Weighted least squares

PBHS 32410 Winter 2024

Bowei kang

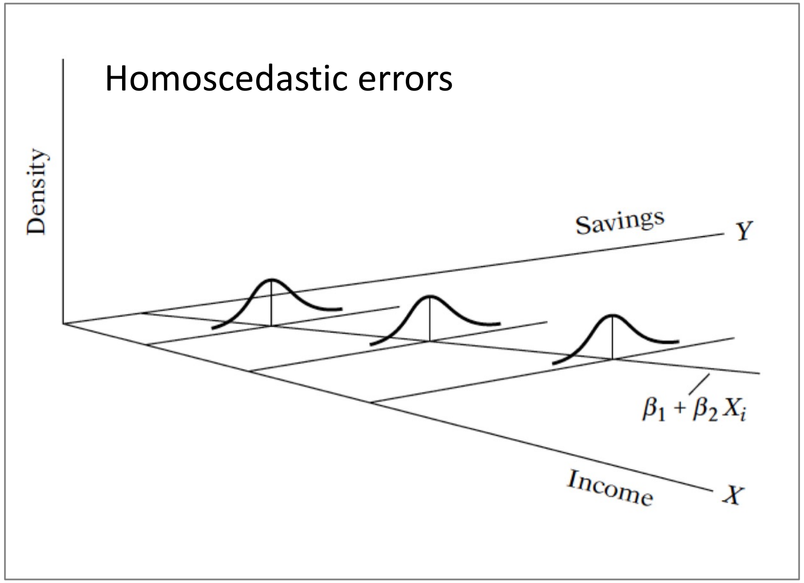
Motivation

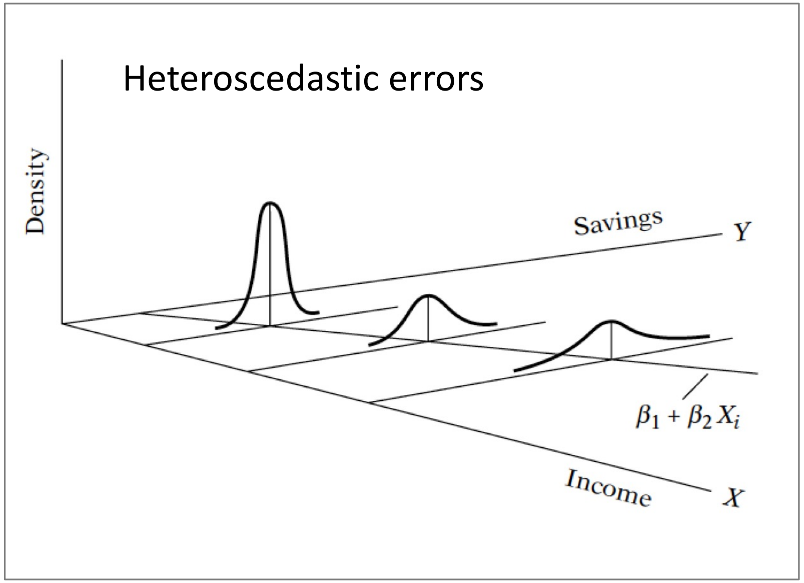
- Let us continue the topic: What do we do when linear regression assumptions are violated?
- We have discussed that transformations are useful tools to “linearize” some non-linear relationship.
- Today, we focus on handling the violation of constant variance assumption via weighted least squares (WLS) method.

Assumption relaxation

- The method of ordinary least squares assumes that there is constant variance in the errors (homoscedasticity)
- The method of weighted least squares can be used when the ordinary least squares assumption of constant variance in the errors is violated (heteroscedasticity).
- WLS is performed based on a relaxed assumption on variance: $Var(\epsilon_i) = \sigma^2$ → $Var(\epsilon_i) = \sigma_i^2$.

Homoscedasticity vs. heteroscedasticity


Homoscedastic errors


Heteroscedastic errors

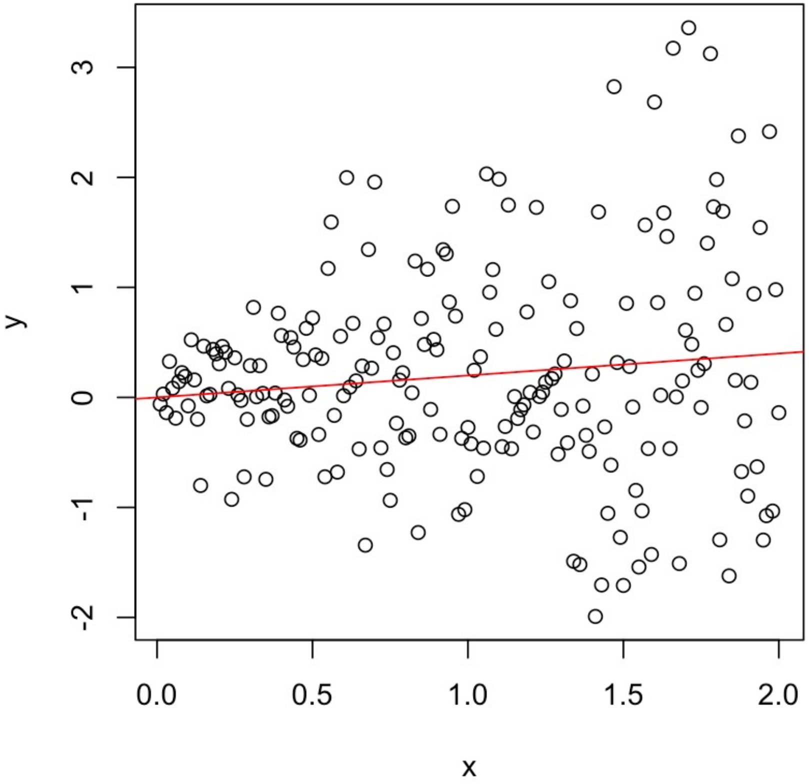
Source: https://hedibert.org/wp-content/uploads/2016/10/heteroskedasticity.pdf

Applying OLS on heteroscedastic data

- Recall OLS only needs the linearity, and do not require assumptions on error terms.
- Thus, OLS estimate for coefficient is still unbiased even when the constant variance assumption is violated.
- Is it necessary to handle heteroscedasticity using tools like WLS?
- What is the consequence of omitting heteroscedasticity by falsely using OLS?

An simulated example

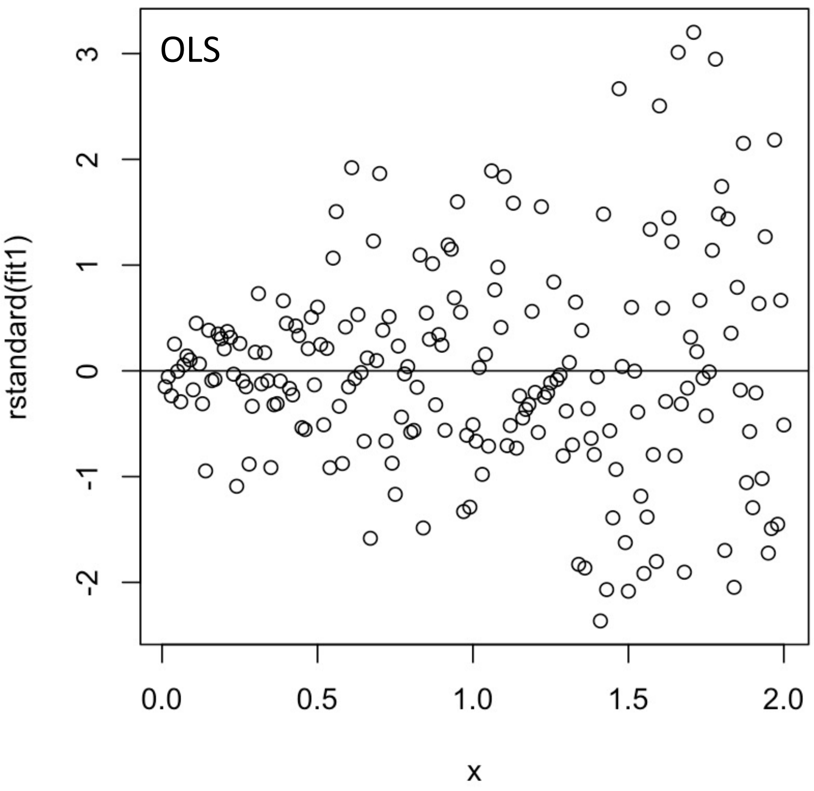
- $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- Assume $\beta_0 = 0, \beta_1 = 0.2$
- X_i is generated from Uniform(0,2)
- $\epsilon_i \sim N(0, \sigma_i^2), \sigma_i^2 = \sigma_0^2 X_i, \sigma_0^2 = 1$
- $n = 200$

