Zurich University
of Applied Sciences

School of
Engineering

IDP Institute of
Data Analysis and
Process Design

WBL Statistik 2024 — Robust Fitting

Half-Day 1: Introduction to Robust Estimation Techniques

Andreas Ruckstuhl
Institut für Datenanalyse und Prozessdesign
Zürcher Hochschule für Angewandte Wissenschaften

Outline:

Half-Day 1
- Regression Model and the Outlier Problem
- Measuring Robustness
- Location M-Estimation
- Inference
- Regression M-Estimation
- Example from Molecular Spectroscopy

Half-Day 2
- General Regression M-Estimation
- Regression MM-Estimation
- Example from Finance
- Robust Inference
- Robust Estimation with GLM

Half-Day 3
- Robust Estimation of the Covariance Matrix
- Principal Component Analysis
- Linear Discriminant Analysis
- Baseline Removal: An application of robust fitting beyond theory

Your Lecturer

Name: **Andreas Ruckstuhl**
Civil Status: Married, 2 children
Education: Dr. sc. math. ETH

Position: Professor of Statistical Data Analysis, ZHAW Winterthur
Lecturer, WBL Applied Statistics, ETH Zürich

Expierence:

1987 – 1991	Statistical Consulting and Teaching Assistant, Seminar für Statistik, ETHZ
1991 – 1996	PhD, Teaching Assistant, Lecturer in NDK/WBL, ETHZ
1996	Post-Doc, Texas A&M University, College Station, TX, USA
1996 – 1999	Lecturer, ANU, Canberra, Australia
Since 1999	Institute of Data Analysis and Process Design, ZHAW

1.1 Regression Model and the Outlier Problem

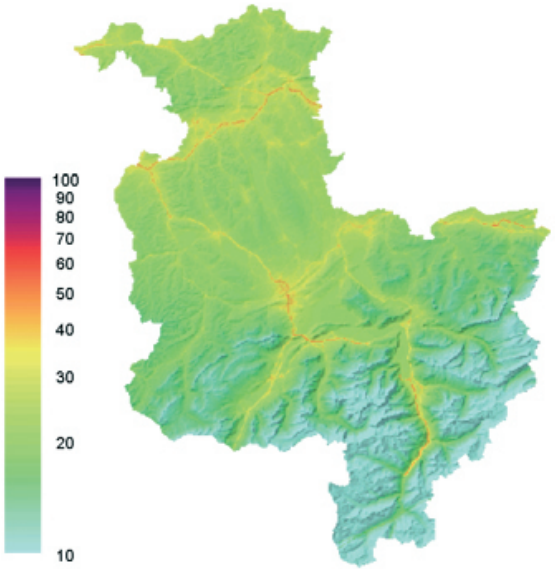
Example Air Quality (I)

False colour maps for NO_2 pollution (with R. Locher)

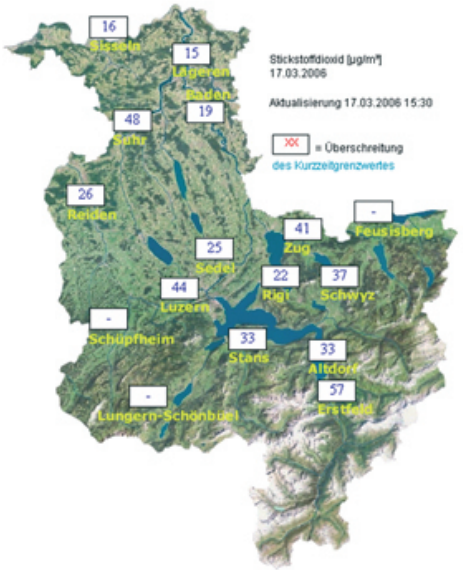
The actual air quality can be assessed insufficiently at a location far away from a measurement station.

More attractive would be a false colour map which shows daily means **highly resolved in time and location**.

today/future



until recently



see also `https://home.zhaw.ch/~lore/OLK/index.html`

Example Air Quality (II)

Available are

- **Daily means at a few sites**
- False colour map of **yearly means** for NO_2 based on a physical model

Idea:

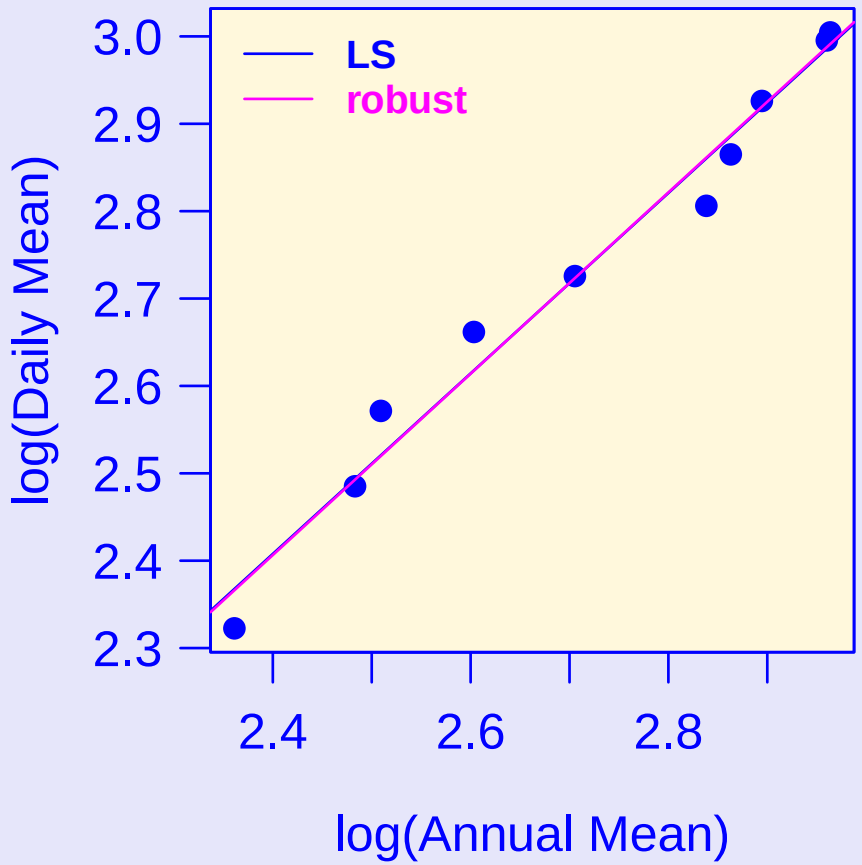
- Construct a false colour map which is highly resolved in **time and space**.

