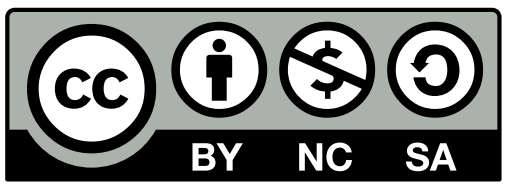
Method Comparison at the SSMT

Dr Giorgio Pioda

Last compiled 18 novembre, 2022



Method Comparison

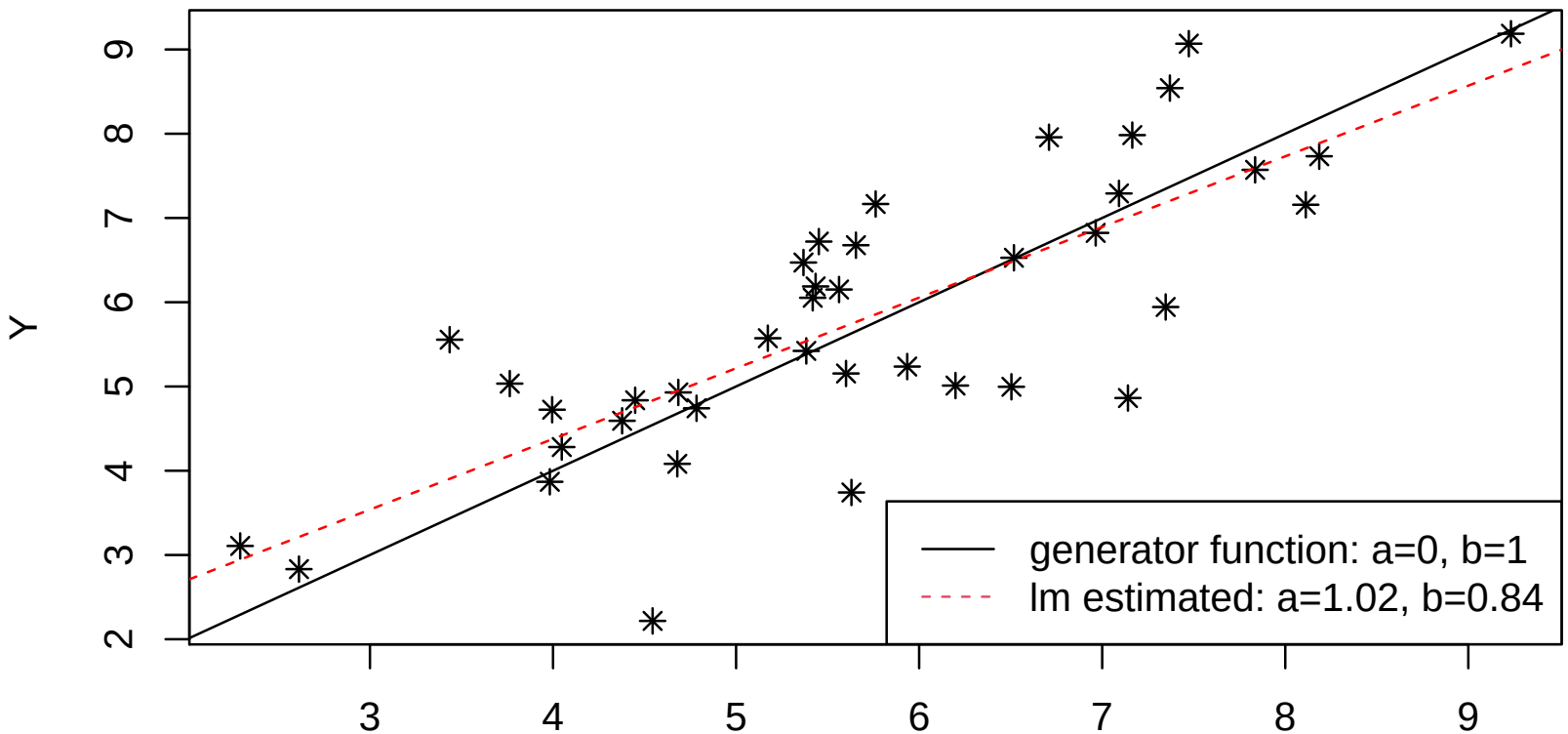
Typical situation: in a laboratory a measurement device/system gets updated with a new one. The new one must be validated against the old one. If validated everything is OK. If not, physicians are informed that the upgrade took place and that a linear transformation of the old patient data is needed to compare them with the new. Data are either:

- Continuous (glucose in blood, ...). Well established methods are available;
- Discrete pos/neg tests or semi quantitative tests (sticks reads, ...). Less established methods.

Reference: ISO 15189 rightlines

The "regression to the mean" problem

- The classical linear regression is not viable.
- The higher the dispersion of the data (the error), the greater the tendency of the lm line to flatten (b<1).
- In absence of data centering also the intercept gets a wrong estimate (a>0)



Continuous data example

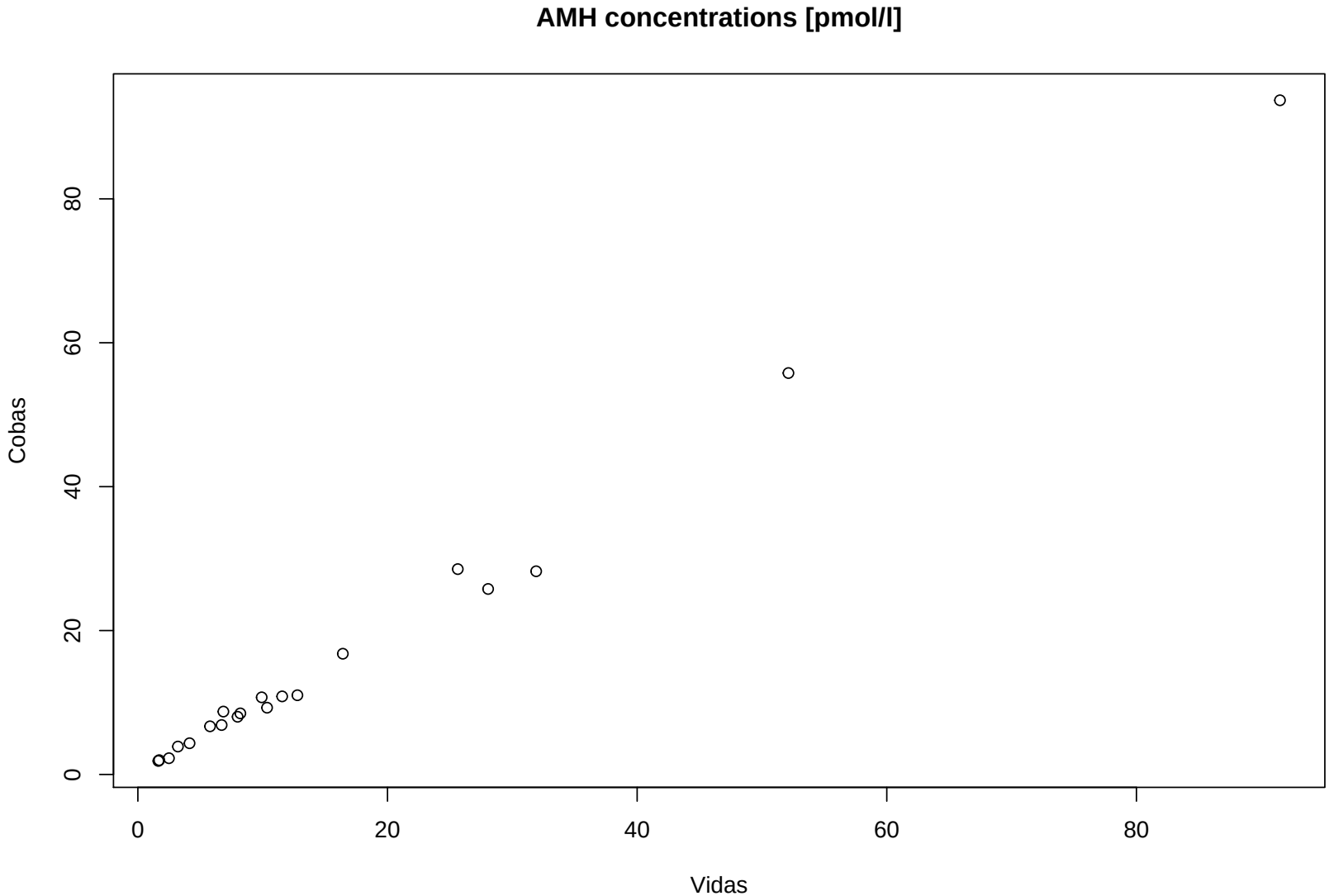
A new device is introduced in the "Centro Cantonale per la fertilità" (part of the EOClab) to measure the AMH (Anti Müllerian Hormone) which is used as biomarker estimator of the ovarian reserves[1]. The new device is compared to an older one. Both use fluorescence based immunoassay methods but with different techniques.

```
head(d.amh)
```

```
##   sample vidas cobas
## 1      1 12.78 11.01
## 2      2 16.42 16.78
## 3      3  2.49  2.26
## 4      4  6.71  6.87
## 5      5  3.21  3.88
## 6      6  1.64  1.88
```

