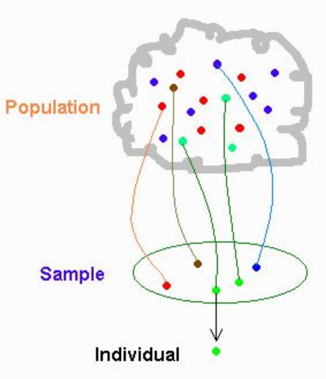
Statistical Inference

Paul Vos

Biostatistics, ECU

Table of contents

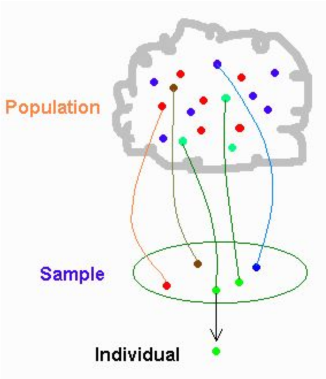
Overview



http://simon.cs.vt.edu/SoSci/converted/Sampling/

- Descriptive statistics are about the sample (or population)

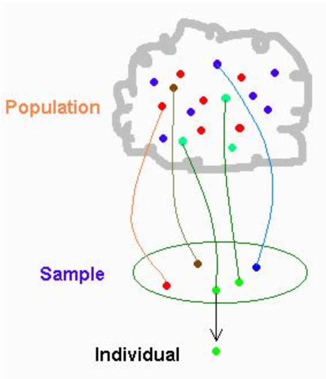
Overview



http://simon.cs.vt.edu/SoSci/converted/Sampling/

- Descriptive statistics are about the sample (or population)
- Inferential statistics relate the known sample to unknown population

Overview



http://simon.cs.vt.edu/SoSci/converted/Sampling/

- Descriptive statistics are about the sample (or population)
- Inferential statistics relate the known sample to unknown population
- How data were produced is crucial to statistical inference

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?
3. What does it mean that the study is significant?

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?
3. What does it mean that the study is significant?
4. What is the difference between practical and statistical significance?

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?
3. What does it mean that the study is significant?
4. What is the difference between practical and statistical significance?

- To answer these questions, we need
  - Descriptive statistics (are assumptions reasonable?)

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?
3. What does it mean that the study is significant?
4. What is the difference between practical and statistical significance?

- To answer these questions, we need
  - Descriptive statistics (are assumptions reasonable?)
  - To know how the data were obtained.

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?
3. What does it mean that the study is significant?
4. What is the difference between practical and statistical significance?

- To answer these questions, we need
  - Descriptive statistics (are assumptions reasonable?)
  - To know how the data were obtained.
  - Randomization

Objectives: To be able to answer questions such as:

1. What does it mean that $(43.2, 47.2)$ is a 95% confidence interval for the mean?
2. The p-value is .03. What does this mean?
3. What does it mean that the study is significant?
4. What is the difference between practical and statistical significance?

- To answer these questions, we need
  - Descriptive statistics (are assumptions reasonable?)
  - To know how the data were obtained.
  - Randomization
  - Probability to describe BOTH randomization and uncertainty.

Examples of Observational Studies

-  Ann Landers Poll: Would you have kids again?

Examples of Observational Studies

-  Ann Landers Poll: Would you have kids again? – 70% said no.

Examples of Observational Studies

-  Ann Landers Poll: Would you have kids again?
  – 70% said no.

"Hey, Pops, what was that letter you sent off to Ann Landers yesterday?"

http://www.stats.uwo.ca/faculty/bellhouse/stat353annlanders.pdf

Examples of Observational Studies (cont)

-  Smoking and lung cancer in humans
  – We observe two groups: smokers and nonsmokers, these groups are as much alike as possible other than smoking.

Examples of Observational Studies (cont)

-  Smoking and lung cancer in humans
  – We observe two groups: smokers and nonsmokers, these groups are as much alike as possible other than smoking.
  – The incidence of lung cancer is much higher in the smoking group, however, ...

Examples of Observational Studies (cont)

-  Smoking and lung cancer in humans
  – We observe two groups: smokers and nonsmokers, these groups are as much alike as possible other than smoking.
  – The incidence of lung cancer is much higher in the smoking group, however, ...
  – Tobacco Co: Cancer in both groups was due to genetics and the gene that causes cancer is also responsible for making smoking especially enjoyable for these individuals.

Examples of Observational Studies (cont)

-  Smoking and lung cancer in humans
  – We observe two groups: smokers and nonsmokers, these groups are as much alike as possible other than smoking.
  – The incidence of lung cancer is much higher in the smoking group, however, ...
  – Tobacco Co: Cancer in both groups was due to genetics and the gene that causes cancer is also responsible for making smoking especially enjoyable for these individuals.
  – Other claims: *real* cause of cancer is something else and this real cause is confounded with smoking.

Examples of Observational Studies (cont)

-  Smoking and lung cancer in humans
  – We observe two groups: smokers and nonsmokers, these groups are as much alike as possible other than smoking.
  – The incidence of lung cancer is much higher in the smoking group, however, ...
  – Tobacco Co: Cancer in both groups was due to genetics and the gene that causes cancer is also responsible for making smoking especially enjoyable for these individuals.
  – Other claims: *real* cause of cancer is something else and this real cause is confounded with smoking.
  – The real cause is a lurking variable.

Examples of Observational Studies (cont)

-  Salk Vaccine Trial (pt I)

Examples of Observational Studies (cont)

-  Salk Vaccine Trial (pt I)

  – Study to find a safe, effective vaccine for polio

  – Severe polio less prevalent among the poor.

  – Second graders received vaccine; 1st and 3rd graders did not; a controlled observational study

Examples of Observational Studies (cont)

-  Salk Vaccine Trial (pt I)
  – Study to find a safe, effective vaccine for polio
  – Severe polio less prevalent among the poor.
  – Second graders received vaccine; 1st and 3rd graders did not; a controlled observational study
- What problems might there be?

Examples of Observational Studies (cont)

-  Salk Vaccine Trial (pt I)
  – Study to find a safe, effective vaccine for polio
  – Severe polio less prevalent among the poor.
  – Second graders received vaccine; 1st and 3rd graders did not; a controlled observational study
- What problems might there be?
  – Diagnosis is somewhat subjective, especially for mild cases, and the study is not blind

Examples of Observational Studies (cont)

-  Salk Vaccine Trial (pt I)
  – Study to find a safe, effective vaccine for polio
  – Severe polio less prevalent among the poor.
  – Second graders received vaccine; 1st and 3rd graders did not; a controlled observational study
- What problems might there be?
  – Diagnosis is somewhat subjective, especially for mild cases, and the study is not blind
  – Not all second graders received the vaccine since it required parental consent.

Random Samples

-  Ann Landers revisited: Same question (Would you have kids again?) but now use a random sample.

Random Samples

-  Ann Landers revisited: Same question (Would you have kids again?) but now use a random sample.
  – now 10% say No.
- What is a (simple) random sample of size $n = 30$?

Random Samples

-  Ann Landers revisited: Same question (Would you have kids again?) but now use a random sample.
  – now 10% say No.
- What is a (simple) random sample of size $n = 30$?

Definition

A simple random sample of size $n = 30$ is a sample of 30 elements chosen from the population such that each sample of size 30 has an equal chance of being selected.

