

# introduzione a G\*Power

Massimo Borelli

UMG School of PhD Programmes



## 1 Introduzione

## 2 Esempi

- recuperare informazioni descrittive

# testi

## Review

### Statistics review 4: Sample size calculations

Elise Whitley<sup>1</sup> and Jonathan Bell<sup>2</sup>

<sup>1</sup>Lecturer in Medical Statistics, University of Bristol, Bristol, UK  
<sup>2</sup>Lecturer in Intensive Care Medicine, St George's Hospital Medical School, London, UK

Correspondence: Editorial Office, Critical Care, editorial@ccforum.com

Published online: 18 May 2002

Critical Care 2002, 6:305–341  
This article is online at <http://ccforum.com/content/6/4/305>  
© 2002 BioMed Central Ltd (Print ISSN 1364-8535; Online ISSN 1469-6596)

## Abstract

The present review introduces the notion of statistical power and the hazard of under-powered studies. The problem of how to calculate an ideal sample size is also discussed within the context of factors that affect power, and specific methods for the calculation of sample size are presented for two common scenarios, along with extensions to the simplest case.

Keywords statistical power, sample size

Previous reviews in this series introduced confidence intervals and *P* values. Both of these have been shown to depend strongly on the size of the study sample in question, with larger samples generally resulting in narrower confidence intervals and smaller *P* values. The notion of how large a study should ideally be is therefore an important one, but it is all too often neglected in practice. The present review provides some simple guidelines on how best to choose an appropriate sample size.

Research studies are conducted with many different aims in mind. A study may be conducted to establish the difference or conversely the similarity, between two groups defined in terms of a particular risk factor or treatment regimen. Alternatively, it may be conducted to estimate some quantity, for example the prevalence of a disease in a specific population, with a given degree of precision. Regardless of the motivation for the study, it is essential that it be of an appropriate size to achieve its aims. The most common aim is probably that of determining whether there is a difference between two groups in some scenario that will be used as the basis for the remainder of the present review. However, the ideas underlying the methods described are equally applicable to all settings.

## Power

The difference between two groups in a study will usually be explored in terms of an estimate of effect, appropriate confidence interval and *P*-value. The confidence interval indicates the likely range of values for the true effect in the population,

while the *P* value determines how likely it is that the observed effect in the sample is due to chance. A related quantity is the statistical power of the study. Put simply, this is the probability of correctly identifying a difference between the two groups in the study sample if one genuinely exists in the populations from which the samples were drawn.

The ideal study for the researcher is one in which the power is high. This means that the study has a high chance of detecting a difference between groups if one exists; consequently, if the study demonstrates no difference between groups, the researcher can be reasonably confident in concluding that none exists in reality. The power of a study depends on several factors (see below), but as a general rule higher power is achieved by increasing the sample size.

It is important to be aware of this because of all too often studies are reported that are simply too small to have adequate power to detect the hypothesized effect. In other words, even when a difference does exist, it may not be detected because the sample size has been recruited. The result of this is that *P* values are higher and confidence intervals wider than would be the case in a larger study, and the erroneous conclusion may be drawn that there is no difference when in fact there is one. This notion is well summed up in the phrase 'absence of evidence is not evidence of absence'. In other words, an apparently null result that shows no differences between groups may simply be due to lack of statistical power, making it extremely unlikely that a true difference will be correctly identified.