[1] D. Sanfilippo, SSMT bachelor thesis, 2019

AHM simple plot



Methdod comparison goal

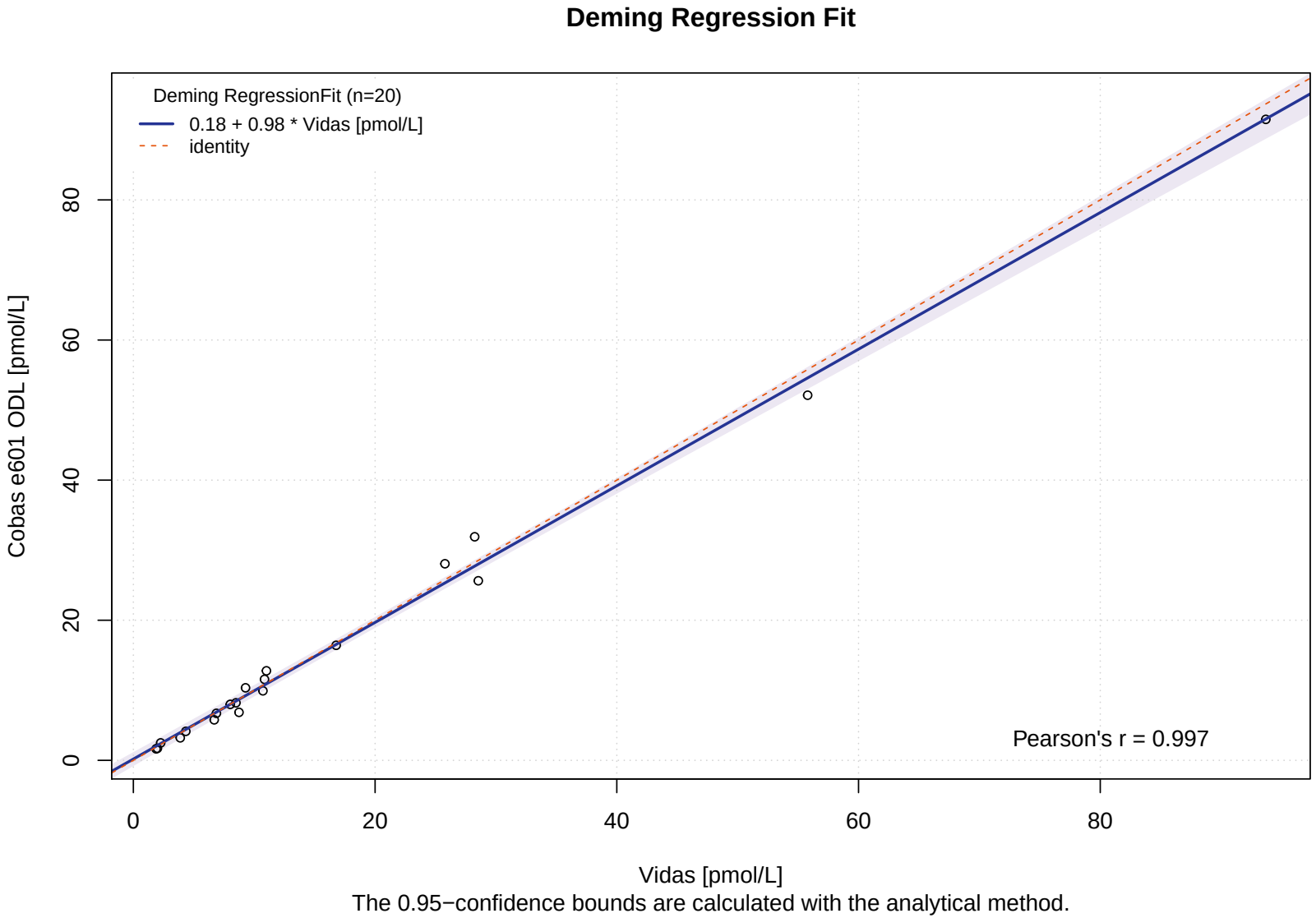
The goal is to confirm that the two devices are equivalent. Several approaches are possible.

- Paired T-test or equivalent procedures (historical method, abandoned)
- Regression methods
- Bland-Altman approach

MethoComp regression methods

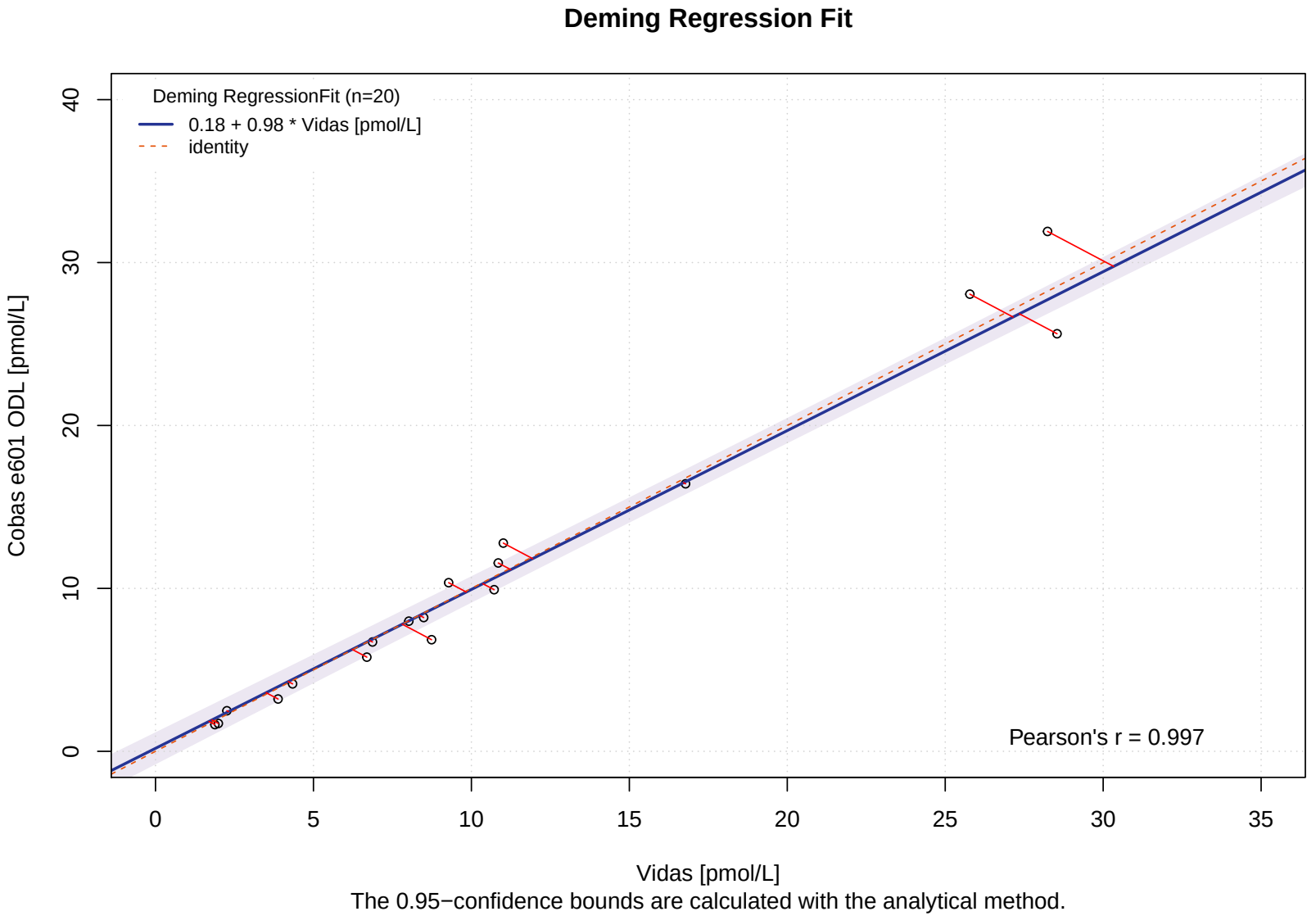
- Both Y_i and X_i are random variable, not only Y_i
- H_0: $\beta_0 = 0$ and $\beta_1 = 1$. Usually tested with confidence intervals at $P < 0.95$
- Deming regression
- Passing Bablok non parametric robust regression
- R packages: mcr, MethComp
- Non ISO advances: mcr with additional functionalities via install_github("piodag/mcr")

Deming example (mcr package)



The 0.95−confidence bounds are calculated with the analytical method.

Deming: the error ratio $\delta = \frac{\sigma_Y^2}{\sigma_X^2}$



The 0.95−confidence bounds are calculated with the analytical method.

