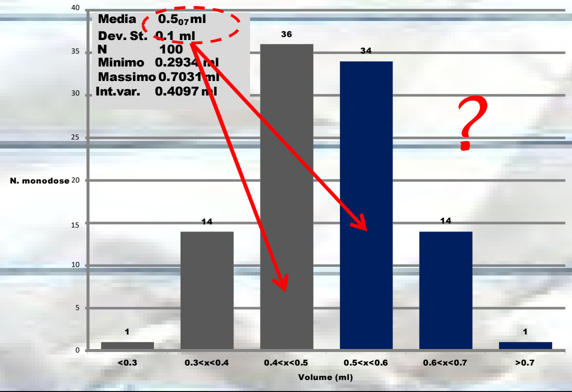
VERIFICA DI IPOTESI (TEST DI IPOTESI)

Distribuzione dei volumi di collirio in contenitori monodose



1.Come decidere se il risultato di una serie ripetuta che differisce molto dagli altri è un dato valido o è un outlier?
Test Q (test di Dixon)

2.Come decidere se la media di una serie di risultati sperimentali è una stima accettabile della media "vera"?
Calcolo dell'intervallo di fiducia della media

3.Come stabilire se il valore misurato è il valore "vero"?
Confronto di una media sperimentale con un valore noto.
Test z, test t di Student

4.Come decidere se i risultati di due analisi sono uguali?
Confronto tra due medie sperimentali.
Applicare il test t

N.B. I test che vedremo si basano sull'assunto che i dati abbiano distribuzione normale

PROVE (TEST) DI IPOTESI

Consistono nel formulare una ipotesi e nel verificare se con i dati a disposizione è possibile accettarla.

Nelle prove di ipotesi, le analisi statistiche del campione sono utilizzate per verificare la validità della ipotesi iniziale ESTESA ALLA POPOLAZIONE

Si tratta di prove (test) eseguite seguendo un procedimento logico predeterminato e basate sulla verifica della validità di una ipotesi statistica.

1. Stabilire l'ipotesi nulla.

L'ipotesi nulla H_0 stabilisce una differenza nulla e quindi un'uguaglianza tra le stime delle proprietà delle due popolazioni in esame.

L'ipotesi H_0 è una ipotesi formulata riguardo ad un parametro della popolazione in esame (ad es. la media).

Nel caso del confronto delle medie μ_1 e μ_2 (ad es. di peso) di due popolazioni da cui sono stati estratti due campioni, l'ipotesi nulla è:

$$H_0: \mu_1 = \mu_2$$
$$(\mu_1 - \mu_2 = 0)$$

5

2. Stabilire l'ipotesi alternativa.

L'ipotesi H_a o H_1 è la negazione di H_0 ossia

$$H_a: \mu_1 \neq \mu_2$$
$$(\mu_1 - \mu_2 \neq 0)$$

Se l'ipotesi nulla viene rigettata, l'ipotesi alternativa viene accettata.

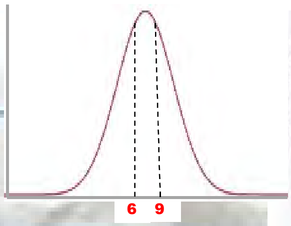
L'ipotesi nulla è unica, le ipotesi alternative possono essere più di una ($\mu_1 < \mu_2$; $\mu_1 > \mu_2$).

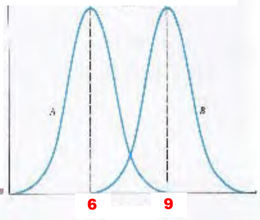
6

VERIFICA DI IPOTESI USANDO LA STATISTICA

La verifica di ipotesi (test di significatività) consiste nello stabilire la accettabilità di una ipotesi sulla base dei dati a disposizione.

H_0: 6 = 9 o H_1: 6 ≠ 9 ?





In statistica l'ipotesi nulla (H_0) postula che due dati siano uguali tra loro (e quindi che la loro differenza sia uguale a zero).

Se vale H_0,

se si osserva una differenza numerica nei dati sperimentali, questa è dovuta all'errore casuale, ad un determinato livello di probabilità.

Se è possibile rifiutare l'ipotesi H_0, allora vale l'ipotesi alternativa (H_1).

Se vale H_1, la differenza numerica osservata nei valori non deriva dall'errore casuale (non è dovuta al caso), ad un determinato livello di probabilità.