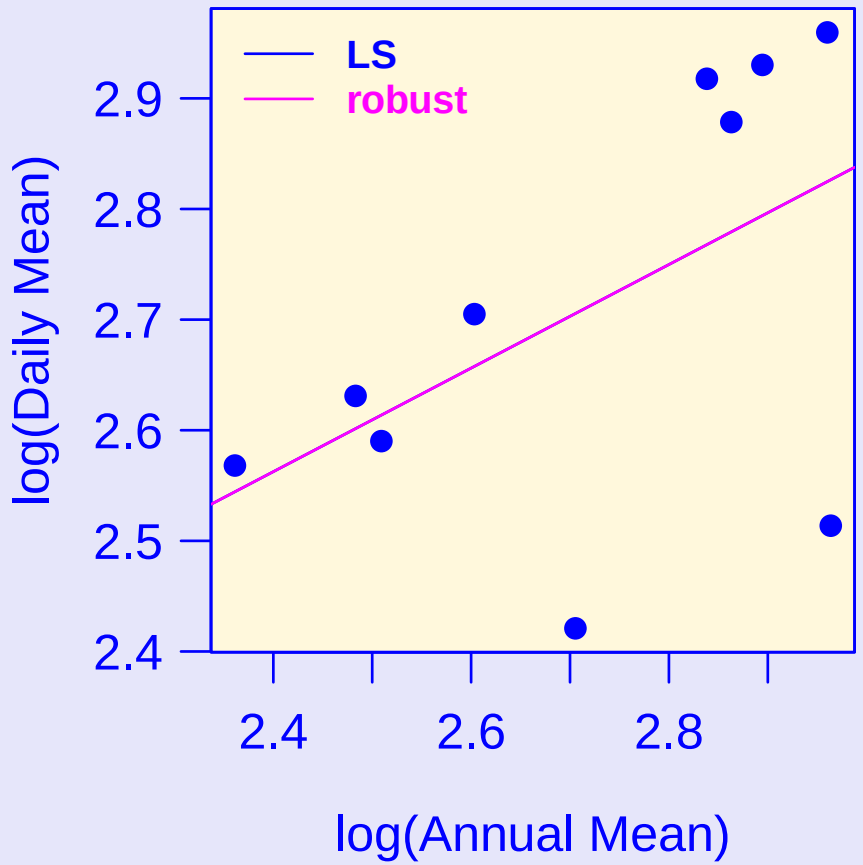
The idea is feasible because there is a nice empirical relationship between the daily and the yearly means:

$$\log \langle \text{daily mean} \rangle \approx \beta_o + \beta_1 \log \langle \text{yearly mean} \rangle$$

β_o and β_1 are weather dependent

Example Air Quality (III)





Left: Stable weather condition over the whole area
Right: The weather condition is different at a few measurement sites.

The Regression Model

In regression, the model is

$$Y_i = \sum_{j=0}^{m} x_i^{(j)} \beta_j + E_i,$$

where the random errors E_i, $i = 1, \ldots, n$ are independent and Gaussian distributed with mean 0 and unknown spread σ.

In residual analysis, we check all assumptions as thoroughly as possible. Among other things, we put a lot of effort in **finding potential outliers and bad leverage points** which we would like to remove from the analysis because they lead to unreliable fits.

Why? - Because **real data** are never exactly Gaussian distributed.

There are gross errors and other perturbations resulting in a distribution which is longer tailed than the Gaussian distribution.

The Oldest Informal Approach to Robustness

(which is still a common practise) is to

1. examine the data for obvious outliers,
2. delete these outliers
3. apply the optimal inference procedure for the assumed model to the *cleaned* data set.

However, this data analytic approach is not unproblematic since

- Even professional statisticians do not always screen the data
- It can be **difficult or even impossible to identify outliers**, particularly in multivariate or highly structured data
- cannot examine the relationships in the data without first fitting a model
- It can be **difficult to formalize this process** so that it can be automatized
- Inference based on applying a standard procedure to the cleaned data will be based on **distribution theory which ignores the cleaning process** and hence will be inapplicable and possibly misleading

1.2 Measuring Robustness

How robust an estimator is, can be investigated by two simple measures:

- **influence function and gross error sensitivity**
- **breakdown point**

Both measures are based on the idea of studying the reaction of the estimator under the influence of gross errors, i.e. **arbitrary added** data.

Gross Error Sensitivity

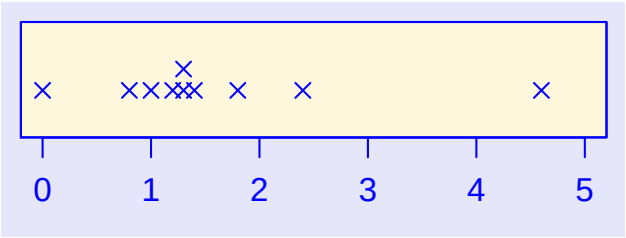
The (gross error) sensitivity is based on the influence function and measures the **maximum effect of a single observation on the estimated value**.

Breakdown Point

The breakdown point gives the **minimum proportion** of data that can be altered without causing **completely unreliable estimates.**

Example: Prolongation of Sleep

(Cushny and Peebles, 1905)



Data on the prolongation of sleep by means of two Drugs. These data were used by Student (1908) as the first illustration of the paired t-test.

$$\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 1.58$$

$$\text{med} = \text{median}\{y_i, i = 1, \ldots, 10\} = 1.3$$

$$\bar{y}_{10\%} = \frac{1}{8}\sum_{i=2}^{9} y_{(i)} = 1.4$$

$$\bar{y}^* = \frac{1}{9}\sum_{i=1}^{9} y_{(i)} = 1.23$$

For $\bar{y}^*$ the rejection rule $\frac{|y_{(n)} - \bar{y}|}{s} > 2.18$ is used, where the value 2.18 depends on the sample size.

Sensitivity Curve - Influence Function

Let $\widehat{\beta}\langle y_1, \ldots, y_n \rangle$ be the estimator. The **sensitivity curve**

$$SC\left\langle y;\, y_1, \ldots, y_{n-1}, \widehat{\beta} \right\rangle = \frac{\widehat{\beta}\langle y_1, \ldots, y_{n-1}, y \rangle - \widehat{\beta}\langle y_1, \ldots, y_{n-1} \rangle}{1/n}$$

describes the **influence of an single observation y on the estimated value**.

Note that $SC\langle \ldots \rangle$ depends on the data $y_1, \ldots, y_{n-1}$.