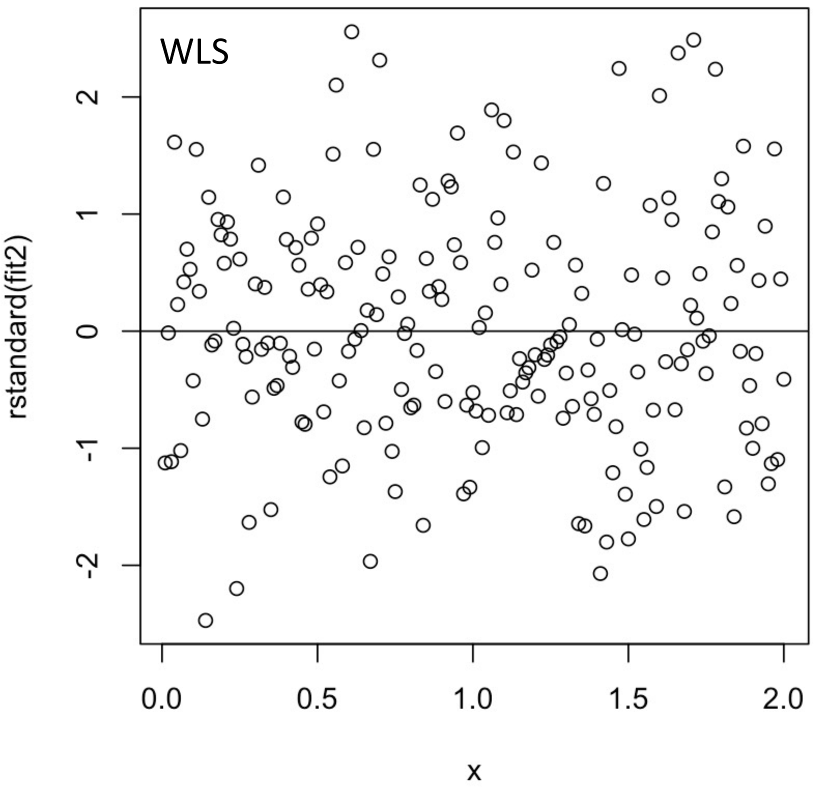


OLS vs. WLS

	OLS	WLS (using true weight $1/\sigma_i^2$)
$\hat{\beta}$	0.2193	0.2295
SE of $\hat{\beta}$	0.1218	0.0864
t-statistic	1.8000	2.6540
P-value	0.0733	0.0086

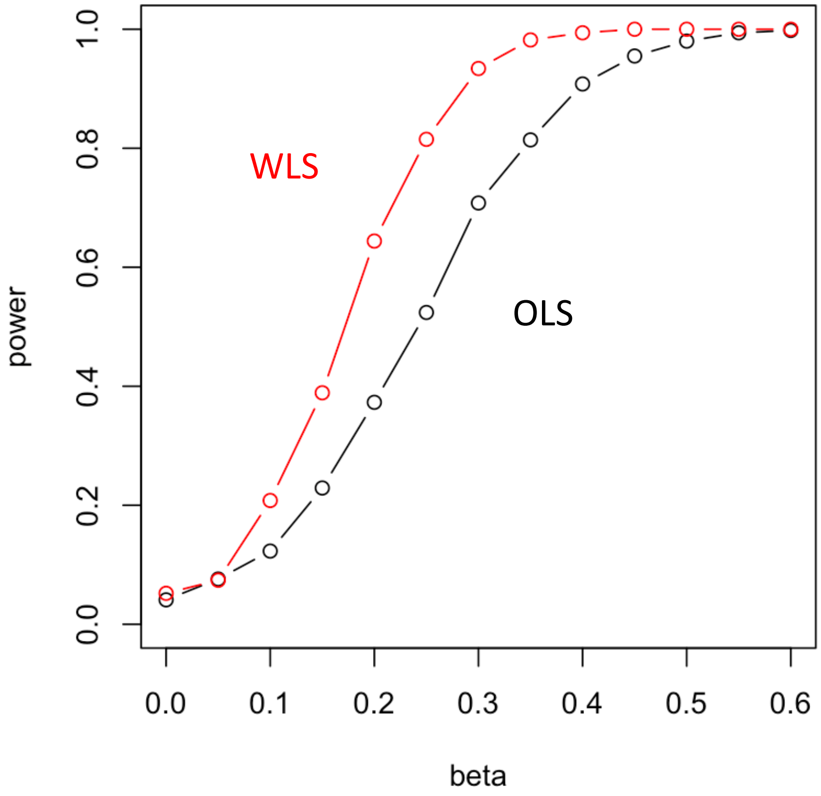
- Answer the question: “is there a linear relationship between X and Y?”
- We want to test H0: $\beta_1 = 0$


OLS
rstandard(fit1)
3
2
1
0
-1
-2
0.0
0.5
1.0
1.5
2.0
x


WLS
rstandard(fit2)
2
1
0
-1
-2
0.0
0.5
1.0
1.5
2.0
x

Difference in power

- Recall power is the proportion of correctly rejecting the null (beta = 0). That is, the "ability" to detect the signal when the signal does exist.



Applying OLS on heteroscedastic data

- If omitting heteroscedasticity and falsely using OLS, the estimators of the variances, $V(\hat{\beta})$, will be biased, and the inference (hypothesis test, confidence interval) based on this variance estimator will be wrong.
- While the ordinary least squares estimator is still unbiased in the presence of heteroscedasticity, it is inefficient and the inference based on the assumption of homoscedasticity is misleading.

WLS model

- In a regression $Y = \beta X + \epsilon$ with unequal variance, covariance matrix of ϵ:
- $\begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$

- If we define the weight, $w_i = 1/\sigma_i^2$, then let matrix W be a diagonal matrix containing these weights:
- $\begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$

- $Var(\epsilon) = W^{-1}$

The logic of using inverse variance as weight

- The error variance reflects the information contained in that observation. A smaller variance means it contains information with less uncertainty and more reliability.
- An observation containing more information (with a smaller error variance) is given a relatively larger weight than an observation containing less information (with a larger error variance).

WLS estimation

- Consider a MLR model, $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i$.
- WLS estimate of $\beta_0, \cdots, \beta_p$ are obtained by minimizing $\sum_i w_i (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_p X_{ip})^2$.
- Here weights are inversely proportional to the individual-specific variance of errors, $w_i = 1/\sigma_i^2$.
- Any observations with a small weight will be severely discounted by WLS in determining the values of $\hat{\beta}$.
- In extreme cases that $w_i = 0$, the effect of WLS is to exclude the ith observation from the estimation process.
- The weight is up to a proportionality constant.

