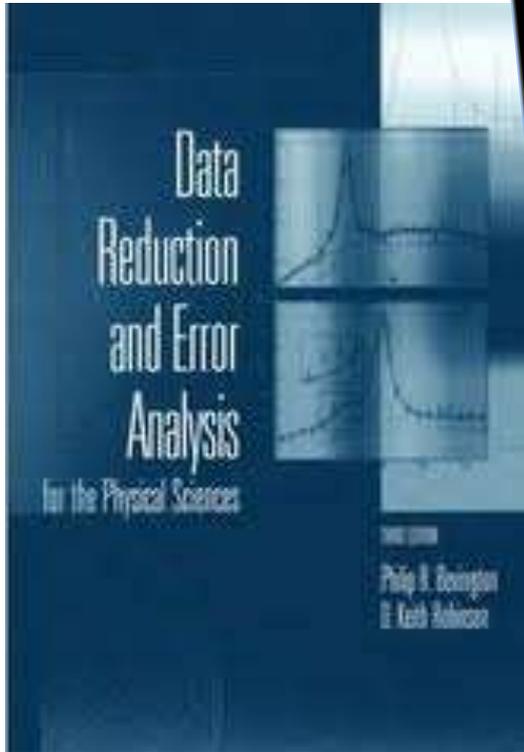
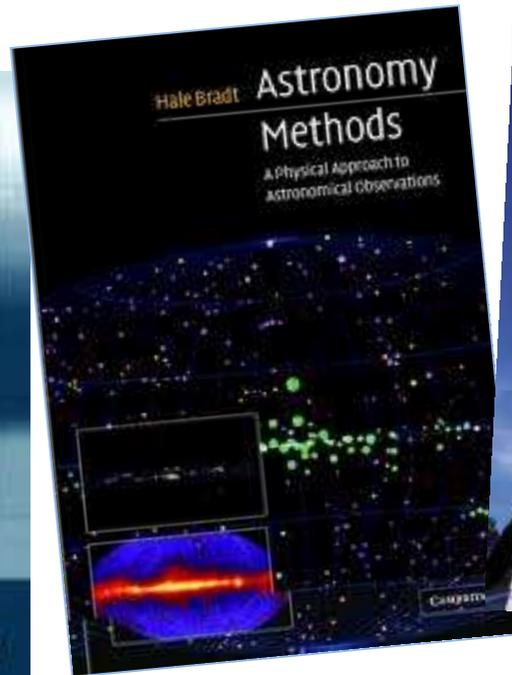


Tratamento de dados e erros

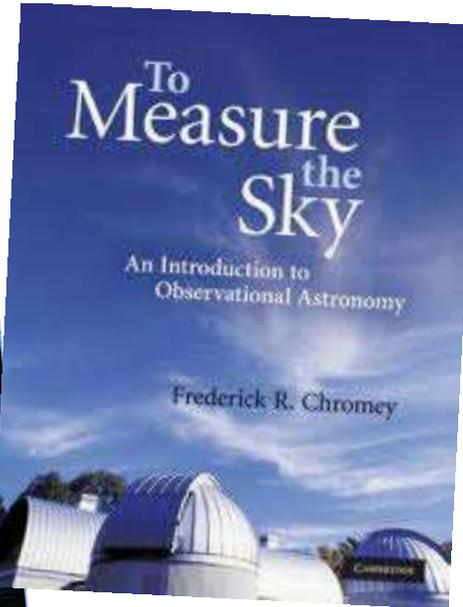
Estatística simples



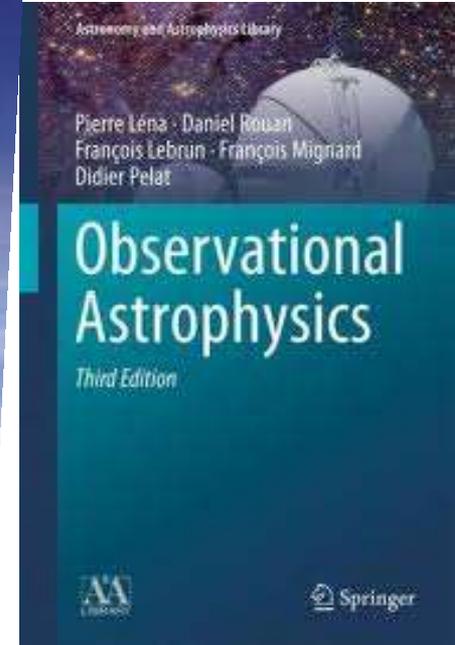
Philip R. Bevington & D. Keith Robinson 2002, 3rd edition



Astronomy Methods. H. Bradt, 2004. Section 6.5, pp. 151 - 172



To Measure the Sky. An Introduction to Observational Astronomy. F. R. Chromey, 2010. Chapter 2.



Observational Astrophysics, 3rd Ed., P. Léna et al., 2012. Appendix B 1

Estatística robusta

<http://www.stats.ox.ac.uk/pub/StatMeth/Robust.pdf>
M.Sc. in Applied Statistics MT2004

Robust Statistics

©1992–2004 B. D. Ripley¹

Economic Statistics

<http://www.scribd.com/doc/75349300/Economic-Statistics>

Medidas (ou Estimativas)

Distance to the Galactic Center

$$R_0 = 8.0 \pm 0.25 \text{ kpc}$$

Estimativa do
valor real

Estimativa da
Incerteza

Unidade



Malkin, Zinovy.

Statistical analysis of the determinations of the Sun's Galactocentric distance

2013, IAU Symp 289, 406

Algarismos Significativos

Distance to the Galactic Center

$$R_0 = \del{8.0 \pm 0.25 \text{ kpc}}$$

$$8,0 \pm 0,3 \text{ kpc} \quad \text{ou}$$

$$8,00 \pm 0,25 \text{ kpc}$$

3 algarismos
significativos

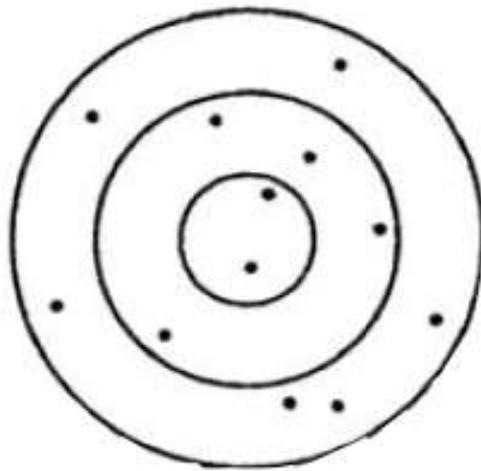
2 algarismos
significativos

Algarismos Significativos

Table 1: Previous average estimates of R_0

Paper	Period covered	R_0 , kpc
Kerr & Lynden-Bell (1986)	1974–1986	8.5 ± 1.1
Reid (1989)	1974–1987	7.7 ± 0.7
Reid (1993)	1974–1992	8.0 ± 0.5
Nikiforov (2004)	1974–2003	7.9 ± 0.2
Avedisova (2005)	1992–2005	7.8 ± 0.32

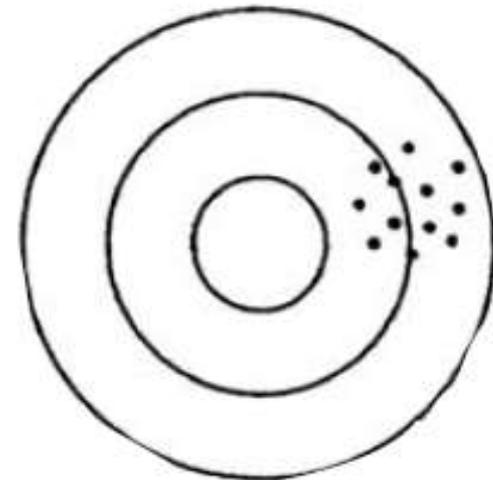
Accuracy and precision



(a)

Precisão: baixa

Exactidão: razoável



(b)

Precisão: elevada

Exactidão: baixa

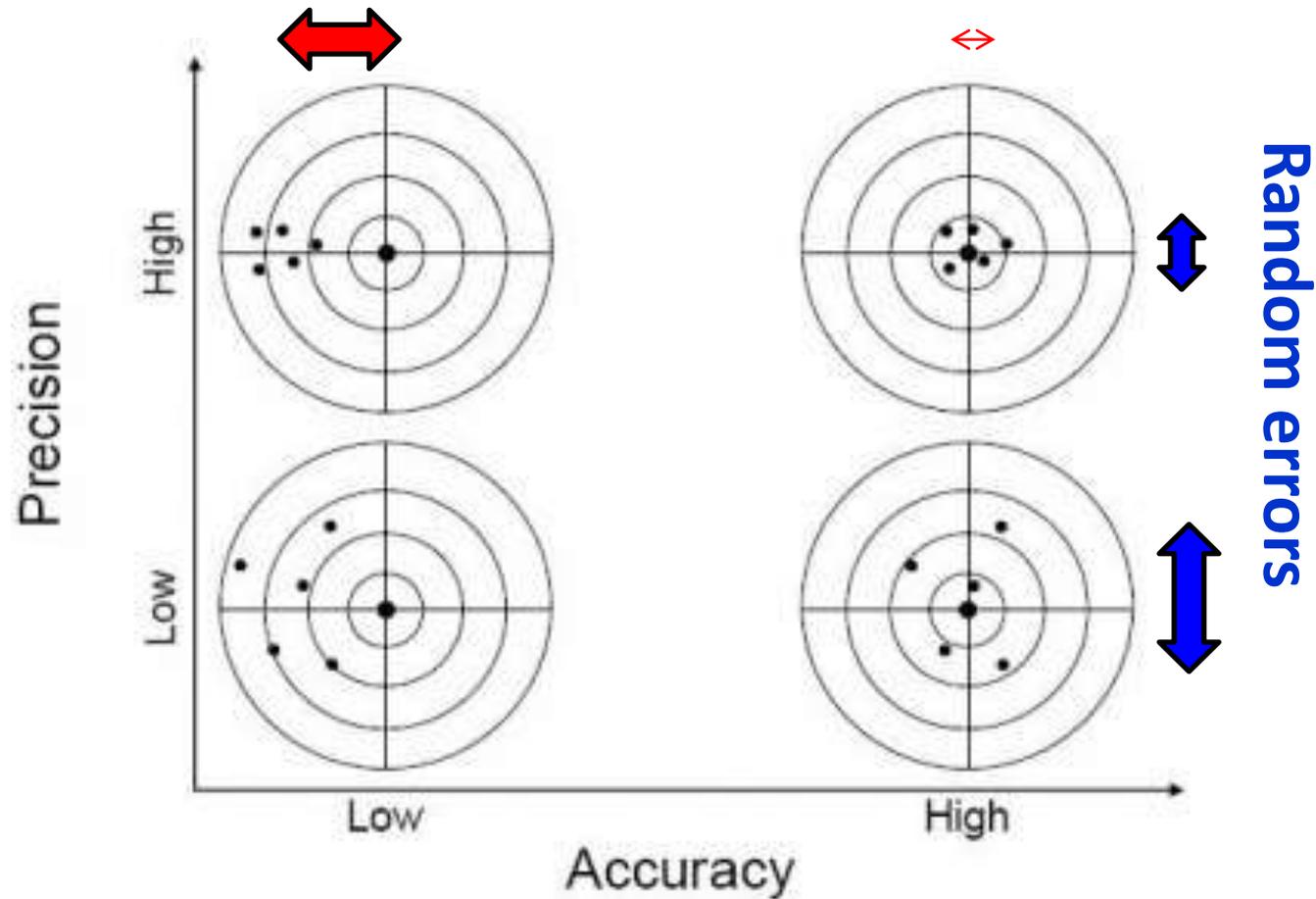
figura 3 Visualização dos conceitos de precisão e exactidão num alvo. Em (a) o conjunto de tiros (resultados) apresenta uma baixa precisão pois apresentam uma dispersão apreciável e uma exactidão razoável visto que não apresentam um desvio sistemático do centro do alvo. Em (b) a precisão é mais elevada (os tiros estão menos dispersos) e a exactidão é mais baixa pois os tiros encontram-se "sistematicamente" afastados para a direita do alvo.

Erros experimentais – uma abordagem pedagógica.