A similar measure, but one which characterises only the **estimator**, is the **influence function**

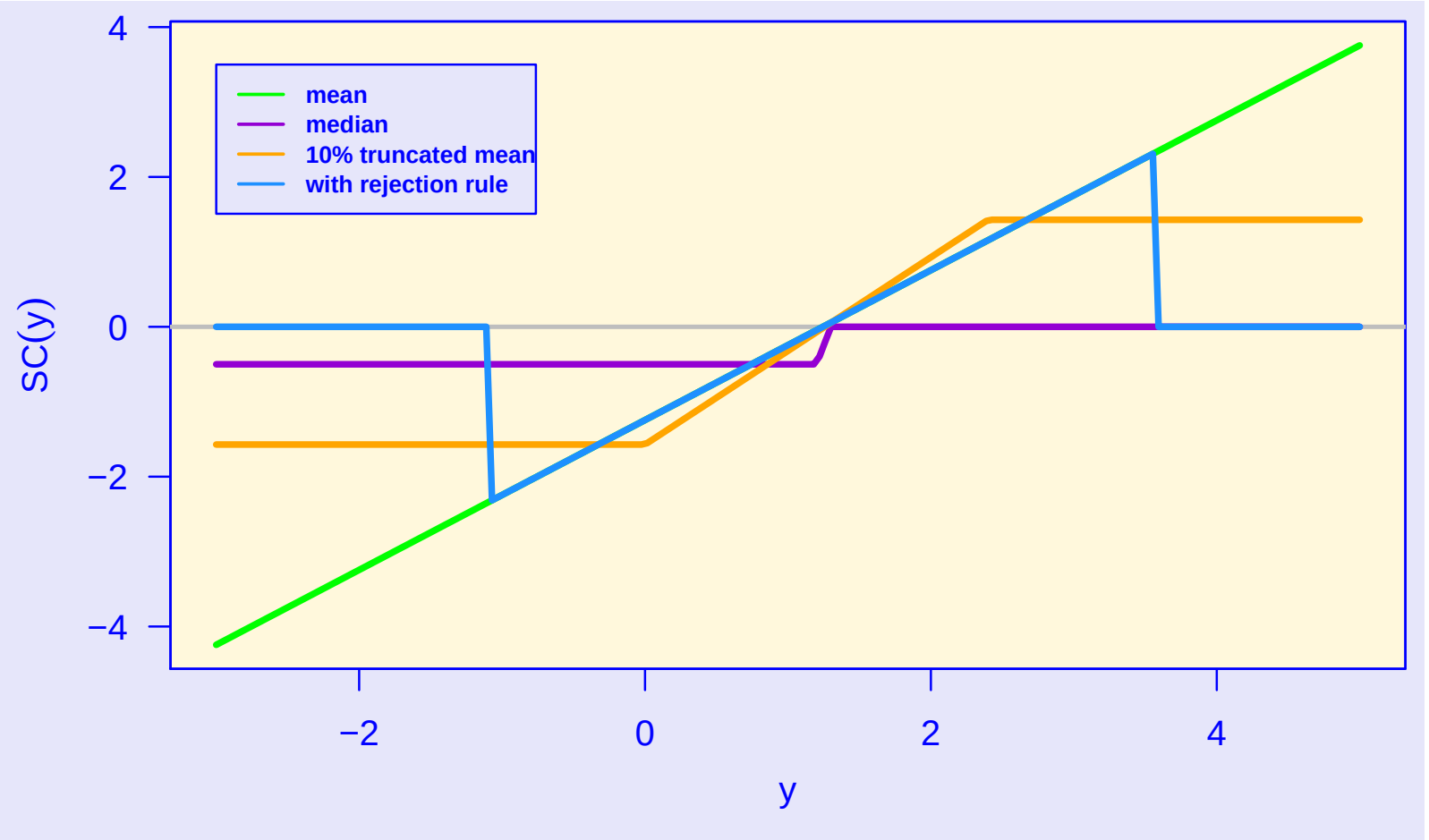
$$SC\left\langle y;\, y_1, \ldots, y_{n-1}, \widehat{\beta} \right\rangle \quad \xrightarrow{n\uparrow\infty} \quad IF\left\langle y; \mathcal{N}, \widehat{\beta} \right\rangle$$

(The data $y_1, \ldots, y_{n-1}$ have been replaced by the model distribution $\mathcal{N}$.)

Sensitivity Curve (Empirical Influence Function)

Sensitivity curve for the prolongation of sleep data.

The "outlier" $y = 4.6$ hours has been varied within $-3 < y < 5$.



Definition of Robustness

A **robust** estimator has a **bounded gross error sensitivity**

$$\gamma^* \left\langle \widehat{\beta}, \mathcal{N} \right\rangle := \max_y \left| IF \left\langle y; \widehat{\beta}, \mathcal{N} \right\rangle \right| < \infty$$

Hence

- $\bar{Y}$ is **not** robust and
- med, $\bar{Y}_{10\%}$, $\bar{Y}^*$ are robust

Breakdown Point

How do the estimation methods respond to two outliers?

We consider the worst case scenario of having two observations moving to infinity: $y_{(10)} = y_{(9)} \to \infty$

$$\bar{y} \to \infty \quad \Rightarrow \quad \varepsilon_n^* \langle \bar{y} \rangle = \frac{0}{n} = 0$$

$$\text{med} = 1.3 \text{ as before} \quad \Rightarrow \quad \varepsilon_n^* \langle \text{med} \rangle = \frac{4}{10} = 0.4 \overset{n \text{ large}}{\longrightarrow} 0.5$$

$$\bar{y}_{10\%} \to \infty \quad \Rightarrow \quad \varepsilon_n^* \langle \bar{y}_{10\%} \rangle = \frac{1}{10} = 0.1$$

$$\bar{y}^* \to \infty \quad \Rightarrow \quad \varepsilon_n^* \left\langle \bar{y}^* \right\rangle = \frac{1}{10} = 0.1$$

The **breakdown point** is the maximum ratio of outlying observation such that the estimator still returns reliable estimates.

More formally, call $\mathcal{X}_m$ the set of all data sets $\underline{y}^* = \{y_1^*, \ldots, y_n^*\}$ of size n having $(n - m)$ elements in common with $\underline{y} = \{y_1, y_2, \ldots, y_n\}$. Then the breakdown point $\varepsilon_n^* \left\langle \widehat{\beta}; \underline{y} \right\rangle$ is equal to m^*/n, where

$$m^* = \max \left\{ m \geq 0 \, : \, \left| \widehat{\beta} \left\langle \underline{y}^* \right\rangle - \widehat{\beta} \left\langle \underline{y} \right\rangle \right| < \infty \text{ for all } \underline{y}^* \in \mathcal{X}_m \right\}.$$

1.3 Location M-Estimation and Inference

Location model $Y_i = \beta + E_i\,, \quad i = 1, \ldots, n\,, \quad$ with E_i i.i.d. $\sim \mathcal{N}, \langle 0,\, \sigma^2 \rangle$

Let us find a weighted mean, wherein the weights are designed to prevent the influence of outliers on the estimator as much as possible:

$$\widehat{\beta}_{\mathrm{M}} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}\,.$$

In order to determine appropriate weights (or weight functions), it helps to consult the corresponding influence function *IF*. Theory tells that the influence function of $\widehat{\beta}_{\mathrm{M}}$ is equal to

$$IF\left\langle y, \widehat{\beta}, \mathcal{N} \right\rangle = const \cdot w \cdot \widetilde{r} \;\; = const \cdot \psi\left\langle \widetilde{r} \right\rangle \quad \text{with } \psi\left\langle \widetilde{r} \right\rangle \stackrel{def}{=} w \cdot \widetilde{r},$$