J. P. Verma  
Priyam Verma

# Determining Sample Size and Power in Research Studies

A Manual for Researchers



Review

## Statistics review 4: Sample size

Elise Whitley<sup>1</sup> and Jonathan Ball<sup>2</sup>

<https://www.biomedcentral.com/collections/CC-Medical>



# Determining Sample Size and Power in Research Studies

J. J.P. Verma

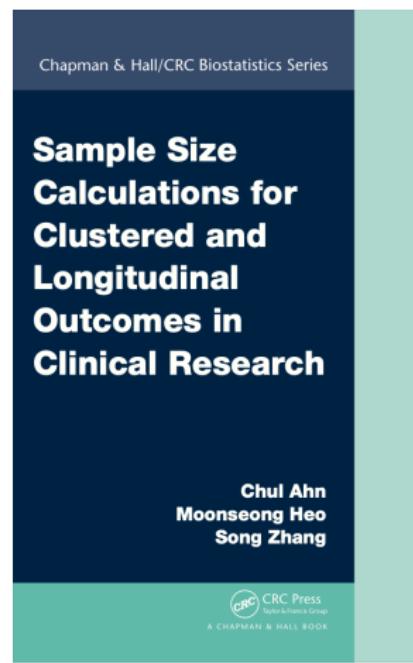
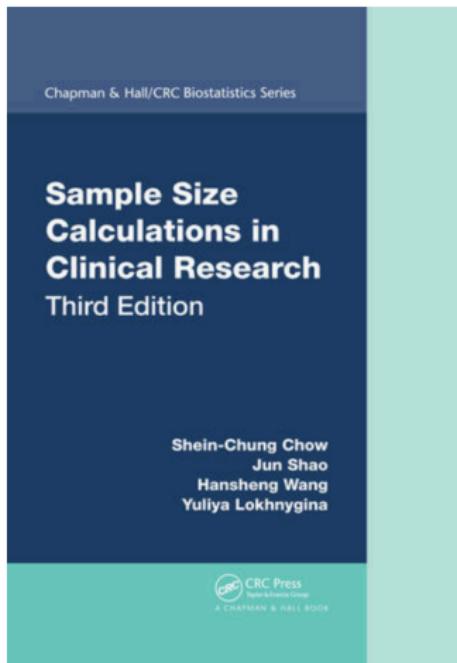
Top 3% 1265 Views 138 Pages 2 Files ▾

