STA302/1001 - Methods of Data Analysis I

(Week 07 lecture notes)

Wei (Becky) Lin

Oct 24, 2016

UNIVERSITY OF TORONTO

Important

- Morning sections: Class on October 27 will take place in ES 1050.
- A1 result is available today
  - Write me mail with your Last name + first name + student ID if you had my permission to waive the lateness penalty. Crowdmark does everything automatically, I need to adjust it back.
- A2 will be posted this week. Waiting for my announcement.
- Make-up midterm
  - Location: BA1170
  - Time: AM 10-12 (90 minutes), Oct. 29. (This Saturday.)
  - A calculator and your student ID.
- Final is on Dec. 12, AM 9-12
  - A to J: BN 2N
  - K to R: BN 2S
  - S to ZHE: BN 3
  - Zho to ZZ: ZZ VLAd

Lecture before Midterm

- Unusual data points: outliers, high levage points, influential points.
- Diagnostics for residual
  - Check linearity by residuals vs fitted values plot or (Scatter plot of Y and X)
  - Check constant variance by residual plots
  - Check Normality by Normal QQ-plot
  - Identify unusual data points.
- Influence Metrics: DFFITS, DFBETAS,COOK's distance
- Case study

Week 07- Learning objectives & Outcomes

- Variable transformations.
- More on logarithmic transformation.
- Box-Cox transformation.
- Interpretation of slope after transformation.
- Chapter 4: Simultaneous Inferences

Variable Transformations

Transformations

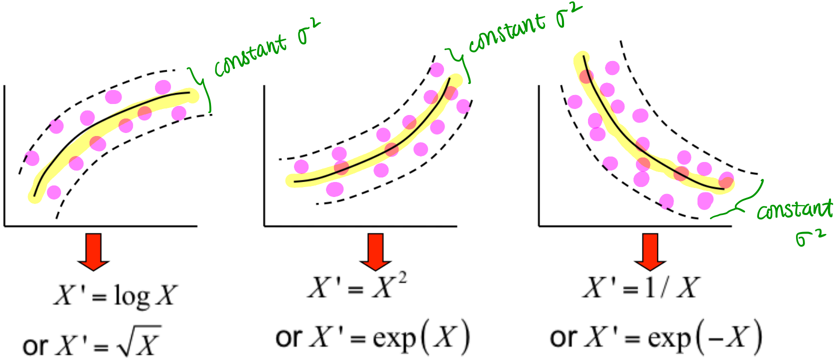
- Why?
  - Satisfy model assumptions.
  - Improve predictive ability.
  - Make it easier to interpret parameters.
- How?
  - First fit a linear regression model to the original variables. Diagnostics indicate
    - Nonlinearity: transformation on X.
    - Nonlinearity, nonconstant variance and non-normality: transformation on Y (transformation on X might also helpful).
    - Box-Cox transformation.
  - Fit linear regression model after the transformations for one or both of the original variables.
  - To make the regression model appropriate for the transformed data.

Transformations (cont.)

- Several alternative transformations may be tried.
- Scatter plot and residual plots based on each transformation should then be prepared and analyzed to decide which transformation is most effective. Check SSE.

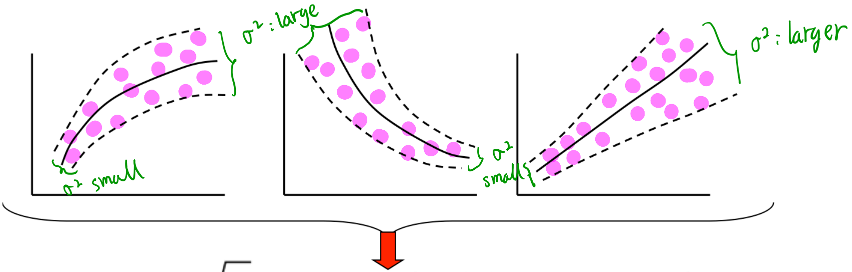
Transformations on X

- $X_i' = f(X_i),\ Y_i = \beta_0 + \beta_1 X_i' + \epsilon_i$
  - Correct nonlinearity, when constant variance in residual is satisfied.



