

Andrea Pisoni, andreapisoni@virgilio.it
Davide Valentinis, davidevalentinis@virgilio.it
Laureandi in Ingegneria Fisica, Politecnico di Milano

Progetto di analisi statistica: misurazione della costante di Hubble e verifica della legge di espansione dell'Universo

Introduzione.

Da decenni ormai la cosiddetta legge di Hubble affascina e coinvolge le menti dei più illustri studiosi di Astrofisica per la sua relativa semplicità, una linearità che tuttavia cela una profonda consistenza con i dati osservativi su scala galattica.

In astronomia e cosmologia la Legge di Hubble, scoperta da Edwin Hubble e Milton Humason nel 1929, afferma che esiste una relazione lineare tra il redshift, ossia il termine anglo – sassone per designare lo "spostamento verso il rosso", della luce emessa dalle galassie e la loro distanza: tanto maggiore è la distanza della galassia e tanto maggiore sarà il suo redshift; di conseguenza risulterà maggiore anche la sua velocità di recessione [4]. Scopo di questo elaborato è la dimostrazione sperimentale della relazione di proporzionalità diretta fra la velocità di recessione delle galassie e la loro distanza dal punto di osservazione; i dati disponibili, corrispondenti alle misurazioni effettuate dallo stesso Hubble nel 1929 [6], sono stati trattati attraverso il software statistico *Essential Regression*, un pacchetto applicativo sviluppato per Microsoft Excel [5].

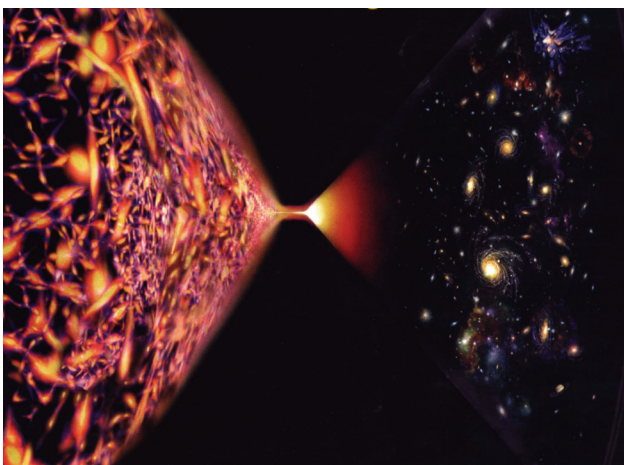


Figura 1. Rappresentazione artistica dell'espansione dell'Universo a partire dal Big Bang

Legge di Hubble e modello del Big Bang.

L'interazione gravitazionale, pur possedendo una minore intensità rispetto alle altre interazioni fondamentali della Natura, rappresenta la componente essenziale per comprendere la composizione attuale dell'Universo e la sua evoluzione temporale. La materia presente nel cosmo, infatti, si presenta organizzata in una gerarchia di strutture di dimensione progressivamente maggiore ed in moto relativo le une rispetto alle altre [3]. Sebbene su scala locale, quale quella dei singoli corpi celesti e dei sistemi stellari, le traiettorie nello spazio vuoto siano dominate dalle forze gravitazionali e risultino sostanzialmente aleatorie, le osservazioni sperimentali indicano che a livello galattico esista un moto di deriva sistematico di separazione o recessione, il quale suggerisce che le galassie si stiano allontanando vicendevolmente nel corso del tempo [4]. La relazione empirica fra la distanza della singola galassia dalla Terra e la velocità di recessione della stessa risulta approssimativamente lineare: tale corrispondenza viene denominata legge di Hubble, dal nome dell'astronomo nordamericano Edwin Hubble che la determinò nel 1929; recenti misurazioni, effettuate attraverso strumenti e tecniche maggiormente potenti e sensibili, confermano la natura proporzionale della relazione e ne estendono la validità a corpi celesti distanti diversi miliardi di anni luce dal nostro pianeta, fornendo un'evidenza sperimentale in favore dell'ipotesi di un Universo in espansione e del modello cosmologico del Big Bang [3].

