

Relazione su Step 3

Misura dell' evoluzione delle proprietà del campione

Descrizione

Il terzo step prevede la realizzazione di grafici che mettono in relazione i dati di massa (l_{gm_tot_p50}), tasso di formazione stellare (sfr_{tot_p50}), magnitudine apparente e assoluta al filtro z (petroMag_z, absMag_Z) e flusso della riga di emissione dell'idrogeno H alpha (h_alpha_flux) con il redshift (z). Le diverse quantità tracciate in funzione del redshift dovranno essere testate per valutare l'intensità della loro correlazione con esso. Questo viene svolto con il coefficiente di Pearson definito

come:
$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$
 che restituisce un coefficiente

numerico da -1 a 1.

se $r \sim 1$, le 2 variabili sono correlate;

se $r \sim -1$, le 2 variabili sono anti correlate;

se $r \sim 0$, non c'è alcuna correlazione tra le 2 variabili.

Come soglia per determinare una correlazione significativa utilizziamo un valore del coefficiente superiore a 0,5 o inferiore a -0,5.

Nel caso in cui un trend sia rilevato sarà necessario dividere il campione in appropriati intervalli di redshift: 0,00 - 0,02; 0,02 - 0,04; 0,04 - 0,06; 0,06 - 0,08; 0,08 - 0,10. In modo da ricreare un istogramma che mostri la distribuzione della quantità correlata nei vari intervalli. Insieme ad essi vogliamo mostrare il valore della media e deviazione standard così come mediana e sedicesimo/ottantaquattresimo percentile per ogni istogramma. Infine gli esatti valori di media, deviazione standard, mediana e percentili devono essere salvati in un file di output.

1 #	Name	mean-std	mean	mean+std
2	0.00-0.02	8.169745	8.634751	9.099757
3	0.02-0.04	8.850512	9.352468	9.854424
4	0.04-0.06	9.524237	9.926364	10.328491
5	0.06-0.08	9.919982	10.280269	10.640555
6	0.08-0.10	10.232164	10.528088	10.824011

Fig. 1 valori di media e deviazione standard per la massa nei vari redshift bin

Metodo

Ho importato le librerie 'os', 'numpy', 'scipy.stats.stats', da cui è stato estratto 'pearsonr', per il calcolo del coefficiente di Pearson, 'matplotlib.pyplot' e 'matplotlib.gridspec' per la realizzazione di scatter plot e istogrammi.

Inizialmente ho estratto le quantità necessarie allo svolgimento dello step ['petroMag_z', 'h_alpha_flux', 'lgm_tot_p50', 'Sfr_tot_p50', 'absMagZ'] creando una nuova lista che le contenesse.

Successivamente ho dovuto togliere i valori non leggibili con il modulo 'isfinite' di numpy e gli outliers con il sigma clipping, ma farlo in maniera tale da ottenere lo stesso numero di dati per l'array del redshift e la proprietà con cui deve essere plottato. Poiché le liste di dati con cui ho avuto a che fare sono numpy array, ho utilizzato i moduli di numpy per la rimozione di elementi non utilizzabili.

Con il metodo sigma clipping è possibile rimuovere i dati che si trovano a più di 4 deviazioni standard dalla distribuzione, mantenendo il 99,99994% di essi.

Ho creato una funzione che contenesse l'intero svolgimento dello step 3 che andrò a chiamare per ogni quantità da plottare in funzione del redshift. Ho calcolato attraverso il modulo 'pearsonr' della libreria scipy.stats.stats il coefficiente di Pearson e attraverso i moduli 'polyfit' e 'polyval' sempre della libreria di numpy ho calcolato i valori per rappresentare la linea di 'Best-fit' che andremo ad inserire nel grafico per avere una prima impressione sull'esistenza di relazioni. Attraverso diversi moduli della libreria matplotlib.pyplot sono andato a creare il grafico, scatter plot. Una volta definite tutte le funzioni che vanno a creare il grafico ho posto il condizionale della correlazione che crea gli istogrammi e gli output per le sole quantità che hanno una correlazione con il redshift $> 0,5$. Ho suddiviso i dati in intervalli di redshift con un ciclo while così da formare un istogramma per ogni intervallo andando a ripetere ciò che è stato fatto nello step 2.

Risultati

I risultati hanno confermato ulteriormente quello che era visibile negli istogrammi; seguendo come esempio le figure 2 e 3 riguardanti l'array della massa, possiamo notare come al crescere dell'intervallo del redshift cresca anche la media degli intervalli della massa. Il contrario lo potremmo vedere nel file della magnitudine assoluta, dove al crescere degli intervalli del redshift decresce la media degli intervalli della magnitudine. Da questo possiamo dedurre che la massa e la magnitudine assoluta hanno una forte correlazione con il redshift, mentre la magnitudine apparente, il flusso h_{α} e il tasso di formazione di stelle sono indipendenti (secondo questi specifici dati) da esso. Da notare dalla figura 3 che negli intervalli di redshift più bassi, gli istogrammi sono meno distribuiti poiché la quantità di valori presenti in quella zona è minore rispetto agli intervalli più alti dove invece riusciamo a riprodurre una distribuzione gaussiana.

