STOR 455
STATISTICAL METHODS I

Jan Hannig

Model Selection (Chapter 7)

- Model selection methods
  - All subset selection
  - Forward selection, backward elimination, stepwise regression

Criteria for Variable Selection

- Variance based: minimize root-MSE
- SSE based criterion: maximize R^2 =1-SSE(model)/SSTO
- MSE based criterion: maximize adjusted-R^2=1-MSE(model)/MSTO

  accounting for d.f. of the model
- Mallow's C_p: minimizing the bias of sub-model
- $PRESS_p$: minimizing prediction error

Automatic search procedures

- When p large, can't do all subset
- Stepwise type procedures
  - Forward selection (Step up)
  - Backward elimination (Step down)
  - Stepwise: combines forward and backward procedure allowing for removal of variables after adding new variables.
- Many other alternatives, but NONE guarantees the optimal solution
  (NP-hard problem)

Forward Selection

- Start with an intercept
- At each step add the “best” variable (using some criteria – R^2, adj R^2, C_p, partial correlation, SSE, …)
- Compare the p-value for the test whether the just added variable is 0 with some pre-selected value.
  - If smaller – add the variable and repeat the procedure
  - If larger – stop. You have arrived at the final model.
- This is purely exploratory – see the book for comments

Backwards Elimination

- Start with the full model
- At each step delete the "worst" variable (using some criteria – smallest increase in SSE, largest p-value,…)
- Compare the p-value for the test whether the just deleted variable is 0 with some pre-selected value.
  – If larger – delete the variable and repeat the procedure
  – If smaller – stop. You have arrived at the final model.

Stepwise Regression

- Combines the forward and backward algorithm
- Starts at some model (empty or full is usual)
- First eliminates as many parameters as possible using backwards rule (using a minimum allowable p-value= alpha-delete)
- Than attempt to add one variable (using a maximum allowable p-value=alpha-add)
- If a variable is added repeat, if not stop (alpha-add<alpha-delete required for convergence).

Summary

- No method is the best for all model selection problems
- Consider more than one criterion
- "Best model" from automatic search procedures should be used as the starting point
- Apply knowledge of the subject matter to make a final selection – use your head!

Model validation

- Three approaches to checking the validity of the model
  - Collect new data, does it fit the model
  - Compare with theory, other data, simulation
  - Use some of the data for the basic analysis and some for validity check, compare SSE with PRESS, MSE with MSPE

Model does not fit?

- If model does not fit we need to do something needs to be done.
- Several remedial measures are available
  - Consider bigger model (add predictors)
  - Delete outliers
  - Transform the data
  - …

Remedial Measures

- Discard outlier
  - or alternatively use robust procedure such as weighted least squares, Generalized linear models, nonparametric methods, …
- Transformation data to
  - Linearize mean response
  - Stabilize variance/Achieve normality

Nonlinear relationships

- As we know, some nonlinear relationships can be approximated by linear models with added predictors
  - Quadratic $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \xi$
  - Higher order polynomials
  - Other functions $Y=\beta_0 + \beta_1 \sin(X) + \beta_2 \cos(X) + \xi$
- Key – non-linear in predictors but linear in parameters!

Nonlinear (2)

- Sometimes we can transform a nonlinear problem into a linear form
  - Transform X: $Y = \beta_0 + \beta_1 \log(X) + \xi$
  - Transform Y: $\log(Y) = \beta_0 + \beta_1 X + \xi$
  - Transform both: $\log(Y) = \beta_0 + \beta_1 \log(X) + \xi$
- Power/log transformation most common
- Note change in assumption about the error

Non constant error variance

- Log/square root transformations often used to stabilize (make constant) variance
- Alternatively, we can model the way in which the error variance changes (it may be linearly related to X) and use weighted least square

Box-Cox Transformations

- Also called power transformations

  $Y' = Y^{\lambda}$

  or $Y' = (Y^{\lambda} - 1)/\lambda$

  - In the second form, the limit as λ approaches zero is the (natural) log

Important Special Cases

- $\lambda = 1$, $Y' = Y^1$, no transformation
- $\lambda = .5$, $Y' = Y^{1/2}$, square root
- $\lambda = -.5$, $Y' = Y^{-1/2}$, one over square root
- $\lambda = -1$, $Y' = Y^{-1} = 1/Y$, inverse
- $\lambda = 0$, $(Y' = (Y^\lambda - 1)/\lambda)$, log is the limit

