

ELEVATING

EVIDENCE



Understanding Education Statistics

Office of Continuous Improvement and Support

Last Updated: April 2019

Objectives

- By the end of this webinar, participants will be able to...
 - recognize common statistical language and
 - interpret the results of statistical tests commonly used in education research.



Agenda

- Introduction and Alignment
- Descriptive Statistics
- Inferential Statistics
- Statistical Significance
- Magnitude of Findings



Introduction

Section 8101(21)(A) of the Every Student Succeeds Act defines evidence-based as “an activity, strategy, or intervention that demonstrates statistically significant effect on improving student outcomes or other relevant outcomes based on...

1. strong evidence from at least 1 well-designed and well-implemented experimental study;
2. moderate evidence from at least 1 well-designed and well-implemented quasi-experimental study; or
3. promising evidence from at least 1 well-designed and well-implemented correlational study with statistical controls for selection bias; or
4. demonstrating a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes and includes ongoing efforts to examine the effects of such activity, strategy, or intervention.”



Alignment

- ❑ [Code of Federal Regulations \(CFR\)](#)
- ❑ [Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments](#)
- ❑ [What Works Clearinghouse Standards Handbook v. 4.0](#)
- ❑ [What Works Clearinghouse Procedures Handbook v. 4.0](#)
- ❑ “Basic Statistics: Tales of Distribution” by Chris Spatz





Part One: Descriptive Statistics

Descriptive Statistics

- ❑ Mathematical tests that provide general information about a set of data.
- ❑ Common mathematical tests used daily by most educators.
- ❑ Provide a foundation for further data analysis.





Measures of Central Tendency

- ❑ Produce one number that describes a set of data
- ❑ Mean – The average of the data set
- ❑ Median – The middle number in a data set
- ❑ Mode – The number that occurs most often in a data set



Measures of Variability

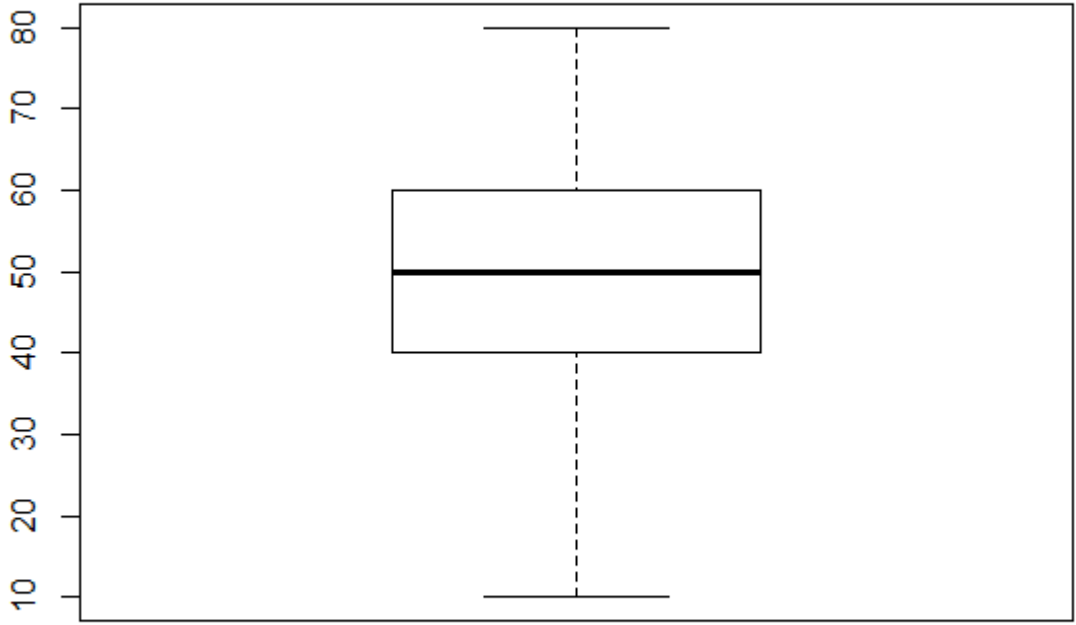
- ❑ Produces one number that describes the spread of a data set
- ❑ Helps educators understand the scope of the data
- ❑ Three main types:
 - ❑ Range
 - ❑ Interquartile Range
 - ❑ Standard Deviation



