

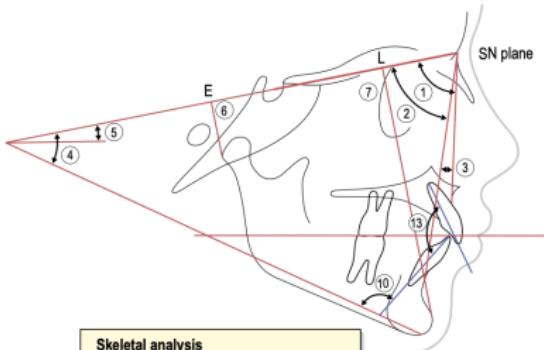
I'Anova

Massimo Borelli

Maggio 2024



differenti etnie



Skeletal analysis

- 1. SNA (82°)
- 2. SNB (80°)
- 3. ANB (2°)
- 4. Mandibular plane to SN (32°)
- 5. Occlusal plane to SN (14.5°)
- 6. Condyle to E point
- 7. Pog-L point

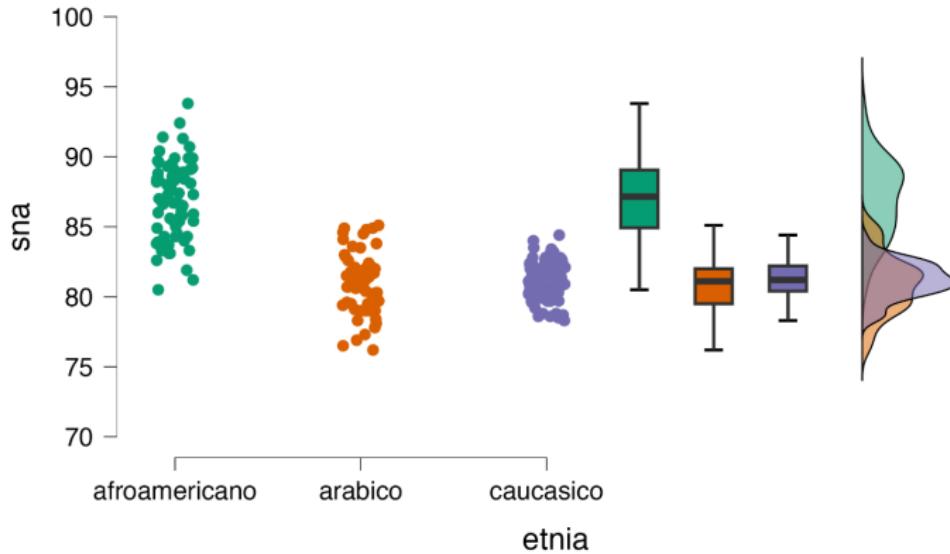
Dental analysis

- 8a. Upper incisor to NA (22°)
 - b. Upper incisor to NA (4 mm)
- 9a. Lower incisor to NB (25°)
 - b. Lower incisor to NB (4 mm)
- 10. Lower incisor to mandibular plane (93°)
- 11. Upper first molar to NA (27mm)
- 12. Lower first molar to NB (23mm)
- 13. Interincisal angle (130°)

differenti etnie

Raincloud plots

sna



oh no!

Independent Samples T-Test

Indep

• The following problem(s) occurred while running the analysis:

- Number of factor levels is ≠ 2 in etnia

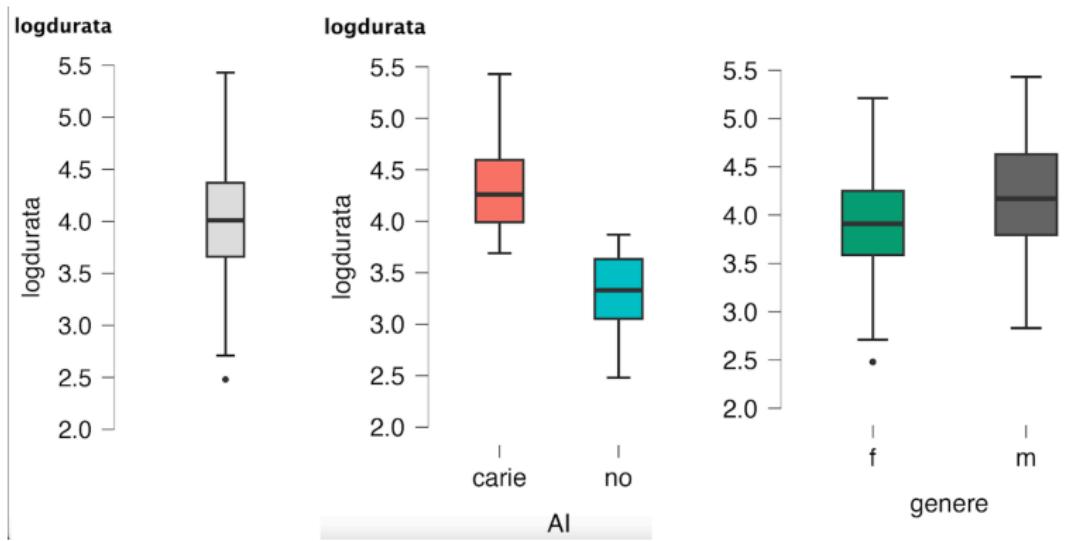
Note. Student's t-test.