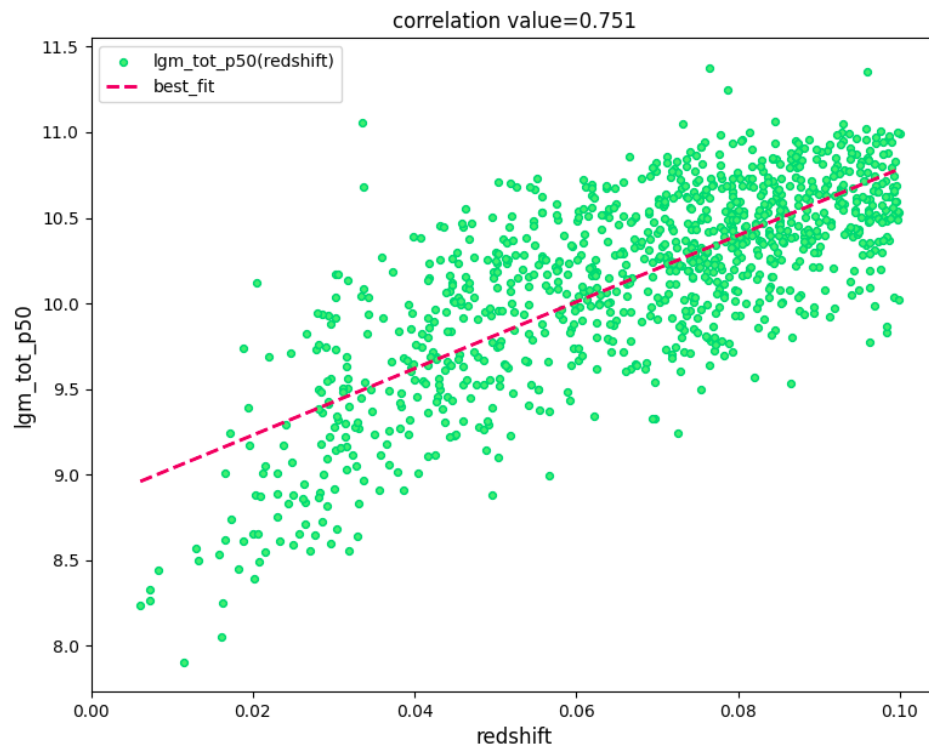


Fig. 2 Massa in funzione di redshift

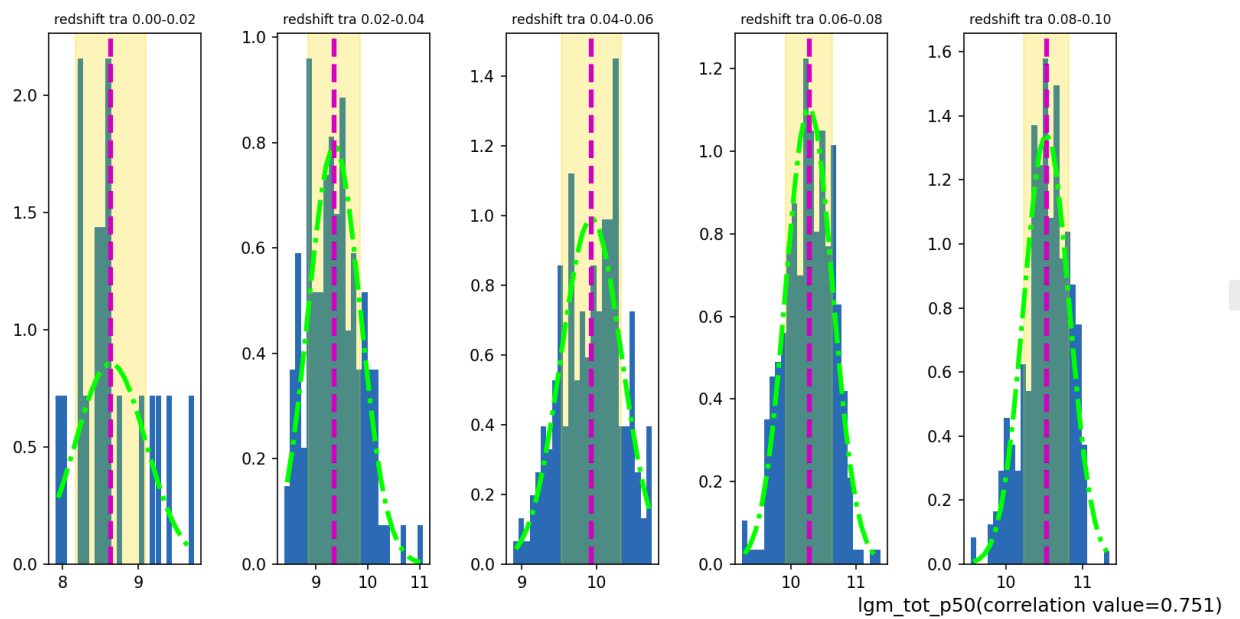


Fig. 3 distribuzione della massa per intervalli di redshift.
in magenta è mostrata la media e in verde la gaussiana