Outline

1. Remedial Measures: Transformation

2. Remedial Measures: Weighted Least Squares

Remedial Measures

- Nonlinearity of regression function:
  - Transformation.
  - Polynomial regression.
  - Nonlinear regression.
- Nonequal error variance:
  - Transformation.
  - Weighted least squares.
- Nonindependence of error terms:
  - Models with correlated error terms.
- Nonnormality of error terms.
  - Transformation.
  - Nonparametric methods.
  - Generalized linear models.
- Presence of outliers:
  - Removal of outliers (with caution).
  - Robust estimation.

Example: Surviving Bacteria

Data consist of number of surviving bacteria after exposure to X-rays for different periods of time.

- Let t denote time (in number of 6-minute intervals)
- let n denote number of surviving bacteria (in 100s) after exposure to X-rays for t time.

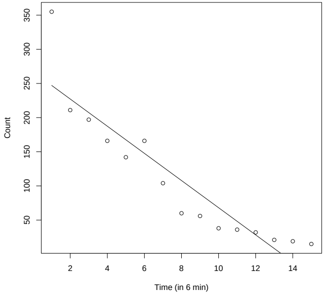
t	1	2	3	4	5	6	7	8
n	355	211	197	166	142	166	104	60

t	9	10	11	12	13	14	15
n	56	38	36	32	21	19	15

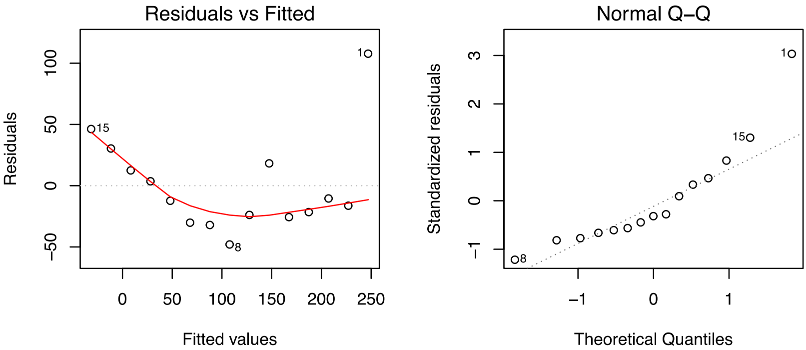
Example: Surviving Bacteria

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	267.010	22.170	12.044	2.0e-08 ***
t	-19.893	2.438	-8.158	1.8e-06 ***

Residual standard error: 40.8 on 13 degrees of freedom
Multiple R-squared: 0.8366, Adjusted R-squared: 0.824
F-statistic: 66.56 on 1 and 13 DF, p-value: 1.804e-06



Example: Surviving Bacteria



Example: Surviving Bacteria

- Here there is a theoretical model:

$$n_t = n_0 e^{\beta t},$$

  where

  - t is time,
  - n_t is the number of bacteria at time t,
  - n_0 is the number of bacteria at the start ($t = 0$), and
  - β is a decay rate with $\beta < 0$.

- Consider a log transformation:

$$\ln(n_t) = \ln(n_0) + \beta t = \alpha + \beta t,$$

  by setting $\alpha = \ln(n_0)$.
  Interpretation: ______________________

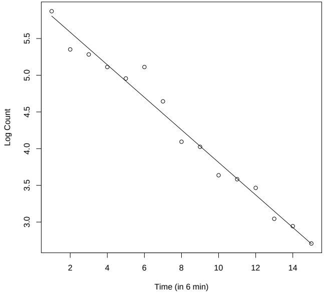
Example: Surviving Bacteria

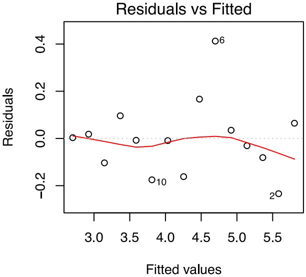
The transformed data are as follows.

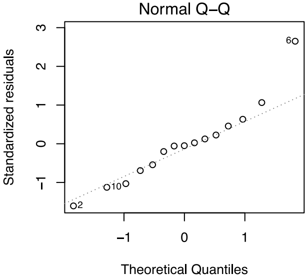
t	1	2	3	4	5	6	7	8
$\ln(n)$	5.87	5.35	5.28	5.11	4.96	5.11	4.64	4.09

t	9	10	11	12	13	14	15
$\ln(n)$	4.03	3.64	3.58	3.47	3.04	2.94	2.71

Example: Surviving Bacteria







Example: Surviving Bacteria

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.028695	0.088259	68.31	< 2e-16 ***
t	-0.221629	0.009707	-22.83	7.1e-12 ***