3. Fissare il livello di significatività α.

Indicato anche come livello α stabilisce il livello di probabilità di accettare l'ipotesi nulla.

Tale livello si esprime come frazione dell'unità, può essere fissato pari ad es. a 0.05 (caso molto comune).

Perciò se la probabilità associata a un evento è $\leq$ 0.05, si può concludere che l'ipotesi nulla non è accettabile e di conseguenza che l'ipotesi alternativa è valida.

Se l'ipotesi nulla è vera, allora il livello di significatività è la probabilità che il risultato ottenuto sia dovuto al caso.

10

La scelta di attribuire ad α il valore numerico di 0.05 è arbitraria anche se è la scelta più comune.

Scegliere il livello di significatività in un test di ipotesi equivale a stabilire la regione di accettazione e di rifiuto dell'ipotesi nulla.

11

Regione di accettazione contiene i risultati dei test che portano ad accettare l'ipotesi nulla

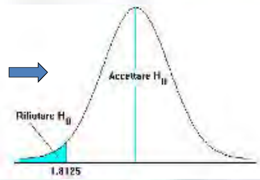
Regione critica contiene i risultati dei test che portano a rifiutare l'ipotesi nulla e accettare l'ipotesi alternativa.

L'ampiezza di queste due regioni dipende da due fattori: il livello di significatività e il numero di code del test.

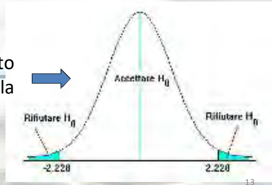
12

4. Stabilire il numero di code associate al piano sperimentale.

Nel test ad una coda, la zona di rifiuto è solamente da una parte della distribuzione (a sinistra quando il segno è negativo, a destra quando è positivo)



Nel test a due code, la zona di rifiuto è distribuita dalle due parti della distribuzione.



Un test di ipotesi è un processo alla ipotesi nulla H_0.

Il test inizia con un atto di fede- supponiamo che H_0 sia vera (come nel caso della presunzione di innocenza nei processi: un imputato è innocente finché non sia provato il contrario al di là di ogni plausibile dubbio)

e con una «scommessa», fissiamo il livello di significatività α

Il campione fornisce una prova che, ad esempio, è contro H_0.

Il p-value misura la plausibilità di H_0.

Più è piccolo il p-value e più sono evidenti le prove contro H_0 e più siamo portati a rifiutare H_0 e a prendere in considerazione l'ipotesi alternativa H_1.

se $p > \alpha$ la prova non è sufficiente per rifiutare l'H_0;
se $p \leq \alpha$ la prova è fortemente contro H_0 che quindi va rifiutata ⇔ i dati osservati sono **statisticamente significativi** (a quel determinato valore di p).

La conclusione di un test di ipotesi è di accettare o respingere l'ipotesi nulla.

La probabilità di fare la scelta sbagliata è calcolabile.

Si calcolano le probabilità

I. di rifiutare l'ipotesi nulla quando è vera;
II. di accettare erroneamente l'ipotesi nulla quando è falsa.

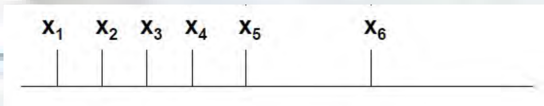
Il risultato di un test statistico è di **significatività** $\Rightarrow$ l'ipotesi nulla è rigettata e dunque esiste una differenza tra i due parametri.
Se l'ipotesi respinta è falsa non si commettono errori, mentre se in realtà H_0 è vera si commette un **errore di I tipo** (α) che consiste nel *rifiutare l'ipotesi nulla quando in realtà è vera*.
La probabilità di commettere un errore di tipo I è uguale al livello di significatività α.

Il risultato di un test statistico è di **non significatività** $\Rightarrow$ l'ipotesi nulla è vera e non esiste differenza tra i due parametri.
Se l'ipotesi formulata è vera non si commettono errori, mentre se in effetti è falsa si commette un **errore di II tipo** (β) che consiste nel ***non rifiutare l'ipotesi nulla quando questa è falsa*.**

16

X_6 è un dato aberrante?

(outlier, errore grossolano)



X_6 è un valore anomalo ma accettabile o è un dato aberrante distante dalle altre osservazioni al punto da non appartenere alla stessa distribuzione di misure?

