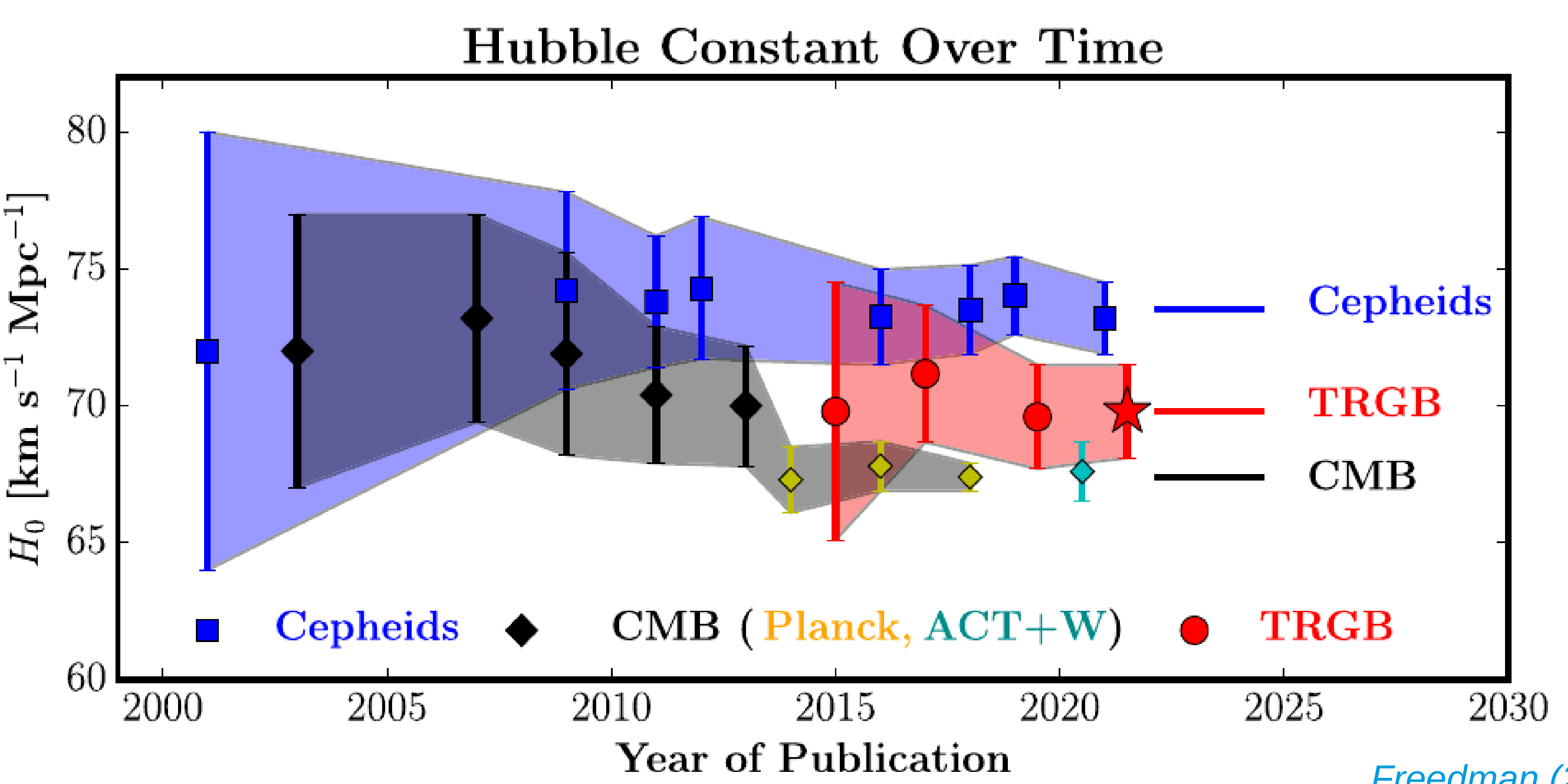
The Hubble tension survey: A statistical analysis of the 2012-2022 measurements

