

CONTROLLI DI QUALITA' E VARIABILITA'

Tra il 1920 e il 1945, si sviluppano le tecniche di controllo statistico della qualità dell'output grazie a George D. Edwards e a Walter A. Shewhart.

Si introdussero tecniche di controllo **sull'intero processo produttivo**, non limitandosi più, a verificare la difettosità dei prodotti solo alla fine del processo dato che i controlli a tappeto su tutti i prodotti stavano iniziando a rivelarsi troppo costosi.

Per effettuare questa nuova tipologia di controlli, si fece sempre più ricorso ai **criteri statistici**. Esaminando pochi prodotti finiti si riusciva a stabilire, mentre si produceva, se il processo presentava delle irregolarità o meno.

I controlli basati su criteri statistici ebbero la massima applicazione durante la seconda guerra mondiale, quando per l'industria bellica diventò necessario utilizzare in modo massiccio manodopera non specializzata e soggetta, quindi, ad un margine di errore maggiore.

CONTROLLI DI QUALITA' E VARIABILITA'

Il termine “controllo di qualità” è ampiamente utilizzato nel linguaggio corrente ed il suo significato è, almeno a grandi linee, noto a molti. La definizione più generale possibile del termine qualità è la seguente:

Il controllo statistico della qualità consiste in una raccolta di strumenti che sono essenziali nelle attività finalizzate al miglioramento della qualità di prodotti e servizi attraverso l'analisi della loro variabilità.

Es: un rivenditore compra delle cassette di frutta da un produttore e si aspetta che siano imballate e sistemate opportunamente in modo da facilitare l'esposizione della merce o la sistemazione in magazzino.

Gli aspetti generali della qualità sono:

- ➔ la qualità di progetto. I beni e servizi sono prodotti con vari gradi di qualità. Tali differenze sono intenzionali
- ➔ la conformità alle normative. Questo aspetto fa riferimento all'aderenza del prodotto alle specifiche e tolleranze assegnategli in fase di progettazione.

CONTROLLI DI QUALITA' E VARIABILITA'

Ogni prodotto possiede un certo numero di elementi misurabili, o comunque percepibili dall'utilizzatore, che contribuiscono congiuntamente alla formazione della qualità del prodotto.



Questi elementi vengono indicati con **CARATTERISTICHE DI QUALITA'** (es. fisiche, sensoriali, comportamento nel tempo).

Quando le caratteristiche di qualità sono misure espresse su una scala continua (es. peso,, concentrazione) si parla di variabili.

Quando invece si utilizzano dati discreti, ad es. dati di conteggio (es. numero di prodotti alterati, ecc.) si parla di attributi.

Il valore desiderato per una caratteristica di qualità è definito **VALORE NOMINALE** oppure **VALORE TARGET**.

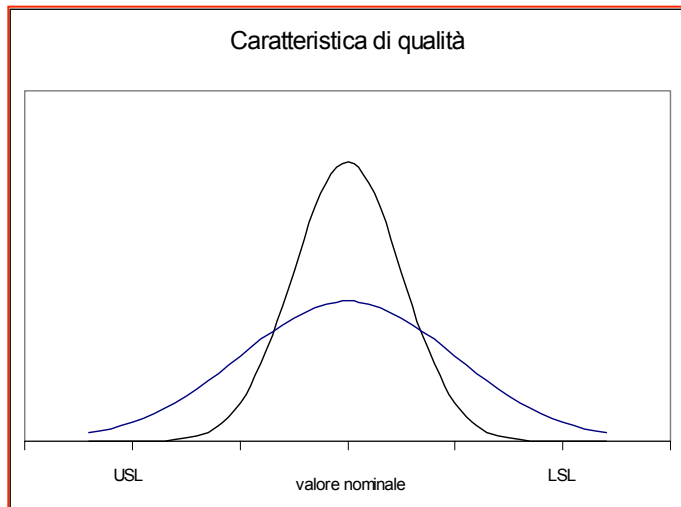
Esiste anche un **INTERVALLO DI VALORI**, (intorno del valore nominale), tale che se il valore della caratteristica di qualità rientra in tale intervallo il prodotto viene ritenuto conforme.

Limite superiore di questo intervallo: **limite di specifica o controllo superiore** (USL, Upper Specification Limit); limite inferiore: **limite di specifica o controllo inferiore** (LSL, Lower Specification Limit).

CONTROLLI DI QUALITA' E VARIABILITA'

La variabilità delle caratteristiche di qualità è un aspetto molto delicato per la qualità del prodotto (grossi investimenti aziende per assicurarsi che i valori delle caratteristiche di qualità dei prodotti siano il più vicino possibile ai valori nominali).

Tuttavia due o più unità di prodotto (o servizio) non sono mai uguali. Pertanto **esiste sempre un livello di variabilità** nelle caratteristiche di un prodotto e la qualità del prodotto dipende dall'ammontare della variabilità.



Esempio di due distribuzioni di caratteristiche di qualità.

Si può notare il diverso livello di variabilità ed è intuitivo comprendere che una **maggiore variabilità aumenta la probabilità** di produrre un elemento che **non rispetta** le specifiche.

CONTROLLI DI QUALITA' E VARIABILITA'

La variabilità può manifestarsi in diversi modi:

- in una unità di prodotto
- tra unità di prodotto
- nel tempo

Inoltre la variabilità dovuta ad almeno quattro cause (4M):

Man, Machine, Methods, Materials

La variabilità non è totalmente eliminabile quindi un certo grado di variabilità può essere ritenuto tollerabile, o fisiologico, per un dato processo produttivo. Questo tipo di variabilità viene indicata anche con il nome di “**variabilità naturale**”.



