

STATISTICA DESCRITTIVA (canale B)

compito n. 1 del 28 marzo 2008

nome e cognome:

corso di laurea: *Astronomia*

matricola:

Parte “teorica”

Si ricorda che la corretta risposta ai quesiti della parte “teorica”

è condizione necessaria, anche se non sufficiente al raggiungimento della sufficienza.

A. Che cosa è una misura di posizione? Si enunci e dimostri almeno una proprietà di un indice di posizione a vostra scelta

B. Si verifichi se le seguenti scritture sono valide

$$(a) \quad \sum_{i=1}^n (x_n - c)(y_n - d) = n(x_n - c)(y_n - d)$$

corretto

errato

$$(b) \quad \prod_{i=1}^n (a^i + b) = a^{\frac{n(n+1)}{2}} + b^n$$

corretto

errato

$$(c) \quad \sum_{i=1}^n \frac{(x_i - a)}{(y_i - b)} = \frac{\sum_{j=1}^n (x_j - a)}{\sum_{j=1}^n (y_j - b)}$$

corretto

errato

$$(d) \quad \sum_{i=1}^n \frac{\log(a)}{x_i} = \frac{n \log(a)}{\sum_{i=1}^n x_i}$$

corretto

errato

Determinare le seguenti somme

$$\sum_{j=6}^{12} 5.1 =$$

$$\sum_{i=1}^5 (i - 10)^2 =$$

--	--

$$\sum_{i=1}^3 \frac{i^5}{3i^6} =$$

--	--

Parte “pratica”

1. L'insieme di dati seguenti riguarda l'astronomia stellare ed è stato estratto da una base di dati raccolta per identificare le differenze tra stelle che ospitano o meno un sistema planetario extrasolare.

È noto che la probabilità di ospitare un pianeta cresce in modo molto ripido con il contenuto in metallo della stella, ma non è chiaro se questo avviene per la metallicità alla nascita o per la successiva crescita del corpo planetario.

Nella tabella seguente si considerano 14 stelle per cui è nota la presenza di pianeti e 12 stelle che viceversa non ospitano alcun pianeta.

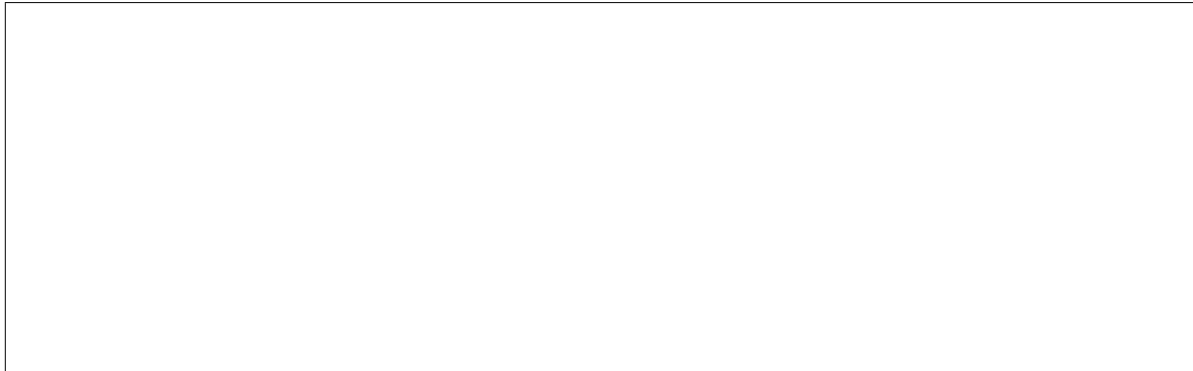
Tra le varie variabili raccolte, sono a disposizione:

- (a) Nome della stella
- (b) Tipo di stella: Tipo=1 indica stelle con pianeti. Tipo=2 è il campione di controllo.
- (c) T_{eff} (in gradi Kelvin) temperatura della stella alla superficie, “temperatura effettiva”
- (d) $\log(Be)$ logaritmo dell'abbondanza di Berillio rispetto alla quantità presente nel sole (cioè il sole ha $\log(Be) = 0.0$).

Si osservi che, per alcune stelle, non si riesce a identificare il valore della la variabile $\log(Be)$, ma solo un suo limite superiore.

