Optional Module
Overview of Statistical Analysis

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Overview of Statistical Analysis

This section will include:

- **What is statistics** and **how it is used**
- **Mean, media, mode** definitions
- **Statistical** vs. **Clinical** Significance
- ***P* values** and **statistical tests**
- **Correlation** vs. **Causation**

Estimated Time Commitment: 30-40 minutes

© COPD Foundation 2021

COPD FOUNDATION

Instructions for Going through the Optional Module

You do not need to go through the module in one sitting. Feel free to take breaks and complete at your own pace!

We have broken the module up into 2 sections with suggested breaks:

Part 1 – What statistics is, how it is used and definitions of mean, median and mode (Slides 5 to 17)

Part 2 – Statistical vs. Clinical Significance, *P* values, statistical tests and Correlation vs. Causation (Slides 19 to 35)

© COPD Foundation 2021

COPD FOUNDATION

Key Terms Used in this Module

- You will hear a lot of **new terms** during this module that you might not know

- We **do not expect you to know what these terms mean right away** or memorize them – just to familiarize yourself with the words

- Here is a **link to the key terms** used in this module that you can reference as you go through – click here

© COPD Foundation 2021

COPD FOUNDATION

Part 1
What statistics is, how it is used and definitions of Mean, Median and Mode (Slides 5 to 17)

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Overview of Statistical Analysis

The purpose of this module is to:
provide a brief overview of some statistical terms and information that you will hear while being a part of a study team.

As a patient/caregiver investigator, **you will NOT be responsible for doing the statistical analysis** for a research study.

Some or all of the concepts included in this **module will likely be included in your research team's analysis of the study's results.**

© COPD Foundation 2021

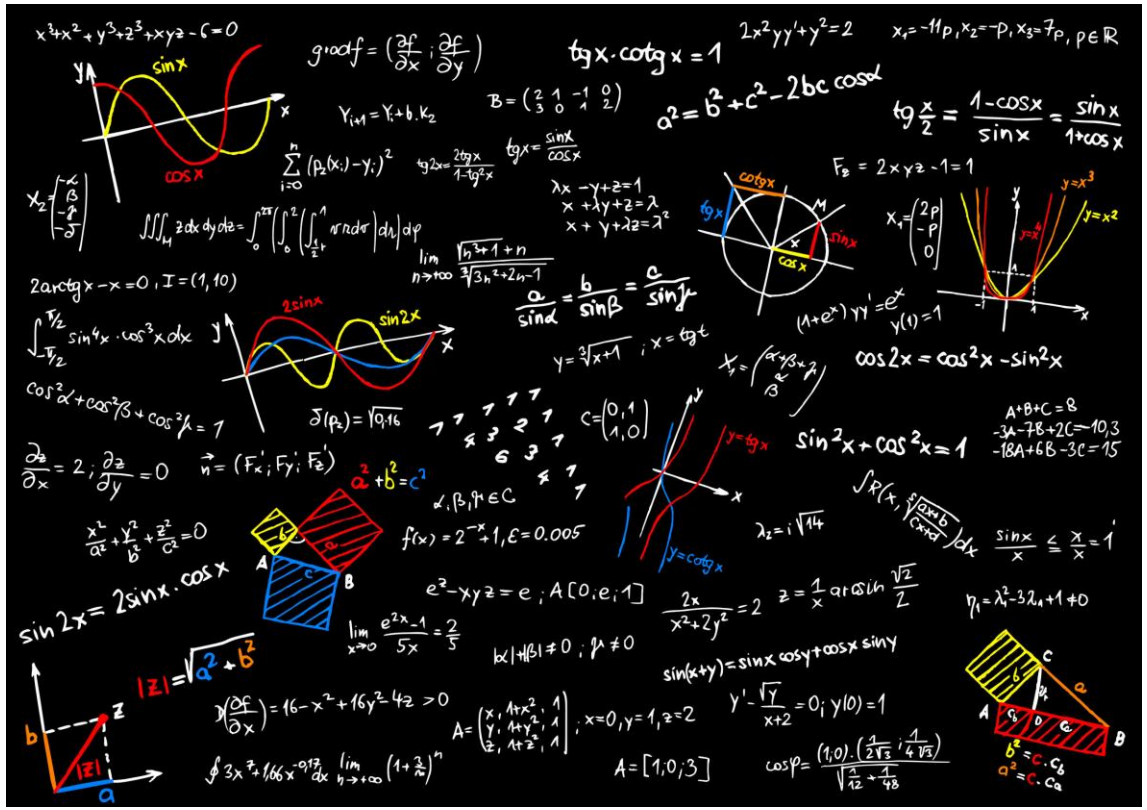
COPD FOUNDATION

Statistics is...