Box-Cox Details

- We can estimate λ by including it as a parameter in a non linear model
- $Y^{\lambda} = \beta_0 + \beta_1 X + \xi$
- Choose λ that give the best fit

Box-Cox Solution

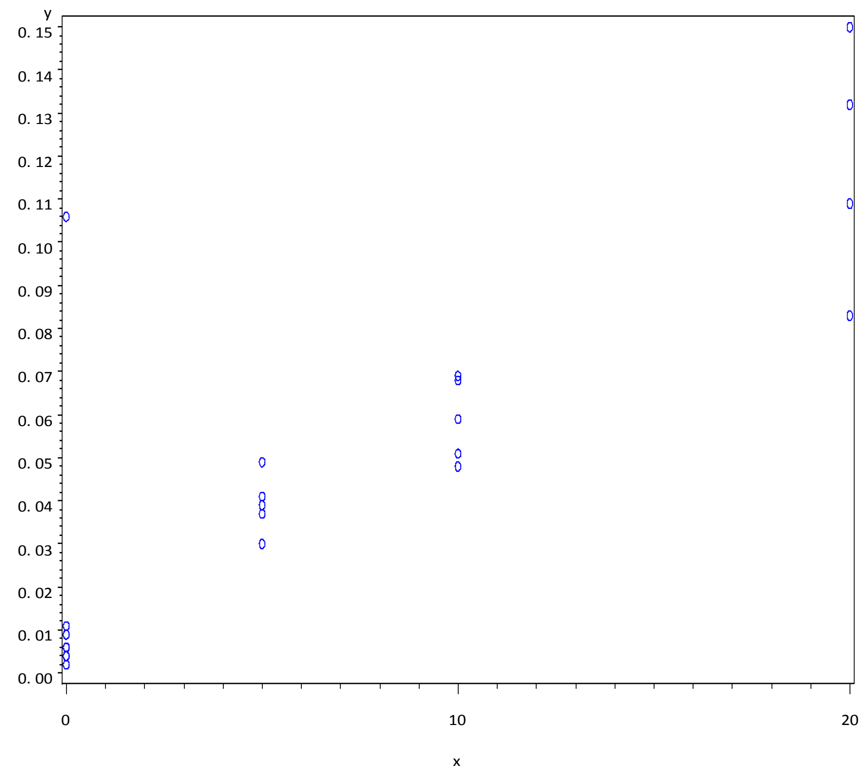
- Standardized transformed Y is
  - $K_1(Y^\lambda - 1)$ if λ neq 0
  - $K_2\log(Y)$ if $\lambda = 0$
- where $K_2 = (\prod Y_i)^{1/n}$, the geometric mean and $K_1 = 1/ (\lambda K_2^{\lambda-1})$
- Run regressions with X as explanatory variable
- Choose λ that minimizes SSE
- SAS code is in macro boxcox.sas or proc transreg

Plutonium Example

- Detecting plutonium 238 using alpha counts
- X: plutonium activity
- Y: observed alpha counts per second
- Relationship depend on measurement device
- Four standard aluminum/plutonium rods tested, each 4 to 10 times

Do it in SAS

```
/* Plutonium example */
data plu;
  infile 'C:\...\CH03TA10.TXT';
  input y x;
run;

symbol1 v=circle h=.8 c=blue i=none;
* scatter plot;
proc gplot data=plu;
  plot y*x;
run;
```



Do it in SAS

* first model with residual/QQ plot;

proc reg data=Plu;

model y = x;

* scatter plot;

plot y*x;

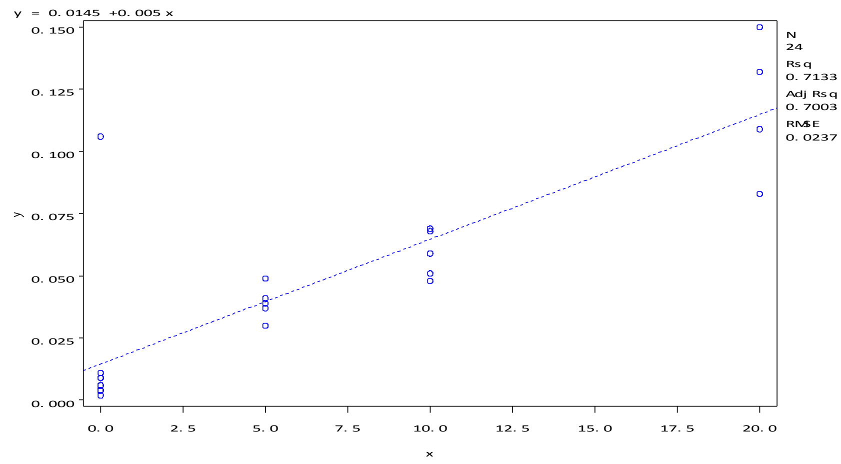
* Plot residuals by predicted values ;

