# STAC67: Regression Analysis

Lecture 21

Thibault Randrianarisoa

November 25, 2026

University of Toronto Scarborough

# Review of Friday

- Search procedures (forward, backward, stepwise, `step( )`) move one predictor at a time; with correlated predictors they can disagree (the cement data).
- Validation: predict cases used neither to choose nor to fit the model, $MSPE = \frac{1}{n^*}\sum_{i=1}^{n^*}(Y_i - \hat{Y}_i)^2$, and compare it with the training $MSE$.
- Unusual cases: outlier in $Y$, the studentized deleted residual $t_i$ with Bonferroni; outlier in $X$, the leverage $h_{ii} > 2p/n$; influence, $DFFITS$, Cook's $D$, $DFBETAS$.
- Multicollinearity: $VIF_k = 1/(1 - R_k^2)$, serious above 10.

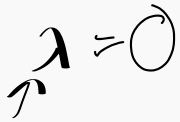
replace $\alpha$ with $\alpha/n$

# Weighted least squares

# Reminder on non-constant variance

Lecture 8: we observed a funnel in the residual plot, confirmed by the Brown-Forsythe test.

Two remedies:



- transform $Y$ (Box-Cox; $\ln Y$ in Lecture 18): this also changes the shape of the mean and the units of the answer;
- weighted least squares (new): keep $Y$ and the regression function, change how much each case counts.

# Reminder on non-constant variance

Lecture 8: we observed a funnel in the residual plot, confirmed by the Brown-Forsythe test.

Two remedies:

- transform $Y$ (Box-Cox; $\ln Y$ in Lecture 18): this also changes the shape of the mean and the units of the answer;
- weighted least squares (new): keep $Y$ and the regression function, change how much each case counts.

The model, with a variance for each case:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i, \qquad \epsilon_i \text{ independent } N(0, \sigma_i^2).$$

# Reminder on non-constant variance

Lecture 8: we observed a funnel in the residual plot, confirmed by the Brown-Forsythe test.

Two remedies:

- transform $Y$ (Box-Cox; $\ln Y$ in Lecture 18): this also changes the shape of the mean and the units of the answer;
- weighted least squares (new): keep $Y$ and the regression function, change how much each case counts.

The model, with a variance for each case:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i, \qquad \epsilon_i \text{ independent } N(0, \sigma_i^2).$$

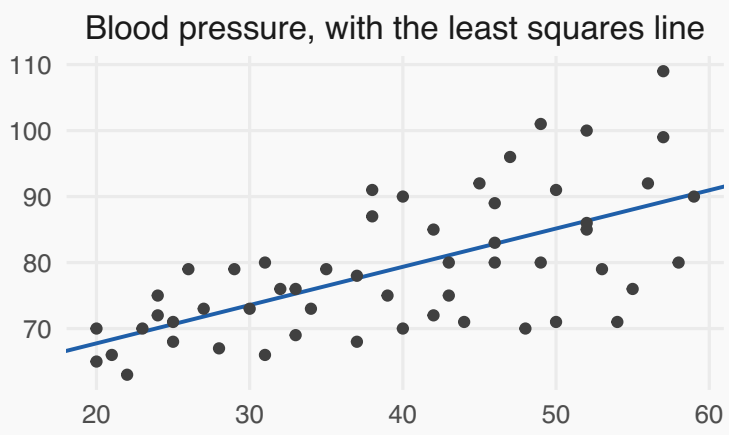
Ordinary least squares is still unbiased, but:

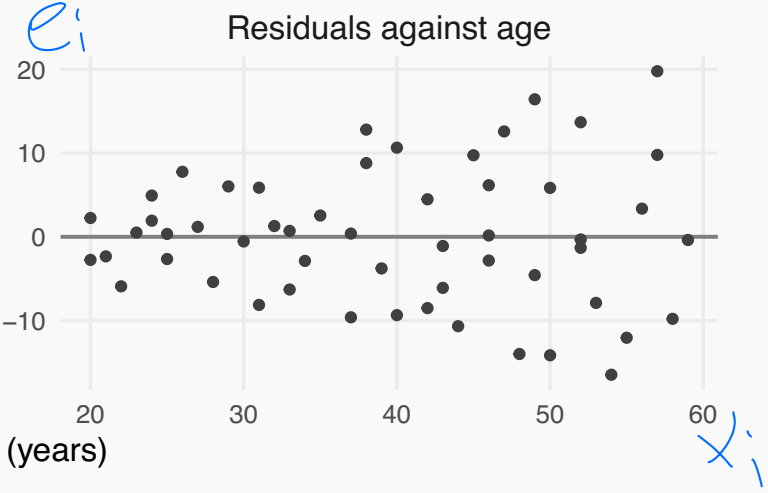
- it gives every case the same weight, the noisy ones included, so it is no longer the most precise estimator;
- its variance is no longer $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, so the standard errors R prints, from $MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$, are wrong, and with them the tests and intervals.

Gauss-Markov not valid anymore

# Example: blood pressure and age

54 healthy women aged 20 to 59: diastolic blood pressure $Y$ (mm Hg) against age $X$ (years).



# Example: blood pressure and age

54 healthy women aged 20 to 59: diastolic blood pressure $Y$ (mm Hg) against age $X$ (years).




The spread grows with age, about linearly. Brown-Forsythe, splitting at the median age (28 and 26 women): $t_{BF} = 2.79$, and $|t_{BF}| > t(0.975;\, 52) = 2.007$: reject equal variances.

# The weighted least squares criterion

Give case $i$ the weight $w_i = 1/\sigma_i^2$, and minimize

$$Q_w = \sum_{i=1}^{n} w_i \left(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1}\right)^2.$$

A precise case (small $\sigma_i^2$) gets a large weight: its residual costs more.

# The weighted least squares criterion

Give case $i$ the weight $w_i = 1/\sigma_i^2$, and minimize

$$Q_w = \sum_{i=1}^{n} w_i \left(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1}\right)^2.$$

A precise case (small $\sigma_i^2$) gets a large weight: its residual costs more.

Why this criterion: with normal errors, the likelihood (Lecture 3) is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{w_i}{2}\left(Y_i - \beta_0 - \cdots - \beta_{p-1} X_{i,p-1}\right)^2\right),$$

largest when $Q_w$ is smallest.

$$\log L(\beta) = -\frac{1}{2} Q_w(\beta) - \sum_{i=1}^{n} \log(2\pi\sigma_i^2)/2$$

# The weighted least squares criterion

Give case $i$ the weight $w_i = 1/\sigma_i^2$, and minimize

$$Q_w = \sum_{i=1}^{n} w_i \left(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1}\right)^2.$$

A precise case (small $\sigma_i^2$) gets a large weight: its residual costs more.

Why this criterion: with normal errors, the likelihood (Lecture 3) is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left( -\frac{w_i}{2} \left(Y_i - \beta_0 - \cdots - \beta_{p-1} X_{i,p-1}\right)^2 \right),$$

largest when $Q_w$ is smallest.

In matrix form, with $\boldsymbol{W} = \text{diag}(w_1, \ldots, w_n)$, $Q_w = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$, and

$$\hat{\boldsymbol{\beta}}_w = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Y}.$$

# Weighted least squares is ordinary least squares

Multiply case $i$ by $\sqrt{w_i}$: its response, and its row of $\boldsymbol{X}$ (the 1 included). With $\boldsymbol{W}^{1/2} = \text{diag}(\sqrt{w_1}, \ldots, \sqrt{w_n})$:

$$\boldsymbol{Y}^* = \boldsymbol{W}^{1/2}\boldsymbol{Y}, \qquad \boldsymbol{X}^* = \boldsymbol{W}^{1/2}\boldsymbol{X}, \qquad \boldsymbol{\epsilon}^* = \boldsymbol{W}^{1/2}\boldsymbol{\epsilon}, \qquad \boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*.$$

