22S:152

More on Transformations
Lecture 15

October 25, 2002

Kate Cowles

374 SH

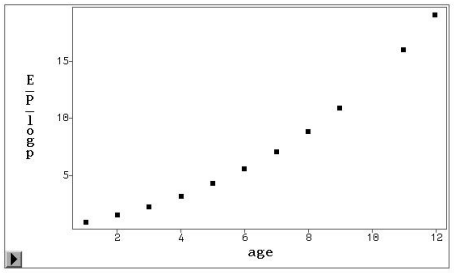
kcowles@stat.uiowa.edu

Plotting the Back Transformed Predicted Values

- To look at the nonlinear relationship between PCB concentration and age
  - transform predicted values from regression equation using transformations back to their original scale

$$\hat{y}_i^* = e^{\hat{y}_i}$$

  - plot these back transformed values vs. age (the untransformed predictor variable)



Variance stabilizing transformations

- Sometimes scientific or statistical theory indicates that the variance of the response variable varies with either
  - value of response variable itself
  - value of predictor
- Example: if response variable is *counts* of something then variance is proportional to expected value
  - square root transformation will stabilize variance
  - $\sqrt{Y}$
  - $\sqrt{Y} + \sqrt{Y+1}$ if some Y_i's are zero or very small

Example: the snow geese data

Aerial survey methods are regularly used to estimate the number of snow geese in their summer range areas west of Hudson Bay in Canada. To obtain estimates small aircraft fly over the range, and when a flock of geese is spotted, an experienced person estimates the number of geese in the flock. To investigate the reliability of this method of counting, an experiment was conducted in which an airplane carrying two observers flew over n = 45 flocks, and each obserer made an independent estimate of the number of birds in each flock. Also, a photograph of the flock was taken so that an exact count of the number of birds in teh flock could be made. The resulting data are given in the dataset "snowgees.dat". (Cooka nd Jacobsen, 1978).

The variables are:

- photo – count based on photo
- obs1 – observer 1's count
- obs2 – observer 2's count

5

```
data geese ;
infile '/group/ftp/pub/kcowles/datasets/snowgees.dat' ;
input photo obs1 obs2 ;
sphoto = sqrt(photo) ;
sobs1 = sqrt(obs1) ;
sobs2 = sqrt(obs2) ;
run ;
```

6

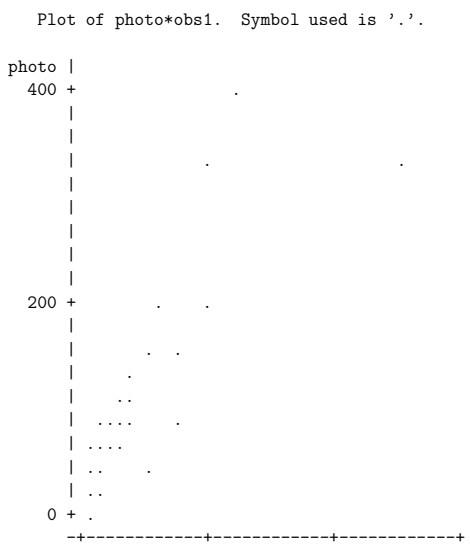
```
 56  50  40
 38  25  30
 25  30  40
 48  35  45
 38  25  30
 22  20  20
 22  12  20
 42  34  35
 34  20  30
 14  10  12
 30  25  30
  9  10  10
 18  15  18
 25  20  30
 62  40  50
 26  30  20
 88  75 120
 56  35  60
 11   9  10
 66  55  80
 42  30  35
 30  25  30
 90  40 120
119  75 200
165 100 200
152 150 150
205 120 200
409 250 300
342 500 500
200 200 300
 73  50  40
123  75  80
150 150 120
 70  50  60
```

7

```
 90  60 100
110  75 120
 95 150 150
 57  40  40
 43  25  35
 55 100 110
325 200 400
114  60 120
 83  40  40
 91  35  60
 56  20  40
```

8

Regression without transformation



```
proc reg ;
model photo = obs1 ;
run ;
```

The REG Procedure
Model: MODEL1
Dependent Variable: photo

Analysis of Variance

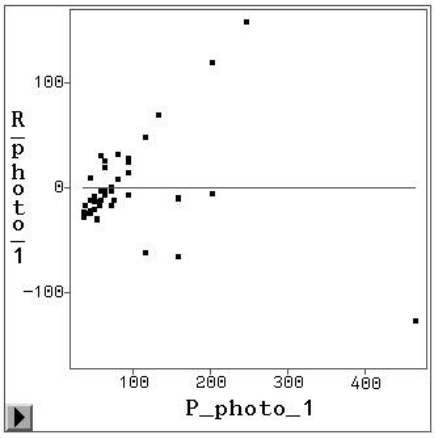
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	254769	254769	129.20	<.0001
Error	43	84790	1971.86476		
Corrected Total	44	339560			

Root MSE	44.40568	R-Square	0.7503
Dependent Mean	89.31111	Adj R-Sq	0.7445
Coeff Var	49.72022		

Parameter Estimates

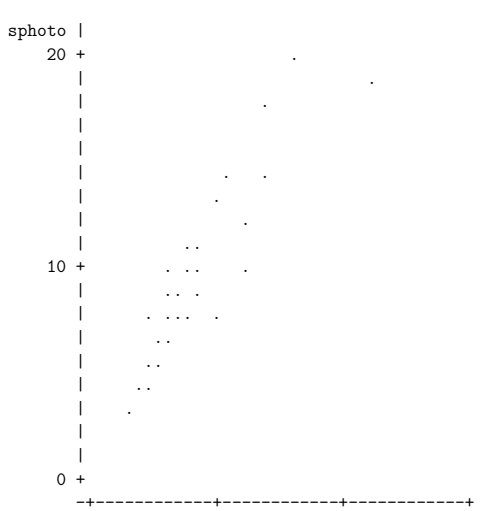
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	26.64957	8.61448	3.09	0.0035
obs1	1	0.88256	0.07764	11.37	<.0001

Residual vs. predicted values



Regression with square root transformations of response variable and predictor

Plot of sphoto*sobs1. Symbol used is '.'.




```
\begin{verbatim}
proc reg ;
model sphoto = sobs1 ;
run ;
```

Dependent Variable: sphoto

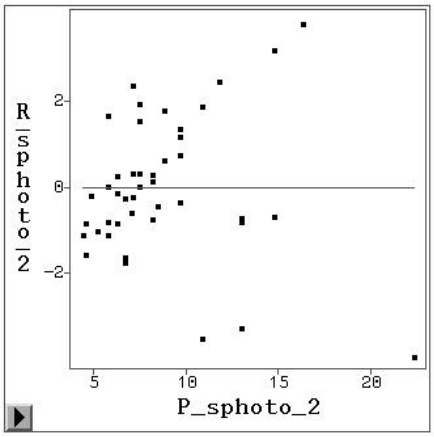
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	572.74519	572.74519	215.15	<.0001
Error	43	114.46762	2.66204		
Corrected Total	44	687.21281			

Root MSE	1.63158	R-Square	0.8334
Dependent Mean	8.60463	Adj R-Sq	0.8296
Coeff Var	18.96159		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.61030	0.53529	3.01	0.0044
sobs1	1	0.93182	0.06353	14.67	<.0001


R_sphoto_2
2
0
-2
5
10
15
20
P_sphoto_2