

A bug on the χ^2 computation of the 21cm Cl.

Pablo Motta
Universidade de São Paulo

Definitions

$$\chi^2 = \Delta^T C^{-1} \Delta \text{ where } \Delta = S_L^{Theo} - S_L^{data}$$

Lets define $C^{-1} = U \Sigma U^T$ (SDV decomposition)

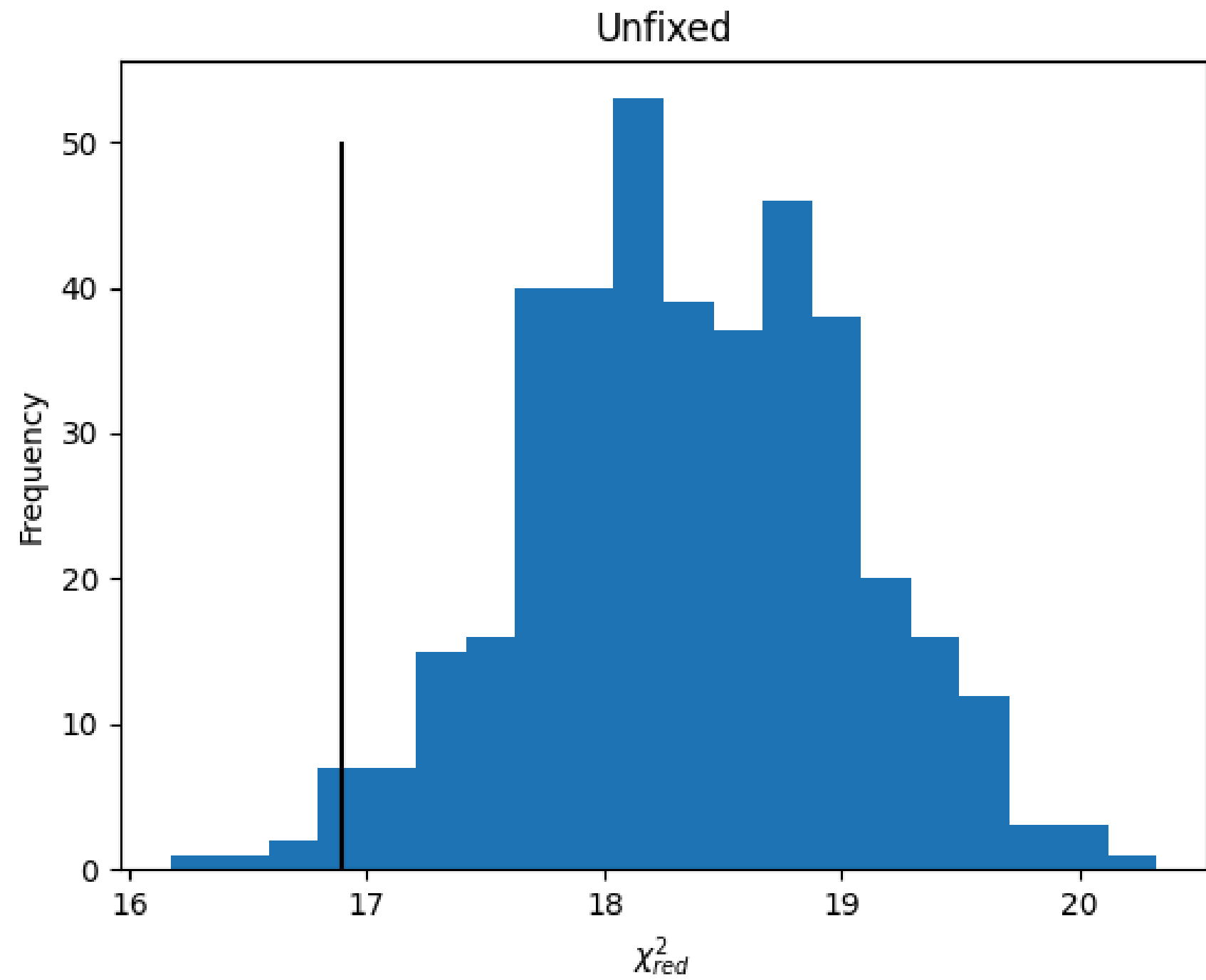
$$\text{and } \chi = (\Delta^T U)^T \Sigma^{1/2}$$

$$\text{we get } \chi^2 = \chi^T \chi = \Delta^T U \Sigma U^T \Delta = \Delta^T C^{-1} \Delta$$

$$\chi_{red}^2 = \chi^2 / (N_{datapoints} - N_{parameters})$$

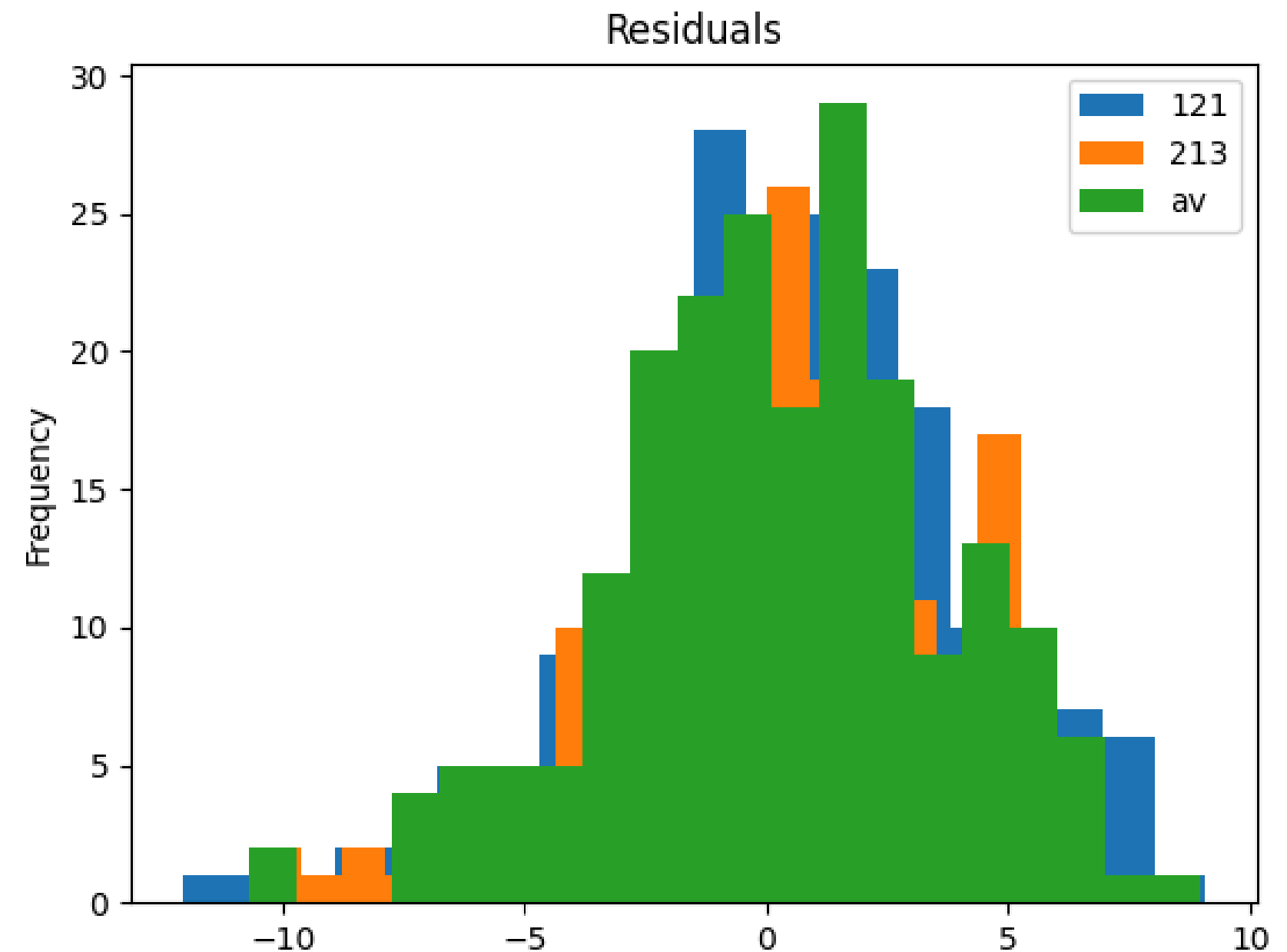
Why chi2 high?

