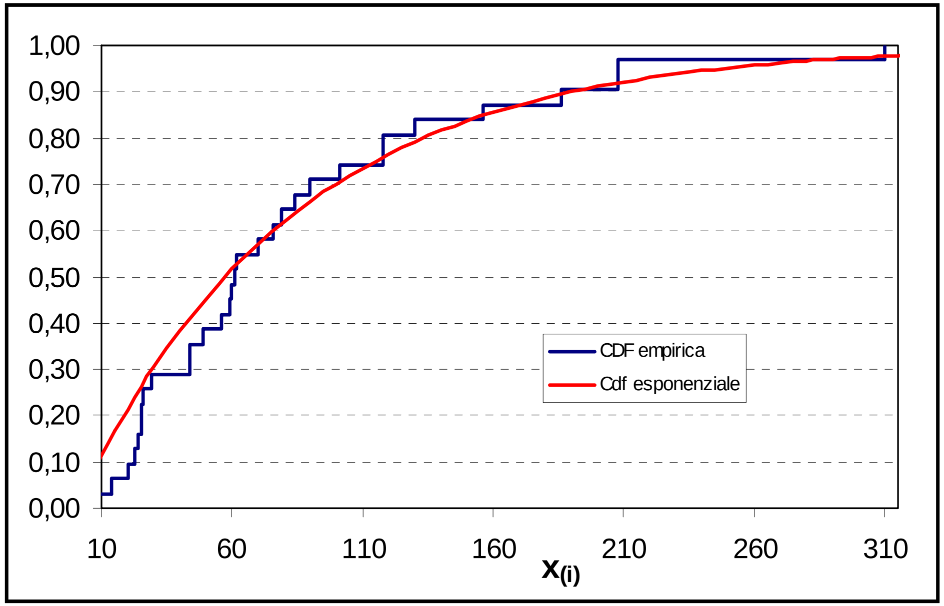
UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

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

STATISTICA PER L'INNOVAZIONE

a.a. 2007/2008




# TEST DI ADATTAMENTO

Prof. Antonio Lanzotti

A cura di: Ing. Antonio Lepore
antonio.lepore@unina.it

# Stima del parametro *p* dai dati sperimentali

Al fine di eseguire il test di adattamento in ipotesi di parametro $p$ non noto, eseguiamo la stima del parametro $p$ della v.a. Binomiale di parametro $n$ direttamente dal set di dati sperimentali.

Prelevato un campione di $N$ determinazioni sperimentali di tale v.a. divise in $m$ classi, se ne calcola la media campionaria secondo la relazione:

$$\bar{Y} = \frac{\sum_{j=1}^{m} y_j \cdot O_j}{N} = \sum_{j=1}^{m} y_j \cdot P_j$$

Essendo $O_j$ la frequenza assoluta associata alla classe $y_j$ ed essendo $P_j$ quella relativa:

$$P_j = O_j / N$$

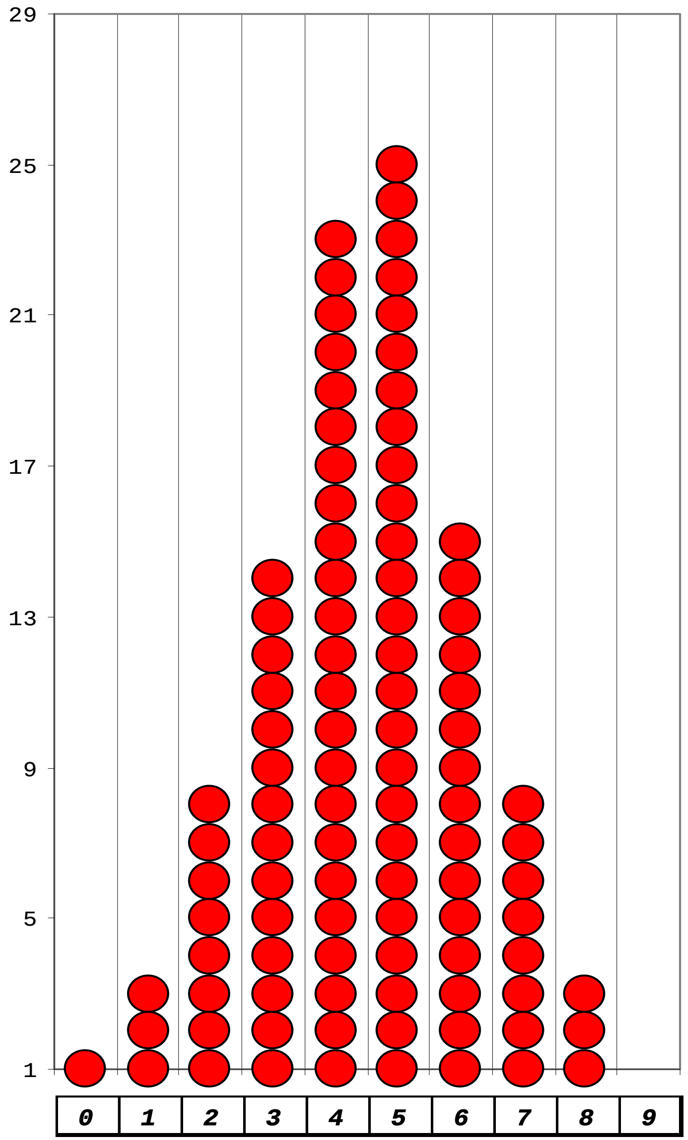
Uguagliando la stima della media alla formulazione della media della v.a. Binomiale (*Metodo dei Momenti*) si trova:

$$\bar{Y} = n \cdot \hat{p} \Rightarrow \hat{p} = \bar{Y} / n$$

# Stima del parametro *p* dai dati sperimentali

| j | $y_j$ | $O_j$ | $P_j$ | $y_j * P_j$ |
|---|---|---|---|---|
| 1 | 0 | 1 | 0,01 | 0 |
| 2 | 1 | 3 | 0,03 | 0,03 |
| 3 | 2 | 8 | 0,08 | 0,16 |
| 4 | 3 | 14 | 0,14 | 0,42 |
| 5 | 4 | 23 | 0,23 | 0,92 |
| 6 | 5 | 25 | 0,25 | 1,25 |
| 7 | 6 | 15 | 0,15 | 0,9 |
| 8 | 7 | 8 | 0,08 | 0,56 |
| 9 | 8 | 3 | 0,03 | 0,24 |
| 10 | 9 | 0 | 0,00 | 0 |
| Totale | | 100 | | 4,48 |

| | |
|---|---|
| Media | 4,48 |
| p | 0,4978 |

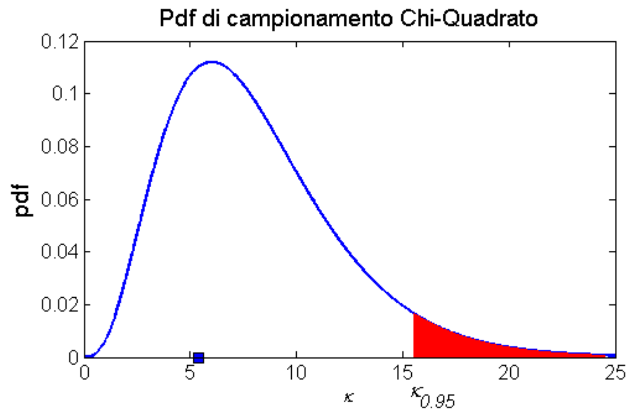


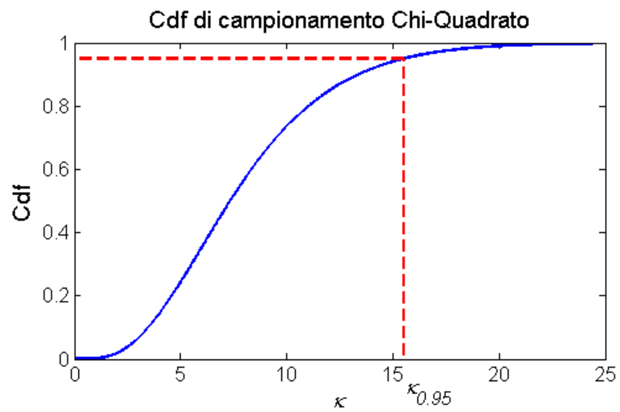
# Test Chi-Quadrato con parametro *p* stimato

