



STAT 571A — Advanced Statistical Regression Analysis

Chapter 3 NOTES Diagnostics and Remedial Measures

© 2015 University of Arizona Statistics GDP. All rights reserved, except where previous rights exist. No part of this material may be reproduced, stored in a retrieval system, or transmitted in any form or by any means — electronic, online, mechanical, photoreproduction, recording, or scanning — without the prior written consent of the course instructor.

Predictor Diagnostics

- It's always a good idea to study the nature of the predictor variable, X_i , and ensure it's not harboring any surprises.
- Useful tools:
 - **Dot plots**
 - **Sequence plots** (scatterplot of X_i vs. i with dots connected — useful if i is a surrogate for time)
 - **Stemplots**
 - **Boxplots**

Residual Diagnostics

- The **residuals** from the LS fit for a Simple Linear Regression (SLR) model are the differences between the observed response and the predicted response:

$$e_i = Y_i - \hat{Y}_i.$$

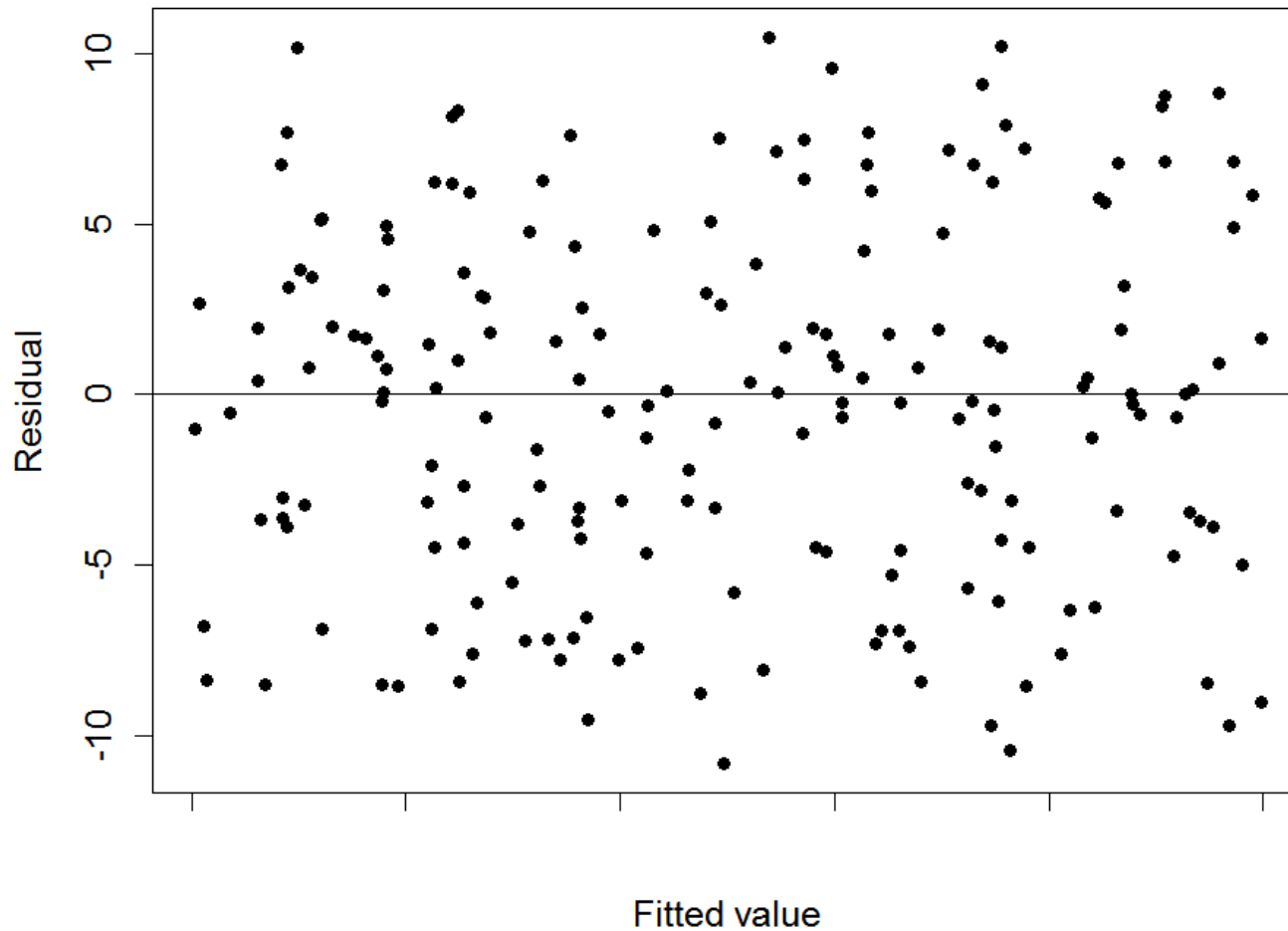
- The e_i 's contain useful info. Recall that $\sum e_i = \sum e_i X_i = \sum e_i \hat{Y}_i = 0$. Thus $\bar{e} = \sum e_i / n = 0$ while

$$s_e^2 = \frac{1}{n-2} \sum (e_i - \bar{e})^2 = \frac{1}{n-2} \sum e_i^2 = \frac{\text{SSE}}{n-2} = \text{MSE}$$

Residual Plots

- A powerful diagnostic tool for assessing model fit is the **Residual Plot**, i.e., a plot of $e_i = Y_i - \hat{Y}_i$ vs. \hat{Y}_i .
- If the SLR model fit is adequate, the residuals, e_i , should cluster around the horizontal line $e = 0$, with no apparent pattern. See, e.g., next slide →

Prototypical Residual Plot with Random Scatter



Residual Plots (cont'd)

Can also use residual plots to study:

- **departure from linearity in $E\{Y\}$**
- **departure from constant σ^2 in $\sigma^2\{Y\}$**
- **departure from indep. assumption**
- **outlying values of Y_i**
- **departure from normality in $\varepsilon_i \sim N(0, \sigma^2)$**
- **other values of X that could be useful
(see Ch. 6)**

Residual Plots (cont'd)

Alternate versions of the residual plot include:

- plot e_i vs. \hat{Y}_i (std. version)
- plot e_i vs. X_i
- plot $|e_i|$ or e_i^2 vs. X_i
- plot e_i vs. i (if sequence is surrogate for time)
→ see Fig. 3.2b
- boxplot and/or histogram of e_i 's → see Fig. 3.2c
- normal probab. plot of e_i 's → see Fig. 3.2d

Residual Plots in R

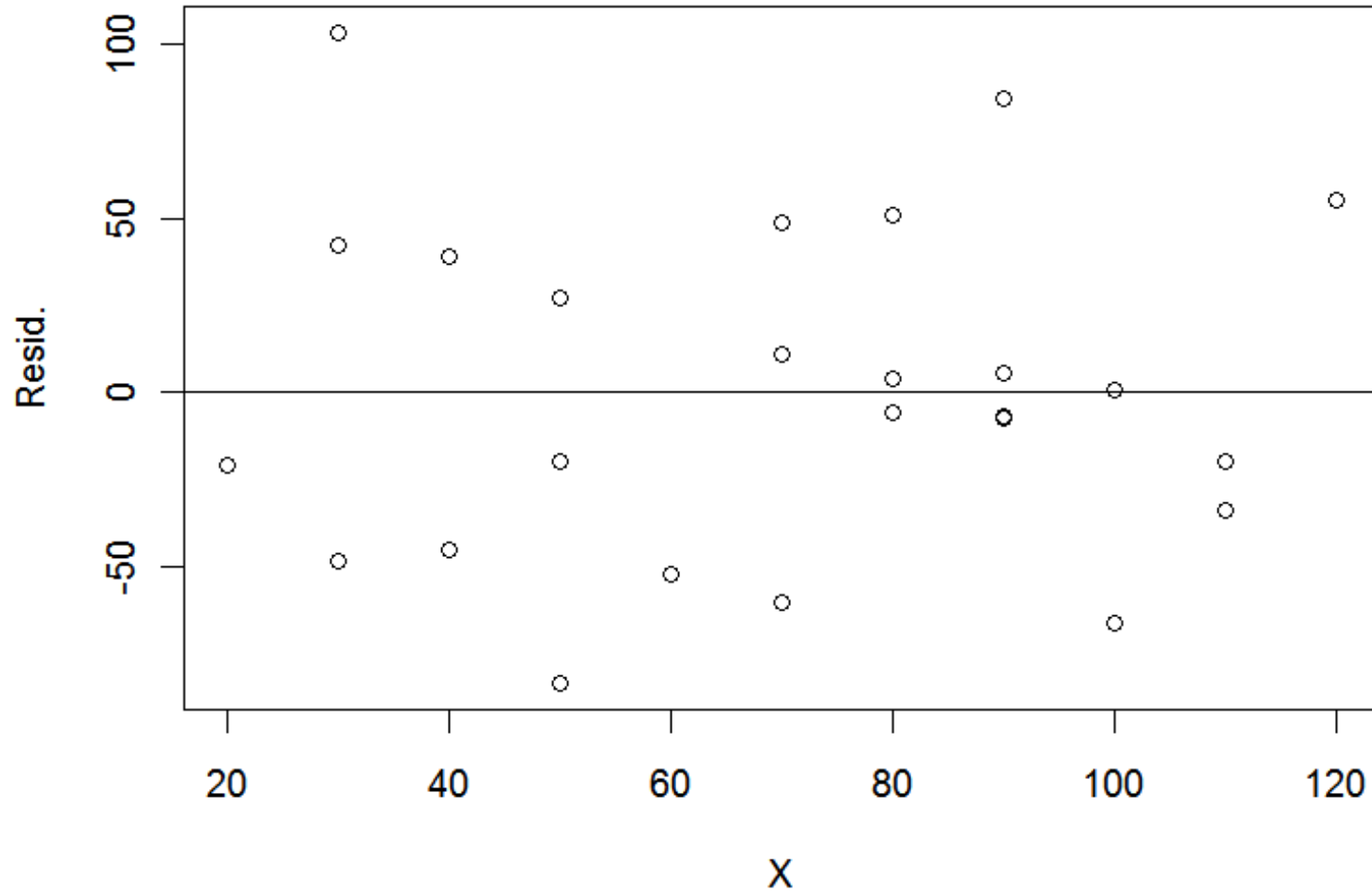
- In R, the `plot()` function can extract the raw residual plot from the `lm()` object.
- Or, use `plot()` directly with the functions `resid()` and `fitted()`.

Toluca Example (cont'd)

In the Toluca Data example (CH01TA01), we produce a residual plot via:

```
> CH01TA01.lm = lm( Y ~ X )  
> plot( resid(CH01TA01.lm) ~ X )  
> abline( h=0 ) #add horiz. e=0 line
```

Toluca Example Residual Plot



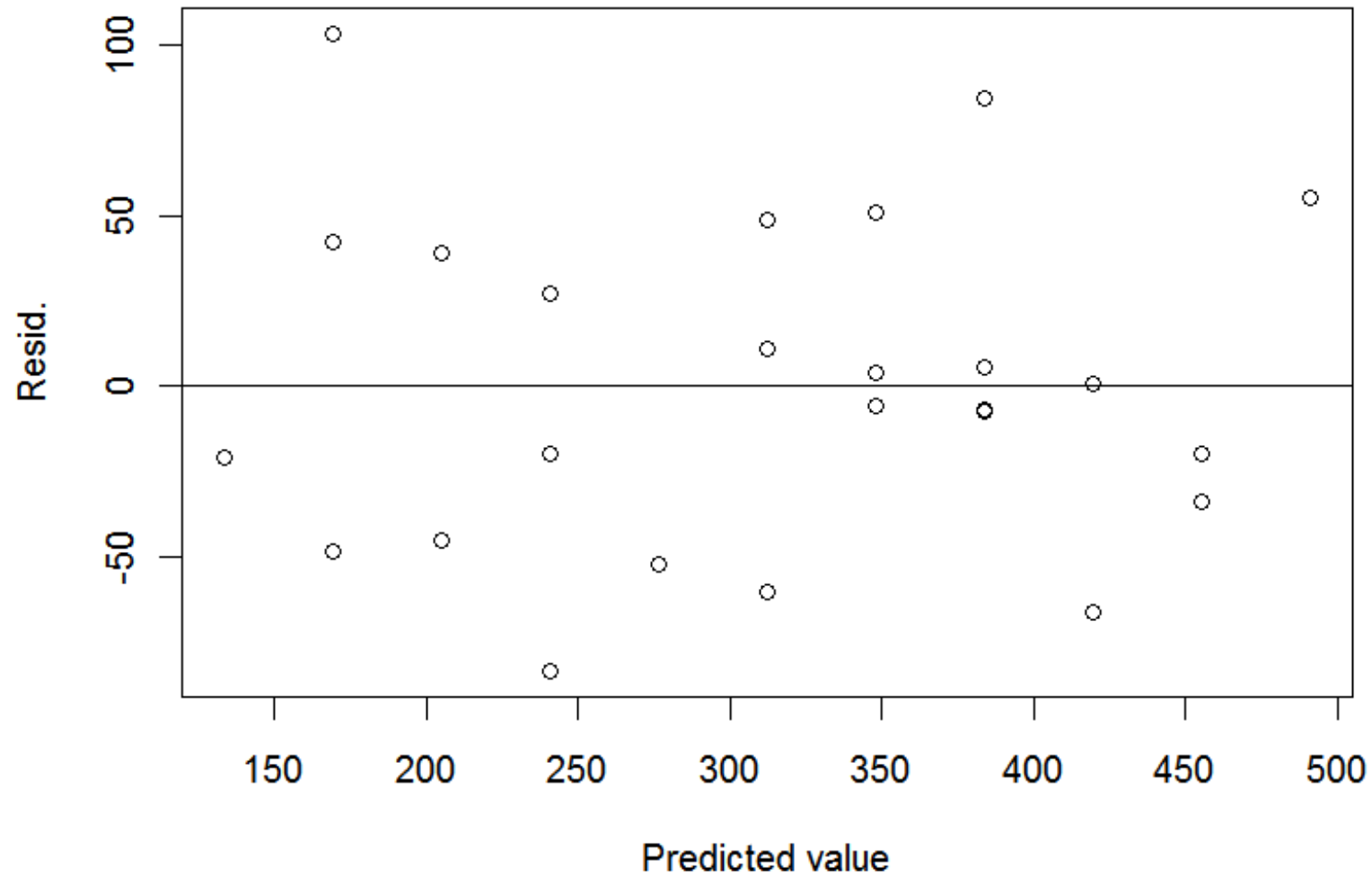
Toluca Resid. Plot (cont'd)