- We computed the reduced chi2 for around 400 realizations.
- The chi2 is computed with respect to the fiducial CI.
- We get values around 18 while we would expect values of 1.
- It is normalized by 141, i.e. number of datapoints (206) minus number of parameters (65).
- Chi2 for the average of simulations is also high (black line).
- We want to understand what is happening here.



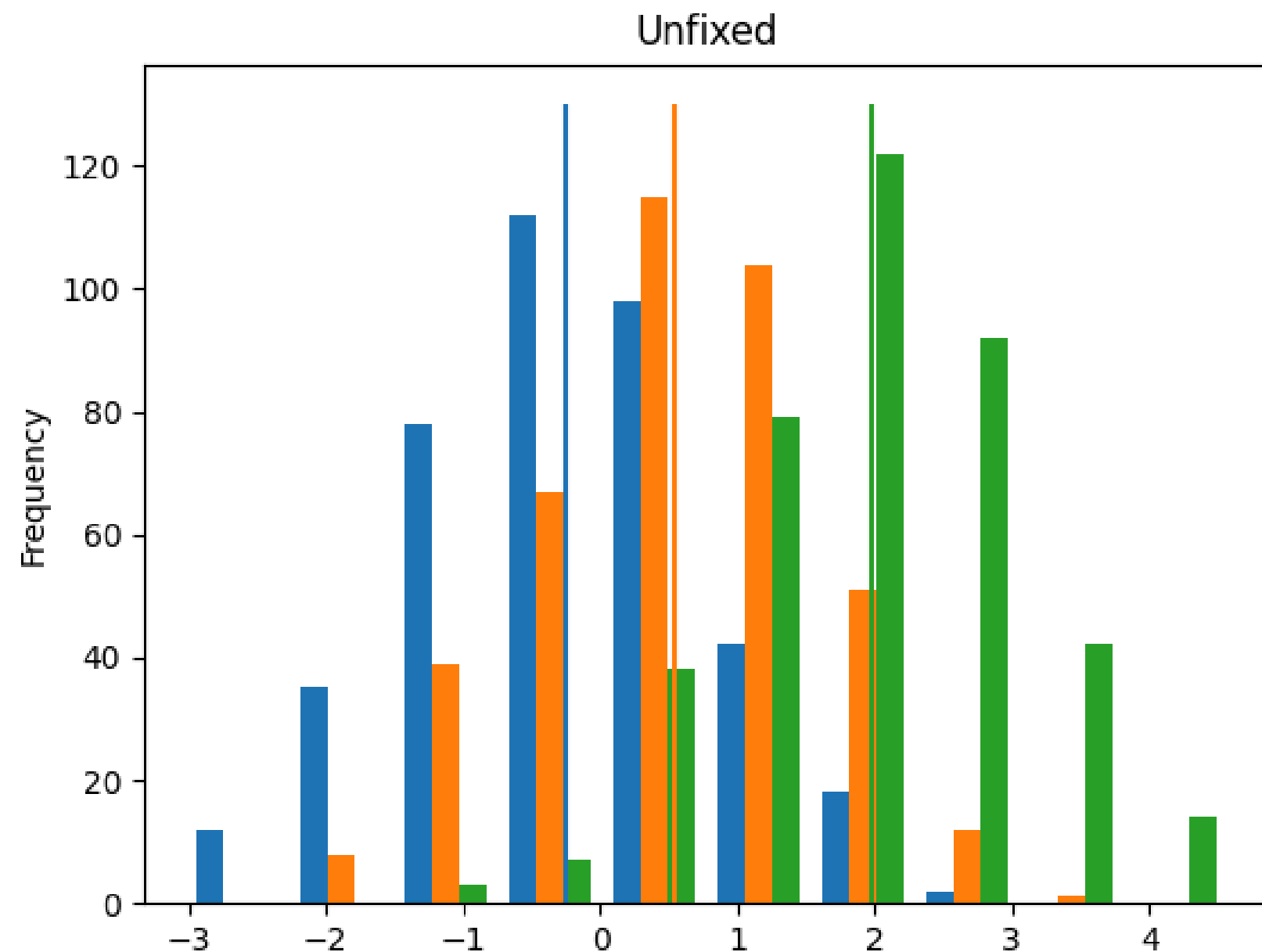
Residuals for all datapoints given one realization

- Here we plot the residuals for all datapoints. Blue and orange are two different realizations and green is the average.
- The histograms represent the distribution over the datapoints.
- These residuals are centered in zero, but too spread, reaching values of e.g. 5 sigma.