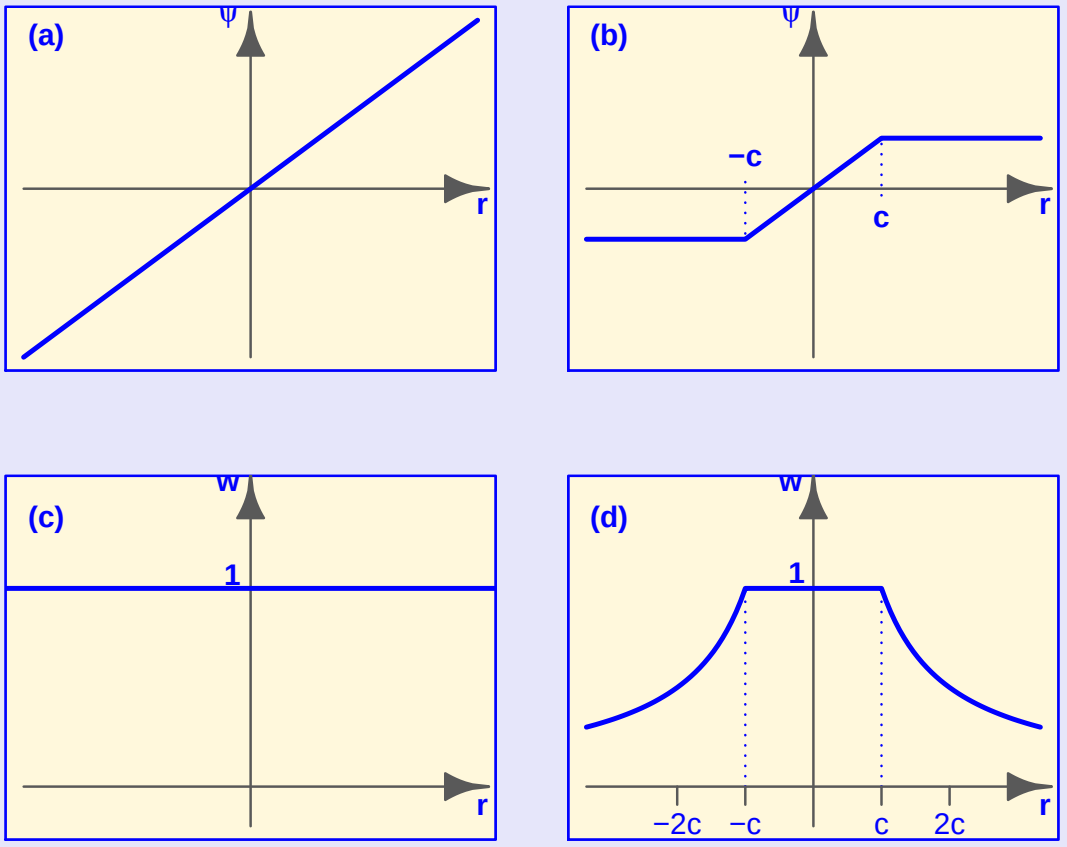
where $\widetilde{r}$ is the scaled residual $\frac{y-\widehat{\beta}}{\widehat{\sigma}}$.

- Hence,
  the influence function of an M-estimator is proportional to the ψ-function.
- For a **robust** M-estimation the ψ-function must be **bounded** – see next page
- The corresponding weight function can be determined by $w_i = \psi(\widetilde{r}_i)/\widetilde{r}_i$

Influence and Weight Function

ψ- (top row9 and weight (bottom row) function for

- ordinary least squares estimation (not robust) – on the left
- a robust M-estimator (ψ-function is also known as Huber's ψ-function) – on the right



- Note that the **weights depend** on the estimation $\widehat{\beta}_M$ and hence is only given implicitly.
- Usually, the **M-Estimator** is defined by an implicit equation,

$$\sum_{i=1}^{n} \psi \left\langle \frac{r_i \langle \widehat{\beta}, \rangle}{\sigma} \right\rangle = 0 \qquad \text{with } r_i \langle \beta \rangle = y_i - \beta \,,$$

  where σ is the scale parameter.

  Note: Basically, this equation corresponds to the normal equation of least-squares estimators.

- A **monotonously increasing, but bounded** ψ-function leads to a **breakdown point of** $\varepsilon^* = \frac{1}{2}$.
- The tuning constant c defining the kink in Huber's ψ-function is determined such that the relative efficiency of the M-estimator is 95% at the Gaussian location model: Hence, $c = 1.345$

Estimation of the Scale Parameter

- The scale parameter σ is usually unknown. But it is needed to determine the position at which unsuitable observation will lose its influence.
- The standard deviation as an estimator for σ however is extremely non-robust.
  (Therefore the rejection rule breaks down with 2 outliers out of 10 obs., cf slide 14.)

- A robust estimator of scale with a breakdown point of $\frac{1}{2}$ is, e.g., the **median of absolute deviation (MAD)**

$$s_{\mathrm{MAD}} \stackrel{def}{=} \frac{\mathrm{median}_i \left\langle \left| y_i - \mathrm{median}_k \langle y_k \rangle \right| \right\rangle}{0.6745}$$

  (The correction factor $1/0.6745$ is needed to obtain a consistent estimate for σ at the Gaussian distribution.)

- There are more proposals of suitable scale estimators. Among them are versions where the scale parameter is estimated in each iteration of the location estimation as e.g. in Huber's Proposal 2.
- A better rejection rule is Huber-type skipped mean:

$$\left| \frac{y_i - \mathrm{median}_k \langle y_k \rangle}{s_{\mathrm{MAD}}} \right| > 3.5\,.$$

Inference

A point estimation without a confidence interval is an incomplete (and often useless) information.
Distribution of the M-estimator: Theory shows that

$$\widehat{\beta} \overset{a}{\sim} \mathcal{N}\left\langle \beta, \frac{1}{n}\tau\,\sigma^2 \right\rangle \qquad \text{with } \tau = \frac{\int \psi^2\left\langle u\right\rangle\ f\left\langle u\right\rangle du}{\left(\int \psi'\left\langle u\right\rangle\ f\left\langle u\right\rangle du\right)^2}$$

The correction factor τ is needed to correct for downweighting good observations. It is larger than 1 and can be

- either calculated using the assumption that $f\left\langle\right\rangle$ is the Gaussian density
- or estimated using the empirical distribution of the residuals:

$$\widehat{\tau} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\psi^2\left\langle \widetilde{r}_i\left\langle\widehat{\beta}\right\rangle\right\rangle}{\left(\frac{1}{n}\sum_{i=1}^{n}\psi'\left\langle \widetilde{r}_i\left\langle\widehat{\beta}\right\rangle\right\rangle\right)^2} \qquad \text{with } \widetilde{r}_i\left\langle\widehat{\beta}\right\rangle = \frac{Y_i - \widehat{\beta}}{s_{\text{MAD}}}$$

Confidence Interval

The confidence interval is calculated based on the z-test because of the asymptotic result.

The confidence interval is formed by all values of β which are not rejected by the asymptotic z-test.