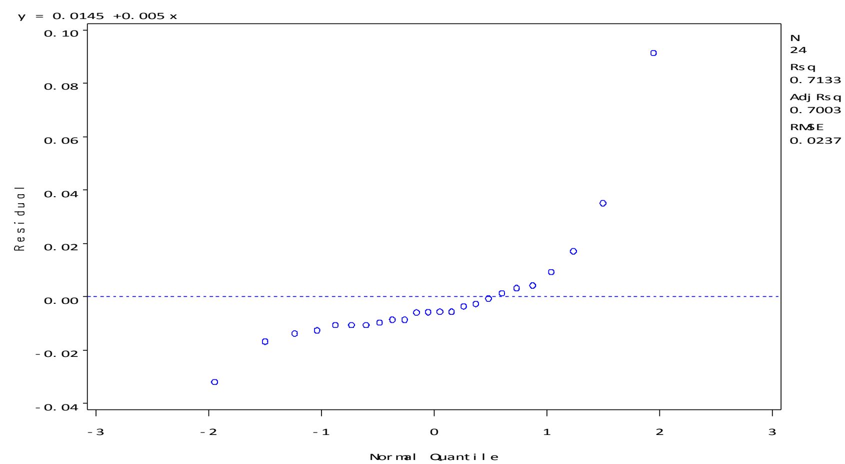
plot (r. student. rstudent.) * p. ;

* Plot Normal quantile plot of residuals ;

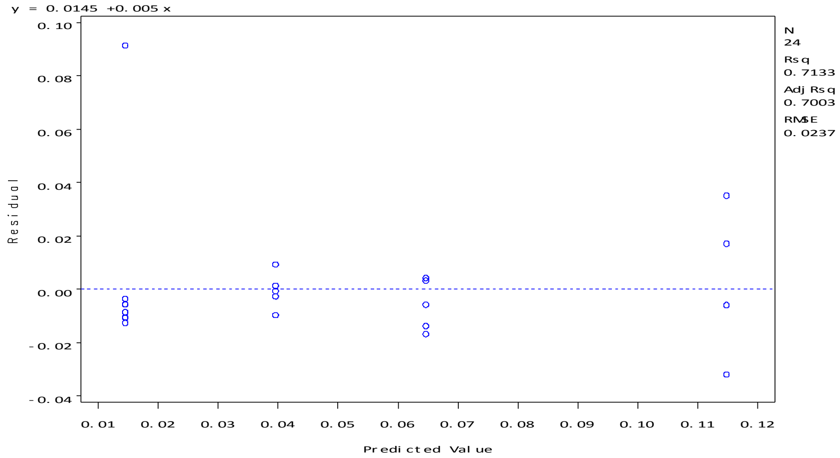
plot r. * nqq. ;

run;