Bao Wang, Martin Lopez-Corredoira, Jun-Jie Wei
2024, MNRAS. 527, 7692

The Hubble Tension

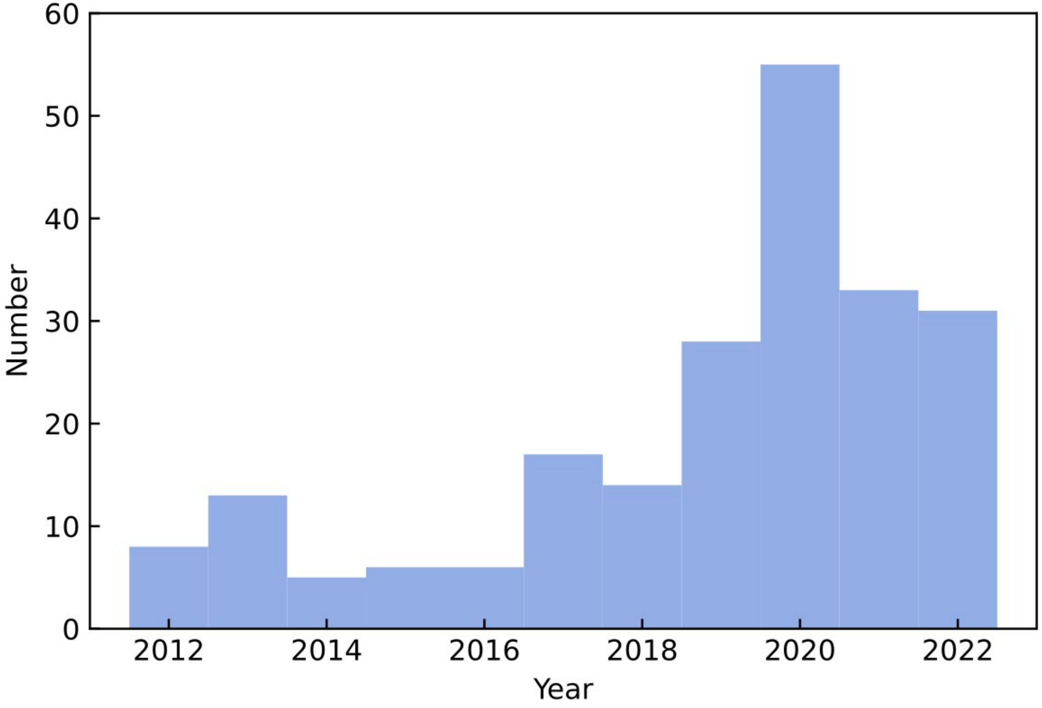


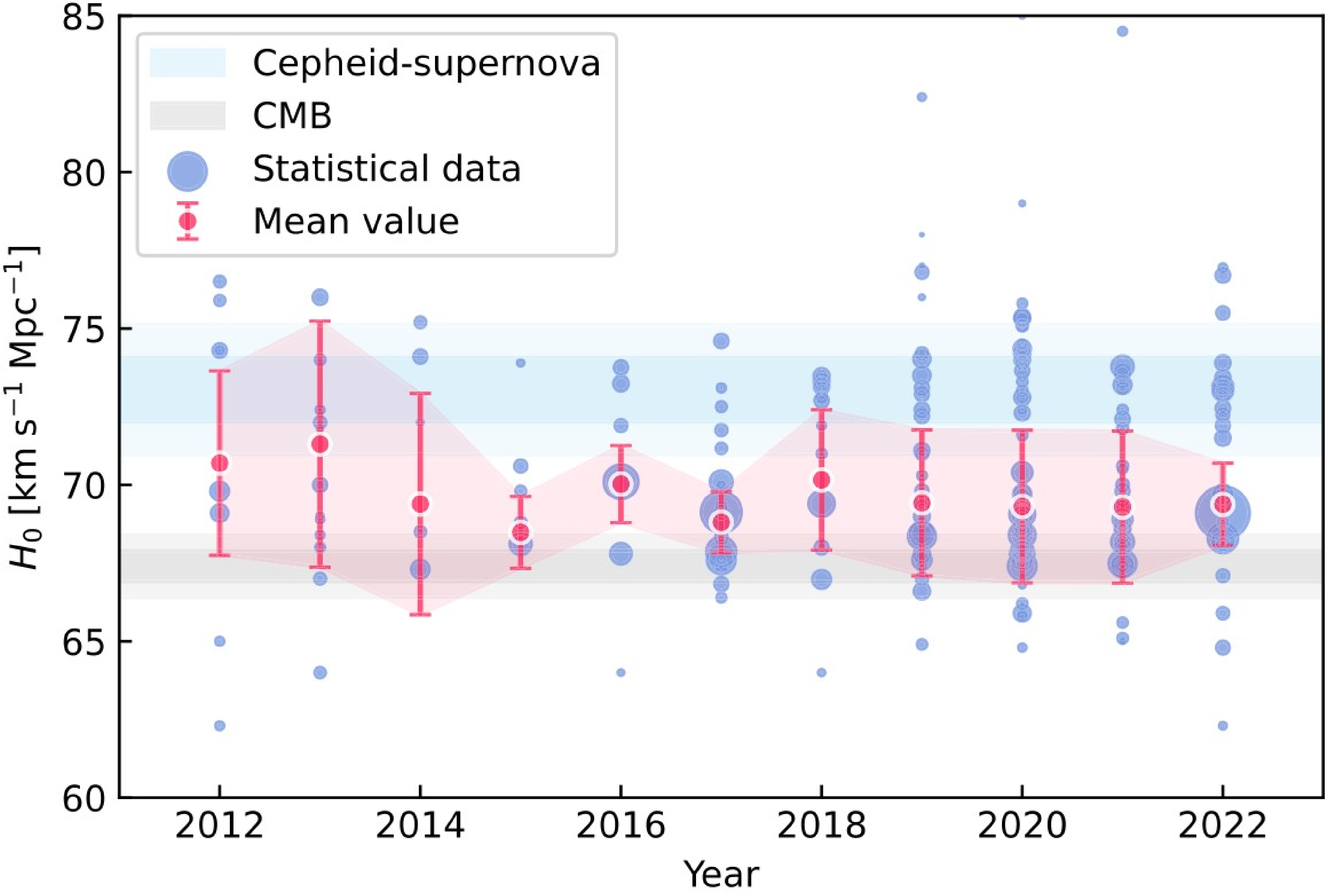
Freedman (2021)

