



UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

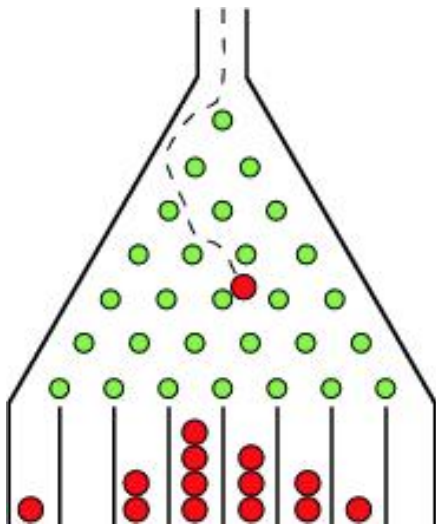
DIPARTIMENTO DI INGEGNERIA AEROSPAZIALE – D.I.A.S.

STATISTICA PER L'INNOVAZIONE

a.a. 2007/2008

QUINCUNX: IL TUBO DI GALTON

Prof. Antonio Lanzotti



A cura di: Ing. Giovanna Matrone

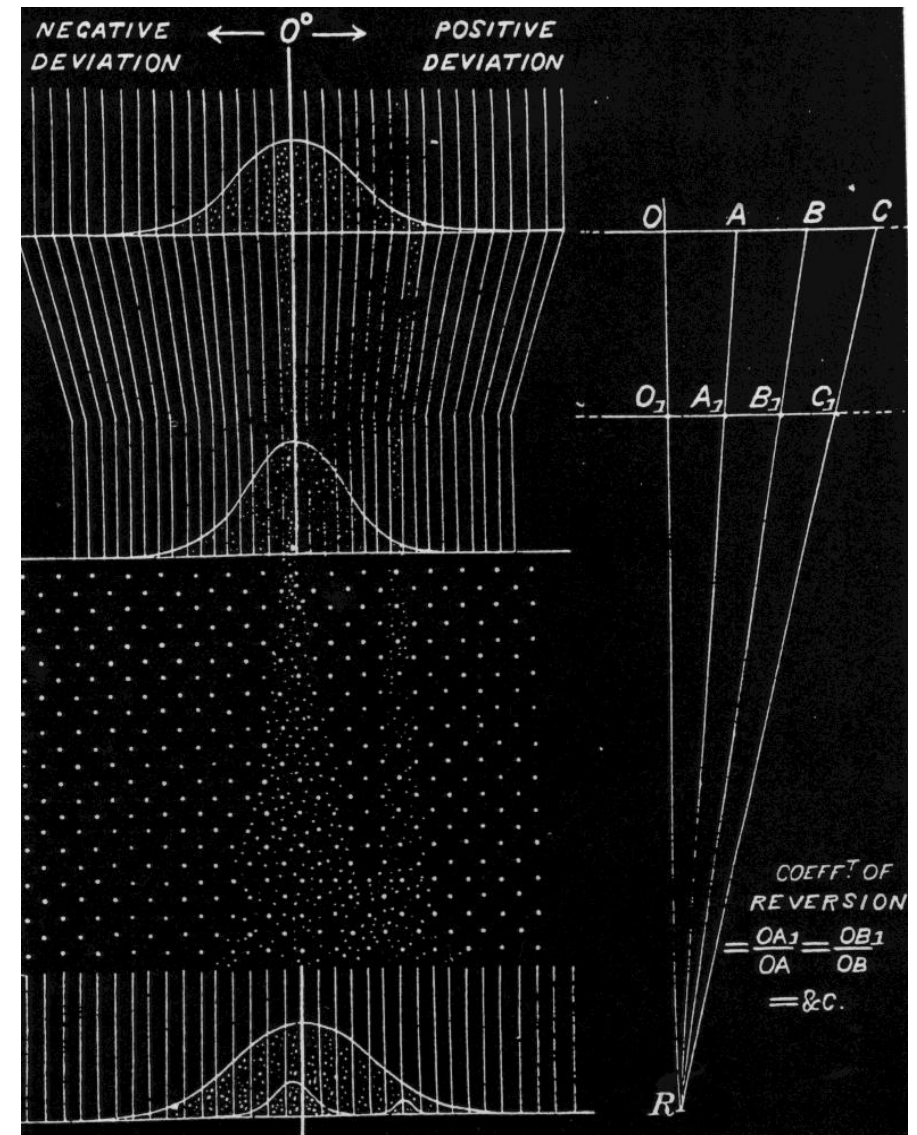
giovanna.matrone@unina.it



Le origini

Il *Quincunx* fu introdotto da Sir *Francis Galton* nel 1877 per i suoi studi di Genetica: egli dimostrò che le deviazioni dalla media di alcune caratteristiche ereditarie (come l'altezza) sono somma di molte “invisibili” cause di disturbo e che la legge della “deviazione genetica” segue la legge della distribuzione Normale.

