Esercitazione del 15 novembre: Simulazioni Montecarlo

Mauro Brunato, mauro.brunato@unitn.it, versione 1.0

Introduzione

In questa esercitazione confronteremo la distribuzione empirica della somma del lancio di due dadi a sei facce con quella teorica. Realizzeremo inoltre un semplice metodo per il calcolo approssimato di Pi greco.

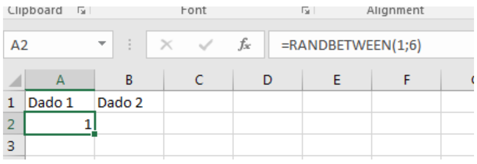
Somma di due dadi

Vogliamo simulare il lancio di due dadi, ripetuto per mille volte. Realizzeremo dunque una tabella con mille righe di due valori casuali compresi fra 1 e 6.

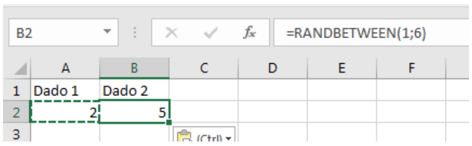
Tabella dei lanci

Iniziamo da un foglio vuoto, e scriviamo le intestazioni "Dado 1" e "Dado 2" nelle celle A1 e B1 del foglio di lavoro. In casella A2 scriviamo la seguente formula, che genera un numero intero casuale fra 1 e 6:

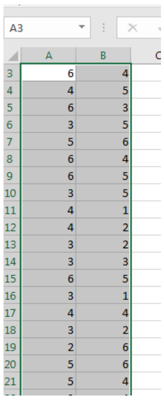
`=RANDBETWEEN(1;6)` (*versione inglese*)
`=CASUALE.TRA(1;6)` (*versione italiana*)



Copiamo (clic su A2, poi Ctrl-C) e incolliamo la formula nella casella B2 (clic su B2, poi Ctrl-V); osserviamo che a ogni ricalcolo i valori generati cambiano:



Vogliamo ora incollare la stessa formula nelle 999 righe successive delle colonne A e B; il metodo più rapido per selezionare un intervallo esteso consiste nello scriverlo nel controllo di selezione (la casella in alto a sinistra della figura precedente); scriviamo dunque "A3:B1001" e premiamo invio. Le caselle indicate risulteranno selezionate, e la pressione di Ctrl-V incollerà la formula su tutte le celle:



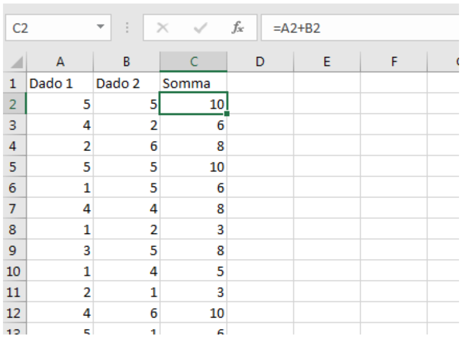
	A	B
3	6	4
4	4	5
5	6	3
6	3	5
7	5	6
8	6	4
9	6	5
10	3	5
11	4	1
12	4	2
13	3	2
14	3	3
15	6	5
16	3	1
17	4	4
18	3	2
19	2	6
20	5	6
21	5	4

Facendo scorrere il foglio verso il basso possiamo verificare che la tabella contiene ora 1000 simulazioni di lanci di due dadi.

Il passo successivo è il calcolo della somma; scriviamo dunque “Somma” nella cella C1, e la formula

`=A2+B2`

nella cella C2. Copiamo poi la formula nelle 999 caselle successive facendo clic su C2, poi Ctrl-C, inserendo l’intervallo C3:C1001 nella casella di selezione, premendo Invio, poi Ctrl-V.



	A	B	C	D	E	F
1	Dado 1	Dado 2	Somma			
2	5	5	10			
3	4	2	6			
4	2	6	8			
5	5	5	10			
6	1	5	6			
7	4	4	8			
8	1	2	3			
9	3	5	8			
10	1	4	5			
11	2	1	3			
12	4	6	10			

Ora la tabella è completa. La colonna C (“Somma”) contiene una sequenza di valori compresi fra 2 e 12.

Conteggio dei risultati

Vogliamo ora realizzare una tabella riassuntiva che conta il numero di esiti per ciascuno dei valori possibili. Sappiamo già come farlo utilizzando le tabelle pivot, ma ora effettueremo il conteggio utilizzando esplicitamente alcune funzioni di Excel.