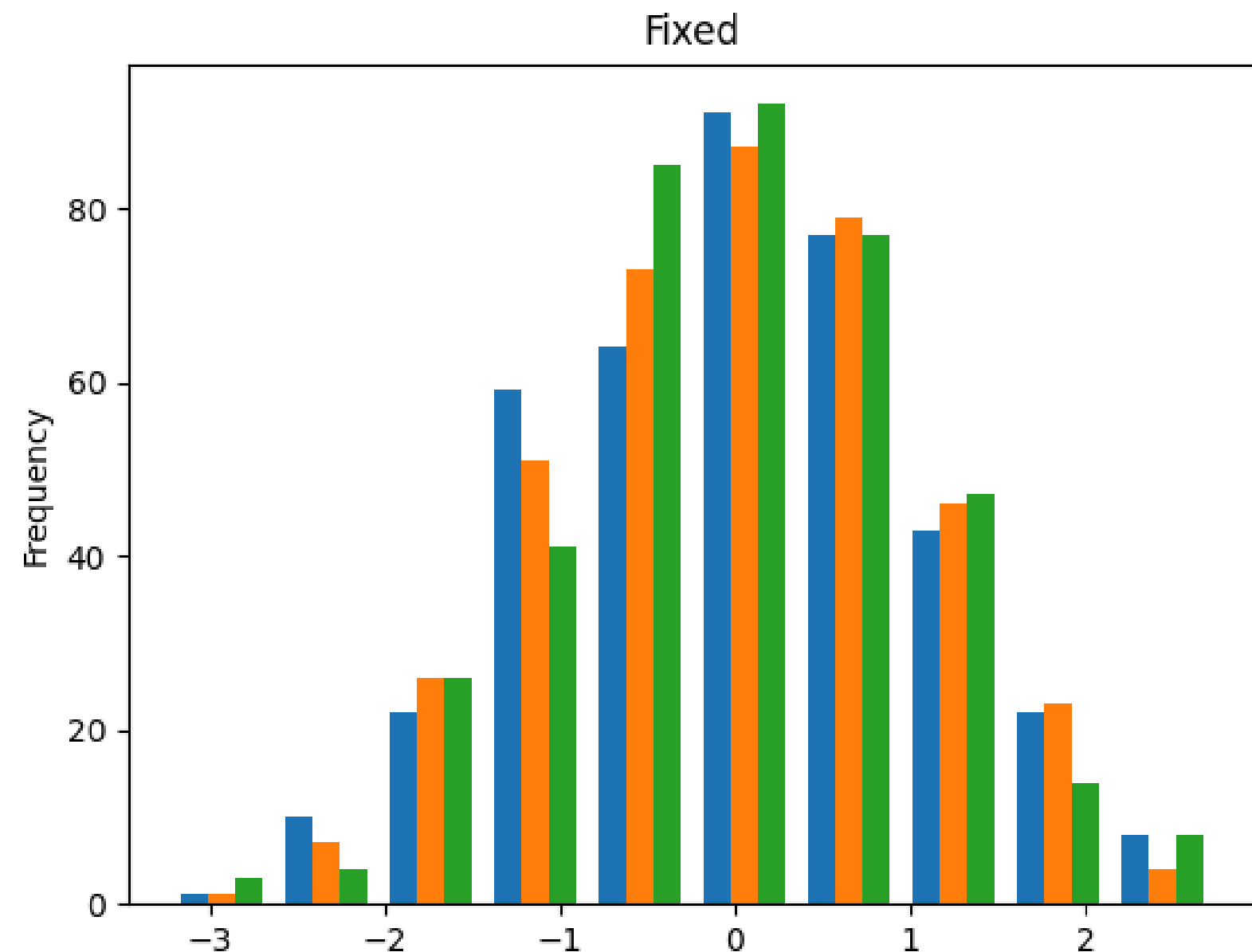
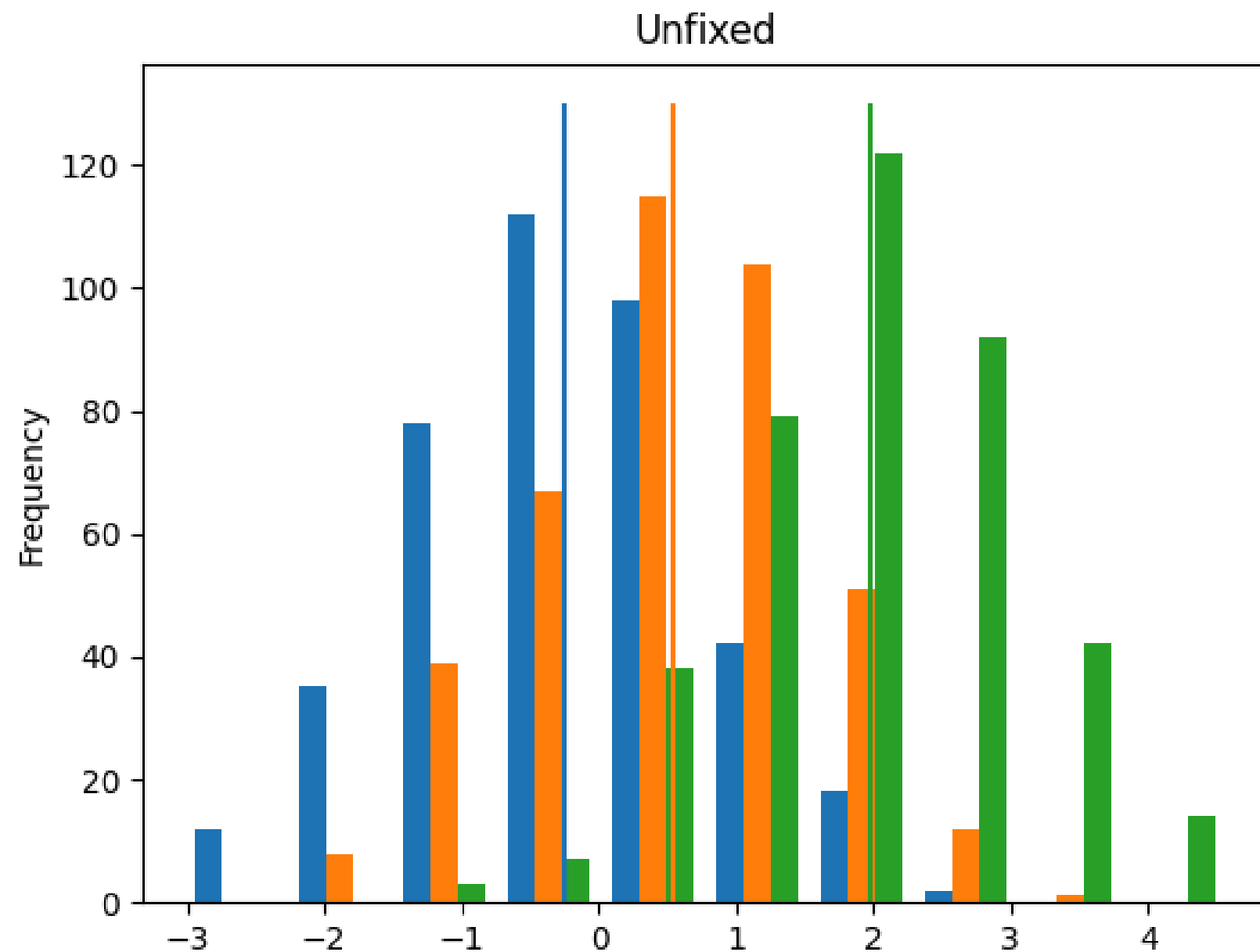
Residuals for all realizations given one datapoint

- Each color is a data point and the histogram represents the distribution over the realizations.
- The thin vertical lines are residuals for the average CI.
- We noticed that the histograms are shifted where the shift depends on datapoint.
- This could explain the problems of histograms of the past slides.