Il Tubo di Galton costituisce un utile strumento per descrivere la distribuzione binomiale.





Le origini

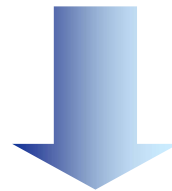
Galton cercò di conciliare
un' evidenza empirica con il teorema del limite centrale



la maggior parte delle misure
fisiche nella popolazione sotto
studio (come l'altezza di un uomo)
si distribuivano normalmente



le cause di disturbo di un oggetto
sono molteplici ma tutte indipendenti
(non esistono fattori dominanti)



QUINCUNX

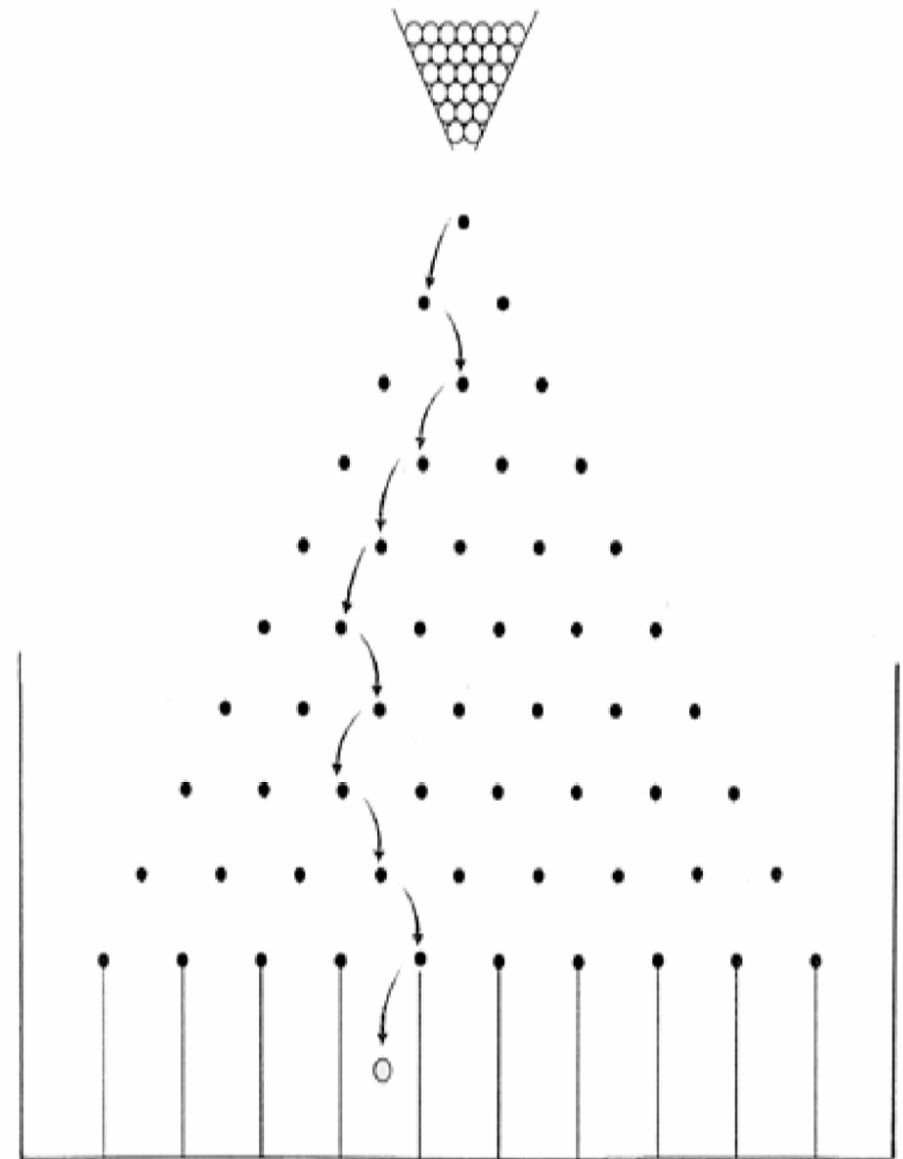


Quincunx

Il dispositivo è caratterizzato da un piano verticale nel quale i punti rossi rappresentano dei chiodi perpendicolari al piano (*pin*), e in basso sono fissate delle celle di raccolta.