Lasciamo un po’ di spazio, e scriviamo “Risultato” nella cella F1. Nelle celle da F2 a F13 scriviamo i numeri da 1 a 12.

Vogliamo ora riportare in colonna G il numero di esiti, fra i nostri 1000 esperimenti, corrispondenti ciascun possibile risultato della colonna F. Scriviamo “Conteggio” in G1. Utilizziamo poi la funzione COUNTIF (CONTA.SE), scrivendo in G2 la seguente formula:

`=COUNTIF(C$2:C$1001;F2)` (*versione inglese*)
`=CONTA.SE(C$2:C$1001;F2)` (*versione italiana*)

Il significato della formula è “Conta quante delle celle nell’intervallo C2:C1001 hanno lo stesso valore della cella F2”. Notiamo che i riferimenti di riga all’intervallo C2:C1001 sono “assoluti” ovvero bloccati da un “$”. Nella copia alle celle successive, infatti, vorremo mantenere fisso l’intervallo, che altrimenti avanzerebbe di riga in riga.

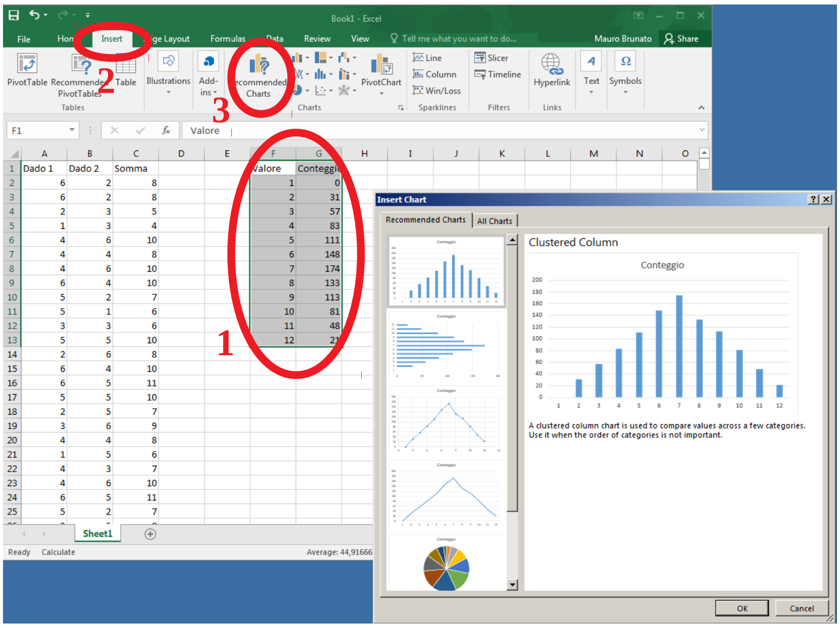
SUM | =COUNTIF(C$2:C$1001;F2)

	A	B	C	D	E	F	G
1	Dado 1	Dado 2	Somma			Valore	Conteggio
2	4	5	9			1	IF(C$2:C$
3	3	4	7			2	
4	6	5	11			3	
5	5	2	7			4	
6	3	2	5			5	
7	5	1	6			6	
8	1	4	5			7	
9	3	2	5			8	
10	6	1	7			9	
11	3	2	5			10	
12	3	3	6			11	
13	5	6	11			12	
14	3	2	5				

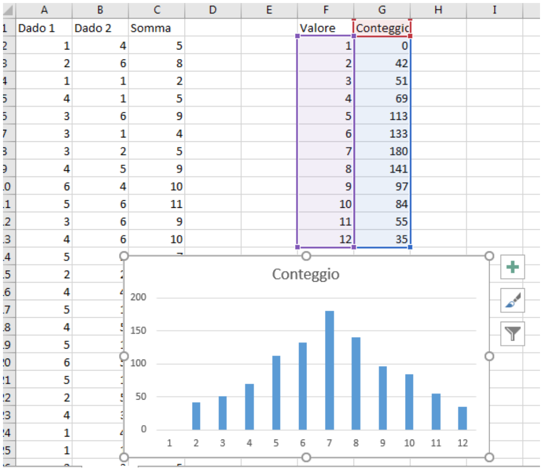
Notiamo che la formula appena inserita vale 0; infatti, non esistono righe il cui esito è 1. Copiamo ora la formula nelle righe successive (clic su G2, Ctrl-C, selezionare da G3 a G13, Ctrl-V):