Statistics is a branch of math focused on

- the collection
- analysis
- interpretation and
- presentation....

....of ***numerical data (numbers)***



© COPD Foundation 2021

COPD FOUNDATION

Statistics in a Research Study

In a research study, statistics....

- is used to ***summarize the data***, to make it quicker and easier to understand
- allows data to be summarized ***the same ways across all studies***
- is used to ***present quantitative data***
- ***helps the reader draw conclusions***

© COPD Foundation 2021

COPD FOUNDATION

Statistical Analysis = Quantitative Data

In a research study:

- survey answers,
- the number, race, age of participants,
- patients' physical traits (weight, and height),
- patients' performance on clinical tests (like lung function or exercise endurance test) and
- even an individual's perceptions of his/her quality of life...

... **can ALL be *measured*** and **represented with *numbers-***

This is ***quantitative data.***

© COPD Foundation 2021

COPD FOUNDATION

For Example....

Examples of statistics seen in all research studies: the reporting of the "***demographics" or statistical descriptions*** (age, race, sex, income, education, etc.) of a study's participants...

All studies use percentages or proportions to describe the demographics of individuals enrolled in the study:
For example:

- **"50% or half of the participants were women"**
- **"66% or two-thirds were white"**
- **"75% or three-fourths were over 65"**

© COPD Foundation 2021

COPD FOUNDATION

Mean, Median and Mode

When reviewing and evaluating results from a study—it is often helpful to take all the results/measurements—the full set of numbers (quantitative data)—and ***find the center***.

Finding the **average** or **center of the data answers the questions**:

- *What was the average answer? (mean)*
- *What is a typical value? (mode)*
- *What is middle of the full range of the data collected? (median)*

The 3 most common ways to measure the center are:

Mean, Median and Mode

© COPD Foundation 2021

COPD FOUNDATION

Mean

The **mean** is also called the **average**

To find the mean:

- Add up all the numbers
- Then, divide that total by the number (quantity) of numbers or answers in the set.

See the example in the graphic...

MEAN

Example: Data gathered is number of years individuals have been diagnosed with COPD.

Data: 5 years, 4 years, 12 years

5+4+12=**21**

21 ÷ 3

(there are 3 answers in the set)

= **7**

7= the MEAN

© COPD Foundation 2021

COPD FOUNDATION

Median

MEDIAN

Example: Data gathered is number of years individuals have been diagnosed with COPD.

Data: 5, 4 and 12 years

4, 5, 12

Median=5

Example 2: (even number of data)

Data: 5, 4 ,4 and 12 years

4,4,5,12

4+5=9

9÷2= **4.5**

Median=4.5

The median is the place that divides the data in half, once the data is ordered from smallest to largest.

To find the median:

- **Put the numbers in order**—smallest to largest
- If the set of numbers has an **odd number of numbers,** then the **one in the exact middle is the median—*see top example in the graphic***
- If the set of numbers has an **even number of numbers,** take the **two numbers in the middle, add them together and divide by 2 to get the median—*see example 2 in the graphic***

© COPD Foundation 2021

COPD FOUNDATION

Mode

In a set of numbers, the mode is the number/value that appears most often.

To find the mode:

- **Place numbers in order, from smallest to largest**
- **Find the number that is listed the most often, this is the mode**

MODE

Example: Data gathered is number of years individuals have been diagnosed with COPD.

Data: 5, 4, 4 and 12 years

4, 4, 5, 12

There are more 4's than any other number

Mode= 4

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 1)

- The next few slides have some questions to help you review and remember what we have presented in this Module.

