Are there 15 different ways of thinking of imprecise regression?

Jochen Einbeck
jochen.einbeck@durham.ac.uk
Durham University

Workshop on Principles and Methods of Statistical Inference with Interval Probability — München, 14th September 2009

What is imprecise in “imprecise regression”

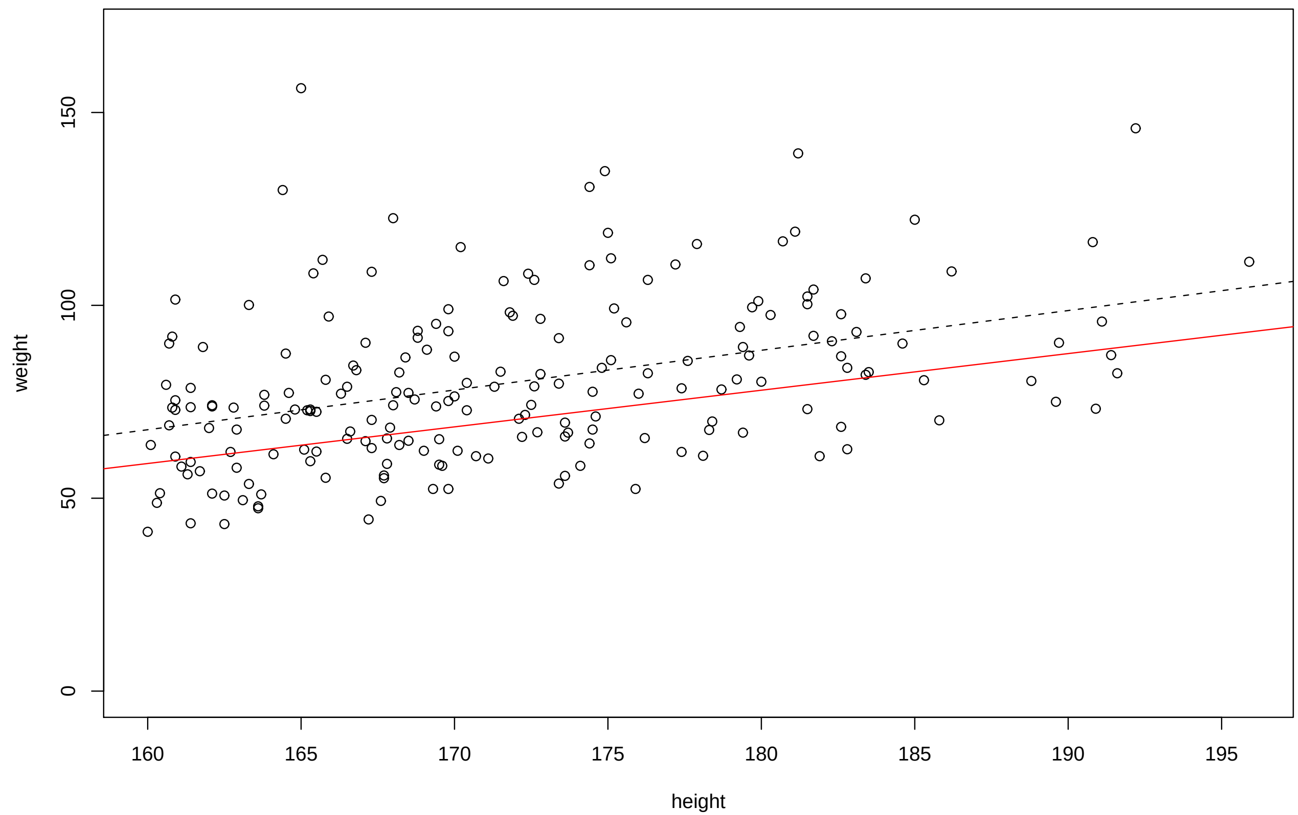
- Thomas Augustin, 13th May 2008, Durham:

  Imprecision in regression can arise in form of ...