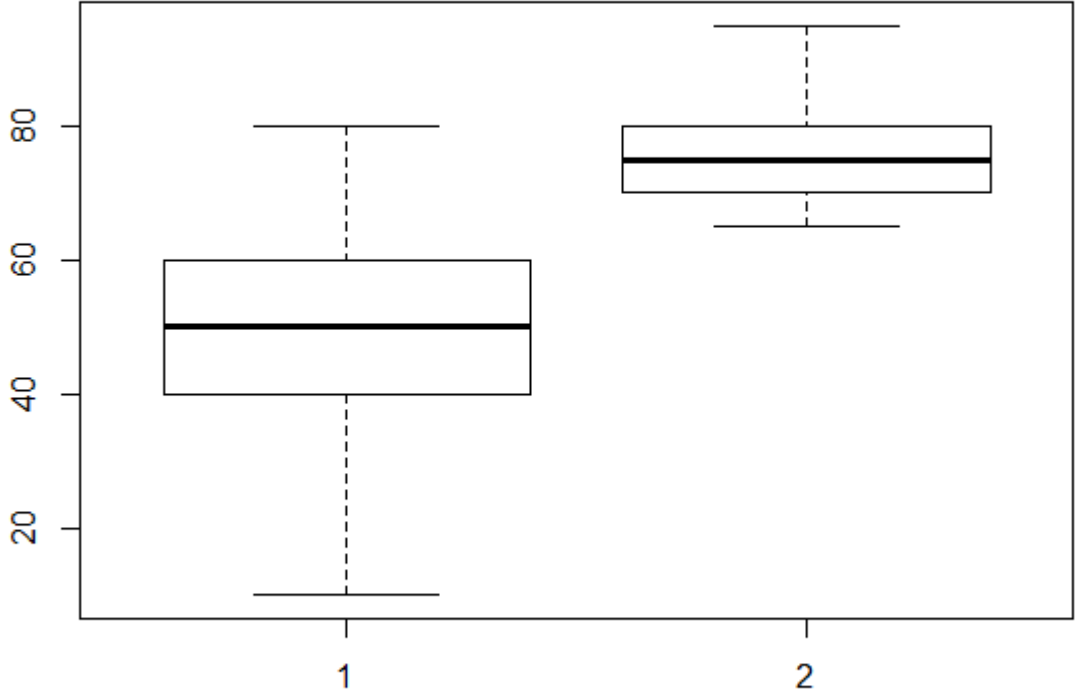
Range and Interquartile Range

- ❑ Range – the distance between the highest score and the lowest score in a data set
- ❑ Interquartile Range – the range of the middle fifty percent of a distribution
- ❑ Often represented together in a box and whisker plot