Random Samples

-  Ann Landers revisited: Same question (Would you have kids again?) but now use a random sample.
  – now 10% say No.
- What is a (simple) random sample of size $n = 30$?

Definition

A simple random sample of size $n = 30$ is a sample of 30 elements chosen from the population such that each sample of size 30 has an equal chance of being selected.

- What is the population in the Ann Landers study?

Randomized Experiments

-  Smoking and lung cancer in mice

Randomized Experiments

-  Smoking and lung cancer in mice
  – Take 100 mice and *randomly* assign 50 to Treatment 1 (Smoke exposure) and the remainder to Treatment 2 (Smoke-free control).
  – Suppose 40% (20) from Treatment 1 have cancer while 4% (2) from Treatment 2 have cancer.

Randomized Experiments

-  Smoking and lung cancer in mice
  – Take 100 mice and *randomly* assign 50 to Treatment 1 (Smoke exposure) and the remainder to Treatment 2 (Smoke-free control).
  – Suppose 40% (20) from Treatment 1 have cancer while 4% (2) from Treatment 2 have cancer.
- How can we respond to the claim: "22 mice were going to get cancer regardless of treatment, we just happened by chance to put only 2 in the control group, the other 20 happened to get assigned to the smoking group."

Randomized Experiments

-  Smoking and lung cancer in mice
  – Take 100 mice and *randomly* assign 50 to Treatment 1 (Smoke exposure) and the remainder to Treatment 2 (Smoke-free control).
  – Suppose 40% (20) from Treatment 1 have cancer while 4% (2) from Treatment 2 have cancer.
- How can we respond to the claim: "22 mice were going to get cancer regardless of treatment, we just happened by chance to put only 2 in the control group, the other 20 happened to get assigned to the smoking group."
- This is possible, but not likely and we can calculate how unlikely (p-value).

Randomized Experiments (cont)

-  Salk Vaccine Trial (pt II)

Randomized Experiments (cont)

-  Salk Vaccine Trial (pt II)
  – A random sample of children selected from those whose parents gave consent.
  – Each child in the sample was randomly assigned to one of two groups: Vaccine or Placebo
  – Neither the child (parents) or the physician knew whether the injection was vaccine or placebo.

Randomized Experiments (cont)

-  Salk Vaccine Trial (pt II)
  – A random sample of children selected from those whose parents gave consent.
  – Each child in the sample was randomly assigned to one of two groups: Vaccine or Placebo
  – Neither the child (parents) or the physician knew whether the injection was vaccine or placebo.
- Randomized, controled, double-blind study (gold standard)
- This design addresses selection bias, diagnosis bias, and lurking variables.

Randomized Experiments (cont)

-  Salk Vaccine Trial (pt II)

  – A random sample of children selected from those whose parents gave consent.

  – Each child in the sample was randomly assigned to one of two groups: Vaccine or Placebo

  – Neither the child (parents) or the physician knew whether the injection was vaccine or placebo.
- Randomized, controled, double-blind study (gold standard)
- This design addresses selection bias, diagnosis bias, and lurking variables.
- The difference between vaccine and placebo is due either to the vaccine or chance. http://wps.aw.com/wps/media/objects/14/15269/projects/ch12_salk/index.html

HIP (Health Insurance Plan) trial

discussion from Freedman 2009

- Does screening for breast cancer save lives?

HIP (Health Insurance Plan) trial

discussion from Freedman 2009

- Does screening for breast cancer save lives?
  - Early 1960s, 62,000 women, age 40-64
  - randomly assigned to treatment or control
  - treatment: offered mammography; control: standard care
  - about 1/3 of treatment group refused screening
  - data: 5 year death count and rate (per 1000)

HIP trial (cont)

Deaths in 5 years of followup

	Group	Breast cancer		All other	
	Size	No.	Rate	No.	Rate
Treatment					
Screened	20,200	23	1.1	428	21
Refused	10,800	16	1.5	409	38
Total	31,000	39	1.3	837	27
Control	31,000	63	2.0	879	28

- Which groups should be compared to determine if mammography save lives?

HIP trial (cont)

Deaths in 5 years of followup

	Group Size	Breast cancer No.	Breast cancer Rate	All other No.	All other Rate
Treatment					
Screened	20,200	23	1.1	428	21
Refused	10,800	16	1.5	409	38
Total	31,000	39	1.3	837	27
Control	31,000	63	2.0	879	28

- Which groups should be compared to determine if mammography save lives?
  – What about Screened vs. Refused?

HIP trial (cont)

Deaths in 5 years of followup

	Group	Breast cancer		All other	
	Size	No.	Rate	No.	Rate
Treatment					
Screened	20,200	23	1.1	428	21
Refused	10,800	16	1.5	409	38
Total	31,000	39	1.3	837	27
Control	31,000	63	2.0	879	28

- Which groups should be compared to determine if mammography save lives?
  - What about Screened vs. Refused?
  - To keep benefits of randomization use Trt vs Cont.

Intention-to-treat analysis

Role of Randomization

Benefits of Randomization

- Addresses the problem of lurking variables
- Allows the quantification of uncertainty

Role of Randomization

Benefits of Randomization

- Addresses the problem of lurking variables
- Allows the quantification of uncertainty

- Statistical inferences (*p*-values, 95% CI) are based on randomization

Role of Randomization

Benefits of Randomization

- Addresses the problem of lurking variables
- Allows the quantification of uncertainty

- Statistical inferences (p-values, 95% CI) are based on randomization
- Many important studies do not use randomization

Role of Randomization

Benefits of Randomization

- Addresses the problem of lurking variables
- Allows the quantification of uncertainty

- Statistical inferences (*p*-values, 95% CI) are based on randomization
- Many important studies do not use randomization
- Statistical inference sometimes proceeds *assuming* the data were obtained using randomization

Role of Randomization

Benefits of Randomization

- Addresses the problem of lurking variables
- Allows the quantification of uncertainty

- Statistical inferences (p-values, 95% CI) are based on randomization
- Many important studies do not use randomization
- Statistical inference sometimes proceeds *assuming* the data were obtained using randomization
- How the data were obtained is v. important to inference, but is not part of formal calculations

Inference without Randomization

- It is better not to do formal inference (p-values and confidence intervals) when there are glaring violations of assumptions (e.g., randomization).

Inference without Randomization

- It is better not to do formal inference (p-values and confidence intervals) when there are glaring violations of assumptions (e.g., randomization).
- Reasonable inferences may still be possible but are better communicated through descriptive statistics.

Inference without Randomization

- It is better not to do formal inference (*p*-values and confidence intervals) when there are glaring violations of assumptions (e.g., randomization).
- Reasonable inferences may still be possible but are better communicated through descriptive statistics.
- "Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials" in *BMJ* (**327**, 2003, pp. 1459-1461). `http://www.bmj.com/content/vol327/issue7429/`

Inference without Randomization (cont)

Evidence based pride and observational prejudice
It is a truth universally acknowledged that a medical intervention justified by observational data must be in want of verification through a randomised controlled

Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

Inference without Randomization (cont)

What is already known about this topic

Parachutes are widely used to prevent death and major injury after gravitational challenge

Parachute use is associated with adverse effects due to failure of the intervention and iatrogenic injury

Studies of free fall do not show 100% mortality

Examples of Random Phenomena

Examples:

- 

Flipping a coin
- 

Drawing from a well-shuffled deck of cards
-  Rolling a pair of dice.
-  Time to first detection of an α particle

Examples of Random Phenomena

Examples:

- 

Flipping a coin
- 

Drawing from a well-shuffled deck of cards
-  Rolling a pair of dice.
- 

Time to first detection of an α particle

What are the common features of these examples?