  I. Prior knowledge on some parameters,
  II. Data,
  III. Regression line.

- Is this list comprehensive?

Robert's imprecise regression line



Hable (2009), ISIPTA Proceedings, Figure 5: Regression lines for the real data set NHANES obtained by the minimum distance estimator (red line) and by the least-squares-estimator (dashed line).

The minimum distance estimator

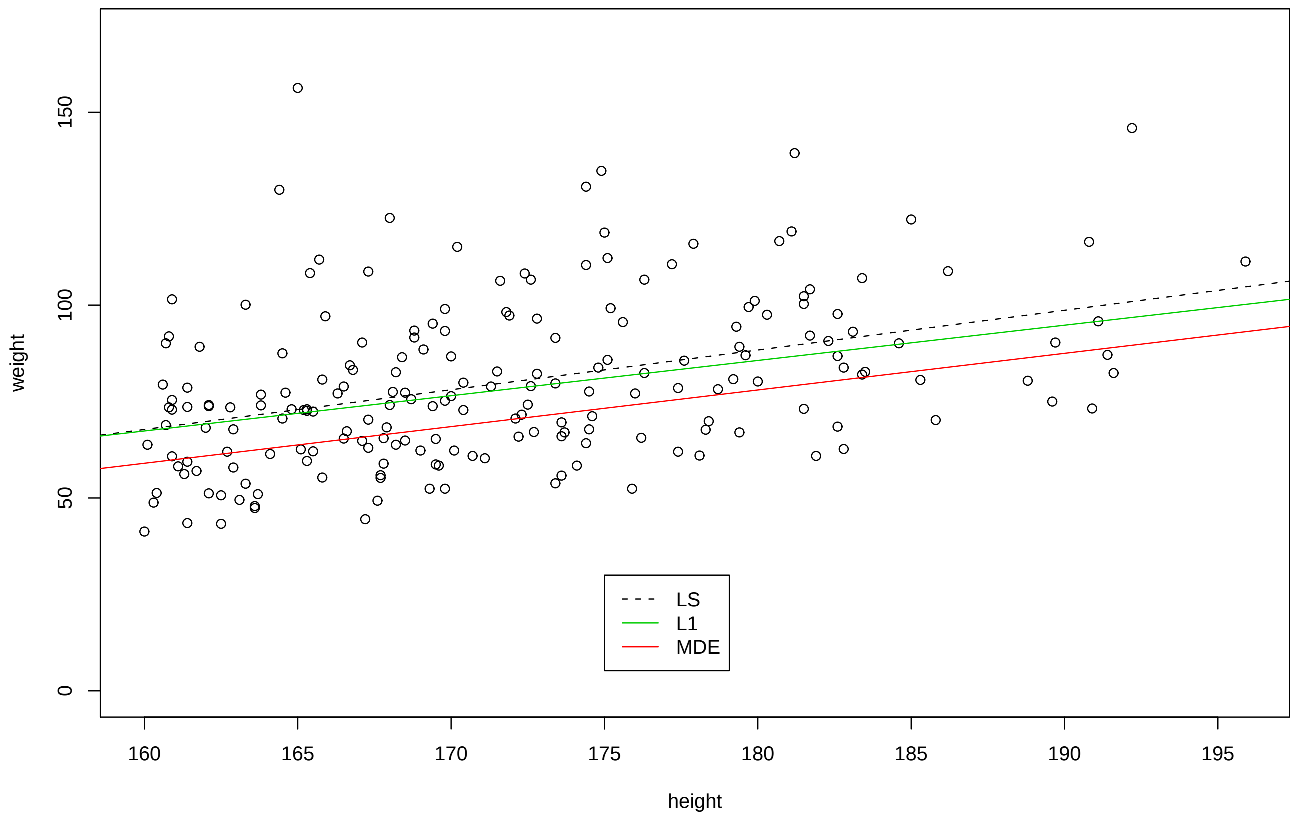
- Given $x_1, \dots, x_n \sim_{iid} P_0$.
- One specifies an imprecise model $(\bar{P}_\theta)_{\theta \in \Theta}$, where each $\bar{P}_\theta$ is a coherent upper provision on a sample space $(\mathcal{X}, \mathcal{B})$ with credal set $\mathcal{M}_\theta$.
- A true parameter is any $\theta_0 \in \Theta$ such that $P_0 \in M_{\theta_0}$.
- Find $\hat{\theta}$ by mimimizing the distance $\inf_{P \in \mathcal{M}_\theta} || \frac{1}{n} \sum \delta_{x_i} - P_\theta ||$ over $\theta \in \Theta$ (linear programming, R package **imprProbEst**).
- Note: The imprecise model has to be set up for every θ. For linear regression this requires to set up a model for every (discrete) combination of intercept and slope that is supposed to be considered. For the Nhanes data, this meant that 55000 (!) list elements had to be passed as arguments to the R function!
- Citation Robert: *“Das ist übrigens ein grundsätzliches Problem, über das man innerhalb der imprecise probability Gemeinde mal diskutieren sollte.”*