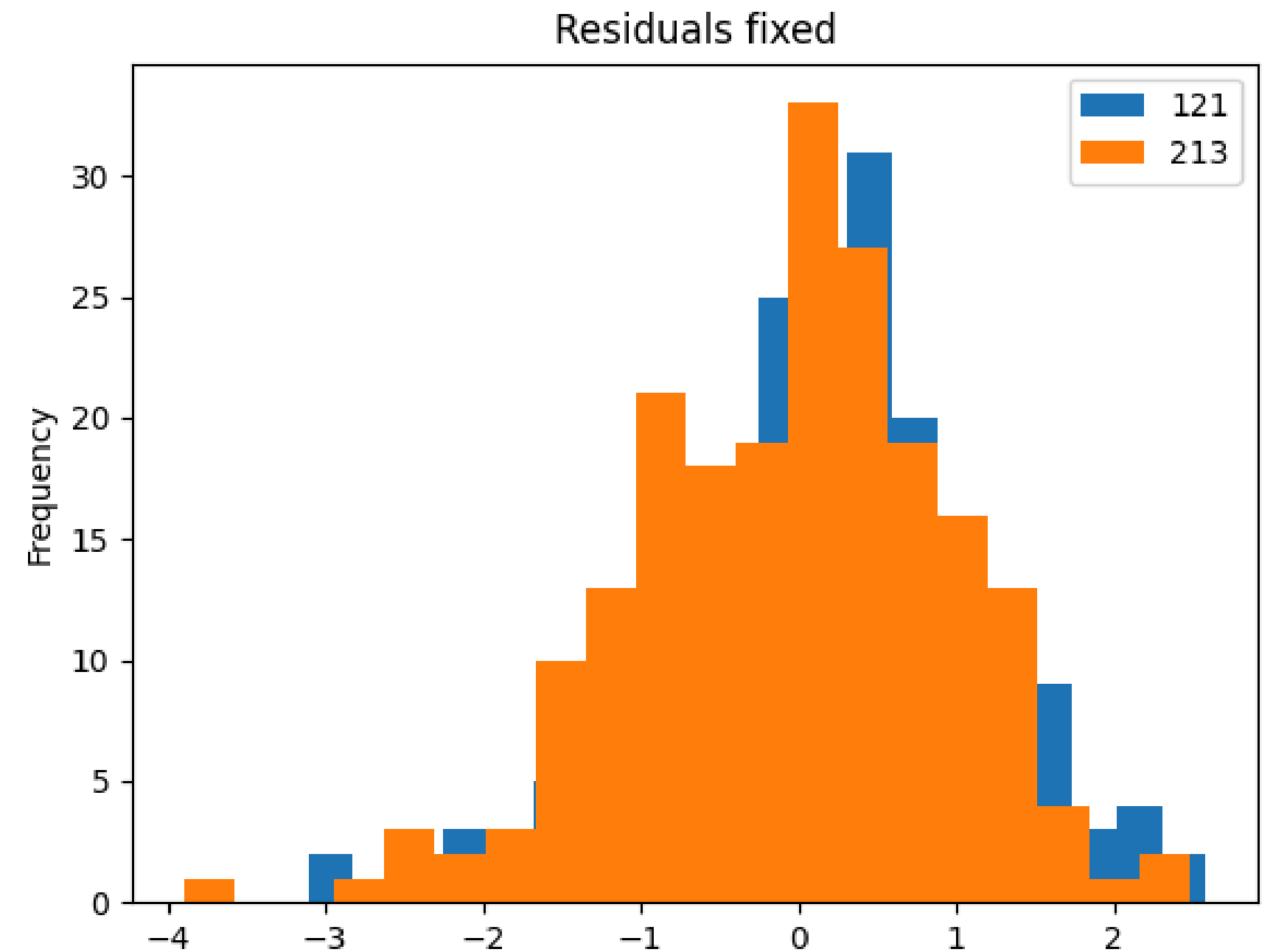
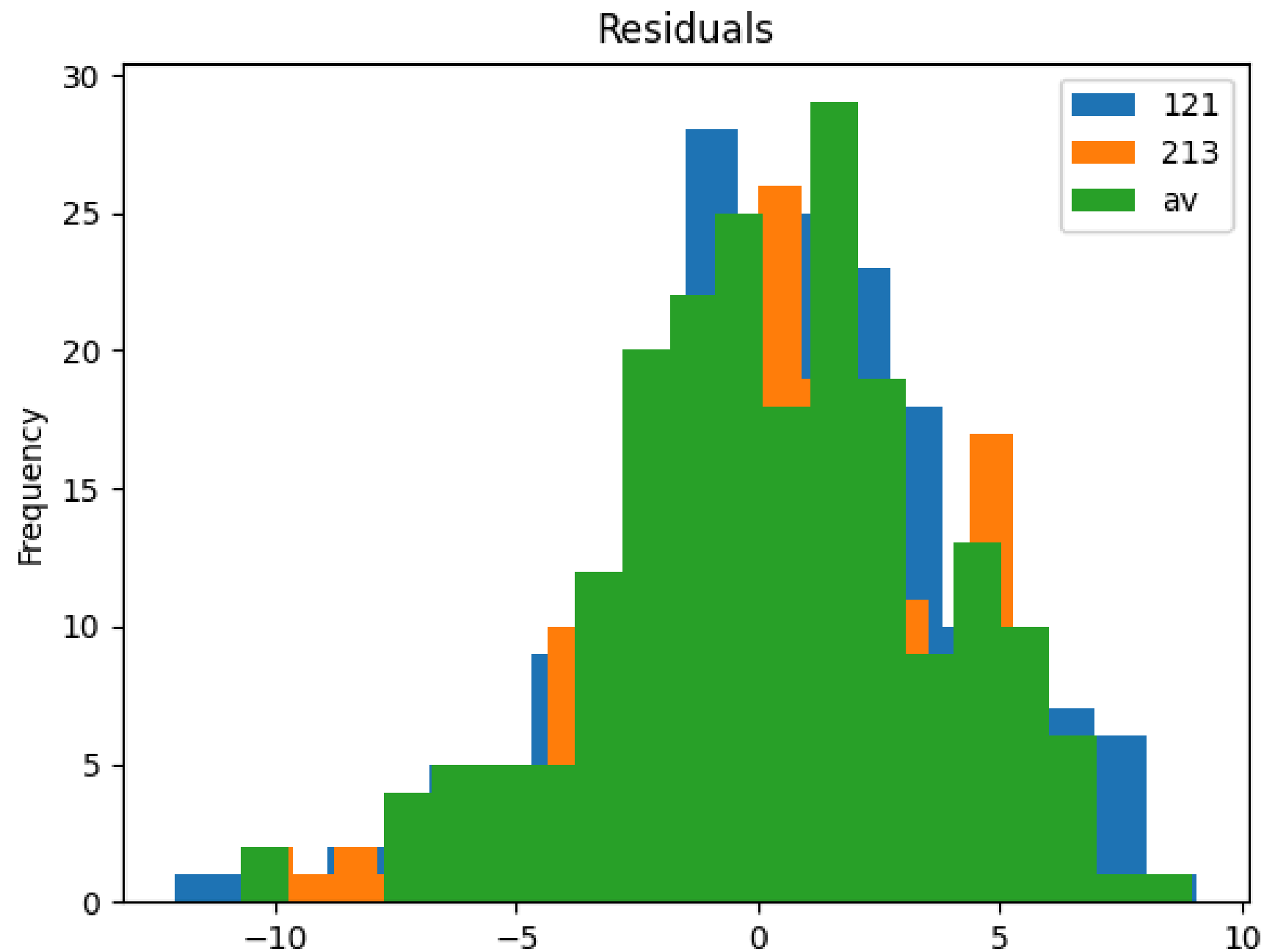
Fixing by hand

- We fix the histogram by removing the residuals computed for the average CI.
- We don't claim to be solving the problem but rather trying to understand it.



