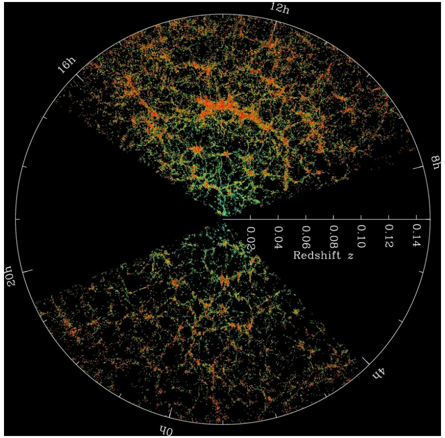
A Median Statistics Estimate of the Distance to M87

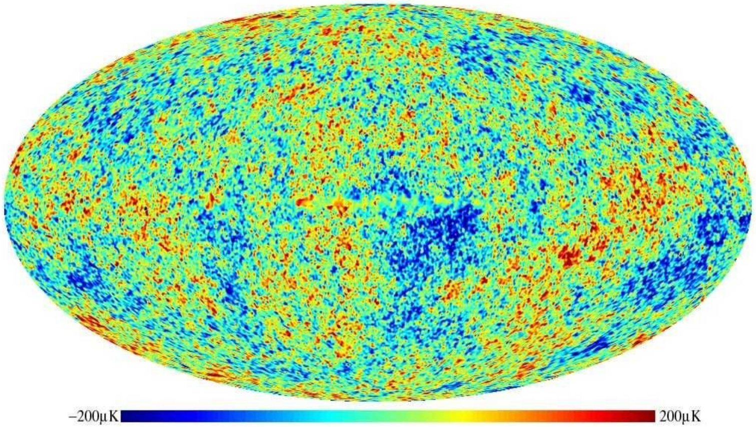
Sofia Splawska with Dr. Bharat Ratra
August 5, 2022
Kansas State University Physics REU 2022

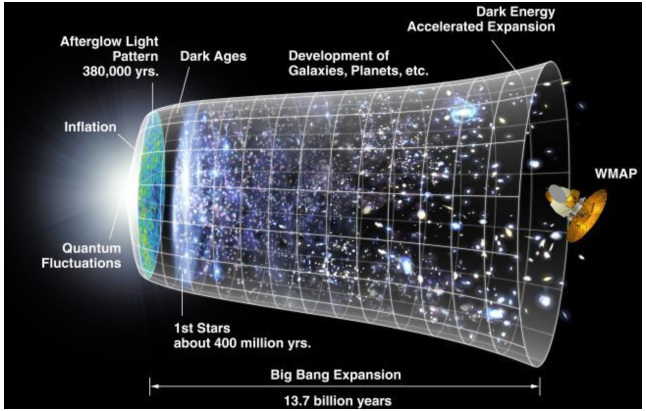
Kansas State University
College of Arts & Sciences
Department of Physics

Cosmology

Study of the universe’s structure, content, and evolution on the largest scales

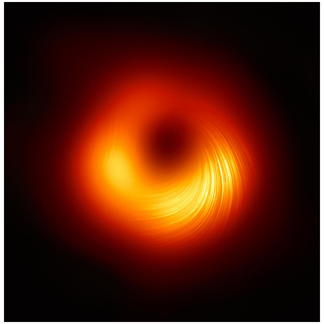






Distances are very important in Cosmology

- Why do we want to know distances?
  - To calculate physical characteristics of objects…
  - … so we can do Cosmology!
- Why do we want the distance to M87?
  - Extend distance framework
  - Study further clusters
  - Study M87