<https://www.academia.edu>

Download PDF



testi



## Chapter 2: Understanding Statistical Inference

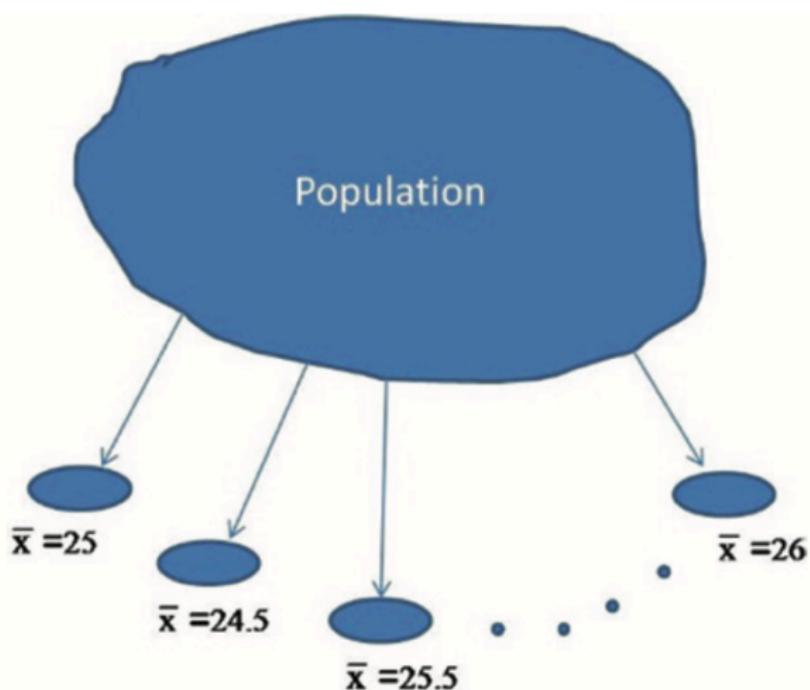


Fig. 2.1 Drawing all possible samples of the same size from the population

# Chapter 2: Understanding Statistical Inference

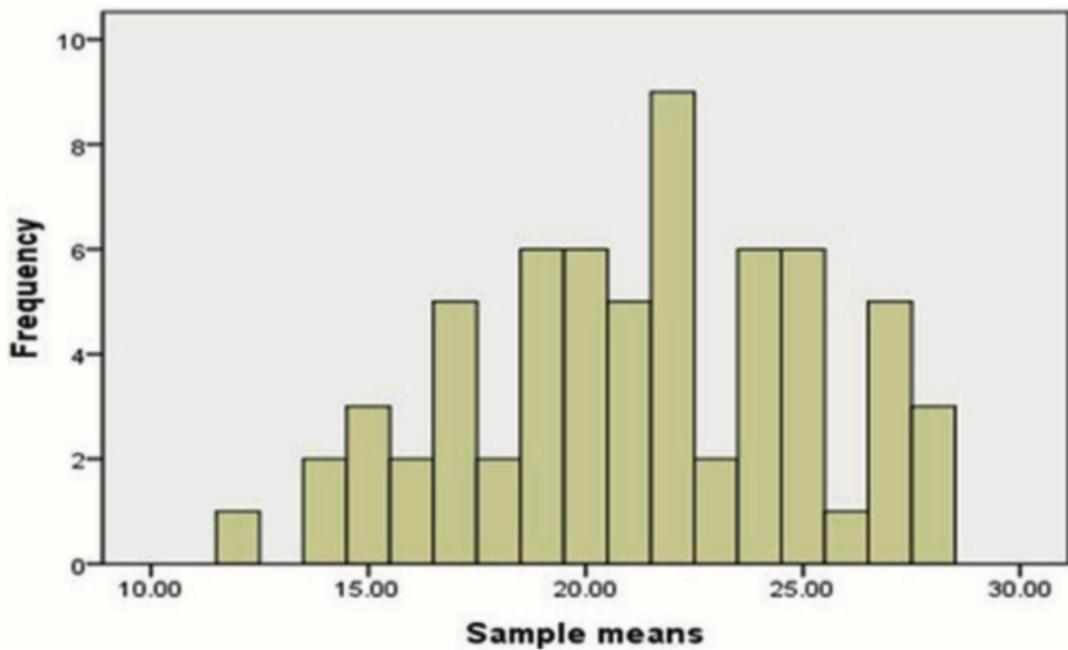


Fig. 2.3 Histogram of the sample means

# Chapter 2: Understanding Statistical Inference

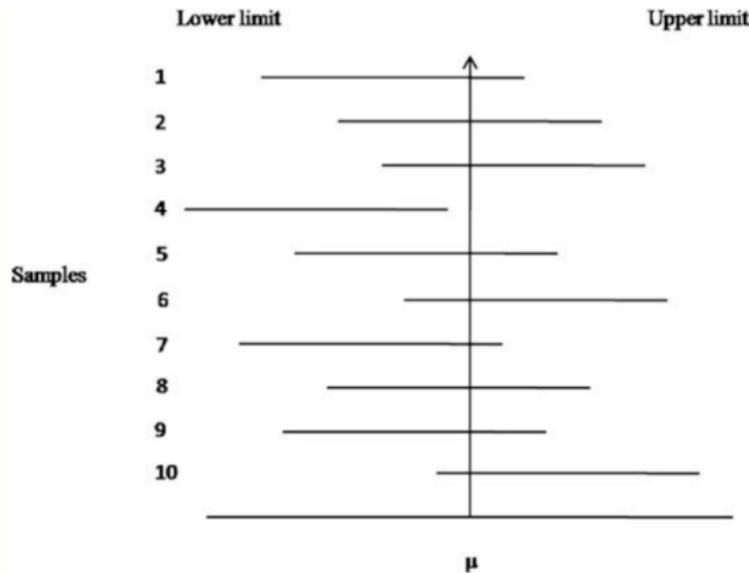


Fig. 2.7 Confidence intervals constructed on the basis of the sample for estimating population mean

