704SM Biostatistica

Massimo Borelli

ottobre 2016

UNIVERSITÀ DEGLI STUDI DI TRIESTE
Dipartimento di Scienze della Vita

SOCIETA' DEI MATEMATICI
E NATURALISTI DI MODENA
www.socnatmatmo.unimore.it

domande

- Cosa è un test statistico?
- Una celebre storia ..
- Invece, oggi ..

Cosa è un test statistico?

DATI (numerici, categorici)

CALCOLO (algebrico, analitico, combinatorio)

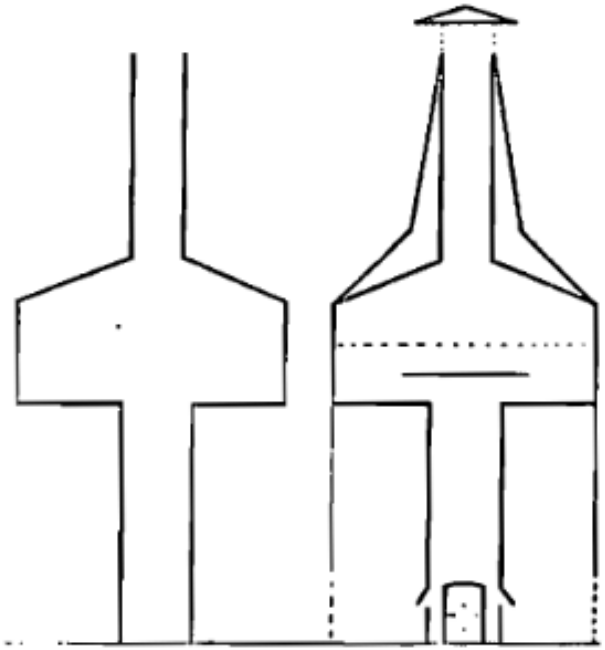
RISULTATO DEL CALCOLO

(consuntivo numerico, valore di probabilità)

DECISIONE

Una celebre storia ..

Una celebre storia ..



The Kiln Drying of Malt.

By H. M. Chubb.

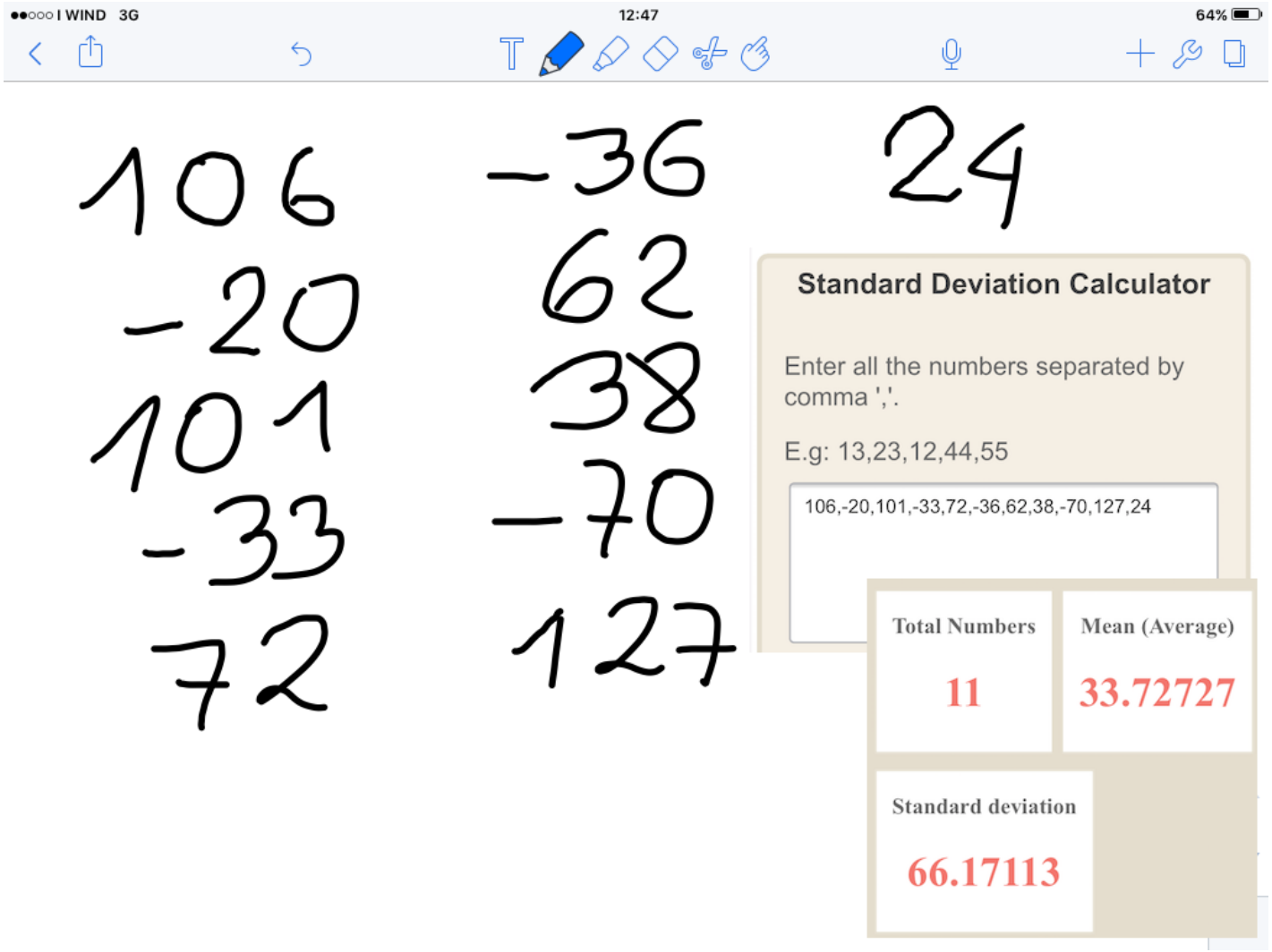
Decisione: conviene o non conviene essicare i semi?

