Topic 15 - Weighted Least Squares

STAT 525 - Fall 2013

Transformation Approach

- Suppose $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}) = \mathbf{W}^{-1}$
- Have linear model but potentially correlated errors and unequal variances
- Consider a transformation based on $\mathbf{W}$

$$\begin{aligned} \mathbf{W}^{1/2}\mathbf{Y} &= \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{1/2}\boldsymbol{\varepsilon} \\ &\downarrow \\ \mathbf{Y}_\mathbf{w} &= \mathbf{X}_\mathbf{w}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{w} \end{aligned}$$

- Can show

$$\mathrm{E}(\boldsymbol{\varepsilon}_\mathrm{w}) = 0 \text{ and } \boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}_\mathrm{w}) = \mathbf{I}$$

- Weighted least squares special case of *generalized least squares* where only variances may differ ($\mathbf{W}$ is a diagonal matrix)

Maximum Likelihood

- Consider

$$Y_i \sim \mathrm{N}(\mathbf{X}_i\boldsymbol{\beta}, \sigma_i^2) \qquad (\sigma_i\text{'s known})$$
$$\downarrow$$
$$f_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2\right\}$$

- Likelihood function $L = f_1 \times f_2 \times \cdots \times f_n$
- Find $\boldsymbol{\beta}$ which maximizes L
- Similar to minimizing

$$Q_w = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2$$

Weighted Least Squares

- Expressed in matrix form

$$Q_w = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

where

$$\mathbf{W} = \begin{bmatrix} 1/\sigma_1^2 & 0 & \cdots & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1/\sigma_{n-1}^2 & 0 \\ 0 & \cdots & \cdots & 0 & 1/\sigma_n^2 \end{bmatrix}$$

- Normal equations: $(\mathbf{X}'\mathbf{W}\mathbf{X})\mathbf{b}_\mathbf{w} = \mathbf{X}'\mathbf{W}\mathbf{Y}$
- Solution: $\mathbf{b}_\mathbf{w} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$

Weighted Least Squares

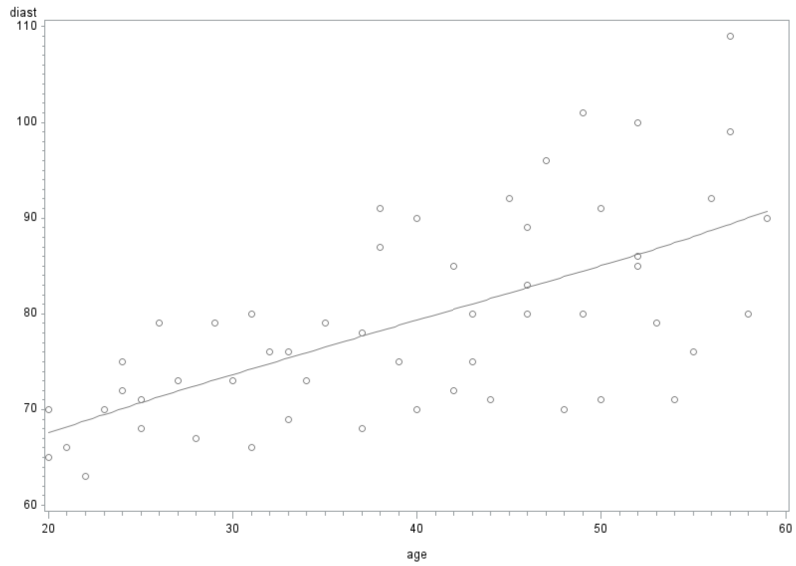
- Can be implemented in SAS using the weight option
- Must determine optimal weights
- Optimal weights $\propto$ 1/variance
- Methods to determine weights
  - Find relationship between the absolute residual and another variable and use this as a model for the standard deviation
  - Instead of the absolute residual, use the squared residual and find function for the variance
  - Use grouped data or approximately grouped data to estimate the variance

Example Page 427

- Interested in the relationship between diastolic blood pressure and age
- Have measurements on 54 adult women
- Age range is 20 to 60 years old
- Issue:
  - Variability increases as the mean increases
  - Appears to be nice linear relationship
  - Don't want to transform X or Y and lose this

SAS Commands I

```
data a1;
  infile 'U:\.www\datasets525\Ch11ta01.txt';
  input age diast;

symbol1 v=circle i=sm70;
proc gplot data=a1;
  plot diast*age/frame;

proc reg data=a1;
  model diast=age;
  output out=a2 r=resid;

proc gplot data=a2;
  plot resid*age;
run;
```