DATA

216 H_0 measurements made between 2012 – 2022:

- **109** from model-independent methods (Cepheids + SNe Ia distance ladder)
- **107** from CMB under the standard ΛCDM model

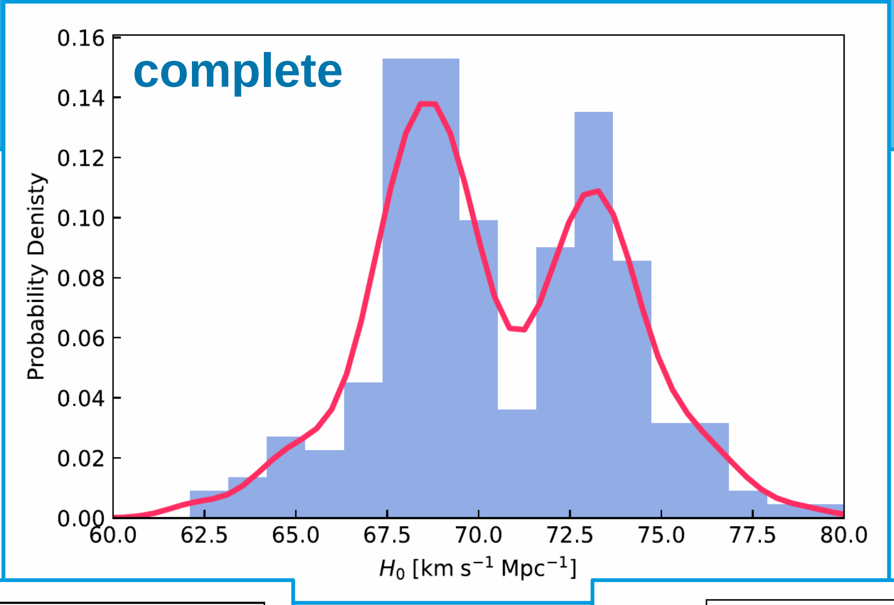


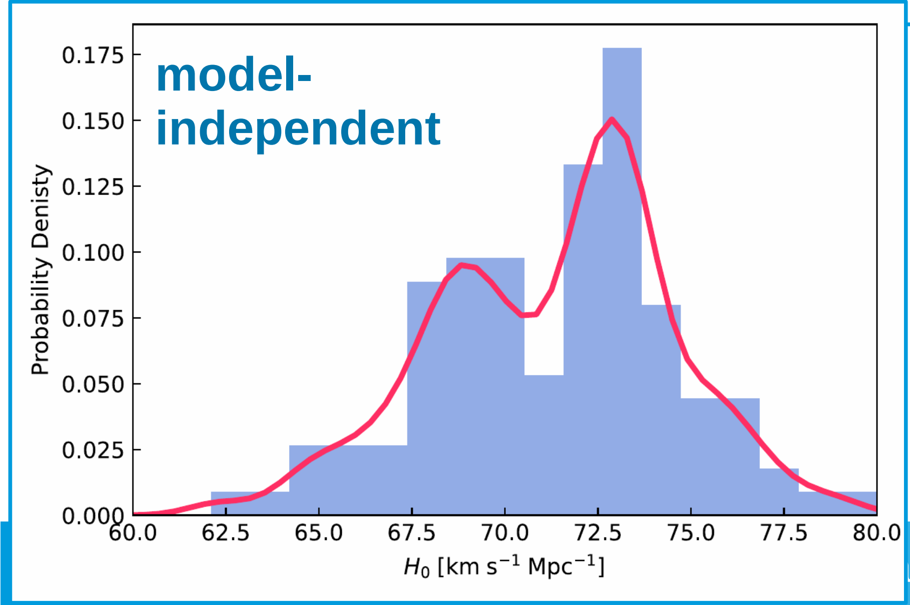

85
80
75
70
65
60
Cepheid-supernova
CMB
Statistical data
Mean value
H_0 [km s^{-1} Mpc^{-1}]
2012
2014
2016
2018
2020
2022
Year

