p Values and Confidence Intervals

Zuber Mulla

Objectives

1. Define a Type I error

2. Define a Type II error.

Objectives (cont.)

3. Use a *p* value to determine if a result is statistically significant.

4. Interpret 95% and 99% confidence intervals and use them to identify a statistically significant result.

Explanations when you Observe or Don’t Observe an Association

- *Truth*

- Chance
- Bias
- Confounding

Hennekens & Buring

Chance

- **The observed relationship may be due to the play of chance, the “luck of the draw,” which can happen any time a sample of a population is examined.**

Evaluating the Role of Chance

- **Tests of statistical significance and confidence intervals (CI) can be used to evaluate the role of chance as an alternative explanation of an observed association between an exposure (e.g., a possible risk factor) and an outcome (e.g., a disease)**

Statistic

- **A measure computed from the data of a sample**

Statistic

- Also known as a
  - Point estimate (sometimes denoted by a “hat”)
  - Parameter estimate

Parameter

- **A measure computed from the data of a population**

Example

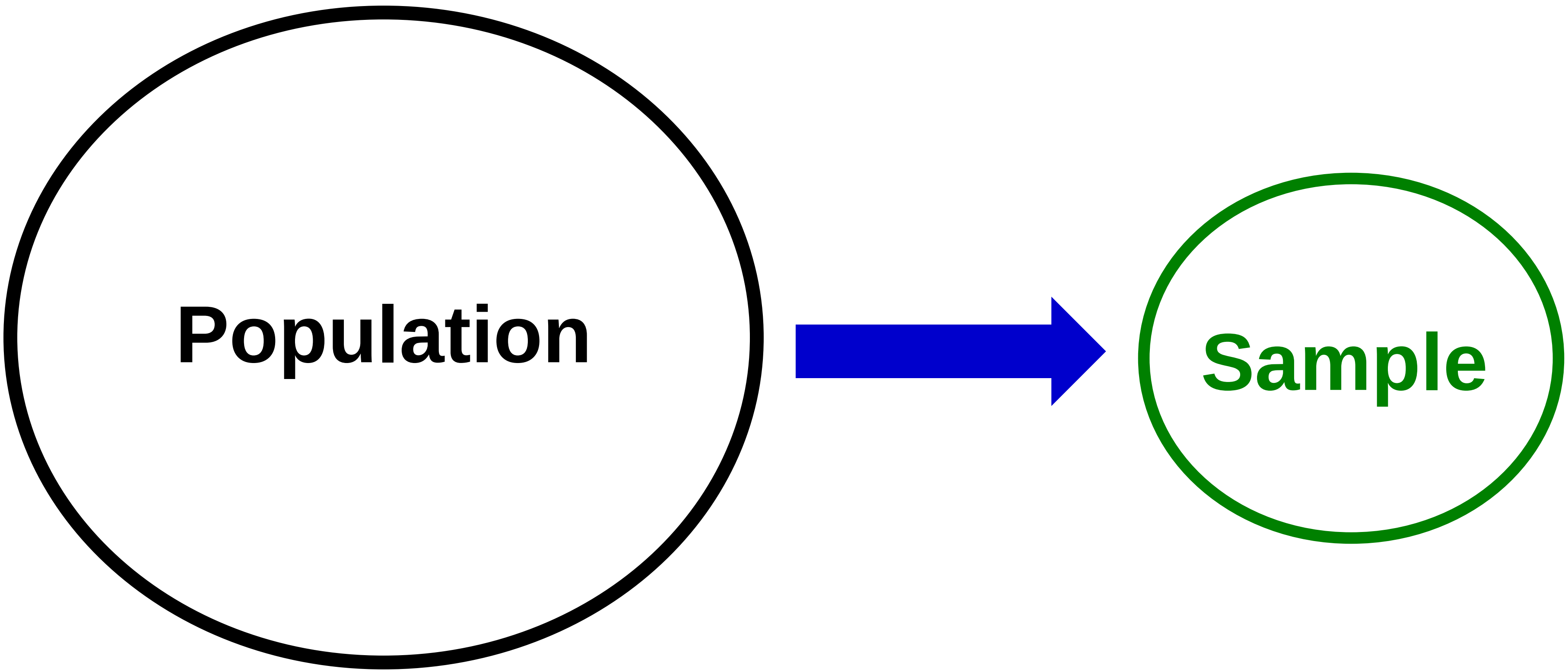
- **Assume your parameter of interest is the average (mean) height of a medical student at PLFSOM**

