ST5202: Applied Regression Analysis

Department of Statistics and Applied Probability
National University of Singapore

05-Feb-2018
Week 4

Announcement

- Assignment #2 available online
  - Due on 12 Feb by 9 pm
  - Submit either in-class or via email (in-class submission preferred)
  - Please write BOTH your name and metric number
- Make-up midterm request due on 26 Feb
  - Better to make request as soon as possible
  - Official supporting document required
  - Request after the due date would not be considered

Week 4

Reviews &
Diagnostics and Remedial Measures (Chapter 3)

Some Reviews & Construction of Confidence Band

- Inference about the mean response $E\{Y_h\}$
- Predicting new observations $Y_{h(new)}$
- Confidence bound for a regression line (new stuff!)
- General linear test approach

Review: Inference about the mean response $E\{Y_h\}$ at $X = X_h$

- $E\{Y_h\}$ is the expected/mean outcome with the given level of X is X_h
    - Estimator for $E\{Y_h\}$ given by: $\hat{Y}_h = b_0 + b_1 X_h$, with sampling distribution:

$$\hat{Y}_h \sim N(\beta_0 + \beta_1 X_h, \sigma^2\{\hat{Y}_h\})$$

with $\sigma^2\{\hat{Y}_h\} = \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right)$

- As with the sampling distribution of the b_i's, σ^2 is unknown and estimated by s^2, which then gives a t-distribution for studentized $\hat{Y}_h$: $\frac{\hat{Y}_h - E\{Y_h\}}{s\{\hat{Y}_h\}} \sim t_{n-2}$ with $s^2\{\hat{Y}_h\} = s^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right)$

Review: Inference about the mean response $E\{Y_h\}$ at $X = X_h$

- Using this sampling distribution, we can construct $(1-\alpha)100\%$ confidence interval for $E\{Y_h\}$:

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\hat{Y}_h\}$$

Review: Prediction of a new observation given $X = X_h$

- We have

$$\begin{aligned} Y_{h(new)} &\sim N(\beta_0 + \beta_1 X_h, \sigma^2) \\ \hat{Y}_h &\sim N(\beta_0 + \beta_1 X_h, \sigma^2\{\hat{Y}_h\}) \end{aligned}$$

- We utilize the distribution of $\left(Y_{h(new)} - \hat{Y}_h\right)$:

$$\begin{aligned} \left(Y_{h(new)} - \hat{Y}_h\right) &\sim N(0, \sigma^2\{pred\}) \text{ where} \\ \sigma^2\{pred\} &= Var\left(Y_{h(new)} - \hat{Y}_h\right) \\ &= Var\left(Y_{h(new)}\right) + Var\left(\hat{Y}_h\right) \\ &= \sigma^2 + \sigma^2\{\hat{Y}_h\} \\ &= \sigma^2\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right) \end{aligned}$$

Review: Prediction of a new observation given $X = X_h$

- We have

$$\frac{(Y_{h(new)} - \hat{Y}_h)}{\sigma\{pred\}} \sim N(0,1)$$

- To construct prediction intervals for $Y_{h(\text{new})}$ based on its distribution, we need to estimate $\sigma\{\text{pred}\}$ by $s\{\text{pred}\}$, which gives:

$$\frac{Y_{h(\text{new})} - \hat{Y}_h}{s\{\text{pred}\}} \sim t_{n-2}$$

  and the $(1-\alpha)100\%$ prediction interval (PI) is given by $\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{pred\}$.
- Interpretation: We are $(1-\alpha)\%$ confident that the PI will contain the new observation
- Note: we can NOT state $Pr(Y_{h(\text{new})} \in (1-\alpha)100\% PI) = 1-\alpha$, because the bounds of the PI have been constructed based on one sample (e.g. based on the estimate $\hat{Y}_h$)

New Stuff

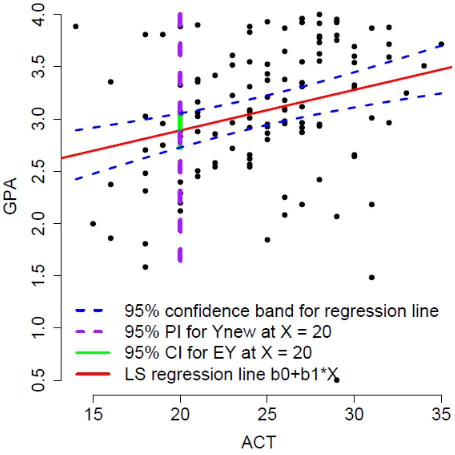
Confidence band for a regression line

- Note: we CANNOT state $Pr(Y_{h(new)} \in (1-\alpha)100\% PI) = 1-\alpha$ because the bounds of the PI have been constructed based one one batch of sample (i.e., based on the estimate $\hat{Y}_h$)
- GPA example:
  - for $\hat{X}_h = 27$, we have $\hat{Y}_h = 3.16238$ and $s\{pred\} = 0.6263652$.
    Is $\frac{Y_{h(new)} - 3.16238}{\sigma\{pred\}} \sim N(0,1)$?
    Is $\frac{Y_{h(new)} - 3.16238}{0.6263652} \sim t_{n-2}$?