Residual standard error: 0.1624 on 13 degrees of freedom
Multiple R-squared: 0.9757, Adjusted R-squared: 0.9738
F-statistic: 521.3 on 1 and 13 DF, p-value: 7.103e-12

How to interpret β ?________________

How to interpret α ?________________

Inference for n_0 is not straightforward.

$$\hat{n}_0 = e^{\hat{\alpha}} = 415.30$$

but

$$E(\hat{n}_0) \neq n_0!$$

Transformation: Remarks

- Ideally, theory should dictate what transformation to use. as is in the bacteria count example
- In practice, transformation is usually chosen empirically. based on data analysis
- Usually it is best to start with a simple transformation and experiment.
  - To meet the linearity assumption, transformation could be that of X, or Y, or both.
  - Common transformations are $\log_{10}, \ln, \sqrt{\cdot}$. Less common transformations are $Y^2, 1/Y, 1/Y^2, \arcsin\sqrt{Y}$.
- Another advantage of transformation is to control unequal variance.

Transformation: Remarks

- Consider a transformation ladder for $Z = X$ or Y.

λ	$\cdots$	-2	-1	-0.5	0	0.5	1	2	$\cdots$
Z^λ	$\cdots$	$\frac{1}{Z^2}$	$\frac{1}{Z}$	$\frac{1}{\sqrt{Z}}$	$\log(Z)$	$\sqrt{Z}$	Z	Z^2	$\cdots$

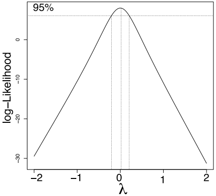
- Transforming Y can affect both linearity and equal variance, but transforming X can affect only linearity.
- Sometimes solving one problem can create another.

Box-Cox Transformation

- **Box-Cox method** is a formal approach to selecting λ to transform Y.
- The idea is to consider

$$Y_i^{\lambda} = \beta_0 + \beta_1 X_i + \epsilon_i.$$

- Estimate λ (along with $\beta_0, \beta_1, \sigma^2$) using ML.
- Choose an interpretable $\hat{\lambda}$ within a 95% CI. In the surviving bacteria example, the Box-Cox method gives $\hat{\lambda} = -0.0202$. Implication: ______________________



- R command: boxcox(object, ...) in the MASS library

Outline

Example: Improving QoL

A study examines the effectiveness of a therapy to life quality. Data consist of the records of 11 patients:

- Let X denote the number of therapy sessions a patient has completed over 6 months;
- Let Y denote an index of quality of life computed from a series of questions the patient responds to at the end of the 6-month period.

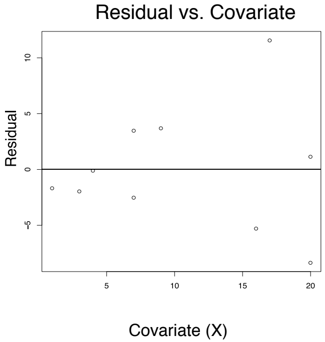
i	1	2	3	4	5	6	7	8	9	10	11
x_i	1	3	4	7	7	9	16	17	20	20	23
y_i	17.5	20.5	24.0	32.5	26.5	36.0	38.5	57.0	51.5	42.0	75.5

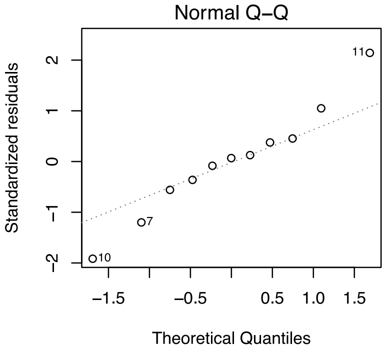
Example: Improving QoL

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	15.0080	4.3651	3.438	0.007412 **
x	2.0190	0.3175	6.359	0.000132 ***

Residual standard error: 7.86 on 9 degrees of freedom
Multiple R-squared: 0.8179, Adjusted R-squared: 0.7977
F-statistic: 40.43 on 1 and 9 DF, p-value: 0.0001315

Example: Improving QoL





Weighted Regression Model

- We extend the usual SLR model to a weighted linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

  where

$$\varepsilon_i \sim \text{ind} \quad N(0, {\sigma_i}^2),$$

  for observations $i = 1, \ldots, n$.

- Consider a simple form for the (unequal) variance σ_i^2.
- For example, suppose $X > 0$ and $c > 0$ is a constant. Some possible forms for the variance are

$$\sigma_i^2 = cX_i, \quad \sigma_i^2 = cX_i^2, \quad \sigma_i^2 = cX_i^3, \quad \sigma_i^2 = cX_i^4.$$

- As X_i increases, the corresponding variance σ_i^2 increases, for the ith observation.

