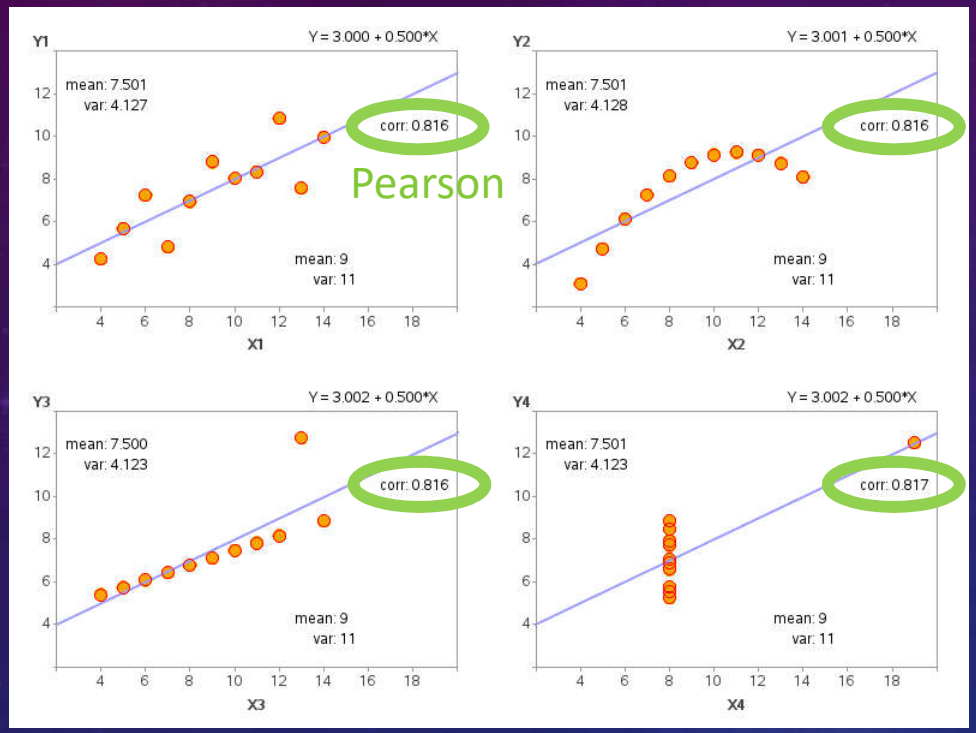
BASIC STATISTICS II

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?



Pearson

Anscombe's Quartet:

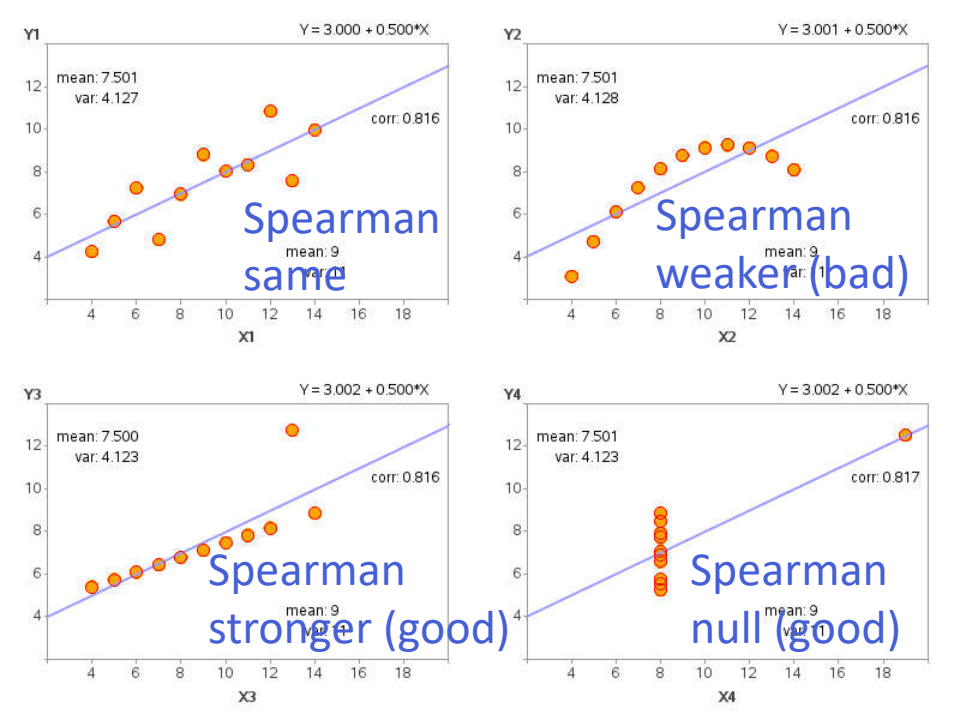
1) illustrates importance of graphical analysis
2) illustrates importance of choosing the right test based on its assumptions

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 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 ≠ causation)
- third parameters/partial correlations

** multi-parameter data sets are often analyzed using "principal component analysis" (PCA, Ivezic text Section 7.3) to find the most fundamental driving parameters – however PCA is most effective for linear correlations, and "hidden" fundamental parameters may not be in the data set*

ALSO THIS WEEK: MONTE CARLO METHODS

- tutorial in lab
- discuss inverse transform sampling method (Figs. 3.23-3.25 in Ivezic text)
- note use of “K-S test”, a hypothesis test discussed in Ch. 4 (null = two samples from same population)