Box and Whisker Plot



Box and Whisker Plot Continued

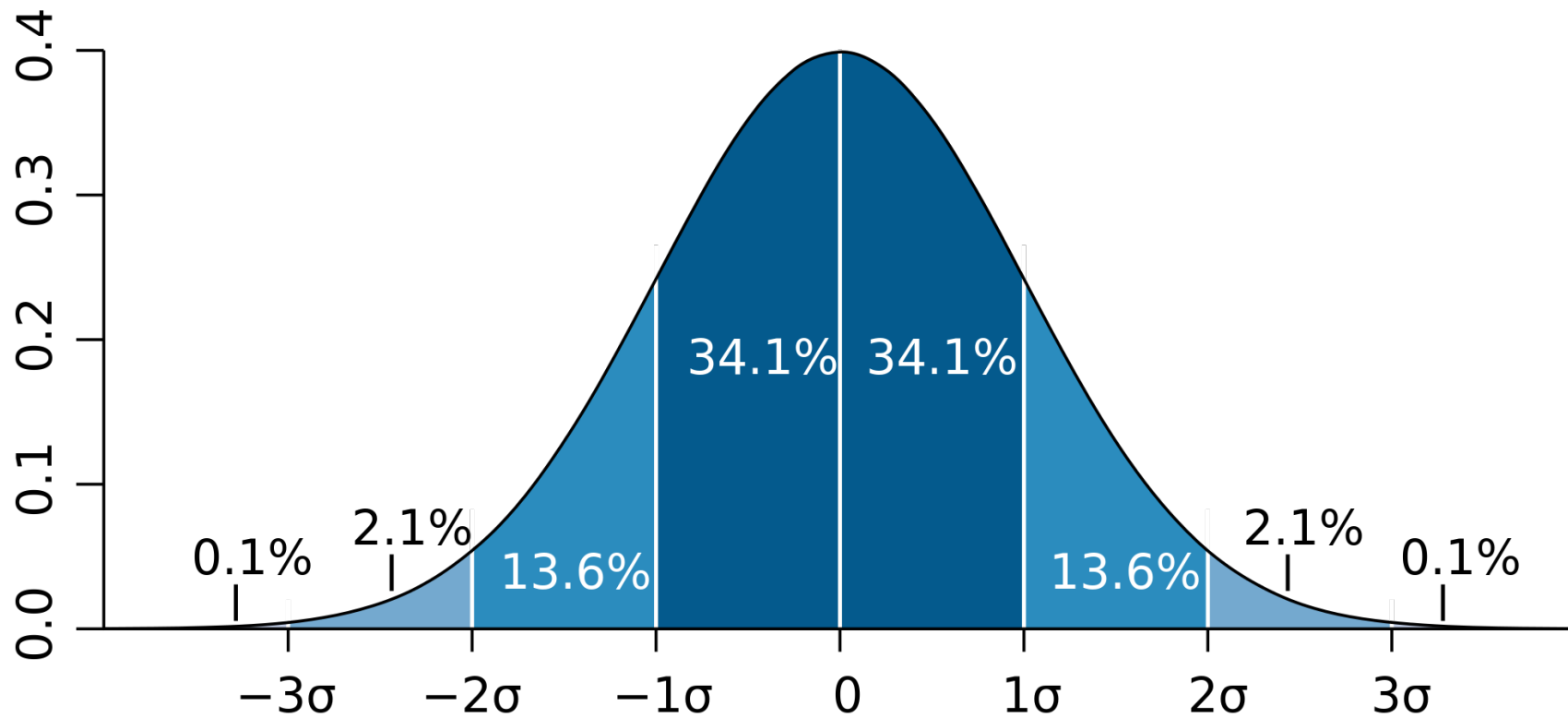




Standard Deviation

- ❑ Describes the dispersion of a set of scores around the mean
- ❑ A low standard deviation means that scores are closer together while a high standard deviation means that scores are farther apart.
- ❑ Typically represented in a bell curve graph

Standard Deviation Continued



Conclusion: Descriptive Statistics

- Descriptive statistics...
 - give us information about a data set,
 - are easy to calculate, and
 - inform instructional decisions.





Skill Check One



Question One

In the chart below you will find information reported from a set of student test scores. What do these measures tell you about the students who took the assessment?

Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Max
18%	25%	52%	54.85%	79.75%	100%



