Statistical Inference by Using Confidence Intervals

General introduction

- HYPOTHESIS TESTING(p value) & ESTIMATION (Confidence Interval)
- Sample is assumed to be representative to the population.
  In research: measurement are always done in the sample, the results will be applied to population.
- (target population ➔ accessible population ➔ intended sample ➔ actual study subjects:
  - **Target population** is determined.
  - **Accessibility of population:** is usually based on practical purposes.
  - The **intended sample** is the sample selected for the study by an appropriate sampling technique.
  - The **actual study subjects** are those who completed the study and excludes the non-response, drop outs, withdrawals, and loss to follow-up)
- There are many journals that you can publish your results to, examples include:
  - -THE BRITISH MEDICAL JOURNAL
  - -THE LANCET
  - -THE MEDICAL JOURNAL OF AUSTRALIA
  - -THE AMERICAN JOURNAL OF PUBLIC HEALTH
  - -THE BRITISH HEART JOURNAL
  - And a lot more
- Statistic and Parameter
  - An observed value drawn from the sample is called a statistic
  - The corresponding value in population is called a parameter
  - We measure, analyze, etc statistics and translate them as parameters
  - Examples of statistics:
    - Proportion
    - Mean
    - Median
    - Mode
    - Difference in proportion/mean
    - OR
    - RR
    - Sensitivity
    - Specificity
    - Kappa
    - LR
    - NNT
- There are 2 ways in inferring statistic into parameter
  - Hypothesis testing → p value
  - Estimation: → Confidence interval (CI)

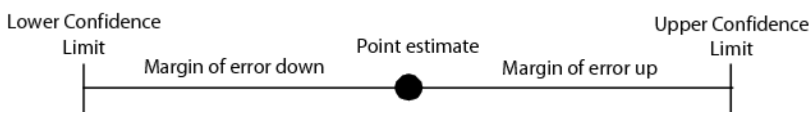
P Value & CI tell the same concept in different ways

P value

- Determines the probability that the observed results are caused solely by chance (probability to obtain the observed results if Ho were true)

Estimation

- Two forms of estimation
  - Point estimation = single value, e.g., x-bar is unbiased estimator of μ
  - Interval estimation = range of values ⇒ confidence interval (CI). A confidence interval consists of:



Confidence Interval

- Estimates the range of values (parameter) in the population using a statistic in the sample (as point estimate)
  - Even a well designed study can give only an idea of the answer sought , because of random variation in the sample.
  - And the results from a single sample are subject to statistical uncertainty, which is strongly related to the size of the sample.
  - The quantities(single mean, proportion, difference in means, proportions, OR, RR, Correlation,…etc) will be imprecise estimate of the values in the overall population, but fortunately the imprecision can itself be estimated and incorporated into the presentation of findings.
  - Example:
    - In a representative sample of 100 observations of heights of men, drawn at random from a large population, suppose the sample mean is found to be 175 cm (sd=10cm) .
    - Can we make any statements about the population mean ?
    - We cannot say that population mean is 175 cm because we are uncertain as to how much sampling fluctuation has occurred.
    - What we do instead is to determine a range of possible values for the population mean, with 95% degree of confidence.
    - **This range is called the 95% confidence interval** and can be an important adjuvant to a significance test.
    - Presenting study findings directly on the scale of original measurement together with information on the inherent imprecision due to sampling variability, has distinct advantages over just giving 'p-values' usually dichotomized into "significant" or "non-significant".
    - "THIS IS THE RATIONALE FOR USING CI"
- In general, the 95% confidence interval is given by:

- Statistic ± confidence factor x S.Error of statistic
    - In the previous example, n =100 ,sample mean = 175, S.D., =10, and the S.Error $=\frac{10}{\sqrt{100}}= 1$.
    - Therefore, the 95% confidence interval is

    175 ± 1.96 * 1 = 173 to 177"

  (That is, if numerous random sample of size 100 are drawn and the 95% confidence interval is computed for each sample, the population mean will be within the computed intervals in 95% of the instances.)

- Different Interpretations of the 95% confidence interval:
  - "We are 95% sure that the TRUE parameter value is in the 95% confidence interval"
  - "If we repeated the experiment many many times, 95% of the time the TRUE parameter value would be in the interval"
  - "the probability that the interval would contain the true parameter value was 0.95."