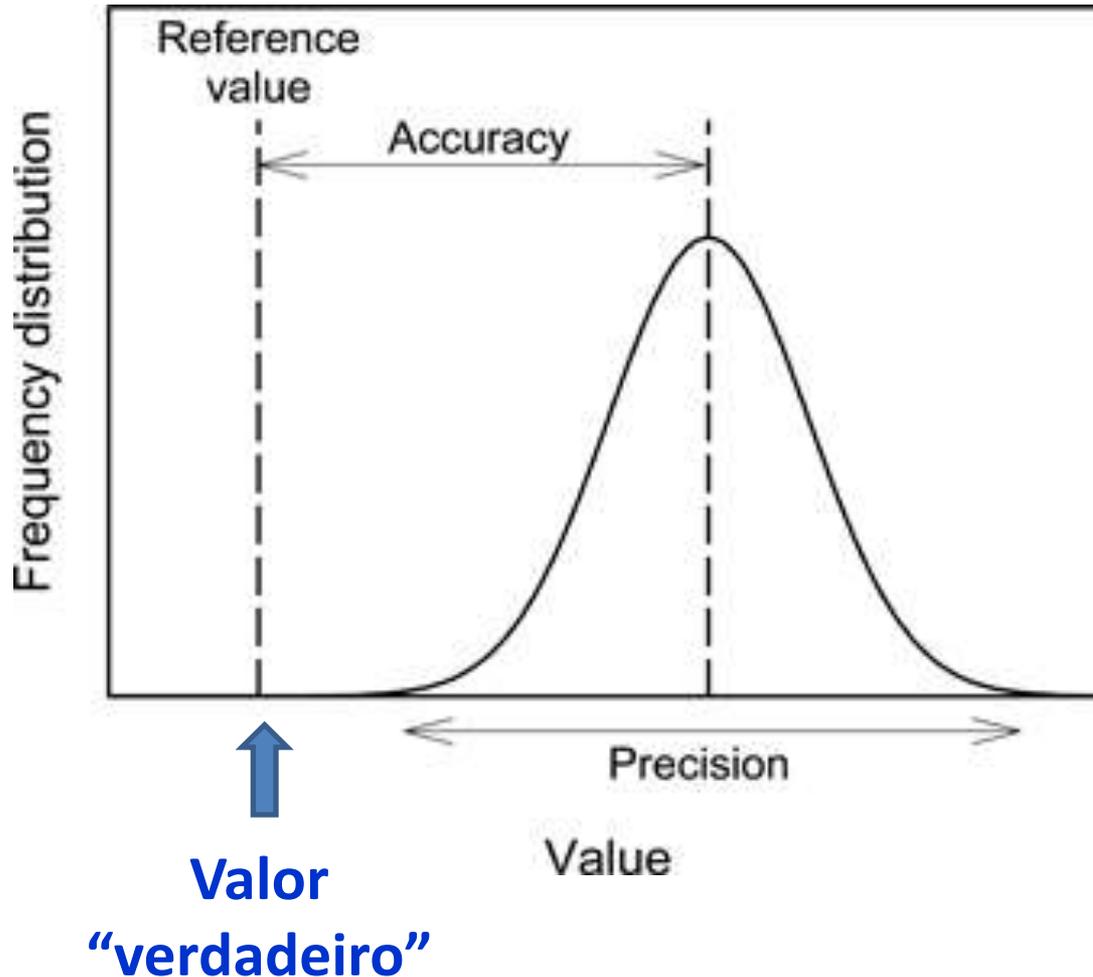
ISABELM.A.FONSECA

Accuracy and precision

Erros sistemáticos



Accuracy and precision



http://dels-old.nas.edu/ilar_n/ilarjournal/49_2/html/v4902Simmons.shtml

Como definir a incerteza (ou erro) ?

- Erro = “Valor Verdadeiro” – Valor medido

- Mas se já conhecemos o valor verdadeiro para que fazer uma medida ???

Definições de erro

- Erro de uma medida x_i (precisão) : $\delta x_i = x_i - \langle x \rangle$
- Erro relativo: $\delta x_i / \langle x \rangle$
- Discrepância (referente à acurácia) = $x_i - x_{\text{true}}$
- Discrepância relativa = $(x_i - x_{\text{true}}) / x_{\text{true}}$
- **Erro sistemático**: tende a desviar o valor medido do valor real, limitando a **acurácia** do resultado.
- **Erro estatístico**: flutuações aleatórias das medidas, limitam a **precisão** do resultado
- **spread = largest result - smallest result**

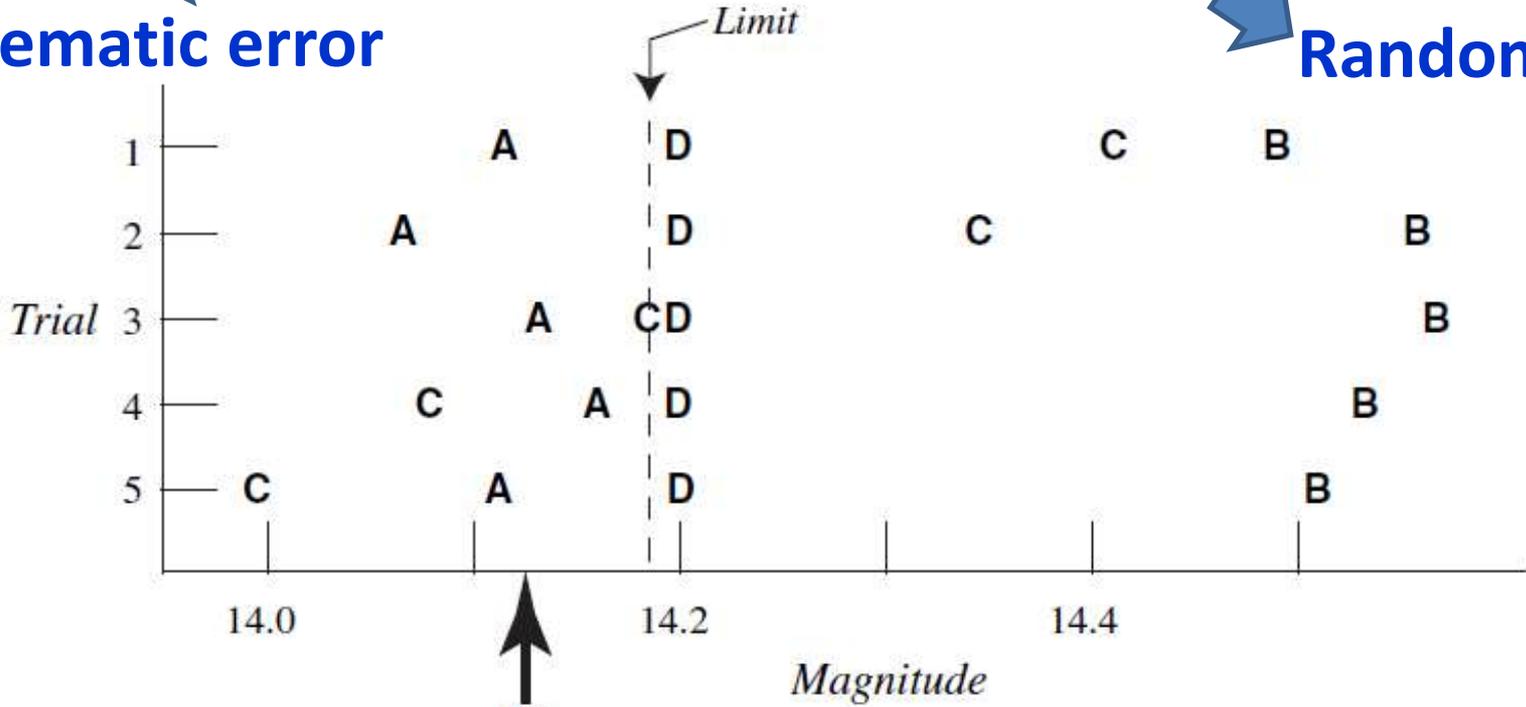
Exemplo

- Valor “verdadeiro” = 100,0 cm
- Medidas : 99,4 99,2 99,5 99,3 99,1 cm
- $\langle x \rangle = 99,3$ cm (média)
- $\delta x_i = 0,1 \quad -0,1 \quad +0,2 \quad 0,0 \quad -0,2$ cm
- $\delta x_i / \langle x \rangle = +0,001 \quad -0,001 \quad +0,002 \quad 0,000 \quad -0,002$
+0,1 % -0,1% +0,2% 0.0% -0,2%
- Discrepância = -0,6 -0,8 -0,5 -0,7 -0,9 cm
- Discrepância relativa = -0,6, -0,8 -0,5 -0,7 -0,9 %
- Erro sistemático = ???

Accuracy and precision

Systematic error

Random error



Astronomer	A	B	C	D
Trial 1	14.115	14.495	14.386	14.2
Trial 2	14.073	14.559	14.322	14.2
Trial 3	14.137	14.566	14.187	14.2
Trial 4	14.161	14.537	14.085	14.2
Trial 5	14.109	14.503	13.970	14.2
Mean	14.119	14.532	14.190	14.2
Deviation from truth	-0.004	+0.409	+0.067	+0.077
Spread	0.088	0.071	0.418	0
σ	0.033	0.032	0.174	0
s	0.029	0.029	0.156	0
Uncertainty of the mean	0.013	0.013	0.070	(0.05)

Valor "real"

Distance to the GC

Random and systematic errors

R_0 determinations with estimation of both statistical and systematic errors

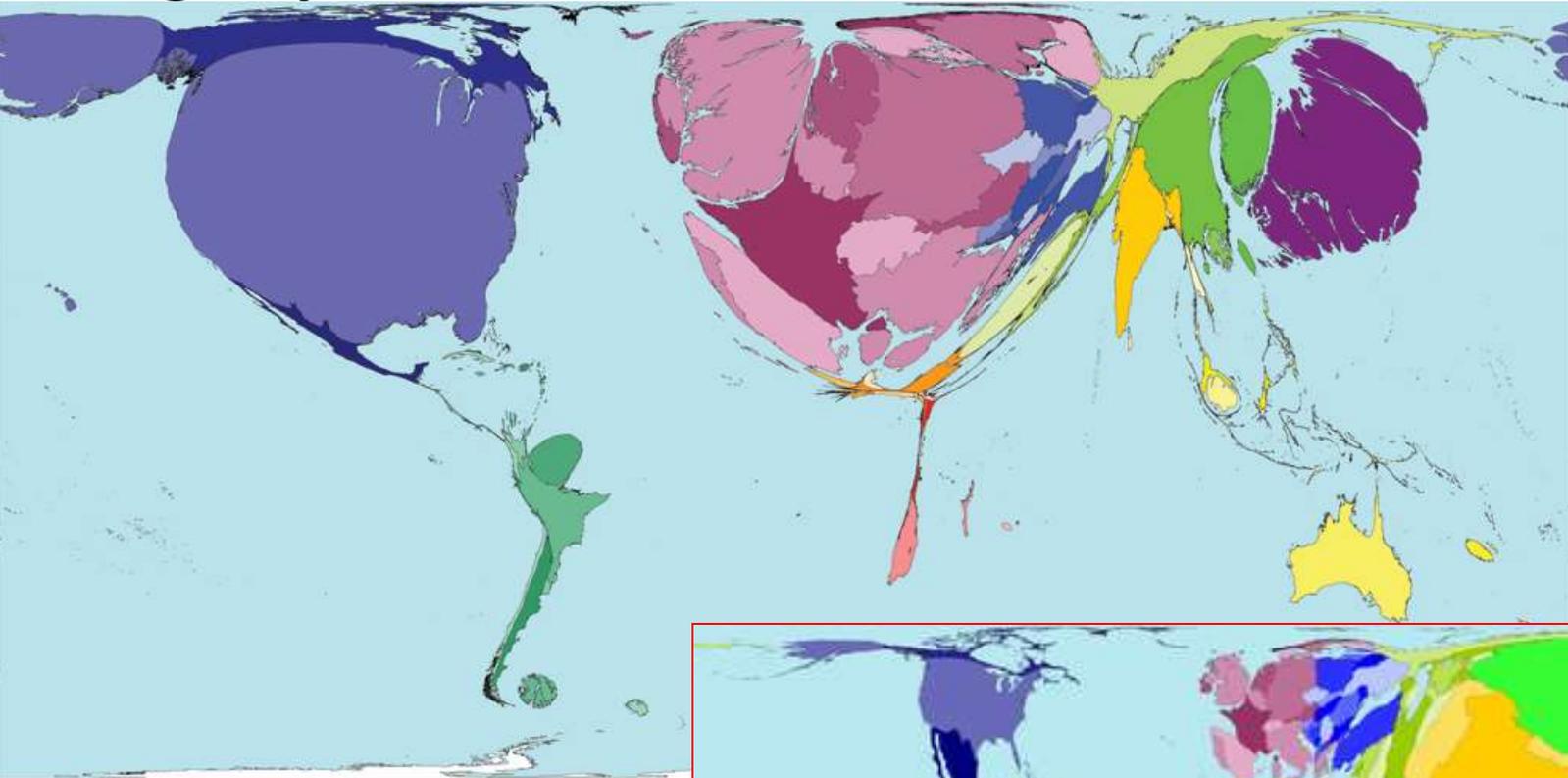
Paper	R_0 , kpc
Nishiyama et al. (2006)	$R_0 = 7.52 \pm 0.10 _{stat} \pm 0.35 _{syst}$
Groenewegen et al. (2008)	$R_0 = 7.94 \pm 0.37 _{stat} \pm 0.26 _{syst}$
Trippe et al. (2008)	$R_0 = 8.07 \pm 0.32 _{stat} \pm 0.13 _{syst}$
Gillessen et al. (2009b)	$R_0 = 8.33 \pm 0.17 _{stat} \pm 0.31 _{syst}$
Gillessen et al. (2009a)	$R_0 = 8.28 \pm 0.15 _{stat} \pm 0.29 _{syst}$
Matsunaga et al. (2009)	$R_0 = 8.24 \pm 0.08 _{stat} \pm 0.42 _{syst}$
Sato et al. (2010)	$R_0 = 8.3 \pm 0.46 _{stat} \pm 1.0 _{syst}$