- This is not a graded test and is meant to only help you retain the information from this Module. There is an answer key at the end.

- Here is a link to the key terms that might help as you go through the review questions: click here

- If you have any questions, please email BRIDGE@copdfoundation.org

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 1)

1. **Which answer below describes statistical analysis?**
   a. Is a branch of math
   b. Is focused on the collection, analysis and presentation of numerical data
   c. Is used in research studies to summarize quantitative data
   d. Allows data to be summarized the same across all research studies
   e. All of the above

Match the term to its definition:

2. Mean	**a. Divides the data in half. After numbers are put in order smallest to largest, it is the number exactly in the middle (ex. "5" of 1, 3, 5, 6, 7)**
3. Median	**b. The number value that appears most often in a set of numbers. (ex. "3" of 1, 3,3, 5, 6, 7)**
4. Mode	**c. The average. Sum of all numbers divided by the quantity of numbers in the set (ex. 8+2+5=15÷3=5)**

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 1) Answer Key

1. e
*Explanation: A*ll of the statements are true about statistics.

2. c
3. a
4.b
*Explanation: M*ean is the average, median divides the numbers in half and mode is the value that appears most often
See slides 11-14 for a review of mean, median and mode.

© COPD Foundation 2021

COPD FOUNDATION

You completed Part 1 - Great job!
Feel free to take a break before going on to Part 2

"This Photo" by Unknown Author is licensed under CC BY-SA-NC

© COPD Foundation 2021

COPD FOUNDATION

Part 2
Statistical Significance, *P* values, Statistical Tests, Clinical Significance and Correlation vs. Causation (Slides 19 to 35)

© COPD Foundation 2021

COPD FOUNDATION

Statistical Significance

***Statistical* significance means:**

the results of the study are *not* **likely to have occurred by chance** but are instead **likely to be attributable to a specific cause or association.**

In other words, statistical significance is:

the **measurement of how confident we are that a difference or relationship exists between two variables** *(such as the treatment group versus the placebo group)* and is not just a result of chance.

Statistically significant data is:

- **Reliable—**it likely reflects what actually happened
- **But not necessarily relevant, important or decision-making worthy—**
  it may not reflect an important difference in what happens to patients or their disease or condition.

© COPD Foundation 2021

COPD FOUNDATION

P value

***P* values** are used to **describe how significant a finding may be**

P value **measures how likely the results could have just occurred from chance**—so the *lower a p value, the more "real" the results, i.e., not a result of "coincidence."*

The *results* of a study are ***statistically significant* when *p value is ≤ 0.05***

This **means there is a *5 percent or less likelihood that the results are due to chance***. Viewed the other way- this means there is a *95% likelihood that the observed differences are real.*