As mentioned, residual plots are usually drawn against the fitted values:

```
> CH01TA01.lm = lm( Y ~ X )  
> plot( resid(CH01TA01.lm) ~  
        fitted(CH01TA01.lm) )  
> abline( h=0 ) #add horiz. e=0 line
```

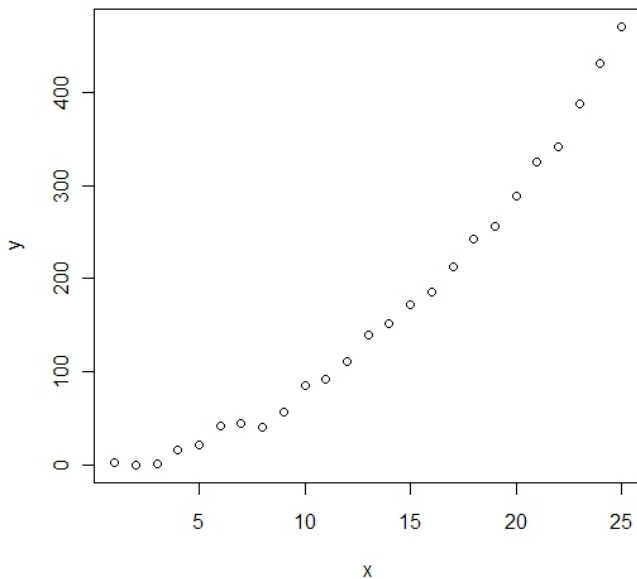
Residual pattern will be identical under an SLR model (reversed if $b_1 < 0$).

Example TA01 Residual Plot

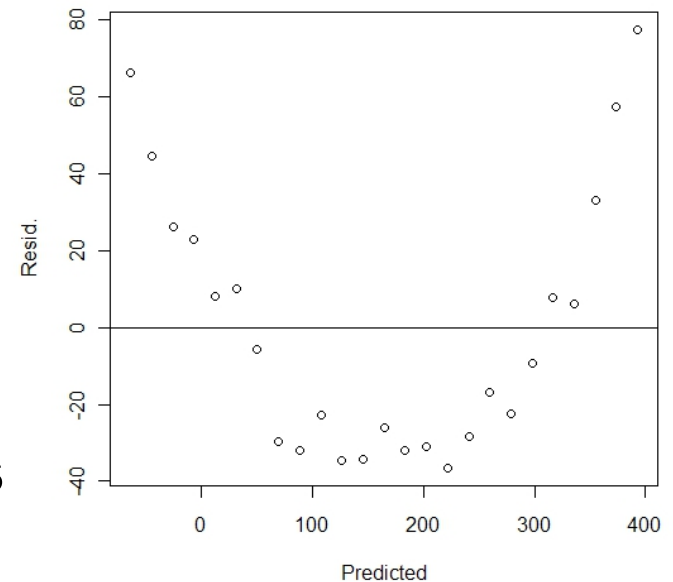


Quadratic Curvature

Curvature in the scatterplot and the resid. plot may indicate a parabolic (quadratic) pattern:



Scatterplot
shows
curvature

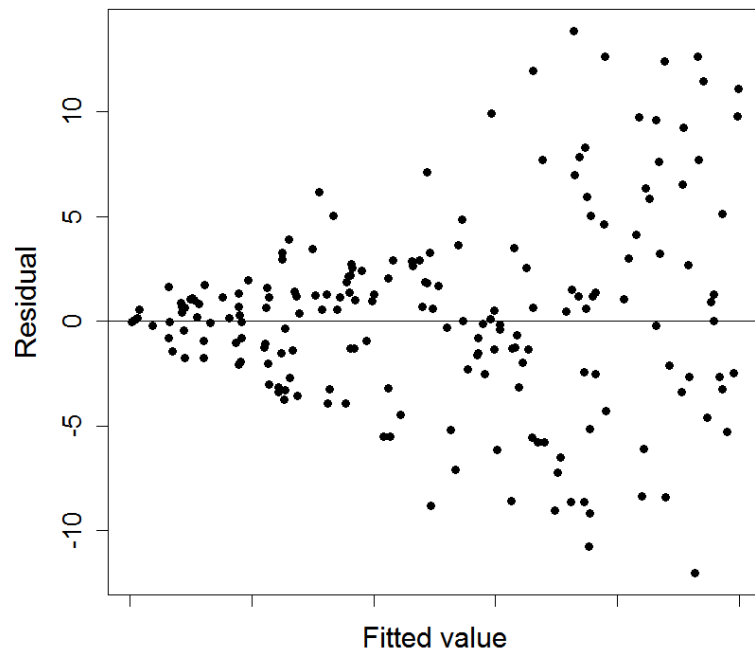


Resid. plot has
striking **down-
then-up** pattern

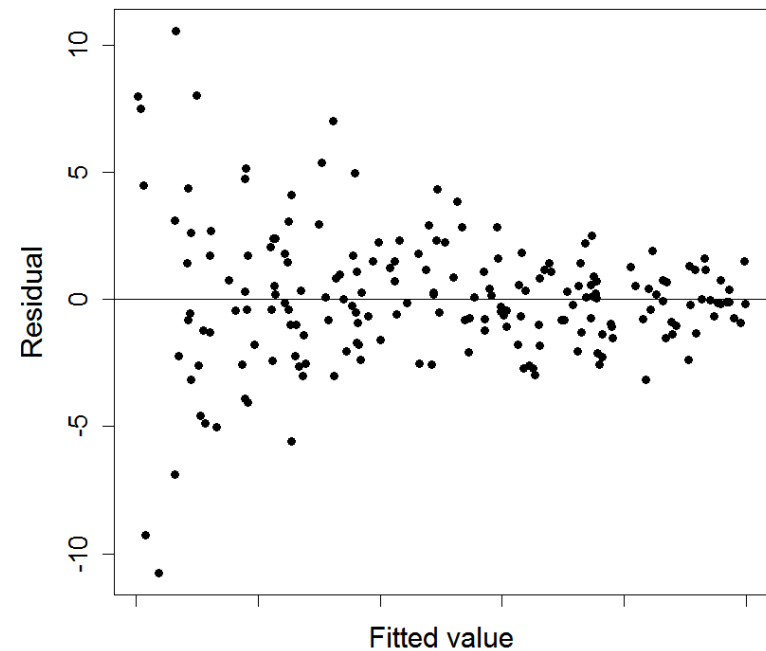
⇒ **Clear parabolic relationship
deviates from linearity assumption**

Variance heterogeneity

Departure from constant σ^2 in $\sigma^2\{Y\}$ will be evident in a widening residual pattern, e.g.

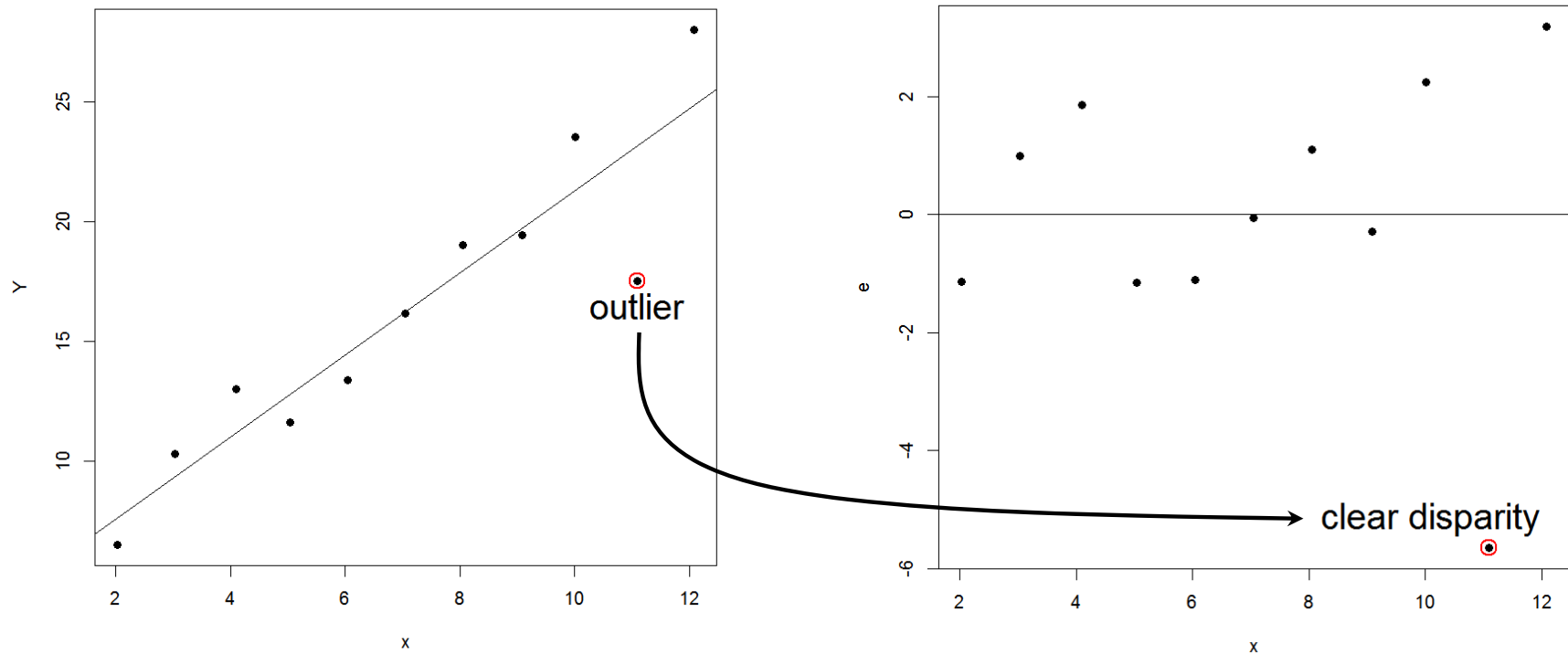


or



Possible Outliers

Unusual observations (possible 'outliers') can be recognized quickly in resid. plots:



Possible Outliers

- **Keep in mind that detection of an unusual observation does not immediately condemn the data point or call for its deletion.**
- **Always examine why the point is flagged: it may involve a simple data entry error or some other explanation.**

Standardizing the Residuals

- A problem with raw residuals is that they are scale/measurement-dependent: in one data set an absolute residual of $|e_i| = 8.2$ may be *less* egregious than a residual of $|e_i| = 0.7$ in another data set.
- We can stabilize residuals across data sets/model fits by standardizing them to similar scales (sort of like a z-score).

Semistudentized Residuals

A ***Semistudentized Residual*** is a raw residual, e_i , divided by its sample std. deviation:

$$e_i^* = \frac{e_i}{s_e} = \frac{e_i}{\sqrt{\text{MSE}}}$$

(Some authors call this a ***Studentized Residual***. Others don't! Be careful of terminology.)

There are **more effective ways** to adjust resid's; we expand on this in Ch. 10.