# Chapter 2: Understanding Statistical Inference

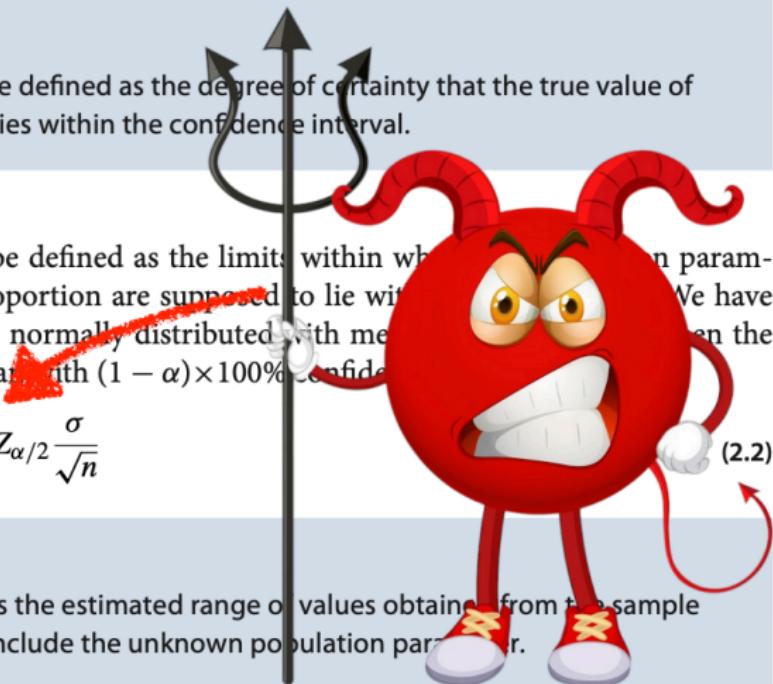
## Definition

Confidence level can be defined as the degree of certainty that the true value of population parameter lies within the confidence interval.

## Confidence Interval

Confidence interval can be defined as the limits within which population parameters such as mean or proportion are supposed to lie with a certain probability. We have seen that if a variable  $x$  is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ , then the confidence interval of mean  $\mu$  with  $(1 - \alpha) \times 100\%$  confidence is:

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



## Definition

A confidence interval is the estimated range of values obtained from the sample data which is likely to include the unknown population parameter.

# Chapter 2: Understanding Statistical Inference

Researcher's decision	Actual state of $H_0$	
	$H_0$ true	$H_0$ false
Reject $H_0$	Type I error ( $\alpha$ ) (False positive)	Correct decision ( $1 - \beta$ )
Do not reject $H_0$	Correct decision	Type II error ( $\beta$ ) (False negative)

- ① L'errore di primo tipo (falso positivo, condannare l'innocente)
- ② L'errore di secondo tipo (falso negativo, assolvere il colpevole)

# Chapter 2: Understanding Statistical Inference

- ① L'errore di primo tipo (falso positivo, condannare l'innocente)
- ② L'errore di secondo tipo (falso negativo, assolvere il colpevole)

## Definition

A Type I error can be defined as the probability of rejecting the null hypothesis when it is true. It is represented by  $\alpha$ .

## Definition

A Type II error is the probability of not rejecting the null hypothesis when it is false. It is represented by  $\beta$ .

# Domanda cruciale

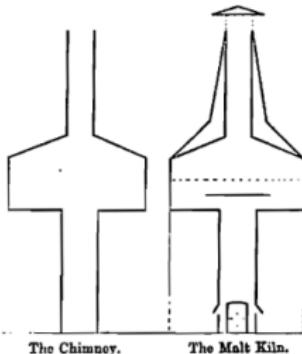
## Definition

A Type I error can be defined as the probability of rejecting the null hypothesis when it is true. It is represented by  $\alpha$ .

## Definition

A Type II error is the probability of not rejecting the null hypothesis when it is false. It is represented by  $\beta$ .

Quando facciamo un esperimento, uno studio clinico, siamo in grado di dire quanto valgono queste probabilità  $\alpha$  e  $\beta$  ?



By H. M. CHUBB.

Not Kiln-Dried	Kiln-Dried	Difference
1903	2009	+106
1935	1915	-20
1910	2011	+101
2496	2463	-33
2108	2180	+72
1961	1925	-36
2060	2122	+62
1444	1482	+38
1612	1542	-70
1316	1443	+127
1511	1535	+24

Tabella: One Sample T-Test

t	df	p	Mean Diff.	95% CI for Mean Diff.	
				Lower	Upper
1.690	10	0.122	33.727	-10.727	78.182

## SPOILER: con G\*Power si può

- basta considerare l'**effect size**  
(la  $d$  di Cohen)

$$d = \frac{m - \mu}{s}$$

t	p	Cohen's d
1.69	0.122	0.510



Tail(s) One

Effect size d 0,51

$\alpha$  err prob 0,05

Total sample size 11



Output parameters	
Noncentrality parameter $\delta$	1,6914786
Critical t	1,8124611
Df	10
Power (1- $\beta$ err prob)	0,4725641

# Domanda cruciale

## Definition

A Type I error can be defined as the probability of rejecting the null hypothesis when it is true. It is represented by  $\alpha$ .

## Definition

A Type II error is the probability of not rejecting the null hypothesis when it is false. It is represented by  $\beta$ .