Una celebre storia ..

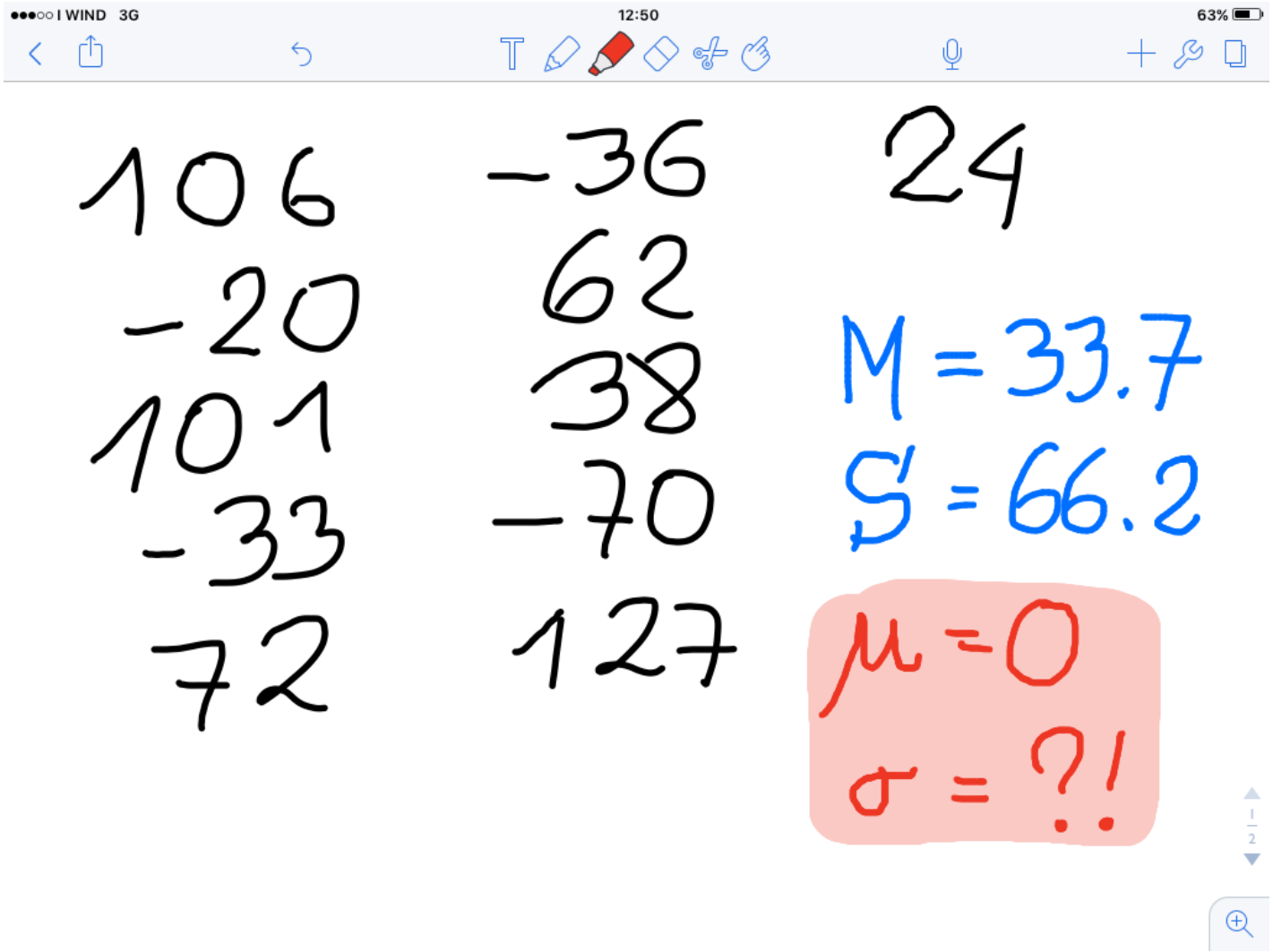
Una celebre storia ..