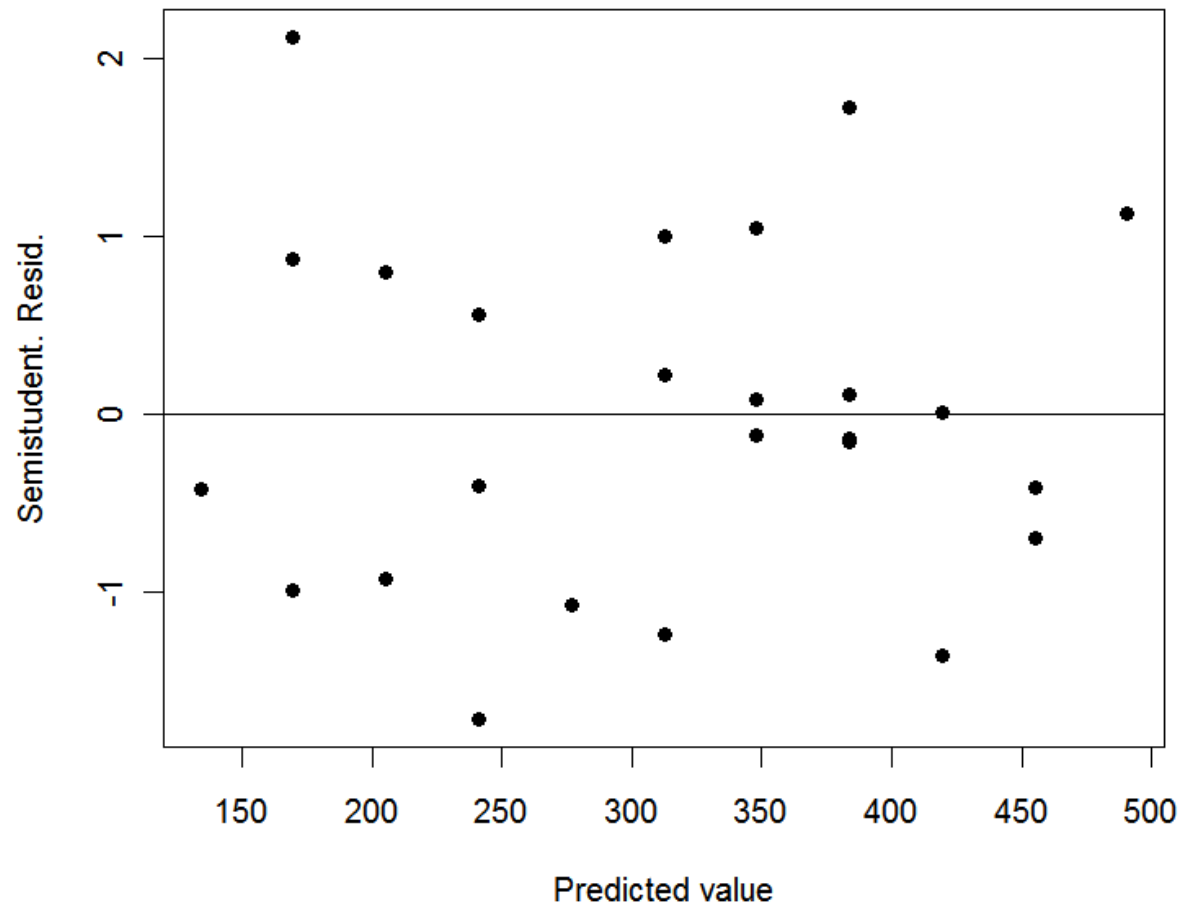
Semistudentized Residuals in R

- In R, we can plot the e_i^* s.
- E.g., with the Toluca data (CH01TA01), use

```
> CH01TA01.lm = lm( Y ~ X )  
> eistar = resid( CH01TA01.lm )/  
          sqrt( anova(CH01TA01.lm)$"Mean Sq"[2] )  
> plot( eistar ~ fitted(CH01TA01.lm) )  
> abline( h=0 )
```

Toluca Example: Semistudentized Residual Plot

Pattern is similar to raw residual plot, but now on a standardized scale.



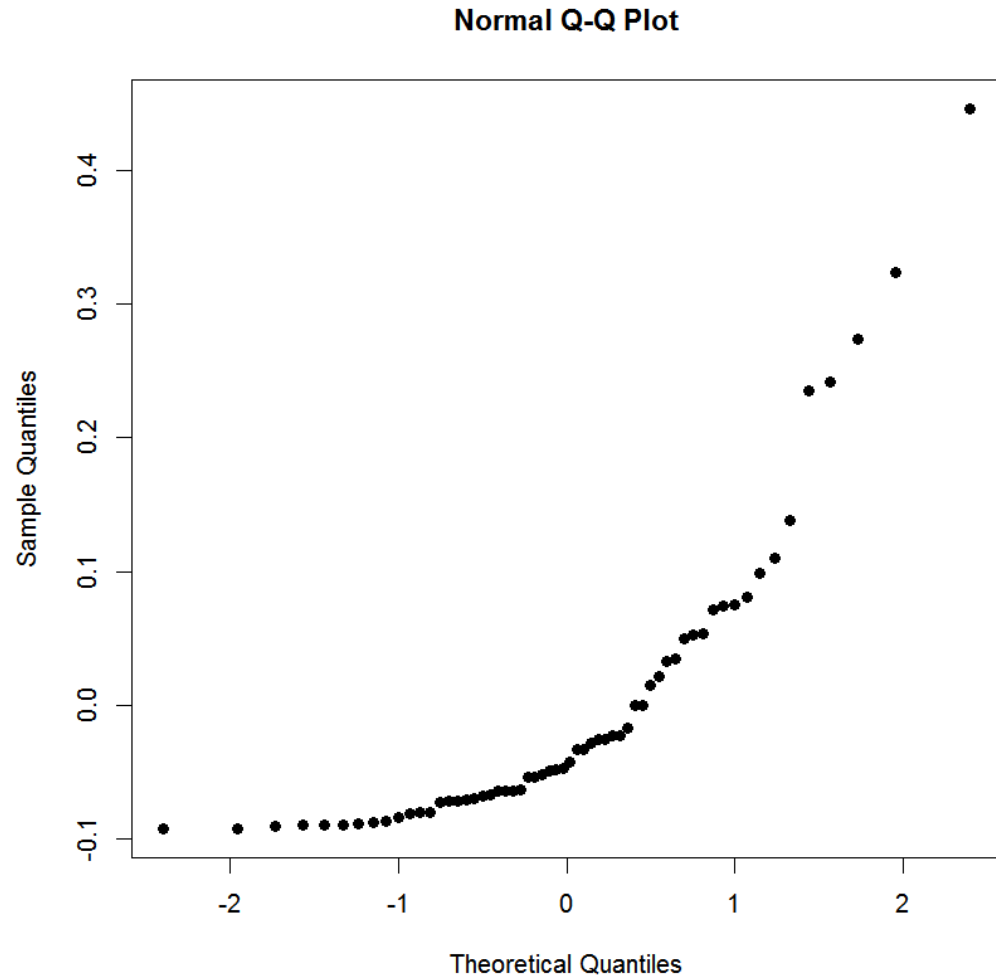
Normal Probability Plots

- To assess departure from normality in $\varepsilon_i \sim N(0, \sigma^2)$, display the e_i 's via a boxplot, histogram, or normal probab. plot (NPP).
- E.g., with NPP's, if normality holds, the NPP will show a 45° line.
- Departure from normality will be evident as deviations from the 45° line.

(See Fig. 3.9)

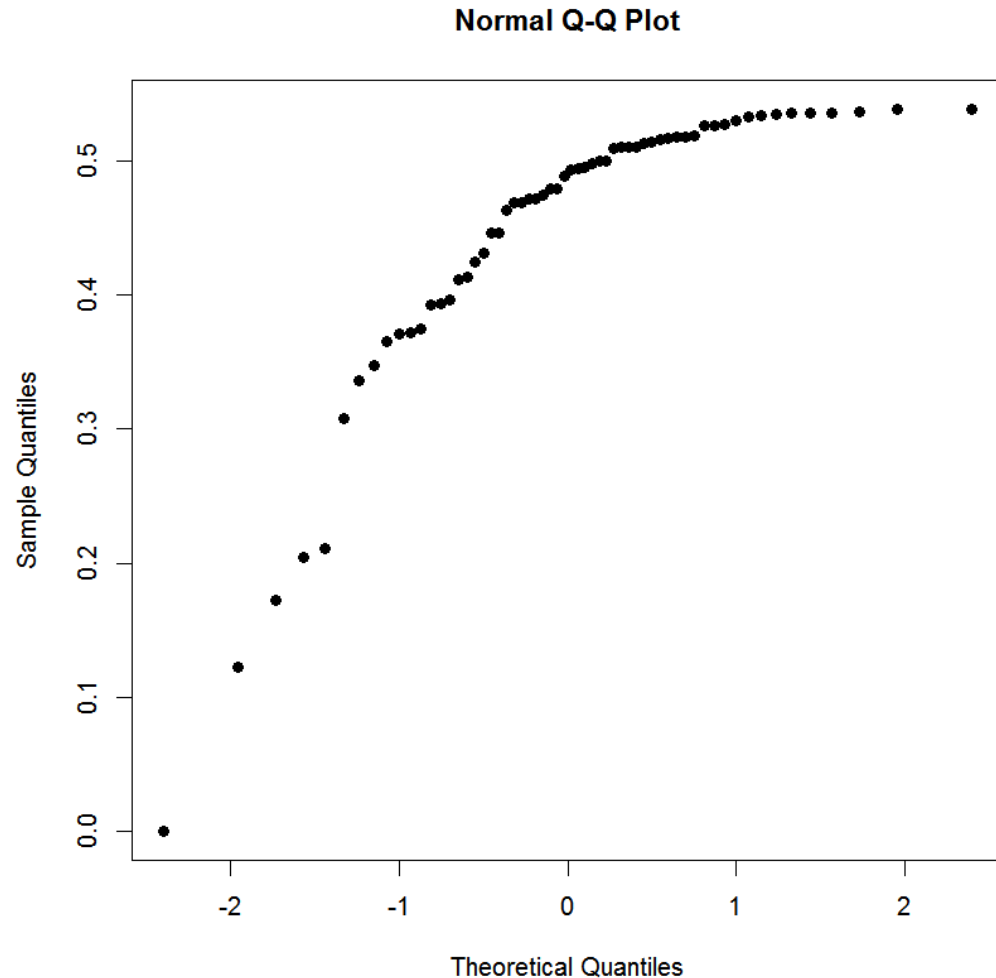
Right Skew in NP Plot

**Right-skewed
residuals
display convex
shape in NPP**



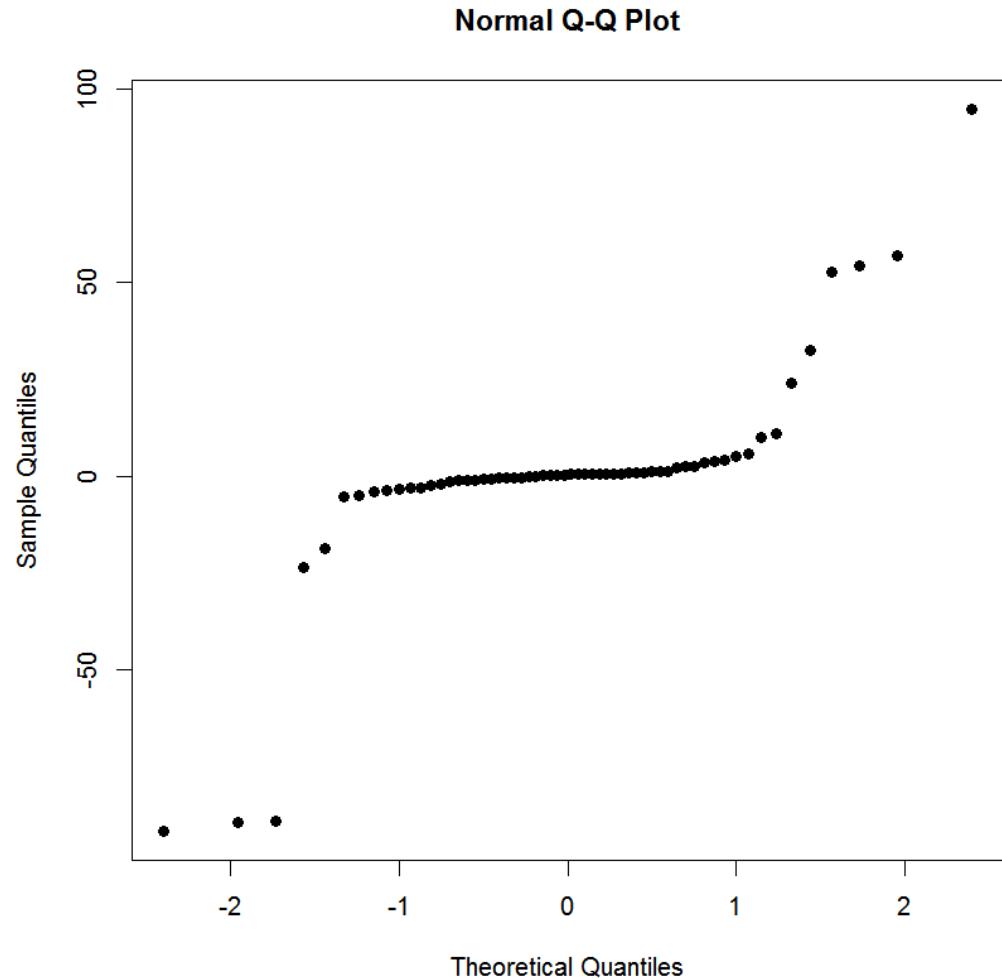
Left Skew in NP Plot

**Left-skewed
residuals
display
concave shape
in NPP**



Thick Tails in NP Plot

**Thick-tailed
(but still
symmetric)
residuals
display as
flattened
shape in NPP**



Test of Normality

- Besides visualizations, we can apply a univariate **test of Normality** to the e_i 's, and assess whether the ε_i 's appear to deviate from the $N(0, \sigma^2)$ assumption.
- The basic test is due to Shapiro and Wilk. Applied to the residuals, it gives a test statistic (W^*) and an associated P-value.
- If P is too small (below a pre-set α -level), reject the null hypothesis of normality.