COSA NON FARE

Far finta di niente e...

bisogna invece studiare il problema!

L'outlier è spesso un dato che porta con sé molte informazioni

Le analisi statistiche di campioni contenenti outliers possono essere fuorvianti perché i dati del campione non hanno distribuzione normale

media e deviazione standard non sono rappresentazioni "valide" del campione

Si replicano le prove per aumentare il numero di dati, in modo tale che il dato outlier abbia minore importanza nei calcoli

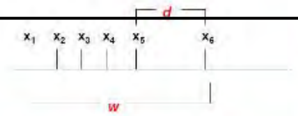
Mediana, intervallo di variazione e percentili per descrivere i risultati

Gli outliers possono essere prodotti anche da errori casuali: un piccolo numero di dati aberranti è sempre presente nei campioni grandi

Possono essere l'indizio che alcuni dati appartengono ad una popolazione differente da quella a cui appartiene la maggior parte dei dati del campione

APPLICO UN TEST STATISTICO
test Q (di Dixon)

Test di Dixon



X_6 e' il risultato sospetto

w = dispersione = differenza tra due valori estremi

d = divario = differenza tra il valore sospetto e quello piu' vicino

Si dispongono i dati di analisi ripetute in ordine crescente per selezionare l'outlier e il suo valore piu' vicino e si calcola $Q_{calcolata}$ secondo l'espressione:

$$Q_{calcolata} = d/w$$

$d = |x_6 - x_5|$ *d* = divario = differenza tra il valore sospetto e quello piu' vicino

$w = |x_6 - x_1|$ *w* = intervallo = differenza tra i valori estremi

$Q_{calcolata}$ deve essere confrontata con il valore critico tabulato $Q_{tabulata}$ in funzione del numero di osservazioni e del livello di fiducia (confidenza).

Se $Q_{calcolata} > Q_{tabulata}$, il dato sospetto puo' essere scartato a quel livello di fiducia

Se $Q_{calcolata} < Q_{tabulata}$, il dato sospetto andrebbe mantenuto

Valori critici di Q

Numero di osservazioni	Q_{crit} (scarto se $Q_{esp} > Q_{crit}$) 90% di confidenza	95% di confidenza	99% di confidenza
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

Se Q_{calc} calcolato dai dati sperimentali è maggiore del valore riportato nella tabella, Q_{crit}, esiste una probabilità pari alla «% di confidenza», che quel dato non appartenga alla stessa popolazione degli altri dati del campione.
Il valore sospetto può essere scartato accettando il rischio di sbagliare con probabilità pari a: 100 – «% di confidenza».

Skoog West Holler, Chimica Analitica, p. 86

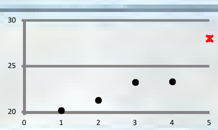
Table I. Critical Values of Dixon's r_{10} (Q) Parameter As Applied to a Two-Tailed Test at Various Confidence Levels, Including the 95% Confidence Level[a]

	confidence level					
N[b]	80% (α = 0.20)	90% (α = 0.10)	95% (α = 0.05)	96% (α = 0.04)	98% (α = 0.02)	99% (α = 0.01)
3	0.886	0.941	**0.970**	0.976	0.988	0.994
4	0.679	0.765	**0.829**	0.846	0.889	0.926
5	0.557	0.642	**0.710**	0.729	0.780	0.821
6	0.482	0.560	**0.625**	0.644	0.698	0.740
7	0.434	0.507	**0.568**	0.586	0.637	0.680
8	0.399	0.468	**0.526**	0.543	0.590	0.634
9	0.370	0.437	**0.493**	0.510	0.555	0.598
10	0.349	0.412	**0.466**	0.483	0.527	0.568
11	0.332	0.392	**0.444**	0.460	0.502	0.542
12	0.318	0.376	**0.426**	0.441	0.482	0.522
13	0.305	0.361	**0.410**	0.425	0.465	0.503
14	0.294	0.349	**0.396**	0.411	0.450	0.488
15	0.285	0.338	**0.384**	0.399	0.438	0.475
16	0.277	0.329	**0.374**	0.388	0.426	0.463
17	0.269	0.320	**0.365**	0.379	0.416	0.452
18	0.263	0.313	**0.356**	0.370	0.407	0.442
19	0.258	0.306	**0.349**	0.363	0.398	0.433
20	0.252	0.300	**0.342**	0.356	0.391	0.425
21	0.247	0.295	**0.337**	0.350	0.384	0.418
22	0.242	0.290	**0.331**	0.344	0.378	0.411
23	0.238	0.285	**0.326**	0.338	0.372	0.404
24	0.234	0.281	**0.321**	0.333	0.367	0.399
25	0.230	0.277	**0.317**	0.329	0.362	0.393
26	0.227	0.273	**0.312**	0.324	0.357	0.388
27	0.224	0.269	**0.308**	0.320	0.353	0.384
28	0.220	0.266	**0.305**	0.316	0.349	0.380
29	0.218	0.263	**0.301**	0.312	0.345	0.376
30	0.215	0.260	**0.298**	0.309	0.341	0.372