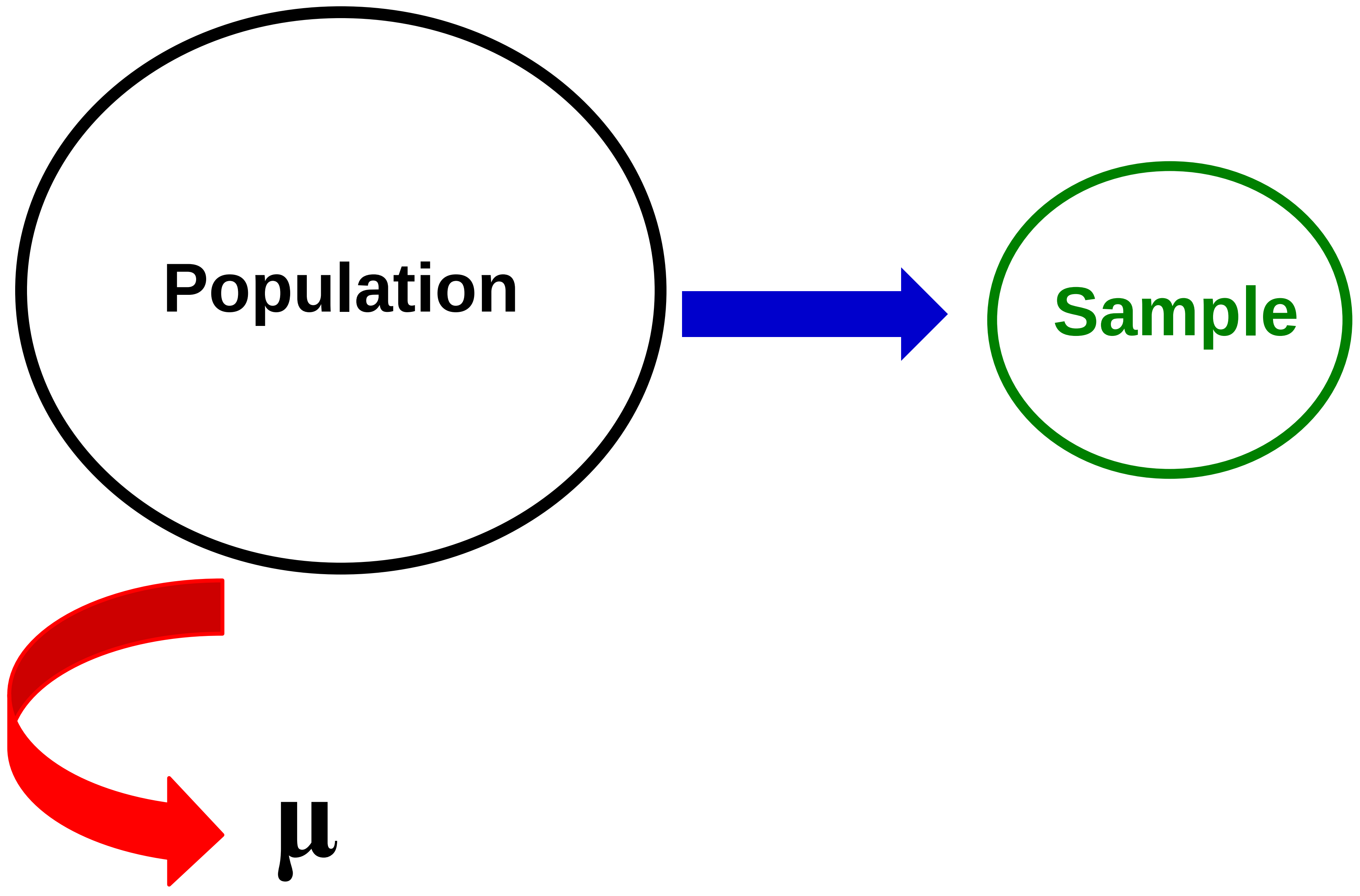
- **Let's use the Greek letter mu to denote this parameter: μ**