	lbs. head corn per acre			Price of head corn in shillings per quarter			cwts. straw per acre		
	N. K. D.	K. D.	Diff.	N. K. D.	K. D.	Diff.	N. K. D.	K. D.	Diff.
1899	1903	2009	+106	26½	26½	0	19¼	25	+5¾
	1935	1915	− 20	28	26½	−1½	22¾	24	+1¼
	1910	2011	+101	29½	28½	−1	23	24	+1
	2496	2463	− 33	30	29	−1	23	28	+5
	2108	2180	+ 72	27½	27	−½	22½	22½	0
	1961	1925	− 36	26	26	0	19¾	19⅝	−¼
	2060	2122	+ 62	29	26	−3	24½	22¼	−2¼
1900	1444	1482	+ 38	29½	28½	−1	15½	16	+½
	1612	1542	− 70	28½	28	−½	18	17¼	−¾
	1316	1443	+127	30	29	−1	14¼	15¾	+1½
	1511	1535	+ 24	28½	28	−½	17	17¼	+¼
Average	1841·5	1875·2	+33·7	28·45	27·55	−·91	19·95	21·05	+1·10
Standard)									

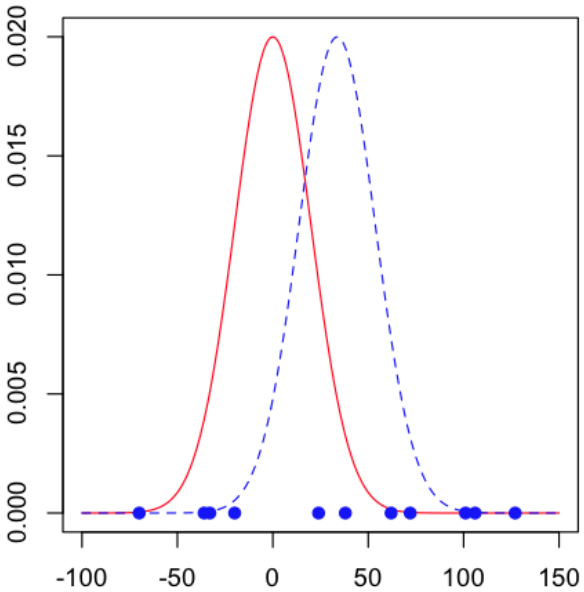
Una celebre storia ..



Una celebre storia ..



Una celebre storia ..



$M = 33.7$

$S = 66.2$

$\mu = 0$

$\sigma = ?!$

la grande scoperta

domanda

- se X è una normale, $z = \frac{X-\mu}{\sigma}$ è una normale?
  - sì

ci ricordiamo che lo standard error è l' *oggetto giusto* per passare dal campione alla popolazione