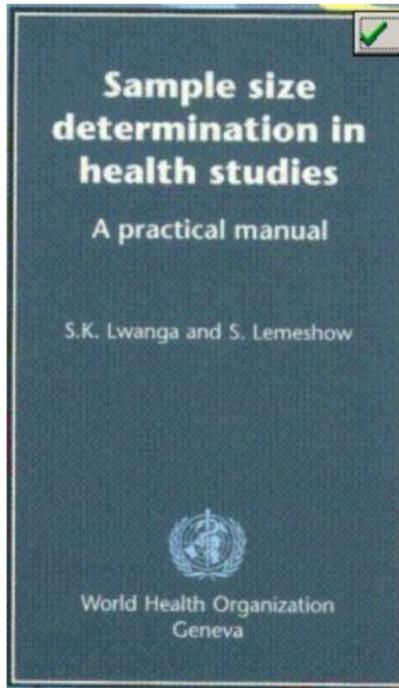
Nell'esperimento di Gossett, se scegliamo la probabilità  $\alpha = 5\%$  allora  $\beta = 53\%$  (poiché  $1 - \beta = 47\%$ )

$\beta = 0.53$ ? Vuol dire 'testa o croce'

Nell'esperimento di Gossett, se scegliamo la probabilità  $\alpha = 5\%$  allora  $\beta = 53\%$

It is important to be aware of this because all too often studies are reported that are simply too small to have adequate power to detect the hypothesized effect. In other words, even when a difference exists in reality it may be that too few study subjects have been recruited. The result of this is that  $P$  values are higher and confidence intervals wider than would be the case in a larger study, and the erroneous conclusion may be drawn that there is no difference between the groups. This phenomenon is well summed up in the phrase, 'absence of evidence is not evidence of absence'. In other words, an apparently null result that shows no difference between groups may simply be due to lack of statistical power, making it extremely unlikely that a true difference will be correctly identified.

# Perché 80% ?



# Perché 80% ?

[“Lehr’s Equation”](#) or the “Rule of 16”:

$$n_{group} = \frac{2(1.96 + 0.84)^2}{(\Delta)^2}$$

$$n_{group} = \frac{16}{(\Delta)^2}$$



PMID: 27688437

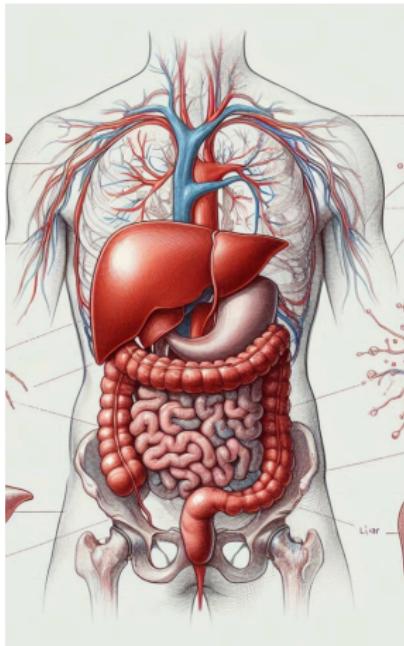
# Perché 80% ?

Martin Bland

La scelta della potenza è arbitraria, nel senso che una scelta ottimale della potenza per un certo studio non esiste; di solito io consiglio di porla pari a 90%, ma anche 80% è una scelta spesso citata. Nel secondo caso otterremo in genere una stima della dimensione campionaria più piccola, ma naturalmente andremo incontro ad un maggiore rischio di fallimento nell'individuare un effetto.

che ci aspettiamo. I metodi descritti in precedenza per la scelta della dimensione campionaria devono dunque rimanere linee guida, e probabilmente è sempre meglio sbagliare sovrastimando la dimensione campionaria richiesta, almeno quando l'obiettivo è giungere ad una decisione definitiva.