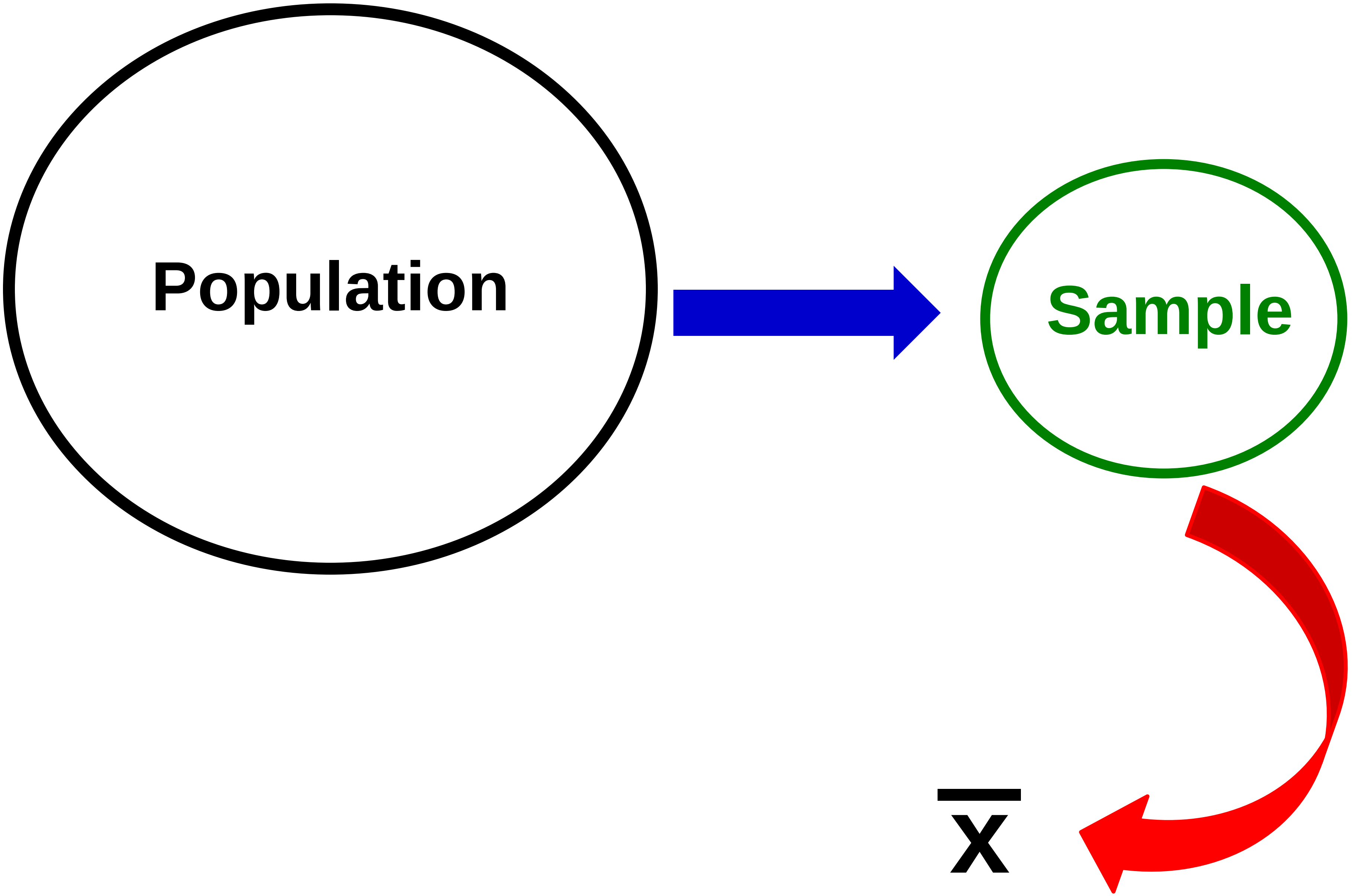
Example

- **Assume there are 180 medical students in your population of interest. You don't have the time and money to measure everyone's height**