Transformations on Y

- $Y_i' = f(Y_i)$, $Y_i' = \beta_0 + \beta_1 X_i + \epsilon_i$
  - Help to fix unequal error variances; non-normality of error terms.
  - Also help to linearize a curvilinear regression.
  - A simultaneous transformation on X may also be helpful or necessary.



$$Y' = \sqrt{Y}, \text{ or } Y' = \log(Y), \text{ or } Y' = 1/Y$$

Example: Transformation on X

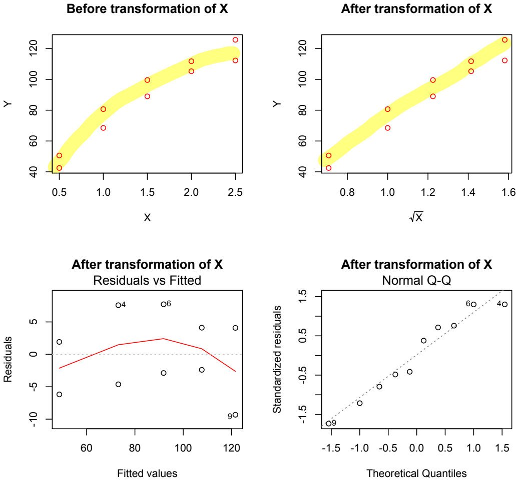
```
# Table 3.7: 10 participants in the study
# X: the number of days of training received;
# Y: performance score in a battery of simulated sales situations
X = c(0.5, 0.5, 1,1,1.5,1.5,2,2,2.5,2.5)
Y=c(42.5,50.6,68.5,80.7,89.0,99.6,105.3,111.8,112.3,125.7)
# Use of Square Root Transformation of X
Xp = sqrt(X)

fit0 = lm(Y~X)
fit1 = lm(Y~Xp)

par(mfrow=c(2,2))
plot(Y~X, type="p",col="red",main="Before transformation of X")
plot(Y~Xp,type="p",,col="red",xlab=expression(paste(sqrt(X))),
     main="After transformation of X" )
plot(fit1,1,main="After transformation of X")
plot(fit1,2,main="After transformation of X")
```

Example: Transformation on X (cont.)

- Diagnostic for $\hat{Y} = -10.33 + 84.35\sqrt{X}$: no evidence of lack of fit or strongly unequal error variances.