| | |
|---|---|
| p | 0,498 |

| j | $y_j$ | $O_j$ | $E_j$ | $O_j$-$E_j$ | $(O_j-E_j)^2$ | $(O_j-E_j)^2/E_j$ |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0,20 | 0,80 | 0,63 | 3,12 |
| 2 | 1 | 3 | 1,81 | 1,19 | 1,41 | 0,78 |
| 3 | 2 | 8 | 7,19 | 0,81 | 0,66 | 0,09 |
| 4 | 3 | 14 | 16,62 | -2,62 | 6,89 | 0,41 |
| 5 | 4 | 23 | 24,72 | -1,72 | 2,95 | 0,12 |
| 6 | 5 | 25 | 24,50 | 0,50 | 0,25 | 0,01 |
| 7 | 6 | 15 | 16,19 | -1,19 | 1,41 | 0,09 |
| 8 | 7 | 8 | 6,88 | 1,12 | 1,26 | 0,18 |
| 9 | 8 | 3 | 1,70 | 1,30 | 1,68 | 0,99 |
| 10 | 9 | 0 | 0,19 | -0,19 | 0,04 | 0,19 |
| | Totale | 100 | | | | 5,98 |

| | | | | | |
|---|---|---|---|---|---|
| K | 5,98 | K (0,95; 8) | 15,51 | p-value | 0,65 |




Poiché ***K<K(0,95;8)*** (ovvero ***p*-value>*α=0,05***) non è possibile rigettare l'ipotesi nulla e pertanto, si conclude che le differenze riscontrate non sono altro che il frutto di fluttuazioni campionarie.

# Test di Kolmogorov Smirnov

Se i gruppi possono essere ordinati con una scala ordinale, il confronto tra distribuzione osservata ed attesa viene realizzato mediante il test di Kolmogorov-Smirnov che si basa sul valore di massima divergenza tra le due distribuzioni cumulate.

Il test si fonda sulla logica secondo la quale, se un campione fosse estratto da una determinata distribuzione teorica, le differenze tra i valori della Cdf ipotizzata e quelli della Cdf campionaria sarebbero di piccola entità.

$$H_0 = \{\underbrace{\text{Cdf}}_{\text{Distribuzione della popolazione}} = \underbrace{F_X(x)}_{\text{Distribuzione teorica}}\}$$

# Test di Kolmogorov Smirnov

La statistica test è:

$$D = \max\left|F_X(x_i) - \hat{F}_X(x_i)\right|$$

dove

$$\hat{F}_X(x_i) = \frac{i}{n}$$

è la CDF empirica stimata dai valori campionari $x_i$
$n$ dimensione del campione
$i$ numero d'ordine crescente dell'osservazione
(in caso di più osservazioni uguali si fa riferimento al n.o. maggiore).

Per $N > 35$ i percentili della CDF di D possono essere valutati con le formule approssimate:

$D = 1.36/\sqrt{n}$ per $\alpha$=0.05

$D = 1.63/\sqrt{n}$ per $\alpha$=0.01

# Test di Kolmogorov Smirnov – Osservazioni

Il test K-S è più potente rispetto al test Chi-Quadrato, soprattutto per elevate dimensioni campionarie

I principali limiti del test K-S:

1. Si può applicare solo a variabili aleatorie continue (non alla Binomiale o alla Poisson, per esempio)

2. L'impiego "classico" della statistica test impone che sia il modello da testare che i suoi parametri devono essere totalmente specificati (e non stimati dai dati)

3. Quando il numero di osservazioni è inferiore a 30, oppure le frequenze attese entro due o più gruppi sono inferiori a 5, si preferisce non utilizzare il test.