Toluca Example (cont'd)

In the Toluca example (CH01TA01), we test the residuals for normality:

```
> CH01TA01.lm = lm( Y ~ X )  
> shapiro.test( resid(CH01TA01.lm) )
```

```
Shapiro-Wilk normality test  
data:  resid(CH01TA01.lm)  
W = 0.9789, p-value = 0.8626
```

P-value is large, so no departure from normality is evidenced.

Test of Variance Homogeneity

- From an SLR model, we **test the residuals for variance homogeneity** (i.e., constant σ^2) via the 'robust' Brown-Forsythe test, a special case of the more-general Levene test.
- The test breaks the e_i 's into 2 groups: (1) e_i 's from small X_i 's vs. (2) e_i 's from large X_i 's. Then, it compares deviations from median between the 2 groups.

Brown-Forsythe Test

- For (small- X) group 1, let $n_1 = \#$ resid's from small X_i 's, and denote the median resid. in that group as \tilde{e}_1 . Compute $d_{i1} = |e_{i1} - \tilde{e}_1|$.
- For (large- X) group 2, let $n_2 = \#$ resid's from large X_i 's, and denote the median resid. in that group as \tilde{e}_2 . Compute $d_{i2} = |e_{i2} - \tilde{e}_2|$.

- Calculate $t_{BF}^* = \frac{\bar{d}_1 - \bar{d}_2}{S_D \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ with

$$S_D^2 = \frac{\sum_{i=1}^{n_1} (d_{i1} - \bar{d}_1)^2 + \sum_{i=1}^{n_2} (d_{i2} - \bar{d}_2)^2}{n - 2}$$

Brown-Forsythe Test (cont'd)

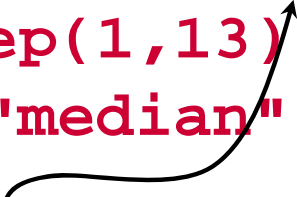
- Then, conclude significant departure from homogeneous variance if $|t_{BF}^*| > t(1 - \frac{\alpha}{2}; n-2)$.
- P-value is $2P[t(n-2) > |t_{BF}^*|]$
- Can do this in R, but must first download and install the *lawstat* package, then load it into the current workspace:

```
> require( lawstat )
```

Toluca Example (cont'd)

In the Toluca Data example (CH01TA01), test the residuals for variance homogeneity:

```
> ei = resid( CH01TA01.lm )  
  
> BF.htest = levene.test( ei[order(X)],  
                           group=c(rep(1,13),rep(2,12)),  
                           location="median" )
```



Note use of `ei[order(X)]` to sort residuals by order of X. (Could also sort by \hat{Y} .)

The `group=` specifier factors into 2 groups for constructing the t^* statistic. (Actually, it gives $F^* = t^{*2}$.)

Toluca Example (cont'd)

Browth-Forsythe output:

modified robust Brown-Forsythe Levene-type test
based on the absolute deviations from the median
data: `ei[order(X)]`

Test Statistic = 1.7331, p-value = 0.201

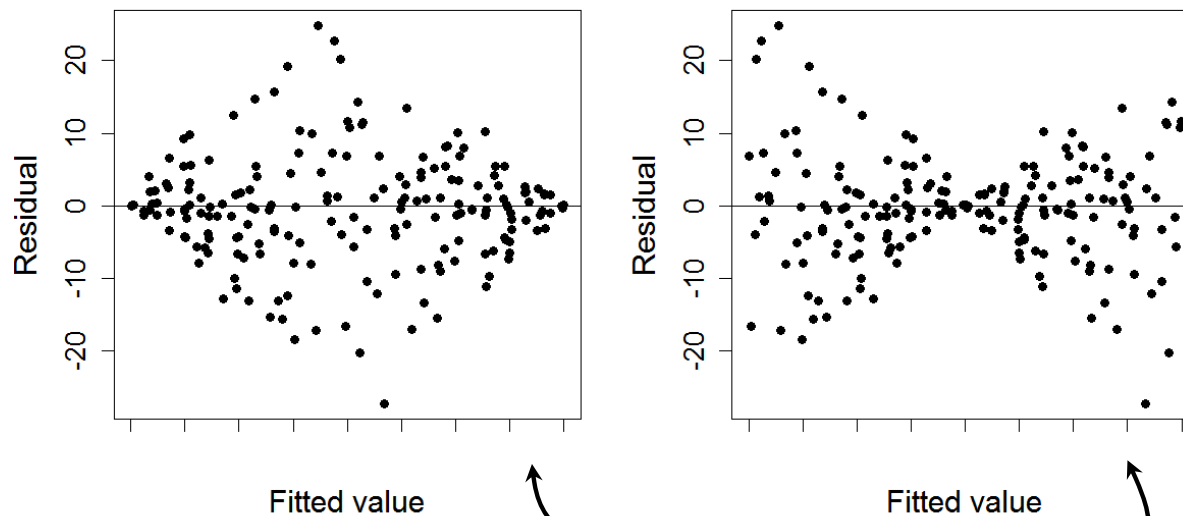
```
> sqrt( BF.htest$statistic ) #for t* statistic
```

Test Statistic
1.316482

**$P = 0.201 > 0.05 = \alpha$. No significant
departure from constant σ^2 is evidenced.**

Brown-Forsythe Test: Caveats

- The Brown-Forsythe test works if variance heterogeneity is of the “megaphone” style (in the textbook’s terminology).
- But (!), it may fail with stranger patterns, such as two megaphones back-to-back:



Brown-Forsythe will fail to detect these patterns

Replication

- When there are multiple Y_i 's observed at the same X_i , we say the data/design exhibits **Replication**.
- The multiple Y_i 's are called **Replicates**.
- In this case, it is possible to cleanly test the 'fit' of an assumed model for $E\{Y_i\}$.

Lack Of Fit (LOF) Testing

Need to expand our notation:

the observations are now indexed as Y_{ij} ,

where $i = 1, \dots, n_j$ (replicates)

and $j = 1, \dots, c$ (unique X values)

Sample size is $n = \sum_{j=1}^c n_j$.

We approach the lack-of-Fit (LOF) test using the FM vs. RM discrepancy measure from §2.7.

LOF Testing (cont'd)

- The Full Model (FM) here is $Y_{ij} = \mu_j + \varepsilon_{ij}$ with $\varepsilon_{ij} \sim \text{i.i.d. } N(0, \sigma^2)$.
- The model allows for $c > 1$ different, unspecified means, μ_j , at each j .
- The Reduced Model (RM) is the SLR:
$$Y_{ij} = \beta_0 + \beta_1 X_j + \varepsilon_{ij}$$
with $\varepsilon_{ij} \sim \text{i.i.d. } N(0, \sigma^2)$
(or whatever posited model is under study).

LOF Testing (cont'd)

- To construct the F-statistic use (2.70): compare $SSE(R)$ under the SLR with $SSE(F)$ under the full, unspecified model.
- If the SSE's differ too much, this implies that significant, explainable variation is still present, even after fitting the SLR model.
- Use the F-statistic to quantify this departure.

LOF Testing (cont'd)

- Recall: the general form of the discrepancy F-statistic is

$$F^* = \frac{\text{SSE}(R) - \text{SSE}(F)}{df_{ER} - df_{EF}} \bigg/ \frac{\text{SSE}(F)}{df_{EF}}$$

- For LOF, the numerator is the mean square for lack-of-fit: MSLF. The denominator is MSE(F).
- Reject the null hypoth. of acceptable fit if $F^* > F(1-\alpha; df_{ER} - df_{EF}, df_{EF})$.

LOF Testing (cont'd)

Notice: The RM SSE is based on the SLR,

$$\begin{aligned}\text{SSE(R)} &= \sum_{j=1}^c \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_{ij}(\text{R}))^2 \\ &= \sum_{j=1}^c \sum_{i=1}^{n_j} (Y_{ij} - b_0 - b_1 X_j)^2\end{aligned}$$

The FM SSE has similar form: $\text{SSE(F)} = \sum_{j=1}^c \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_{ij}(\text{F}))^2$, but what is $\hat{Y}_{ij}(\text{F})$??

Answer: The FM makes no specification on $E[Y_{ij}]$, so the LS estimators are simply the replicate means $\hat{Y}_{ij}(\text{F}) = \bar{Y}_j$.

Pure Error

With this, find $SSE(F) = \sum_{j=1}^c \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$. This is called the **Pure Error**, and denoted as **SSPE**. Then,

$$MSE(F) = MSPE = \sum_{j=1}^c \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2 / (n-c).$$

The F-statistic becomes

$$F_{LOF}^* = \frac{SSE(R) - SSPE}{(n-2) - (n-c)} / MSPE.$$

Reject the null hypoth. of acceptable fit if $F^* > F(1-\alpha; c-2, n-c)$.

LOF ANOVA Table

We collect all these terms together into a special **LOF ANOVA table**:

Source	d.f.	SS	MS
Regr.	1	$SSR = \sum \sum (\hat{Y}_{ij} - \bar{Y})^2$	MSR
Error	n-2	$SSE(R) = \sum \sum (Y_{ij} - \hat{Y}_{ij})^2$	MSE(R)
LOF	c-2	$SSLF = \sum \sum (\bar{Y}_j - \hat{Y}_{ij})^2$	MSLF
Pure Error	n-c	$SSPE = \sum \sum (Y_{ij} - \bar{Y}_j)^2$	MSPE
Total	n-1	$SSTO = \sum \sum (Y_{ij} - \bar{Y})^2$	

The F-test employs the LOF statistic

$$F_{LOF}^* = MSLF/MSPE.$$

Example: Lack of Fit (LOF) test

Example: Bank Data (CH03TA04)

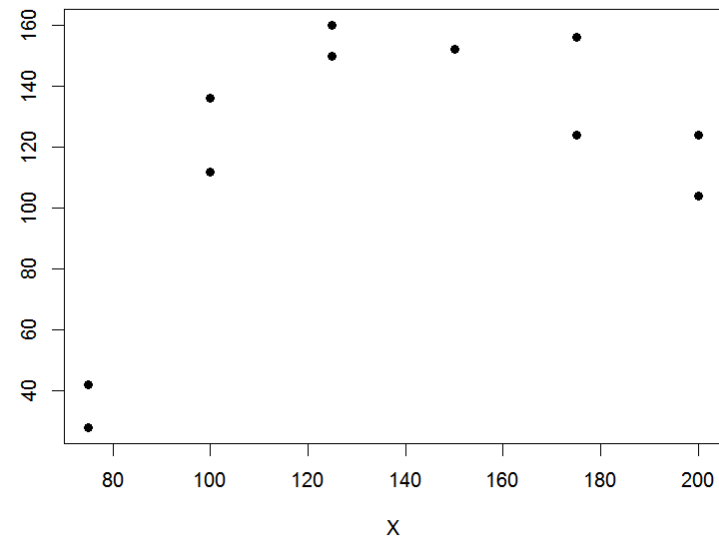
Y = Number of new accts.,

X = Min. deposit.

```
> X = c(125, 100, ... , 100)
```

```
> Y = c(160, 112, ... , 136)
```

```
> plot( Y~X, pch=19 )
```



Example CH03TA04 (cont'd)

To test for LOF with the Bank Data (CH03TA04), can use:

```
> # fit reduced model
> rmCH03TA04.lm = lm( Y ~ X )
>
> # fit full model via factor() command
> fmCH03TA04.lm = lm( Y ~ factor(X) )
```

Example CH03TA04 (cont'd)

```
> #SLR anova component
```

```
> anova( rmCH03TA04.lm )
```

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	1	5141.3	5141.3	3.1389	0.1102
Resid	9	14741.6	1638.0		

Example CH03TA04 (cont'd)

```
> #SSPE with F* for LOF
```

```
> anova( rmCH03TA04.lm, fmCH03TA04.lm )
```

Analysis of Variance Table

Model 1: Y ~ X

Model 2: Y ~ factor(X)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	9	14742				
2	5	1148	4	13594	14.801	0.005594

F* LOF statistic is given, with corresponding P-value.