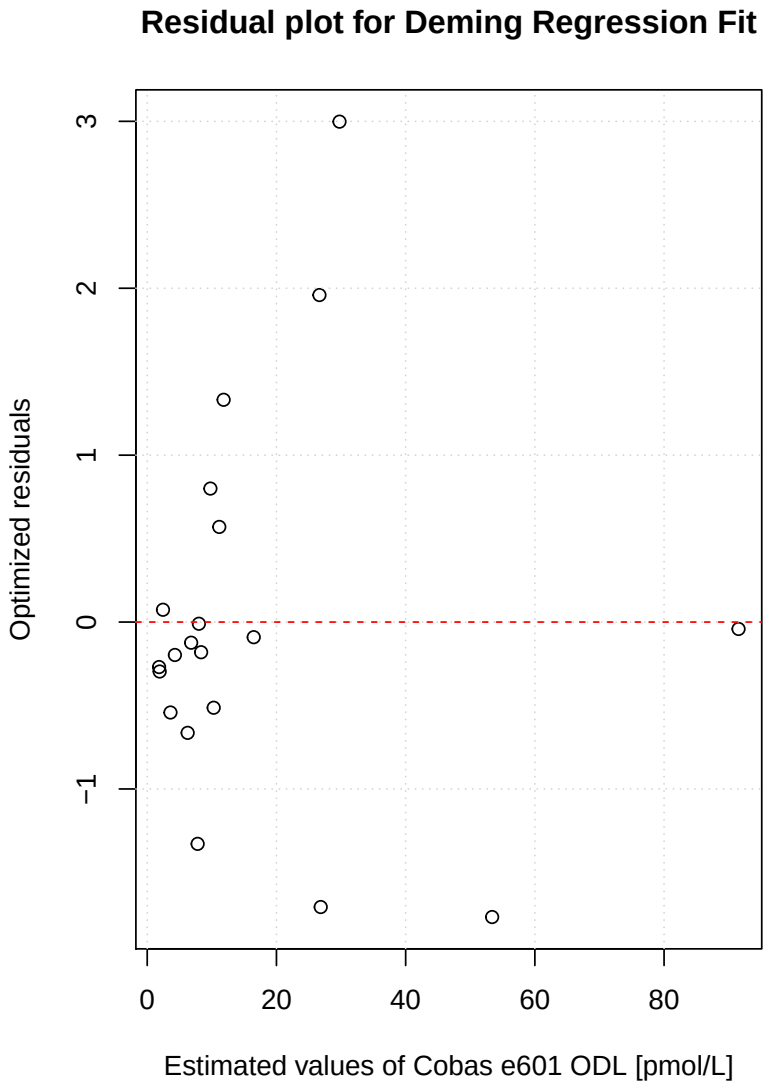
Testing the hypotesis

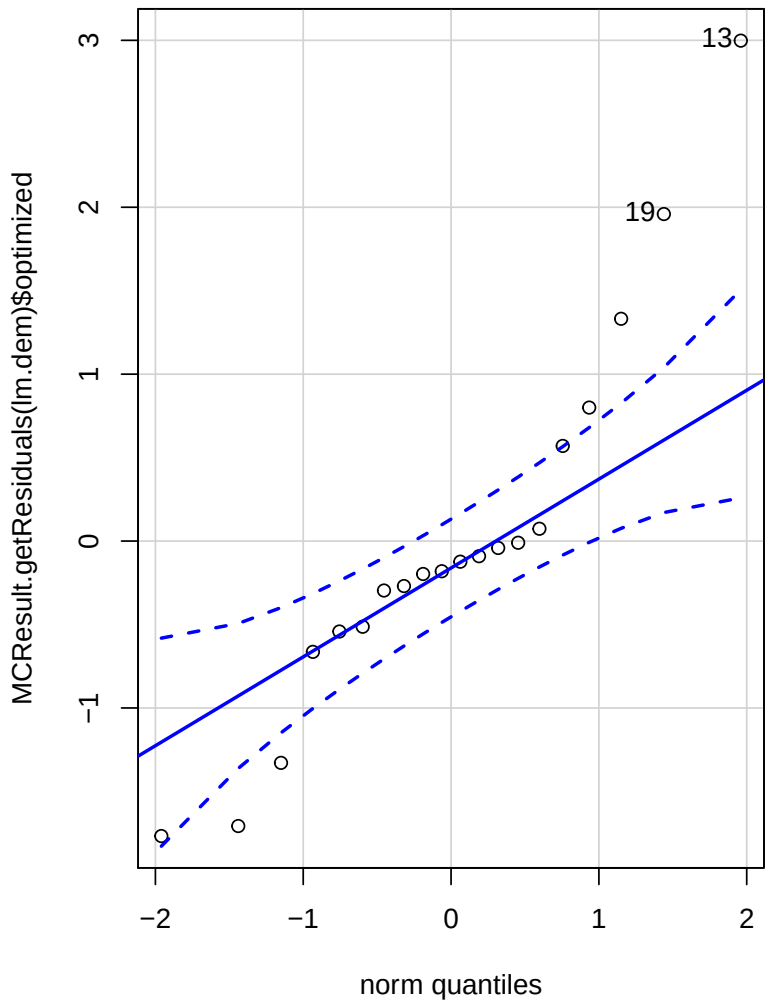

```
library("knitr")
lm.dem.res<-MCResult.getCoefficients(lm.dem)
kable(lm.dem.res)
```

	EST	SE	LCI	UCI
Intercept	0.18318	0.46790	-0.79985	1.1662
Slope	0.97518	0.01695	0.93957	1.0108

In this case the new Vidas method is validated.

Residual plots

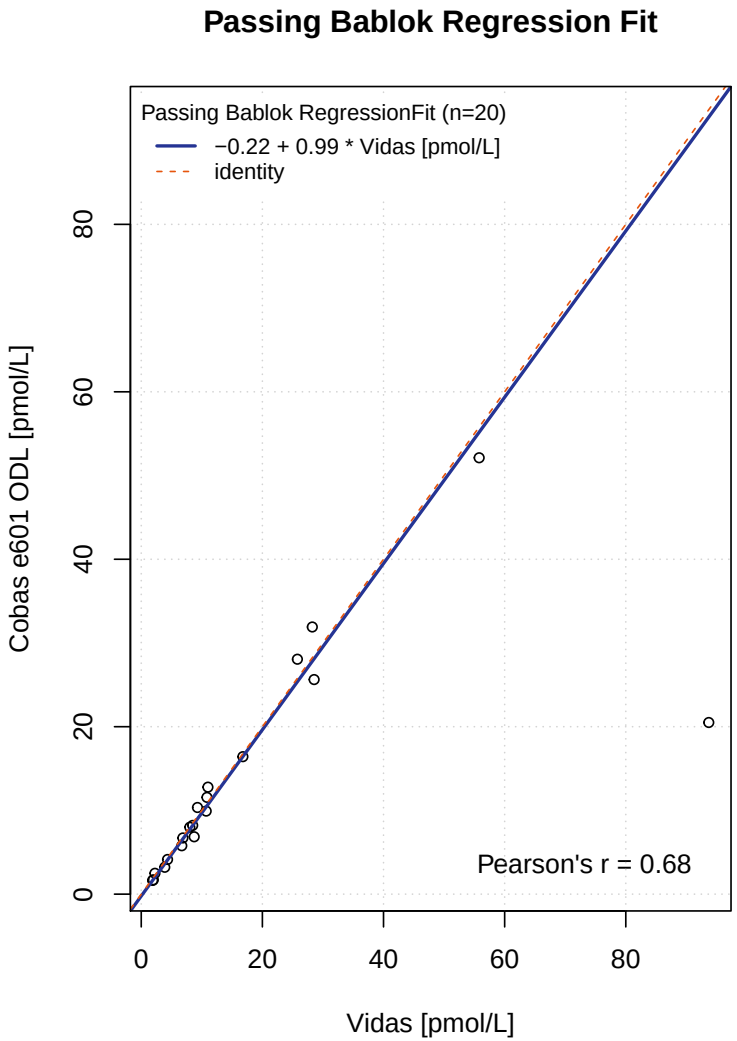


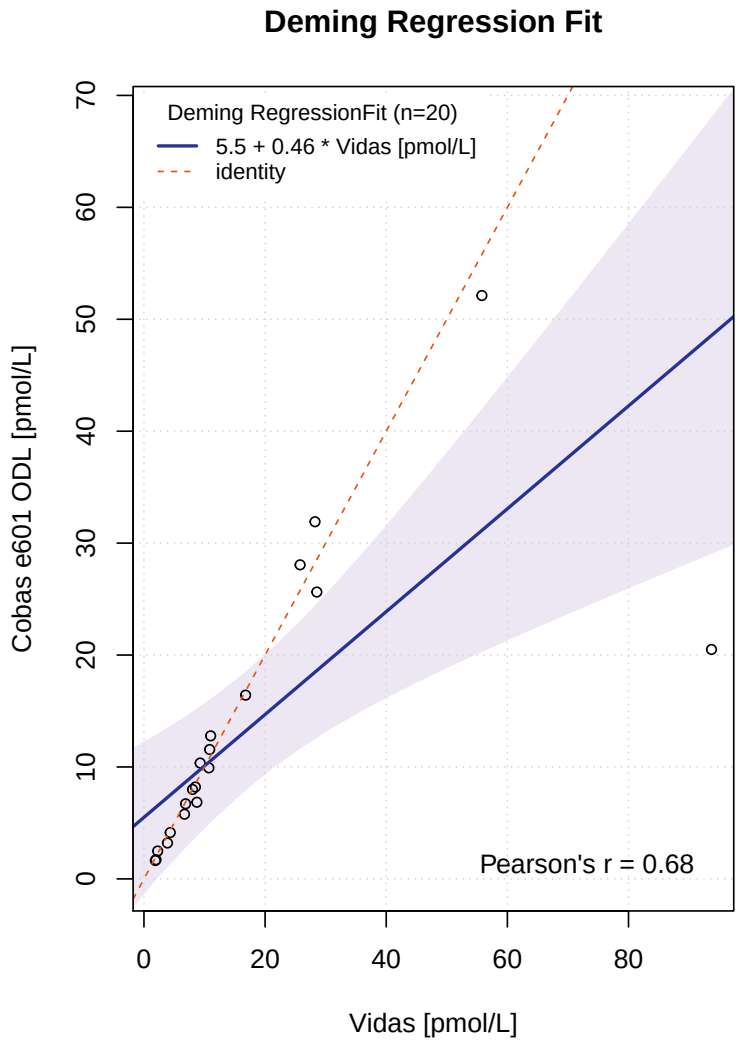


Some violation of the normality of the residuals. Unfortunately the log transformation is not a good solution (because of the intercept).