$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

quale è l'idea chiave?

Se c'è differenza in media, si riduce anche la dispersione



$$\sigma = 43$$

$$\sigma = 42, \sigma = 10$$

$$\sigma = 30, \sigma = 51$$

1 la one-way Anova

- consideriamo come un **Fixed Factor** la etnia
 - in ordine: afroamericano < arabico < caucasico
- come **Dependent Variable** la sna

iniziamo con l'approccio frequentista

Tabella: ANOVA - sna

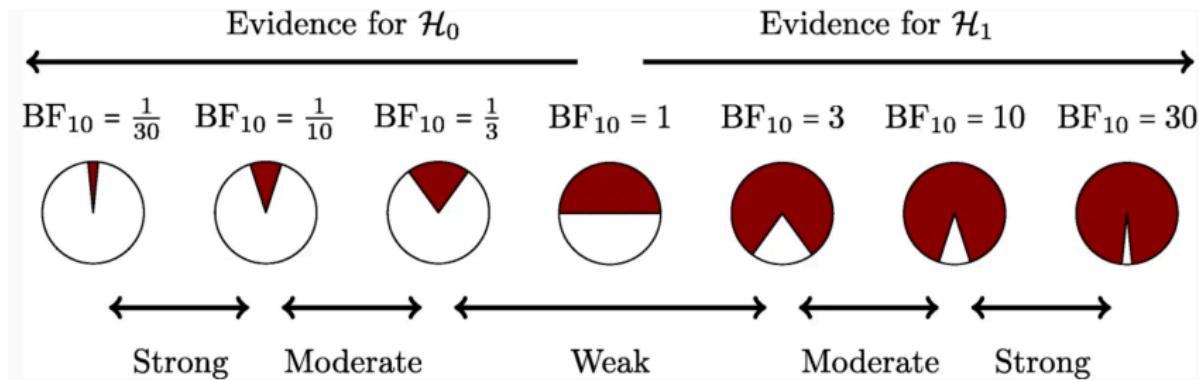
Cases	Sum of Squares	df	Mean Square	F	p
etnia	1620.094	2	810.047	196.673	< .001
Residuals	967.905	235	4.119		

Tabella: Descriptives - sna

etnia	N	Mean	SD	SE	Coef. of variation
afroamericano	66	86.927	2.793	0.344	0.032
arabico	67	80.957	2.104	0.257	0.026
caucasico	105	81.197	1.274	0.124	0.016

ora con l'approccio bayesiano

Models	$P(M)$	$P(M data)$	BF_M
Null model	0.500	5×10^{-48}	5×10^{-48}
etnia	0.500	1.000	$2 \times 10^{+47}$



La diagnostica del modello: 'ipotesi di normalità'

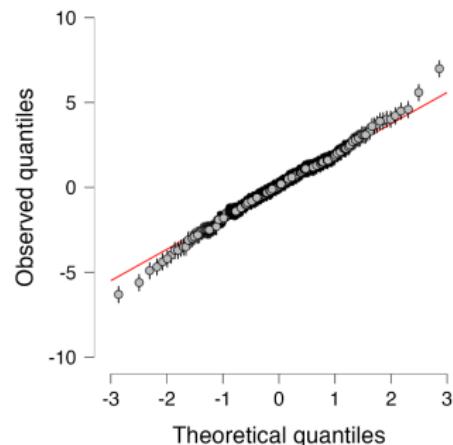
Tabella: Equality of Variances (Levene's)

F	df1	df2	p
26.462	2.000	235.000	< .001

Tabella: Kruskal-Wallis Test

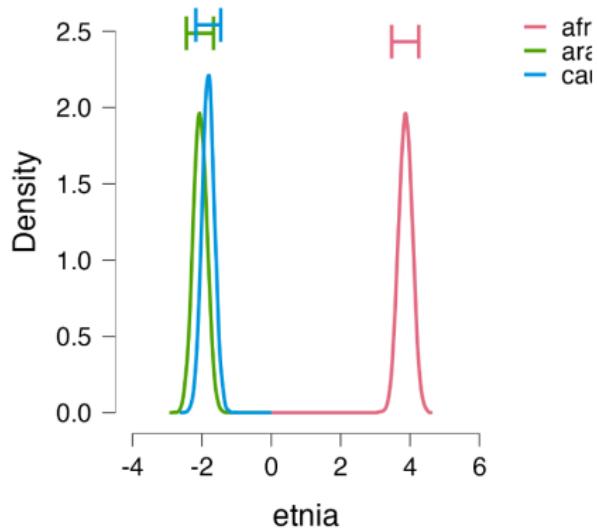
Factor	Statistic	df	p
etnia	123.336	2	< .001

Model Averaged Q-Q Plot



$p < 0.001$; ma chi è diverso da chi?