D.B. Rorabacher, Analytical Chemistry (1991) 63:139-146

Esempio. Verificare se ci sono dati aberranti (outliers) nel gruppo di misure sotto riportato.

Dataset iniziale: 23.23; 21.29; 20.16; 28.05; 23.33
H_0: 28.05 appartiene al dataset,
H_1: il dato non appartiene al dataset.



Dispongo in ordine crescente i dati
20.16; 21.29; 23.23; 23.33; 28.05;

d = 28.05 - 23.33 = 4.72
w = 28.05 – 20.16 = 7.89
$Q_{calcolato}$ = d/w = 4.72/7.89 = 0.598226…

Il valore critico per 5 dati al 95% di livello di fiducia, è $Q_{critico}$ (n=5, 95%) = 0.710.

$Q_{calcolato} < Q_{critico}$
Non si può rifiutare l'ipotesi nulla ✂ non ho prove per escludere il dato "strano" e devo considerarlo probabile come gli altri 4 al livello di fiducia del 95%.

Esempio. Verificare se ci sono outliers nel gruppo di misure sotto riportato relativo al contenuto di glucosio in un campione di "latte" di soia.

Dataset: 6.265; 6.234; 6.236; 6.246; 6.380; 6.222 (g/100 ml)
H_0: 6.380 appartiene al dataset,
H_1: il dato non appartiene al dataset.

Dispongo in ordine crescente i dati
6.222; 6.234; 6.236; 6.246; 6.265; 6.380

d = 6.380 – 6.265 = 0.115
w = 6.380 – 6.222 = 0.158
$Q_{calcolato}$ = d/w = 0.115/0.158 = 0.72785

Il valore critico per 6 dati al 95% di livello di fiducia, è $Q_{critico}$ (n=6, 95%) = 0.625.

$Q_{calcolato} > Q_{critico}$

Si può rifiutare l'ipotesi nulla al livello del 95% di confidenza
se escludo il dato 6.380 accetto il rischio pari al 5% di probabilità di commettere un errore di giudizio.

INTERVALLO DI CONFIDENZA PER LA MEDIA DI DATI DI CUI E' NOTA LA DEVIAZIONE STANDARD DI POPOLAZIONE

INTERVALLO DI CONFIDENZA

Il valore esatto della media μ per una popolazione di dati non può mai essere determinato precisamente perché richiederebbe un numero infinito di misure.

E' possibile definire un intervallo di valori che contenga il valore μ con una certa probabilità.

L'analisi statistica permette di fissare dei limiti attorno alla media sperimentale Xφ entro i quali cade il valore vero di μ con un certo grado di probabilità.

Questi limiti sono detti **limiti di confidenza** e l'intervallo da essi definito è noto come **intervallo di confidenza** o **intervallo di fiducia**

27

La probabilità che il valore atteso di un parametro stimato sia incluso in un intervallo di confidenza (o di fiducia) chiama **livello di fiducia**, e si indica con $1 - \alpha$.

Il livello di fiducia è espresso da un numero tra 0 e 1 (o in percento). La quantità complementare, α, si chiama **livello di significatività**.

Quindi la scelta di un dato livello di fiducia non esclude la possibilità di fare previsioni sbagliate:

Es. se per la stima di un parametro si sceglie il livello di fiducia $1 - \alpha = 95\%$

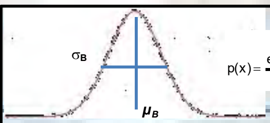
esistono 5 possibilità su cento ($5\% = \alpha$) che il valore vero del parametro sia diverso dalla stima (cada in modo significativo al di fuori dell'intervallo di fiducia).