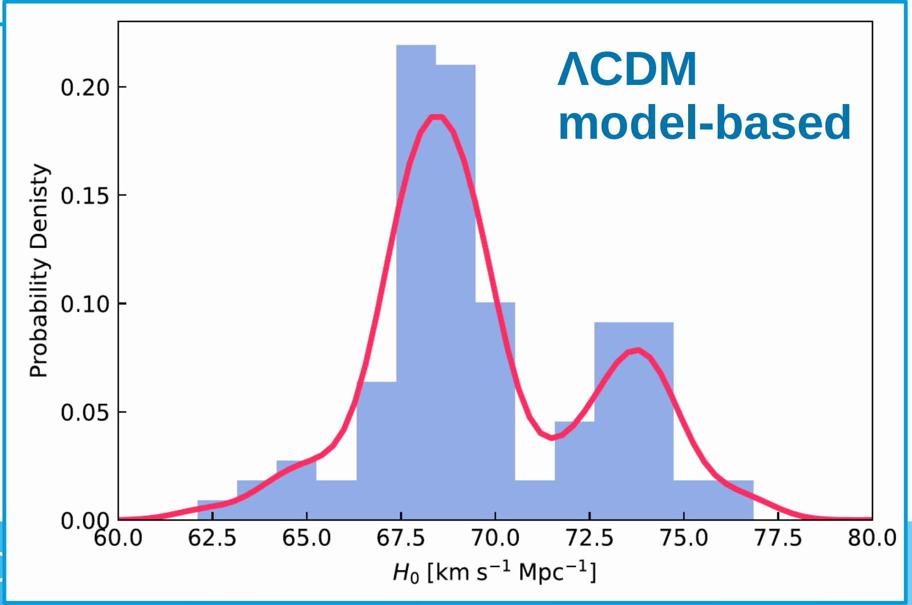
DATA

Data divided into 3 samples:

- **complete** (216)
- **model-independent** (109)
- **ΛCDM model-based** (107)


complete
Probability Denisty
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5
80.0
H_0 [km s^{-1} Mpc^{-1}]


model-
independent
Probability Denisty
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5
80.0
H_0 [km s^{-1} Mpc^{-1}]


ΛCDM
model-based
Probability Denisty
0.00
0.05
0.10
0.15
0.20
60.0
62.5
65.0
67.5
70.0
72.5
75.0
77.5
80.0
H_0 [km s^{-1} Mpc^{-1}]

1. Statistical significance of bimodality

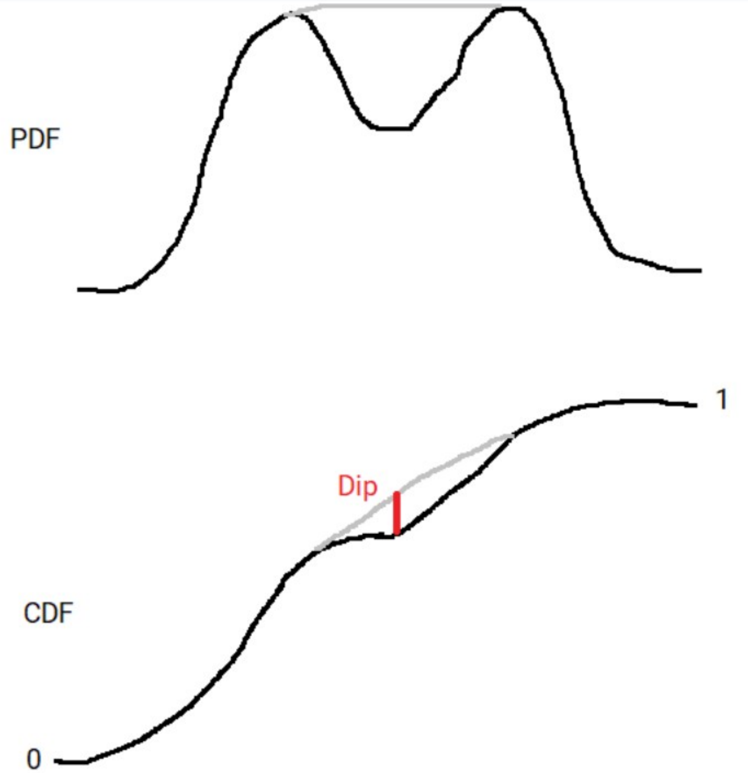
The **„Dip Test”**

- used to test multimodality of distributions (Hartigan & Hartigan 1985) by calculating the **discrepancy** between:
  - the **empirical distribution function** and
  - the **unimodal distribution that minimizes the maximum discrepancy**

1. Statistical significance of bimodality

The **„Dip Test"**

- the distribution can be deformed into a unimodal one by moving the CDF **by at most the dip** at each point
- the **Dip** is the **smallest number** for which this is true



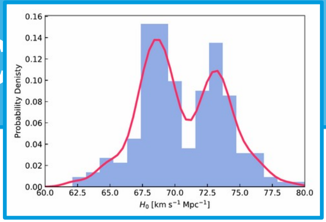
https://skeptric.com/dip-statistic/index.html