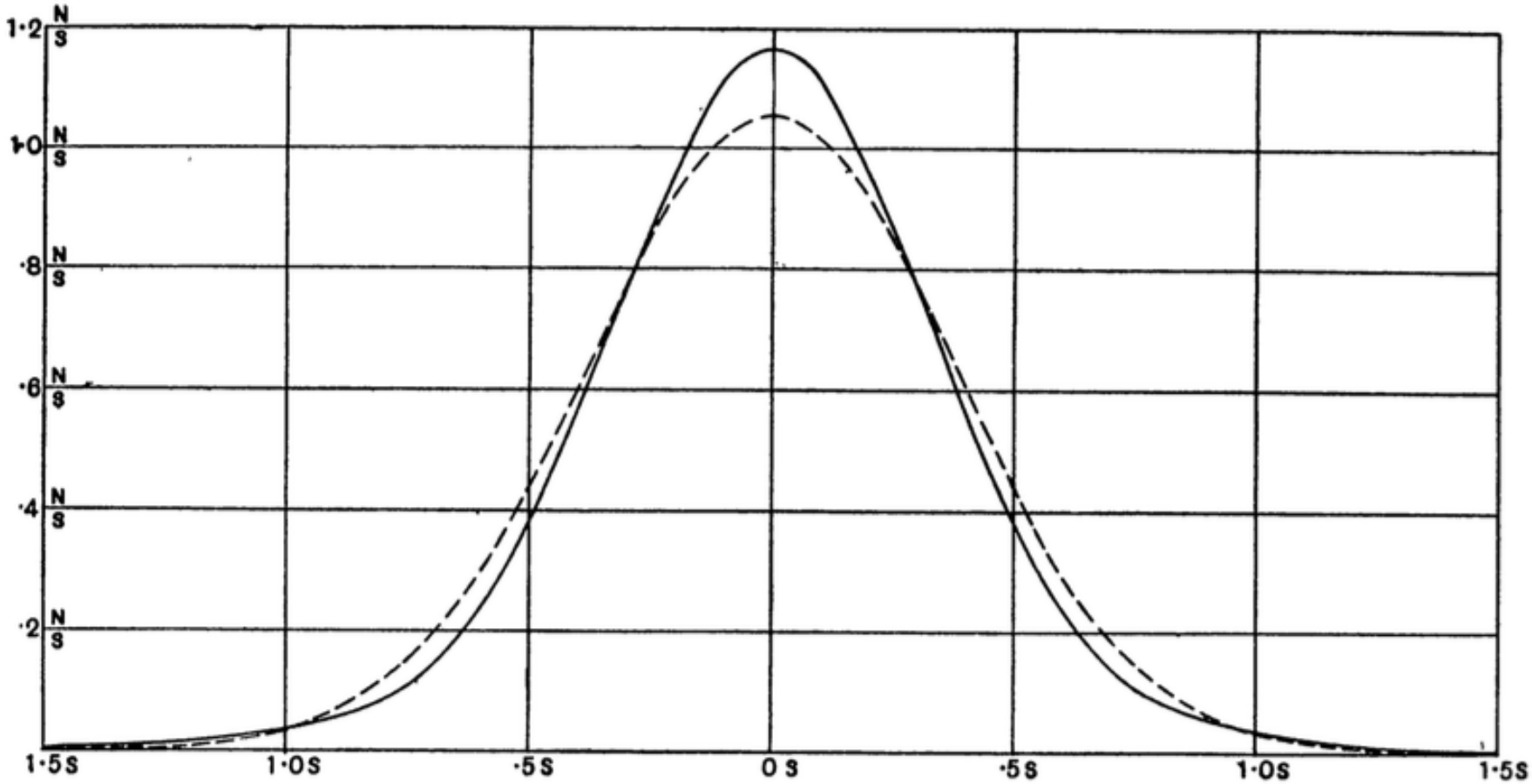
domanda

- se X è una normale, $\frac{X-M}{\frac{s}{\sqrt{N}}}$ è una normale?
  - no!

Una celebre storia ..

$$y = \frac{\sqrt{\frac{n}{2\pi}}}{\sigma} \cdot \frac{\int_0^\infty s^{n-1} e^{-\frac{ns^2(1+z^2)}{2\sigma^2}} ds}{\int_0^\infty s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds}$$

Una celebre storia ..



I have tabled the area for the normal curve with standard deviation $1/\sqrt{7}$ so a to compare with my curve for $n = 10$*. It will be seen that odds laid accordin

Una celebre storia ..

The paper is divided into the following nine sections:

I. The equation is determined of the curve which represents the frequency distribution of standard deviations of samples drawn from a normal population.

II. There is shown to be no kind of correlation between the mean and the standard deviation of such a sample.

III. The equation is determined of the curve representing the frequency distribution of a quantity z, which is obtained by dividing the distance between the mean of a sample and the mean of the population by the standard deviation of the sample.