WLS estimation

- WLS method is equivalent to performing OLS on the transformed variables.
- $Y = X\beta + \epsilon$, where $Var(\epsilon) = W^{-1}\sigma^2$
- Let $W^{1/2}$ be a diagonal matrix with diagonal entries equal to $\sqrt{w_i}$.
- Then we have $Var(W^{1/2}\epsilon) = \sigma^2 I_n$
- Consider the transformation $\tilde{Y} = W^{1/2}Y, \tilde{X} = W^{1/2}X, \tilde{\epsilon} = W^{1/2}\epsilon$
- This gives rise to the OLS model: $\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon}$
- Using the results from OLS we then get the solution: $\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\,\tilde{X}'\tilde{Y} = (X'WX)^{-1}X'WY$
- WLS estimate is also unbiased and approximately normal: $\hat{\beta} \sim N(\beta, (X'WX)^{-1}\sigma^2)$ as $n \to \infty$.

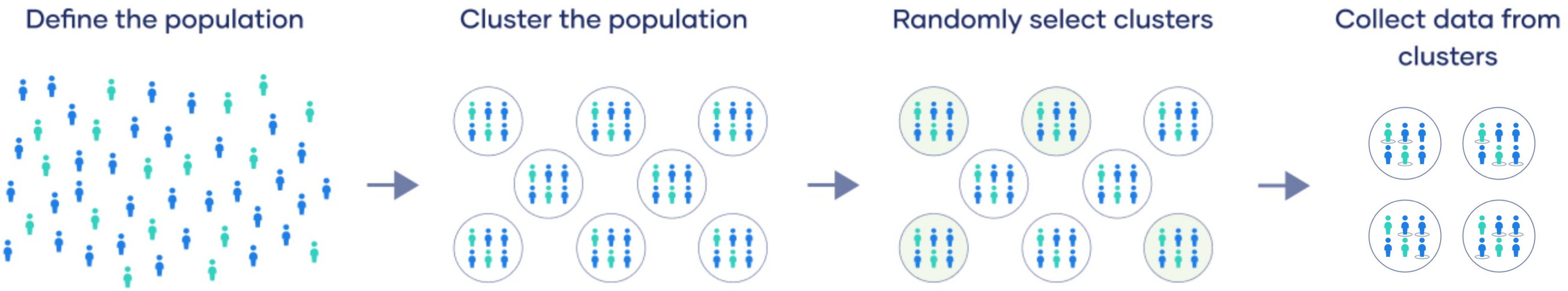
Determine weights

- Weights are known.
- Weights are unknown but the source of heteroscedasticity can be identified
- Weights are unknown and the source of heteroscedasticity cannot be identified.

Weight known a priori

- Circumstances where the weights are known:
- If the i-th response is an average of equally variable observations, then $Var(y_i) = \sigma^2/n_i$ and $w_i = n_i$.
- Recall: $Var(\bar{y}) = Var\left(\frac{1}{n}\sum_i y_i\right) = \frac{1}{n^2}\sum_i Var(y_i) = \frac{1}{n^2}n\sigma^2 = \sigma^2/n$
- If the i-th response is a total of equally variable observations, $Var(y_i) = n_i\,\sigma^2$ and $w_i = 1/n_i$.
- Recall: $Var(\sum_i y_i) = \sum_i Var(y_i) = n\sigma^2$

Cluster survey



https://www.scribbr.com/methodology/cluster-sampling/

Weight known a priori

- Data are collected by selecting a set of schools at random and interviewing a prescribed number of randomly selected students at each school.
- The response variable, Y, is the average expenditure at the ith school.
- The predictors are characteristics of the school, like size of city or town where school is located, type of school (public or private), distance to nearest urban center, etc.

y_{ij}	$\bar{y}_{i\cdot}$	x_i	$w_i = n_i$
y_{11}	$\bar{y}_{1\cdot} = \sum_j y_{1j} / n_1$	x_1	n_1
y_{12}			
$\vdots$			
y_{1n_1}			
y_{21}	$\bar{y}_{2\cdot} = \sum_j y_{2j} / n_2$	x_2	n_2
y_{22}			
$\vdots$			
y_{2n_2}			
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{m1}	$\bar{y}_{m\cdot} = \sum_j y_{mj} / n_m$	x_m	n_m
y_{m2}			
$\vdots$			
y_{mn_m}			

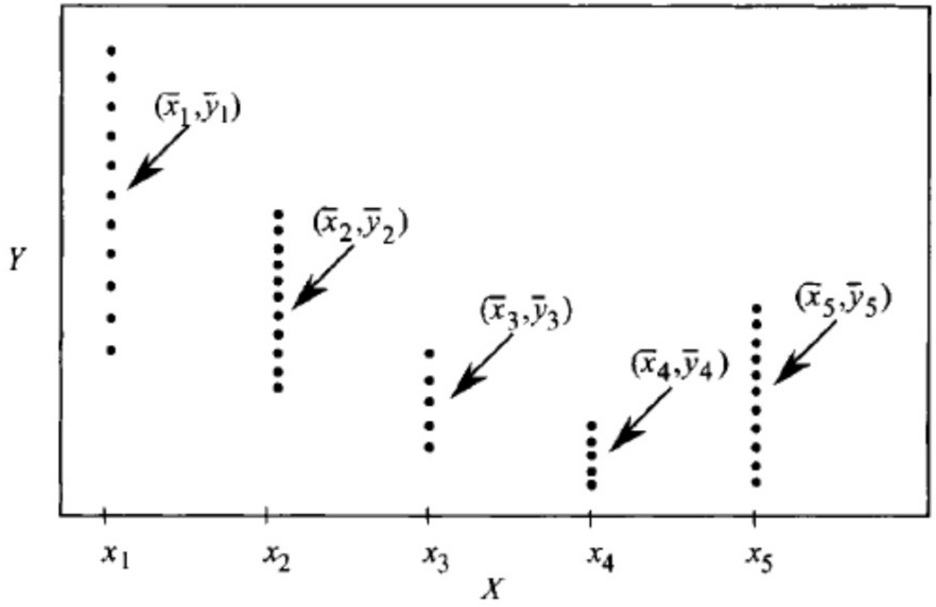
Weight known a priori

- Note that the procedure implicitly recognizes that observations from institutions where a large number of students were interviewed as more reliable and should have more weight in determining the regression coefficients than observations from institutions where only a few students were interviewed.
- Estimation can be performed by using WLS or by OLS on transforming variables:
- $\sqrt{n_i} y_i = \sqrt{n_i}\beta_0 + \sqrt{n_i}\beta_1 x_{i1} + \cdots + \sqrt{n_i}\beta_p x_{ip} + \sqrt{n_i}\epsilon_i$
- $\sqrt{n_i}\epsilon_i \sim N(0, \sigma^2)$
- Technical note: regard $\sqrt{n_i}$ as a new predictor and perform linear regression without intercept.
- In R, let y, x1, x2, n be vectors of data or weights, generate yy = sqrt(n)*y, xx0 = sqrt(n), xx1 = sqrt(n)*x1,), xx2 = sqrt(n)*x2. Here, * is the element-wise multiplication.
- lm(yy ~ 0 + xx0 + xx1 + xx2)

Weight known a priori

- This assumes y_{ij}'s are identically distributed: $Var(y_{ij}) = \sigma^2$ for all i, j.
- This is too restrictive to be realistic in many cases.
- One possible solution in experiment context is to replicate measurements on the response variable corresponding to a set of fixed values of the predictor variables. That is, the first column in the table is available.
- The variance of response variable for the ith group can be directly used in estimating the weight.
- $w_{ij} = 1/s_i^2$ where $s_i^2 = \sum_j \left(y_{ij} - \bar{y}_i\right)^2 / (n_i - 1)$