Hence, the $(1-\alpha)$ confidence interval is

$$\widehat{\beta} \pm q^{\mathcal{N}}_{1-\alpha/2} \sqrt{\widehat{\tau}} \frac{s_{\mathrm{MAD}}}{\sqrt{n}} ,$$

where $q^{\mathcal{N}}_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard Gaussian distribution.

To adjust for estimating the scale parameter in a heuristic manner,

$\mathcal{N}$ may be replaced by t_{n-1} in $q^{\mathcal{N}}_{1-\alpha/2}$.

Example Prolongation of Sleep

```
> y # data
> n <- length(y)
> library(robustbase)
> (y.hM <- huberM(y, se=T, k=1.345))
$mu # 1.371091
$s # 0.59304
$SE # 0.2302046
## Confidence interval
> h1 <- qt(0.975, n-1) * y.hM$SE
> y.hM$mu + c(-1,1)*h1
## 0.8503317 1.8918499

## Classical estimation
> t.test(y) # One Sample t-test
95 percent confidence interval:
0.7001142 2.4598858
```

- Using Huber's M-estimator with $c = 1.345$ results in $\widehat{\beta}_M = 1.37$

  Remember, $\bar{Y} = 1.58$

- The robust scale estimator results in $s_{\text{MAD}} = 0.59$ and $\sqrt{\widehat{\tau}}$ is $\sqrt{1.507} = 1.227$

- The 95% confidence interval is

  $1.37 \pm 2.26 \cdot 0.59/\sqrt{10} \cdot 1.227 = [0.85, 1.89]$

  The classical one is [0.70, 2.46]

2.1 Regression M-Estimation

The linear regression model is

$$Y_i = \beta_1 \cdot 1 + \beta_2 \cdot x_i^{(2)} + \ldots + \beta_p \cdot x_i^{(p)} + E_i \qquad \text{with } E_i \text{ i.i.d.}$$

- If the errors E_i are **exactly Gaussian** distributed, then the least-square estimator is optimal.

- If the errors E_i are only **approximately Gaussian** distributed, then the least-square estimator is **not approximately optimal** but rather inefficient already at slightly longer tailed distributions.

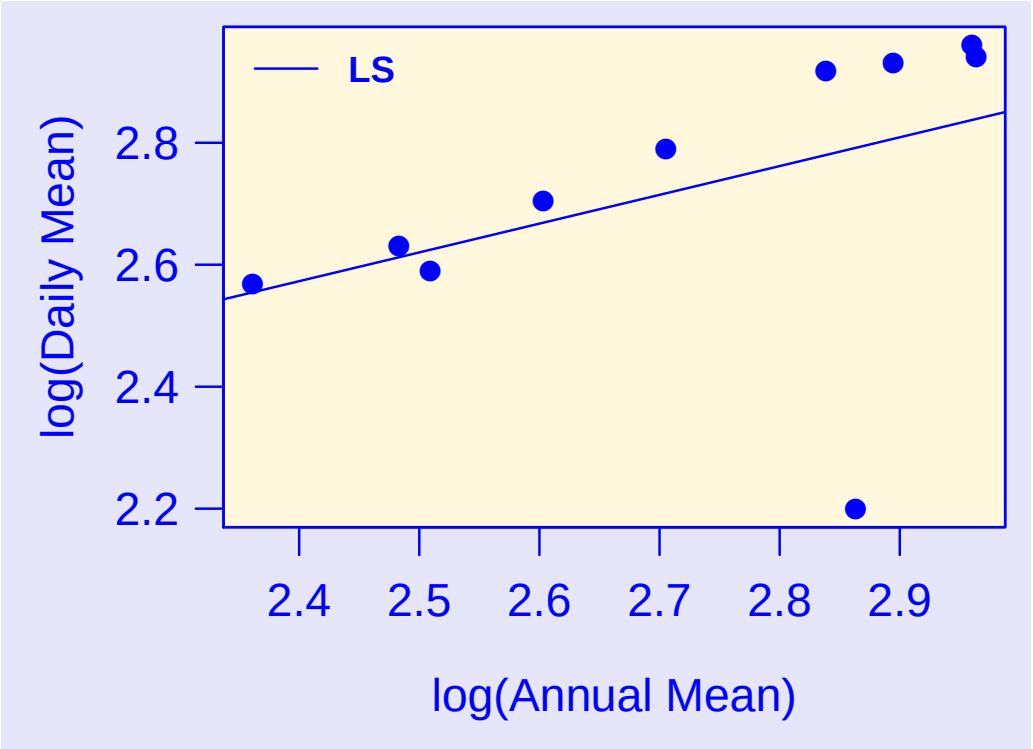
Real data are never exactly Gaussian distributed. There are gross errors and its distribution is often longer tailed than the Gaussian distribution.

Example Modified Air Quality (IV)

Recall the Example Air Quality:

Two outliers suffice to give unusable least squares lines

To keep the world simple (at least at the moment) we modify the data set such that it contains just one outlier:



From OLS to M-estimation

Analogously to the location model, we will replace the ordinary least squares (OLS) estimator by a weighted least squares(WLS) estimator, where the weights should reduce the influence of large residuals.

Normal equations of the WLS estimator: $\boldsymbol{X}^T \boldsymbol{W} \underline{r} = \underline{0}$ or equivalently,

$$\sum_{i=1}^{n} w_i \cdot r_i \cdot \underline{x}_i = \underline{0}\,,$$

where $r_i\left\langle \underline{\beta} \right\rangle := y_i - \sum_{j=0}^{m} x_i^{(j)} \beta_j$ are the residuals.

Replacing again $(w_i \cdot r_i)$ by $\psi\left\langle r_i/\sigma \right\rangle$ yields the formal definition of a **regression M-estimator** $\widehat{\underline{\beta}}_{\mathrm{M}}$:

$$\sum_{i=1}^{n} \psi\left\langle \frac{r_i\left\langle \underline{\beta}_{\mathrm{M}} \right\rangle}{\sigma} \right\rangle \underline{x}_i = \underline{0}\,.$$

The influence function of a regression M-estimator

In order to determine appropriate weights (or weight functions), it again helps to consult the corresponding influence function *IF* of the regression M-estimator:

$$IF\left\langle \underline{x}, y; \underline{\widehat{\beta}}_{\mathrm{M}}, \mathcal{N} \right\rangle = \psi\left\langle \frac{r}{\sigma} \right\rangle \boldsymbol{M}\,\underline{x}\,.$$

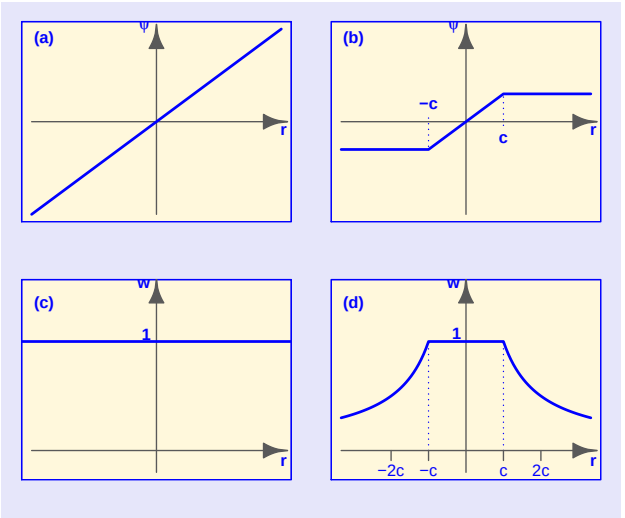
$\boldsymbol{M}$ is a matrix, which only depends on the design matrix X.