(bayesian | plots |
Model averaged posteriors)



la questione dei 'multiple comparison' /1

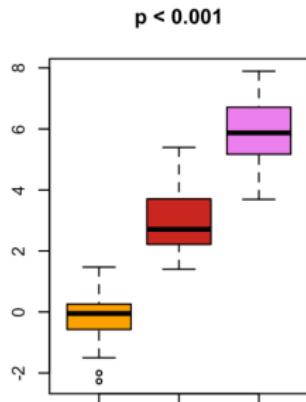
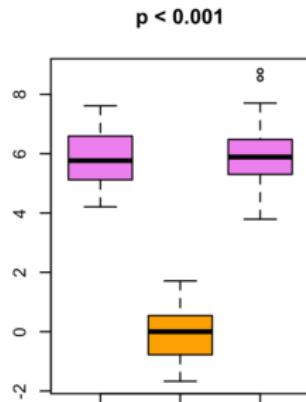
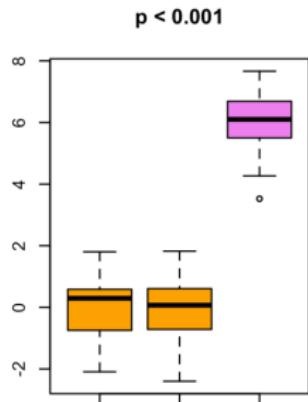
Consideriamo tre simulazioni. Per esempio:

```
a = rnorm(n = 30, mean = 0, sd = 1)
```

```
b = rnorm(n = 30, mean = 0, sd = 1)
```

```
c = rnorm(n = 30, mean = 6, sd = 1)
```

la questione dei 'multiple comparison' /2



la questione dei 'multiple comparison' /3

cattiva idea: fare molti t-tests, ciascuno per ogni coppia di medie.

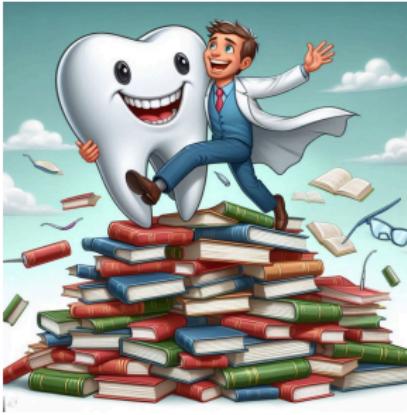
$$\alpha = 0.05$$

$$\begin{aligned}1 - \left(1 - \frac{5}{100}\right) \cdot \left(1 - \frac{5}{100}\right) \cdot \left(1 - \frac{5}{100}\right) &= \\&= 1 - \left(1 - \frac{5}{100}\right)^3 = 0.143\end{aligned}$$

occhio all'errore

I test multipli gonfiano la probabilità di un errore di primo tipo ('mandare in galera un innocente').

la questione dei 'multiple comparison' /4

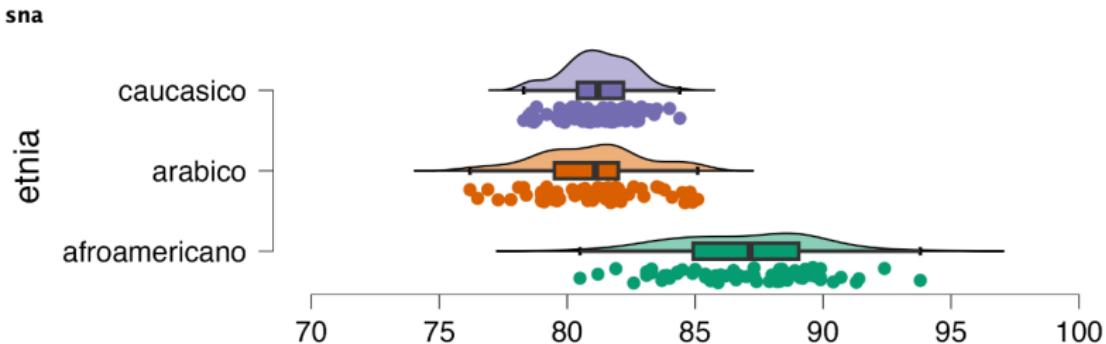


Carlo Bonferroni

soluzione 'radicale' (Bernoulli: $1 + nh < (1 + h)^n$)

Se $n = 3$ gruppi, allora $n \cdot (n - 1)/2 = 3$ confronti, quindi
 $h = \alpha/3 = 0.05/3 = 0.017$.

la soluzione dei 'multiple comparison': John Tukey



		Mean Difference	SE	t	p_{tukey}
afroam	arabico	5.971	0.352	16.964	< .001
	caucasico	5.730	0.319	17.974	< .001
arabico	caucasico	-0.240	0.317	-0.758	0.729

un cenno alla two-way Anova

Cases	...	df	Mean Square	F	p
etnia	...	2	816.609	199.215	< .001
genere	...	1	8.334	2.033	0.155
etnia * genere	...	2	6.054	1.477	0.230
Residuals	...	232	4.099		