## esempio: angiosarcoma

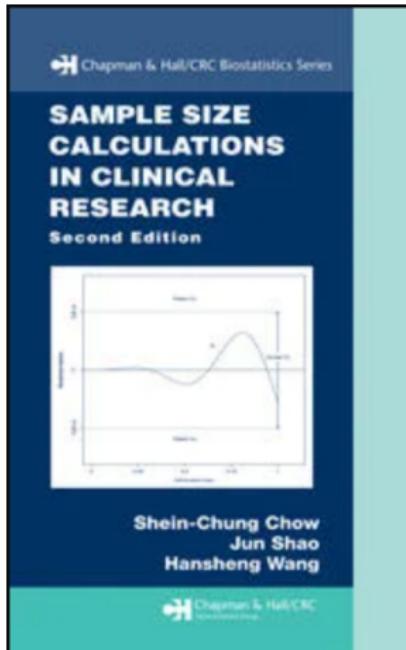


- 5 yrs survival rate: ~ 10%
- studio preclinico: 2/6 ok
  - ~ 30%



i test esatti per le proporzioni

il test binomiale: un solo campione



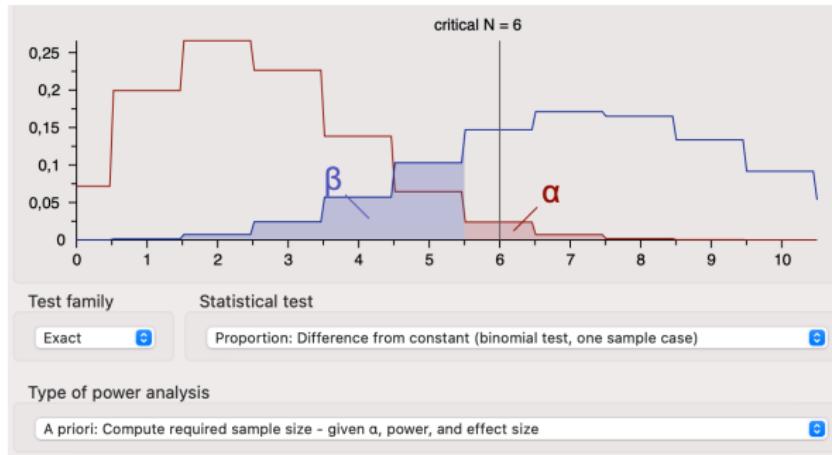
# i test esatti per le proporzioni

## il test binomiale: un solo campione

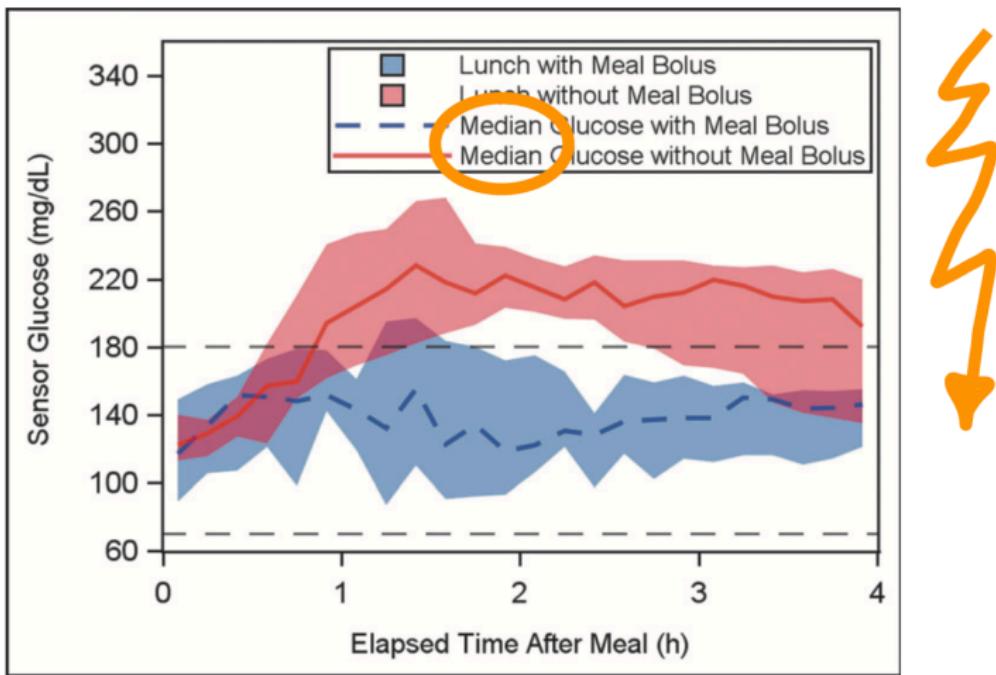
### 5.1.3 An Example

Suppose the investigator is interested in conducting a trial to study the treatment effect of a test compound in curing patients with certain types of cancer. The responder is defined to be the subject who is completely cured by the study treatment. According to literature, the standard therapy available on the market can produce a cure rate of 10% ( $p_0 = 10\%$ ). A pilot study of the test compound shows that the test compound may produce a cure rate of 30% ( $p_1 = 30\%$ ). The objective of the planning trial is to confirm such a difference truly exists. It is desirable to have a sample size, which can produce 80% power at 5% level of significance. According to **Table 5.1.1**, the total sample size needed is given by 25. The null hypothesis should be rejected if there are at least 5 subjects who are classified as responders.

