More on linear regression analysis

Tuan V. Nguyen
Professor and NHMRC Senior Research Fellow
Garvan Institute of Medical Research
University of New South Wales
Sydney, Australia

What we are going to learn

- Interaction effects

Birthweight data

- File: birthsmokers.txt
- N = 32 births
- Baby's birth weight is related mother's smoking habits during pregnancy
- Response (y): birth **weight** in grams of baby
- Potential predictor (x_1):
  - **smoking** status of mother (yes or no)
  - Potential predictor (x_2): length of **gestation** in weeks

Birthweight data

- The model:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$$

- y_i is the birth weight of baby *i* in grams
- x_{i1} is the length of gestation of baby *i* in weeks
- $x_{i2} = 1$, if baby *i*'s mother smoked and $x_{i2} = 0$, if not
- ε_i error terms follow a normal distribution with mean 0 and equal variance σ^2.

Birthweight data

- **The model:**

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$$

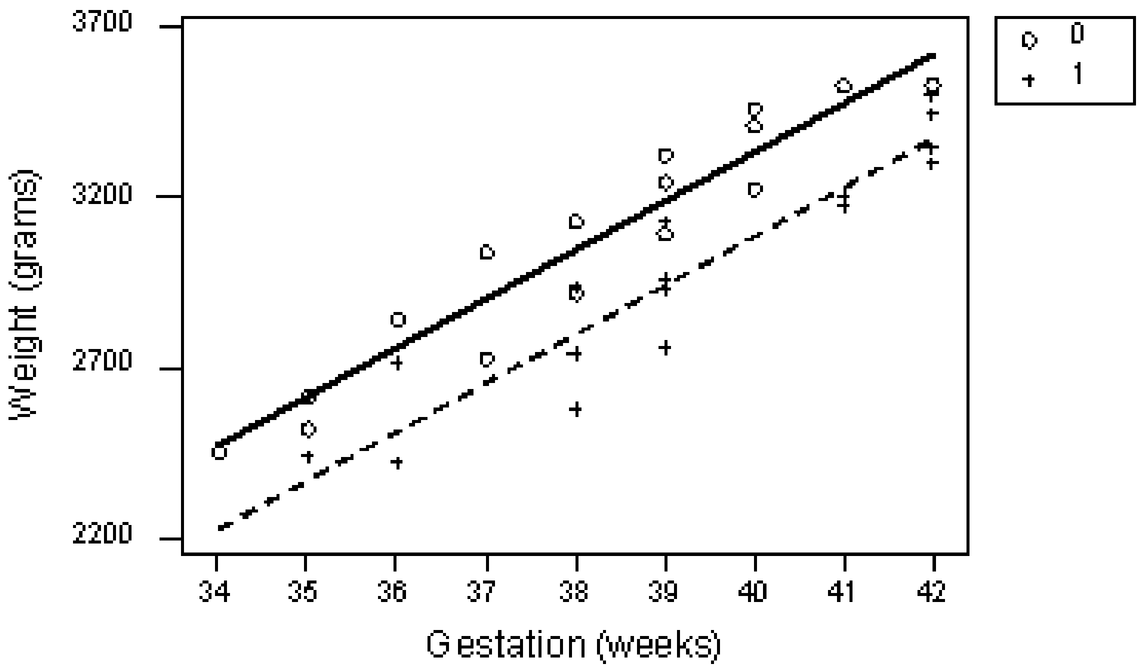
- **The rehression equation based on actual data**