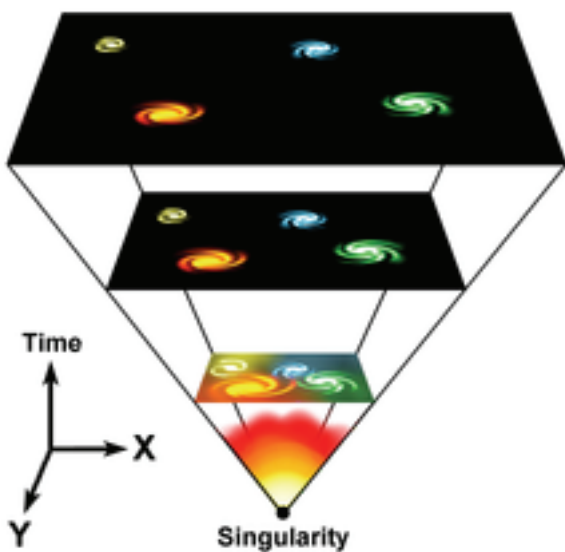


Figura 2. L'evoluzione e la recessione delle galassie secondo il modello cosmologico del Big Bang

Derivazione sperimentale della legge.

La luce, nello stesso modo delle radiazioni sonore. È soggetta all'effetto Doppler relativistico, secondo il quale la radiazione elettromagnetica emessa da una sorgente in allontanamento da un osservatore presenta una frequenza minore di quella reale di emissione propria della sorgente stessa; se la galassia si avvicina allo spettatore la frequenza di emissione della prima risulta aumentata. La teoria discende dalla natura ondulatoria della luce, secondo la quale la radiazione luminosa risulta interpretabile quale un'onda dotata di opportuna frequenza. Essendo una sorgente luminosa, anche una galassia si rivela soggetta a tale fenomeno fisico, ammesso che essa si muova rispetto ad un osservatore terrestre [4].

Osservando lo spettro di emissione delle galassie, si nota che esso possiede la configurazione caratteristica di un qualsiasi ammasso stellare, ma con un sistematico spostamento delle linee spettrali verso le frequenze minori; tale effetto viene detto spostamento verso il rosso o *redshift*, in quanto la radiazione di colore rosso corrisponde alla frequenza minore presente nella gamma della luce visibile.

Il *redshift* delle galassie, dunque, supporta l'ipotesi secondo cui le galassie si allontanano reciprocamente dalla Via Lattea. Il moto, tuttavia, essendo relativo, non dipende dalla posizione dell'osservatore nello spazio, ossia da qualunque oggetto celeste si constaterrebbe il medesimo fenomeno di recessione delle altre nebulose galattiche rispetto al sistema di riferimento; geometricamente la dinamica equivale a quella di una sfera in espansione con raggio crescente, per cui il centro della dilatazione non si trova sulla superficie esterna e da ogni punto origine di quest'ultima gli altri

punti superficiali si allontanano con velocità proporzionale alla distanza dall'origine di riferimento [4].

Sperimentalmente il *redshift* può essere misurato correlando lo spettro di emissione della galassia in esame con quello prodotto dai corrispondenti elementi chimici in analisi di laboratorio; uno spettrografo valuta lo spostamento Doppler, quindi la velocità è dedotta dalla seguente legge valida in prima approssimazione:

$$\text{shift} = \lambda_{\text{riposato}} \cdot \frac{\text{velocità radiale}}{c} \quad (1)$$

