Welcome back to "Name That Statistic"

Types of variables	Levels of Measurement				
Independent variable	Categorical (nominal or ordinal)	Categorical (two independent groups)	Categorical (two dependent groups)	Categorical (more than two groups)	Continuous (ratio or interval)
Dependent variable	Categorical (nominal or ordinal)	Continuous (Ratio or interval)	Continuous (Ratio or interval)	Continuous (Ratio or interval)	Continuous (ratio or interval)
Statistical procedure:	_____?	_____?	_____?	_____?	Correlation (or Pearson *r*)

What is “Correlation”?

- The Pearson *r* “correlation coefficient” is a summary statistic that indicates both the *strength* and *direction* of the relationship between two variables
- It has a value of between -1 and +1
  - Values less than zero (e.g -0.8) indicate a negative correlation
  - Values greater than zero (e.g. +0.8) indicate a positive correlation

Examples of correlational research hypotheses

- The number of outpatient therapy sessions utilized is positively correlated with the extent of depression as measured by the total score of the Beck Depression Inventory
- The 242 final exam score is positively correlated with the number of hours students spend preparing for the exam
- The number of behavioral incidents by children in residential care is negatively correlated with the number of strength-based supportive comments from staff

Compare to other statistical hypotheses…

- Chi-Square: “The variables are associated…”
- *t*-test: “The group means differ…”
- One-way ANOVA: “The group means differ…”
- Correlation: “There is a positive [or negative] correlation between the two variables.”
- (Preview to next week) Multiple Regression: “The independent variable is predictive of the dependent variable, controlling for…”

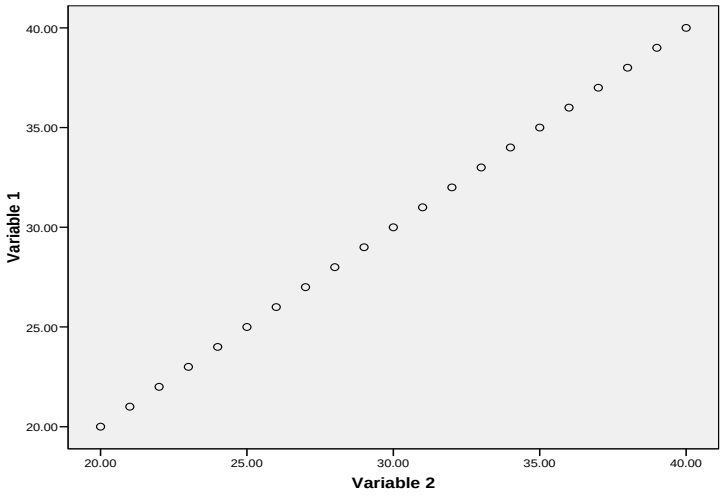
The Logic of Correlation

Subject #	Variable X	Variable Y	Correlation
1	X_1	Y_1	$r_{xy} = \frac{1}{n-1}\sum\left(\frac{x - mean_x}{sd_x}\right)\left(\frac{y - mean_y}{sd_y}\right)$
2	X_2	Y_2	
3	X_3	Y_3	
.	.	.	Where X and Y are interval or ratio variables, *n* is the sample size and *sd* is the variable's standard deviation
.	.	.	
.	.	.	
N	X_N	Y_N	$r\,(\mathrm{XY})$

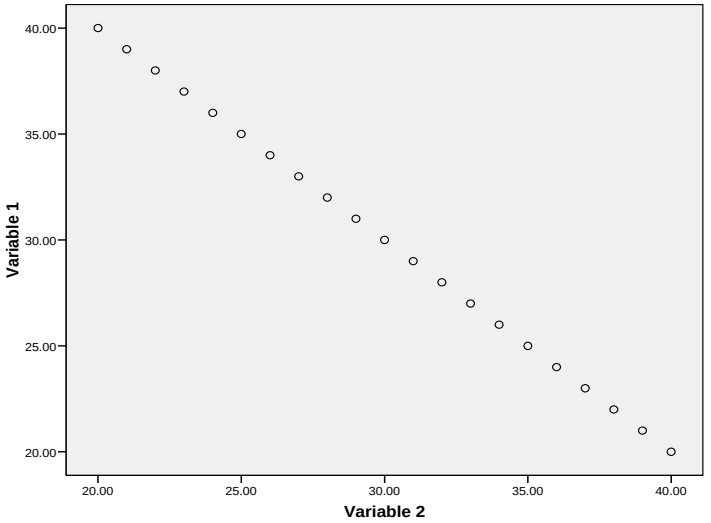
Scatter plots (a.k.a. scattergrams)

- Scatter plot: the graphic representation of the relationship between two ratio or interval variables, plotting the value of one variable against another with one dot
- Useful as a preliminary step in visually inspecting data
  - Seeing *strength and direction* of relationship
  - Seeing how *linearly* the variables are related
  - Seeing *outliers*

