

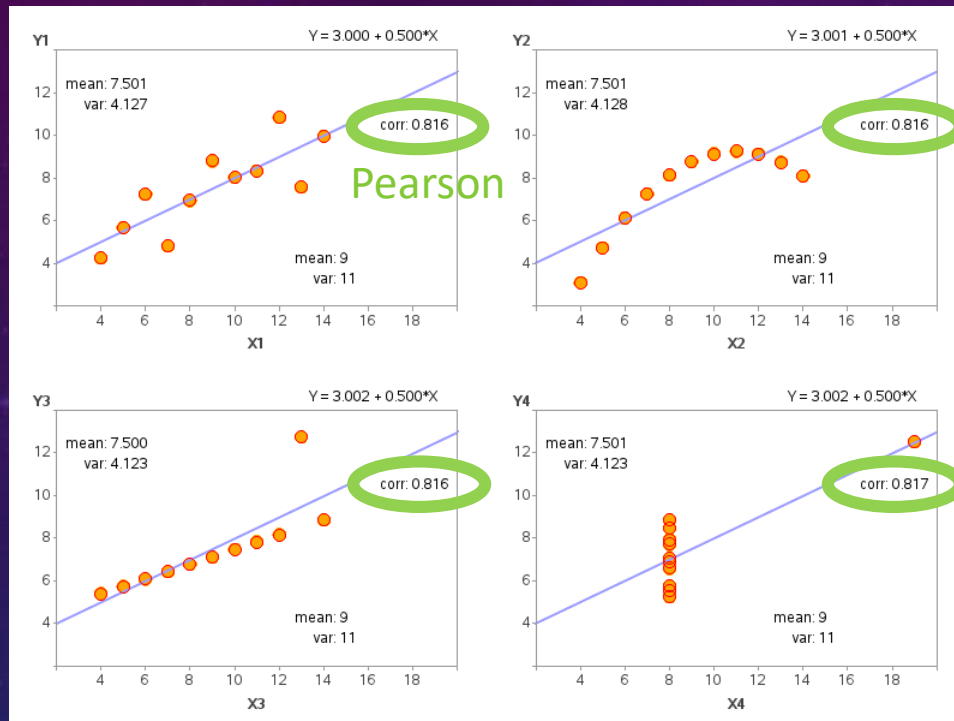
The background is a gradient of dark blue and purple, speckled with small white dots. On the left side, there are several concentric circular patterns. One large circle has a scale around its perimeter with numbers ranging from 150 to 260. Other smaller circles and arcs are scattered across the left and top-left areas, some with arrows indicating a clockwise direction. The overall aesthetic is technical and scientific.

# BASIC STATISTICS II

# CORRELATION TESTS ARE A SPECIAL CASE OF HYPOTHESIS TESTS

- Need not involve a model; may be “non-parametric”
- Return the probability of the null hypothesis:  
for correlation tests, the null hypothesis is that the two data sets have no association

# DOES IT MATTER WHICH TEST I CHOOSE?



## Anscombe's Quartet:

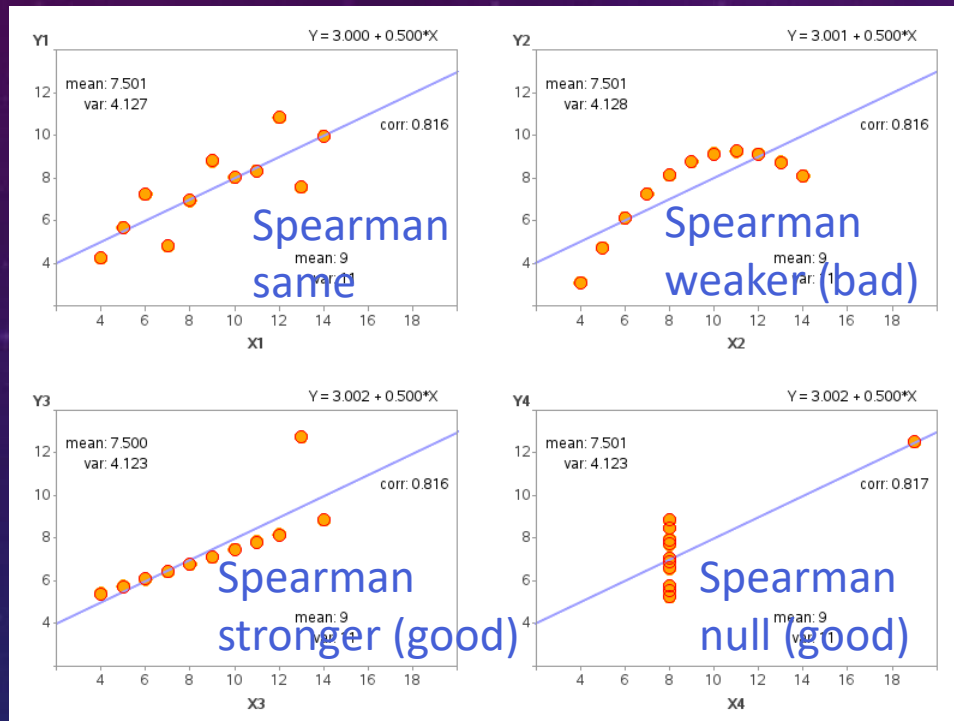
- 1) illustrates importance of graphical analysis
- 2) illustrates importance of choosing the right test based on its assumptions

## Most common correlation/association tests:

- **Pearson** (parametric): assumes linear relation + Gaussian scatter
- **Spearman rank** (non-parametric): assumes monotonic relation
- **Kendall's tau** (non-parametric): based on pairwise point analysis

# DOES IT MATTER WHICH TEST I CHOOSE?

## Spearman-Pearson comparison



What about Spearman vs. Kendall?

Spearman = most popular/widespread

Kendall's tau = better statistical properties per Ivezic et al. textbook

Add Kendall's tau to the plots and decide for yourself!

# ISSUES FOR CORRELATIONS:

- selection bias (luminosity vs. distance)
- covariance (color vs. mass)
- causality\* (correlation  $\neq$  causation)
- third parameters/partial correlations

*\* multi-parameter data sets are often analyzed using “principal component analysis” (PCA) to find the most fundamental driving parameters – however PCA is most effective for linear correlations, and “hidden” fundamental parameters may not be included in the data set*