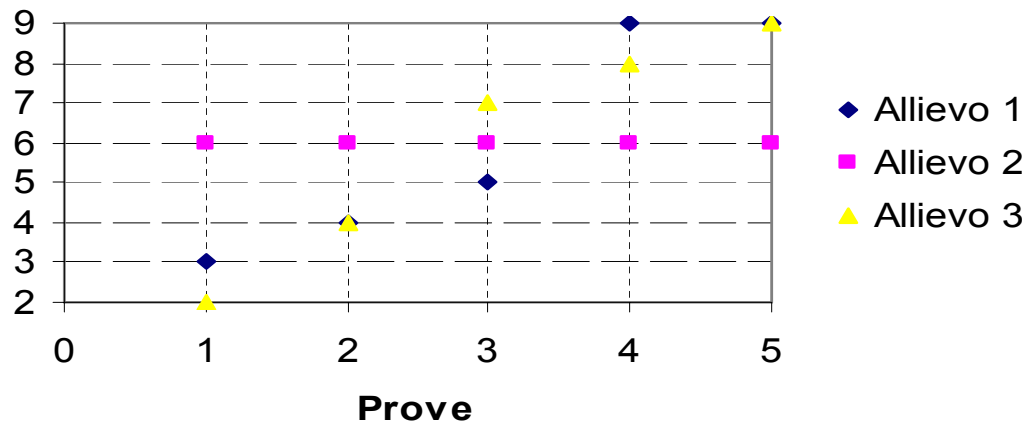
Il controllo di qualità ha l'obiettivo di mantenere la variabilità nel processo e nel prodotto ad un livello naturale. Il miglioramento della qualità mira ad una riduzione della variabilità nel processo e nel prodotto.

VARIABILITA'

Il calcolo della media ci permette di sintetizzare una quantità di dati da una parte, ma dall'altra riduce l'informazione racchiudendo tanti valori in un solo "dato", rendendo simili situazioni che sono molto diverse, in realtà (vedi anche lezioni precedenti sulla precisione analitica)

	1^a prova	2^a prova	3^a prova	4^a prova	5^a prova	MEDIA
Allievo 1	3	4	5	9	9	6
Allievo 2	6	6	6	6	6	6
Allievo 3	2	4	7	8	9	6

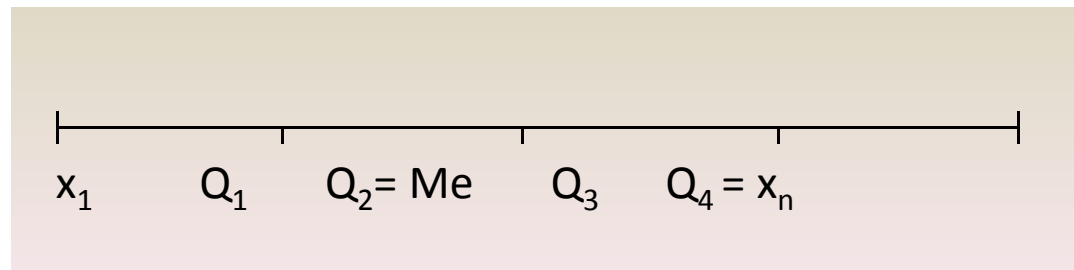
Pr ridurre la perdita di informazioni si ricorre allo studio della variabilità



MEDIANA E QUARTILI

Il concetto di “mediana” si può facilmente generalizzare ottenendo altri valori divisori, tra cui i più usati sono i **quartili** e i **percentili**. Tali indici di posizione si fondano sempre sul concetto di divisione della distribuzione.

I **Quartili** dividono la serie ordinata in quattro parti contenendo ciascuna lo stesso numero di dati.



I **Percentili** dividono la serie ordinata in cento parti contenendo ciascuna lo stesso numero di dati.

Ovviamente il Q_1 (primo quartile) corrisponde al 25° percentile

QUARTILI: esempio di utilizzo

Punteggio	Frequenza	Frequ. relativa %	Frequ. Cumulata	
0	1	4,2%	4,2%	
2	2	8,3%	12,5%	
4	1	4,2%	16,7%	
5	1	4,2%	20,8%	
7	1	4,2%	25,0%	= Q1
9	1	4,2%	29,2%	
10	1	4,2%	33,3%	
11	1	4,2%	37,5%	
12	2	8,3%	45,8%	
13	1	4,2%	50,0%	=Q2
14	1	4,2%	54,2%	
16	1	4,2%	58,3%	
19	1	4,2%	62,5%	
21	4	16,7%	79,2%	=Q3
22	1	4,2%	83,3%	
24	1	4,2%	87,5%	
25	1	4,2%	91,7%	
26	1	4,2%	95,8%	
29	1	4,2%	100,0%	=Q4

Primo quartile (Q_1): si trova esattamente sul valore 7, dato che la % cumulata corrispondente a tale punto è 25.0%

Secondo quartile (Q_2): si trova esattamente sul valore 13, dato che la % cumulata corrispondente a tale punto è 50.0%

Coincide sempre con la mediana

Terzo quartile (Q_3): si trova all'incirca sul valore 21, dato che la % cumulata corrispondente a tale punto è 79.2% (75%)

Quarto quartile (Q_4): si trova sempre sull'ultimo valore, in questo caso 29, dato che la % cumulata corrispondente a tale punto è 100%

INDICI DI VARIABILITA'

- Campo di variazione o range R
- Deviazione standard
- Varianza
- Scarto quadratico medio
- ...

Permettono di valutare le diseguaglianze dei dati in relazione al loro scostamento o dispersione da una media

Campo di variazione o range R di un insieme di valori osservati è la differenza tra il valore massimo e il valore minimo

$$R = x_{\max} - x_{\min}$$

Questo indice presenta due limiti importanti:

1. dipende **esclusivamente dai valori di massimo e di minimo** osservati, senza considerare la loro distribuzione
2. su di esso influisce pesantemente la presenza anche di **un solo** valore anomalo

VARIANZA

Varianza di un insieme di valori osservati è la media aritmetica degli scarti dalla media al quadrato (σ^2 , sigma quadrato)

$$\sigma^2 = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}$$

$$\text{Es. 1}^\circ \text{ allievo: } \sigma^2 = \frac{(3-6)^2 + (4-6)^2 + (5-6)^2 + (9-6)^2 \cdot 2}{5} = 8$$

