How to statistically distinguish several distributions ?

KRUSKAL WALLIS TEST

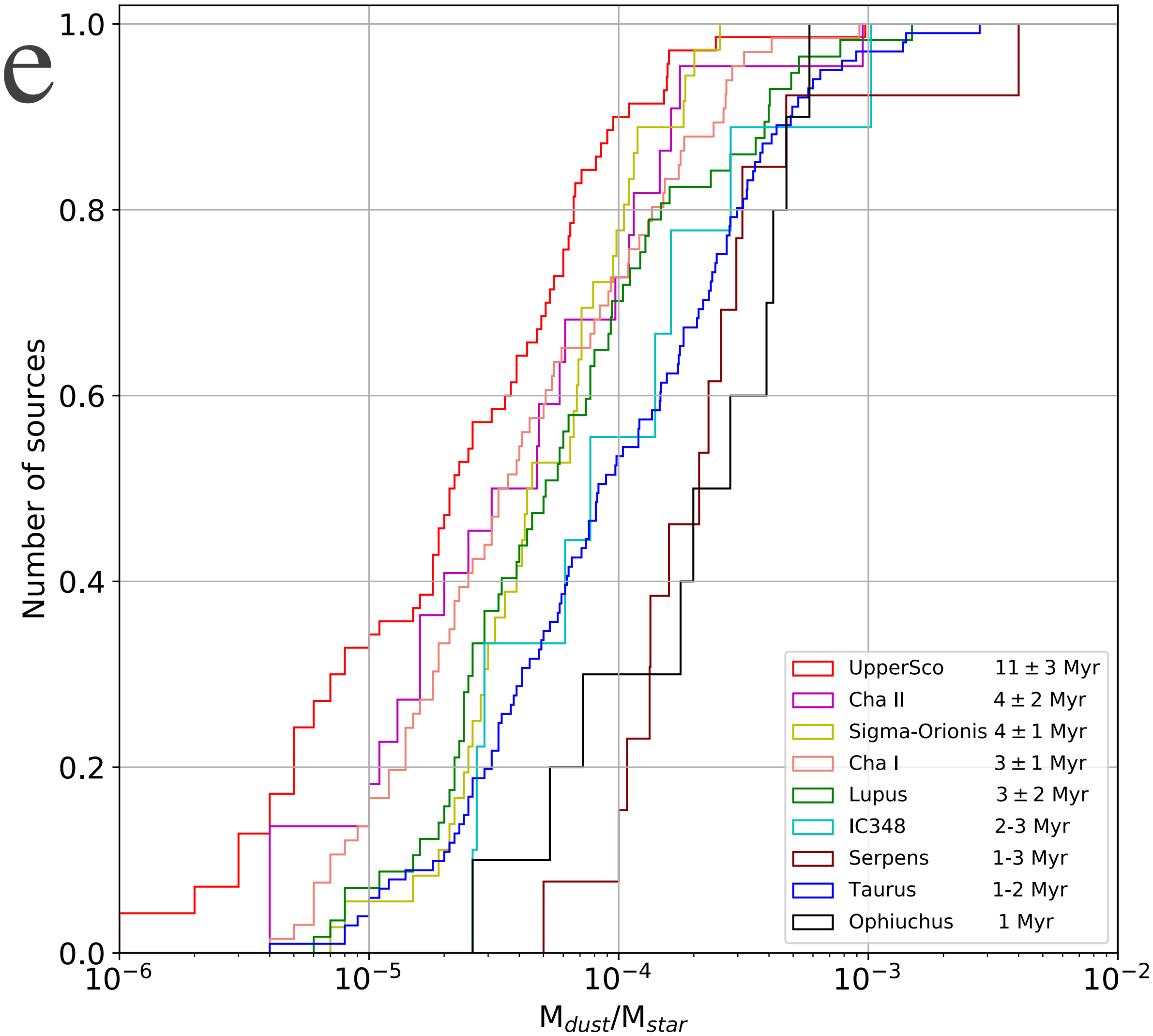
MARION VILLENAVE - 17 NOVEMBER 2017 – MCMC COFFEE

Science case

Context

- Protoplanetary disks
- Planet formation
- **Disk evolution**

-> 9 Dust mass distributions



Are those regions statistically distinct from each other ?