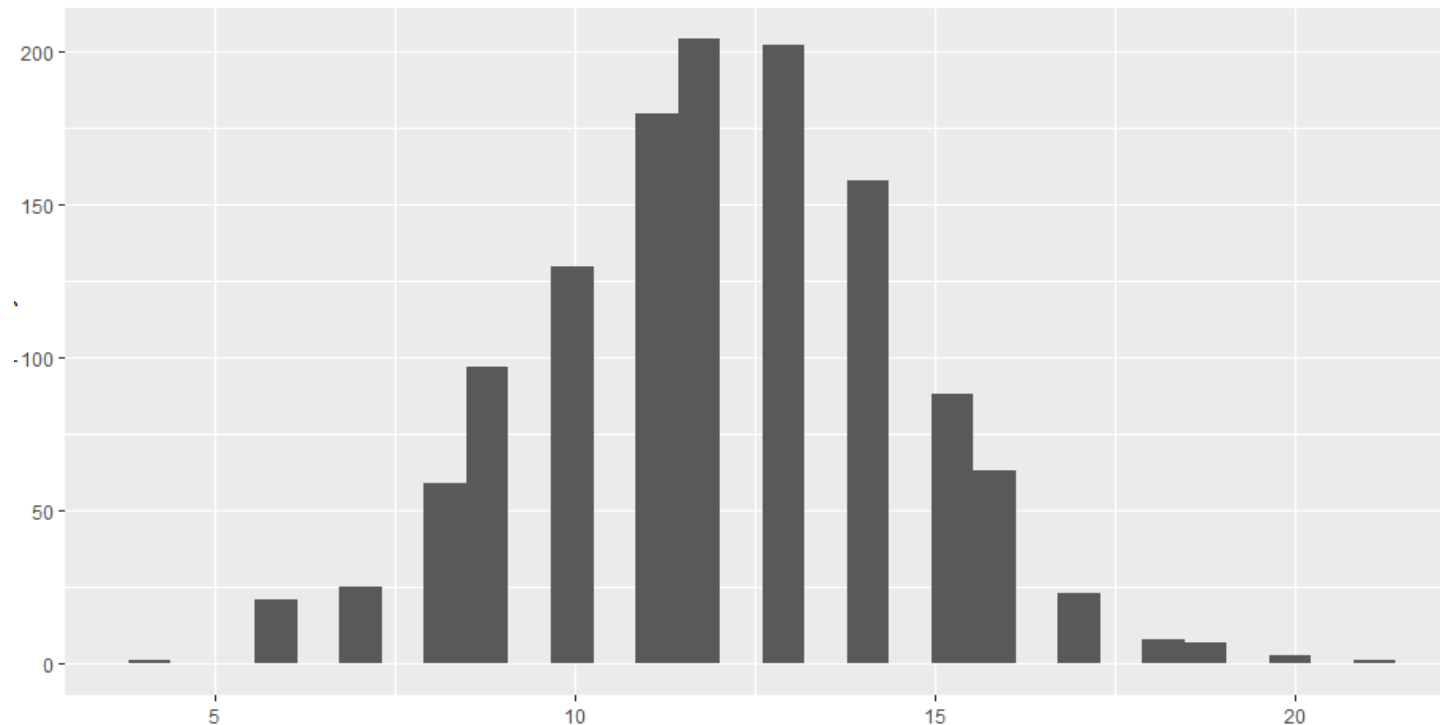
Question One: Discussion

- ❑ Student scores are fairly wide spread, meaning some students demonstrated mastery while others need more help.
- ❑ The median and mean score are pretty close together, this would suggest that the majority of student scores fell in the middle of the distribution.
- ❑ Half of the students in class scored between 25% and 79.75%.
- ❑ Some re-teaching is required before we move on.



Question Two

The histogram below shows the distribution of the average years of teaching experience in Kentucky public schools. What information can you learn from the distribution?