	1 ^a prova	2 ^a prova	3 ^a prova	4 ^a prova	5 ^a prova	<i>MEDIA</i>	Varianza
Allievo 1	3	4	5	9	9	6	8
Allievo 2	6	6	6	6	6	6	0
Allievo 3	2	4	7	8	9	6	8,5

DEVIAZIONE STANDARD

Lo scarto quadratico medio (sqm) o deviazione standard è la radice quadrata (positiva) della varianza

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}}$$

	1^a prova	2^a prova	3^a prova	4^a prova	5^a prova	<i>MEDIA</i>	Varianza	sqm o Deviazione standard
Allievo 1	3	4	5	9	9	6	8	2,83
Allievo 2	6	6	6	6	6	6	0	0,00
Allievo 3	2	4	7	8	9	6	8,5	2,92

NORMALIZZAZIONE

La **normalizzazione** è una procedura statistica che permette di mettere a confronto distribuzioni diverse

Esempio

Avendo due prove il cui punteggio “raw” (“grezzo”) massimo raggiungibile dagli studenti è diverso, 30 nella prima prova e 45 nella seconda prova, non è possibile confrontare i risultati ottenuti.

Si ricorre quindi alla normalizzazione, che si basa su una **proporzione** rispetto al punteggio **massimo raggiungibile**

$$(Punti\ studente) : (p.ti\ totali) = (P.ti\ studente\ normalizzati) : 100$$

Allievi	Punti 1 ^a prova	Punti 1 ^a prova	P.ti 1 ^a p normalizzati	P.ti 2 ^a p normalizzati
A1	25	40	83,3	88,9
A2	15	43	50,0	95,6
A3	28	38	93,3	84,4
A4	19	33	63,3	73,3
A5	22	31	73,3	68,9
A6	30	20	100,0	44,4
A7	27	26	90,0	57,8
A8	18	45	60,0	100,0

Per A1 $1^a p \quad 25 : 30 = x : 100 \quad x = 25/30 * 100 = 83,3$
 $2^a p \quad 40 : 45 = x : 100 \quad x = 40/45 * 100 = 88,9$

COEFFICIENTE DI VARIAZIONE

Il **coefficiente di variazione (CV)** è dato dal **rapporto tra deviazione standard e media** moltiplicato per 100.

Generalmente, se il CV supera il 35% indica una variabilità eccessiva, fuori dai parametri accettabili

E' un indice di variabilità **relativo** (non dipende dall'unità di misura)



L'uso del CV si rende necessario ogni qualvolta si vogliono confrontare le misure di variabilità relative a distribuzioni le cui **modalità sono espresse in unità di misure diverse** (es. confronto tra variabilità dell'altezza e del peso) oppure sono espresse nella stessa unità di misura ma il loro valore medio risulta molto diverso (es. confronto delle variabilità dei pesi fra un campione di neonati ed uno di adulti).

CONTROLLO STATISTICO

Qualsiasi azione ripetuta è soggetta ad una certa variabilità.

Ad esempio, i processi di trasformazione dalle materie prime nell'industria alimentare non producono lo stesso identico prodotto da un ciclo lavorativo ad un altro.

L'importante è **misurare e monitorare** questa variabilità di processo.

Il **controllo della qualità** è fondato proprio sulla stima della variabilità dei processi.

Un processo può essere soggetto a due tipi di variazioni:

Variazioni casuali. Queste sono imprevedibili. Esprimono la naturale variazione dei risultati di un processo.

Si stimano con media e deviazione standard.

Variazioni sistematiche. Sono prevedibili. Mostrano un andamento o trend. Esprimono l'effetto di una causa, che si ripresenta sistematicamente durante il processo di lavorazione.

CONTROLLO STATISTICO

I 7 strumenti del controllo statistico di qualità

Il “controllo statistico di un processo” è un insieme di strumenti per il “problem solving” che comprende:

1. Istogrammi
2. Diagrammi di Pareto
3. Diagrammi di causa ed effetto
4. Schede di controllo
5. Carte di controllo
6. Diagrammi di correlazione
7. Test

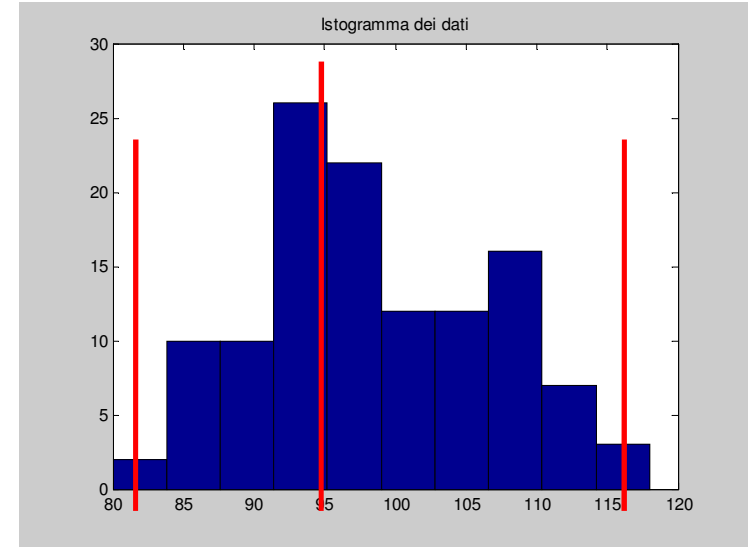
Esempio: Un'azienda farmaceutica decide di effettuare un controllo sul processo di iniezione di un farmaco, all'interno di appositi flaconi. L'azienda assume come tollerabile un quantitativo minimo di medicinale nei flaconi pari a 82 mL e uno massimo di 118 mL e in fase di progetto stabilisce un quantitativo obiettivo (target) di 95 mL.

Gli operatori addetti a tale compito hanno a disposizione le misure del contenuto dei flaconi del prodotto medicinale riportate nella tabella

CONTROLLO STATISTICO: ISTOGRAMMA esempio

Un primo approccio al problema può essere la costruzione di un istogramma dalla tabella dei dati.

