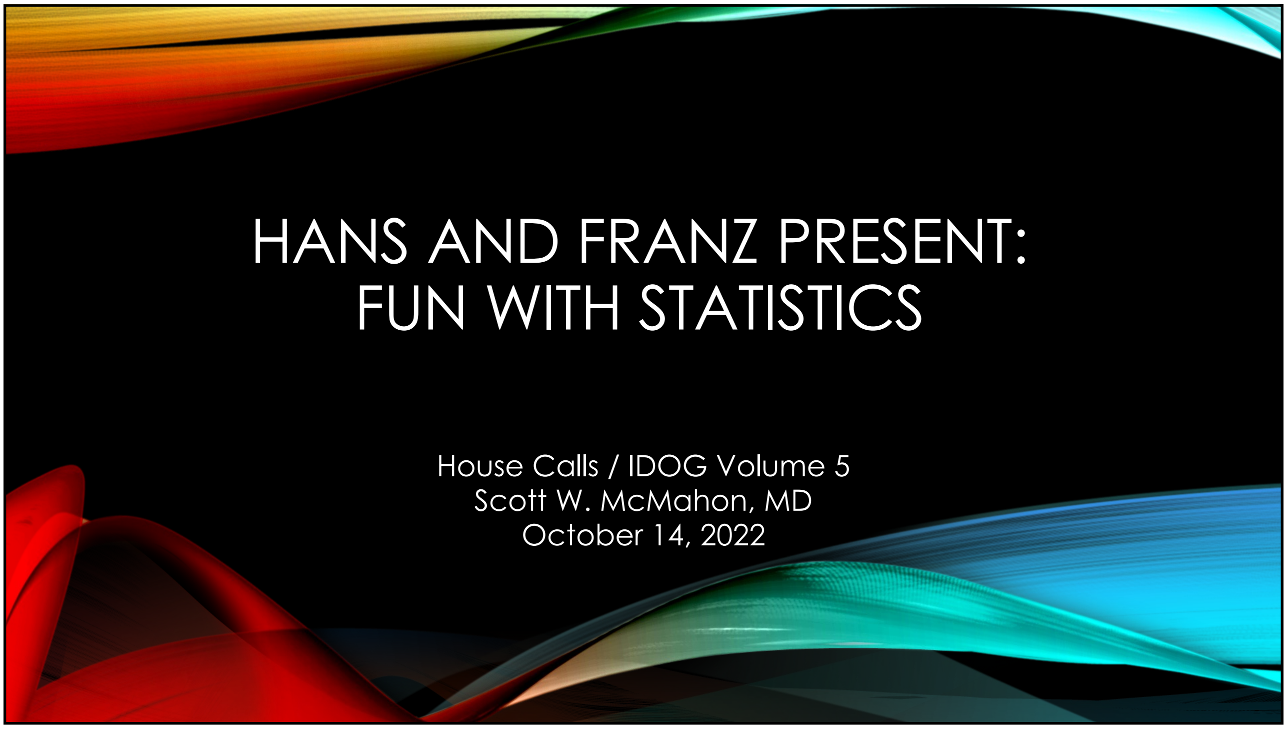
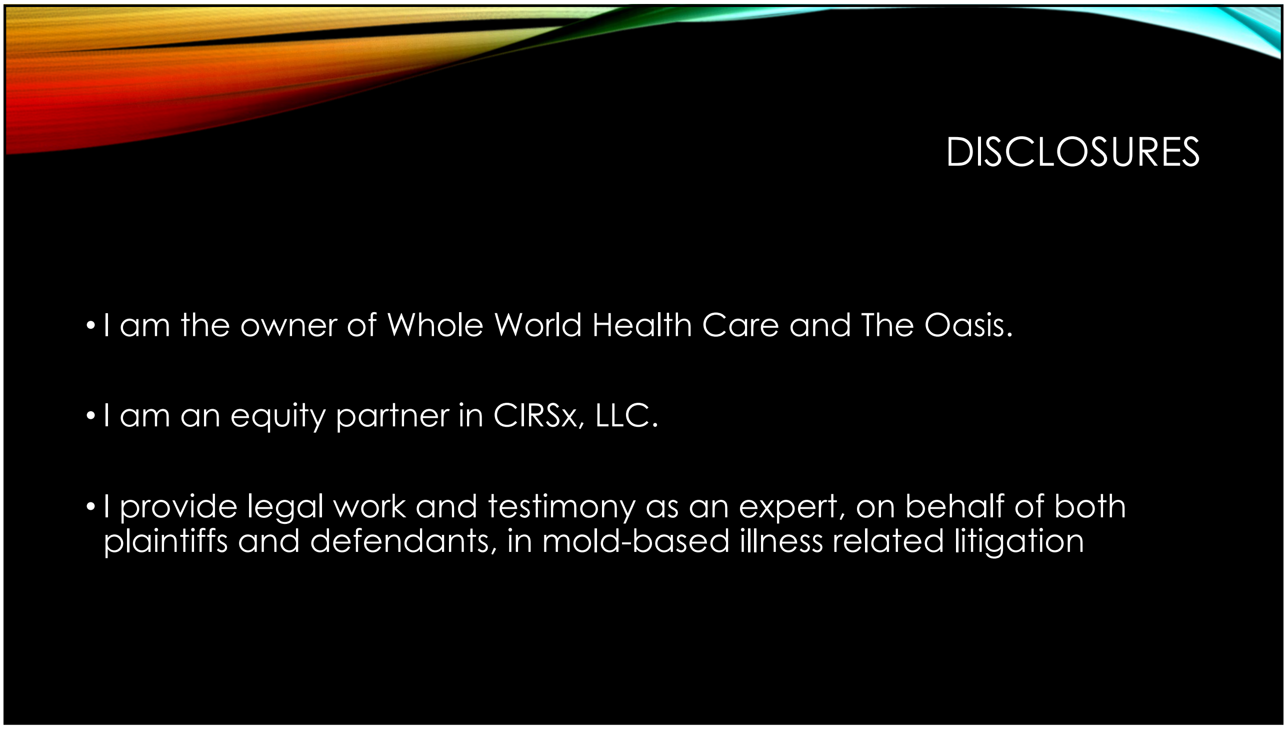


HANS AND FRANZ PRESENT: FUN WITH STATISTICS

House Calls / IDOG Volume 5
Scott W. McMahon, MD
October 14, 2022

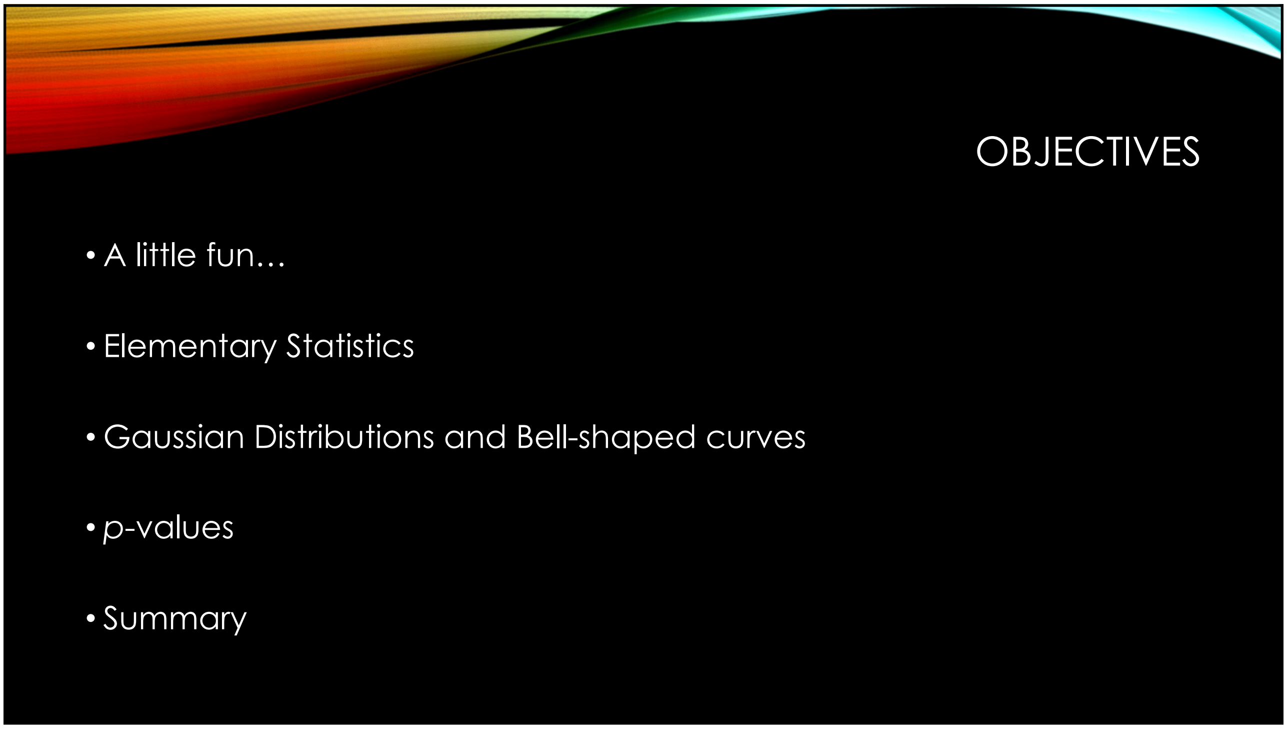
1



DISCLOSURES

- I am the owner of Whole World Health Care and The Oasis.
- I am an equity partner in CIRSx, LLC.
- I provide legal work and testimony as an expert, on behalf of both plaintiffs and defendants, in mold-based illness related litigation

2



OBJECTIVES

- A little fun…
- Elementary Statistics
- Gaussian Distributions and Bell-shaped curves
- *p*-values
- Summary

3

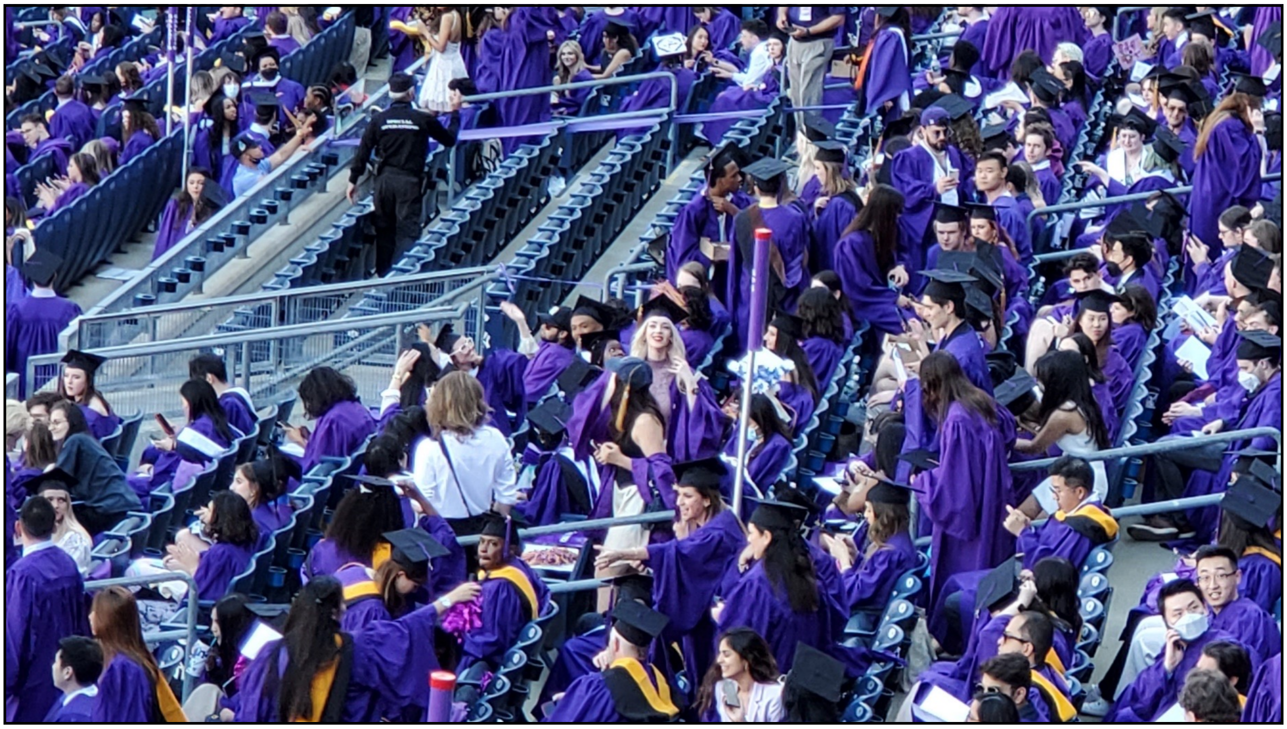
A LITTLE FUN…

4

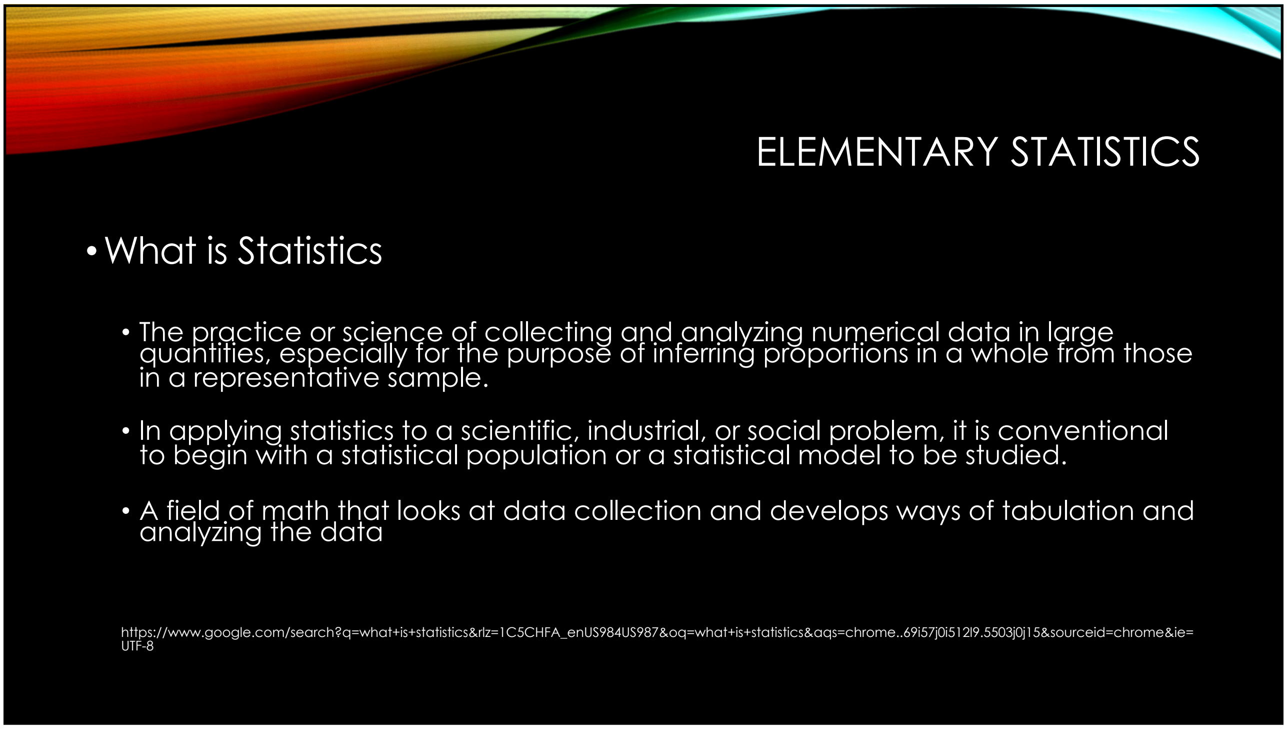
PARTY DECK
PARTY DECK
NYU Welcome Graduates, Family & Friends #NYU2020 #NYU2021
27-Time World Champions
DELTA
Nathan's
AVIS
YANKEE STADIUM

5

6



7

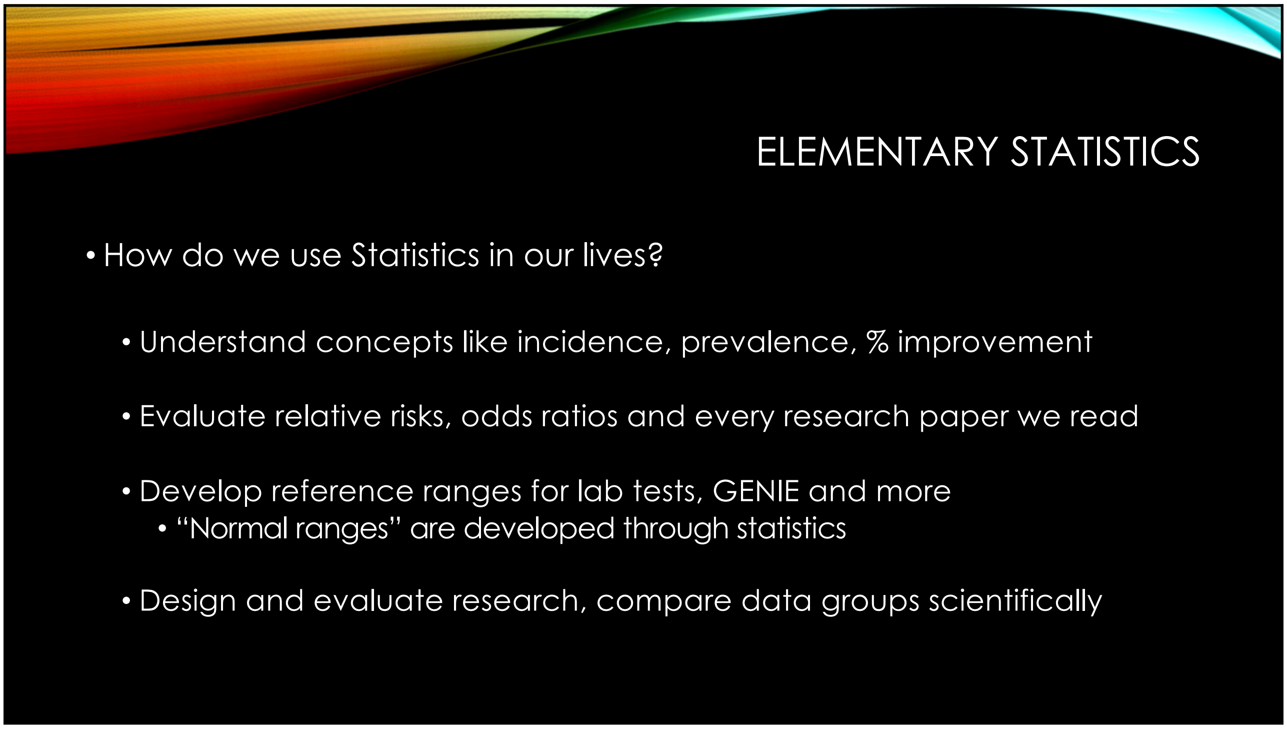


ELEMENTARY STATISTICS

- What is Statistics
  - The practice or science of collecting and analyzing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.
  - In applying statistics to a scientific, industrial, or social problem, it is conventional to begin with a statistical population or a statistical model to be studied.
  - A field of math that looks at data collection and develops ways of tabulation and analyzing the data

https://www.google.com/search?q=what+is+statistics&rlz=1C5CHFA_enUS984US987&oq=what+is+statistics&aqs=chrome..69i57j0i512l9.5503j0j15&sourceid=chrome&ie=UTF-8