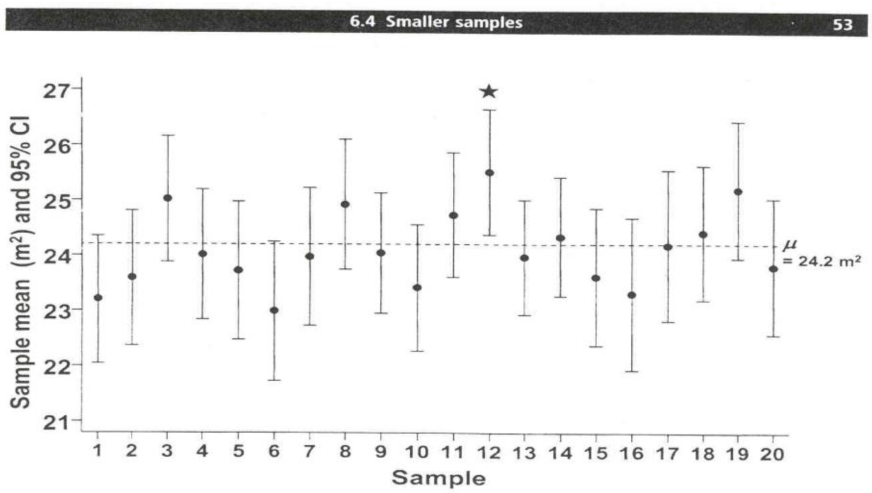


Fig. 6.2 Mean sprayable areas, with 95% confidence intervals, from 20 samples of 100 houses in a rural area. The star indicates that the CI does not contain the population mean

- But why do we always see 95% CI's?
  - "[Hypothesis tests] are sometimes overused and their results misinterpreted."
  - "Confidence intervals are of more than philosophical interest, because their broader use would help eliminate misinterpretations of published results."
  - "Frequently, a significance level or p-value is reduced to a 'significance test' by saying that if the level is greater than 0.05, then the difference is 'not significant' and the null hypothesis is 'not rejected'….
- Confidence Intervals for Reporting Results
  - "Duality" between confidence intervals and p-values
  - **Example**: Assume that we are testing that for a significant change in QOL due to an intervention, where QOL is measured on a scale from 0 to 50.
    - 95% confidence interval: (-2, 13)
    - P value = 0.07

- If the 95% confidence interval overlaps 0, then a t-test testing that the treatment effect is 0 will be insignificant at the alpha = 0.05 level.
- If the 95% confidence interval does not overlap 0, then a t-test testing that the treatment effect is 0 will be significant at the alpha = 0.05 level.

- Most commonly used CI:
  - CI 90% corresponds to p 0.10
  - CI 95% corresponds to p 0.05
  - CI 99% corresponds to p 0.01
- Notes:
  - p value → only for analytical studies
  - CI for descriptive and analytical studies
- How to calculate CI
  - General Formula:
  - CI = p $\pm$ Zα x SE
    - p = point of estimate, a value drawn from sample (a statistic)
    - Z_α = standard normal deviate for, if = 0.05→ Zα = 1.96 (~ 95% CI)
- Example 1: Confidence Interval of the Proportion
  - 100 KKUH students → 60 do daily exercise (p=0.6)
  - What is the proportion of students do daily exercise in the KKUH ?
    - Confidence interval (CI) = p $\pm$ Zα x SE
      - p=0.6
      - Z_α=1.96 (because confidence interval 95%)
      - SE standard error= $\sqrt{\frac{pq}{n}}$ = $\sqrt{\frac{0.6\times0.4}{100}}$= 0.05

      Confidence Interval of 95%(CI) = 0.6 ± 1.96 ×0.05 = 0.6 $\pm$ 1

      Confidence Interval of 95% (CI) ≈ 0.5 ; 0.7
    - We are 95% sure that 50% to 70% of KKUH students do daily exercise
- Example 2: Confidence Interval of the Mean
  - 100 newborn babies, mean body weight = 3000 (SD = 400) grams, what is 95% CI?

