

Roadmap to Data Analysis

VI. Comparing Two Continuous Variables

Pearson's r Correlation

Learning Objectives

- Understand the type of data that are relevant for a correlation analysis
- Understand the basics of how correlation works
- Understand how to use a correlation calculator or Excel to conduct a correlation analysis with sample data
- Understand how to interpret results

Type of Data for Correlation Analysis

- Correlation tests the relationship between a continuous independent variable and a continuous dependent variable
- Continuous variables are those that increase or decrease steadily, rather than differ by categories. Correlation can answer, for example, the relationship between
 - Total score of the PTSD checklist and total score of the CES-D depression scale
 - Number of months in the US and weekly income
 - The number of available community clinicians and the number of clients in treatment
 - Number of months on an agency's caseload and total score of the Self-Sufficiency matrix

How does the Pearson's r Correlation work?

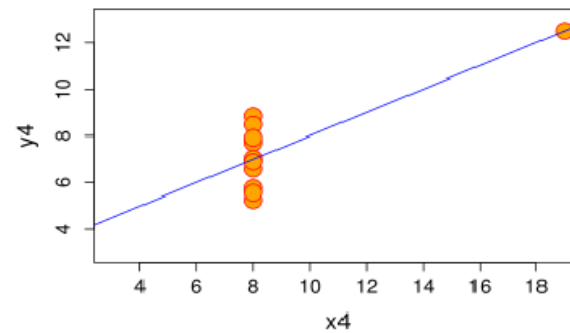
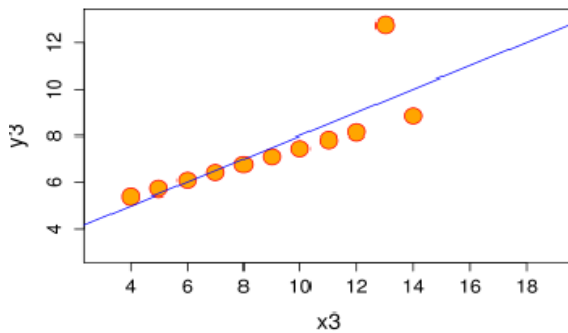
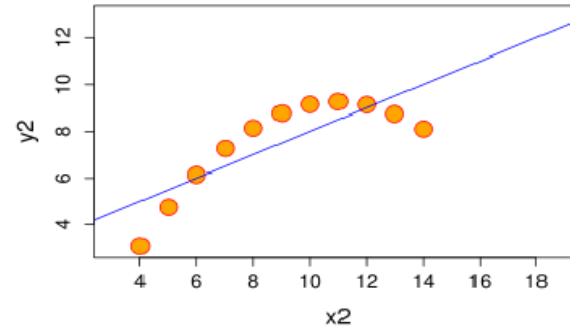
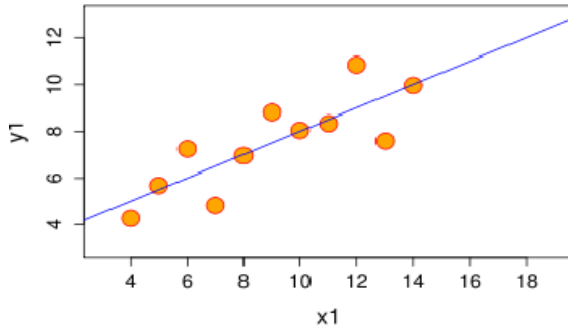
- Correlation can explore the relationship between two variables *with different units of measurement*
- Variables can be positively correlated (as one variable increases or decreases, the other variable changes in the same direction), or
- Variables can be negatively correlated (as one variable increases or decreases, the other variable changes in the opposite direction)
- Variables can be strongly, weakly, or non-correlated

What is the “ r ”?

- The r (or the Greek letter *rho*, “ ρ ”) is a statistic that shows the direction and strength of the correlation
 - A minus r indicates a negative correlation
 - A positive r indicates a positive correlation
- The r is a number between -1 (a perfect negative correlation) and +1 (a perfect positive correlation)
- The closer to zero the r , the weaker the correlation (“0” means not at all correlated)

Caution when interpreting r

$r = 0.81$ (a high positive correlation),
but— these are very different relationships!