Example: Transformation on X (cont.)

```
X = c(0.5, 0.5, 1,1,1.5,1.5,2,2,2.5,2.5)
Y=c(42.5,50.6,68.5,80.7,89.0,99.6,105.3,111.8,112.3,125.7)
Xp = sqrt(X)

fit0 = lm(Y~X); fit1 = lm(Y~Xp)
anova(fit0); anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X          1 6397.5  6397.5  99.464 8.66e-06 ***
## Residuals  8  514.6    64.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value     Pr(>F)
## Xp         1 6597.3  6597.3  167.72 1.197e-06 ***
## Residuals  8  314.7    39.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Transformation on X (cont.)

- Find confidence interval for β_0, β_1
- Find confidence interval for $E(Y)$ when $\sqrt{X} = 1.2$
- Find Prediction interval for Y when $\sqrt{X} = 1.2$
- Find Prediction interval for Y when $\sqrt{X} = c(1, 1.2)$

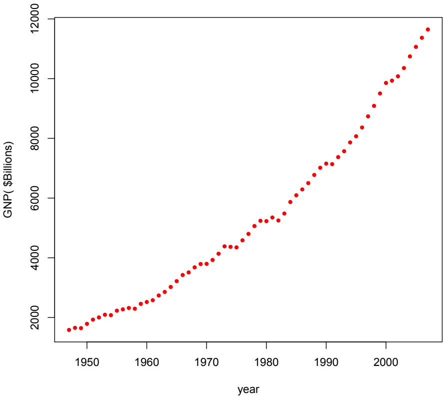
```
>
> confint(fit1)
                2.5 %    97.5 %
(Intercept) -28.52761  7.871247
Xp           68.59288 98.312426
> predict.lm(fit1,newdata=data.frame(Xp=1.2),interval="confidence")
     fit      lwr     upr
1 89.815 85.23631 94.3937
> predict.lm(fit1,newdata=data.frame(Xp=1.2),interval="prediction")
     fit      lwr      upr
1 89.815 74.64462 104.9854
> predict.lm(fit1,newdata=data.frame(Xp=c(1,1.2)),interval="prediction")
       fit      lwr       upr
1 73.12447 57.70737  88.54158
2 89.81500 74.64462 104.98539
>
```

PI for $X_h=1$ and $X_h=1.2$

Example: Transformation on Y

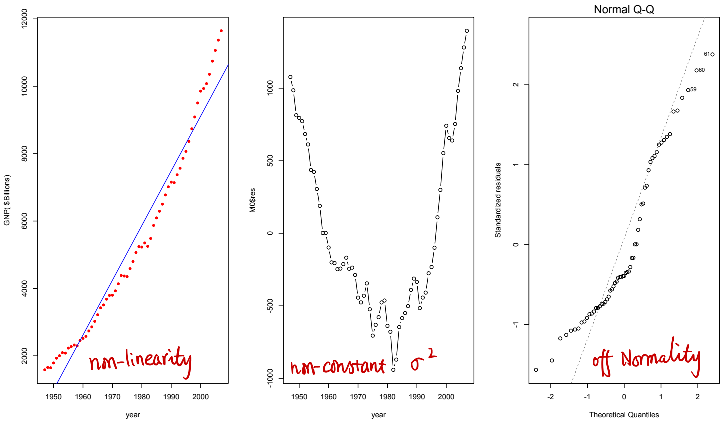
Annual US GNP data analysis

- US GNP data (1947-2007)
  - Y: annual (adjusted) US GNP (Gross National Product) (in $Billions).
  - X: years



Annual US GNP data analysis (cont.)

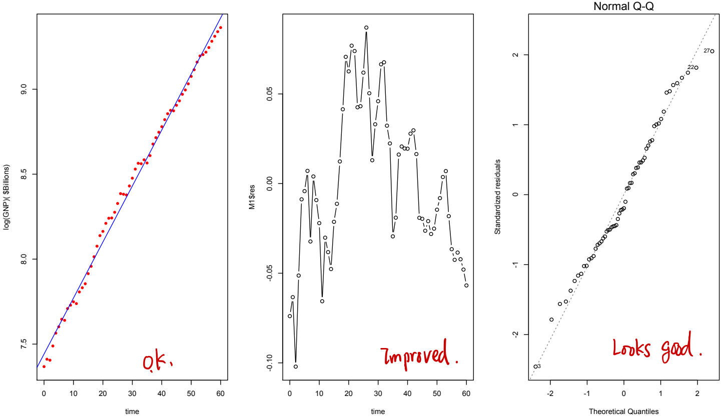
- Fitted model, M0: $GNP_t = -315741.23 + 162.43\ Year_t + \epsilon_t$