	A	B	C	D	E	F	G
1	Dado 1	Dado 2	Somma			Valore	Conteggio
2	6	2	8			1	0
3	6	2	8			2	31
4	2	3	5			3	57
5	1	3	4			4	83
6	4	6	10			5	111
7	4	4	8			6	148
8	4	6	10			7	174
9	6	4	10			8	133
10	5	2	7			9	113
11	5	1	6			10	81
12	3	3	6			11	48
13	5	5	10			12	21
14	2	6	8				

Possiamo ora creare un grafico a barre dei conteggi. Selezioniamo le celle da F1 a G13 (clic su F1 e trascinare fino a G13), poi inseriamo un grafico selezionando il nastro “Insert” (“Inserisci”) e il tasto “Recommended chart”; si apre una finestra di dialogo che suggerisce correttamente il tipo di grafico da noi desiderato:



Accettiamo il grafico suggerito e disponiamolo sotto la tabella dei conteggi:

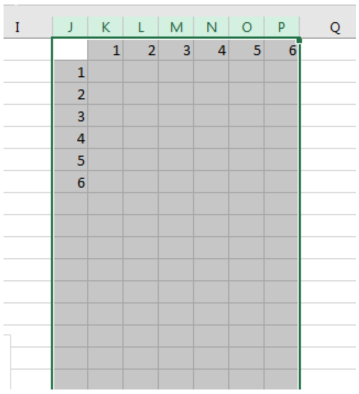


Confronto fra la distribuzione empirica e quella teorica

Vogliamo ora calcolare, sempre utilizzando Excel, quali sono i valori attesi (teorici) dei conteggi.

Come prima cosa, realizziamo una tabella che riporta, per ogni possibile combinazione del valore dei due dadi, il valore della loro somma (una “tabellina delle addizioni” dei numeri da 1 a 6).

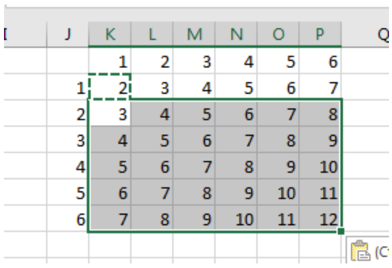
Inseriamo i numeri da 1 a 6 nelle celle da J2 a J7, e nelle celle da K1 a P1. Dato che la tabella dovrà contenere numeri piccoli, possiamo anche selezionare le colonne da J a P (clic sull’intestazione J e trascinamento fino a P), poi ridimensioniamo la colonna J (trascinando il bordo destro dell’intestazione); le altre colonne selezionate si ridimensioneranno anch’esse:



Vogliamo ora che ogni casella della matrice contenga la somma dei due valori ai bordi (colonna J e riga 1). Inseriamo in K2 la seguente formula:

```
=$J2+K$1
```

Notiamo che la formula fissa il rigerimento alla colonna J del primo addendo e alla riga 1 del secondo. Copiamo ora la formula in tutta la tabella (clic in K2, Ctrl-C, selezionare L2:P2, Ctrl-V, selezionare K3:P7, Ctrl-V).



Ogni cella di questa matrice (intervallo K2:P7) rappresenta una delle 36 possibili combinazioni equiprobabili del lancio di due dadi (da “1,1” a “6,6”) con la corrispondente somma. Osserviamo che esiste solo una combinazione il cui esito è 2 (entrambi i dadi danno 1), ci sono due combinazioni il cui risultato è 2, sei il cui risultato è 7, e così via. La probabilità che un lancio dia somma s è data da

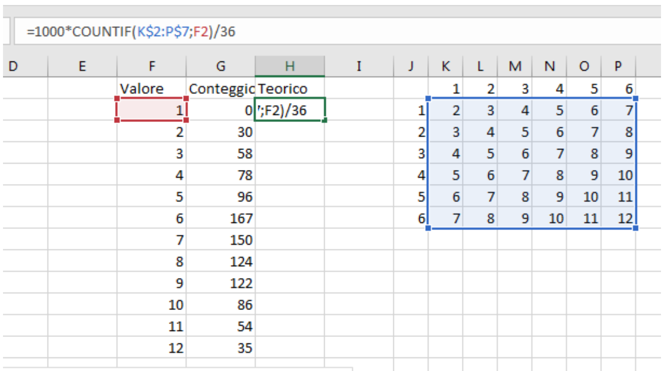
$$\Pr(s) = \frac{\text{Numero di combinazioni la cui somma è } s}{36},$$

quindi il numero atteso di eventi è la probabilità moltiplicata per il numero di lanci (nel nostro caso 1000).

