  xlab="Lag 2 Residuals")
> cor(res[3:81],res[1:79])
[1] 0.1232966
```



## Effect of Dependent Errors

Positive serial correlation make the value of one measurement closer to that of the previous measurement and hence reduces the within group variability (SSE) and enlarges the  $F$ -statistic, so it is easier to reject the null.

# Spatial Dependence

*Spatial dependence* can arise when the experimental units are arranged in space, like plants in a farm. Spatial dependence occurs when units that are closer together are more similar than units farther apart.