Examples of Random Phenomena

Examples:

- 

Flipping a coin
- 

Drawing from a well-shuffled deck of cards
- 
Rolling a pair of dice.
- 

Time to first detection of an α particle

What are the common features of these examples?
What about: Guessing game between two people?
– Rock, Paper, Scissors: Win or Lose.

Examples of Random Phenomena

Examples:

- 

Flipping a coin
- 

Drawing from a well-shuffled deck of cards
-  Rolling a pair of dice.
-  Time to first detection of an α particle

What are the common features of these examples?
What about: Guessing game between two people?

– Rock, Paper, Scissors: Win or Lose.
– How are the outcomes (Win/Lose) not random?

Characteristics of Randomness

Properties of Random Phenomena

1. Individual outcomes are *uncertain*
2. There is *structure* in the aggregate of all outcomes

Characteristics of Randomness

Properties of Random Phenomena

1. Individual outcomes are *uncertain*
2. There is *structure* in the aggregate of all outcomes

- Statistical inference utilizes this structure to quantify uncertainty.
- This structure allows us to quantify how uncertain outcomes are/were.
- The structure applies before data have been observed **AND** after.

Probability of Random Events

- The structure in random phenomena allow assignment of numeric/objective probabilities.

Probability of Random Events

- The structure in random phenomena allow assignment of numeric/objective probabilities.
- Examples of events:
  - Coin lands Heads up.
  - First card drawn is red.
  - The dice roll summed to 7.
  - The first α particle was detected b/t .80 and .85 ms.

Probability of Random Events

- The structure in random phenomena allow assignment of numeric/objective probabilities.
- Examples of events:
  - Coin lands Heads up.
  - First card drawn is red.
  - The dice roll summed to 7.
  - The first α particle was detected b/t .80 and .85 ms.
- The probability of an event is the *proportion* of outcomes in the aggregate of all outcomes that result in the event occurring.
- Sometimes this is called the limiting relative frequency interpretation of probability.

Properties of Probability

$$P(A) = \frac{\text{number of outcomes in } A}{\text{Total number of outcomes}}$$

- A and B are events; e.g. $A =$ card is $\heartsuit$, $B =$ card is $\spadesuit$ or $\clubsuit$
- Probability of any event is between 0 and 1: $0 \leq P(A) \leq 1$
- $P(\text{not}A) = 1 - P(A)$
- $P(A \text{ or } B) = P(A) + P(B)$ if A and B disjoint.

Conditional Probability

- $P(A|B)$ is the probability of A occurring if B has occurred.
- $P(A|B)$ can be very different from $P(B|A)$
- Consider choosing an individual randomly and $A =$ 'US senator' and $B =$ 'male'

Conditional Probability

- $P(A|B)$ is the probability of A occurring if B has occurred.
- $P(A|B)$ can be very different from $P(B|A)$
- Consider choosing an individual randomly and $A =$ 'US senator' and $B =$ 'male'
  – $P(A|B)$ is very small but $P(B|A) = 83/100$

Conditional Probability

- $P(A|B)$ is the probability of A occurring if B has occurred.
- $P(A|B)$ can be very different from $P(B|A)$
- Consider choosing an individual randomly and $A =$ 'US senator' and $B =$ 'male'
  – $P(A|B)$ is very small but $P(B|A) = 83/100$
- Or, $A =$ 'pregnant' and $B =$ 'female'
  – $P('pregnant'|'female')$ vs. $P('female'|'pregnant')$

Conditional Probability

- $P(A|B)$ is the probability of A occurring if B has occurred.
- $P(A|B)$ can be very different from $P(B|A)$
- Consider choosing an individual randomly and $A =$ 'US senator' and $B =$ 'male'
  - $P(A|B)$ is very small but $P(B|A) = 83/100$
- Or, $A =$ 'pregnant' and $B =$ 'female'
  - $P('pregnant'|'female')$ vs. $P('female'|'pregnant')$
- ACI: testing for a rare disease and $A =$ 'test is positive' and $B =$ 'disease'
  - $P('test\ is\ positive'|'disease') = .95$
  - $P('disease'|'test\ is\ positive') =??$

http://myweb.ecu.edu/vosp/ofe/Click&Clack-Epi.mp3

Dealing a straight flush

- Modified Example attributed to John Maynard Keynes

  if Billy Graham were playing poker, the probability he would deal himself a straight flush given honest play on his part is not the same as the probability of honest play on his part given that he as dealt himself a straight flush. [Finkelstein, 2009]

- $P(A|B) \neq P(B|A)$; $A =$ 'straight flush', $B =$ 'fair play'

Dealing a straight flush

- Modified Example attributed to John Maynard Keynes

  > *if Billy Graham were playing poker, the probability he would deal himself a straight flush given honest play on his part is not the same as the probability of honest play on his part given that he as dealt himself a straight flush. [Finkelstein, 2009]*

- $P(A|B) \neq P(B|A)$; $A =$ 'straight flush', $B =$ 'fair play'
- But, A and B are different types of "events" (B is not an event).
- $P(A|B)$ can be calculated $\frac{36}{2{,}598{,}960}$. Determining $P(B|A)$, the probability that Rev. Graham cheated, is more difficult.

Dealing a straight flush

- Modified Example attributed to John Maynard Keynes

  > *if Billy Graham were playing poker, the probability he would deal himself a straight flush given honest play on his part is not the same as the probability of honest play on his part given that he as dealt himself a straight flush. [Finkelstein, 2009]*

- $P(A|B) \neq P(B|A)$; $A =$ 'straight flush', $B =$ 'fair play'
- But, A and B are different types of "events" (B is not an event).
- $P(A|B)$ can be calculated $\frac{36}{2{,}598{,}960}$. Determining $P(B|A)$, the probability that Rev. Graham cheated, is more difficult.
- For us, $A =$ 'unlikely data', $B =$ 'two treatments are the same'. We calculate $P(A|B)$ but we are interested in $P(B|A)$.

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):
  - Trump will still be president in 2018.

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):
    - Trump will still be president in 2018.
    - I will probably get selected for jury duty next year.

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):
  - Trump will still be president in 2018.
  - I will probably get selected for jury duty next year.
  - There probably is a God who created the universe (Pascal).

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):
  - Trump will still be president in 2018.
  - I will probably get selected for jury duty next year.
  - There probably is a God who created the universe (Pascal).
- DOB probability differs from stochastic probability (of random events):
  - Stochastic probabilities are obtained from the aggregate structure of random phenomena.

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):
    - Trump will still be president in 2018.
    - I will probably get selected for jury duty next year.
    - There probably is a God who created the universe (Pascal).
- DOB probability differs from stochastic probability (of random events):
    - Stochastic probabilities are obtained from the aggregate structure of random phenomena.
    - Stochastic probabilities are objective (all people would agree to numeric value).