Passing Bablok regression

Good in the presence of outliers. Here the highest value is modified creating a fancy outlier with very high leverage.





The 0.95−confidence bounds are calculated with the analytical met

Passing Bablok - General infos

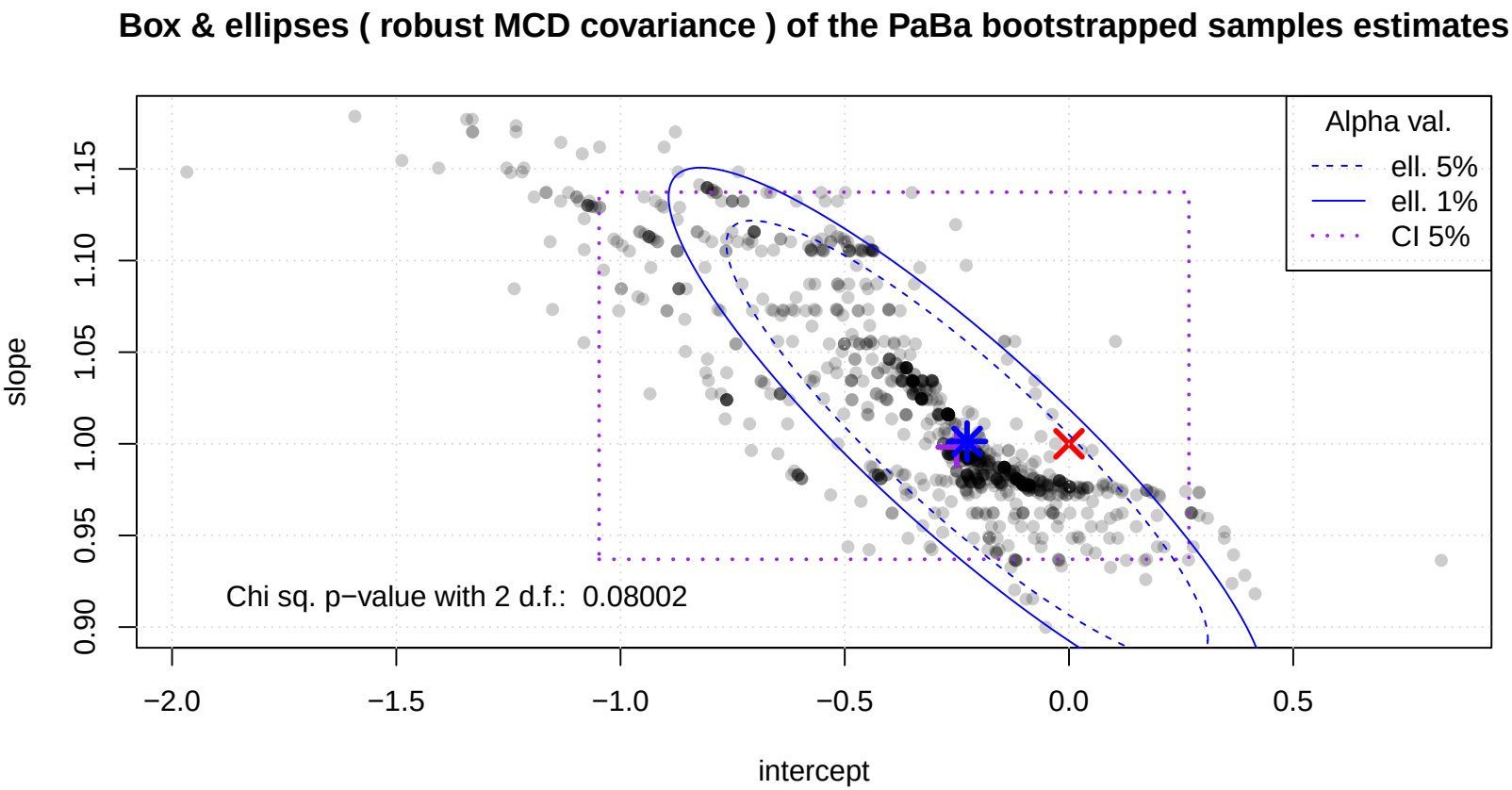
- Non parametric method
- All pairs of possible slopes calculated. The slope is the median of all the slopes (with a special offset that cut out values lower than -1).
- Intercept calculated via slope and x_{mean} and y_{mean}
- Good power with $n > 100$. For preliminary studies accepted $n >= 40$
- For less data the power is low and the nonrobust Deming method is preferred.
- Ugly with low data precision (and ties in the slope pairs): a 4 digit precision is warmly suggested to avoid bias!

Non ISO advances: a visual analysis of the bootstrapped pairs

- Visual inspection of bootstrapped data.
- Problems with ties and bias (PaBa method) easily detected. Accumulation points visible
- Single H_0 testing possible with Mahalanobis distances and robust covariance matrix methods.
- The single H_0 testing has much more power. Good solution for datasets with narrow value range (like K^+ in plasma)

The BoxEllipses plot

- Box of conventional CI
- Ellipses calculated with robust covariance matrix. Chi-squared probability of the MD plotted as help.



BoxEllipses 2

- In ideal cases the cloud of bootstrapped paird should look homogeneous and have an elliptic shape with negative slope.
- Accumulation points are alarming
- Reason for distortions:
  - dataset to small (like above)
  - data precision to low (for PaBa regression)
  - presence of outliers, presence of heteroscedasticity

In the example above the limited dataset is highlighted. CI validation succeeds but the plot looks pretty bad. 20 pairs of measurement are not enough. Ask for more data!

Passing & Bablock ugly side

An example with PoC Hb method comparison on a 100 sample pairs dataset

- Blind bootstrapped PaBa gives a positive slope validation and a weak difference for the intercept.
- Blind analytical PaBa gives the same result.
- BoxEllipses plot of the PaBa bootstrap samples shows how bad the problem of the ties is.

Blind PaBa, analytical and bootstrapped

Table 2: Bootstrap BCa

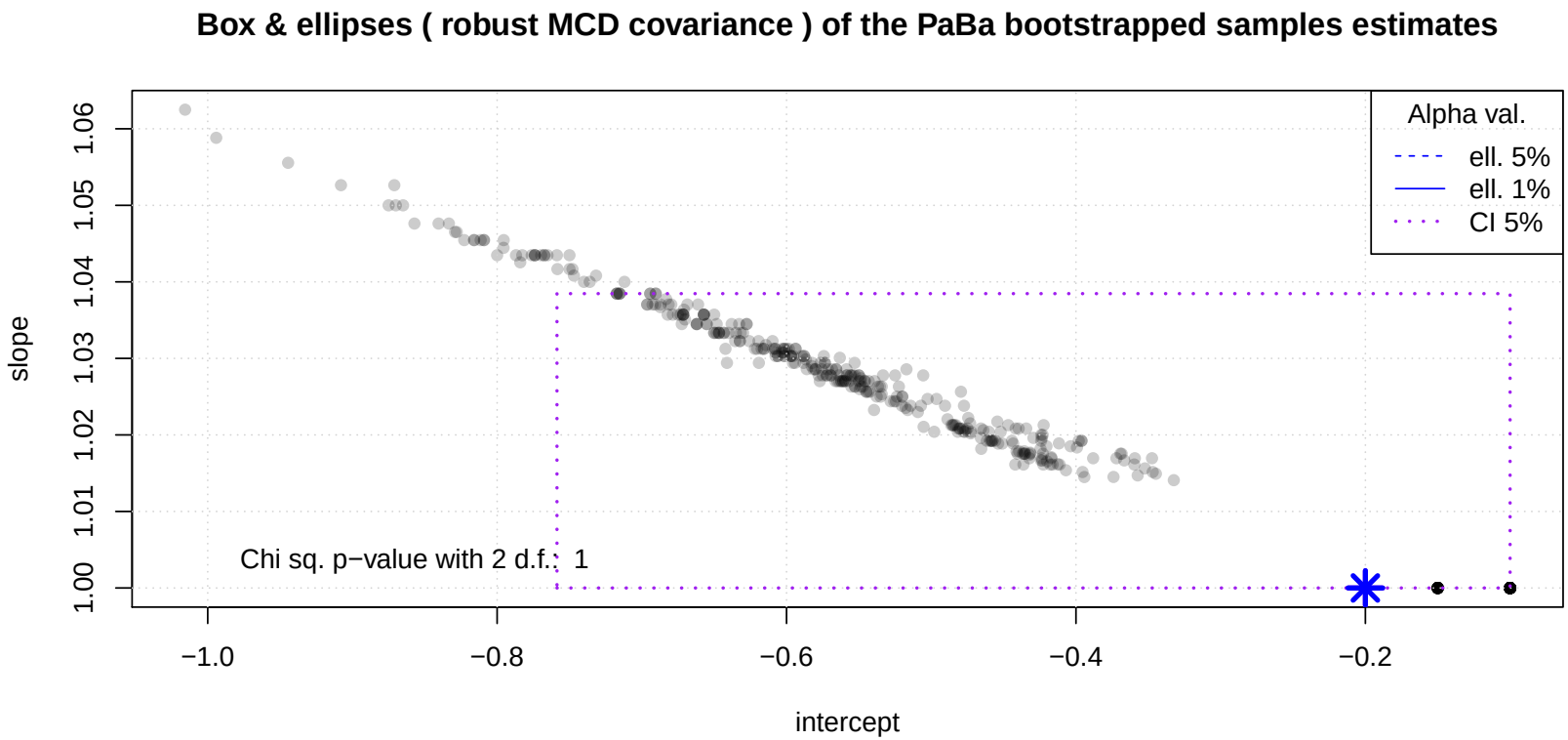
	EST	SE	LCI	UCI
Intercept	-0.2	NA	-0.7586	-0.1000
Slope	1.0	NA	1.0000	1.0385