New Stuff
Confidence band for a regression line

- Regression line $\beta_0 + \beta_1 X$ is estimated by $\hat{Y} = b_0 + b_1 X$
- Construct a confidence band for $\beta_0 + \beta_1 X$:
  The band (area) is expected to contain the true regression line 95/100 repeated samples
- The bounds of the band will be (slightly) wider than the individual CIs at each level of X_h, because the band has to include the entire regression line 95/100 times, instead of just one expected value
- The "Working-Hotelling" confidence band for the regression line is given by:
  $$\hat{Y}_h \pm W \cdot s\{\hat{Y}_h\},$$
  with
  $$W = \sqrt{2F(1-\alpha; 2, n-2)}$$
- $F(1-\alpha; 2, n-2)$ is the $(1-\alpha)$ percentile of the F-distribution with 2 and $(n-2)$ degrees of freedom

Confidence band for a regression line
GPA example



Review: General Linear Test Approach

- Three steps: 1) full Model, 2) reduced model, and 3) test statistic
- Error sum of squares of the full model (SSE(F)) measures the variability of the Y_i observations around the fitted regression line from the full model
- Error sum of squares of the reduced model (SSE(R)) is the variability of the observation Y_i around the fitted regression line from the reduced model
- IDEA: if SSE(F) is not much less than SSE(R), then it implies that full model does not explain the data much better than the reduced model

General Linear Test Approach

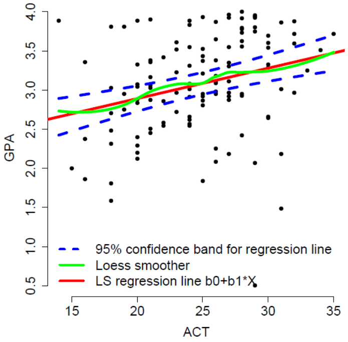
- The test statistic
$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \sim F(df_R - df_F, df_F) \text{ when } H_0 \text{ holds}$$
where df_R and df_F are the degrees of freedom associated with the reduced model and the full model respectively
- The decision rule:
$$\text{If } F^* \leq F(1-\alpha; df_R - df_F, df_F), \text{ conclude } H_0$$
$$\text{If } F^* > F(1-\alpha; df_R - df_F, df_F), \text{ conclude } H_a$$

Diagnostics and Remedial Measures

Exploration of Shape of Regression Function: Smoothing (Nonparametric Regression Curves)

- Fit a smooth curve without any constraints on the regression function to the data
  $\rightarrow$ Helpful to explore the nature of the regression relationship, if any, by fitting a smoothed curve
- Nice method to find such a smoothed curve:
  Lowess method (or simply loess)
  - Stands for "Locally Weighted Regression Scatter Plot Smoothing"
  - Fits a regression function locally;
    Span parameter determines size of the neighborhood that is used to fit the curve (thus how smooth the curve is)
  - Sometimes an iterative procedure is used, to down-weight outliers
- R command: "loess"

Exploration of Shape of Regression Function: Smoothing (GPA example)



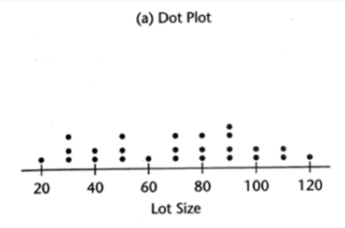
Linear relation seems appropriate

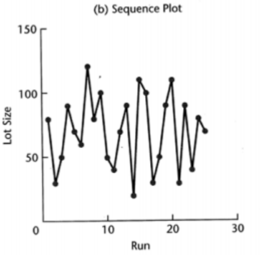
Week 4: Diagnostics and Remedial Measures

- How can we tell that our regression model is appropriate?
  $\rightarrow$ Graphic diagnostics
  $\rightarrow$ Tests
- What do we do if not?
  $\rightarrow$ Depends on our data
  (transformation of variables, perform weighted least squares, etc)

Graphical Diagnostics for Predictor Variable

- Dot plot
  - Useful for visualizing distributions of inputs when the data points are not too many
- Sequence plot
  - Useful for visualizing pattern (if there exists any)
- Stem-and-leaf-plot
  - Similar to histogram
- Box plot
  - Useful for visualizing distribution of inputs



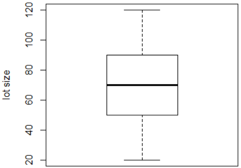


(c) Stem-and-Leaf Plot

```
The decimal point is 1 digit(s) to the right of the |

 2 | 0
 3 | 000
 4 | 00
 5 | 000
 6 | 0
 7 | 000
 8 | 000
 9 | 0000
10 | 00
11 | 00
12 | 0
```