The minimum distance estimator

- The MDE looks more “robust” than the LS estimator.
- Actually, there are two differences between LS and MDE
  - one is precise, the other one imprecise.
  - one uses squared, the other one absolute differences.
- Is the difference between the two lines maybe only due to the latter (robustness) effect?
- Need to compare to “robust” L1 estimator

$$\sum |y_i - a - bx_i| \longrightarrow \min$$

MDE and L1 regression



- We observe that the imprecise line is indeed intrinsically different to the precise lines (whether robust or not).

A fourth notion of imprecision in regression

- In which sense is the MDE estimator imprecise?
  - I. Priors? ↯ There are no priors.
  - II. Data? ↯ We have to use precise data.
  - III. Regression line? ↯ The MDE yields (normally) *one* regression line.

  Hence, a fourth class is needed, in which all of the above are precise, but estimation is based on an

  - IV. Imprecise probability model.
- Denotational ambuigity here: Models with imprecise prior (impr. Dirichlet, iLUCK, etc.) have often been labeled as "imprecise probability models" as well.

Imprecise, and imprecise prior, models

- Useful citation: Utkin, Zatenko & Coolen (ISIPTA 09 proc.) write over imprecise Bayesian models:

"Typically, a precise parametric model is assumed, with imprecision following though the use of sets of conjugated priors"

- Hence, we distinguish in what follows beteeen
  - imprecise probability models, where imprecision enters through imprecise modelling (via credal set) of the model parameter.
  - imprecise prior (or imprecise Bayesian) models, where imprecision enters (usually) through sets of conjugate priors.

Four notions on two levels

- We classify the four notions of imprecise regression:
  - I. (Prior) and IV. (Model) concern the probability level.
  - II. (Data) and III. (Regression line) concern the statistics level.
- In total, we have 2^4 possible combinations of I-IV (one of which entirely precise), which can be visualized in form a 4×4 grid:

The imprecise regression grid

Probability level		Imprecise probability model: no		Imprecise probability model: yes	
	Statistics level	**Imprecise data: no**	**Imprecise data: yes**	**Imprecise data: no**	**Imprecise data: yes**
Imprecise Prior	**Imprecise Regression Line**				
no	no	Frequentist regression		*Hable* (MDE)	
no	yes	*Einbeck* (cropped loss)		*Skulj* (NPI) *Augustin* (Credal ML)	
yes	no	*Utkin/Zat./Coolen* (ML+impr.Bayes)	*Utkin (?)*		
yes	yes	*Walter/Augustin* (iLUCK) *Vernon/Goldst.* (Bayes Lin.)			