População e amostra de uma população

- **População** é o conjunto **completo** de medições
- **Amostra** é uma **parte** (representativa ou não) da população
- Populações pequenas podem ser estudadas completamente. *Ex.: estatura dos alunos de física*
- Populações grandes ou indefinidamente grandes apenas podem ser estudadas via amostras. *Ex.: peso de cada pessoa na Terra (7 bilhões !!!)*

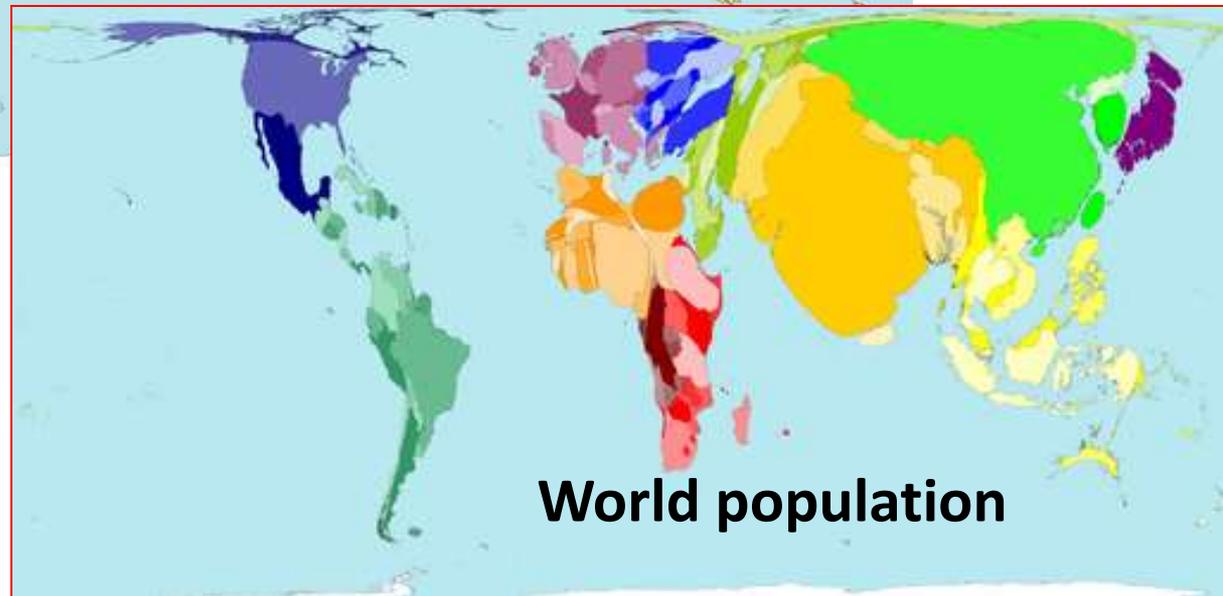
População de artigos/pessoa

Artigos publicados



**Be careful about bias
in your sample !!!**

<http://www.worldmapper.org/>



World population

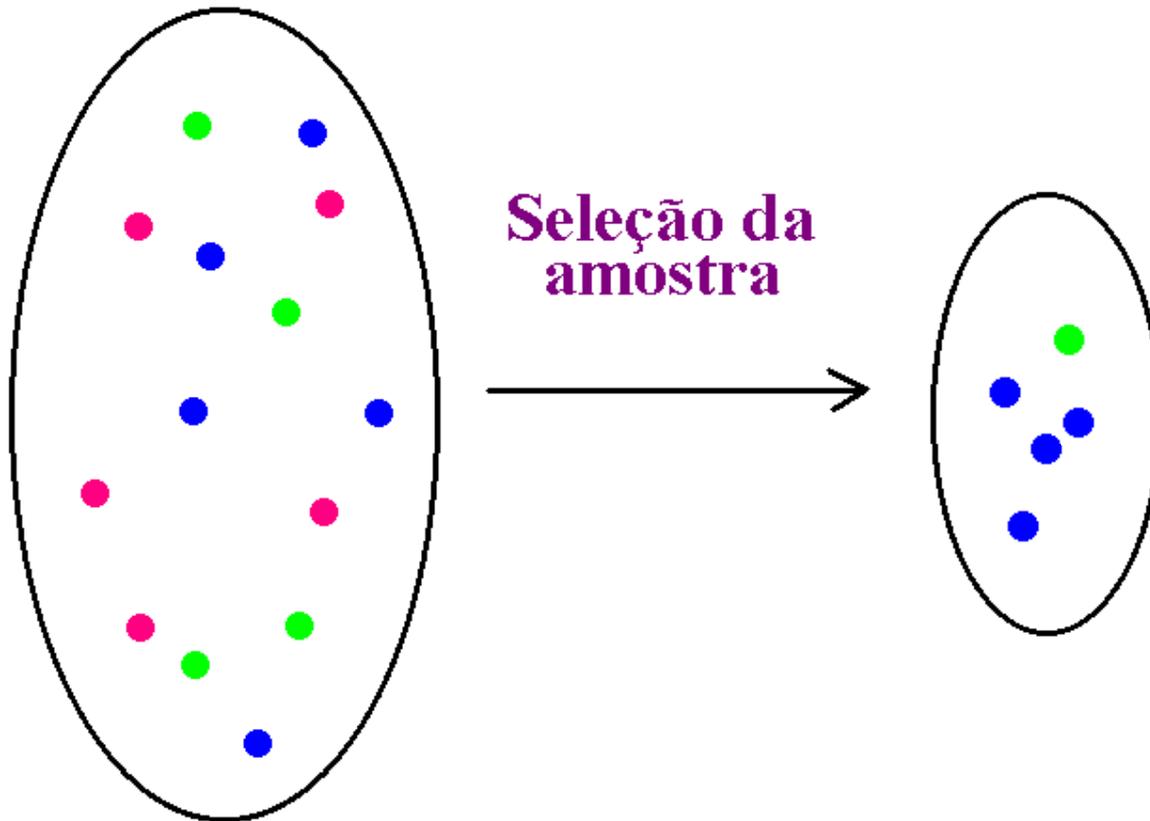
População de estaturas (adultos) na Terra

Como selecionar uma amostra representativa (sem *bias*) ?

Country	Average male height	Average female height	Ratio (male to female)	Sample population / age range	Share of pop. over 15 covered ^[3]	Methodology	Year
Argentina	1.73 m	1.60 m	1.08	17 (healthy)	N/A	Measured	2000
Bolivia – Aymara	1.600 m	1.422 m	1.13	20–29	N/A	Measured	1970
Brazil	1.71 m	1.59 m	1.07	18+	93.2%	Measured	2008–2009
China	1.66 m	1.57 m	1.06	Rural, 17	N/A	Measured	2002
China	1.70 m	1.59 m	1.07	Urban, 17	N/A	Measured	2002
Germany	1.81 m	1.68 m	1.08	18–25	N/A	Self-reported	2009
Germany	1.78 m	1.65 m	1.08	18+	96.5%	Self-reported	2009
Netherlands	1.83 m	1.70 m	1.08	20–30	N/A	Self-reported	2010
Peru	1.64 m	1.510 m	1.09	20+	85.4%	Measured	2005 ¹⁶

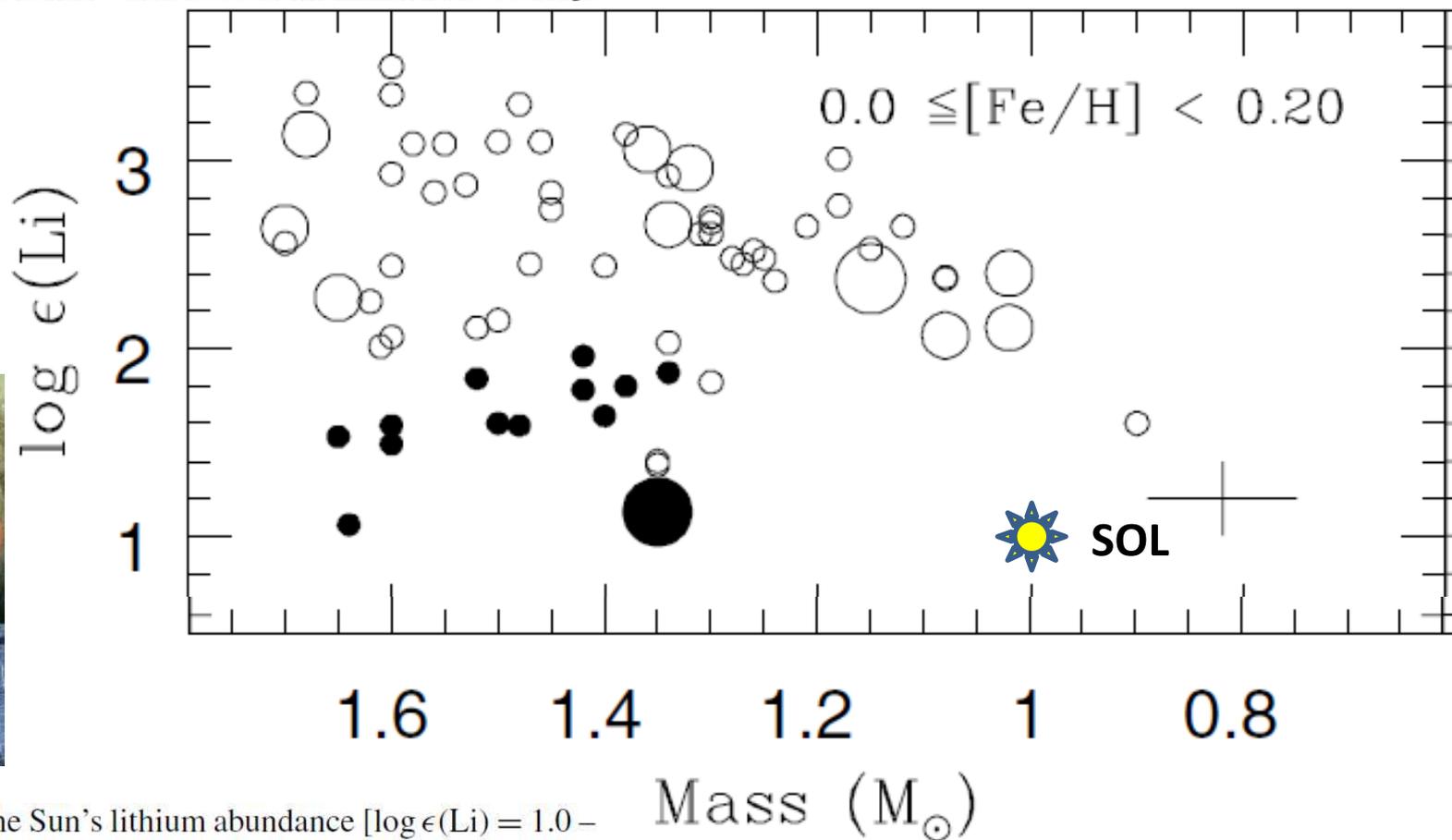
Bias (vies) da amostragem ou *vies de seleção, vies de cobertura, etc*

