STATISTICS & DRG

Some statistical methodological problems in the analysis of hospital cost and length of stays classified by DRG and their solutions

Unité de statistique

IUMSP

Outline

1. PCS and DRG : a short introduction
2. The statistical problems
3. Classical solutions
4. Robust solutions
5. Examples
6. Reimbursement: a proposal
7. Conclusions

1. PCS and DRG : a short introduction

PCS: *Patient Classification System*

- a tool of hospital (system) management,
- rules to assign each stay to a group, e.g. a DRG, using a grouper; groups are defined according to diagnosis and treatment, groups are cost homogeneous.

→ the hospital customer population is a mixture of DRG-cases: a casemix.

Main goals: budgeting, reimbursing

- Pay each stay according to the "mean cost" of the corresponding DRG → payments fairly distributed among hospitals (according to their casemix).
- Compare hospital performance standardized by casemix, stimulate an economically competitive behaviour among hospitals.

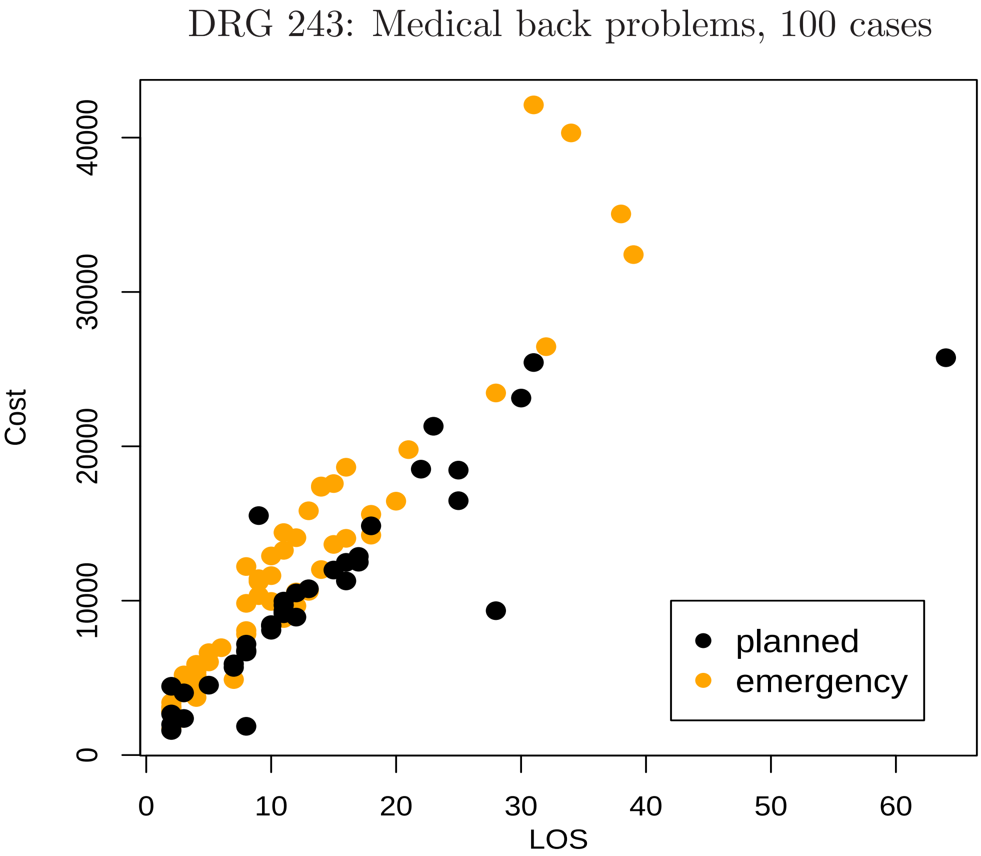
Hystory:

- In the 60's : design of the first DRG system at Yale University
- Migration from USA, to Western EU, Australia, East EU, Asia
- Huge variety of PCS (APDRG, IRDRG, APRDRG, ANDRG, ...)
- From ~ 500 to ~ 1500 groups
- A Swiss DRG system will be implemented starting from 2012.

In theory

- Diagnosis and treatement ensure the economical homogeneity
- Length of stay (LOS) is a good proxy of Cost
- Annual LOS means / Cost means (e.g., on a national basis) are reliable statistics