Table 3: Analytical

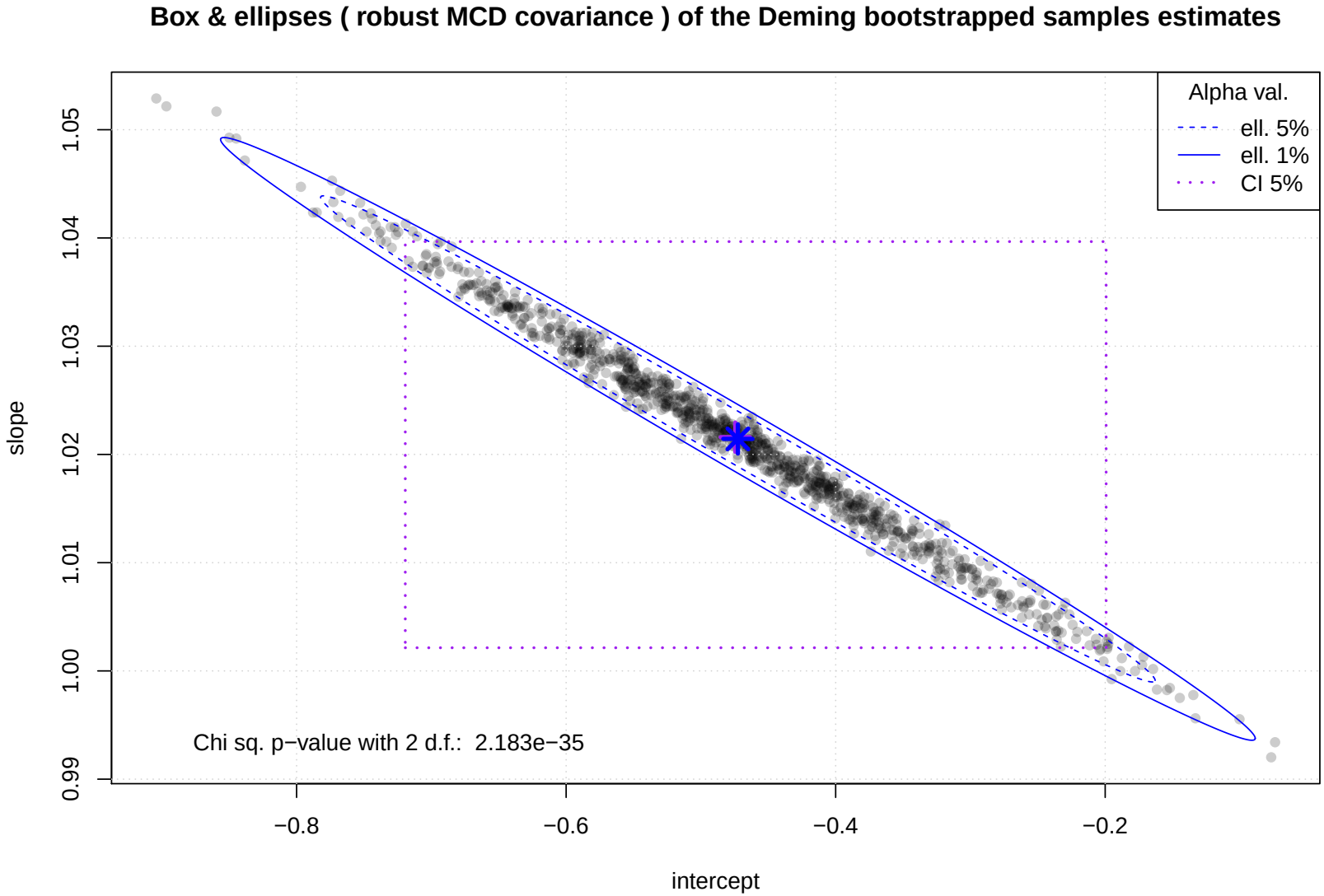
	EST	SE	LCI	UCI
Intercept	-0.2	NA	-0.7563	-0.2000
Slope	1.0	NA	1.0000	1.0417

Ugly Paba BoxEllipses plot

It's not even possible to calculate the covariance matrix and draw the ellipses with MCD algorithm. The box of "classical" has the H_O on the boundary for the slope.



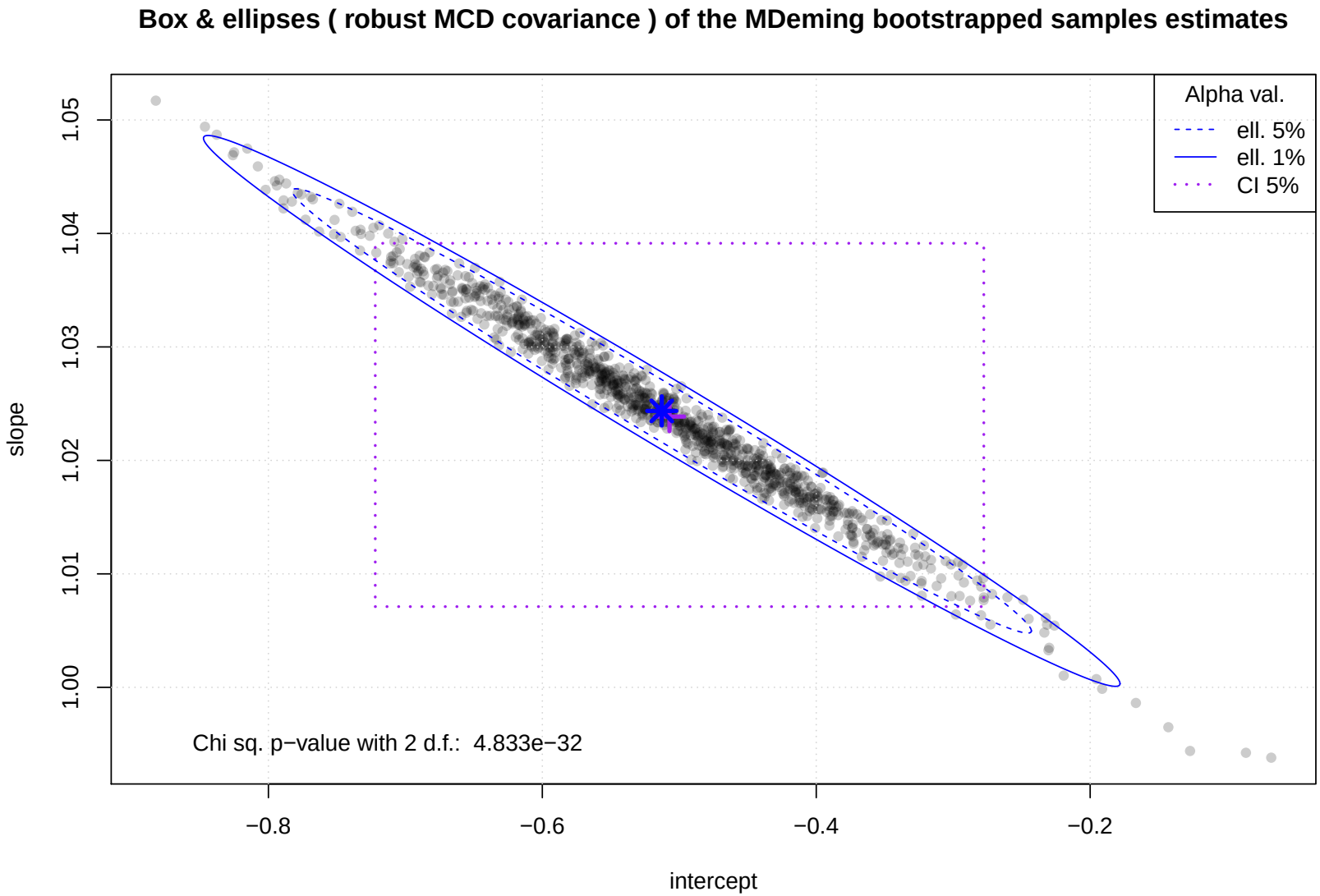
Good (nonrobust) Deming plot



Non ISO: bootstrapped M-Deming (and MM-Deming) regression

- The M-Deming analytical method has problems with the definition of the d.f. and thus for the CI determination.
- The bootstrap procedure circumvent these limits.
- The M-Deming has no bias with low precision datasets. Ideal for thae validation of PoC devices.
- Combination with the BoxEllipses method and the single H_0 testing integrated.
- In presence of extreemely leveraged outliers, consider MM-Deming (also available) with redescending weight function .

