A Median Statistic Analysis of the Distance to M87

~Nicholas Rackers

Black Hole at the center of M87

What is Messier 87

Galaxy in the center of the Virgo Cluster, the closest cluster to our local cluster.

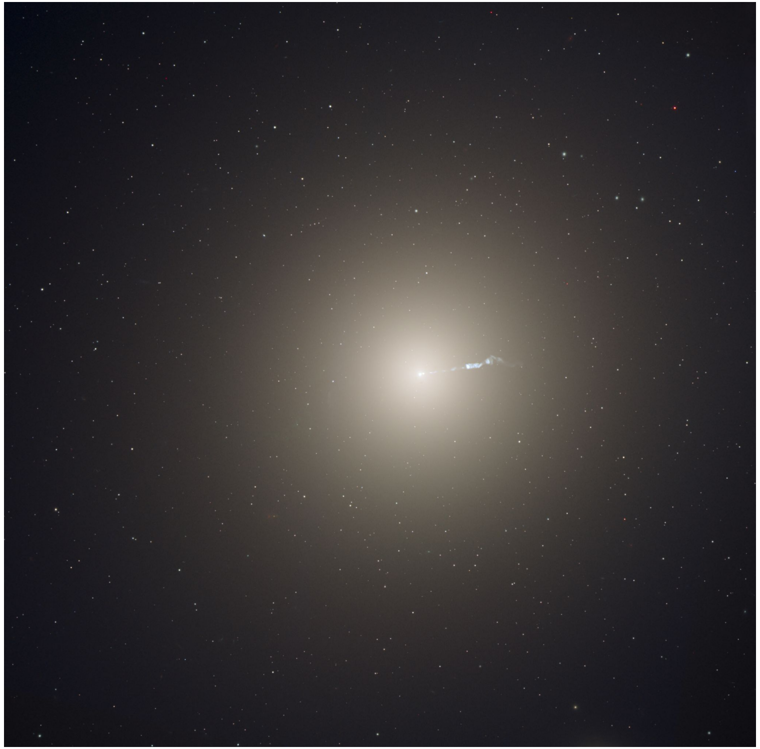
M87:

- Several Trillion Stars
- 15,000 surrounding Globular Clusters
- Giant (1,500 pc) plasma Jet tail
- 6.5 Billion M☉ Black Hole

VS

Milky Way:

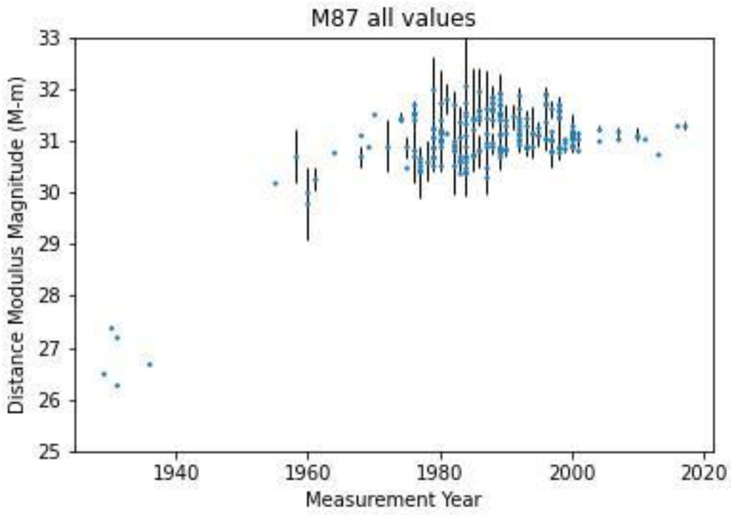
- Several Billion Stars
- 150 surrounding Globular Clusters
- Home to Earth
- 4.1 Million M☉ Black Hole



Galaxy M87

The Data Set

- 211 Independent measurements compiled by De Grijs and Bono
- 15 different measurement methods (tracers)



Distance Modulus:

(M-m) = 5 log(d/10)

[d] = [parsecs]