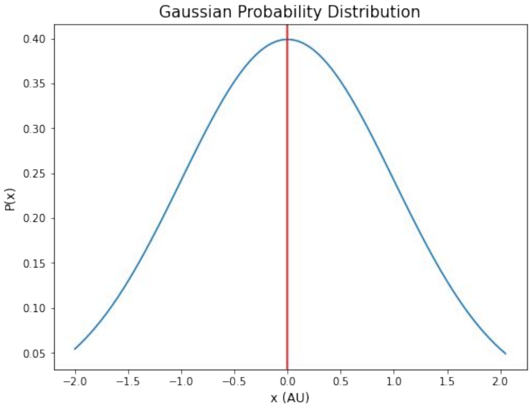
M87’s supermassive black hole, Pōwehi

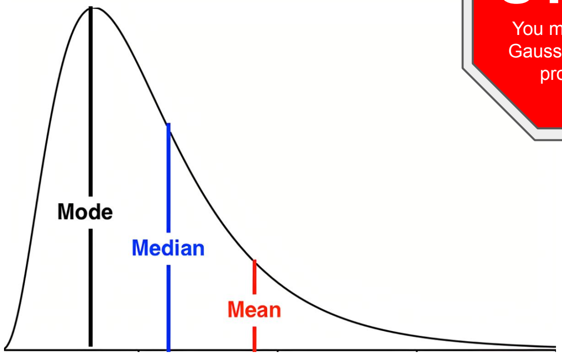
Messier 87 (M87)

The Virgo Cluster, home to M87

Statistics in Cosmology

- We characterize our data with a central estimate:
  - Ex: Mean, median, mode, weighted mean
- Intrinsic Gaussianity





The distribution of data is not always Gaussian!

- Solution: ~~Mean~~ Median Statistics
- Median statistics provides an accurate central estimate without assuming Gaussianity
- The price (larger error bars) is worth it

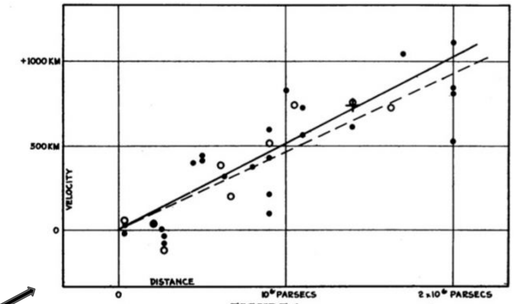
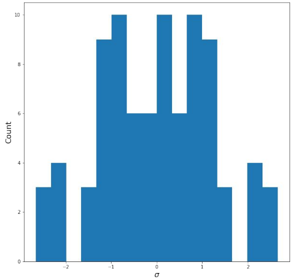


FIGURE 1
Velocity-Distance Relation among Extra-Galactic Nebulae.*

The LMC

*Hubble, Proceedings of the National Academy of Sciences, 1929, 15, 168

Procedure



Median, weighted mean

p-values?

Choose data → Calculate Central Estimates → Create Error Distribution → KS test → Report values

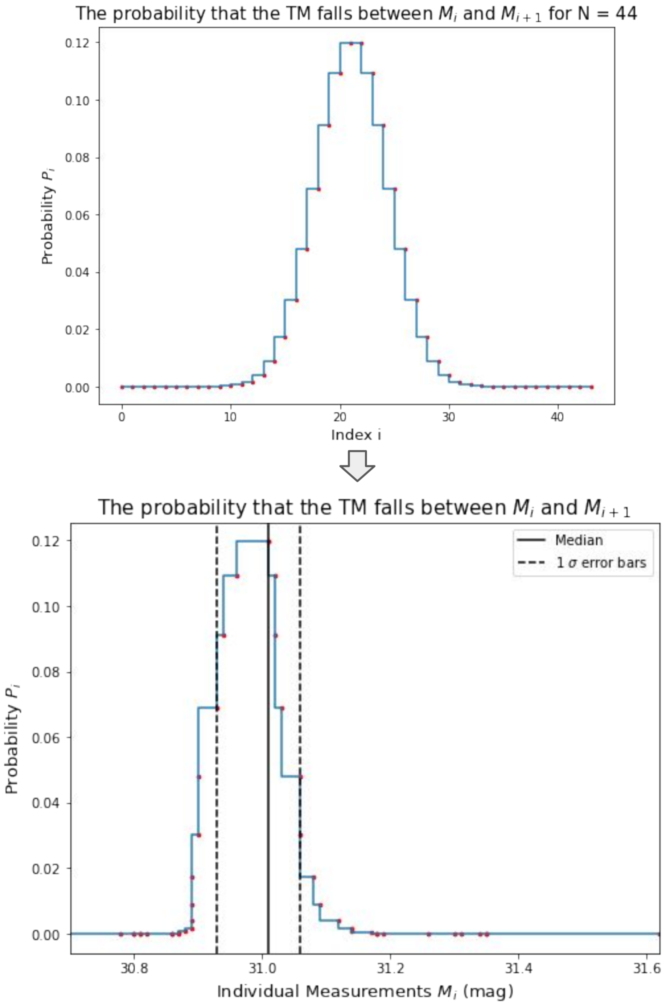
Test Gaussianity to verify that median statistics is the best choice