Question: Which methods are “statistically precise”, i.e. from a “black-box” point of view, they take precise data and produce a single regression line?

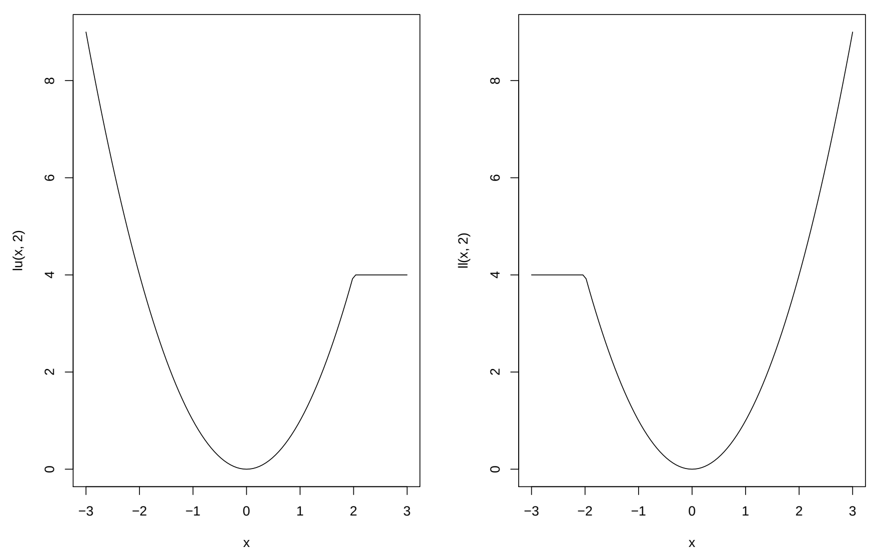
Probability level		Imprecise probability model			
		no		yes	
	Statistics level	Imprecise data			
		no	yes	no	yes
Imprecise Prior	**Imprecise Regression Line**				
no	no	Frequentist regression		*Hable* (MDE)	
no	yes	*Einbeck* (cropped loss)		*Skulj* (NPI) *Augustin* (Credal ML)	
yes	no	*Utkin/Zat./Coolen* (ML+impr.Bayes)	*Utkin (?)*		
yes	yes	*Walter/Augustin* (iLUCK) *Vernon/Goldst.* (Bayes Lin.)			

Question: Which methods are “probabilistically precise”, i.e. they use only techniques based on precise probability?

Probability level		Imprecise probability model: no		Imprecise probability model: yes	
	Statistics level	Imprecise data: no	Imprecise data: yes	Imprecise data: no	Imprecise data: yes
Imprecise Prior	**Imprecise Regression Line**				
no	no	Frequentist regression		*Hable* (MDE)	
no	yes	*Einbeck* (cropped loss)		*Skulj* (NPI) *Augustin* (Credal ML)	
yes	no	*Utkin/Zat./Coolen* (ML+impr.Bayes)	*Utkin (?)*		
yes	yes	*Walter/Augustin* (iLUCK) *Vernon/Goldst.* (Bayes Lin.)			