Probability and Uncertainty

- Probability assigned to random events describes the uncertainty in outcomes
- The term probability is also used to describe a degree-of-belief (DOB):
  - Trump will still be president in 2018.
  - I will probably get selected for jury duty next year.
  - There probably is a God who created the universe (Pascal).
- DOB probability differs from stochastic probability (of random events):
  - Stochastic probabilities are obtained from the aggregate structure of random phenomena.
  - Stochastic probabilities are objective (all people would agree to numeric value).
  - DOB probabilities are subjective/difficult to assign numeric value

Probability and Uncertainty (cont)

- DOB probability is related to stochastic probability
  - Observing an unlikely event elicits surprise
  - Discovering that a belief held with high certainty is false elicits surprise

Probability and Uncertainty (cont)

- DOB probability is related to stochastic probability
  - Observing an unlikely event elicits surprise
  - Discovering that a belief held with high certainty is false elicits surprise
- Stochastic probability and DOB probability are related but distinct concepts.

Probability and Uncertainty (cont)

- DOB probability is related to stochastic probability
  - Observing an unlikely event elicits surprise
  - Discovering that a belief held with high certainty is false elicits surprise
- Stochastic probability and DOB probability are related but distinct concepts.
- Do not assign the numeric stochastic probability to a DOB probability.

Probability and Uncertainty (cont)

- DOB probability is related to stochastic probability
  - Observing an unlikely event elicits surprise
  - Discovering that a belief held with high certainty is false elicits surprise
- Stochastic probability and DOB probability are related but distinct concepts.
- Do not assign the numeric stochastic probability to a DOB probability.
- Statistical inference exploits the stochastic structure in data obtained by randomization to address the DOB uncertainty regarding claims about nature/population.

Duality of Probability

Worksheet: http://myweb.ecu.edu/vosp/7021/wsjanus.pdf

Duality of Probability

Worksheet: http://myweb.ecu.edu/vosp/7021/wsjanus.pdf

Stochastic/DOB Duality

Both types of probability play a role in statistical inference, but great care needs to be exercised to keep these from being confused.

Duality of Probability

Worksheet: http://myweb.ecu.edu/vosp/7021/wsjanus.pdf

Stochastic/DOB Duality

Both types of probability play a role in statistical inference, but great care needs to be exercised to keep these from being confused.

- A *p*-value of .04 **DOES NOT** mean that the probability is .04 that the "treatment has no effect" (null hypothesis) is true.

Duality of Probability

Worksheet: http://myweb.ecu.edu/vosp/7021/wsjanus.pdf

Stochastic/DOB Duality

Both types of probability play a role in statistical inference, but great care needs to be exercised to keep these from being confused.

- A *p*-value of .04 **DOES NOT** mean that the probability is .04 that the "treatment has no effect" (null hypothesis) is true.
- A *p*-value of .04 **DOES** mean that the data observed are unlikely, values this extreme occur by chance with probability .04 **IF** the null is true (as well as additional assumptions).

Motivating Example – Head Circumference

Head circumference at birth linked to cancer in childhood

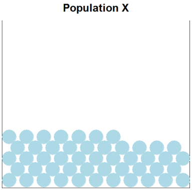
The Lancet Oncology - Volume 7, Issue 1 (January 2006) - Copyright © 2006 Elsevier -

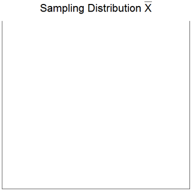
Head circumference could predict risk of brain cancer

In this issue of *The Lancet Oncology*, Samuelsen and co-workers [1] report a positive association between head circumference at birth and incidence of brain tumours in childhood (0–15 years). The investigators recorded a relative risk of brain cancer of 1·16 (95% CI 1·09–1·23) per 1 cm increase in head circumference in unadjusted analyses, and of 1·27 (1·16–1·38) per 1 cm increase after adjustment for birthweight, gestational age, and sex. The analyses were based on data from a large Norwegian medical birth registry linked with N...

– We are interested in the Mean (and distribution) of circumference in a target population ◂ Return

Urn model – Construction

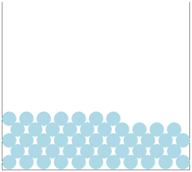


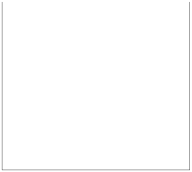


- For each head circum., select a ball , write the value on the ball, and place in the urn X. For example, head circum. 19.2 is represented by 

.

Urn model – Construction

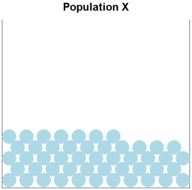


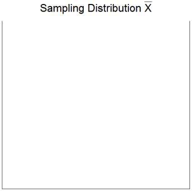


- For each head circum., select a ball , write the value on the ball, and place in the urn X. For example, head circum. 19.2 is represented by 

.
- Take a sample of $n = 30$ from the population, calculate the sample mean $\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$, and place this numeric value on a dark blue ball .

Urn model – Construction





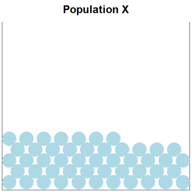
- For each head circum., select a ball , write the value on the ball, and place in the urn X. For example, head circum. 19.2 is represented by 

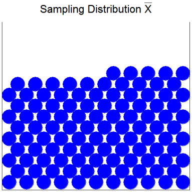
.
- Take a sample of $n = 30$ from the population, calculate the sample mean $\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$, and place this numeric value on a dark blue ball .
- Repeat for *all* possible samples of size $n = 30$.

Urn model – Properties

- Mean μ
- Standard Deviation σ
- Shape: Normal/Other

Urn model – Properties

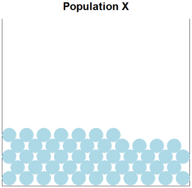


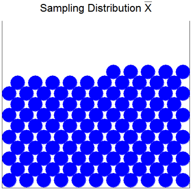


- Mean μ
- Standard Deviation σ
- Shape: Normal/Other

- Mean μ

Urn model – Properties

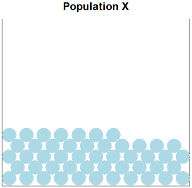


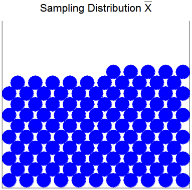


- Mean μ
- Standard Deviation σ
- Shape: Normal/Other

- Mean μ
- Standard Deviation $\sigma/\sqrt{n}$ (aka *standard error*)

Urn model – Properties

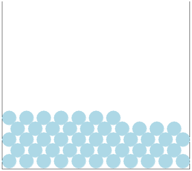


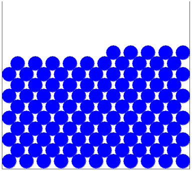


- Mean μ
- Standard Deviation σ
- Shape: Normal/Other

- Mean μ
- Standard Deviation $\sigma/\sqrt{n}$ (aka *standard error*)
- Shape: Normal*

Urn model – Properties





- Mean μ
- Standard Deviation σ
- Shape: Normal/Other

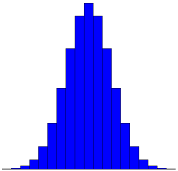
- Mean μ
- Standard Deviation $\sigma/\sqrt{n}$ (aka *standard error*)
- Shape: Normal*

$*$ if n is large enough.