IV. The curve found in I. is discussed.

V. The curve found in III. is discussed.

VI. The two curves are compared with some actual distributions.

VII. Tables of the curves found in III. are given for samples of different size.

VIII and IX. The tables are explained and some instances are given of their use.

X. Conclusions.

Una celebre storia ..

VOLUME VI MARCH, 1908 No. 1

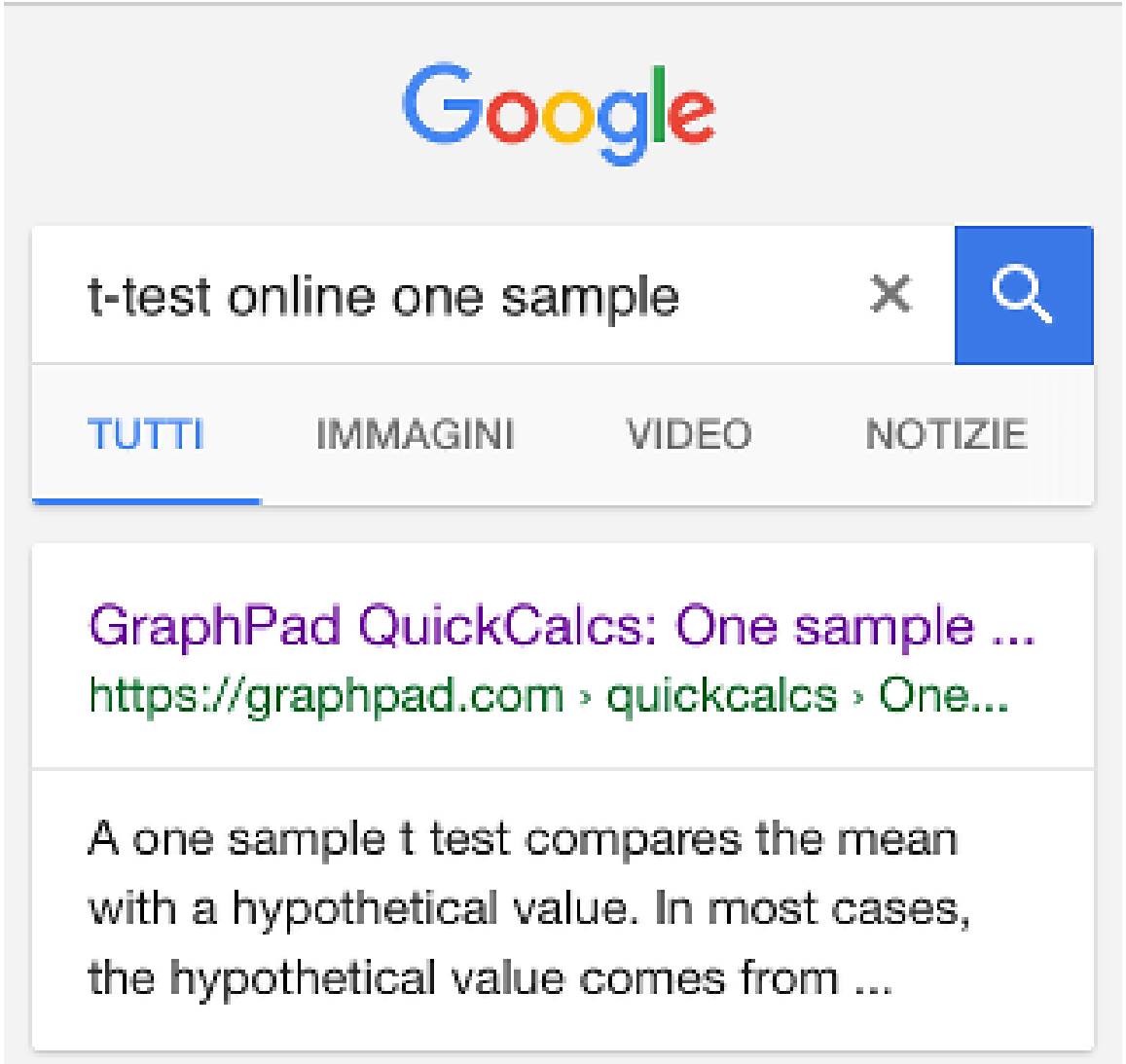
BIOMETRIKA.