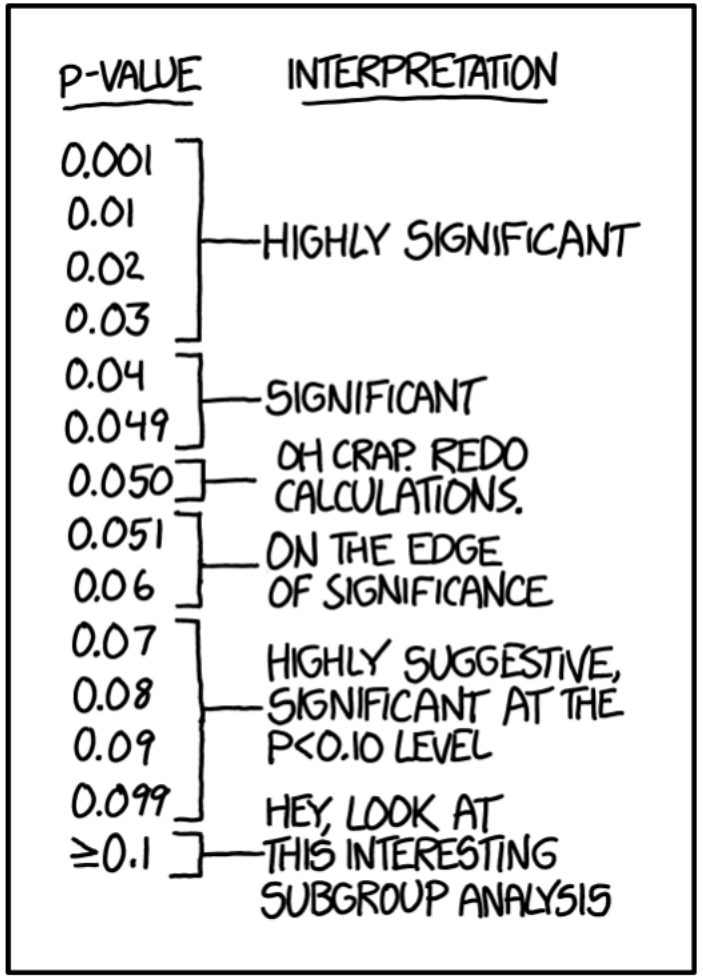


Image credit: https://xkcd.com/1478

© COPD Foundation 2021

COPD FOUNDATION

Statistical Tests

Several statistical tests are used to ***compare two or more groups of data/information (and p values are typically used to determine whether the differences observed are significant)***

Don't worry, you do not need to completely understand these tests. We listed them here because these are terms that you will see often in study statistical analysis

- **Chi-Square Test**—these tests tell you how **significant the differences between groups are when using proportions.** In other words, it lets you know if those differences comparing proportions could have happened by chance (is a group with a 30% response different from a group with a 60% response).

© COPD Foundation 2021

COPD FOUNDATION

Statistical Tests continued...

Don't worry, you do not need to completely understand these tests. We listed them here because these are terms that you will see often in study statistical analysis

- **T-Tests**—these tests tell you how **significant the differences between groups are when using means or averages.** In other words, it lets you know if those differences could have happened by chance. For example, is an FEV_1 of 50% different from an FEV_1 of 60%.

- **Analysis of Variance** (ANOVA)—used to determine whether there are any **statistically significant differences** between the **means of three or more** independent (unrelated) groups.

- **Regression Analysis**—is used to compare two or more groups for an outcome (such as hospitalization or lung function) that also controls for other factors (such as age, sex, smoking history, etc.)

© COPD Foundation 2021

COPD FOUNDATION

Clinical Significance

<u>*Clinical* significance is different from statistical significance</u>

If a study has *clinical significance*, its **results or findings will be of practical use to health care providers and patients.**

<u>Clinical significance answers the question:</u>

"Is the 'intervention' studied *(i.e., new medicine, surgical procedure, use of action plan, etc.)* **effective enough—will make enough of a difference to patients' lives or their condition**—to consider implementing it in the future?

© COPD Foundation 2021

COPD FOUNDATION

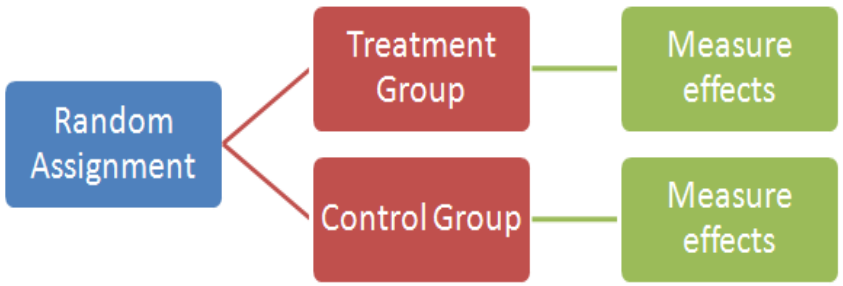
Statistical Significance vs. Clinical Significance

Statistical Significance	Clinical Significance
Signifies that results are ***unlikely to be from chance or coincidence***	Signifies that results can have ***an impact on health or health care***
Measures the ***reliability*** of the results	Measures the ***importance*** of the results
"There was a **difference between the two groups studied that did NOT occur from random chance"**	"What we learned from the **difference in results in the two groups studied is important and has practical applications**—we should consider using this in the future"

© COPD Foundation 2021

COPD FOUNDATION

Statistical Significance vs. Clinical Significance: An Example



Let's imagine a study in which...