- **You take a random sample, of, say, 30 students, and calculate the average height. This is the sample mean designated as $\bar{x}$**


Population
Sample


Population
Sample
μ


Population
Sample
$\bar{x}$

Introduction to the *p* Value

- The null hypothesis (H_0) is usually (but not always) the hypothesis of no difference

- You conducted a study and rejected the null hypothesis because $p = 0.01$

- Does this mean that there was a 1% probability that the null hypothesis is correct? No.

Definition of *p* Value

“The *p* value for a hypothesis test is the probability of obtaining, when H_0 is true, a value of the test statistic as extreme as or more extreme (in the appropriate direction) than the one actually computed.” **–Daniel (1991)**

Definition of *p* Value

“The *p* value for a hypothesis test is the probability of obtaining, when H_0 is true, a value of the *test statistic* as extreme as or more extreme (in the appropriate direction) than the one actually computed.” –Daniel (1991)

Test Statistic

$$= \frac{\text{Point estimate}}{\text{Standard error}}$$

More on the *p* Value from Rothman (1st edition)

"…when the data are very discrepant with the null hypothesis, the *p* value is small, and when the data are concordant with the null hypothesis, the *p* value is large. Nonetheless, the *p* value is not the probability that the null hypothesis is correct. The *p* value is calculated only after assuming that the null hypothesis is correct; it refers to the probability

More on the *p* Value from Rothman

"that data could deviate from the null hypothesis as much as they did or more. It can thus be viewed as a measure of consistency between the data and the null hypothesis…The null hypothesis may be the most reasonable hypothesis for the data even if the *p* value is low, or it may be implausible or just incorrect even if the *p* value is high." - K. Rothman (*Epidemiology: An Introduction*, page 117, published 2002)

α is the probability of committing a Type 1 error

β is the probability of committing a Type II error

- Power = 1 - β

- The ability of a study to detect an association if one exists.

- Power of a test is the probability of rejecting the null hypothesis if it is false

- Power usually set at 80%: 1 - β

α is the probability of committing a Type 1 error

False positive

β is the probability of committing a Type II error

False negative

Result: You reject the null

- Made correct decision

or

- Committed a Type I error if the null is true

Result: Fail to reject the null

- **Made correct decision**

or

- **Committed a Type II error if null is false**

In the following slides "RR" represents "relative risk"

Before the test is performed…

H_o: RR = 1

H_A: RR ≠ 1

$\alpha = 0.05$

Alpha must be pre-specified

The test has been performed

H_o: RR = 1

H_A: RR ≠ 1

α = 0.05

p = 0.01

The test has been performed

H_o: RR = 1

H_A: RR ≠ 1

α = 0.05

$p = 0.01$

Decision: Since the *p* value is less than or equal to alpha, we reject the null hypothesis.

- **Rejecting this null hypothesis means the result is statistically significant (may not be clinically significant)**

- **In plain English, random chance is not a likely explanation of this result**

- **What if the *p* was 0.06?**

- **If the *p* value is greater than the alpha, then we "fail to reject the null hypothesis" (don't say "we accept the null hypothesis")**

- **Random chance is a likely explanation of the results**

- **What if the *p* was 0.05?**

- **Then we reject the null hypothesis**

- **However, a lot of epidemiologists, unlike some statisticians, tend to be wary of relying solely on significance testing (*P* values) when assessing the results of a study.**

- **For example, a well-conducted study reporting an RR of 12 and *P* = 0.06 may be dismissed as a result that is not statistically significant by some researchers even when that may not be advisable**

Confidence Intervals

H_o: RR = 1

H_A: RR ≠ 1

α = 0.05

p value is not needed to assess statistical significance.