Hence, the ψ **function** again reflects the **influence** of a residual (and therefore of an observation) onto the estimator.

To obtain a robust regression M-estimator the ψ-function must be bounded
as, e.g., Huber's ψ-function

Figure on the right: ψ- and weight function for

- ordinary least squares estimation (not robust) – on the left
- Huber's M-estimator – on the right



Asymptotic Distribution

The regression M-estimator is asymptotically normally distributed with covariance matrix $\sigma^2 \tau \boldsymbol{C}^{-1}$, where $\boldsymbol{C} := \sum_i \underline{x}_i \underline{x}_i^T$.

Up to the correction factor τ, this covariance matrix corresponds to the covariance matrix of the least squares estimation.

Thus the covariance matrix of the estimator $\widehat{\underline{\beta}}$ can be estimated by

$$\widehat{\boldsymbol{V}} = (\widehat{\sigma}^2)\, \widehat{\tau}\, \widehat{\boldsymbol{C}}^{-1}$$

- where

$$\widehat{\boldsymbol{C}} = \frac{\sum_i \widetilde{w}_i \underline{x}_i \underline{x}_i^T}{\frac{1}{n}\sum_i \widetilde{w}_i} \qquad \text{with } \widetilde{w}_i \text{ either } \widetilde{w}_i := 1 \text{ or } \widetilde{w}_i := \psi\left\langle \widetilde{r}_i \right\rangle / \widetilde{r}_i.$$

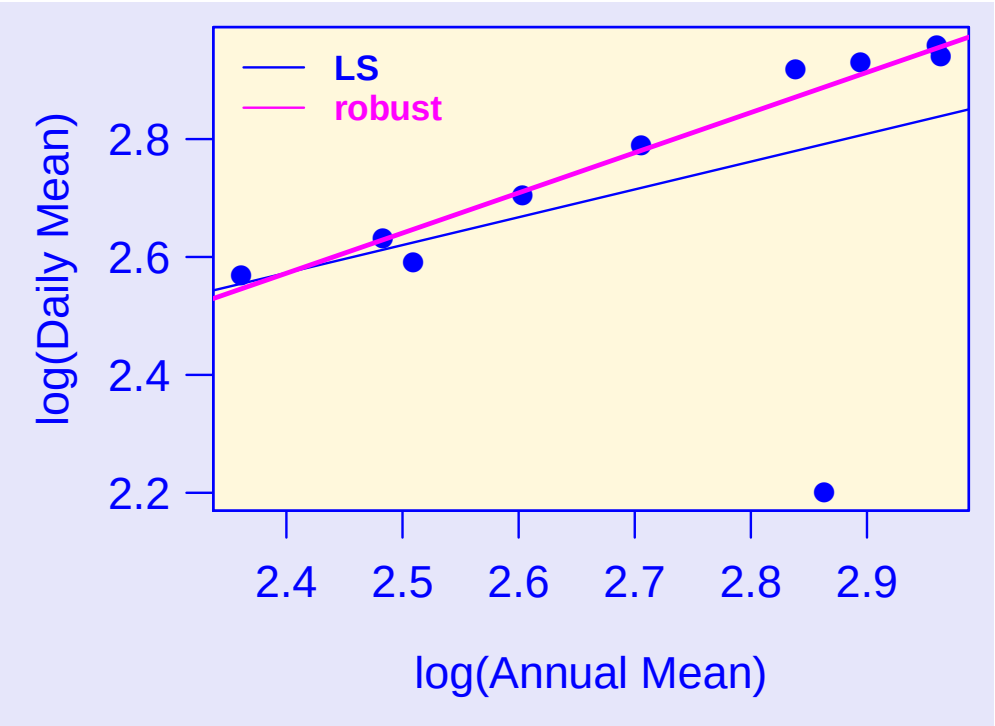
  The $\widetilde{r}_i$ are again the appropriate scaled residuals $\widetilde{r}_i = r_i \left\langle \widehat{\underline{\beta}} \right\rangle / \widehat{\sigma}$.

Weights of the form $\widetilde{w}_i = \psi\left\langle \widetilde{r}_i \right\rangle / \widetilde{r}_i$ are sensible in the calculation of $\widehat{\boldsymbol{C}}$, since contaminated observations should be down-weighted too in the estimation of the covariance matrix according to their potential degree of contamination, measured by $\widetilde{w}_i$.

Example Modified Air Quality (V)

```
>
> f1 <- lm(...) # LS
> coef(f1)
Intercept Slope
1.440561 0.471861
```

```
> ## library("MASS")
> f2 <- rlm(...) # robust
> coef(f2)
Intercept Slope
0.9369407 0.6814527
```



What should be done with outliers?

Diagnose	*Treatment*
Transcription error	Correct
Different population (e.g. diseased, pregnant, ...)	Remove observation/population
Implausible or dubious value	Find explanation, ...
	they can be the gateway for new insights

Prerequisite: Detect outliers!

Robust methods are called for, particularly in high dimensional data, because **robust methods**

- **are not deceived by outliers,**
- **help us detect outliers more easily and faster.**
- **may help to automatize a statistical analysis in a safe way.**

2.2 Example From Molecular Spectroscopy

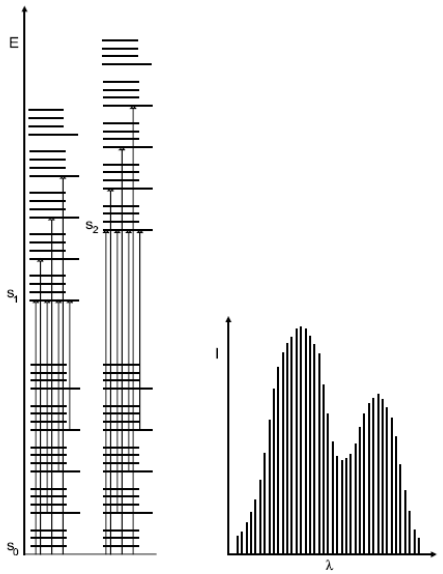
Model for the spectrum data:

