Variance models and generalized least squares

Peter Hoff

STAT 423

Applied Regression and Analysis of Variance

University of Washington

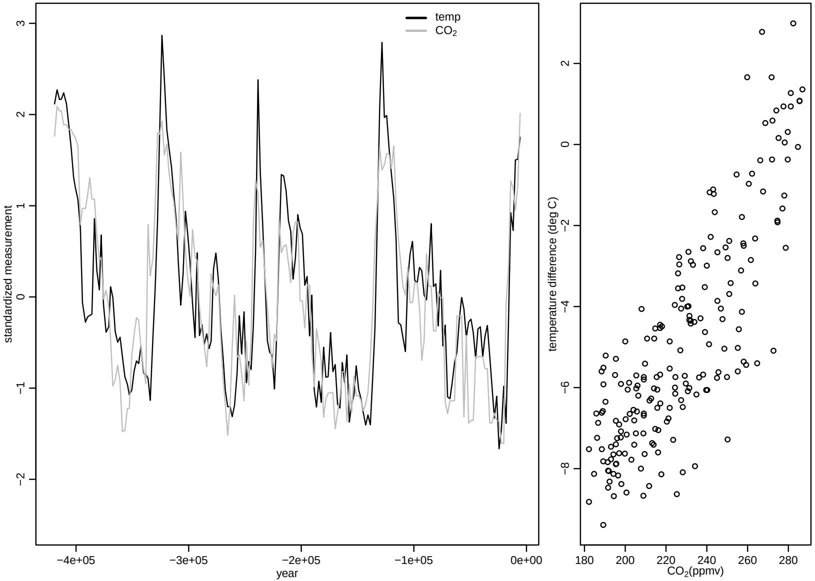
Vostok ice core data

Ice core data from East Antarctica:

`year` year at which ice was frozen (-5000 to -420000 years)

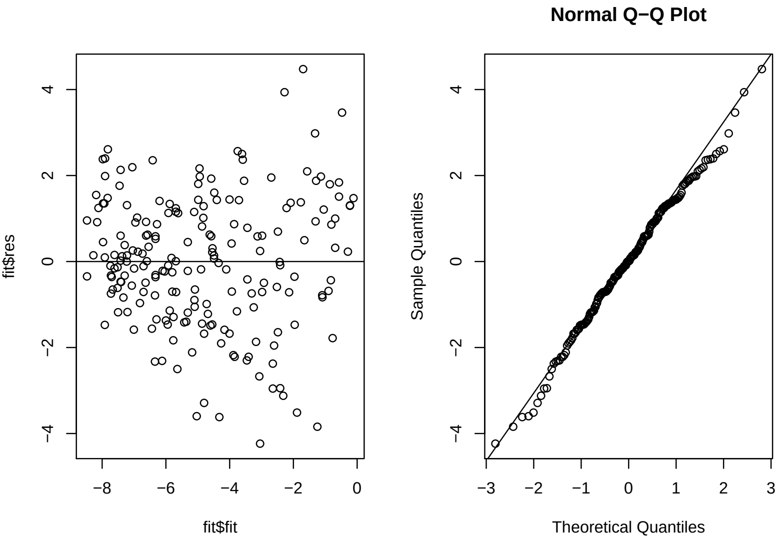
`co2` CO_2 concentration from ice sample (ppm)

`tmp` temperature estimate (difference from current), 1000 years after `co2`


temp
CO2
standardized measurement
year
−4e+05
−3e+05
−2e+05
−1e+05
0e+00
temperature difference (deg C)
CO2(ppmv)
180
200
220
240
260
280

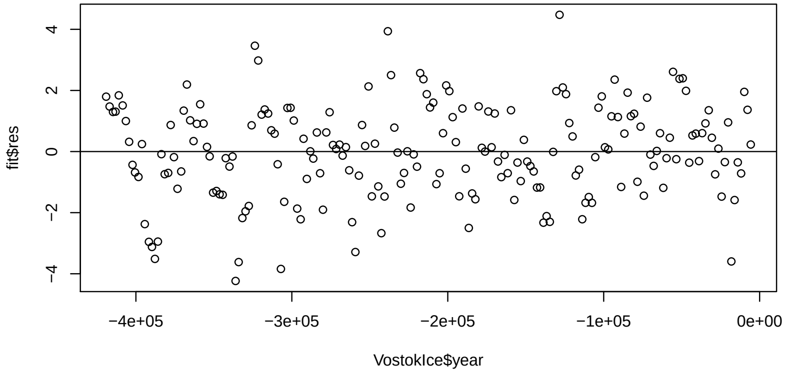
$$\mathsf{tmp} = \beta_0 + \beta_1 \times \mathsf{co}_2 + \epsilon$$

```r
fit<-lm( tmp ~ co2 , data=VostokIce )

summary(fit)$coef

##                 Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) -23.02414043 0.879542928 -26.17739 3.272183e-66
## co2           0.07985291 0.003833728  20.82905 8.766508e-52
```

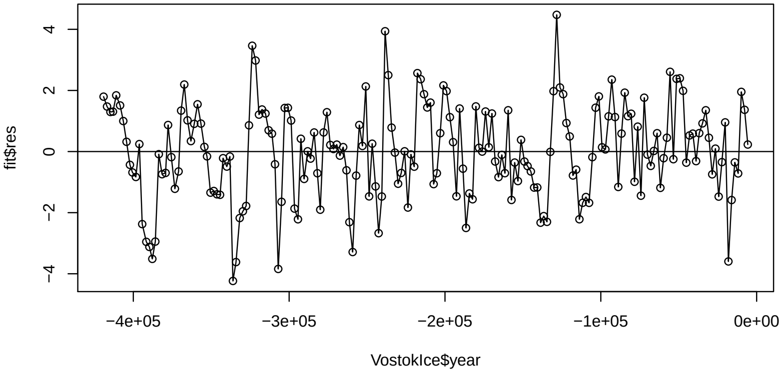


Normal Q−Q Plot
fit$res
fit$fit
Sample Quantiles
Theoretical Quantiles

```r
plot(fit$res~VostokIce$year)

abline(h=0)
```



```r
plot(fit$res~VostokIce$year)
lines(fit$res~VostokIce$year)
abline(h=0)
```

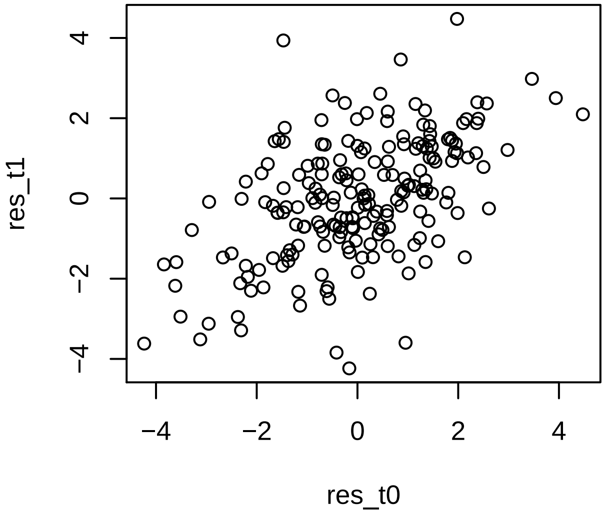


```r
res_t1<-fit$res[-1]
res_t0<-fit$res[-length(fit$res)]

cbind( res_t0 , res_t1 )[1:4,]

##     res_t0   res_t1
## 1 1.796974 1.474342
## 2 1.474342 1.296136
## 3 1.296136 1.306136
## 4 1.306136 1.839503
```

```r
res_t1<-fit$res[-1]
res_t0<-fit$res[-length(fit$res)]

cbind( res_t0 , res_t1 )[1:4,]

##     res_t0   res_t1
## 1 1.796974 1.474342
## 2 1.474342 1.296136
## 3 1.296136 1.306136
## 4 1.306136 1.839503
```

```r
cor( cbind( res_t0 , res_t1 ) )

##           res_t0    res_t1
## res_t0 1.0000000 0.5183877
## res_t1 0.5183877 1.0000000
```

```
plot( res_t0, res_t1 )
```



