# Introduction to
# Medical Statistics II
## For Medical Students


Phichayut Phinyo

MD, Dip(ClinEpidemio) Dip(ClinStat)

Center for Clinical Epidemiology and Clinical Statistics

Department of Family Medicine

Faculty of Medicine, Chiang Mai University

# Class objectives

- After the end of this session, (hopefully) the students would be able to understand and correctly describe the concept of the following topics:

# Topics outline

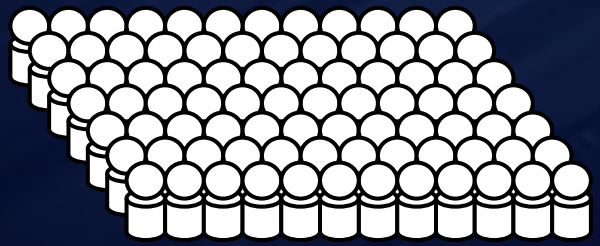
- Medical statistics II
  - Statistical inference
  - Hypothesis generation and testing
  - The alpha level and p value
  - Type I error and type II error
  - Test of statistical significance
  - Test of clinical importance
  - Confidence intervals
  - Impact (size of effect)

# Inferential Statistics

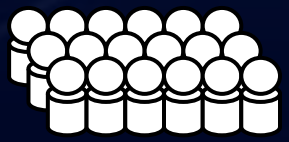
- Learning objectives
  - Statistical inference
  - Confidence intervals
  - Hypothesis generation
  - Hypothesis testing
  - The alpha level and p value
  - Type I error and type II error
  - Test of statistical significance, clinical importance
  - Impact (size of effect)

# Statistical Inference

ประชากร (Population)



Sampling
(สุ่มตัวอย่าง)



กลุ่มตัวอย่าง (Sample)

ความจริง (Truth)

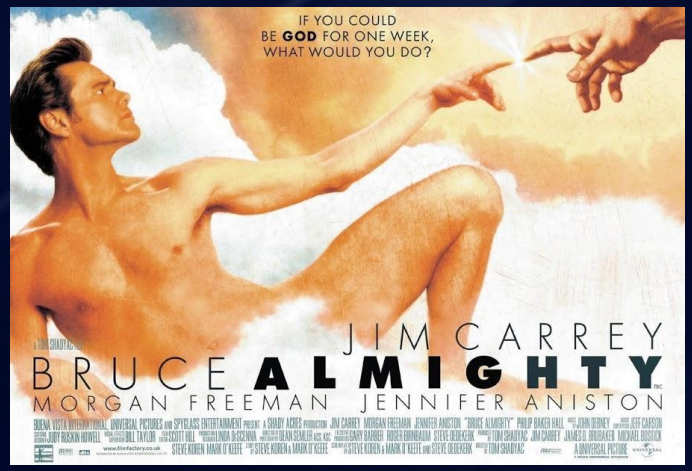

วิจัย วิเคราะห์ สรุปข้อมูล

พารามิเตอร์ของประชากร
(population parameter)

สถิติในกลุ่มตัวอย่าง
(sample statistics)

# Statistical Inference

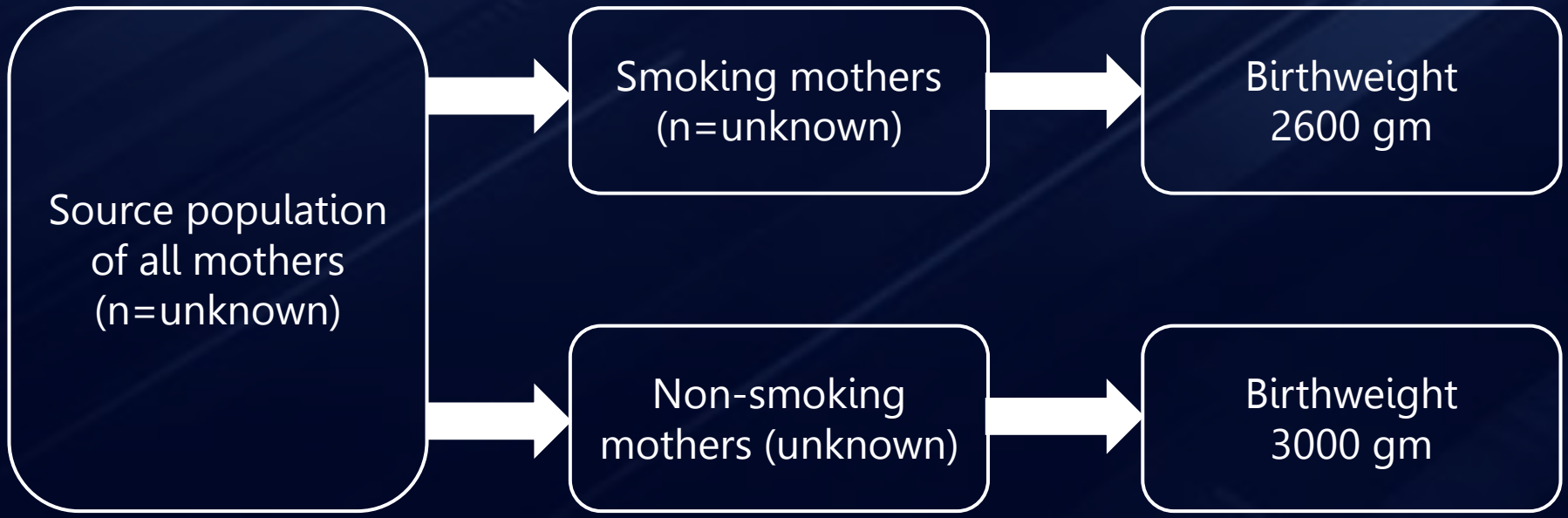
If we want to know the birthweight of newborns from smoking mothers and non-smoking mothers, how should we design and analyze our study?




Given that you are ...

# Statistical Inference


So you just log-in to your infinite database…

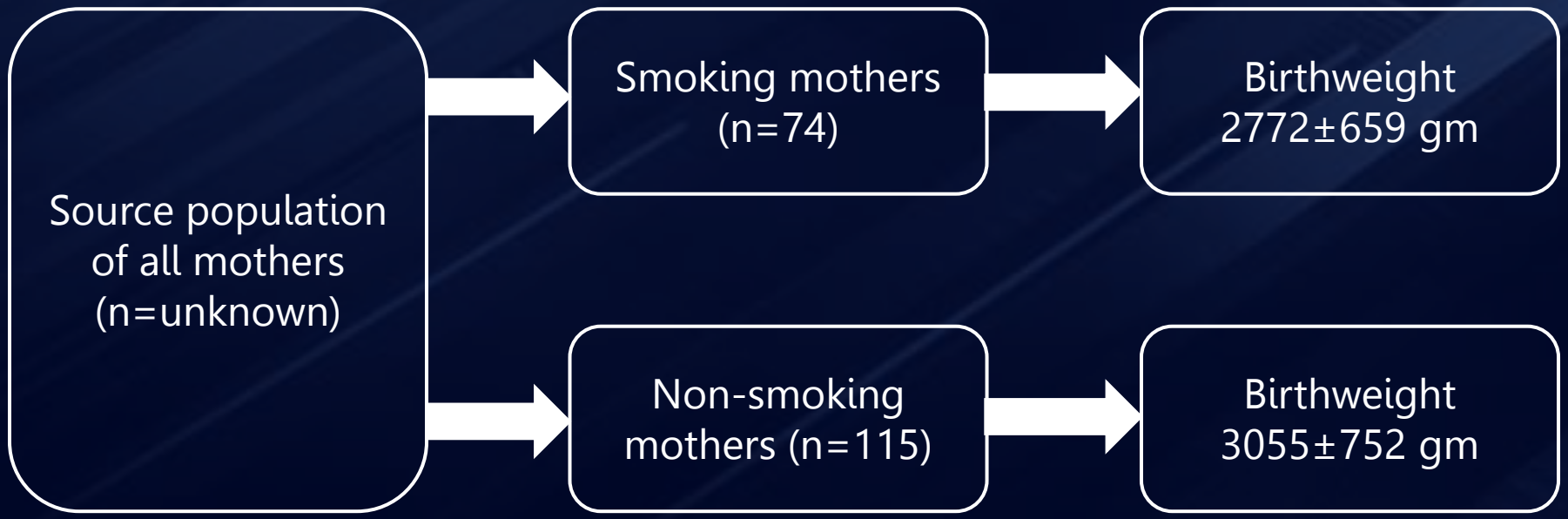


# Statistical Inference

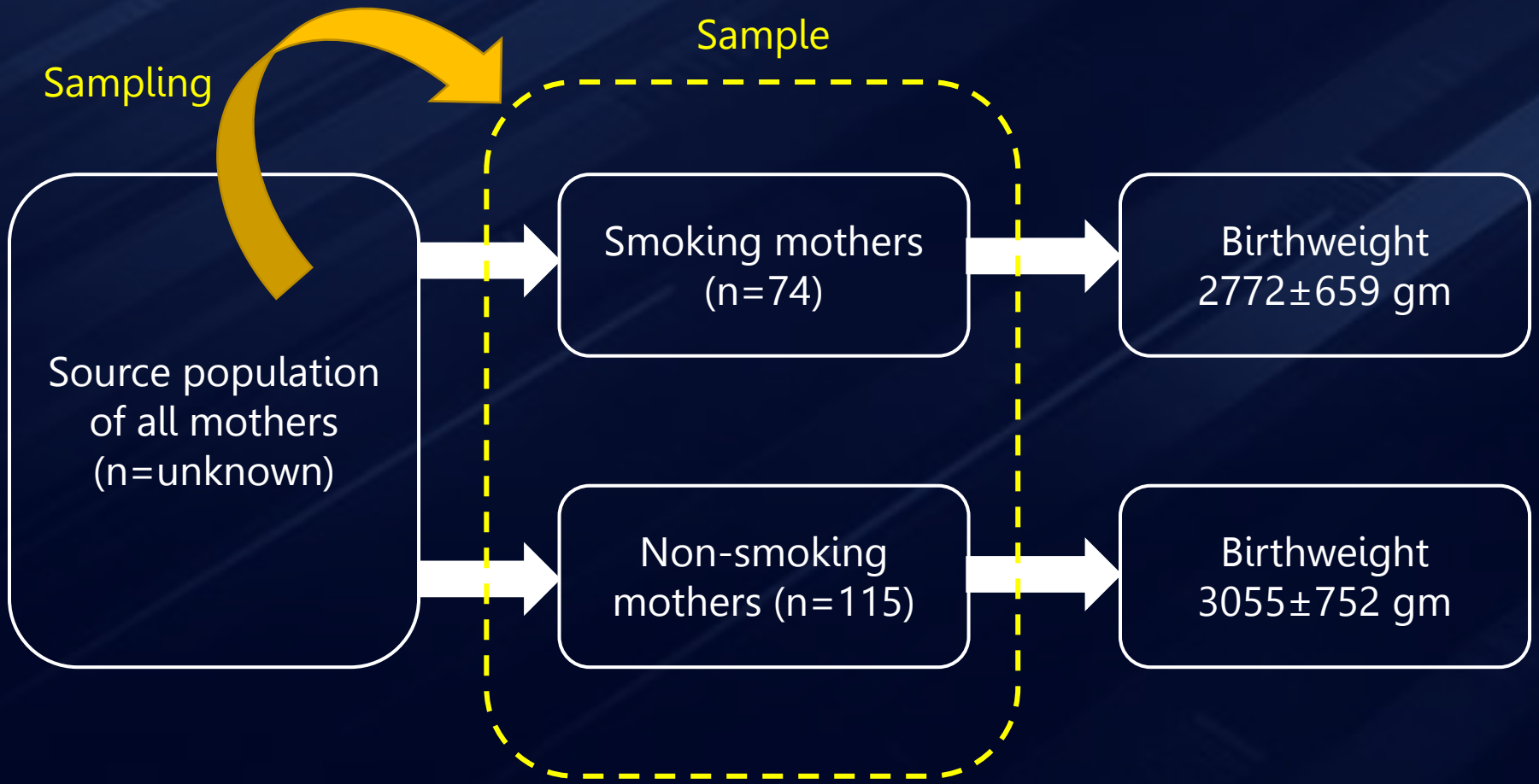
If we want to know the birthweight of newborns from smoking mothers and non-smoking mothers, how should we design and analyze our study?

But if you're not … Then, what should you do?

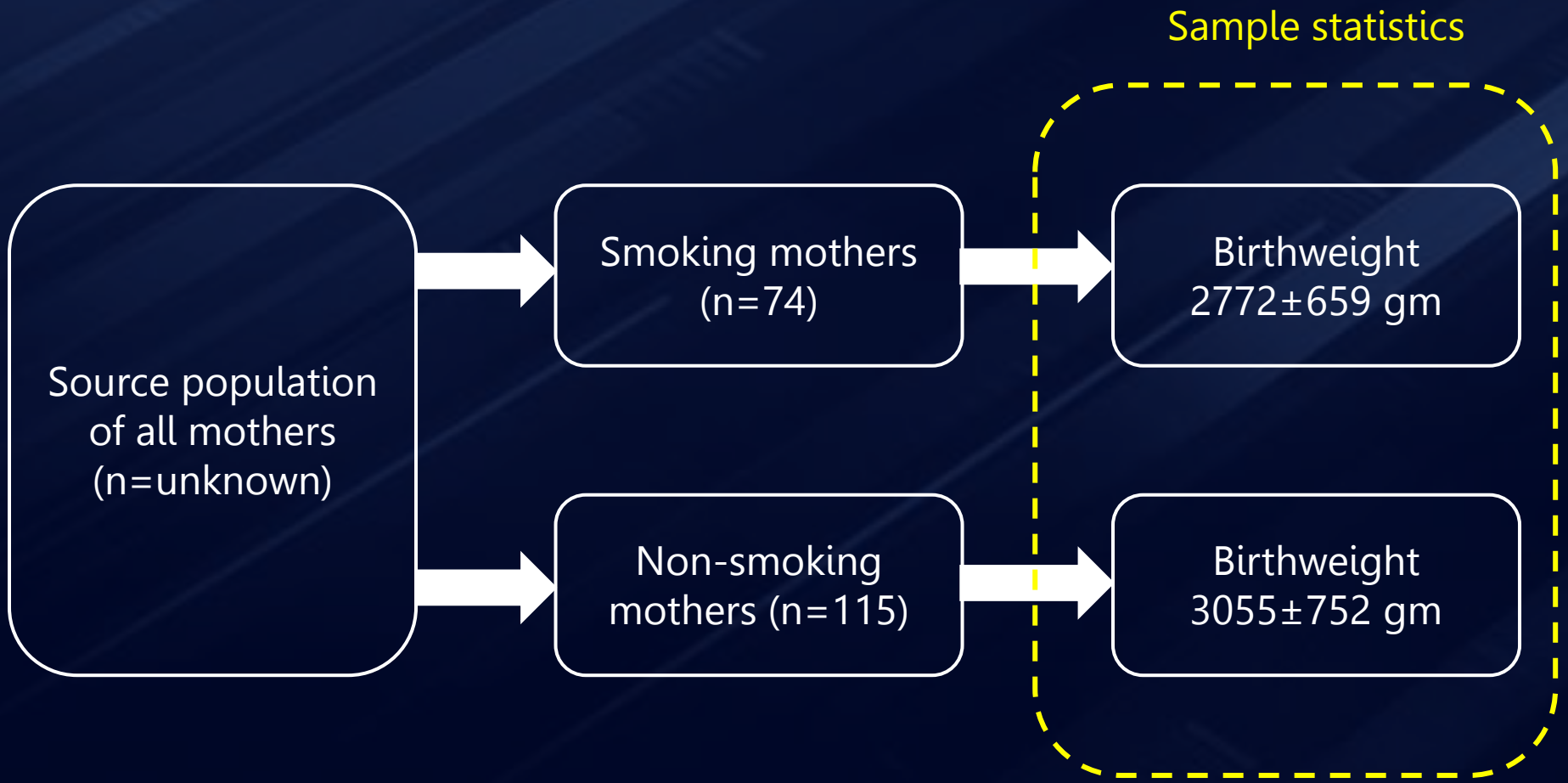
# Statistical Inference



# Statistical Inference



# Statistical Inference



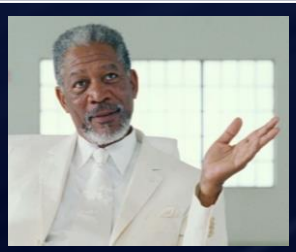
# Statistical Inference

Point estimate!

Sample statistics

Is it 2772??

True BW of smoking moms



Smoking mothers (n=74)

Birthweight 2772±659 gm

True BW of non-smoking moms

Non-smoking mothers (n=115)

Birthweight 3055±752 gm

Is it 3055??

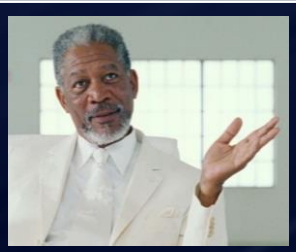
# Statistical Inference

Point estimate!

Why is there a difference??

Δ= 2772-2600

BW smoking moms= 2600



BW non-smoking moms = 3000

Δ=3055-3000

Smoking mothers (n=74)

Birthweight 2772±659 gm

Non-smoking mothers (n=115)

Birthweight 3055±752 gm

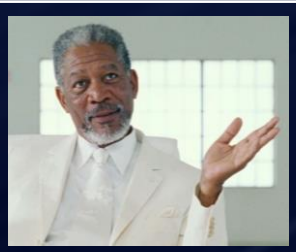
# Statistical Inference

Point estimate!

Difference between the point estimates and the unknown population mean is due to sampling error!

Δ= 2772-2600

BW smoking moms= 2600



BW non-smoking moms = 3000

Δ=3055-3000

Smoking mothers (n=74)

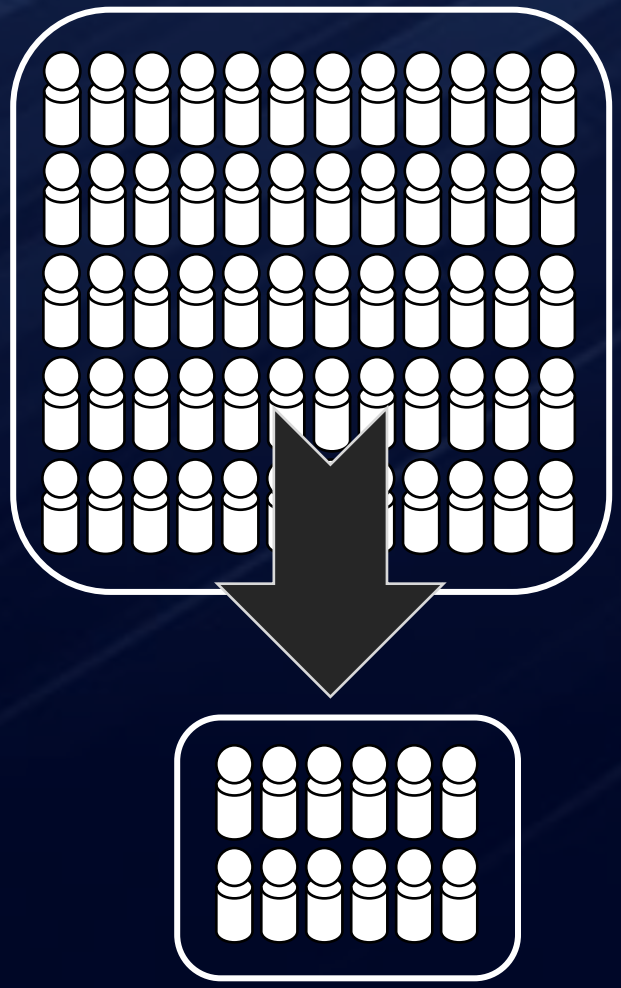
Birthweight 2772±659 gm

Non-smoking mothers (n=115)

Birthweight 3055±752 gm

Sampling error – error introduced because we are working with a sample and not the population itself.