Experimental group: **400 individuals receive a new medicine for treating their COPD**

Control group: **400 other individuals continue to use the medicine they always have used.**

Study Outcomes: Lung function tests results—how well the individuals move air in and out of their lungs—of the 400 individuals getting the new medicine *(the experimental group)* will be compared to the lung function test results of 400 individuals taking the medicines they usually take *(the control group).* **The lung function test results are the study's outcomes.**

Results: After 2 years, the results show that the individuals taking the new medicine can breathe out, on average, **an additional 30 ccs of air in the first second.**

Statistical Significance**:** After using the appropriate statistical tests we find **there is a *p* value of 0.001 which is less than 0.05 and therefore, "statistically significant."**

Clinical Significance: Most doctors would say that **this small amount of improvement cannot be felt by the patient**, would not improve activities or decrease rates of exacerbation. **Therefore, it is NOT "clinically significant".**

© COPD Foundation 2021

COPD FOUNDATION

Study Impact

Statistics are used to decide:

- **How should we report the results** or outcomes of the study?
- **Should we recommend changes** to care, life-style or practice based on the results?
- **Who needs to learn about the results?**
  - ***Statistically significant results*** may be **useful to researchers** to suggest a new direction or type of study or medication to study.
  - ***Clinically significant results*** may be **useful to doctors and other clinicians, health systems, insurers, and researchers.**