- $\sqrt{MSE} = 606.4$, $R^2 = 0.9583$



Annual US GNP data analysis (cont.)

$\log GNP$, $t_i = year - 1947$

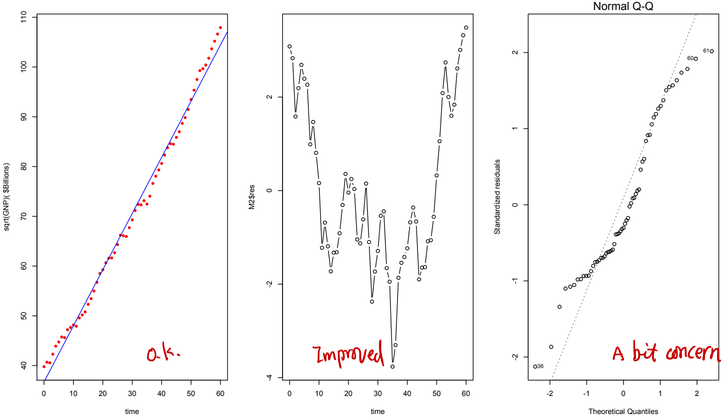
- Fitted model, M1: $log(GNP_i) = 7.44134 + 0.03297\, t_i + \epsilon_i$
- $\sqrt{MSE} = 0.04279$, $R^2 = 0.9948$



Annual US GNP data analysis (cont.)

$\sqrt{GNP}$, $t_i = year - 1947$

- Fitted model,M1: $\sqrt{GNP} = 36.70997 + 1.1288\, t_i + \epsilon_i$
- $\sqrt{MSE} = 1.786$, $R^2 = 0.9923$



Annual US GNP data analysis (cont.)

Which transformation is best?

Y	X	$\sqrt{MSE}$	R^2
GNP	year	606.4	0.9583
$\log(GNP)$	t=year-1947	0.04279	0.9948
$\sqrt{GNP}$	t=year-1947	1.786	0.9923

Compare SSE's in terms of original units

- $SSE_{linear} = \sum_i (Y_i - \hat{Y}_i)^2 = 21,697,236$
- $SSE_{log} = \sum_i (Y_i - \exp \widehat{\log Y_i})^2 = 3,005,365$
- $SSE_{sqrt} = \sum_i (Y_i - \widehat{\sqrt{Y_i}}^2)^2 = 4,628,230$

all fitted values can be obtained in R

In this case, logarithmic transformation also offers better interpretation: exponential growth model for GNP (b_1 is the estimated annual growth rate).

$= GNP_0 (e^{\tilde{\beta}_1})^t e^{\varepsilon_i} \doteq GNP_0 (1+\tilde{\beta}_1)^t e^{\varepsilon_i}$

$$GNP_i = GNP_0 \exp(\widetilde{\beta}_1 t + \epsilon_i) \approx GNP_0 (1 + \widetilde{\beta}_1)^t e^{\epsilon_i}$$

Taylor Expansion

$e^x \doteq 1+x$

$$\log(GNP_i) = \log(GNP_0) + \beta_1 t + \epsilon_i$$

Annual US GNP data analysis (cont.)

Based on the logarithmic transformation model,

- Find confidence interval for β_0, β_1
- Find confidence interval for $E(GNP)$ when $time = 50$, i.e. at year 1997.
- Find Prediction interval for GNP when $time = 63$, i.e, at year 2010.

$$\log(Y_h) : [L, U] = \widehat{\log(Y_h)} \pm t_{1-\alpha/2,n-2} s(pred)_{\log(Y_h)}$$