L'esperimento consiste nel lasciar cadere delle palline (*biglie*) dal vertice del 'triangolo' formato dai chiodi. Ogni biglia segue un percorso che è casualmente determinato dagli urti contro i *pin* per finire in uno dei compartimenti verticali sottostanti.

Il *quincunx* è simmetrico, con l'asse di simmetria rappresentato da una linea verticale che passa per il chiodo di partenza.



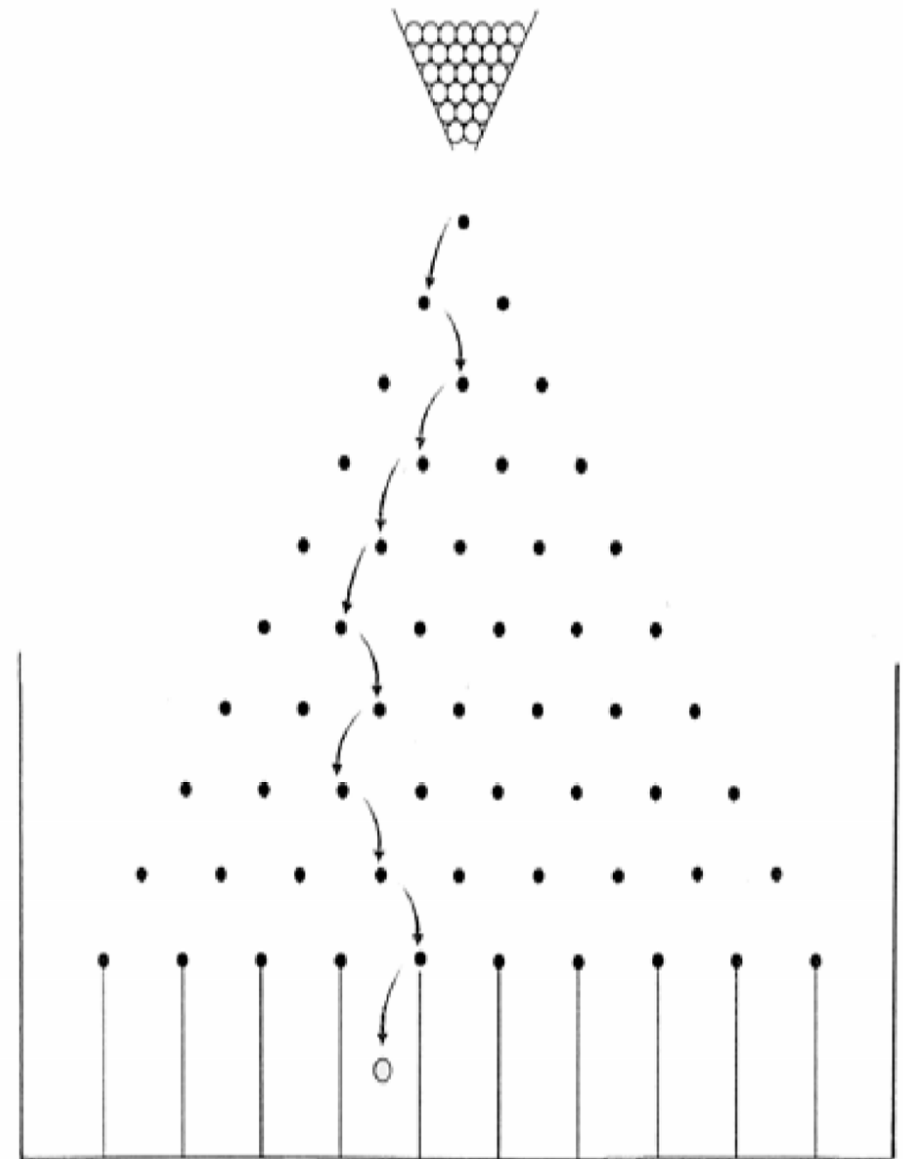


Quincunx

Le premesse dell'esperimento sono che gli spazi fra i chiodi sono di poco maggiori della biglia che cade, e che questa, nel suo tragitto, compia un solo rimbalzo per riga, con la sola possibilità di cadere a destra o a sinistra del chiodo.