Nonconstant variance with replicated observations



Limitations of repeated measurement method

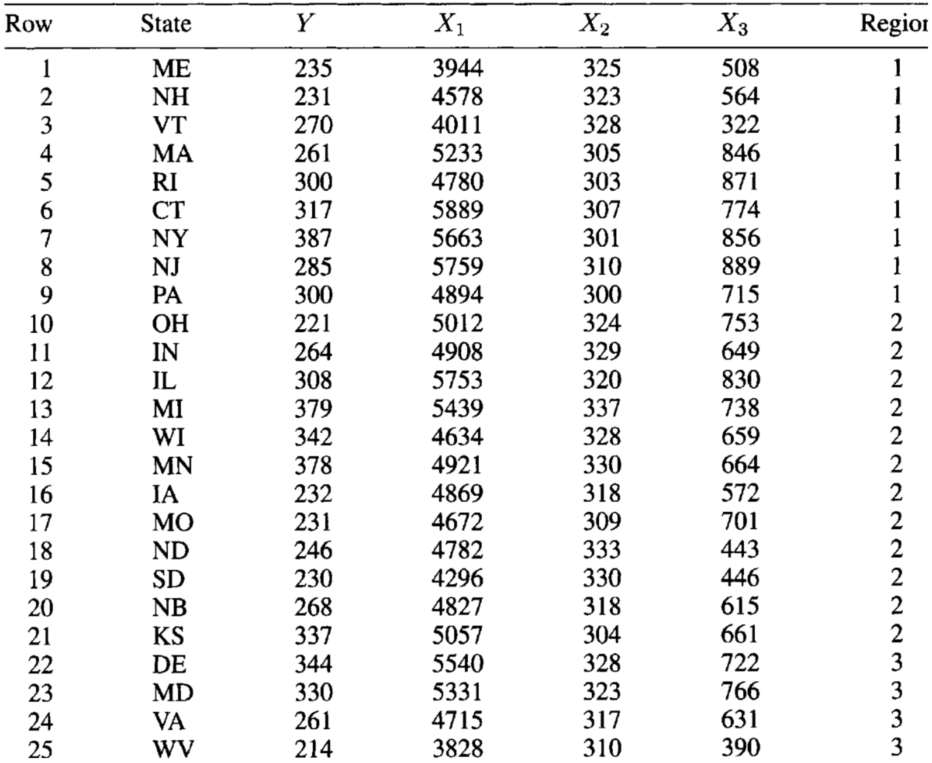
- When the data are collected in a controlled laboratory setting, the researcher can choose to replicate the observations at any values of the predictor variables. But in most observational studies, the presence of replications on the response variable for a given value of X is rather uncommon.
- When there is only one predictor variable, it is possible that some replications will occur. If there are many predictor variables, it is virtually impossible to imagine coming upon two observations with identical values on all predictor values.
- We seldomly are able to estimate variance of each Y directly using the repeated measurements.

Summary so far

- In the first simulated example, variance of error is proportional to some predictor, $\sigma_i^2 = \sigma_0^2 X_i$.
- In the second cluster survey example, the method of data collection indicates heteroscedasticity. And the variance of error is proportional to the sample size of each cluster, $Var(y_i) = \sigma^2/n_i$.
- In the third repeated measure example, the variance of error and weight can be estimated reliably estimated by replicates, $s_i^2 = \sum_j \left(y_{ij} - \bar{y}_i\right)^2 /(n_i - 1)$.
- In all 3 cases, heteroscedasticity was expected at the outset, and homogeneity of variance can be accomplished by a transformation. This transformation is constructed directly from the information in the raw data.
- What if we do not know the structure of heteroscedasticity in advance?
- How to decide a proper weight?

Education expenditure

- Y – Per capita expenditure on education in 1975
- X1 – per capita income in 1973
- X2 – number of residents per thousand under 18 years of age in 1974
- X3 – number of residents per thousand living in urban areas in 1970
- Each row represents a state, n = 50
- The data are grouped by geographical region: (1) Northeast, (2) North Central, (3) South, and (4) West.



Row	State	Y	X_1	X_2	X_3	Regior
1	ME	235	3944	325	508	1
2	NH	231	4578	323	564	1
3	VT	270	4011	328	322	1
4	MA	261	5233	305	846	1
5	RI	300	4780	303	871	1
6	CT	317	5889	307	774	1
7	NY	387	5663	301	856	1
8	NJ	285	5759	310	889	1
9	PA	300	4894	300	715	1
10	OH	221	5012	324	753	2
11	IN	264	4908	329	649	2
12	IL	308	5753	320	830	2
13	MI	379	5439	337	738	2
14	WI	342	4634	328	659	2
15	MN	378	4921	330	664	2
16	IA	232	4869	318	572	2
17	MO	231	4672	309	701	2
18	ND	246	4782	333	443	2
19	SD	230	4296	330	446	2
20	NB	268	4827	318	615	2
21	KS	337	5057	304	661	2
22	DE	344	5540	328	722	3
23	MD	330	5331	323	766	3
24	VA	261	4715	317	631	3
25	WV	214	3828	310	390	3

Objective and assumption

- The objective is to get the best representation of the relationship between expenditure on education and the other variables using data for all 50 states.
- Our assumption is that, although the relationship is structurally the same in each region, the coefficients and residual variances may differ from region to region.
- Note that we can use indicator variables to look for specific effects associated with the regions or to formulate tests for the equality of regressions across regions. However, our objective here is to develop ONE relationship that can serve as the best representation for all regions and all states.
- This goal is accomplished by taking regional differences into account through the method of WLS.

WLS scheme

- It is assumed that there is a unique residual variance associated with each of the four regions, $(c_1\sigma)^2$, $(c_2\sigma)^2$, $(c_3\sigma)^2$, $(c_4\sigma)^2$, where σ is the common part and the c_i's are unique to the regions.
- According to the principle of weighted least squares, the regression coefficients should be determined by minimizing $S = S_1 + S_2 + S_3 + S_4$, where $S_i = \sum_j \frac{1}{{c_i}^2}(y_j - x_j\beta)^2$, $i = 1, 2, 3, 4$.
- The factors $1/{c_i}^2$ are the weights that determine how much influence each observation has in estimating the regression coefficients.
- The weighting scheme is intuitively justified by arguing that observations with large error variance should have little influence in determining the coefficients.

WLS scheme

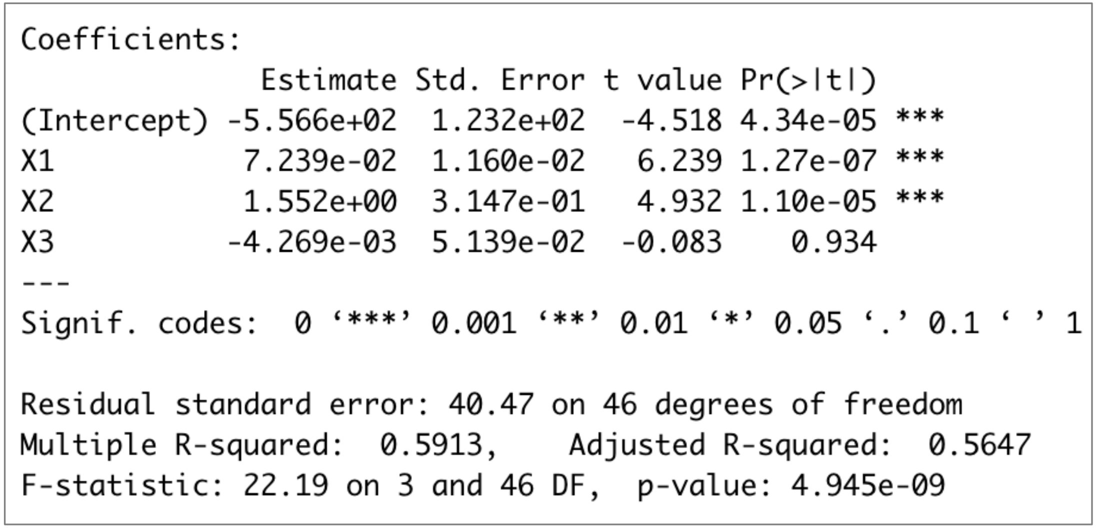
- Alternatively, WLS scheme can be justified by transforming variables.
- Transform the data such that we regress y/c_i on x/c_i.
- Then the transformed error term ϵ/c_i will have a common variance σ^2, and the estimated coefficients have all the standard least square properties.