(d) Box Plot



Diagnostics of Residuals: Residuals

- The residual e_i can be regarded as the observed error:

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

- The true error ϵ_i is

$$\epsilon_i = Y_i - E\{Y_i\}$$

- Semistudentized residual

$$e_i^* = \frac{e_i - \bar{e}}{\sqrt{MSE}} = \frac{e_i}{\sqrt{MSE}}$$

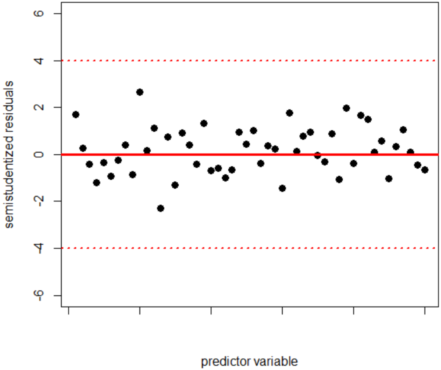
Diagnostics of Residuals: Departures from Model to be Studied by Residuals

- The regression function is not linear
- The error term do not have constant variance
- The error terms are not independent
- The model fits all but one or a few outlier observations
- The error terms are not normally distributed
- One or several important predictor variables have been omitted from the model

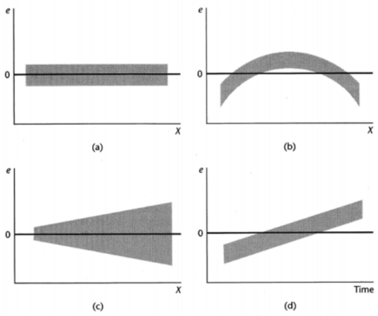
Diagnostics plots of residuals

- Informal diagnostic plots
- Provides information on whether departure from the simple linear regression model exists

Diagnostics plots of residuals: Prototypes

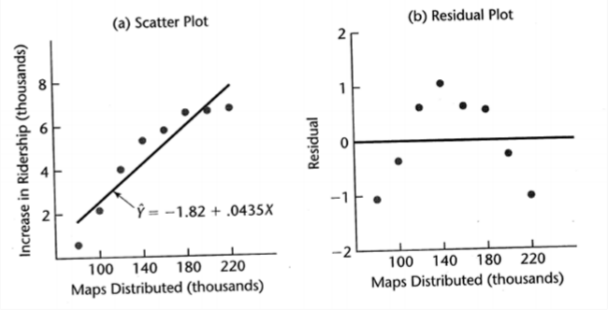


Diagnostics plots of residuals: Prototypes



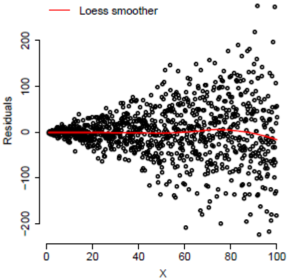
Diagnostics plots of residuals: Nonlinearity of Regression Function

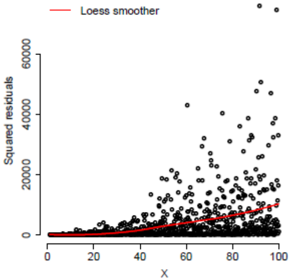
- Can be studied from a residual plot against the predictor variable or, equivalently, a residual plot against the fitted values
- Systematic patterns suggests points out the lack of linearity in true regression function



Diagnostics plots of residuals: Nonconstancy of Error Variance

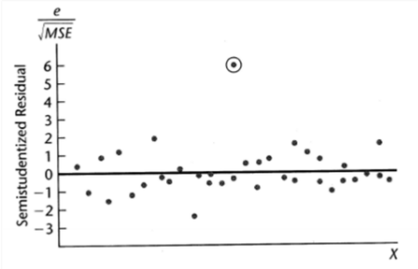
- Can be studied from "e_i vs. X"
- Can be studied from "$|e_i|$ vs X" or "e_i^2 vs. X"
- "Megaphone" type as below suggests the variance increases as the values of the predictor variable increases



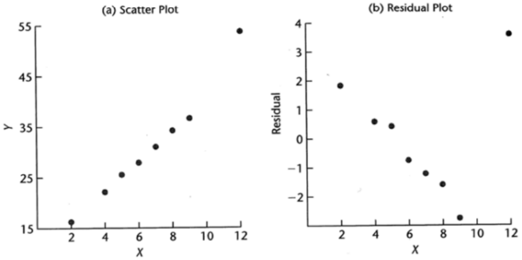


Diagnostics plots of residuals: Presence of Outliers

- We can use box plots, stem-and-leaf plots, dot plots, and residual plots against X or $\hat{Y}$
- Using a rule of thumb, semistudentized residuals with absolute value of four or more can be considered as outliers ($\frac{|e_i|}{\sqrt{MSE}} > 4$)