Example

- Is there a correlation between length of treatment and improvement in depression? (Note – these data are fictitious for illustration)
- Sample: 20 clients in treatment for depression
- Independent variable—number of treatment sessions
- Dependent variable—change from pre- to post-treatment measurement of the Center for Epidemiological Studies Depression Scale (CES-D). In the CES-D, the higher the score, the more serious the depression*

*Note – some questions are “reverse scored” and have to be adjusted before the analysis. The CES-D form provides the scoring system

Excerpt from CES-D

During the past week:	Rarely or none of the time (less than 1 day)	Some or a little of the time (1–2 days)	Occasionally or a moderate amount of the time (3–4 days)	Most or all of the time (5–7 days)
1. I was bothered by things that usually don't bother me.				
2. I did not feel like eating; my appetite was poor.				
3. I felt that I could not shake off the blues even with help from my family or friends.				
4. I felt that I was just as good as other people.				
5. I had trouble keeping my mind on what I was doing.				
6. I felt depressed.				

Our data

Client ID	CES-D Pre-test score	CES-D Post-test score	Change score (Post – Pre)	Number of treatment sessions
1	26	25	-1	5
2	27	22	-5	5
3	35	34	-1	20
4	22	25	3	5
5	25	32	7	1
6	18	18	0	1
7	12	10	-2	8
8	14	14	0	2
9	45	28	-17	15
10	25	20	-5	10
11	15	16	1	5
12	54	30	-24	12
13	26	25	-1	5
14	28	26	-2	6
15	38	30	-8	10
16	54	35	-19	18
17	50	20	-30	18
18	45	25	-20	24
19	42	32	-10	15
20	36	22	-14	18

Calculating the Correlation

- You can use Excel to calculate the Pearson's r correlation. Go to

<http://www.gifted.uconn.edu/siegle/research/correlation/excel.htm>

(Also accessible at

<http://www.healtorture.org/content/correlation-calculation>)

The step-by-step instructions are easy to follow for someone familiar with Excel. Excel will provide the Pearson's r , but not the p value.

- For both the r and p value, go to

<http://www.socscistatistics.com/pvalues/pearsondistribution.aspx>

We will use this calculator for our example.

(Also accessible at <http://www.healtorture.org/content/pearson-r-and-p-value-calculators>)

Calculator steps—Calculate R

1. For this calculator, there are two steps – calculating the correlation R and then calculating the p value
2. Go to tab “Statistical test calculators” and click on [Pearson Correlation Coefficient Calculator](#), then “Take me to the calculator”
3. Enter (or cut and paste from a spreadsheet) “X values” – X values are those for the independent variable, in this case the 20 values for “Number of treatment sessions”
4. Enter “Y values” – Y values are those for the dependent variable, in this case the 20 values for “Change score”
5. Click “Calculate R”

Step 2 – Calculate p value

- Note the R value from the results screen – the value of R in this example is -0.7503
- Click on the tab “Quick P Value Calculators”
- Click the link “P Value from Pearson (R) Calculator”
- Enter the R value
- Enter the N : (N is the sample size, the number of pairs) – in this example 20
- Click “Calculate”
- Note the results statement:

The P-Value is 0.000139. The result is significant at $p < 0.05$.

Interpret Results

- The R of $-.7503$ indicates a high negative correlation – the higher the number of treatment sessions, the lower the CES-D score (indicating improvement in depression as a result of more treatment sessions)
- This p value is less than $.001$, indicating that the R value we obtained is not just by chance--which indicates that it is very likely a high correlation would also be found in other similar samples
- Reporting results: *“In this sample there is a high negative correlation between the number of treatment sessions and improvement in depression ($r = -0.75$, $p < .001$). In other words, there is a relationship between the number of treatment sessions and lower depression symptoms”.*

Important Caveats

- Correlation does not imply causality – we don't really know from this analysis that more treatment sessions will result in lower depression (there may be other explanations, or outside factors that affect depression, such as improved integration into a community, or some other basic needs being met)
- As shown in slide 6, the R value alone doesn't tell us whether the relationship between X and Y is *linear*. For example, it could be that the effect of many treatment sessions wears off at some point, at which point further treatment sessions will contribute very little to improvement in depression.