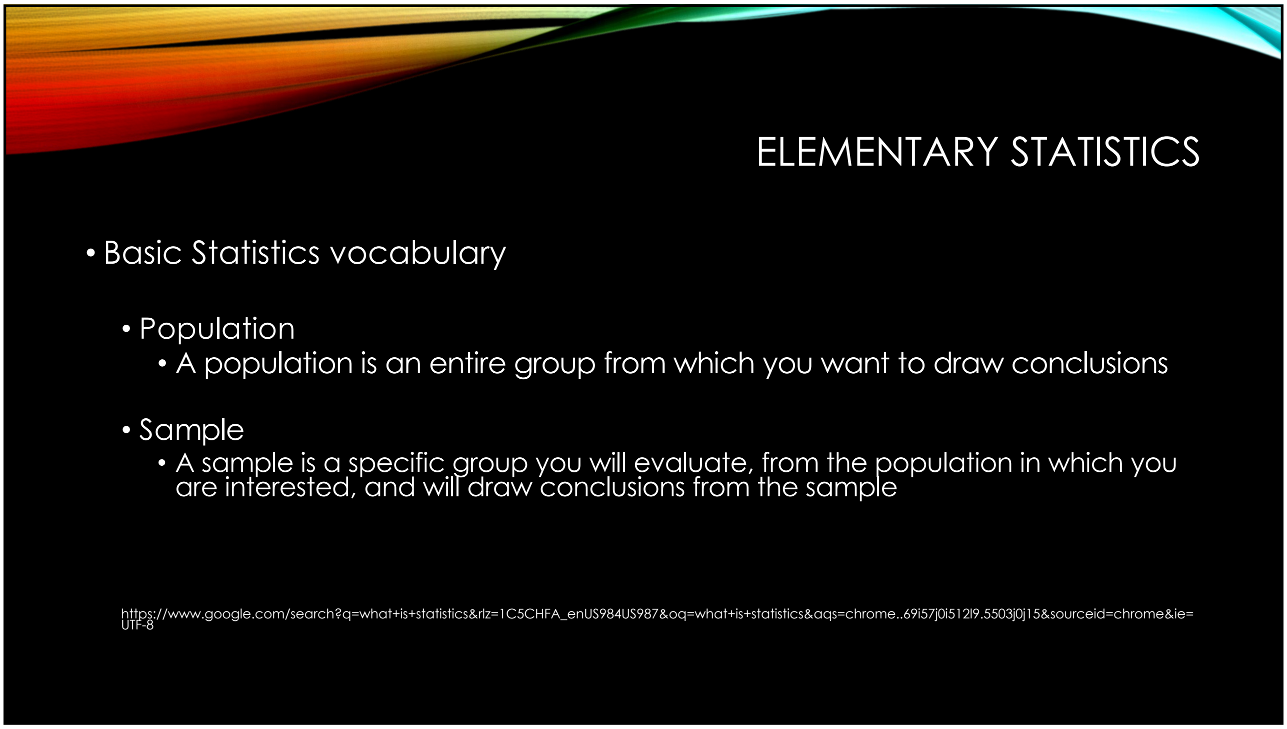
8



ELEMENTARY STATISTICS

- How do we use Statistics in our lives?
  - Understand concepts like incidence, prevalence, % improvement
  - Evaluate relative risks, odds ratios and every research paper we read
  - Develop reference ranges for lab tests, GENIE and more
    - "Normal ranges" are developed through statistics
  - Design and evaluate research, compare data groups scientifically

9

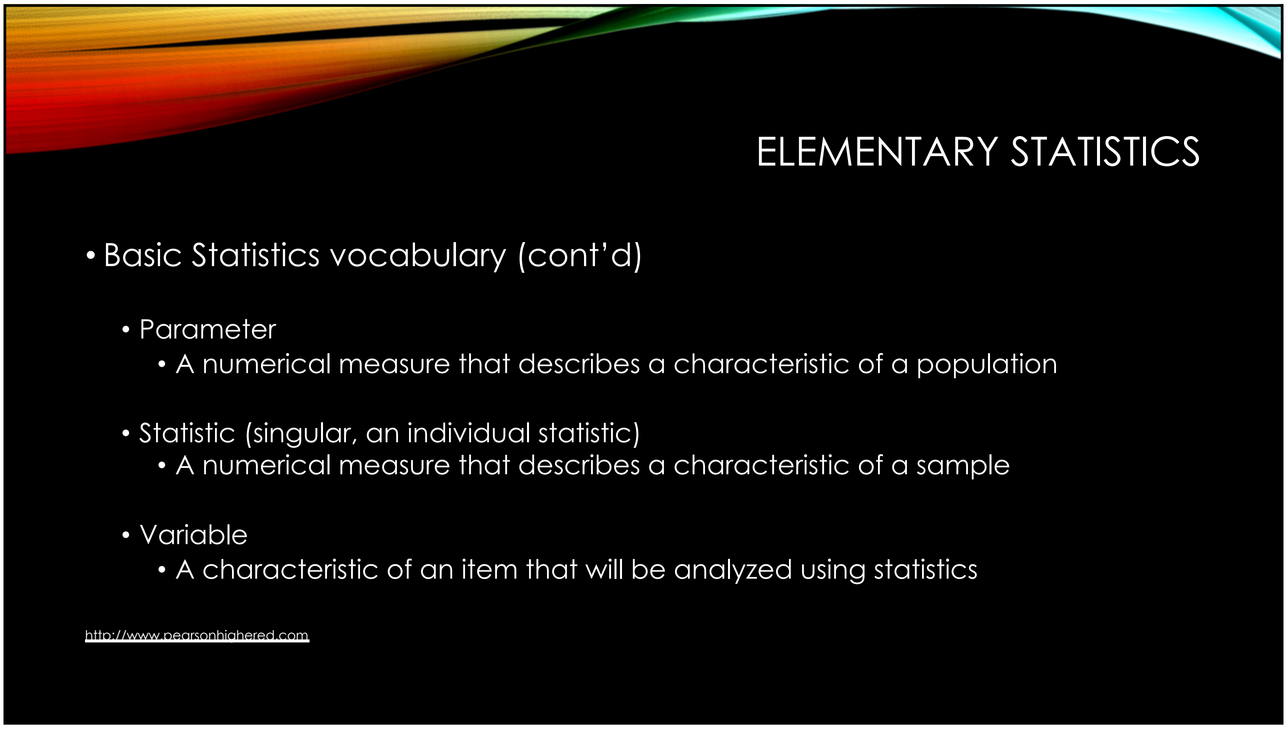


ELEMENTARY STATISTICS

- Basic Statistics vocabulary
  - Population
    - A population is an entire group from which you want to draw conclusions
  - Sample
    - A sample is a specific group you will evaluate, from the population in which you are interested, and will draw conclusions from the sample

https://www.google.com/search?q=what+is+statistics&rlz=1C5CHFA_enUS984US987&oq=what+is+statistics&aqs=chrome..69i57j0i512l9.5503j0j15&sourceid=chrome&ie=UTF-8

10

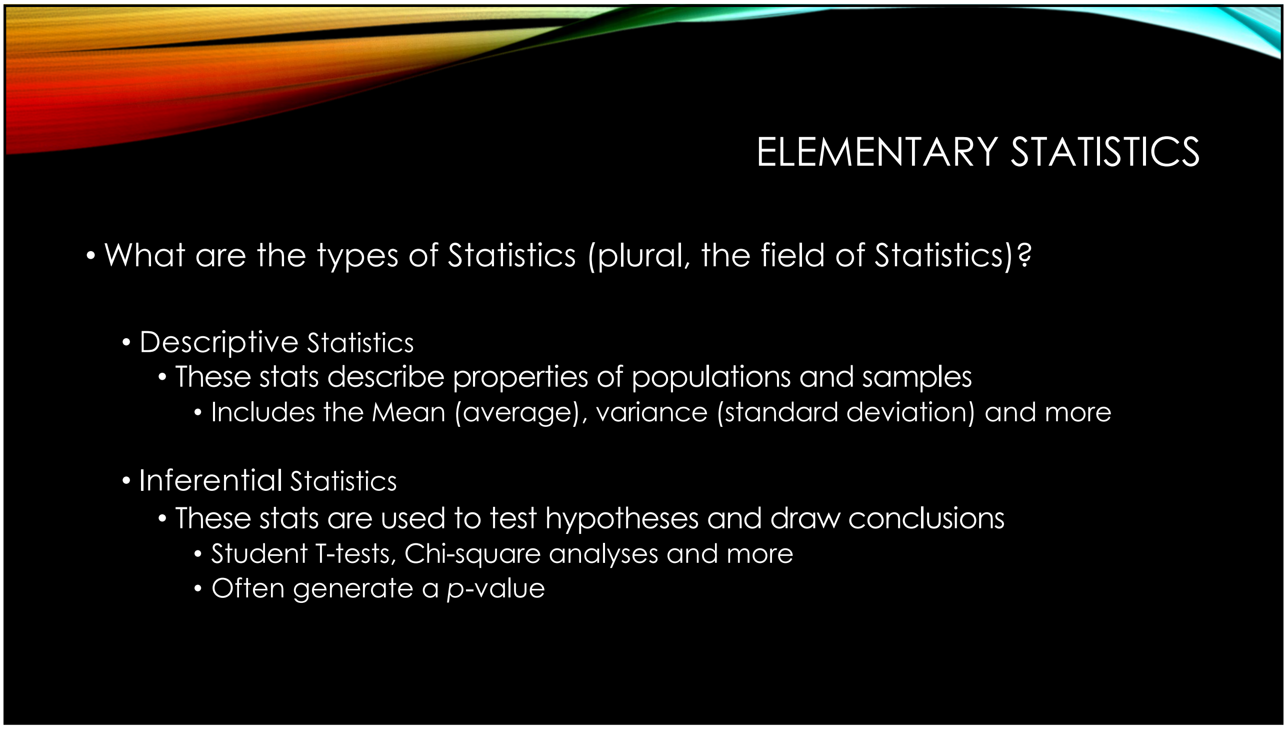


ELEMENTARY STATISTICS

- Basic Statistics vocabulary (cont'd)
  - Parameter
    - A numerical measure that describes a characteristic of a population
  - Statistic (singular, an individual statistic)
    - A numerical measure that describes a characteristic of a sample
  - Variable
    - A characteristic of an item that will be analyzed using statistics

http://www.pearsonhighered.com

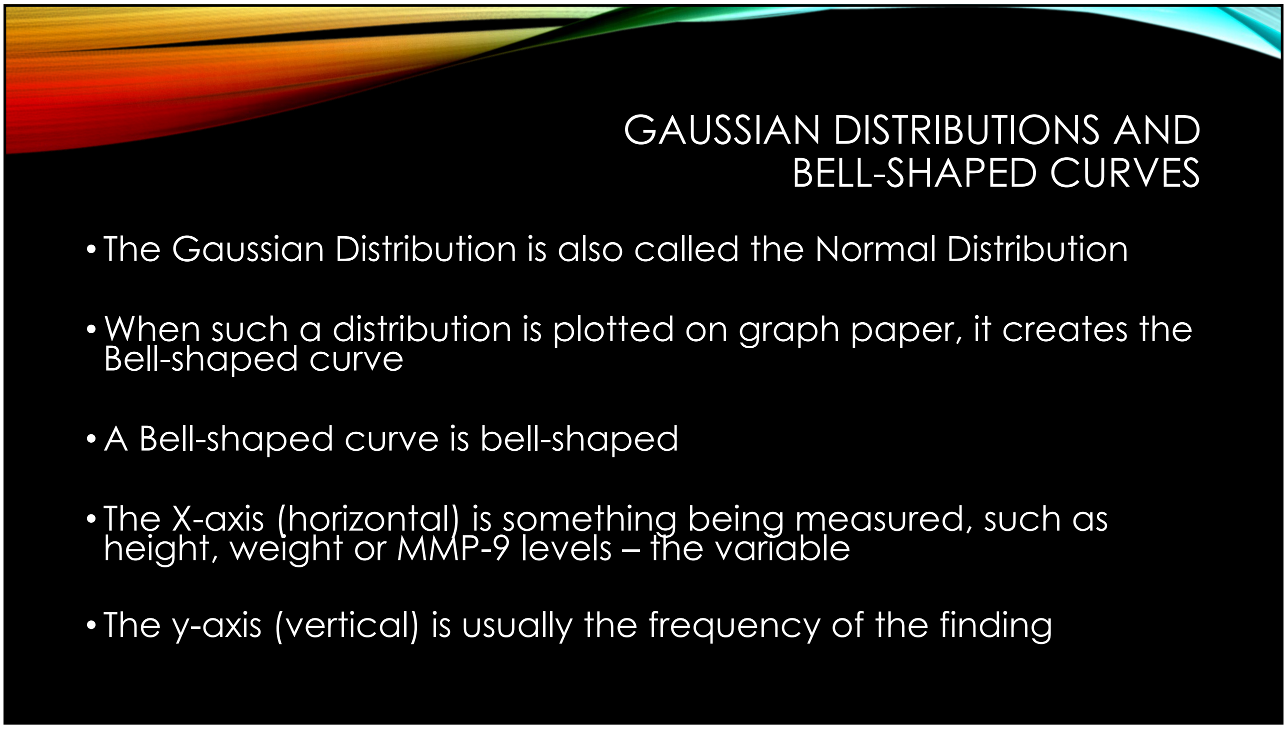
11



ELEMENTARY STATISTICS

- What are the types of Statistics (plural, the field of Statistics)?
  - Descriptive Statistics
    - These stats describe properties of populations and samples
      - Includes the Mean (average), variance (standard deviation) and more
  - Inferential Statistics
    - These stats are used to test hypotheses and draw conclusions
      - Student T-tests, Chi-square analyses and more
      - Often generate a *p*-value

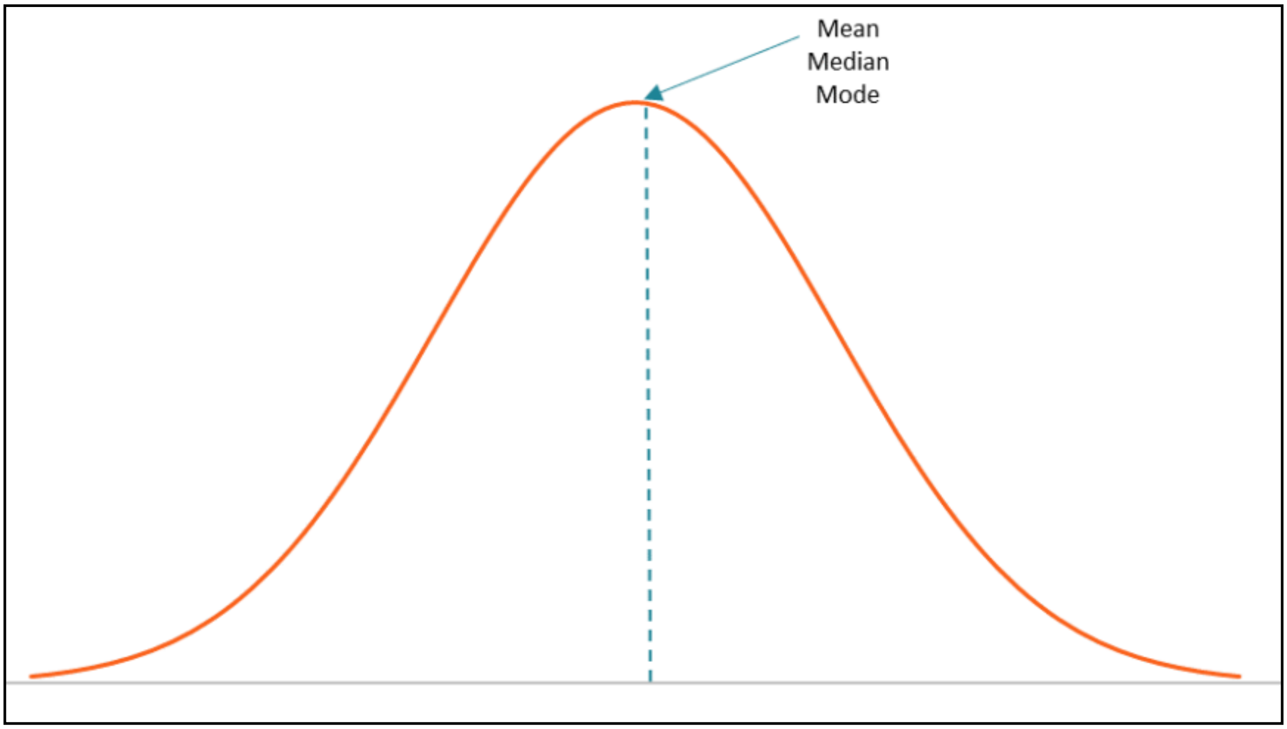
12



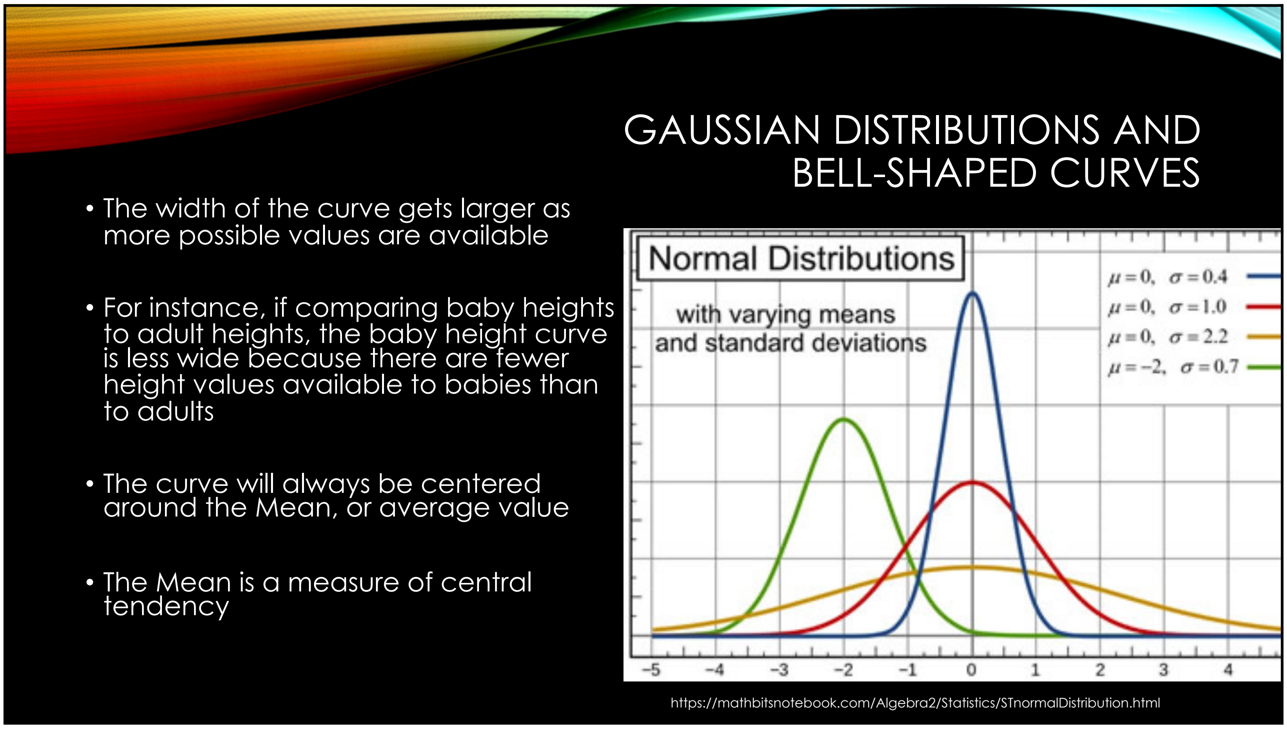
GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- The Gaussian Distribution is also called the Normal Distribution
- When such a distribution is plotted on graph paper, it creates the Bell-shaped curve
- A Bell-shaped curve is bell-shaped
- The X-axis (horizontal) is something being measured, such as height, weight or MMP-9 levels – the variable
- The y-axis (vertical) is usually the frequency of the finding