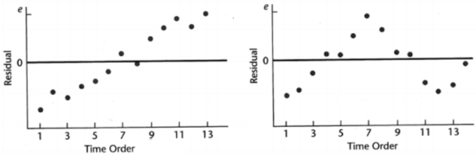
Diagnostics plots of residuals: Presence of Outliers



- Distorting effect on residuals caused by an outlier when remaining data follow linear regression

Diagnostics plots of residuals: Nonindependence of Error Terms

- Investigate the residuals against some type of sequence regarding X if there is any (e.g., time, geographical location, etc)
- Any systematic trends implies correlation between error terms that are near each other in the sequence
- (Note: e_i's are not independent unlike ϵ_i's, but for large sample size, the dependency effect among e_i can be ignored)



Diagnostics plots of residuals: Nonnormality of Error Terms

- Boxplot: graphical summary of important numbers (median, quartiles and outlies), good to check symmetry
- Histogram
- Quantile-quantile plot (QQ-plot)
- (Note: the number of samples must be reasonably large)

Diagnostics plots of residuals: Quantile-quantile plot

- Graphical tool to determine whether a sample is consistent with a certain theoretical distribution
  (in this case, it is a standard normal distribution $N(0, 1)$)
- Each point in a QQ-plot corresponds to a probability p:
  - x-coordinate: p^{th} quantile of theoretical distribution
  - y-coordinate: p^{th} quantile of sample
- p^{th} quantile (=percentile) of a distribution:
  point x such that $P(X \leq x) = p$
- p^{th} quantile of sample:
  point x such that $\frac{\#obs \leq x}{n} \approx p$

Diagnostics plots of residuals: Quantile-quantile plot

- If the sample is drawn from the compared theoretical distribution, then
  $\rightarrow$ the sample quantiles and the theoretical quantiles are approximately equal $\rightarrow$ hence the x and y coordinates of points in QQ-plot are approximately equal
  $\rightarrow$ hence the QQ-plot lies close to the line $y = x$

Diagnostics plots of residuals: Quantile-quantile plot

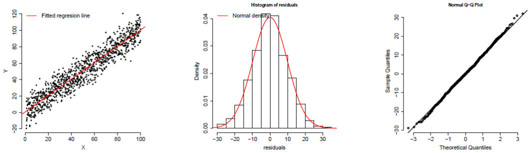


Figure: No visible violation of normality assumption

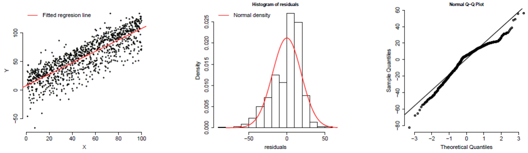


Figure: Residuals are left skewed

Diagnostics plots of residuals: Quantile-quantile plot

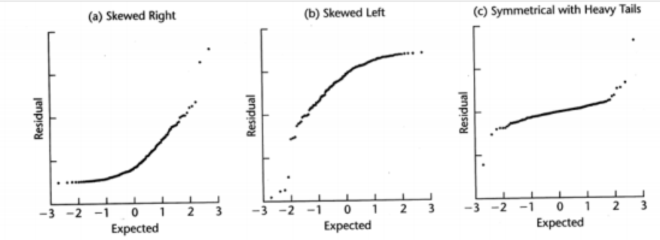
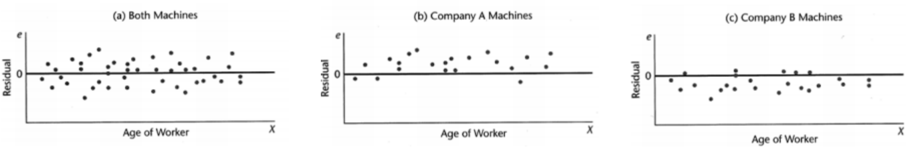


Figure: QQ-plots when normality does not hold

Diagnostics plots of residuals: Omission of Important Predictor Variables

- Example–partitioned the data set with respect to type of machine
- Partitioning data can reveal dependence on omitted variable
- Can suggest that inclusion of other inputs is important



Overview of Tests Involving Residuals

- Tests of randomness (run test, Durbin-Watson test, Chapter 12)
- Tests for constancy of variance (Brown-Forysthe test, Breusch-Pagan test, Section 3.6)
- Tests for Outliers (Chapter 10)
- Tests for normality (Correlation test, Section 3.5)

Correlation Test for Normality

- Test statistic: correlation between sample quantiles and theoretical (normal) quantiles
  $\rightarrow$ A high value of correlation is indicative of normality
- Table B.6 in the textbook provides critical values for a given level α and various sample sizes
  - If, the observed coefficient of correlation is at least as large as the provided critical value, then conclude that the error terms are reasonably normally distributed