Per la misurazione delle distanze galattiche Hubble utilizzò il metodo delle Cefeidi [6]: queste ultime rappresentano stelle la cui luminosità è oscillante con un periodo ben definito, il quale dipende soltanto dalla magnitudine assoluta della stella analizzata [4]. Dato che risulta possibile determinare la magnitudine assoluta di Cefeidi relativamente prossime alla Via Lattea, la tecnica consente di misurare la magnitudine assoluta di qualunque Cefeide quantificando il periodo di variazione della sua luminosità ed applicando la relazione seguente per calcolarne la distanza [4]:

$$m_{\text{apparente}} - M_{\text{assoluta}} = 5 \log_{10} r - 5 \quad (2)$$

Nell'equazione (2) compaiono la magnitudine assoluta della galassia in esame, rispettivamente $m_{\text{apparente}}$ ed M_{assoluta} , invece r rappresenta la distanza radiale del sistema galattico analizzato dall'osservatore. Una volta determinate le velocità e le distanze delle galassie è possibile rappresentare la relazione fra le due grandezze in un grafico: il coefficiente angolare della retta che interpola le misure effettuate rappresenta la costante di Hubble [4].

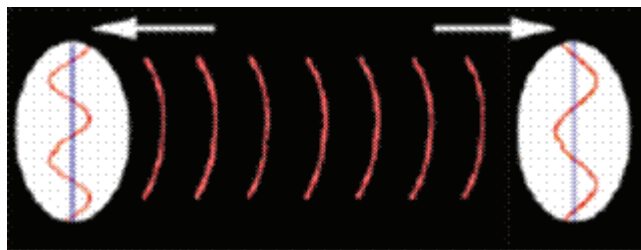


Figura 3. Il *redshift* delle galassie, corrispondente allo spostamento verso il rosso della frequenza della radiazione emessa dalle galassie in recessione, mostra che i sistemi galattici si allontanano reciprocamente al ritmo determinato dalla costante di Hubble

Il valore ottenuto da Hubble corrisponde a circa 492 km/(s Mparsec), il quale si rivela distante dalle misure effettuate con le tecniche odierne [6]; queste ultime forniscono un coefficiente di circa 72 km/(s Mparsec) [7]. In realtà gli errori commessi nelle prime analisi di Hubble alterano il risultato sperimentale; in particolare:

- Hubble aveva confuso due qualità di stelle variabili Cefeidi utilizzate per valutare le distanze astronomiche [6]
- L'astronomo aveva considerato molto luminosi degli astri provenienti da regioni III lontane [6]
- Il metodo di valutazione delle distanze si fonda sull'osservazione della magnitudine stellare apparente utilizzando un semplice telescopio ottico; l'errore sperimentale conseguente alle valutazioni dell'occhio umano si ripercuote sui dati impiegati per le analisi

La correzione di queste imprecisioni condusse ad un ridimensionamento graduale verso il basso del valore della costante di Hubble nei decenni successivi. I dati di Hubble del 1929 rivelano notevoli lacune statistiche. Infatti le singole galassie hanno velocità peculiari di diverse centinaia di km/s e corrispondono a recessioni massime soltanto di 1.200 km/s [6]. In passato le imprecisioni hanno indotto persino a proporre una legge quadratica redshift – distanza, ma i dati illustrati di seguito sulle supernovae di tipo Ia, raccolti da Riess, Press e Kirshner nel 1996, si estendono oltre 30.000 km/s e forniscono un'evidente conferma della legge lineare di Hubble,

$$v = \frac{dD_{AB}}{dt} = HD_{AB} \quad (3)$$

La retta interpolante nel grafico seguente ha una pendenza di 65 km/s/Mpc [7].

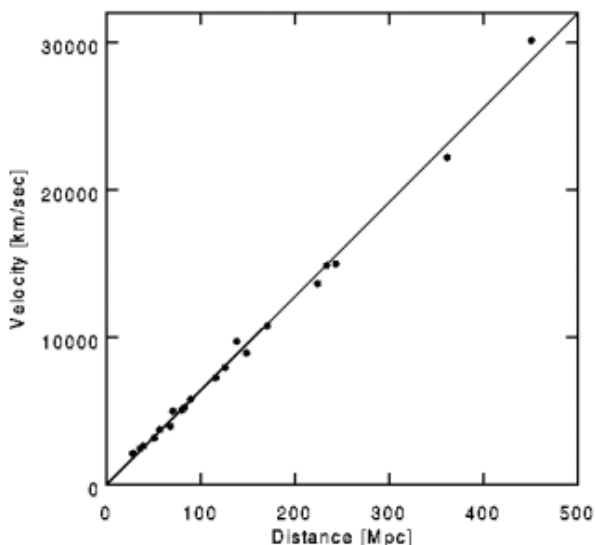


Figura 4. Grafico delle velocità, in MegaParsec, rapportate alle distanze, in Km/sec, per le supernovae di qualità Ia analizzate nel 1996 da Riess, Press e Kirshner