Weight = -2390 + 143*Gest – 245*Smoking

Birthweight data

- **The rehression equation based on actual data**

Weight = -2390 + 143*Gest – 245*Smoking

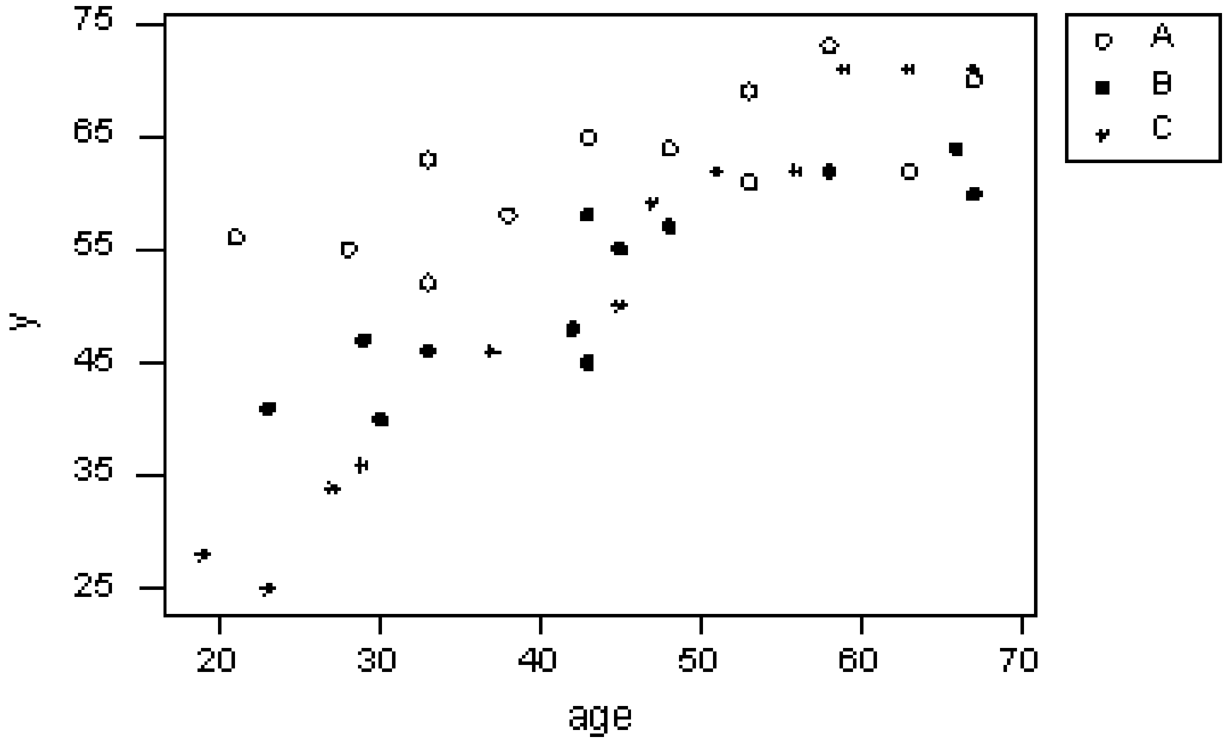


No interaction effects!

Interaction effects

- Depression.txt, 36 patients with depression
- Comparing the effectiveness of 3 treatments for severe depression (A, B, and C)
- y_i = measure of the effectiveness of the treatment for individual i
- x_{i1} = age (in years) of individual i
- x_{i2} = 0 if individual i received treatment A, 1 if treatment B, 2 if treatment C

Interaction effects



Interaction model

- **The "main effect" regression model**

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$$

- **The "interaction effect" regression model**

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i1} x_{i2} + e_i$$

R codes

```
dep =
read.table("/Users/tuannguyen/Documents/_Viet
nam2012/Can Tho /Datasets/depression.txt",
header=T)

attach(dep)

res = lm(y ~ age+TRT)

summary(res)
```