$\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$

$Y_i^* = \sqrt{w_i}\, Y_i \qquad X_{ip}^* = \sqrt{w_i}\, X_{ip}$

$\varepsilon_i^* \sim \mathcal{N}(0, \underbrace{w_i \sigma_i^2}_{=1})$

$Y = X\beta + \varepsilon \Rightarrow Y^* = W^{1/2} X \beta + \varepsilon^*$

$= X^* \beta + \varepsilon^*$

# Weighted least squares is ordinary least squares

Multiply case $i$ by $\sqrt{w_i}$: its response, and its row of $\boldsymbol{X}$ (the 1 included). With $\boldsymbol{W}^{1/2} = \text{diag}(\sqrt{w_1}, \ldots, \sqrt{w_n})$:

$$\boldsymbol{Y}^* = \boldsymbol{W}^{1/2}\boldsymbol{Y}, \qquad \boldsymbol{X}^* = \boldsymbol{W}^{1/2}\boldsymbol{X}, \qquad \boldsymbol{\epsilon}^* = \boldsymbol{W}^{1/2}\boldsymbol{\epsilon}, \qquad \boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*.$$

The new errors have variance 1, so Lectures 11 to 20 apply to $\boldsymbol{Y}^*$ and $\boldsymbol{X}^*$: $\hat{\boldsymbol{\beta}}_w$ is unbiased, with $\boldsymbol{\sigma}^2\{\hat{\boldsymbol{\beta}}_w\} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$; the $t$ and $F$ tests, the intervals and the diagnostics work as before.

$\sigma^2 (X^T X)^{-1} = (X^T W X)^{-1}$

# Variances known up to a constant

More often, only the relative sizes are known: $\sigma_i^2 = \sigma^2 / w_i$, with the $w_i$ known and $\sigma^2$ unknown.

we know $\sigma_i^2 / \sigma_j^2$, but not any of $\sigma_i^2$

# Variances known up to a constant

More often, only the relative sizes are known: $\sigma_i^2 = \sigma^2/w_i$, with the $w_i$ known and $\sigma^2$ unknown.

$$Y_i = \frac{1}{n_i}\sum_{R=1}^{n_i} z_{iR} \qquad Var(z_{iR}) = \sigma^2$$

**Example**

Each $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$: $Var(\bar{Y}_i) = \sigma^2/n_i$, so $w_i = n_i$.

# Variances known up to a constant

More often, only the relative sizes are known: $\sigma_i^2 = \sigma^2/w_i$, with the $w_i$ known and $\sigma^2$ unknown.

**Example**

Each $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$: $Var(\bar{Y}_i) = \sigma^2/n_i$, so $w_i = n_i$.

- The constant does not matter for $\hat{\boldsymbol{\beta}}_w$: weights $c\,w_i$ multiply $Q_w$ by $c$, with the same minimum.
- $\boldsymbol{\sigma}^2\{\hat{\boldsymbol{\beta}}_w\} = \sigma^2(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$, estimated by $MSE_w\,(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$, with

$$MSE_w = \frac{\sum_i w_i e_i^2}{n-p}, \qquad e_i = Y_i - \hat{Y}_i.$$

# Variances known up to a constant

More often, only the relative sizes are known: $\sigma_i^2 = \sigma^2/w_i$, with the $w_i$ known and $\sigma^2$ unknown.

**Example**

Each $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$: $Var(\bar{Y}_i) = \sigma^2/n_i$, so $w_i = n_i$.

- The constant does not matter for $\hat{\boldsymbol{\beta}}_w$: weights $c\,w_i$ multiply $Q_w$ by $c$, with the same minimum.
- $\boldsymbol{\sigma}^2\{\hat{\boldsymbol{\beta}}_w\} = \sigma^2(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$, estimated by $MSE_w\,(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$, with

$$MSE_w = \frac{\sum_i w_i e_i^2}{n-p}, \qquad e_i = Y_i - \hat{Y}_i.$$

In R: `lm(y ~ x, weights = w)`. Then `summary`, `confint` and `anova` use the weights, and the "residual standard error" is $\sqrt{MSE_w}$.