Stella	Tipo	T_{eff}	$\log(Be)$	Stella	Tipo	T_{eff}	$\log(Be)$
HD9826	1	6212	1.05	HD1461	2	5768	1.14
HD17051	1	6252	1.03	HD3823	2	5948	1.02
HD22049	1	5073	0.77	HD14412	2	5368	0.80
HD46375	1	5268	< 0.80	HD20010	2	6275	1.01
HD108147	1	6248	0.99	HD20794	2	5444	0.91
HD114762	1	5884	0.82	HD23249	2	5074	< 0.15
HD117176	1	5560	0.86	HD23484	2	5176	< 0.70
HD120136	1	6339	< 0.25	HD36435	2	5479	0.99
HD130322	1	5392	0.95	HD38858	2	5752	1.02
HD145675	1	5311	< 0.65	HD43162	2	5633	1.08
HD179949	1	6260	1.08	HD43834	2	5594	0.94
HD187123	1	5845	1.08	HD76151	2	5803	1.02
HD202206	1	5752	1.04	HD189567	2	5765	1.06
HD210277	1	5532	0.91	HD192310	2	5069	< 0.60
HD217014	1	5804	1.02	HD26965A	2	5126	0.76
HD217107	1	5646	0.96				
HD222582	1	5843	1.14				

- (a) Attraverso un'analisi grafica si confrontino le distribuzioni di $\log(Be)$ tra le stelle con e senza pianeta



- (b) Si fornisca un'opportuno indice di posizione per confrontare le due distribuzioni. Si giustifichi e si discuta la scelta.

- (c) Si ottenga un indice per misurare la variabilità delle due distribuzioni e si confrontino i risultati ottenuti per i due gruppi di stelle.

2. Facendo riferimento ai dati presentati nell'esercizio precedente, si considerino ora solo le stelle per cui abbiamo a disposizione il valore di $\log(Be)$.

[Alcune quantità che potrebbero essere utili allo studente sono state calcolate e vengono fornite di seguito. Considerando le stelle per cui abbiamo osservato $\log(Be)$, si indichino con $i = 1, \dots, 14$ quelle con pianeti e con $i = 15, \dots, 26$ quelle senza pianeti]

$$\begin{aligned} \sum_{i=1}^{14} x_i &= 81303, & \sum_{i=15}^{26} x_i &= 67955, & \sum_{i=1}^{14} x_i^2 &= 473802535, & \sum_{i=15}^{26} x_i^2 &= 385791093, \\ \sum_{i=1}^{14} x_i^3 &= 2770494650145, & \sum_{i=15}^{26} x_i^3 &= 2195722248611, & \sum_{i=1}^{14} y_i &= 13.7, & \sum_{i=15}^{26} y_i &= 11.75, \\ \sum_{i=1}^{14} y_i^2 &= 13.555, & \sum_{i=15}^{26} y_i^2 &= 11.6403, & \sum_{i=1}^{14} y_i^3 &= 13.54733, & \sum_{i=15}^{26} y_i^3 &= 11.6516, \\ \sum_{i=1}^{14} \sum_{j=1}^{14} x_i x_j &= 6610177809, & \sum_{i=1}^{15} \sum_{j=1}^{26} x_i x_j &= 4617882025, & \sum_{i=1}^{14} x_i y_i &= 79873.63, \\ & & \sum_{i=15}^{26} x_i y_i &= 66787.64, & \sum_{i=1}^{14} x_i \sum_{j=15}^{26} x_j &= 5524945365, \end{aligned}$$

- (i) Si ottenga un indice che misuri se è presente una dipendenza in media tra quantità di Berillio osservato e presenza di pianeti. Come si interpreta il risultato?

- (ii) A causa di processi stellari interni, è noto che abbondanze di Berillio sono correlate con la massa stellare che viene tracciata dalla temperatura stellare in superficie T_{eff} .

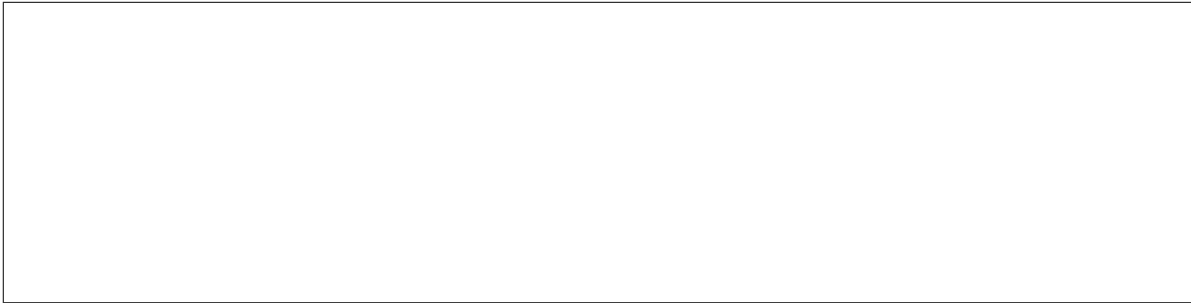
Si disegni un opportuno grafico per verificare la relazione tra la temperatura della stella in superficie (T_{eff}) e la quantità di Berillio ($\log(Be)$). Si faccia attenzione a tracciare il grafico in modo tale che si riesca a confrontare il comportamento delle stelle con pianeta da quelle senza.