Comments on LOF Approach

- We don't need replication at every X_j . As long as at least 1 X_j has replicated Y 's, the test can proceed, if poorly (& assuming the replicated Y 's are different).
- Can show that $E\{\text{MSPE}\} = \sigma^2$ (always). For that matter,
$$E\{\text{MSLF}\} = \sigma^2 + \sum n_j \{\mu_j - (\beta_0 + \beta_1 X_j)\}^2 / (c-2).$$
- Any p -parameter model can be tested in this manner, as long as some replication is present and $c > p$.
- (Also see pp. 126-127)

Remediation

- What if our diagnostics uncover failure in the SLR model?
- If it's just heterogeneous variance, we can move to Weighted Least Squares (WLS) → discussed in Ch. 11.
- If the linearity is in question, we might be able to **transform** Y (or X) to “linearize” the regression. (This can also stabilize non-constant variance.) See next slide →

Transformations

- When the Y-vs.-X relationship does not conform to a simple linear model, it may be possible to **transform** either the X_i 's or the Y_i 's (or both!) to at least approximately satisfy the SLR requirements.
- We don't know this in advance, of course, so always
 - (a) plot the data, and
 - (b) plot/examine the residuals for guidance.

Transforming X to \sqrt{X}

- When a curvilinear relationship is evidenced in the data, a **transform of X** to \sqrt{X} might be called for.

- That is, replace the SLR model with

$$Y_i = \beta_0 + \beta_1 X_i^{1/2} + \varepsilon_i \quad (i = 1, \dots, n).$$

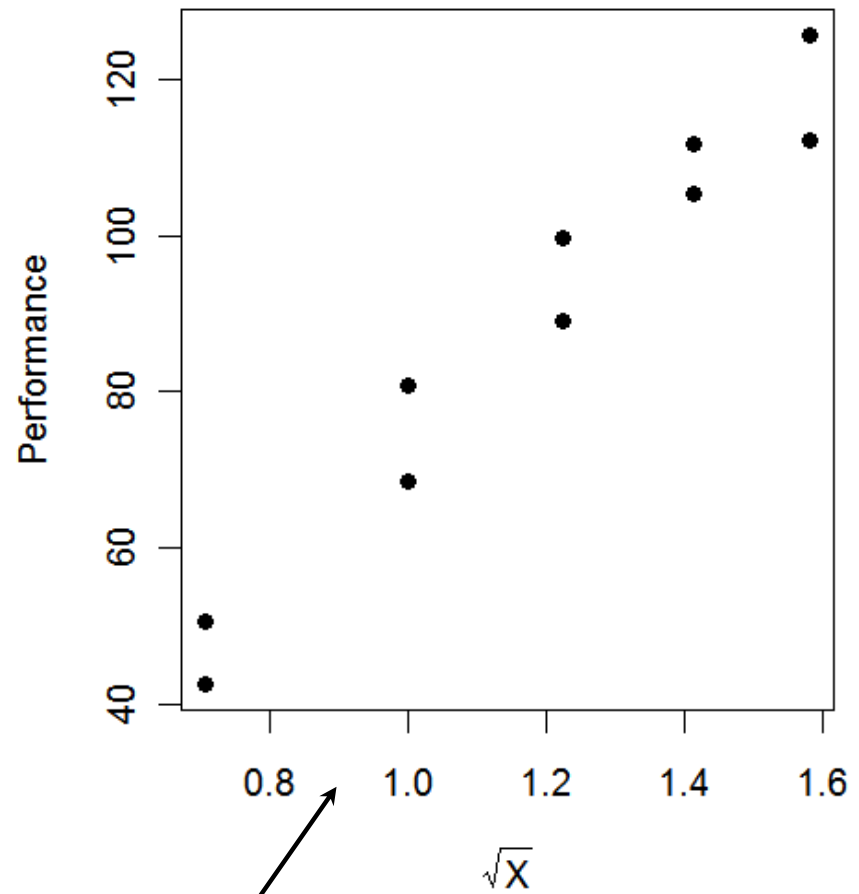
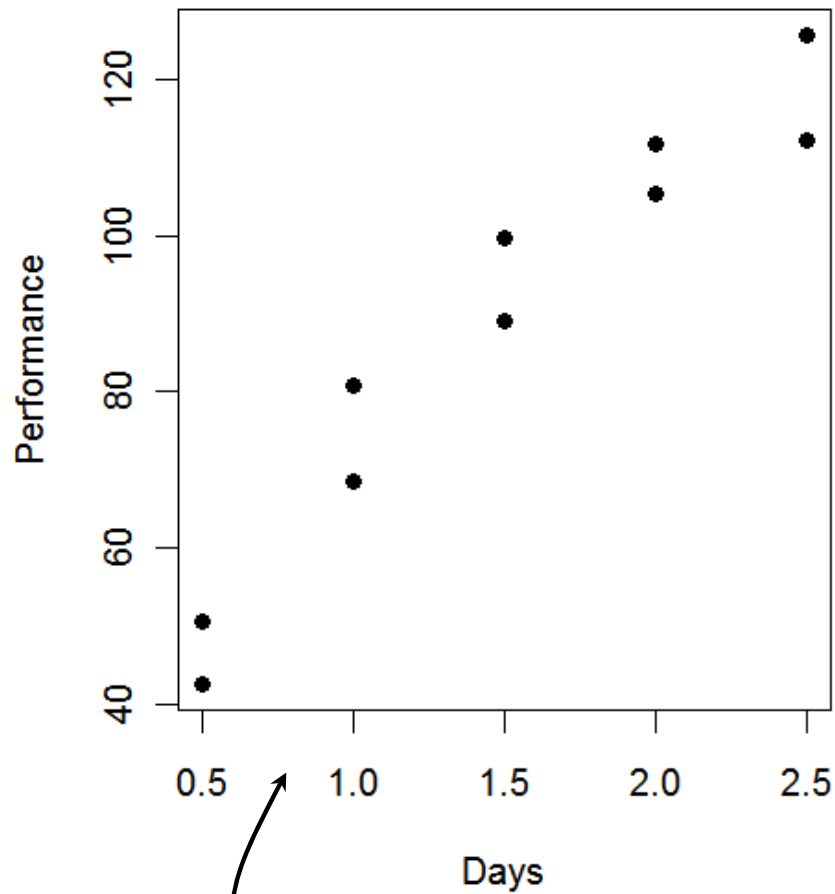
- Sales Training example (CH03TA07):

```
> par( mfrow=c(1,2) )
```

```
> plot( Y ~ X )
```

```
> plot( Y ~ sqrt(X) )
```


Sales Training (CH03TA07)



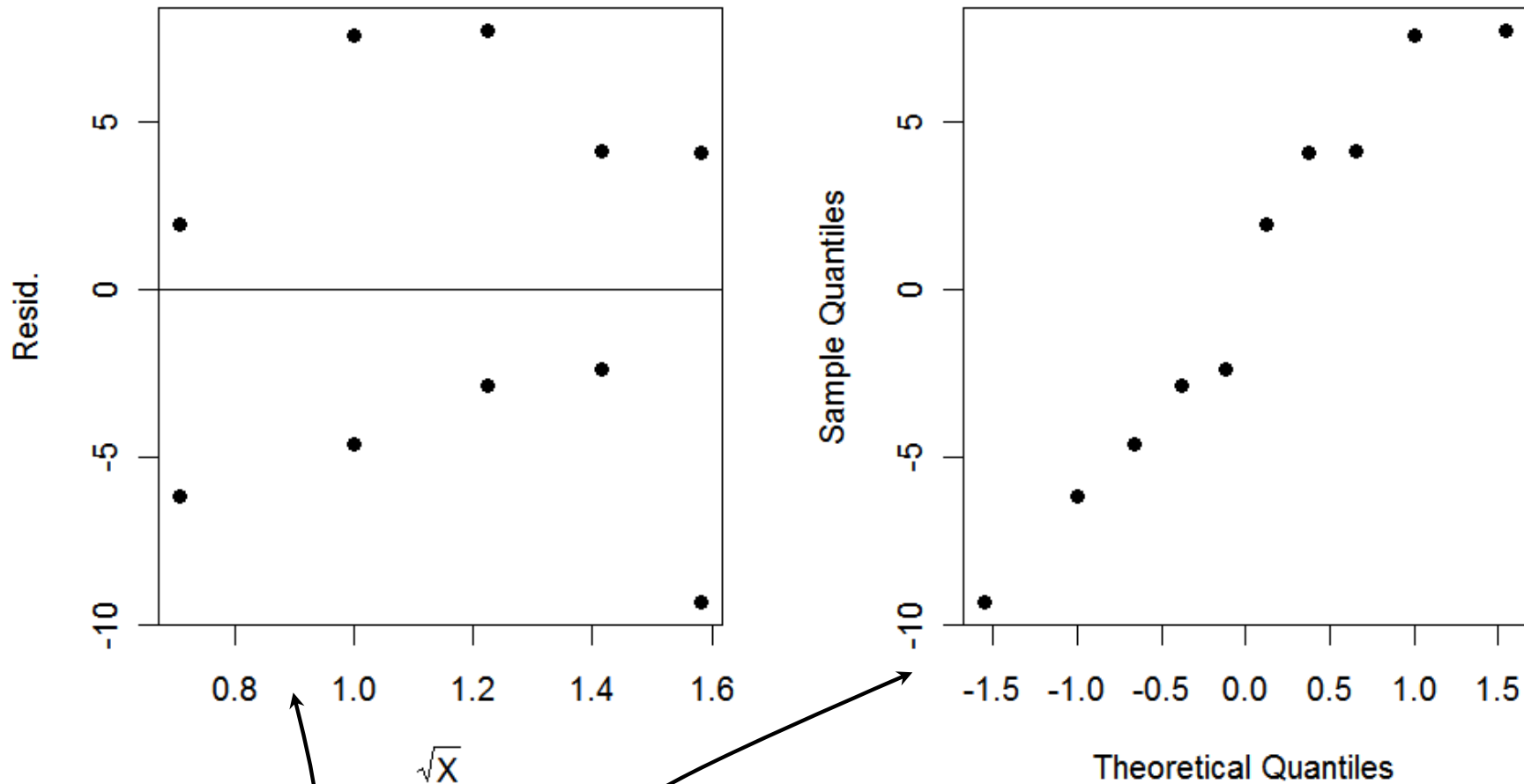
Orig. curvilinear pattern linearized by \sqrt{X} transform.

Sales Train'g (CH03TA07) (cont'd)

Resid. plot and normal probability plots seem reasonable after \sqrt{X} transform:

```
> sqrtX = sqrt(X)
> trCH03TA07.lm = lm( Y ~ sqrtX )
> par( mfrow=c(1,2) )
> plot( resid(trCH03TA07.lm) ~
        sqrtX, xlab=expression(sqrt(X)) )
> abline( h=0 )
> qqnorm( resid(trCH03TA07.lm), main="" )
```

Sales Train'g (CH03TA07) (cont'd)



Resid. plot and NPP after \sqrt{X} transform (cf. Fig. 3.14).

Transforming X to $\ln(X)$

- Sometimes, a few X_i -values are very far away from the bulk of the data, or the X 's are geometrically or exponentially spaced (e.g., $X = 1, 3, 10, 30, 100, 300, \dots$)
- If so, and if all the X_i 's are positive ($X > 0$), then a **logarithmic transform** may be called for; i.e., replace the SLR model with

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i \quad (i = 1, \dots, n).$$