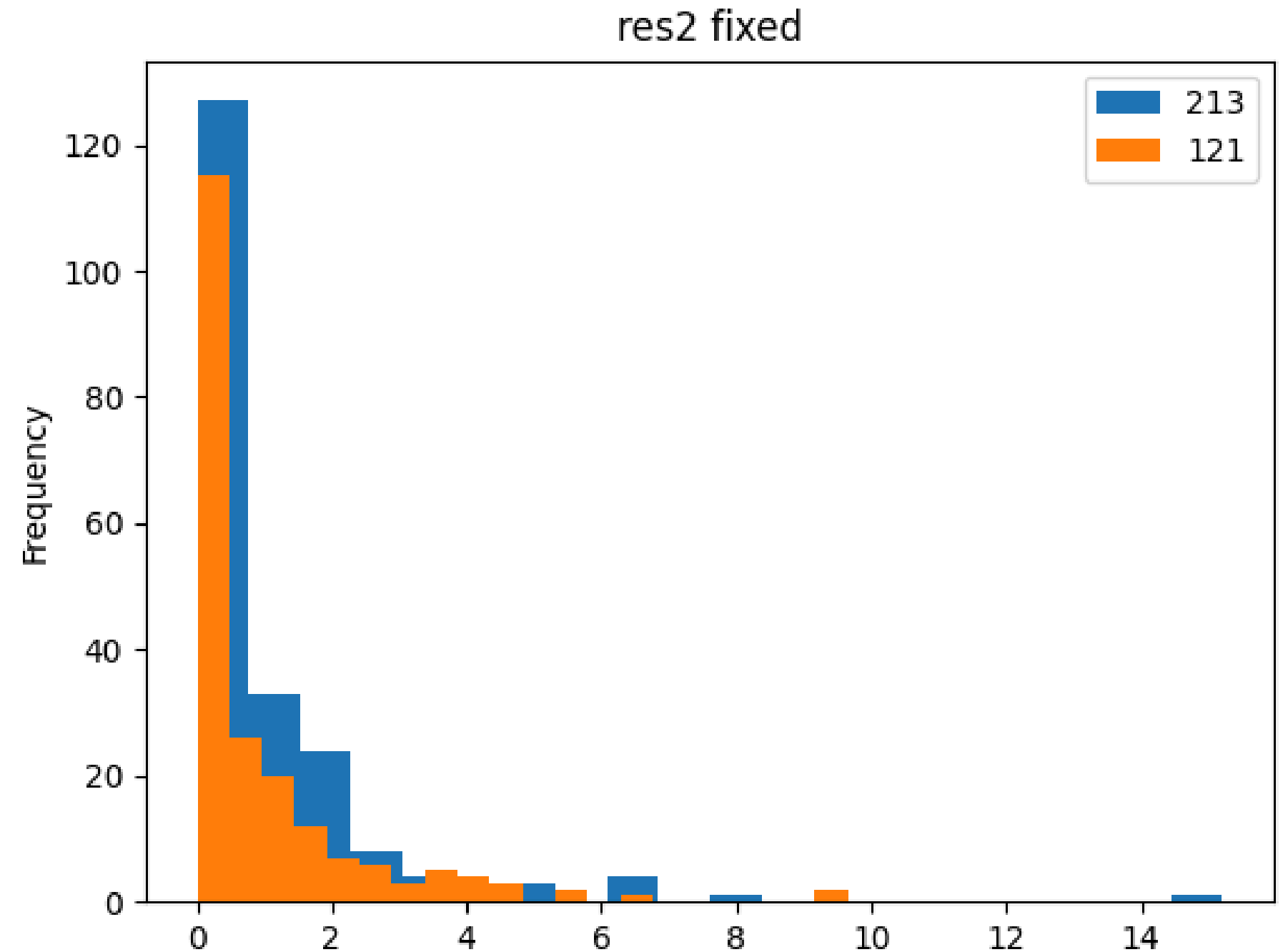
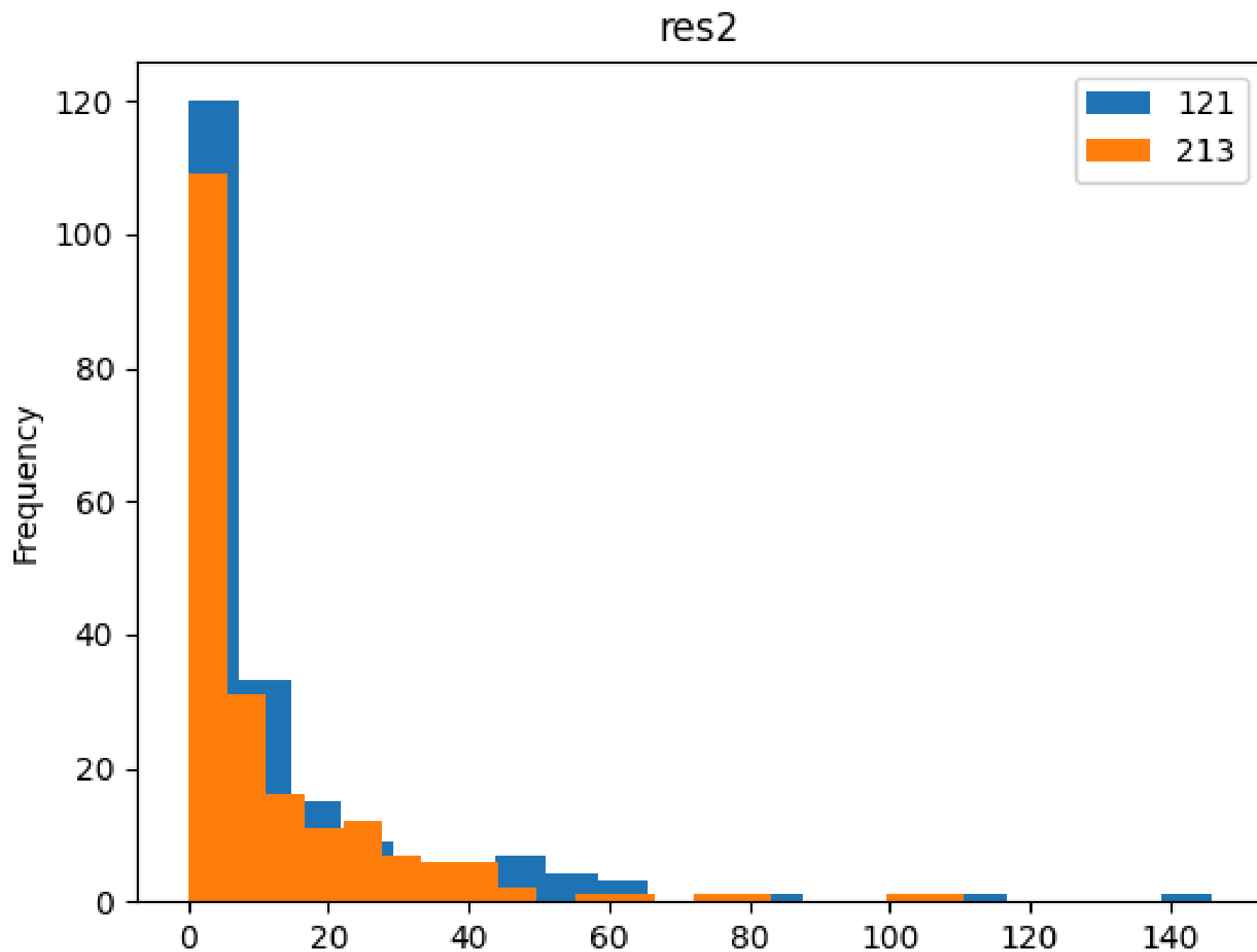
Fixing by hand

- We apply the same “fixing-by-hand” method on the other histogram.
- We see that now most of the values are between -1 and 1.



