P values and Confidence Intervals Friends or Foe

Dr.L.Jeyaseelan
Dept. of Biostatistics
Christian Medical College
Vellore – 632 002, India

JPGM WriteCon March 30-31, 2007, KEM Hospital, Mumbai

Confidence Interval

- The mean or proportion observed in a sample is the best estimate of the true value in the population.
- We can combine these features of estimates from a sample from a sample with the known properties of the normal distribution to get an idea of the uncertainty associated with a single sample estimate of the population value.

Confidence Interval

- The interval between Mean - 2SE and Mean + 2 SE will be 95%.
- That is, we expect that 95% CI will not include the true population value 5% of the time.
- No particular reason for choosing 95% CI other than convention.

Confidence Interval

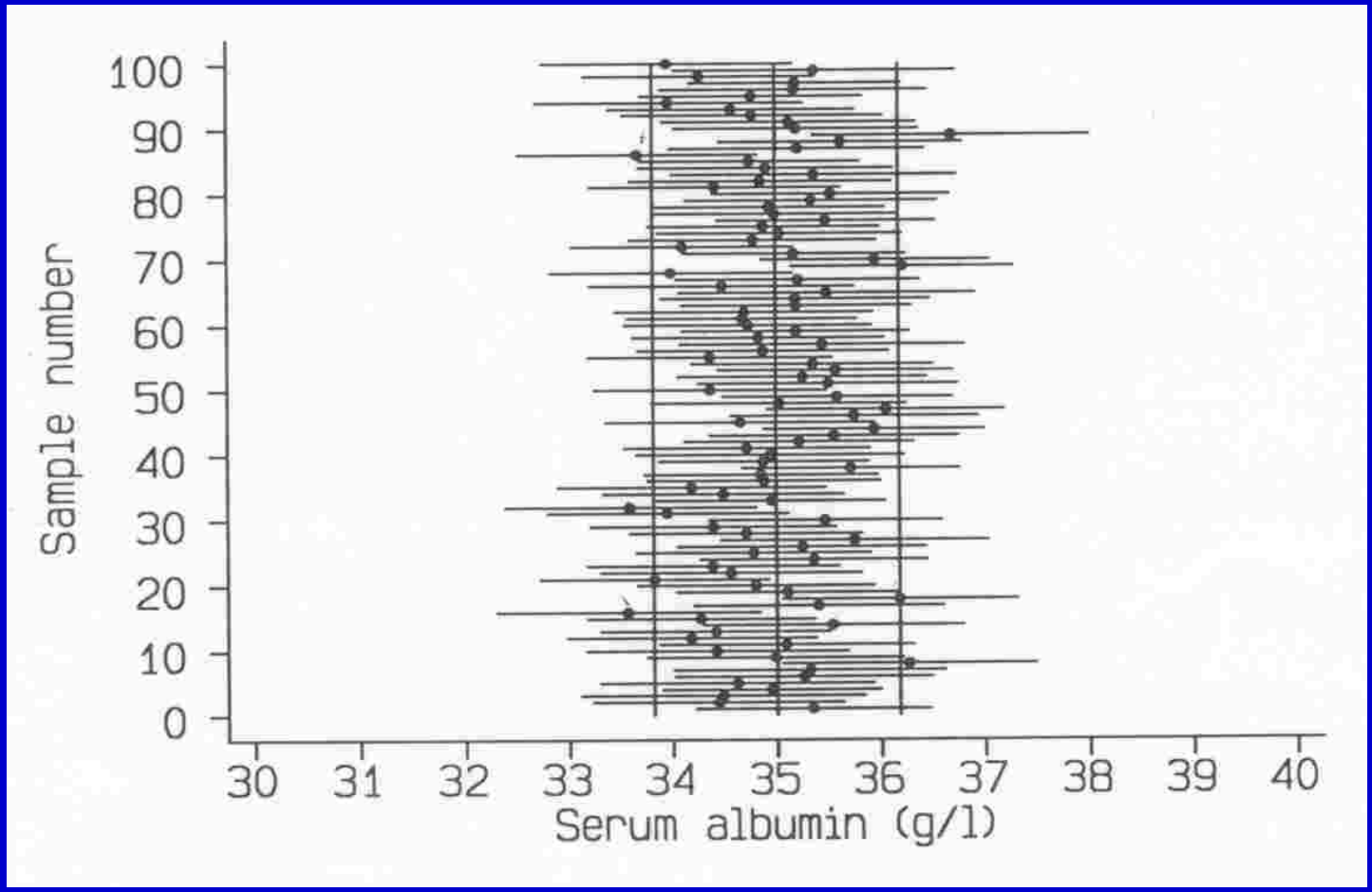
Interpretation:

The 95% CI for the sample mean is interpreted as a range of values which contains the true population mean with probability 0.95%.

Ex: Mean serum albumin of the population (source) was 35 g/l.

Calculate 95% CI for the mean serum albumin using each of the 100 random samples of size 25.

Scatter Plot of 95% CI:



Confidence intervals for mean serum albumin constructed from 100 random samples of size 25. The vertical lines show the range within which 95% of sample means are expected to fall.

Confidence Interval

- CI expresses a summary of the data in the original units or measurement.
- It reflects the variability in the data, sample size and the actual effect size
- Particularly helpful for non-significant findings.

Relative Risk	1.5
95% CI	0.6 – 3.8

Confidence Interval for Low Prevalence

- Asymptotic Formula:

  $p \pm 1.96\ SE$

  provides symmetric CI.

- When the prevalence is small, negative lower limit can occur.

  Ex: The observed proportion is 0, the symmetric method produces the interval 0 to 0.