	LUN	MAR	MER	GIO	VEN	SAB	LUN	MAR	MER	GIO	VEN	SAB
8.00	94	97	92	94	106	108	95	98	111	85	109	110
9.00	108	118	92	100	109	92	105	111	96	110	108	97
10.00	105	97	101	102	93	99	97	109	95	96	103	88
11.00	85	96	93	93	94	92	108	99	95	91	88	96
12.00	93	103	95	99	101	80	98	101	106	95	103	83
13.00	111	100	90	98	110	85	111	109	104	97	115	93
14.00	109	92	108	89	103	95	91	99	95	93	105	97
15.00	102	99	86	96	110	92	94	99	87	114	100	102
16.00	99	115	84	89	110	85	93	101	84	89	113	91
17.00	93	104	84	86	109	99	100	100	94	91	113	109

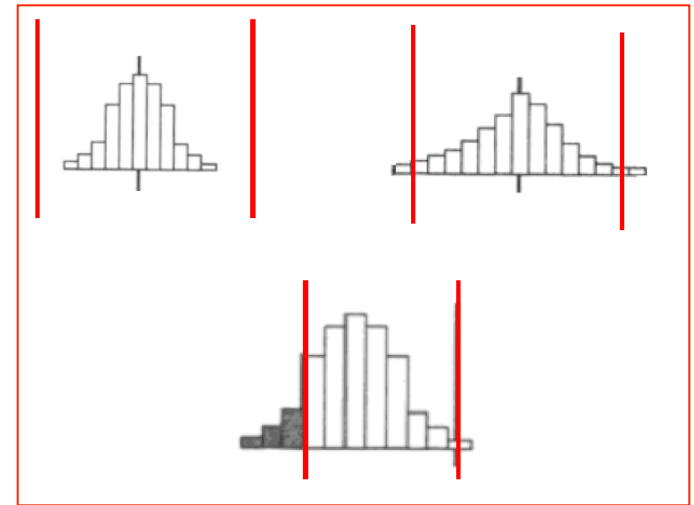
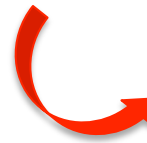


Dall'istogramma si nota come i dati seguano approssimativamente una distribuzione **normale**, con una variabilità dei dati piuttosto accentuata.

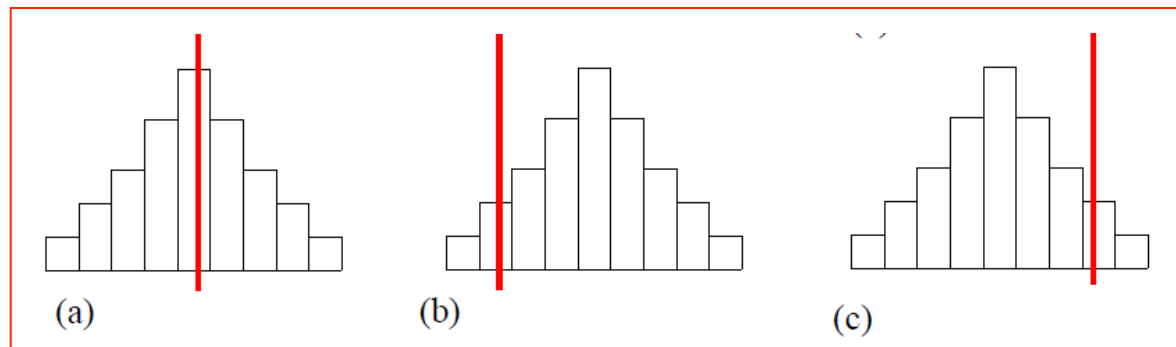
Rispetto al target aziendale, il processo è abbastanza centrato, ma la variabilità risulta eccessiva per cui potrebbe essere necessaria un'azione correttiva sulla variabilità del processo

CONTROLLO STATISTICO: ISTOGRAMMA esempio

Un istogramma consente di valutare la precisione del processo produttivo tramite l'analisi di dispersione della distribuzione dei dati, anche in relazione ai **limiti di tolleranza**



Dalla sovrapposizione dell'istogramma con la retta del valore obiettivo si può verificare il posizionamento del valore centrale dei dati rispetto al target assegnato (**tendenza centrale**).



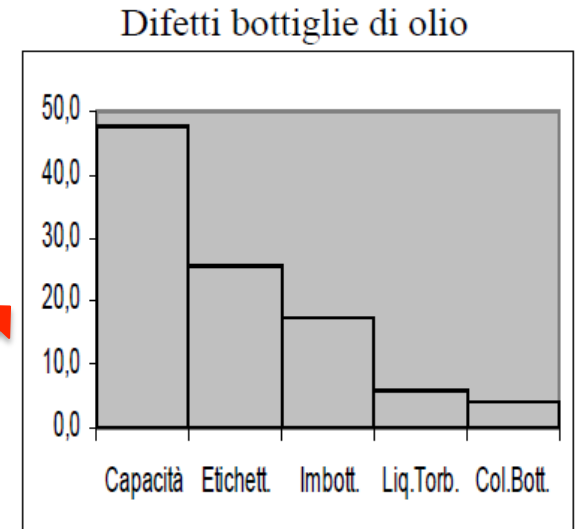
CONTROLLO STATISTICO: DIAGRAMMA DI PARETO es.

Si supponga che nel controllo di qualità effettuato su 2165 bottiglie di olio, se ne siano riscontrate 416 difettose, come dalla tabella

Tipi di difetti (classe di difetti)	Numero pezzi difettosi per tipo	% per tipo di difetto
Imbottigliamento	72	17.3
Etichettatura	103	24.7
Liquido torbido	18	4.3
Capacità	198	47.6
Colore bottiglia	25	6.0

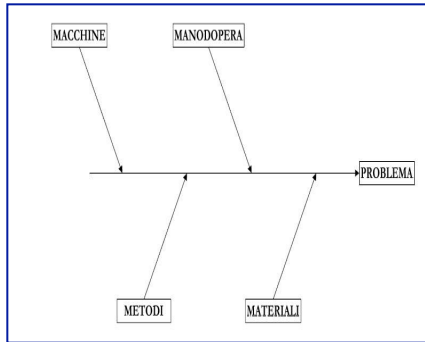
Un **diagramma di Pareto** è un istogramma in cui ogni colonna rappresenta una voce di “difettosità” e le cui colonne sono ordinate in ordine decrescente di valore.