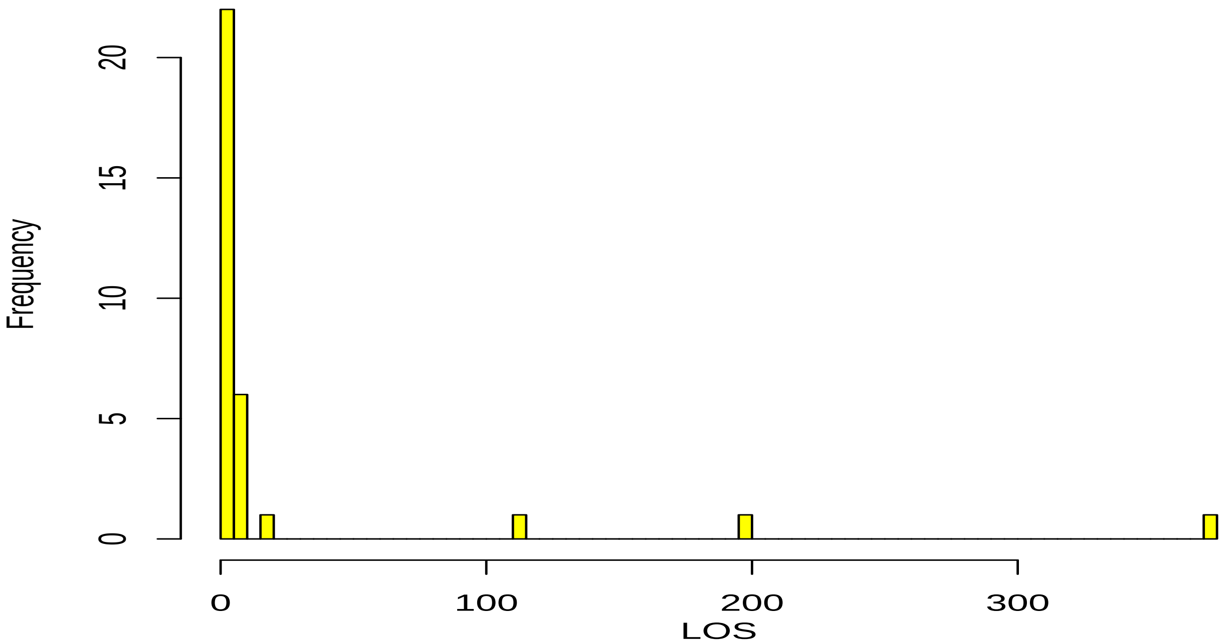
In practice



- Cost may depend on other covariates, e.g., LOS, admission type, ...
- There are atypical cases: outliers (low outliers; high outliers)
- LOS outliers are not necessarily Cost outliers

DRG 34: disorders of the nervous system, 32 cases

LOS : 1,1,2,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4,5,5,5,5,6,6,7,7,8,9,16,115,198,374



Mean of complete data = 25.5 jours

Mean removing 3 outliers = 4.4 jours

The usual mean is very sensitive to outliers

2. Statistical problems

Develop automatic procedures to:

- detect outliers,
- compute a cost mean (LOS mean) robust (resistant) wrt outliers,
- taking into account available covariates,
- taking into account the asymmetrical shapes of the distributions,
- without necessarily increasing the number of groups.

(Marazzi-Paccaud-Ruffieux-Beguin, 1998)

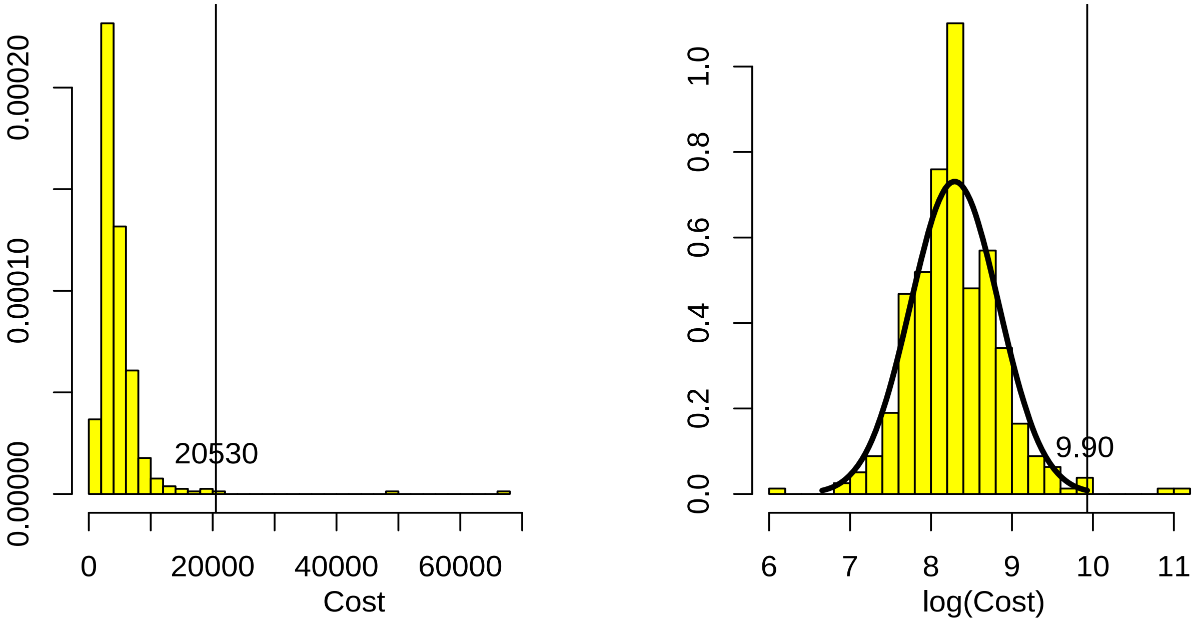
Related economical problems

Improve the reimbursement rules :

- pay typical cases according to robust mean cost of associated DRG,
- inspect outliers,
- pay outliers according to a special rule,
- and minimize “bad incentives”
  e.g., patient selection, refusing those that might be very expensive.