Misspecified variance/covariance

Estimation

A1: $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$

Misspecified variance/covariance

Estimation

A1: $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$

R1: $\mathrm{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

Misspecified variance/covariance

Estimation

A1: $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$

R1: $\mathrm{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

Inference

A2: $\mathrm{Var}[\mathbf{y}] = \sigma^2\mathbf{I}$

Misspecified variance/covariance

Estimation

A1: $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$

R1: $\mathrm{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

Inference

A2: $\mathrm{Var}[\mathbf{y}] = \sigma^2\mathbf{I}$

R2:

- standard errors: $\widehat{\mathrm{Var}[\hat{\boldsymbol{\beta}}]} = \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}$
- hypothesis tests
- confidence intervals

Misspecified variance/covariance

Estimation

A1: $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$

R1: $\mathrm{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

Inference

A2: $\mathrm{Var}[\mathbf{y}] = \sigma^2\mathbf{I}$

R2:

- standard errors: $\widehat{\mathrm{Var}[\hat{\boldsymbol{\beta}}]} = \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}$
- hypothesis tests
- confidence intervals

If we assume $\mathrm{Var}[\mathbf{y}] = \sigma^2\mathbf{I}$ and it is not, our SEs, CIs and HTs could be misleading.

Alternative covariance model

For the `VostokIce` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

where $\rho \in (-1, 1)$ is an unknown quantity, to be estimated.

Alternative covariance model

For the `VostokIce` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

where $\rho \in (-1, 1)$ is an unknown quantity, to be estimated.

This covariance model is said to have *first-order autoregressive structure*.

Alternative covariance model

For the `VostokIce` data, it might be more reasonable to assume

$$\mathrm{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

where $\rho \in (-1, 1)$ is an unknown quantity, to be estimated.

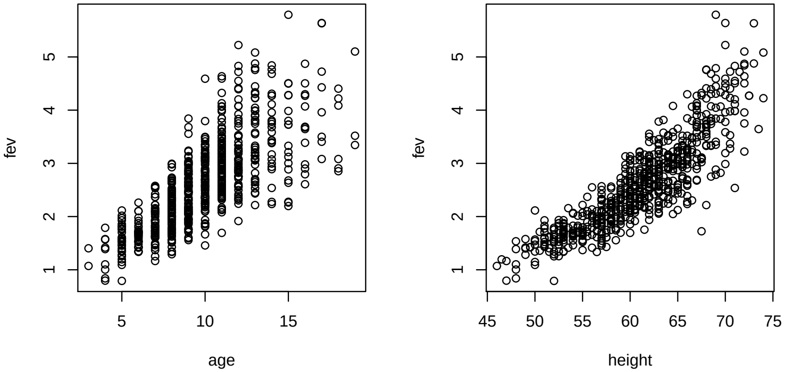
This covariance model is said to have *first-order autoregressive structure*.

Q:

- What are the SEs of the OLS estimates for autocorrelated data?
- Are the OLS estimates appropriate if we have autocorrelated data?

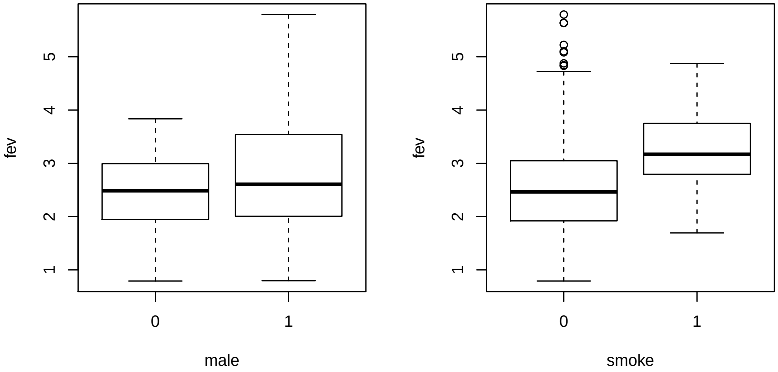
FEV data

Data on lung capacity by age, height, sex and smoking status

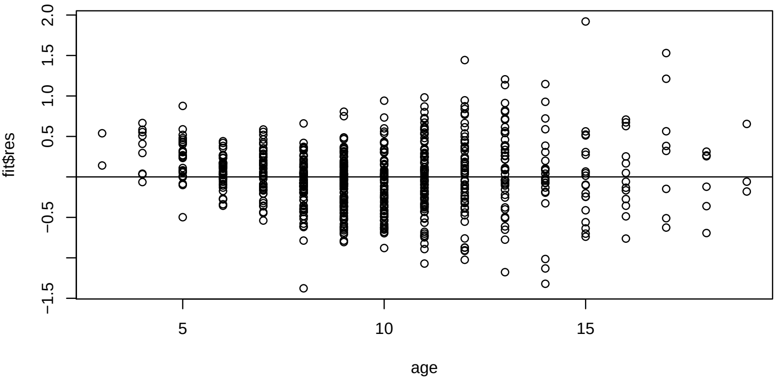


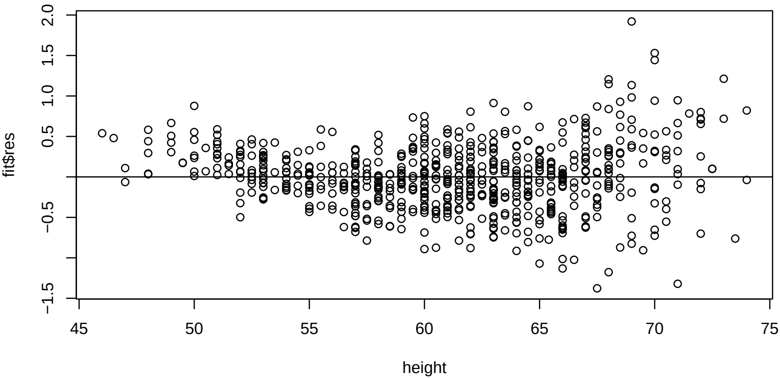
FEV data

Data on lung capacity by age, height, sex and smoking status



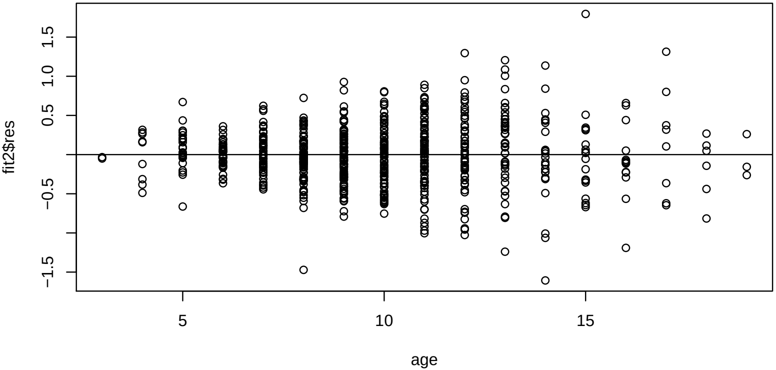
```
fit<-lm( fev ~ age + height + male + smoke, data=fev)

summary(fit)

##
## Call:
## lm(formula = fev ~ age + height + male + smoke, data = fev)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37656 -0.25033  0.00894  0.25588  1.92047
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.456974   0.222839 -20.001  < 2e-16 ***
## age          0.065509   0.009489   6.904 1.21e-11 ***
## height       0.104199   0.004758  21.901  < 2e-16 ***
## male         0.157103   0.033207   4.731 2.74e-06 ***
## smoke       -0.087246   0.059254  -1.472    0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4122 on 649 degrees of freedom
## Multiple R-squared:  0.7754, Adjusted R-squared:  0.774
## F-statistic:   560 on 4 and 649 DF,  p-value: < 2.2e-16
```