Sotto dette ipotesi possiamo affermare che se le righe del sistema sono n , la biglia, per arrivare in qualsiasi buca, dovrà fare n urti contro i chiodi.

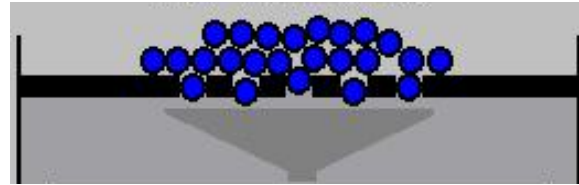
Nell'apparato si può inoltre notare che le celle sono pari al numero delle righe incrementato di uno; questo significa che se le righe sono n , le celle sono $n+1$, fattore molto utile nei calcoli.





Quincunx

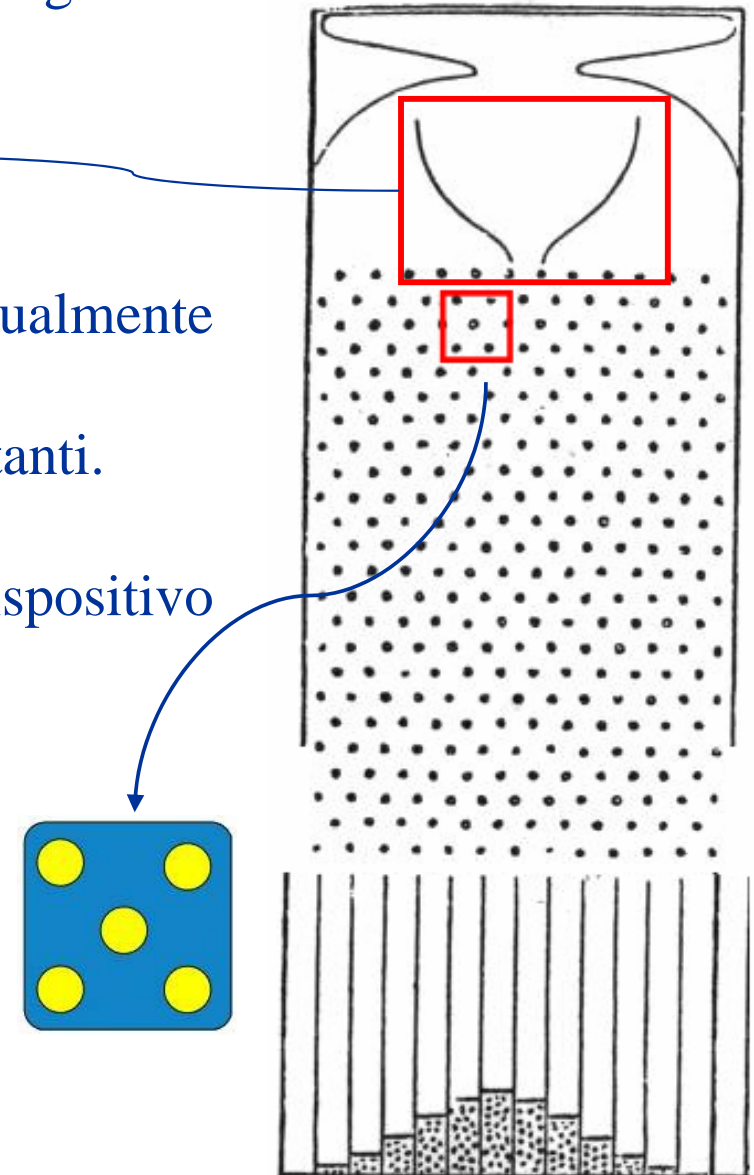
Dall'imbuto viene fatto cadere un certo numero di biglie.



Ogni biglia segue un percorso che è casualmente determinato dagli urti contro gli aghi (*pin*) per finire in uno dei compartimenti verticali sottostanti.

La disposizione delle biglie alla base del dispositivo richiama una curva normale.

Il termine *Quincunx* deriva dalla particolare disposizione a *quinconce* degli aghi, in maniera tale che fra due aghi di una riga venga a trovarsi un solo pin nella riga successiva, ossia come la disposizione del numero cinque espresso sulle facce dei dadi, così definita.