# Applicazione

Supponiamo di disporre di alcune misure del tempo di funzionamento (tra due guasti successivi) dell'impianto di condizionamento dell'aria di 13 esemplari del modello di aeromobile Boeing 720 (Lawless, 1982).

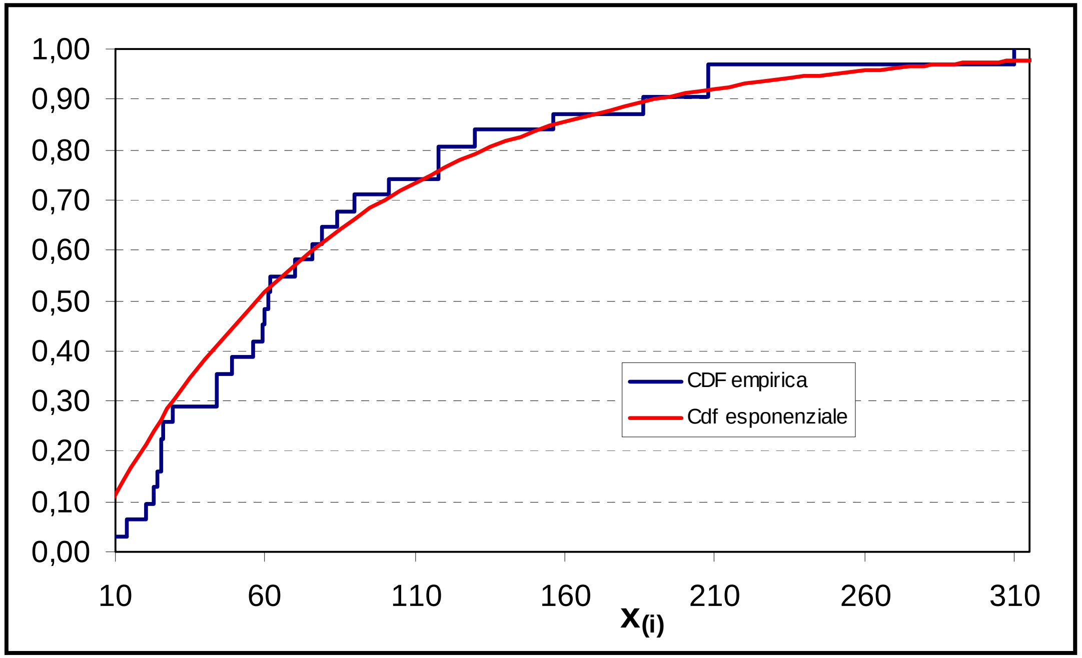
- I 31 dati seguenti sono relativi all'aeromobile n°3 e sono espressi in ore di funzionamento:

{ 90, 10, 60, 186, 61, 49, 14, 24, 56, 20, 79, 84, 44, 59, 29, 118,
25, 118, 25, 156, 310, 76, 26, 44, 23, 62, 130, 208, 70, 101, 208 }

Ipotizzando che il *modello Esponenziale* di parametro $\theta$=82,74 $h^{-1}$ ben interpreti i dati, vogliamo verificarne la bontà di adattamento mediante il test di *Kolmogorov-Smirnov*.