2.0
1.5
1.0
0.5
−0.5
−1.5
fit$res
5
10
15
age

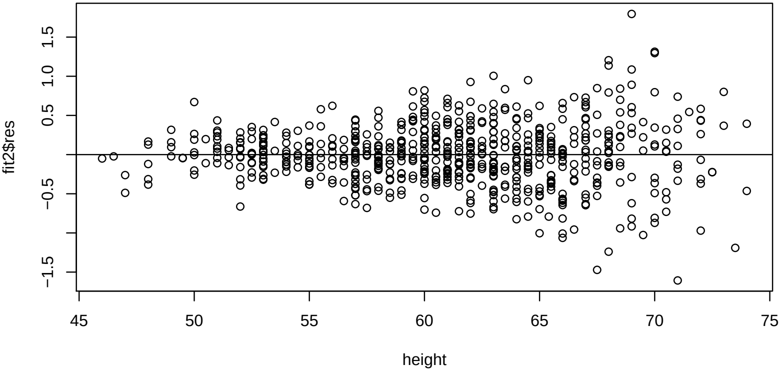


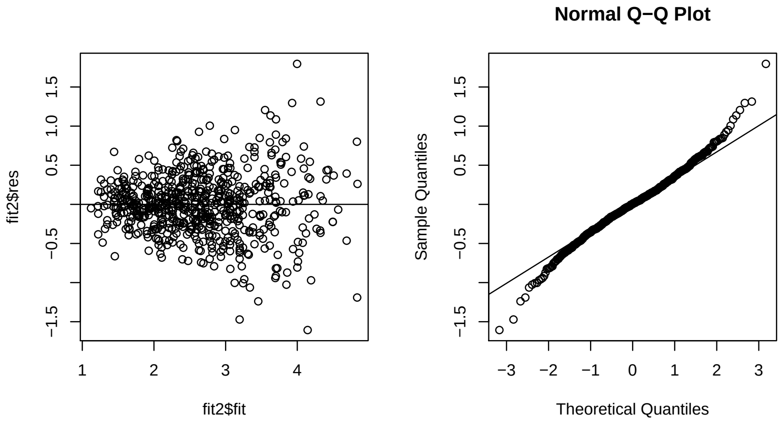
fit$res
2.0
1.5
1.0
0.5
−0.5
−1.5
45
50
55
60
65
70
75
height

```r
fit2<-lm(fev~ age+I(age^2) + height+I(height^2) + male + smoke, data=fev)

summary(fit2)$coef

##                  Estimate   Std. Error    t value     Pr(>|t|)
## (Intercept)  5.92227124 1.7832742991  3.3210097 9.473601e-04
## age          0.02800588 0.0421609352  0.6642614 5.067597e-01
## I(age^2)     0.00176656 0.0017540528  1.0071304 3.142486e-01
## height      -0.23734507 0.0617263805 -3.8451156 1.324046e-04
## I(height^2)  0.00284376 0.0004949071  5.7460481 1.407141e-08
## male         0.09515148 0.0328666266  2.8950790 3.918433e-03
## smoke       -0.13959064 0.0574575045 -2.4294588 1.539252e-02
```



fit2$res
−1.5
−0.5
0.5
1.0
1.5
5
10
15
age


fit2$res
−1.5
−0.5
0.5
1.0
1.5
45
50
55
60
65
70
75
height



Normal Q−Q Plot
fit2$res
fit2$fit
1.5
1.0
0.5
−0.5
−1.5
1
2
3
4
Sample Quantiles
Theoretical Quantiles
−3
−2
−1
0
1
2
3

Alternative covariance model

For the `fev` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} v_1 & 0 & 0 & \cdots & 0 \\ 0 & v_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & v_n \end{pmatrix}$$

where v_i is increasing in $\mathbf{x}_i^T \boldsymbol{\beta}$.

Alternative covariance model

For the `fev` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} v_1 & 0 & 0 & \cdots & 0 \\ 0 & v_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & v_n \end{pmatrix}$$

where v_i is increasing in $\mathbf{x}_i^T \boldsymbol{\beta}$.

This covariance model is said to have *mean-variance relationship*.

Alternative covariance model

For the `fev` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} v_1 & 0 & 0 & \cdots & 0 \\ 0 & v_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & v_n \end{pmatrix}$$

where v_i is increasing in $\mathbf{x}_i^T \boldsymbol{\beta}$.

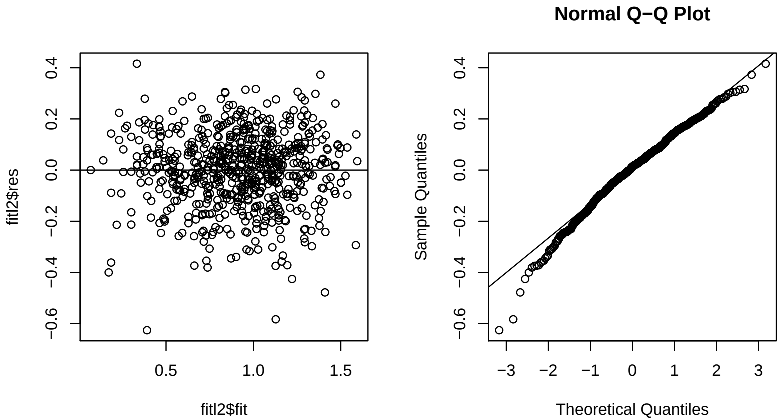
This covariance model is said to have *mean-variance relationship*.

Q:

- What are the SEs of the OLS estimates for heteroscedastic data?
- Are the OLS estimates appropriate if we have heteroscedastic data?

Alternative approach: variance stabilizing transformation

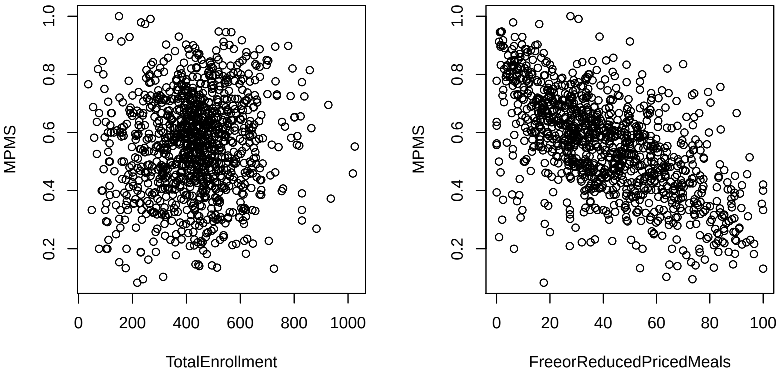
```
fitl2<-lm(log(fev)~age+I(age^2)+height+I(height^2)+male+smoke,data=fev)

summary(fitl2)$coef

##                  Estimate   Std. Error     t value     Pr(>|t|)
## (Intercept) -2.403151e+00 0.6571994572 -3.65665425 0.0002762954
## age          2.466185e-02 0.0155377912  1.58721748 0.1129519557
## I(age^2)    -6.163316e-05 0.0006464303 -0.09534385 0.9240712320
## height       5.794673e-02 0.0227483477  2.54729422 0.0110864759
## I(height^2) -1.259285e-04 0.0001823907 -0.69043290 0.4901695920
## male         3.201560e-02 0.0121125108  2.64318464 0.0084119141
## smoke       -4.384840e-02 0.0211751164 -2.07075125 0.0387780566
```


Normal Q−Q Plot
fitl2$res
fitl2$fit
Sample Quantiles
Theoretical Quantiles

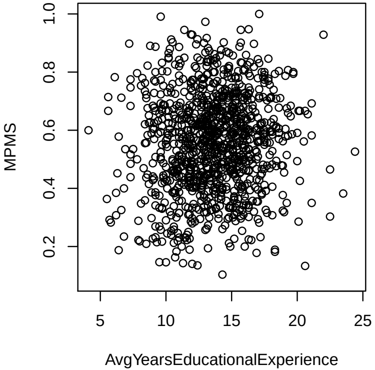
Educational testing

```
colnames(WASL)

##  [1] "County"
##  [2] "School"
##  [3] "MathMetStandard"
##  [4] "ReadingMetStandard"
##  [5] "WritingMetStandard"
##  [6] "ScienceMetStandard"
##  [7] "MathTotalTested"
##  [8] "ReadingTotalTested"
##  [9] "WritingTotalTested"
## [10] "ScienceTotalTested"
## [11] "TotalEnrollment"
## [12] "PercentWhite"
## [13] "StudentsPerClassroomTeacher"
## [14] "FreeorReducedPricedMeals"
## [15] "AvgYearsEducationalExperience"
## [16] "PercentTeachersWithAtLeastMasterDegree"
```