Calculating the Median

True median (TM): the median of the dataset as the number of measurements N goes to infinity

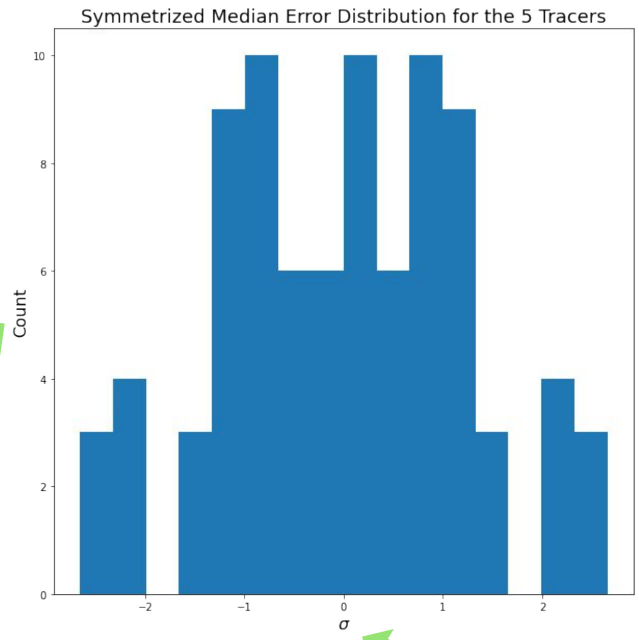
The probability that the TM falls between measurements M_i and M_{i+1} is:

$$P = \frac{2^{-N} N!}{i!(N-i!)}$$



Gott et. al. (2001)