Simulazione

http://cirdis.stat.unipg.it/files/macchina_galton/macchina_galton/esperimento.htm

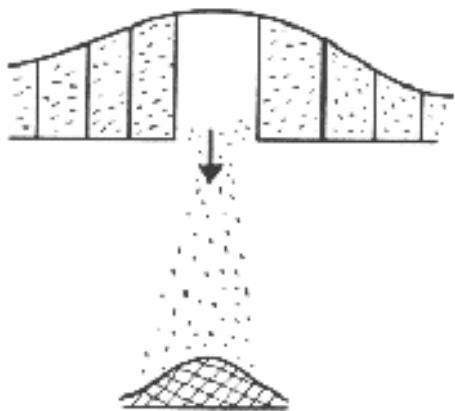
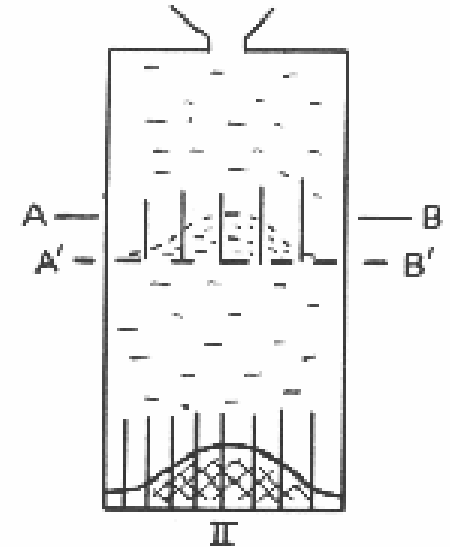


Quincunx

1) Cosa succede se le biglie sono arrestate ad un livello intermedio?



La disposizione delle biglie nel compartimento intermedio richiama ancora una curva normale meno dispersa



2) Cosa succede se le biglie di un singolo compartimento intermedio sono lasciate libere?



Esse formeranno una “collinetta” simile alla curva normale immediatamente sotto il compartimento

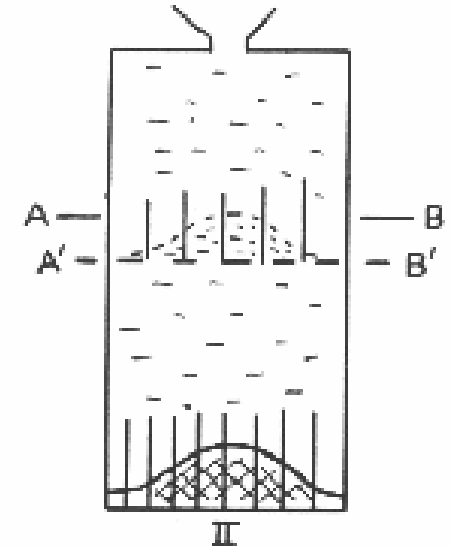


Quincunx

3) Cosa succede se tutte le biglie dei compartimenti intermedi sono lasciate libere?



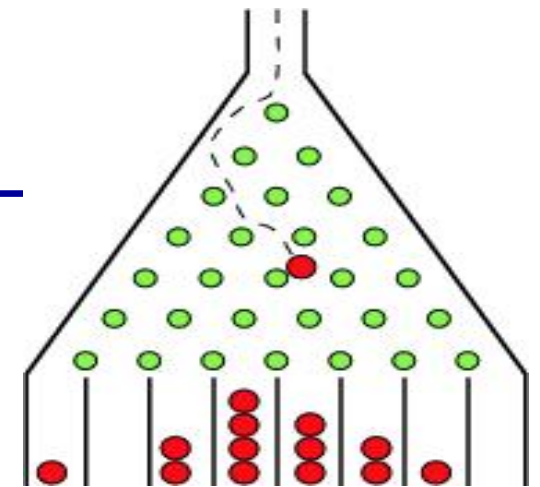
Si formeranno tante piccole “collinette” che aggregate avranno una forma simile alla curva normale ottenuta senza stato intermedio





Considerazioni statistiche

Ad ogni urto, la probabilità che la biglia subisca una deviazione a sinistra (o a destra) è pari al 50%.



Se indichiamo con X_j la v.a. *deviazione* subita dalla biglia alla j -esima riga:

$$X_j \sim \text{Bernoulliana } (p = 0,5) \quad \begin{cases} X_j = 1 \text{ se la biglia viene deviata a destra} \\ X_j = 0 \text{ se la biglia viene deviata a sinistra} \end{cases}$$

Se indichiamo con Y la v.a. *posizione finale* occupata dalla biglia, essa sarà somma di n v.a. Bernoulliane i.i.d.

$$Y = \sum_{j=1}^n X_j \sim \text{Binomiale } (n = \text{righe di pins}, p)$$

Una realizzazione di tale v.a. è il numero di deviazioni a sinistra (ovvero a destra) subite dalla biglia.