Results of main effect model

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   32.54335    3.58105   9.088 2.23e-10 ***
age            0.66446    0.06978   9.522 7.42e-11 ***
TRTB          -9.80758    2.46471  -3.979 0.000371 ***
TRTC         -10.25276    2.46542  -4.159 0.000224 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Residual standard error: 6.035 on 32 degrees of freedom
Multiple R-squared: 0.784,    Adjusted R-squared: 0.7637
F-statistic: 38.71 on 3 and 32 DF,  p-value: 9.287e-11
```

Interpretation of the main model

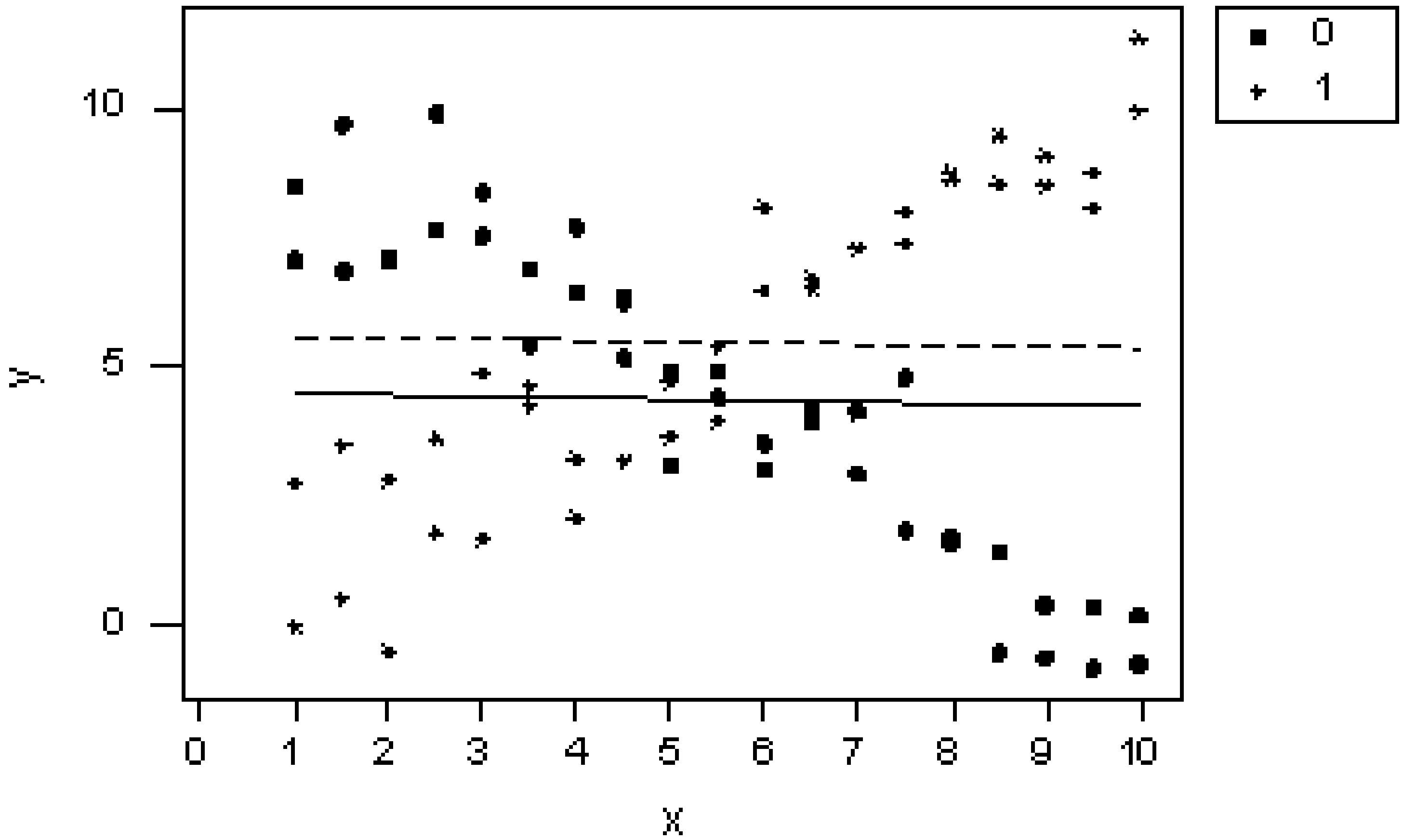
```
(Intercept)  32.54335    3.58105   9.088 2.23e-10 ***
age           0.66446    0.06978   9.522 7.42e-11 ***
TRTB         -9.80758    2.46471  -3.979 0.000371 ***
TRTC        -10.25276    2.46542  -4.159 0.000224 ***
```

Treatment A: y = 32.54 + 0.66*age

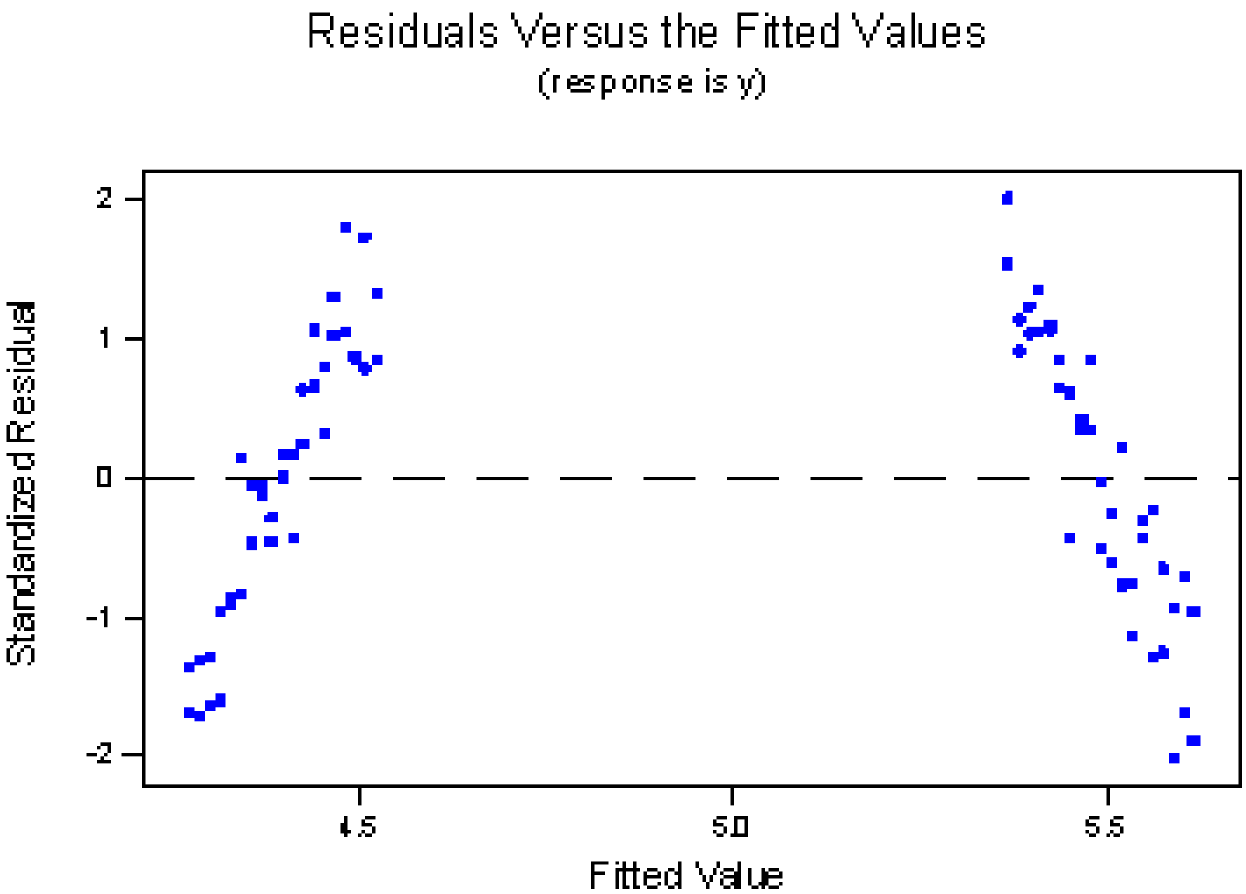
Treatment B: y = 32.54 + 0.66*age – 9.81
= 22.73 + 0.66*age

Treatment C: y = 32.54 + 0.66*age – 10.25
= 22.29 + 0.66*age

The model in graphical format



Residual analysis of "main effect" model



R codes for interaction model