    95% CI = x $\pm$ 1.96 x SEM (SEM: Standard Error of Mean = $\frac{SD}{\sqrt{N}}$)

    = 3000 $\pm$ 1.96$\times\frac{400}{\sqrt{100}}$

    = 3000 $\pm$ 80 = (3000-80);(3000+80)

    = 2920 ; 3080
- Examples 3: CI of difference between proportions (p1-p2)
  - 50 patients with drug A, 30 cured (p1=0.6)
  - 50 patients with drug B, 40 cured (p2=0.8)
    - 95% CI (p_1-p_2) = (p_1-p_2) ± Z_α × **SE (p_1-p_2)** (SE: standard error)

      SE (p_1-p_2) = $\sqrt{\frac{P1Q1}{n1}+\frac{P2Q2}{N2}}$ = $\sqrt{\frac{0.6\times0.4}{50}+\frac{0.8\times0.2}{50}}$ = $\sqrt{\frac{0.4}{50}}$ = **0.09**
    - 95% CI (p_1-p_2) = (0.6-0.8) ± 1.96 × **0.09** ≈ -0.2 ± 0.18
    - 95% CI (p_1-p_2) = - 0.02 ; - 0.38

- Example 4: CI for difference between 2 means
    - Compare the Mean systolic BP between smokers and non-smokers: 50 smokers (mean=146.4, SD=18.6 mmHg) and 50 non-smokers (mean=140.4, SD=16.2 mmHg)

      SOLUTION:
        - 95% CI (x1-x2) = (x1-x2) $\pm$ 1.96 x **SE (x1-x2)**

      SE (x1-x2) = **3.53** (equations are hard but put here for reference)

$$s = \sqrt{\frac{(n_1-1)s_1^{\,2}+(n_2-1)s_2^{\,2}}{(n_1+n_2-2)}}$$

$$s = \sqrt{\frac{(49\times 18.6)+49\times 16.2}{98}} = 17.7$$

$$SE(x_1-x_2) = 17.7\times\sqrt{\left(\frac{1}{50}+\frac{1}{50}\right)} = \mathbf{3.53}$$

        - 95% CI (x1-x2) = 6.0 $\pm$ 1.96 x **3.53** ≈ 6.0 ± 7
        - 95% CI (x1-x2) = -1 ; 13.

Proportions

- Proportion
- Binary outcome (e.g. yes/no)
- Number between 0 and 1
- 2x2 table

	Group 1	Group 2
+ve	P1	P2
-ve	N1	N2

- Effect sizes
    - Risk Difference (RD); Relative Risk (RR); Odds Ratio (OR)
- Other commonly supplied CI
    - Relative risk (RR)
    - Odds ratio (OR)
    - Sensitivity, specificity (Se, Sp)
    - Likelihood ratio (LR)
    - Relative risk reduction (RRR)
    - Number needed to treat (NNT)
- Examples
    - RR = 5.6 (95% CI = 1.2 ; 23.7)
    - OR = 12.8 (95% CI = 3.6 ; 44,2)
    - NNT = 12 (95% CI = 9 ; 26)

- If p value <0.05, then 95% CI:
    - exclude 0 (for difference), because if A=B then A-B = 0 → p>0.05
    - exclude 1 (for ratio), because if A=B then A/B = 1, → p>0.05

APPLICATION OF CONFIDENCE INTERVALS

Example 1: comparing two treatments

- The following finding of non-significance in a clinical trial on 178 patients.

Treatment	Success	Failure	Total
A	76 (75%)	25	101
B	51(66%)	26	77
Total	127	51	178

- Chi-square value = 1.74 (p > 0.1)
- (non –significant)
- i.e. there is no difference in efficacy between the two treatments.
- The observed difference is: 75% - 66% = 9%
- and the 95% confidence interval for the difference is: 4% to 22%
- This indicates that compared to treatment B, treatment A has, at best an appreciable advantage (22%) and at worst , a slight disadvantage (- 4%).
- This inference is more informative than just saying that the difference is non significant.

Example: comparing two treatments

Treatment	Success	Failure	Total
A	49 (82%)	11	60
B	33 (60%)	22	55
Total	82	33	115

- The chi-square value = 6.38
- p = 0.01 (highly significant)
- The observed difference in efficacy is
- 82 % - 60% = 22%
- 95% C .I. = 6 % to 38%
- This indicates that changing from treatment B to treatment A can result in 6% to 38% more patients being cured.
- Again, this is more informative than just saying that the two treatments are significantly different.

Example 3: Disease Control Program

- Consider the following findings pertaining to case-holding in the National TB control program

Year	Program	Completed Treatment	Failed to Complete	Total
1987	Routine	276 (46%)	324	600
1988	Special	312 (52%)	288	600

- The chi square = 4.32, p-value = 0.04 (significant)
- The impact of special motivation program
  = 52 % - 46% = **6% in terms of improved**
- **case-holding.**
  - The 95% C.I. = 0.4 % to 11.6%, which indicates that the benefit from the special motivation program is not likely to be more than 11.6%.
  - This information may helps the investigator to conclude that the special program was not really worthwhile, and that other strategies need to be explored, to provide a greater magnitude of benefit.