```
> confint(M1) # log(GNP)~ (year-1947)
                  2.5 %     97.5 %
(Intercept) 7.41968355 7.46300624
time        0.03234954 0.03359498
> # time=64 ->  1947+50=year 1997; 1947+63= year 2010,
> predict.lm(M1,newdata=data.frame(time=50),interval="confidence",level=0.90)
       fit      lwr      upr
1 9.089958 9.076101 9.103815
> exp(predict.lm(M1,newdata=data.frame(time=50),interval="confidence",level=0.90))
       fit      lwr      upr
1 8865.811 8743.803 8989.522
> predict.lm(M1,newdata=data.frame(time=63),interval="prediction",level=0.95)
       fit      lwr      upr
1 9.518597 9.429854 9.607341
> exp(predict.lm(M1,newdata=data.frame(time=63),interval="prediction",level=0.95))
      fit     lwr      upr
1 13610.5 12454.7 14873.56
>
```

$P(lwr < E\log Y_h < upr) = 0.90$

$P(e^{lwr} < EY_h < e^{upr}) = 0.90$

$P(9.43 < \log Y_h < 9.61) = 0.95$

$P(e^{9.43} < Y_h < e^{9.61}) = 0.95$

More on logarithmic transformation

- The default logarithmic transformation merely involves taking the natural logarithm — denoted ln or loge or simply log10 — of each data value.
- One could consider taking a different kind of logarithm, such as log base 10, or log base 2. In R, log10(), log2().
- However, the natural logarithm, log base e where e is the constant 2.718282. . . , is the most common used in practical.

Why consider the natural logarithmic transformation:

- Small values that are close together are spread further out.
- Large values that are spread out are brought closer together.

Why Might Logarithms Work?

Logarithms are often used because they are connected to common exponential growth and power curve relationships.

- The exponential growth equation for variables y and x

$$y = a * e^{bX} \Rightarrow \log(y) = \log(a) + bX$$

- A general power curve equation is

$$y = a * x^{b} \Rightarrow \log(y) = \log(a) + b\log(X)$$

  This regression equation is sometimes referred to as a log-log regression equation.

Example: logarithmic transformation

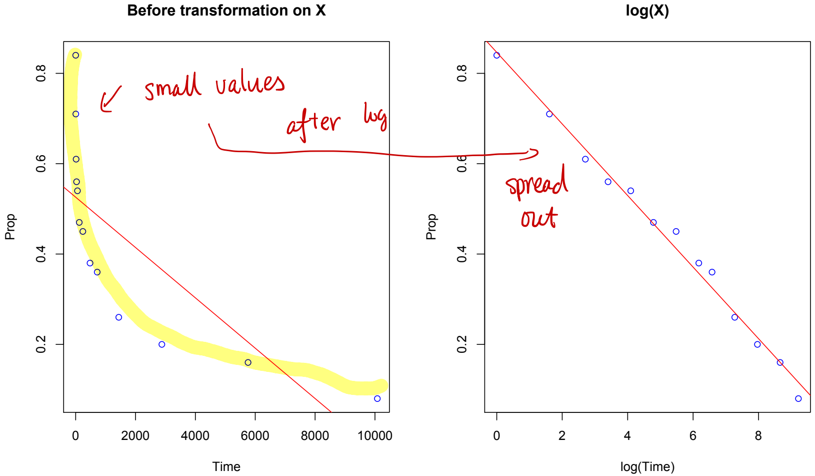
Data:

- A memory retention experiment in which 13 subjects were asked to memorize a list of disconnected items. The subjects were then asked to recall the items at various times up to a week later.
- The proportion of items (y = prop) correctly recalled at various times (x = time, in minutes) since the list was memorized

