Extra: weighted regression

CS 6140
Machine Learning
Professor Olga Vitek

March 16, 2017

A Linear Model

- Consider $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Sigma}$

  where

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

  - heterogeneous variance or *heteroschedastcity*
  - (before: homogeneous variance or *homoschedasticity*)

- A linear model, but unequal variances

  - transformation of $\mathbf{X}$ or $\mathbf{Y}$ alone may unduly affect the relationship between $\mathbf{X}$ and $\mathbf{Y}$
  - error variance is often a function of $\mathbf{X}$ or $\mathbf{Y}$
  - can incorporate correlated errors

A Transformation Approach

- Suppose $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{W}^{-1}$

  where

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

  - notation change from $\boldsymbol{\Sigma}$ to $\mathbf{W}$

- Consider a transformation based on $\mathbf{W}$

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

- Can show $\mathrm{E}(\boldsymbol{\varepsilon}_w) = 0$ and $\boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}_w) = \sigma^2\mathbf{I}$

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

Weighted Least Squares

$$\mathbf{W}^{1/2}\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{1/2}\boldsymbol{\varepsilon} \quad \rightarrow \quad \mathbf{Y}_w = \mathbf{X}_w\boldsymbol{\beta} + \boldsymbol{\varepsilon}_w$$

- The least squares method minimizes

$$\begin{aligned} Q_w &= (\mathbf{Y}_w - \mathbf{X}_w\boldsymbol{\beta})'(\mathbf{Y}_w - \mathbf{X}_w\boldsymbol{\beta}) \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

- By taking a derivative of Q_w, obtain normal equations:

$$(\mathbf{X}_w'\mathbf{X}_w)\mathbf{b} = \mathbf{X}_w'\mathbf{Y}_w \quad \rightarrow \quad (\mathbf{X}'\mathbf{W}\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{W}\mathbf{Y}$$

- Solution of the normal equations:
  $(\mathbf{X}_w'\mathbf{X}_w)^{-1}\mathbf{X}_w'\mathbf{Y}_w \quad \rightarrow \quad b = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$

- Difficulty: need to determine the weight matrix $\mathbf{W}$ in advance

Alternative View: Maximum Likelihood

- Consider

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

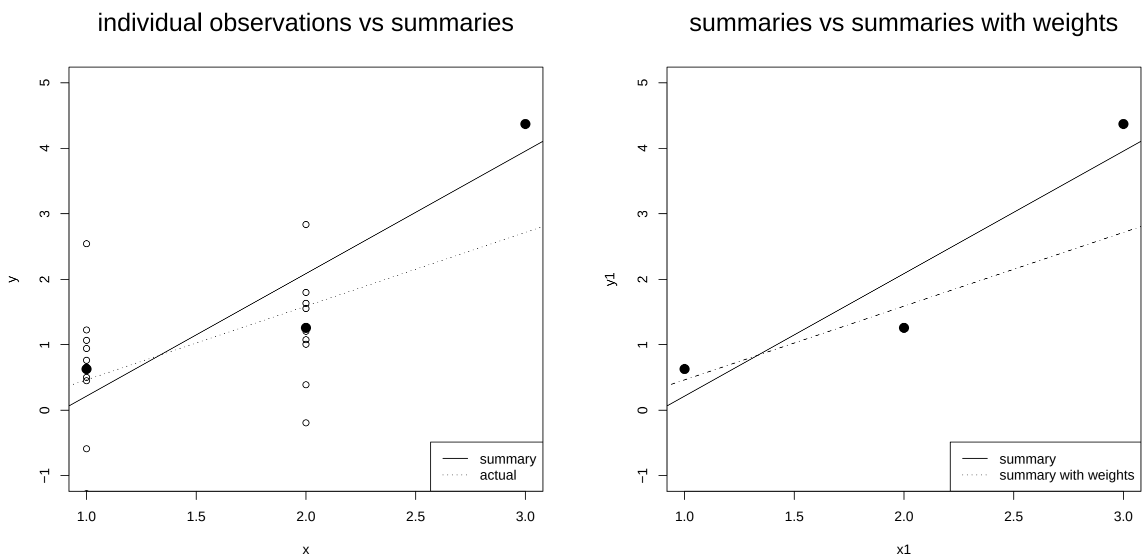
- Likelihood function $L = f_1 \times f_2 \times \cdots \times f_n$
  - if σ_i^2 is known, maximize wrt $\boldsymbol{\beta}$