# Statistical Inference



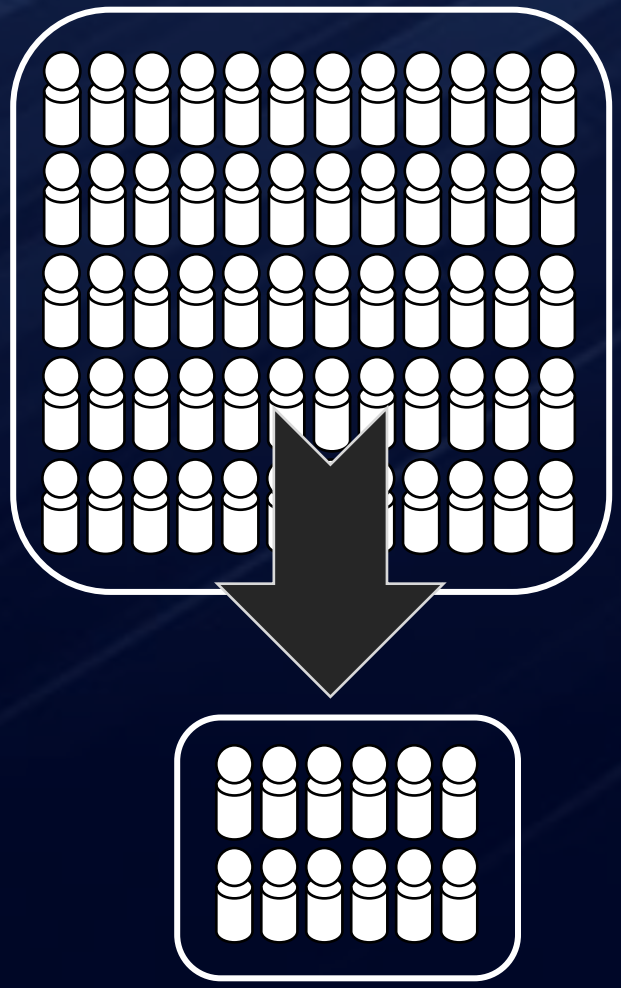
Population parameter

BW = 2600 gm??

Sample statistics

BW = 2772 gm

# Statistical Inference



Population parameter

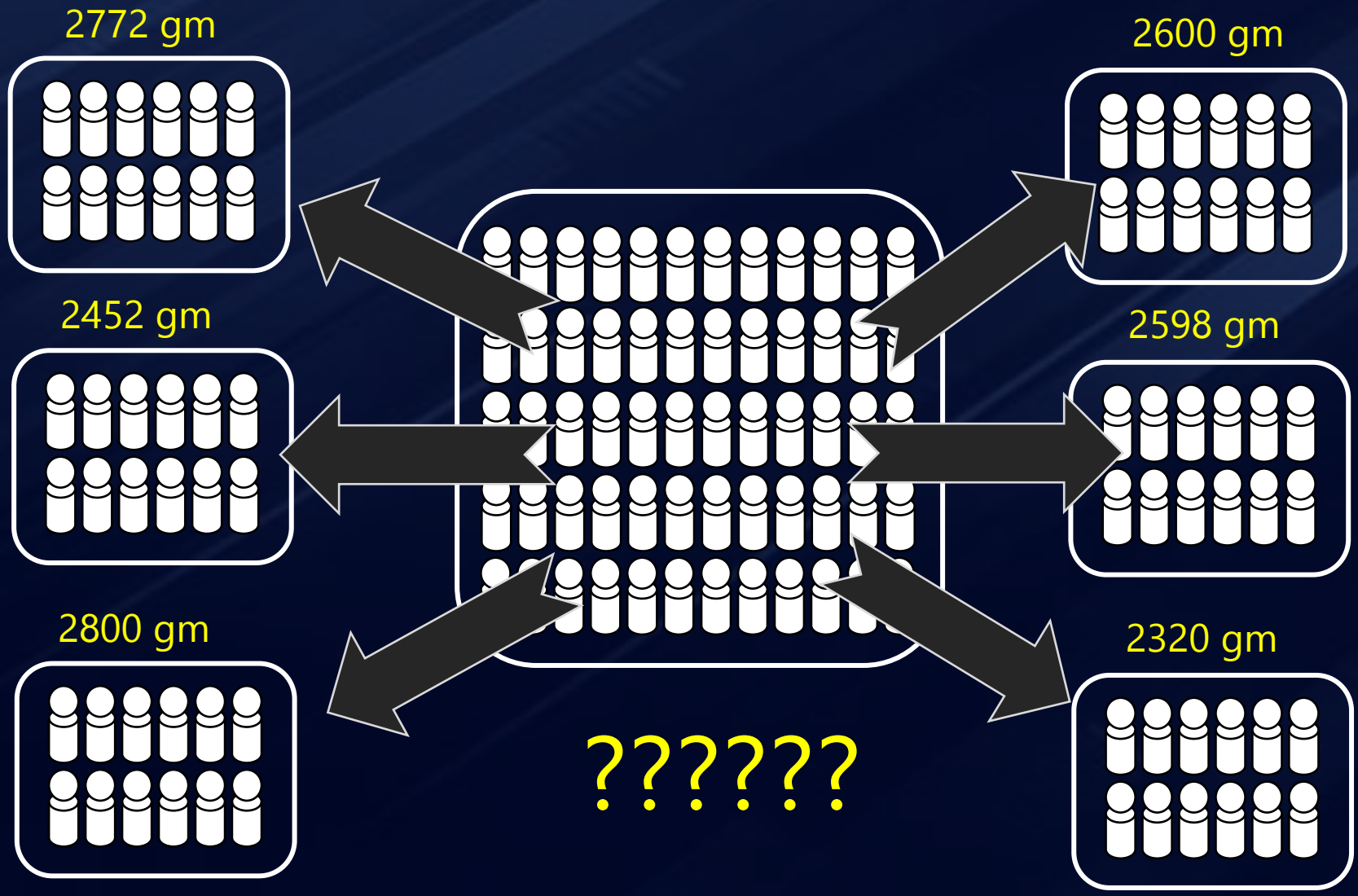
BW = 2600 gm??

**Certainty??**

Sample statistics

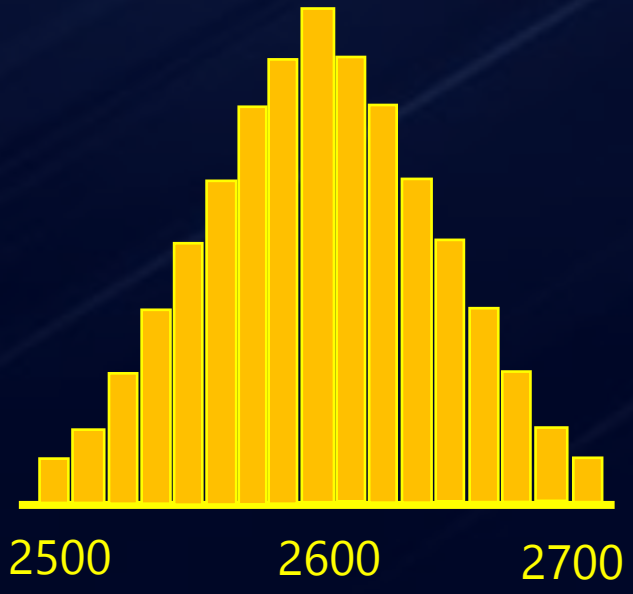
BW = 2772 gm

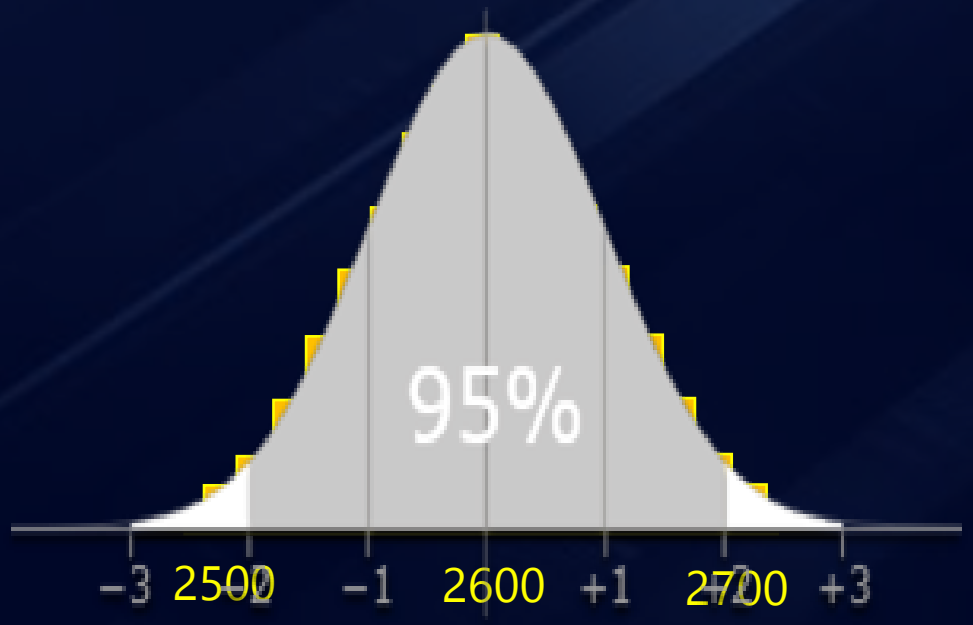
# Statistical Inference



# Statistical inference

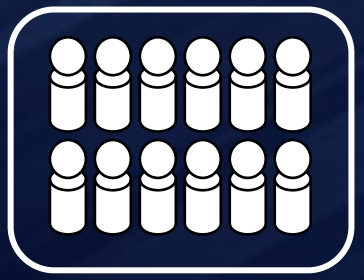
- If the number of trials are large enough then the sample means will follow a Normal distribution because of the central limit theorem



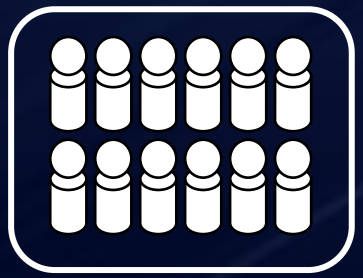



# Statistical Inference

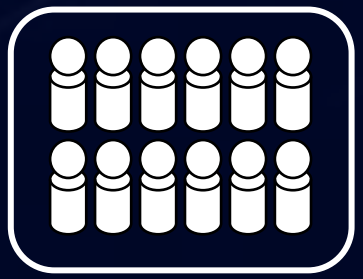
2772 gm

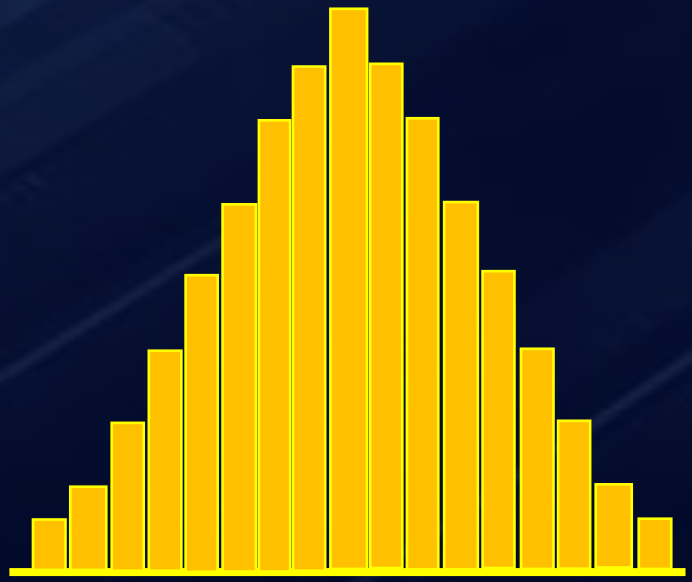


2452 gm



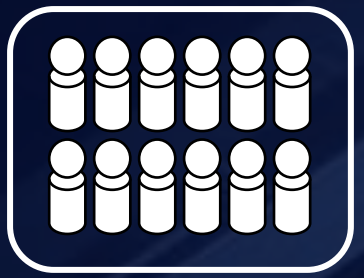
2800 gm



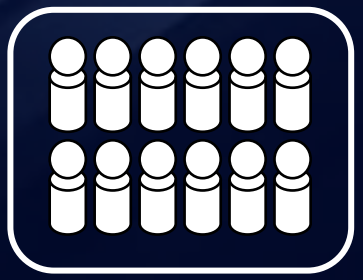



Sampling Distribution of the Mean (SEM)

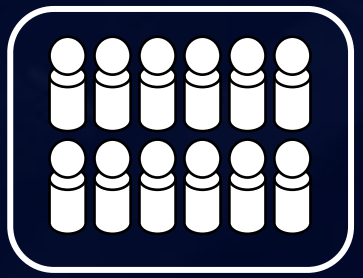
2600 gm



2598 gm



2320 gm



# Standard error

- Standard error of the mean
- SD of the mean

$$SE = \frac{SD}{\sqrt{n}}$$

As the sample size increases, the estimated standard error decreases. The bigger the sample, the smaller is the error in our estimate of population mean.

# Standard error

- Standard error of the mean
  - Why divided by $\sqrt{n}$ ?
    - Scenario 1: Mean 100 SD 10 N=10
    - Scenario 2: Mean 100 SD 10 N=100
    - Scenario 3: Mean 100 SD 10 N=1000
    - Scenario 4: Mean 100 SD 10 N=10000
    - Scenario 5: Mean 100 SD 10 N=100000
    - Calculate the SE for each scenario

# Standard error

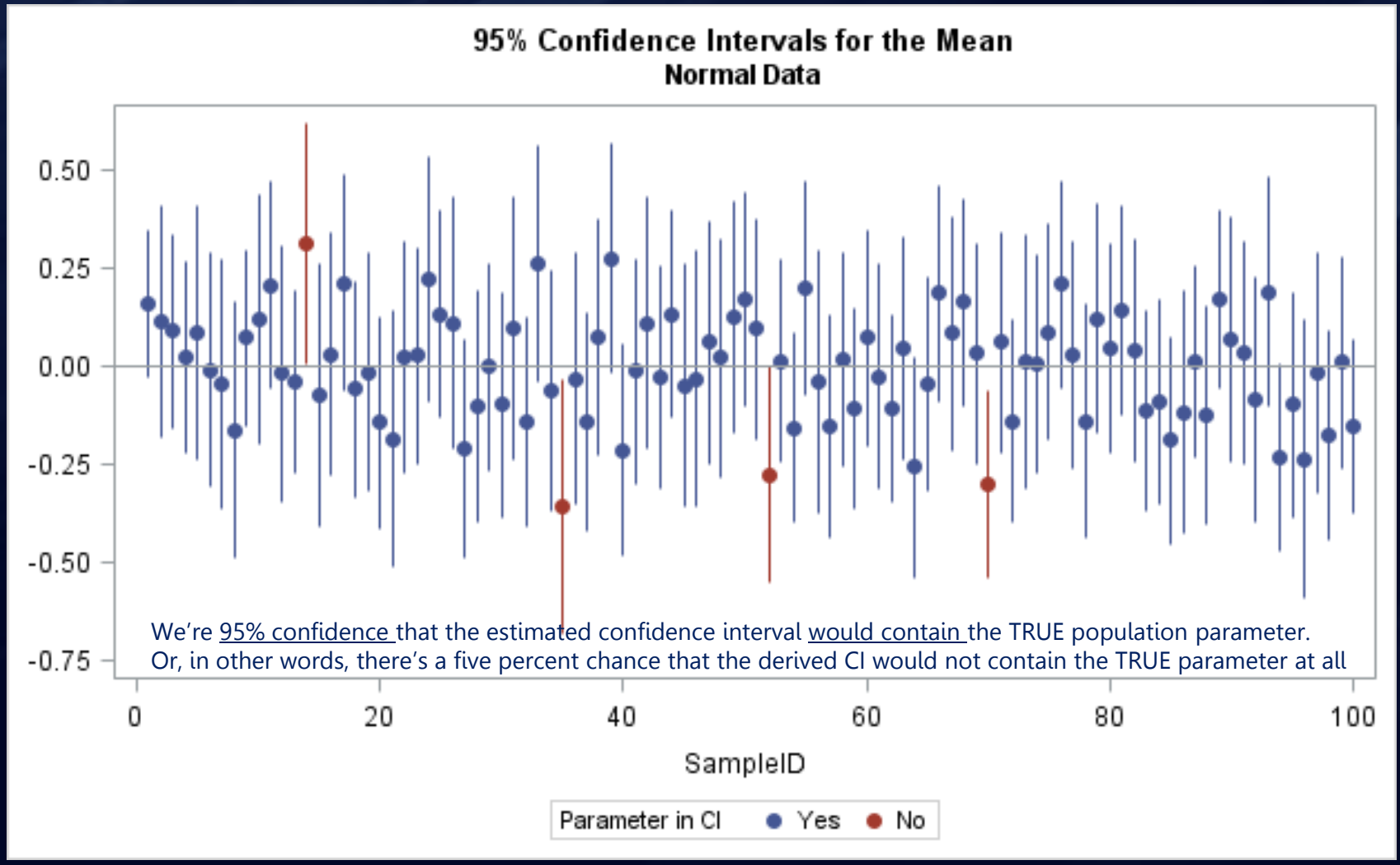
- Standard error of the mean
  - Why divided by $\sqrt{n}$ ?
    - Scenario 1: Mean 100 SD 10 N=10 SE 3.16
    - Scenario 2: Mean 100 SD 10 N=100 SE 1.00
    - Scenario 3: Mean 100 SD 10 N=1000 SE 0.32
    - Scenario 4: Mean 100 SD 10 N=10000 SE 0.10
    - Scenario 5: Mean 100 SD 10 N=100000 SE 0.03
    - Calculate the SE for each scenario

As sample size increases, the standard error of the mean decreases and will continue to approach zero as your sample size increases infinitely (closer to the TRUE population size)

# Confidence interval

- 95% confidence interval
  - Upper bound
    Mean + 1.96 (SE)
  - Lower bound
    Mean – 1.96 (SE)