Visualizing the Sampling Distribution

Area Histogram:

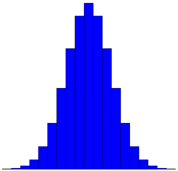
- A bargraph where the area of each bar gives relative frequency of data falling in the range specified by the bars width
- The number of values in the sampling distribution is HUGE, so many bars can be used:

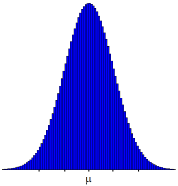


Visualizing the Sampling Distribution

Area Histogram:

- A bargraph where the area of each bar gives relative frequency of data falling in the range specified by the bars width
- The number of values in the sampling distribution is HUGE, so many bars can be used:

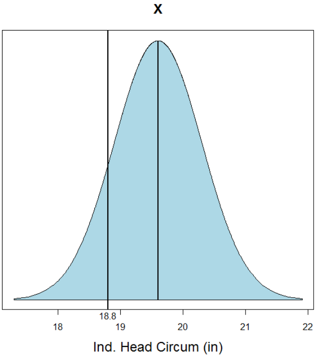




First Steps

Example - Head Sizes

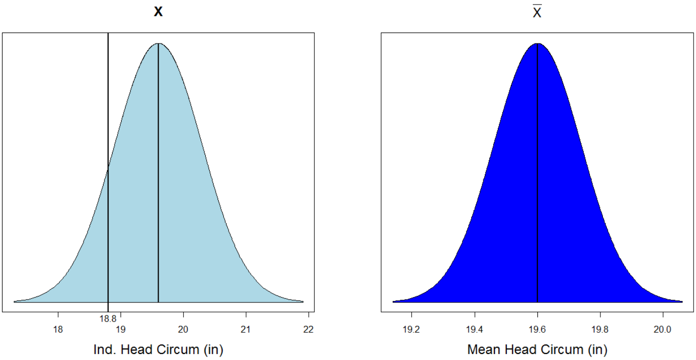
SRS of $n = 25$ from target pop. gives $\bar{x} = 18.8$ inches
Compare with US (36 mos): $\mu = 19.6$, $\sigma = .7$ Return



First Steps

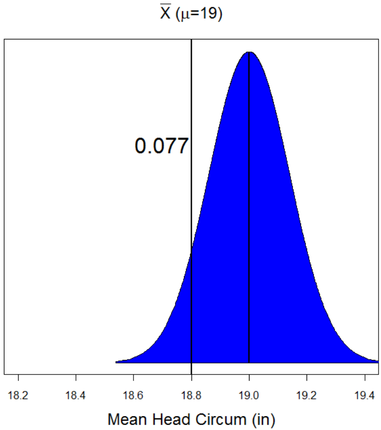
Example - Head Sizes

SRS of $n = 25$ from target pop. gives $\bar{x} = 18.8$ inches
Compare with US (36 mos): $\mu = 19.6$, $\sigma = .7$ Return



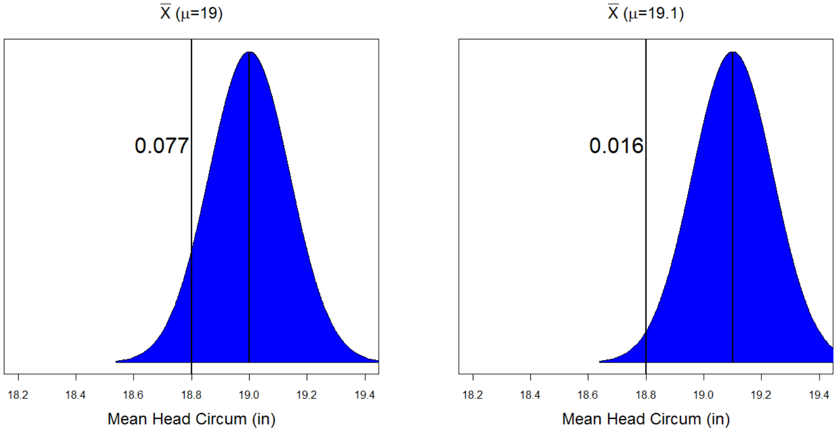
First Steps (cont)

Could $\mu = 19$? $\mu = 19.1$ (still assuming $\sigma = .7$)



First Steps (cont)

Could $\mu = 19$? $\mu = 19.1$ (still assuming $\sigma = .7$)



First Steps - Summary

- We conceptually construct all possible sample means of size $n = 25$: $\overline{X}$

First Steps - Summary

- We conceptually construct all possible sample means of size $n = 25$: $\overline{X}$
- The distribution of $\overline{X}$ depends on the value assigned to population mean μ

First Steps - Summary

- We conceptually construct all possible sample means of size $n = 25$: $\overline{X}$
- The distribution of $\overline{X}$ depends on the value assigned to population mean μ
- The observed sample mean $\bar{x}$ is compared to the value assigned to μ using $\overline{X}$

First Steps - Summary

- We conceptually construct all possible sample means of size $n = 25$: $\overline{X}$
- The distribution of $\overline{X}$ depends on the value assigned to population mean μ
- The observed sample mean $\bar{x}$ is compared to the value assigned to μ using $\overline{X}$
- Same ideas hold for other parameters, such as success proportion p

First Steps - Summary

- We conceptually construct all possible sample means of size $n = 25$: $\overline{X}$
- The distribution of $\overline{X}$ depends on the value assigned to population mean μ
- The observed sample mean $\bar{x}$ is compared to the value assigned to μ using $\overline{X}$
- Same ideas hold for other parameters, such as success proportion p
- $\widehat{P}$ is the collection of all sample proportions (n fixed)
- The distribution of $\widehat{P}$ depends on the value assigned to population proportion p
- The observed sample proportion $\hat{p}$ is compared to the value assigned to p using $\widehat{P}$

Proof by *reductio ad absurdum*

Euclid's proof regarding largest prime number.

Theorem

There is no largest prime number.

Proof.

1. Suppose p were the largest prime number.

4. Thus, $q + 1$ is prime and greater than p which is impossible (probability $= 0$). □

Proof by *reductio ad absurdum*

Euclid's proof regarding largest prime number.

Theorem

There is no largest prime number.

Proof.

1. Suppose p were the largest prime number.
2. Let q be the product of the first p numbers.

4. Thus, $q + 1$ is prime and greater than p which is impossible (probability $= 0$). □

Proof by *reductio ad absurdum*

Euclid's proof regarding largest prime number.

Theorem

There is no largest prime number.

Proof.

1. Suppose p were the largest prime number.
2. Let q be the product of the first p numbers.
3. Then $q + 1$ is not divisible by any of them.
4. Thus, $q + 1$ is prime and greater than p which is impossible (probability $= 0$).

□

Argument by reduction to an unlikely observation

Inference

Claim

Mean head size in target pop. differs from US (19.6 in.)

Argument

1. Suppose $H_0 : \mu = 19.6$ is true.

4. Thus, observing $\bar{x} = 18.8$ is unlikely if H_0 is true.

Argument by reduction to an unlikely observation

Inference

Claim

Mean head size in target pop. differs from US (19.6 in.)

Argument

1. Suppose $H_0 : \mu = 19.6$ is true.
2. H_0 together with other assumptions (SRS, $\sigma = .7, \ldots$) allow construction of $\bar{X}$.

4. Thus, observing $\bar{x} = 18.8$ is unlikely if H_0 is true.

Argument by reduction to an unlikely observation

Inference

Claim

Mean head size in target pop. differs from US (19.6 in.)

Argument

1. Suppose $H_0 : \mu = 19.6$ is true.
2. H_0 together with other assumptions (SRS, $\sigma = .7, \ldots$) allow construction of $\bar{X}$.
3. $\bar{x} = 18.8$ is an extreme observation in $\bar{X}$. Goto
4. Thus, observing $\bar{x} = 18.8$ is unlikely if H_0 is true.