Il problema della misura della costante di Hubble è stato sostanzialmente risolto solo negli ultimi anni, quando misure effettuate con il Telescopio Spaziale Hubble hanno fornito una stima di $H_0 = 72 \pm 8$ (km/s)/Mpc (Maggio 2001), confermata poi da diverse altre osservazioni indipendenti: ad esempio, le osservazioni della radiazione cosmica di fondo condotte col satellite WMAP (2003) hanno fornito un valore di H_0 pari a 71 ± 4 (km/s)/Mpc [7]. Per quanto riguarda il campo di validità della legge di Hubble, la relazione non può essere applicata per distanze astronomiche molto piccole, generalmente inferiori a qualche milione di anni luce, e per distanze molto grandi, generalmente al di sopra di alcuni miliardi di anni luce [4].

Per quanto riguarda le distanze molto piccole, applicando la relazione matematica alle galassie della nebulosa di Andromeda, queste si dovrebbero allontanare, mentre invece presentano un moto di avvicinamento [4]. Si pensa che l'inversione della norma espansionistica sia dovuta a certi effetti gravitazionali locali che si estrinsecano più intensamente quando gli oggetti celesti sono relativamente vicini [3]. Per le grandi distanze, invece, la ridotta applicabilità della legge dipende dal fatto che errori e incertezze tendono ad aumentare man mano che le distanze si fanno più grandi [3].

Legge di Hubble e modelli teorici di Universo.

La formulazione della relazione proporzionale fra velocità di recessione delle galassie e relative distanze dall'osservatore costituisce un supporto fondamentale all'ipotesi di un cosmo in espansione [4].

Ripercorrendo in senso inverso la traiettoria radiale descritta dalle singole galassie, dunque, si giunge ad un istante in corrispondenza del quale la distanza tra di esse risulta nulla. Secondo questa concezione l'Universo sarebbe stato originato dall'esplosione di una singolarità iniziale denominata comunemente big bang [4].

Le medesime congetture di un cosmo dinamico in espansione sono state dedotte teoricamente dalle equazioni della Relatività generale di Albert Einstein, fornendo un ulteriore riferimento per la teoria moderna del modello cosmologico standard [3].

La legge lineare redshift – distanza trovata da Hubble si rivela compatibile con la visione copernicana dell'Universo: la nostra collocazione nel cosmo non è affatto speciale. Innanzitutto si nota che la velocità di allontanamento è una quantità simmetrica: se A vede B allontanarsi, allora B vede A separarsi; tale proprietà non è valida se la relazione considerata non è lineare. Per tale ragione, se ammettessimo una relazione quadratica tra velocità e distanza, allora un osservatore

in un'altra galassia scoprirebbe una legge di genere differente ed inoltre tale relazione dipenderebbe dalla direzione; dato che tali proprietà non sussistono, si conclude che la relazione fra recessione e distanza radiale è di proporzionalità diretta [4]. Inoltre la legge di Hubble genera un'espansione omologa che non altera la forma degli oggetti, invece altre possibili relazioni velocità – distanza producono distorsioni durante l'espansione [3].

La legge di Hubble fornisce un procedimento efficace per la determinazione delle distanze galattiche: conoscendo l'entità del redshift è possibile derivare la velocità radiale e dunque la distanza di un corpo celeste [4].