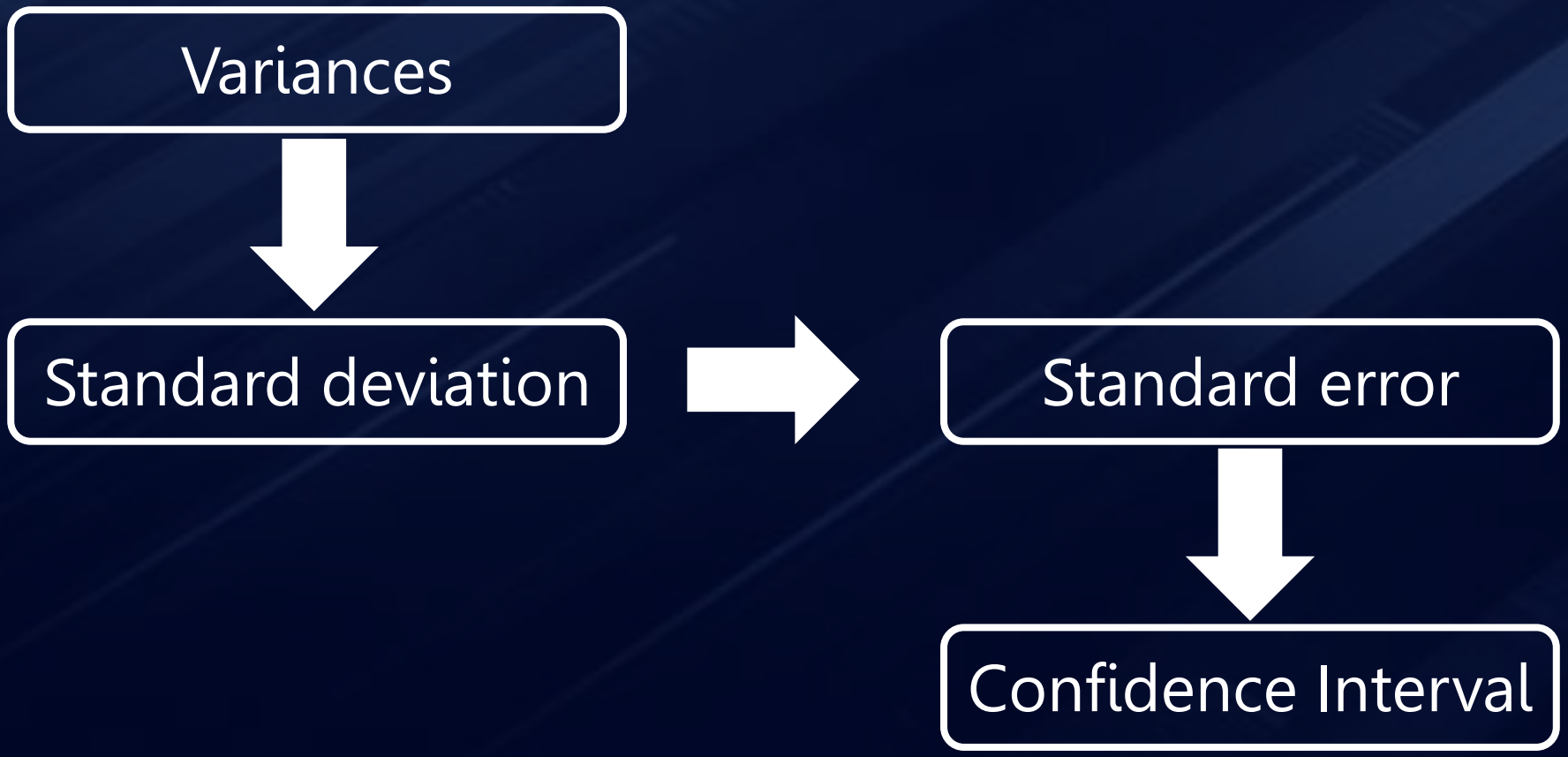
# 95% Confidence interval



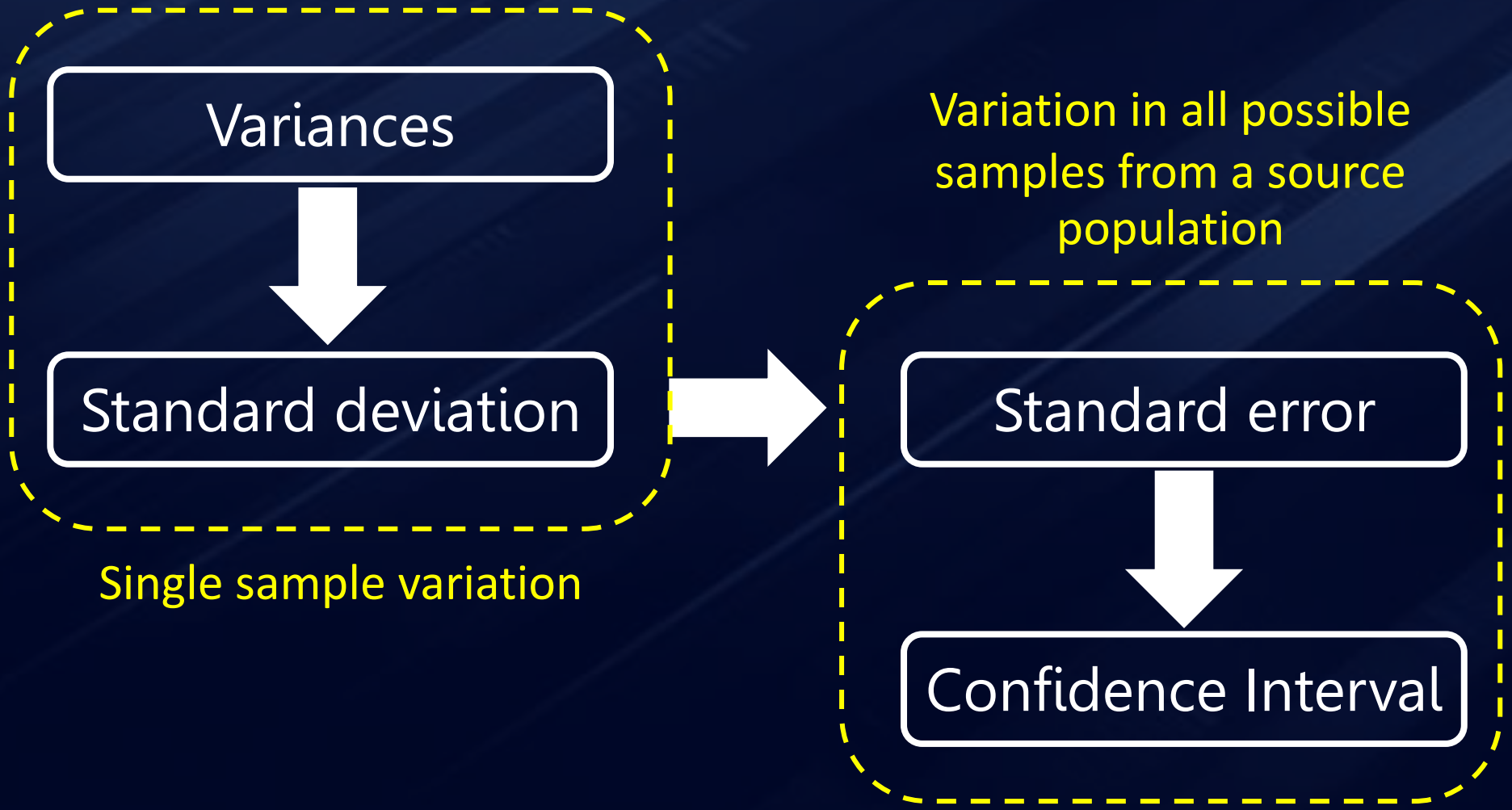
# Confidence interval

| Confidence Level | Critical Values (z) |
|---|---|
| 80% | 1.28 |
| 85% | 1.44 |
| 90% | 1.64 |
| 95% | 1.96 |
| 99% | 2.57 |

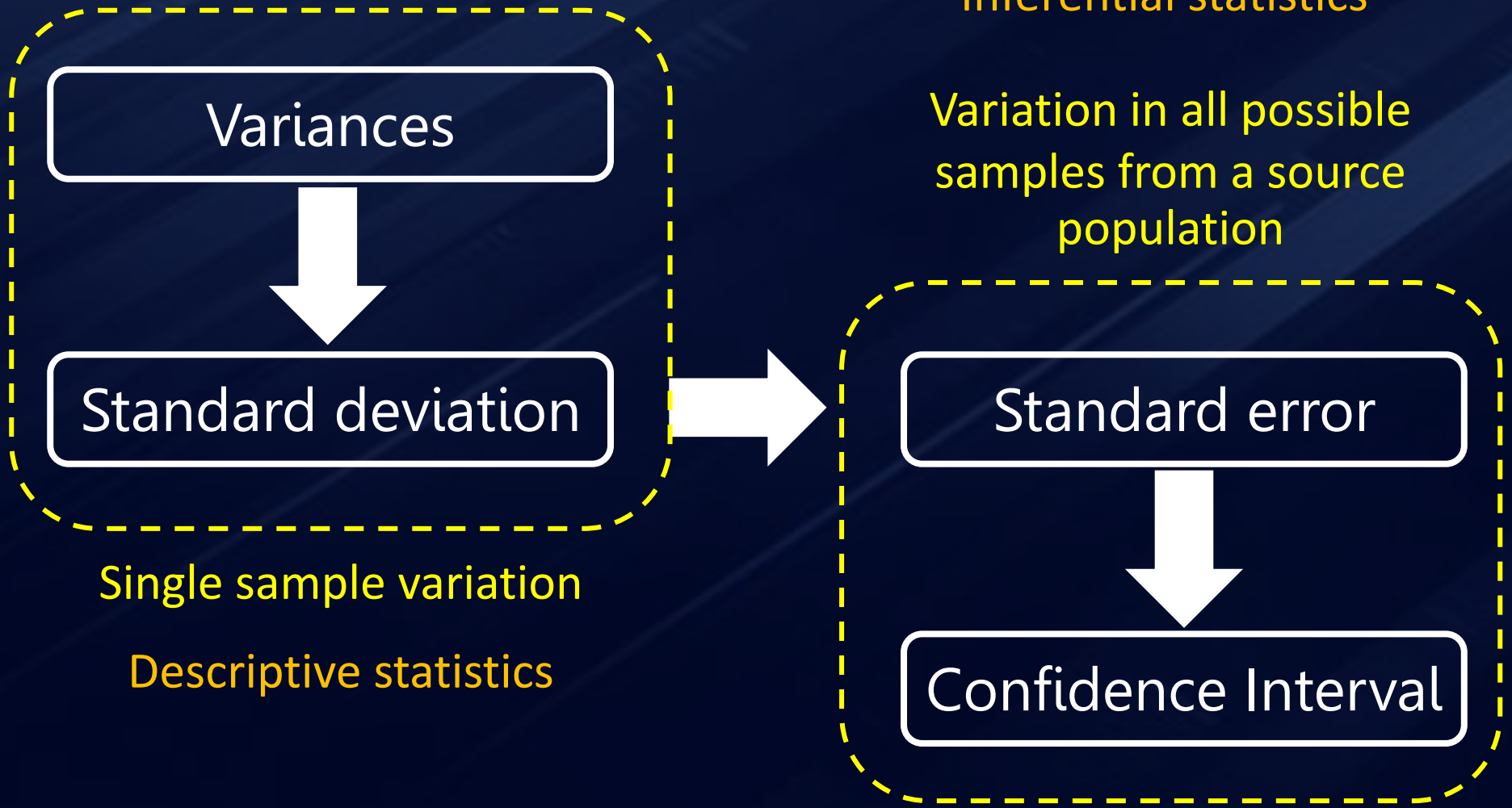
# Continuum


Variances
Standard deviation
Standard error
Confidence Interval

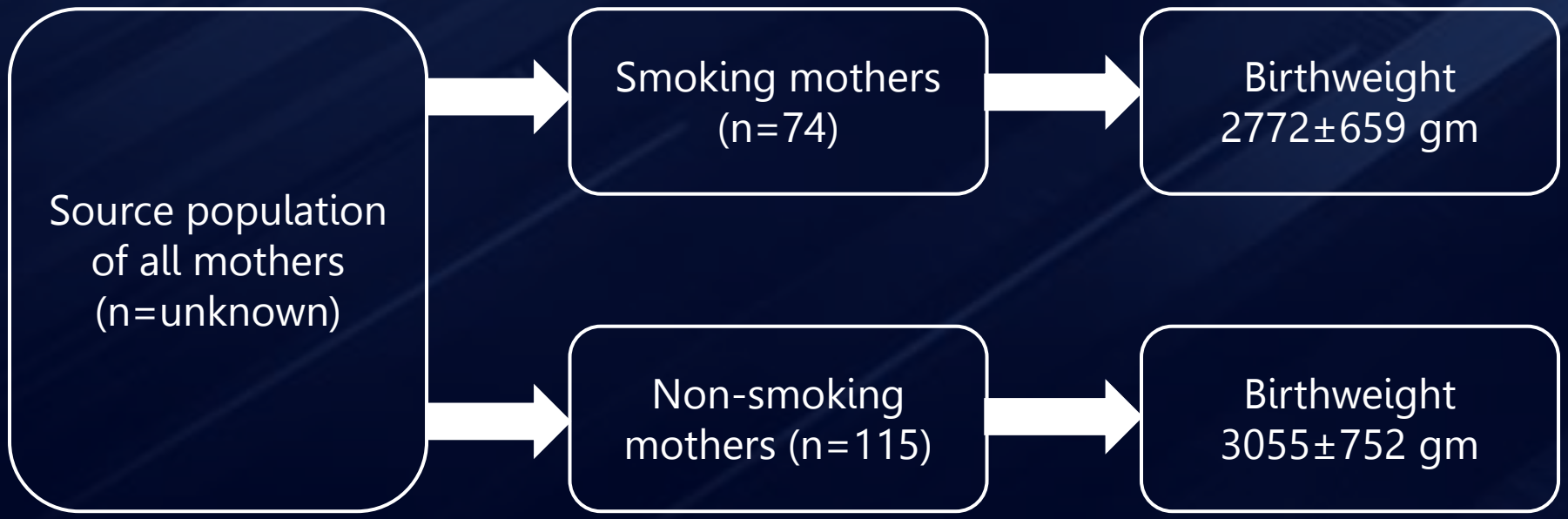
# Continuum


Variances
Standard deviation
Single sample variation
Variation in all possible samples from a source population
Standard error
Confidence Interval

# Continuum


Variances
Standard deviation
Single sample variation
Descriptive statistics
Inferential statistics
Variation in all possible samples from a source population
Standard error
Confidence Interval

# Calculate the 95%CI



# 95%CI

- Calculation formula
  - Mean±1.96(SE)
  - For smoking moms:

    SE=659/ $\sqrt{74}$ = 76.6

    1.96xSE = 150.2 => 2621.8 to 2922.2

  - For non-smoking moms:

    SE=752/$\sqrt{115}$ = 70.1

    1.96xSE = 137.4 => 2917.6 to 3192.4

```
. tabstat birthwt,by(smoke) stat(mean sd n median p25 p75)

Summary for variables: birthwt
     by categories of: smoke (smoking during pregnancy: 0=no, 1=yes)
```

| smoke | mean | sd | N | p50 | p25 | p75 |
|---|---|---|---|---|---|---|
| 0 | 3054.957 | 752.409 | 115 | 3100 | 2495 | 3629 |
| 1 | 2772.297 | 659.8075 | 74 | 2775.5 | 2367 | 3260 |
| Total | 2944.286 | 729.016 | 189 | 2977 | 2414 | 3475 |

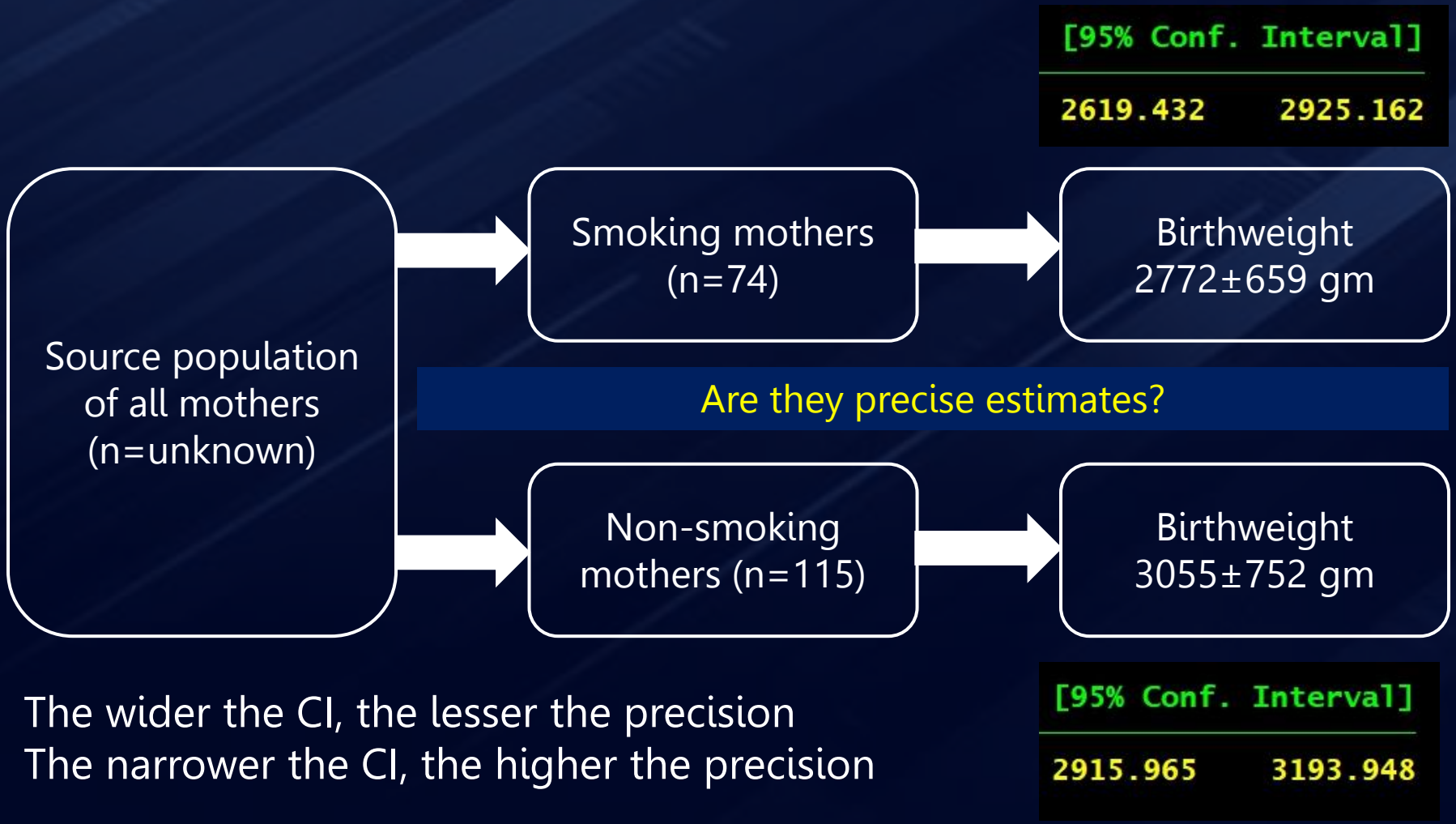
```
. ci means birthwt if smoke==1
```

| Variable | Obs | Mean | Std. Err. | [95% Conf. | Interval] |
|---|---|---|---|---|---|
| birthwt | 74 | 2772.297 | 76.70106 | 2619.432 | 2925.162 |

```
. ci means birthwt if smoke==0
```

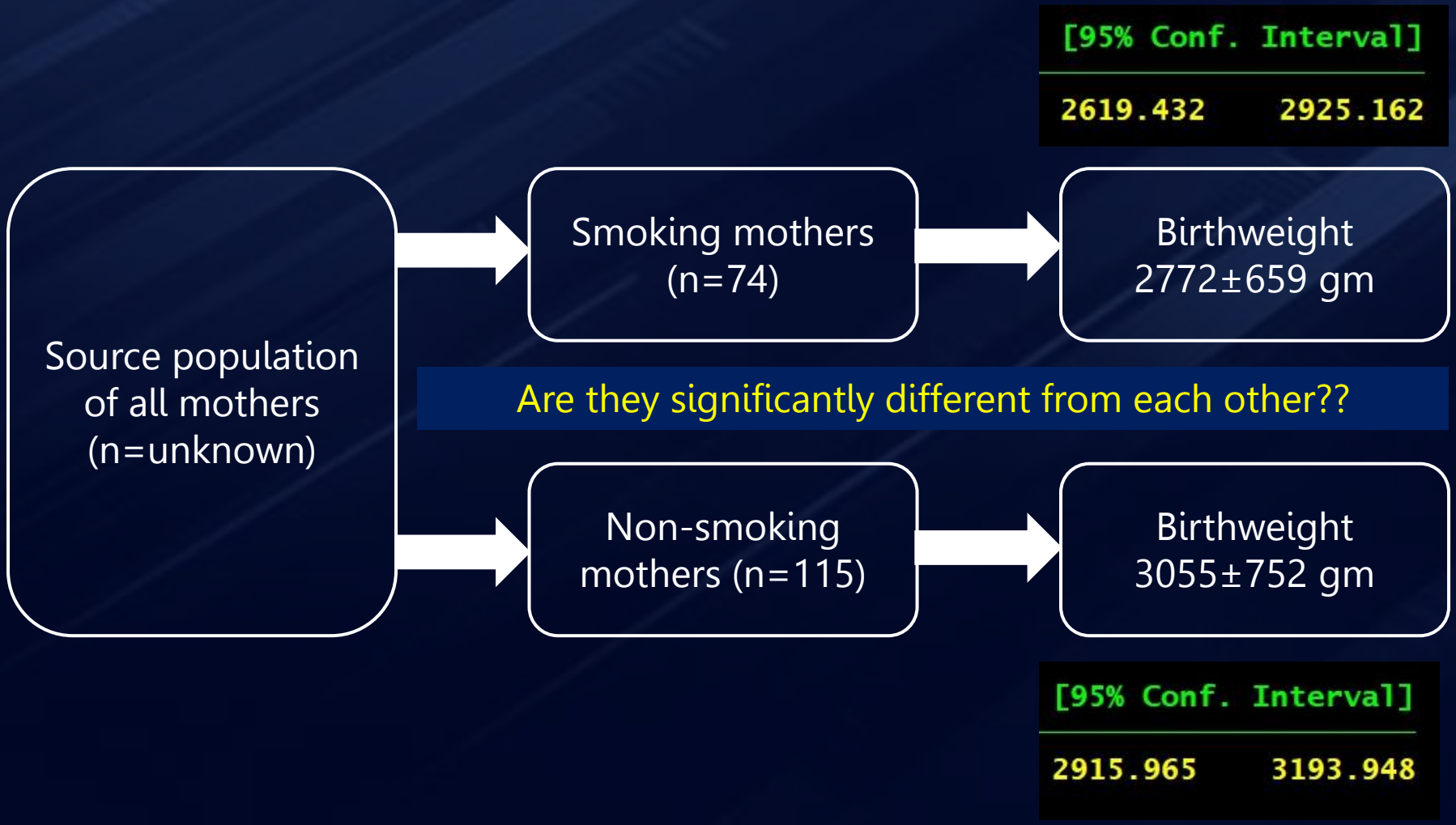
| Variable | Obs | Mean | Std. Err. | [95% Conf. | Interval] |
|---|---|---|---|---|---|
| birthwt | 115 | 3054.957 | 70.1625 | 2915.965 | 3193.948 |

# Statistical inference




The wider the CI, the lesser the precision
The narrower the CI, the higher the precision