A amostra é representativa da população?
Se houver bias → conclusões erradas



Lithium abundances of the local thin disc stars

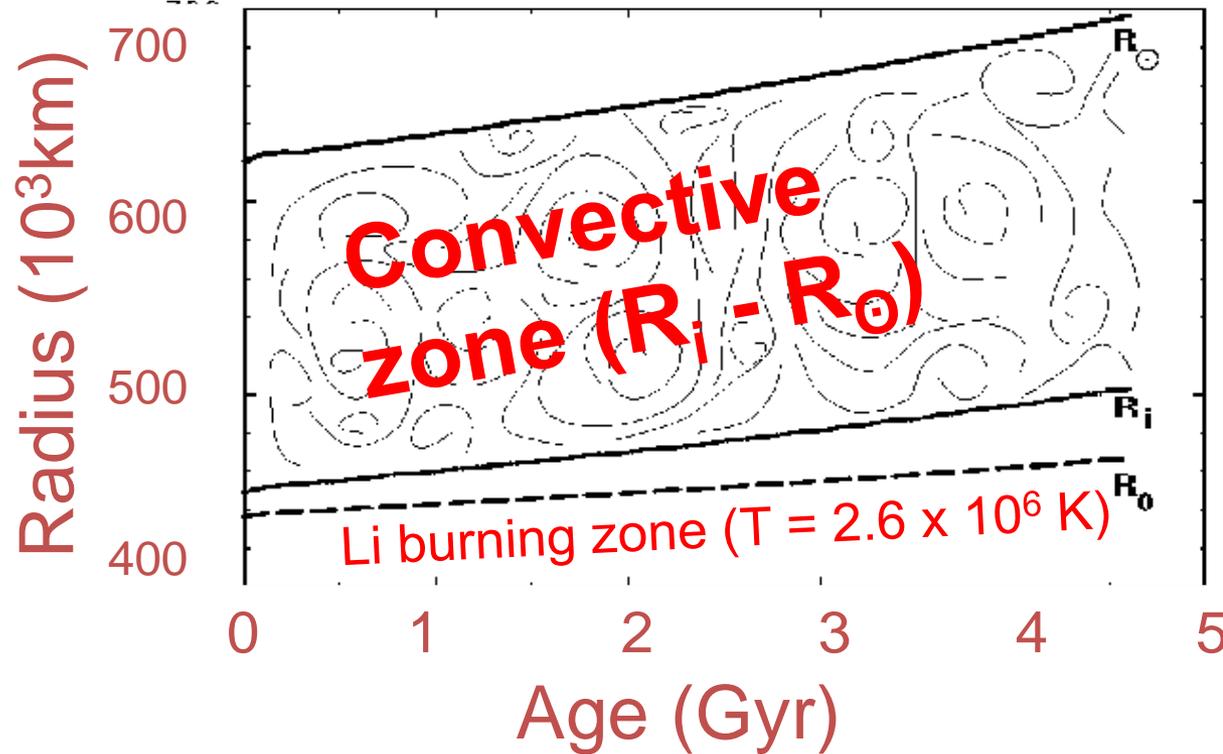
David L. Lambert¹ and Bacham E. Reddy^{1,2★}



A curiosity is that the Sun's lithium abundance [$\log \epsilon(\text{Li}) = 1.0$ – Müller, Peytremann & de la Reza 1975] appears to fall by more than 1 dex below the trend defined by the field stars (see Fig. 3). If placed among NGC 188's stars, the Sun would be deemed very Li-poor. Among M67's stars, the Sun would be one of the most Li-poor stars. This hint that the Sun may be 'peculiar' as regards the depletion of lithium weakens its value as a calibrator for prescriptions of non-standard modes of lithium astration.

Estudaram 450 anãs de tipo F e G em diferentes faixas de massa e metalicidade

Is the solar Li abundance peculiar ?



Astrophys Space Sci (2010) 328: 193–200

The solar, exoplanet and cosmological lithium problems

J. Meléndez · I. Ramírez · L. Casagrande · M. Asplund ·

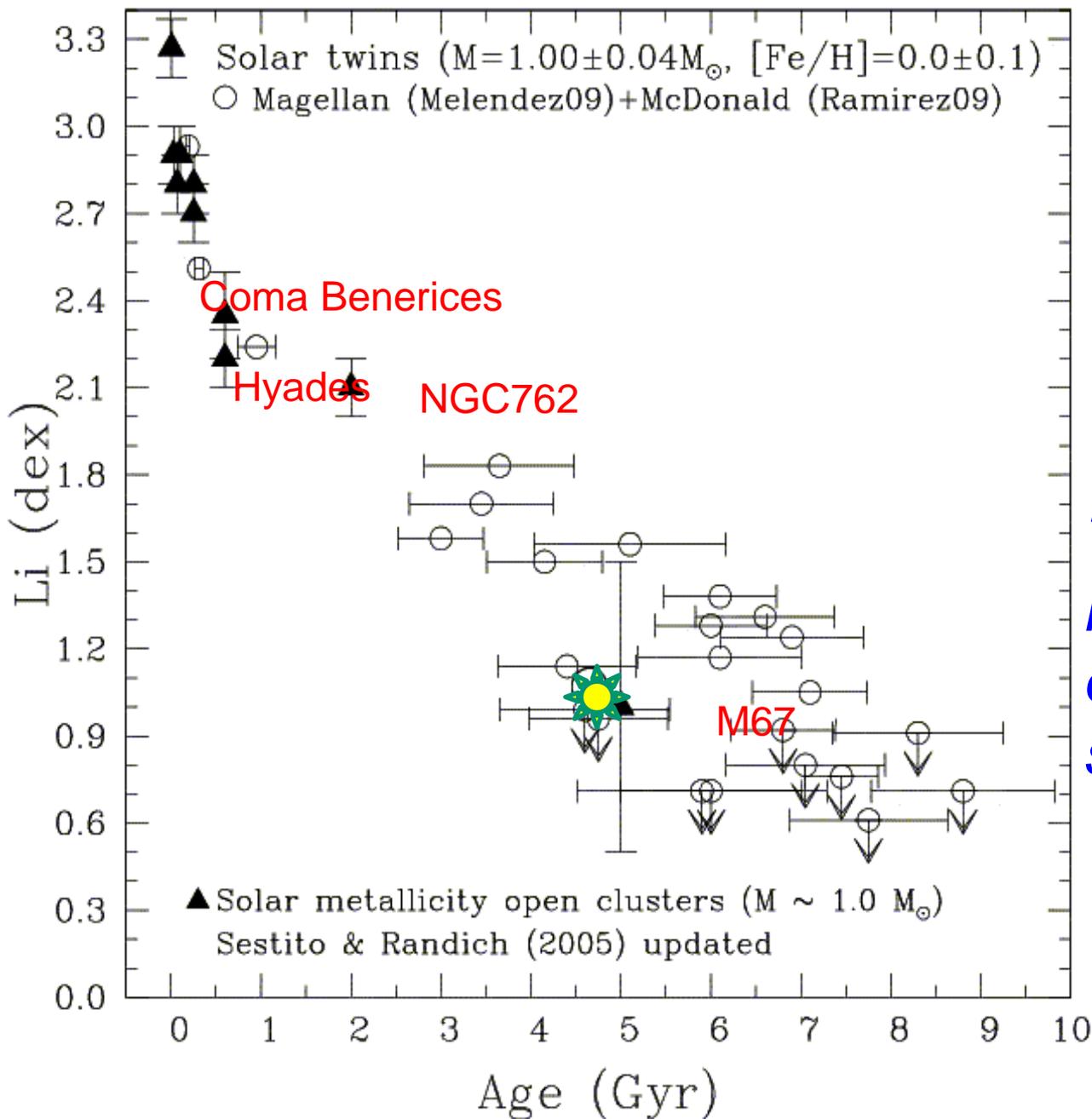
B. Gustafsson · D. Yong · J.D. do Nascimento Jr. ·

M. Castro · M. Bazot

A&A 519, A87 (2010)

Lithium depletion in solar-like stars: no planet connection

P. Baumann¹, I. Ramírez¹, J. Meléndez², M. Asplund¹, and K. Lind³



— Solar twins
 — in open
 — cluster and
 — field stars

*The Sun is normal in
 lithium compared to
 others 1-solar-mass
 stars at 4.6 Gyr*

(Melendez et al. 2010;
 Baumann et al. 2010)