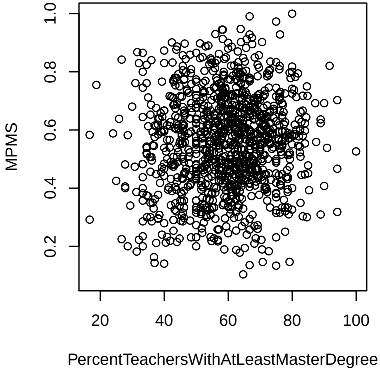
```
MPMS<-WASL$MathMetStandard/WASL$MathTotalTested
```



MPMS
TotalEnrollment
MPMS
FreeorReducedPricedMeals



MPMS
1.0
0.8
0.6
0.4
0.2
5
10
15
20
25
AvgYearsEducationalExperience



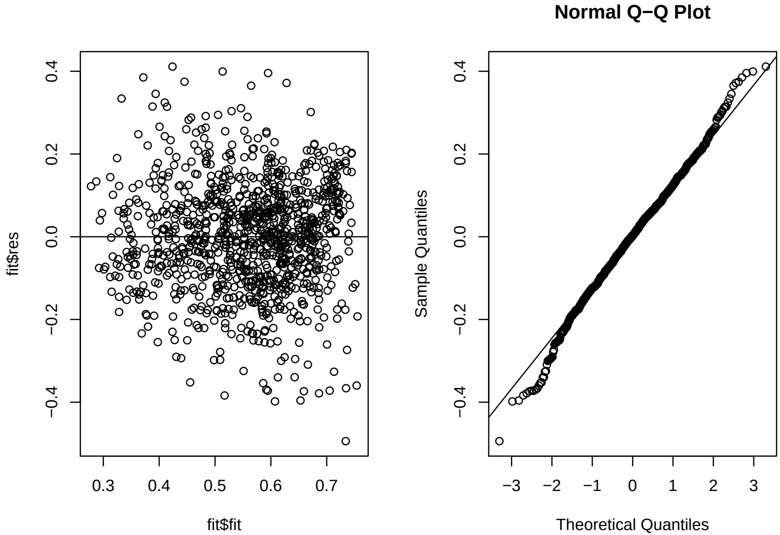
MPMS
1.0
0.8
0.6
0.4
0.2
20
40
60
80
100
PercentTeachersWithAtLeastMasterDegree

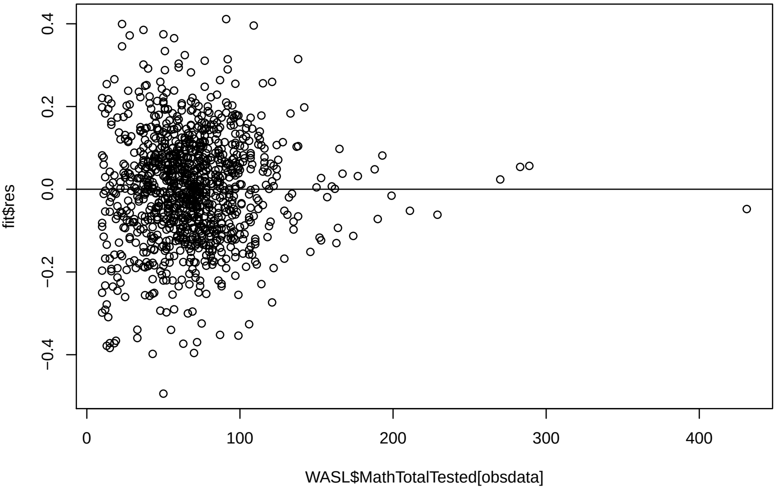
```r
summary(lm(MPMS ~ TotalEnrollment + FreeorReducedPricedMeals +
                 AvgYearsEducationalExperience + PercentTeachersWithAtLeastMasterDegree,
```

```
##                                              Estimate   Std. Error
## (Intercept)                              6.968472e-01 3.160130e-02
## TotalEnrollment                         -3.928079e-06 2.771583e-05
## FreeorReducedPricedMeals                -4.589542e-03 1.828598e-04
## AvgYearsEducationalExperience            2.474453e-03 1.425301e-03
## PercentTeachersWithAtLeastMasterDegree   2.919762e-04 3.080947e-04
##                                            t value      Pr(>|t|)
## (Intercept)                             22.0512212  1.293163e-88
## TotalEnrollment                         -0.1417269  8.873234e-01
## FreeorReducedPricedMeals               -25.0986867 6.373824e-109
## AvgYearsEducationalExperience            1.7360918  8.284576e-02
## PercentTeachersWithAtLeastMasterDegree   0.9476831  3.435125e-01
```

```r
fit<-lm(MPMS~FreeorReducedPricedMeals+AvgYearsEducationalExperience,data=WASL)

summary(fit)$coef
```

```
##                                   Estimate   Std. Error    t value
## (Intercept)                    0.711660884 0.0210808699  33.758611
## FreeorReducedPricedMeals      -0.004595451 0.0001793026 -25.629590
## AvgYearsEducationalExperience  0.002556806 0.0013908766   1.838269
##                                    Pr(>|t|)
## (Intercept)                   3.737758e-169
## FreeorReducedPricedMeals      1.274005e-112
## AvgYearsEducationalExperience  6.630862e-02
```


Normal Q−Q Plot
fit$res
fit$fit
Sample Quantiles
Theoretical Quantiles
0.4
0.2
0.0
−0.2
−0.4
0.3
0.4
0.5
0.6
0.7
−3
−2
−1
0
1
2
3


fit$res
WASL$MathTotalTested[obsdata]
0.4
0.2
0.0
−0.2
−0.4
0
100
200
300
400

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\mathrm{E}[y_i] =$

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] =$

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] = p_i(1 - p_i)/n_i$.

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] = p_i(1 - p_i)/n_i$.

Thus we expected that the errors in a linear model will be heteroscedastic because

- the variance depends on p_i, *and*
- the variance depends on n_i.

Alternative covariance model

For these data

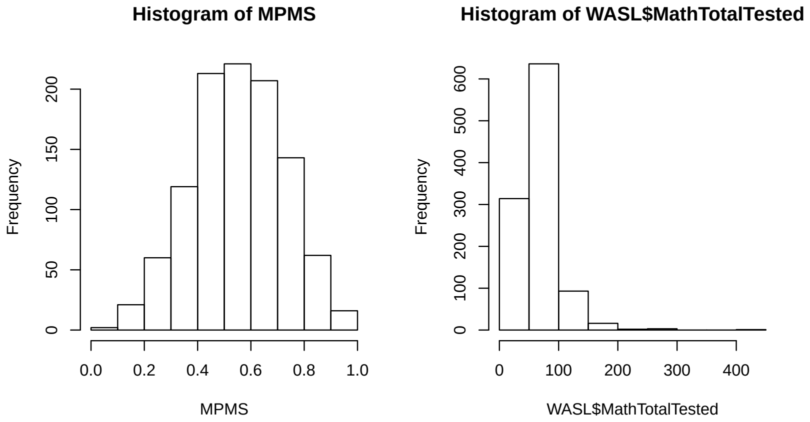
$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] = p_i(1 - p_i)/n_i$.

Thus we expected that the errors in a linear model will be heteroscedastic because

- the variance depends on p_i, *and*
- the variance depends on n_i.


Histogram of MPMS
Frequency
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
MPMS
Histogram of WASL$MathTotalTested
Frequency
0
100
200
300
400
500
600
0
100
200
300
400
WASL$MathTotalTested

Alternative covariance model

For the `WASL` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix}$$

where w_i is the number of students tested in school i.

Alternative covariance model

For the `WASL` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix}$$

where w_i is the number of students tested in school i.

This covariance is a result of the responses being sample means with different sample sizes.