Good (robust) M-Deming plot

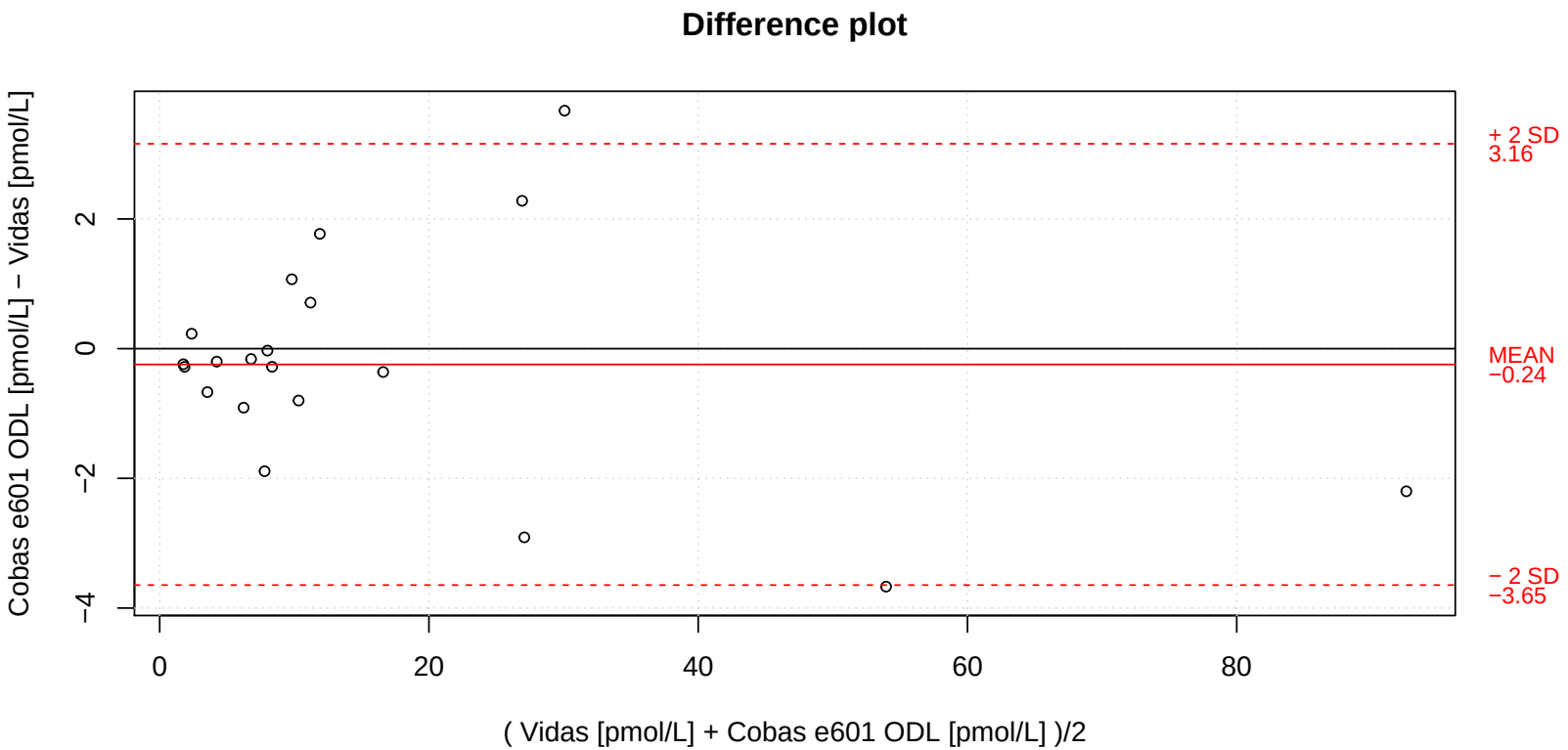


Bland Altman method

- Combines a graphical analysis with a T test. A kind of "pairwise T test on steroids".
- Data is recalculated forming the pairwise mean $m_i = \frac{y_i + x_i}{2}$ of the samples and the pairwise differences $d_i = y_i - x_i$.
- The graphical analysis is similar to the analysis of Tukey-Anscombe plot

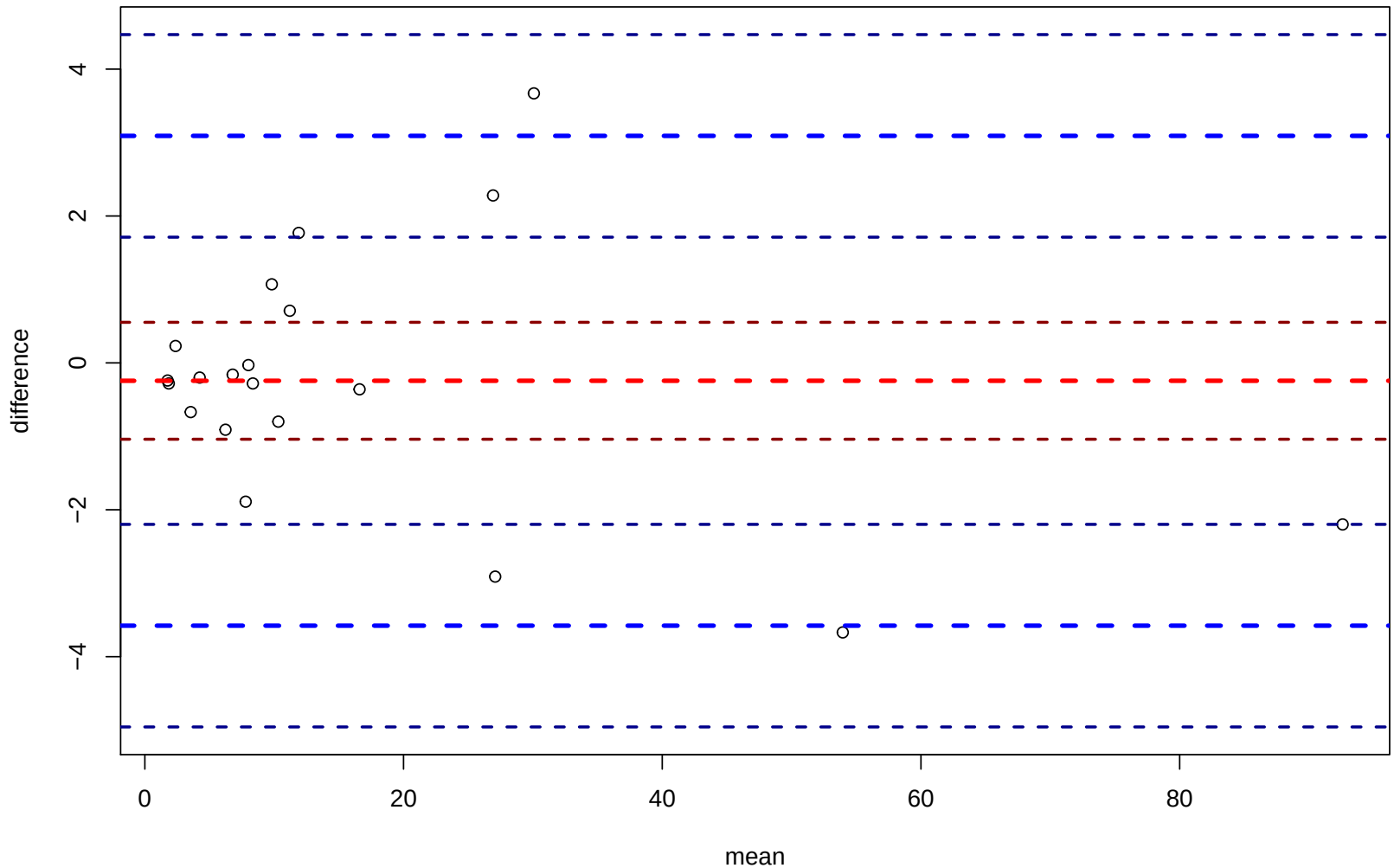
BA example (mcr package)

Back to AMH example



C.I. for the sample mean (SE) are missing; in red the C.I. for the SD.