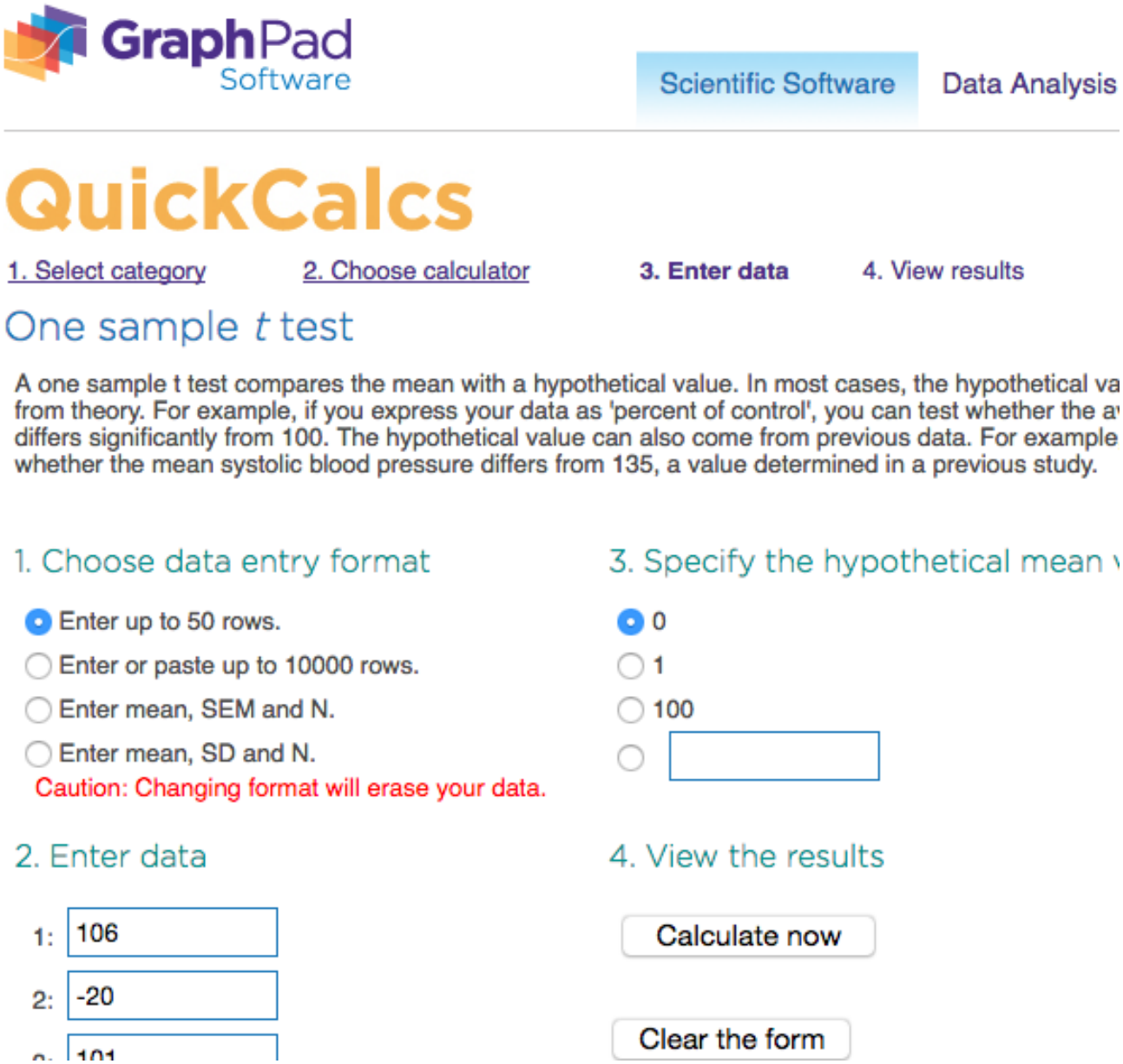
THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