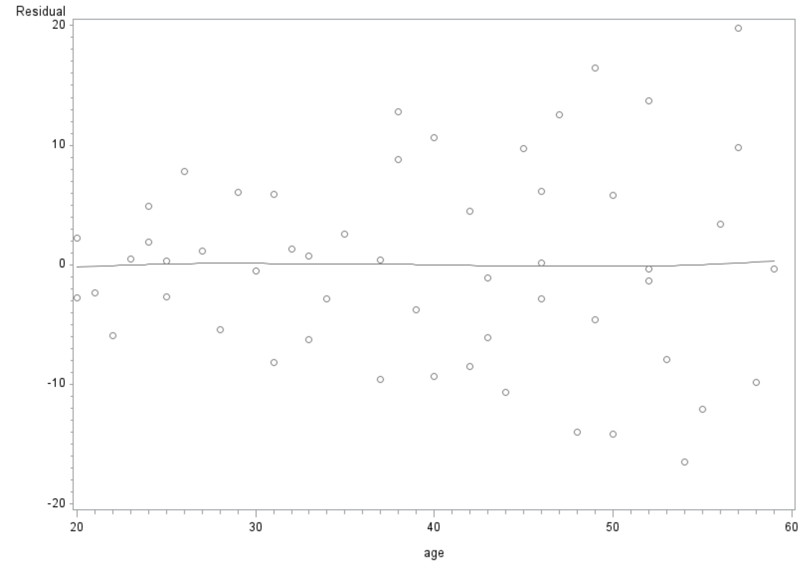
Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2374.96833	2374.96833	35.79	<.0001
Error	52	3450.36501	66.35317		
Corrected Total	53	5825.33333			

Root MSE	8.14575	R-Square	0.4077
Dependent Mean	79.11111	Adj R-Sq	0.3963
Coeff Var	10.29659		

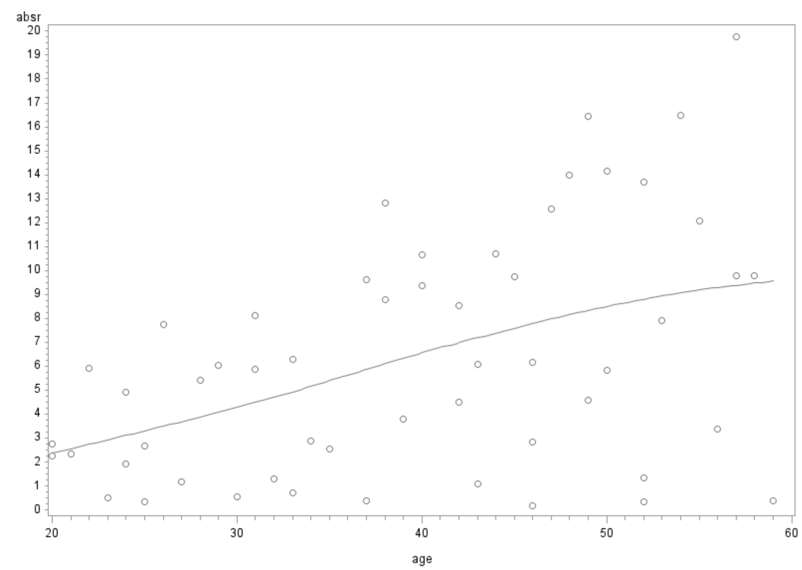
Parameter Estimates

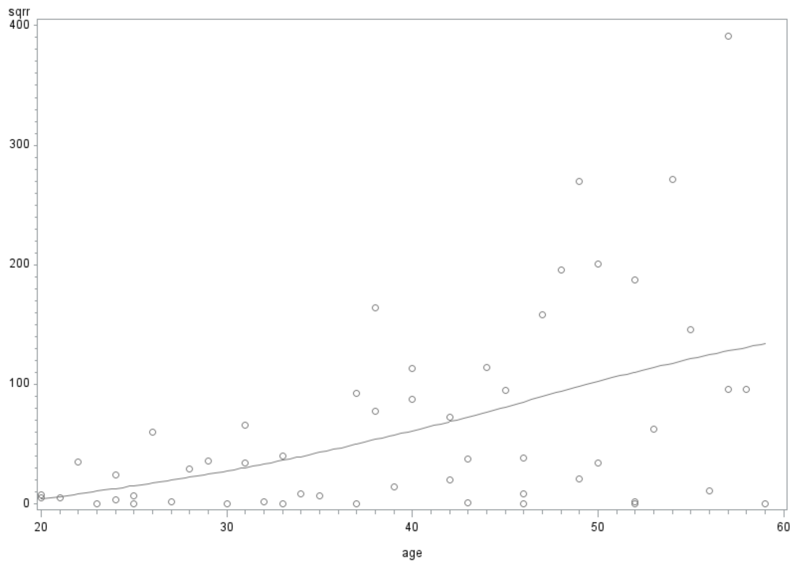
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	56.15693	3.99367	14.06	<.0001
age	1	0.58003	0.09695	5.98	<.0001