Test di adattamento del Chi-Quadrato

Il test del Chi-Quadrato rientra tra quelle tecniche d'inferenza statistica che non richiedono ipotesi a priori sul tipo e sulle caratteristiche della distribuzione (*'distribution free'* o test non parametrici).

Esso consente di valutare se le differenze tra le *frequenze assolute osservate* O_j di elementi appartenenti alla categoria j e quelle *attese* E_j (che ci si aspetta di riscontrare in base alla teoria che assumiamo) siano il frutto di fisiologiche fluttuazioni campionarie.

L'ipotesi nulla è: $H_0 = \{Y \text{ è una v.a. Binomiale di parametro } p\}$

La statistica utilizzata, fondata su N osservazioni, è la seguente:

$$K = \sum_{j=1}^m \frac{(O_j - E_j)^2}{E_j}$$

Per N sufficientemente elevato e sotto l'ipotesi nulla la statistica in esame tende a distribuirsi secondo una χ^2 con $\nu = m - 1$ g.d.l..



Test di adattamento con parametro p noto

Test di adattamento alla Binomiale di parametro $p = 0,5$

Per la distribuzione Binomiale la pmf è:

$$P_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

Essa esprime la probabilità della concomitanza di esattamente y successi su n lanci (indipendentemente dal loro ordine di accadimento).

Nel caso dell'esperimento di Galton:

- n = righe di *pin*
- y = determinazioni di $Y = y_j$
($j = 1, 2, \dots, n$ perché Y somma di n Bernoulliane; d'altra parte, siccome la v.a. Bernoulliana può assumere solo valori '0' e '1', la v.a. Binomiale definita come somma di n v.a. Bernoulliane, può assumere valori da '0' a ' n ').



Test di adattamento con parametro p noto

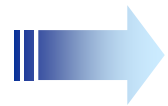
Inoltre la media di Y rappresenterà la frequenza attesa degli elementi che ricadono nella categoria j -esima:

$$E_j \{Y\} = P_Y(y_j) \cdot N$$

Dove N è la dimensione del campione che, nel caso specifico, è rappresentato dal numero totale di biglie raccolte sul fondo del dispositivo di Galton.

ESEMPIO:

$$\begin{cases} Y \sim \text{Bin}(n, p) \\ p = 0,5 \\ n = 9 \\ N = 100 \end{cases}$$