# Example: a cholesterol drug

A dose-response study in patients with high cholesterol: for each of six doses, only the mean change in cholesterol (%) after 12 weeks, over the $n_i$ patients of that dose.

| Dose (mg) | 1 | 2.5 | 5 | 10 | 20 | 40 |
|---|---|---|---|---|---|---|
| $n_i$ | 15 | 17 | 12 | 14 | 18 | 13 |
| $\bar{Y}_i$ (%) | -35.8 | -45.0 | -52.7 | -49.7 | -58.2 | -66.0 |

```
fitw <- lm(ldl ~ log(dose), data = chol, weights = n)
```

$(n_1, \ldots, n_6)$

# Example: a cholesterol drug

A dose-response study in patients with high cholesterol: for each of six doses, only the mean change in cholesterol (%) after 12 weeks, over the $n_i$ patients of that dose.

| Dose (mg) | 1 | 2.5 | 5 | 10 | 20 | 40 |
|---|---|---|---|---|---|---|
| $n_i$ | 15 | 17 | 12 | 14 | 18 | 13 |
| $\bar{Y}_i$ (%) | -35.8 | -45.0 | -52.7 | -49.7 | -58.2 | -66.0 |

```
fitw <- lm(ldl ~ log(dose), data = chol, weights = n)
```

| | $b_0$ | $s\{b_0\}$ | $b_1$ | $s\{b_1\}$ |
|---|---|---|---|---|
| Ordinary | -36.99 | 2.38 | -7.42 | 1.04 |
| Weighted, $w_i = n_i$ | -36.96 | 2.24 | -7.38 | 0.99 |

# Example: a cholesterol drug

A dose-response study in patients with high cholesterol: for each of six doses, only the mean change in cholesterol (%) after 12 weeks, over the $n_i$ patients of that dose.

| Dose (mg) | 1 | 2.5 | 5 | 10 | 20 | 40 |
|---|---|---|---|---|---|---|
| $n_i$ | 15 | 17 | 12 | 14 | 18 | 13 |
| $\bar{Y}_i$ (%) | -35.8 | -45.0 | -52.7 | -49.7 | -58.2 | -66.0 |

```
fitw <- lm(ldl ~ log(dose), data = chol, weights = n)
```

| | $b_0$ | $s\{b_0\}$ | $b_1$ | $s\{b_1\}$ |
|---|---|---|---|---|
| Ordinary | -36.99 | 2.38 | -7.42 | 1.04 |
| Weighted, $w_i = n_i$ | -36.96 | 2.24 | -7.38 | 0.99 |

The $n_i$ are close (12 to 18), so the fit barely moves. Doubling the dose lowers the mean change by about $7.38 \ln 2 = 5.1$ percentage points. $\sqrt{MSE_w} = 11.6$ estimates $\sigma$, the standard deviation for one patient.

# Which weights?

Give $w_i$, up to a constant.

1. $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$.

$$Y_i = \frac{1}{n_i} \sum_{R=1}^{n_i} Z_{iR} \to \mathrm{Var}(X_i) = \frac{\sigma^2}{n_i} \qquad w_i = n_i$$

2. $Y_i$ is the total of $n_i$ such measures.

$$Y_i = \sum_{R=1}^{n_i} Z_{iR} \to \mathrm{Var}(Y_i) = n_i \sigma^2 \qquad w_i = 1/n_i$$

3. The standard deviation of $\epsilon_i$ is proportional to $X_i > 0$: a funnel that starts at the origin.

$$\mathrm{Var}(Y_i) = c X_i^2 = \sigma^2 / w_i \to w_i \propto \frac{1}{X_i^2}$$

# Which weights?

Give $w_i$, up to a constant.

1. $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$.

   $w_i = n_i$.

2. $Y_i$ is the total of $n_i$ such measures.

3. The standard deviation of $\epsilon_i$ is proportional to $X_i > 0$: a funnel that starts at the origin.

# Which weights?

Give $w_i$, up to a constant.

1. $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$.

   $w_i = n_i$.

2. $Y_i$ is the total of $n_i$ such measures.