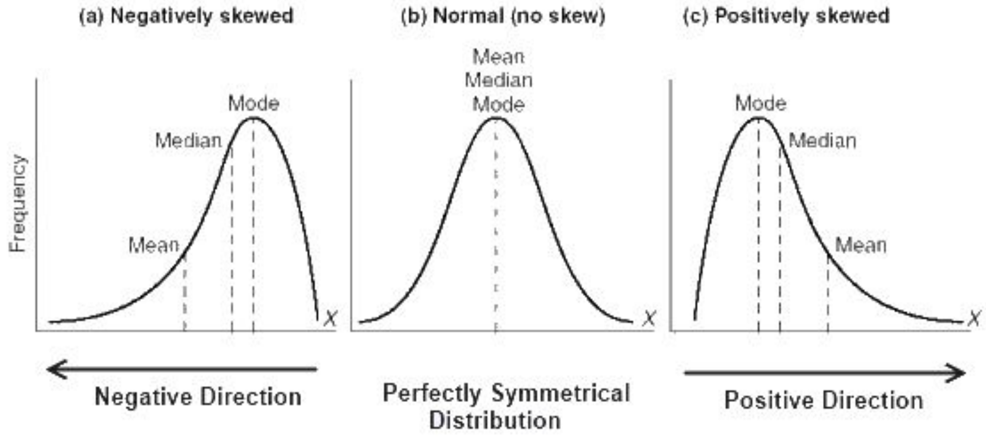
Median Statistics is resilient

Mean Assumptions

1. Statistically Independent
2. No systematic Effects
3. Errors are Gaussian
4. We know standard deviation

Median Assumptions

1. Statistically Independent
2. No systematic Effects
3. Not necessary
4. Not necessary



If our Distribution is non-Gaussian (non-Normal), the Median is a better estimate than the Mean.

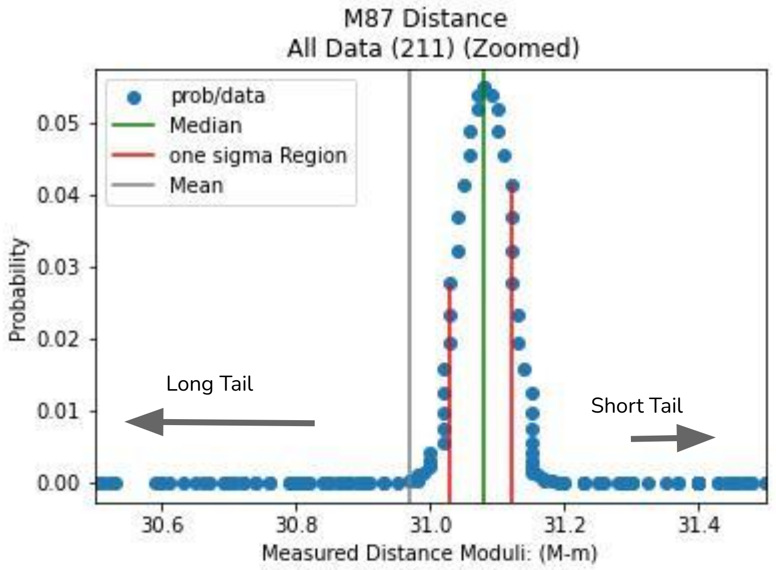
Finding the Median

- Median is the middle measurement
- 50% greater than median
- 50% less than median

Binomial Distribution!!!

Reported Values:

- Median 31.08 $\pm^{0.04}_{0.05}$
- Mean: 30.97 ± 0.07



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

P statistics (Choosing the Central Estimate)

- P < 0.95: Report the Median
- Optimal Scale >1: errors were underestimated as a whole

Reported Value:

- Median 31.08 $\pm^{0.04}_{0.05}$

Kolmogorov-Smirnov (KS) Test

Central Estimate	Scale (S)	P-value
Median	1	~0.0
Optimized Median	2.928	0.276
Weighted mean	1	~0.0
Optimized Weighted mean	2.296	0.983

Note: Weighted mean ignores any data with zero reported uncertainty

P ≥ 0.95 means the distribution is gaussian.
S > 1 means the errors were underestimated
S < 1 means the errors were overestimated

Systematic Effects

- Different tracers have different medians
- Statistical uncertainty is the variance in each tracer type
- Systematic uncertainty is the difference between the medians

Systematic Uncertainty:

$\pm^{0.04}_{0.06}$

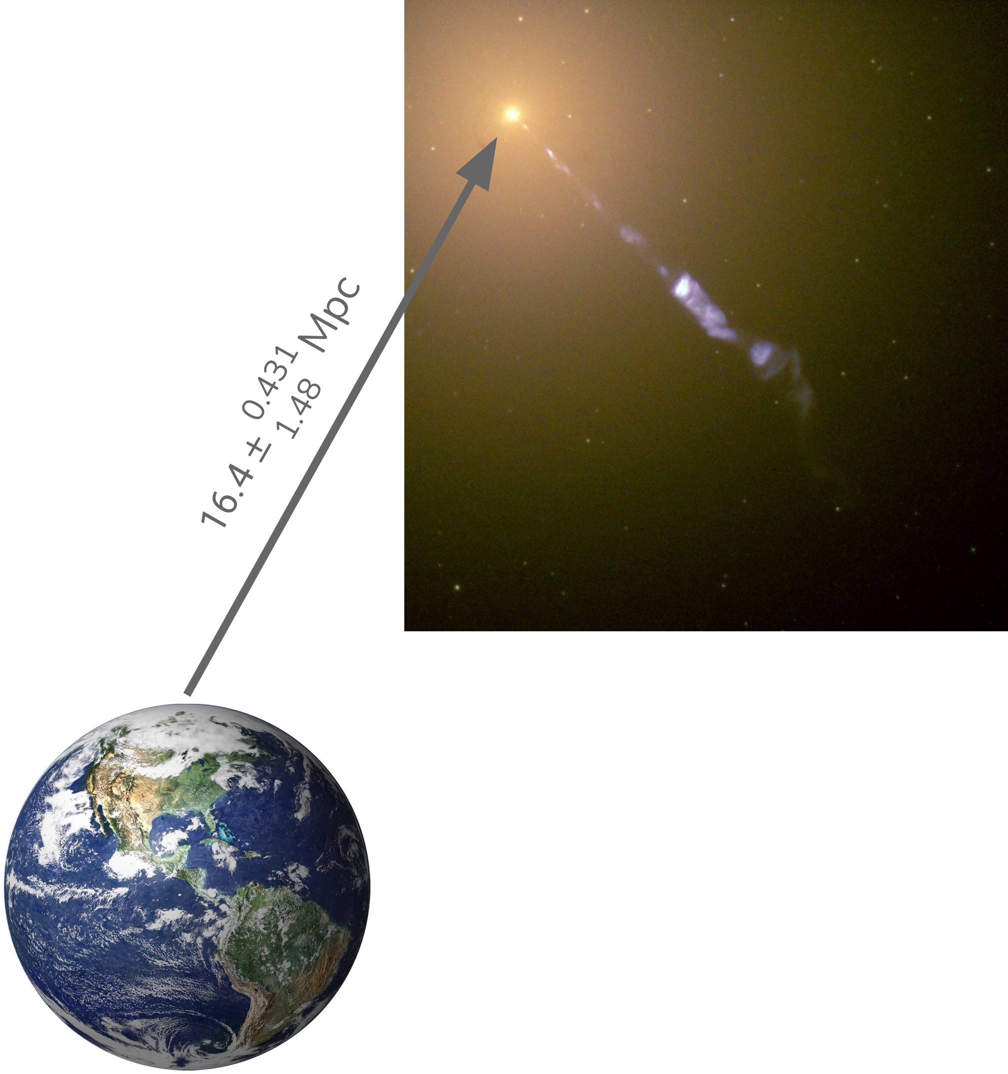
Tracer Type	#	Median (Error)	1σ Uncertainty (Width)
All Data	211	31.08	-0.05, +0.04(0.09)
TFR............	36	31.24 (0.16)	-0.16, +0.16 (0.32)
GCLF...........	32	31.11 (0.03)	-0.21, +0.12 (0.33)
Averages........	21	31.08 (0.0)	-0.06, +0.25 (0.31)
SBF............	18	31.12 (0.04)	-0.09, +0.03 (0.12)
SNe............	18	31.64 (0.56)	-0.12, +0.09 (0.21)
Other Methods...	15	30.9 (-0.18)	-0.1, +0.25 (0.35)
PNLF...........	12	30.86 (-0.22)	-0.03, +0.03 (0.06)
Faber-Jackson...	11	31.14 (0.06)	-0.3, +0.5 (0.8)
Color-magnitude.	11	30.84 (-0.24)	-0.17, +0.2 (0.37)
Novae..........	8	31.4	...
Hubble law......	8	27.3	...
Cepheids........	7	31.02	...
HII............	6	31.2	...
Group Membership	5	30.5	...
TRGB...........	3	31.05	...
Subgroup Medians	15	31.08	-0.06, +0.4 (0.1)