1. Statistical significance of bimodality

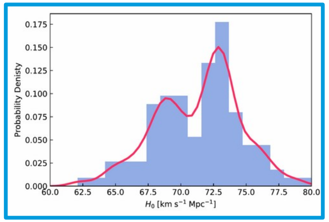
The **„Dip Test"**

- the ***p* value** for the dip test – the probability of unimodality:
  - $p < 0.05$ – significant multimodality
  - $0.05 < p < 0.10$ – marginal multimodality
  - $p \gg 0.05$ – unimodality

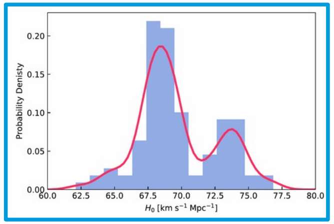
1. Statistical signific … bimodality



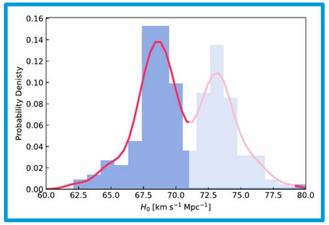
Data divided into 5 samples:



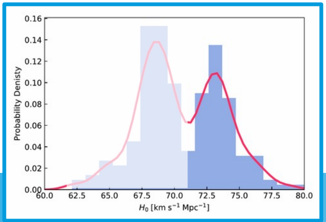
- **complete** (216)



- **model-independent** (109)
- **ΛCDM model-based** (107)



- **H_0 < 71 km s^{-1} Mpc^{-1}** (118)

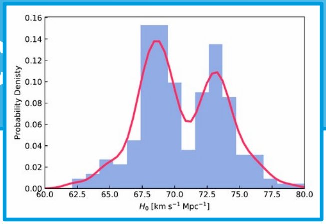


- **H_0 ≥ 71 km s^{-1} Mpc^{-1}** (98)

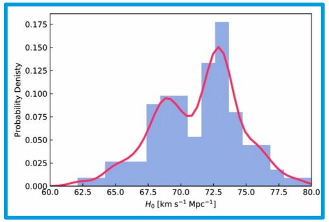
1. Statistical signific… bimodality

Data divided into 5 samples:

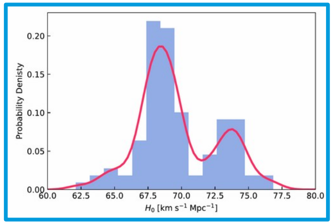
- **complete** (216)
- **model-independent** (109)
- **ΛCDM model-based** (107)
- **H_0 < 71 km s^{-1} Mpc^{-1}** (118)
- **H_0 ≥ 71 km s^{-1} Mpc^{-1}** (98)



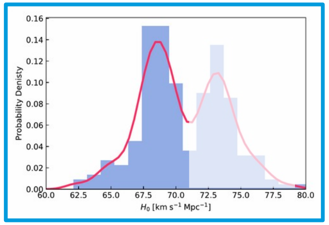
***p* = 0.01**



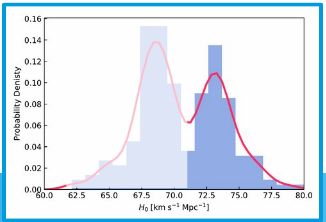
***p* = 0.46**



***p* = 0.16**



***p* = 0.96**



***p* = 0.98**

2. Statistical significance of measurements

2. Statistical significance of measurements

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(H_{0,i} - \overline{H_0}\right)^2}{\sigma_i^2}$$

calculated for the same 5 subsamples

2. Statistical significance of measurements

Table 1. The results of dip tests, weighted averages and χ^2 fittings in three categories: complete, model-independent, and ΛCDM model-based measurements. The Q values measure the statistical significance for χ^2 fittings, and the p values measure the significance of the unimodality for dip tests.

	Number	$\overline{H_0}$ (σ) (km s^{-1} Mpc^{-1})	χ^2	Q	p
Complete	216	69.35 ± 0.12	515.99	8.85×10^{-27}	0.01
Model-independent	109	70.82 ± 0.22	181.48	1.23×10^{-5}	0.46
ΛCDM model-based	107	68.94 ± 0.13	237.56	4.36×10^{-12}	0.16
$H_0 < 71$ km s^{-1} Mpc^{-1}	118	68.78 ± 0.07	95.09	0.93	0.96
$H_0 \geq 71$ km s^{-1} Mpc^{-1}	98	73.53 ± 0.13	30.05	$\sim$1.00	0.98

Q > 0.05 means „statistically significant”

Q measures the probability that trends come from chance erros
==> it is unlikely that the trends in 3 groups are due to chance errors

3. Outlier rejection

How many outliers should we remove to to reach $Q > 0.05$?

3. Outlier rejection

How many outliers should we remove to to reach Q > 0.05 ?

define x – the number of σ deviations between the measurements $H_{0,i}$ and the average $\overline{H_0}$

$$x = \frac{\left| H_{0,i} - \overline{H_0} \right|}{\sigma_i}$$

then exclude data with $x > x_{min}$ and repeat the analysis

3. Outlier rejection

Table 2. The results of weighted averages and χ^2 fittings in three categories after removing the outliers.The number of outliers and minimal deviations are also displayed.