Error distribution

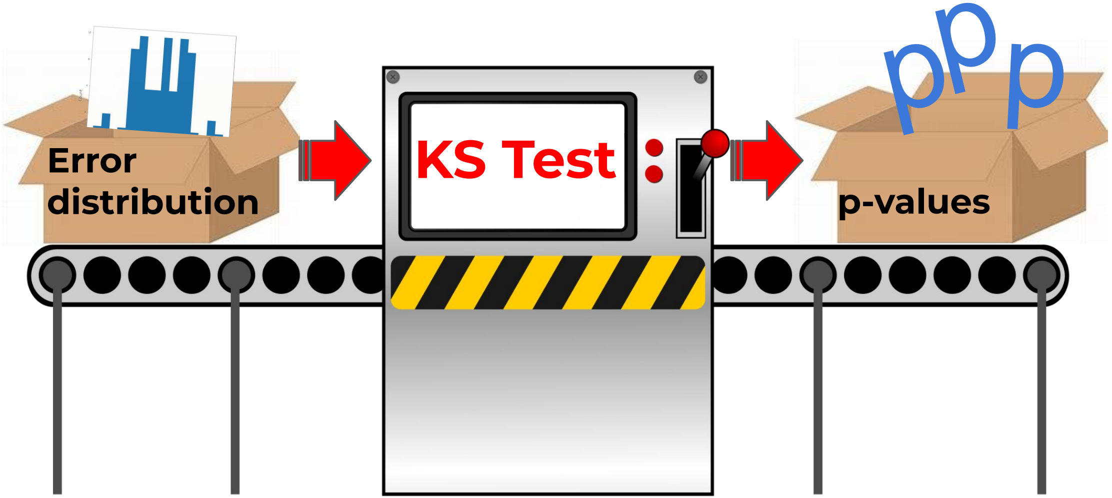


Number of measurements at that “distance” from the central estimate

Number of standard deviations away from central estimate

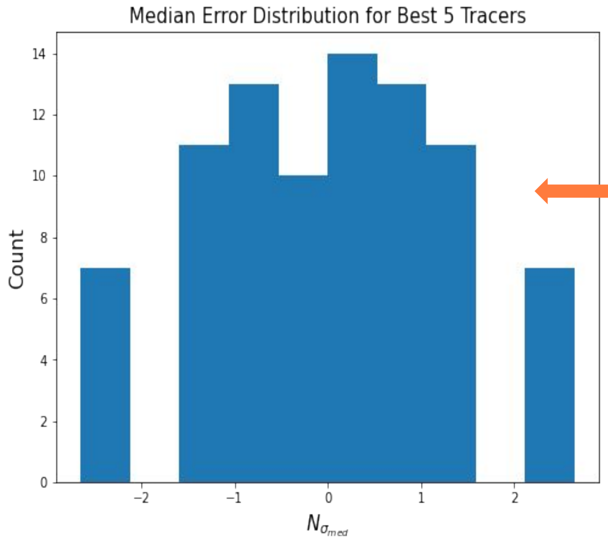
The Kolmogorov-Smirnov Test (KS Test)

The KS test measures the similarity between an empirical error distribution and a given continuous probability distribution (in this case, the Gaussian) by calculating a p-value

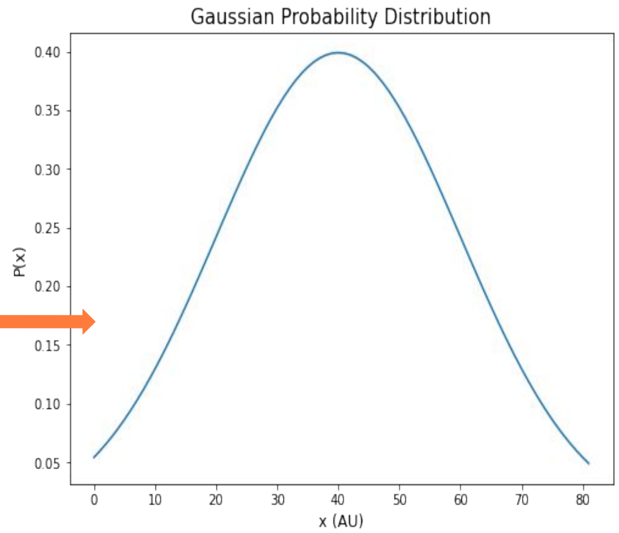


What does the KS test tell us?