   $Var(Y_i) = n_i\sigma^2$, so $w_i = 1/n_i$.

3. The standard deviation of $\epsilon_i$ is proportional to $X_i > 0$: a funnel that starts at the origin.

# Which weights?

Give $w_i$, up to a constant.

1. $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$.

   $w_i = n_i$.

2. $Y_i$ is the total of $n_i$ such measures.

   $Var(Y_i) = n_i\sigma^2$, so $w_i = 1/n_i$.

3. The standard deviation of $\epsilon_i$ is proportional to $X_i > 0$: a funnel that starts at the origin.

   $\sigma_i = cX_i$, so $w_i = 1/X_i^2$.

# Which weights?

Give $w_i$, up to a constant.

1. $Y_i$ is the mean of $n_i$ measures of variance $\sigma^2$.

   $w_i = n_i$.

2. $Y_i$ is the total of $n_i$ such measures.

   $Var(Y_i) = n_i\sigma^2$, so $w_i = 1/n_i$.

3. The standard deviation of $\epsilon_i$ is proportional to $X_i > 0$: a funnel that starts at the origin.

   $\sigma_i = cX_i$, so $w_i = 1/X_i^2$.

In the third case, $Q_w = \sum_i \left(Y_i/X_i - \beta_0/X_i - \beta_1\right)^2$: ordinary least squares of $Y_i/X_i$ on $1/X_i$, where the intercept and the slope swap roles.

# Variances estimated from the residuals

Usually the $\sigma_i^2$ are unknown. So we estimate how the spread changes, from a first fit:

1. Fit by ordinary least squares, and plot the residuals.
2. A funnel: regress $|e_i|$ on the predictor (or on $\hat{Y}_i$). The fitted values $\hat{s}_i$ estimate $\sigma_i$: the standard deviation function. Or regress $e_i^2$, for the variance function $\hat{v}_i$.
3. Take $w_i = 1/\hat{s}_i^2$ (or $1/\hat{v}_i$).
4. Refit with these weights.
5. If the coefficients changed a lot, repeat steps 2 to 4 with the new residuals.

# Variances estimated from the residuals

Usually the $\sigma_i^2$ are unknown. So we estimate how the spread changes, from a first fit:

1. Fit by ordinary least squares, and plot the residuals.
2. A funnel: regress $|e_i|$ on the predictor (or on $\hat{Y}_i$). The fitted values $\hat{s}_i$ estimate $\sigma_i$: the standard deviation function. Or regress $e_i^2$, for the variance function $\hat{v}_i$.
3. Take $w_i = 1/\hat{s}_i^2$ (or $1/\hat{v}_i$).
4. Refit with these weights.
5. If the coefficients changed a lot, repeat steps 2 to 4 with the new residuals.

The weights are estimated, so the tests and intervals are approximate: fine unless $n$ is small. Check that every $\hat{s}_i$ is positive.

# Blood pressure: in R

The 54 women of slide 3, in the data frame `bp`: Y is the diastolic blood pressure (mm Hg), X the age (years).

```
fit  <- lm(Y ~ X, data = bp)
# the standard deviation function, then the weights
sfit <- lm(abs(resid(fit)) ~ X, data = bp)
fitw <- lm(Y ~ X, data = bp, weights = 1 / fitted(sfit)^2)
```

# Blood pressure: in R

The 54 women of slide 3, in the data frame `bp`: Y is the diastolic blood pressure (mm Hg), X the age (years).

```
fit  <- lm(Y ~ X, data = bp)
# the standard deviation function, then the weights
sfit <- lm(abs(resid(fit)) ~ X, data = bp)
fitw <- lm(Y ~ X, data = bp, weights = 1 / fitted(sfit)^2)
```

$\hat{s} = -1.55 + 0.198\,X$: from 2.4 at age 20 to 10.1 at age 59. An older woman gets about $(10.1/2.4)^2 \approx 18$ times less weight.