```
>
> t(word)
     [,1] [,2]  [,3]  [,4]  [,5]   [,6]   [,7]   [,8]   [,9]   [,10]  [,11]   [,12]    [,13]
time 1.00 5.00 15.00 30.00 60.00 120.00 240.00 480.00 720.00 1440.00 2880.0 5760.00 10080.00
prop 0.84 0.71  0.61  0.56  0.54   0.47   0.45   0.38   0.36    0.26    0.2    0.16     0.08
>
```

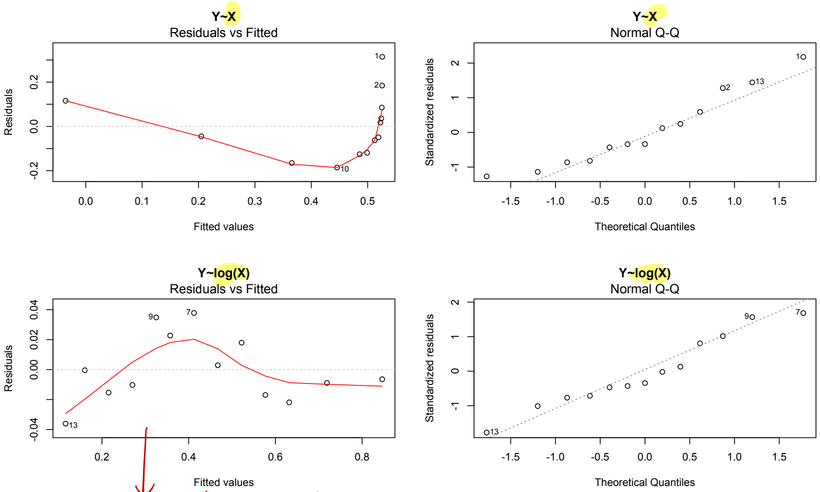
Example: logarithmic transformation (cont.)

- Scatter plot



Example: logarithmic transformation (cont.)

- Diagnostics: residual plot and Normal QQ-plot



Good to me since n=10.

A summary on variable transformation

- Using transformations is part of an iterative process where all the linear regression assumptions are re-checked after each iteration.
  - We transform the predictor (X) values only.
  - We transform the response (Y) values.
  - We transform both the predictor (X) values and response (Y) values.
- W try a transformation and then check to see if the transformation eliminated the problems with the model. If it doesn't help, we try another transformation and so on.

Model building

- Model formulation
- Model estimation
- Model evaluation *

Model use

Box-Cox transformation

- It is often difficult to determine from diagnostic plots which transformation of Y is most appropriate.
  - Correcting skewness of the distribution of error terms.
  - unequal error variances.
  - nonlinearity of the regression function.
- Box-Cox procedure: automatically identifies a transformation from the family of power transformations on Y.

Box-Cox transformation (cont.)

- The family of power transformations:

$$Y' = Y^{\lambda}$$

- λ: a parameter to be determined from the data.

$$
\begin{array}{ll}
\lambda = 2 & Y' = Y^2 \\
\lambda = 0.5 & Y' = \sqrt{Y} \\
\lambda = 0 & Y' = \log_e(Y) \quad (\textit{by definition}) \\
\lambda = -0.5 & Y' = 1/\sqrt{Y} \\
\lambda = -1 & Y' = 1/Y
\end{array}
$$

Box-Cox transformation (cont.)

- The model becomes

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

- λ : need to be estimated by MLE as well as $\beta_0, \beta_1, \sigma^2$
- Simple procedure to obtained $\hat{\lambda}$: search in a range of potential λ
  - λ grid on : -2, 1.75, . . ., 1.75,2
  - Each λ: standardized the $\hat{Y}_i^{\lambda}$.

Box-Cox transformation (cont.)

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

$$K_1 = \frac{1}{\lambda K_2^{\lambda - 1}}, \quad K_2 = (\prod_{i=1}^{n} Y_i)^{1/n}$$

Note that K_2 is the geometric mean of the Y_i observations.

- W_i is the standardized observation of Y_i so that the magnitude of the error sum of squares does not depend on λ
- $W_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- The MLE $\hat{\lambda}$ is that value of λ for which SSE is a minimum.
- Scatter and residual plots should be utilized to examine the appropriateness of the transformation identified by the Box-Cox procedure.

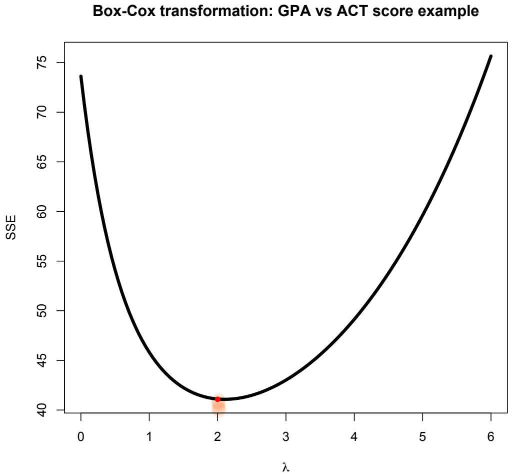
Box-Cox transformation (cont.)

- Theoretical or a priori considerations can be utilized to help in choosing an appropriate transformation.
- When transformation models are employed, the estimators b_0, b_1 have the least squares properties with respect to the transformation observations, not the original ones.
- The MLE of λ with the Box-Cox procedure is subject to sampling variability.
    - SSE is fairly stable in a neighbourhood around the estimate.
    - a nearby λ: easy to understand. EX: $\lambda = 0$ instead of $\hat{\lambda} = 0.13$.
- when $\lambda = 1 \Leftarrow$ no transformation of Y may be needed.

Example: Box-Cox transformation

GPA and ACT score

- R code to generate the plot and data are available on portal



Interpretation after Transformations

Interpretation of slope (β_1)

Model	Response	predictor
Level-level	Y	X
Level-log	Y	$\log(X)$
Log-Level	$\log(Y)$	X
Log-log	$\log(Y)$	$\log(X)$

- For Level-level model, $E(Y|X) = \beta_0 + \beta_1 X$. The interpretation of β_1: on average, Y changes by β_1 as X increases by 1 unit.

$$E(Y|X+1) - E(Y|X) = (\beta_0 + \beta_1(X+1)) - (\beta_0 + \beta_1 X) = \beta_1$$

Interpretation of slope (β_1): Level-log model

$$E(Y|\log(X)) = \beta_0 + \beta_1 \log(X)$$

- For level-log model, the interpretation of β_1: associated with each two-fold increase (i.e doubling) of X, there is a $\beta_1 \log(2)$ change in the mean of Y.