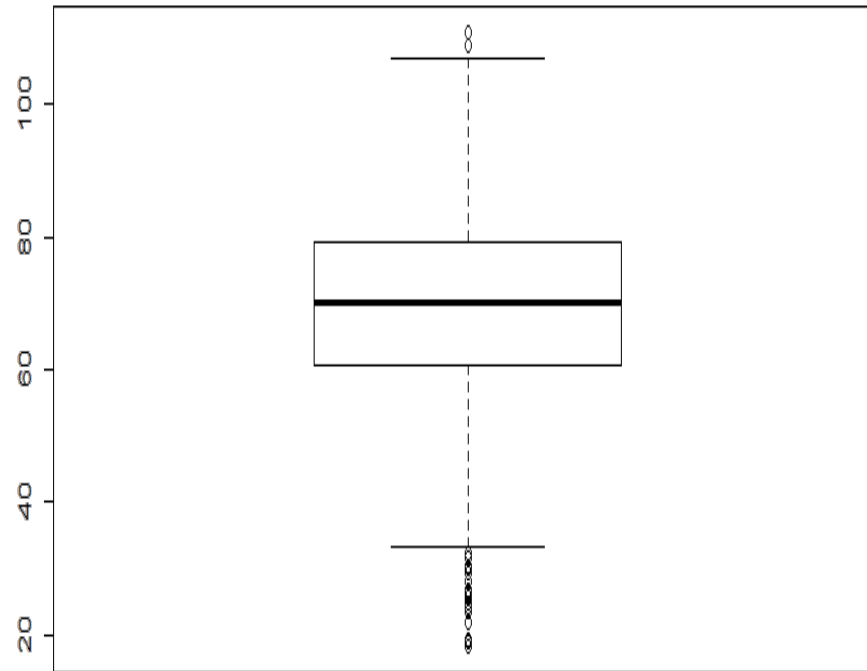
Question Two: Discussion

- ❑ The distribution follows a normal distribution pattern.
- ❑ The majority of schools in Kentucky have an average of 10-15 years of teaching experience.
- ❑ Teaching experience sharply declines after 15 years.



Question Three

The box and whisker plot to the right represents the distribution of proficiency rate in Kentucky's accountability system in the Fall of 2018. What insights can you glean from the plot?





Question Three: Discussion

- ❑ The median proficiency rate for Kentucky schools is around 70.
- ❑ Half of Kentucky's schools have a proficiency rate between 60 and 80, a range of 20 points.
- ❑ There are a handful of outliers who scored far above the majority of schools.



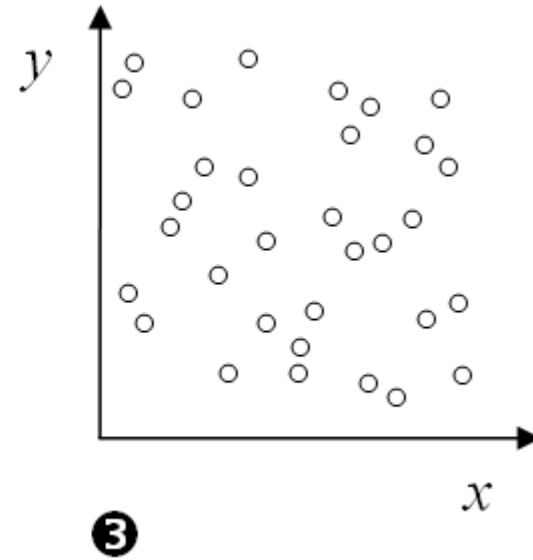
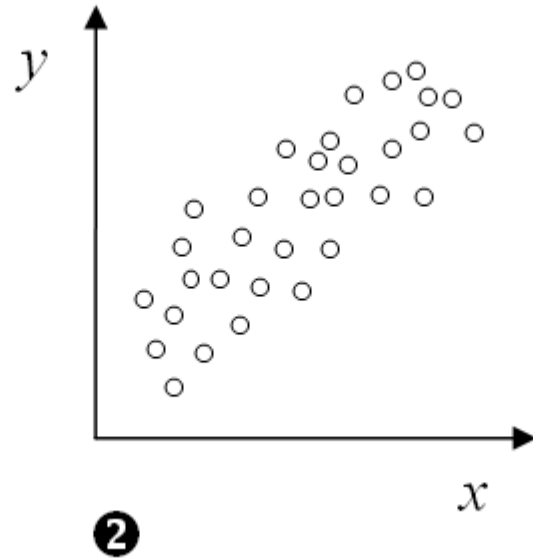
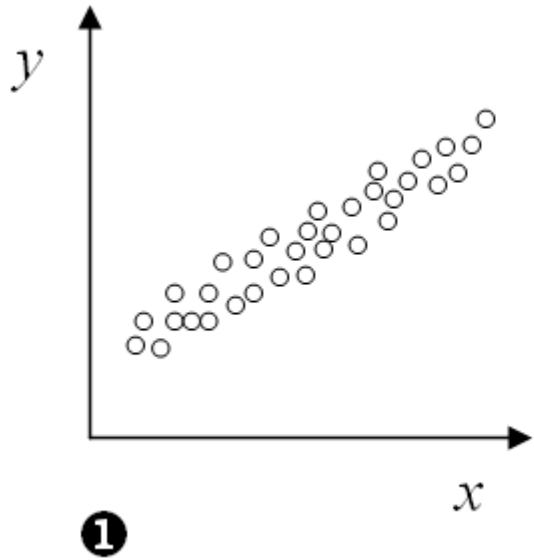
Part Two: Inferential Statistics