# Blood pressure: in R

The 54 women of slide 3, in the data frame `bp`: Y is the diastolic blood pressure (mm Hg), X the age (years).

```
fit  <- lm(Y ~ X, data = bp)
# the standard deviation function, then the weights
sfit <- lm(abs(resid(fit)) ~ X, data = bp)
fitw <- lm(Y ~ X, data = bp, weights = 1 / fitted(sfit)^2)
```

$\hat{s} = -1.55 + 0.198\,X$: from 2.4 at age 20 to 10.1 at age 59. An older woman gets about $(10.1/2.4)^2 \approx 18$ times less weight.

| | $b_1$ | $s\{b_1\}$ | 95% interval for $\beta_1$ |
|---|---|---|---|
| Ordinary | 0.580 | 0.0970 | (0.385, 0.775) |
| Weighted | 0.596 | 0.0792 | (0.437, 0.755) |

# Blood pressure: in R

The 54 women of slide 3, in the data frame `bp`: Y is the diastolic blood pressure (mm Hg), X the age (years).

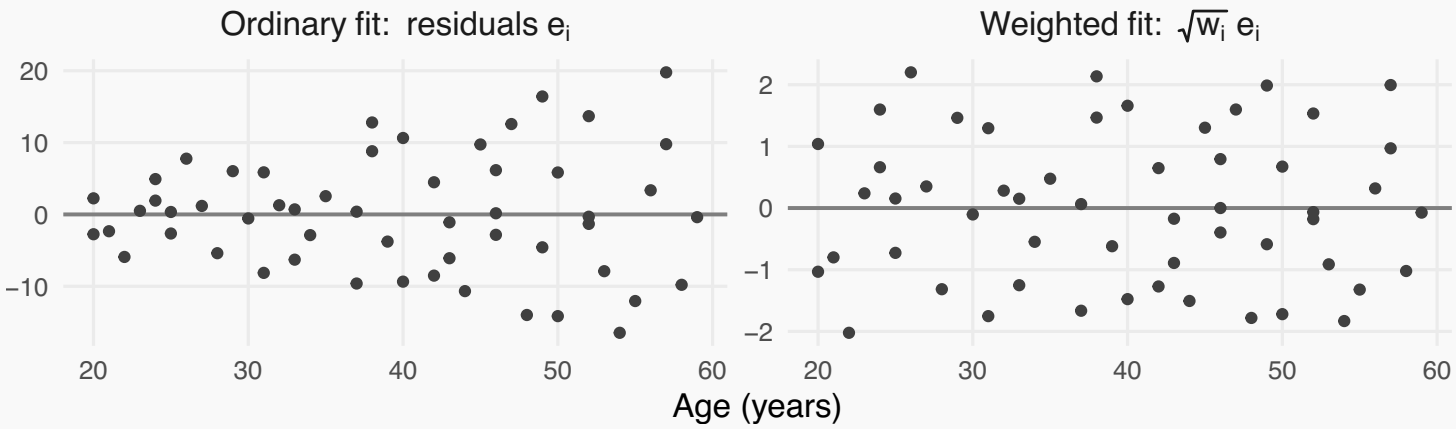
```
fit  <- lm(Y ~ X, data = bp)
# the standard deviation function, then the weights
sfit <- lm(abs(resid(fit)) ~ X, data = bp)
fitw <- lm(Y ~ X, data = bp, weights = 1 / fitted(sfit)^2)
```

$\hat{s} = -1.55 + 0.198\,X$: from 2.4 at age 20 to 10.1 at age 59. An older woman gets about $(10.1/2.4)^2 \approx 18$ times less weight.