$$Y_i = \theta_{u_i} - \theta_{\ell_i} + E_i$$

which can be also written as

$$\underline{Y} = \boldsymbol{X}\underline{\theta} + \underline{E}\,,$$

where

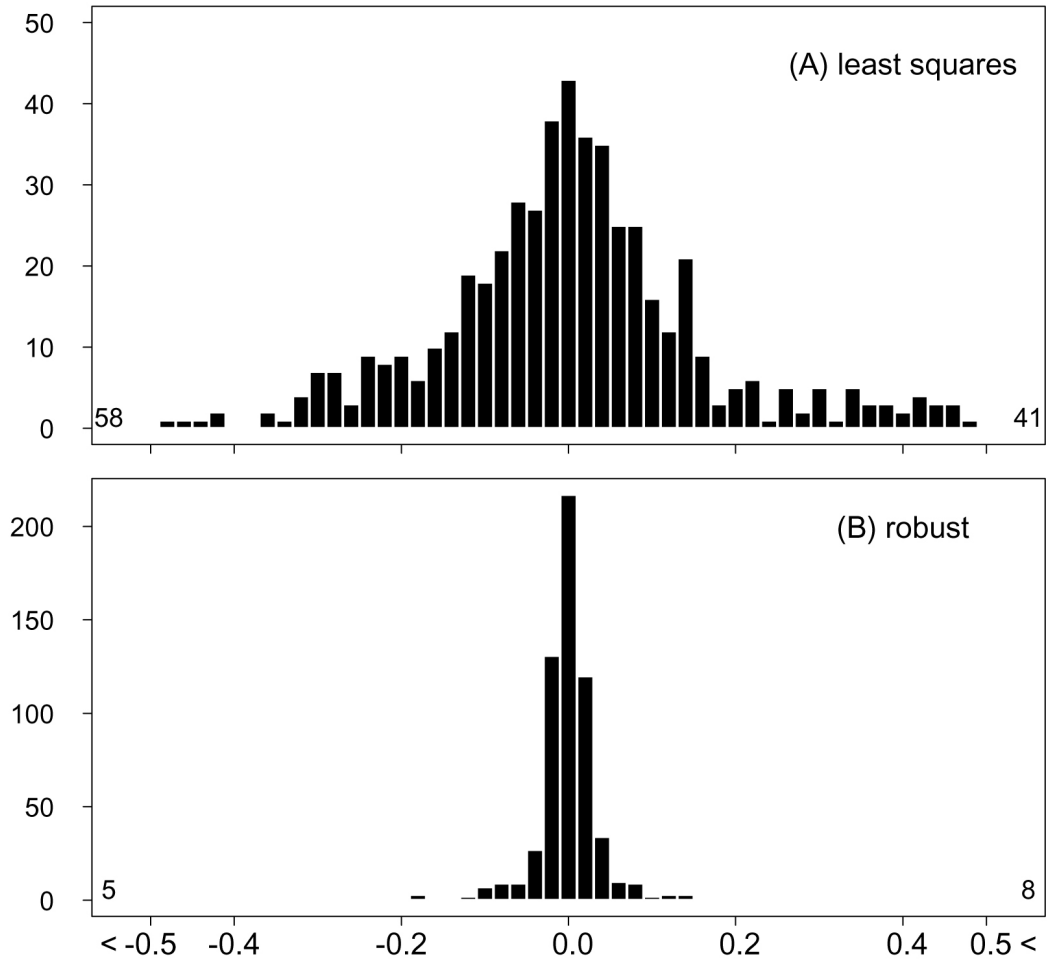


$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}, \underline{\theta} = \begin{pmatrix} \theta_A \\ \theta_B \\ \dots \\ \dots \\ \theta_a \\ \theta_b \\ \dots \end{pmatrix}, \text{ and } \boldsymbol{X} = \left[\begin{array}{cccc|cccc} 0 & 1 & 0 & \dots & 0 & 0 & -1 & 0 & \dots \\ 1 & 0 & 0 & \dots & 0 & -1 & 0 & 0 & \dots \\ & & \vdots & & & & & \vdots & \\ & & \vdots & & & & & \vdots & \\ & & \vdots & & & & & \vdots & \\ & & \vdots & & & & & \vdots & \end{array}\right]$$

See Sec. 2.2 in the Lecture Notes for more details

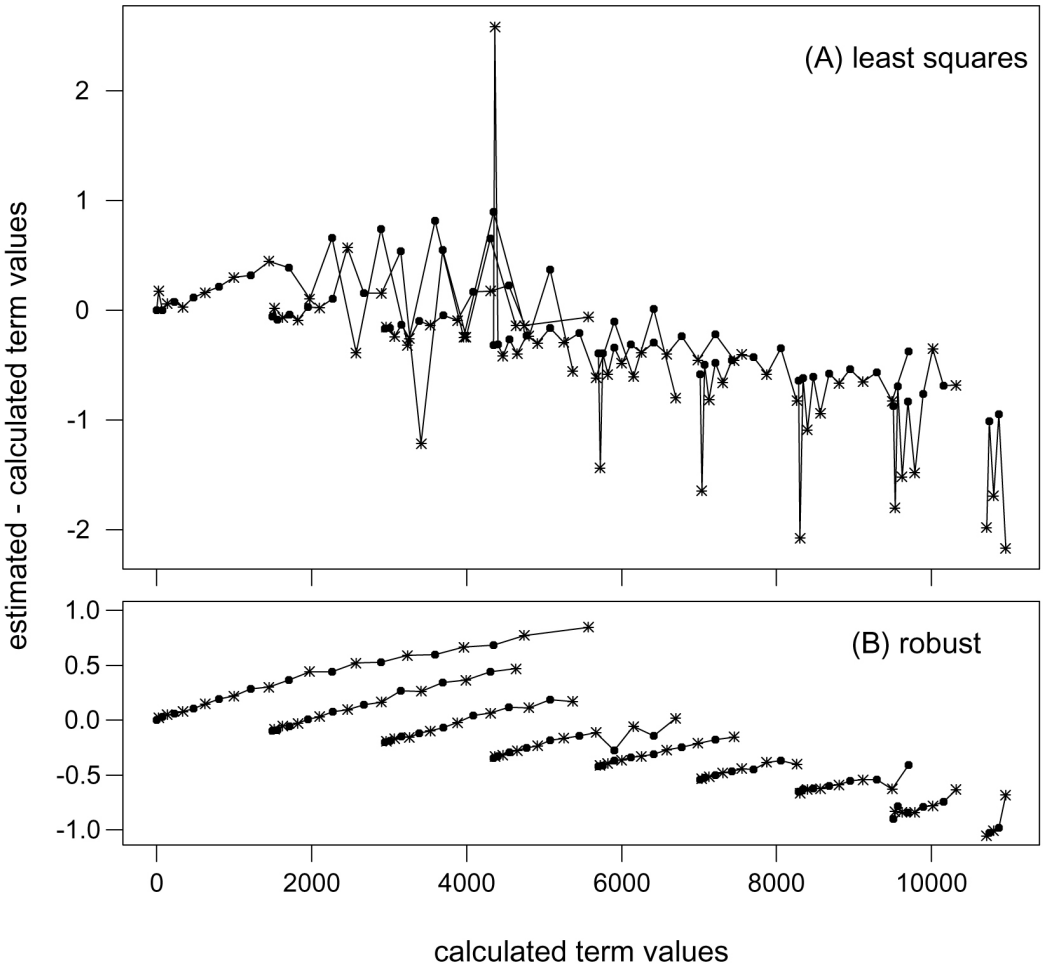
Example Tritium Molecule:

Histograms of the residuals



Example Tritium Molecule:

Estimated coefficients compared with theoretical ones (involves heavy computation and many approximations)



Appendix: Computational Aspects

A regression M-estimator is defined implicitly (cf. SLIDE 24) by

$$\sum_{i=1}^{n} \psi\left\langle r_i\left\langle \widehat{\underline{\beta}}_{\mathrm{M}}\right\rangle / \sigma\right\rangle \cdot x_i^{(k)} = 0\,, \qquad k = 1,2,\ldots,\mathrm{p}. \tag{1}$$

or with weights

$$w_i \overset{def}{=} \frac{\psi\left\langle r_i\left\langle \widehat{\underline{\beta}}_{\mathrm{M}}\right\rangle / \sigma\right\rangle}{r_i\left\langle \widehat{\underline{\beta}}_{\mathrm{M}}\right\rangle / \sigma} \tag{2}$$

the system of quations in (1) can be written as

$$\sum_{i=1}^{n} w_i \cdot r_i\left\langle \widehat{\underline{\beta}}_{\mathrm{M}}\right\rangle \cdot x_i^{(k)} = 0\,, \qquad k = 1,2,\ldots,\mathrm{p}. \tag{3}$$

(These are the normal equations of a weighted regression
but the weights depend on the solution $\widehat{\underline{\beta}}_{\mathrm{M}}$!)

Iterated Re-Weighted Least Squares

1. Let $\underline{\widehat{\beta}}^{(m=0)}$ be an initial solution for (3)
2. Calculate $r_i^{(m)} = y_i - \sum_{j=1}^{p} x_i^{(j)} \cdot \widehat{\beta}_j^{(m)}$,
   $\sigma^{(m)} = \text{median}\{|r_i^{(m)}|\}/0.6745$ (MAV)
   and $w_i^{(m)}$ according to (2)
3. Solve weighted normal equations (3)
4. Repeat steps (ii) and (iii) until convergence.

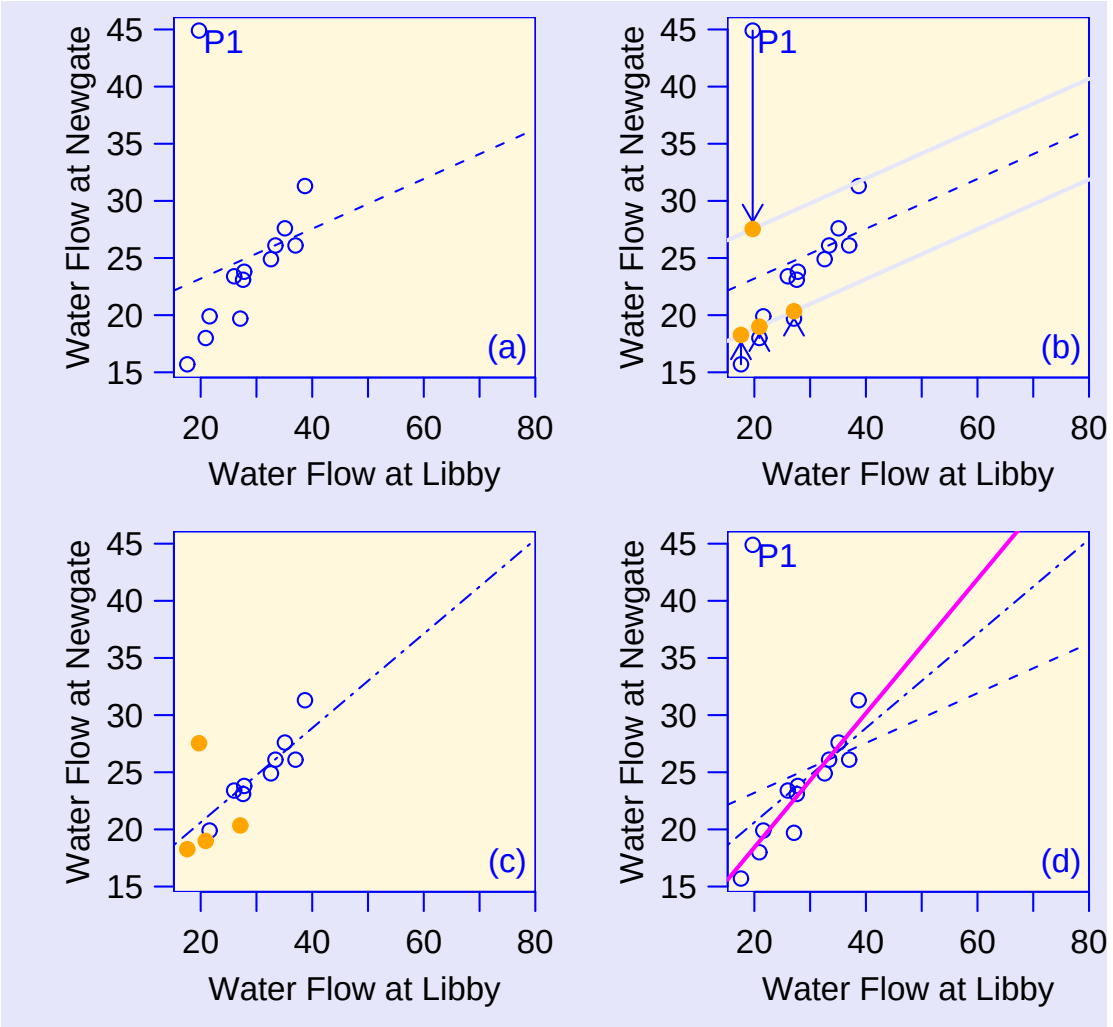
Algorithm Based on Modified Observations

Replace step (3) above by solving the normal equation from the least-squares problem

$$\sum_{i=1}^{n} \left(y_i^{(m)} - \sum_{j=1}^{p} x_i^{(j)} \cdot \widehat{\beta}_j \right)^2 \stackrel{!}{=} \min_{\underline{\widehat{\beta}}}$$

where $y_i^{(m)} \stackrel{def}{=} \sum_{j=1}^{p} x_i^{(j)} \cdot \widehat{\beta}_j^{(m)} + w_i^{(m)} \cdot r_i^{(m)}$ are pseudo-observations.

Illustration of the "Modified Observations" Algorithm



Take Home Message Half-Day 1

- The least-squares estimator is the optimal estimator when the data is normally distributed
- However, the least-squares estimator is unreliable if contaminated observations are present

  see, e.g., the example from molecular spectroscopy
- There are better (=efficient, intuitively "correct") estimators.
  With **M-Estimators**, the influence of gross errors can be controlled very smoothly.

Generally, **influence function** and **breakdown point** are two very important measure for assessing the robustness properties of an estimator.

Be aware: Outliers are only define with respect to a model