# Test di Kolmogorov Smirnov - Applicazione

| n.o. | $x_{(i)}$ | i/n | CDF exp(1/θ) | 1/θ | D | $D_{max}$ |
|---|---|---|---|---|---|---|
| 1 | 10 | 0,032 | 0,114 | 0,01209 | 0,082 | **0,118** |
| 2 | 14 | 0,065 | 0,156 | | 0,091 | |
| 3 | 20 | 0,097 | 0,215 | | 0,118 | |
| 4 | 23 | 0,129 | 0,243 | | 0,114 | |
| 5 | 24 | 0,161 | 0,252 | | 0,090 | |
| 6 | 25 | | | | | |
| 7 | 25 | 0,226 | 0,261 | | 0,035 | |
| 8 | 26 | 0,258 | 0,270 | | 0,012 | |
| 9 | 29 | 0,290 | 0,296 | | 0,005 | |
| 10 | 44 | | | | | |
| 11 | 44 | 0,355 | 0,412 | | 0,058 | |
| 12 | 49 | 0,387 | 0,447 | | 0,060 | |
| 13 | 56 | 0,419 | 0,492 | | 0,072 | |
| 14 | 59 | 0,452 | 0,510 | | 0,058 | |
| 15 | 60 | 0,484 | 0,516 | | 0,032 | |
| 16 | 61 | 0,516 | 0,522 | | 0,005 | |
| 17 | 62 | 0,548 | 0,527 | | 0,021 | |
| 18 | 70 | 0,581 | 0,571 | | 0,010 | |
| 19 | 76 | 0,613 | 0,601 | | 0,012 | |
| 20 | 79 | 0,645 | 0,615 | | 0,030 | |
| 21 | 84 | 0,677 | 0,638 | | 0,040 | |
| 22 | 90 | 0,710 | 0,663 | | 0,047 | |
| 23 | 101 | 0,742 | 0,705 | | 0,037 | |
| 24 | 118 | | | | | |
| 25 | 118 | 0,806 | 0,760 | | 0,047 | |
| 26 | 130 | 0,839 | 0,792 | | 0,047 | |
| 27 | 156 | 0,871 | 0,848 | | 0,023 | |
| 28 | 186 | 0,903 | 0,894 | | 0,009 | |
| 29 | 208 | | | | | |
| 30 | 208 | 0,968 | 0,919 | | 0,049 | |
| 31 | 310 | 1,000 | 0,976 | | 0,024 | |