WLS estimation procedure

- c_i's are unknown and must be estimated in the same sense that β's and σ^2 must be estimated.
- This can be done by a two-stage estimation procedure.
- Stage 1:
- Run OLS on raw data by ignoring the region information.
- Using empirical residuals grouped by region to compute an estimate of the reginal residual variance: $\hat{\sigma}_i^2 = \sum_j e_j^2 / (n_i - 1), i = 1, 2, 3, 4, j = 1, \cdots, n_i$.
- Stage 2:
- Estimate $\hat{c}_i^2 = \hat{\sigma}_i^2 / (\hat{\sigma})^2$, where $\hat{\sigma} = \sum_{i,j} e_{ij}^2 / n, n = \sum_i n_i$.
- Apply WLS using weights $1/\hat{c}_i^2$.

Stage 1: OLS

```
fit1 = lm(Y ~ X1 + X2 + X3, data = dat)
summary(fit1)
```

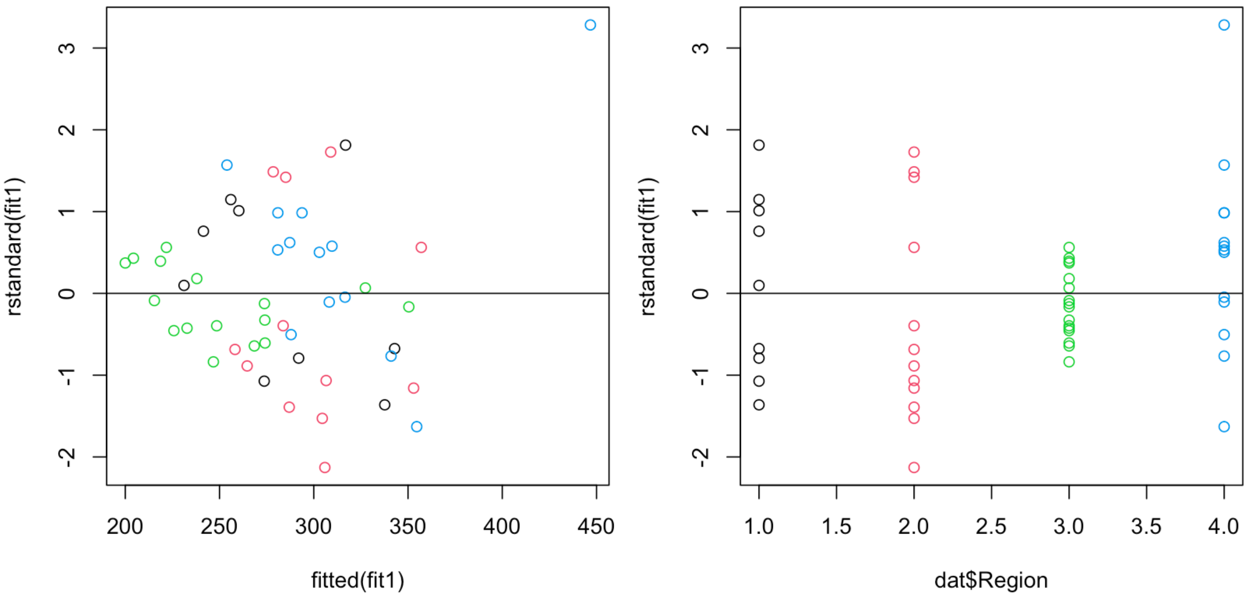


```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.566e+02  1.232e+02  -4.518 4.34e-05 ***
X1           7.239e-02  1.160e-02   6.239 1.27e-07 ***
X2           1.552e+00  3.147e-01   4.932 1.10e-05 ***
X3          -4.269e-03  5.139e-02  -0.083    0.934    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.47 on 46 degrees of freedom
Multiple R-squared:  0.5913,    Adjusted R-squared:  0.5647 
F-statistic: 22.19 on 3 and 46 DF,  p-value: 4.945e-09
```

Stage 1: OLS

- The residual vs. fitted value plot has a funnel shape, indicating heteroscedasticity.
- The spread of residuals in the right figure is different across regions, which also indicates unequal variance.

1-Northeast, 2-North Central, 3-South, 4-West

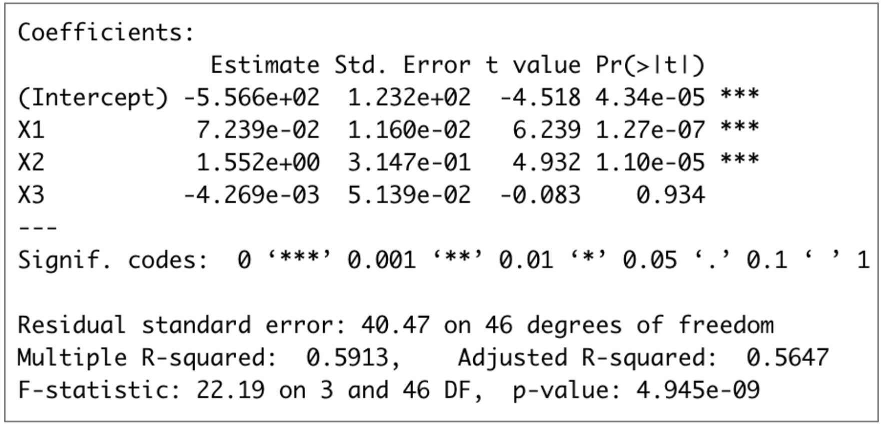


Stage 2: WLS

	n_i	$\hat{\sigma}_i^2$	$\hat{c}_i^2$	w_i
Region 1	9	1936	1.28	0.61
Region 2	12	2871	1.91	0.28
Region 3	16	299	0.20	25.37
Region 4	13	1507	1.32	0.58