	$x_{\min}$	Outliers	Number	$\overline{H_0}$ (σ) $(\mathrm{km\,s^{-1}\,Mpc^{-1}})$	χ^2	Q
Complete	2.4	27	189	69.17 ± 0.09	216.42	0.08
Model-independent	3.6	1	108	72.45 ± 0.21	84.20	0.95
ΛCDM model-based	2.6	13	94	68.85 ± 0.10	106.33	0.16

3. Outlier rejection

Table 2. The results of weighted averages and χ^2 fittings in three categories after removing the outliers.The number of outliers and minimal deviations are also displayed.

	$x_{\min}$	Outliers	Number	$\overline{H_0}$ (σ) (km s^{-1} Mpc^{-1})	χ^2	Q
Complete	2.4	27	189	69.17 ± 0.09	216.42	0.08
Model-independent	3.6	1	108	72.45 ± 0.21	84.20	0.95
ΛCDM model-based	2.6	13	94	68.85 ± 0.10	106.33	0.16

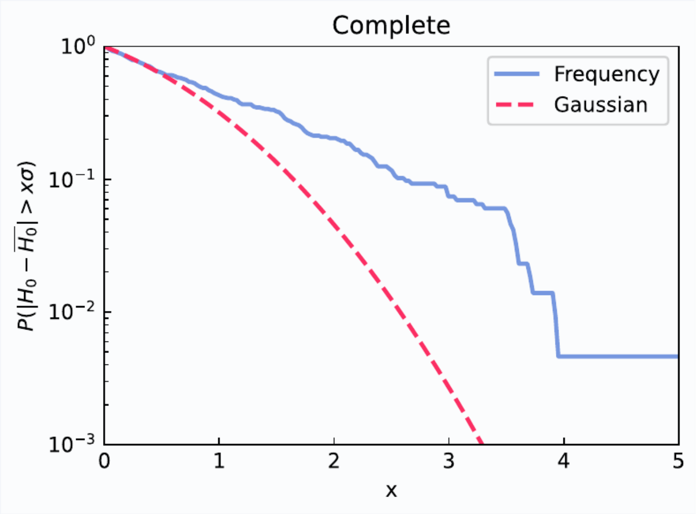
a sign of a possible tension

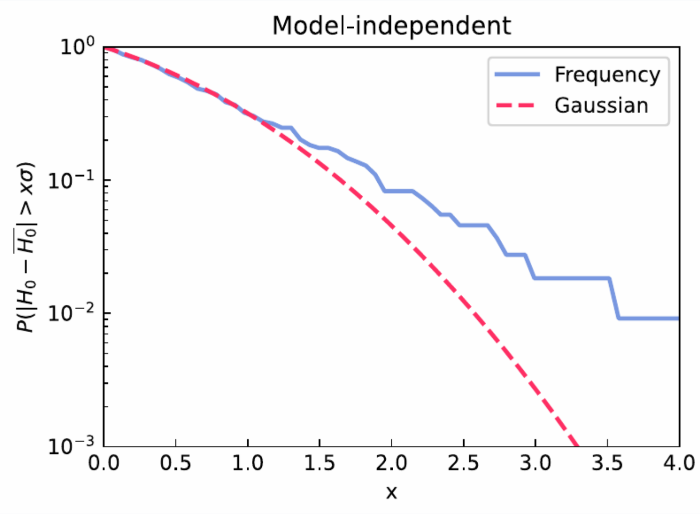
==> what is the degree of the tension?

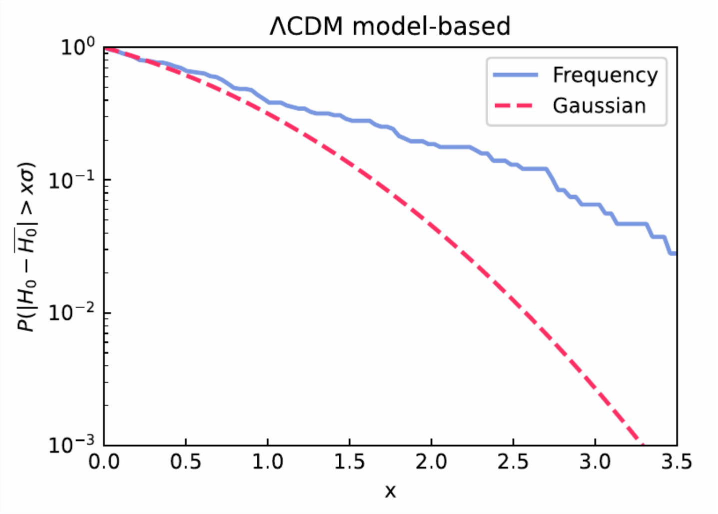
4. Estimating the real degree of tension