*Error is the difference between a subgroup median and the median of all data
*Subgroup Medians is the set of all previous medians

Final Reported value

We recommend a new Distance Modulus to M87 of

$31.08 \pm^{0.04}_{0.05}$ (statistical) $\pm^{0.04}_{0.06}$ (systematic)



Acknowledgements

KANSAS STATE
UNIVERSITY

College of Arts & Sciences
Department of Physics

- Sofia Splawska for working with me
- Dr. Bharat Ratra for guiding us
- Jacob Peyton and Aman Singal for helping us with the code

- Kansas State University Physics Department for hosting the REU
- National Science Foundation for providing funding

Sources:

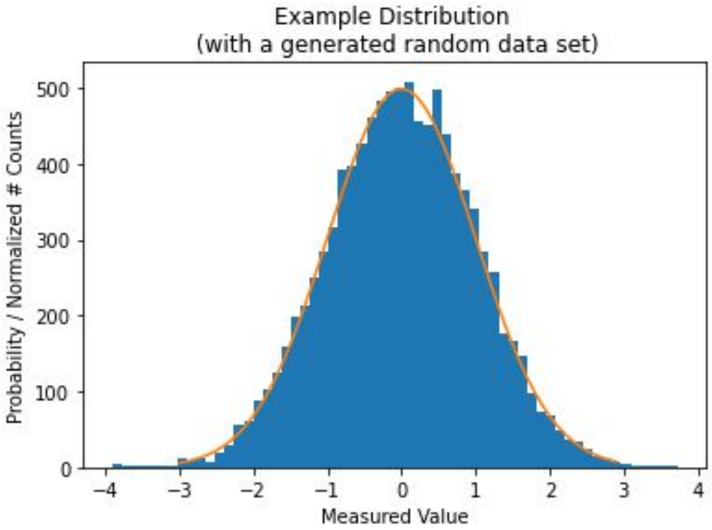
- R. de Grijs and G. Bono, The Astrophysical Journal Supplement Series 246, 3 (2019).
- J. R. G. III, M. S. Vogeley, S. Podariu, and B. Ratra, The Astrophysical Journal 549, 1 (2001).
- G. Chen and B. Ratra, Publications of the Astronomical Society of the Pacific 123, 1127 (2011).
- S. Crandall and B. Ratra, The Astrophysical Journal 815, 87 (2015).
- J. Penton, J. Peyton, A. Zahoor, and B. Ratra, Publications of the Astronomical Society of the Pacific 130, 114001 (2018).

Image Links:

- https://www.eso.org/public/images/eso2105a/
- https://www.nasa.gov/feature/goddard/2017/messier-87
- https://www.zarantech.com/blog/interview-questions/data-science-statistics-interview-questions-answers/
- https://www.flickr.com/photos/donkeyhotey/5679642883

A Typical Data Analysis (non-median)

- Take data
  - Measure values and estimate uncertainties
- Average the Values, compute standard deviation
- Finished



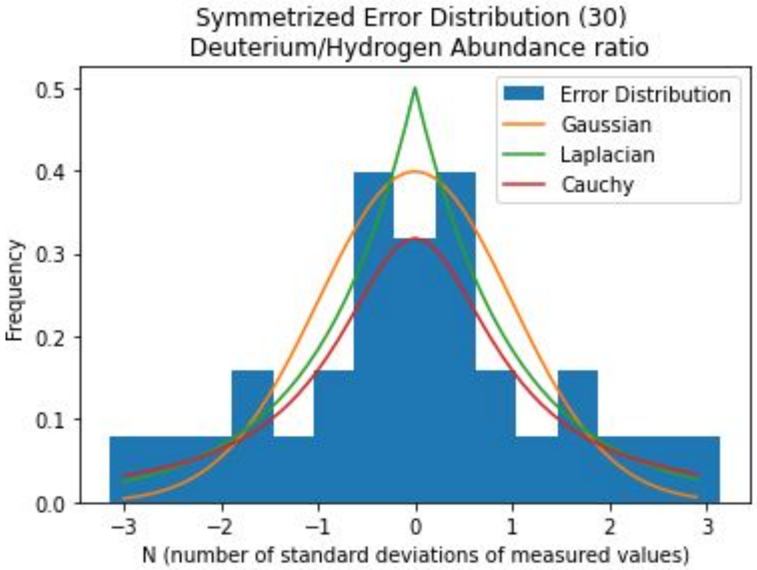
Data Assumptions

1. Statistically Independent
2. No systematic Effects
3. Errors are Gaussian (normal)
4. We know the standard deviation of errors

Creating Error Distribution

N is the number of standard deviations from the median, it is a measure of how much each measurement deviates from the central estimate (CE). Where M is the measurement and σ is the uncertainty.