13



14

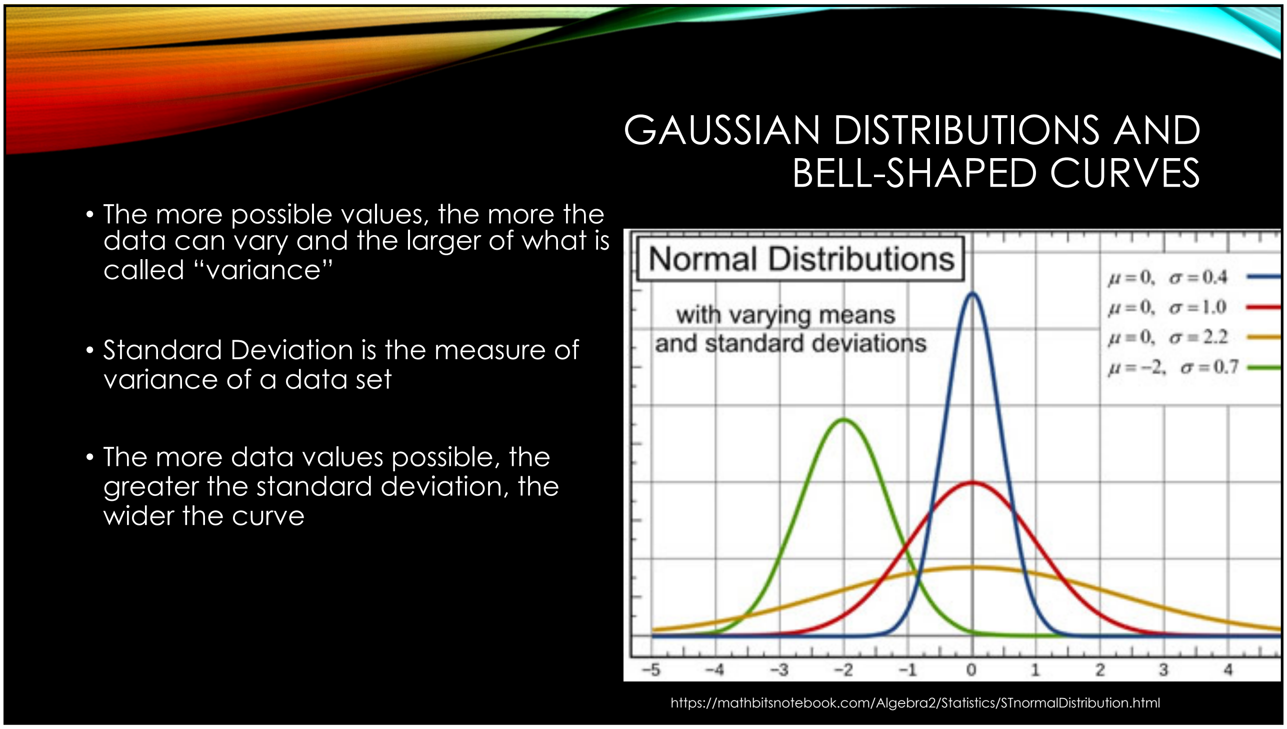


GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- The width of the curve gets larger as more possible values are available
- For instance, if comparing baby heights to adult heights, the baby height curve is less wide because there are fewer height values available to babies than to adults
- The curve will always be centered around the Mean, or average value
- The Mean is a measure of central tendency

https://mathbitsnotebook.com/Algebra2/Statistics/STnormalDistribution.html

15

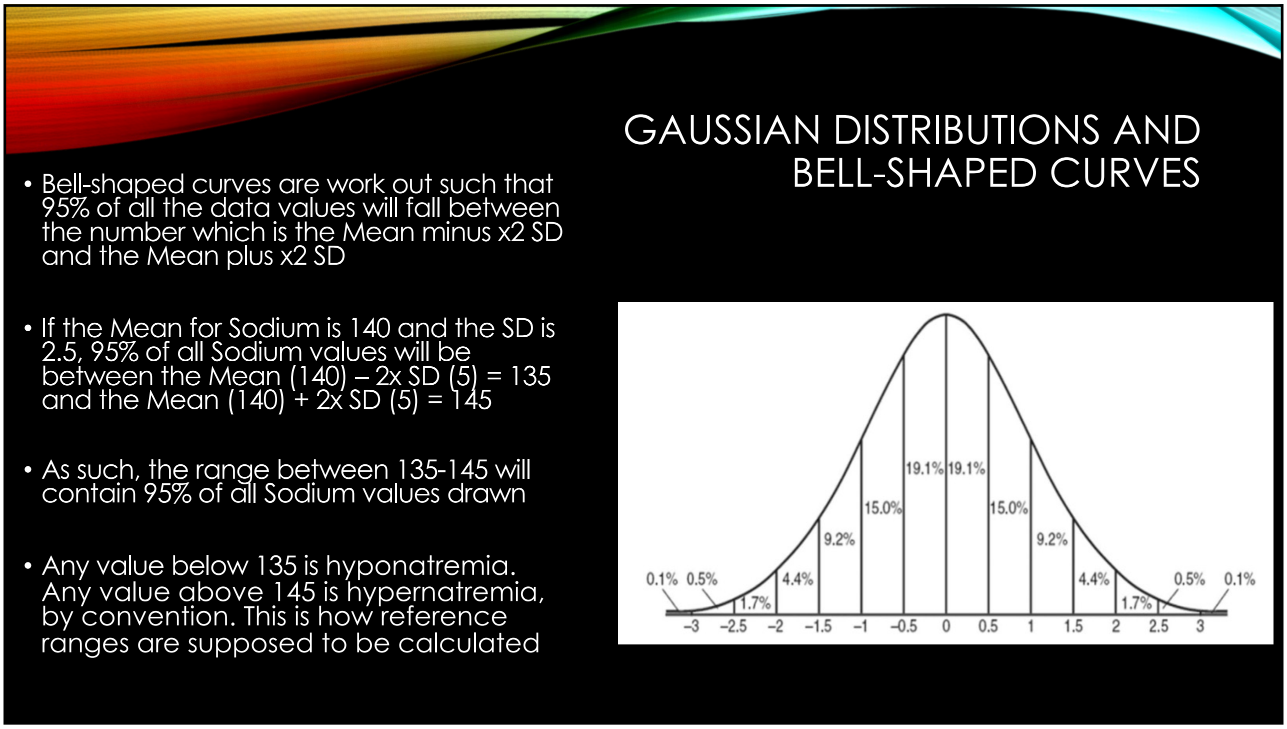


GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- The more possible values, the more the data can vary and the larger of what is called “variance”
- Standard Deviation is the measure of variance of a data set
- The more data values possible, the greater the standard deviation, the wider the curve

https://mathbitsnotebook.com/Algebra2/Statistics/STnormalDistribution.html

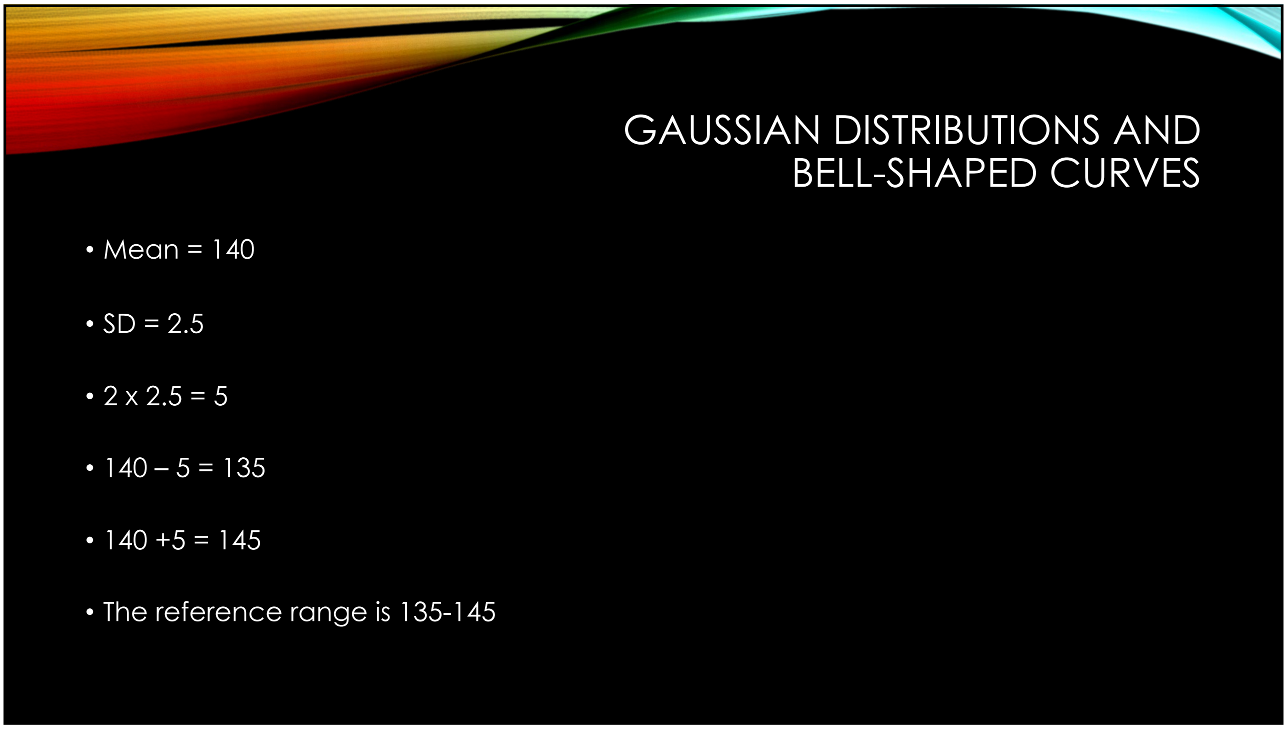
16



GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- Bell-shaped curves are work out such that 95% of all the data values will fall between the number which is the Mean minus x2 SD and the Mean plus x2 SD
- If the Mean for Sodium is 140 and the SD is 2.5, 95% of all Sodium values will be between the Mean (140) – 2x SD (5) = 135 and the Mean (140) + 2x SD (5) = 145
- As such, the range between 135-145 will contain 95% of all Sodium values drawn
- Any value below 135 is hyponatremia. Any value above 145 is hypernatremia, by convention. This is how reference ranges are supposed to be calculated

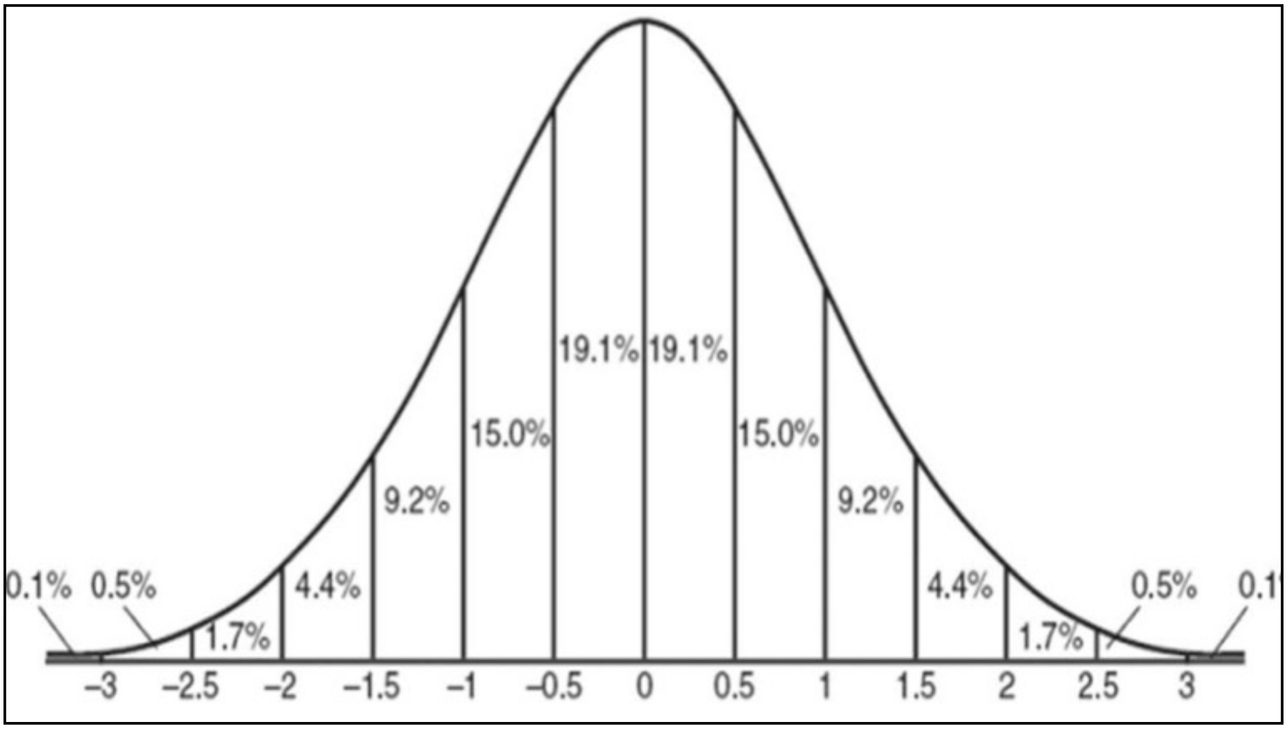
17



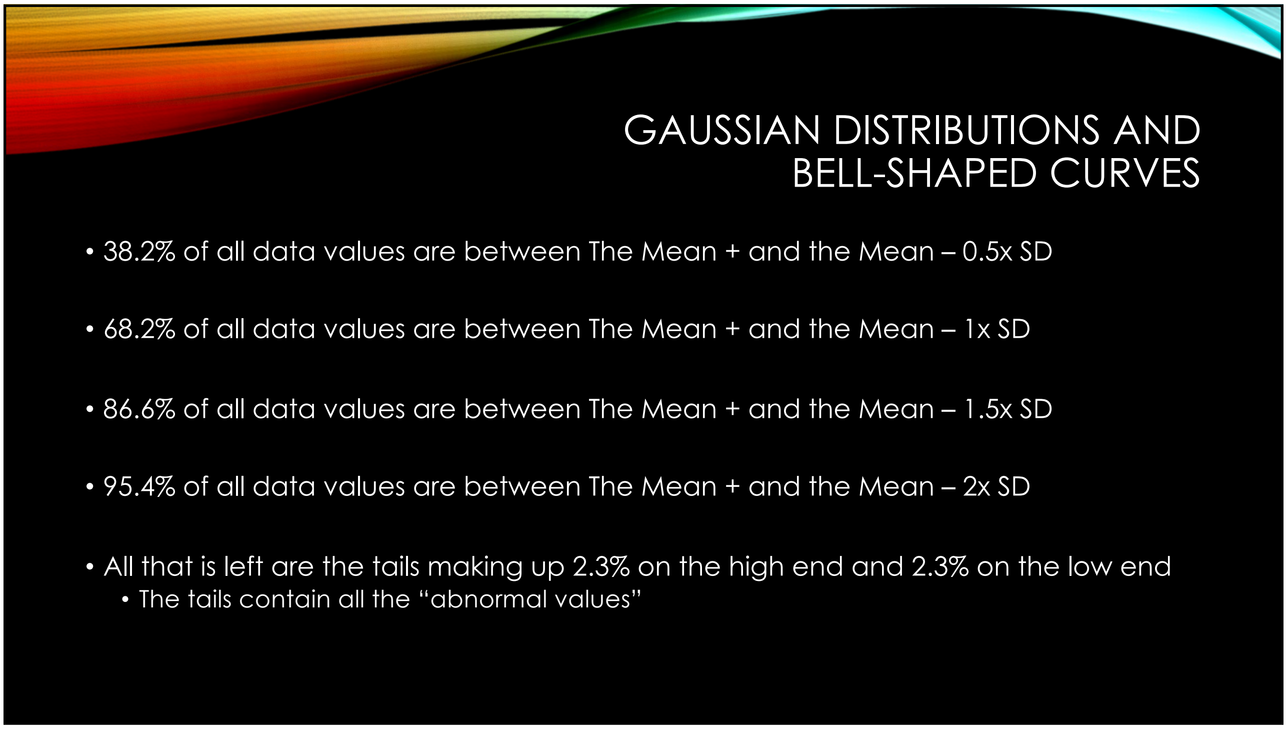
GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- Mean = 140
- SD = 2.5
- 2 x 2.5 = 5
- 140 – 5 = 135
- 140 +5 = 145
- The reference range is 135-145