Alternative covariance model

For the `WASL` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix}$$

where w_i is the number of students tested in school i.

This covariance is a result of the responses being sample means with different sample sizes.

Q:

- What are the SEs of the OLS estimates for heteroscedastic data?
- Are the OLS estimates appropriate if we have heteroscedastic data?

Alternative model for the data

Logistic regression:

$$p_i = \frac{\exp\left(\beta^T x_i\right)}{1 + \exp\left(\beta^T x_i\right)}$$

$$\log \tfrac{p_i}{1-p_i} = \beta^T x_i$$

OLS and GLS

OLS: $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

- $\mathrm{E}[\hat{\boldsymbol{\beta}}_{OLS}] = \boldsymbol{\beta}$
- $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{OLS}] = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\Sigma\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}$

OLS and GLS

OLS: $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

- $\mathrm{E}[\hat{\boldsymbol{\beta}}_{OLS}] = \boldsymbol{\beta}$
- $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{OLS}] = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\Sigma\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}$

GLS: $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}\mathbf{y}$

- $\mathrm{E}[\hat{\boldsymbol{\beta}}_{GLS}] = \boldsymbol{\beta}$
- $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{GLS}] = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}$

OLS and GLS

OLS: $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

- $\mathrm{E}[\hat{\boldsymbol{\beta}}_{OLS}] = \boldsymbol{\beta}$
- $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{OLS}] = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\Sigma\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}$

GLS: $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}\mathbf{y}$

- $\mathrm{E}[\hat{\boldsymbol{\beta}}_{GLS}] = \boldsymbol{\beta}$
- $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{GLS}] = (\mathbf{X}^T\Sigma^{-1}\mathbf{X})^{-1}$

$\mathrm{Var}[\hat{\boldsymbol{\beta}}_{GLS}] < \mathrm{Var}[\hat{\boldsymbol{\beta}}_{OLS}]$, in a matrix sense.

Feasible GLS

If we use $\hat{\boldsymbol{\beta}}_{OLS}$ to infer $\boldsymbol{\beta}$, we need Σ to compute standard errors.

- Typically, some aspect of Σ is unknown.
- In practice, we estimate Σ from the data.

Feasible GLS

If we use $\hat{\boldsymbol{\beta}}_{OLS}$ to infer $\boldsymbol{\beta}$, we need Σ to compute standard errors.

- Typically, some aspect of Σ is unknown.
- In practice, we estimate Σ from the data.

We can't even compute $\hat{\boldsymbol{\beta}}_{GLS}$ without knowledge of Σ

- For WLS, we know "enough" to obtain $\hat{\boldsymbol{\beta}}_{GLS}$.
- For other problems, we might iteratively estimate Σ and $\hat{\boldsymbol{\beta}}_{GLS}$.

This later approach is called *feasible GLS*.

Feasible GLS

If we use $\hat{\boldsymbol{\beta}}_{OLS}$ to infer $\boldsymbol{\beta}$, we need Σ to compute standard errors.

- Typically, some aspect of Σ is unknown.
- In practice, we estimate Σ from the data.

We can't even compute $\hat{\boldsymbol{\beta}}_{GLS}$ without knowledge of Σ

- For WLS, we know "enough" to obtain $\hat{\boldsymbol{\beta}}_{GLS}$.
- For other problems, we might iteratively estimate Σ and $\hat{\boldsymbol{\beta}}_{GLS}$.

This later approach is called *feasible GLS*.

Feasible GLS: Letting $\hat{\boldsymbol{\beta}}_{GLS}^{(0)} = \hat{\boldsymbol{\beta}}_{OLS}$, iterate as follows:

1. Estimate $\hat{\Sigma}_{(k+1)}$ from $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}^{(k)}$;
2. Let $\hat{\boldsymbol{\beta}}_{GLS}^{(k+1)} = (\mathbf{X}^T\hat{\Sigma}_{(k+1)}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\Sigma}_{(k+1)}^{-1}\mathbf{y}$

OLS for ice core data with corrected SEs

```
fit<-lm( tmp ~ co2 , data=VostokIce )

s2<-sum(fit$res^2)/(n-2)

e1<-fit$res[-1]
e0<-fit$res[-n]
rho<-cor(e1,e0)

c( s2, rho )

## [1] 2.3498446 0.5183877
```

```
Sig<- s2* ( rho^abs(outer( 1:n , 1:n , "-" ) ) )

Sig[3,8]

## [1] 0.08796556

s2*rho^5

## [1] 0.08796556
```

OLS for ice core data with corrected SEs

$$\text{Var}[\hat{\beta}_{OLS}] = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\Sigma\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}$$

```
X<-cbind(1,VostokIce$co2)

tXX<-t(X)%*%X

itXX<-solve(tXX)

VBols<-  itXX %*% (t(X)%*%Sig%*%X) %*% itXX

VBols

##              [,1]          [,2]
## [1,]  2.081738848 -0.0089660853
## [2,] -0.008966085  0.0000393115
```

OLS for ice core data with corrected SEs

```
se_b1<-sqrt( VBols[2,2] )

se_b1

## [1] 0.006269889

t_b1<-fit$coef[2]/se_b1

t_b1

##      co2
## 12.73594
```

```
summary(fit)$coef

##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) -23.02414043 0.879542928 -26.17739 3.272183e-66
## co2           0.07985291 0.003833728  20.82905 8.766508e-52
```

Feasible GLS

```
y<-VostokIce$tmp

beta<- solve( t(X)%*%X )%*%t(X)%*%y

Theta<-NULL
for(iter in 1:15)
{
  ## update Sigma
  e<-y-X%*%beta
  s2<-sum(e^2)/(n-2)
  rho<-cor(e[-1], e[-n])
  Sig<- s2* ( rho^abs(outer( 1:n , 1:n , "-" ) ) )

  ## update beta
  beta<- solve( t(X)%*%solve(Sig)%*%X ) %*% t(X) %*% solve(Sig) %*% y

  ## save parameter estimates
  Theta<-rbind(Theta, c(beta,s2,rho) )
}
```

Feasible GLS

```
Theta
```

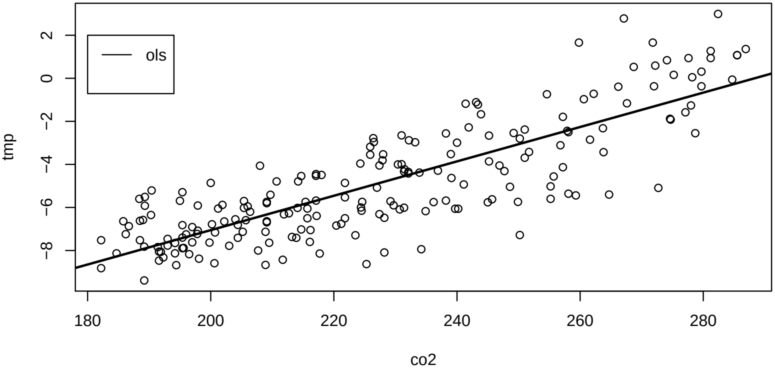
```
##              [,1]       [,2]     [,3]      [,4]
##  [1,] -18.83769 0.06155813 2.349845 0.5183877
##  [2,] -16.55915 0.05163925 2.620561 0.6257848
##  [3,] -14.79284 0.04398281 2.994345 0.6927016
##  [4,] -13.44266 0.03816190 3.393018 0.7395576
##  [5,] -12.52059 0.03421227 3.761456 0.7708074
##  [6,] -11.96593 0.03185155 4.044637 0.7896868
##  [7,] -11.66375 0.03057188 4.227341 0.8000821
##  [8,] -11.50913 0.02991926 4.330819 0.8054454
##  [9,] -11.43274 0.02959743 4.384851 0.8081085
## [10,] -11.39568 0.02944142 4.411819 0.8094040
## [11,] -11.37785 0.02936645 4.424970 0.8100278
## [12,] -11.36932 0.02933056 4.431309 0.8103267
## [13,] -11.36524 0.02931341 4.434347 0.8104695
## [14,] -11.36330 0.02930523 4.435799 0.8105376
## [15,] -11.36237 0.02930133 4.436493 0.8105702
```