La costante di proporzionalità viene denominata H_0 e determina fisicamente il ritmo dell'espansione del cosmo. Calcolando il reciproco di tale grandezza si ottiene un'unità di misura temporale, corrispondente all'età attuale dell'Universo in relazione all'istante del Big Bang. Il pedice "0" in H_0 indica il valore istantaneo di una quantità variabile nel tempo. Dal momento che $1/H_0$ o t_0 è approssimativamente l'età del cosmo, anche il valore di H_0 dipende dal tempo. Infine, dato che la luce si muove nel vuoto con velocità c , moltiplicando questa grandezza per il valore di t_0 si ottiene la misura approssimata del raggio attuale dell'universo visibile [4].

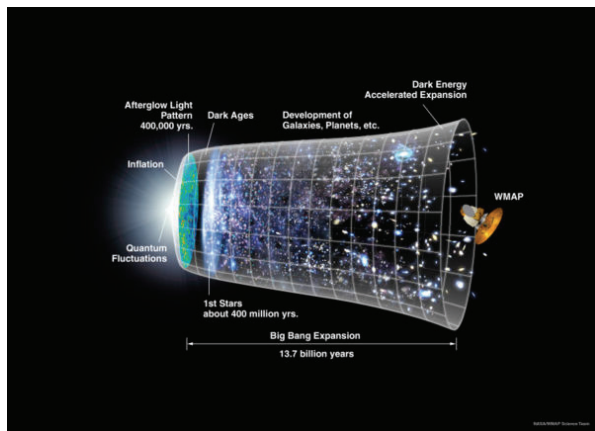


Figura 5. Evoluzione dell'Universo secondo la moderna teoria cosmologica dell'Universo inflazionario: l'espansione inizia da una singolarità primordiale detta Big Bang, in corrispondenza della quale la totalità della materia presente nel cosmo era compressa in un volume ridottissimo di spazio.

Parte sperimentale.

Riferendoci a dati provenienti dall'articolo originale redatto da Hubble nel 1929 [8], abbiamo effettuato inizialmente un'analisi di genere qualitativo, per assicurarci dell'adeguatezza dei dati rispetto al modello proposto.

Tabella 1. Set di dati relativo alle distanze delle galassie, esaminate da Hubble, espresse in MegaParsec e delle corrispondenti velocità di recessione, espresse in Km/sec.

DISTANZE (Megaparsec)	VELOCITA' DI RECESSIONE (Km/s)
0,032	170
0,034	290
0,45	200
0,5	290
0,5	270
0,63	200
0,8	300
0,9	650
0,9	150
0,9	500
1	920
1,1	450
1,1	500
1,4	500
1,7	960
2	500
2	850
2	800
2	1090

A tale proposito, abbiamo verificato che il set di dati a nostra disposizione soddisfacesse il requisito detto di normalità: questo significa che la variabilità di ogni misurazione, ossia il grado di errore sperimentale sul valore della misura, deve essere distribuita statisticamente secondo una curva nota quale gaussiana [1]; usufruendo di appositi strumenti forniti da Excel, è possibile rappresentare la variabilità dei dati osservati rispetto ai valori corrispondenti di una curva gaussiana, ed il grafico risultante è denominato *Normal Probability Plot*. [2]. Se l'andamento risultante si approssima ad una retta, l'approssimazione gaussiana può considerarsi verificata e dunque si è in grado di procedere con l'applicazione del modello di regressione lineare [1].

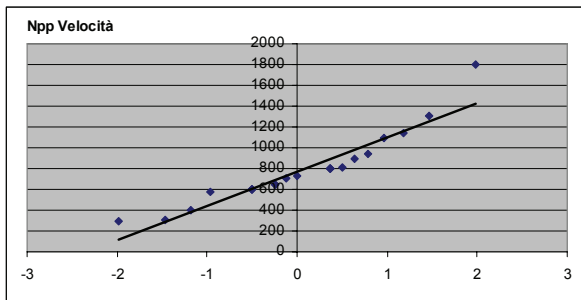


Figura 2. Normal Probability Plot per le velocità delle galassie osservate da Hubble.