18



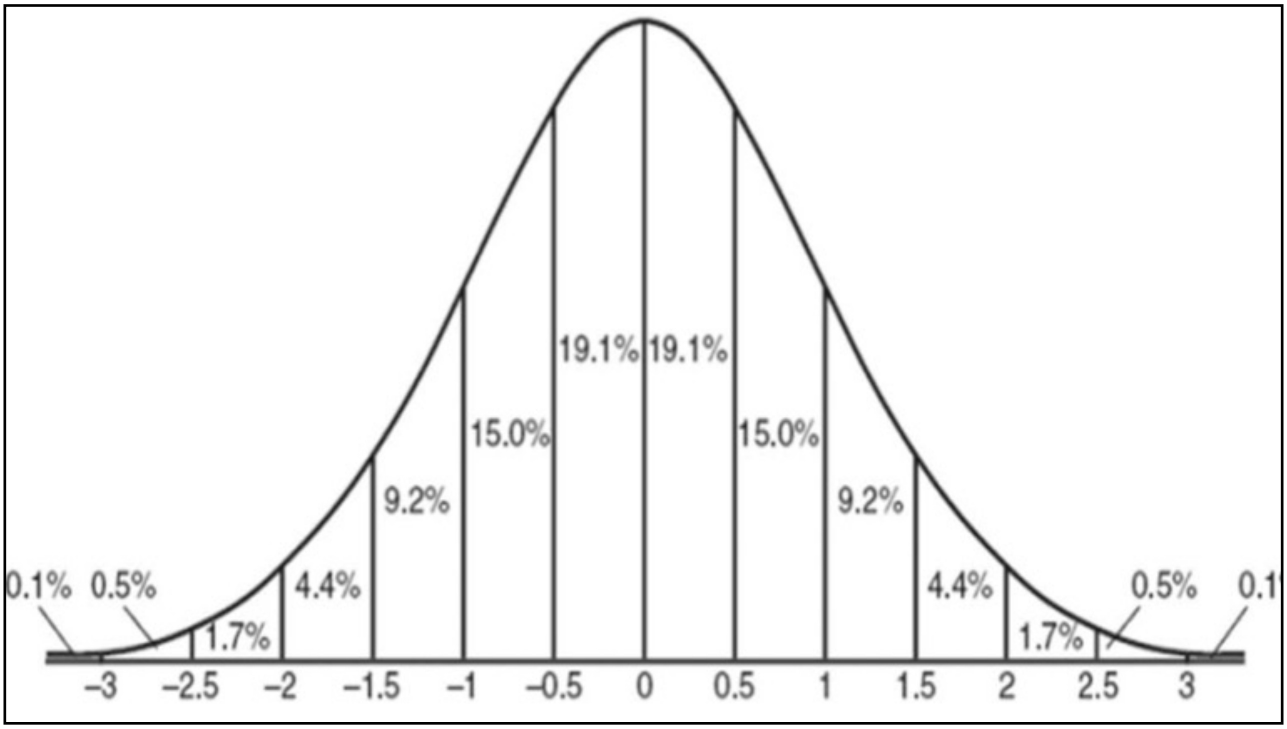
19



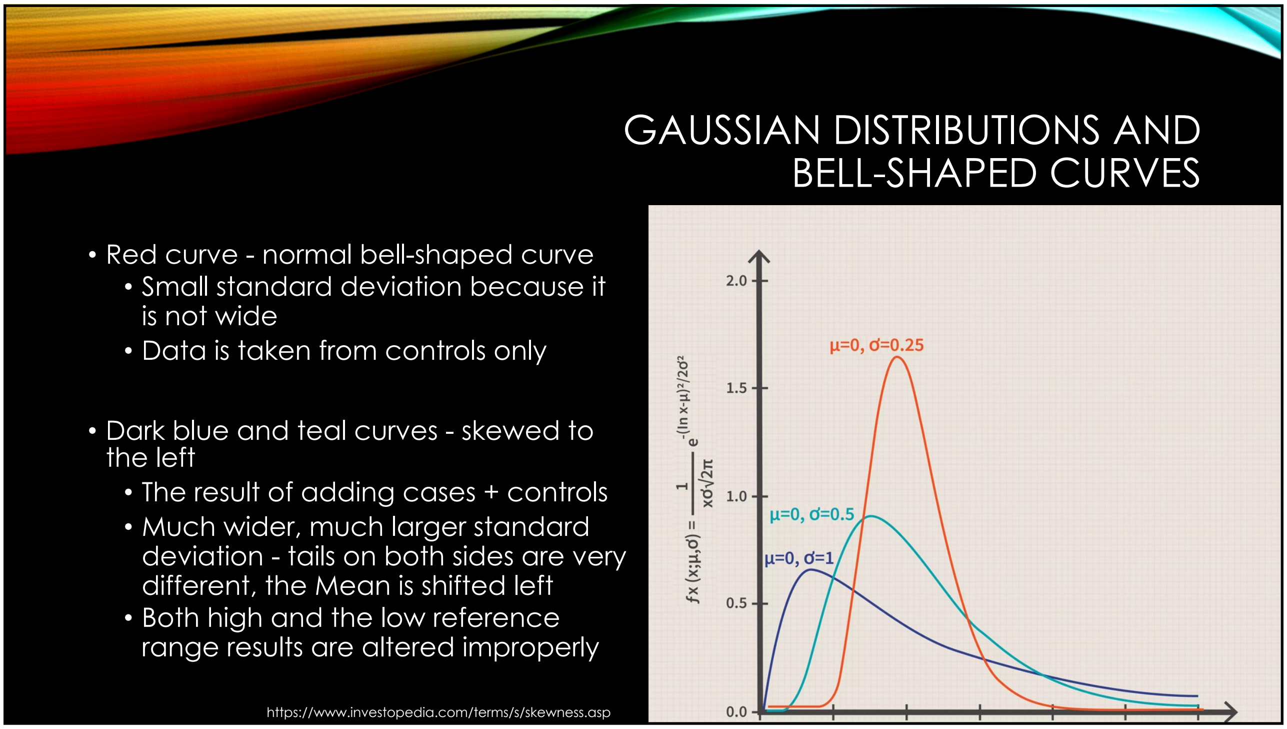
GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- 38.2% of all data values are between The Mean + and the Mean – 0.5x SD
- 68.2% of all data values are between The Mean + and the Mean – 1x SD
- 86.6% of all data values are between The Mean + and the Mean – 1.5x SD
- 95.4% of all data values are between The Mean + and the Mean – 2x SD
- All that is left are the tails making up 2.3% on the high end and 2.3% on the low end
  - The tails contain all the "abnormal values"

20



21



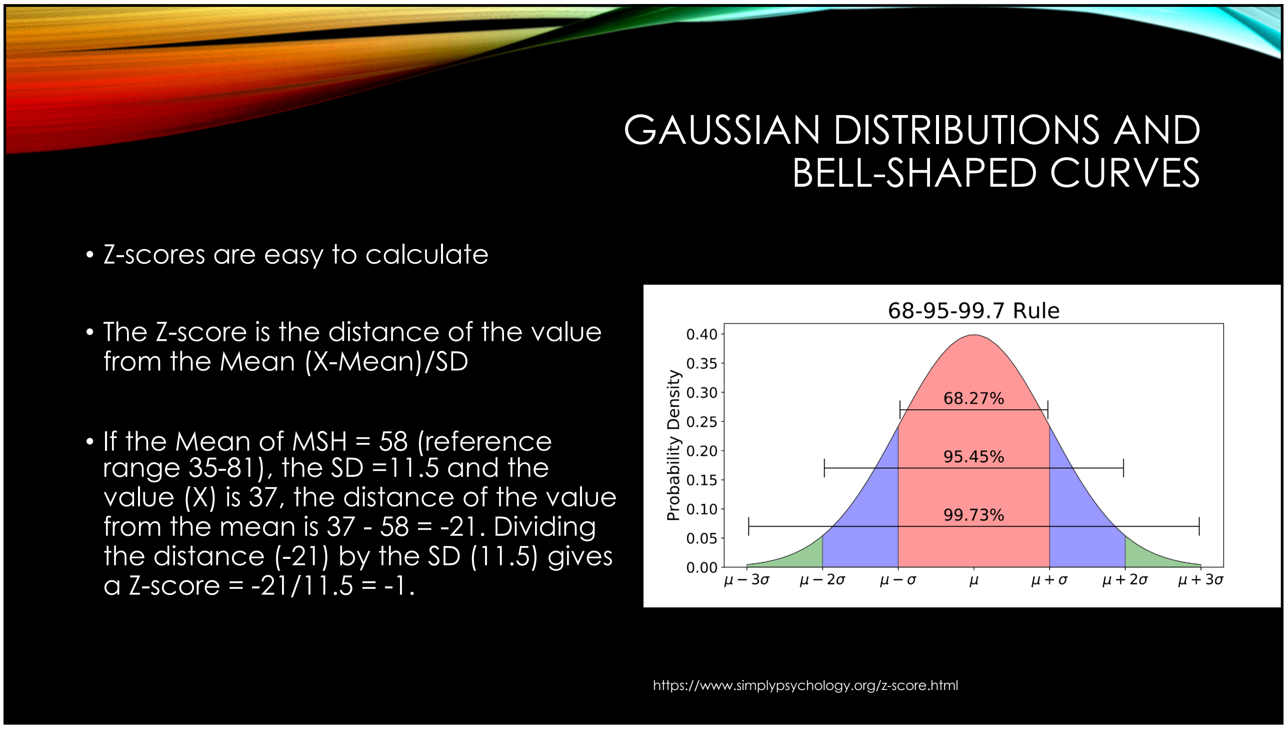
GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- Red curve - normal bell-shaped curve
  - Small standard deviation because it is not wide
  - Data is taken from controls only
- Dark blue and teal curves - skewed to the left
  - The result of adding cases + controls
  - Much wider, much larger standard deviation - tails on both sides are very different, the Mean is shifted left
  - Both high and the low reference range results are altered improperly

https://www.investopedia.com/terms/s/skewness.asp

$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x-\mu)^2/2\sigma^2}$
2.0
1.5
1.0
0.5
0.0
μ=0, σ=0.25
μ=0, σ=0.5
μ=0, σ=1

22

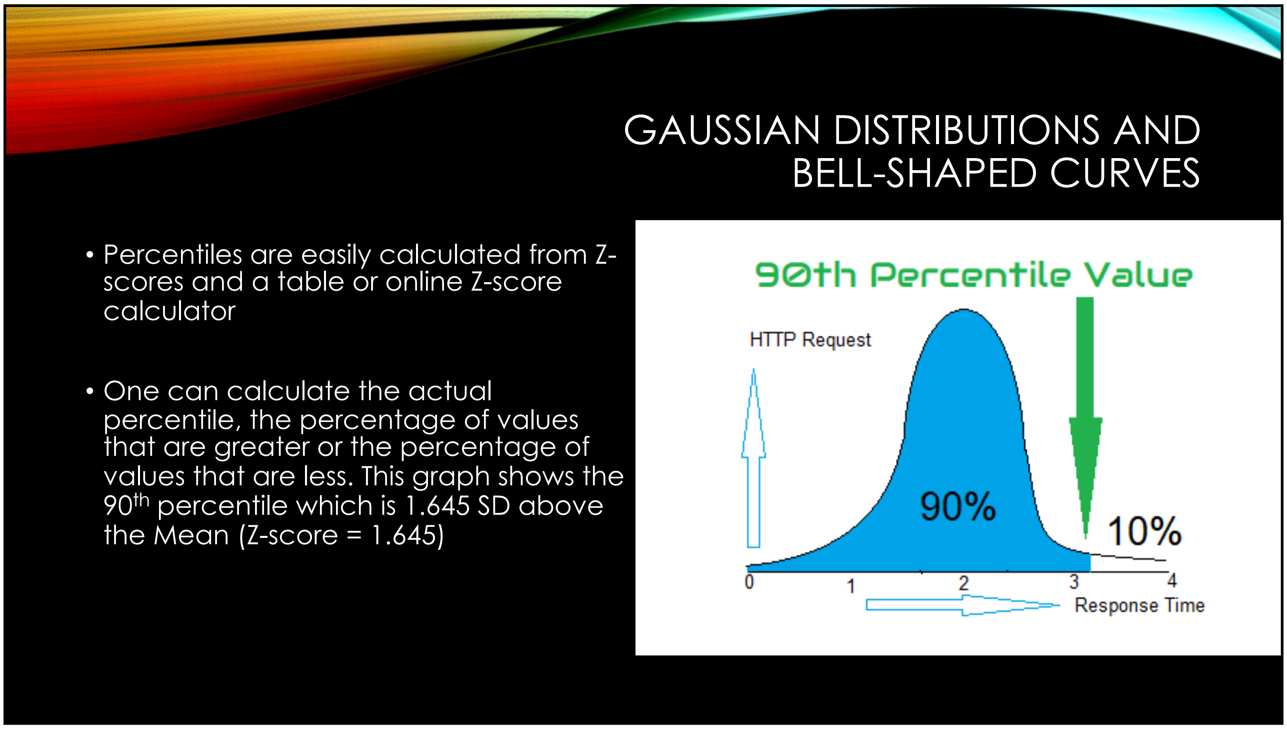


GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- Z-scores are easy to calculate
- The Z-score is the distance of the value from the Mean (X-Mean)/SD
- If the Mean of MSH = 58 (reference range 35-81), the SD =11.5 and the value (X) is 37, the distance of the value from the mean is 37 - 58 = -21. Dividing the distance (-21) by the SD (11.5) gives a Z-score = -21/11.5 = -1.

https://www.simplypsychology.org/z-score.html

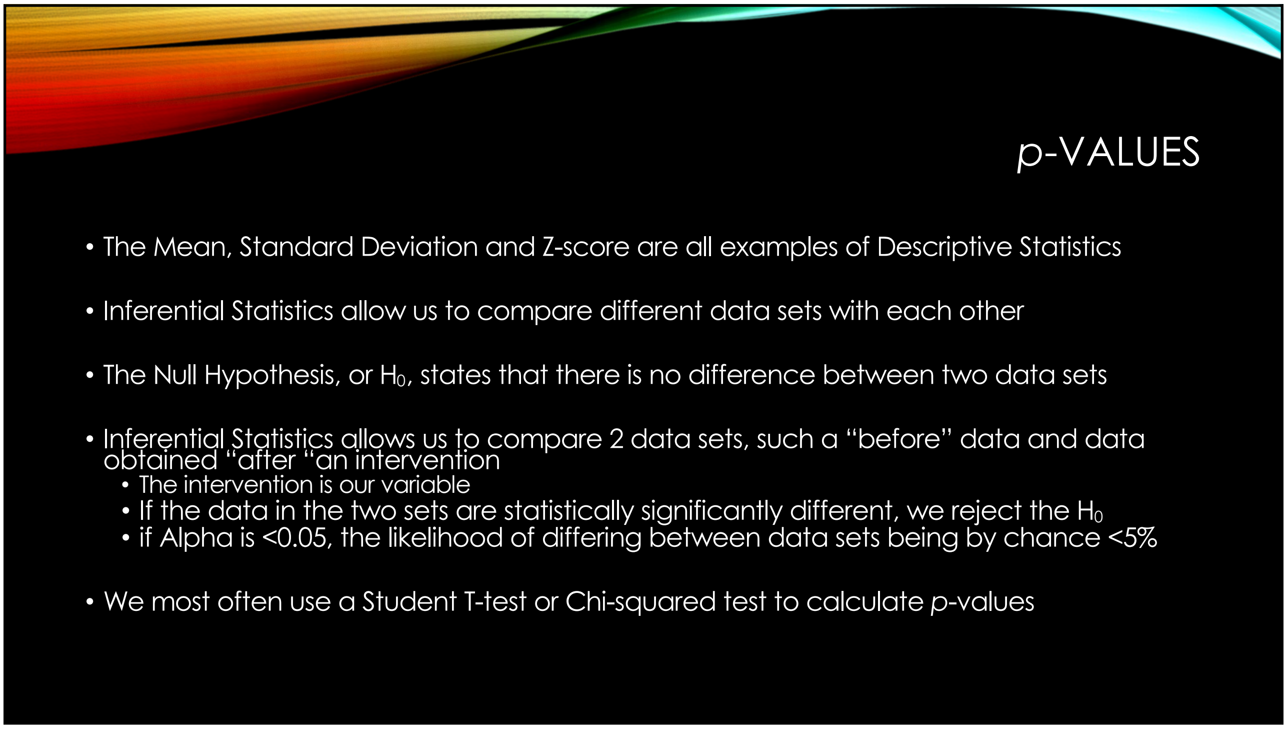
23



GAUSSIAN DISTRIBUTIONS AND BELL-SHAPED CURVES

- Percentiles are easily calculated from Z-scores and a table or online Z-score calculator
- One can calculate the actual percentile, the percentage of values that are greater or the percentage of values that are less. This graph shows the 90th percentile which is 1.645 SD above the Mean (Z-score = 1.645)

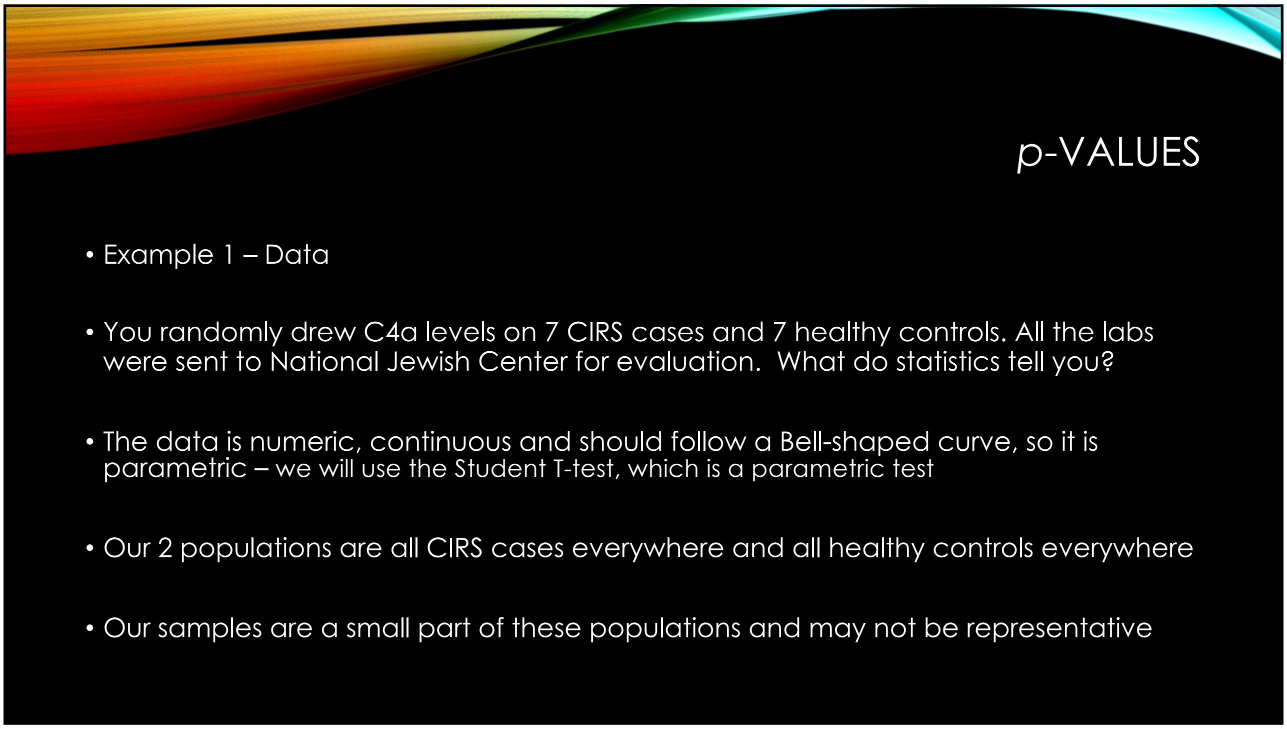
24



p-VALUES

- The Mean, Standard Deviation and Z-score are all examples of Descriptive Statistics
- Inferential Statistics allow us to compare different data sets with each other
- The Null Hypothesis, or H_0, states that there is no difference between two data sets
- Inferential Statistics allows us to compare 2 data sets, such a “before” data and data obtained “after “an intervention
  - The intervention is our variable
  - If the data in the two sets are statistically significantly different, we reject the H_0
  - if Alpha is <0.05, the likelihood of differing between data sets being by chance <5%
- We most often use a Student T-test or Chi-squared test to calculate *p*-values