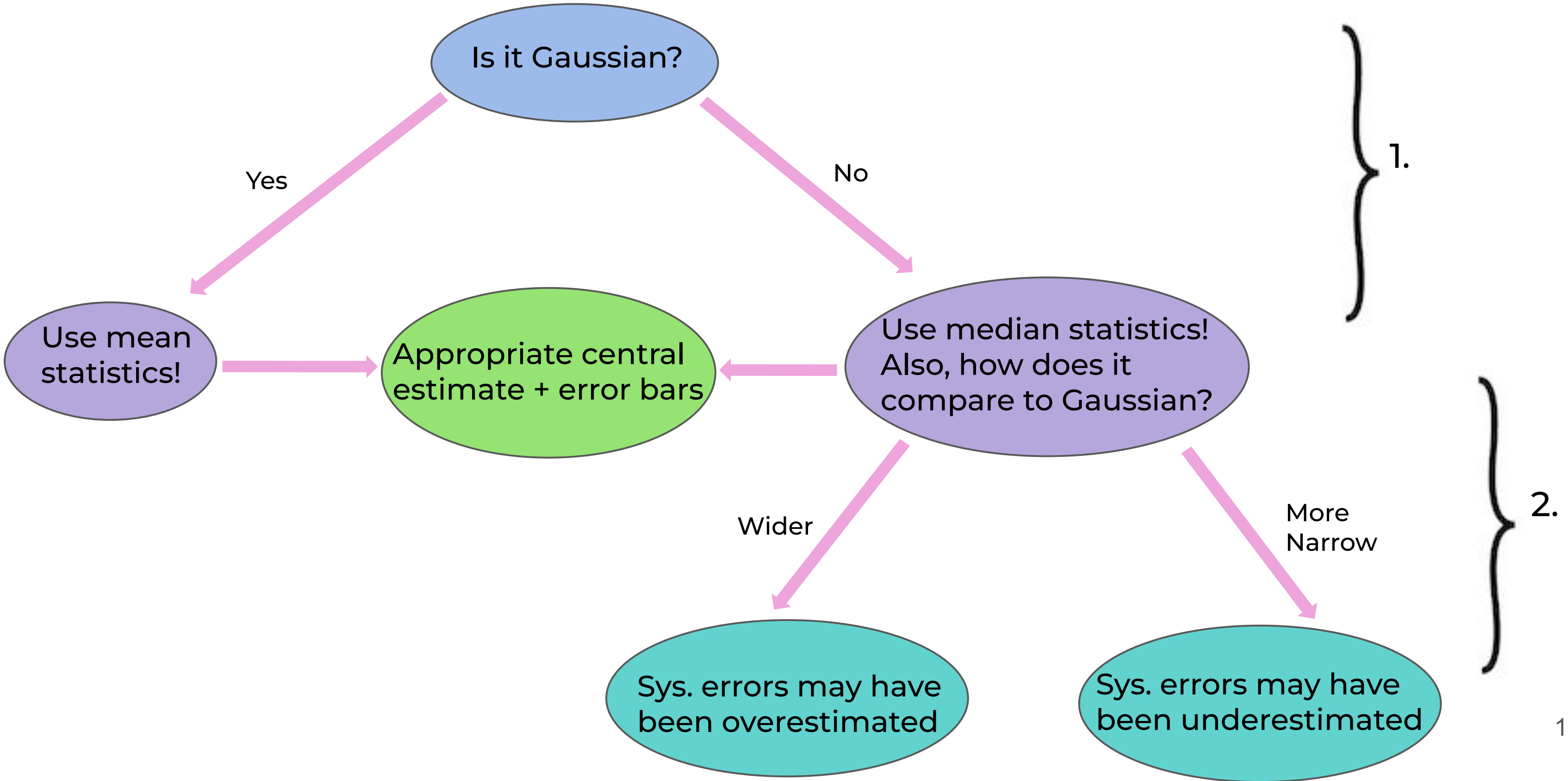
p-value: the probability that we can reject the hypothesis that the data do not come from the tested distribution



What is the probability that this data was **not** drawn from a distribution **other** than this one?

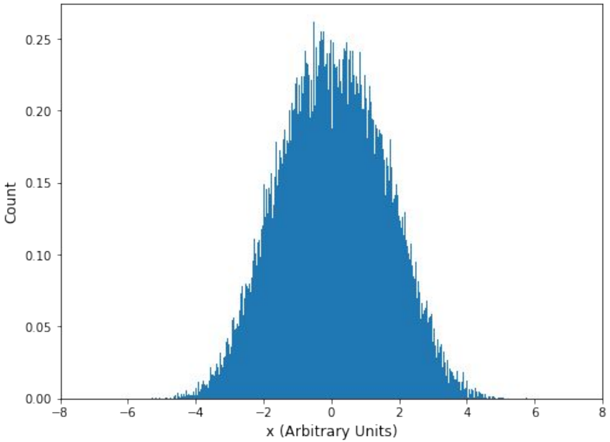


Why is this helpful?