BA example (package BlandAltmanLeh)



BA stats

- H_0: $\bar{d}_i = 0$
- Tested with confidence intervals

	diff
lower.limit.ci.lower	-4.47011
lower.limit.ci.upper	-1.71191
mean.diff.ci.lower	-0.55272
mean.diff.ci.upper	1.03972
upper.limit.ci.lower	2.19891
upper.limit.ci.upper	4.95711

Discrete, sempiquantitative example

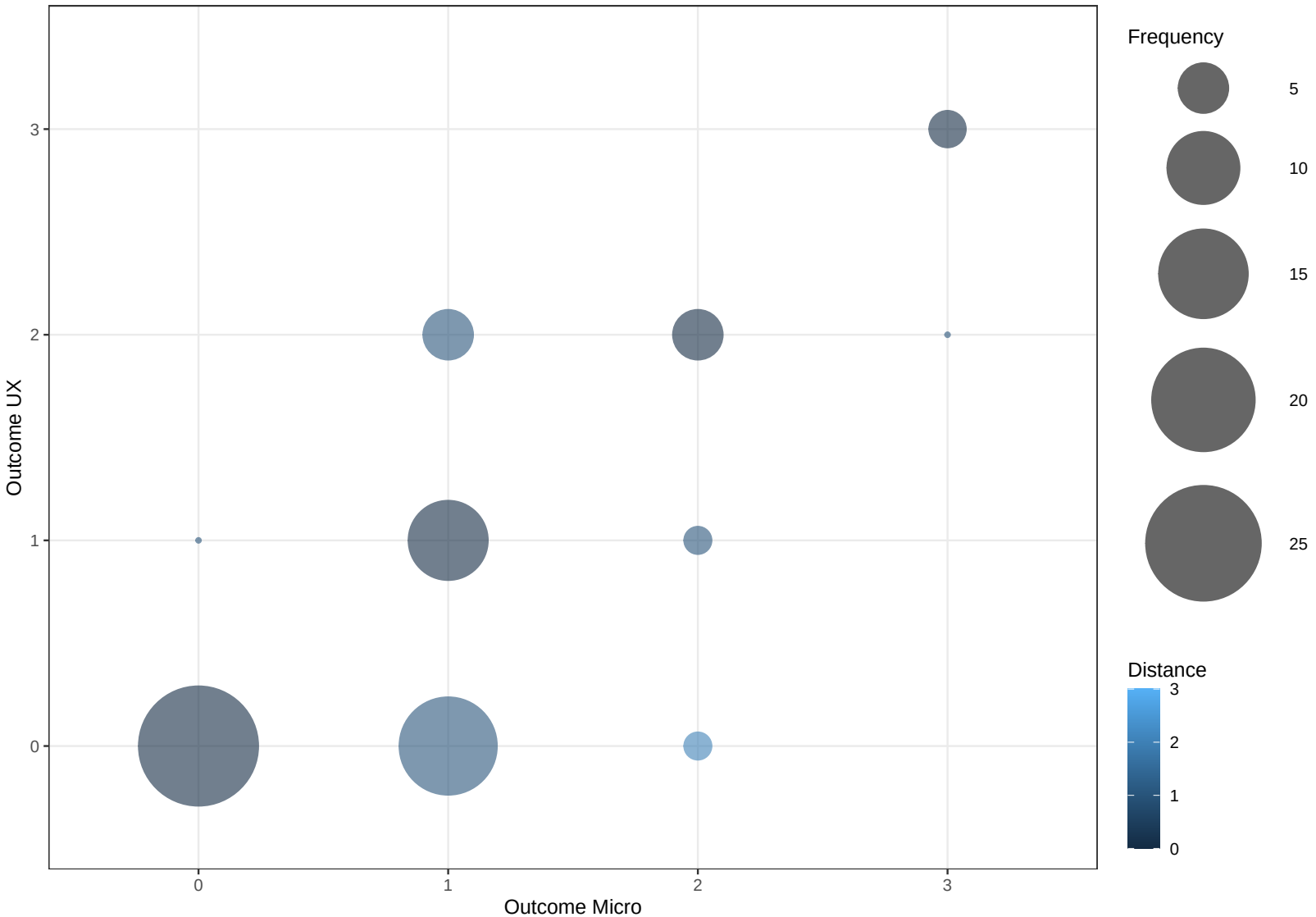
Discrete, ordered outcomes of the detection of bacteria in urine with two different stick sets from two different suppliers[2].

```
d.bact <- read.csv("Bact.csv", sep=";")
head(d.bact)
```

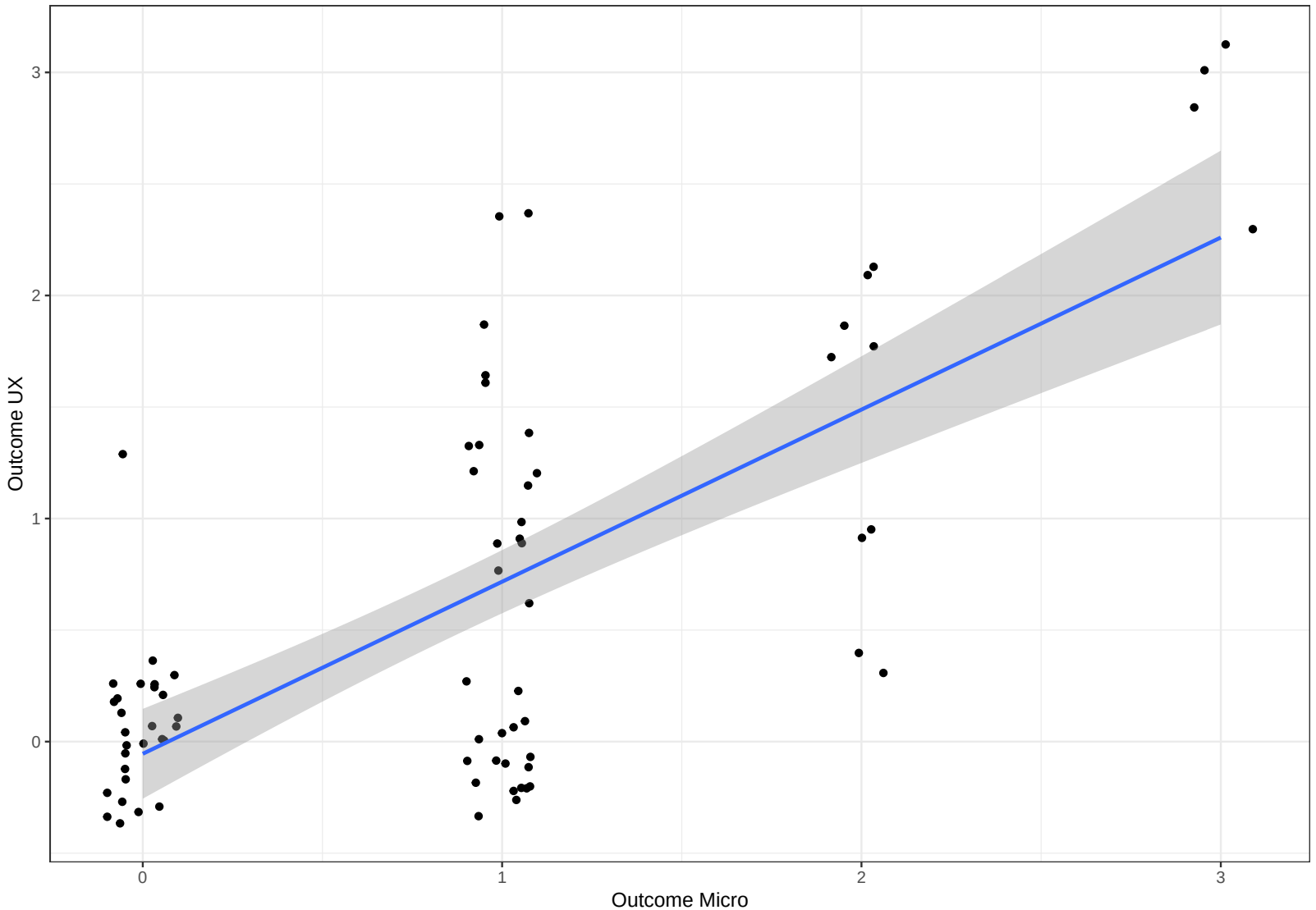
```
##   BactMicro  BactUx BaMicro BaUX
## 1   assenti assenti       0    0
## 2   assenti assenti       0    0
## 3    2croci assenti       2    0
## 4    1croce  1croce       1    1
## 5   assenti assenti       0    0
## 6  assenti  assenti       0    0
```

[2] S. Damiano, SSMT bachelor thesis, 2017.