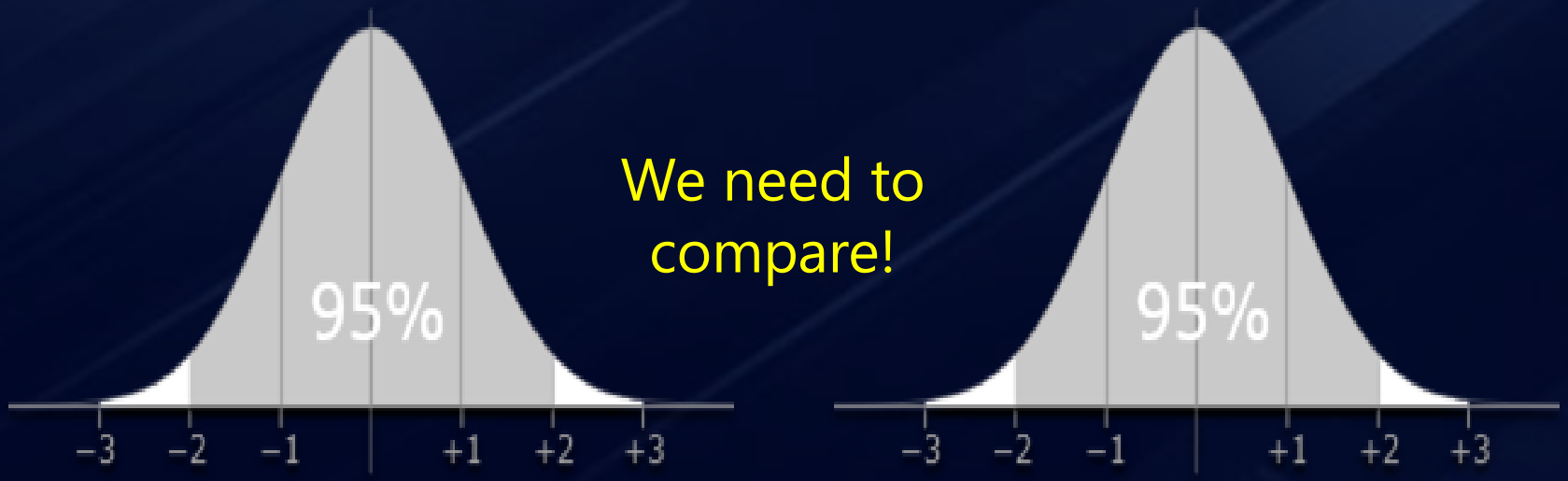
# Statistical inference



# Comparative question

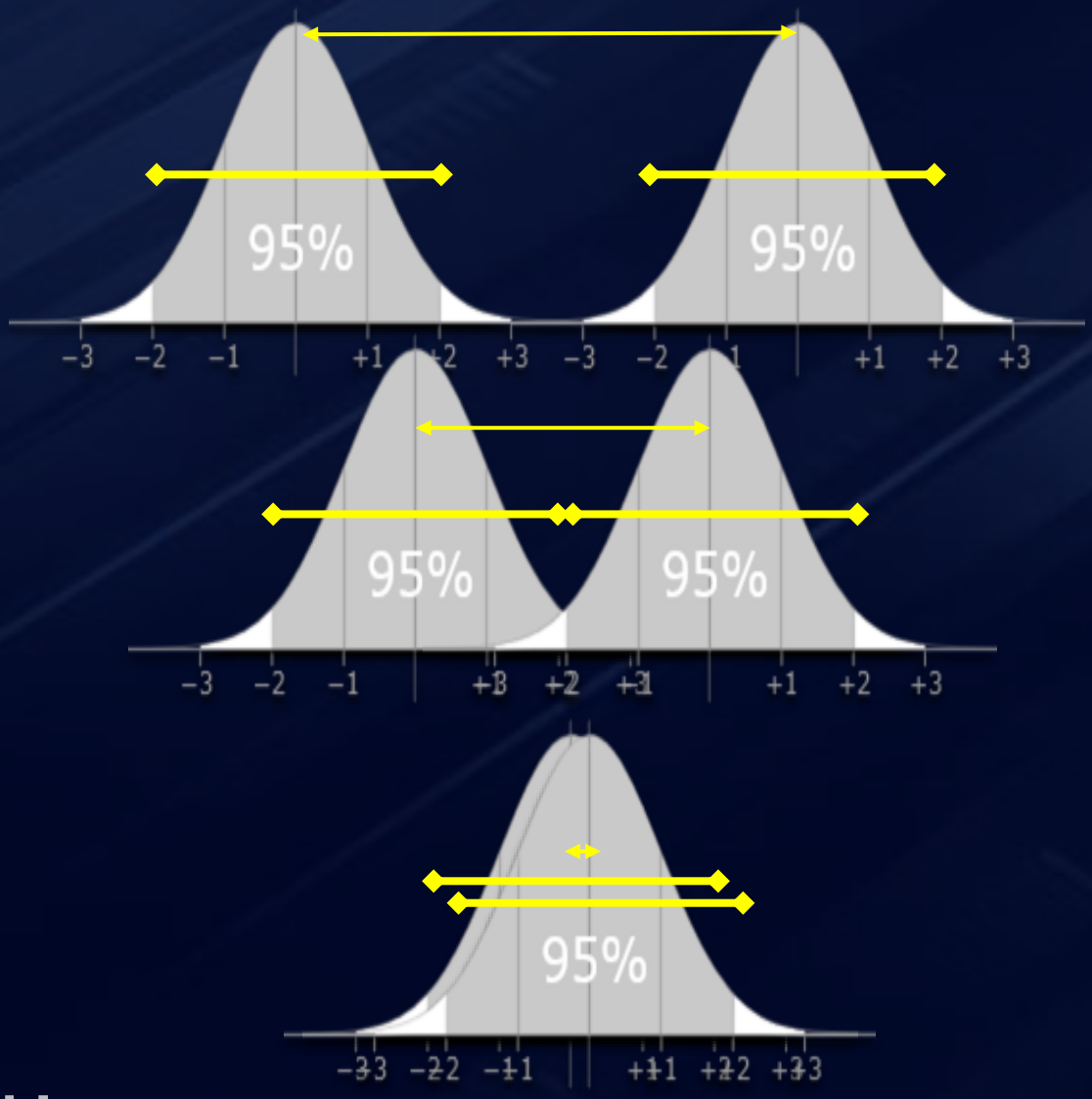
- Are they significantly different from each other?

Are the mean birthweight different between the 2 groups?

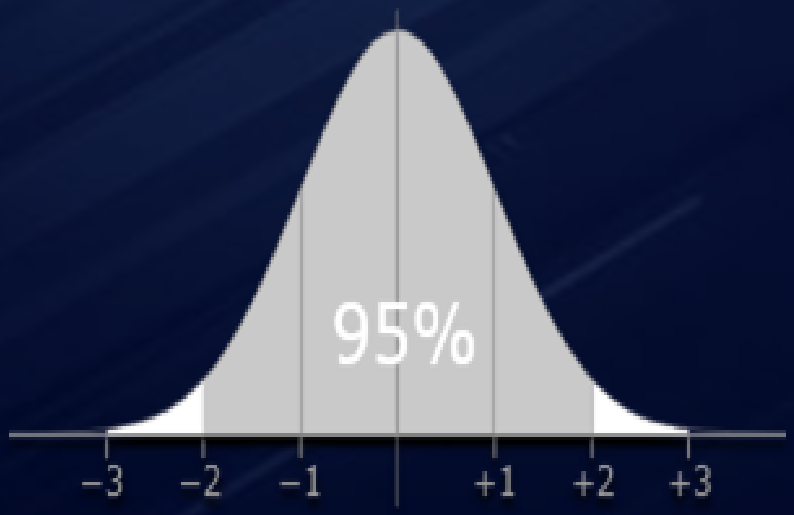



Is 2,772 equal to 3,055 ?

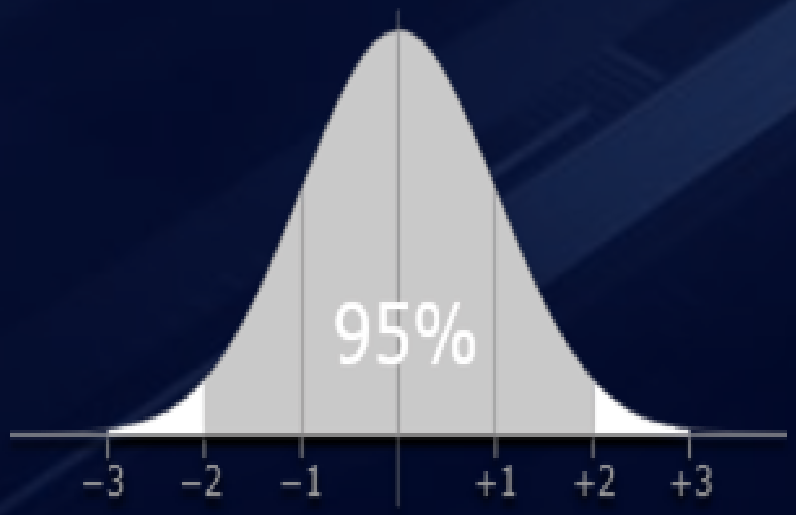
# Statistical inference



# Statistical inference

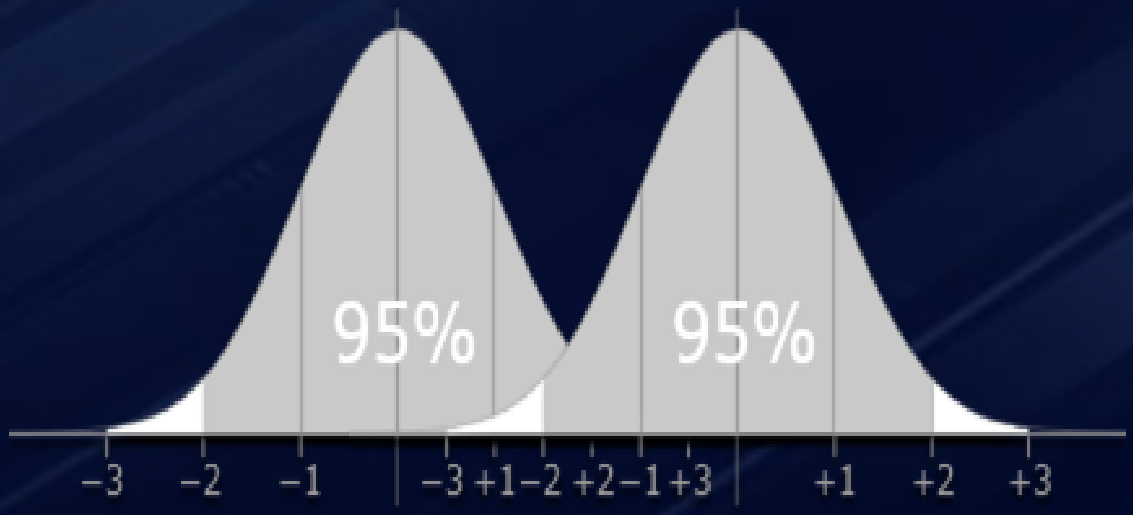



| [95% Conf. | Interval] |
|---|---|
| 2619.432 | 2925.162 |




| [95% Conf. | Interval] |
|---|---|
| 2915.965 | 3193.948 |

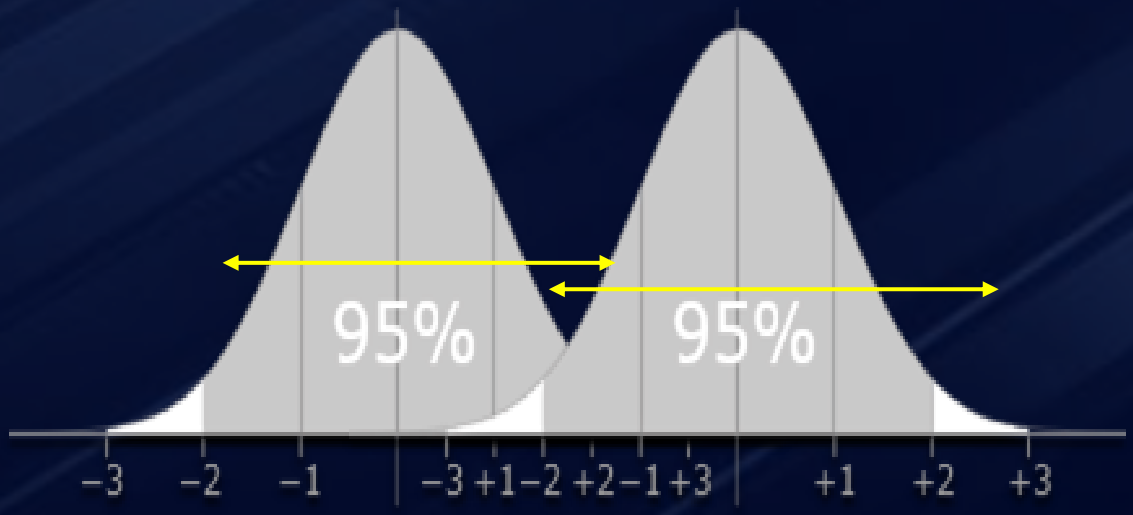
# Statistical inference




| [95% Conf. | Interval] | [95% Conf. | Interval] |
|---|---|---|---|
| 2619.432 | 2925.162 | 2915.965 | 3193.948 |

Are they significantly different from each other??

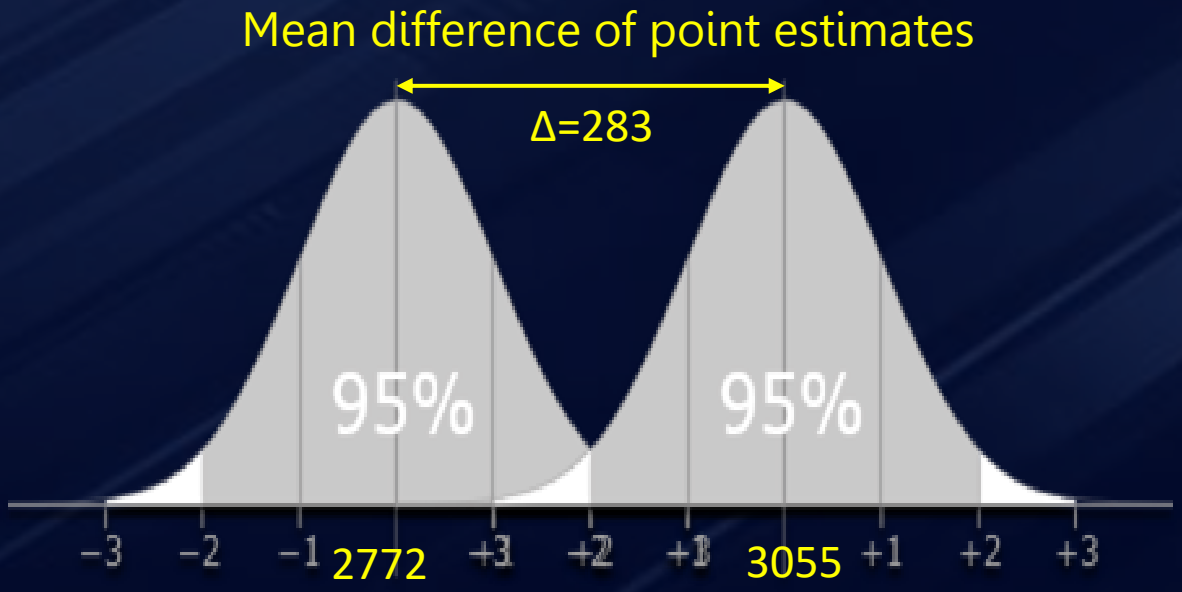
# Statistical inference




| [95% Conf. Interval] | | [95% Conf. Interval] | |
|---|---|---|---|
| 2619.432 | 2925.162 | 2915.965 | 3193.948 |

Are they significantly different from each other??

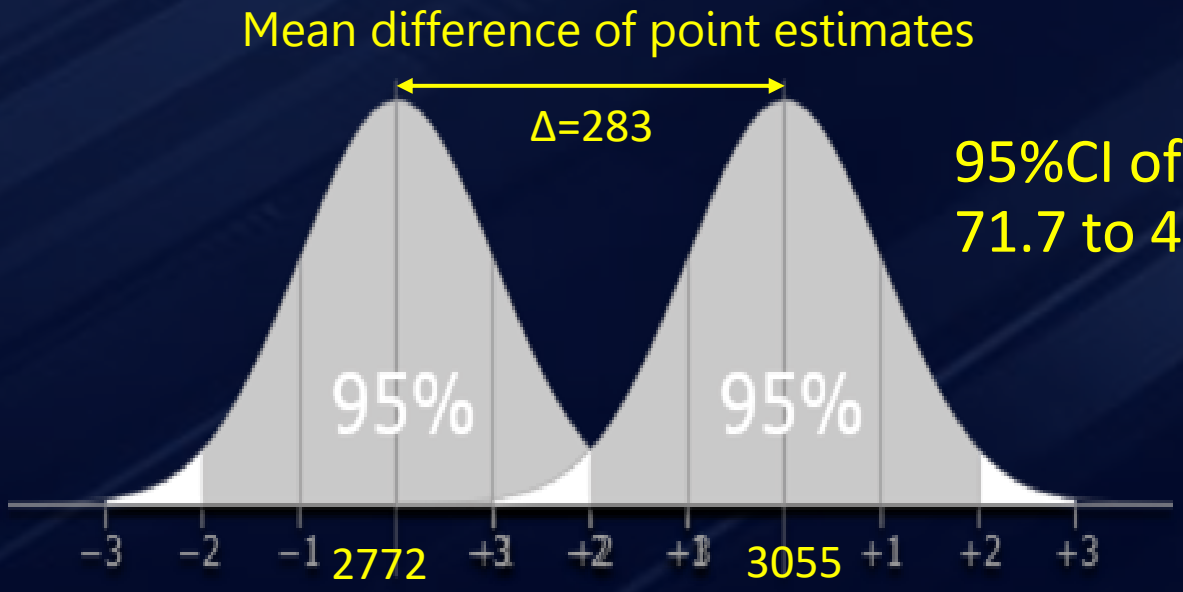
# Statistical inference




| [95% Conf. Interval] | | [95% Conf. Interval] | |
|---|---|---|---|
| 2619.432 | 2925.162 | 2915.965 | 3193.948 |

Are they significantly different from each other??

# Statistical inference




| [95% Conf. Interval] | | [95% Conf. Interval] | |
|---|---|---|---|
| 2619.432 | 2925.162 | 2915.965 | 3193.948 |

Are they significantly different from each other??

# Hypothesis testing

- General concept
  1. Making an initial assumption (hypothesis)
  2. Collect the data (evidence)
  3. Based on the available data (evidence), decide whether to reject or not reject the initial assumption (hypothesis)

# Hypothesis testing

**Criminal Trial Analogy**

"the defendant is innocent until proven guilty by evidence"

# Hypothesis generation

- Null hypothesis ($H_0$)
  - Hypothesis of no difference (contradicts to the investigators' belief)
  - Defendant is not guilty (innocent)

- Alternative hypothesis ($H_a$)
  - Hypothesis of difference (complies with the investigators' belief)
  - Defendant is guilty (criminal)

# Hypothesis generation

- Null hypothesis ($H_0$)
  - Hypothesis of no difference (contradicts to the investigators' belief)
  - Defendant is not guilty (innocent)
- Alter
  - Hypothesis of difference (complies with the investigators' belief)
  - Defendant is guilty (criminal)

No statistically testable!