Intervallo di fiducia al 90%

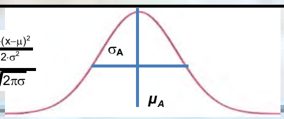
La barra indica il valore della media del campione, la regione invece che si estende da Inf a Sup è l'intervallo di fiducia del 90% di probabilità.

L'intervallo di fiducia viene calcolato per stabilire il livello di probabilità che il valore vero della media della popolazione si trovi all'interno dell'intervallo compreso tra Inf e Sup.

29

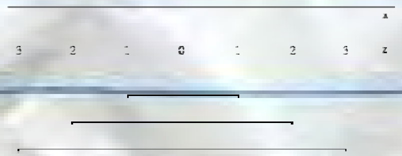


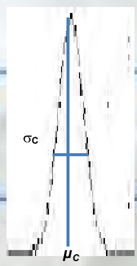
$$p(x) = \frac{e^{\frac{-(x-\mu)^2}{2\cdot\sigma^2}}}{\sqrt{2\pi\sigma}}$$



$$p(z) = \frac{e^{-\frac{1}{2} z^2}}{\sqrt{2\pi}}$$

$$Z = \frac{X - \mu}{\sigma}$$





x è una variabile normale: la popolazione a cui appartiene ha μ, σ

$z = \frac{\bar{x} - \mu}{\sigma}$ **Variabile z standardizzata**

$\bar{x}$ **Media del campione**

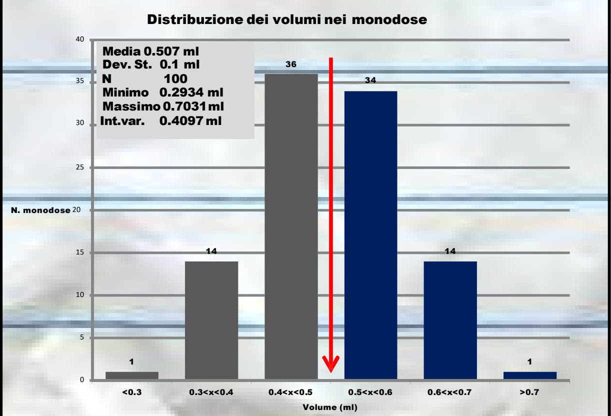
μ **Media della popolazione**

σ **Deviazione standard della popolazione**

livello di conf.	90%	95%	99%
α	0.10	0.05	0.01
z_α	1.65	1.96	2.58

www.dima.unige/pls_statistica

VERIFICA DI IPOTESI (TEST DI IPOTESI)



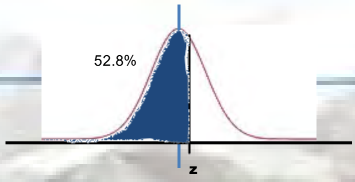
Esempio di uso della variabilez

Valori **prima** della regolazione della macchina dosatrice

Media $\bar{x} = 0.507$ ml, $s \approx \sigma = 0.1$ ml, $N = 100$

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{0.507 - 0.5}{0.1} = \frac{0.007}{0.1} = 0.07$$

Proporzione della popolazione
che si trova a sinistra di $z = 0.07$ (tabulata) $= 0.52790$.

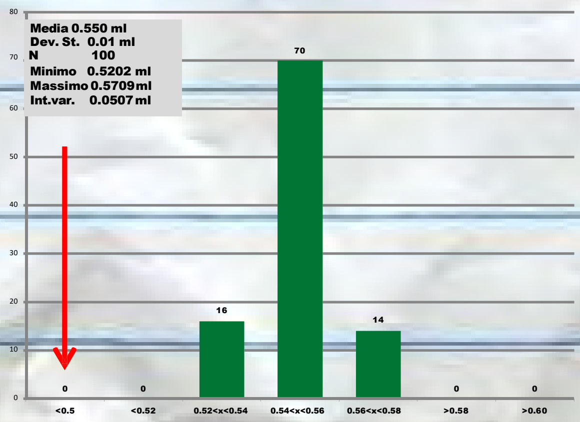


	Seconda cifra decimale di z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	[illegible]	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	[illegible]	[illegible]	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	[illegible]	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	[illegible]	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900

z = 0.07

?

