Chapter 11: Weighted Least Squares

Department of Statistics, University of South Carolina

Stat 704: Data Analysis I

11.1: Weighted least squares

- Chapters 3 and 6 discuss transformations of $x_1, \ldots, x_k$ and/or Y.
- This is "quick and dirty" but may not solve the problem.
- Or can create an uninterpretable mess (book: "inappropriate").
- More advanced remedy: *weighted least squares* regression.
- Model is as before

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_k x_{ik} + \epsilon_i,$$

but now

$$\epsilon_i \stackrel{ind.}{\sim} N(0, \sigma_i^2).$$

Note the subscript on σ_i...

- Here $\text{var}(Y_i) = \sigma_i^2$. Give observations with higher variance *less weight* in the regression fitting.
- Say $\sigma_1, \ldots, \sigma_n$ are known. Let $w_i = 1/\sigma_i^2$ and define the weight matrix

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} \sigma_1^{-2} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^{-2} \end{bmatrix}.$$

- Maximizing the likelihood (pp. 422-423) gives the estimates for $\boldsymbol{\beta}$:

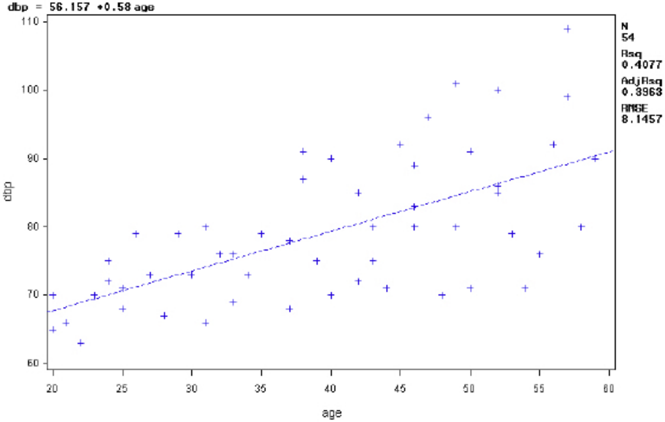
$$\mathbf{b}_w = (\mathbf{XWX}')^{-1}\mathbf{X}'\mathbf{WY}.$$

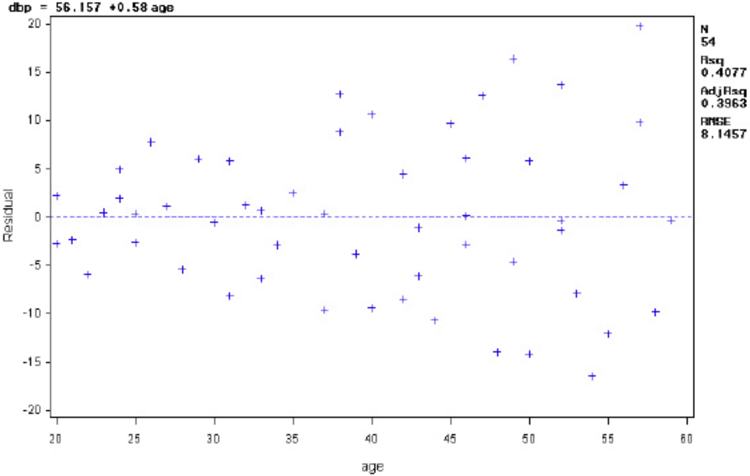
- However, $\sigma_1, \ldots, \sigma_n$ are almost always unknown.
- If the mean function is appropriate, then $E(e_i^2) = \sigma_i^2(1 - h_{ii})$ where e_i is obtained from ordinary least squares, so e_i^2 estimates σ_i^2 and $|e_i|$ estimates σ_i (pp. 424-425) as $h_{ii} \to 0$ as $n \to \infty$.
- Look at plots of $|e_i|$ from a normal fit against predictors in the model and the fitted values $\hat{Y}_i$ to see how σ_i changes with predictors or fitted values.
- For example, if $|e_i|$ increases linearly with $\hat{Y}_i = \mathbf{x}_i'\mathbf{b}$, then we'll fit $|e_i| = \alpha_0 + \alpha_1 x_{i1} + \cdots + \alpha_k x_{ik} + \delta_i$ and obtain the fitted values $\widehat{|e_i|}$.
- If e_i^2 increases linearly with only x_{i4}, then we'll fit $e_i^2 = \alpha_0 + \alpha_4 x_{i4} + \delta_i$ and obtain the fitted values $\widehat{e_i^2}$.

1. Regress Y against predictor variable(s) as usual (OLS) & obtain $e_1, \ldots, e_n$ & $\hat{Y}_1, \ldots, \hat{Y}_n$.
2. Regress $|e_i|$ against predictors $x_1, \ldots, x_k$ or fitted values $\hat{Y}_i$.
3. Let $\hat{s}_i$ be the fitted values for the regression in 2.
4. Define $w_i = 1/\hat{s}_i^2$ for $i = 1, \ldots, n$.
5. Use $\mathbf{b}_w = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$ as estimated coefficients – automatic in SAS. SAS will also use the correct $\text{cov}(\mathbf{b}_w) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$ (p. 423). This is developed formally in linear models.