# Hypothesis generation

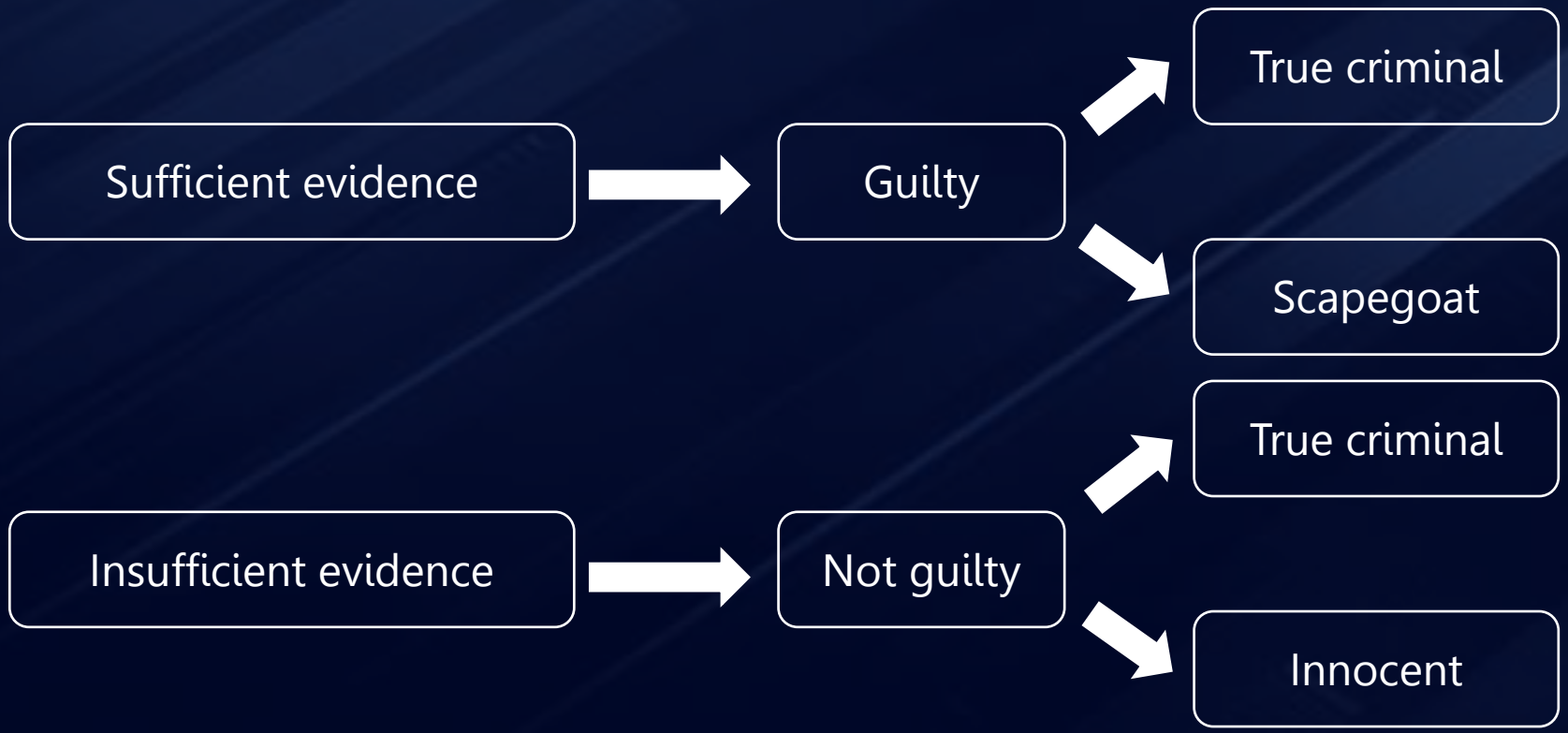
- Null hypothesis ($H_0$)
  - Hypothesis of no difference (contradicts to the investigators' belief)
  - Defendant was not at the crime scene
- Alternative hypothesis ($H_a$)
  - Hypothesis of difference (complies with the investigators' belief)
  - Defendant was at the crime scene

# Hypothesis testing

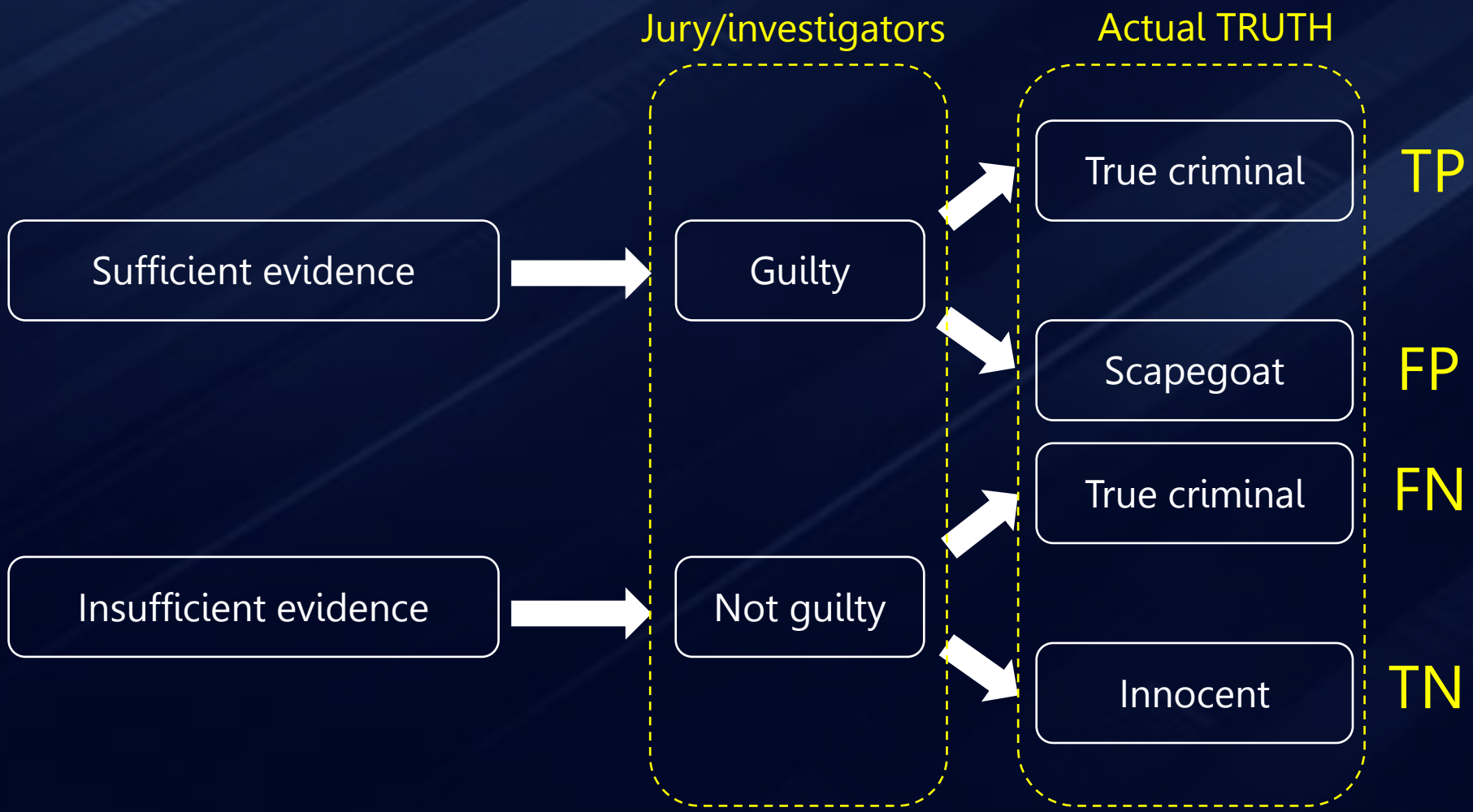
In statistics, try to think of yourself (the investigators) as the jury. The jury makes a decision based on the available evidence (data).

| | | |
|---|---|---|
| Sufficient evidence | Able to reject the null hypothesis → | Guilty and punished |
| Insufficient evidence | Not able to reject the null hypothesis → | Not guilty, behave as if the defendant is innocent |

# Hypothesis testing



# Hypothesis testing



# Hypothesis testing

| **Trial Result** | **Actual Truth** | |
|---|---|---|
| | Drug works | Drug does not work |
| Drug works | Correct | Type I error (False positive) |
| Drug does not work | Type II error (False negative) | Correct |

# Hypothesis testing

| **Trial Result** | **Actual Truth** | |
|---|---|---|
| | Drug works | Drug does not work |
| Drug works | Correct | Type I error (False positive) |
| Drug does not work | Type II error (False negative) | Correct |

# Hypothesis testing

| **Trial Result** | **Actual Truth** | |
|---|---|---|
| | Drug works | Drug does not work |
| Drug works | Correct | Type I error (False positive) |
| Drug does not work | Type II error (False negative) | Correct |

# Hypothesis generation

- Alpha error (type I error)
  - When the investigators rejected the null hypothesis, when, in fact, the null hypothesis is TRUE.
  - Accept: at critical value 0.05, 0.01
- Statistical power (type II error, Beta error)
  - When the investigators were not able to reject the null hypothesis, when, in fact, the null hypothesis is FALSE.
  - Accept: power 80-90%, Beta 10-20%

# Hypothesis generation

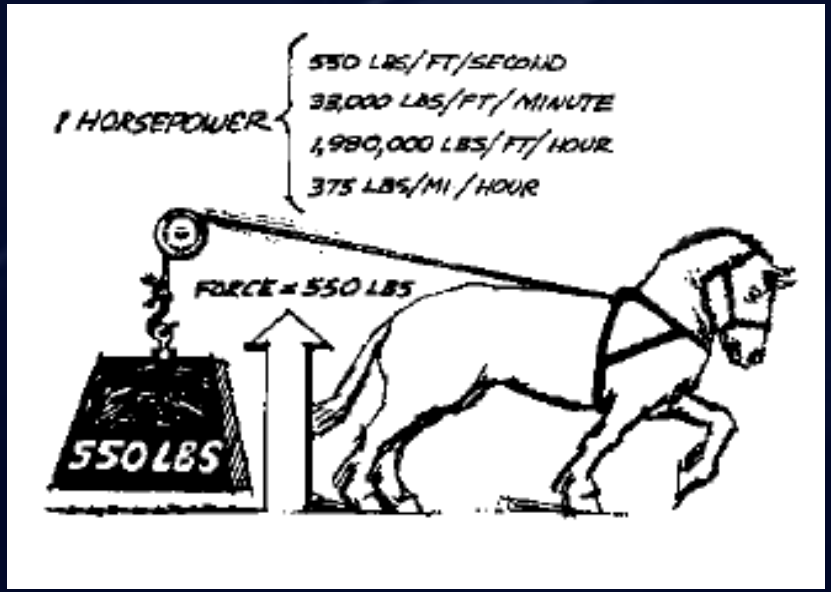
- To minimize the possibility of both type I and type II error, researchers calculate the appropriate study size for hypothesis testing!

- Factors that influence the study size
  - Statistical power
  - Alpha error
  - Variability
  - Effect size

# Hypothesis generation

- To minimize the possibility of both type I and type II error, researchers calculate the appropriate study size for hypothesis testing!

- Increase study size
- Increase statistical power
- Reduced type II error



# Hypothesis generation

- To minimize the possibility of both type I and type II error, researchers calculate the appropriate study size for hypothesis testing!

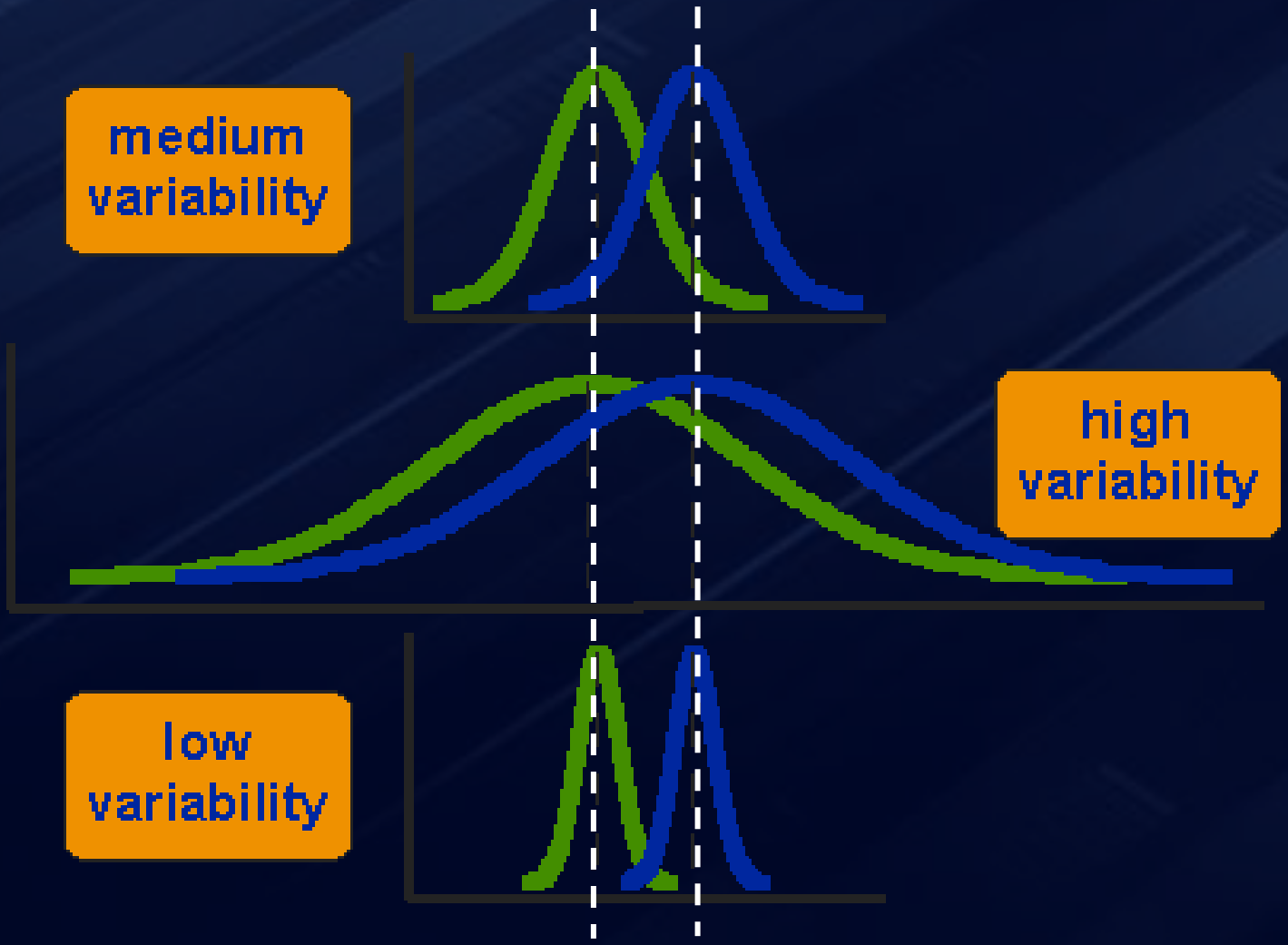
- Increase study size
- Reduce type I error
- 0.10 – small study size
- 0.05 – moderate study size
- 0.01 – highest study size

# Hypothesis generation

- To minimize the possibility of both type I and type II error, researchers calculate the appropriate study size for hypothesis testing!

- Variability
  - Large variability – needs large study size
  - Small variability – needs small study size

# Hypothesis generation



# Hypothesis generation

- To minimize the possibility of both type I and type II error, researchers calculate the appropriate study size for hypothesis testing!

- Effect size
  - Large effect – needs small study size
  - Small effect – needs large study size

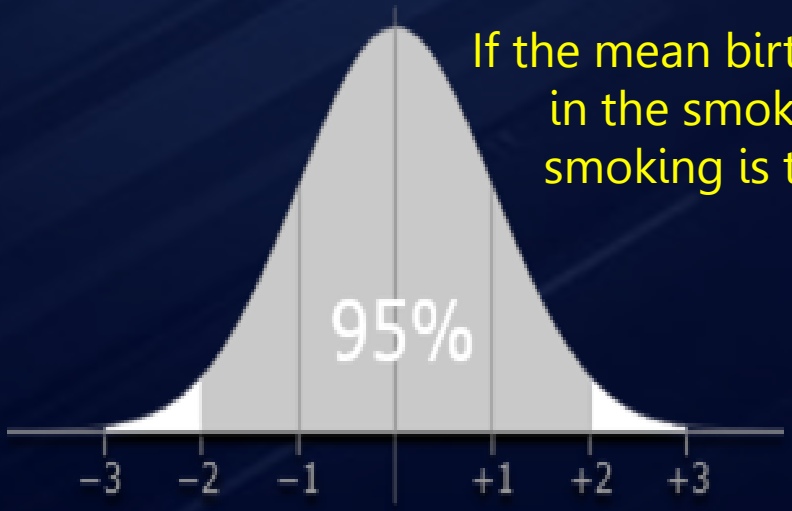
# Hypothesis generation

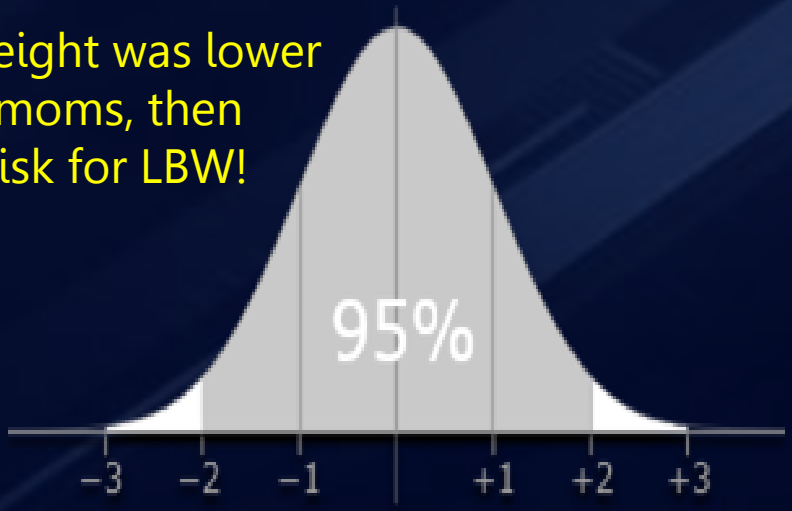
# Hypothesis testing

Smoking moms

Non-smoking moms

If the mean birth weight was lower in the smoking moms, then smoking is the risk for LBW!




| [95% Conf. | Interval] |
|---|---|
| 2619.432 | 2925.162 |

| [95% Conf. | Interval] |
|---|---|
| 2915.965 | 3193.948 |

Testing association between maternal smoking and neonatal birthweight

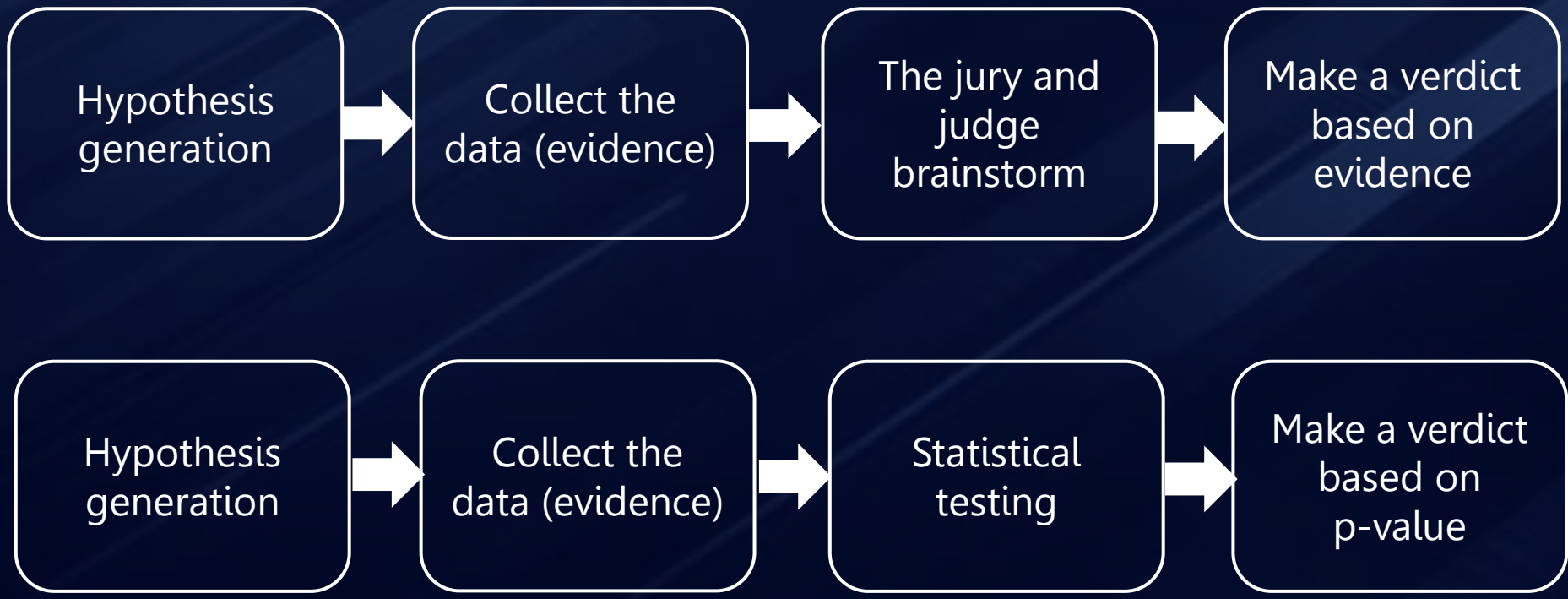
# Hypothesis generation

- Null hypothesis ($H_0$)
  - Hypothesis of no difference (contradicts to the investigators' belief)
  - BW in smokers = BW in non-smokers
- Alternative hypothesis ($H_a$)
  - Hypothesis of difference (complies with the investigators' belief)
  - BW in smokers ≠ BW in non-smokers

# Hypothesis generation

- Null hypothesis ($H_0$)
  - Hypothesis of no difference (contradicts to the investigators' belief)
  - BW in smokers = BW in non-smokers
  - BW difference = 0
- Alternative hypothesis ($H_a$)
  - Hypothesis of difference (complies with the investigators' belief)
  - BW in smokers ≠ BW in non-smokers
  - BW difference ≠ 0

# Hypothesis generation



# Hypothesis testing

Smoking moms

| [95% Conf. | Interval] |
|---|---|
| 2619.432 | 2925.162 |

Non-smoking moms

| [95% Conf. | Interval] |
|---|---|
| 2915.965 | 3193.948 |

Find the statistical testing to compare two independent mean

Independent or two-sample t-test

# Hypothesis testing with t-test

```
. ttest birthwt,by(smoke)

Two-sample t test with equal variances
```

| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 0 | 115 | 3054.957 | 70.1625 | 752.409 | 2915.965 | 3193.948 |
| 1 | 74 | 2772.297 | 76.70106 | 659.8075 | 2619.432 | 2925.162 |
| combined | 189 | 2944.286 | 53.02811 | 729.016 | 2839.679 | 3048.892 |
| diff | | 282.6592 | 106.9544 | | 71.66693 | 493.6515 |

```
    diff = mean(0) - mean(1)                                  t =   2.6428
Ho: diff = 0                     degrees of freedom =      187

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9955         Pr(|T| > |t|) = 0.0089          Pr(T > t) = 0.0045
```

# Hypothesis testing with t-test

```
. ttest birthwt,by(smoke)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     115    3054.957    70.1625     752.409    2915.965    3193.948
       1 |      74    2772.297   76.70106     659.8075   2619.432    2925.162
---------+--------------------------------------------------------------------
combined |     189    2944.286   53.02811     729.016    2839.679    3048.892
---------+--------------------------------------------------------------------
    diff |            282.6592   106.9544                71.66693    493.6515
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   2.6428
Ho: diff = 0                     degrees of freedom =              187

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9955         Pr(|T| > |t|) = 0.0089          Pr(T > t) = 0.0045
```

P-value

# What is the P-value?

- P stands for probability.
- Thus, P-value is the probability value.
- It expresses the weight of evidence in favor of against the stated null hypothesis.