```
dep =
read.table("/Users/tuannguyen/Documents/_Vietnam2012/Can Tho /Datasets/depression.txt",
header=T)

attach(dep)

int = lm(y ~ age+TRT+age:TRT)

summary(int)
```

Results of interaction effect model

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     47.51559    3.82523  12.422 2.34e-13 ***
age              0.33051    0.08149   4.056 0.000328 ***
TRTB           -18.59739    5.41573  -3.434 0.001759 **
TRTC           -41.30421    5.08453  -8.124 4.56e-09 ***
age:TRTB         0.19318    0.11660   1.657 0.108001
age:TRTC         0.70288    0.10896   6.451 3.98e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Residual standard error: 3.925 on 30 degrees of freedom
Multiple R-squared: 0.9143,   Adjusted R-squared: 0.9001
F-statistic: 64.04 on 5 and 30 DF,  p-value: 4.264e-15
```

“Interpretation” of model

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)  47.51559    3.82523  12.422 2.34e-13 ***
age           0.33051    0.08149   4.056 0.000328 ***
TRTB        -18.59739    5.41573  -3.434 0.001759 **
TRTC        -41.30421    5.08453  -8.124 4.56e-09 ***
age:TRTB      0.19318    0.11660   1.657 0.108001
age:TRTC      0.70288    0.10896   6.451 3.98e-07 ***