3. Classical solution suggested by the Normal model

APDRG 25: Seizure & headache age > 17 w/o cc, 395 cases



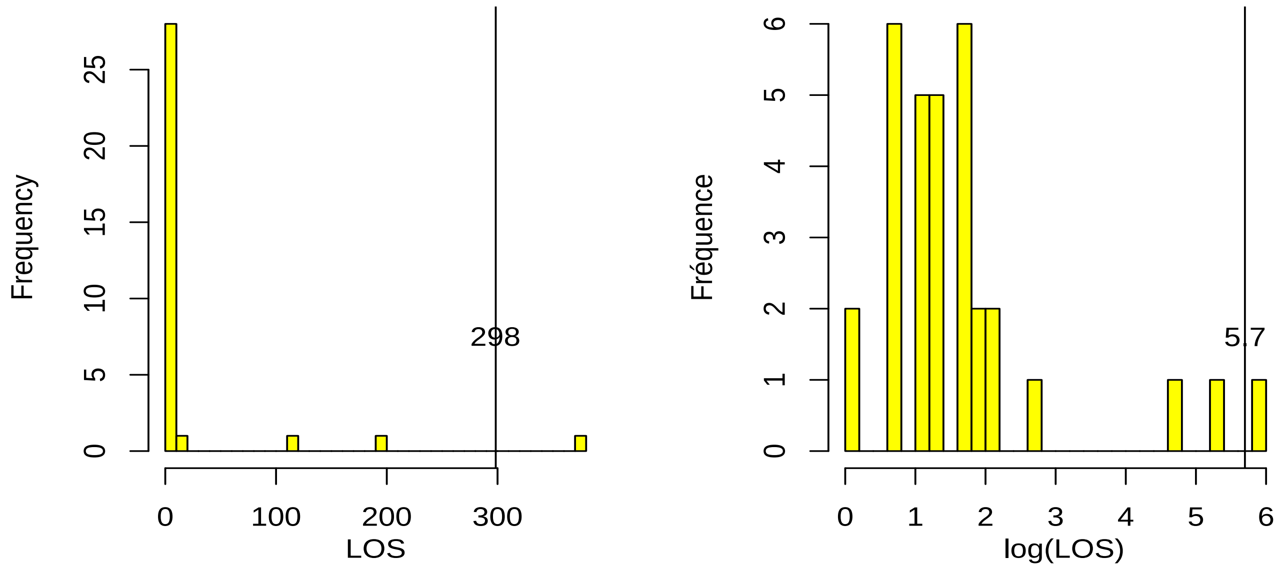
trim point = mean + 3 · standard deviation

Complete mean = 4785 FR; mean without outliers = 4470 Fr

This method is efficient* when the model is adequate for (almost) all cases

* precise, subject to very small sampling fluctuations

DRG 34: disorders of the nervous system, 32 cases



Complete data: mean = 25.5 standard deviation of log-s = 1.34

Cases < 298 : mean = 14.2

Cases < 20 = 4.4 standard deviation of log-s = 0.63

The classical method is very sensitive to outliers

4. Robust solution

- Assume that a parametric model, e.g.,
  Normal, Weibull, Gamma, Log-Normal, Log-Weibull, Log-Gamma, ...
  provides and adequate description of the majority of data,
  not necessarily the totality.
- Fit an adequate model to the majority using a robust procedure.

$$\text{robust mean} = \text{mean of the adjusted model}$$

Technical difficulty: how to combine robustness and efficiency ?