Kruskal-Wallis test

- Testing for differences between several independent groups
- Testing if samples come from the same population or if at least one come from a different population

- Alternative to ANOVA test :
  - When data don’t come from a normal distribution
  - No hypothesis on variances

- If 2 samples only :
  - equivalent to Mann-Whitney tests

Kruskal-Wallis test

- Non-parametric test :
  - Based on ranked data

- Hypothesis :
  - H_0 : All populations have the same **median** yield
  - Ha : Not all medians yields are equals

Kruskal-Wallis test

Groupe 1
$m_{1\,0} \ldots m_{1\,n1}$

Groupe 2
$m_{2\,0} \ldots m_{2\,n2}$

Groupe 3
$m_{3\,0} \ldots m_{3\,n3}$

Rank all data from all groups together

Tied values
Get the mean value of their rank

Rank	**Data**
1	$m_{2\,0}$
2	$m_{1\,7}$
3	$m_{3\,2}$
4	$m_{2\,4}$
5	$m_{2\,5}$
6	$m_{2\,1}$
7	$m_{1\,0}$
…	…

Kruskal-Wallis test

Groupe 1
$m_{1\,0} \ldots m_{1\,n1}$

Groupe 2
$m_{2\,0} \ldots m_{2\,n2}$

Groupe 3
$m_{3\,0} \ldots m_{3\,n3}$

n_i	Number of observations in group i
r_{ij}	Rank of observation j from group i
N	Total number of observations
$\bar{r_{i.}} = \frac{\sum_{j=0}^{n_i} r_{ij}}{n_i}$	Average rank of all observations in group i
$\bar{r} = \frac{1}{2}(N+1)$	Average of all r_{ij}

$$H = \frac{12}{N(N+1)} \sum_{groups} n_i \bar{r_{i.}}^2 - 3(N+1)$$

Calculating the associated p-value

xlstat options

- **Assymptotic method**

  p-value obtained from approximation of the law by a χ^2 law with g-1 degrees of freedom

  Reliable except if N small

- **Exact method**

  Using the real H distribution

  Numerically intense

- **MC method**

  Random ressampling and estimation of a confidence interval around p-value

  Better results for large number of simulations

Interpretation

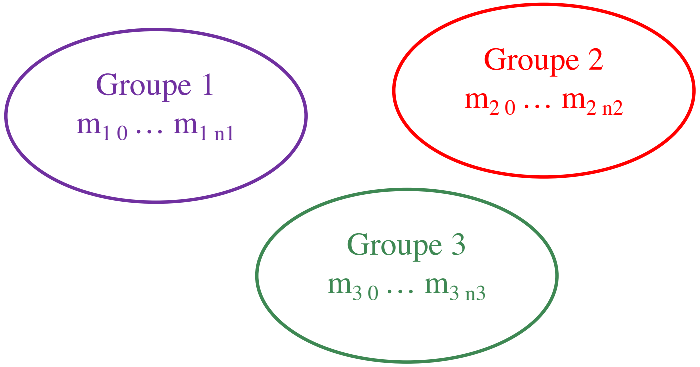
α : critical value to interpret the test (usually 0.05)

- **p-value < α** : **H_0 rejected** – Distributions are statistically different
  (large H)
- p-value > α : H_0 accepted – Distributions are similar
  (smaller H)

	UpperSco	Cha II	σ-Ori	Cha I	Lupus	Taurus
UpperSco		.	.	.	Yes	Yes
Cha II	0.113		.	.	.	.
σ-Ori	0.002	0.362		.	.	.
Cha I	0.008	0.784	0.387		.	Yes
Lupus	< 0.0001	0.190	0.7	0.148		.
Taurus	< 0.0001	0.004	0.026	0.0001	0.035	

α = 0.014

Bonferroni correction



Probability of falsely rejecting null hypothesis increase with the number of samples compared
(inflation of type I error)

$$\alpha_{corrected} = \alpha \,/\, \text{nb of tests conducted}$$

Example :

- For 4 regions
- Number of tests : 3 + 2 + 1 = 6
- So for $\alpha = 0.05$, $\alpha_{corrected} = 0.05/6 = 0.008$

Changes the interpretation of the results because we compare several samples all together

Bibliography

Original Article
Use of Ranks in One-Criterion Variance Analysis
W. H. Kruskal and W. A. Wallis, 1952
Journal of the American Statistical Association
http://www.tandfonline.com/doi/abs/10.1080/01621459.1952.10483441

Wikipedia
https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance

Use the test ?

XLSTAT
https://www.xlstat.com/en/solutions/features/kruskal-wallis-test

Python
scipy.stats.kruskal -> statistic + p-value for 2 or more distributions