4. Estimating the real degree of tension

the frequency of deviations larger than $x\sigma$ from $\overline{H_0}$



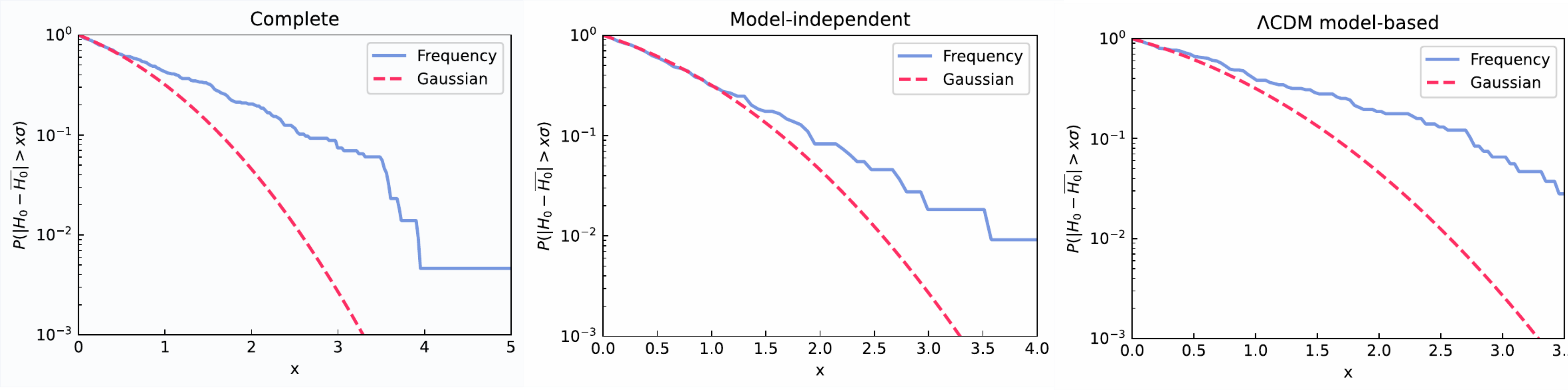




measures the degree to which the sample deviates from the Gaussian

4. Estimating the real degree of tension

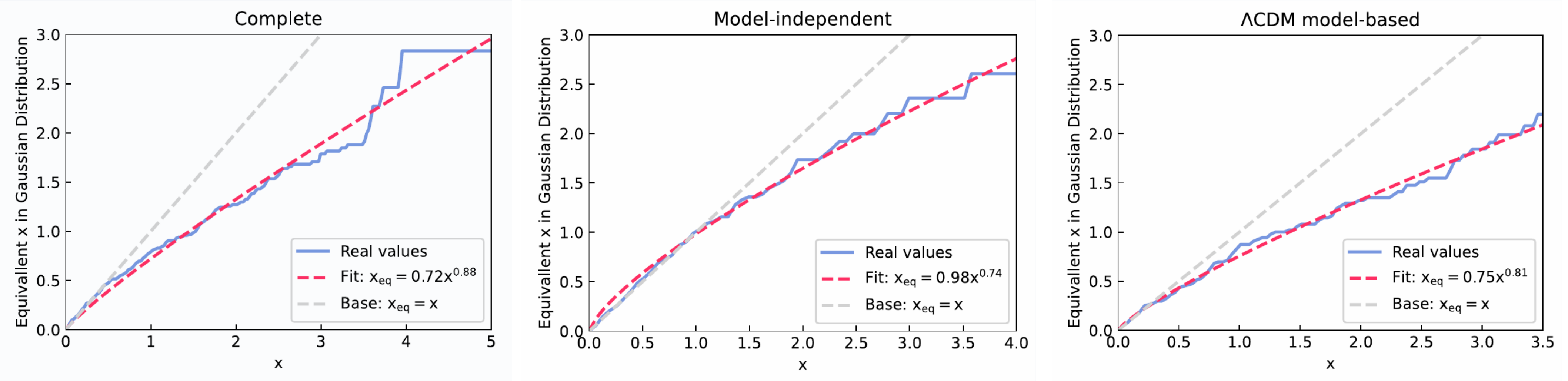
the frequency of deviations larger than $x\sigma$ from $\overline{H_0}$



==> define x_{eq} – the equivalent deviation between the probabilities of the the Gaussian distribution and the real frequency

4. Estimating the real degree of tension

fit a function $x_{eq} = a * x^b$



i.e. when the deviation of $H_{0,i}$ from $\overline{H_0}$ is $x\sigma$, it is actually $x_{eq}\sigma$

4. Estimating the real degree of tension

$$x_{\rm eq} = (0.719 \pm 0.013) x^{0.879 \pm 0.014} \quad \text{(Complete)}$$

$$x_{\rm eq} = (0.983 \pm 0.012) x^{0.744 \pm 0.011} \quad \text{(Model-independent)}$$

$$x_{\rm eq} = (0.750 \pm 0.007) x^{0.818 \pm 0.009} \quad (\Lambda\text{CDM model-based})$$