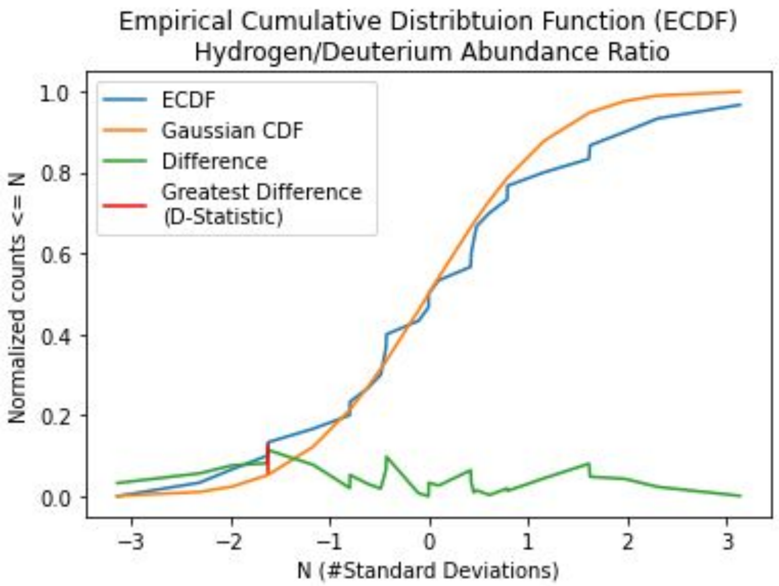
$$N_{\sigma_i} = \frac{M_i - M_{\text{CE}}}{\sqrt{\sigma_i^2 + \sigma_{\text{CE}}^2}} \; *$$



*Depending on the central estimate (mean, weighted mean, median) this equation slightly differs, the paper will contain every possibility

Compare Data to Gaussian

- We use the Kolmogorov-Smirnov (KS) test to compare the Error distribution (EDF) to the Probability Density Function (PDF).
- This compares the Cumulative Distribution Functions, finding the D-Statistic, being the maximum of the difference between the ECDF and the CDF
- The D-Statistic yields a p-value (0-1). If p>= 0.95 we can say with 95% confidence that the Data is Gaussian.*
- If the data is not gaussian, we expect Mean statistics to fail us, so median is a better value to report.



D-Statistic = 0.114
P-value = 0.809 < 0.95

*Technically we should say we can reject the Null Hypothesis which supposes that the data doesn't come from the PDF and thus we say that the data doesn't not come from a gaussian distribution

More detailed Analyses

- More Central Estimates (Arithmetic Mean, Median, Weighted Mean)
  - For Gaussian Distribution Weighted Mean would give best error estimates
  - For non-Gaussian Distribution Median gives best estimate
- More random distributions to check (Gaussian, Laplacian, Cauchy, Student's t)
  - We expect all error distributions to be gaussian
  - Seeing if errors are over/under estimated
- Accounting for Different Tracers (ways that the distance was measured)
  - Each would have their own systematic error
  - We can combine different tracers to discover their systematic errors

Outline Clean up the slides

- Describe M87 and its importance
- Outline the benefits of a Median analysis over a mean analysis
  - Probably reword the slide and stuff
- Introduce the Data (If too long I can just audially introduce?)
- Show the probability plot (I think it is helpful)
- Give some info on pstatistics (show the p values that are useful, scaling)
  - Conclusions from scaling and stuff
- Describe the systematic effects from different tracers and why we can conjecture the systematic uncertainty
  - Show the Medians table with pertinent values
  - Do I want fancy thing to show table later?
- Finiteime?
  - Include cool picture of M87 and report the distance (arrow with distance thingy?)
- Citations
  - Double check i have cited everything