Aggiungiamo una colonna alla tabella dei conteggi scrivendo l'intestazione "Teorico" nella cella H1. Nella cella H2 inseriamo la seguente formula:

`=1000*COUNTIF(K$2:P$7;F2)/36` (*versione inglese*)
`=1000*CONTA.SE(K$2:P$7;F2)/36` (*versione italiana*).

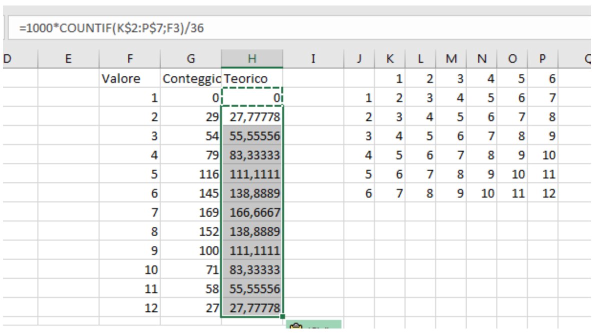
La formula calcola la probabilità vista sopra per il valore di s contenuto nella cella F2, e la moltiplica per il numero di esperimenti. Osserviamo che i numeri di riga che fanno riferimento alla matrice K2:P7 sono assoluti per impedire che la copia della formula nelle celle sottostanti li modifichi.



=1000*COUNTIF(K$2:P$7;F2)/36

F	G	H	I	J	K	L	M	N	O	P
Valore	Conteggic	Teorico			1	2	3	4	5	6
1	0	';F2)/36		1	2	3	4	5	6	7
2	30			2	3	4	5	6	7	8
3	58			3	4	5	6	7	8	9
4	78			4	5	6	7	8	9	10
5	96			5	6	7	8	9	10	11
6	167			6	7	8	9	10	11	12
7	150									
8	124									
9	122									
10	86									
11	54									
12	35									

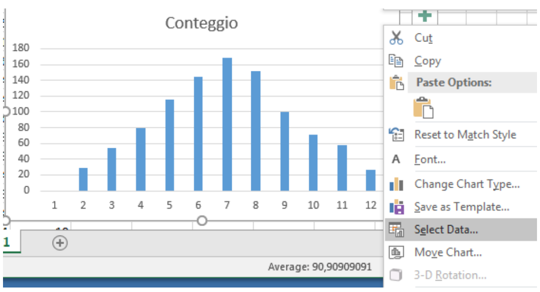
Copiamo ora la formula da H2 a H3:H13, ottenendo i risultati attesi:



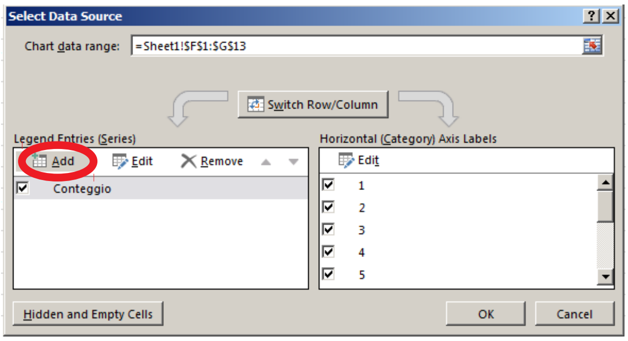
=1000*COUNTIF(K$2:P$7;F3)/36

F	G	H	I	J	K	L	M	N	O	P
Valore	Conteggic	Teorico			1	2	3	4	5	6
1	0	0		1	2	3	4	5	6	7
2	29	27,77778		2	3	4	5	6	7	8
3	54	55,55556		3	4	5	6	7	8	9
4	79	83,33333		4	5	6	7	8	9	10
5	116	111,1111		5	6	7	8	9	10	11
6	145	138,8889		6	7	8	9	10	11	12
7	169	166,6667								
8	152	138,8889								
9	100	111,1111								
10	71	83,33333								
11	58	55,55556								
12	27	27,77778								