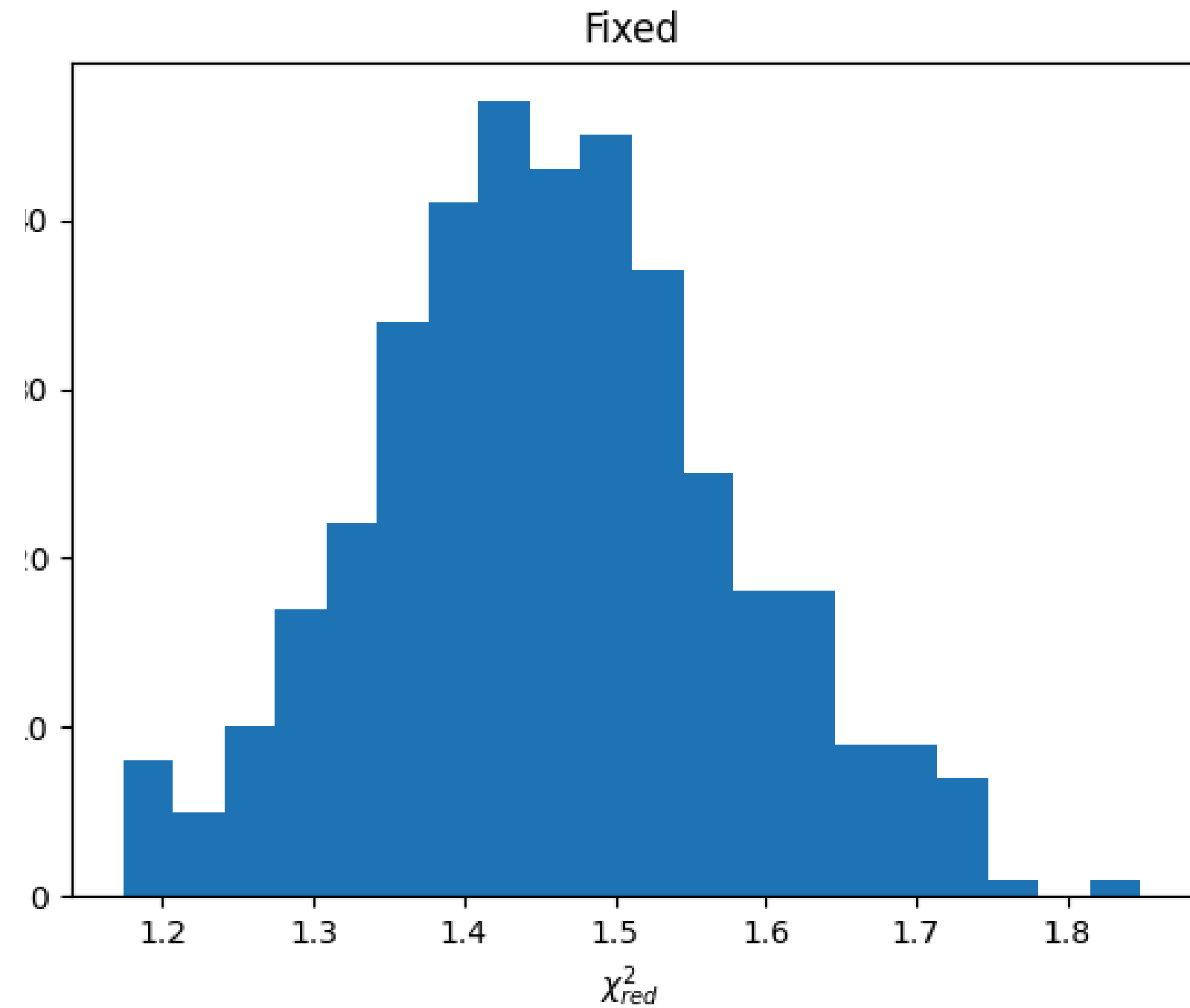
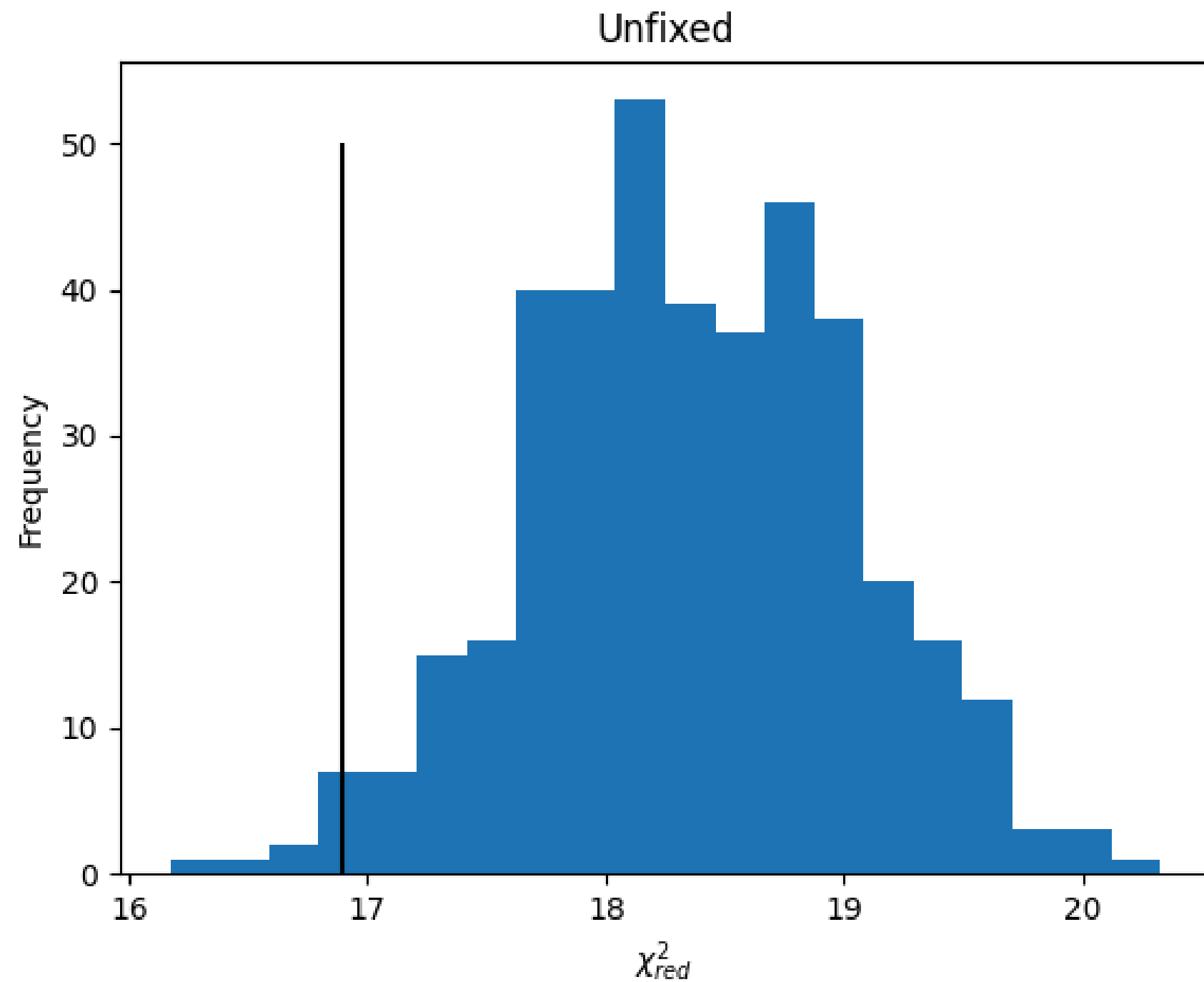
Fixing by hand

- Here we show the histogram for the residuals squared.
- The summation of this values gives the chi2.