- Same as minimizing

$$\begin{aligned} Q_w &= \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(Y_i - \mathbf{X}_i'\boldsymbol{\beta}))^2 \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

- Difficulty: need to determine the weight matrix $\mathbf{W}$ in advance

Weighted Regression: Known Weights




```
--------------Full dataset-----------------
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.6635     0.6364  -1.043  0.31024
x             1.1262     0.3797   2.966  0.00793 **

------------Unweighted summary -----------
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.6572     1.5492   -1.07    0.479
x1            1.8715     0.7171    2.61    0.233

------------Weighted summary -------------
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.6635     1.2721  -0.522    0.694
x1            1.1262     0.7590   1.484    0.378
```

Impact of Weights on b: Ex. in Simple Regression

Assume simple regression $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ where X has c levels

$$
\begin{aligned}
b_1 &= \frac{\sum\limits_{i=1}^{n} (X_i - \bar{X}_i) Y_i}{\sum\limits_{i=1}^{n} (X_i - \bar{X})^2} \\
&= \frac{\sum\limits_{j=1}^{c} (X_j - \bar{X}_j) \sum\limits_{k=1}^{n_j} Y_{jk}}{\sum\limits_{j=1}^{c} n_j (X_j - \bar{X})^2} \\
&= \frac{\sum\limits_{j=1}^{c} (X_j - \bar{X}_j) n_j \bar{Y}_j}{\sum\limits_{j=1}^{c} n_j (X_j - \bar{X})^2}
\end{aligned}
$$

How to Determine Unknown Weights?

- $\boldsymbol{\sigma}^2(\varepsilon)$ is often a function of $\mathbf{X}$
  - 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
  - Usually can estimate $\boldsymbol{\sigma}^2(\varepsilon)$ up to a constant
    * since residual variation in estimating $\boldsymbol{\sigma}^2(\varepsilon)$ versus $\mathbf{X}$ persists

- This changes the model slightly to $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ where $\boldsymbol{\sigma}^2(\varepsilon) = \sigma^2 \cdot \mathbf{W}^{-1}$