SAS Commands II

```
data a2; set a2;
  absr=abs(resid);  sqrr=resid*resid;

proc gplot data=a2;
  plot (resid absr sqrr)*age;

proc reg data=a2;
  model absr=age;
  output out=a3 p=shat;

data a3; set a3;
  wt=1/(shat*shat);

proc reg data=a3;
   model diast=age / clb;
   weight wt;
run;
```



Construction of Weights

- Will assume abs(res) is linearly related to age
- Fit least squares model to predict SD_i

```
proc reg data=a2;
  model absr=age;
  output out=a3 p=shat;
```

- Weight is $= 1/\mathrm{SD}_i^2$

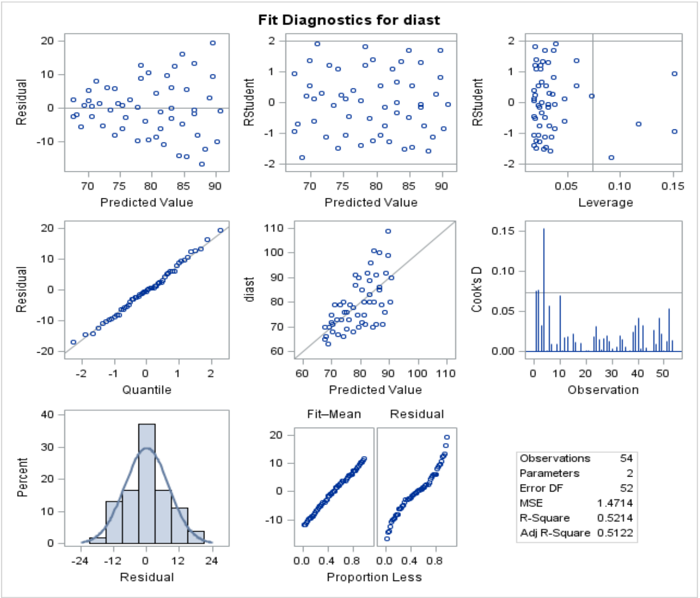
```
data a3; set a3;
  wt=1/(shat*shat);

proc reg data=a3;
   model diast=age / clb;
   weight wt;
```

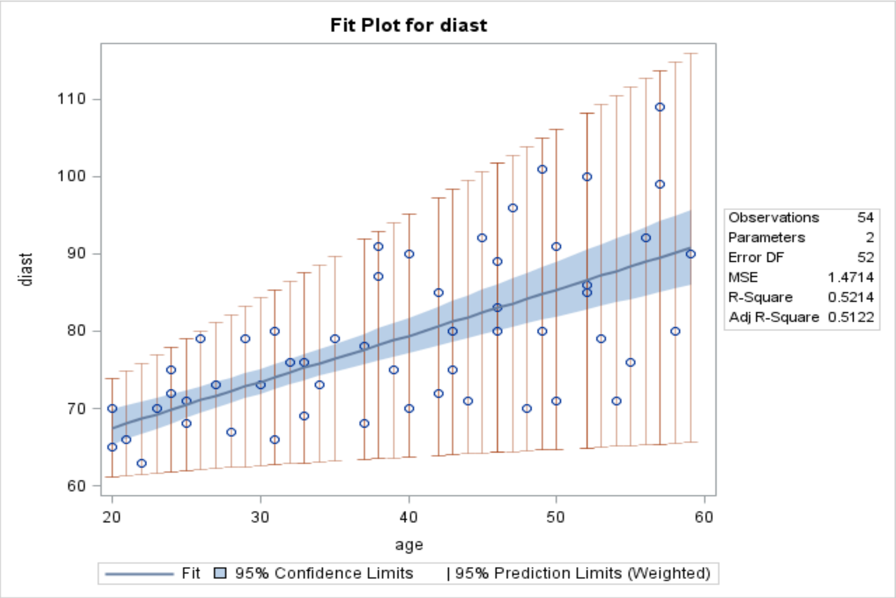

Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	83.34082	83.34082	56.64	<.0001
Error	52	76.51351	1.47141		
Corrected Total	53	159.85432			