Inference based on GLS

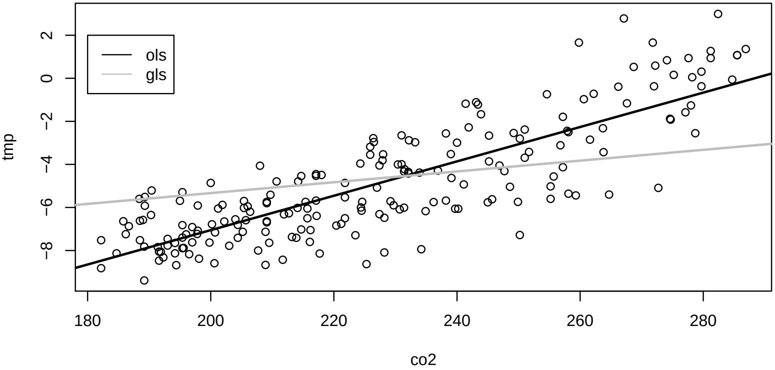
$$\text{Var}[\hat{\beta}_{GLS}] = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1}$$

```
VBgls<-solve( t(X)%*%solve(Sig)%*%X )

se_b1<-sqrt( VBgls[2,2] )

se_b1

## [1] 0.006986359

t_b1<-beta[2]/se_b1

t_b1

## [1] 4.194077
```

```r
library(nlme)

fit_gls <- gls(tmp~co2, correlation=corARMA(p=1),data=VostokIce)

summary(fit_gls)

## Generalized least squares fit by REML
##   Model: tmp ~ co2
##   Data: VostokIce
##        AIC      BIC    logLik
##   664.8527 678.0058 -328.4264
##
## Correlation Structure: AR(1)
##  Formula: ~1
##  Parameter estimate(s):
##       Phi
## 0.8470009
##
## Coefficients:
##                  Value Std.Error   t-value p-value
## (Intercept) -10.355955 1.7231321 -6.009960   0e+00
## co2           0.025119 0.0070807  3.547493   5e-04
##
##  Correlation:
##     (Intr)
## co2 -0.947
##
## Standardized residuals:
##        Min         Q1         Med          Q3         Max
## -1.7007966 -0.7472771 -0.3278727  0.4786885  2.7789435
##
## Residual standard error: 2.312641
## Degrees of freedom: 200 total; 198 residual
```



ols
tmp
2
0
−2
−4
−6
−8
180
200
220
240
260
280
co2



ols
gls
tmp
co2
2
0
−2
−4
−6
−8
180
200
220
240
260
280

OLS vs GLS

Q: Why does the OLS fit seem more visually reasonable?

OLS vs GLS

Q: Why does the OLS fit seem more visually reasonable?

A: Recall what $\hat{\beta}_{OLS}$ and $\hat{\beta}_{GLS}$ are minimizing.

$$\hat{\beta}_{OLS} = \arg\min(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$
$$\hat{\beta}_{GLS} = \arg\min(\mathbf{y} - \mathbf{X}\beta)^T\hat{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

OLS vs GLS

Q: Why does the OLS fit seem more visually reasonable?

A: Recall what $\hat{\beta}_{OLS}$ and $\hat{\beta}_{GLS}$ are minimizing.

$$\hat{\beta}_{OLS} = \arg\min(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$
$$\hat{\beta}_{GLS} = \arg\min(\mathbf{y} - \mathbf{X}\beta)^T\hat{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

By definition, $\hat{\beta}_{OLS}$ will provide a better fit to the observed data.

OLS vs GLS

Q: Why does the OLS fit seem more visually reasonable?

A: Recall what $\hat{\beta}_{OLS}$ and $\hat{\beta}_{GLS}$ are minimizing.

$$\hat{\beta}_{OLS} = \arg\min(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$
$$\hat{\beta}_{GLS} = \arg\min(\mathbf{y} - \mathbf{X}\beta)^T\hat{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

By definition, $\hat{\beta}_{OLS}$ will provide a better fit to the observed data.

Do you believe that β_{GLS} is closer to the truth?

Simulation study

IDEA:

- Simulate data under a particular correlation ρ;
- Compare OLS and GLS estimates.

Simulation study

IDEA:

- Simulate data under a particular correlation ρ;
- Compare OLS and GLS estimates.

```
beta_true<-c(-11,.05) ; s2_true<-4 ; rho_true<-.85

hSig_true<- mhalf( s2_true* ( rho_true^abs(outer( 1:n , 1:n , "-" ) ) ) )
```

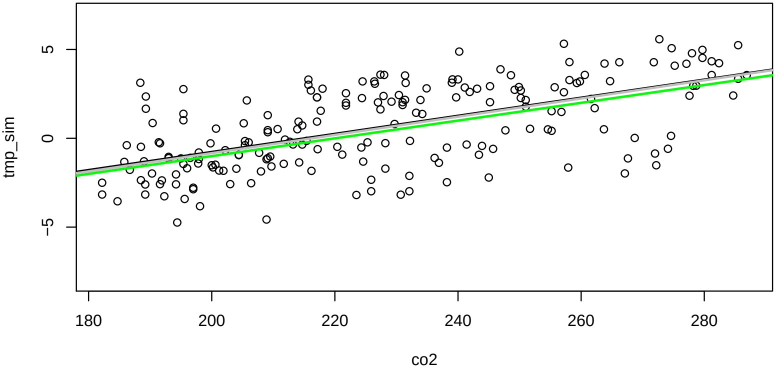
Simulation study

IDEA:

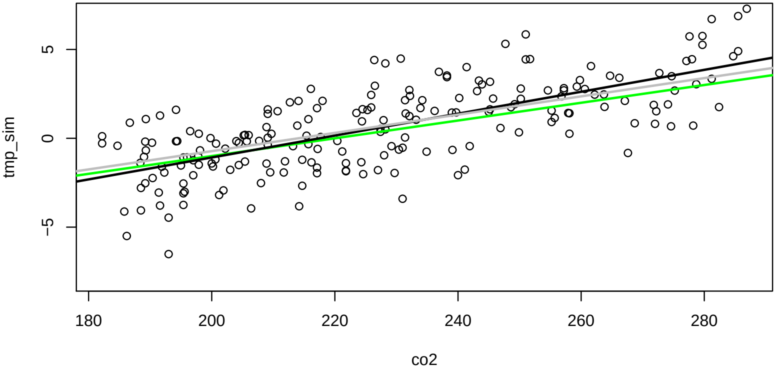
- Simulate data under a particular correlation ρ;
- Compare OLS and GLS estimates.

```
beta_true<-c(-11,.05) ; s2_true<-4 ; rho_true<-.85

hSig_true<- mhalf( s2_true* ( rho_true^abs(outer( 1:n , 1:n , "-" ) ) ) )
```

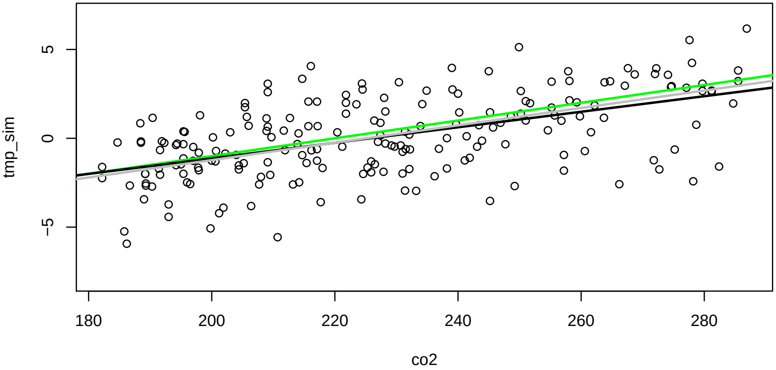
```
tmp_sim<- beta_true[1] + beta_true[2]*co2 + hSig_true%*%rnorm(n)

beta_ols<-lm(tmp_sim~co2)$coef

beta_gls<-gls(tmp_sim~co2, correlation=corARMA(p=1))$coef
```