Argument by reduction to an unlikely observation

Inference

Claim

Mean head size in target pop. differs from US (19.6 in.)

Argument

1. Suppose $H_0 : \mu = 19.6$ is true.
2. H_0 together with other assumptions (SRS, $\sigma = .7, \ldots$) allow construction of $\bar{X}$.
3. $\bar{x} = 18.8$ is an extreme observation in $\bar{X}$. Goto
4. Thus, observing $\bar{x} = 18.8$ is unlikely if H_0 is true.

Either H_0 is false or the data observed were unlikely.

Argument when observation is NOT unlikely
Inference

Claim

Mean head size in target pop. is less than 19 inches

Argument

1. Suppose $H_0 : \mu = 19$ ($H_0 : \mu \geq 19$) is true.

4. Thus, observing $\bar{x} = 18.8$ provides little evidence against H_0.

Argument when observation is NOT unlikely
Inference

Claim

Mean head size in target pop. is less than 19 inches

Argument

1. Suppose $H_0 : \mu = 19$ ($H_0 : \mu \geq 19$) is true.
2. H_0 together with other assumptions (SRS, $\sigma = .7, \ldots$) allow construction of $\bar{X}$.

4. Thus, observing $\bar{x} = 18.8$ provides little evidence against H_0.

Argument when observation is NOT unlikely

Inference

Claim

Mean head size in target pop. is less than 19 inches

Argument

1. Suppose $H_0 : \mu = 19$ ($H_0 : \mu \geq 19$) is true.
2. H_0 together with other assumptions (SRS, $\sigma = .7, \ldots$) allow construction of $\bar{X}$.
3. $\bar{x} = 18.8$ is not that unlikely to observe in $\bar{X}$.
4. Thus, observing $\bar{x} = 18.8$ provides little evidence against H_0.

Argument when observation is NOT unlikely

Inference

Claim

Mean head size in target pop. is less than 19 inches

Argument

1. Suppose $H_0 : \mu = 19$ ($H_0 : \mu \geq 19$) is true.
2. H_0 together with other assumptions (SRS, $\sigma = .7, \ldots$) allow construction of $\bar{X}$.
3. $\bar{x} = 18.8$ is not that unlikely to observe in $\bar{X}$.
4. Thus, observing $\bar{x} = 18.8$ provides little evidence against H_0.

If H_0 is true, observing $\bar{x} = 18.8$ is not that unlikely.

The null hypothesis

- **Research Hypothesis:** What one believes or fears to be true.
  - Men are paid more than women.
  - Grocer is cheating consumers.
  - New treatment is better than existing treatment.
  - New treatment is equivalent to the existing treatment.

The null hypothesis

- **Research Hypothesis:** What one believes or fears to be true.
  - Men are paid more than women.
  - Grocer is cheating consumers.
  - New treatment is better than existing treatment.
  - New treatment is equivalent to the existing treatment.
- **Null Hypothesis:** Negation of the research hypothesis after a suitable quantification (e.g., what constitutes a better treatment).
  - $H_0 : \mu_{\text{men}} = \mu_{\text{women}}$
  - $H_0 : \mu \neq 16$
  - $H_0 : \mu_{\text{new}} = \mu_{\text{old}}$
  - $H_0 : |\mu_{\text{new}} - \mu_{\text{old}}| \geq \delta$

The null hypothesis

- **Research Hypothesis:** What one believes or fears to be true.
  - Men are paid more than women.
  - Grocer is cheating consumers.
  - New treatment is better than existing treatment.
  - New treatment is equivalent to the existing treatment.
- **Null Hypothesis:** Negation of the research hypothesis after a suitable quantification (e.g., what constitutes a better treatment).
  - $H_0 : \mu_{\text{men}} = \mu_{\text{women}}$
  - $H_0 : \mu \neq 16$
  - $H_0 : \mu_{\text{new}} = \mu_{\text{old}}$
  - $H_0 : |\mu_{\text{new}} - \mu_{\text{old}}| \geq \delta$
- **Caution:** Null specifies **Exact** equality
  - **Not** H_0 : There is no *significant* difference between the mean salary of men and women.

The p-value

Definition

The p-value is the probability of observing data as *extreme* as what was observed if H_0 is true.

- Extreme data can mean, small values, large values, or both.

The p-value

Definition

The p-value is the probability of observing data as *extreme* as what was observed if H_0 is true.

- Extreme data can mean, small values, large values, or both.
- p-value is a measure of how unlikely the observed data are if H_0 is true.

The p-value

Definition

The p-value is the probability of observing data as *extreme* as what was observed if H_0 is true.

- Extreme data can mean, small values, large values, or both.
- p-value is a measure of how unlikely the observed data are if H_0 is true.
- When the p-values is smaller than a certain value (called the significance level α; usually $\alpha = .05$) the data are said to be statistically significant, and, too often, simply as significant.
- Statistical significance means the observed data were unlikely if the null is true;

The p-value

Definition

The p-value is the probability of observing data as *extreme* as what was observed if H_0 is true.

- Extreme data can mean, small values, large values, or both.
- p-value is a measure of how unlikely the observed data are if H_0 is true.
- When the p-values is smaller than a certain value (called the significance level α; usually $\alpha = .05$) the data are said to be statistically significant, and, too often, simply as significant.
- Statistical significance means the observed data were unlikely if the null is true; we reject H_0.

The p-value

Definition

The p-value is the probability of observing data as *extreme* as what was observed if H_0 is true.

- Extreme data can mean, small values, large values, or both.
- p-value is a measure of how unlikely the observed data are if H_0 is true.
- When the p-values is smaller than a certain value (called the significance level α; usually $\alpha = .05$) the data are said to be statistically significant, and, too often, simply as significant.
- Statistical significance means the observed data were unlikely if the null is true; we reject H_0.
- Court room analogy: H_0 is on trial.

Statistical vs Practical Significance

Consider $H_0 : \mu_{men} = \mu_{women}$. Suppose p-value$= .04$.

- Statistical significance means that the observed data are unlikely (probability less than .05) if the null is true, that is, if the mean salaries for men and women are the same.

Statistical vs Practical Significance

Consider $H_0 : \mu_{men} = \mu_{women}$. Suppose p-value$= .04$.

- Statistical significance means that the observed data are unlikely (probability less than .05) if the null is true, that is, if the mean salaries for men and women are the same.
- Statistical significance also means the null is not likely to be true (a DOB probability, so cannot assign .05 to this).

Statistical vs Practical Significance

Consider $H_0 : \mu_{men} = \mu_{women}$. Suppose p-value$= .04$.

- Statistical significance means that the observed data are unlikely (probability less than .05) if the null is true, that is, if the mean salaries for men and women are the same.
- Statistical significance also means the null is not likely to be true (a DOB probability, so cannot assign .05 to this).
- Statistical significance does NOT mean the difference between men and women is large or important.

Statistical vs Practical Significance

Consider $H_0 : \mu_{men} = \mu_{women}$. Suppose p-value$= .04$.

- Statistical significance means that the observed data are unlikely (probability less than .05) if the null is true, that is, if the mean salaries for men and women are the same.
- Statistical significance also means the null is not likely to be true (a DOB probability, so cannot assign .05 to this).
- Statistical significance does NOT mean the difference between men and women is large or important.
- Statistical significance need not be the same as practical or clinical significance.