- (iii) Si stimino le rette ai minimi quadrati: (a) di tutte le stelle insieme; (b) solo delle stelle con pianeti; (c) delle stelle senza pianeti. Si traccino le rette nel grafico precedente.



- (iv) Si osservano differenze? come si possono descrivere i risultati ottenuti?



- (v) Si fornisca un indicatore della bontà di adattamento delle rette. Quale retta risulta maggiormente adattarsi ai dati?



- (vi) Possiamo considerare i risultati ottenuti validi per tutte le stelle osservate nell'esercizio precedente? perché?

3. Le stelle sono in genere classificate secondo la loro temperatura superficiale, che può essere stimata utilizzando la loro emissione luminosa. La tabella seguente descrive una classificazione comune detta Classe Spettrale, dove la temperatura è misurata in gradi Kelvin.

Classe Spettrale	Temperatura	Colore
O	30.000 ° K - 60.000 ° K	Azzurro
B	10.000 ° K - 30.000 ° K	Bianco-Azzurro
A	7.500 ° K - 10.000 ° K	Bianco
F	6.000 ° K - 7.500 ° K	Giallo-bianco
G	5.000 ° K - 6.000 ° K	Giallo
K	3.000 ° K - 5.000 ° K	Arancione
M	≤ 3.000 ° K	Rosso

Per un insieme di 301 stelle sono disponibili alcune informazioni tra cui la Classe Spettrale e la distanza dalla Terra in anni luce, raggruppata in due grandi classi, distanza minore o uguale a 200 anni luce e distanza maggiore di 200 anni luce. Di seguito viene fornita la tabella (prime due righe) con la distribuzione congiunta delle frequenze assolute delle due variabili in esame.

	Classe Spettrale						
	O	B	A	F	G	K	M
≤ 200 anni luce	0	17	50	16	28	39	7
> 200 anni luce	7	73	5	11	10	27	11
distanza media	977.14	486.02	227.61	427.56	211.68	252.3	260.39

- (a) Si ottengano le distribuzioni marginali delle due variabili considerate.

- (b) Si consideri la variabile distanza dalla terra rispetto ai 200 anni luce. È presente maggiore mutabilità nella sua distribuzione marginale o nella sua distribuzione condizionata alla classe spettrale G? E rispetto alla classe spettrale O? Si risponda utilizzando un opportuno indice statistico.

- (c) Attraverso un opportuno indice e con commenti adeguati si dica se, per le stelle considerate, la Classe Spettrale è in qualche modo legata alla variabile disponibile che misura la distanza dalla terra. È presente una qualche forma di indipendenza?

- (d) L'ultima riga della tabella sopra mostra la distanza media dalla terra delle stelle considerate, presenti in ciascuna classe spettrale. Sfruttando questa informazione è possibile dire se è presente dipendenza in media? Come si potrebbe calcolare un indice che misuri questo tipo di dipendenza?

4. Si consideri una coppia di variabili quantitative (X e Y). Supponiamo che X possa assumere solo un numero limitato di modalità distinte (per fissare le idee pensiamo che X possa assumere solo valori uguali a 1, 2, 3 e 4).

Cosa si può dire in generale della relazione tra i due indicatori η^2 che misura la dipendenza in media e ρ^2 il quadrato del coefficiente di correlazione?

i ☐ $\eta^2 \geq \rho^2$

ii ☐ $\eta^2 \leq \rho^2$

iii ☐ $\eta^2 = \rho^2$

iv ☐ non esiste una regola generale che fornisca un ordinamento tra i due indicatori.

Si spieghi il motivo della scelta precedente.

Qual è il livello massimo della differenza tra η^2 e ρ^2 ? sotto quali condizioni si manifesta?