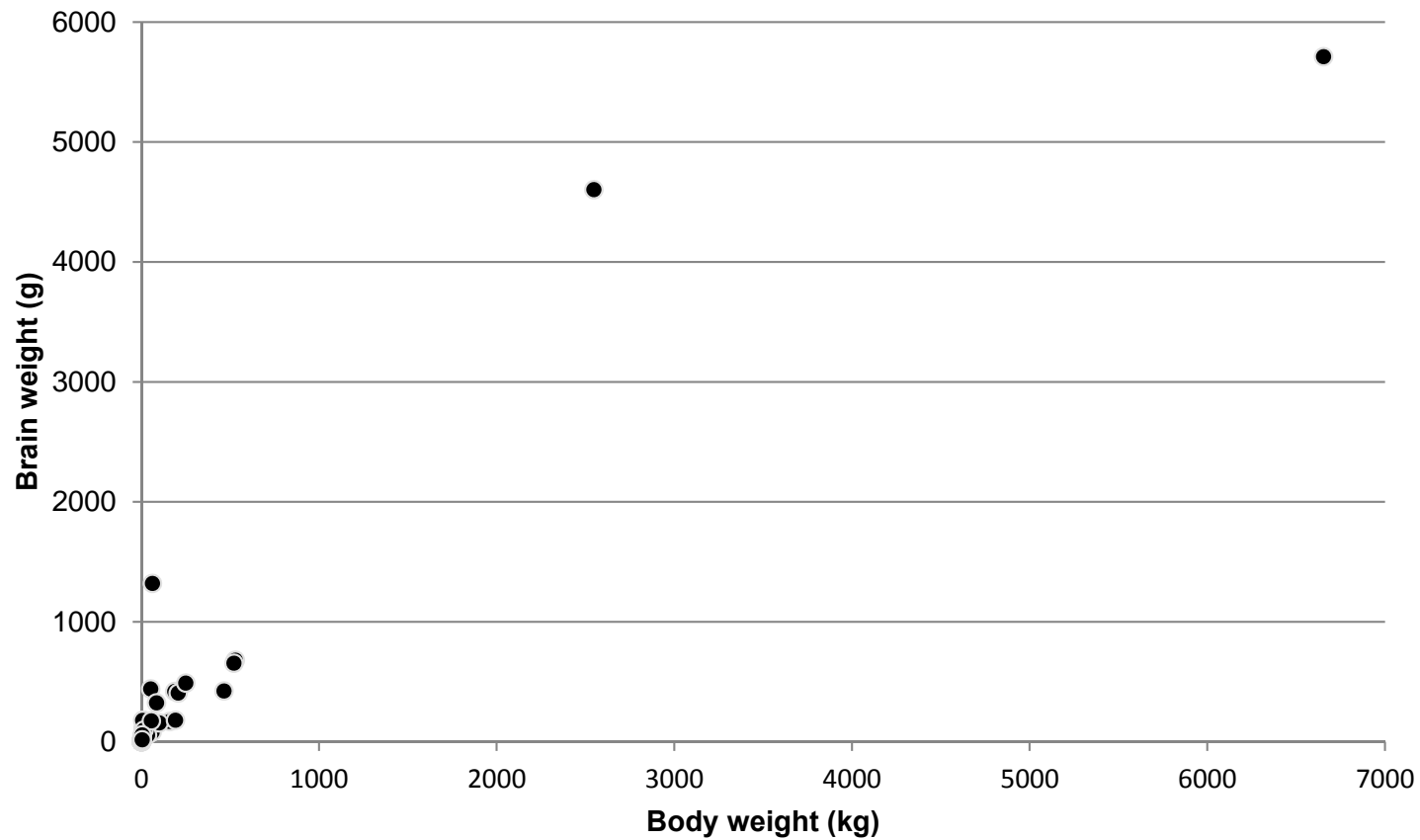
Example

Brain Weight/Body Weight

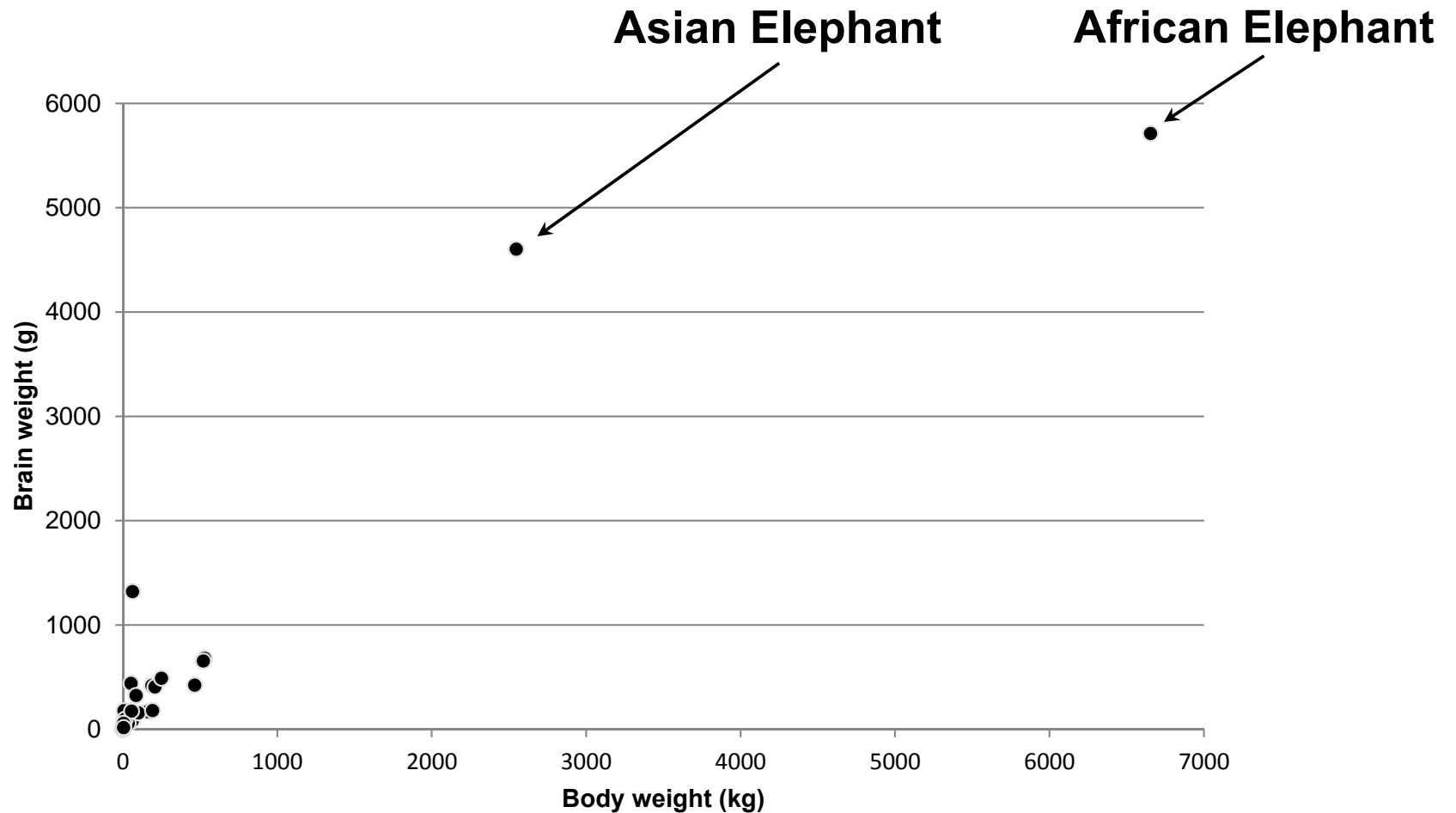
In a study of mammalian body features,
 X = avg. body weight (kg) and
 Y = avg. brain weight (g)
were determined for $n = 62$ terrestrial
mammals.

On the original scale, a scatterplot shows
little distinguishability, due to the extreme
observations at high X (next slide \rightarrow)

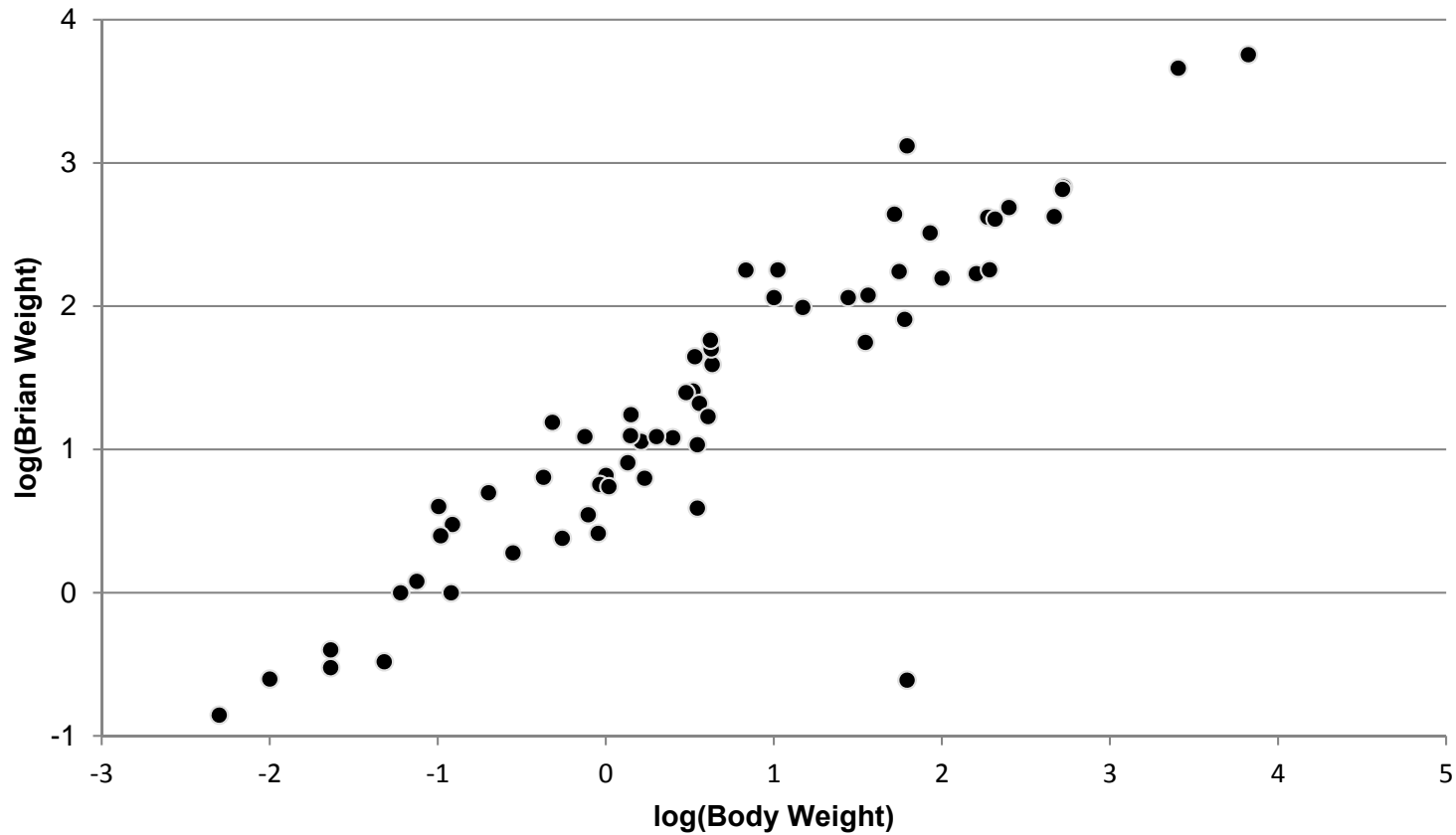
Brain Weight Scatterplot



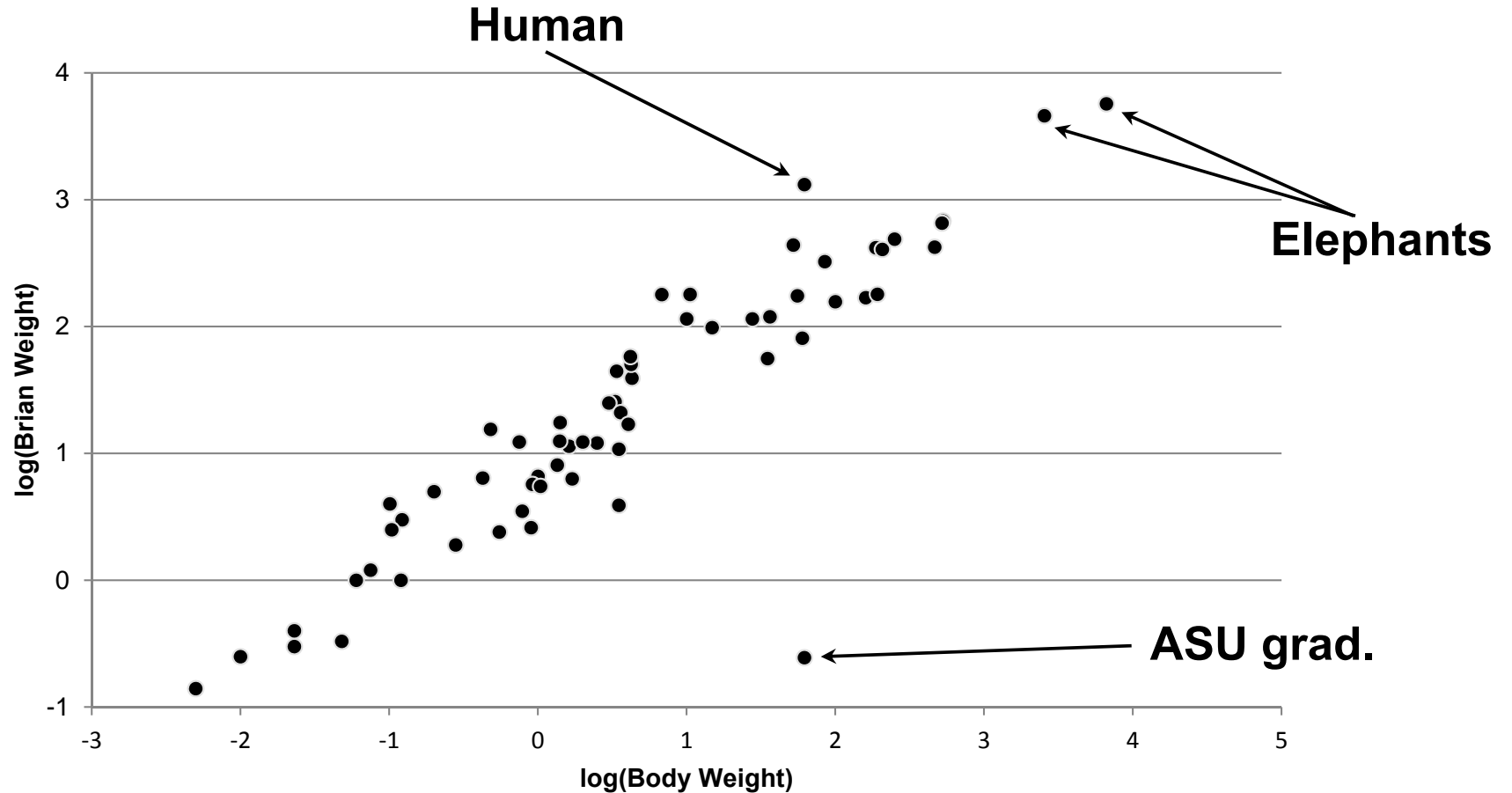
Brain Weight Scatterplot (cont'd)