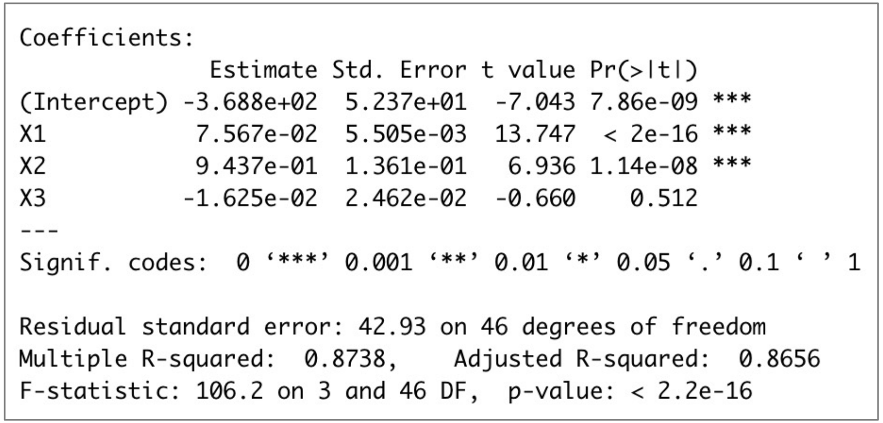
Stage 2: WLS

OLS



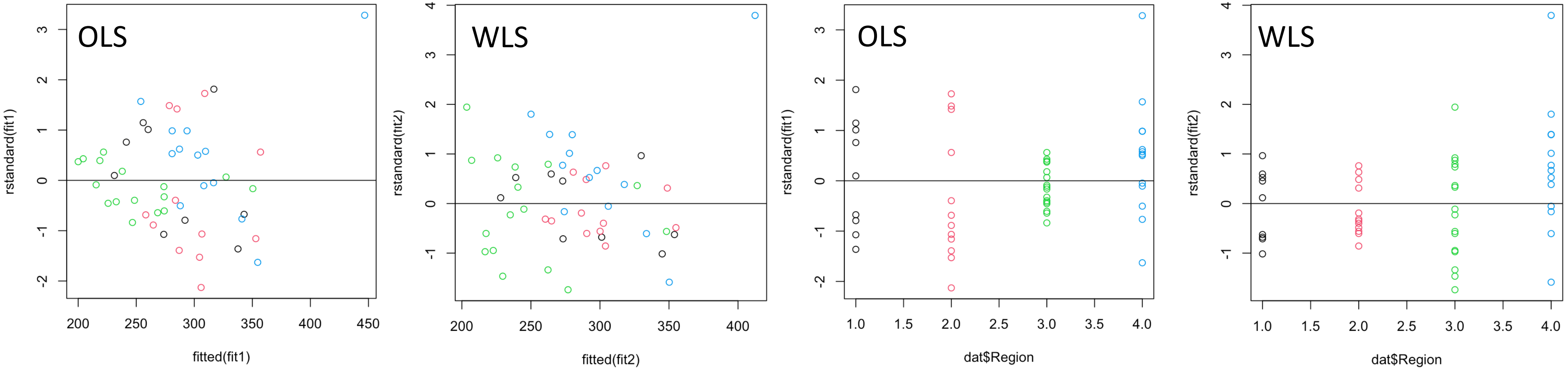
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.566e+02  1.232e+02  -4.518 4.34e-05 ***
X1           7.239e-02  1.160e-02   6.239 1.27e-07 ***
X2           1.552e+00  3.147e-01   4.932 1.10e-05 ***
X3          -4.269e-03  5.139e-02  -0.083    0.934    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.47 on 46 degrees of freedom
Multiple R-squared:  0.5913,    Adjusted R-squared:  0.5647 
F-statistic: 22.19 on 3 and 46 DF,  p-value: 4.945e-09
```

WLS

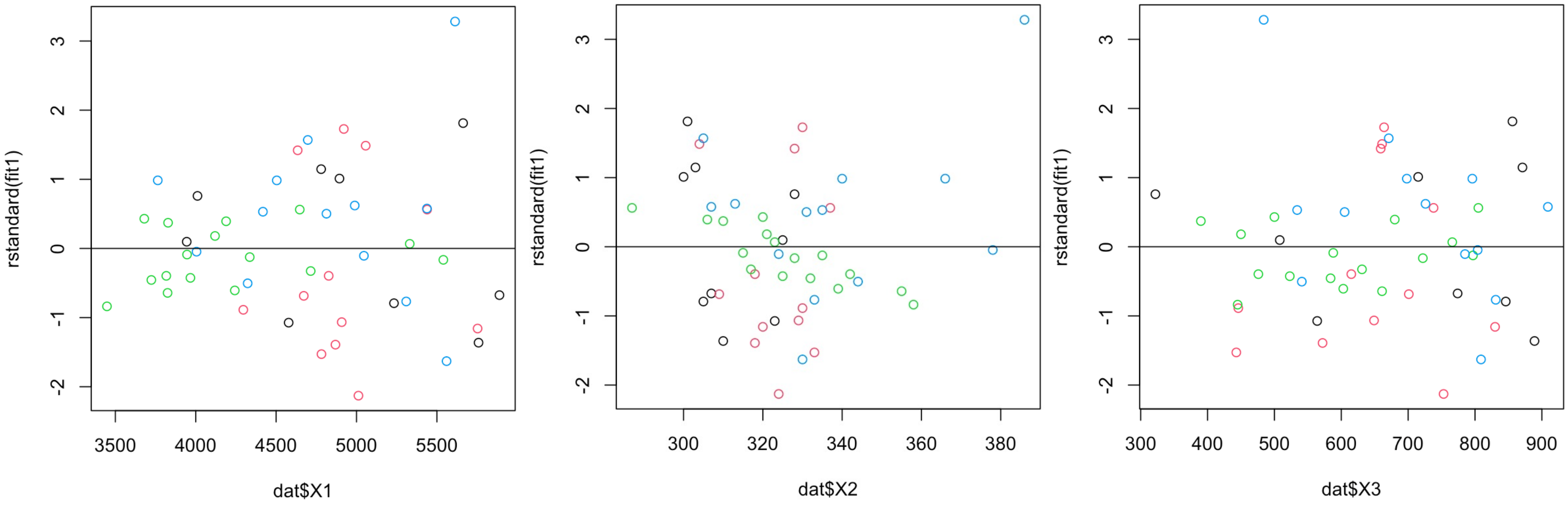


```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.688e+02  5.237e+01  -7.043 7.86e-09 ***
X1           7.567e-02  5.505e-03  13.747  < 2e-16 ***
X2           9.437e-01  1.361e-01   6.936 1.14e-08 ***
X3          -1.625e-02  2.462e-02  -0.660    0.512    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42.93 on 46 degrees of freedom
Multiple R-squared:  0.8738,    Adjusted R-squared:  0.8656 
F-statistic: 106.2 on 3 and 46 DF,  p-value: < 2.2e-16
```

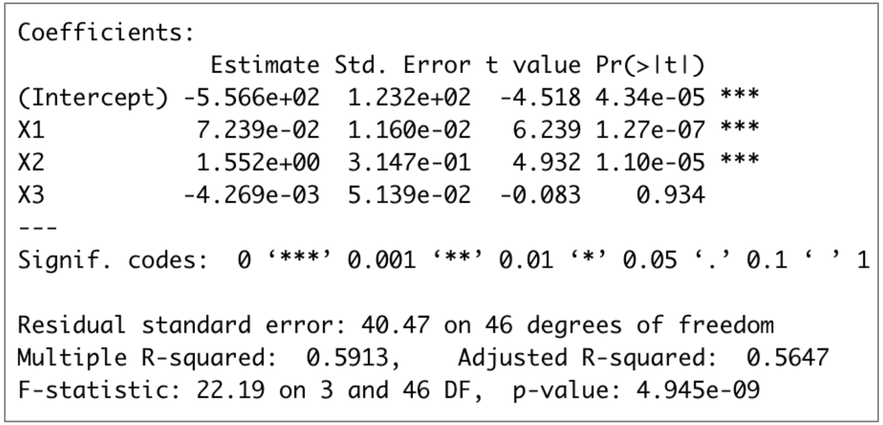
Stage 2: WLS


OLS
rstandard(fit1)
-2
-1
0
1
2
3
200
250
300
350
400
450
fitted(fit1)
WLS
rstandard(fit2)
-1
0
1
2
3
4
200
250
300
350
400
fitted(fit2)
OLS
rstandard(fit1)
-2
-1
0
1
2
3
1.0
1.5
2.0
2.5
3.0
3.5
4.0
dat$Region
WLS
rstandard(fit2)
-1
0
1
2
3
4
1.0
1.5
2.0
2.5
3.0
3.5
4.0
dat$Region

If we do not use region information

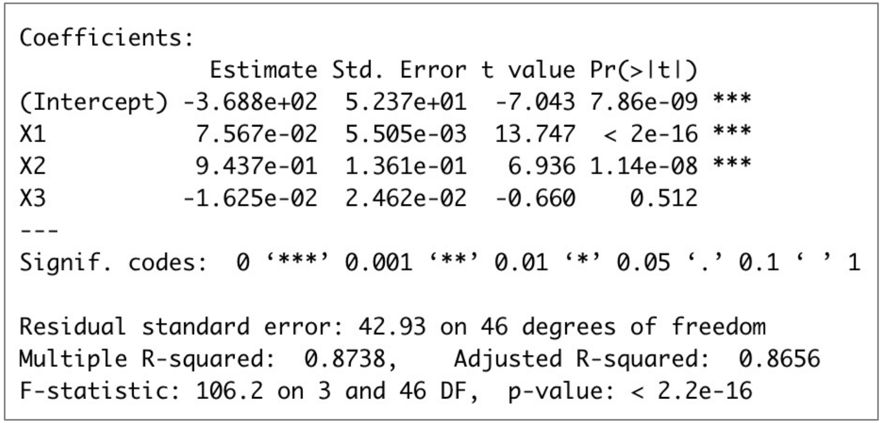


OLS



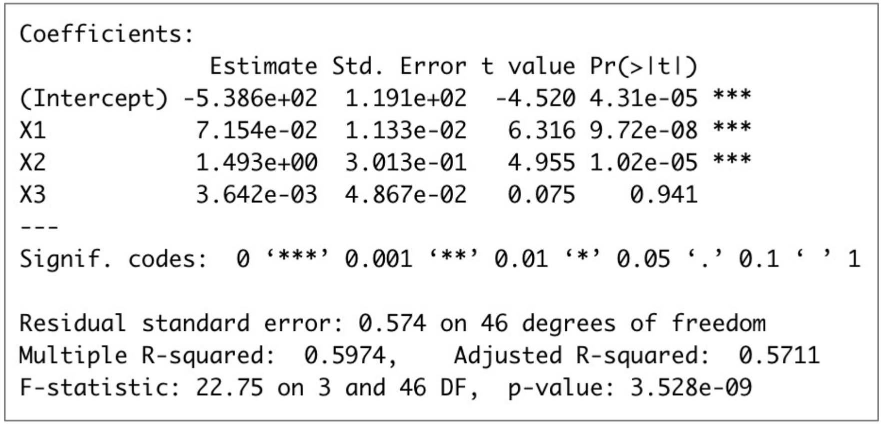
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.566e+02  1.232e+02  -4.518 4.34e-05 ***
X1           7.239e-02  1.160e-02   6.239 1.27e-07 ***
X2           1.552e+00  3.147e-01   4.932 1.10e-05 ***
X3          -4.269e-03  5.139e-02  -0.083    0.934    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.47 on 46 degrees of freedom
Multiple R-squared:  0.5913,	Adjusted R-squared:  0.5647 
F-statistic: 22.19 on 3 and 46 DF,  p-value: 4.945e-09
```