CHARACTERISTICS OF CI'S

- The (im) precision of the estimate is indicated by the width of the confidence interval.
- The wider the interval the less precision
- THE WIDTH OF C.I. DEPENDS ON:
  - SAMPLE SIZE
  - VAIRABILITY
  - DEGREE OF CONFIDENCE

CONFIDENCE INTERVALS RATHER THAN P VALUES

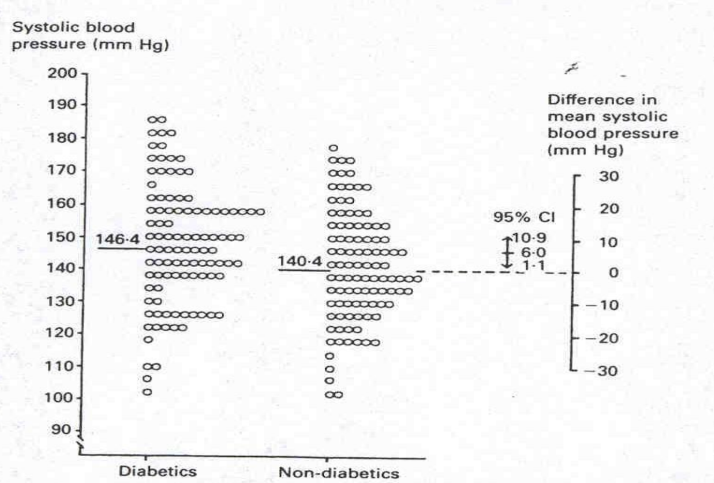


FIG 2.1—Systolic blood pressures in 100 diabetics and 100 non-diabetics with mean levels of 146·4 and 140·4 mm Hg respectively. The difference between the sample means of 6·0 mm Hg is shown to the right together with the 95% confidence interval from 1·1 to 10·9 mm Hg.

CONFIDENCE INTERVALS RATHER THAN P VALUES

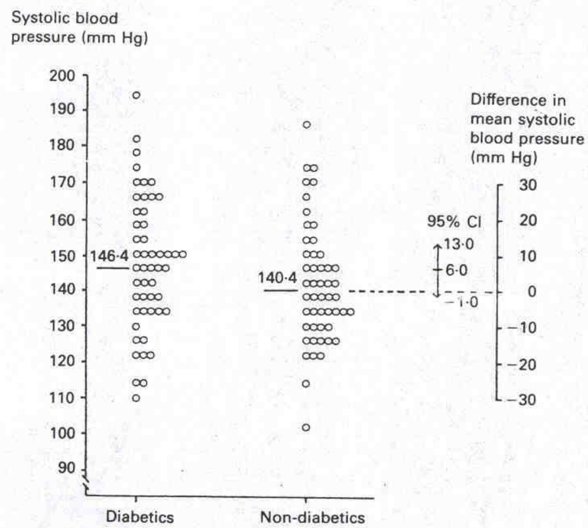


FIG 2.2—As fig 2.1 but showing results from two samples of half the size—that is, 50 subjects each. The means and standard deviations are as in fig 2.1, but the 95% confidence interval is wider, from −1·0 to 13·0 mm Hg, owing to the smaller sample sizes.

Clinical importance vs. statistical significance

Sometimes tests produce very highly statistical difference that is not clinically significant. (this usually occurs in large sample size) and vice versa.

Example 1: After comparing 5000 people with intervention of a drug to 5000 controls for reducing cholesterol level, there was highly statistical significant that the drug decrease the cholesterol level 2 mmHg (p=0.0023, significant). However, reduction for 2 mmHg is not important.