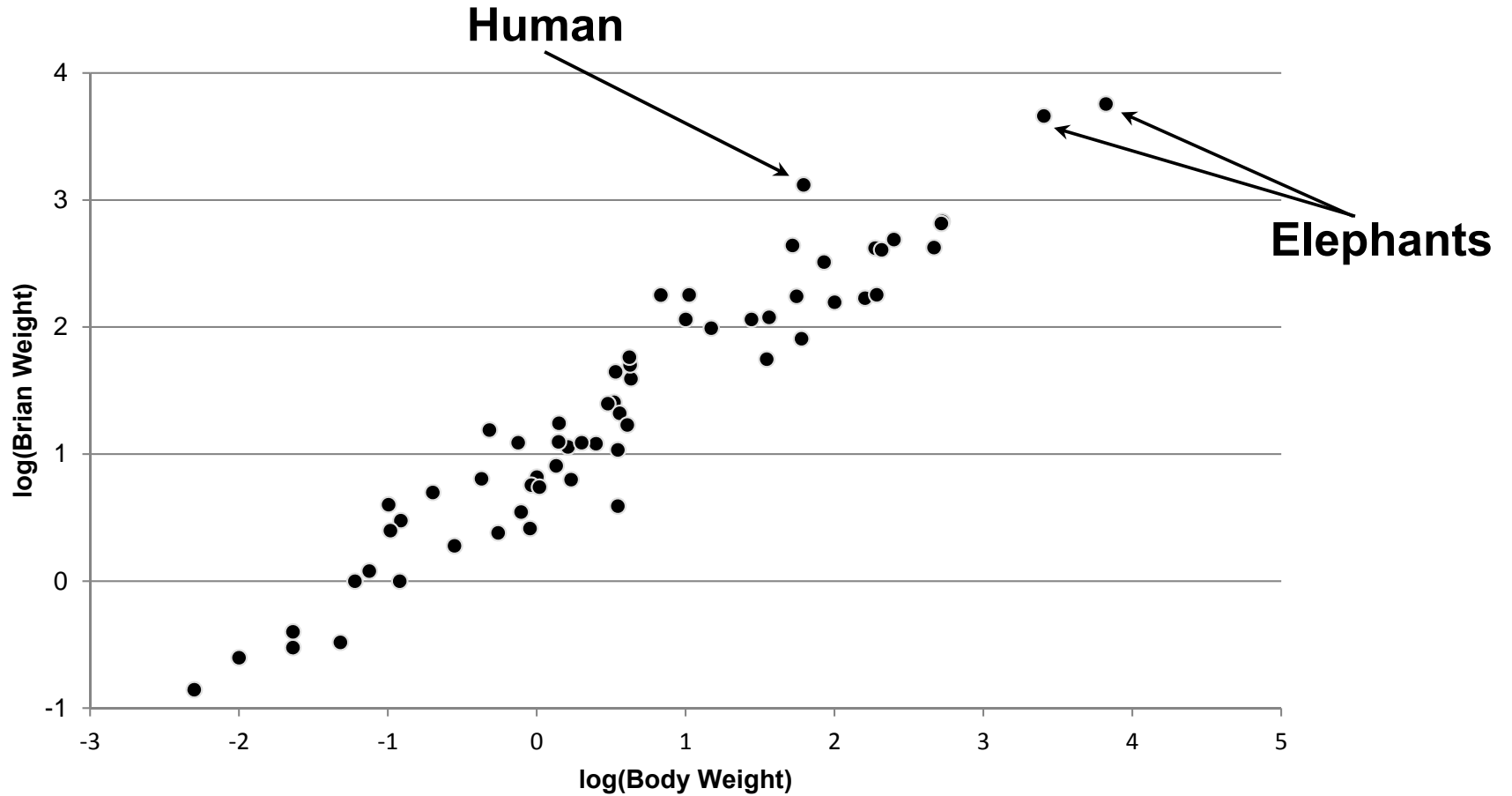
Brain Weight Scatterplot after $\ln(X) - \ln(Y)$ transform



Brain Weight Scatterplot after $\ln(X) - \ln(Y)$ transform



Brain Weight Scatterplot after $\ln(X) - \ln(Y)$ transform



Brain Weight Example (cont'd)

- **We see that the logarithmic transform applied to X (and here, also to Y) can increase distinguishability in the scatterplot.**
- **In this example, it also “reduced” the outlying effect of the extreme observations (elephants), while also enhancing our ability to identify other possible extreme observations (maybe humans?).**

Other Useful Transformations

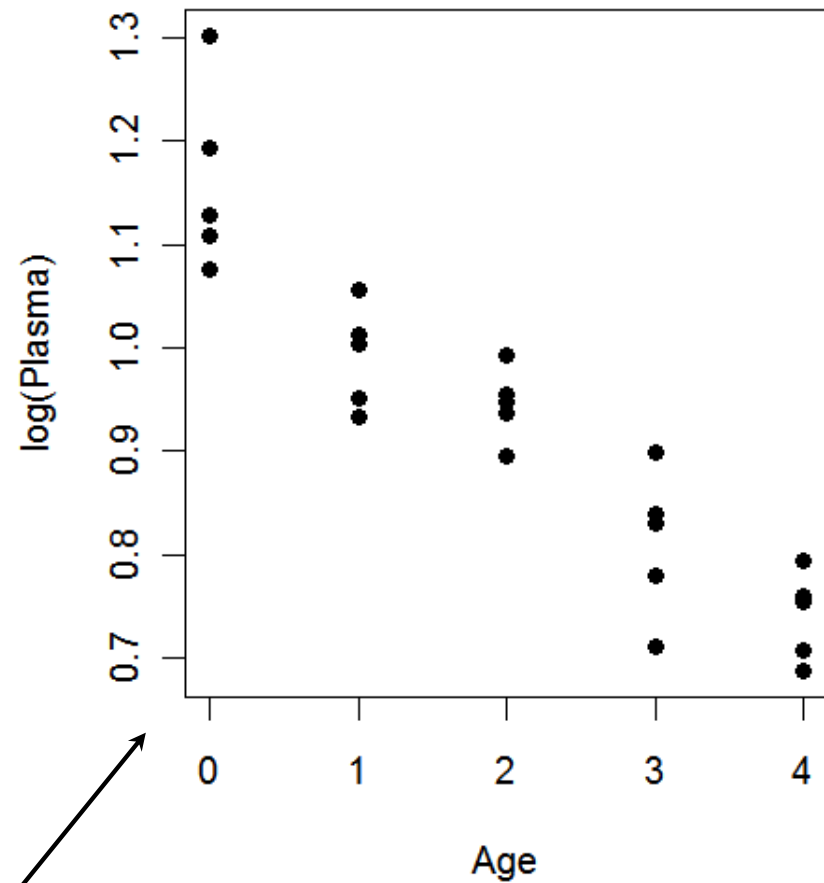
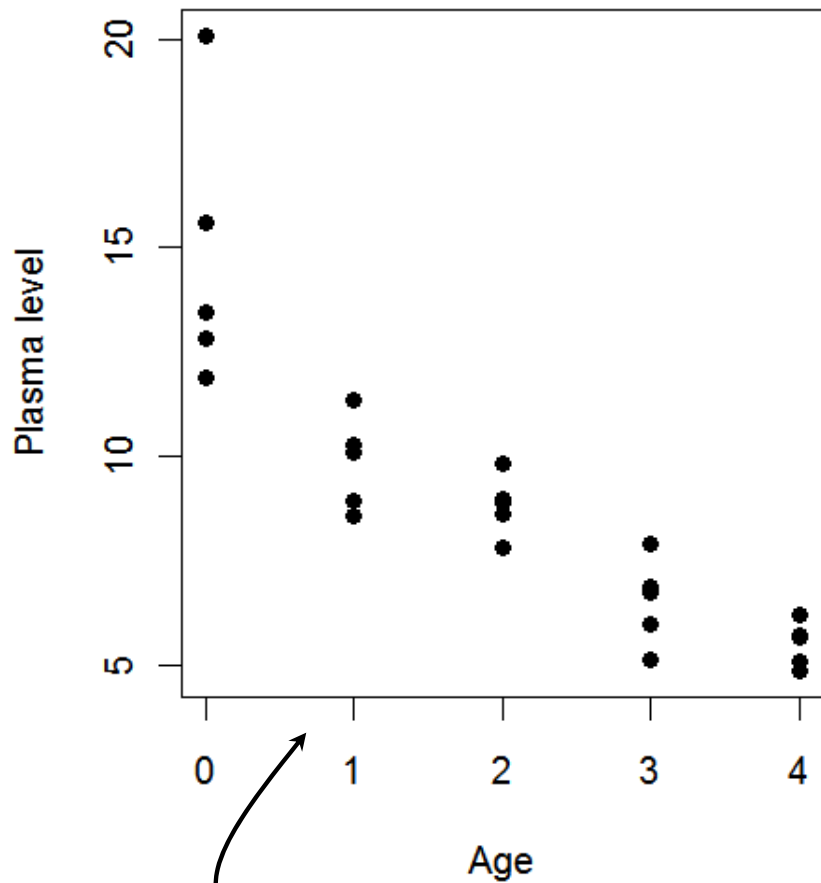
Some other useful transformations of the X variable include

- $X' = \log_{10}(X)$
- $X' = 1/(X + k)$ (for some $k \geq 0$)
- $X' = X - (1/X)$

Transforming Y to $\log_{10}Y$

- When a curvilinear relationship is evidenced in the data, a **transform of Y** to $\log_{10}Y$ might be called for.
- That is, replace the SLR model with
$$\log_{10}Y_i = \beta_0 + \beta_1X_i + \varepsilon_i \quad (i = 1, \dots, n).$$
- **Blood plasma example (CH03TA08):**
 - > `par(mfrow=c(1,2))`
 - > `plot(Y ~ X)`
 - > `plot(log10(Y) ~ X)`

Blood Plasma (CH03TA08)



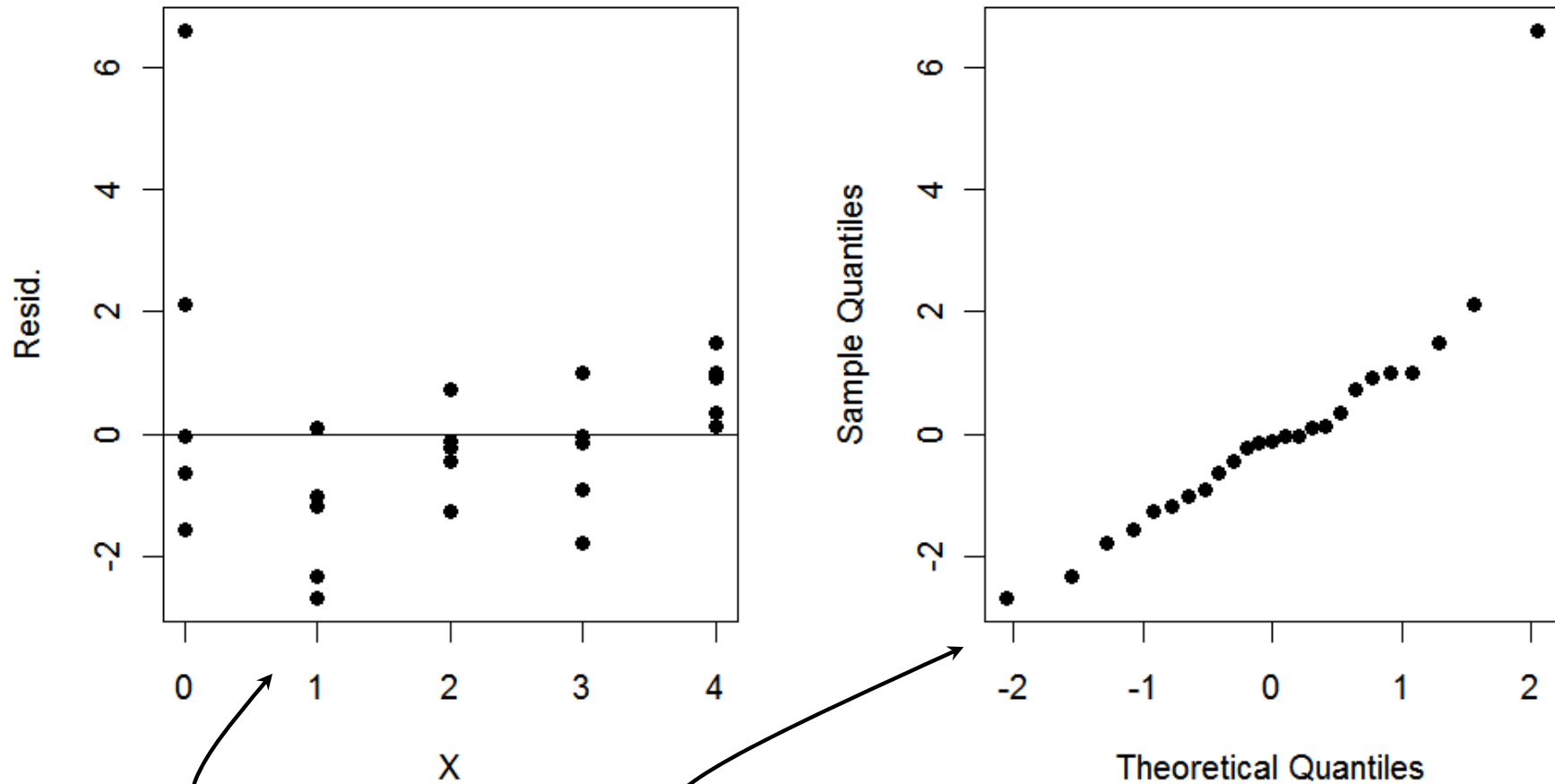
Curvilinear pattern linearized by $\log_{10}(Y)$ transform