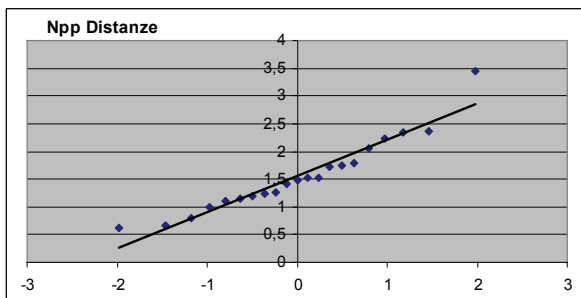


Figura 3. Normal Probability Plot per le distanze delle galassie incluse nel modello di regressione.

I grafici mostrano che l'elaborazione dei dati sperimentali assume un comportamento assimilabile a quello delle corrispondenti previsioni teoriche, data la distribuzione approssimativamente lineare mostrata dai grafici di cui sopra. Il requisito di normalità è necessario per assicurare una corretta esecuzione del modello di regressione; per i dettagli si rimanda alla letteratura specifica [1].

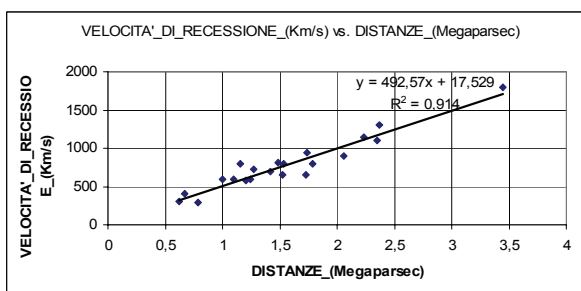


Figura 4. Grafico di regressione prodotto da Essential Regression per la relazione fra velocità e distanza delle galassie [5]: la distribuzione dei dati si approssima in misura notevole ad una retta passante per l'origine degli assi.

Il modello di regressione che si andrà ad applicare è una versione cosiddetta lineare semplice, poiché la variabile dipendente o risposta, contrassegnata dalla lettera Y in figura, è posta in relazione con un'unica variabile indipendente X, denominata regressore, tramite un'equazione del tipo:

$$Y = b_0 + b_1 X \quad (4)$$

I parametri b_0 e b_1 presenti nell'equazione (4) sono detti coefficienti di regressione e rappresentano

rispettivamente la pendenza e l'intercetta della retta rappresentativa della relazione lineare [2]. L'intercetta costituisce il punto in cui il grafico interseca l'asse delle ordinate, ossia il valore della variabile dipendente Y corrispondente al valore nullo della variabile indipendente X. Il significato fisico del parametro b_0 è costituito dalla velocità che presenterebbe una galassia situata a distanza 0 dalla Terra; le valutazioni fisiche impongono che un tale corpo, essendo situato nello stesso punto dell'osservatore, non si muova rispetto a quest'ultimo, dunque il valore di b_0 dovrebbe rivelarsi nullo secondo la concezione teorica adottata [4].

Nel nostro caso in particolare notiamo che un modello così strutturato rispecchia l'equazione proposta da Hubble, assumendo $b_0 = 0$ e $b_1 = H$, nonché identificando X con la distanza dalla Terra di una delle galassie osservate e Y con la velocità di recessione propria della galassia stessa. Nostro intento sarà quindi quello di verificare effettivamente tali uguaglianze e di conseguenza l'equazione di Hubble; tale obiettivo può essere conseguito tramite un opportuno programma che, partendo dai dati in input immessi dall'utente (in questo caso quelli relativi alle tabelle precedenti), sviluppa un modello di regressione lineare [5]; tale modello è rappresentato graficamente nella figura 4. In particolare viene fornita una stima del valore dei coefficienti di regressione e viene prodotto quello che si suole dire un intervallo di confidenza per i medesimi con un relativo livello di significatività. La regione di confidenza è definita calcolando gli estremi di un intervallo probabilistico entro cui si ritiene che giaccia il valore di una singola misura da noi cercato; nel nostro caso, le analisi sono state effettuate con un livello di confidenza pari al 95% [1]. Inoltre, assieme alla stima dei coefficienti di regressione, è fondamentale stabilire quanto questi valori siano rilevanti e significativi all'interno del modello proposto, ossia è necessario valutare la probabilità statistica che tali coefficienti non siano nulli [1]. Per far questo ci si riferisce ad un parametro detto p-value; basti sapere che tanto più questo parametro è piccolo tanto più il coefficiente è rilevante per spiegare il modello e dunque tanto meno può essere ignorato o addirittura eliminato [1].