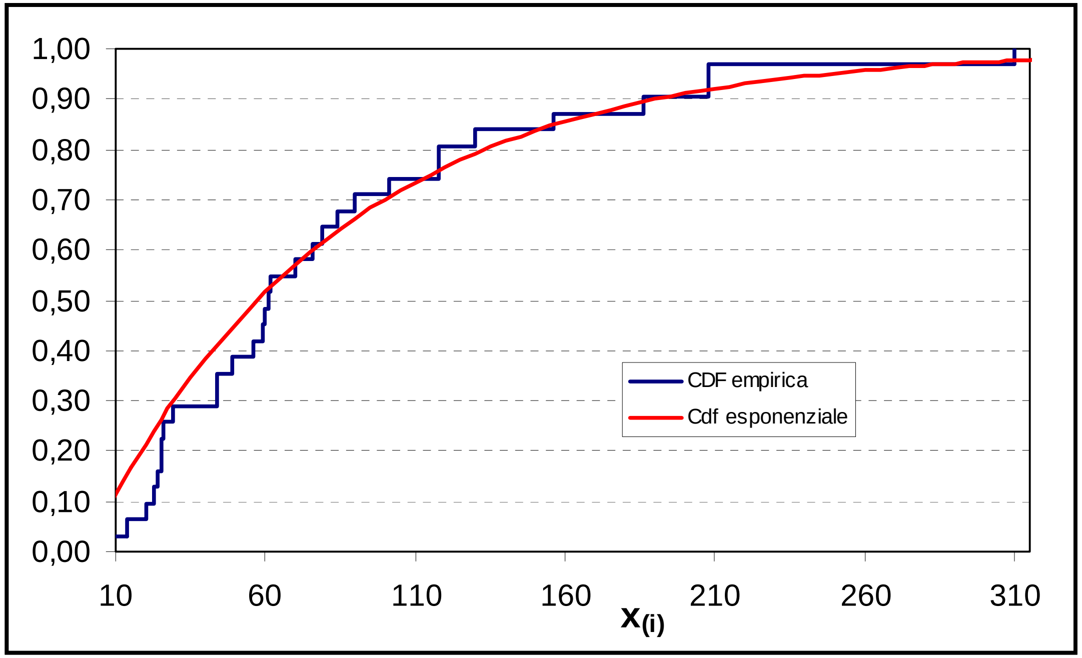



**Cdf ipotizzata:** $$F_X(x_i) = 1 - e^{-\frac{x_i}{\theta}}$$

**Cdf empirica:** $$\hat{F}_X(x) = \begin{cases} 0 & x < x_1 \\ \dfrac{i + O_i - 1}{n} & x_{(i)} \le x < x_{(i+1)} \\ 1 & x \ge x_n \end{cases}$$

# Test di Kolmogorov Smirnov - Applicazione

| | |
|---|---|
| Dmax | 0,118 |
| α=0,1 | 0,22 |
| α=0,05 | 0,24 |
| α=0,01 | 0,29 |




La massima differenza *D* riscontrata è minore del valore soglia della zona di rigetto, per tutti i livelli di rischio riportati. Pertanto si può affermare che non è possibile rigettare l'ipotesi nulla ad un livello di significatività inferiore a 0.9.

## 1) Test Chi-Quadrato

Un produttore di un componente per aerei informa i suoi clienti che la qualità dei suoi prodotti non è uniforme e che ogni componente può essere di qualità A,B,C,D,E indipendentemente l'una dall'altra con probabilità 0.15, 0.25, 0.35, 0.20, 0.05 rispettivamente. Tuttavia uno dei clienti acquistando grossi volumi di merce ha l'impressione di ricevere troppi pezzi di qualità E (la peggiore) e quindi decide di verificare l'affermazione del produttore investendo tempo e denaro per stabilire il livello qualitativo di 30 componenti. Dopo la verifica di questi pezzi egli ne trova 3 di qualità A, 6 di qualità B, 9 di qualità C, 7 di qualità D, 5 di qualità E. Al 95% di significatività cosa si decide?
Verificare cosa accade al 90% e al 99% di significatività.

# (Suggerimenti)

Il procedimento logico che deve essere seguito nell'applicazione del test Chi-Quadrato comprende diverse fasi:

1. Stabilire l'ipotesi nulla $H_0$ e l'eventuale ipotesi alternativa
2. Specificare il livello di significatività, l'ampiezza del campione e i gradi di libertà
3. Calcolare il valore del test statistico sulla base dei dati sperimentali, stimando il valore di probabilità ad esso associato
4. Sulla base della probabilità trarre le conclusioni: se la probabilità risulta superiore a quella prefissata, concludere che non è possibile rifiutare l'ipotesi nulla $H_0$ , viceversa se la probabilità risulta inferiore a quella prefissata rifiutare l'ipotesi nulla $H_0$ e quindi implicitamente accettare l'ipotesi alternativa $H_1$

# Esercizi proposti

## 2) Test Chi-Quadrato e test di Kolmogorov-Smirnov

Per il seguente campione casuale verificare se è plausibile supporre che la popolazione da cui proviene è gaussiana. Si osservi che la media campionaria risulta 0.05 e la deviazione campionaria è 2.54.

| | |
|---|---|
| -3,37 | -0,38 |
| -3,24 | 0,40 |
| -2,69 | 0,53 |
| -2,38 | 1,49 |
| -1,56 | 3,19 |
| -1,17 | 3,40 |
| -0,96 | 3,55 |
| -0,55 | 4,47 |

Cosa accade se invece si vuole testare che il campione segue una popolazione normale di media 0 e deviazione standard 2.5?

**Suggerimento: Costruire le classi con il metodo delle frequenze teoriche costanti cioè fissare le classi con valori superiori pari a** $x_i = \sigma z_{\alpha_i} + \mu$ **con** $\alpha_i = 0.2, 0.4, 0.6, 0.8$