Report 95% confidence interval (CI)

- **Point estimate = Your sample relative risk (RR) or sample odds ratio (OR)**

- **Interval estimate = The CI for the population RR or population OR**

General format of a CI

Point estimate ± (*t*) (Standard error)

Standard Error

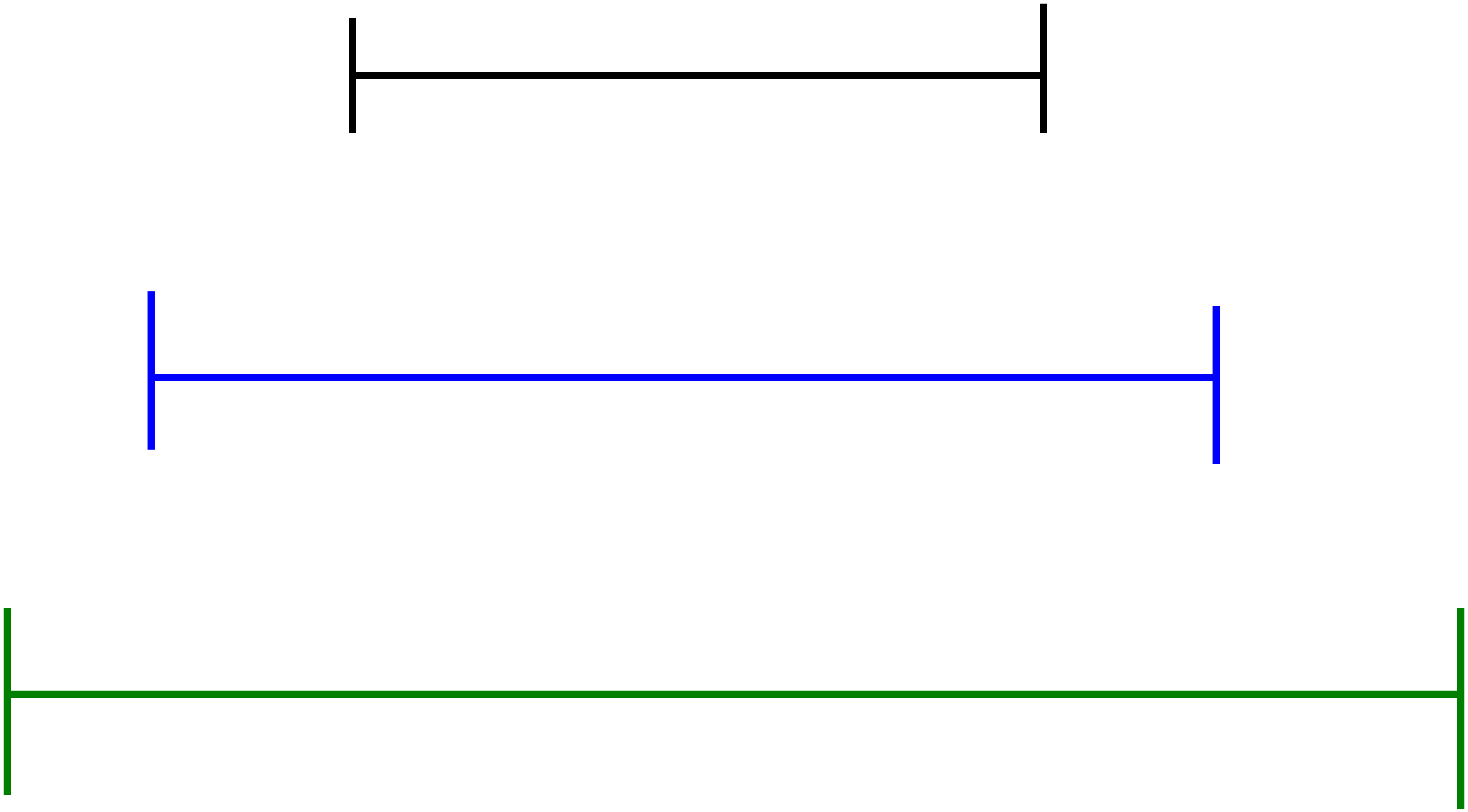
- **Definition from the text *Epidemiology An Introduction 2nd Edition* by Rothman (p. 165): A measure of the statistical variability of the estimate.**