Dai grafici elaborati (figura 4) si nota come in maniera approssimata i dati in nostro possesso evidenzino una certa linearità, in conformità con quanto sostenuto da Hubble; per confermare questa affermazione ci si può riferire ad un parametro, detto R^2 , che mostra quanto il modello proposto dall'utente riesca a spiegare la variabilità dei dati raccolti, ossia quanto il valore teorico calcolato dal modello lineare si avvicini a quello ricavato sperimentalmente: tanto questo indice si avvicina a 1, tanto il modello si dimostra aderente alla realtà fisica [1]. Nel nostro caso si calcola $R^2 = 0,914$, considerabile un ottimo risultato considerando anche l'incertezza e gli errori sperimentali contenuti nei dati in dotazione.

Tuttavia il valore dell'indice R^2 non è sufficiente a dimostrare la bontà del modello: bisogna innanzitutto riferirsi alla cosiddetta diagnostica, ovvero al grafico dei residui: i residui rappresentano il discostamento tra

ogni valore calcolato dal modello e la misura sperimentale corrispondente [1]. Per assicurarci una corretta analisi dobbiamo verificare che questi residui si distribuiscano in maniera completamente casuale (in gergo si suole dire “a nuvola”), in modo da non mostrare alcun comportamento notevole e quindi compromettere la *randomizzazione* dei dati; qualora si constatasse un andamento ordinato, quale una parabola o una curva, si potrebbero avanzare dubbi sull’effettiva linearità della relazione fra le variabili in esame [1]. Altri andamenti notevoli sono le distribuzioni dei residui ad imbuto, le quali evidenziano una variabilità delle misure che dipende dai dati stessi; tale fenomeno è denominato *eteroschedasticità*. Infine una distribuzione ordinata dei residui è in grado di evidenziare anomalie negli strumenti sperimentali: per esempio un errore può avvenire se lo strumento di misurazione è soggetto a surriscaldamento: si noterebbe in questo caso uno scostamento maggiore dei dati osservati, rispetto alle previsioni teoriche, per misure più recenti [1]. Il grafico relativo ai residui (fig 5 e fig 6) soddisfa in modo consistente le nostre aspettative. Per quanto riguarda la stima dei coefficienti di regressione l’ output è il seguente (fig 7).

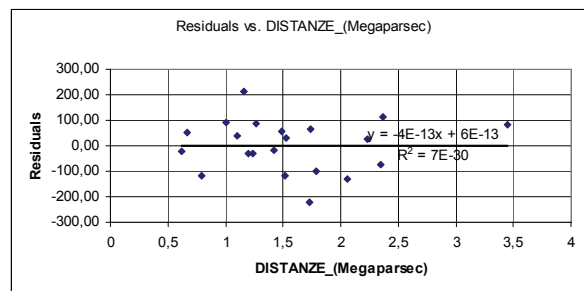


Figura 6. Distribuzione dei residui per le misure relative alla distanza delle galassie osservate dalla Terra; la configurazione non evidenzia alcun trend notevole, quindi anche in questo caso i dati soddisfano i requisiti di variabilità necessari per lo svolgimento delle analisi successive.