$$E(Y|2X) - E(Y|X) = \beta_0 + \beta_1 \log(2X) - [\beta_0 + \beta_1 \log(X)] = \beta_1 \log(2)$$

- For example
  - Y = pH
  - X = time after slaughter (hrs.)
  - Estimated model: $\hat{Y} = 6.98 - 0.73 \log(X)$
  - Interpretation of b_1: it is estimated that for each doubling of time after slaughter (between 0 and 8 hours) the mean pH decreases by 0.5 $= 0.73 * \log(2)$.

Interpretation of slope (β_1): Log-level model

$$E(\log(Y)|X) = \beta_0 + \beta_1 X \Leftarrow E(Y) = e^{\beta_0 + \beta_1 X}$$

As X increases by 1, what happens?

$$\frac{E(Y|X+1)}{E(Y|X)} = \frac{e^{\beta_0 + \beta_1 (X+1)}}{e^{\beta_0 + \beta_1 X}} = e^{\beta_1}$$

Interpretation:

- As X increases by 1, the mean of Y changes by the multiplicative factor of e^{β_1}.
- If $\beta_1 > 0$: As X increases by 1, the mean of Y increases by $(e^{\beta_1} - 1) * 100\%$
- If $\beta_1 < 0$: As X increases by 1, the mean of Y decreases by $(1 - e^{\beta_1}) * 100\%$

Example:

- Estimated model: $\hat{\log}(Y) = 18.96 \underbrace{-0.50}_{<0} X$
- $1 - e^{-0.5} = 0.4$
- Interpretation: it is estimated that, on average, Y decreases by 40% with each one unit increases in X.

Interpretation of slope (β_1): both Y and X logged

$$E(\log(Y)|X) = \beta_0 + \beta_1 \log(X)$$ ← log-log regression model

Interpretation:

- Associated with each doubling of X, the mean of Y changes by the the multiplicative factor of $e^{\beta_1 \log(2)}$.
- If $\beta_1 > 0$: As X is doubled, the mean of Y increases by $(e^{\beta_1 \log(2)} - 1) * 100\%$
- If $\beta_1 < 0$: As X is doubled, the mean of Y decreases by $(1 - e^{\beta_1 \log(2)}) * 100\%$

Example:

- Y: number of bird species on an island
- X: island area
- Estimated model: $E(\log(\widehat{Y)|\log}(X)) = 1.94 - 0.25 \log(X)$
  - since $1 - e^{0.25 * \log(2)} = 0.19$
  - Associated with each doubling of island area, it is estimated that there is a 19% decreases in the mean number of bird species.

CH4: Simultaneous Inference

Joint Estimation of β_0 and β_1

- The $100(1-\alpha)\%$ CI of β_0 is

$$[L_0, U_0] = b_0 \pm t_{1-\alpha/2;n-2}s(b_0), \quad s^2(b_0) = MSE(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}})$$

- The $100(1-\alpha)\%$ CI of β_1 is

$$[L_1, U_1] = b_1 \pm t_{1-\alpha/2;n-2}s(b_1), \quad s^2(b_1) = \frac{MSE}{S_{XX}}$$

- What is the confidence coefficient of their joint intervals?
  - $P(L_0 \leq \beta_0 \leq U_0, L_1 \leq \beta_1 \leq U_1) = ?$
  - Let $1-\alpha = 0.95$: Not provide 95% C.I.s for β_0 and β_1 since $(0.95)^2$ if the inferences were independent.

Joint Estimation of β_0 and β_1 (cont.)

- Let A_0 denote the event that the first confidence interval does not cover β_0. Then $P(A_0) = \alpha$
- Let A_1 denote the event that the first confidence interval does not cover β_1. Then $P(A_1) = \alpha$
- Here $A_0^c \bigcap A_1^c$ is the event which indicates that both of the confidence intervals cover β_0 and β_1.

$$P(A_0^c \bigcap A_1^c) = ?$$

Bonferroni inequality: $P(A_0^c \cap A_1^c) \geq 1 - 2\alpha$

$P(A_0^c \cap A_1^c) = P(\overline{A_0 \cup A_1})$

$= 1 - P(A_0 \cup A_1)$

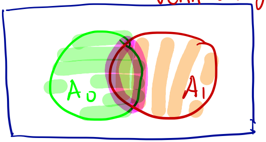
$= 1 - [P(A_0) + P(A_1) - P(A_0 \cap A_1)]$

$= 1 - P(A_0) - P(A_1) + \underbrace{P(A_0 \cap A_1)}_{\geq 0}$

$\Rightarrow P(A_0^c \cap A_1^c) \geq 1 - P(A_0) - P(A_1) = 1 - 2\alpha$

Q.E.D.

Venn diagram



$P(A_0 \cup A_1)$

$= \underbrace{P(A_0) + P(A_1)}_{P(A_0 \cap A_1) \text{ twice}}$

$- P(A_0 \cap A_1)$

Joint Estimation of β_0 and β_1 (cont.)

$\to P(A_0) = 5\%$ $\to P(A_1) = 5\%$ $\Rightarrow P(A_0^c \cap A_1^c) \geq 1 - 5\% \times 2 = 90\%$

- β_0 and β_1 are separately estimated with 95% C.I.. The Bonferroni inequality guarantees us a family confidence coefficient of at least 90% the that both intervals based on the same sample are correct.
- The $1-\alpha$ family confidence limits for β_0 and β_1 for SLR model by the Bonfirroni procedure:

$$b_i \pm Bs\{b_i\}, B = t_{1-\alpha/4;n-2}$$

where i=0 for b_0 and i=1 for b_1.

$\Downarrow$

to ensure $P(A_0^c \cap A_1^c) \geq 1 - 2 \times \frac{\alpha}{2}$

$= 1 - \alpha$

i.e. find $\boxed{1 - \frac{\alpha}{2}}$ CI for β_0 / β_1 respectively.

Example: Joint Estimation of β_0 and β_1