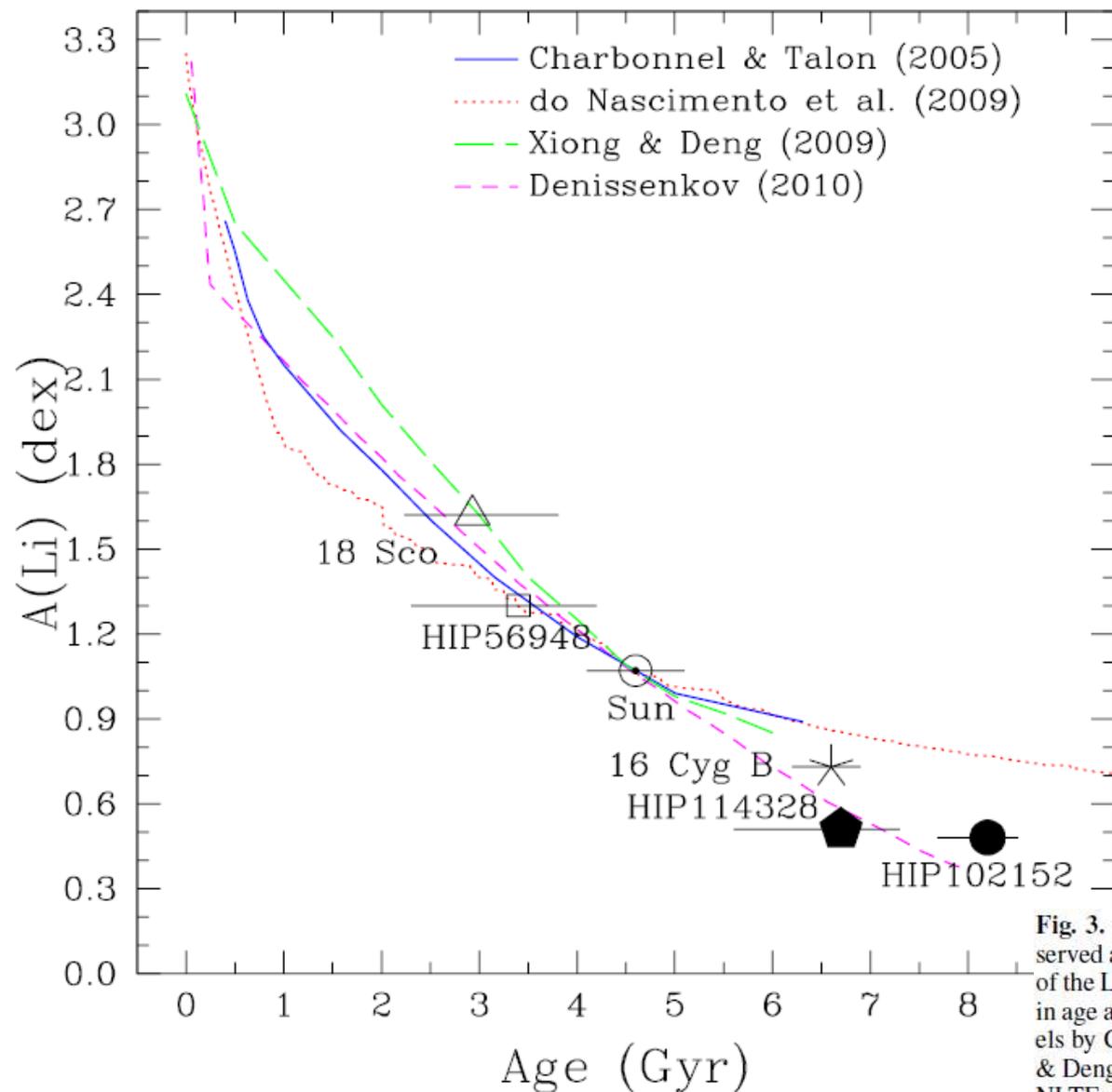
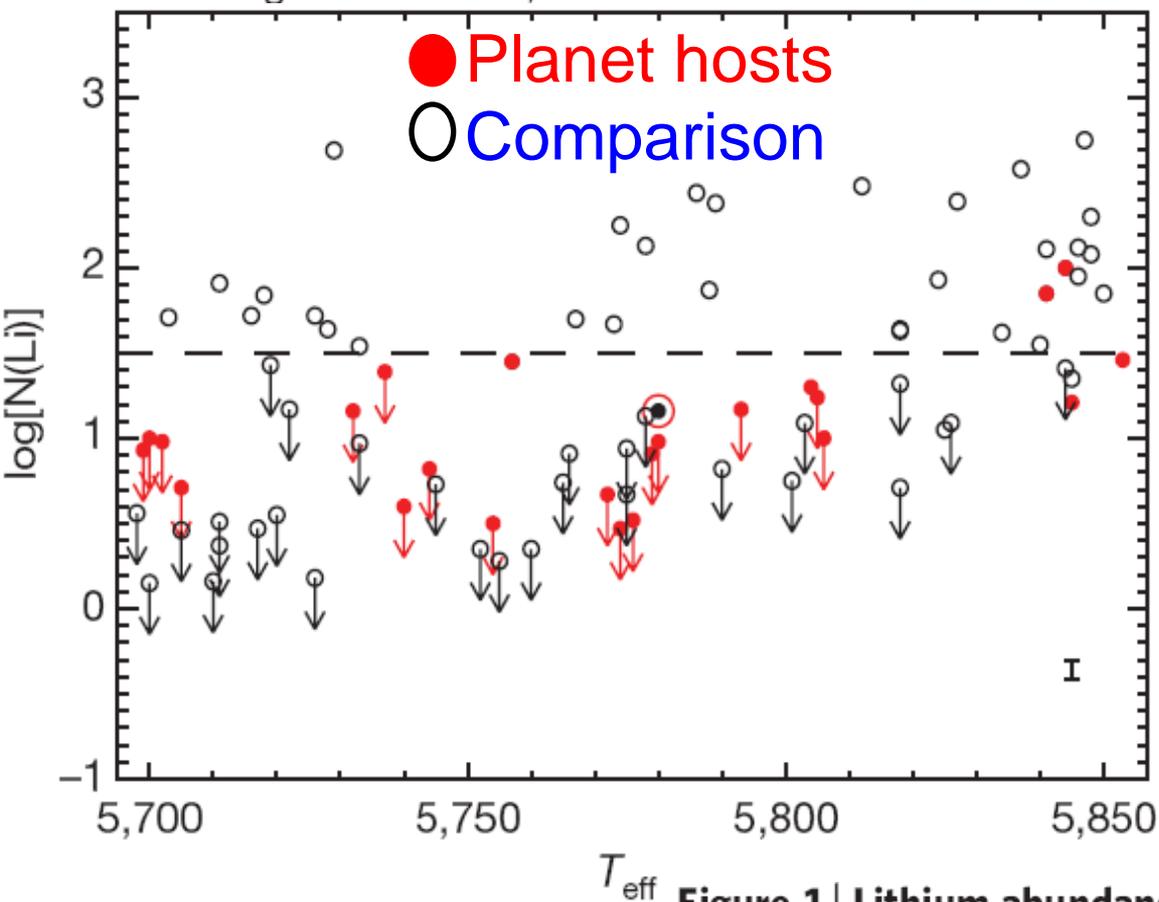


Fig. 3. NLTE Li abundances vs. age for the Sun and solar twins observed at high spectral resolution and high S/N. The total error bar ($\pm\sigma$) of the Li abundance is about the size of the symbols, while the error bars in age are shown by horizontal lines. For comparison we show the models by Charbonnel & Talon (2005); do Nascimento et al. (2009); Xiong & Deng (2009); Denissenkov (2010), shifted to reproduce our observed NLTE solar Li abundance. The model with initial rotation velocity of 50 km s^{-1} was adopted for Charbonnel & Talon (2005). HIP 114328 is shown by a pentagon, and helps to define a clear Li-age correlation.

HIP 114328: a new refractory-poor and Li-poor solar twin [★]

Enhanced lithium depletion in Sun-like stars with orbiting planets

Garik Israelian^{1,2}, Elisa Delgado Mena^{1,2}, Nuno C. Santos^{3,4}, Sergio G. Sousa^{1,3}, Michel Mayor⁴, Stephane Udry⁴, Carolina Domínguez Cerdeña^{1,2}, Rafael Rebolo^{1,2,5} & Sofia Randich⁶

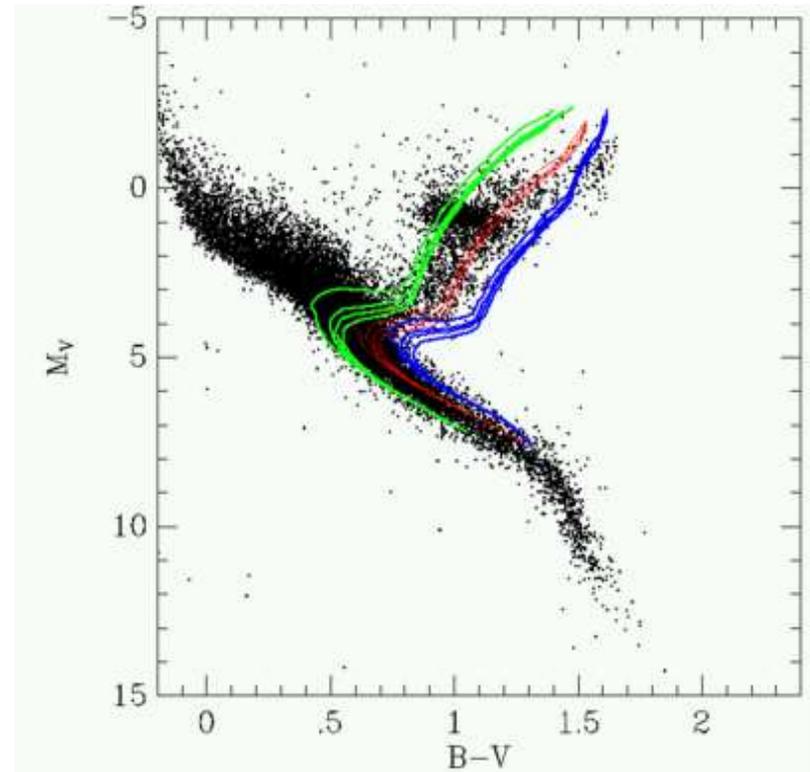
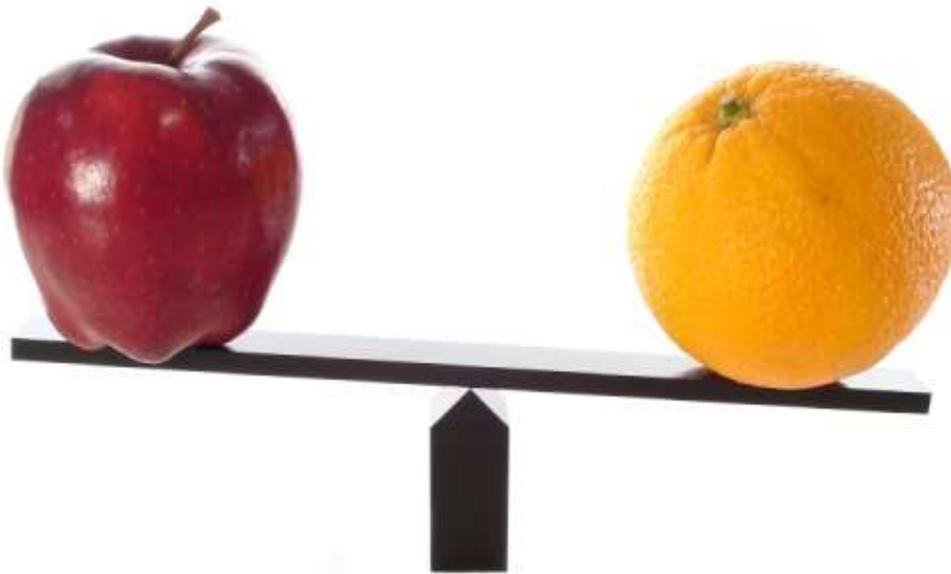


Planet-host stars around solar T_{eff} seem depleted in Li

Figure 1 | Lithium abundance plotted against effective temperature in solar-analogue stars with and without detected planets. The planet-

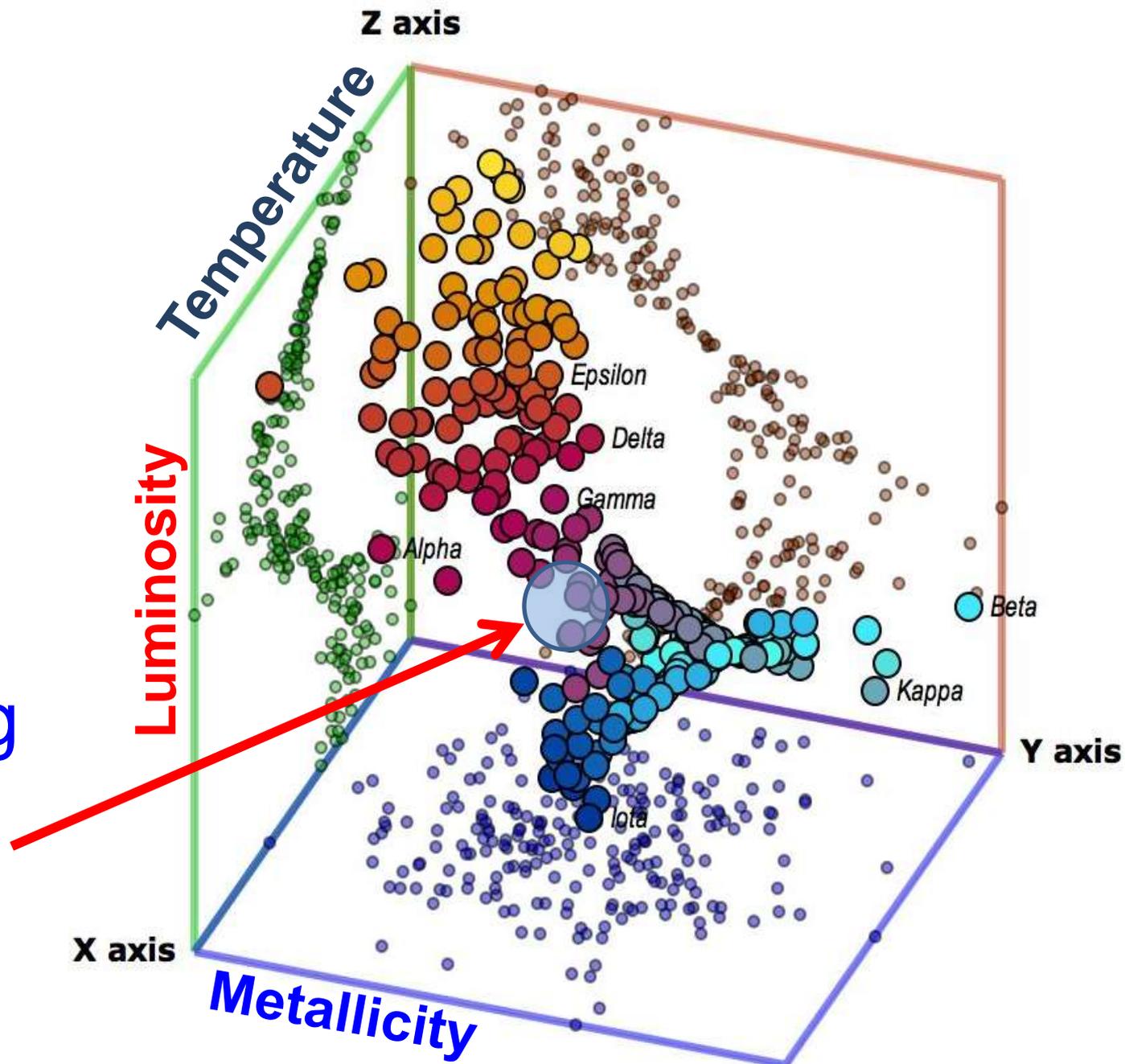
You cannot compare apples and oranges ...