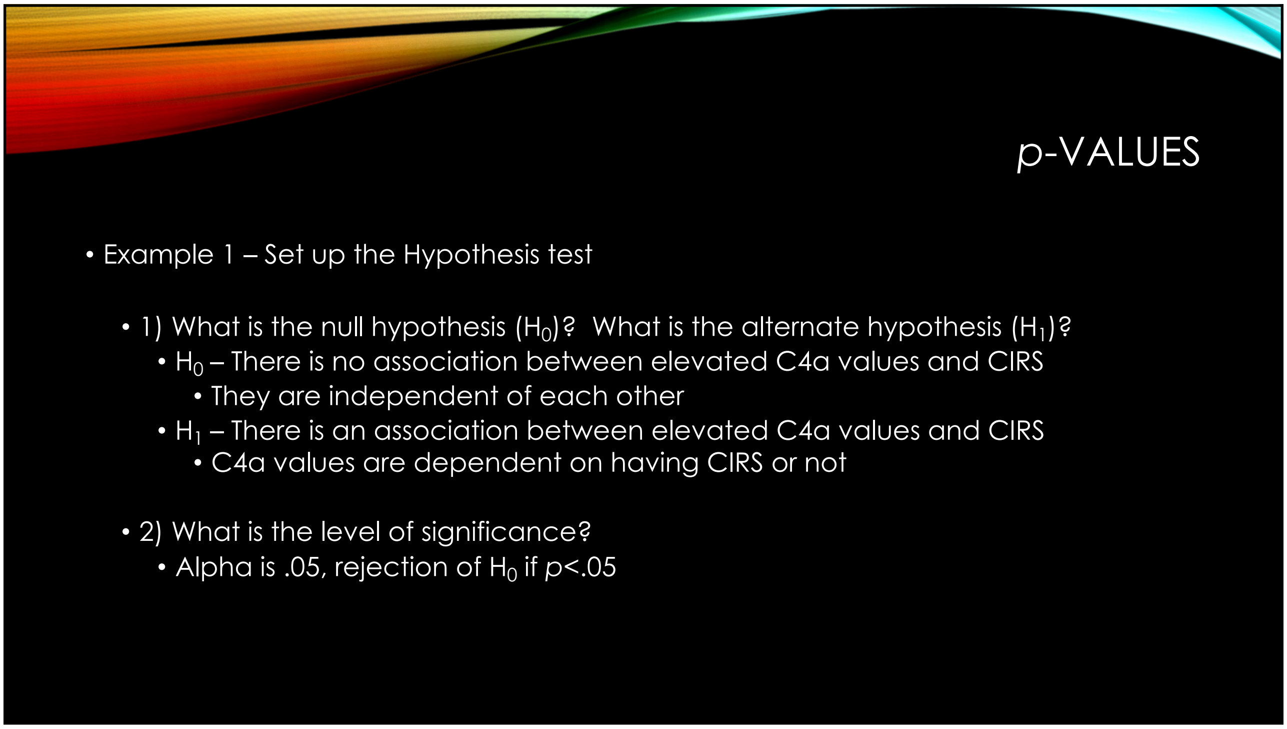
25



p-VALUES

- Example 1 – Data
- You randomly drew C4a levels on 7 CIRS cases and 7 healthy controls. All the labs were sent to National Jewish Center for evaluation. What do statistics tell you?
- The data is numeric, continuous and should follow a Bell-shaped curve, so it is parametric – we will use the Student T-test, which is a parametric test
- Our 2 populations are all CIRS cases everywhere and all healthy controls everywhere
- Our samples are a small part of these populations and may not be representative

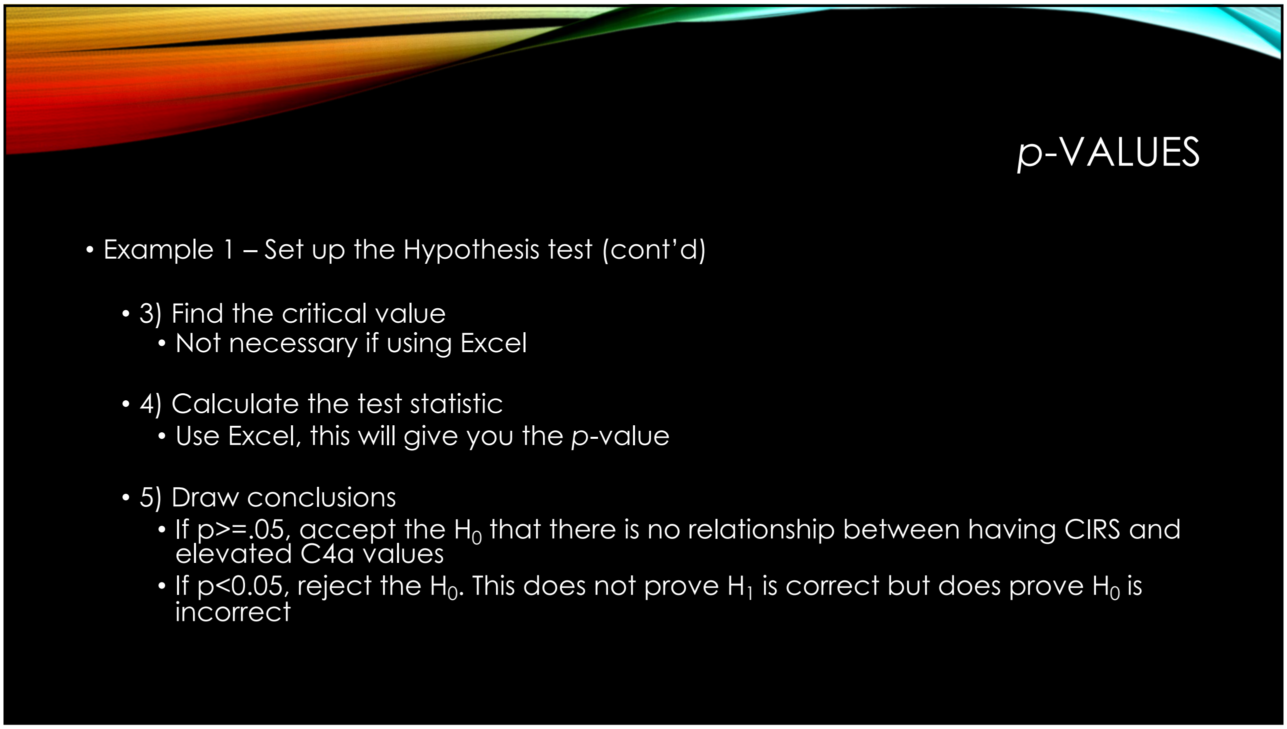
26



p-VALUES

- Example 1 – Set up the Hypothesis test
  - 1) What is the null hypothesis (H_0)? What is the alternate hypothesis (H_1)?
    - H_0 – There is no association between elevated C4a values and CIRS
      - They are independent of each other
    - H_1 – There is an association between elevated C4a values and CIRS
      - C4a values are dependent on having CIRS or not
  - 2) What is the level of significance?
    - Alpha is .05, rejection of H_0 if $p<.05$

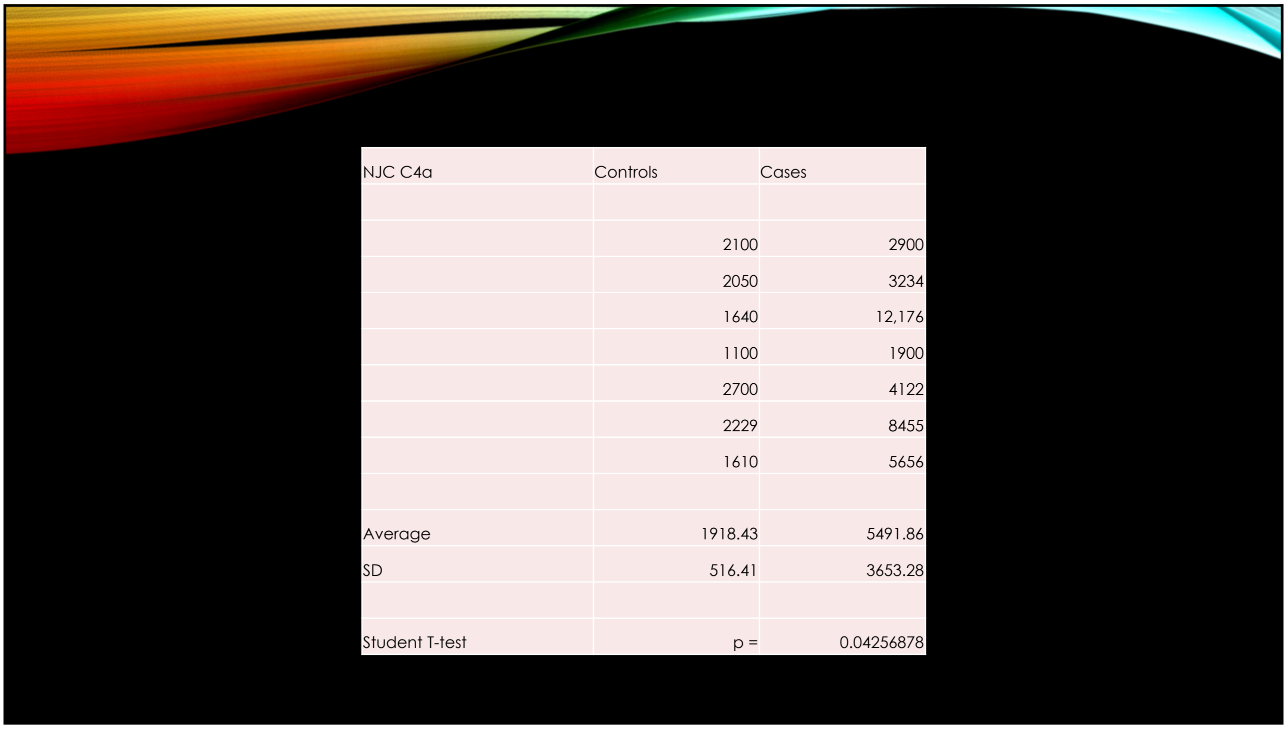
27



p-VALUES

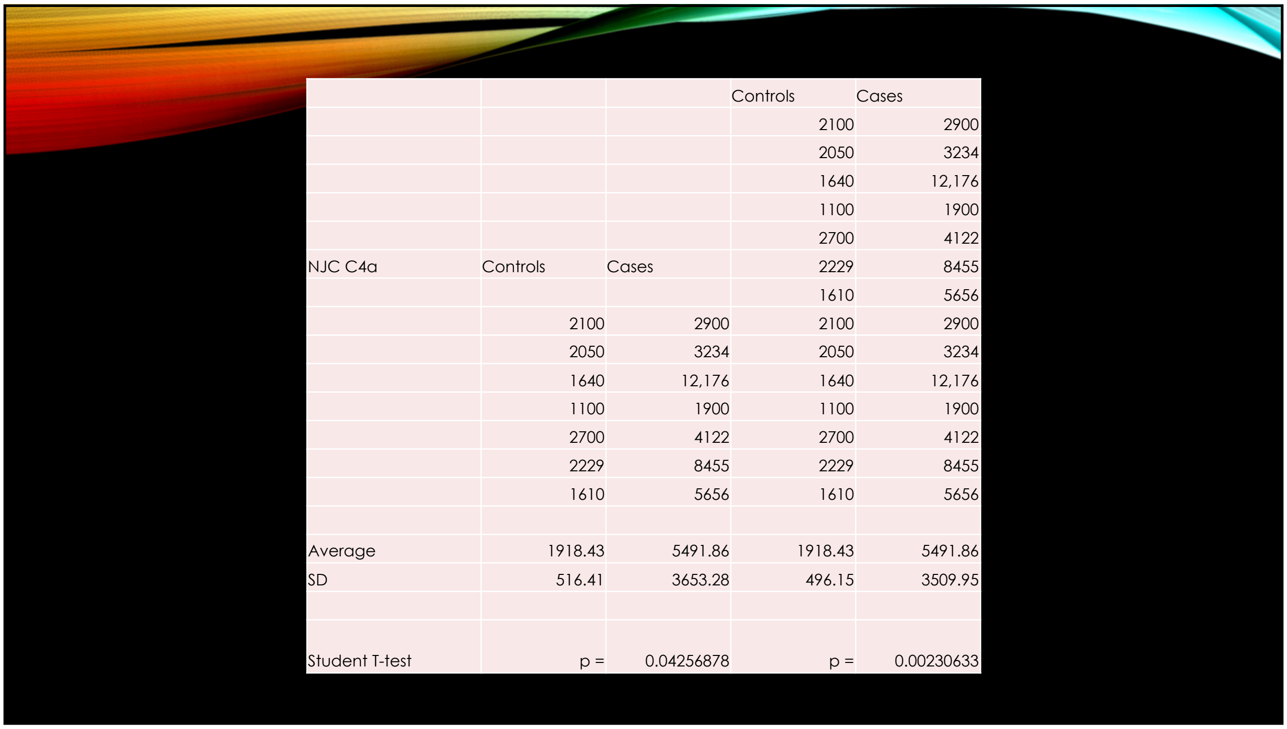
- Example 1 – Set up the Hypothesis test (cont'd)
  - 3) Find the critical value
    - Not necessary if using Excel
  - 4) Calculate the test statistic
    - Use Excel, this will give you the *p*-value
  - 5) Draw conclusions
    - If $p>=.05$, accept the H_0 that there is no relationship between having CIRS and elevated C4a values
    - If $p<0.05$, reject the H_0. This does not prove H_1 is correct but does prove H_0 is incorrect

28



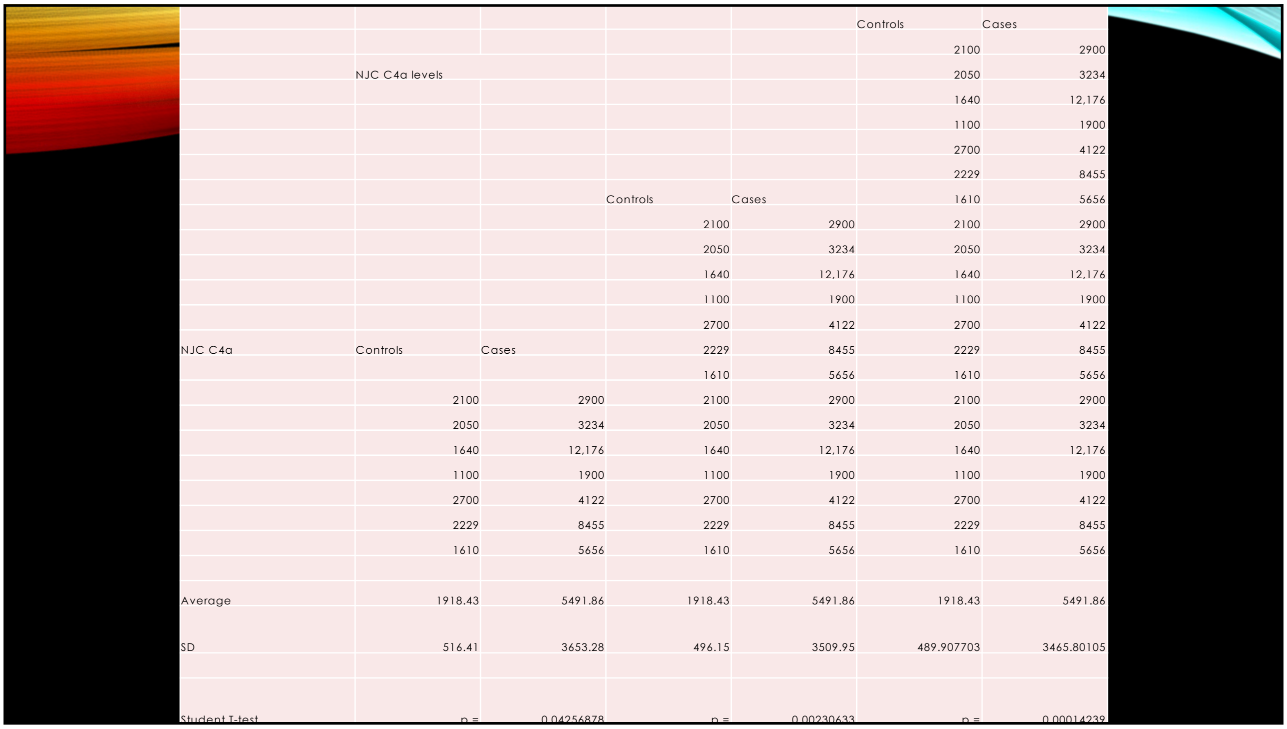
NJC C4a	Controls	Cases
	2100	2900
	2050	3234
	1640	12,176
	1100	1900
	2700	4122
	2229	8455
	1610	5656
Average	1918.43	5491.86
SD	516.41	3653.28
Student T-test	p =	0.04256878

29



			Controls	Cases
			2100	2900
			2050	3234
			1640	12,176
			1100	1900
			2700	4122
NJC C4a	Controls	Cases	2229	8455
			1610	5656
	2100	2900	2100	2900
	2050	3234	2050	3234
	1640	12,176	1640	12,176
	1100	1900	1100	1900
	2700	4122	2700	4122
	2229	8455	2229	8455
	1610	5656	1610	5656
Average	1918.43	5491.86	1918.43	5491.86
SD	516.41	3653.28	496.15	3509.95
Student T-test	p =	0.04256878	p =	0.00230633