Blood Plasma (CH03TA08) (cont'd)

Residual plot and normal probability plot
under 'usual' SLR model:

```
> CH03TA08.lm = lm( Y ~ X )  
> par( mfrow=c(1,2) )  
> plot( resid(CH03TA08.lm) ~ X )  
> abline( h=0 )  
> qqnorm( resid(CH03TA08.lm), main='' )
```

Blood Plasma (CH03TA08) (cont'd)



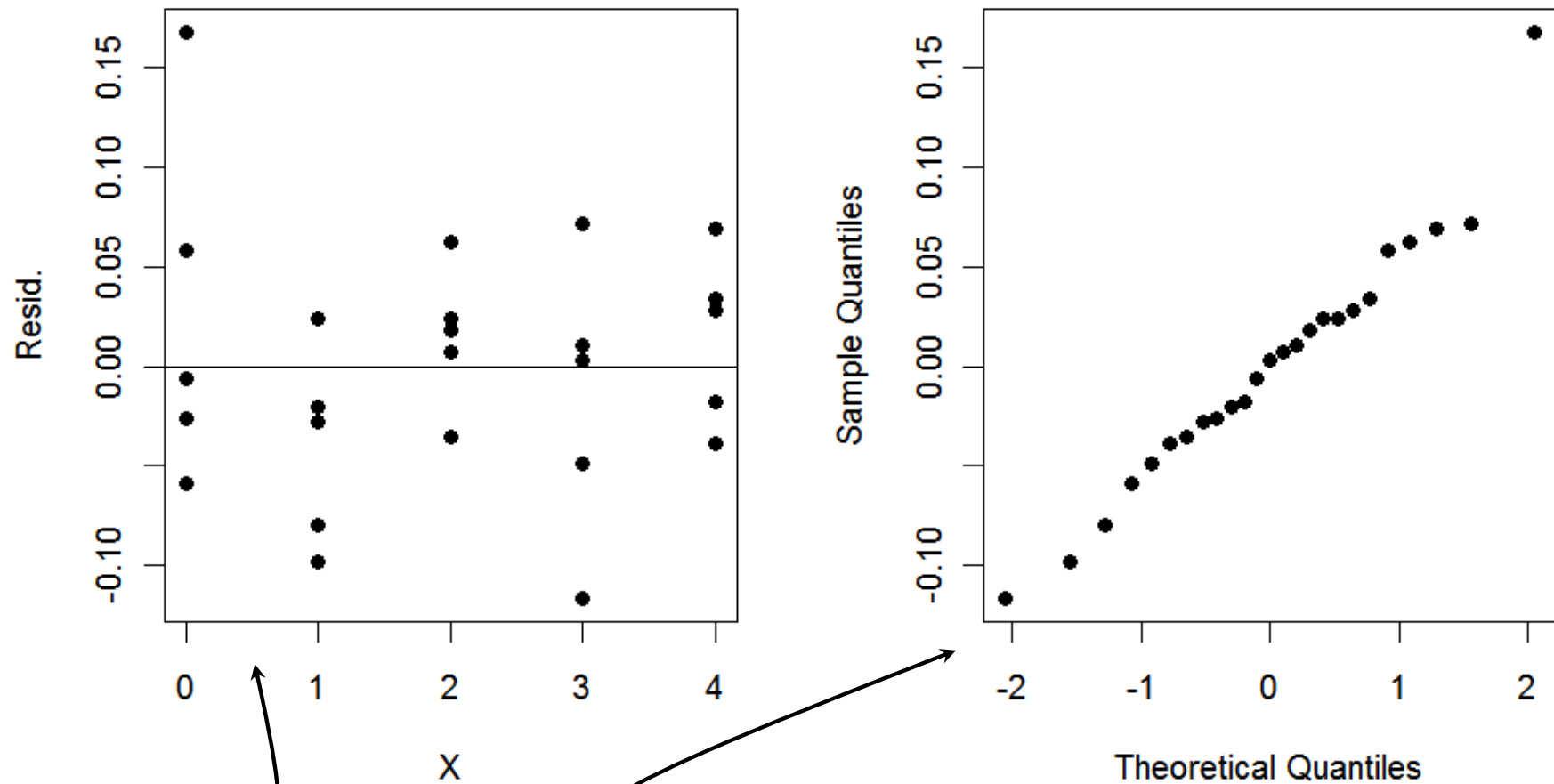
Resid. plot and NPP under SLR model show multiple model violations

Blood Plasma (CH03TA08) (cont'd)

Residual plot and normal probability plot
under $\log_{10}(Y)$ transform:

```
> trCH03TA08.lm = lm( log10(Y) ~ X )  
> par( mfrow=c(1,2) )  
> plot( resid(trCH03TA08.lm) ~ X )  
> abline( h=0 )  
> qqnorm( resid(trCH03TA08.lm),  
          main="" )
```

Blood Plasma (CH03TA08) (cont'd)



Resid. plot and NPP after $\log_{10}(Y)$ transform (cf. Fig. 3.14).

Box-Cox Power Transform

- A general class of transformations that includes the square root and the log is the **Box-Cox Power Transformation**:

$$W_i = \frac{Y_i^\lambda - 1}{\lambda K_2^{\lambda-1}}$$

- where λ is a transform parameter and K_2 is the **geometric mean** of the Y_i s:

$$K_2 = \left(\prod_{i=1}^n Y_i \right)^{1/n}$$

Box-Cox Transform (cont'd)

- If $\lambda = 0$, use the continuity-preserving logarithm for the Box-Cox transform:

$$W_i = K_2 \ln(Y_i) \text{ at } \lambda = 0.$$

- We can apply maximum likelihood (ML) to *estimate* λ from the data. (Alternatively, can minimize the regression SSE over a series λ values; see p. 135.)
- In R, this is performed via the `boxcox()` function in the *MASS* package.

Blood Plasma (CH03TA08) (cont'd)

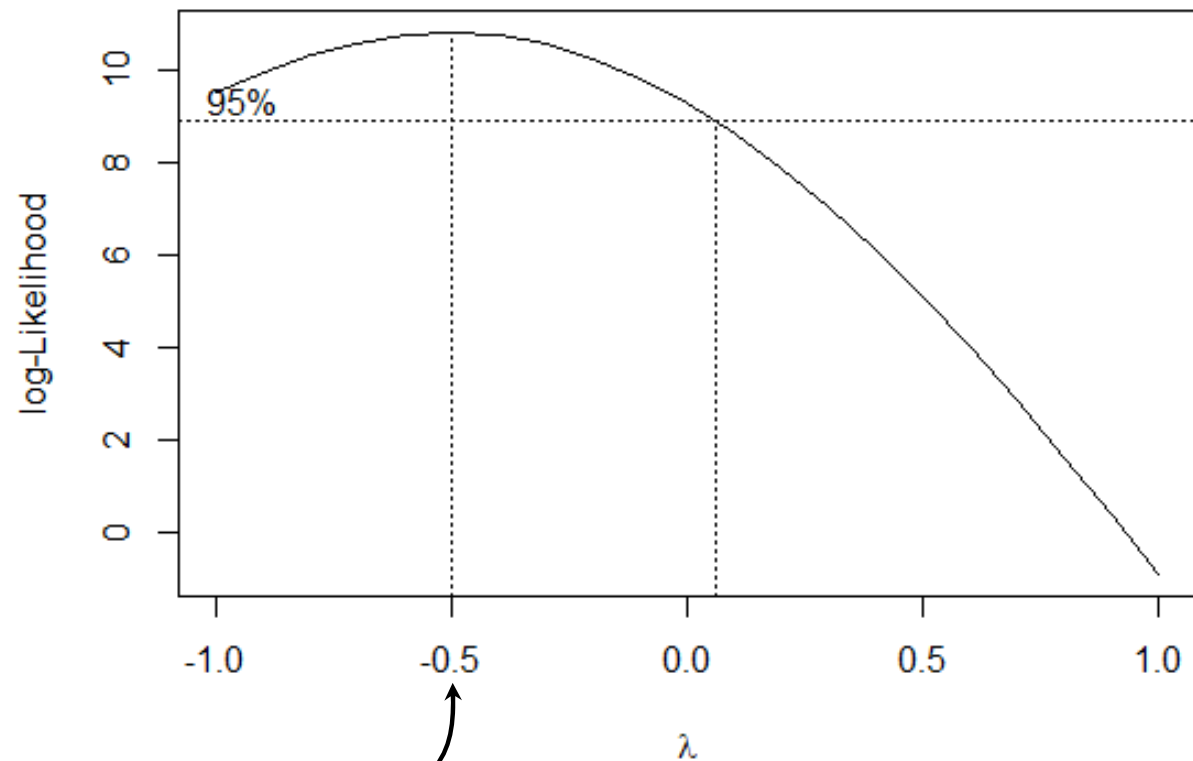
Box-Cox transform analysis (mimics presentation in Neter *et al.*, p.136):

```
> require( MASS )  
  
> CH03TA08.lm = lm( Y ~ X )  
  
> CH03TA08.bc = boxcox( CH03TA08.lm,  
  lambda=seq(-1, 1, 0.1), interp=F )  
  
> cbind( CH03TA08.bc$x, CH03TA08.bc$y )
```

Blood Plasma (CH03TA08) (cont'd)

`boxcox()` function produces plot of log-likelihood (cf. to Fig. 3.17 in Neter *et al.*)

Max. log-likelihood
(also min. SSE) at
 $\lambda = -0.5$



Blood Plasma (CH03TA08) (cont'd)

Call to `cbind()` prints out λ ('x' col.) and calculated values of log-likelihood ('y' col.):

[1,]	-1.0	9.5036602	[12,]	0.1	8.6280273
[2,]	-0.9	9.9557422	[13,]	0.2	7.8716970
[3,]	-0.8	10.3202421	[14,]	0.3	7.0196583
[4,]	-0.7	10.5882760	[15,]	0.4	6.0806054
[5,]	-0.6	10.7517109	[16,]	0.5	5.0634353
[6,]	-0.5	10.8036292	[17,]	0.6	3.9768915
[7,]	-0.4	10.7387696	[18,]	0.7	2.8292907
[8,]	-0.3	10.5538884	[19,]	0.8	1.6283340
[9,]	-0.2	10.2479881	[20,]	0.9	0.3809955
[10,]	-0.1	9.8223822	[21,]	1.0	-0.9065263
[11,]	0.0	9.2805821			

Transforming Y: Caveats

We need to **be careful** when transforming the response, Y .

Say you use $W = \ln(Y)$, so that $Y = \exp\{W\}$.

- The model becomes $E[W] = \beta_0 + \beta_1 X$.
- The fitted values have the form $\hat{W}_h = b_0 + b_1 X_h$. But don't stop there!
- Should reverse the transform to end with $\hat{Y}_h = \exp\{b_0 + b_1 X_h\}$.
- Use similar reverse-transforms for other transforming functions.

Transforming X: Caveats

- By contrast, if you transform just the X-variable/predictor, you only change the input scale.
- At that point, stay on the chosen, transformed, input scale throughout all the calculations.

Example: if you transform X_i to $\sqrt{X_i}$ the model becomes $E[Y_i] = \beta_0 + \beta_1\sqrt{X_i}$. Then stay on the $\sqrt{X_i}$ scale for estimates, inferences, predictions, etc.