*comparer des pommes avec des oranges
comparer des pommes et des poires*

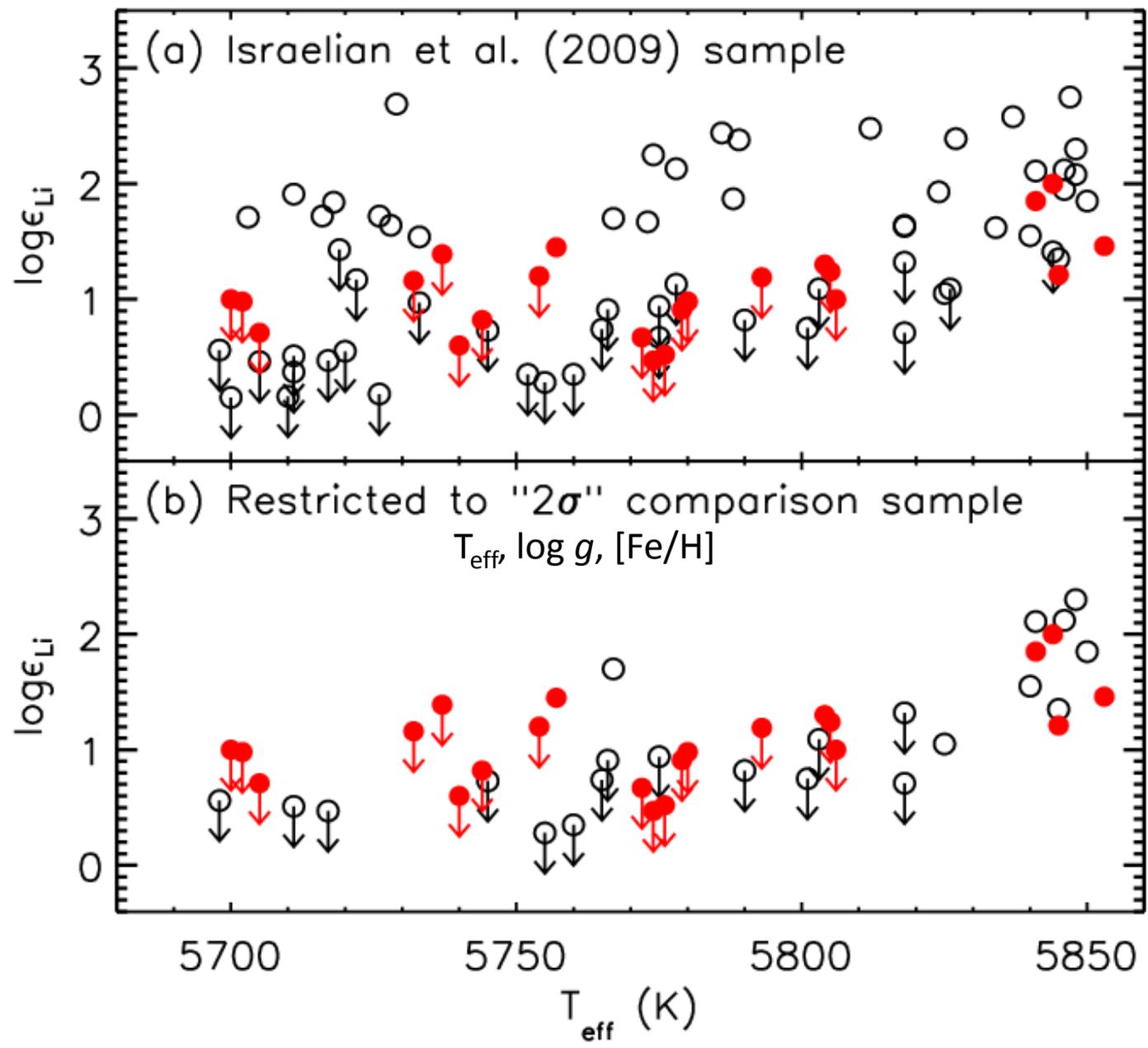


*You cannot add pears and apples ...
No puedes sumar peras con manzanas*

Comparing
apples &
apples



Li depletion is not enhanced in planet hosts !



Apples & oranges

Israelian et al. 2009, Nature

● Planet hosts
○ Comparison

Apples & apples

Baumann, Ramírez, Meléndez, & Asplund 2010, A&A, 519, A87

*Conclusion on lithium in stars
with and without planets:
(sem viés na comparação)*

there is no difference in Li
abundance between stars
with and without planets

Exemplo de populações e amostras

Population	Sample	Better sample
1000 colored marbles mixed in a container: 500 red, 499 blue, 1 purple	5 marbles drawn at random from the container	50 marbles drawn at random
The luminosities of each star in the Milky Way galaxy (about 10^{11} values)	The luminosities of each of the nearest 100 stars (100 values)	The luminosities of 100 stars at random locations in the galaxy (100 values)
The weights of every person on Earth	The weights of each person in this room	The weights of 100 people drawn from random locations on Earth

Idade de cada estrela da Galáxia

- **Aglomerado aberto ?**
 - **Aglomerado globular ?**
 - **Braços espirais ?**
 - **Halo ?**
 - **Bojo ?**
- 20 estrelas em 50 lugares “aleatórios” (na verdade, representativos) na Galáxia (1000 valores)**

Valor Central e Desvio Padrão de uma População x_i de M elementos (no total)

Valor central :

Média :
$$\mu = \frac{1}{M} \sum_{i=1}^M x_i$$

Mediana : value that divides the population exactly in half

Mode : is the value that occurs most often.

Desvio padrão :

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_i - \mu)^2}$$

Valor Central e Desvio Padrão **estimados** de uma população usando uma **amostra de N elementos**

Média $\langle x \rangle$:
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Desvio Padrão Estimado:
$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Usamos $N - 1$ ao invés de N pois $\langle x \rangle$ é apenas um estimado de μ .

Exemplo, população de estaturas de 5 meninas:

População : 149, 151, 153, 152, 169 cm. $\mu = 154,8\text{cm}$. $\sigma = 7,22\text{cm}$

Sample : 151, 152 cm. $\langle x \rangle = 151,5\text{ cm}$, $\sigma = 0,5\text{cm}$, $s = 0,70\text{cm}$

Sample : 149, 169 cm. $\langle x \rangle = 159\text{ cm}$, $\sigma = 10\text{cm}$, $s = 14\text{cm}$

Sample : 149,151,169 cm. $\langle x \rangle = 156,3\text{ cm}$, $\sigma = 9\text{cm}$, $s = 11\text{cm}$

Desvio padrão da média de uma amostra de n elementos σ_n

$$\sigma_{\mu}(n) = \frac{s}{\sqrt{n}}$$

O desvio padrão da média é também chamado de *standard error*

Exemplo, população de estaturas de 5 meninas:

População : 149, 151, 153, 152, 169 cm. $\mu = 154,8$ cm. $\sigma = 7,22$ cm

$\sigma_n = 3,2$ cm. $\mu = 154,8 \pm 3,2$ cm

Sample : 151, 152 cm. $\langle x \rangle = 151,5$ cm, $\sigma = 0,5$ cm , $s = 0,70$ cm,

$\sigma_n = 0,5$ cm $\langle x \rangle = 151,5 \pm 0,5$ cm

Sample : 149, 169 cm. $\langle x \rangle = 159$ cm, $\sigma = 10$ cm , $s = 14$ cm ,

$\sigma_n = 10$ cm $\langle x \rangle = 159 \pm 10$ cm

Sample : 149,151,169 cm. $\langle x \rangle = 156,3$ cm, $\sigma = 9$ cm , $s = 11$ cm ,

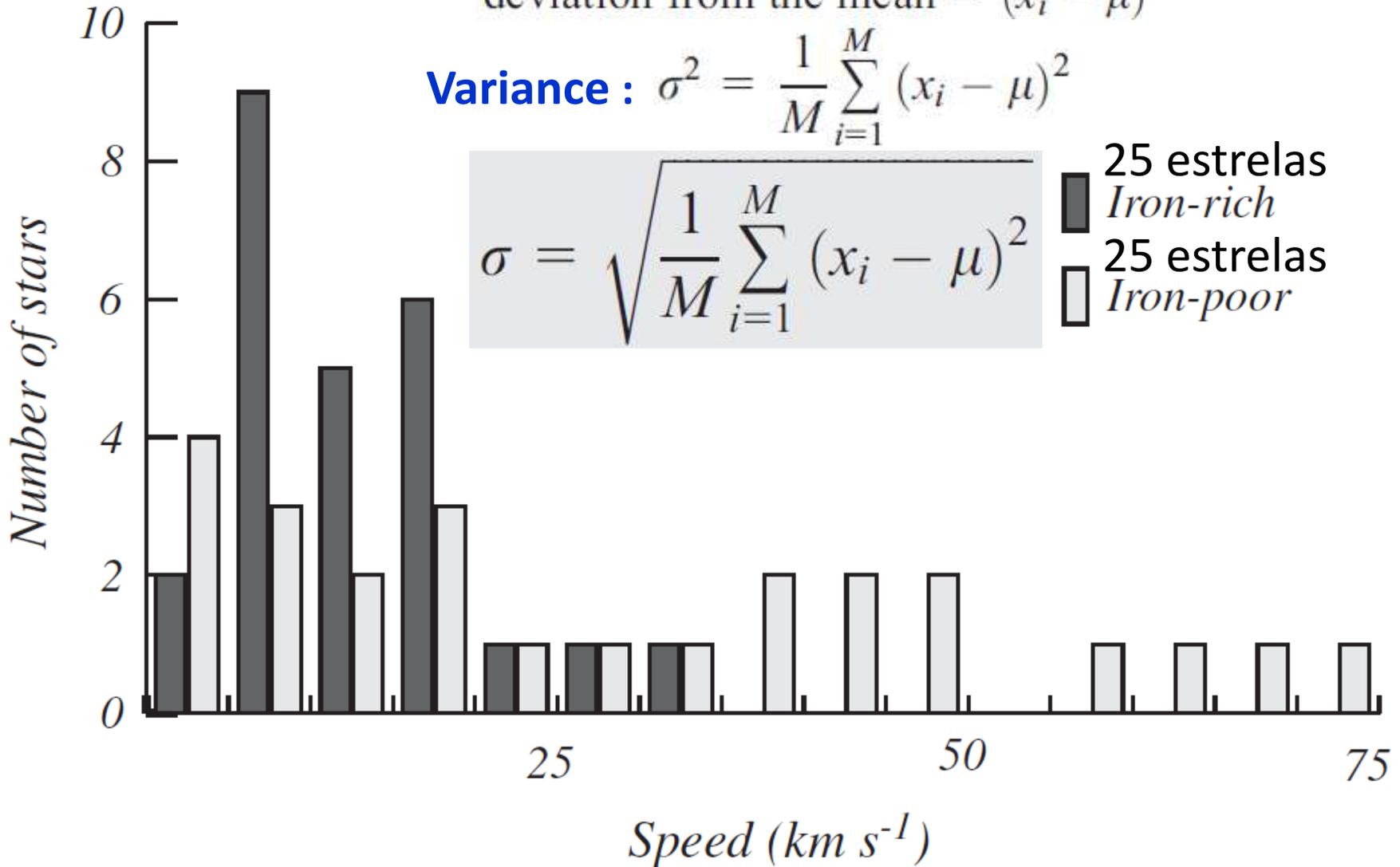
$\sigma_n = 6,3$ cm $\langle x \rangle = 156,3 \pm 6,3$ cm

Populações estelares em nossa Galáxia

deviation from the mean = $(x_i - \mu)$

Variance : $\sigma^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \mu)^2$

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_i - \mu)^2}$$



GROUP A: Variance = $57.25 \text{ km}^2 / \text{s}^2$, $\sigma = 7.57 \text{ km/s}$. Mean = 12.85 km/s

Chemical evolution of the Galactic bulge as traced by microlensed dwarf and subgiant stars^{★,★★}