```

The model is:

If treatment=A, then y = 47.51 + 0.33*age

If treatment=B, then y = 47.51 + 0.33*age – 18.60 + 0.19*age

= 28.91 + 0.52*age

If treatment=C, then y = 47.51 + 0.33*age – 41.30 + 0.70*age

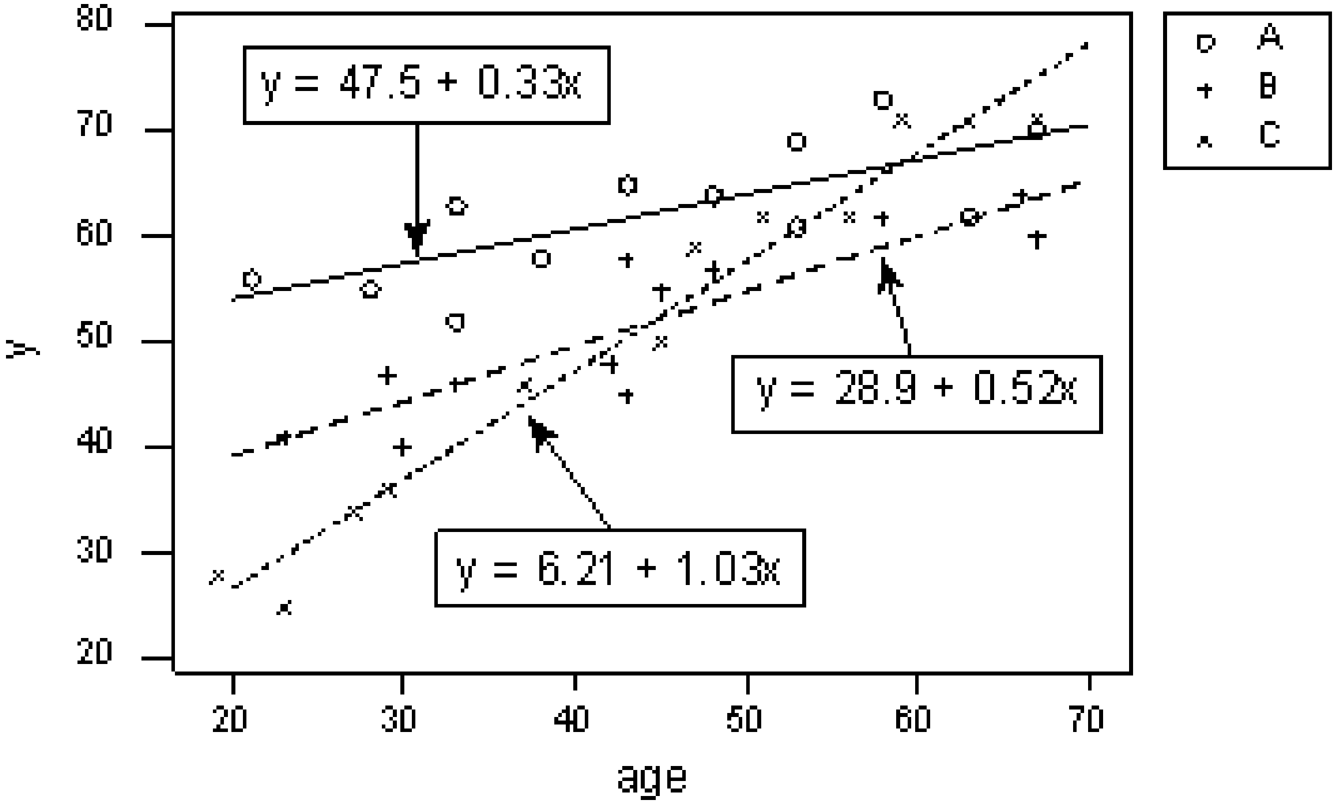
= 6.21 + 1.03*age

"Interpretation" of model

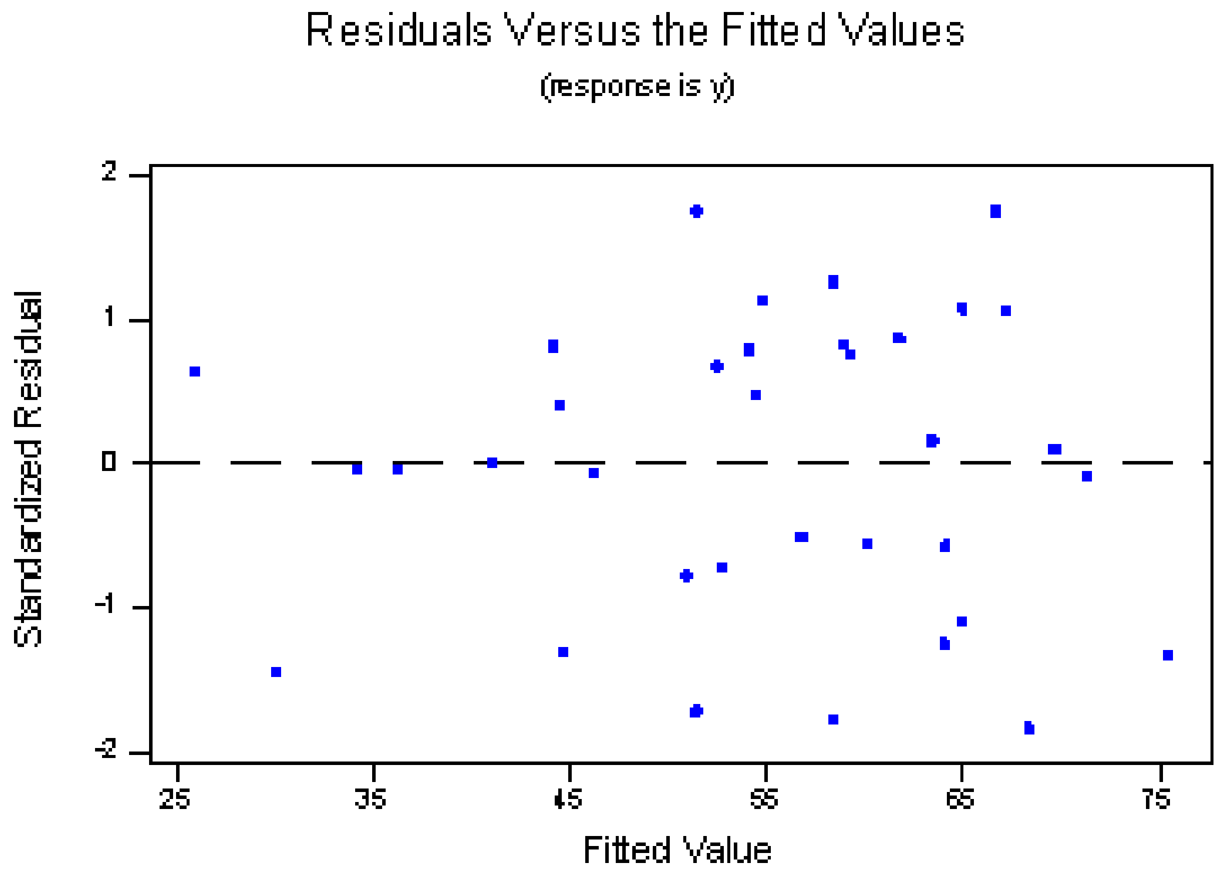
$y_A = 47.51 + 0.33 * age$

$y_B = 28.91 + 0.52 * age$

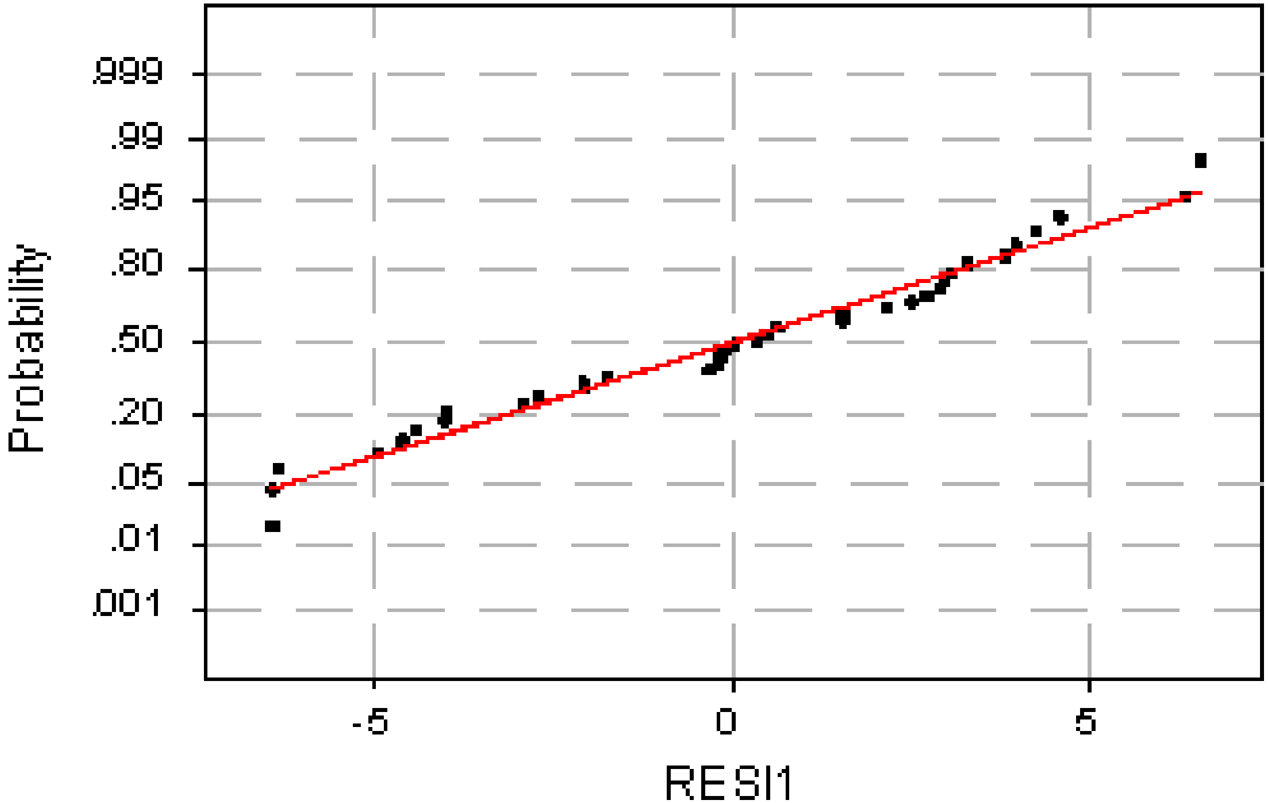
$y_C = 6.21 + 1.03 * age$



Residual analysis



Normal plot



Average: -0.0000000
StDev: 3.63376
N: 36

W-test for Normality
R: 0.9866
P-Value (approx): > 0.1000