Construction of Confidence Intervals

- The sample mean $\bar{x}$ is an *estimate* for the population mean μ
- An estimate is an educated guess.

Construction of Confidence Intervals

- The sample mean $\bar{x}$ is an *estimate* for the population mean μ
- An estimate is an educated guess.
- The *standard error* is an estimate for the standard deviation of the sampling distribution for the estimator.

Construction of Confidence Intervals

- The sample mean $\bar{x}$ is an *estimate* for the population mean μ
- An estimate is an educated guess.
- The *standard error* is an estimate for the standard deviation of the sampling distribution for the estimator.
- The *standard error* describes the variability of $\bar{x}$ across all possible samples (i.e., sampling distribution).

$$95\% \text{ CI for } \mu \quad : \quad \bar{x} \pm t_{.025}\widehat{se}_{\bar{X}}, \quad \widehat{se}_{\bar{X}} = \frac{s}{\sqrt{n}}$$

Construction of Confidence Intervals

- The sample mean $\bar{x}$ is an *estimate* for the population mean μ
- An estimate is an educated guess.
- The *standard error* is an estimate for the standard deviation of the sampling distribution for the estimator.
- The *standard error* describes the variability of $\bar{x}$ across all possible samples (i.e., sampling distribution).

$$\begin{aligned} 95\% \text{ CI for } \mu &: \quad \bar{x} \pm t_{.025}\widehat{se}_{\bar{X}}, \quad \widehat{se}_{\bar{X}} = \frac{s}{\sqrt{n}} \\ 95\% \text{ CI for } \theta &: \quad \hat{\theta} \pm t_{.025}\widehat{se}_{\hat{\theta}} \end{aligned}$$

- θ is a parameter (e.g., risk, slope, etc) estimated by $\hat{\theta}$
- $\widehat{se}_{\hat{\theta}}$ measures the variability of $\hat{\theta}$ across all samples.

Procedural Interpretation

- The formula for a confidence interval defines a procedure for transforming the data into an interval.
- IF certain assumptions for how the data were generated hold true, then this procedure can be given the following interpretation:

Procedural Interpretation

- The formula for a confidence interval defines a procedure for transforming the data into an interval.
- IF certain assumptions for how the data were generated hold true, then this procedure can be given the following interpretation:

  "95% of the intervals generated by this procedure correctly cover the true mean (parameter)."

- Assumptions include:
  - Data obtained from a random sample.
  - Data are approximately normal / sample size is large enough.
  - No outliers.
  - More assumptions for more complicated models.

Data Interpretation

- Confidence intervals can be constructed by inverting statistical tests – consider many null hypotheses, one for each value of the parameter.

Data Interpretation

- Confidence intervals can be constructed by inverting statistical tests – consider many null hypotheses, one for each value of the parameter.
- Formally, the 95% confidence interval consists of all values for the parameter (mean, relative risk) that would not be rejected at the $\alpha = .05$ level.

Data Interpretation

- Confidence intervals can be constructed by inverting statistical tests – consider many null hypotheses, one for each value of the parameter.
- Formally, the 95% confidence interval consists of all values for the parameter (mean, relative risk) that would not be rejected at the $\alpha = .05$ level.
- More simply, either the true value of the parameter is in the 95% confidence interval or the data that were observed were unlikely, happening by chance with probability at most .05.

Data Interpretation

- Confidence intervals can be constructed by inverting statistical tests – consider many null hypotheses, one for each value of the parameter.
- Formally, the 95% confidence interval consists of all values for the parameter (mean, relative risk) that would not be rejected at the $\alpha = .05$ level.
- More simply, either the true value of the parameter is in the 95% confidence interval or the data that were observed were unlikely, happening by chance with probability at most .05.
- The required assumptions are exactly the same as for the procedural interpretation.

Revisit Lancet Study on brain tumor and head size

Go back to Lancet study

- The Lancet study on head size and risk of brain tumor states the

  "relative risk of brain cancer of 1.16 (95% CI 1.09 – 1.23) per 1 cm increase in head circumference ...

Revisit Lancet Study on brain tumor and head size

Go back to Lancet study

- The Lancet study on head size and risk of brain tumor states the

  "relative risk of brain cancer of 1.16 (95% CI 1.09 – 1.23) per 1 cm increase in head circumference . . .

- Other notation: 95% CI for relative risk is $(1.09, 1.23)$.
- How do we interpret this interval?

CI from Lancet Study

Proc The procedure that generated the interval $(1.09, 1.23)$ covers the true relative risk 95% of the time.

CI from Lancet Study

Proc The procedure that generated the interval $(1.09, 1.23)$ covers the true relative risk 95% of the time.

Wrong The probability is 95% that the true RR is in the interval $(1.09, 1.23)$. – Why is this wrong?

CI from Lancet Study

Proc: The procedure that generated the interval $(1.09, 1.23)$ covers the true relative risk 95% of the time.

Wrong: The probability is 95% that the true RR is in the interval $(1.09, 1.23)$. – Why is this wrong?

- The procedural interpretation works better *before* data have been collected.

Data: Either the true RR is between 1.09 and 1.23 or the observed data were unlikely, happening by chance with probability at most .05

CI from Lancet Study

Proc The procedure that generated the interval $(1.09, 1.23)$ covers the true relative risk 95% of the time.

Wrong The probability is 95% that the true RR is in the interval $(1.09, 1.23)$. – Why is this wrong?

- The procedural interpretation works better *before* data have been collected.

Data Either the true RR is between 1.09 and 1.23 or the observed data were unlikely, happening by chance with probability at most .05

Data If the true RR is less than 1.09 (greater than 1.23), then the probability of observing a RR as large (small) as 1.16 happens by chance with probability less than .05.

Confidence Intervals and Practical Significance

- Determination of practical significance is discipline specific.

Confidence Intervals and Practical Significance

- Determination of practical significance is discipline specific.
- If RR greater than 1.05 are considered clinically important, then 95% CI $(1.09, 1.23)$ shows that the increased risk is both statistically and practically significant.

Confidence Intervals and Practical Significance

- Determination of practical significance is discipline specific.
- If RR greater than 1.05 are considered clinically important, then 95% CI $(1.09, 1.23)$ shows that the increased risk is both statistically and practically significant.
- Suppose the 95% CI for the difference in mean salaries is $(800, 1300)$ dollars per year. If differences greater than 1000 are considered important (actionable), then the difference may or may not be of practical importance.

Confidence Intervals and Practical Significance

- Determination of practical significance is discipline specific.
- If RR greater than 1.05 are considered clinically important, then 95% CI $(1.09, 1.23)$ shows that the increased risk is both statistically and practically significant.
- Suppose the 95% CI for the difference in mean salaries is $(800, 1300)$ dollars per year. If differences greater than 1000 are considered important (actionable), then the difference may or may not be of practical importance.
- Confidence intervals address both statistical and practical significance.

NEJM article: Breast Cancer Recurrence

The 2007 NEJM article (link) "MRI Evaluation of the Contralateral Breast in Women with Recently Diagnosed Breast Cancer" (29 March,**356**, pp. 1295-1303) found that 6 out of 101 women with a particular type of cancer in one breast, had cancer in the other breast.