Weighted Least Squares

- The method of **weighted least squares (WLS)** finds β_0, β_1 that minimize the **weighted error sum of squares**:

$$Q_W(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i \left\{ Y_i - (\beta_0 + \beta_1 X_i) \right\}^2$$

where $w_i \propto 1/\sigma_i^2$.

- For example, suppose $\sigma_i^2 = cX_i^2$. Then

$$Q_W(\beta_0, \beta_1) = \sum_{i=1}^{n} \frac{1}{X_i^2} \left\{ Y_i - (\beta_0 + \beta_1 X_i) \right\}^2 .$$

That is, the weight is inversely proportional to X_i^2.

Weighted Least Squares

- Recall that the LS estimates of β_0, β_1 are:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2},\\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}.\end{aligned}$$

- The WLS estimates of β_0, β_1 are:

$$\begin{aligned}\hat{\beta}_1^w &= \frac{\sum_{i=1}^n w_i (Y_i - \bar{Y}^w)(X_i - \bar{X}^w)}{\sum_{i=1}^n w_i (X_i - \bar{X}^w)^2},\\ \hat{\beta}_0^w &= \bar{Y}^w - \hat{\beta}_1^w \bar{X}^w,\end{aligned}$$

  where

$$\bar{X}^w = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i} \quad \text{and} \quad \bar{Y}^w = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}$$

  are the weighted averages of X_i's and Y_i's.

Weighted Least Squares

More generally, consider a general linear model

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \quad \text{where} \quad \varepsilon \sim \mathcal{MVN}(\mathbf{0}, \sigma^2\Sigma),$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times 2}$ and Σ is a known matrix. (What is Σ in SLR? How about in WLS?).
Proof: the least-square estimates are

$$\hat{\beta} = (\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{Y})$$

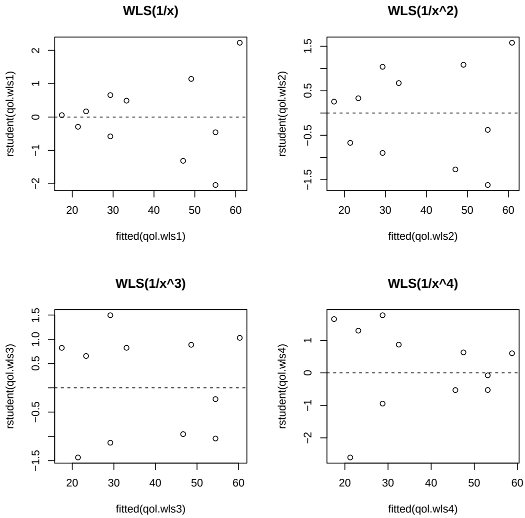
and

$$\hat{\sigma}^2 = \frac{1}{n-2}\boldsymbol{Y}^T\left(\Sigma^{-1} - \Sigma^{-1}\boldsymbol{X}(\boldsymbol{X}^T\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\Sigma^{-1}\right)\boldsymbol{Y}$$

Proof in class.
HW: What is the sampling distribution of $\widehat{\beta}$?

Example: Improving QoL

Standardized residual plots with $\sigma_i^2 = cX_i$, cX_i^2, cX_i^3, and cX_i^4.



Example: Improving QoL

WLS (X^3):

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	15.5380	0.2252	68.99	1.43e-13 ***
x	1.9454	0.1548	12.57	5.20e-07 ***

Residual standard error: 0.1379 on 9 degrees of freedom
Multiple R-squared: 0.9461, Adjusted R-squared: 0.9401
F-statistic: 157.9 on 1 and 9 DF, p-value: 5.196e-07

Source	df	SS	MS
x	1	3.00061	3.0006
Error	9	0.17103	0.0190
Total	10	3.17164	–

$$SS_T = \sum_{i=1}^{n}(Y_i - \bar{Y})^2, SS_E = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2, SS_X = SS_T - SS_X$$

Weighted Least Squares: Remarks

- The purpose of weighted regression is to meet the model assumptions as transformation, but the approach is different.
- Standardized residuals are recommended for weighted regression.
- In the QoL example, with $\sigma_i^2 = cX_i^3$, the studentized residual plot shows that the fit improves the most.

Weighted Least Squares: Remarks

- Both WLS and transformation provide an approach to stabilize variance.
- Statistical software packages give sums of squares and coefficient of determination (R^2) for WLS, but their interpretation is not straightforward.
  Formulas for the $s.e.(\hat{\beta}_0)$ and $s.e.(\hat{\beta}_1)$ are more complicated than SLR.
- When to use transformation and when to use WLS?
  In practice, consider WLS if a linear-line relationship between X and Y is obvious and unequal variance is a problem.