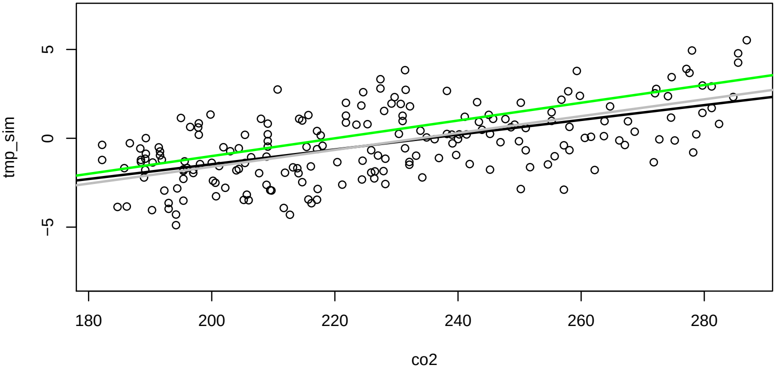
5
0
−5
tmp_sim
180
200
220
240
260
280
co2

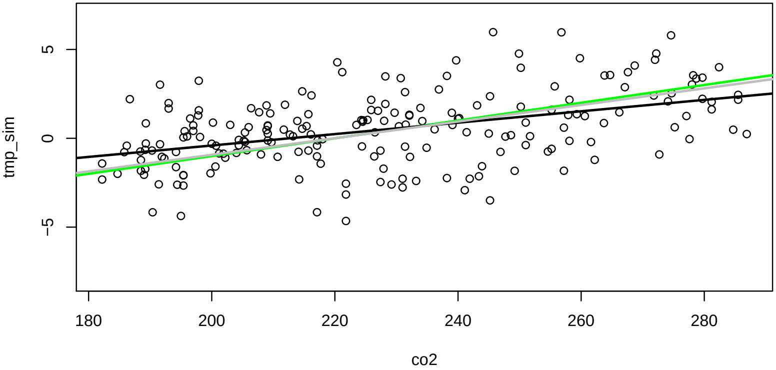


tmp_sim
co2
5
0
−5
180
200
220
240
260
280



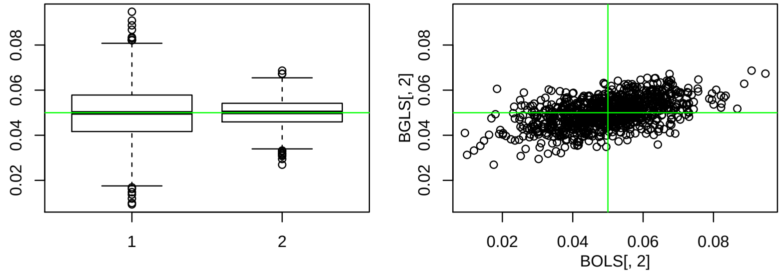
tmp_sim
co2
5
0
−5
180
200
220
240
260
280


tmp_sim
co2
5
0
−5
180
200
220
240
260
280

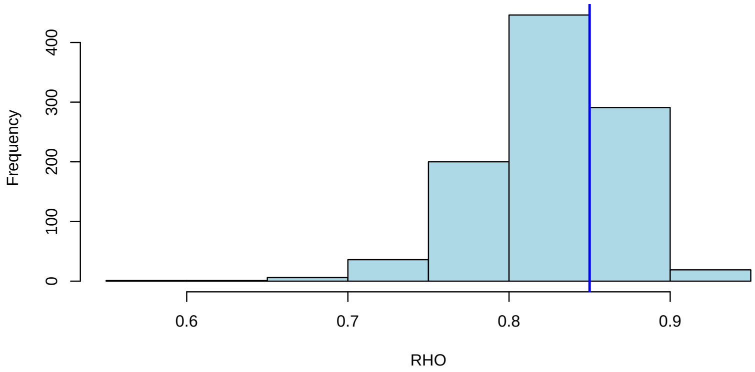


5
0
−5
tmp_sim
180
200
220
240
260
280
co2

```
BOLS<-BGLS<-NULL
for(s in 1:1000)
{
  set.seed(s)

  tmp_sim<- beta_true[1] + beta_true[2]*co2 + hSig_true%*%rnorm(n)

  BOLS<-rbind(BOLS,lm(tmp_sim~co2)$coef)
  BGLS<-rbind(BGLS,gls(tmp_sim~co2,correlation=corARMA(p=1))$coef)
}
```



Estimates of ρ



Abundance of caution

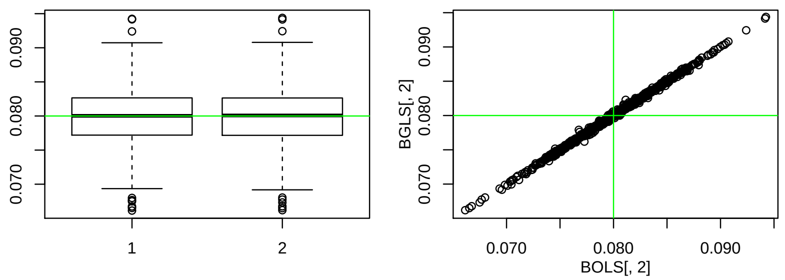
What if the data are not correlated, but we use GLS?

Abundance of caution

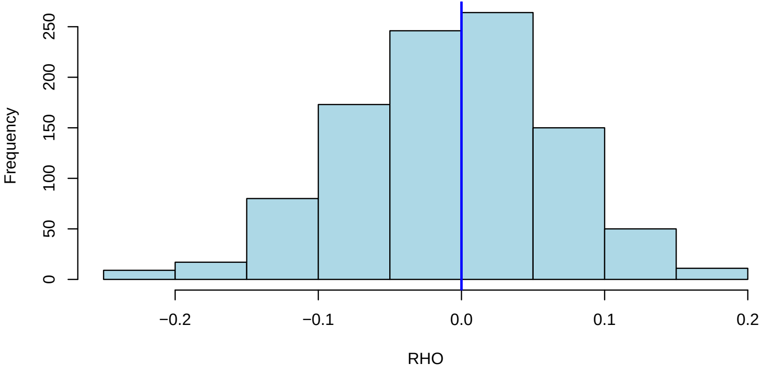
What if the data are not correlated, but we use GLS?

- Both estimators are still unbiased.
- Which will have lower variance?

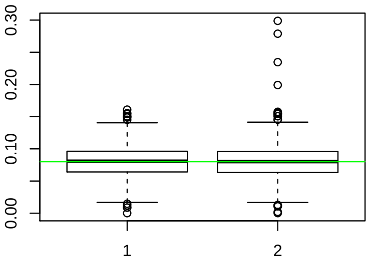
```r
beta_true<-c(-23,.08) ; s2_true<-2.5 ; rho_true<-0

hSig_true<- mhalf( s2_true* ( rho_true^abs(outer( 1:n , 1:n , "-" ) ) ) )

BOLS<-BGLS<-NULL
for(s in 1:1000)
{
  set.seed(s)

  tmp_sim<- beta_true[1] + beta_true[2]*co2 + hSig_true%*%rnorm(n)

  BOLS<-rbind(BOLS,lm(tmp_sim~co2)$coef)
  BGLS<-rbind(BGLS,gls(tmp_sim~co2,correlation=corARMA(p=1))$coef)
}
```