- Precise definition: P value is the probability, given the null hypothesis is true, of obtaining the data as extreme or more extreme than that of the observed.

# What is the P-value?

- P stands for probability.
- Thus, P-value is the probability value.
- It expresses the weight of evidence in favor of against the stated null hypothesis.

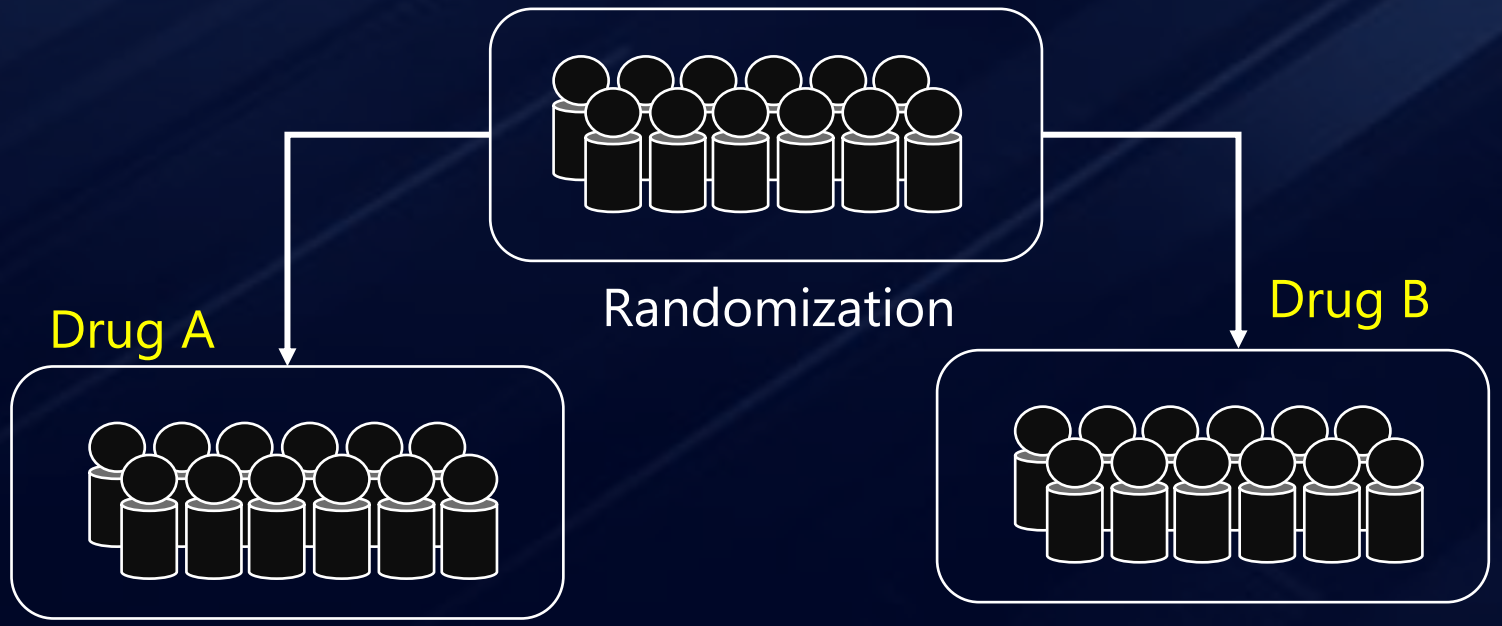
- Precise definition: P value is the probability, given the null hypothesis is true, of obtaining the data as extreme or more extreme than that of the observed.

โอกาสพบความแตกต่างบนสมมติฐานของความไม่แตกต่าง

# What is the P-value?

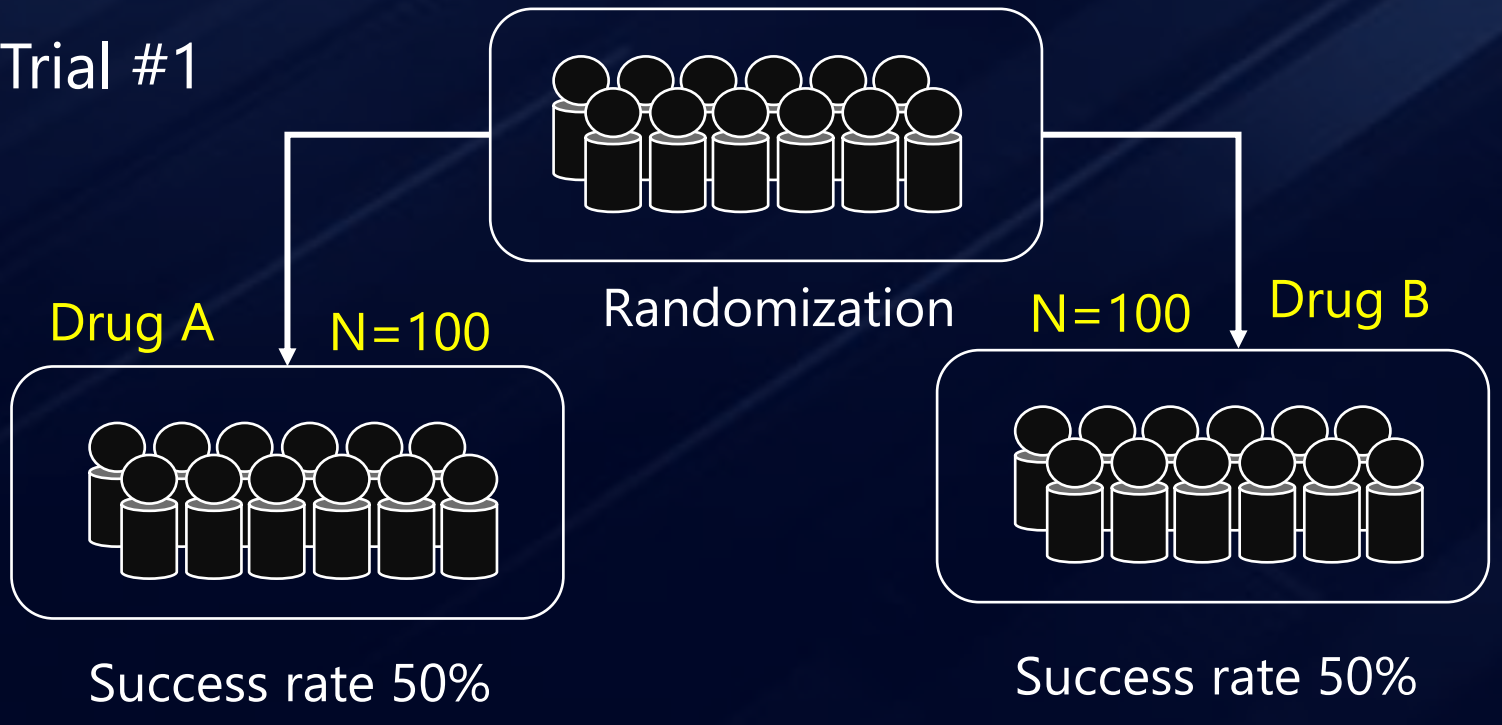
# What is the P-value?

- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B



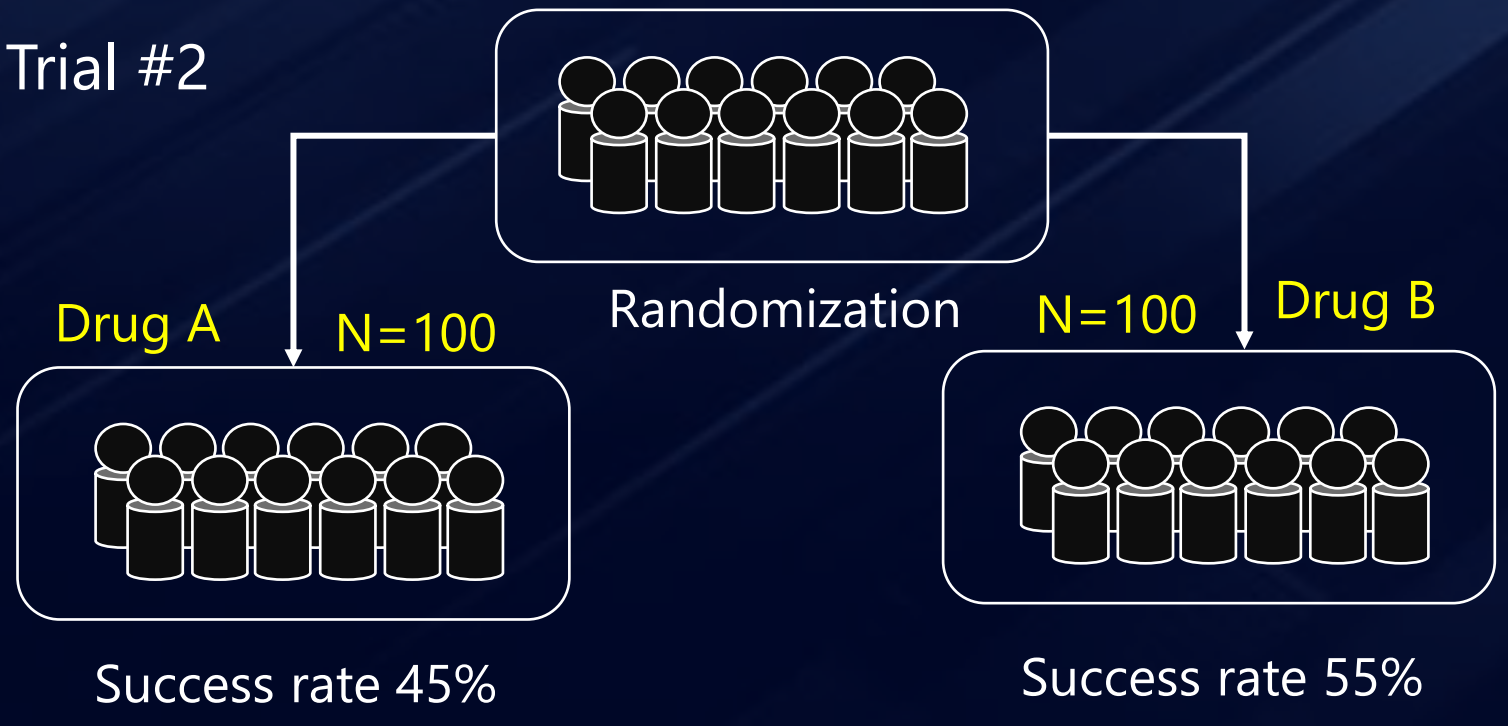
# What is the P-value?

- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B



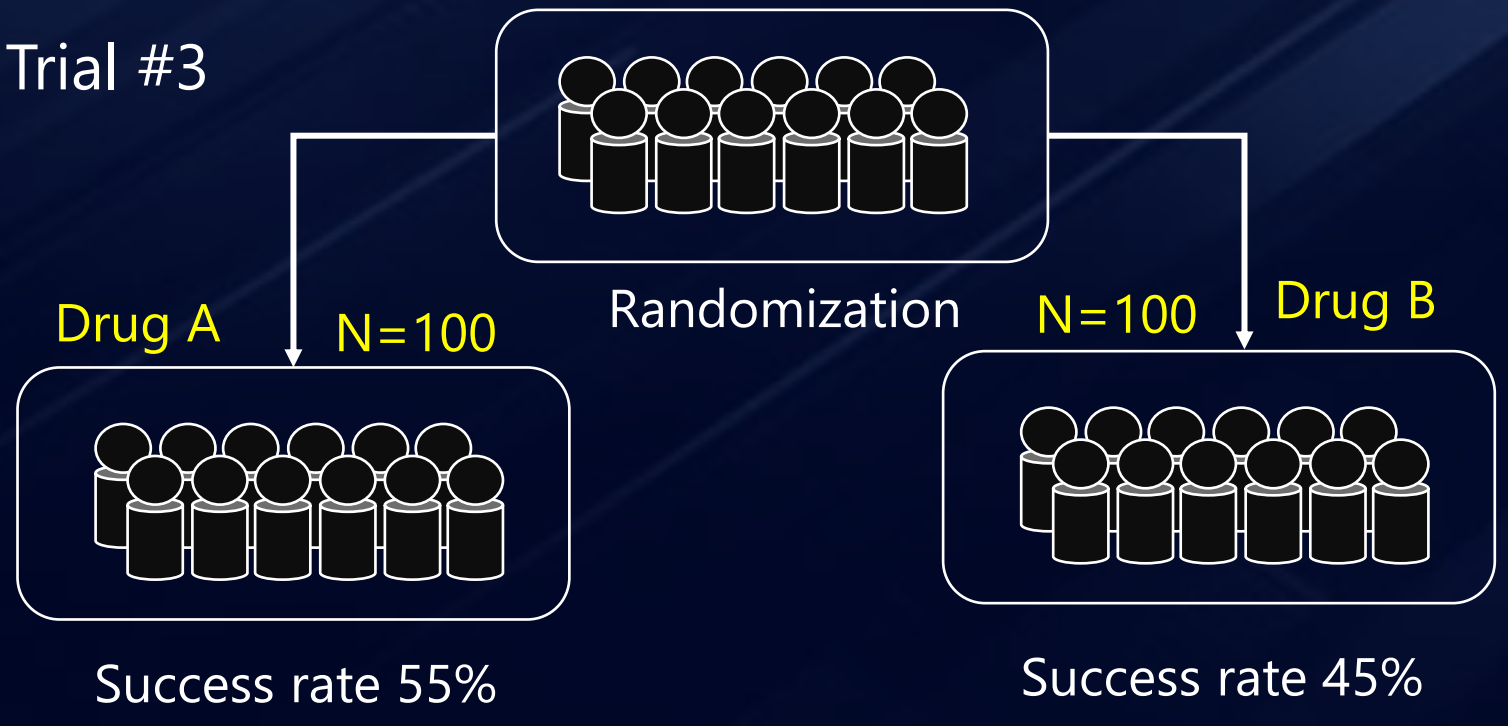
# What is the P-value?

- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B



# What is the P-value?

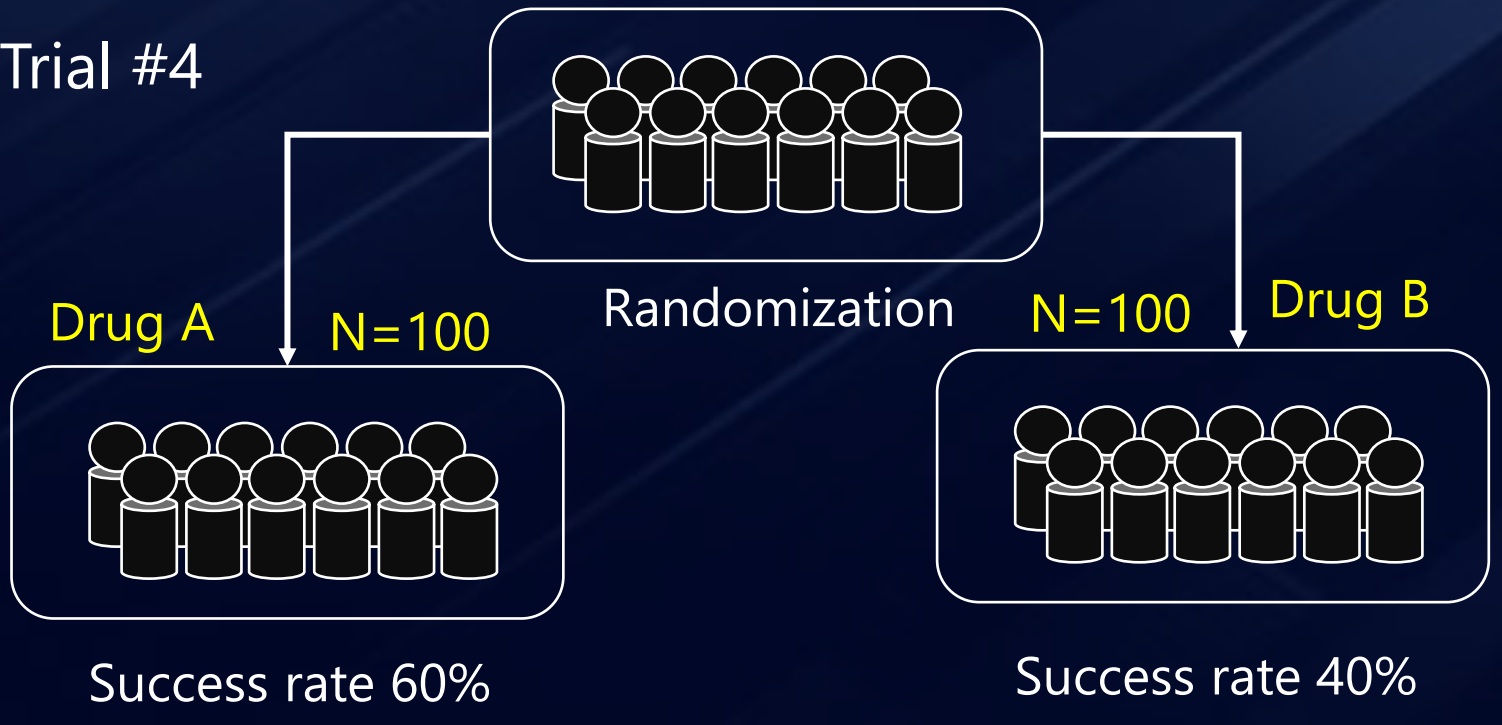
- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B




Diff=10, p-value=0.200

# What is the P-value?

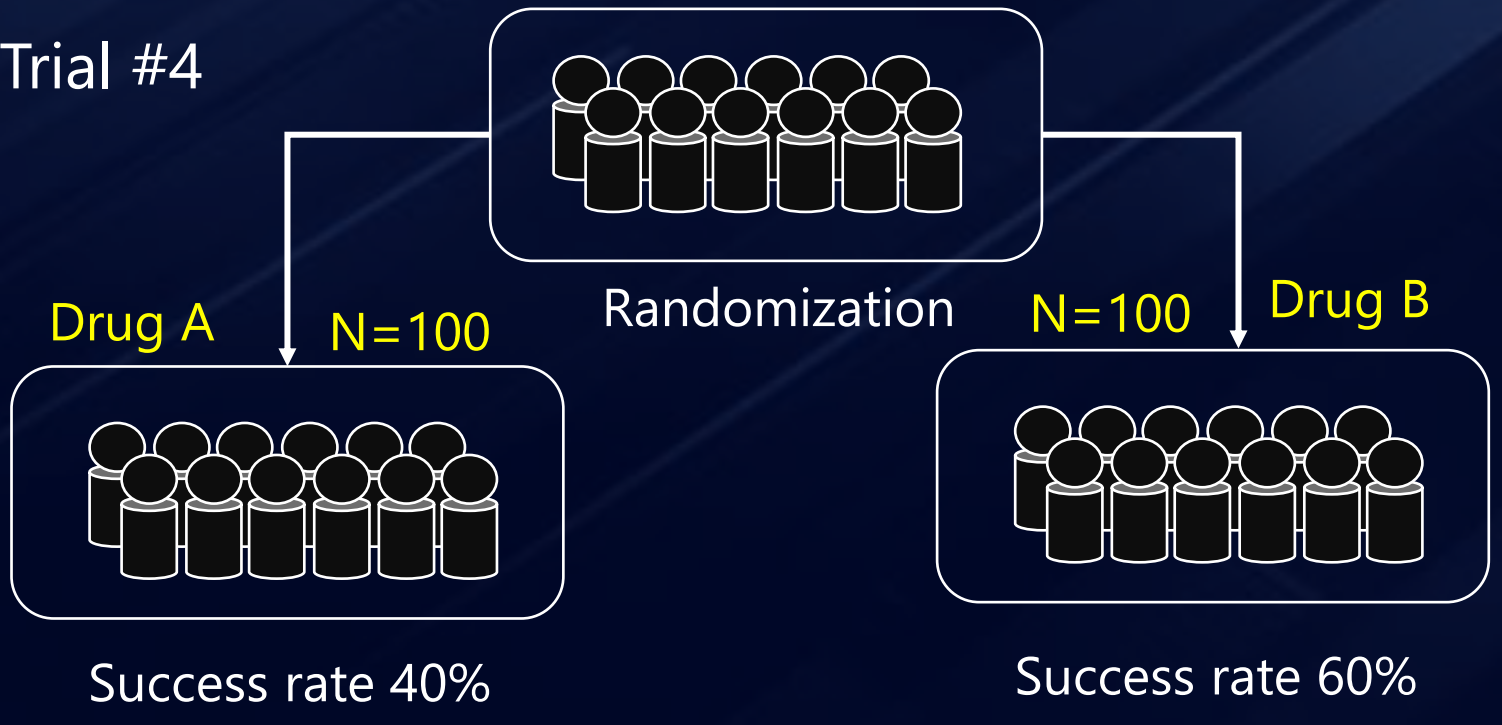
- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B