```
toluca=read.table(
  "/Users/Wei/TA/Teaching/0-STA302-2016F/Week07-Oct24/toluca.txt",
                 col.names = c("lotsize", "workhrs"))
# plot(toluca$lotsize,toluca$workhrs)

modt = lm(lotsize~workhrs,data=toluca)
confint(modt)
```

```
##                   2.5 %     97.5 %
## (Intercept) -17.1880966 13.4715943
## workhrs       0.1838466  0.2763702
```

```
confint(modt,level=1-0.05/2) # Bonferroni C.I.
```

to ensure

$P(L_0 < \beta_0 < U_0, L_1 < \beta_1 < U_1) = 0.95$

```
##                  1.25 %    98.75 %
## (Intercept) -19.6277718 15.9112696
## workhrs       0.1764842  0.2837326
```

Joint Estimation of β_0 and β_1 (cont.)

since $P(A_0^c \cap A_1^c) \geq 1-2\alpha$

- The Bonferroni $1-\alpha$ family confidence coefficient is actually a lower bound on the true family confidence coefficient.
- If g interval estimates are desired with family confidence coefficient $1-\alpha$, constructing each interval estimate with statement confidence coefficient $1-\alpha/g$ will suffice.
- The Bonferroni technique is ordinarily most useful when the number of simultaneous estimates is not too large.
- It is not necessary with the Bonferroni procedure that the C.I. have the same statement confidence coefficient. $(P(A_1) + P(A_2) = \alpha)$

If g is large, $1-\alpha/g \to 1$; the CI is too wide to be useful.

Simultaneous Estimation of mean response

- The mean response at a number of X levels need to be estimated.
- Two procedures for simultaneous estimation of a number of different mean responses: Working-Hotelling procedure, Bonferroini procedure,
    - Working-Hotelling procedure
    - Bonferroni procedure

Simultaneous Estimation of mean response

Working-Hotelling procedure:

- Based on the confidence band for the regression line (Chap. 2.6).
- The confidence band contains the entire regression line, so it contains the mean responses at all X levels.
- The simultaneous confidence limits for g mean responses $E\{Y_h\}$

$$\hat{Y}_h + \pm W s\{\hat{Y}_h\}, W^2 = 2F(1-\alpha; 2, n-2)$$

where

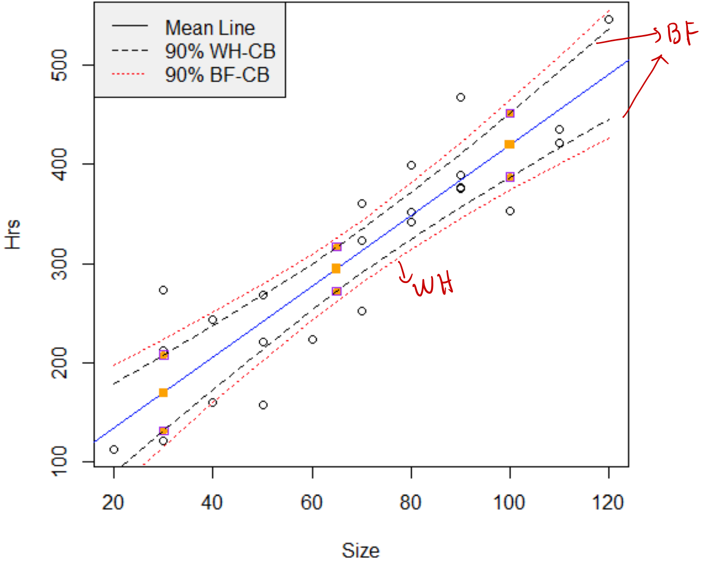
$$\hat{Y}_h = b_0 + b_1 X_h, s\{\hat{Y}_h\} = MSE[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{XX}}]$$

Bonferroin procedure:

- The Bonferroni confidence limits for $E\{Y_h\}$ at g levels X_h with $1-\alpha$ family confidence coefficient:

$$\hat{Y}_h + \pm B s\{\hat{Y}_h\}, B = t_{1-\alpha/(2g); n-2}$$

Example: Simultaneous Estimation of mean response

- 

Simultaneous Prediction Intervals for New observation

- The simultaneous predictions of g new observations on Y in g independent trials at g different levels of X.
- Two procedure:
  - Scheffe Procedure: using the F distribution

$$\hat{Y}_h + \pm Ss\{pred\}, S^2 = gF(1-\alpha; g, n-2)$$

  - Bonferroni procedure: using the t distribution

$$\hat{Y}_h + \pm Bs\{pred\}, B = t_{1-\alpha/(2g);n-2}$$

  - $s^2\{pred\} = MSE[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{XX}}]$
- Reference(http://rstudio-pubs-static.s3.amazonaws.com/5218_61195adcdb7441f7b08af3dba795354f.html)

Good.

Practice problems and upcoming topics

- Practice problems after today's lecture: Chapter 3: 3.9, 3.18, 3.19, 3.20. Chapter 4: 4.1, 4.3, 4.4, 4.8, 4.19, 4.21, 4.24, 4.25.
- Upcoming topics
  - Review on matrices.
  - Ch5: Simple Linear Regression Model in Matrix Terms.
- Reading for upcoming topics: Ch5.9 - Ch5.13.