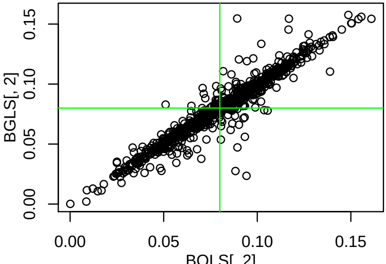


Estimates of ρ

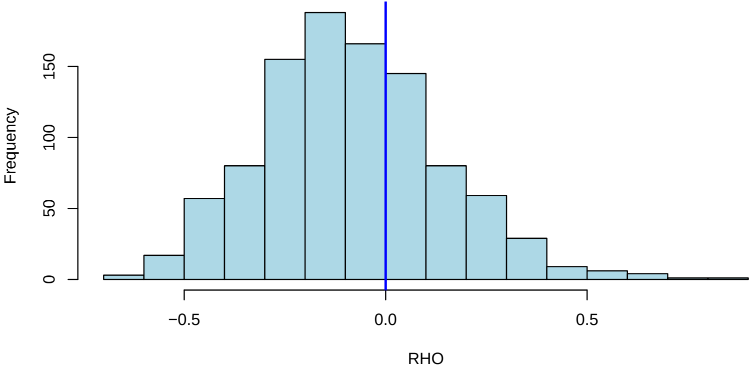


```
n<-20
co2<-co2[1:n]
beta_true<-c(-23,.08) ; s2_true<-2.5 ; rho_true<-0

hSig_true<- mhalf( s2_true* ( rho_true^abs(outer( 1:n , 1:n , "-" ) ) ) )

BOLS<-BGLS<-NULL
for(s in 1:1000)
{
  set.seed(s)

  tmp_sim<- beta_true[1] + beta_true[2]*co2 + hSig_true%*%rnorm(n)

  BOLS<-rbind(BOLS,lm(tmp_sim~co2)$coef)
  BGLS<-rbind(BGLS,gls(tmp_sim~co2,correlation=corARMA(p=1))$coef)
}
```





Estimates of ρ



Summmary of GLS for correlated data

- If there is correlation, GLS is more reliable.
- If there is no correlation, $\hat{\rho} \approx 0$ and $\hat{\beta}_{GLS} \approx \hat{\beta}_{OLS}$.

This latter behavior worked in this example becuase there was plenty of data ($n = 200$) to get a good estimate of ρ.

Summmary of GLS for correlated data

- If there is correlation, GLS is more reliable.
- If there is no correlation, $\hat{\rho} \approx 0$ and $\hat{\beta}_{GLS} \approx \hat{\beta}_{OLS}$.

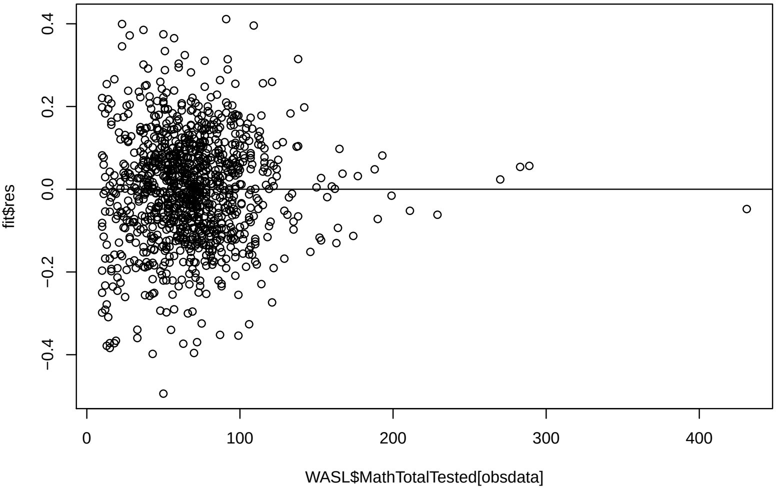
This latter behavior worked in this example becuase there was plenty of data ($n = 200$) to get a good estimate of ρ.

If there is less data, or the correlation model is more complicated, the ability of the GLS estimate to reduce to the OLS estimate is diminished.

Weighted lest squares

WASL data revisited:

```
fit<-lm(MPMS~FreeorReducedPricedMeals+AvgYearsEducationalExperience,data=WASL)

round(summary(fit)$coef,4)

##                               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)                     0.7117     0.0211  33.7586   0.0000
## FreeorReducedPricedMeals       -0.0046     0.0002 -25.6296   0.0000
## AvgYearsEducationalExperience   0.0026     0.0014   1.8383   0.0663
```



fit$res
0.4
0.2
0.0
−0.2
−0.4
0
100
200
300
400
WASL$MathTotalTested[obsdata]

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] =$

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] =$

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] = p_i(1 - p_i)/n_i$.

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\text{E}[y_i] = p_i$
- $\text{Var}[y_i] = p_i(1 - p_i)/n_i$.

Thus we expected that the errors in a linear model will be heteroscedastic because

- the variance depends on p_i, *and*
- the variance depends on n_i.

Alternative covariance model

For these data

$$y_i = \text{percent met standard}_i = \frac{\text{total met standard}_i}{\text{number tested}_i}$$

Let p_i be the probability a student tested in school i will meet standard.

- $\mathrm{E}[y_i] = p_i$
- $\mathrm{Var}[y_i] = p_i(1 - p_i)/n_i$.

Thus we expected that the errors in a linear model will be heteroscedastic because

- the variance depends on p_i, *and*
- the variance depends on n_i.

Alternative covariance model

For the `WASL` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix}$$

where w_i is the number of students tested in school i.

Alternative covariance model

For the `WASL` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix}$$

where w_i is the number of students tested in school i.

This covariance is a result of the responses being sample means with different sample sizes.

Alternative covariance model

For the `WASL` data, it might be more reasonable to assume

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix}$$

where w_i is the number of students tested in school i.

This covariance is a result of the responses being sample means with different sample sizes.

Q:

- What does the GLS estimate look like?
- How to implement this in R?

WLS

$$\mathrm{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix} = \sigma^2 \mathbf{W}^{-1},$$

where **W** is a matrix of weights.

WLS

$$\text{Var}[\mathbf{y}] = \Sigma = \sigma^2 \begin{pmatrix} 1/w_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/w_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1/w_n \end{pmatrix} = \sigma^2 \mathbf{W}^{-1},$$

where $\mathbf{W}$ is a matrix of weights.

The GLS estimator is

$$\begin{aligned} \hat{\beta}_{GLS} &= (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{y} \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X}/\sigma^2)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}/\sigma^2 \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \end{aligned}$$

WLS

$$\hat{\beta}_{WLS} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$$

Note 1: We don't need to know σ^2, only the weights $\mathbf{W}$.

WLS

$$\hat{\beta}_{WLS} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$$

Note 1: We don't need to know σ^2, only the weights $\mathbf{W}$.

Note 2: The GLS/WLS estimator is basically putting "more weight" on observations with smaller variances, less weight on observations with big variances.

WLS

$$\hat{\beta}_{WLS} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$$

Note 1: We don't need to know σ^2, only the weights $\mathbf{W}$.

Note 2: The GLS/WLS estimator is basically putting "more weight" on observations with smaller variances, less weight on observations with big variances.

This second note is made clear by considering what β_{WLS} is optimizing:

$$\begin{aligned}
\hat{\beta}_{WLS} &= \arg\min_{\beta}(\mathbf{y} - \mathbf{X}\beta)^T\Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta) \\
&= \arg\min_{\beta}(\mathbf{y} - \mathbf{X}\beta)^T\mathbf{W}(\mathbf{y} - \mathbf{X}\beta) \\
&= \arg\min_{\beta}\sum w_i(y_i - \mathbf{x}_i^T\beta)^2
\end{aligned}$$

WLS implementation

```
fit_ols<-lm( MPMS ~ FreeorReducedPricedMeals +
                    AvgYearsEducationalExperience, data=WASL)

fit_wls<-lm( MPMS ~ FreeorReducedPricedMeals +
                    AvgYearsEducationalExperience, weights=MathTotalTested, dat
```

WLS implementation

```
round(summary(fit_ols)$coef,4)

##                                 Estimate Std. Error  t value Pr(>|t|)
## (Intercept)                       0.7117     0.0211  33.7586   0.0000
## FreeorReducedPricedMeals         -0.0046     0.0002 -25.6296   0.0000
## AvgYearsEducationalExperience     0.0026     0.0014   1.8383   0.0663
```

```
round(summary(fit_wls)$coef,4)

##                                 Estimate Std. Error  t value Pr(>|t|)
## (Intercept)                       0.7304     0.0206  35.4357   0.0000
## FreeorReducedPricedMeals         -0.0048     0.0002 -28.3886   0.0000
## AvgYearsEducationalExperience     0.0021     0.0014   1.5162   0.1298
```