Diff=20, p-value=0.100

# What is the P-value?

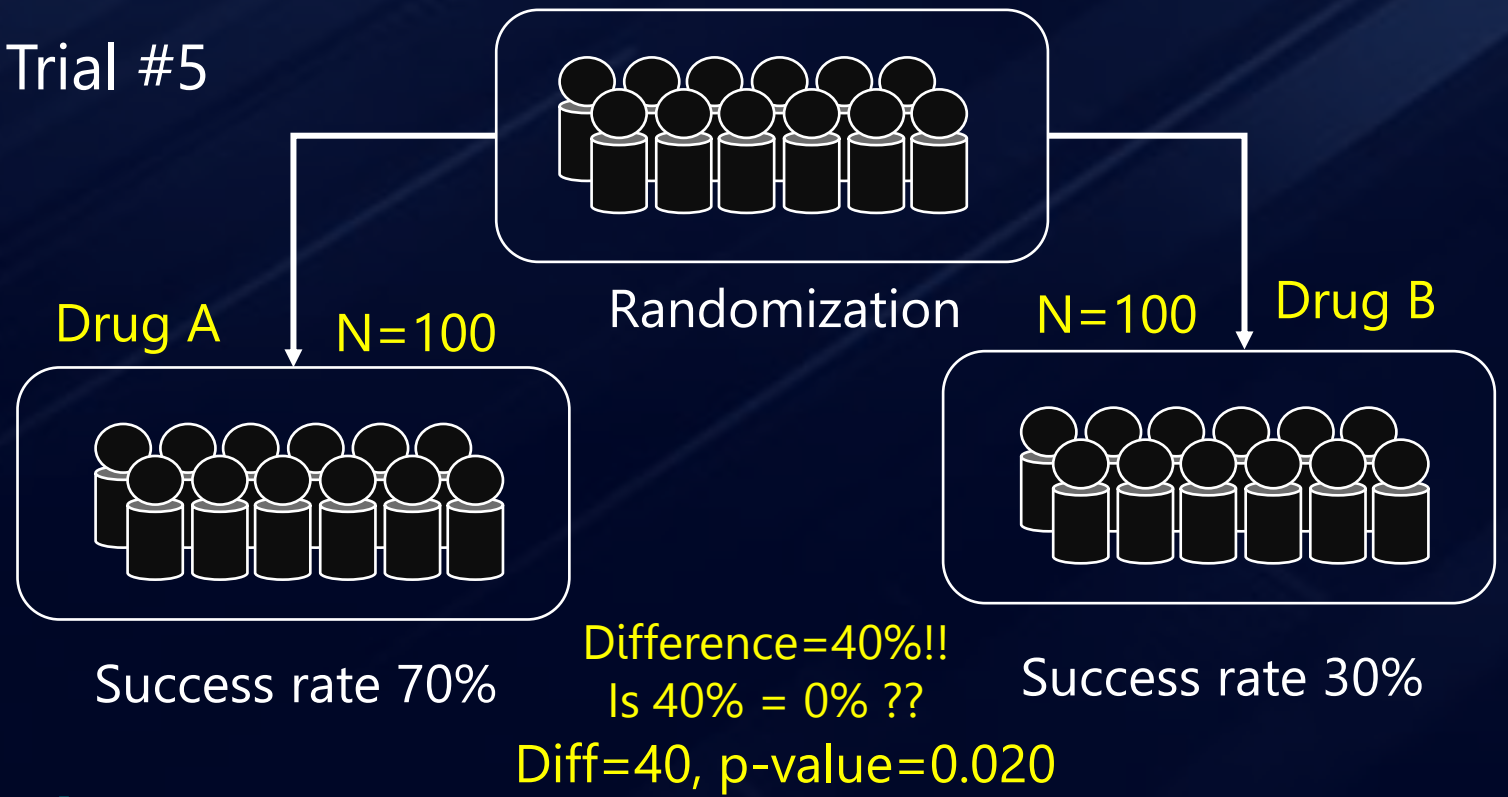
- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B




Diff=20, p-value=0.100

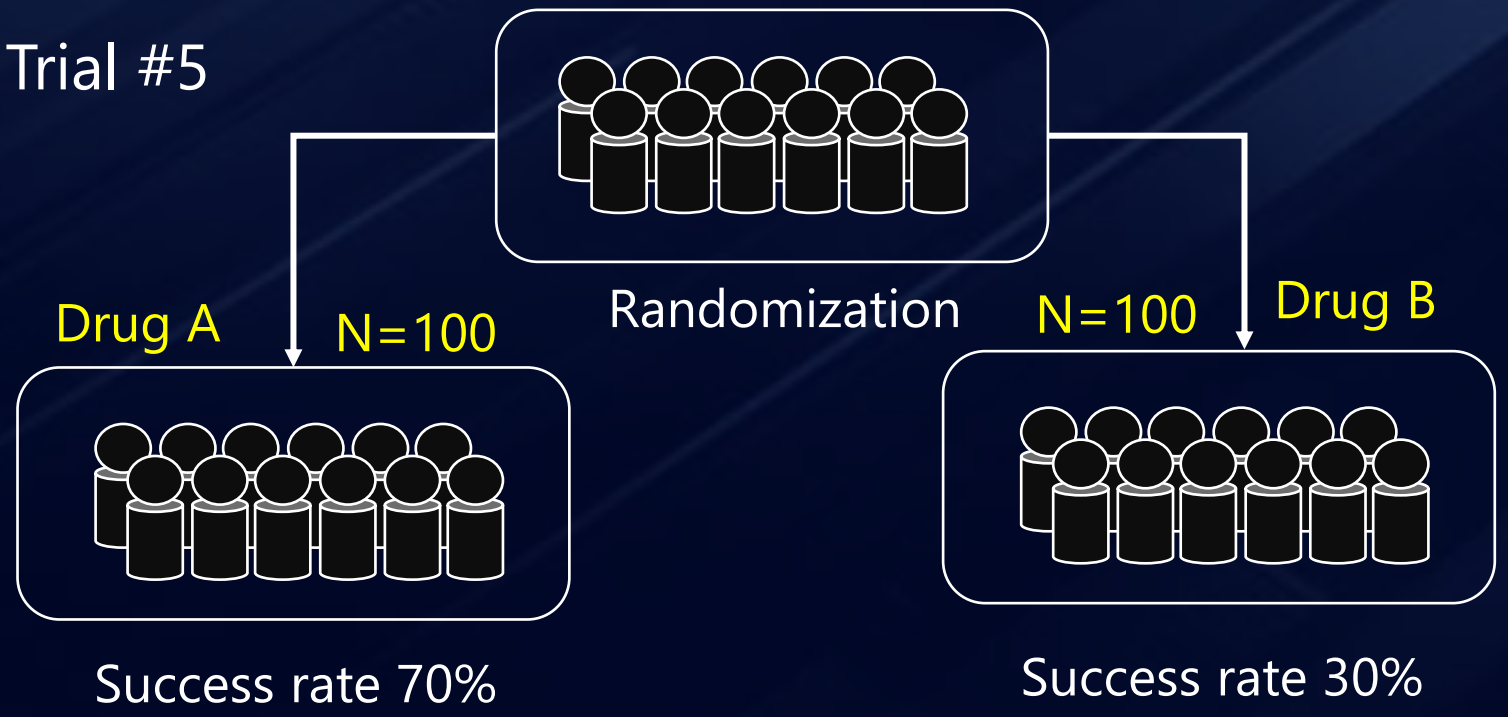
# What is the P-value?

- Suppose we want to test the efficacy of the 2 drugs
- Outcomes = success rate of something
- Given the null hypothesis is really TRUE.
- H0: success rate in drug A = success rate in drug B



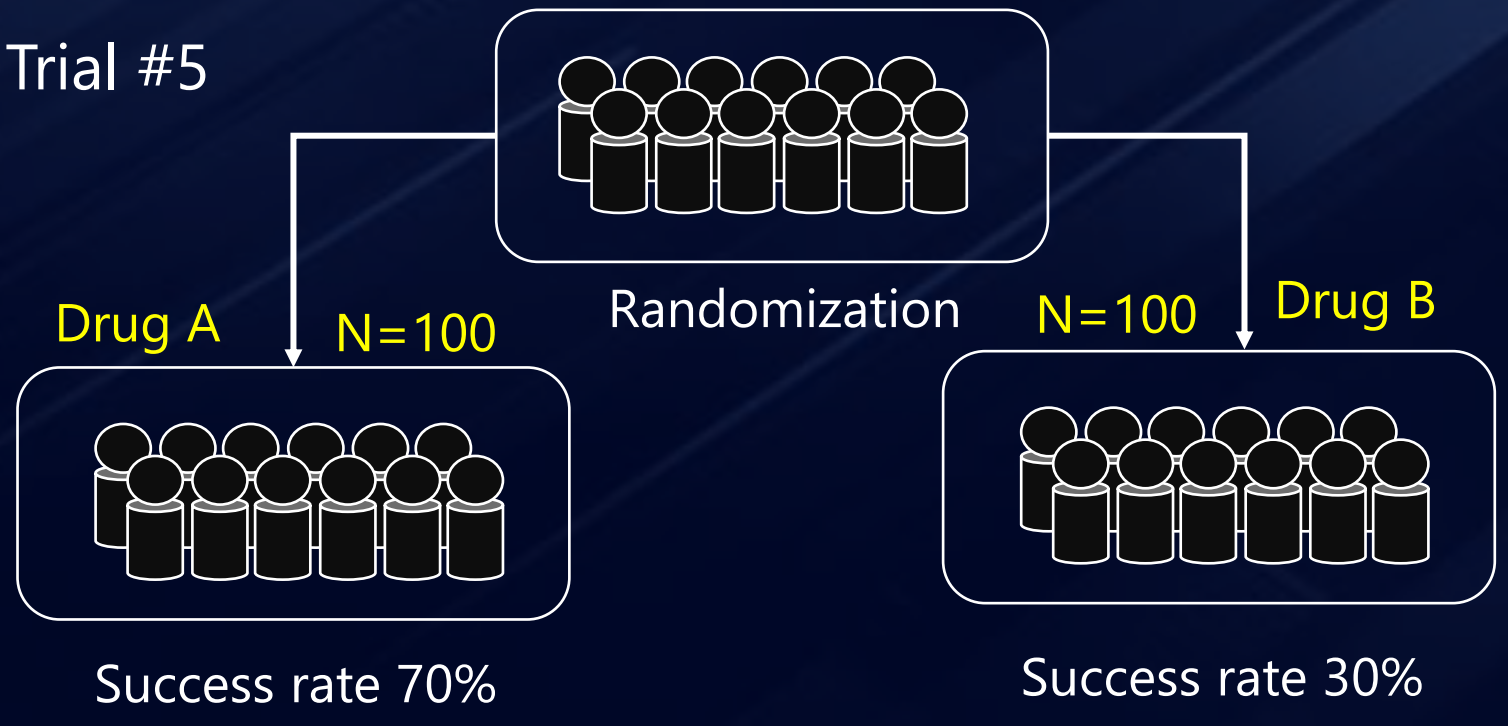
# What is the P-value?

- Observed difference = 40%
- What are the chances to get a result like this, when the null hypothesis is TRUE?
- P-value <0.05 means?



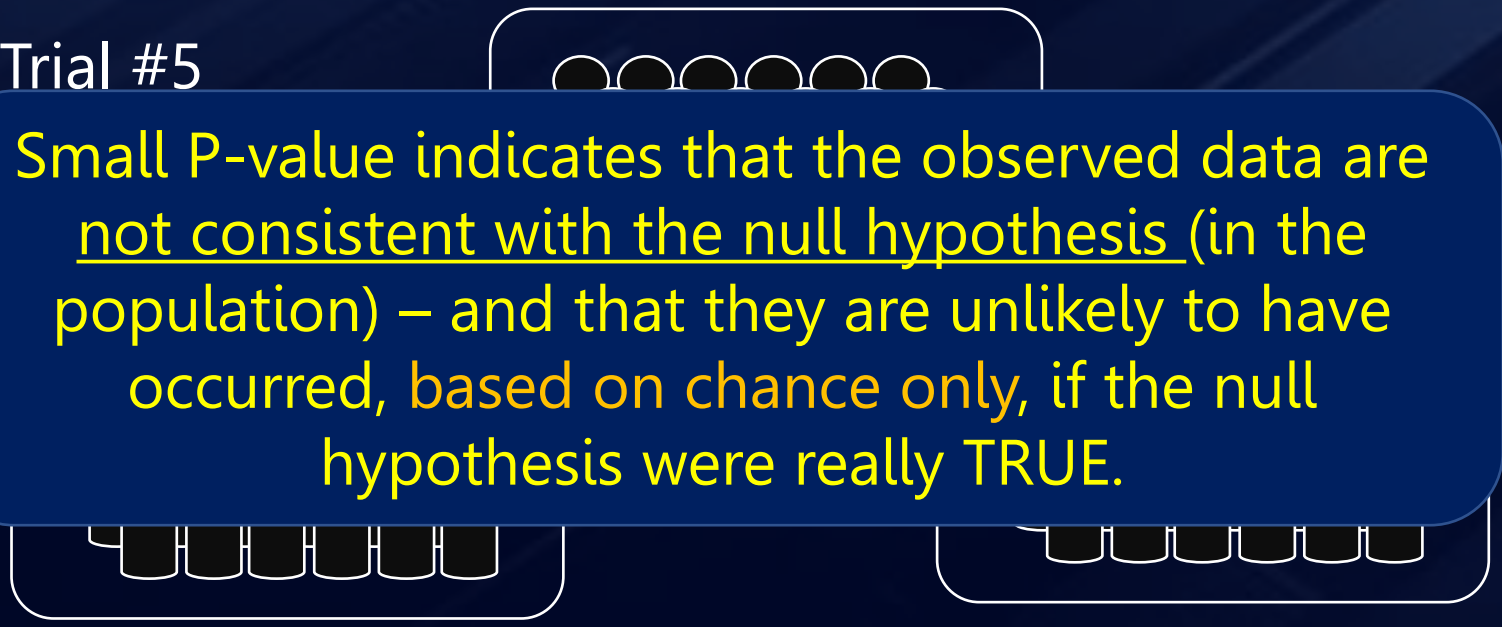
# What is the P-value?

- P-value <0.05 means there is less than 5% chance to identify the difference of 40% or more in any sample if the null hypothesis is TRUE.



# What is the P-value?

- P-value <0.05 means there is less than 5% chance that a difference this large or more could occur by chance (random error) alone.

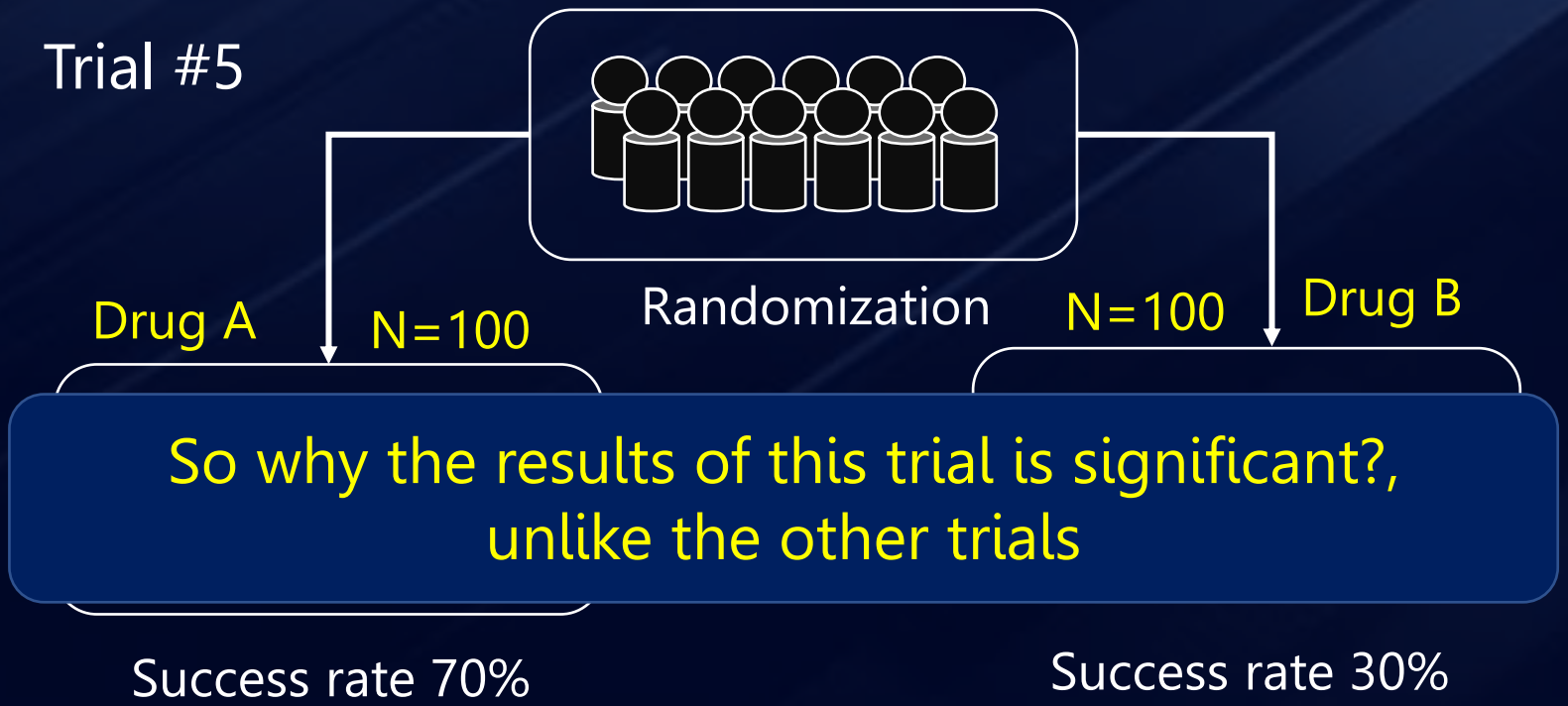



Small P-value indicates that the observed data are not consistent with the null hypothesis (in the population) – and that they are unlikely to have occurred, based on chance only, if the null hypothesis were really TRUE.