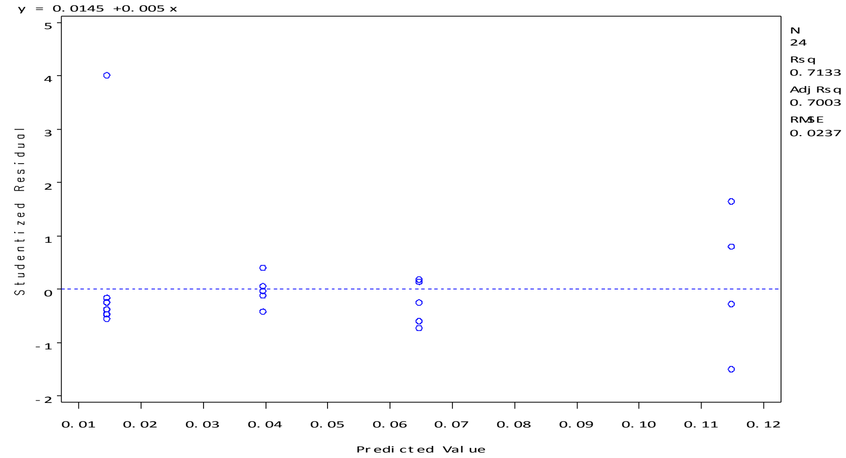




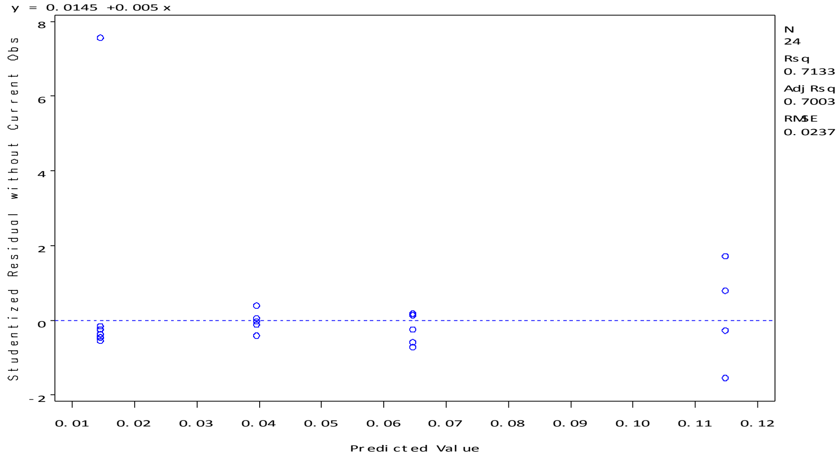
Do it in SAS



y = 0.0145 +0.005 x
Residual
Predicted Value
N
24
Rsq
0.7133
Adj Rsq
0.7003
RMSE
0.0237



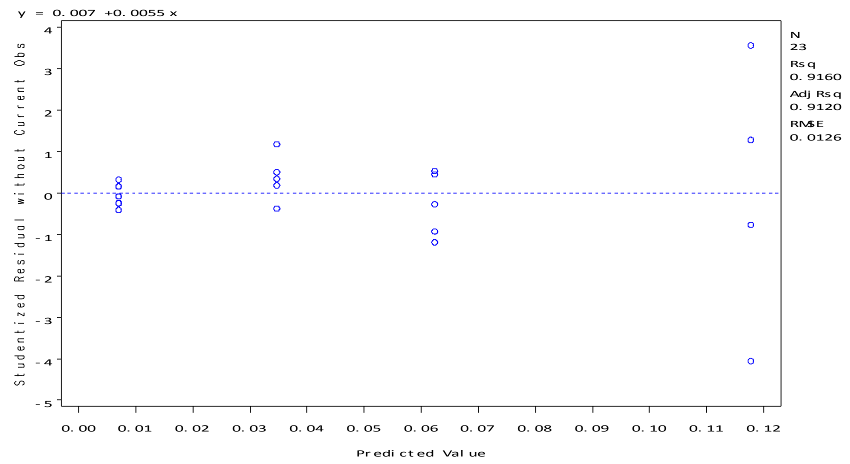
y = 0.0145 +0.005 x
Studentized Residual
Predicted Value
N
24
Rsq
0.7133
Adj Rsq
0.7003
RMSE
0.0237

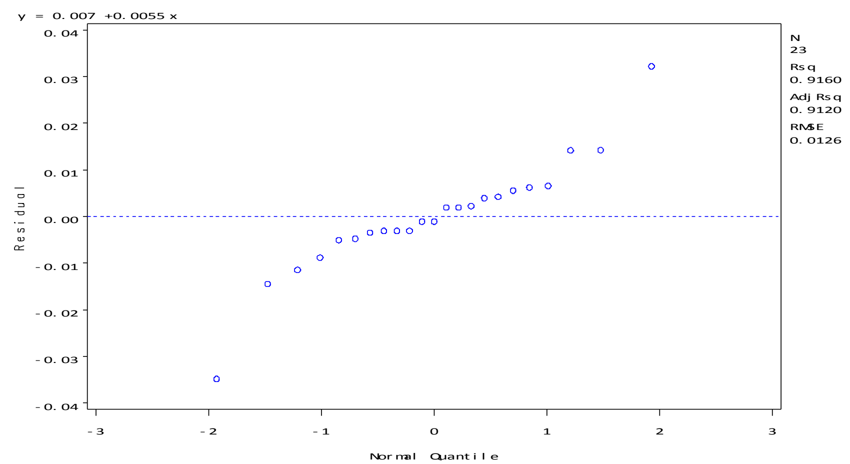


y = 0.0145 +0.005 x
Studentized Residual without Current Obs
Predicted Value
N
24
Rsq
0.7133
Adj Rsq
0.7003
RMSE
0.0237