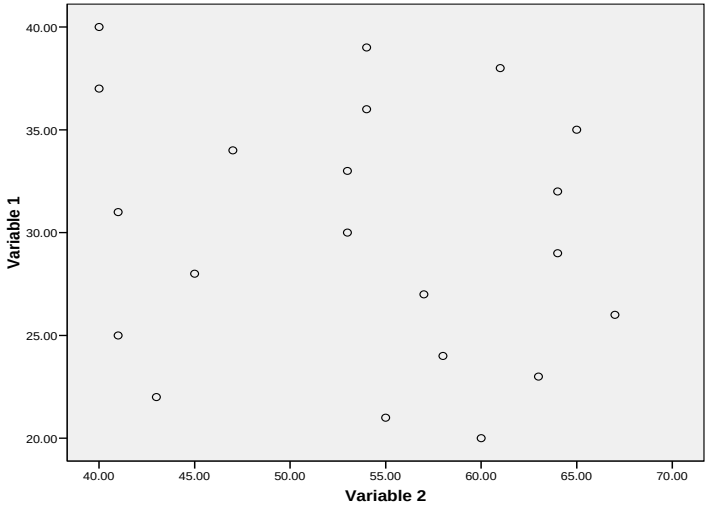
Perfect Positive Correlation:
$r = 1.0$



Perfect Negative Correlation:
$r = -1.0$

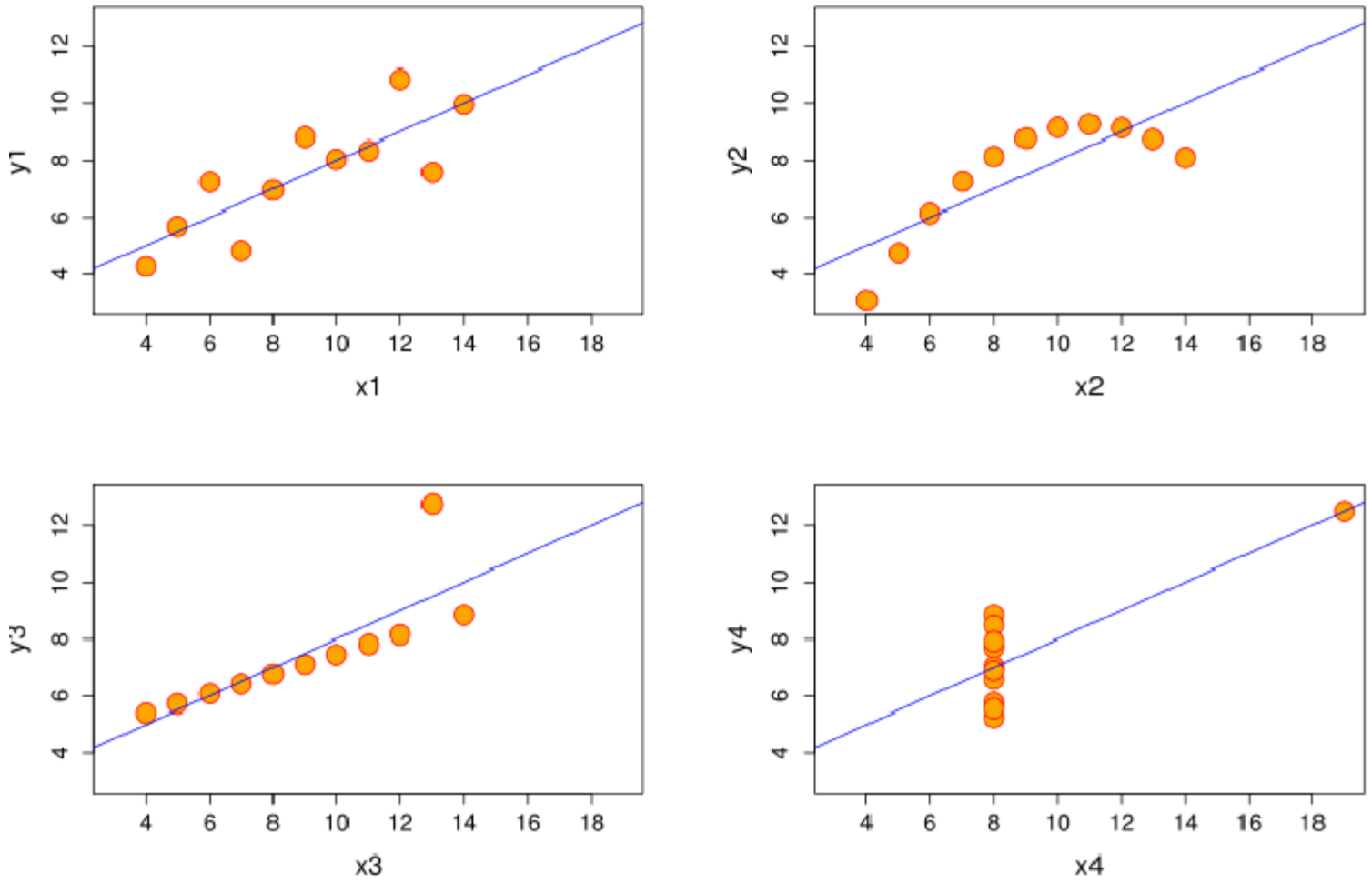


No Correlation:
$r \approx 0.0$



r = 0.81

But– these are very different distributions!



What are the requirements to use *r*?

- The Pearson *r* is a *parametric* statistic. Why?
  - The variables should be normally distributed in the population
    - For larger samples there can be some relaxation of this requirement
    - There are non-parametric tests for non-normally distributed variables, and those other than continuous
  - Also, the variables should be related *linearly*—either positively or negatively
    - Why is this important? (Hint: see previous slide)

Inference testing in correlation

- The *r* statistic can be located in a table of critical values
- The logic of inference testing is the same as other statistics:
  - If the *p* value given by SPSS is equal to or less than the alpha, then we reject the Null Hypothesis
  - We also need to interpret the correlation coefficient (*r)* and inspect the scatter plot:
    - Is it in the *same direction* as hypothesized?
    - Does the *strength* of the correlation support the alternate hypothesis?
    - Are the variables linearly related?

Can we imply causality from correlation?

- Remember the requirements for causality:
  1. Time ordering (The IV should precede the DV chronologically)
  2. Correlation between variables
  3. No other rival hypothesis (effect of 3rd variable)
- So—what's missing in correlation?