WLS – use region

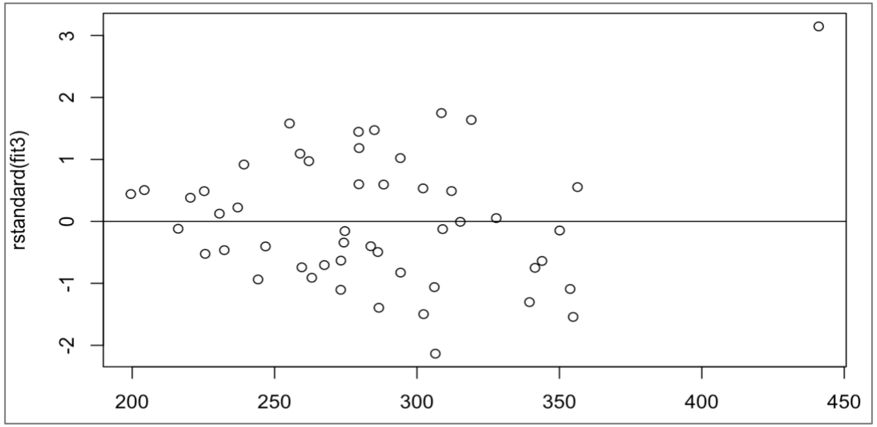


```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.688e+02  5.237e+01  -7.043 7.86e-09 ***
X1           7.567e-02  5.505e-03  13.747  < 2e-16 ***
X2           9.437e-01  1.361e-01   6.936 1.14e-08 ***
X3          -1.625e-02  2.462e-02  -0.660    0.512    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42.93 on 46 degrees of freedom
Multiple R-squared:  0.8738,	Adjusted R-squared:  0.8656 
F-statistic: 106.2 on 3 and 46 DF,  p-value: < 2.2e-16
```

WLS – use 1/X1 as weight, not use region



```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.386e+02  1.191e+02  -4.520 4.31e-05 ***
X1           7.154e-02  1.133e-02   6.316 9.72e-08 ***
X2           1.493e+00  3.013e-01   4.955 1.02e-05 ***
X3           3.642e-03  4.867e-02   0.075    0.941    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.574 on 46 degrees of freedom
Multiple R-squared:  0.5974,	Adjusted R-squared:  0.5711 
F-statistic: 22.75 on 3 and 46 DF,  p-value: 3.528e-09
```

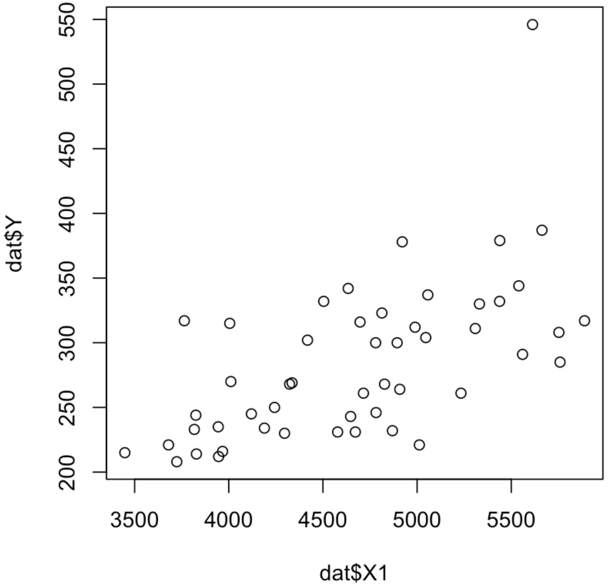


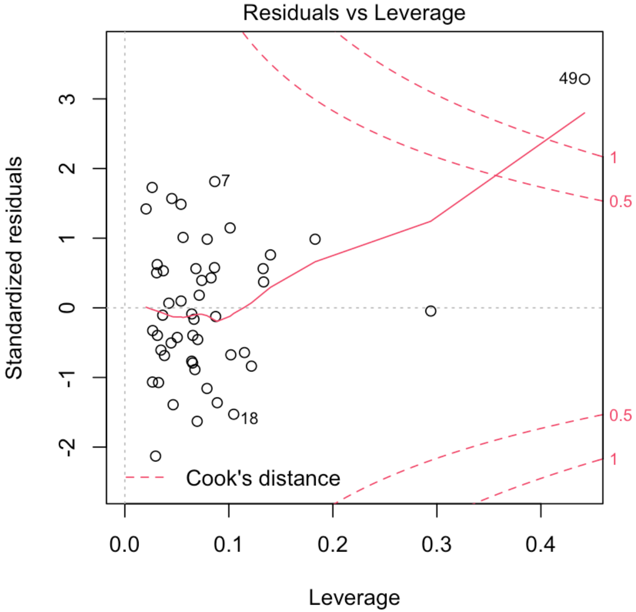
Summary for the education expense example

- Detection of heteroscedasticity in multi-variate regression is not a simple task.
- If heteroscedasticity is present, it is often discovered as a result of some good intuition on the part of the analyst on how observations may be grouped or clustered (e.g., regions).
- For multiple regression models, the plots of standardized residuals vs. fitted values and predictors can serve as the first step.
- This plot, however, does not necessarily indicate why the variance differ (source of unequal variance).

Outlier

- Observation 49 (AK) is an outlier with high standardized residual of 3.28.
- It is also influential with Cook's distance of 2.13.
- Compared to other states, Alaska represents a very special situation that has considerable influence on the regression results.
- We exclude AK in the following analysis.