The frequentist sandpit to imprecise regression

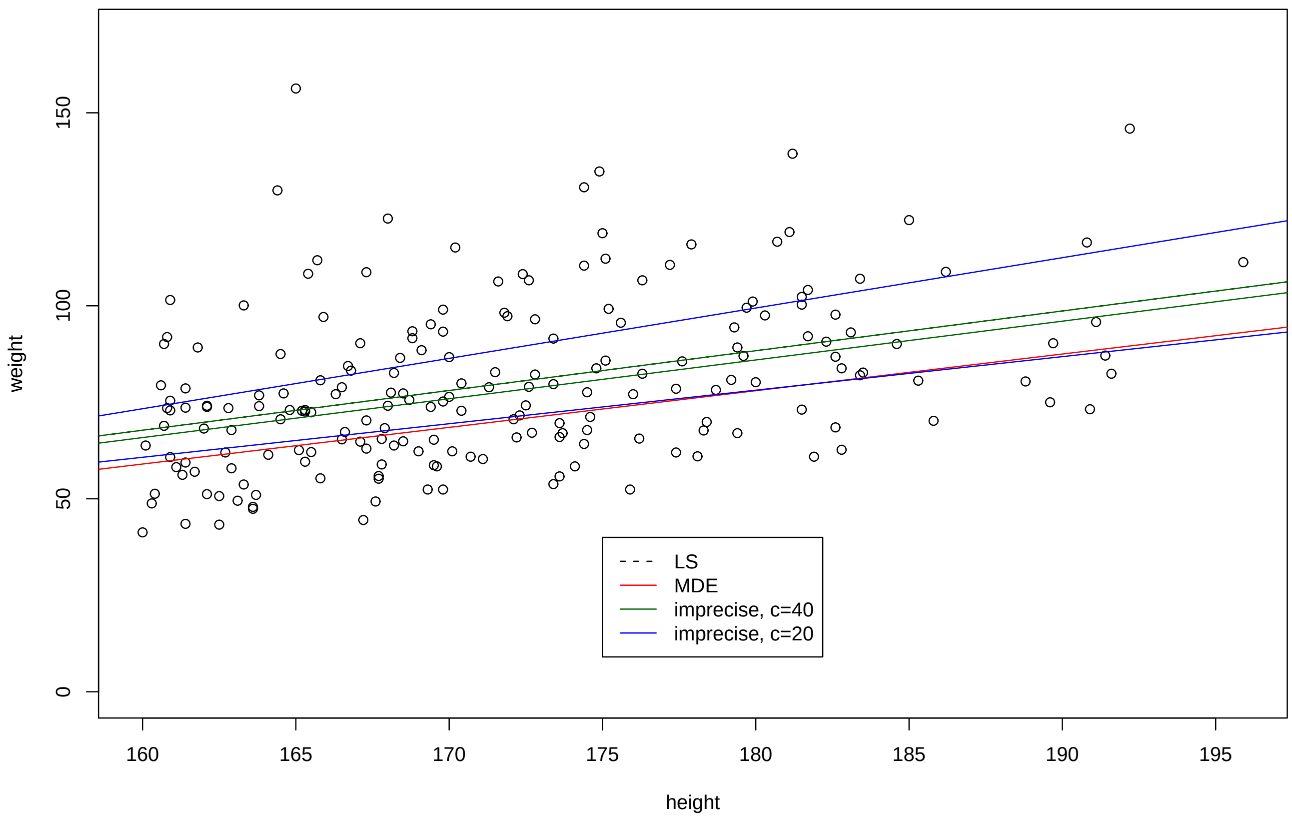
- Einbeck, Durham Workshop, 2008: Notionally, imprecision is seen as a consequence of lack of trust into (some part of) the data at hand.
- Essentially, what is used is a pair of cropped loss functions, each of which distrusts (not dismisses!) outliers on either side of the data cloud:



MDE and imprecise “frequentist” regression

Consider imprecise regression with cropping at

- c=40 (almost no cropping – no “lack of trust”)
- c=20 (strong cropping – strong “lack of trust”)



Back to the imprecise probability grid:

Probability level		Imprecise probability model: no		Imprecise probability model: yes	
	Statistics level	Imprecise data			
Imprecise Prior	Imprecise Regression Line	no	yes	no	yes
no	no	Frequentist regression		*Hable* (MDE)	
no	yes	*Einbeck* (cropped loss)		*Skulj* (NPI) *Augustin* (Credal ML)	
yes	no	*Utkin/Zat./Coolen* (ML+impr.Bayes)	*Utkin (?)*		
yes	yes	*Walter/Augustin* (iLUCK) *Vernon/Goldst.* (Bayes Lin.)			