30



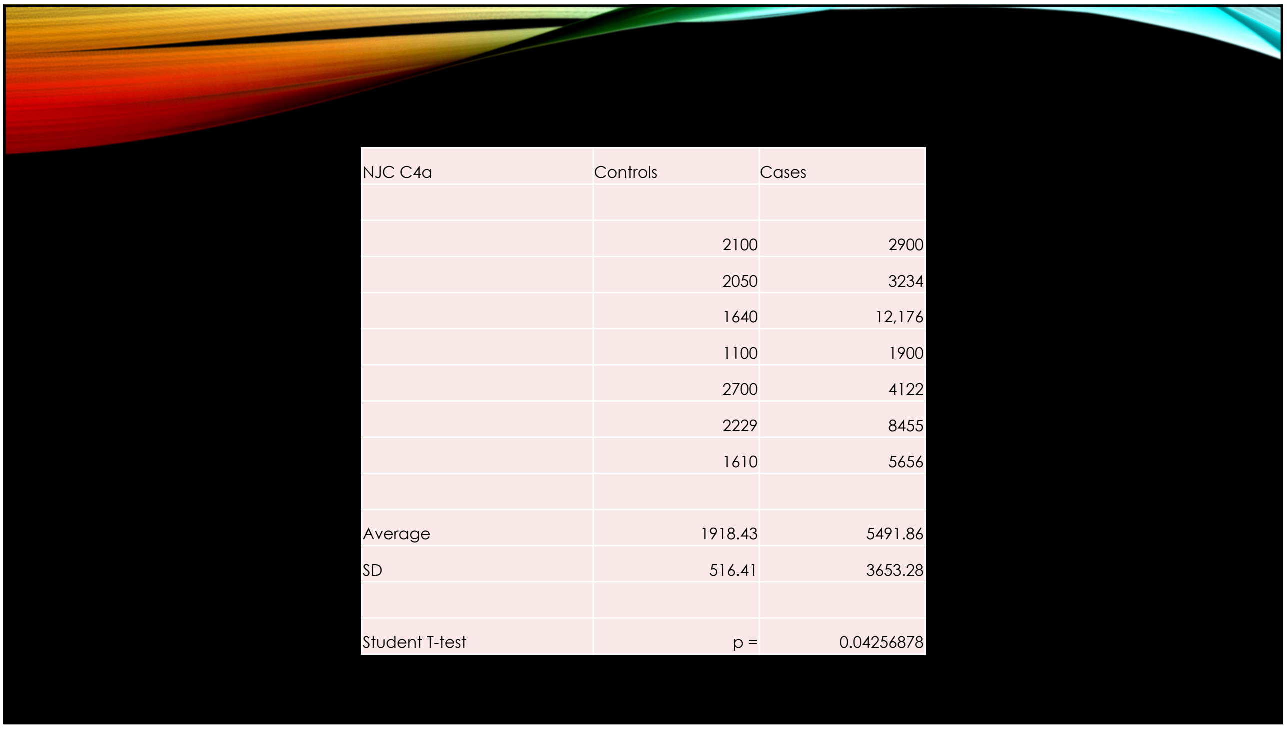
					Controls	Cases
					2100	2900
	NJC C4a levels				2050	3234
					1640	12,176
					1100	1900
					2700	4122
					2229	8455
			Controls	Cases	1610	5656
			2100	2900	2100	2900
			2050	3234	2050	3234
			1640	12,176	1640	12,176
			1100	1900	1100	1900
			2700	4122	2700	4122
NJC C4a	Controls	Cases	2229	8455	2229	8455
			1610	5656	1610	5656
	2100	2900	2100	2900	2100	2900
	2050	3234	2050	3234	2050	3234
	1640	12,176	1640	12,176	1640	12,176
	1100	1900	1100	1900	1100	1900
	2700	4122	2700	4122	2700	4122
	2229	8455	2229	8455	2229	8455
	1610	5656	1610	5656	1610	5656
Average	1918.43	5491.86	1918.43	5491.86	1918.43	5491.86
SD	516.41	3653.28	496.15	3509.95	489.907703	3465.80105
Student T-test	p =	0.04256878	p =	0.00230633	p =	0.00014239

31

p-VALUES

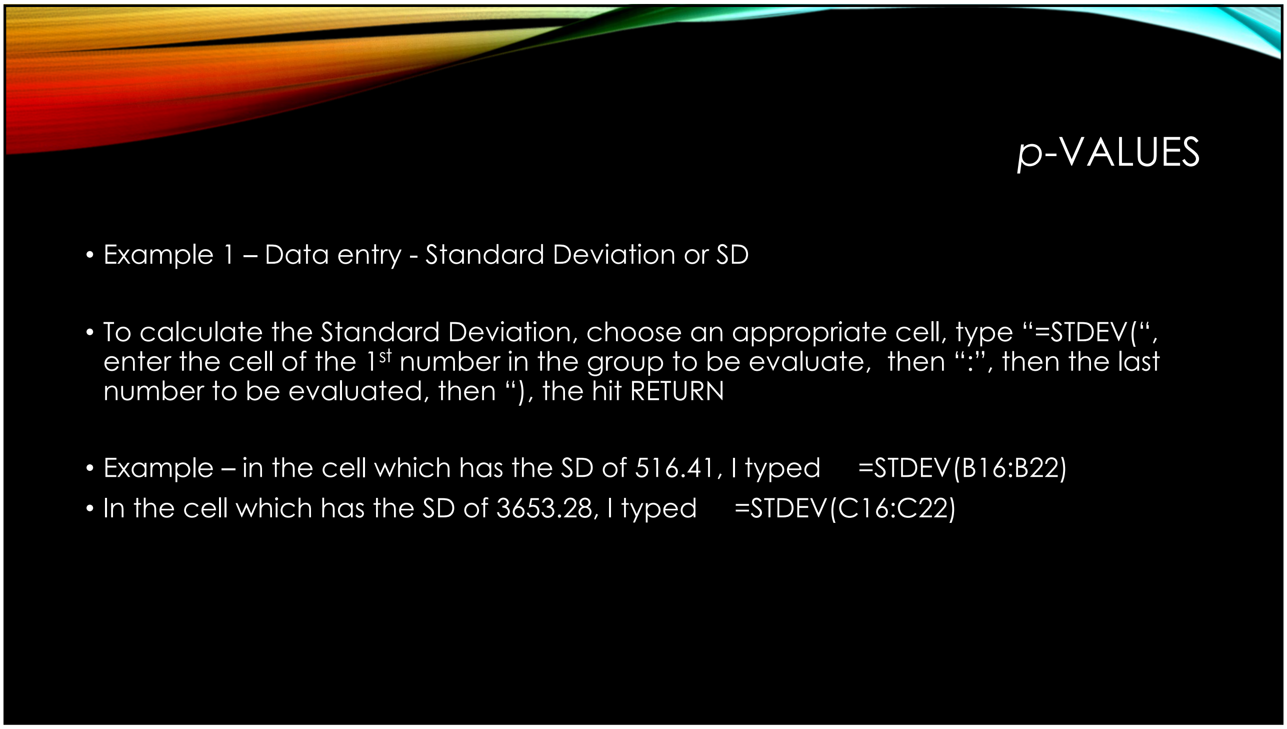
- Example 1 – Data entry – Average or Mean
- To calculate the Average, or Mean, choose an appropriate cell, type "=sum(", enter the cell of the 1st number in the group to be summed, then ":", then the last number to be summed, then ")/" then the number of data points. Then hit RETURN.
- Example – in the cell which has the Average of 1918.43, I typed =sum(B16:B22)/7
- In the cell which has the Average of 5491.86, I typed =sum(C16:C22)/7
- There are other ways to do this, for me, this is the easiest

32



NJC C4a	Controls	Cases
	2100	2900
	2050	3234
	1640	12,176
	1100	1900
	2700	4122
	2229	8455
	1610	5656
Average	1918.43	5491.86
SD	516.41	3653.28
Student T-test	p =	0.04256878

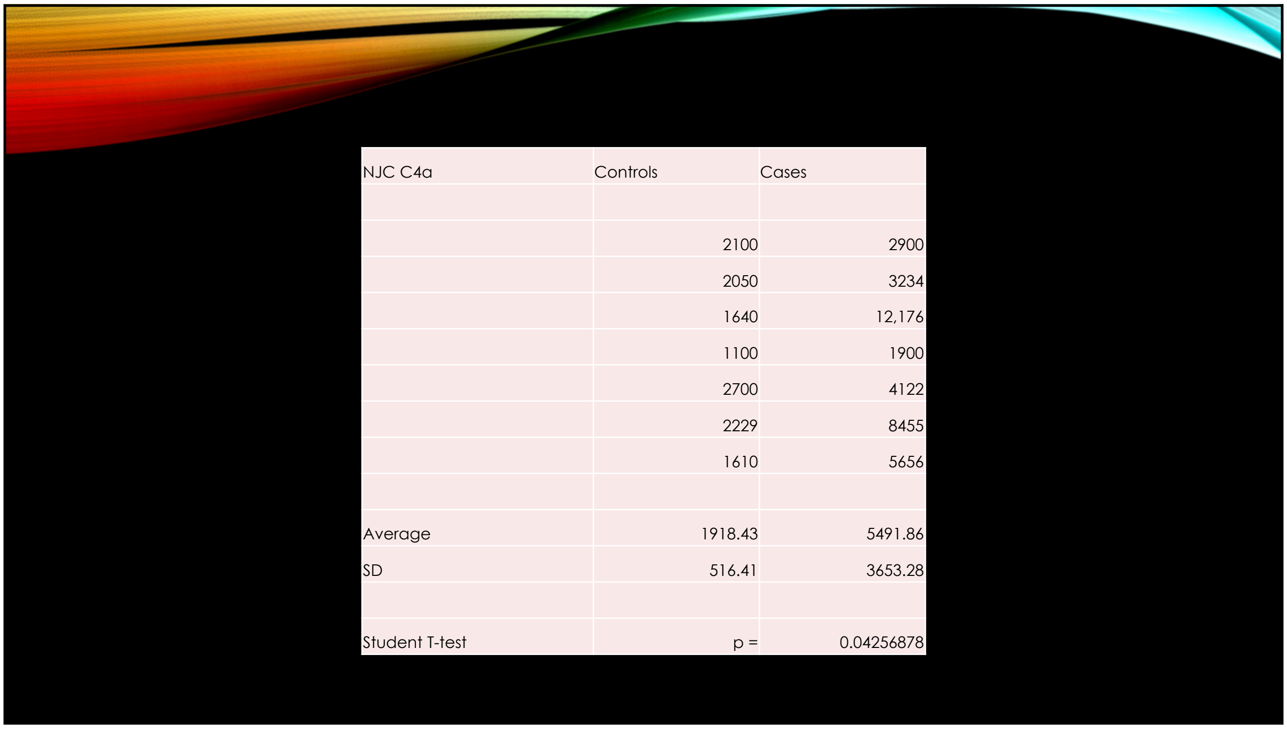
33



p-VALUES

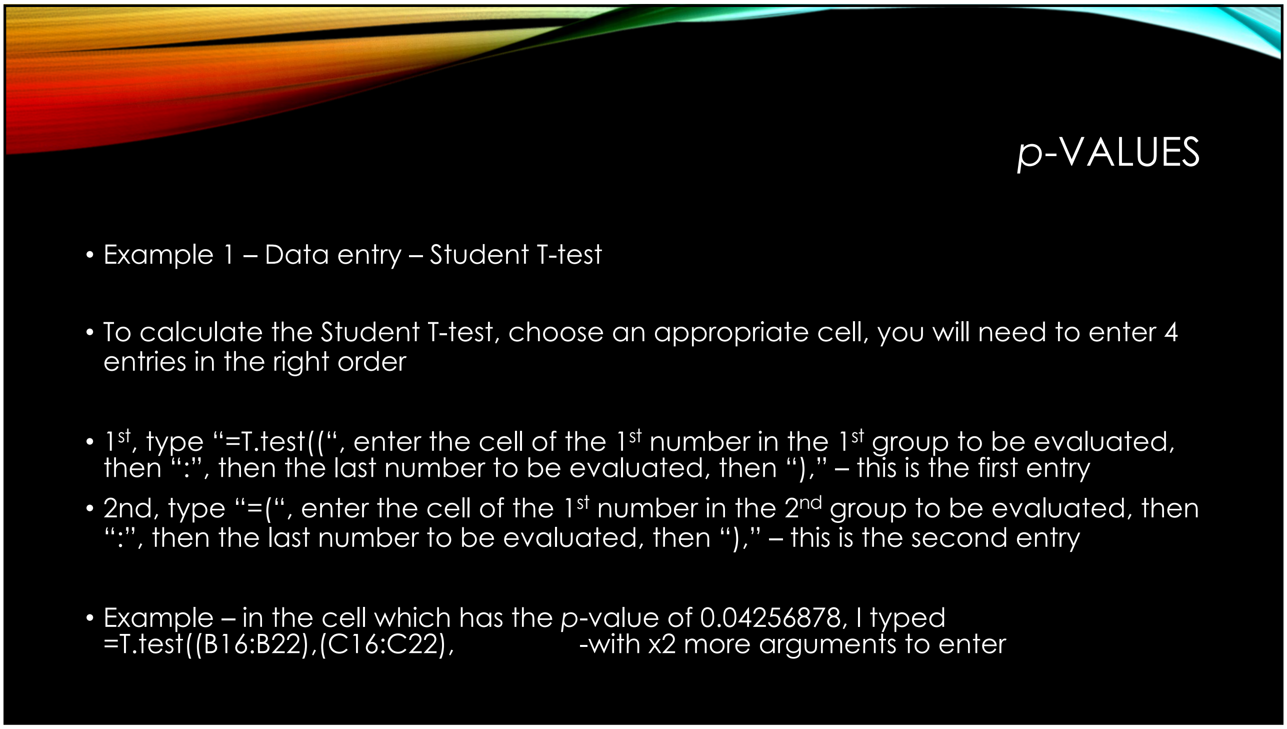
- Example 1 – Data entry - Standard Deviation or SD
- To calculate the Standard Deviation, choose an appropriate cell, type "=STDEV(", enter the cell of the 1st number in the group to be evaluate, then ":", then the last number to be evaluated, then "), the hit RETURN
- Example – in the cell which has the SD of 516.41, I typed =STDEV(B16:B22)
- In the cell which has the SD of 3653.28, I typed =STDEV(C16:C22)

34



NJC C4a	Controls	Cases
	2100	2900
	2050	3234
	1640	12,176
	1100	1900
	2700	4122
	2229	8455
	1610	5656
Average	1918.43	5491.86
SD	516.41	3653.28
Student T-test	p =	0.04256878

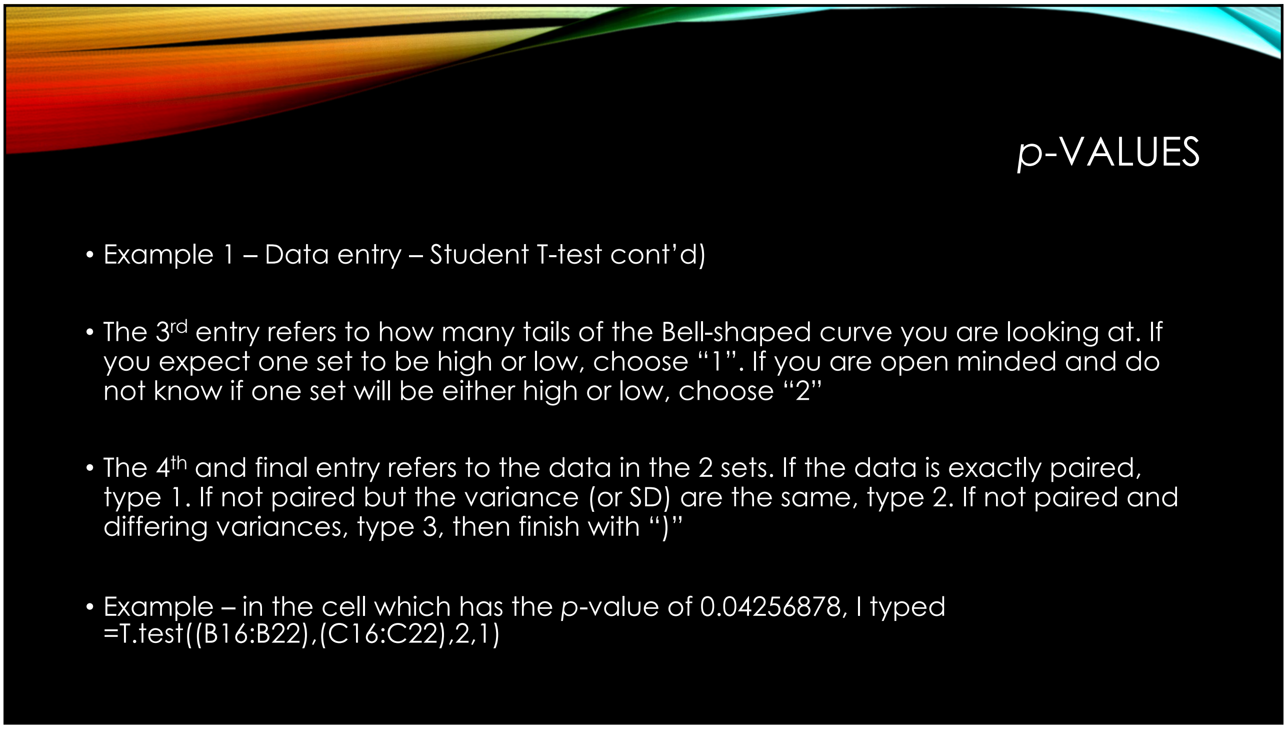
35



p-VALUES

- Example 1 – Data entry – Student T-test
- To calculate the Student T-test, choose an appropriate cell, you will need to enter 4 entries in the right order
- 1st, type "=T.test((", enter the cell of the 1st number in the 1st group to be evaluated, then ":", then the last number to be evaluated, then ")," – this is the first entry
- 2nd, type "=(", enter the cell of the 1st number in the 2nd group to be evaluated, then ":", then the last number to be evaluated, then ")," – this is the second entry
- Example – in the cell which has the *p*-value of 0.04256878, I typed =T.test((B16:B22),(C16:C22), -with x2 more arguments to enter