SAS code: initial fit

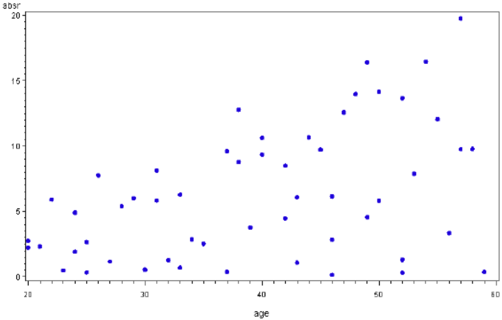
```
* SAS example for Weighted Least Squares ;
* Blood pressure data in Table 11.1      ;
data bloodp; input age dbp @@; datalines;
  27   73  21   66  22   63  24   75  25   71  23   70  20   65
  20   70  29   79  24   72  25   68  28   67  26   79  38   91
  32   76  33   69  31   66  34   73  37   78  38   87  33   76
  35   79  30   73  31   80  37   68  39   75  46   89  49  101
  40   70  42   72  43   80  46   83  43   75  44   71  46   80
  47   96  45   92  49   80  48   70  40   90  42   85  55   76
  54   71  57   99  52   86  53   79  56   92  52   85  50   71
  59   90  50   91  52  100  58   80  57  109
; run;

* Regress the response, dbp, against the predictor, age ;
* The plots show definite nonconstant error variance    ;
proc reg data=bloodp;
 model dbp=age;
 output out=temp r=residual;
 plot dbp*age r.*age;
run;
```


dbp = 56.157 +0.58 age
N
54
Rsq
0.4077
AdjRsq
0.3963
RMSE
8.1457
dbp
age
60
70
80
90
100
110
20
25
30
35
40
45
50
55
60


dbp = 56.157 +0.58 age
Residual
age
N
54
Rsq
0.4077
AdjRsq
0.3963
RMSE
8.1457

SAS code: determining w_i

```
* Plot of absolute residuals against age shows that
  absolute residuals increase approximately linearly;
data temp; set temp; absr = abs(residual); run;
symbol1 v=dot h=0.8;
axis1 order=(0 to 20 by 5);
proc gplot data=temp; PLOT absr*age/ vaxis=axis1; run;
```



SAS code: WLS fit

```
* Regress absolute residuals against the age           ;
* This second regression is done on the data set temp  ;
proc reg data=temp;
  model absr=age;
  output out=temp1 p=s_hat ;
run;

* Define weights using the fitted values from this second regression ;
data temp1; set temp1; w=1/(s_hat**2); run;

* Using the WEIGHT option in PROC REG to get the WLS estimates ;
* This last regression is done on the data set temp1           ;
proc reg data=temp1;
  weight w;
  model dbp=age / clb;
  output out=temp2 r=residual;
  plot dbp*age r.*age;
run;
```

SAS output: WLS fit

The REG Procedure
Dependent Variable: dbp

Weight: w

Analysis of Variance

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		

Parameter Estimates

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

- se(b_1) reduced from 0.097 (OLS) to 0.079 (WLS). WLS verified via bootstrap on pp. 462–463 (just FYI).
- R^2 no longer interpreted the same way in terms of amount of total variability explained by model.
- In WLS, standard inferences about coefficients may not be valid for small sample sizes when weights are estimated from the data.
- If MSE of the WLS regression is near 1, then our estimation of the "error standard deviation" function is okay. Here it's 1.21.

Fitting the model directly...

- A drawback of this approach is that the weights $w_i = 1/\hat{s}_i^2$ have associated variability that is not reflected in $\text{cov}(\mathbf{b}_w)$.
- It is possible to fit the implied model

$$Y_i = \beta_0 + \beta_1 a_i + \epsilon_i, \quad \epsilon_i \sim N(0, \tau_0 + \tau_1 a_i),$$

  directly in SAS. One option is to have SAS maximize the associated likelihood in PROC NLMIXED.
- Note that a similar, and possibly more appropriate, model

$$Y_i = \beta_0 + \beta_1 a_i + \epsilon_i, \quad \epsilon_i \sim N(0, e^{\tau_0 + \tau_1 a_i}),$$

  was used for the Breusch-Pagan test $H_0 : \tau_1 = 0$ described in Sections 3.6 and 6.8. This model can also be fit easily in PROC NLMIXED.
- However, things like F-tests go out the window and everything relies on asymptotics (which for large enough samples work fine).

SAS code: fitting model directly

```
* Model fit directly using PROC NLMIXED                    ;
* Starting values obtained from regressions 1 and 2 ;
proc nlmixed data=bloodp;
 parms beta0=50 beta1=0.5 tau0=-1 tau1=0.2;
 mu=beta0+beta1*age; sigma=tau0+tau1*age;
 model dbp ~ normal(mu,sigma*sigma);
run;
```

With abridged output

The NLMIXED Procedure

Fit Statistics

-2 Log Likelihood	362.5
AIC (smaller is better)	370.5
BIC (smaller is better)	378.5

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	55.5317	2.4689	54	22.49	<.0001	0.05	50.5819	60.4815	3.678E-6
beta1	0.5973	0.07811	54	7.65	<.0001	0.05	0.4407	0.7539	0.000108
tau0	-2.0367	1.7585	54	-1.16	0.2519	0.05	-5.5622	1.4889	4.053E-6
tau1	0.2414	0.05557	54	4.34	<.0001	0.05	0.1300	0.3528	0.000067