Diff=40, p-value=0.020

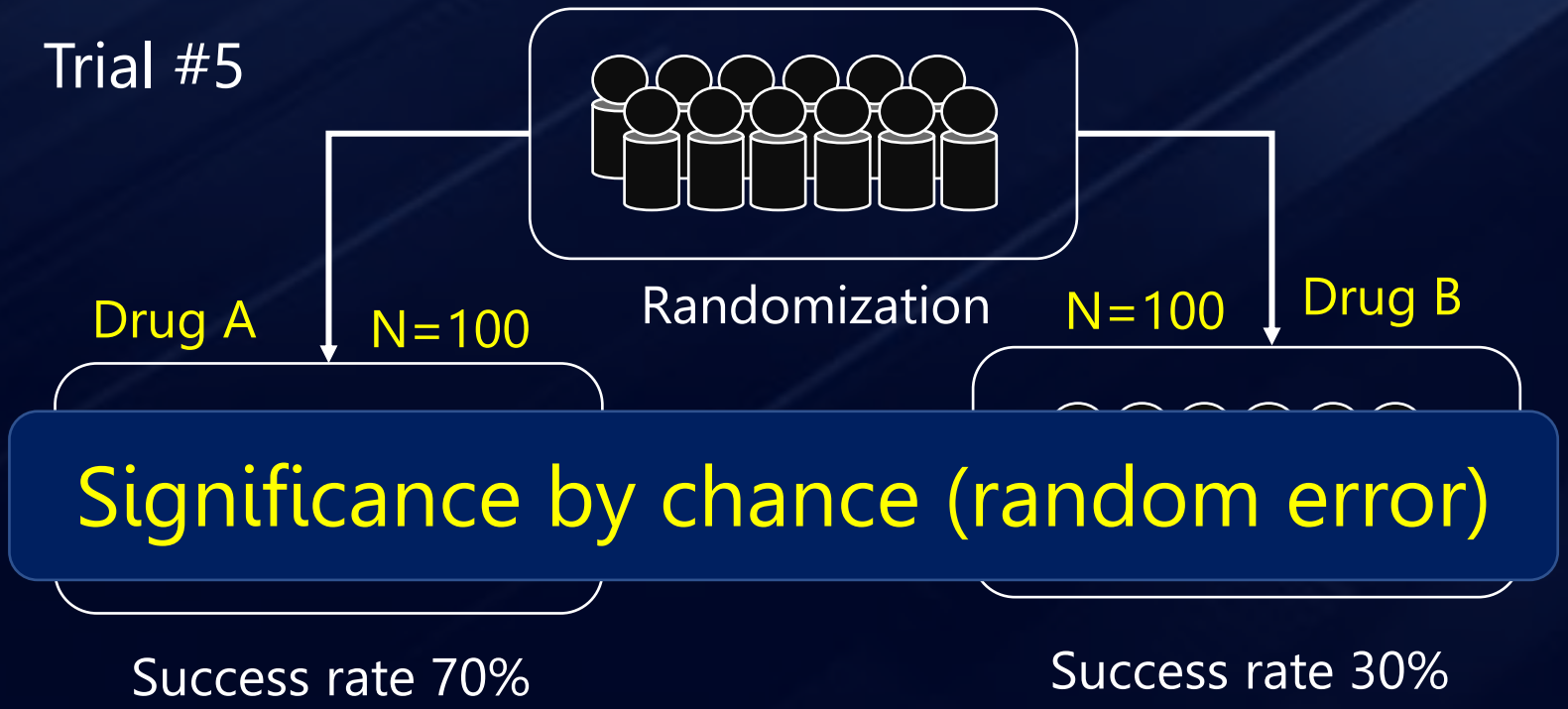
# What is the P-value?

- P-value <0.05 means there is less than 5% chance that a difference this large or more could occur by chance (random error) alone.



# What is the P-value?

- P-value <0.05 means there is less than 5% chance to identify the difference of 40% or more in any sample if the null hypothesis is TRUE.



# What is the P-value?

- 0.05 or 5% is commonly used as a cut-off.
- $P < 0.05$ is commonly described as statistically significant difference.
- $P \geq 0.05$ is commonly described as not statistically significant difference.

# What is the P-value?

Common misconception about the P-value

- A P-value is the probability that the null hypothesis is TRUE.
- A P-value quantifies the probability to falsely reject the null hypothesis.
- The smaller the P-value, the larger or the more importance is the clinical effect.

# What is the P-value?

Common misconception about the P-value

- The null hypothesis is already and always TRUE based on P-value definition!
- A P-value quantifies the statistical evidence that the results identified in the sample are different from the null hypothesis (in population) or not?
- The smaller the P-value, the more statistical evidence to reject the null hypothesis.

# Hypothesis testing with t-test

```
. ttest birthwt,by(smoke)

Two-sample t test with equal variances
```

| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 0 | 115 | 3054.957 | 70.1625 | 752.409 | 2915.965 | 3193.948 |
| 1 | 74 | 2772.297 | 76.70106 | 659.8075 | 2619.432 | 2925.162 |
| combined | 189 | 2944.286 | 53.02811 | 729.016 | 2839.679 | 3048.892 |
| diff | | 282.6592 | 106.9544 | | 71.66693 | 493.6515 |

```
    diff = mean(0) - mean(1)                      t =   2.6428
Ho: diff = 0                     degrees of freedom =      187

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9955         Pr(|T| > |t|) = 0.0089          Pr(T > t) = 0.0045
```

P-value ??

# Hypothesis testing with t-test

```
. ttest birthwt,by(smoke)

Two-sample t test with equal variances
```

| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 0 | 115 | 3054.957 | 70.1625 | 752.409 | 2915.965 | 3193.948 |
| 1 | 74 | 2772.297 | 76.70106 | 659.8075 | 2619.432 | 2925.162 |
| combined | 189 | 2944.286 | 53.02811 | 729.016 | 2839.679 | 3048.892 |
| diff | | 282.6592 | 106.9544 | | 71.66693 | 493.6515 |

```
    diff = mean(0) - mean(1)                      t =   2.6428
Ho: diff = 0                     degrees of freedom =      187

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9955         Pr(|T| > |t|) = 0.0089          Pr(T > t) = 0.0045
```

The probability is less than 1 in 100 that a difference this large could be due to chance (random error) alone (there must be some effect from smoking on birthweight for sure).

# Hypothesis testing

- Comparative questions
  - Compare different between groups (independent group) = อิสระต่อกัน
  - Compare different within groups (dependent group) = ไม่อิสระต่อกัน

# Hypothesis testing

- Comparative questions
  - เปรียบเทียบระดับน้ำตาลในเลือดที่ 3 เดือนระหว่างผู้ป่วยที่ได้รับการรักษาด้วยยาเบาหวานใหม่ กับผู้ป่วยที่ได้รับการรักษาด้วยยาเบาหวานแบบเดิม

  - Two independent groups (ใหม่ vs. เดิม)
  - Mean FBS in ใหม่ vs. Mean FBS in เดิม
  - Compare two independent means

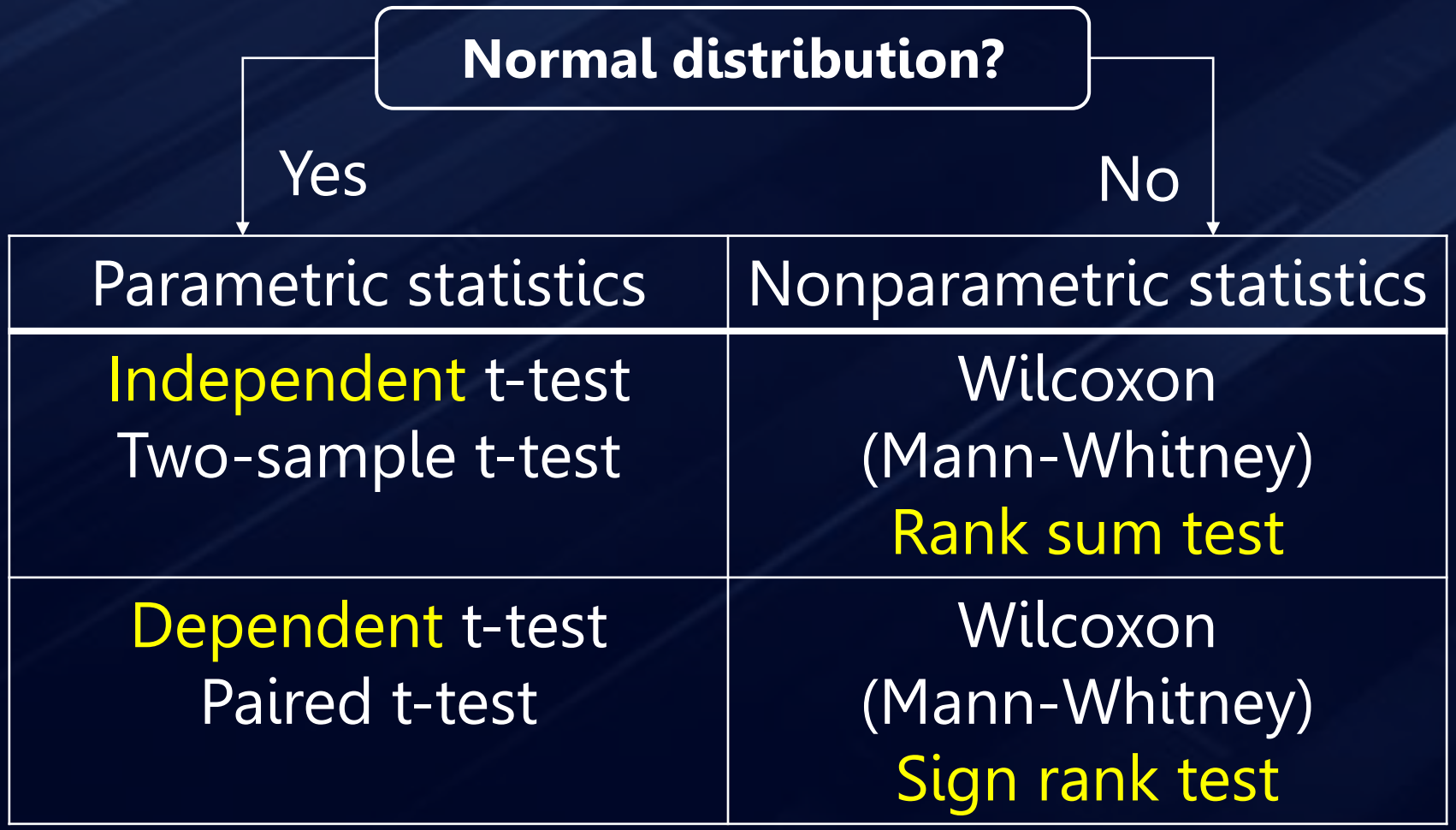
# Hypothesis testing

- Comparative questions
  - เปรียบเทียบระดับน้ำตาลในเลือดที่ 3 เดือนระหว่างผู้ป่วยที่ได้รับการรักษาด้วยยาเบาหวานใหม่ กับผู้ป่วยที่ได้รับการรักษาด้วยยาเบาหวานแบบเดิม
    Compare two independent means
  - Null hypothesis
    - Mean FBS in ใหม่ = Mean FBS in เก่า
  - Alternative hypothesis
    - Mean FBS in ใหม่ ≠ Mean FBS in เก่า

# Hypothesis testing

- Continuous variables
  - 2 Types of statistics
    - Parametric statistics
      - Based on normal distribution
      - Based on parameters
    - Non-parametric statistics
      - Not based on distribution
      - Based on rank and order

# Two independent: numerical




| Parametric statistics | Nonparametric statistics |
|---|---|
| Independent t-test<br>Two-sample t-test | Wilcoxon<br>(Mann-Whitney)<br>Rank sum test |
| Dependent t-test<br>Paired t-test | Wilcoxon<br>(Mann-Whitney)<br>Sign rank test |

# t-test

- Independent t-test
  - Compares two independent means
  - Based on normal distribution
  - Null hypothesis:
    Mean 1 = Mean 2
    (Two sample come from population with the same mean)
- Assumption
  - Continuous data, normally distributed
  - Variances are the same

# Ranksum test

- Rationale of test
  - It is the analogue of t-test
  - It is based on the ranks of the data
  - Suitable for larger samples
  - Does not test the medians, it tests the whole distribution
  - Assumption
    - The data are in two groups and can be ranked

# Ranksum test

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | **Wilcoxon Rank-Sum Test** | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | Original data | | | Ranks | | | | | |
| 4 | | | | | | | | Control | Drug |
| 5 | Control | Drug | | Control | Drug | | count | 12 | 12 |
| 6 | 11 | 34 | | 4 | 22.5 | | rank sum | 119.5 | 180.5 |
| 7 | 15 | 31 | | 10 | 20.5 | | | | |
| 8 | 9 | 35 | | 2 | 24 | | α | 0.05 | |
| 9 | 4 | 29 | | 1 | 17.5 | | tails | 2 | |
| 10 | 34 | 28 | | 22.5 | 16 | | W | 119.5 | |
| 11 | 17 | 12 | | 11 | 5.5 | | W-crit | 115 | |
| 12 | 18 | 18 | | 12.5 | 12.5 | | sig | no | |
| 13 | 14 | 30 | | 8.5 | 19 | | | | |
| 14 | 12 | 14 | | 5.5 | 8.5 | | | | |
| 15 | 13 | 22 | | 7 | 14 | | | | |
| 16 | 26 | 10 | | 15 | 3 | | | | |
| 17 | 31 | 29 | | 20.5 | 17.5 | | | | |
| 18 | 17 | 24.33333 | | 119.5 | 180.5 | | | | |

# Hypothesis testing

| | Group A (n=XX) | Group B (n=YY) | P-value | Statistical Test used |
|---|---|---|---|---|
| Age | 50±10 | 40±10 | 0.030 | |
| Parity | 1±1 | 3±2 | 0.045 | |
| Gestational age | 32±3 | 36±3 | 0.005 | |
| Hemoglobin | 10.2±1.6 | 12.5±1.8 | 0.009 | |
| Creatinine | 2±1.5 | 0.7±1.1 | 0.067 | |
| Blood loss | 352±250 | 658±387 | 0.042 | |
| Length of stay | 5±10 | 2±5 | 0.033 | |

Fictitious data

# Hypothesis testing

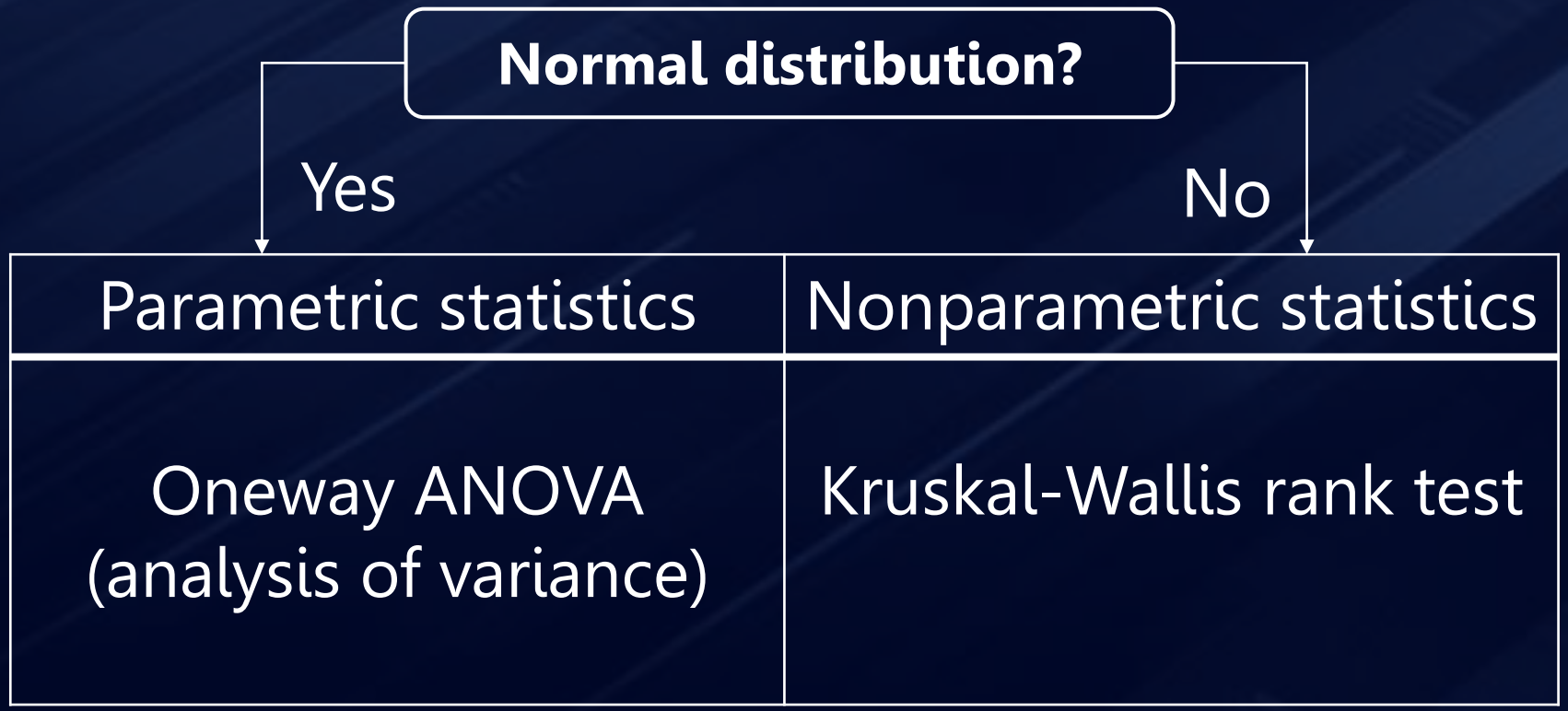
| | Group A (n=XX) | Group B (n=YY) | P-value | Statistical Test used |
|---|---|---|---|---|
| Age | 50±10 | 40±10 | 0.030 | Independent t-test |
| Parity | 1±1 | 3±2 | 0.045 | Rank sum test |
| Gestational age | 32±3 | 36±3 | 0.005 | Independent t-test |
| Hemoglobin | 10.2±1.6 | 12.5±1.8 | 0.009 | Independent t-test |
| Creatinine | 2±1.5 | 0.7±1.1 | 0.067 | Rank sum test |
| Blood loss | 352±250 | 658±387 | 0.042 | Rank sum test |
| Length of stay | 5±10 | 2±5 | 0.033 | Rank sum test |

Fictitious data

# Hypothesis testing

- Continuous variables
  - Comparing >2 groups
    - Multiple pairwise t-test
    - Oneway ANOVA
    - Kruskal-Wallis rank test

# >2 independent: numerical



# Hypothesis testing

- Categorical variables
  - Comparing ≥2 groups
    - Chi-squared test
    - Fisher's exact probability test

# Chi-squared test

- Rationale of test
  - Calculated the expected frequencies if there were no association.
  - Null Hypothesis: No association
    Expected freq = Observed freq
  - If observed frequencies are very different to the expected values, there is high possibility of association.

# Hypothesis testing

- Hypothesis-driven question?

  Does obese people carry higher risk of coronary heart disease than non-obese people?

  Obese, category [yes, no] = [0,1]
  CHD, category [yes, no] = [0,1]

# Hypothesis testing

- Comparative questions
  - เปรียบเทียบสัดส่วนของการเกิดโรคหลอดเลือดหัวใจในคนที่ตุ้ยนุ้ยกับคนที่ไม่ตุ้ยนุ้ยว่าแตกต่างกันหรือไม่?

  - Two independent groups (ตุ้ยนุ้ย vs. ไม่ตุ้ย)
  - %CHD in obese vs. %CHD in non-obese
  - Compare two independent proportions
    - %CHD (event of CHD/total at risk)

# Hypothesis testing

- Example
  - Null hypothesis
    Proportion CHD in obese =
    Proportion CHD in non-obese

  - Alternative hypothesis
    Proportion CHD in obese ≠
    Proportion CHD in non-obese

# Hypothesis testing

| | CHD+ve | CHD-ve | Total |
|---|---|---|---|
| Obese | 50 (0.5) | 50 (0.5) | 100 |
| Non-obese | 50 (0.5) | 50 (0.5) | 100 |
| Total | 100 | 100 | 200 |

# Hypothesis testing

| | CHD+ve | CHD-ve | Total |
|---|---|---|---|
| Obese | 70 (0.7) | 30 (0.3) | 100 |
| Non-obese | 30 (0.3) | 70 (0.7) | 100 |
| Total | 100 | 100 | 200 |

# Hypothesis testing

```
. tabi 70 30 \ 30 70, chi2
```

| row | col 1 | 2 | Total |
|---|---|---|---|
| 1 | 70 | 30 | 100 |
| 2 | 30 | 70 | 100 |
| Total | 100 | 100 | 200 |

Pearson chi2(1) = 32.0000 Pr = 0.000

# Fisher's exact test

- Exact probability test
- Complex calculation
- It is useful for small samples
- Used for only 2x2 tables
- Based on exact probabilities
- Tests the same null hypothesis as Chi2
- Assumption of test: None

# Hypothesis testing

```
. tabi 70 30 \ 30 70, chi2 exact
```

| row | col 1 | 2 | Total |
|---|---|---|---|
| 1 | 70 | 30 | 100 |
| 2 | 30 | 70 | 100 |
| Total | 100 | 100 | 200 |

```
          Pearson chi2(1) =  32.0000   Pr = 0.000
           Fisher's exact =                 0.000
   1-sided Fisher's exact =                 0.000
```

# Take Home Knowledge

- What's inferential statistics?
- What's the difference between SD, SE, and Confidence interval?
- What's hypothesis testing?
- What's null and alternative hypothesis?

# Take Home Knowledge

- The importance of study size estimation?
- Factors that influence study size?

# Take Home Knowledge

- Compare two normally distributed numerical data?
- Compare two non-normally distributed numerical data?
- Compare two or more independent categorical data?