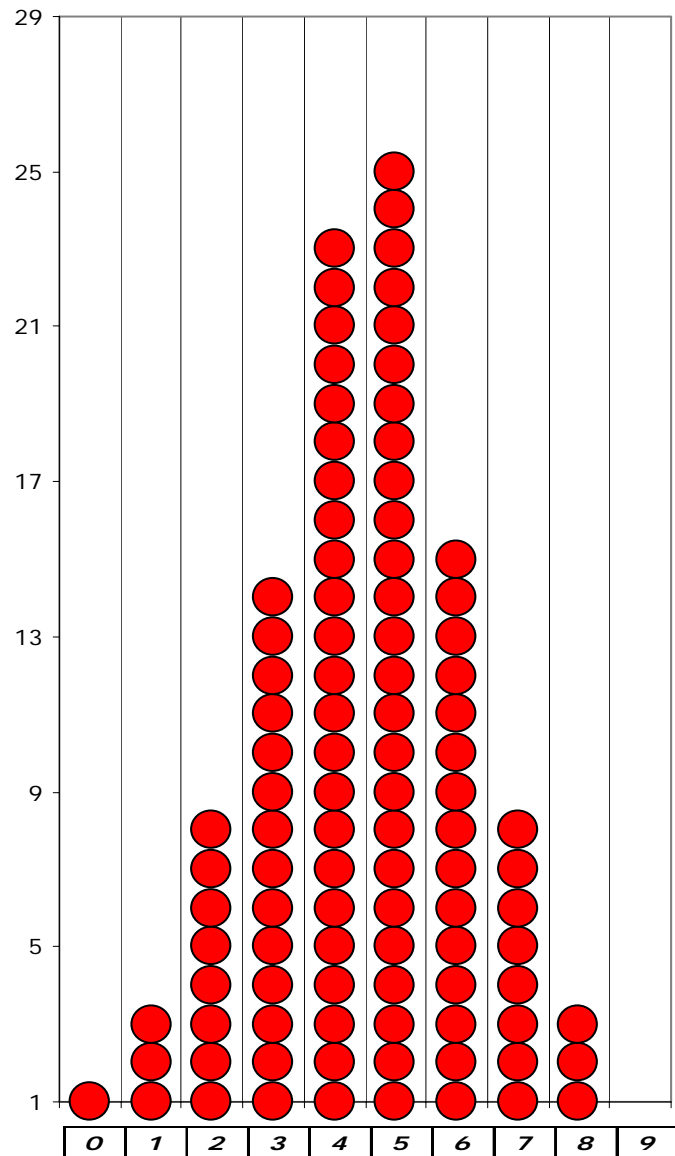


$$P_Y(y_2) = \binom{9}{1} 0,5^1 (1 - 0,5)^8 = 0,0176$$

$$E_2 = 0,0176 \cdot 100 = 1,76$$



Risultati della simulazione del tubo di Galton



j	y_j	O_j
1	0	1
2	1	3
3	2	8
4	3	14
5	4	23
6	5	25
7	6	15
8	7	8
9	8	3
10	9	0

Numero di pallina lanciate: $N = 100$

Numero di file di *pins*: $n = 9$

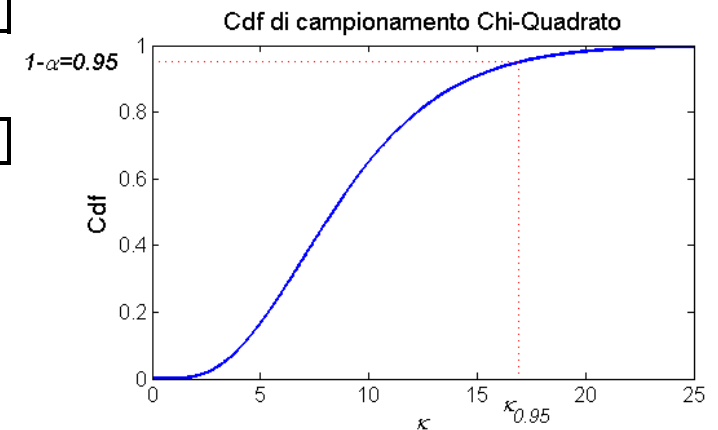
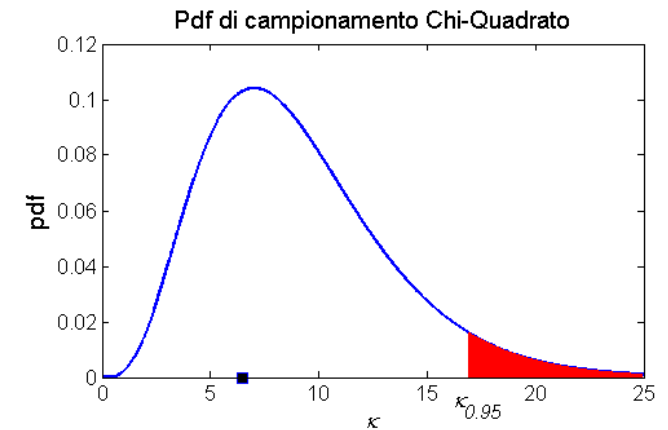
Numero di classi di raccolta: $m = 10$



Test di adattamento con parametro p noto

y_j	O_j	E_j	$O_j - E_j$	$(O_j - E_j)^2$	$((O_j - E_j)^2)/E_j$
0	1	0,20	0,80	0,65	3,32
1	3	1,76	1,24	1,54	0,88
2	8	7,03	0,97	0,94	0,13
3	14	16,41	-2,41	5,79	0,35
4	23	24,61	-1,61	2,59	0,11
5	25	24,61	0,39	0,15	0,01
6	15	16,41	-1,41	1,98	0,12
7	8	7,03	0,97	0,94	0,13
8	3	1,76	1,24	1,54	0,88
9	0	0,20	-0,20	0,04	0,20
Totale	100				6,12

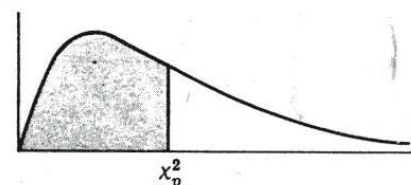
K	6,12	K (0,95; 9)	16,92	p-value	0,73
---	------	-------------	-------	---------	------



Poiché $K < K (0,95;9)$, ovvero $p\text{-value} > \alpha = 0,05$, non è possibile rigettare l'ipotesi nulla H_0 e, pertanto, si conclude che le differenze riscontrate non sono altro che il frutto di fluttuazioni campionarie.