- Indica i difetti in ordine di importanza
- L'esperienza insegna che è più facile ridurre una barra “alta” che non azzerare una barra “corta”

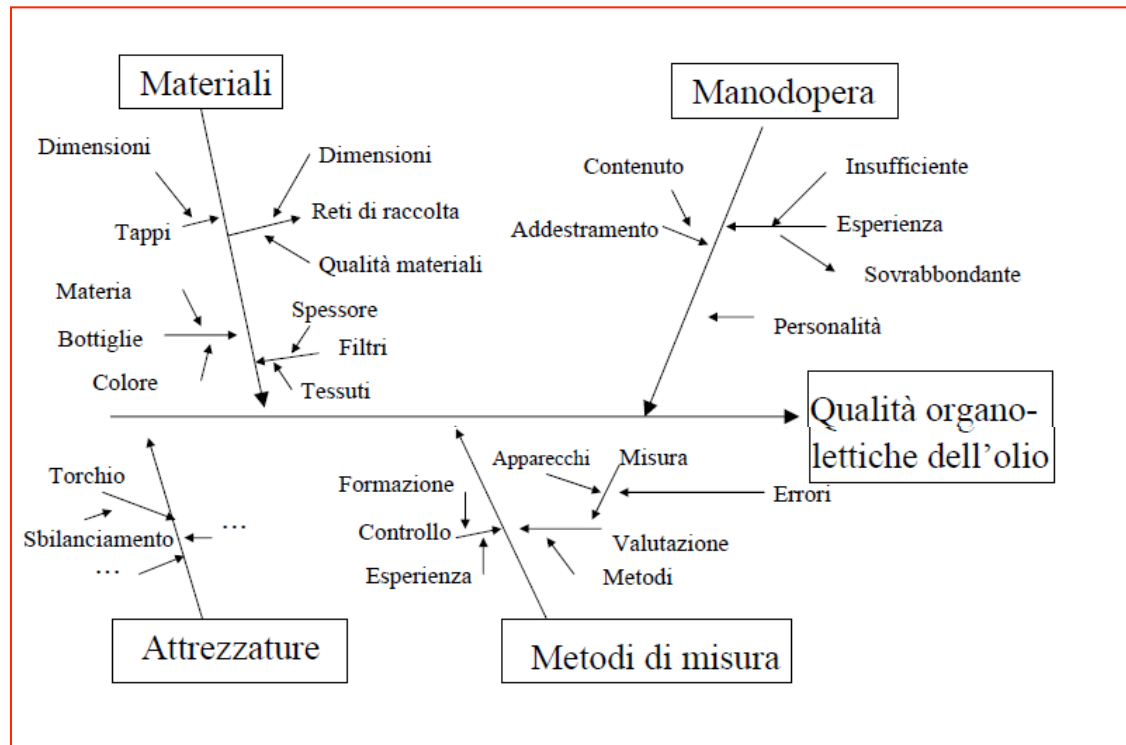


Il principio di Pareto è un risultato di tipo statistico-empirica che si osserva in molti sistemi complessi dotati di una struttura di causa-effetto e afferma che circa il 20% delle cause provoca l'80% degli effetti.

CONTROLLO STATISTICO: DIAGRAMMA CAUSA-EFFETTO O DIAGRAMMA DI ISHIKAWA (“a lisca di pesce” 😊)



Strumento grafico che permette di identificare, riunire e mostrare facilmente le **cause possibili** che hanno originato un problema o una certa caratteristica.



Solitamente utilizzato quando si devono analizzare i processi ad un livello approfondito per una loro **riprogettazione**.

Non ci aiuta ad individuare la causa principale che genera il nostro effetto ma restringe il campo d'azione originando un insieme di possibili cause.

È molto utile quando le cause possono essere **molte**

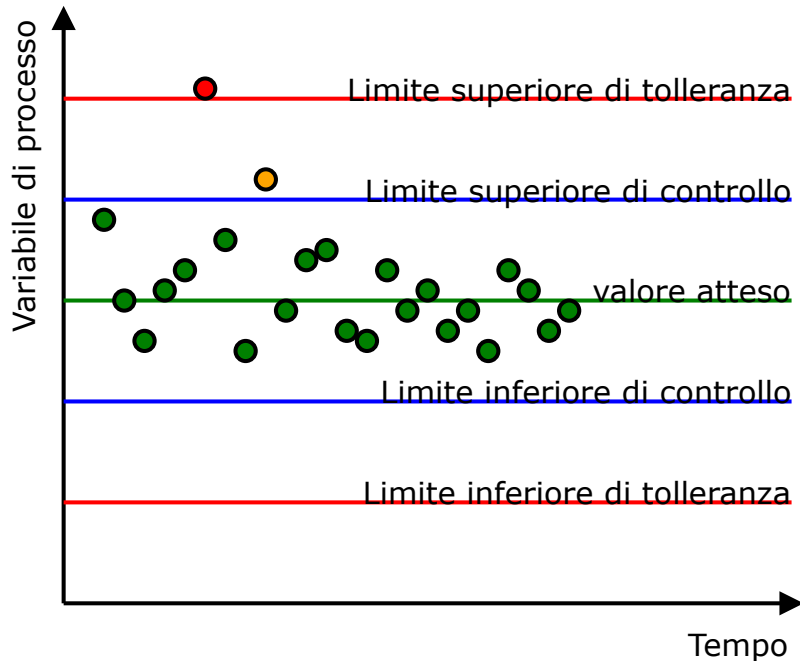
CONTROLLO STATISTICO: SCHEDE DI CONTROLLO

Lo scopo delle schede è quello di:

Stimare la variabilità del processo

Ridurre o eliminare le fonti di variazione sistematiche

Consentono di riportare i dati in modo semplice e con una disposizione tale da rendere agevole il loro uso