Tests for Constancy of Error Variance

- Brown-Forsythe Test
- Breusch-Pagan Test

Tests for Constancy of Error Variance: Brown-Forsythe Test

- Works well when the variance of the error terms either increase or decreases with X
- Works well when the sample size should be large enough to ignore dependencies between the residuals

Tests for Constancy of Error Variance: Brown-Forsythe Test

- Procedure:

  1. Select a cut-off value X_0 for X
     - Group 1 consists of n_1 samples with $X_i \leq X_0$. Associated residuals are denoted by e_{i1}
     - Group 1 consists of n_2 samples with $X_i > X_0$. Associated residuals are denoted by e_{i2}

  2. Calculate the absolute deviation of the residuals of medians in each group.
     - e.g., for group 1: $d_{i1} = |e_{i1} - \tilde{e}_1|$, with $\tilde{e}_1 = \mathrm{median(e_{i1})}$

Tests for Constancy of Error Variance: Brown-Forsythe Test

- Procedure (continued):
  3. Run a two-sample t-test for the d_{i1}'s and d_{i2}'s to test whether their means are equal:

$$t^*_{BF} = \frac{\bar{d}_1 - \bar{d}_2}{s\sqrt{1/n_1 + 1/n_2}}$$

where $\bar{d}_k$ denotes the group mean in group k, and

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

(note the different definition of s!)

Tests for Constancy of Error Variance: Brown-Forsythe Test

- Procedure (continued):
  4. Approximately, $t^*_{BF} \sim t(n-2)$ holds under H_0. Therefore, with confidence level α,

$$\begin{aligned} \text{If } |t^*_{BF}| &\leq t(1-\alpha/2; n-2), \text{ conclude the error variance is constant} \\ \text{If } |t^*_{BF}| &> t(1-\alpha/2; n-2), \text{ conclude the error variance is NOT constant} \end{aligned}$$

Tests for Constancy of Error Variance: Breush-Pagan Test

- Assume error terms are independently and normally distributed and

$$\log \sigma_i^2 = \gamma_0 + \gamma_1 X_i$$

- Tests

$$H_0 : \gamma_1 = 0 \text{ vs. } H_a : \gamma_1 \neq 0$$

Tests for Constancy of Error Variance: Breush-Pagan Test

- Procedure
  1. Regress Y on X, and get SSE
  2. Regress e_i^2 on X_i, and get the regression sum of squares SSR^*
  3. Get test statistic $X_{BP}^2 = \frac{SSR^*}{2} \div \left(\frac{SSE}{n}\right)^2$
  4. When n is reasonably large, $X_{BP}^2 \sim \chi^2(1)$ under H_0. Therefore,

$$\begin{aligned} \text{If } X_{BP}^2 &> \chi^2(1-\alpha;1), \text{conclude } H_a \\ \text{If } X_{BP}^2 &\leq \chi^2(1-\alpha;1), \text{conclude } H_0 \end{aligned}$$

F-Test for Lack of Fit

- Tests whether a linear function adequately fits the data
- Main idea: Assess if linear relation is appropriate by comparing the fits of a linear regression model to "just estimating the mean $E\{Y\}$ at each X level"
- Assumes $Y_i|X_i$ are 1) independent, 2) normally distributed, and 3) have the same variance σ^2
- The tests requires replicates of Y at some X levels (or, if there are no replicates, group observations with similar values of X's)

F-Test for Lack of Fit

- Notation: Y_{ij}'s are then Y's at level X_j,
  with $j = 1, \cdots, c$ (c=number of X levels),
  and $i = 1, \cdots, n_j$ (n_j=number of outcomes at level X_j)
- The test is based on a general linear test approach
  - Full model: $Y_{ij} = \mu_j + \epsilon_{ij}$
  - Reduced model: $Y_{ij} = \beta_0 + \beta_1 X_j + \epsilon_{ij}$
    (ϵ_{ij} are independent $N(0, \sigma^2)$)

GPA example: F-test for lack of fit

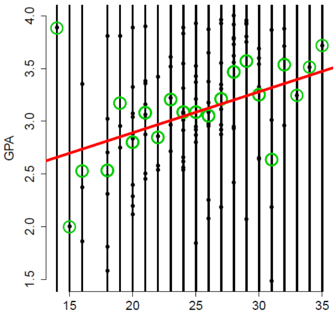


Figure: Fitted regression line (red, $\hat{Y}$ under reduced model) and means at each X level (green, $\hat{Y}$ under full model

F-test for Lack of Fit

- General linear test approach to compare the linear regression model (reduced model under H_0) to "just estimation the mean $E\{Y\}$ at each X level" (full model)
- Test

$$\begin{aligned} H_0 : E\{Y\} &= \beta_0 + \beta_1 X \\ H_a : E\{Y\} &\neq \beta_0 + \beta_1 X \end{aligned}$$