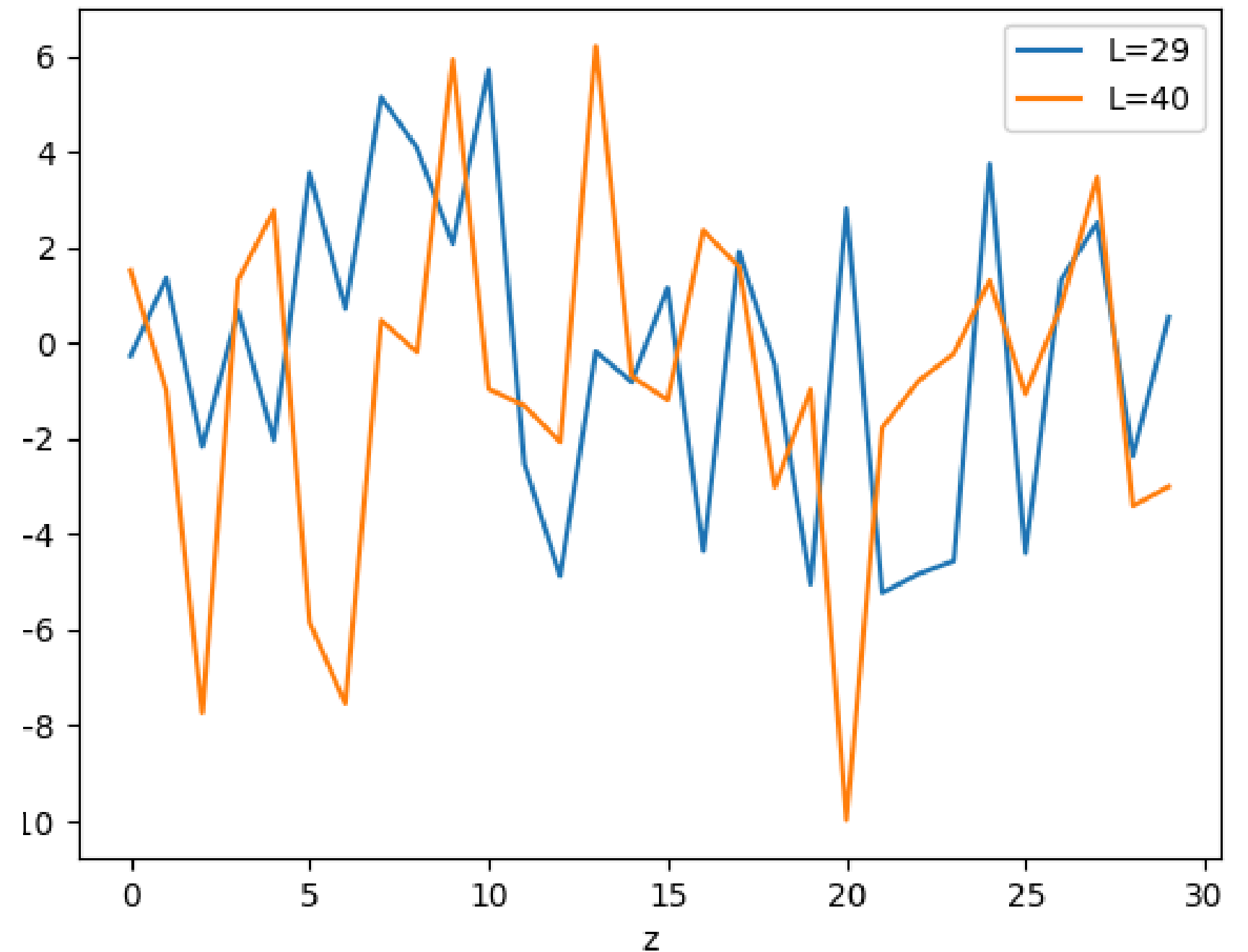
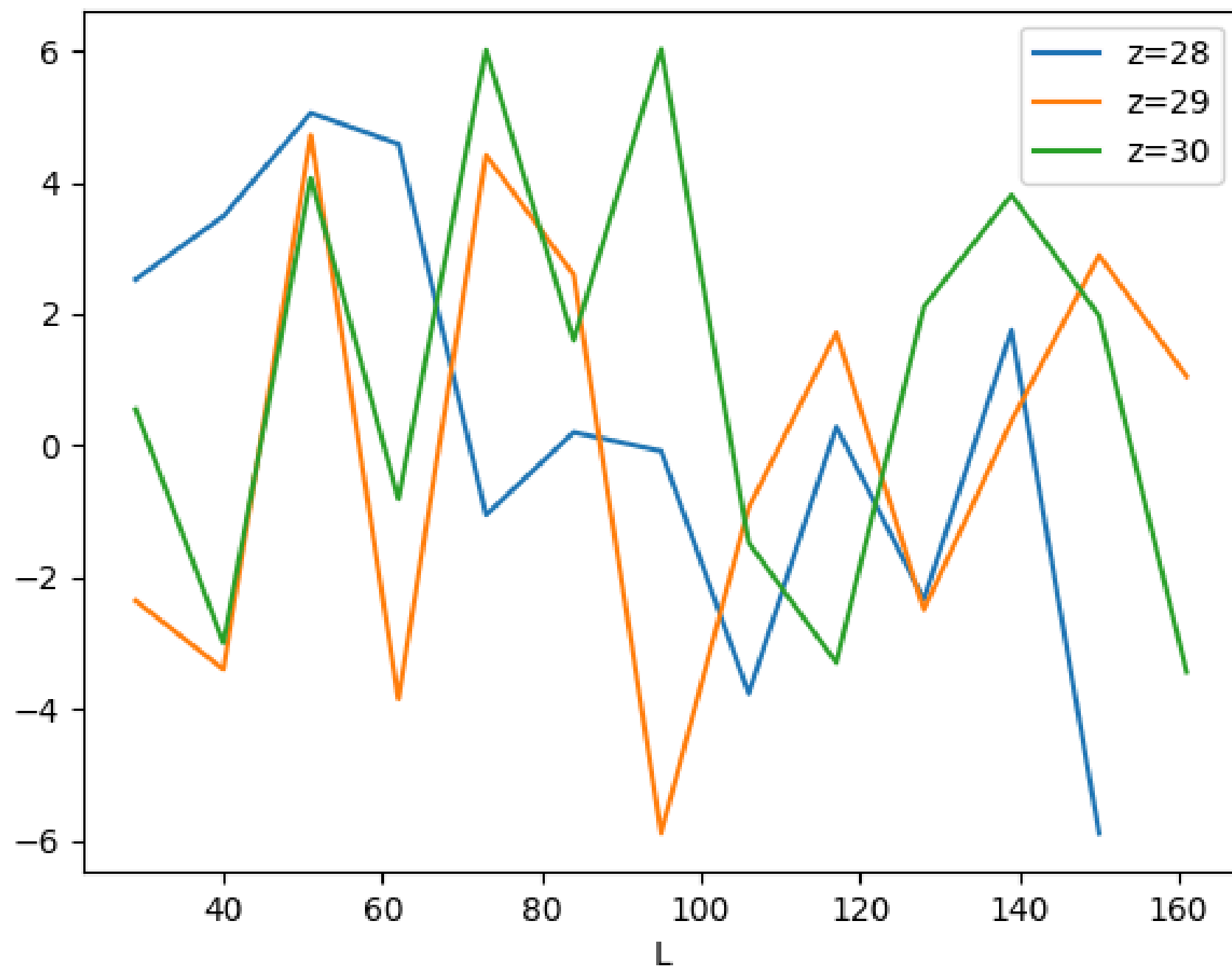
Fixing by hand - chi2

- Finally we show what happen with the chi2 after the “fixing-by-hand”.
- The residuals decrease from 18 to 1.4.