- Need to use exact or more complex method to compute the CI

Dose-response relationship between severity of hiccups and hyponatraemia.

	Control group	Hiccups group		
		Mild	Moderate	Severe
n	50	23	12	15
Number with serum sodium <130 mEq/L	4	9	10	15
Crude odds ratio	-	7 (CI: 2-34)	58 (CI: 8-587)	320 (CI: 20-829)
Mean (SD) serum sodium	136 (4)	131 (6)	125 (9)	119 (6)

(Geore J et al. Natl. Med.J India 1996)

Distribution of Hyponatraemia and Severe Hiccups:

	Severe Hiccups	Control
Low sodium	15	4
Normal sodium	0	46
Total	15	50

OR	320
95% CI	20 – 829
p value	< 0.001

Thalidomide Disaster

		Congenital Malformation Yes	No	Total
Thalidomide	Yes	41	0	41
	No	5	300	305
Total		46	300	346

OR = 61
95% CI = 26 - 146
p <.001

P Value

P value

- **Analyses involve comparison (Testing Hypothesis):**
  - Serum cholesterol is higher in men than women
  - Certain combination of temperature and light optimizes cell growth in a laboratory experiment.
  - New treatment for Psoriasis relieves the symptoms of patients in men than women.

P value

- Null Hypothesis:

  No difference between two or more groups.

  Ex: the serum cholesterol is the same on average for men and women.

P value

- **Alternate Hypothesis:**

  There is a difference between two or more groups.

  (basically the study was initiated based on this expectation).

P value

- Probability of having observed our data (or more extreme data) when the null hypothesis is true.
    - In clinical trial scenario the above refers to the difference between the treatment groups.
    - *Therefore, the likely variation in a sample due to chance when the null hypothesis is true in the population.*

P value

- **Large P value:**

  If $p > 0.2$

  The data could occur often when the null hypothesis is true.

  Does it imply that the treatments are equally effective?

- **Small P value:**

  If $p < .001$

  The null hypothesis appears implausible. Our data could hardly arise purely by chance when the null hypothesis is true.

P value

- Cut offs are arbitrary and have no specific importance ($p<.05$, $p<.01$ ..)

$p=0.055$ $p=0.045$

Now it is common to see p values as

$p=0.02$ $p=0.15$

Reaction to Abuses Due to misinterpretation

- A movement that aimed at improving interpretation caused a disaster.

  "The author of one submission to a journal that publishes articles on Epidemiology was asked by an editor that "all references to statistical hypothesis testing and statistical significance be removed from the paper"

  The author of another submission was to that journal was told "We are trying to discourage use of the concept of 'statistical significance' which is in our view, outmoded and potentially misleading. Thus, I would ask that you delete references to p values and statistical significance.

Fleiss JL. AJPH 1986.

P value – When?

- **Confounders and Model Building:**

  To label a variable as confounder, a subject matter expert and statistician work together.

  In model building, we select potentially significant variables using $p<0.10$ or $p<.20$ etc using bi-variate analyses.

  Then we asses the significance of the hypothesized risk factor on outcome, keeping all the potential confounders and the risk factors in the model.

P value – When?

- Nonparametric survival analysis:

  Outcome: Time to event data

  The most popular procedure for testing whether two sample survival curves

  Log Rank test

P value – When?

- **Psychological Research:**

- The experimental hypothesis differs from the theoretical construct of interest to the researchers.

  "The CI is not a satisfactory alternative to the use of significance tests for decision making in the frame work of theory corroboration or refutation."

Chow SL. Psychological Bull 1988.

Analyses of Contingency Table

- The interaction between 2 or more variables each of which is observed at several levels, can be reported by a single significance test.

  Ex: Chi-square test with 16 DF for over all association between the row and column variables of a 5x5 table.

  “You can not express the association with 16 CI”

Statistical Vs Clinical Significance

Ex: In a study to compare the blood pressure in the left and the right arms, a small difference of about 1 mm Hg (both systolic and diastolic) was found. This difference was highly statistically significant, but of no importance clinically.

Is the blood pressure same in both arms?
Gould et al. Clin. Cardiology 1985.

Antenatal Steroids for Preventing Neonatal Respiratory Distress Syndrome.

(Am J Obste Gynecol 1981)

	Pre – Eclampsia			No Eclampsia		
	RDS		Total	RDS		Total
	Yes	No		Yes	No	
Steroid	7 (21.2)	26	33	21 (7.9)	246	267
Placebo	9 (27.3)	24	33	37 (14.1)	225	262
	p = 0.51			p = 0.021		

Mathews and Altman. BMJ 1996

Pre – Eclampsia:

Difference in percentage 21.2 – 27.3 = 6.1%

Eclampsia:

Difference in percentage 7.9 – 14.1 = 6.2%

Reasons:

* Mothers who had pre-eclampsia were small.
* (No evidence to conclude that Rx is effective)
* Sample size was not designed to study

Current Scenario

- “Several prominent medical or epidemiological journals have moved to require authors to put greater emphasis on estimation of effect sizes in results, particularly through the use of CI….”

- BMJ, Lancet, Annals of Internal Medicine now specify that one should calculate CI whenever, the data warrant this approach

Walter SD, Am J Epidemiol 1995.

Summary

- From the methodological viewpoint, no single method of analysis and data presentation is to be preferred in all situations.

- A change in editorial policy alone is unlikely to eliminate future statistical errors.

Summary (contd..)

- Mandate from editors to use one for other is unsatisfactory or coercive.

- Instead, further work is needed on how consumers of medical research interprets studies and how they are influenced by the authors' method of summarizing the data.

Walter SD, Am J Epidemiol 1995.


Thank You