F-test for Lack of Fit

- Full model: $E\{Y_{ij}\} = \mu_j$
    - Predicted values $\hat{Y}_{ij} = \hat{\mu}_j = \bar{Y}_j$
    - $SSE(F) = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$, and $df_R = n - c$
    - $SSE(F)$ is called the pure error sum of squares (SSPE)
    - Based on the best fit under all possible regression relations
- Reduced model under $H_0 : E\{Y_{ij}\} = \beta_0 + \beta_1 X_j$
    - Predicted values $\hat{Y}_{ij} = b_0 + b_1 X_j$
    - $SSE(R) = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_{ij})^2$, and $df_R = n - 2$

F-test for Lack of Fit

- Test statistic:

$$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{SSE - SSPE}{c-2} \div \frac{SSPE}{n-c} \\ &\sim F(c-2, n-c) \text{ under } H_0 \end{aligned}$$

- Decision rule

$$\begin{aligned} \text{If } F^* &> F(1-\alpha; c-2, n-c), \text{ conclude } H_a \\ \text{If } F^* &\leq F(1-\alpha; c-2, n-c), \text{ conclude } H_0 \end{aligned}$$

F-test for Lack of Fit: GPA example

```
Console C:/Users/Yunjin/Dropbox/teaching/ST5202/Week3/
> colnames(gpa.example) = c("Y", "X")
> full.model = lm(Y ~ factor(X), data = gpa.example)
> reduced.model = lm(Y ~ X, data=gpa.example)
> anova(reduced.model, full.model)
Analysis of Variance Table

Model 1: Y ~ X
Model 2: Y ~ factor(X)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    118 45.818
2     99 39.332 19    6.4857 0.8592 0.6324
> 
```

- We don't reject H_0, as there is no statistical evidence that the linear model is inappropriate.

F-test for Lack of Fit: Interpretation and extended ANOVA table

- What's the difference between this test and the F-test for $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$?
  - F-test for lack of fit is used to determine if a linear model is appropriate, rejecting H_0 means that the linear model is not appropriate
  - F-tset for slope is used to determine if the linear association between X and Y is significant, this F-test is not useful if a linear model is not appropriate!

F-test for Lack of Fit: Interpretation and extended ANOVA table

- Extended ANOVA table:
  - $SSE = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_{ij})^2 = SSPE + SSLF$
  - $SSPE$ = Pure error sum of squares $= \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$
  - $SSLF$ = "Lack of fit" sum of squares of linear regression model

$$SSLF = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (\hat{Y}_{ij} - \bar{Y}_j)^2$$

- Degrees of freedom:

$$\begin{aligned} df(SSE) &= df(SSPE) + df(SSLF) \\ n-2 &= (n-c) + (c-2) \end{aligned}$$

Overview of Remedial Measures: What do we do if the linear regression model is inappropriate?

- If simple regression model is not appropriate, then we have two choices:
  - Abandon simple regression model, then develop and use a more appropriate model
  - Employ some transformation on the data so that linear regression model is appropriate for the transformed data

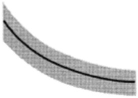
Overview of Remedial Measures: What do we do if the linear regression model is inappropriate?

- Nonlinearity of regression function $\rightarrow$ Transformations (Section 3.9)
- nonconstancy of error variance $\rightarrow$ weighted least squares (Chapter 11) or transformations (Section 3.9)
- Nonindependence of Error terms $\rightarrow$ work with a model that calls for correlated error term (Chapter 12)
- Nonnormality of error terms $\rightarrow$ Transformations (Section 3.9)
- Omission of important predictor variables $\rightarrow$ modify the model (Multiple regression analysis in Chapter 6 and forward)
- Outlying Observations $\rightarrow$ robust regression (Chapter 11)

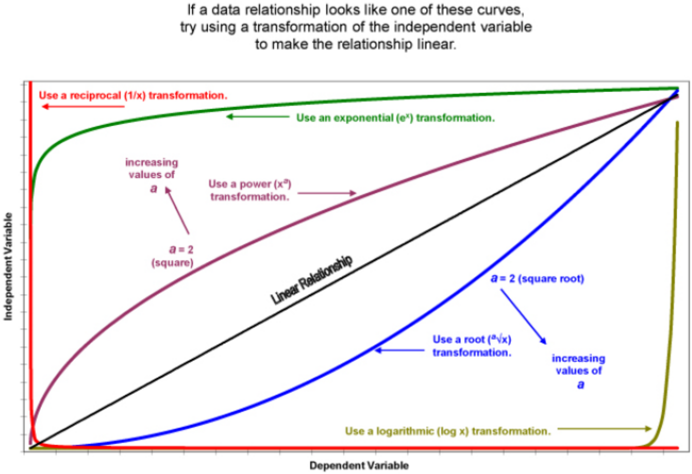
Transformations:
For nonlinearity relation only

- When the distribution of the error terms is reasonably close to a normal distribution and the error terms have approximately constant variance.
- Transformation of X should be attempted
- Transformationof Y should be refrained since it will affect the distribution of the error terms