Do it in SAS

```
* remove outlier;
proc reg data=Plu;
* following command removes
   the outlier;
where NOT ( x EQ 0 AND y GE
   0.09 );
  model y = x;
* Plot residuals by predicted
   values ;
  plot rstudent. * p. ;
* Plot Normal quantile plot of
   residuals ;
  plot r. * nqq. ;
run;
```





Do it in SAS

```
proc transreg data=plu;
model boxcox(y)=identity(x);
run;
```

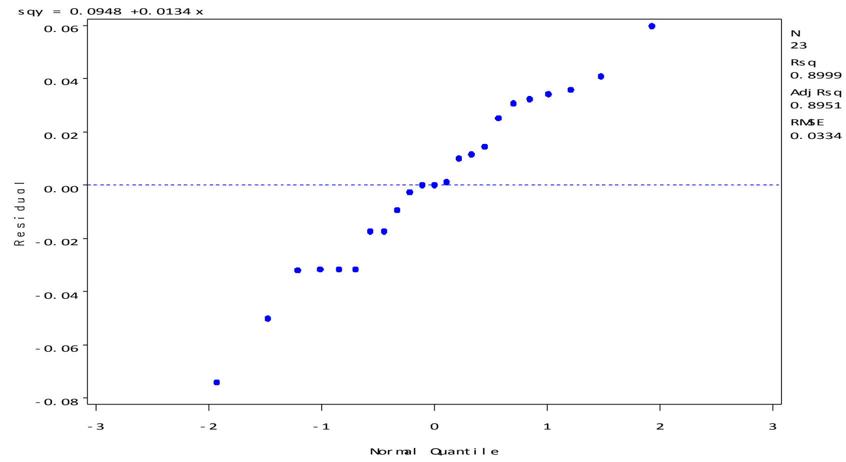
The TRANSREG Procedure
Box-Cox Transformation Information for y

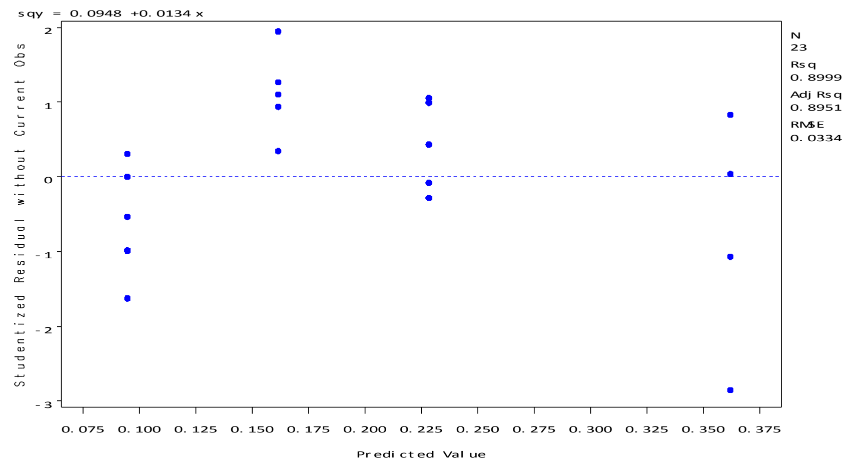
Lambda	R-Square	Log Like
-3.00	0.09	-22.8128
-2.75	0.10	-10.5162
-2.50	0.12	1.5873
-2.25	0.14	13.4530
-2.00	0.17	25.0222
-1.75	0.20	36.2187
-1.50	0.25	46.9457
-1.25	0.31	57.0871
-1.00	0.38	66.5173
-0.75	0.47	75.1246
-0.50	0.57	82.8497
-0.25	0.67	89.7200
0.00	0.76	95.8371
0.25	0.84	101.2188
0.50 +	0.90	105.2423 *
0.75	0.92	105.7181 <
1.00	0.92	100.6558
1.25	0.89	91.9689
1.50	0.84	82.2649
1.75	0.79	72.5284
2.00	0.74	62.9473
2.25	0.69	53.5043
2.50	0.65	44.1498
2.75	0.61	34.8407
3.00	0.58	25.5459