Scheda di controllo n. _____

Prodotto	Data
Fase	Stabilimento
N.pezzi	Reparto
N.lotto	N.ordine
Compilatore	

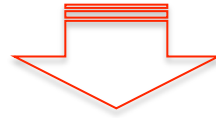
Tipo di difetto	Esito controllo	Subtotale
Difetti superficiali	////////////////////	32
Tagli	////////////////////	23
Incompleto	////////////////////	48
Deformato	////	4
Altri	////////	8

Totale pezzi	115
Totale scartati	86

CONTROLLO STATISTICO: SCHEDE DI CONTROLLO

In ogni processo di produzione c'è un “rumore di fondo” dovuto a cause inevitabile ma di effetto ridotto

Un processo la cui fonte di variabilità è dovuta solo al “rumore di fondo” è detto “sotto controllo statistico”



Altre fonti di variabilità:

- macchinari non funzionanti adeguatamente
- errori degli operatori
- difetti nelle materie prime

Invece, un processo la cui fonte di variabilità **NON** è dovuta solo al “rumore di fondo” è detto “fuori controllo statistico”

Le carte di controllo sono lo strumento più idoneo a segnalare queste “anomalie” nel monitoraggio di un processo di produzione

CONTROLLO STATISTICO: SCHEDE DI CONTROLLO

Come si usa un carta di controllo?



Sia W una statistica campionaria che misura qualche caratteristica di qualità e sia $E[W]$ il valore atteso (expected) di W e $Var[W]$ la sua varianza

Allora: $E[W] = \mu_W$ e $Var[W] = \sigma^2_W$

Limite superiore: $\mu_W + k\sigma^2_W$

Limite centrale: μ_W

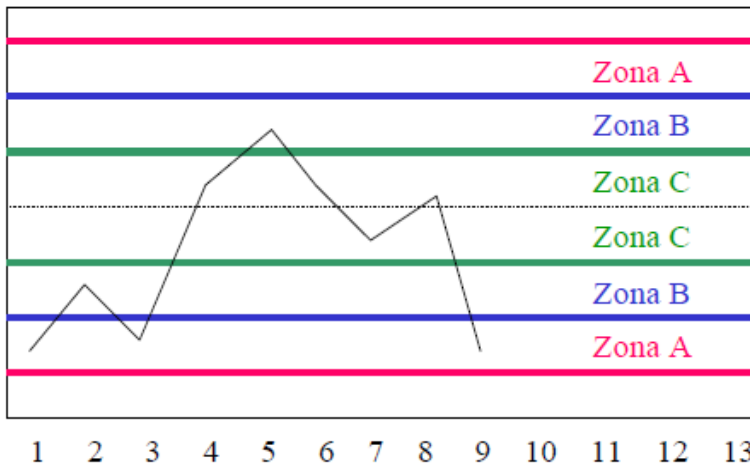
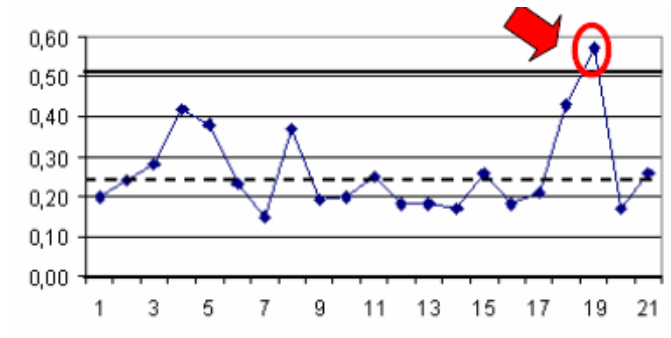
Limite inferiore: $\mu_W - k\sigma^2_W$

Per un certo k

CONTROLLO STATISTICO: CARTA DEI 3 σ

Se dovesse essere disponibile una valutazione teorica (storica o di progetto) della varianza della popolazione e della media, usando il teorema del limite centrale è possibile sostituire il parametro k con **3** e la varianza $\sigma_w = \sigma$ e per media quella della popolazione.

Viene dunque selezionato un campione casuale, costruita la carta di controllo



$3\hat{\sigma}_X$
 $2\hat{\sigma}_X$
 $1\hat{\sigma}_X$
 $1\hat{\sigma}_X$
 $2\hat{\sigma}_X$
 $3\hat{\sigma}_X$

Se i valori medi misurati cadono all'interno dei limiti di controllo e/o non mostrano particolari andamenti, il processo è considerato **sotto controllo statistico** al livello indicato dalla carta

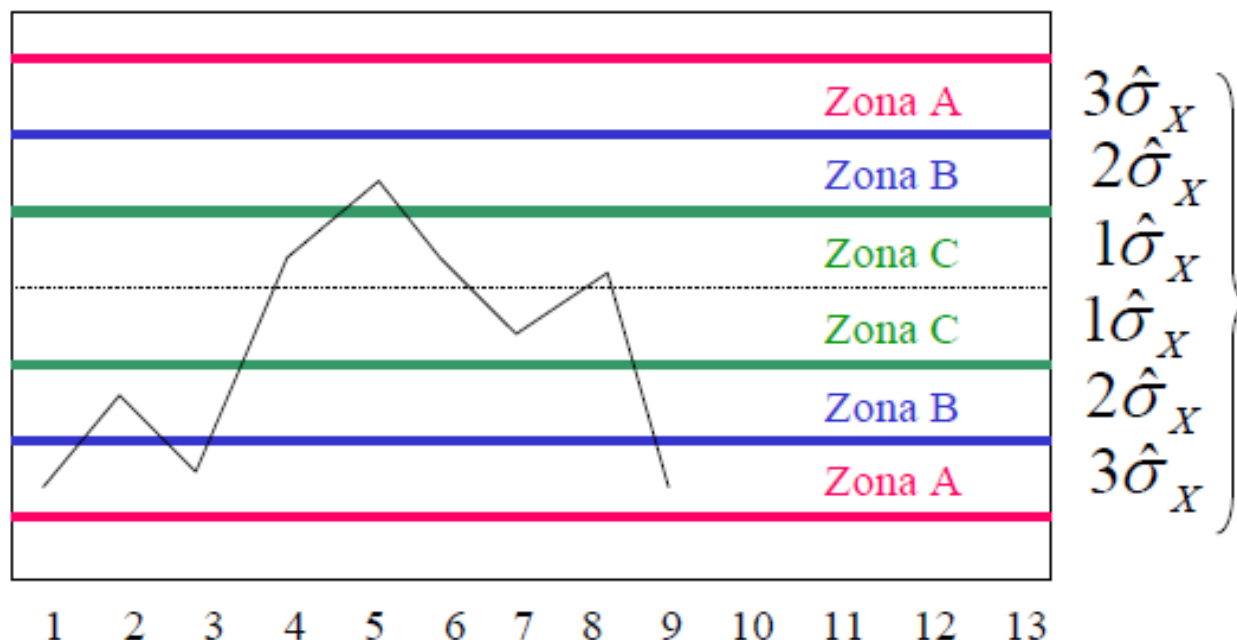
[Il Teorema del Limite Centrale afferma che la somma o la media di un grande numero di variabili aleatorie indipendenti dotate della stessa distribuzione è approssimativamente normale, indipendentemente dalla distribuzione sottostante]