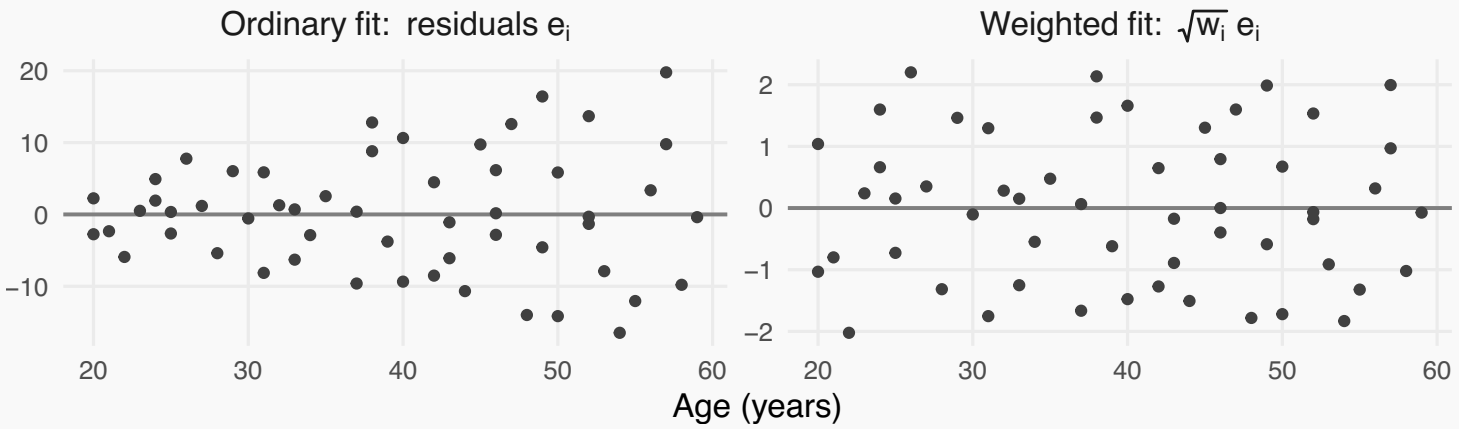
| | $b_1$ | $s\{b_1\}$ | 95% interval for $\beta_1$ |
|---|---|---|---|
| Ordinary | 0.580 | 0.0970 | (0.385, 0.775) |
| Weighted | 0.596 | 0.0792 | (0.437, 0.755) |

The slope barely moves; the interval is narrower, and now valid: the mean blood pressure rises by 0.44 to 0.76 mm Hg per year of age. A second pass gives $b_1 = 0.5964$: no need to iterate.

# Check again


Ordinary fit: residuals $e_i$
Weighted fit: $\sqrt{w_i}\, e_i$
20
10
0
−10
2
1
0
−1
−2
20
30
40
50
60
Age (years)

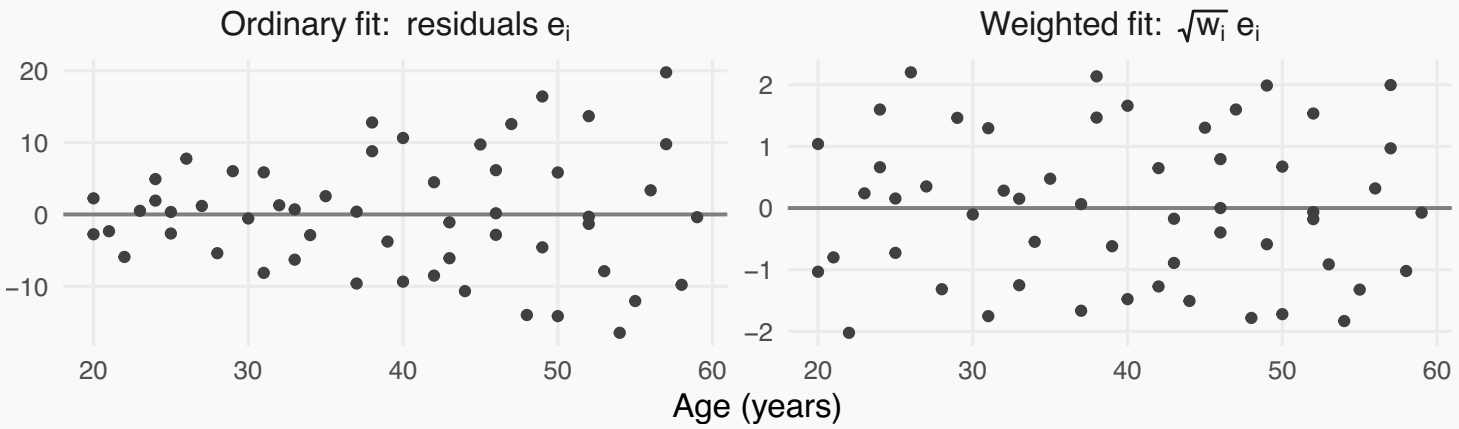
# Check again




The weighted residuals $\sqrt{w_i}\, e_i$ have a constant spread. Brown-Forsythe on them: $t_{BF} = -0.18$, no evidence left.