http://www.matapp.unimib.it/~fcaraven/did0607/tavola_normale.pdf



Valori dopo la regolazione della macchina dosatrice

Media $\bar{x} = 0.550\,\text{ml}$

$s \approx \sigma = 0.01\,\text{ml}$

$N = 100$

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{0.550 - 0.5}{0.01} = 5$$

z	Φ(z)
3,84	0,99994
3,85	0,99994
3,86	0,99994
3,87	0,99995
3,88	0,99995
3,89	0,99995
3,90	0,99995
3,91	0,99995
3,92	0,99996
3,93	0,99996
3,94	0,99996
3,95	0,99996
3,96	0,99996
3,97	0,99996
3,98	0,99997
3,99	0,99997
4,00	0,99997

Per z = 4 la proporzione della popolazione è

La proporzione della popolazione di pezzi con V al di sopra di 0.500ml è superiore al 99.997%.

Rischio OOS < 3 pezzi ogni 100'000

Per questi stessi dati, si può calcolare l'Intervallo di Confidenza prima della regolazione della macchina
Media= 0.507 ml, s ~ σ = 0.1ml, N=100

$$x - z \cdot \frac{\sigma}{\sqrt{N}} < \mu < x + z \cdot \frac{\sigma}{\sqrt{N}}$$

$$\text{Errore standard della media del campione} = \frac{\sigma}{\sqrt{N}} = \frac{0.1}{\sqrt{100}} = \frac{0.1}{10} = 0.01$$

$$0.507 - 2.58\text{x}0.01 < \mu < 0.507 + 2.58\text{x}0.01$$

Intervallo di confidenza al 99% di probabilità

$$0.507 - 0.0258 < \mu < 0.507 + 0.0258 \Rightarrow 0.481 < \mu < 0.533$$

Limiti di confidenza al 99% di probabilità

Livello di confidenza al 99% di probabilità, z = 2.58

livello di conf.	90%	95%	99%
α	0.10	0.05	0.01
z_α	1.65	1.96	2.58

... e dopo la regolazione della macchina:
Media= 0.550 ml, s ~ σ = 0.01ml, N=100

$$\text{Errore standard della media} = \frac{\sigma}{\sqrt{N}} = \frac{0.01}{\sqrt{100}} = \frac{0.01}{10} = 0.001$$

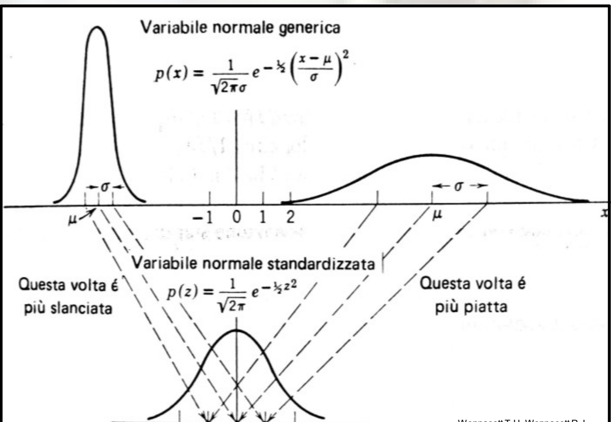
Livello di confidenza 99%, z = 2.58

Intervallo di confidenza (99%)

$$x - z \cdot \frac{\sigma}{\sqrt{N}} < \mu < x + z \cdot \frac{\sigma}{\sqrt{N}}$$

$$0.550 - 2.58\text{x}0.001 < \mu < 0.550 + 2.58\text{x}0.001$$

$$0.54742 < \mu < 0.55258 \Rightarrow 0.547 < \mu < 0.553$$



Wonnacott T.H.-Wonnacott R.J., Introduzione alla stastistica. FrancoAngeli Ed., 18a edizione, Milano 2002. p.91

INTERVALLO DI CONFIDENZA PER LA MEDIA DI DATI DI CUI NON E' NOTA LA DEVIAZIONE STANDARD DI POPOLAZIONE

Quando si ha un campione poco numeroso, costituito da N misurazioni indipendenti (x_1, x_2, ..., x_N, $N < 30$) estratte da una popolazione normale con media μ e deviazione standard (stimata) s, allora la quantità

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{N}}\right)}$$

ha una distribuzione t di Student con N-1 gradi di libertà, indicata con t_{N-1}.
Quando il valore di N è sufficientemente grande (N >> 30), la distribuzione della quantità t è molto prossima alla distribuzione normale, così la curva normale può essere utilizzata al posto della t di Student.