Step 1: Use a very resistant but inefficient procedure to detect outliers

Step 2: Apply a classical and efficient procedure to the cleaned data

Examples of very robust but inefficient statistics

Le MED (median) is a measure of the “centre” of the data

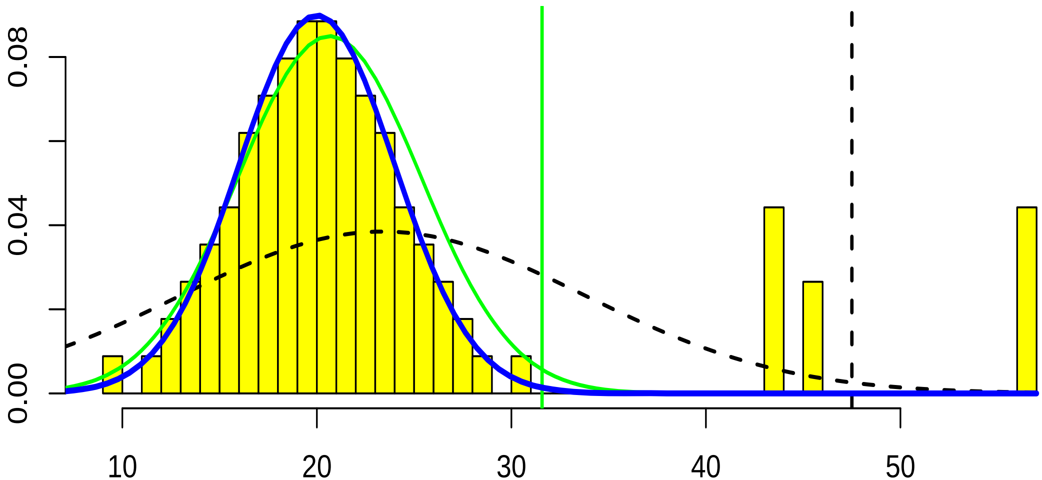
$$\mathrm{MED}(1,2,3,1000,2000) = 3$$

Le MAD (median absolute deviation) is a measure of “variability”

$$\begin{aligned}&\mathrm{MAD}(1,2,3,1000,2000) = \\ &\mathrm{MED}(|1-3|,|2-3|,|3-3|,|1000-3|,|2000-3|) = \\ &\mathrm{MED}(2,1,0,997,1997) = 2\end{aligned}$$

MED and MAD resist to $\sim 50\%$ outlier contamination !

Robust fit of a Normal model with mean μ and stand.dev. σ



1. Very robust but inefficient fit : solve

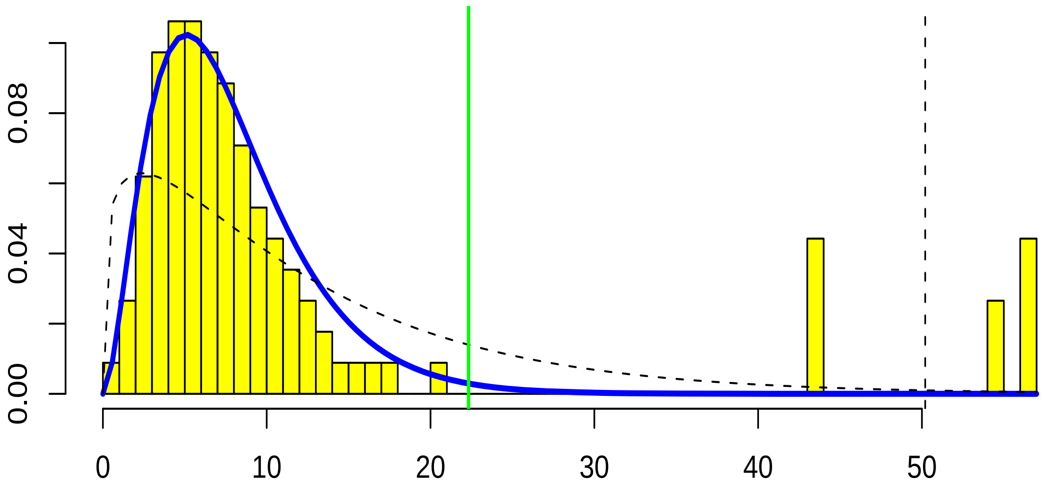
$$\mu = \text{MED(data)} ,$$
$$0.6745 \cdot \sigma = \text{MAD(data)}$$

Trim point = percentile 99% of the very robust model

2. Robust and efficient model : Max-likelihood(data < trim point)

Robust mean = mean of the robust and efficient model
$\approx$ mean(data < trim point)

Robust fit of Gamma or Weibull model (Marazzi-Ruffieux, 1999)



1. Very robust but inefficient fit : solve

$$\text{MED(model)} = \text{MED(data)},$$
$$\text{MAD(model)} = \text{MAD(data)},$$

Trim point = percentile 99% of the very robust model

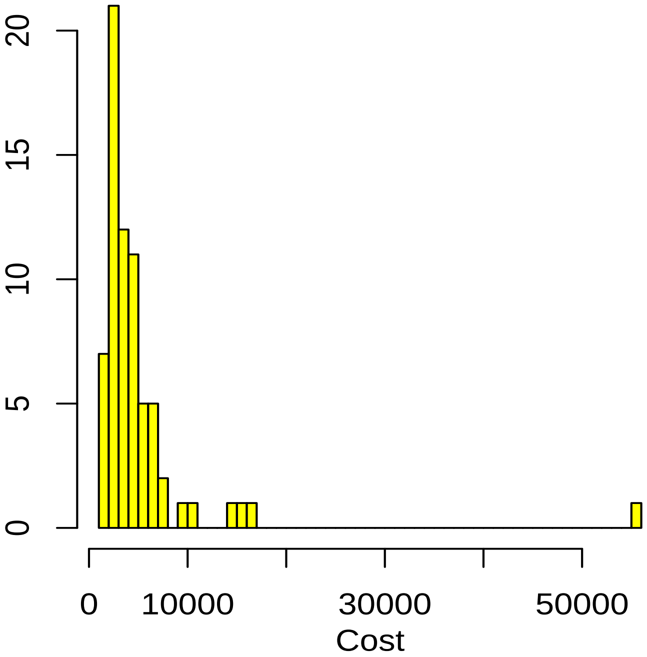
Robust and efficient model : Max-likelihood(data $<$ trim point)