4. Estimating the real degree of tension

$$x_{\rm eq} = (0.719 \pm 0.013)x^{0.879\pm0.014} \quad \text{(Complete)}$$

$$x_{\rm eq} = (0.983 \pm 0.012)x^{0.744\pm0.011} \quad \text{(Model-independent)}$$

$$x_{\rm eq} = (0.750 \pm 0.007)x^{0.818\pm0.009} \quad (\Lambda\text{CDM model-based})$$

5σ tension reported by Riess et al. (2022) becomes:

$$x_{eq} = 0.719 * x^{0.879} = 0.719 * 5^{0.879} = (2.96 \pm 0.12)\sigma \approx 3\sigma$$

4. Estimating the real degree of tension

x_{eq} calibration should be independent of of the data that are tested, so authors recalibrate x_{eq} with data from 1976-2019 and get:

$$x_{eq} = 0.83 * x^{0.62} = 0.83 * 5^{0.62} = 2.25\sigma$$

5. The effect of correlated data

Data divided into 5 samples after removing correlated datasets:

- **complete** (216 → 152)
- **model-independent** (109 → 71)
- **ΛCDM model-based** (107 → 81)
- **$H_0 < 71$ km s^{-1} Mpc^{-1}** (118 → 85)
- **$H_0 \geq 71$ km s^{-1} Mpc^{-1}** (98 → 67)

5. The effect of correlated data

Table 4. The results of dip tests, weighted averages and χ^2 fittings in three categories after removing some of the correlated measurements. The Q values measure the statistical significance for χ^2 fittings, and the p values measure the significance of the unimodality for dip tests.

	Number	$\overline{H_0}$ (σ) (km s^{-1} Mpc^{-1})	χ^2	Q	p
Complete	152	69.25 ± 0.13	357.49	8.68×10^{-19}	0.12
Model-independent	71	70.32 ± 0.25	123.61	8.20×10^{-5}	0.62
ΛCDM model-based	81	68.99 ± 0.15	193.89	1.91×10^{-11}	0.13
$H_0 < 71$ km s^{-1} Mpc^{-1}	85	68.84 ± 0.09	81.13	0.57	0.69
$H_0 \geq 71$ km s^{-1} Mpc^{-1}	67	73.65 ± 0.15	18.35	$\sim$1.00	0.98

5. The effect of correlated data

Table 4. The results of dip tests, weighted averages and χ^2 fittings in three categories after removing some of the correlated measurements. The Q values measure the statistical significance for χ^2 fittings, and the p values measure the significance of the unimodality for dip tests.

	Number	$\overline{H_0}$ (σ) (km s^{-1} Mpc^{-1})	χ^2	Q	p
Complete	152	69.25 ± 0.13	357.49	8.68×10^{-19}	0.12
Model-independent	71	70.32 ± 0.25	123.61	8.20×10^{-5}	0.62
ΛCDM model-based	81	68.99 ± 0.15	193.89	1.91×10^{-11}	0.13
$H_0 < 71$ km s^{-1} Mpc^{-1}	85	68.84 ± 0.09	81.13	0.57	0.69
$H_0 \geq 71$ km s^{-1} Mpc^{-1}	67	73.65 ± 0.15	18.35	$\sim$1.00	0.98

no more multimodality in the **complete** sample (before: p=0.01)

5. The effect of correlated data

Table 4. The results of dip tests, weighted averages and χ^2 fittings in three categories after removing some of the correlated measurements. The Q values measure the statistical significance for χ^2 fittings, and the p values measure the significance of the unimodality for dip tests.

	Number	$\overline{H_0}$ (σ) (km s^{-1} Mpc^{-1})	χ^2	Q	p
Complete	152	69.25 ± 0.13	357.49	8.68×10^{-19}	0.12
Model-independent	71	70.32 ± 0.25	123.61	8.20×10^{-5}	0.62
ΛCDM model-based	81	68.99 ± 0.15	193.89	1.91×10^{-11}	0.13
$H_0 < 71$ km s^{-1} Mpc^{-1}	85	68.84 ± 0.09	81.13	0.57	0.69
$H_0 \geq 71$ km s^{-1} Mpc^{-1}	67	73.65 ± 0.15	18.35	$\sim$1.00	0.98

Q measures the probability that trends come from chance erros
==> it is unlikely that the trends in 3 groups are due to chance errors

TAKEAWAY

- the Hartigans' **Dip Test** is good to test multimodality
- the **bimodal distribution of H_0** is also present in a model-independent sample and a ΛCDM model-based sample
- the deviation of H_0 measurements with respect to $\overline{H_0}$ are larger than expected from their error bars if they follow a Gaussian distribution ==> **5σ tension may in fact be a 3σ tension**
- there is a **compatibility with a Gaussian** distribution for samples where of $H_0 < 71$ km s^{-1} Mpc^{-1} and where $H_0 \geq 71$ km s^{-1} Mpc^{-1} without the need to remove outliers ==> error underestimation related to methodology (systematics)