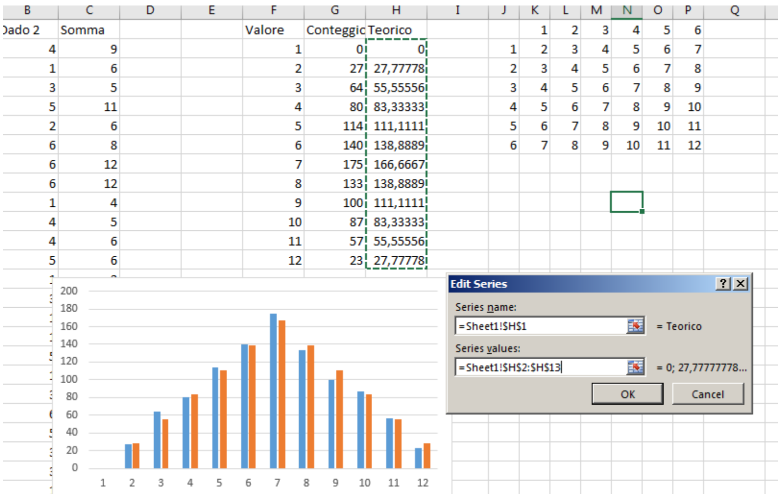
Aggiungiamo la colonna all'istogramma. Facciamo clic nel grafico con il tasto destro e selezioniamo "Select data" ("Seleziona dati"):



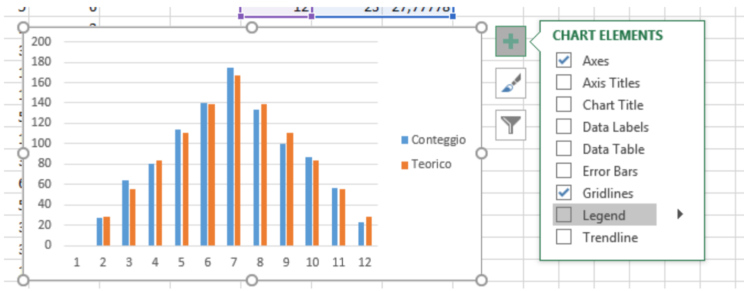
Si apre una finestra di dialogo che ci permette di inserire una nuova serie di valori nel grafico esistente:



Premiamo “Add” (“Aggiungi”) nella finestra delle serie; nella nuova finestra di dialogo, selezioniamo la casella “Series name” (“Nome della serie”) e facciamo clic sulla cella H1; poi selezioniamo la casella “Series values” (“valori della serie”) e selezioniamo l’intervallo H2:H13 sul foglio. A questo punto, l’istogramma mostra due serie di barre.



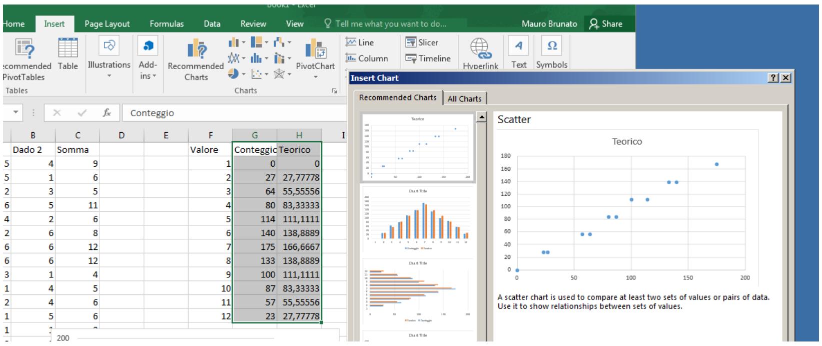
Accettiamo il tutto premendo OK nei due dialoghi. Infine, aggiungiamo una legenda che ci permetta di distinguere la seria empirica da quella teorica. Facciamo clic sul grafico, poi sul pulsante “+” sul lato destro del grafico. Si apre un menu sul quale potremo selezionare “Legend” (“legenda”):



Concordanza fra la serie empirica e quella teorica

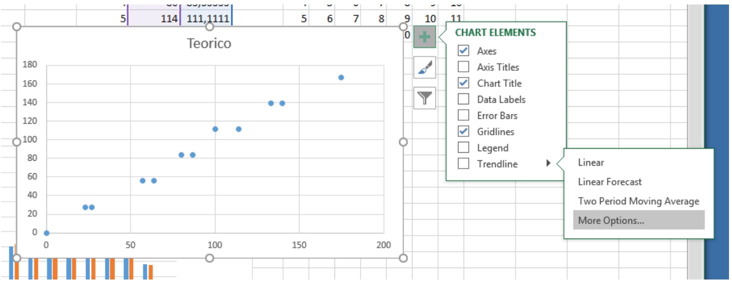
Osserviamo che i due istogrammi sono abbastanza simili (quello arancione, teorico, è simmetrico), ma che vi sono alcune differenze. Ad ogni ricalcolo, che possiamo ottenere andando a modificare una qualunque cella inutilizzata del foglio, la distribuzione empirica viene ricalcolata e risulta leggermente diversa.