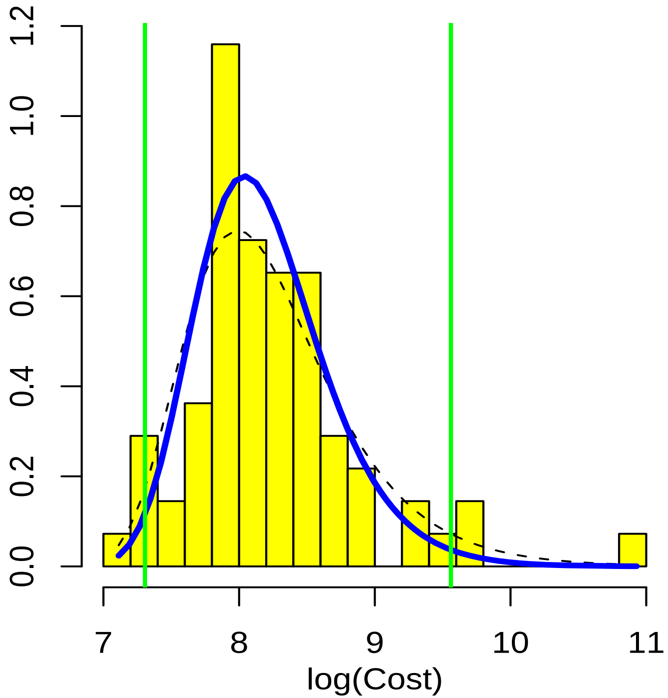
Robust meaan = mean of the robust and efficient model

5. Exemples

Robust fit of Generalized Loggamma model (Marazzi-Yohai, 2011)

DRG185 : Dental & oral dis exc exctract & restorat, Age > 17; 69 cases





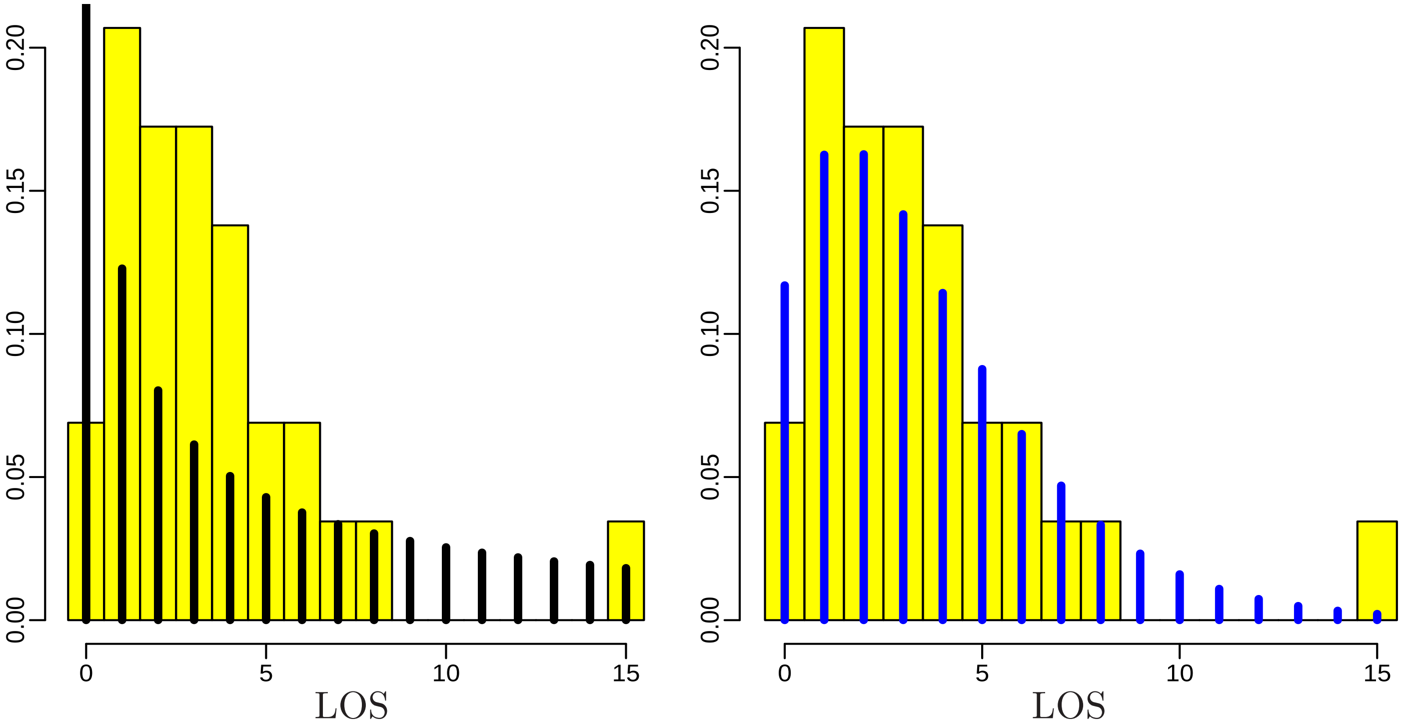
Classical mean = 5060 Fr

Two trim points

Robust mean = 4200 Fr

Robust fit of Negative Binomial model (Marazzi-Yohai 2010; Amiguet 2011)