For the project: the rubric's Validation criterion asks for remedial measures "taken where needed and checked again".

# Check again




The weighted residuals $\sqrt{w_i}\, e_i$ have a constant spread. Brown-Forsythe on them: $t_{BF} = -0.18$, no evidence left.

For the project: the rubric's Validation criterion asks for remedial measures "taken where needed and checked again".

**Careful**

`resid(fitw)` returns $e_i = Y_i - \hat{Y}_i$, which still fan out. Plot `weighted.residuals(fitw)` or `rstandard(fitw)`, which use $\sqrt{w_i}\, e_i$; `plot(fitw)` uses them too.

# What weighting changes, and what it does not

- Unchanged: the response and its units, the regression function, how a coefficient is read.
- Changed: the estimates, a little; the standard errors, tests and intervals, which become valid.
- The $R^2$ of a weighted fit has no simple meaning (here 0.52 against 0.41): do not compare it with the ordinary fit's.

# What weighting changes, and what it does not

- Unchanged: the response and its units, the regression function, how a coefficient is read.
- Changed: the estimates, a little; the standard errors, tests and intervals, which become valid.
- The $R^2$ of a weighted fit has no simple meaning (here 0.52 against 0.41): do not compare it with the ordinary fit's.

Weights or a transformation?

- the mean looks right in the units of $Y$, and only the spread changes: weights;
- the mean is curved too, or the errors are skewed: transform $Y$ (Lecture 18), then check again.

# What weighting changes, and what it does not

- Unchanged: the response and its units, the regression function, how a coefficient is read.
- Changed: the estimates, a little; the standard errors, tests and intervals, which become valid.
- The $R^2$ of a weighted fit has no simple meaning (here 0.52 against 0.41): do not compare it with the ordinary fit's.

Weights or a transformation?

- the mean looks right in the units of $Y$, and only the spread changes: weights;
- the mean is curved too, or the errors are skewed: transform $Y$ (Lecture 18), then check again.

For the project: weighted least squares is now a tool from the course. If you use it, show the standard deviation function and the weighted residuals.

# From question to final model

# The report, step by step

| Step | What the reader should find | Tools |
|---|---|---|
| Question | What you explain or predict, and why it matters | |
| Description | Only what the model uses: the response, its relations with the predictors, unusual cases, data issues | Scatterplots, boxplots, correlations |
| Model | Predictors considered (following from the question): factors, transformations, interactions, all justified | Lectures 8, 15 to 18 |
| Selection | How you reached the final model | $R^2_a$, $C_p$, AIC, BIC, PRESS, `step()` |
| Validation | Residuals (L, I, N, E), unusual cases, VIF, remedies checked again | Lectures 8, 9, 19 to 21 |
| Interpretation | Coefficients and intervals in context; the answer, and its limits | Lectures 15 to 18 |

# Summary

- Unequal variances: weights $w_i = 1/\sigma_i^2$, $\hat{\boldsymbol{\beta}}_w = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{Y}$: ordinary least squares on the data multiplied by $\sqrt{w_i}$.
- Variances known up to a constant (means of $n_i$ measures: $w_i = n_i$), or estimated by regressing $|e_i|$ on a predictor.
- Check again on the weighted residuals $\sqrt{w_i}\, e_i$, not on `resid(fitw)`.
- The report follows the analysis: question, description, model, selection, validation, interpretation.

Report, R Markdown file, dataset, contribution statement and slides: Tuesday 1 December, 23:59. No late submission.