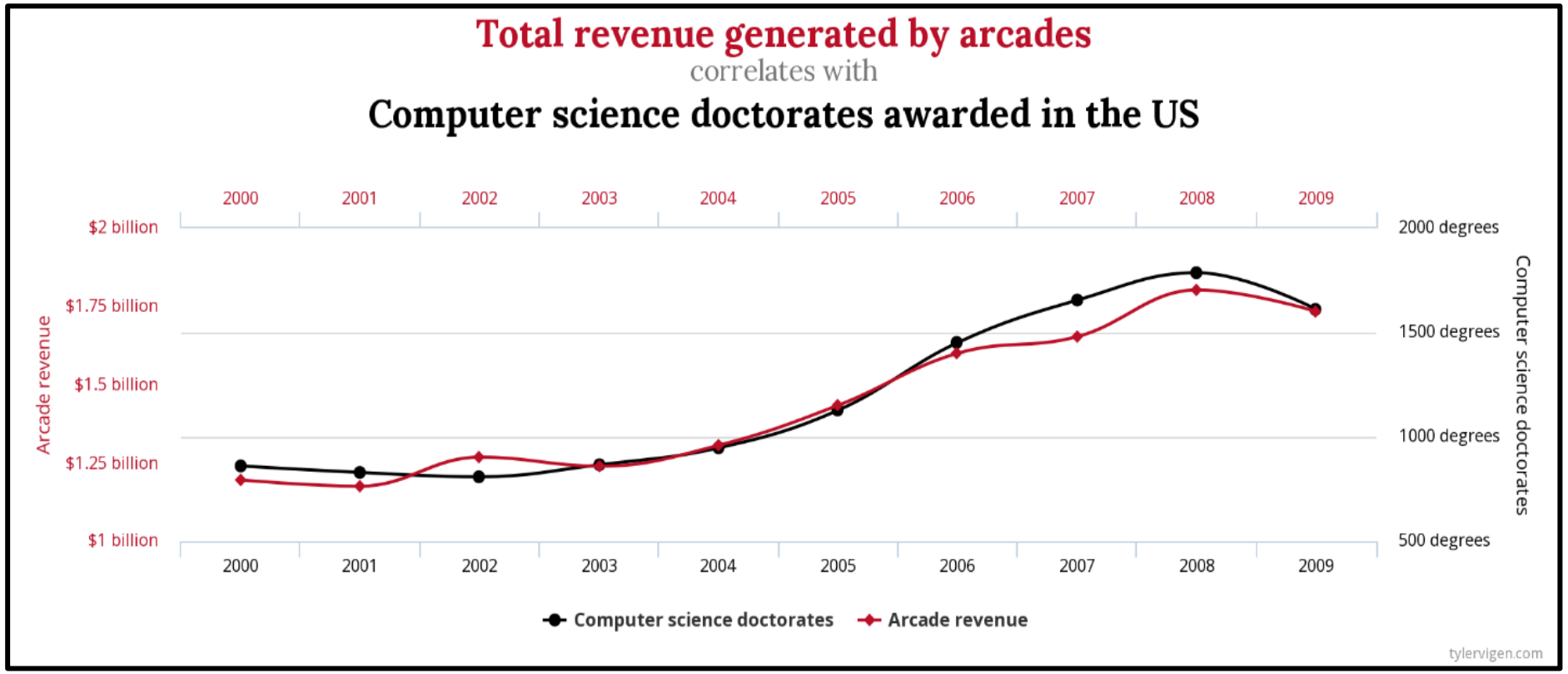
© COPD Foundation 2021

COPD FOUNDATION

Correlation vs. Causation

There is a famous saying in statistics:

“Correlation does not imply causation”



© COPD Foundation 2021

COPD FOUNDATION

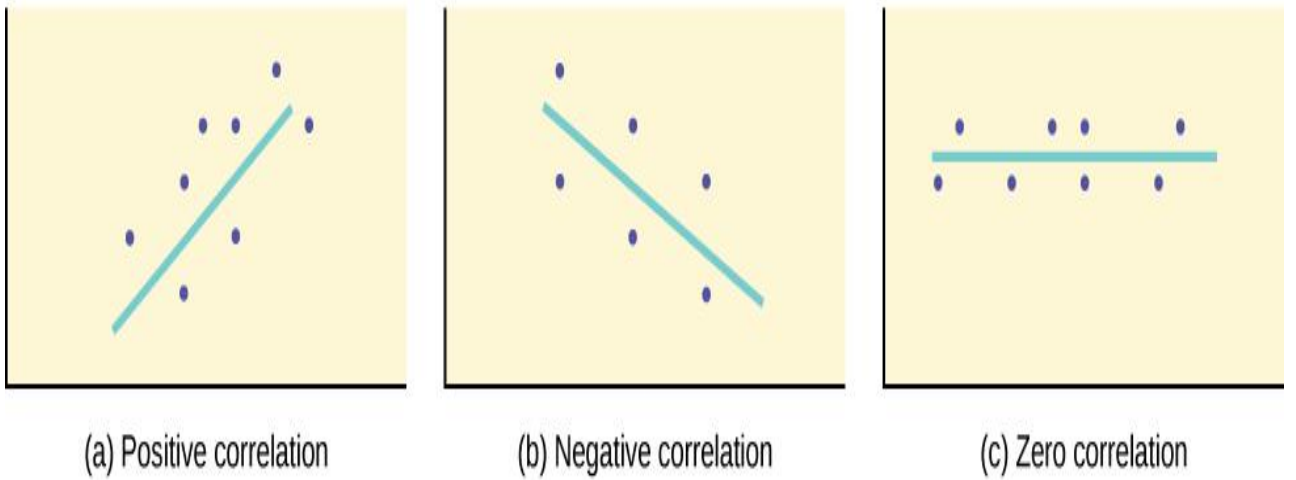
Correlation

Correlation is a **measure of how things are related**.

Specifically, in statistics, it **is a statistical calculation** of ***if and how much*** **two paired variables are related.**

- Two variables can have a **"positive" correlation--** Both variables move in the same direction—increasing or decreasing together, in parallel.
  Example: Tall people have bigger shoe sizes.

- Or two variables can have a **"negative" correlation—one variable increases while the other decreases.**
  Example: Students with increased absences have lower grades.



Correlation works for *quantitative data*—as we learned earlier, this is data that can be *measured and summarized with numbers*.

© COPD Foundation 2021

COPD FOUNDATION

Causation

Causation means **one variable or event is directly responsible for another event** or variable.

Example: If you go out in the rain, you will get wet.

In statistics, **causation is extremely hard to prove**.