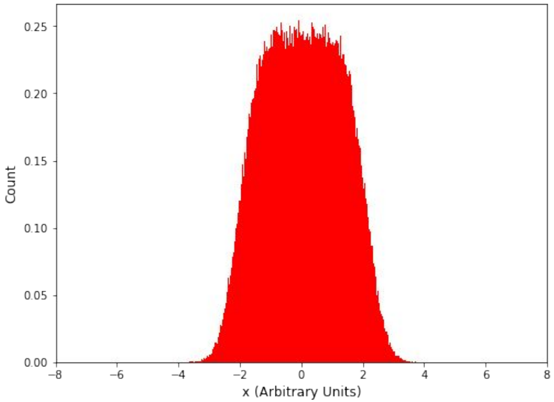


How does scaling work?

We divide the error distribution by S and increment S from 0 to 10 until we optimize p.

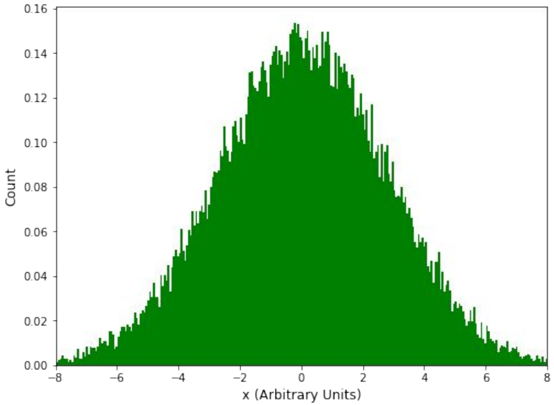


S>1



Best fit scaling makes the distribution more narrow → errors may have been overestimated

S<1



Best fit scaling makes the distribution wider → errors may have been underestimated

Dataset

Clustering of Local Group Distances: Publication Bias or Correlated Measurements? VI. Extending to Virgo Cluster Distance

Richard de Grijs and Giuseppe Bono 2020 *ApJS* 246 3

7922 hits for “M87” in the NASA/Astrophysics Data System

↓

A 213 post-1929 independent measurements ← Unadjusted

↓

B 44 “internally consistent” & “tight averages” (5 tracers) ← Adjusted to LMC distance

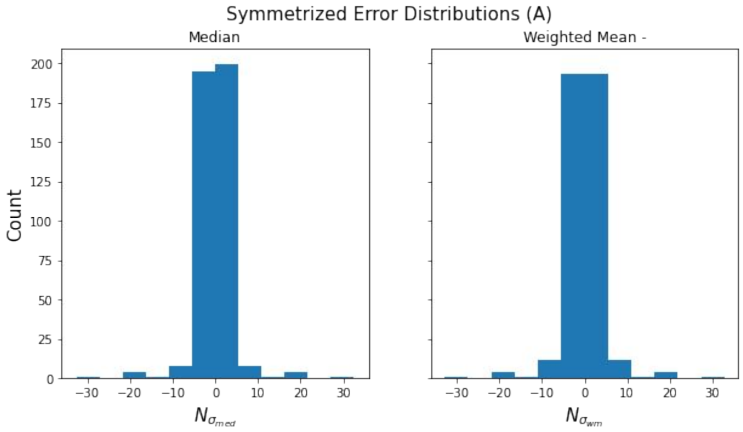
↓

…

de Grijs and Bono narrow it further

Dataset A

A 213 post-1929 independent measurements (15 tracers)



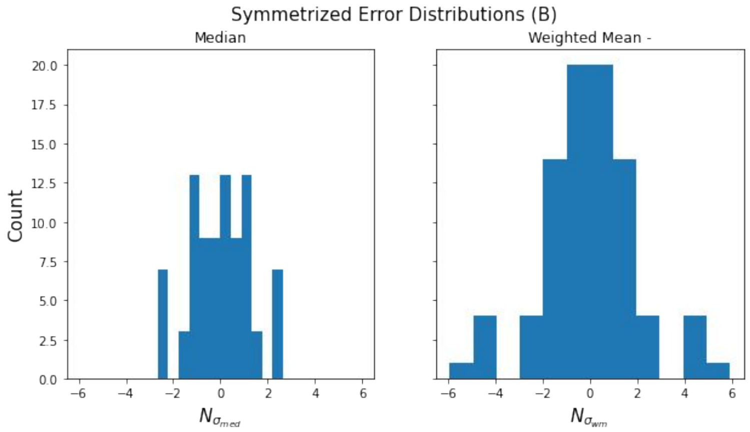
Error Distribution	Gaussian p-value
Median	<.001
Weighted Mean	<.001