Transformations:
For nonlinearity relation only

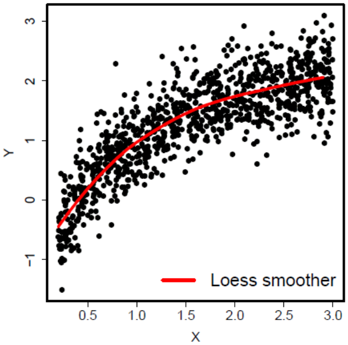
	Prototype Regression Pattern	Transformations of X
(a)		$X' = \log_{10} X$ $X' = \sqrt{X}$
(b)		$X' = X^2$ $X' = \exp(X)$
(c)		$X' = 1/X$ $X' = \exp(-X)$

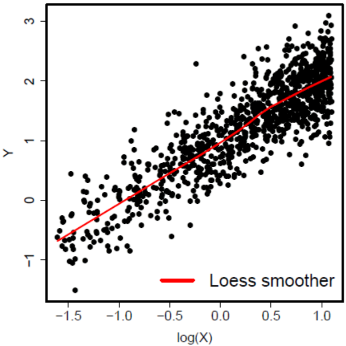
Transformations: For nonlinearity relation only



Transformations: For nonlinearity relation only

- Transformations can help satisfy the assumption of a linear regression model
- Transform X to linearize a nonlinear regression function

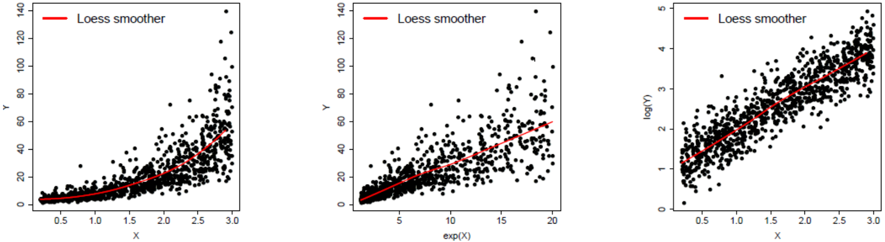




Transformations: For Nonnormality and Unequal Error Variances

- Non-normality and unequal variances of error terms frequently appear together
- To remedy these in the normal regression moidel, we need to transform Y
    - Shapes and spreads of distributions of Y need to be changed
    - May help linearize a curvlinear regression relation
    - Disadvantage: interpretations are on the transformed scale, so they can be more difficult
- Can be combined with transformation on X

Transformations: For Nonnormality and Unequal Error Variances–example



Transformations: Transforming Y's using Box-Cox Transformations

- Sometimes, it can be difficult to determine from diagnostic plots which transformation on Y is most appropriate
- The Box-Cox procedure automatically identifies a transformation from the family of power transformations on Y.

Transformations: Transforming Y's using Box-Cox Transformations

- In Box-Cox transformations, a power tranform $Y' = Y^\lambda$ is used as the response variable:

$$Y' = \begin{cases} K_1(Y^\lambda - 1) & \lambda \neq 0 \\ K_2(\log_e Y) & \lambda = 0 \end{cases}$$

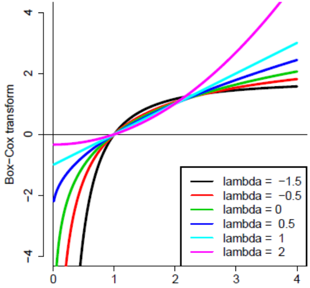
where $K_1 = \frac{1}{\lambda K_2^{\lambda-1}}$ and $K_2 = (\prod_{i=1}^n Y_i)^{1/n}$
(standardized so that the magnitude of the error sum of squares does not depend on the value of λ)

- Or simply:

$$Y' \propto \begin{cases} Y^\lambda & \lambda \neq 0 \\ \log_e Y & \lambda = 0 \end{cases}$$

Transformations:
Transforming Y's using Box-Cox Transformations

$$Y' \propto \begin{cases} Y^{\lambda} & \lambda \neq 0 \\ \log_e Y & \lambda = 0 \end{cases}$$



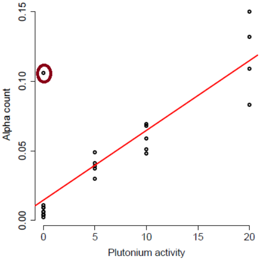
- $\lambda > 1$ spreads out large values of Y and compress small values
- $\lambda < 1$ compress larges values of Y and spreads out small values

Transformations: Transforming Y's using Box-Cox Transformations

- Select optimal λ with maximum likelihood estimation (plug in Y' as dependent variable instead of Y)
- Often likelihood is relatively flat around optimal λ, so choose a number that is easier to interpret, like $0.5, 2, -0.5$
- R command for finding λ: "boxcox($Y \sim X$, plotit=T)"

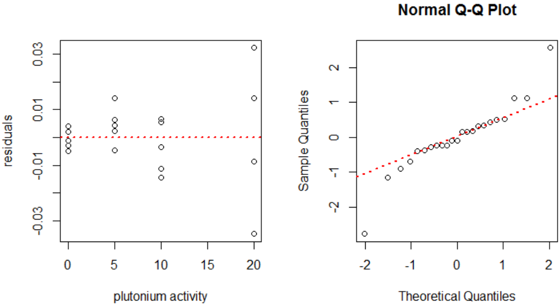
Plutonium example (Section 3.11)

Examine the relation between plutonium activity and the number of alpha particles that it submits per second.