IV. Two bulge populations

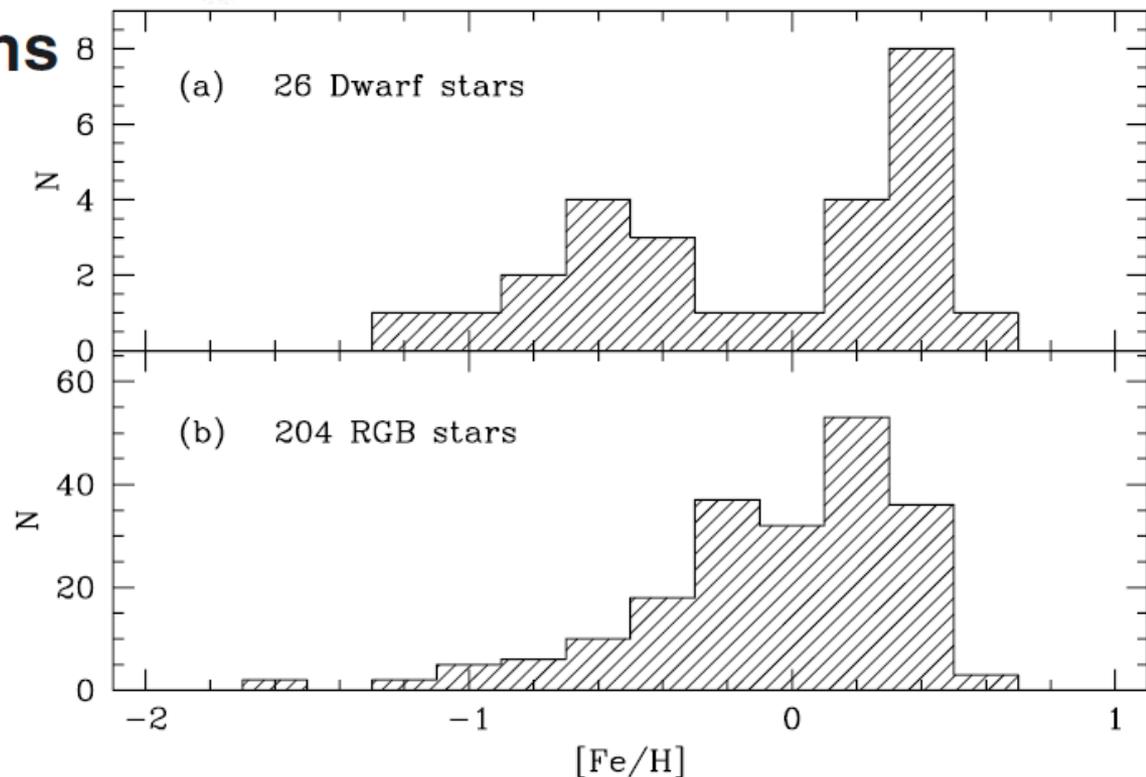


Fig. 6. **a)** MDF for our 26 microlensed dwarf stars in the bulge. **b)** MDF for 204 RGB stars in the bulge, from [Zoccali et al. \(2008\)](#). **c)** The significance for the failure of reproduce a gap vs. position of a gap in the dwarf star MDF (see explanation in Sect. 4.1).

T. Bensby^{1,2}, D. Adén¹, J. Meléndez³, A. Gould⁴, S. Feltzing¹, M. Asplund⁵, J. A. Johnson^{4,★★★}, S. Lucatello⁶, J. C. Yee⁴, I. Ramírez⁷, J. G. Cohen⁸, I. Thompson⁷, I. A. Bond⁹, A. Gal-Yam¹⁰, C. Han¹¹, T. Sumi¹², D. Suzuki¹², K. Wada¹², N. Miyake¹³, K. Furusawa¹³, K. Ohmori¹³, To. Saito¹⁴, P. Tristram¹⁵, and D. Bennett¹⁶

A&A 533, A134 (2011)

T. Bensby¹, J. C. Yee², S. Feltzing¹, J. A. Johnson², A. Gould², J. G. Cohen³, M. Asplund⁴, J. Meléndez⁵, S. Lucatello⁶, C. Han⁷, I. Thompson⁸, A. Gal-Yam⁹, A. Udalski¹⁰, D. P. Bennett¹¹, I. A. Bond¹², W. Kohei¹³, T. Sumi¹³, D. Suzuki¹³, K. Suzuki¹⁴, S. Takino¹⁴, P. Tristram¹⁵, N. Yamai¹⁶, and A. Yonehara¹⁶

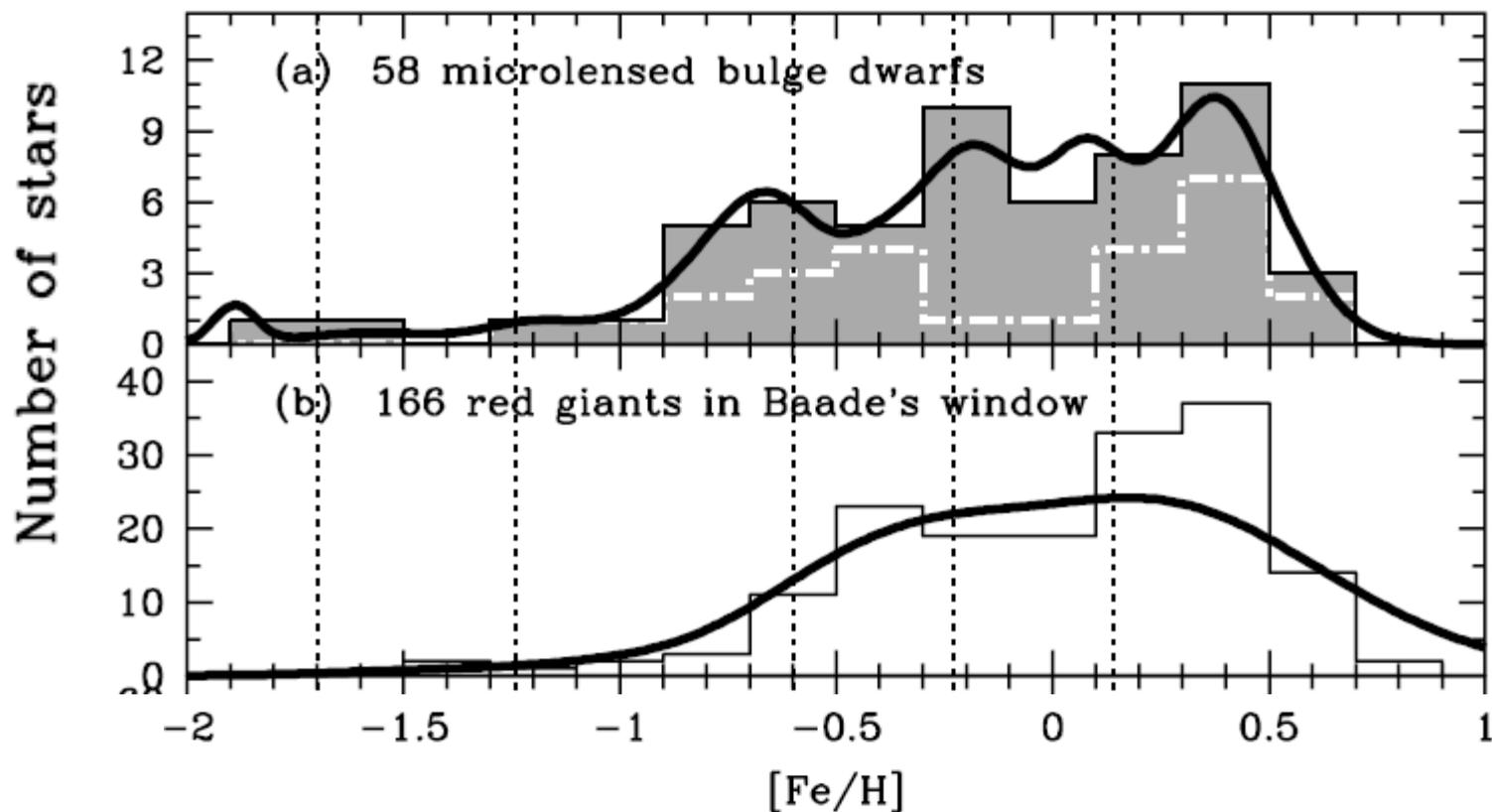


Fig. 12. a) Metallicity distribution for the microlensed dwarf sample (white dashed line shows the distribution of the 26 microlensed dwarf stars from [Bensby et al. 2011](#)); b) 166 red giant stars in Baade's window from [Hill et al. \(2011\)](#); c) 220 red clump stars in Baade's window from

Valor central **estimado (amostra)**
e valor central “verdadeiro” (μ) de
uma população

$$\mu = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i = \lim_{N \rightarrow \infty} \bar{x}$$

$$\mu \approx \bar{x}$$

Variança e desvio padrão **estimados** e “verdadeiros” de uma população são similares para $N \gg 1$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$\sigma^2 \approx s^2$$

$$\begin{aligned} \sigma^2 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \lim_{N \rightarrow \infty} \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \\ &= \lim_{N \rightarrow \infty} s^2 \end{aligned}$$

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \approx \sigma$$

Média ponderada

$$y_c = \sigma_c^2 \sum_{i=1}^n (y_i / \sigma_i^2)$$

$$1/\sigma_c^2 = \sum_{i=1}^n (1/\sigma_i^2)$$

- Exemplo:
- $y_1 = 18 \pm 3$ cm, $y_2 = 16 \pm 4$ cm
- $y_c = (3) * 144/25 = 17,3$ cm $\pm 2,4$ cm ?
- $1/s_c^2 = (1/16) + (1/9) = 25/144$, $s_c^2 = 144/25$

Distribuição das medidas

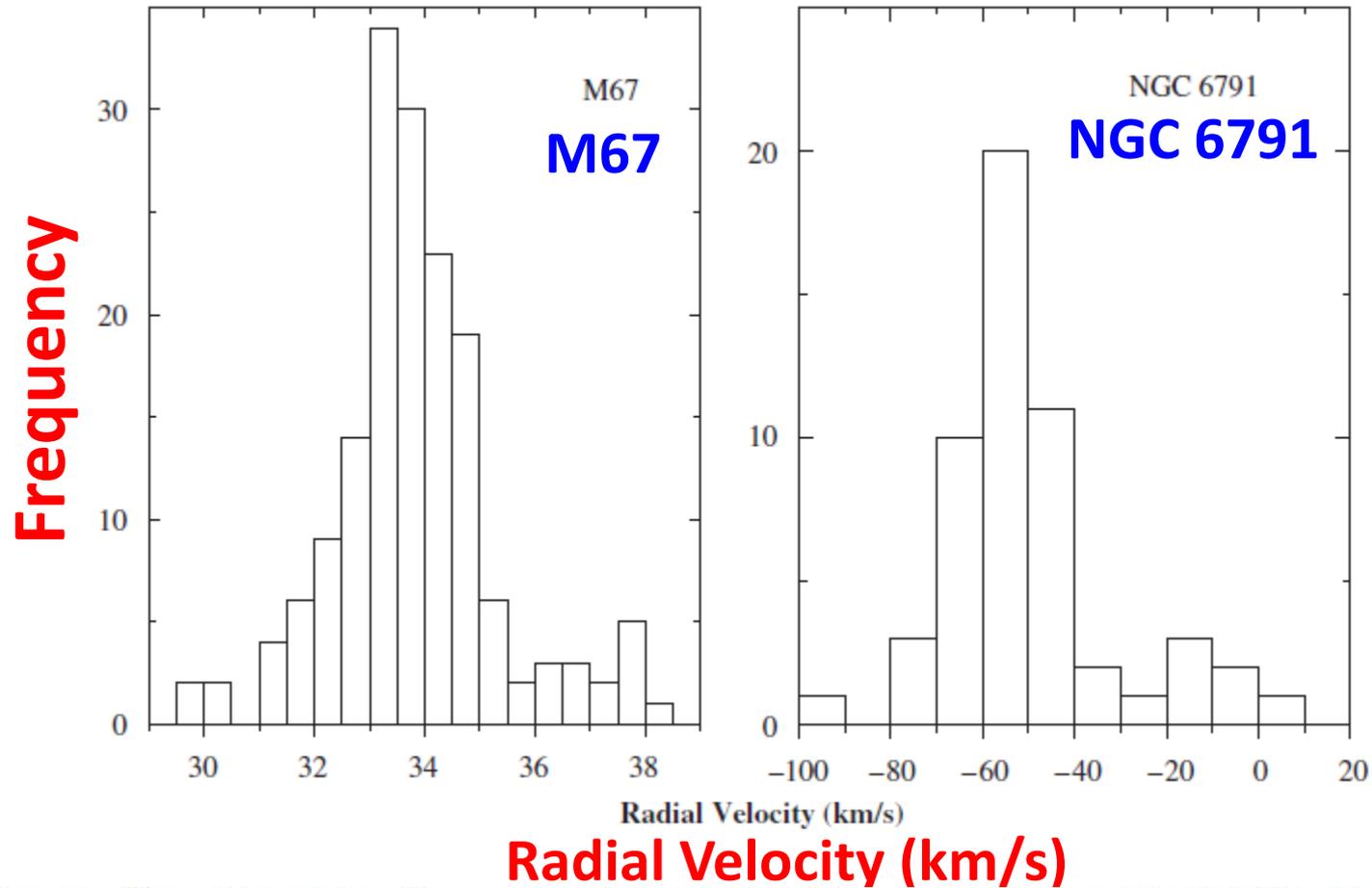


Fig. 2. Radial velocity histogram of four stellar clusters after removing stars according to the second selection method, that is, star velocity difference to the central peak of the distribution larger than 3σ .