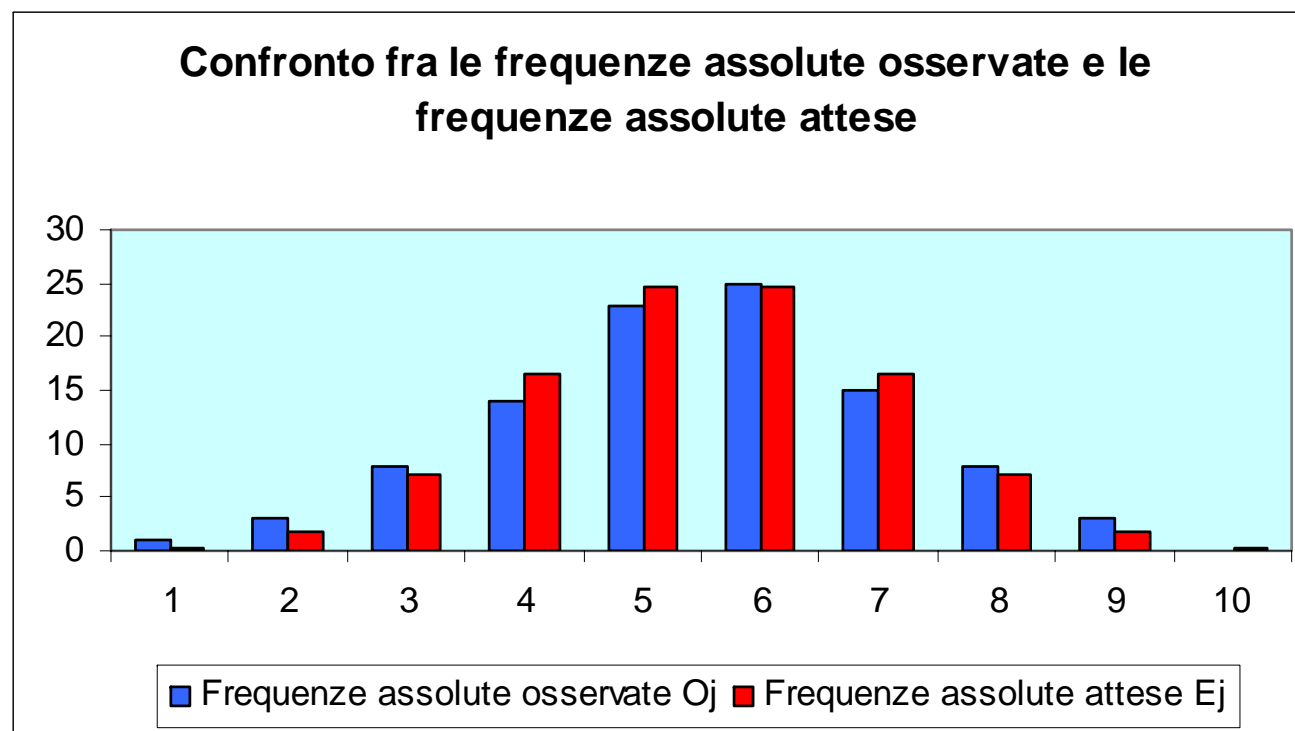
VALORI DEI PERCENTILI
per la
DISTRIBUZIONE DELLA VARIABILE
CASUALE CHI-QUADRATO
con ν gradi di libertà
(area ombreggiata = p)



ν	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$	$\chi^2_{0,50}$	$\chi^2_{0,25}$	$\chi^2_{0,10}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0,102	0,0158	0,0039	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,34	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	23,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7	29,1	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	166,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3



Confronto grafico tra le frequenze assolute



Il risultato del test avvalora con evidenza statistica l'ipotesi fatta riguardo alla natura aleatoria degli scostamenti rilevabili (vedi grafico) tra le frequenze assolute osservate e quelle attese.

È però doveroso ricordare che, a valle del risultato del test, si corre comunque un rischio, seppur molto contenuto, nel non imputare la presenza di tali scostamenti all'effetto di causa deterministica a noi ignota.