Anything wrong?

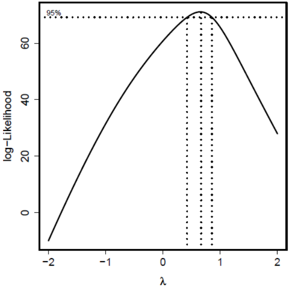
Plutonium example (Section 3.11)



Anything wrong?

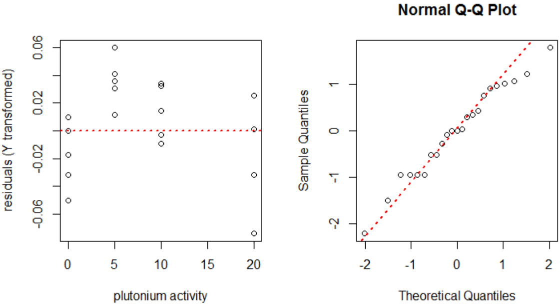
Plutonium example (Section 3.11)

Box-Cox transformation to find λ



Plutonium example (Section 3.11)

Transform Y so that $Y' = \sqrt{Y}$ $(\lambda = .5)$



Anything wrong?

Plutonium example (Section 3.11)

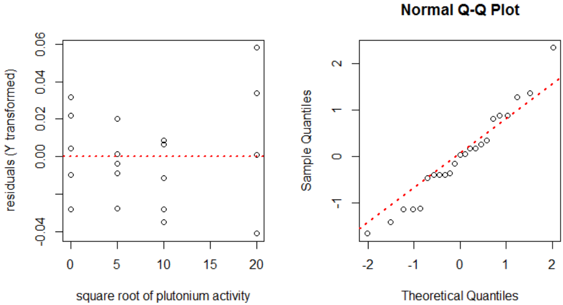
Transform Y so that $Y' = \sqrt{Y}$ ($\lambda = .5$);
Lack of fit F-test

```
Analysis of Variance Table

Model 1: sqrt(Y) ~ X
Model 2: sqrt(Y) ~ factor((X))
  Res.Df      RSS Df Sum of Sq      F   Pr(>F)
1     21 0.023453
2     19 0.011346  2  0.012106 10.136 0.00101 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plutonium example (Section 3.11)

Transform both X and Y so that $X' = \sqrt{X}$ and $Y' = \sqrt{Y}$ ($\lambda = .5$)



Anything wrong?

Plutonium example (Section 3.11)

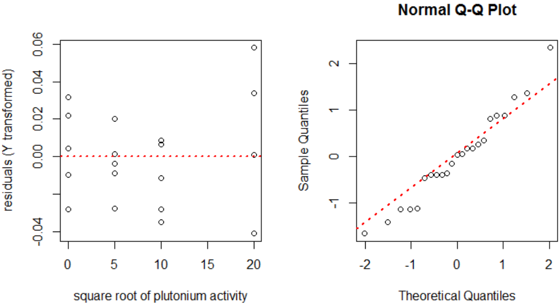
Transform both X and Y so that $X' = \sqrt{X}$ and $Y' = \sqrt{Y}$ ($\lambda = .5$);
Lack of fit F-test

```
Analysis of Variance Table

Model 1: sqrt(Y) ~ sqrt(X)
Model 2: sqrt(Y) ~ factor(sqrt(X))
  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1     21 0.012883
2     19 0.011346  2 0.0015368 1.2868 0.2992
> 
```

Plutonium example (Section 3.11)

Transform both X and Y so that $X' = \sqrt{X}$ and $Y' = \sqrt{Y}$ ($\lambda = .5$)



Anything wrong?

Diagnostics and Remedial Measures Summary

- Non-linear regression function:
  - Diagnose with residual plots (e_i versus X_i) and F-test for lack of fit
  - Solve it with:
    - Transformations of X (not Y, why?)
    - Polynomial regression (chapter 8)
    - Non-linear regression (part III)
- Non-constancy of error variance
  - Diagnose with residual plots ($|e_i|$ versus X_i) and Brown-Forsythe/Breusch-Pagan test
  - Solve it with:
    - Transformations of Y
    - Weighted least-squares (chapter 11)
- Outliers and influential points (chapter 11)
- Non-independence of errors (chapter 12)

Reading: Section 2.6 & whole Chapter 3