CONTROLLO STATISTICO: CARTA DEI 3 σ

Alcune regole per il riconoscimento di sequenze non casuali

Un processo è “fuori controllo” se:

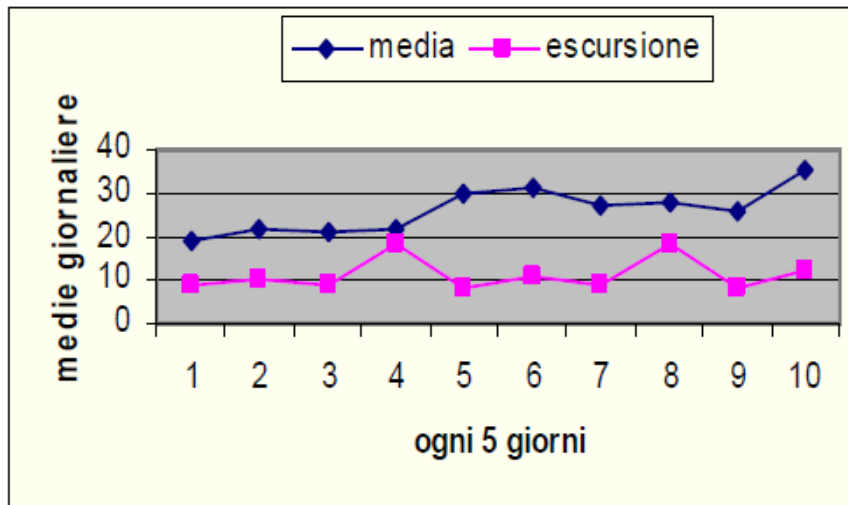
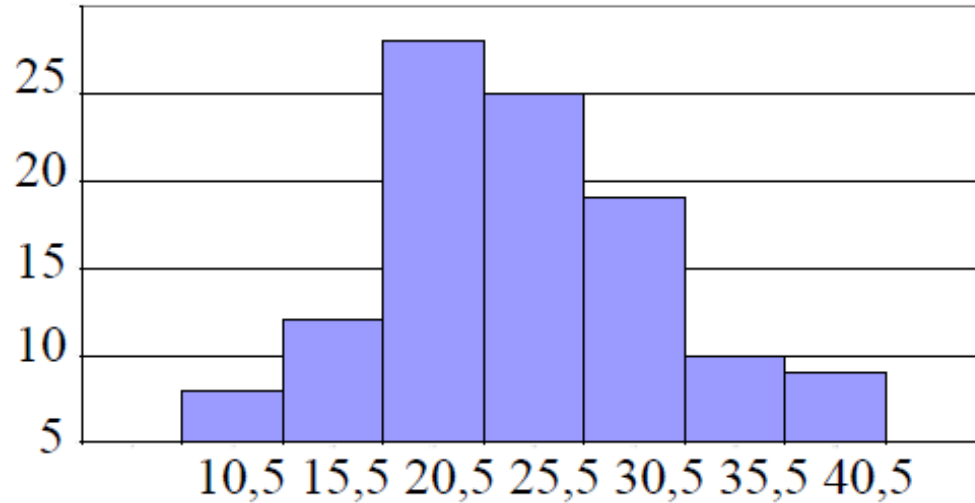
- un punto cade al di sopra (linea rossa superiore) o al di sotto (linea rossa inferiore) della zona A
- due punti consecutivi cadono all'interno della zona B
- 4 o 5 punti consecutivi cadono nella zona centrale C
- 8 punti consecutivi cadono al di sopra o al di sotto della zona centrale C



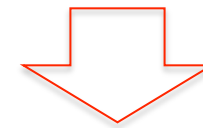
CONTROLLO STATISTICO: CARTA DEI 3 σ - ESEMPIO

ESEMPIO:

Lunghezza di un certo materiale di plastica registrata 5 volte al giorno.



C'e' una tendenza della media a crescere nel tempo. I punti del diagramma possono considerarsi nella norma?



Deriva temporale ma simile distribuzione: es. utensile che si degrada nel tempo?

SIGNIFICATIVITA' STATISTICA

Quando si effettua un test di significatività statistica, inizialmente si assume la cosiddetta «ipotesi zero» (o «**ipotesi nulla**»), secondo la quale **non esiste nessuna differenza tra i gruppi riguardo al parametro considerato.**

Secondo l'ipotesi zero, i gruppi (es. due serie di misure analitiche nel nostro caso) sono fra loro uguali e la differenza osservata va attribuita al caso.

Ovviamente l'ipotesi zero può essere vera o falsa. Ora occorre decidere: si accetta o si rifiuta l'ipotesi nulla?

Per decidere è necessario analizzare i dati con un **test statistico.**

⇒ Se il test «consiglia» di **rifiutare l'ipotesi zero**, allora la differenza osservata viene dichiarata **statisticamente significativa.**

⇒ Se invece il test «consiglia» di **accettare l'ipotesi zero**, allora la differenza è **statisticamente non significativa.**

SIGNIFICATIVITA' STATISTICA

Come sempre avviene, i risultati di un test statistico non hanno un valore di assoluta e matematica certezza, ma soltanto di **probabilità**.