Standard Error

- **From *A Dictionary of Epidemiology 5th edition*: The standard deviation of an estimate. Used to calculate confidence intervals.**

- **Standard deviation: A measure of dispersion or variation…**

Top to bottom: 90%, 95%, and 99% CIs (not drawn to scale)

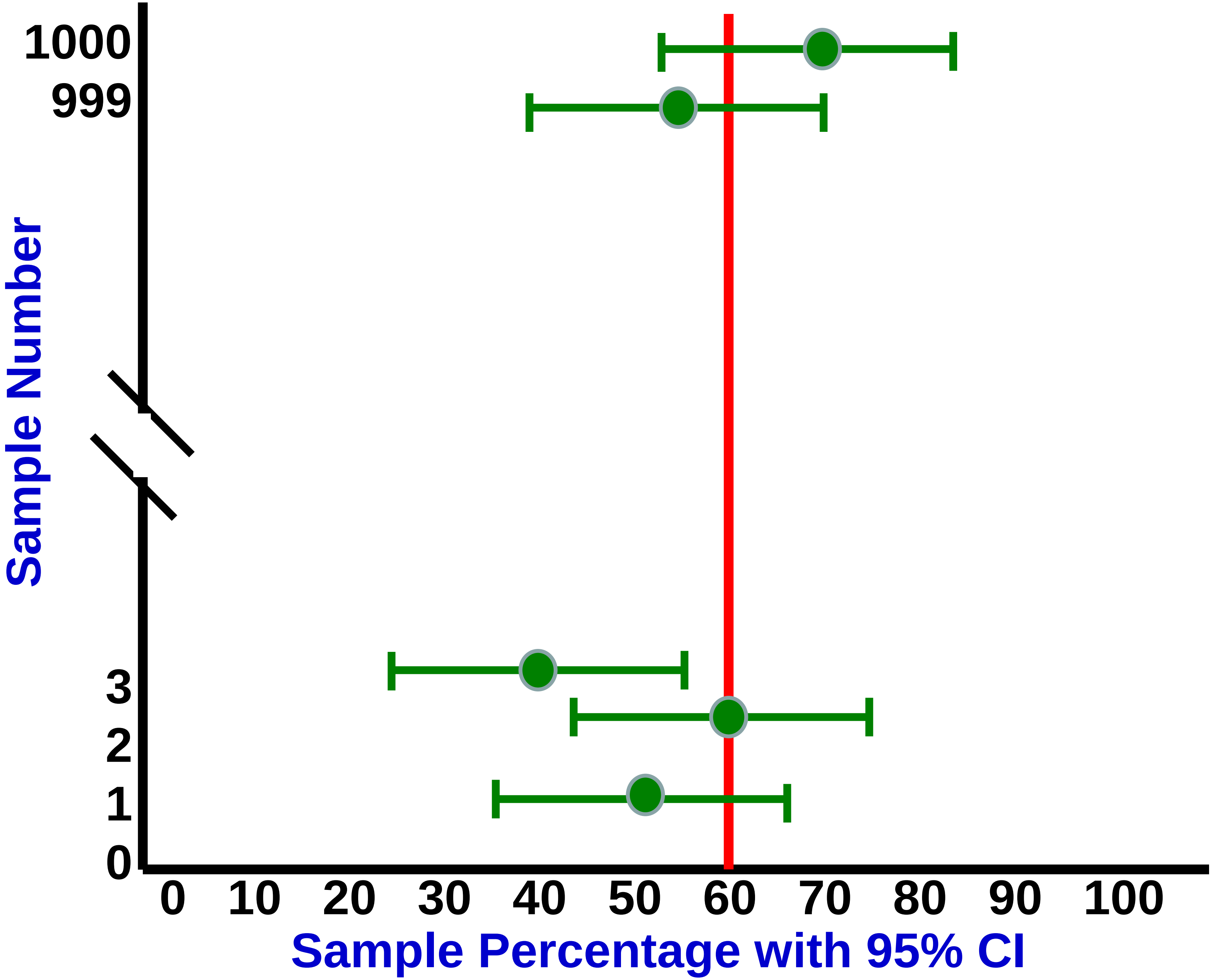


- **The following example and the figure on the next slide) are adapted from p. 81 of the text by Huntsberger and Leaverton.**