Bubble plot

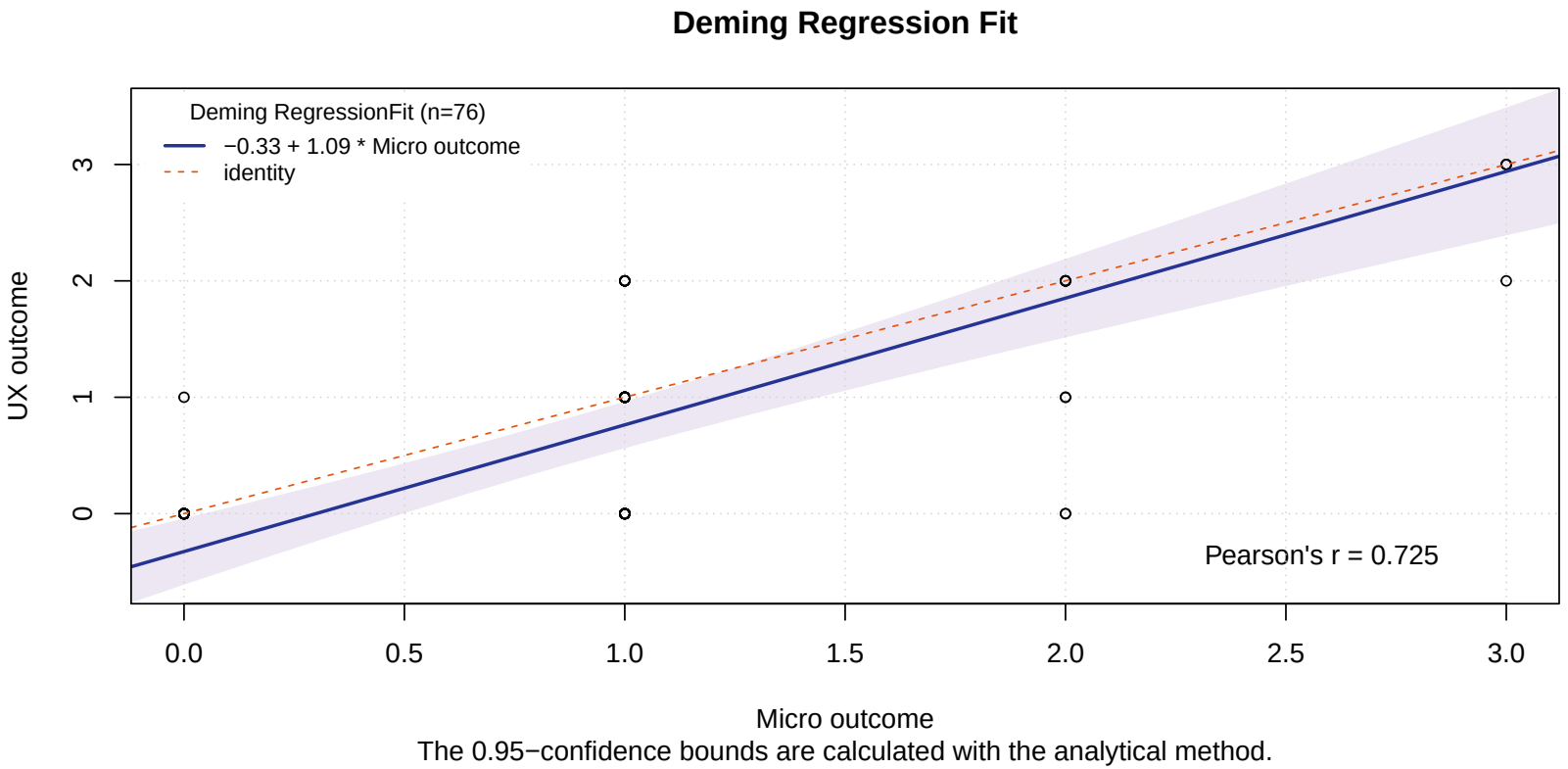

Outcome UX
Outcome Micro
0
1
2
3
Frequency
5
10
15
20
25
Distance
3
2
1
0

Discrete data with regression - ggplot2



Deming with mcr package on discrete data

- The C.I. are very large.
- PB is method not viable because of the high number of ties (repeated values). M-Deming neither.



The 0.95−confidence bounds are calculated with the analytical method.

Solution: Coen's Kappa

- Concordance index. Basically used in psychiatry.
- Evaluate how far the concordance goes beyond the pure random concordance by chance.
- $\kappa = 1$ for perfect concordance. $\kappa = 0$ for a totally random concordance. $\kappa < 0$ for discordance.
- Coen's proposal has a scale of value.
- Analytic form for the confidence intervals available.

Kappa details

	A	B	Observer1
A	a	b	$a+b$
B	c	d	$c+d$
Observer2	$a+c$	$b+d$	$a+b+c+d$

$$P_{obs} = \frac{a+d}{a+b+c+d}$$

$$P_{ch} = \frac{(a+c)\cdot(a+b)}{a+b+c+d} + \frac{(b+d)\cdot(c+d)}{a+b+c+d}$$

$$\kappa = \frac{P_{obs}-P_{ch}}{1-P_{ch}}$$

For H_0: $\kappa = 0$ test $Z = \frac{\kappa}{\sqrt{Var(\kappa)}}$

$$Var(\kappa) = \frac{P_{obs}-{P_{obs}}^2}{N\cdot(1-P_{ch})}$$

Kappa scale of value

According to Cohen

kappa	Agreement
< 0	Poor
$0 - 0.2$	Slight
$0.21 - 0.4$	Fair
$0.41 - 0.6$	Moderate
$0.61 - 0.8$	Substantial
$0.81 - 1$	Almost perfect

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3165924/

Kappa CI - bootstrap is mandatory

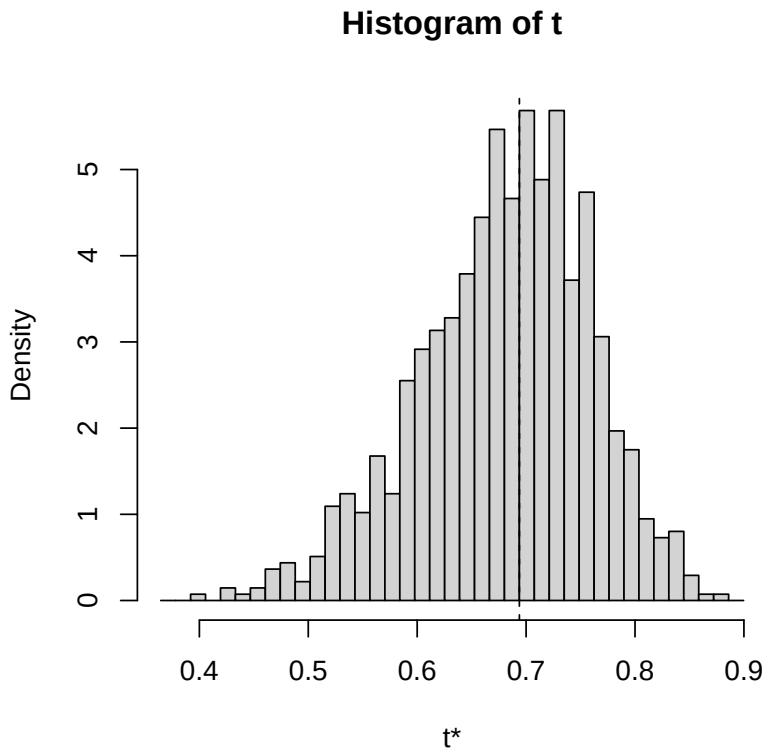

```
lkappa.boot <- function(data,x) {
    lkappa(data[x,], type="weighted")}
k.print<-lkappa(d.bact[,3:4],
           type="weighted",weights = "squared")
names(k.print)<-"Kappa"
res<-boot(d.bact[,3:4],lkappa.boot,1000)
res.print<-as.data.frame(c(k.print,quantile(res$t,c(0.025,0.05,0.95,0.975))))
colnames(res.print)<-"Estimate"
kable(res.print)
```

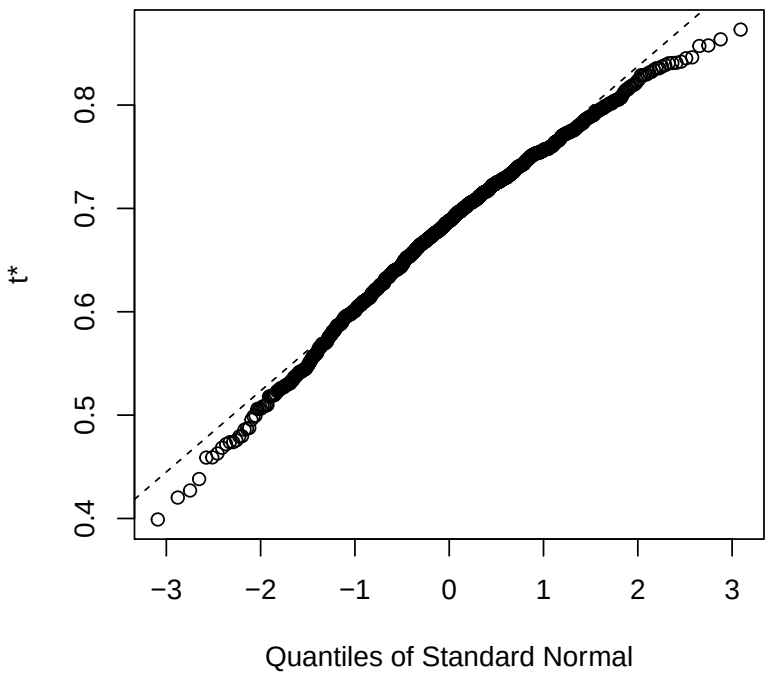
	Estimate
Kappa	0.69383
2.5%	0.50924
5%	0.53698
95%	0.79713
97.5%	0.81914

According to ISO 15189 the validation in this case fails since a lower c.i (5%) > 0.75 is requested.

Kappa bootstrap plot

- The distribution of the t^* is usually not symmetric.
- Using the quantiles for testing can be a good idea.





Literature

- ISO 15189 rightlines
- J. Pum, Advances in Clinilibrarycal Chemistry, Volume 90, ISSN 0065-2423, pag 215-281
- G. Pioda, arXiv:2105.04628, 2021
- H. Passing, W. Bablok, J. Clin. Chem. Clin. Biochem. Vol. 21, 1983, pp. 709-720
- J.M. Bland, D.G. Altman, Stat, 1983, 32, 307-17
- J.M. Bland, D.G. Altman, Stat. Met. Med. Res., 1998, 8, 135-60