Example:

- A study compares individuals who drink alcohol and those who don't.
- The drinkers' group has a higher rate of pancreatic cancer than the non-drinkers.
- But we cannot say if the one variable, drinking, is the cause of the cancer. There could be something else that caused the drinkers to be different from the non-drinkers.
- For example, many people who are smokers, also drink alcohol. So, did the drinkers' smoking cause the higher risk of cancer?

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 2)

- The next few slides have some questions to help you review and remember what we have presented in this Module.

- This is not a graded test and is meant to only help you retain the information from this Module. There is an answer key at the end.

- Here is a link to the key terms that might help as you go through the review questions: click here

- If you have any questions, please email BRIDGE@copdfoundation.org

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 2)

5. Which of the following statements is NOT necessarily true about "statistically significant" data?

a. Is not likely to have occurred by chance
b. Is likely to be attributable to a specific cause
c. Is reliable data
d. Is important and worthy of using to make new decisions

6. Which of these statements is true about *p* value?

a. It is the way in which significance is reported in statistics
b. It measures how likely the results could have just occurred by chance.
c. The lower the *p* value, the more "real" the results or the less likely to be from coincidence.
d. All of the above

7. True or False: The results of a study are statistically significant when *P value is* ≤ 0.05.

a. True
b. False

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 2) continued...

8. Which of the following statements about clinical significance is NOT true?

a. Clinical significance signifies that results can have an impact on health care
b. Clinical significance measures the importance of the results
c. All studies that have statistical significance also have clinical significance.
d. Clinical significance means "what we learned from the different results in the two groups studied is important and has practical applications—we should change how we provide care"

9. Which of the statements below is an example of correlation?

a. Tall people have bigger shoe sizes
b. If you go outside in the rain, you will get wet

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 2) Answer Key

5.d

Explanation: The statement "Is important and worthy of using to make new decisions" is a statement that would be made about data that is "clinically significant." Data that is statistically significant is not necessarily important or decision-worthy. To be statistically significant means the numbers are reliable and the results are not just the result of a coincidence. See slide 20 for a quick review or slides 24-26 for a more in-depth review.

6.d

Explanation: All of the statements are true about *p* value

7.a-true

Explanation: This is the measurement for statistically significant data: *p* value < 0.05. The lower the *p* value, the less likely the data is the result of chance or coincidence and therefore more significant.

© COPD Foundation 2021

COPD FOUNDATION

Optional Module: Review Questions (Part 2) Answer Key continued...

8.c—"All studies that have statistical significance also have clinical significance."
Explanation: This is the one statement that is NOT true about clinical significance. A study can be statistically significant without being clinically significant. If a study is statistically significant, it means the data is reliable and likely NOT the result of chance or a coincidence. BUT that does not mean the findings are important and clinically significant. To be clinically significant, the data must be significant and important enough that health care decisions would be altered or changed because of it. And that doesn't always happen. See slides 24 and 25 along with the example on slide 26 for a review.

9.a
Explanation: "Tall people have bigger shoes" is an example of a correlation. There is a relationship between height and big feet—therefore the example is a correlation. "If you go out in the rain, you will get wet" is an example of causation—rain causes wetness. See slides 28-30 for a review.

© COPD Foundation 2021

COPD FOUNDATION

Links to Key Resources for the Optional Module

Links to Online Videos on Statistics & Resources	**Khan Academy Videos:** • Mean, Median and Mode – https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data • Confidence Intervals – https://www.khanacademy.org/math/ap-statistics/estimating-confidence-ap • Significance Testing and P Values – https://www.khanacademy.org/math/ap-statistics/tests-significance-ap • Standard Deviation and Bias – https://www.khanacademy.org/math/ap-statistics/summarizing-quantitative-data-ap/measuring-spread-quantitative/v/sample-standard-deviation-and-bias
Takeaway Documents	• Glossary of key terms used: click here • Optional Module Summary Document: click here • Optional Module Review Questions with Answers and Explanations: click here

© COPD Foundation 2021

COPD FOUNDATION

Next Steps

Congratulations!
You have completed the Optional Module!

When you are ready, please go on to ***Module 5 - Specific COPD-Related Research Information***.

© COPD Foundation 2021

COPD FOUNDATION