Correlation

- ❑ Describes the degree of relationship between two variables
- ❑ Commonly used to establish reliability in academic tests
- ❑ Represented by the Correlation Coefficient and reported with a lower-case r

Correlation Cont.



Correlation Continued

Correlation does not
equal causation.



t Test

- ❑ Used in hypothesis testing
- ❑ Multiple varieties
- ❑ Uses previously discussed statistics in its calculation
- ❑ Are reported as a lower-case t





Degrees of Freedom

- ❑ Measure of variability within the results of a statistical test
- ❑ Important for determining statistical significance
- ❑ Reported with a lower-cased *df*



Analysis of Variance

- ❑ Commonly called ANOVA
- ❑ Compares means across more than two variables
- ❑ Is reported as a capital F
- ❑ Common variations include the Multiple Analysis of Variance and Analysis of Covariance tests
- ❑ Almost always include post-hoc tests



Chi Square Test

- ❑ Compares expected outcomes to actual outcomes

Table One		
	Pizza	Ice Cream
Male	50%	50%
Female	50%	50%

Table Two		
	Pizza	Ice Cream
Male	10%	90%
Female	85%	15%

Statistical Significance and Magnitude of Findings





Statistical Significance

- ❑ Provides context to the results of a statistical test
- ❑ Reported with a lower-cased p
- ❑ Results are considered significant if the p -value is lower than 0.05.

Distribution Chart



The t Distribution

df	Two-Tailed		One-Tailed	
	0.05	0.01	0.05	0.01
1	12.706	63.656	6.314	31.821
2	4.303	9.925	2.920	6.965
3	3.182	5.841	2.353	4.541
4	2.776	4.604	2.132	3.747
5	2.571	4.032	2.015	3.365
6	2.447	3.707	1.943	3.143
7	2.365	3.499	1.895	2.998



Effect Size

- ❑ Provides context for the magnitude of a finding
- ❑ Reported as either a lower-cased d or g
- ❑ Considered substantively important when d or g is equal to or greater than 0.25