The Right Honourable
The Earl of Iveagh
Edward Cecil Guinness

proviamoci da soli




GraphPad Software
Scientific Software
Data Analysis
QuickCalcs
1. Select category
2. Choose calculator
3. Enter data
4. View results
One sample t test
A one sample t test compares the mean with a hypothetical value. In most cases, the hypothetical va
from theory. For example, if you express your data as 'percent of control', you can test whether the a
differs significantly from 100. The hypothetical value can also come from previous data. For example
whether the mean systolic blood pressure differs from 135, a value determined in a previous study.
1. Choose data entry format
Enter up to 50 rows.
Enter or paste up to 10000 rows.
Enter mean, SEM and N.
Enter mean, SD and N.
Caution: Changing format will erase your data.
3. Specify the hypothetical mean
0
1
100
2. Enter data
1: 106
2: -20
4. View the results
Calculate now
Clear the form

proviamoci da soli

1:	106
2:	-20
3:	101
4:	-33
5:	72
6:	-36
7:	62
8:	38
9:	-70
10:	127
11:	24

One sample *t* test results

P value and statistical significance:
The two-tailed P value equals 0.1218
By conventional criteria, this difference is considered to be not statistically significant.

Confidence interval:
The hypothetical mean is 0.00
The actual mean is 33.73
The difference between these two values is 33.73
The 95% confidence interval of this difference:
From -10.73 to 78.18

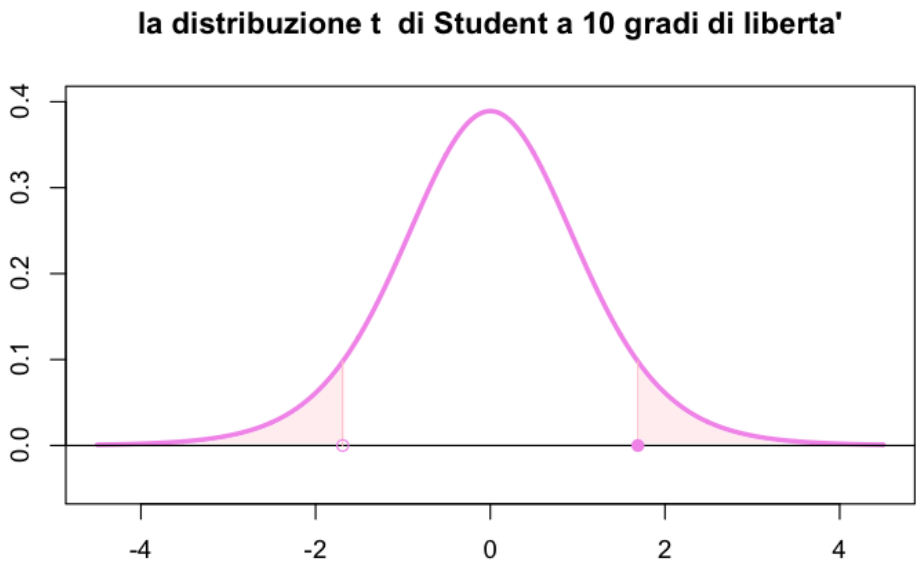
Intermediate values used in calculations:
t = 1.6905
df = 10
standard error of difference = 19.951

Review your data:

Mean 33.73
SD 66.17
SEM 19.95
N 11

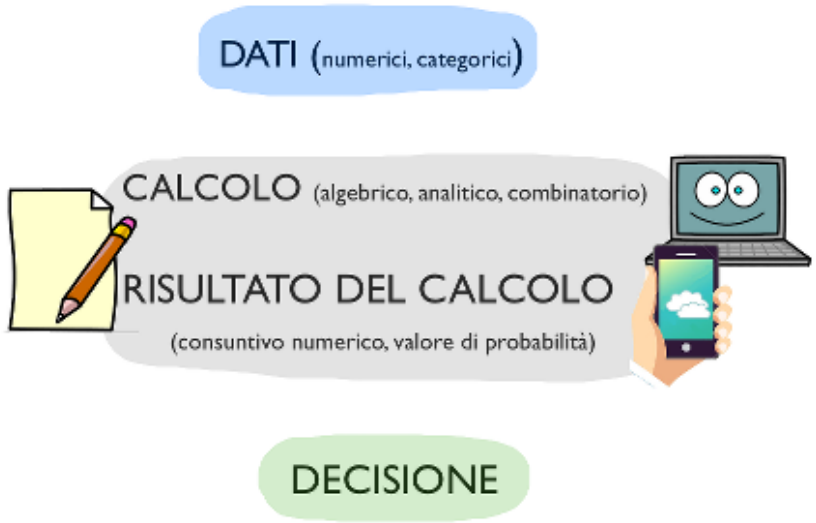
Una celebre storia ..

$$\frac{33.7 - 0}{\frac{66.2}{\sqrt{11}}} = 1.69$$



$$P(-1.69) = 0.06$$

Una celebre storia ..