DRG 34: disorders of the nervous system, 32 cases



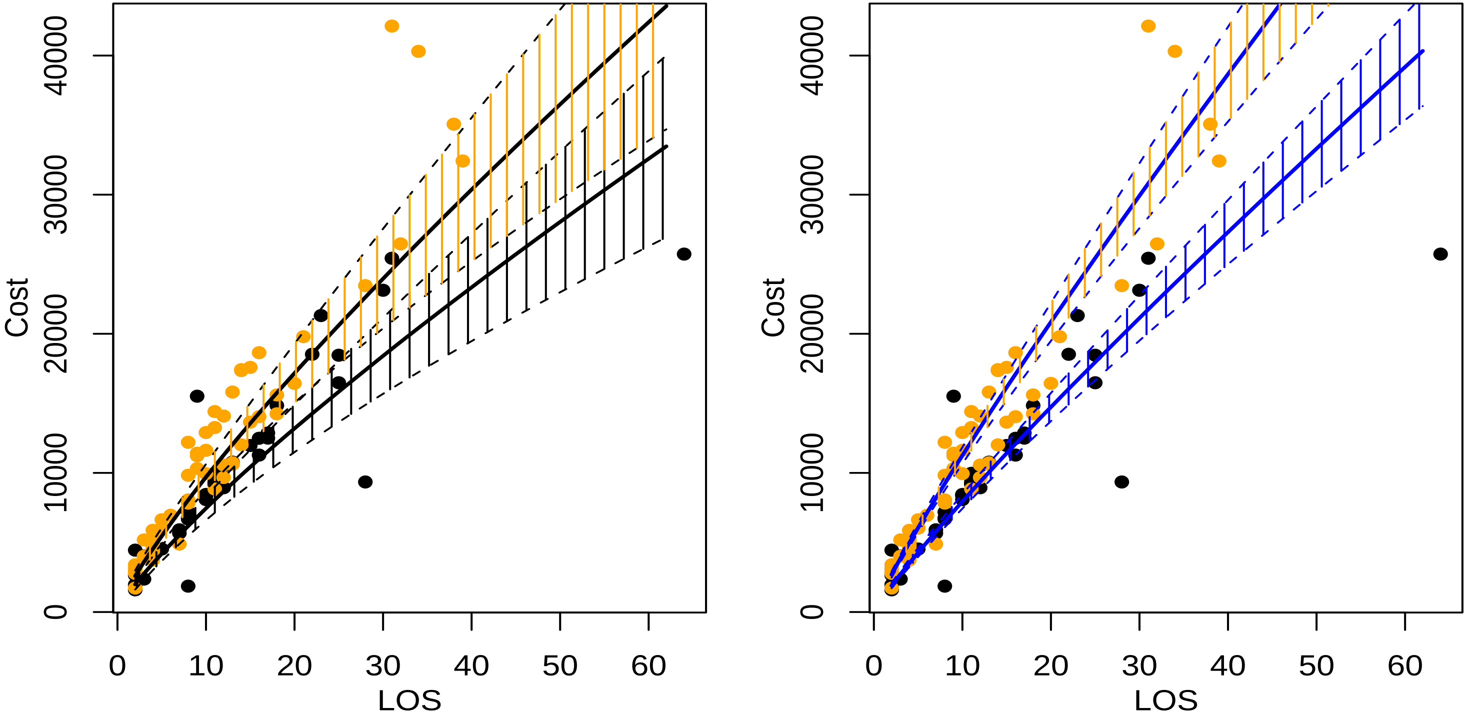
Classical mean = 24.4 days

Robust mean = 3.6 days

Mean cost depending on LOS and Admission type

Robust regression with asymmetric response (Marazzi-Yohai, 2004)