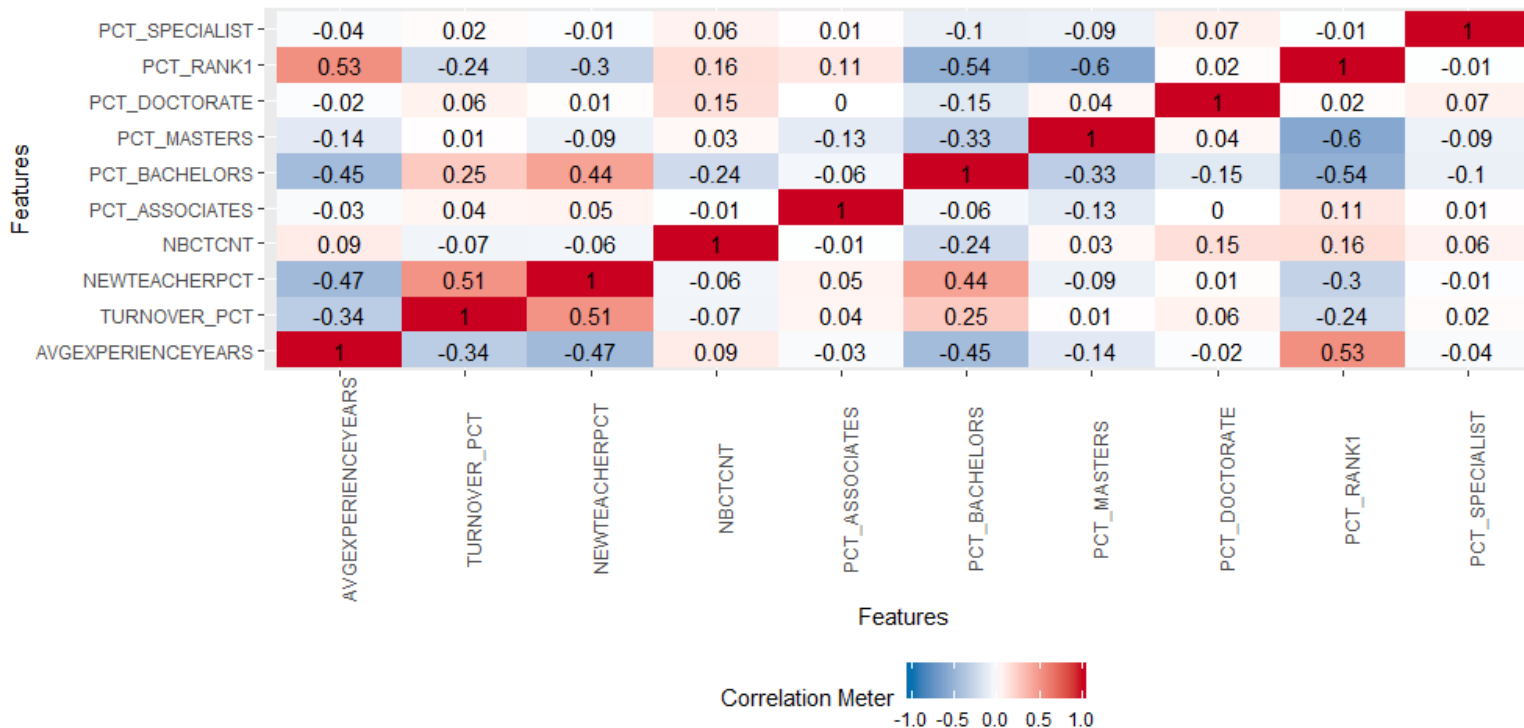


Skill Check Two



Question Four

The chart on the screen presents the correlation of various teacher characteristics in Kentucky public schools. Which traits are strongly correlated?





Question Four: Discussion

- ❑ There is a positive correlation between the percentage of new teachers in a school and the percentage of teacher turnover in Kentucky schools.
- ❑ There is a positive correlation between the average years of teaching experience and the percentage of teachers with a Rank I.
- ❑ There is a slight, yet positive correlation between the percentage of teacher with a national board certification and the percentage of teachers with doctorate degrees.



Question Five

Below are the results of an analysis of variance test that compares the rate of emergency certified teachers in schools identified as TSI, CSI, and Other. Can you describe the outcome?

$$[F(2, 1267) = 2.022, p = 0.1333]$$



Question Five: Discussion

- ANOVA tests look for relationships between the means of multiple variables.
- Check the p value to look for significance.
- There is no significant relationship between the rate of emergency certified teachers in schools identified as TSI, CSI, or Other.



Question Six

You are designing an intervention class for third graders who are falling behind in reading. You need to select a teaching strategy to deploy in the intervention class. Below are the effect sizes of several teaching strategies. Which strategies are you going to deploy in your class?

Microteaching: $d=0.88$	Student/Teacher Relationships: $d = 0.72$
Concept Matching: $d=0.57$	Worked Examples: $d=0.57$
Detailed Feedback: $d=0.73$	Frequent Testing: $d=0.34$
Peer Tutoring: $d=0.55$	Metacognitive Strategies: $d=0.69$



Question Six: Discussion

- Effect size should not be the sole descriptor used to inform your decision.
- All of the strategies have an effect size that is substantively important.
- Any combination of these strategies should help your struggling learners.

Wrap-up

- Topics Discussed:
 - Descriptive Statistics
 - Inferential Statistics
 - Statistical Significance
 - Magnitude of Findings





If you have questions regarding evidence-based practices or educational statistics, please contact the District 180 branch in the Office of Continuous Improvement and Support at (502) 564-2116.