Per studiare la concordanza fra il conteggio empirico e teorico, creiamo un diagramma di dispersione fra le due serie di dati. Selezioniamo l’intervallo G2:H13, andiamo al nastro “Insert” (“Inserisci”) e premiamo “Recommended charts” (“Grafici consigliati”):

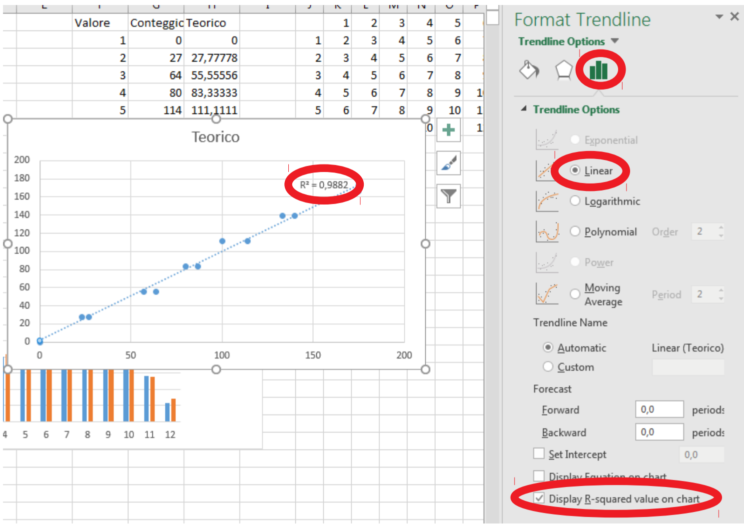


Vediamo che Excel suggerisce automaticamente il grafico da noi desiderato: un diagramma di dispersione fra la serie empirica e quella teorica. Premiamo OK per inserire il nuovo grafico nel foglio di lavoro.

Se la concordanza fosse perfetta, i punti sarebbero tutti perfettamente allineati. Per studiare meglio l'allineamento, aggiungiamo una linea di tendenza. Facciamo clic sul grafico, poi sul pulsante "+" sul lato destro, poi sulla freccia a lato della voce "Trendline" ("Linea di tendenza") e alla voce "More options" ("Altre opzioni"):

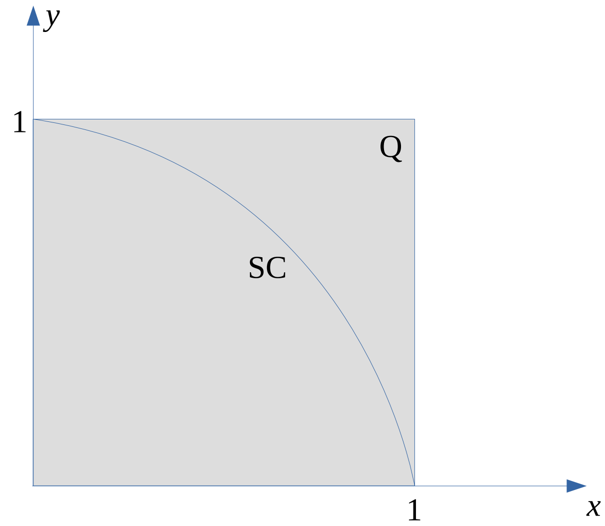


Nella colonna di destra compariranno varie opzioni di configurazione della linea di tendenza. Selezioniamo le opzioni di struttura facendo clic sull'icona del grafico, poi selezioniamo la funzione "Linear" ("Lineare"), e attiviamo la visualizzazione del coefficiente di regressione R^2. Nel grafico, accanto alla retta di regressione, compare un valore (compreso fra 0 e 1) la cui vicinanza a 1 rappresenta la bontà dell'allineamento:



Stima del valore di pi greco

Consideriamo la seguente figura composta da un settore circolare retto SC inscritto in un quadrato di lato unitario Q:



L'area del quadrato è ovviamente $A_Q = 1$, mentre il settore circolare ha un'area pari a un quarto del cerchio di raggio 1: $A_{SC} = \frac{\pi}{4}$.