APDRG 243: Medical back problems, 100 cases

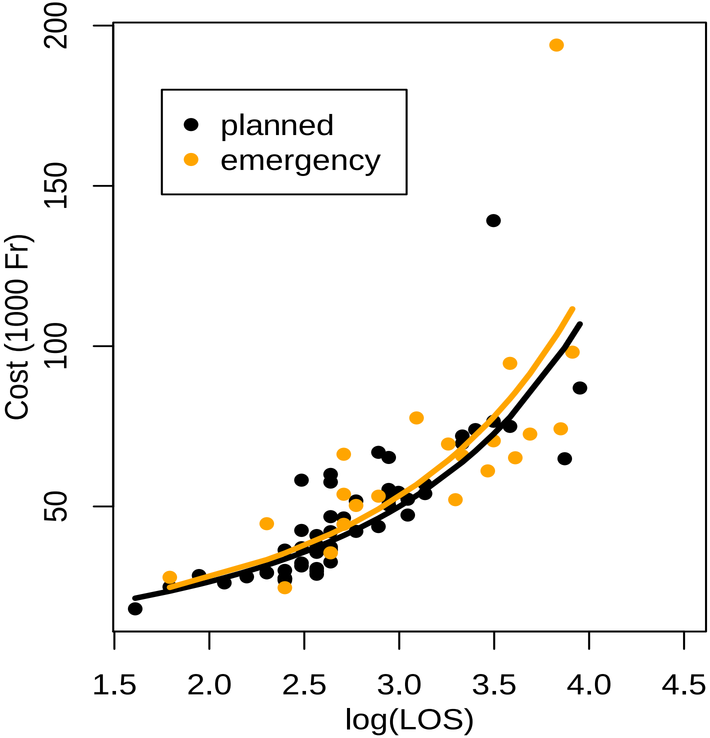


Classical regression   Robust regression

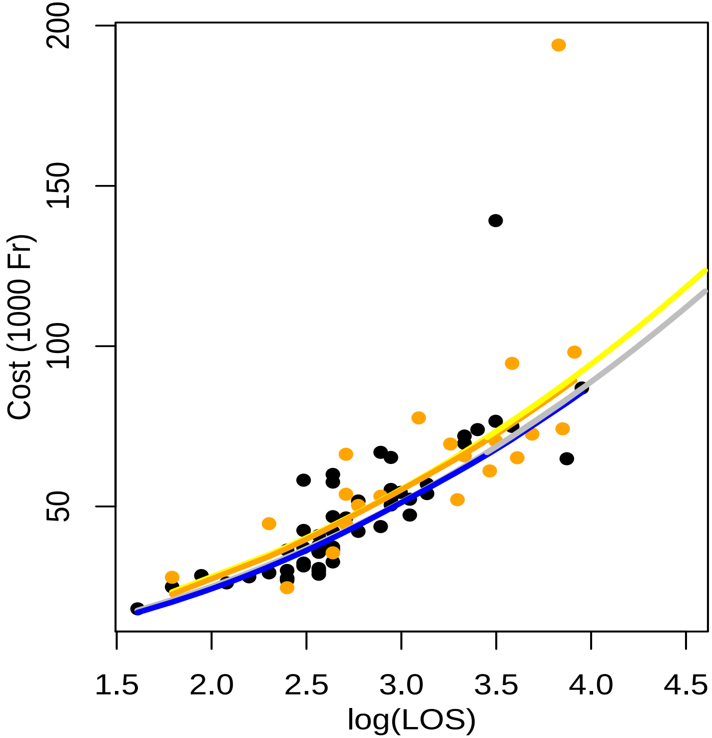
Mean cost depending on LOS and Admission type

Robust Box-Cox transformations and robust smearing (Marazzi-Yohai, 2009)