< - Best Lambda
* - 95% Confidence Interval
+ - Convenient Lambda

Do it in SAS

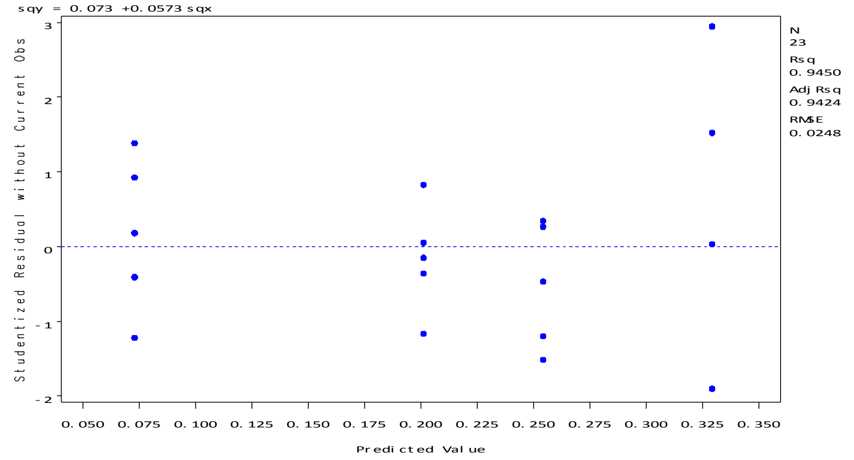
```
*Data transformation;
data plu1;
    set plu;
     where NOT ( x EQ 0 AND y GE
    0.09 );
     sqy=sqrt(y);
     sqx=sqrt(x);
     run;
proc print data=plu1;
run;

* Transformed y;
proc reg data=Plu1;
  model sqy = x;
  plot sqy*x rstudent.*p. r.*nqq.;
run;
```

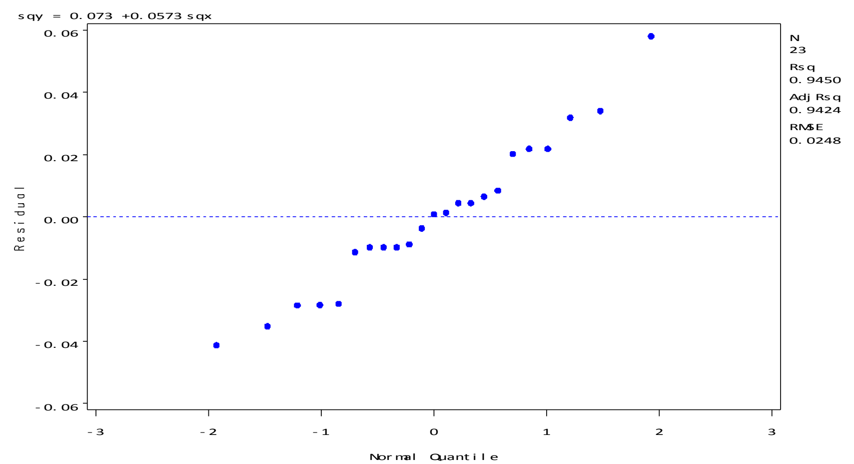




Do it in SAS

```
* transform x;
proc reg data=Plu1;
  model sqy = sqx;
  plot sqy*sqx rstudent.*p. r.*nqq.;
run;
```





Do it in SAS

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.22142	0.22142	360.92	<.0001
Error	21	0.01288	0.00061348		
Corrected Total	22	0.23430			

Root MSE	0.02477	R-Square	0.9450
Dependent Mean	0.18483	Adj R-Sq	0.9424
Coeff Var	13.40098		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.07301	0.00783	9.32	<.0001
sqx	1	0.05731	0.00302	19.00	<.0001

Do it in SAS

- The final model we fit is $\text{sqrt}(y)=0.07301+0.05731*\text{sqrt}(x)+\xi$
- Be careful when interpreting – the predicted values are sqrt(y), to get a predicted value for y need to square!