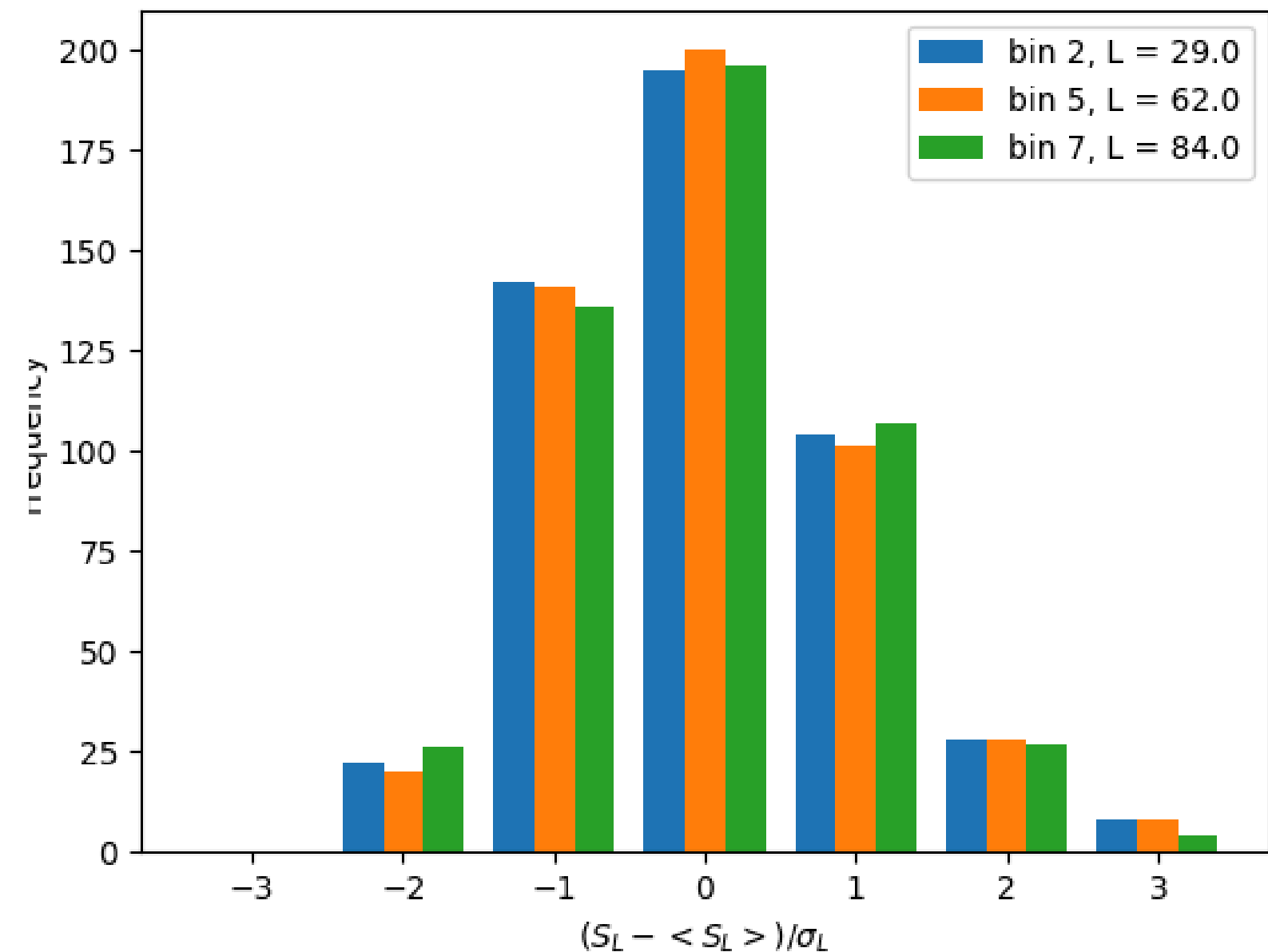
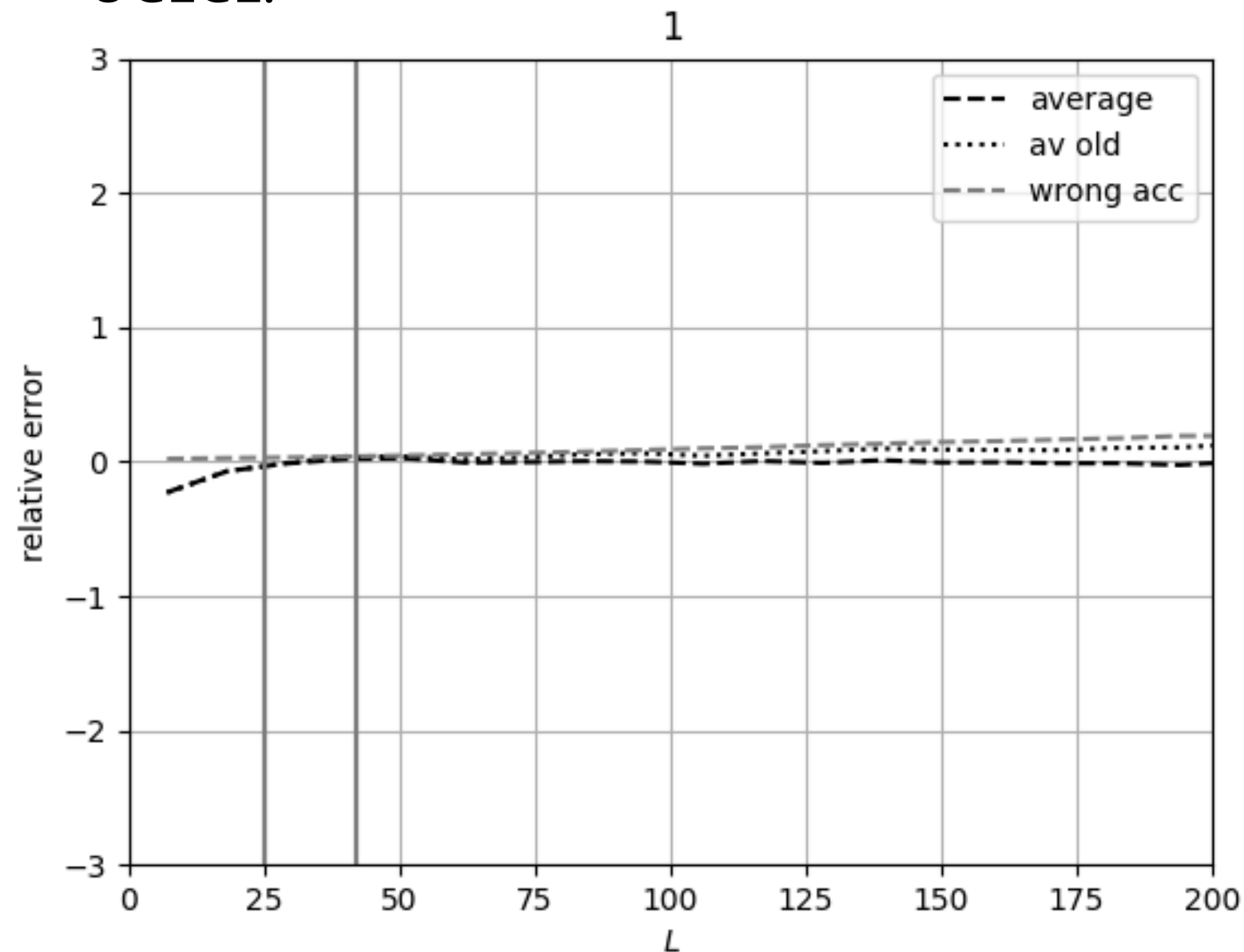
Residuals for the average CI over L and z

- We proofed there is a bug causing systematic shifts on the residuals on each datapoint, which causes a high chi2 with respect to the fiducial CI.
- Let's look for residuals for the average. We don't see any clear pattern.



Sigma normalization

- We have shown before (see figs.) that both average CI and histogram of the distribution over the realizations are well-behaved when it's applied the normalization with respect to sigma (sqrt of the diagonal of covariance).
- The bug appears when the residuals are computed with the full covariance matrix within UCLCL.



Conclusions

- The reduced χ^2 of the data CIs with respect to the fiducial are around 1.4. The χ^2 for the average CI is also too high (approx. 17).
- We proofed there is a bug causing systematic shifts on the residuals on each datapoint, which causes the high χ^2 .
- Using the residuals of the average CI as a reference we fixed by hand the values of the residuals per datapoint of a given realization.
- With the “fixing-by-hand”, the histograms get well-behaved and the χ^2 becomes 1.4.
- We don’t know yet what is causing this bug.
- One hypothesis is that there is a problem on the inversion of the covariance matrix. The inversion is done inside UCLCL and the code should be able to check it. I will check if the inverse is good.
- If you have another ideas please let me know.