From a viewpoint of effectiveness, the diagonal seems to be attractive?

Probability level		Imprecise probability model: no		Imprecise probability model: yes	
	Statistics level	Imprecise data: no	Imprecise data: yes	Imprecise data: no	Imprecise data: yes
Imprecise Prior	**Imprecise Regression Line**				
no	no	Frequentist regression		*Hable* (MDE)	
no	yes	*Einbeck* (cropped loss)		*Skulj* (NPI) *Augustin* (Credal ML)	
yes	no	*Utkin/Zat./Coolen* (ML+impr.Bayes)	*Utkin (?)*		
yes	yes	*Walter/Augustin* (iLUCK) *Vernon/Goldst.* (Bayes Lin.)			

A word on residuals

- Hardly any mentioning of the word “residual” so far in imprecise regression papers.
  - What is an imprecise residual?

A word on residuals

- Hardly any mentioning of the word "residual" so far in imprecise regression papers.
  - What is an imprecise residual?
- Firstly, let's go one step back:
  - What is a precise residual?

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \hat{E}(y_i|x_i)$$

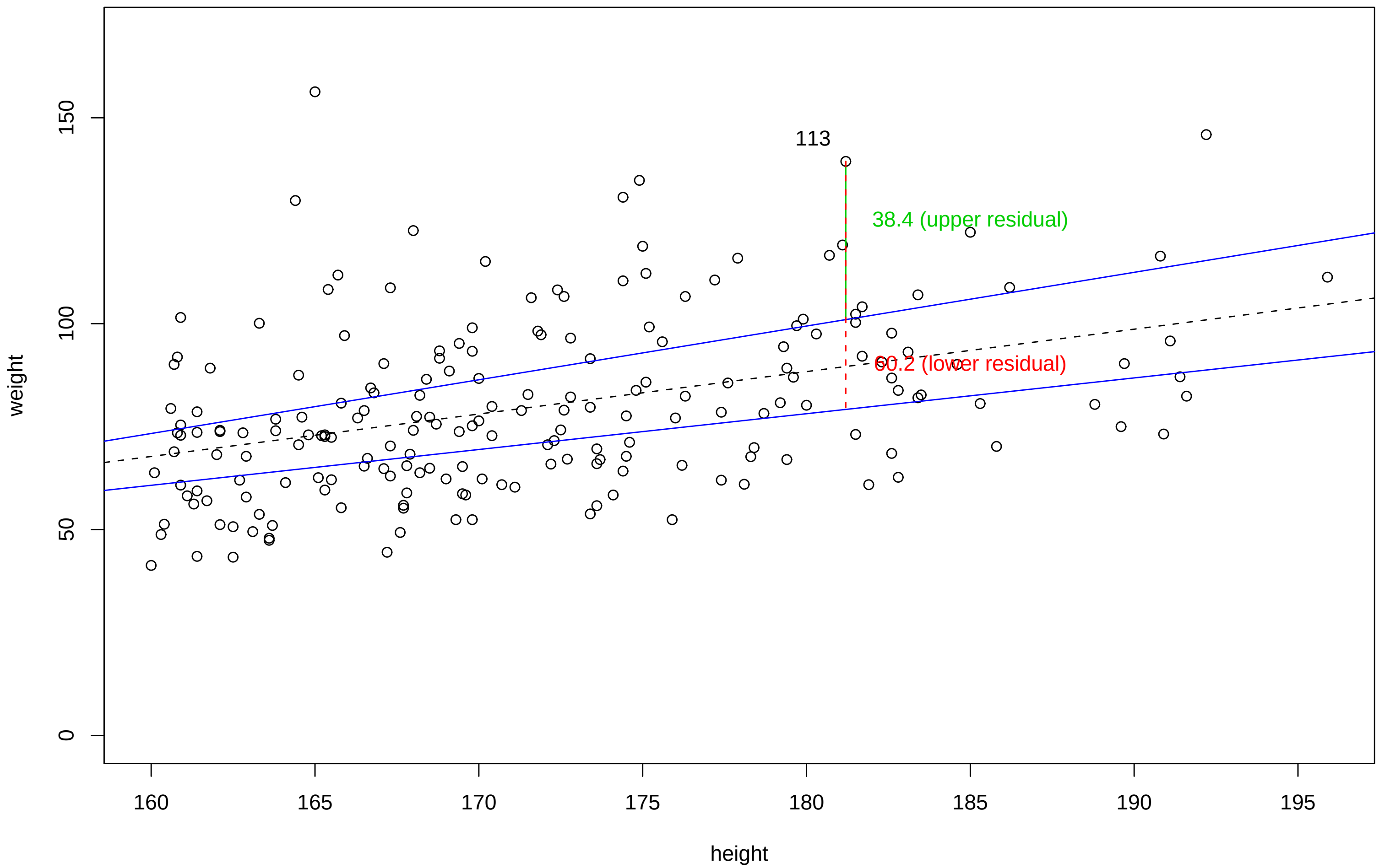
  - Here imprecision enters straightforwardly through the use of upper and lower expectations, giving upper and lower residuals:

$$\begin{aligned} \bar{\epsilon}_i &= y_i - \bar{E}(y_i|x_i) \\ \underline{\epsilon}_i &= y_i - \underline{E}(y_i|x_i) \end{aligned}$$

Upper and lower residuals

- At each point x_i, the upper residual gives the distance to the upper (highest) estimate, and the lower residual gives the distance to the lower (smallest) estimate.

regression with cropped loss functions, c=20

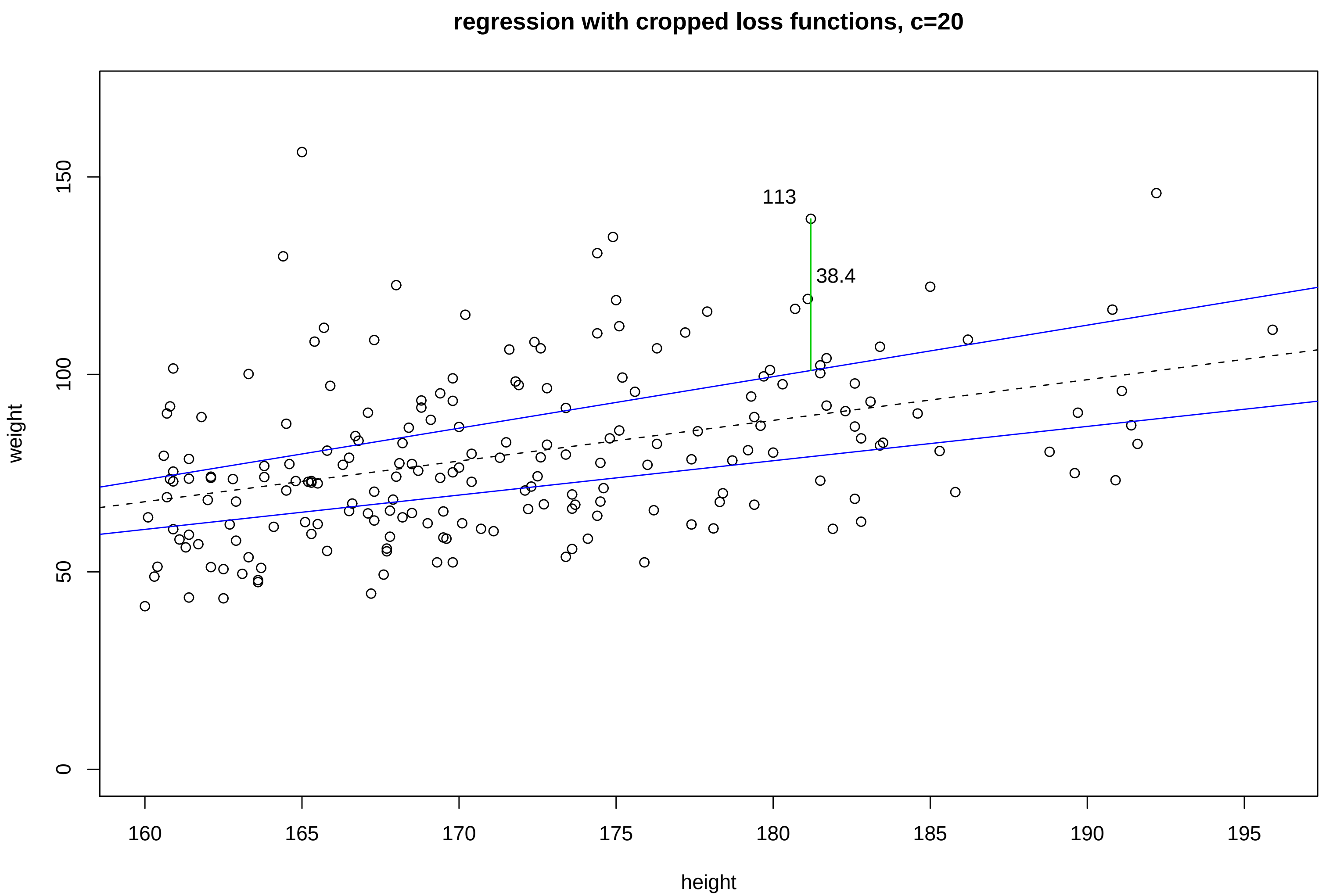


A word on residuals (2)

- Actually, this notion of residuals seem sort of strange:
  - For any data position, either the lower and the upper residual (or both) are not really meaningful.
  - In fact, do we want two residuals?
- Note that for each data point i there are three cases to consider:
  - y_i is above all lines. Then $0 < \bar{\epsilon}_i < \underline{\epsilon}_i$.
  - y_i is below all lines. Then $0 > \underline{\epsilon}_i > \bar{\epsilon}_i$.
  - y_i is somewhere in between.
- Define naive residuals

$$\tilde{\epsilon}_i = \begin{cases} \bar{\epsilon}_i & \text{if} \quad \bar{\epsilon}_i > 0 \\ 0 & \text{if} \quad \text{sign}(\bar{\epsilon}_i \underline{\epsilon}_i) < 0 \\ \underline{\epsilon}_i & \text{if} \quad \underline{\epsilon}_i < 0 \end{cases}$$

Naive residuals



- Here, the naive and the upper residuals are the same.

Some questions for today

- Based on the imprecise probability grid, can we reach an agreement
  - which boxes are particularly worthwhile to follow?
  - which boxes (columns/rows) seem to be rather useless?
  - which box(es) would correspond to **the** colloquial or scientific meaning of *imprecise regression*?
- When using an imprecise probability model, do we have an idea how we can avoid discretization of the parameter space?
- Is there a natural definition/interpretation for imprecise residuals?
- Do we know what we man with *imprecise data* ?

References

Augustin, T. (2009). Towards a notion of Credal Maximum Likelihood — Imprecise Probability-based Mixed models and unobserved heterogeneity. ISIPTA '09, Poster / Book of Abstracts p. 18.

Hable, R. (2009). A Minimum Distance Estimator in an Imprecise Probability Model — Computational Aspects and Computation. ISIPTA '09 Proceedings, 209–218.

Utkin, L.V, Zatenko, S.I., and Coolen, F.P.A (2009). Combining imprecise Bayesian and maximum likelihood estimation for reliability growth models. ISIPTA '09 Proceedings, 421–430.

Walter, G., Augustin, T., and Peters, A. (2007). Linear regression analysis under sets of conjugated priors. ISIPTA '07 Proceedings, 445–455.