Error Distribution	Gaussian p-value	Scale Factor
Median	.805	2.194
Weighted Mean	.619	2.336

Unscaled p value is low + optimal p requires high scaling → **errors may have been overestimated**

Dataset B

B 44 “internally consistent” & “tight averages” (5 tracers)



Error Distribution	Gaussian p-value
Median	.470
Weighted Mean	.089

Error Distribution	Gaussian p-value	Scale Factor
Median	.998	1.291
Weighted Mean	.992	1.791

Unscaled p value is low **+** optimal p requires high scaling **→ errors may have been overestimated**

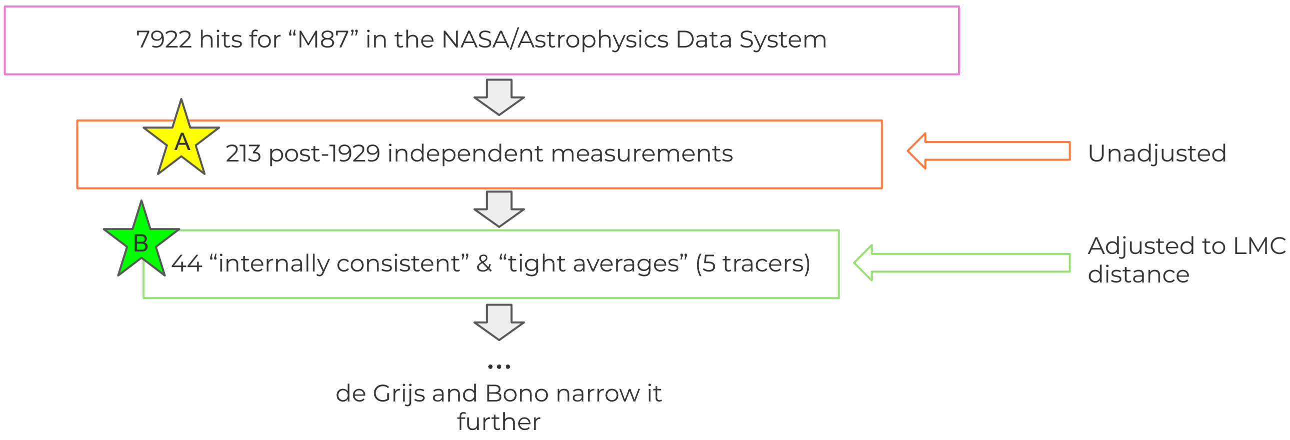
Recommended values

Dataset A: d = $31.08^{+0.04}_{-0.05}$(stat) → 16.44±0.53 Mpc (median)

Dataset B: d = $31.01^{+0.05}_{-0.08}$(stat) → 15.92±0.48 Mpc (median)

$$distance\ in\ pc = 10^{\frac{d}{5}+1}$$

De Grijs & Bono: d = 31.03± 0.14 (stat) → 16.07± 1.03Mpc (mean)

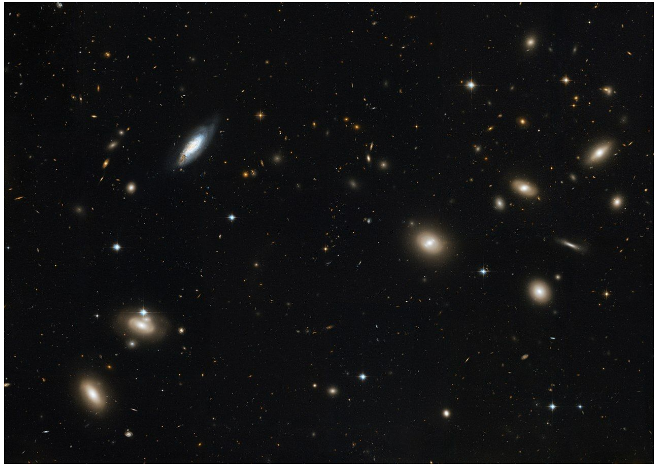


Conclusions

- Median statistics is a powerful alternative to mean statistics when the distribution of error-affected measurements is non-Gaussian
- Refine distance framework to more distant clusters