All 50 states

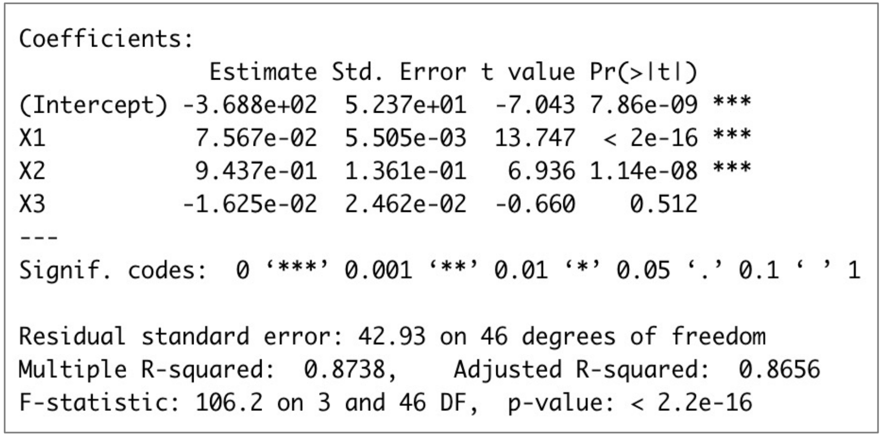
	n_i	$\hat{\sigma}_i^2$	$\hat{c}_i^2$	w_i
Region 1	9	1936	1.28	0.61
Region 2	12	2871	1.91	0.28
Region 3	16	299	0.20	25.37
Region 4	13	1507	1.32	0.58

49 states (exclude AK)

	n_i	$\hat{\sigma}_i^2$	$\hat{c}_i^2$	w_i
Region 1	9	1633	1.39	0.52
Region 2	12	2659	2.26	0.20
Region 3	16	266	0.23	19.59
Region 4	12	1178	0.88	1.29

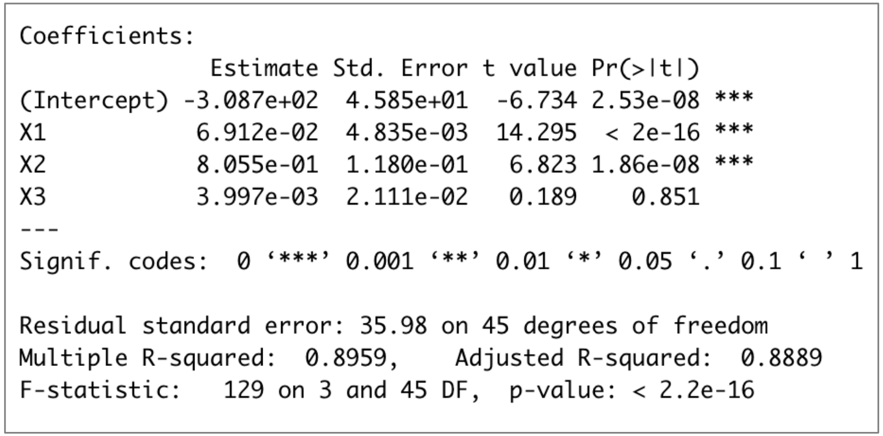
WLS results

WLS, 50 states



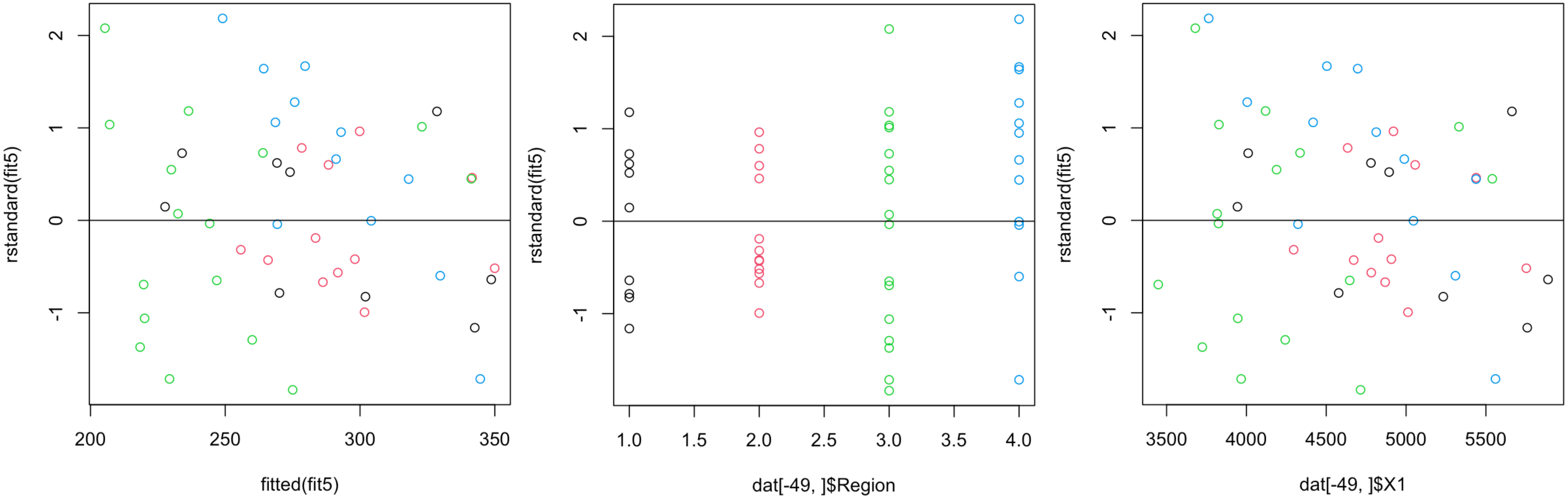
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.688e+02  5.237e+01  -7.043 7.86e-09 ***
X1           7.567e-02  5.505e-03  13.747  < 2e-16 ***
X2           9.437e-01  1.361e-01   6.936 1.14e-08 ***
X3          -1.625e-02  2.462e-02  -0.660    0.512    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42.93 on 46 degrees of freedom
Multiple R-squared:  0.8738,	Adjusted R-squared:  0.8656 
F-statistic: 106.2 on 3 and 46 DF,  p-value: < 2.2e-16
```

WLS, 49 states (exclude AK)



```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.087e+02  4.585e+01  -6.734 2.53e-08 ***
X1           6.912e-02  4.835e-03  14.295  < 2e-16 ***
X2           8.055e-01  1.180e-01   6.823 1.86e-08 ***
X3           3.997e-03  2.111e-02   0.189    0.851    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.98 on 45 degrees of freedom
Multiple R-squared:  0.8959,	Adjusted R-squared:  0.8889 
F-statistic:   129 on 3 and 45 DF,  p-value: < 2.2e-16
```

Unequal variance is "corrected"



Generalized least squares (GLS)

- Relax the traditional assumptions on error term further.
- Error terms between individuals may be correlated.
- WLS is a special case of GLS.
- Covariance matrix of ϵ:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \cdots & \sigma_{n1} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

Take-away messages

- Some key points regarding weighted least squares are:
- Point estimate of the coefficients from WLS will usually be nearly the same as OLS (in theory, they are both unbiased and consistent).
- The use of weights will (legitimately) impact the widths of statistical intervals and efficiency of inference .
- The difficulty, in practice, is determining estimates of the error variances.
- In designed experiments with large numbers of replicates, weights can be estimated directly from sample variances of the response variable at each combination of predictor variables.
- Use theory or prior research results to instruct the weights construction.
- Weights can be estimated empirically. But this is not a easy task especially in the multiple regression.

Reference

- Suarez E, Perez CM, Rivera R, & Martinez MN Application of Regression Models in Epidemiology, 2017, John Wiley and Sons, Hoboken, NJ
- Chatterjee & Hadi. Regression Analysis by Example, 5th Edition 2005, Wiley Interscience, New York.
- https://ms.mcmaster.ca/canty/teaching/stat3a03/Lectures7.pdf
- https://online.stat.psu.edu/stat501/lesson/13
- https://hedibert.org/wp-content/uploads/2016/10/heteroskedasticity.pdf