Example

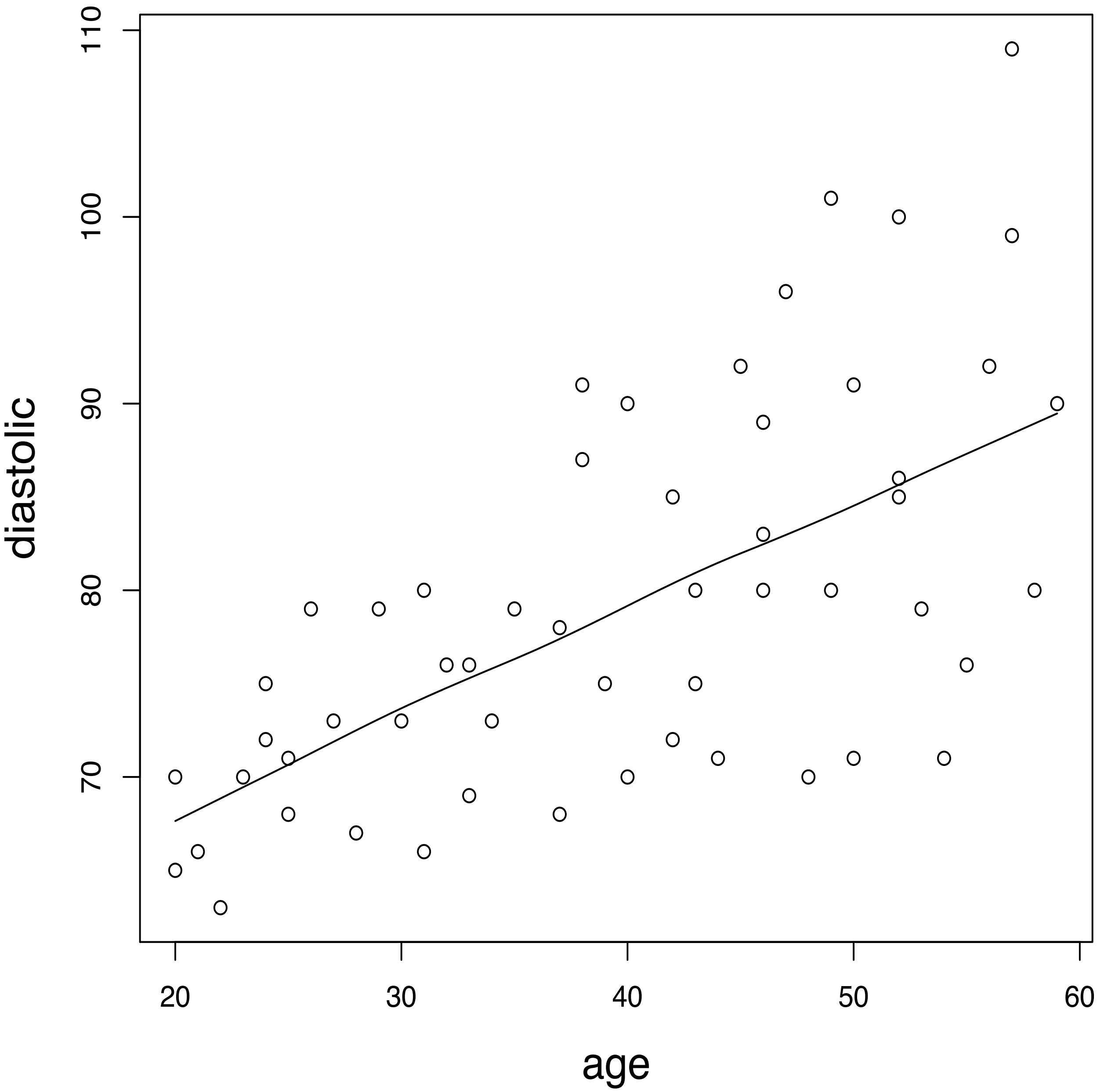
- Interested in the relationship between diastolic blood pressure and age

- 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

Scatterplot



Unweighted regression

```
> m.unweighted <- lm(diast ~ age, data=X)
> summary(m.unweighted)
...

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 56.15693    3.99367  14.061  < 2e-16 ***
age          0.58003    0.09695   5.983 2.05e-07 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 8.146 on 52 degrees of freedom
Multiple R-squared:  0.4077,Adjusted R-squared:  0.3963
F-statistic: 35.79 on 1 and 52 DF,  p-value: 2.05e-07
```

Weighted Regression

See option 'weights' in `lm` in R

```
> # Estimate the weights
> #---------------------
> w <- predict(lm(abs(m.unweighted$res) ~ X$age))

> # Fit weighted model
> #---------------------
> m.weighted <- lm(diast ~ age, data=X, weights=1/(w^2))
> summary(m.weighted)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 55.56577    2.52092  22.042  < 2e-16 ***
age          0.59634    0.07924   7.526 7.19e-10 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 1.213 on 52 degrees of freedom
Multiple R-squared:  0.5214,Adjusted R-squared:  0.5122
F-statistic: 56.64 on 1 and 52 DF,  p-value: 7.187e-10
```

*** Not much difference in the estimates but a slight reduction in the standard errors. Should not interpret R^2 in this situation.

Real difference would be seen in prediction intervals