- **Assume that you want to sample a population in order to estimate the prevalence of diabetes. Assume that the parameter (the prevalence of diabetes in the population) is 60% (typically the value of parameters are unknown).**

- **You then take 1000 samples from your population of interest. You then calculate 1000 point estimates (in this case, the prevalence of diabetes) and calculate1000 95% confidence intervals (CI).**

- **We would expect that approximately 950 (95%) of the CIs to enclose the true population prevalence (the parameter) of 60% (see the figure on the following slide)**


1000
999
3
2
1
0
0
10
20
30
40
50
60
70
80
90
100
Sample Number
Sample Percentage with 95% CI

t values

- **For 90% CI, use 1.645**
- **For 95% CI, use 1.96**
- **For 99% CI, use 2.576**

Assessing Statistical Significance Using 95% CI

- **If the 95% CI excludes the null value (which for RR is 1) then the result is statistically significant at the 0.05 level and**

- **the *p* value will typically be 0.05 or less**

Our example

$\widehat{RR}$ = 2.49, 95% CI: 1.21 – 5.13

- Interpretation of the CI: I am 95% confident that the true RR is between 1.21 and 5.13.

- This CI is for the *population* RR

Assessing Statistical Significance Using 99% CI

- If the 99% CI excludes the null value (which for RR is 1) then the result is statistically significant at the 0.01 level and

- the *p* value will typically be 0.01 or less

Don't always look for 1!

- The null value for a statistical test of a treatment difference may be 0 as in the following example:

H_o: Cure Rate Tx 1 - Cure Rate Tx 2 = 0

H_A: Cure Rate Tx 1 - Cure Rate Tx 2 ≠ 0

- **So if the 95% CI for this tx difference excludes 0, then the result is statistically significant at the 0.05 level and**

- **the *p* value from an appropriate test would likely be 0.05 or less**

- **Statistical significance does not necessarily equal clinical significance**

- **The *p* value is influenced by two things (p. 32, Hennekens & Buring):**
  - **Strength of the association**
  - **Sample size**

Width of CIs

- **As your sample size increases your estimate becomes more stable and the CI then will be narrower.**

- **Algebraically evident: the increase in sample size leads to a lower standard error which leads to a narrower CI**

Wide CIs

- **If you rejected the null, then there is a 5% chance that you made the wrong decision, so who cares if the CI may be wide**
- **Yes, a wide CI indicates a small sample size**

Wide CIs

- Wide CIs are more important if you fail to reject the null

From Hennekens and Buring

“The information provided by the confidence interval is particularly important when interpreting the results of studies that are not statistically significant (i.e., null findings).

Hennekens and Buring

A narrow confidence interval will add support to the belief that there is actually no true increased risk, whereas a wide interval suggests that the data are compatible with a true increased (or decreased) risk but that the sample size was simply not sufficient to have adequate statistical power to exclude chance as a likely explanation of the findings.”

“A statistically significant result does not mean that chance cannot have accounted for the findings, only that such an explanation is unlikely.”

Hennekens and Buring, *Epidemiology in Medicine* 1987

Cited References

Daniel WW. *Biostatistics: A Foundation for Analysis in the Health Sciences 5th edition*. John Wiley & Sons, Inc., New York; 1991.

Hennekens CH, Buring JE. *Epidemiology in Medicine*. Little, Brown and Company, Boston; 1987.

Huntsberger DV, Leaverton PE. Statistical Inference in the Biomedical Sciences. Allyn and Bacon, Inc., Boston; 1970.

Cited References

Porta M. *A Dictionary of Epidemiology* Fifth Edition. Oxford University Press, New York; 2008.

Rothman KJ. Epidemiology an introduction. Oxford University Press, New York; 2002.

Rothman KJ. Epidemiology an introduction 2nd edition. Oxford University Press, New York; 2012.