Pertanto, una decisione di respingere l'ipotesi zero (presa sulla base del «consiglio» del test statistico) è probabilmente giusta, ma potrebbe essere errata.

La misura di questo rischio di cadere in errore si chiama «**livello di significatività**» del test.

LIVELLO DI SIGNIFICATIVITA'

Il **livello di significatività** di una prova può essere scelto a piacere dallo sperimentatore.

Di solito si sceglie un livello di probabilità di 0.05 (5%) o di 0.01 (1%). Questa probabilità (detta “**valore P**”, p-value) rappresenta una stima quantitativa della probabilità che le differenze osservate siano dovute al caso.

Più precisamente, il valore P è "la probabilità di ottenere un risultato altrettanto estremo o più estremo di quello osservato se la diversità è interamente dovuta alla sola variabilità campionaria, assumendo quindi che l'ipotesi iniziale nulla sia vera".

Notare che P è una probabilità e quindi può assumere solo valori compresi fra 0 e 1. Un valore P che si avvicina a 0 testimonia una bassa probabilità che la differenza osservata possa essere attribuita al caso.

LIVELLO DI SIGNIFICATIVITA'

Esempio

E' stata svolta una sperimentazione su due gruppi di animali nutriti con due mangimi diversi. Un gruppo è stato nutrito con il mangime A, e l'altro con il mangime B. Si è osservato che gli animali nutriti con A crescono di più rispetto a quelli nutriti con B.

Però questo effetto potrebbe essere dovuto al caso... possiamo essere sicuri che, ripetendo l'esperimento, si otterrebbero ancora dati che indicano che il mangime A è migliore di B? Ovviamente no!



A partire dai dati, si calcola il valore P: si ottiene una **stima quantitativa della probabilità** che le differenze osservate siano dovute al caso.

P è la risposta alla seguente domanda: «se in realtà non ci fossero differenze fra A e B, e se la sperimentazione fosse eseguita molte volte, quale proporzione di sperimentazioni condurrebbe alla conclusione che A è migliore di B?»

LIVELLO DI SIGNIFICATIVITA'

Il livello di significatività **5%** viene adottato molto frequentemente in quanto si ritiene che il rapporto $1/20$ (cioè 0.05) sia sufficientemente piccolo da poter concludere che sia «piuttosto improbabile» che la differenza osservata sia dovuta al semplice caso. In effetti, la differenza potrebbe essere dovuta al caso, e lo sarà 1 volta su 20.

Tuttavia, questo evento è «improbabile». Se si vuole escludere con maggiore probabilità l'effetto del caso, si userà un livello di significatività inferiore (es. 1%).



- se l'ipotesi nulla viene respinta al livello di significatività 5%, allora abbiamo il 5% di probabilità di respingere un'ipotesi nulla che, in effetti, era vera;
- se l'ipotesi nulla viene respinta al livello di significatività 1%, allora abbiamo l'1% di probabilità di respingere un'ipotesi nulla che - in effetti - era vera;

In generale, se l'ipotesi nulla viene respinta al **livello di significatività n%**, abbiamo **n% di probabilità di respingere un'ipotesi nulla che, in effetti, era vera.**

SIGNIFICATIVITA' STATISTICA

E' necessario sottolineare un concetto molto importante: «**statisticamente significativo**» non vuol dire importante, o di grande interesse, o rilevante. «Statisticamente significativo» significa semplicemente che ciò è stato osservato è «difficilmente dovuto al caso».



Test statistici di comune impiego

Numerosi test statistici vengono usati per determinare con un certo grado di probabilità l'esistenza (o l'assenza) di differenze significative nei dati in esame o meglio, più in generale, di accettare o rigettare una ipotesi nulla.

Nella tabella che segue vengono elencati alcuni dei test più comunemente impiegati

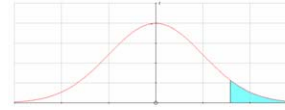
TEST DI SIGNIFICATIVITA' STATISTICA

Scopo	Test	Confronto fra
verificare la significatività statistica di una differenza	Test Chi quadrato	due (o più) percentuali (con molte osservazioni)
	Test esatto di Fisher	due (o più) percentuali (con poche osservazioni)
	Test U di Mann-Whitney	due mediane
	Test di Wilcoxon	due mediane
	Test t di Student	due medie
	Analisi della varianza	più di due medie
	Test t di Bonferroni	più di due medie
descrivere il grado di una associazione	Coefficiente di Regressione	una variabile dipendente ed una indipendente
	Test r di Pearson	due variabili

I test statistici consentono di stimare il grado di certezza di affermazioni e la precisione di misure effettuate su un campione casuale di una determinata popolazione.

Per ottenere risultati validi, occorre scegliere, fra le tante disponibili, le procedure statistiche adatte all'indagine o all'esperimento effettuato.

Tavole t di Student :
 $P(T > t_{\alpha})$



GL	0.100	0.050	0.025	0.010	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
infinity	1.282	1.645	1.960	2.326	2.576	3.091	3.291

ANALISI DELLA VARIANZA (ANOVA)

L'analisi della varianza (ANOVA - Analysis of Variance, σ^2) è un insieme di tecniche statistiche facenti parte della statistica inferenziale che permettono di confrontare **due o più gruppi** di dati confrontando la variabilità interna a questi gruppi con la variabilità tra i gruppi.

L'ipotesi nulla solitamente prevede che i dati di tutti i gruppi abbiano la stessa origine, ovvero la stessa distribuzione stocastica, e che le differenze osservate tra i gruppi siano dovute solo al caso.

Si usano queste tecniche quando le variabili esplicative sono di tipo nominale (discreto). E' possibile usare queste tecniche anche in presenza di variabili esplicative di tipo ordinale o continuo, ma in tal caso sono meno efficienti delle tecniche alternative (ad esempio: regressione lineare).