Risultato dei calcoli

- t = 1.69
- P = 0.12

interpretazione

- c'è un 12% di probabilità che la discrepanza tra la media della popolazione $\mu = 0$ e la media del campione $M = 33.7$ sia dovuta al caso e non all'efficacia del trattamento

decisione

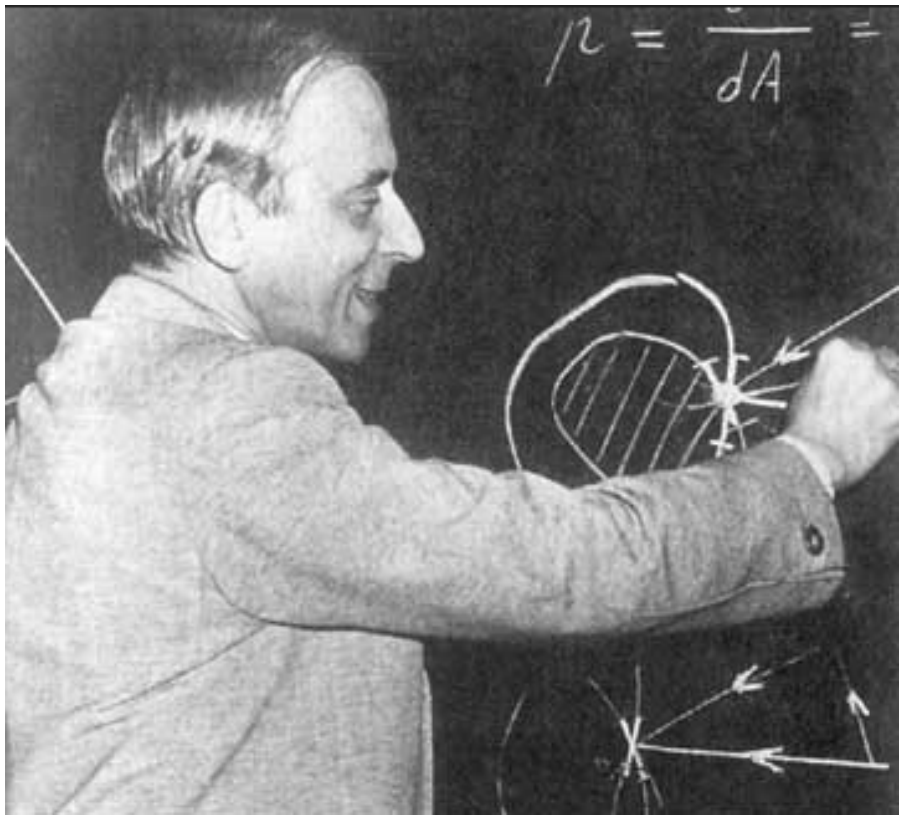
- non abbiamo evidenze sufficienti per ritenere che l'essicazione sia efficace.

domande

- Invece, oggi ..

- un consuntivo (italiota = 'statistica test')
- un p-value
- un intervallo di fiducia (italiota = 'intervallo di confidenza')

Invece oggi: approccio bayesiano vs. approccio frequentista



Invece oggi: frequentisti

```
> resa = c(106, -20, 101, -33, 72, -36, 62, 38, -70, 127, 24)
> t.test(resa, mu = 0)

   One Sample t-test

data:  resa
t = 1.6905, df = 10, p-value = 0.1218
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -10.72710  78.18164
sample estimates:
mean of x
 33.72727
```

Invece oggi: bayesiani

```
> formula = resa ~ 1
> output = inla(formula, data = Gosset)
> summary(output)

Call:
"inla(formula = formula, data = Gosset)"

Time used:
 Pre-processing    Running inla Post-processing
         0.3595          0.1301           0.0721

Fixed effects:
               mean     sd 0.025quant 0.5quant 0.975quant
(Intercept) 33.7273 19.933    -6.0463  33.7265    73.4669

The model has no random effects

Model hyperparameters:
                                        mean    sd 0.025quant 0.5quant 0.975quant  mode
Precision for the Gaussian observations 3e-04 1e-04      1e-04    3e-04      5e-04 2e-04

Expected number of effective parameters(std dev): 1.00(0.00)
Number of equivalent replicates : 11.00

Marginal log-Likelihood:  -75.47
```