NEJM article: Breast Cancer Recurrence

The 2007 NEJM article (link) “MRI Evaluation of the Contralateral Breast in Women with Recently Diagnosed Breast Cancer” (29 March, **356**, pp. 1295-1303) found that 6 out of 101 women with a particular type of cancer in one breast, had cancer in the other breast.

- Suppose there is a claim that the recurrence rate is higher than 6%. Say 10%, 12%, or 15%.

NEJM article: Breast Cancer Recurrence

The 2007 NEJM article (link) "MRI Evaluation of the Contralateral Breast in Women with Recently Diagnosed Breast Cancer" (29 March, **356**, pp. 1295-1303) found that 6 out of 101 women with a particular type of cancer in one breast, had cancer in the other breast.

- Suppose there is a claim that the recurrence rate is higher than 6%. Say 10%, 12%, or 15%.
- We calculate the probability of observing as few as 6 recurrences when $p = .10, .12, .15$.
- 6-out-of-101-applet

Breast Cancer Recurrence (cont)

- We want Prob(Claim|Data); we address this using Prob(Data|Claim)

Breast Cancer Recurrence (cont)

- We want Prob(Claim|Data); we address this using Prob(Data|Claim)
- Prob(6 or fewer recurrences | true rate is 12.48%) $= .0501$
- Prob(6 or fewer recurrences | true rate is 12.49%) $= .0498$

Breast Cancer Recurrence (cont)

- We want Prob(Claim|Data); we address this using Prob(Data|Claim)
- Prob(6 or fewer recurrences | true rate is 12.48%) = .0501
- Prob(6 or fewer recurrences | true rate is 12.49%) = .0498
- So 12.48% is in the 95% CI but 12.49% is not.
- Similar steps can be taken for claims that the true rate is *lower* than 6%

Breast Cancer Recurrence (cont)

- We want Prob(Claim|Data); we address this using Prob(Data|Claim)
- Prob(6 or fewer recurrences | true rate is 12.48%) = .0501
- Prob(6 or fewer recurrences | true rate is 12.49%) = .0498
- So 12.48% is in the 95% CI but 12.49% is not.
- Similar steps can be taken for claims that the true rate is *lower* than 6%
- Prob(6 or more recurrences | true rate is 2.22%) .0509
- Prob(6 or more recurrences | true rate is 2.21%) .0499

Breast Cancer Recurrence (cont)

- We want Prob(Claim|Data); we address this using Prob(Data|Claim)
- Prob(6 or fewer recurrences | true rate is 12.48%) = .0501
- Prob(6 or fewer recurrences | true rate is 12.49%) = .0498
- So 12.48% is in the 95% CI but 12.49% is not.
- Similar steps can be taken for claims that the true rate is *lower* than 6%
- Prob(6 or more recurrences | true rate is 2.22%) .0509
- Prob(6 or more recurrences | true rate is 2.21%) .0499
- 95% CI for true rate is 2.22% to 12.48%

Breast Cancer Recurrence (cont)

- We want Prob(Claim|Data); we address this using Prob(Data|Claim)
- Prob(6 or fewer recurrences | true rate is 12.48%) = .0501
- Prob(6 or fewer recurrences | true rate is 12.49%) = .0498
- So 12.48% is in the 95% CI but 12.49% is not.
- Similar steps can be taken for claims that the true rate is *lower* than 6%
- Prob(6 or more recurrences | true rate is 2.22%) .0509
- Prob(6 or more recurrences | true rate is 2.21%) .0499
- 95% CI for true rate is 2.22% to 12.48%
- Paper (using normal approximation) 1% to 11%.

NEJM article: MRI sensitivity in Breast Cancer

- This study also calculated the sensitivity for this cancer.
- Sensitivity is the proportion of true positives (cancer) among all those who test positive.
- In this case, six tested positive and each had cancer so the sensitivity is $\frac{6}{6}$ or 100%.
- But what about the true sensitivity of MRI for all patients?
- 6-out-of-6 applet

MRI sensitivity in Breast Cancer (cont)

- Prob(6 out of 6 correct | true sensitivity is 54.1%) $= .0501$
- Prob(6 out of 6 correct | true sensitivity is 54.0%) $= .0496$

MRI sensitivity in Breast Cancer (cont)

- Prob(6 out of 6 correct | true sensitivity is 54.1%) = .0501
- Prob(6 out of 6 correct | true sensitivity is 54.0%) = .0496
- So 54.1% is in the 95% CI for the sensitivity but 54.0% is not.
- 95% CI for sensitivity is 54.1% to 100%

MRI sensitivity in Breast Cancer (cont)

- Prob(6 out of 6 correct | true sensitivity is 54.1%) = .0501
- Prob(6 out of 6 correct | true sensitivity is 54.0%) = .0496
- So 54.1% is in the 95% CI for the sensitivity but 54.0% is not.
- 95% CI for sensitivity is 54.1% to 100%
- Paper reports 95% CI as 100% to 100% !

MRI sensitivity in Breast Cancer (cont)

- Prob(6 out of 6 correct | true sensitivity is 54.1%) = .0501
- Prob(6 out of 6 correct | true sensitivity is 54.0%) = .0496
- So 54.1% is in the 95% CI for the sensitivity but 54.0% is not.
- 95% CI for sensitivity is 54.1% to 100%
- Paper reports 95% CI as 100% to 100% !
- Calculations were done using SAS, but the real problem was not paying attention to the assumptions involved (normal approximation cannot be used when $n = 6$) nor to the data interpretation of the CI.

MRI sensitivity in Breast Cancer (cont)

- Prob(6 out of 6 correct | true sensitivity is 54.1%) = .0501
- Prob(6 out of 6 correct | true sensitivity is 54.0%) = .0496
- So 54.1% is in the 95% CI for the sensitivity but 54.0% is not.
- 95% CI for sensitivity is 54.1% to 100%
- Paper reports 95% CI as 100% to 100% !
- Calculations were done using SAS, but the real problem was not paying attention to the assumptions involved (normal approximation cannot be used when $n = 6$) nor to the data interpretation of the CI.
- How were the data collected? Was randomization used?

Statistical Inference Summary

- Statistical inference requires knowing the type of data we have and the story behind the data.

Statistical Inference Summary

- Statistical inference requires knowing the type of data we have and the story behind the data.
- Statistical inference is motivated by DOB probability (e.g., "new treatment is better"), but

Statistical Inference Summary

- Statistical inference requires knowing the type of data we have and the story behind the data.
- Statistical inference is motivated by DOB probability (e.g., "new treatment is better"), but
- Numerical probabilities do not directly address the DOB probability (the p-value is not the probability that the two treatments are identical).

Statistical Inference Summary

- Statistical inference requires knowing the type of data we have and the story behind the data.
- Statistical inference is motivated by DOB probability (e.g., "new treatment is better"), but
- Numerical probabilities do not directly address the DOB probability (the p-value is not the probability that the two treatments are identical).
- Studies that are significant (or, highly significant) need not be important or of practical significance.

Statistical Inference Summary

- Statistical inference requires knowing the type of data we have and the story behind the data.
- Statistical inference is motivated by DOB probability (e.g., "new treatment is better"), but
- Numerical probabilities do not directly address the DOB probability (the p-value is not the probability that the two treatments are identical).
- Studies that are significant (or, highly significant) need not be important or of practical significance.
- Confidence intervals address both statistical and practical significance.