Root MSE	1.21302	R-Square	0.5214
Dependent Mean	73.55134	Adj R-Sq	0.5122
Coeff Var	1.64921		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	55.56577	2.52092	22.04	<.0001	50.50718	60.62436
age	1	0.59634	0.07924	7.53	<.0001	0.43734	0.75534





Summary

- Not much change in the parameter estimates
- Slight reduction in the parameters' standard errors
- Be wary:
  - R^2 does not have usual meaning
  - Interpretation of residual plots
  - Construction of confidence and prediction intervals
- Since weights based on residuals, can take iterative approach and re-estimate weights based on new residuals and repeat.

Iterative Approach

- Usually converges quite quickly
- For this example:

Iteration	b_0	SE(b_0)	b_1	SE(b_1)
1	55.56577	2.52092	0.59634	0.07924
2	55.56264	2.51851	0.59643	0.07922
3	55.56261	2.51849	0.59643	0.07922
4	55.56261	2.51849	0.59643	0.07922

- Usually changes within level of accuracy so run only once

Mixed model approach(?)

- We're assuming the variance/covariance matrix is a diagonal matrix whose values along the main diagonal (the variances) are either a
  - Linear function of age
  - Quadratic function of age
- This relationship along with the estimation of parameters can be done simultaneously using the lin(q) covariance structure
- Need to create appropriate diagonal matrices and specify reasonable starting values...sample code provided

Output - Linear Relationship

Cov Parm	Subject	Estimate
LIN(1)	Intercept	-51.0389
LIN(2)	Intercept	2.8707

Fit Statistics

-2 Res Log Likelihood	365.0
AIC (smaller is better)	369.0
AICC (smaller is better)	369.3
BIC (smaller is better)	372.9

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	tValue	Pr > |t|
Intercept	55.3831	2.5720	17.8	21.53	<.0001
age	0.5996	0.07855	39.2	7.63	<.0001

Output - Quadratic Relationship

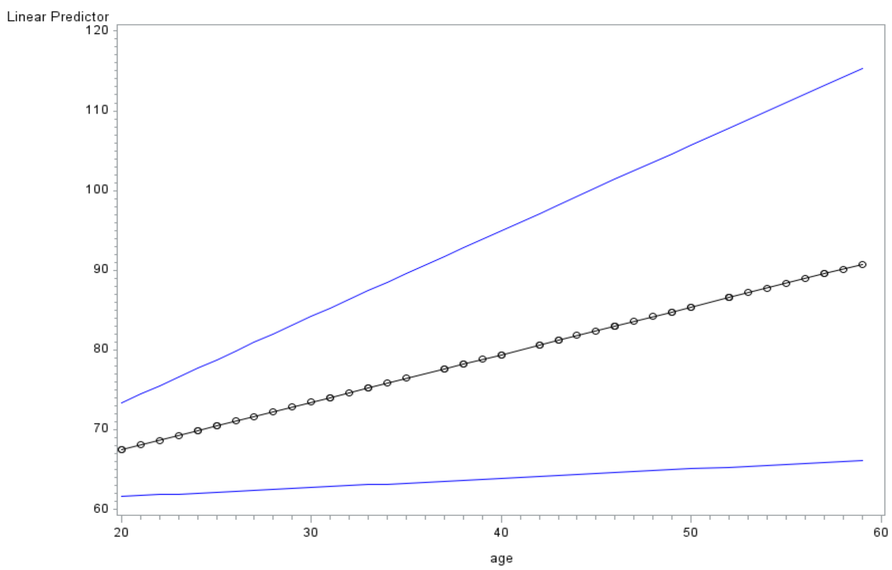
Cov Parm	Subject	Estimate
LIN(1)	Intercept	9.4647
LIN(2)	Intercept	-1.2853
LIN(3)	Intercept	0.06331

Fit Statistics

-2 Res Log Likelihood	364.4
AIC (smaller is better)	370.4
AICC (smaller is better)	370.9
BIC (smaller is better)	376.2

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	tValue	Pr > |t|
Intercept	55.6087	2.8634	14.8	19.42	<.0001
age	0.5954	0.08666	25.4	6.87	<.0001



Background Reading

- KNNL Section 11.1
- knnl427.sas
- KNNL Sections 11.2-11.6