Example 2: This is a table to compare between standard and new treatment effect:

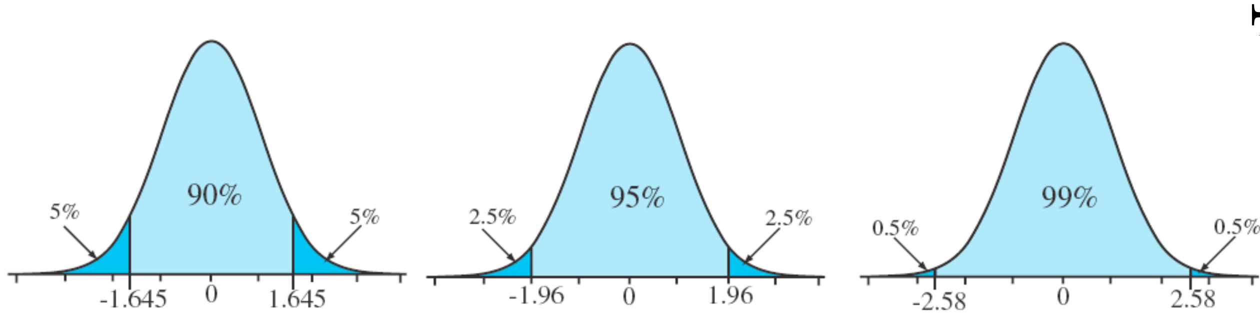
	Yes	No	Total
Standard	0	10	10
New	3	7	10

Clinically; there was 30% risk reduction which is good, but stastically (Fischer's exact test: p=0.211) it was insignificant, but still it makes you think that it might give considerable effect so larger sample size should be taken.

EFFECT OF VARIABILITY

- Properties of error
    - **Error increases with smaller sample size:** For any confidence level, large samples reduce the margin of error
    - **Error increases with larger standard Deviation:** As variation among the individuals in the population increases, so does the error of our estimate
    - **Error increases with larger z values:** Tradeoff between confidence level and margin of error
- Precision
    - The margin of error is a function of:
        - the population standard deviation
        - the confidence level
        - the sample size.
- If everything else remains the same, then
    - The ***larger*** the sample size, the ***narrower*** the CI.
    - The ***higher*** the confidence level, the ***wider*** the CI;
    - The ***larger*** the population SD, the ***wider*** the CI.
- Not only 95%….
    - 90% confidence interval:
        - mean ± 1.65 standard error of mean
        - NARROWER than 95%
    - 99% confidence interval:
        - mean ± 2.58 standard error of mean
        - WIDER than 95%
    - Other degrees of confidence

C.I. (degree of confidence) 1-α	Alpha level (α)	Z -value
90%	0.1	1.64
95%	0.05	1.96
98%	0.02	2.33
99%	0.01	2.58



CONFIDENCE INTERVALS RATHER THAN P VALUES

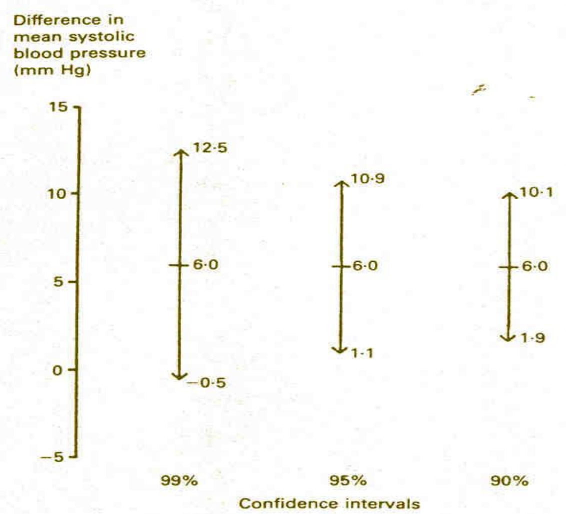


FIG 2.3—Confidence intervals associated with differing degrees of "confidence" using the same data as in fig 2.1.

Other Confidence Intervals

- Hazard ratios
- median survival
- difference in median survival
- Correlation, Regression
- Adjusted OR, RR
- Non – parametric tests

Recap

- 95% confidence intervals are used to quantify certainty about parameters of interest.
- Confidence intervals can be constructed for any parameter of interest (we have just looked at some common ones).
- The general formulas shown here rely on the central limit theorem
- You can choose level of confidence (does not have to be 95%).
- Confidence intervals are often preferable to p-values because they give a "reasonable range" of values for a parameter.

Concluding remarks

- P value & confidence interval
    - p values (hypothesis testing) gives you the probability that the result is merely caused by chance or not by chance, it does not give the magnitude and direction of the difference
    - Confidence interval (estimation) indicates estimate of value in the population given one result in the sample, it gives the magnitude and direction of the difference
    - p value alone tends to equate statistical significance and clinical importance
    - CI avoids this confusion because it provides estimate of clinical values and also statistical significance
    - ➔ whenever applicable, provide CI especially for the main results of study
- In appraising study results, focus should be on confidence intervals rather than on p values.