Supponiamo di generare un punto a caso all'interno di Q. La probabilità che il punto si trovi in SC è intuitivamente pari al rapporto fra le aree delle due figure:

$$\Pr((x, y) \in \mathrm{SC}) = \frac{A_{SC}}{A_Q} = \frac{\pi}{4}.$$

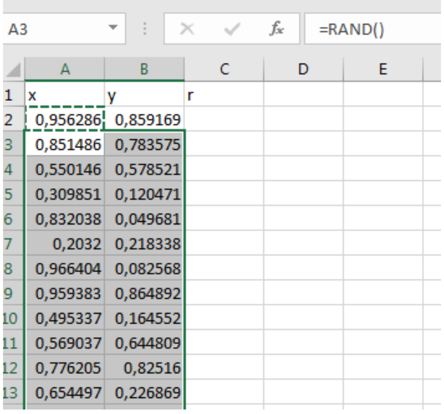
Il nostro obiettivo è dunque la generazione di mille punti (x, y) con le coordinate comprese fra 0 e 1 e contare quale frazione si trova nel settore circolare, ovvero ha distanza minore o uguale a 1 dall'origine:

$$(x, y) \in \mathrm{SC} \Leftrightarrow \sqrt{x^2 + y^2} \leqslant 1.$$

Questo conteggio fornirà una stima di $\frac{\pi}{4}$, quindi sarà sufficiente moltiplicarla per 4 per ottenere una stima di π.

Creiamo dunque un nuovo foglio di lavoro e scriviamo le intestazioni "x", "y" e "r" nelle celle A1, B1 e C1. Nella cella A2 inseriamo la seguente formula, che genera un valore casuale distribuito uniformemente nell'intervallo [0,1]:

`=RAND()`	(*versione inglese*)
`=CASUALE()`	(*versione italiana*)



A3 =RAND()

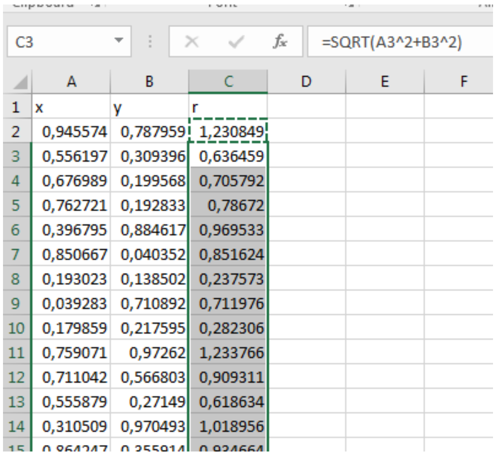
	A	B	C	D	E
1	x	y	r		
2	0,956286	0,859169			
3	0,851486	0,783575			
4	0,550146	0,578521			
5	0,309851	0,120471			
6	0,832038	0,049681			
7	0,2032	0,218338			
8	0,966404	0,082568			
9	0,959383	0,864892			
10	0,495337	0,164552			
11	0,569037	0,644809			
12	0,776205	0,82516			
13	0,654497	0,226869			

Copiamo la formula anche nella cella B2. Poi selezioniamo l'intervallo A3:B1001 e incolliamo la formula in tutte le celle:

Ora inseriamo la formula della distanza dall'origine in C2:

`=SQRT(A2^2+B2^2)` (*versione inglese*)
`=RADQ(A2^2+B2^2)` (*versione italiana*)

Copiamo la formula da C2 all'intervallo C3:C1001, così tutti i punti saranno corredati dalla loro distanza dall'origine.



C3 =SQRT(A3^2+B3^2)

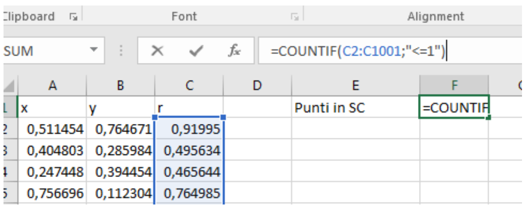
	A	B	C	D	E	F
1	x	y	r			
2	0,945574	0,787959	1,230849			
3	0,556197	0,309396	0,636459			
4	0,676989	0,199568	0,705792			
5	0,762721	0,192833	0,78672			
6	0,396795	0,884617	0,969533			
7	0,850667	0,040352	0,851624			
8	0,193023	0,138502	0,237573			
9	0,039283	0,710892	0,711976			
10	0,179859	0,217595	0,282306			
11	0,759071	0,97262	1,233766			
12	0,711042	0,566803	0,909311			
13	0,555879	0,27149	0,618634			
14	0,310509	0,970493	1,018956			

Osserviamo che alcuni punti hanno una distanza minore di 1, altri maggiore.