APDRG 545: cardiac valve procedure with major cc, 78 cases



Classical smearing

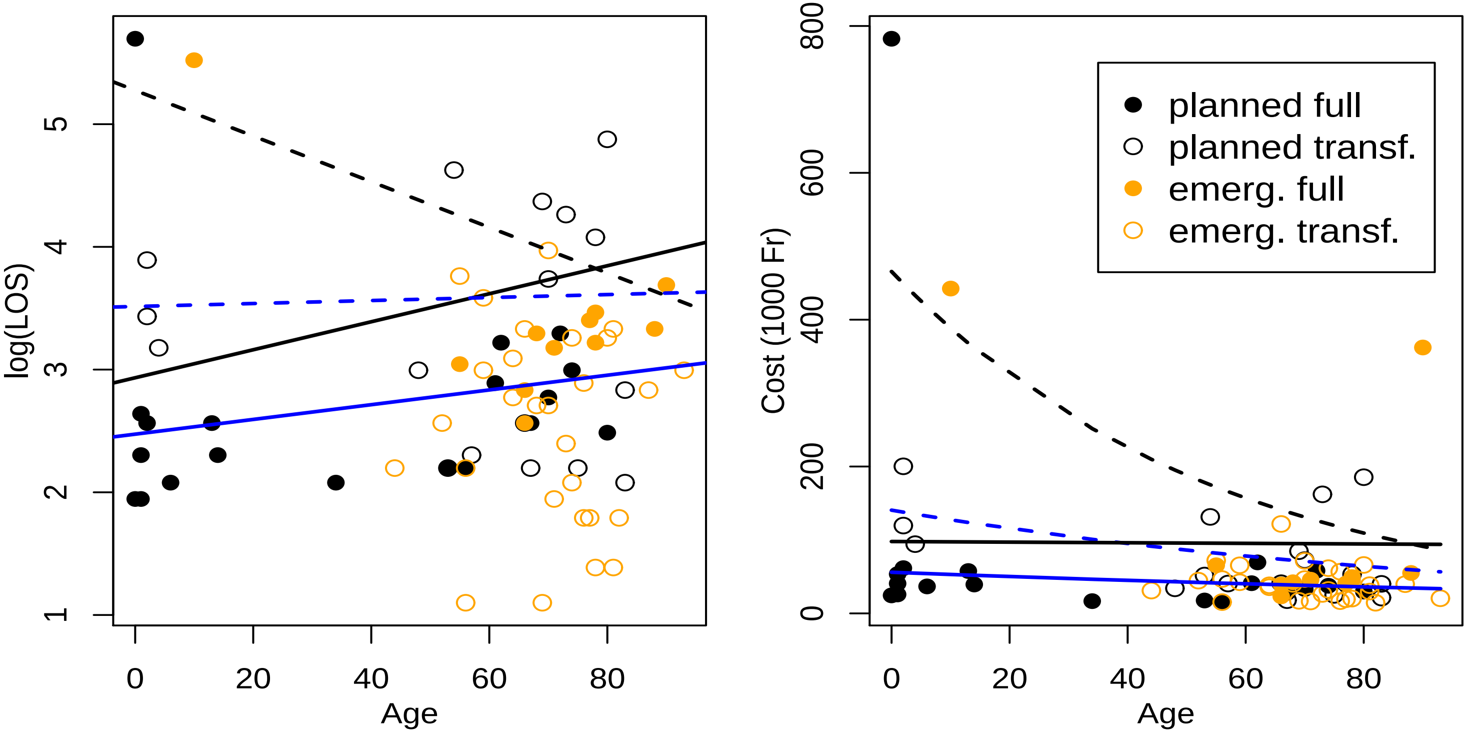


Robust smearing
and Parametric prediction

Mean LOS/Cost depending on Age and Admission type

Robust AFT with censored data (Locatelli-Marazzi-Yohai, 2010; 2011)

APDRG 549: major cardiovascular proc. with major cc, 75 cases, 45 transfers



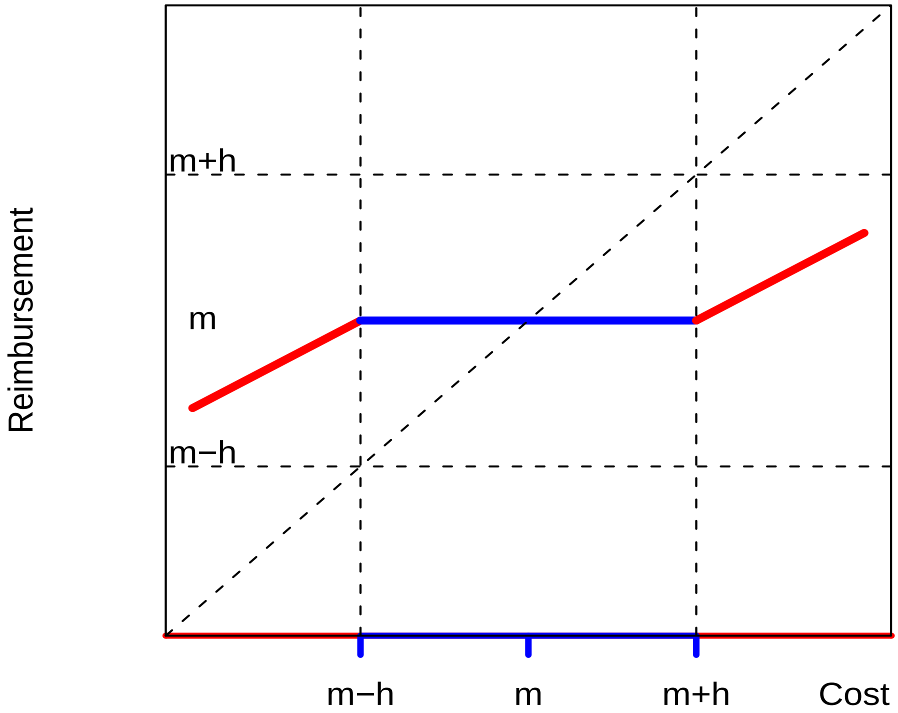
– Class.plann; - - Class.emerg.
– Rob. plann; - - Rob. emerg.

– Class.plann; - - Class.emerg
– Rob. plann; - - Rob. emerg.

6. **Reimbursement** of exceptional cases (Marazzi-Gardiol-Duong, 2007)

Let m be the mean Cost

- if $m - h < \text{Cost} < m + h$ pay m
- else pay $0.9 \times (\text{Cost} - m)$



Adapted to partial reimbursement of LOS

7. Conclusions

- Several robust statistical procedures for the analysis of

  positive responses with asymmetric distribution

  and for the computation of their robust means with/without covariates, with/without censoring have been developed.

  These are general statistical procedures for responses like:
  cost, expenses, survival data, production measures, ...

- Robust statistics is a very active area of statistical research.

- In the analysis of DRG data we have shown that robust statistics is necessary for:
  - the automatic detection of outliers;
  - the computation of LOS and cost means not affected by outliers taking into account:
    the shape of the distributions,
    the available covariates,
    censoring, e.g., transfers to another hospital;
  - to discover and quantify certain types of heterogeneities.

- Introduction of this technology in practice ?
  SwissDRG ?