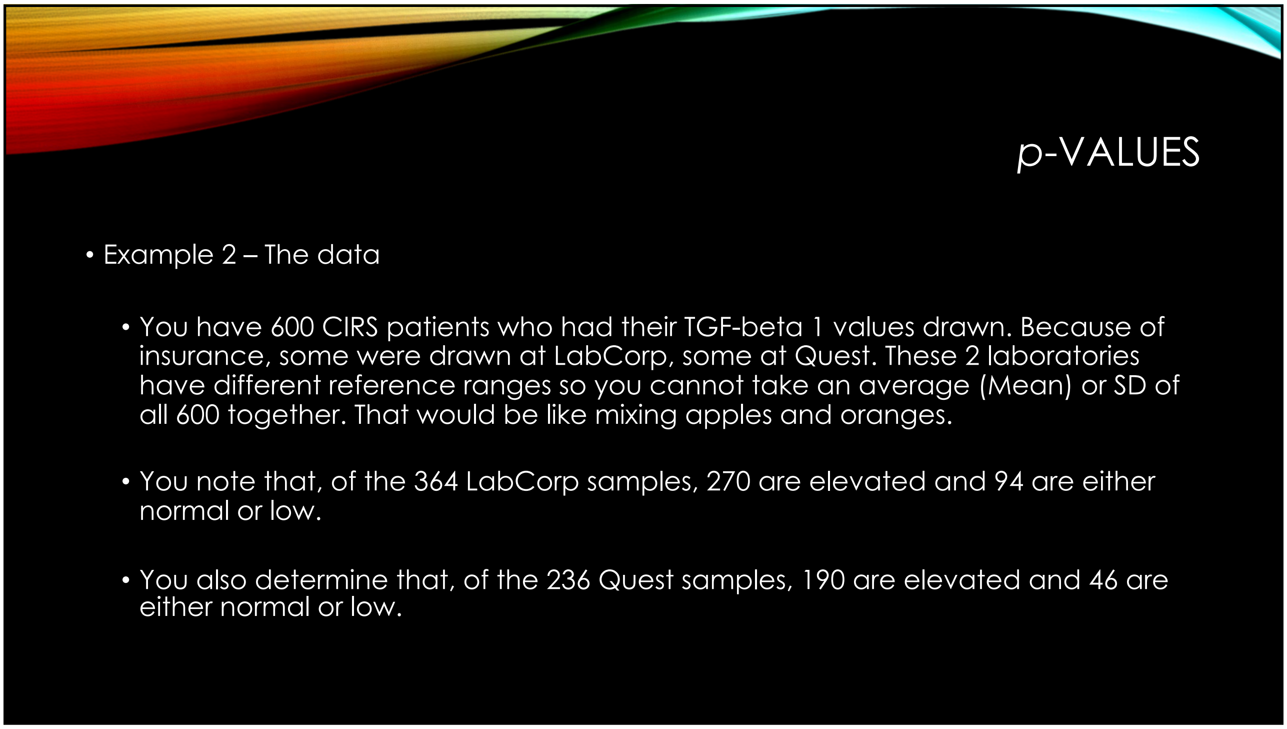
36



p-VALUES

- Example 1 – Data entry – Student T-test cont'd)
- The 3rd entry refers to how many tails of the Bell-shaped curve you are looking at. If you expect one set to be high or low, choose "1". If you are open minded and do not know if one set will be either high or low, choose "2"
- The 4th and final entry refers to the data in the 2 sets. If the data is exactly paired, type 1. If not paired but the variance (or SD) are the same, type 2. If not paired and differing variances, type 3, then finish with ")"
- Example – in the cell which has the *p*-value of 0.04256878, I typed =T.test((B16:B22),(C16:C22),2,1)

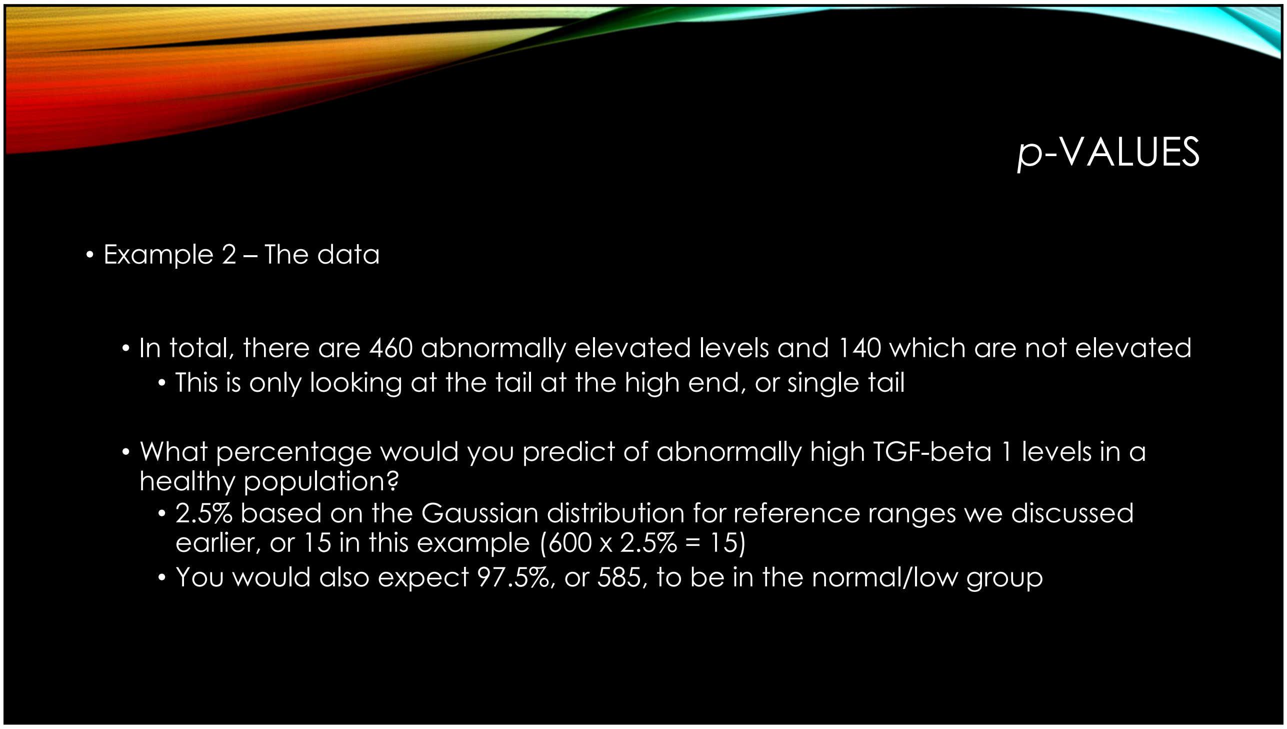
37



p-VALUES

- Example 2 – The data
  - You have 600 CIRS patients who had their TGF-beta 1 values drawn. Because of insurance, some were drawn at LabCorp, some at Quest. These 2 laboratories have different reference ranges so you cannot take an average (Mean) or SD of all 600 together. That would be like mixing apples and oranges.
  - You note that, of the 364 LabCorp samples, 270 are elevated and 94 are either normal or low.
  - You also determine that, of the 236 Quest samples, 190 are elevated and 46 are either normal or low.

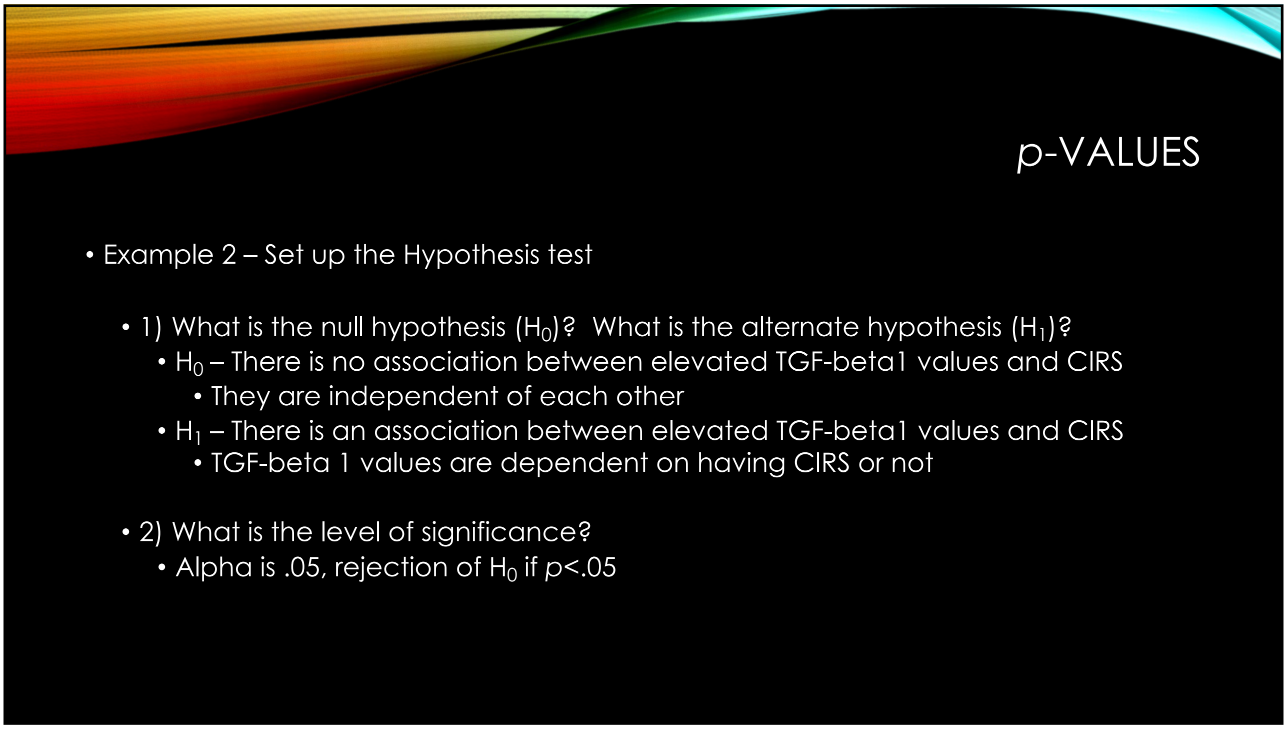
38



p-VALUES

- Example 2 – The data
  - In total, there are 460 abnormally elevated levels and 140 which are not elevated
    - This is only looking at the tail at the high end, or single tail
  - What percentage would you predict of abnormally high TGF-beta 1 levels in a healthy population?
    - 2.5% based on the Gaussian distribution for reference ranges we discussed earlier, or 15 in this example (600 x 2.5% = 15)
    - You would also expect 97.5%, or 585, to be in the normal/low group

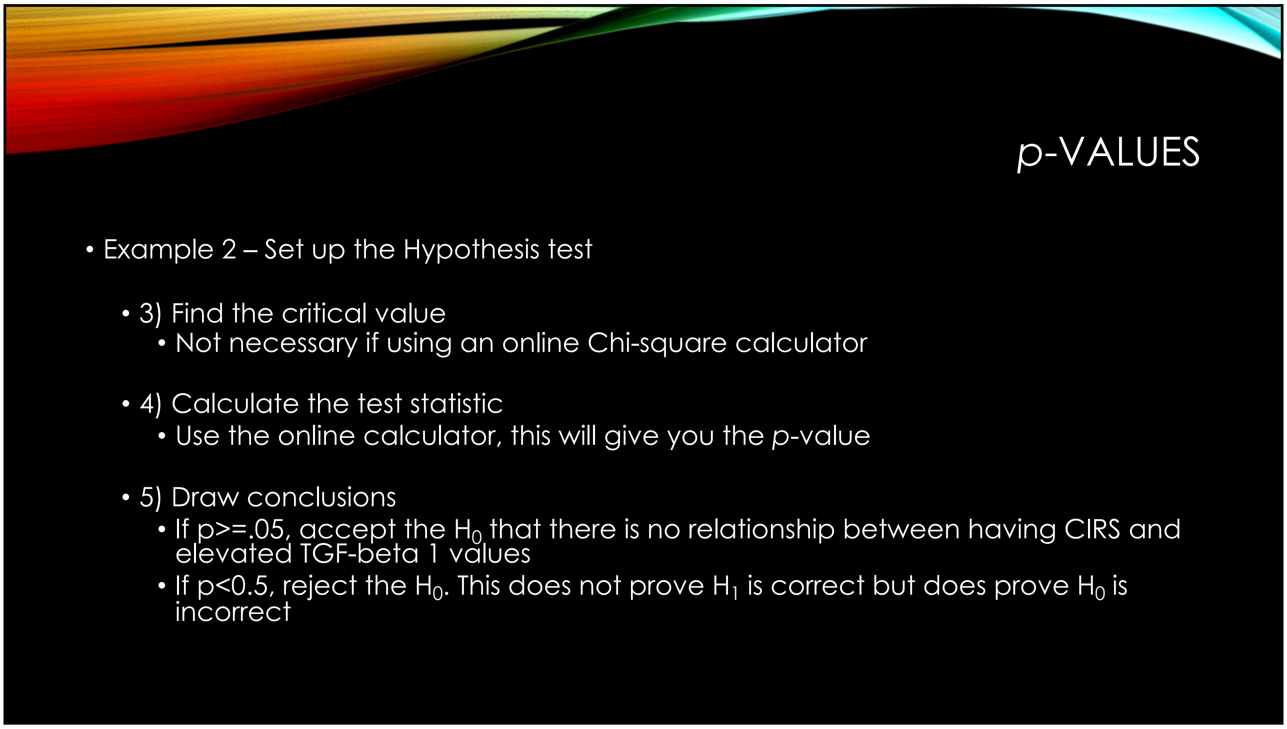
39



p-VALUES

- Example 2 – Set up the Hypothesis test
  - 1) What is the null hypothesis (H_0)? What is the alternate hypothesis (H_1)?
    - H_0 – There is no association between elevated TGF-beta1 values and CIRS
      - They are independent of each other
    - H_1 – There is an association between elevated TGF-beta1 values and CIRS
      - TGF-beta 1 values are dependent on having CIRS or not
  - 2) What is the level of significance?
    - Alpha is .05, rejection of H_0 if p<.05

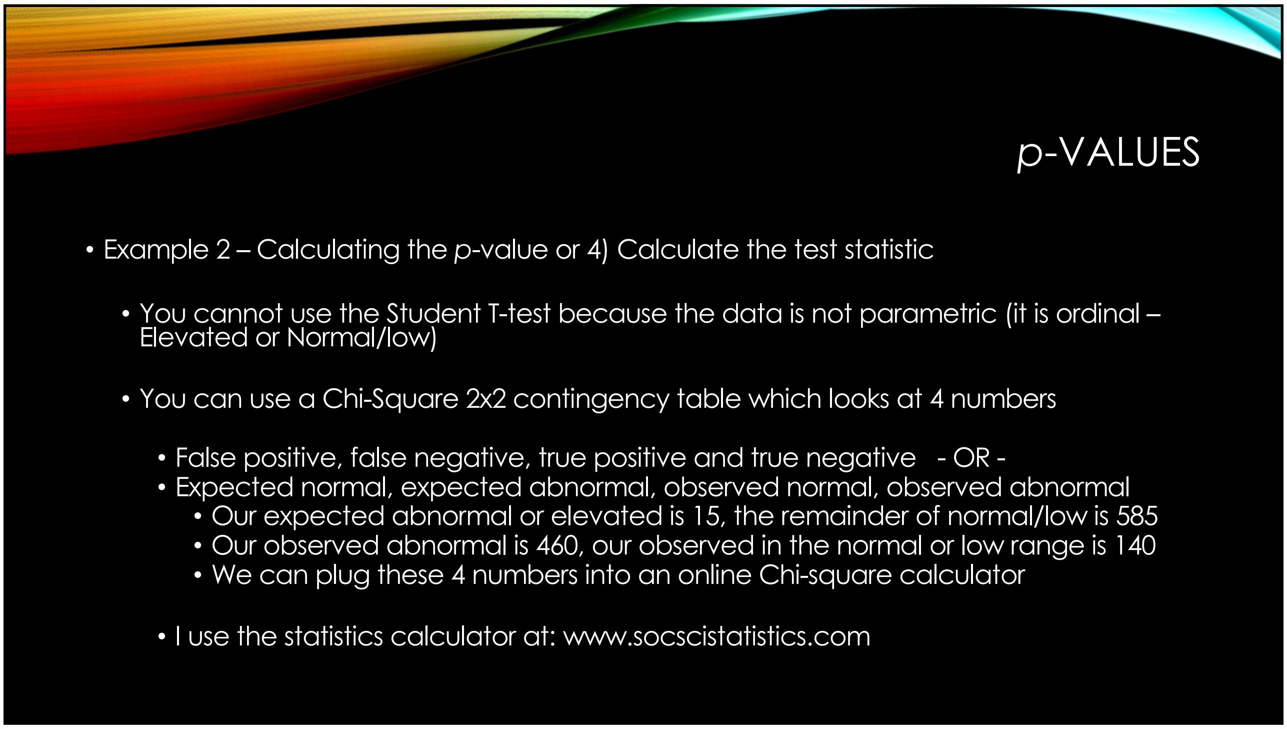
40



p-VALUES

- Example 2 – Set up the Hypothesis test
  - 3) Find the critical value
    - Not necessary if using an online Chi-square calculator
  - 4) Calculate the test statistic
    - Use the online calculator, this will give you the *p*-value
  - 5) Draw conclusions
    - If p>=.05, accept the H_0 that there is no relationship between having CIRS and elevated TGF-beta 1 values
    - If p<0.5, reject the H_0. This does not prove H_1 is correct but does prove H_0 is incorrect

41

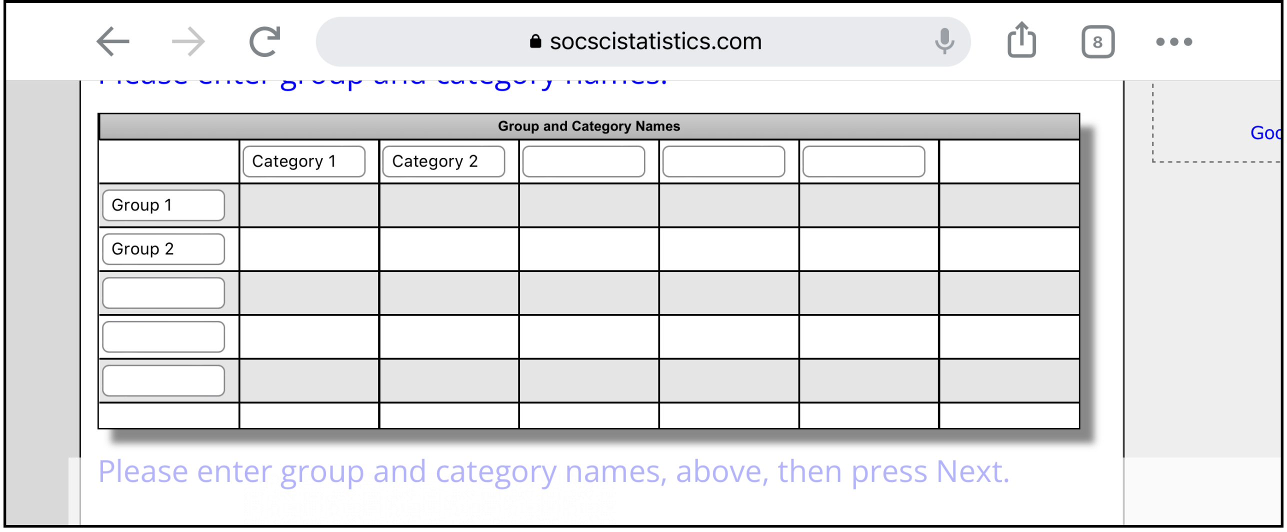


p-VALUES

- Example 2 – Calculating the *p*-value or 4) Calculate the test statistic
  - You cannot use the Student T-test because the data is not parametric (it is ordinal – Elevated or Normal/low)
  - You can use a Chi-Square 2x2 contingency table which looks at 4 numbers
    - False positive, false negative, true positive and true negative - OR -
    - Expected normal, expected abnormal, observed normal, observed abnormal
      - Our expected abnormal or elevated is 15, the remainder of normal/low is 585
      - Our observed abnormal is 460, our observed in the normal or low range is 140
      - We can plug these 4 numbers into an online Chi-square calculator
    - I use the statistics calculator at: www.socscistatistics.com

42

p-VALUES



socscistatistics.com

Please enter group and category names.