Ora inseriamo il conteggio dei punti di distanza minore o uguale a 1. Nella casella E1 scriviamo "Punti in SC" (se necessario allarghiamo la colonna E). In F1 scriviamo la formula che conta questi punti:

`=COUNTIF(C2:C1001;"<=1")` (*versione inglese*)
`=CONTA.SE(C2:C1001;"<=1")` (*versione italiana*)

Osserviamo che il criterio di conteggio è rappresentato da una stringa fra virgolette. Il significato della formula è "Conta tutti i valori nell'intervallo C2:C1001 che sono minori o uguali a 1".

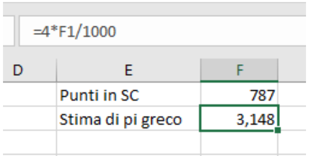


SUM =COUNTIF(C2:C1001;"<=1")

	A	B	C	D	E	F
1	x	y	r		Punti in SC	=COUNTIF
2	0,511454	0,764671	0,91995			
3	0,404803	0,285984	0,495634			
4	0,247448	0,394454	0,465644			
5	0,756696	0,112304	0,764985			

A questo punto, possiamo procedere con la stima di Pi greco, dividendo il conteggio per 1000 (il numero di punti generati) e moltiplicando per 4. Nella casella E2 scriviamo “Stima di pi greco”, mentre in F2 inseriamo la formula

```
=4*F1/1000
```



Osserviamo che la stima può variare di esperimento in esperimento, e ci basterà modificare il contenuto di qualunque cella non utilizzata per forzare un ricalcolo. Per valutare la bontà dell’approssimazione, inseriamo il valore corretto di pi greco così come viene fornito da Excel scrivendo “Valore esatto” nella cella E3 e la formula

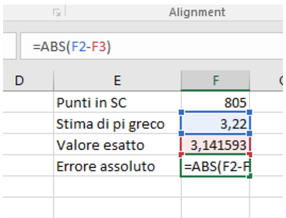
```
=PI()
```

in F3 (PI è una funzione senza argomenti che restituisce sempre il valore di Pi greco preciso alla sedicesima cifra decimale).

Calcoliamo l’errore assoluto alla riga successiva, inserendo “Errore assoluto” nella cella E4 e la formula

```
=ABS(F2-F3)        (versione inglese)
=ASS(F2-F3)        (versione italiana)
```

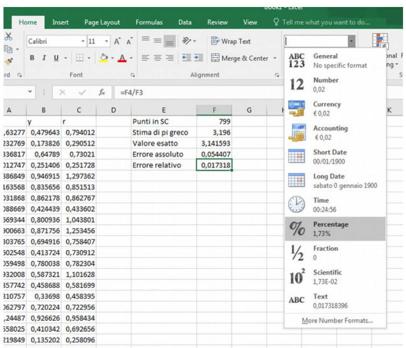
nella cella F4:



Infine, possiamo calcolare l’errore relativo dividendo l’errore assoluto per il valore esatto. Nella cella E5 scriviamo “Errore relativo”, e in F5 inseriamo la formula

```
=F4/F3
```

Infine, con la cella F5 selezionata, possiamo modificarne il formato per rappresentare l’errore in percentuale andando al nastro “Home” e selezionando “Percent” (“Percentuale”) nella casella dei formati numerici:



Otterremo quindi la seguente tabella sperimentale (ovviamente, i valori esatti differiranno di esperimento in esperimento):

A	B	C	D	E	F	G
x	y	r		Punti in SC	799	
0,63277	0,479643	0,794012		Stima di pi greco	3,196	
0,232769	0,173826	0,290512		Valore esatto	3,141593	
0,336817	0,64789	0,73021		Errore assoluto	0,054407	
0,012747	0,251406	0,251728		Errore relativo	1,73%	
0,886849	0,946915	1,297362				
0,163568	0,835656	0,851513				
0,031868	0,862178	0,862767				
0,088669	0,424439	0,433602				
0,669344	0,800936	1,043801				
0,900663	0,871756	1,253456				
0,303765	0,694916	0,758407				