VELOCITA' DI RECESSIONE_(Km/s) = b0 + b1*DISTANZE_(Megaparsec)					
		P value	Std Error	-0,95	0,95
b0	17,5294117	0,768022	58,58399	-105,088	140,1472
b1	492,5720243	1,42E-11	34,66618	420,0148	565,1292

Figura 7. Stima dei coefficienti di regressione per il modello lineare proposto per la legge di Hubble

Il valore b_0 si discosta da zero non in maniera significativa, infatti è pari a 17, invece b_1 , pari a 492,6, si avvicina al valore sperimentalmente ricavato: 498,7. Ancora più interessante è l’intervallo di confidenza per il termine noto (-105.088;140.1472); questa regione risulta includere ampiamente lo zero, conformemente al modello proposto. Se si osservano i relativi p – value si nota come quello di b_0 (circa 0,77) sia alto se confrontato a quello di b_1 ($1,42 \cdot 10^{-11}$); tale misura sottolinea la notevole sussistenza di una relazione di tipo lineare: essendo b_1 così significativo per la correttezza del modello, viene garantita una pendenza non nulla della retta di regressione, mentre b_0 non sembra essere indispensabile ai fini dell’interpretazione delle osservazioni effettuate.

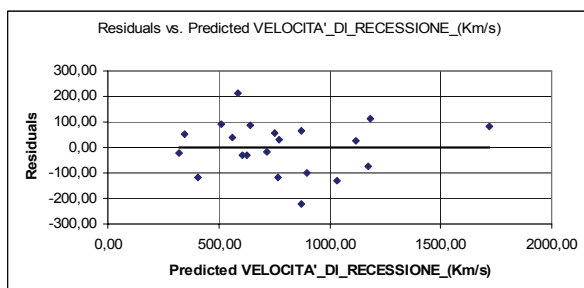


Figura 5. Distribuzione dei residui per i dati relativi alla velocità di recessione delle galassie; la configurazione non presenta alcun andamento particolare e dunque la randomizzazione delle misure risulta confermata.

Risultati e discussione.

Le elaborazioni statistiche considerate consentono di confermare, entro l’errore sperimentale, la consistenza del modello di regressione lineare per la tematica trattata. Le misurazioni, analizzate attraverso strumenti informatici di calcolo e l’utilizzo del computer, avvalorano l’ipotesi avanzata da Hubble nel 1929 [8]: fra la velocità di recessione delle galassie considerate dall’astronomo americano e la corrispondente distanza approssimata dalla Terra sussiste effettivamente una relazione matematica, la quale aderisce in misura soddisfacente al modello di proporzionalità diretta. Il coefficiente di regressione lineare ottenuto concorda sufficientemente con il valore proposto da Edwin Hubble nel corso delle sue analisi. Sebbene questa misurazione sovrastimi il valore della costante di recessione, tale anomalia è dovuta essenzialmente all’incertezza sperimentale delle osservazioni ed agli errori di valutazione commessi durante l’interpretazione dei dati empirici. La proporzionalità fra velocità e distanza non viene affatto invalidata dalle approssimazioni descritte; al contrario, gli indici statistici analizzati evidenziano l’accuratezza del modello di regressione lineare, la quale è stata confermata, nel corso dei decenni successivi alle misure di Hubble, attraverso strumentazioni tecnologiche e metodi analitici maggiormente sensibili ed efficaci.

Bibliografia

- [1] Montgomery D., Runger G., Hubele N., *Statistica per ingegneria*, Egea, Milano, 2004
- [2] Paganoni A. M., Pontiggia L., *Laboratorio di statistica con Excel*, Pearson Education, Milano, 2007
- [3] Romano G. *La complessità dell'Universo*, Gremese, Roma, 2007
- [4] Bergamaschini M., Magazzini P., Mazzoni L., *L'indagine del mondo fisico*, Carlo Signorelli Editore, Milano, 2001, volume F
- [5] Sito Internet dal quale è possibile reperire Essential Regression:
<http://www.geocities.com/SiliconValley/Network/1032>
- [6] Sito Internet da cui si può reperire un utile set di dati per l'analisi statistica:
<http://lib.stat.cmu.edu/DASL/Stories/Hubble'sConstant.html>
- [7] <http://arxiv.org/abs/astro-ph/9604143>
- [8] http://antwrp.gsfc.nasa.gov/diamond_jubilee/d_1996/hub_1929.html