Fornax
~19 Mpc

Coma
~99 Mpc

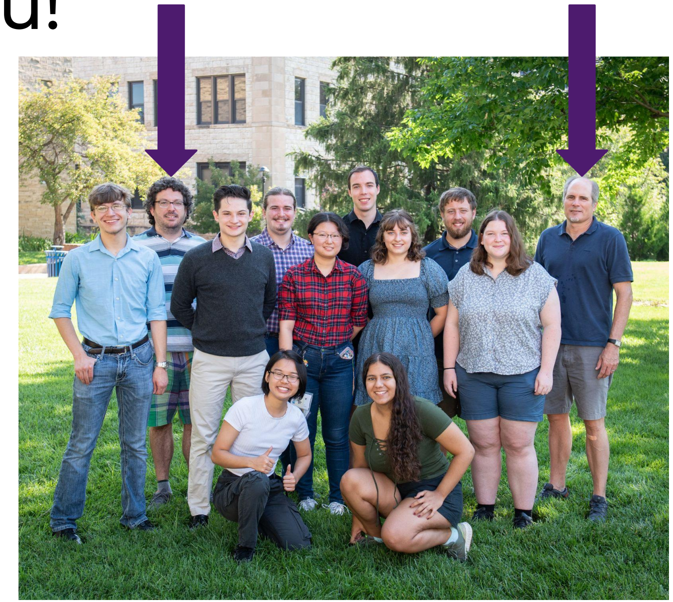


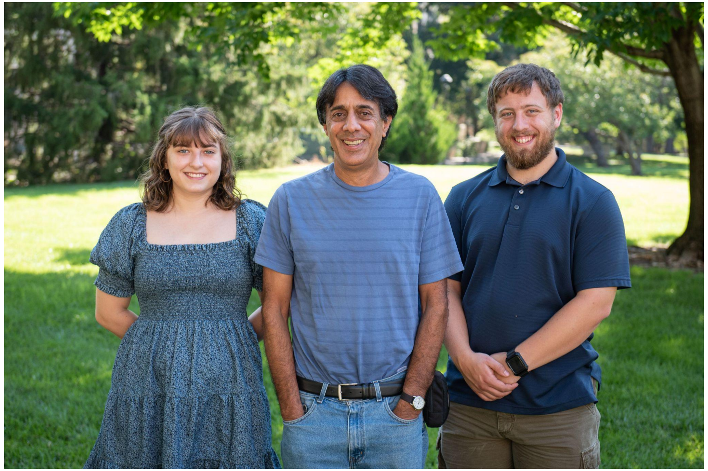
Acknowledgements–Thank you!

Dr. Bharat Ratra, Nicholas Rackers, Dr. Bret Flanders, Dr. Loren Greenman, Kim Coy

Kansas State University

The National Science Foundation





KANSAS STATE UNIVERSITY

College of Arts & Sciences

Department of Physics

Citations

J. Richard Gott III, Michael S. Vogeley, Silviu Podariu, & Bharat Ratra (2001). Median Statistics, H_0, and the Accelerating Universe*. The Astrophysical Journal, 549(1), 1–17.*

Richard de Grijs, & Giuseppe Bono (2019). Clustering of Local Group Distances: Publication Bias or Correlated Measurements? VI. Extending to Virgo Cluster Distances*. The Astrophysical Journal Supplement Series, 246(1), 3.*

Hubble, *Proceedings of the National Academy of Sciences*, 1929, 15, 168