Goo

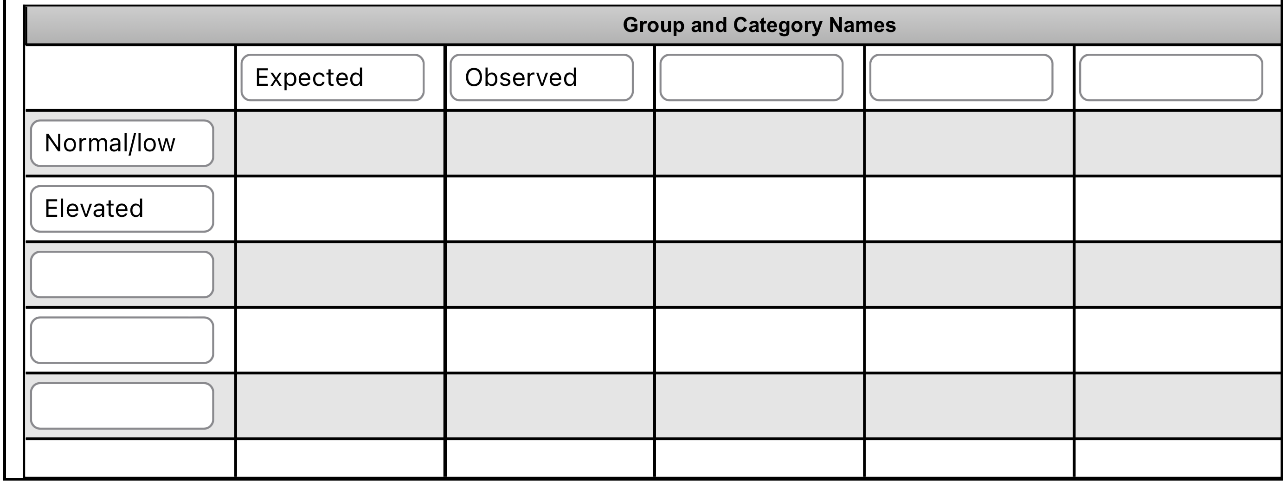
Group and Category Names

	Category 1	Category 2				
Group 1						
Group 2						

Please enter group and category names, above, then press Next.

43

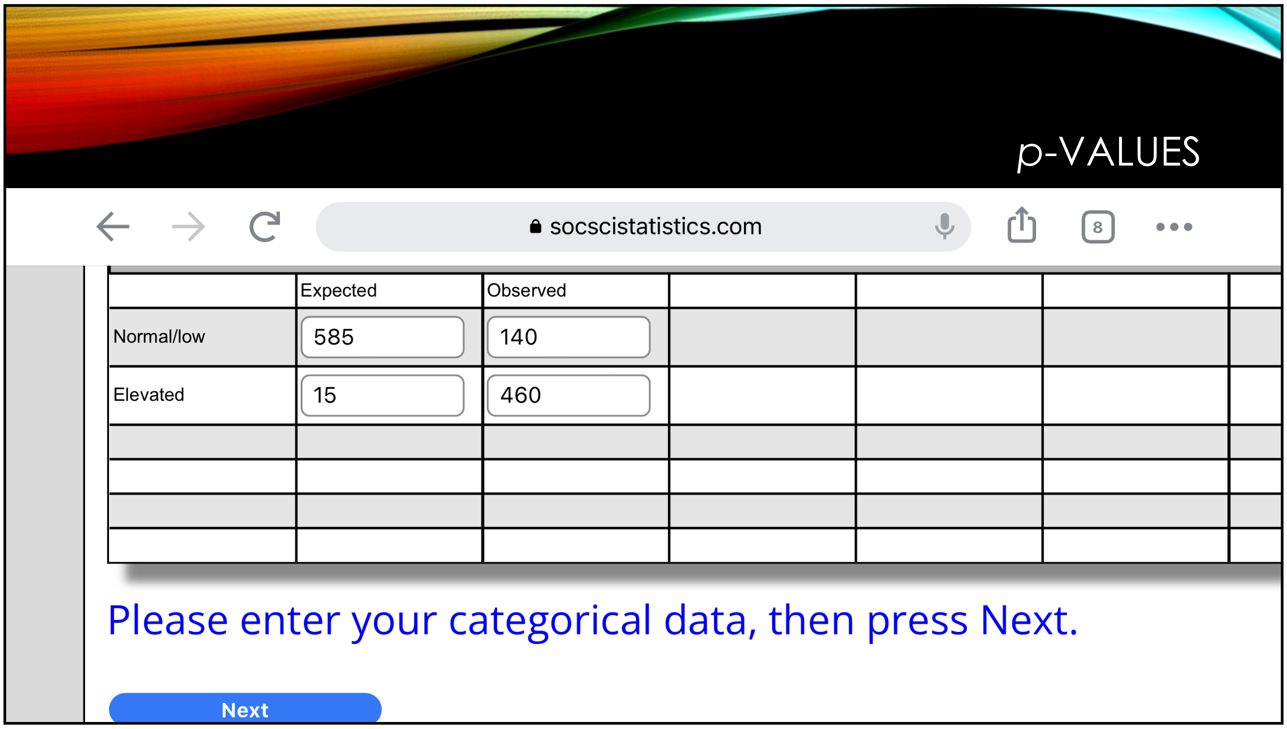
p-VALUES



Group and Category Names

	Expected	Observed			
Normal/low					
Elevated					

44

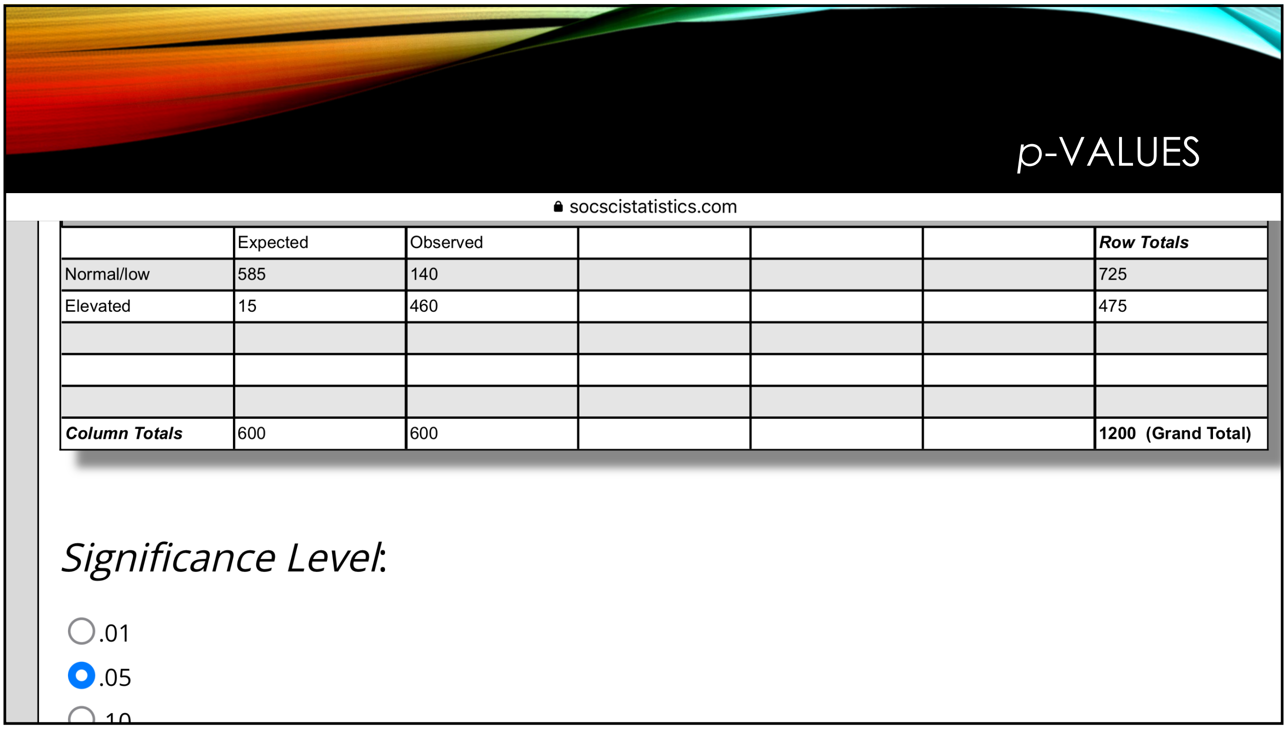


p-VALUES

	Expected	Observed				
Normal/low	585	140				
Elevated	15	460				

Please enter your categorical data, then press Next.

45



p-VALUES

	Expected	Observed				***Row Totals***
Normal/low	585	140				725
Elevated	15	460				475
Column Totals	600	600				**1200 (Grand Total)**

Significance Level:

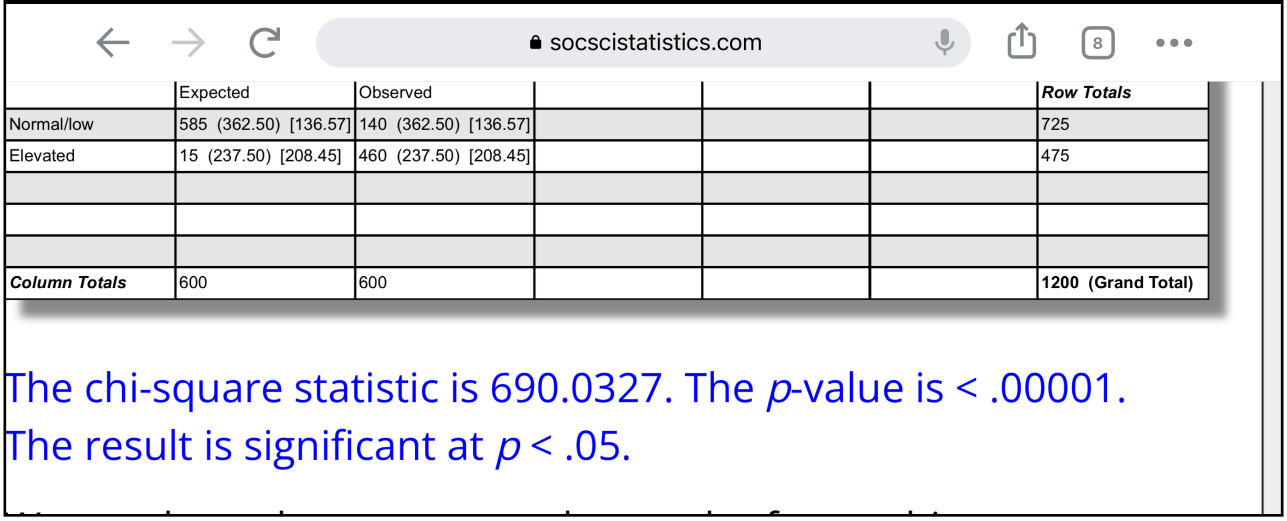
○ .01

● .05

○ .10

46

p-VALUES



	Expected	Observed				Row Totals
Normal/low	585 (362.50) [136.57]	140 (362.50) [136.57]				725
Elevated	15 (237.50) [208.45]	460 (237.50) [208.45]				475
Column Totals	600	600				**1200 (Grand Total)**

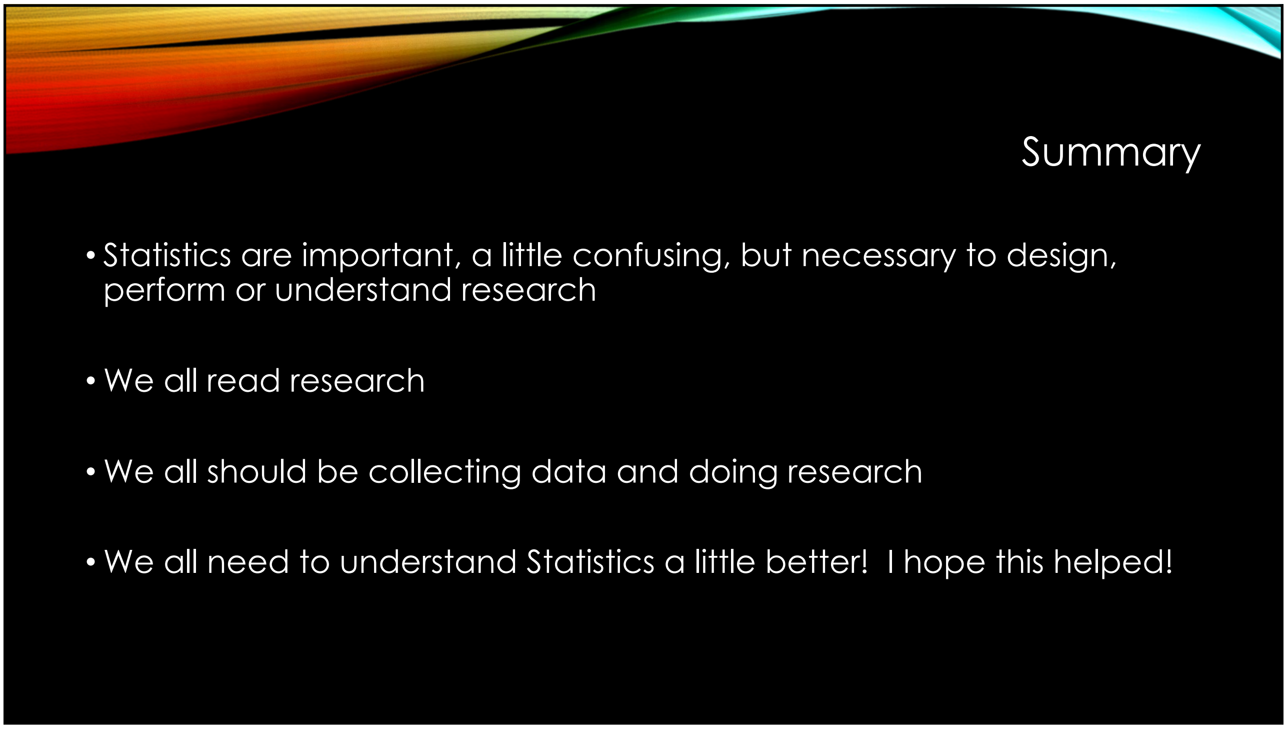
The chi-square statistic is 690.0327. The *p*-value is < .00001. The result is significant at $p < .05$.

47

p-values

- The Chi-square statistic calculated was 690.0327
- The *p*-value was <.00001 which is statistically significant when Alpha is set at .05
- We reject the null hypothesis which said the CIRS patients are no more likely than the normal population to have elevated TGF-beta1 levels
- Another way to look at the *p*-value is that for $p<.00001$, if we repeated the data collection 100,000 more times, we would likely have the same result (rejecting H_0) 99,999 times and a different result (accepting H_0) just 1 time.
- The difference between the 2 data groups is unlikely to be from chance

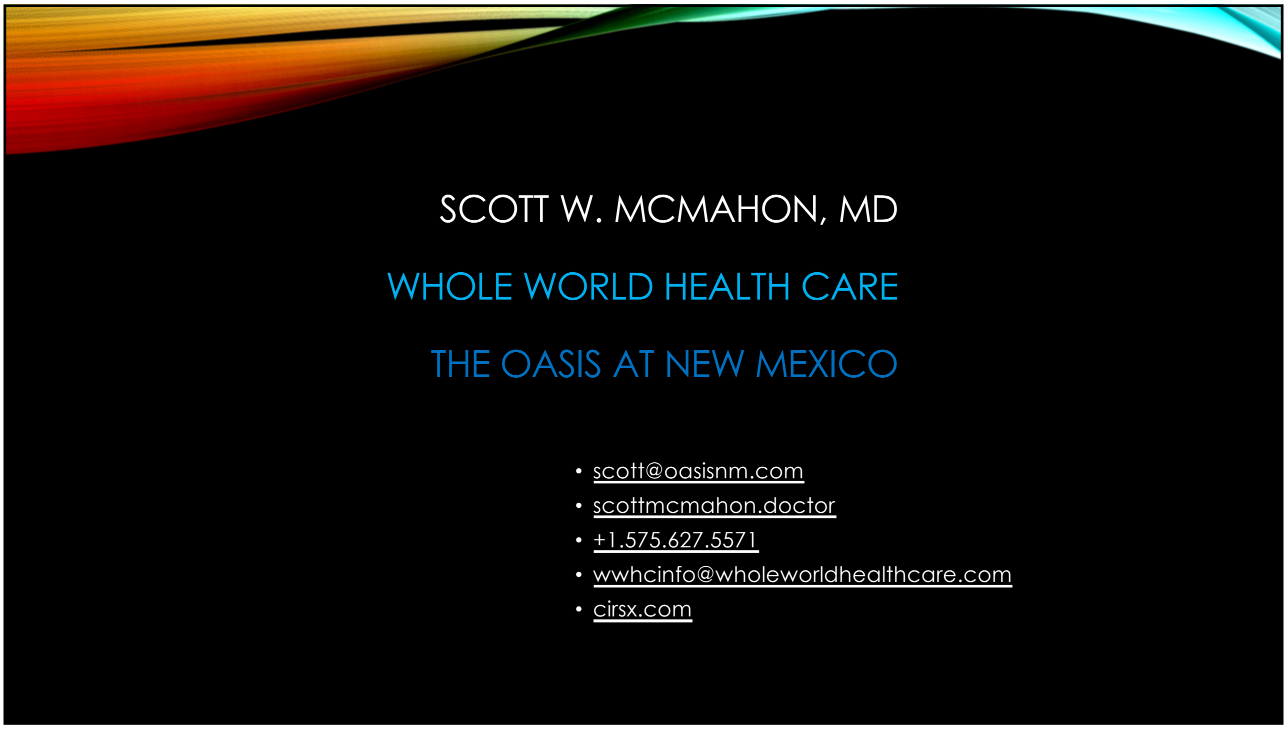
48



Summary

- Statistics are important, a little confusing, but necessary to design, perform or understand research
- We all read research
- We all should be collecting data and doing research
- We all need to understand Statistics a little better! I hope this helped!

49



SCOTT W. MCMAHON, MD

WHOLE WORLD HEALTH CARE

THE OASIS AT NEW MEXICO

- scott@oasisnm.com
- scottmcmahon.doctor
- +1.575.627.5571
- wwhcinfo@wholeworldhealthcare.com
- cirsx.com

50