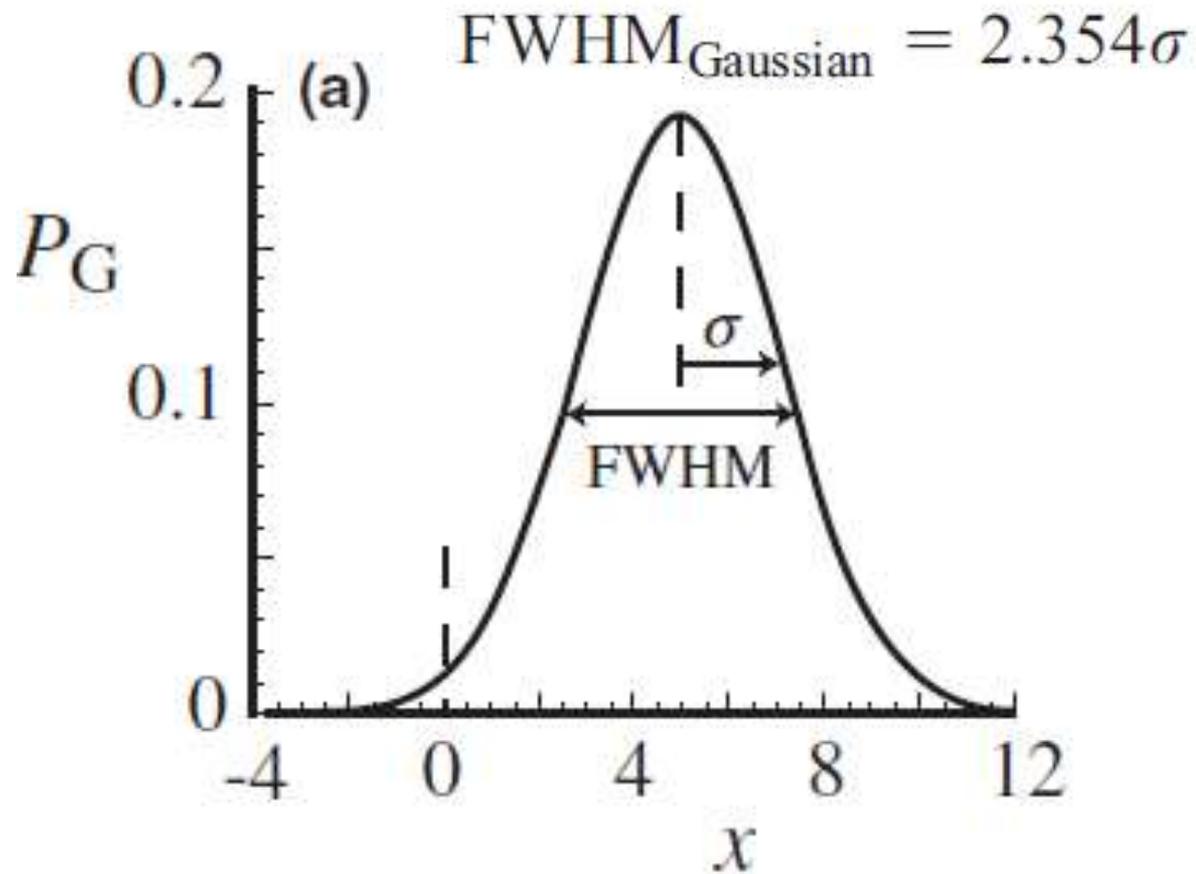
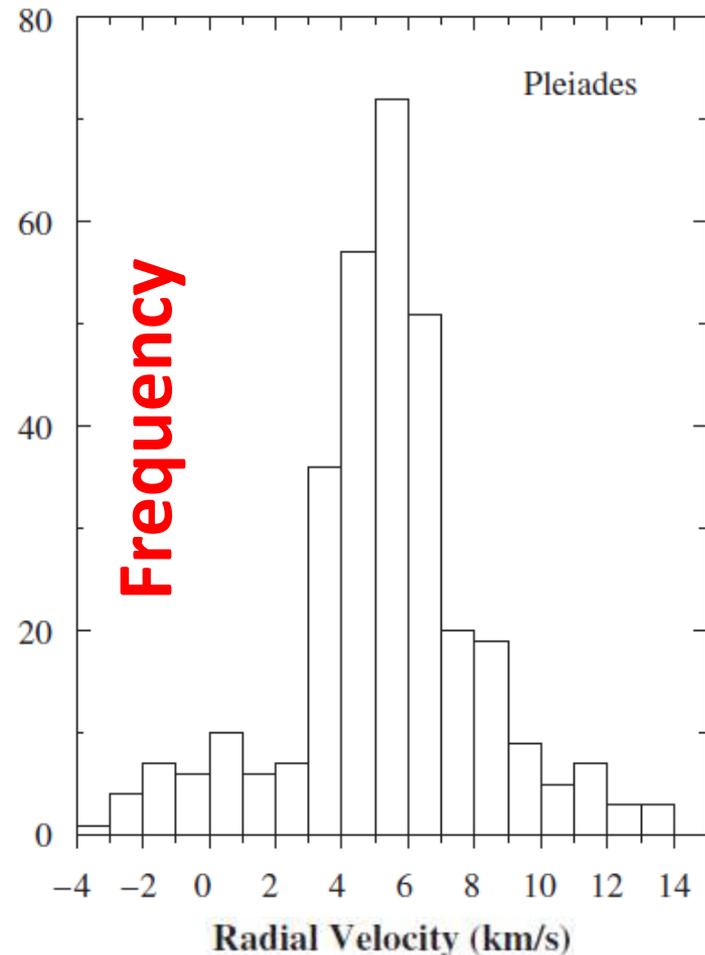
Physica A 384 (2007) 507–515

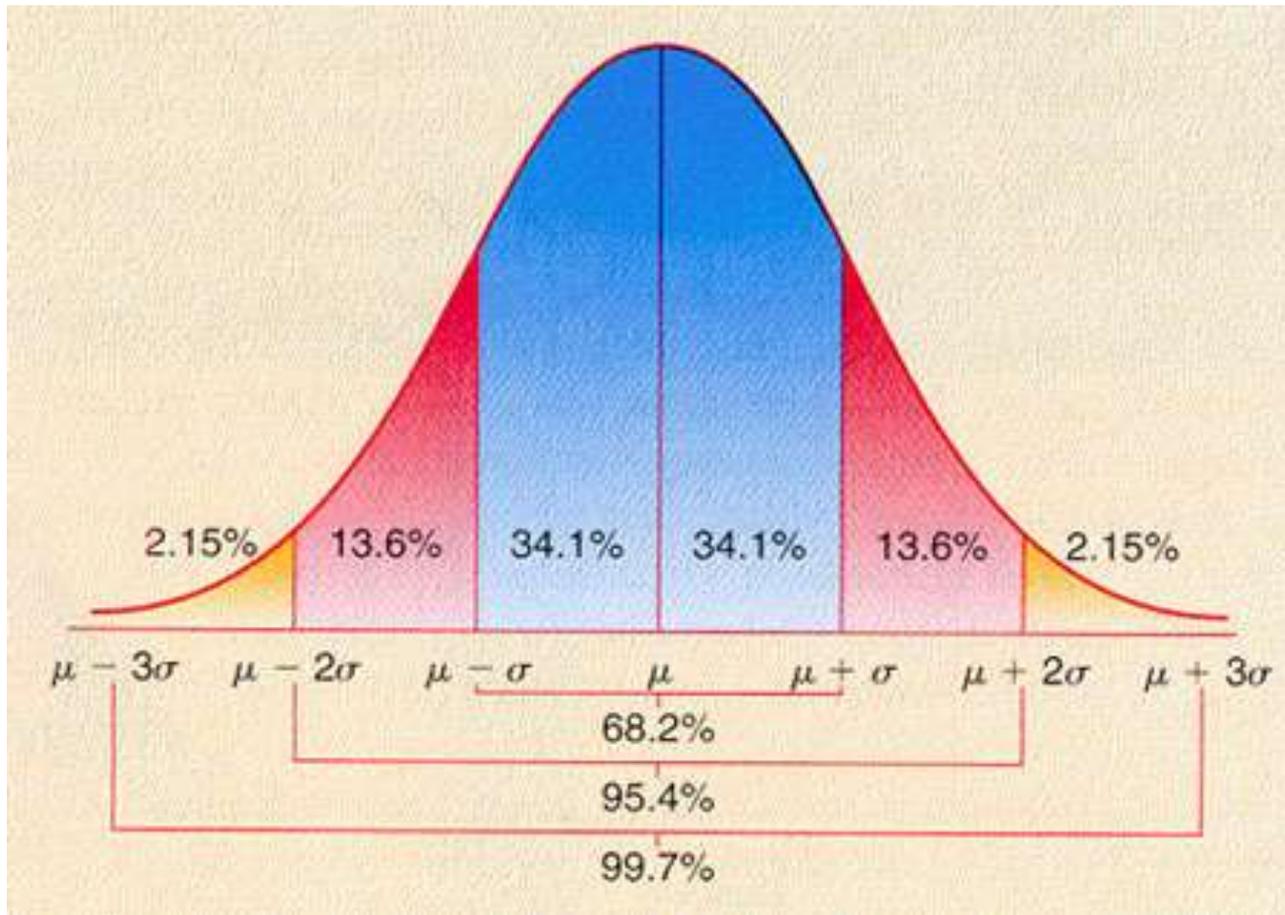
Radial velocities of open stellar clusters: A new solid constraint favouring Tsallis maximum entropy theory

J.C. Carvalho^a, B.B. Soares^a, B.L. Canto Martins^a, J.D. do Nascimento Jr.^a,
A. Recio-Blanco^b, J.R. De Medeiros^{a,*}

The Gaussian, or normal, distribution

$$P_G(x, \mu, \sigma)dx = \frac{dx}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$$





Sum of the variance

$$\sigma^2 = \sigma_1^2 + \sigma_2^2$$

- Exemplo $\sigma_1 = 3$ cm, $\sigma_2 = 4$ cm
- Erro total ? $\sigma = 5$ cm

Robust statistics (ordem)

- Trimean = $(Q1 + 2 \text{ Mediana} + Q3)/4$
- interquartile deviation: $IQ = Q3 - Q1$
- quartile deviation : $QD = IQ/2$
- $MAD = \text{median} \{ |x_i - \text{median}| \}$

pseudo- σ :

- $\sigma_{MAD} = 1,4826 \text{ MAD}$
- $\sigma_{QD} = 1,4826 \text{ QD}$
- $\sigma_{QD} = IQ/1,349$

Exemplo

- 2 5 **5** 6 6 **6** 9 9 **9** 9 150
- 11 elementos
- $\langle x \rangle = 19,6$ $\sigma = 41,3$
- Moda = 9
- Q2 = Mediana (50% da população) = 6
- Q1 (25% da população) = 5
- Q3 (75% da população) = 9
- Trimean = 6,5 IQ = 4 QD = 2 $\sigma(\text{QD}) = 3,0$
- NOTA: se eliminarmos o último ponto (150) obtemos $\langle x \rangle = 6,6$ e $\sigma = 2,2$