# i test esatti per le proporzioni



Input parameters		Calc P2 from		Proportions	
<input checked="" type="radio"/> Difference $P_2 - P_1$ <input type="radio"/> Ratio $P_2 / P_1$ <input type="radio"/> Odds ratio		<input type="checkbox"/> 0,1 <input type="checkbox"/> 0,2 <input type="checkbox"/> 0,3 <input type="checkbox"/> Sync values		$P_1$ <input type="text" value="0,1"/> $P_2$ <input type="text" value="0,30"/>	
Tail(s): <input type="radio"/> One <input type="radio"/> Two <input type="radio"/> Both <input type="checkbox"/> Determine <input type="checkbox"/> Effect size g <input type="text" value="0,2"/> <input type="checkbox"/> $\alpha$ err prob <input type="text" value="0,05"/> <input type="checkbox"/> Power (1- $\beta$ err prob) <input type="text" value="0,8"/> <input type="checkbox"/> Constant proportion <input type="text" value="0,1"/>		<input type="button" value="Calculate"/>		Effect ? <input type="button" value="Calculate and transfer to main window"/>	



## recuperare informazioni descrittive / 1 of 2

Possiamo dedurre la media  $\mu$  dalla mediana, dai quartili e dal range?

$$\mu \approx \frac{a + 2m + b}{4} + \frac{a - 2m + b}{4n} \approx \frac{a + 2m + b}{4}$$

$$\mu \approx \frac{a + 2Q_1 + 2m + 2Q_3 + b}{8}$$

$$\mu \approx \frac{Q_1 + m + Q_3}{3}$$

## recuperare informazioni descrittive / 2 of 2

Possiamo dedurre la deviazione standard  $\sigma$  dalla mediana, dai quartili e dal range?

$$\sigma \approx \frac{b - a}{\xi(n)}$$

$$\sigma \approx \frac{Q_3 - Q_1}{\eta(n)}$$

$$\sigma \approx \frac{1}{2} \left( \frac{b - a}{\xi(n)} + \frac{Q_3 - Q_1}{\eta(n)} \right)$$

# Un esempio



- the Reference Data Set
  - dal 2003 al 2014, King's College London Dental Institute
- LL8Gf
  - Lower Left Third Molar
  - Stage G female
  - 18 years old
  - caucasian UK

# Un esempio

<https://www.dentalage.co.uk/rds-uk-caucasian/>

$n = 114$

$$\mu \approx \frac{a + 2Q_1 + 2m + 2Q_3 + b}{8}$$
$$\frac{14.9 + 34.2 + 36.2 + 39.9 + 22.8}{8}$$
$$\mu \approx 18.5$$

0th%ile	0.5%ile	5%ile	10%ile
14.96	15.73	15.97	16.32
25%ile	50%ile	75%ile	90%ile
17.07	18.13	18.99	20.64
95%ile	99.5%ile	100th%ile	
21.35	21.91	22.78	

# Un esempio

<https://www.dentalage.co.uk/rds-uk-caucasian/>

$n = 114$

$$\sigma \approx \frac{1}{2} \left( \frac{b - a}{\xi(n)} + \frac{Q_3 - Q_1}{\eta(n)} \right)$$

$$\frac{1}{2} \left( \frac{22.8 - 14.9}{4.55} + \frac{19.0 - 17.1}{1.35} \right)$$

$$\sigma \approx 1.6$$

0th%ile	0.5%ile	5%ile	10%ile
14.96	15.73	15.97	16.32
25%ile	50%ile	75%ile	90%ile
17.07	18.13	18.99	20.64
95%ile	99.5%ile	100th%ile	
21.35	21.91	22.78	

## Un esempio

$$\mu \approx 18.5$$

$$\sigma \approx 1.6$$

TDS	n	mean	sd
LL8Gf	114	18.25	1.63