Psychol. Statistics using R

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Correlation Analysis

Changwei W. Wu, Ph.D.

Graduate Institute of Mind, Brain and Consciousness
Research Center of Brain and Consciousness
Taipei Medical University

Statistics

1. **Parametric correlation**
   - Pearson correlation coefficients
2. **Concerns of correlation**
   - Considerations from the scatter plot
3. **Non-parametric correlation**
   - Spearman & Kendall

Theories

Practice

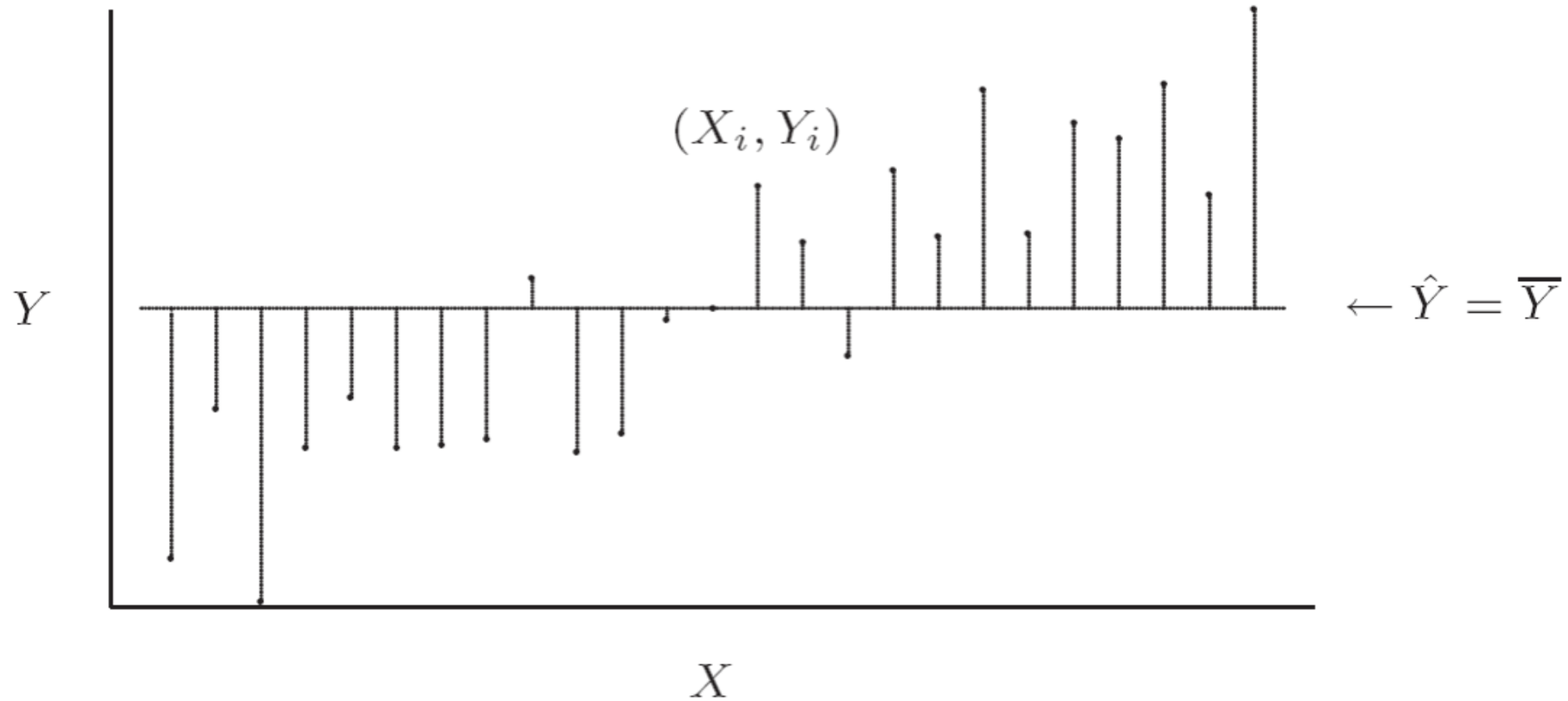
Assignment

TAIPEI MEDICAL UNIVERSITY
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

One Variable

- **Deviation**: differences between observed values and the average.
- **Sum of Squares (SS)**
- **Variance**: SS of the variable divided by degree of freedom.

$$\mathrm{SS}_{\mathrm{Total}} \quad \sum_i (Y_i - \overline{Y})^2 = \text{total sum of squares for } Y.$$



$$\text{Variance}(s^2) = \frac{\sum (x_i - \bar{x})^2}{N-1} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N-1}$$

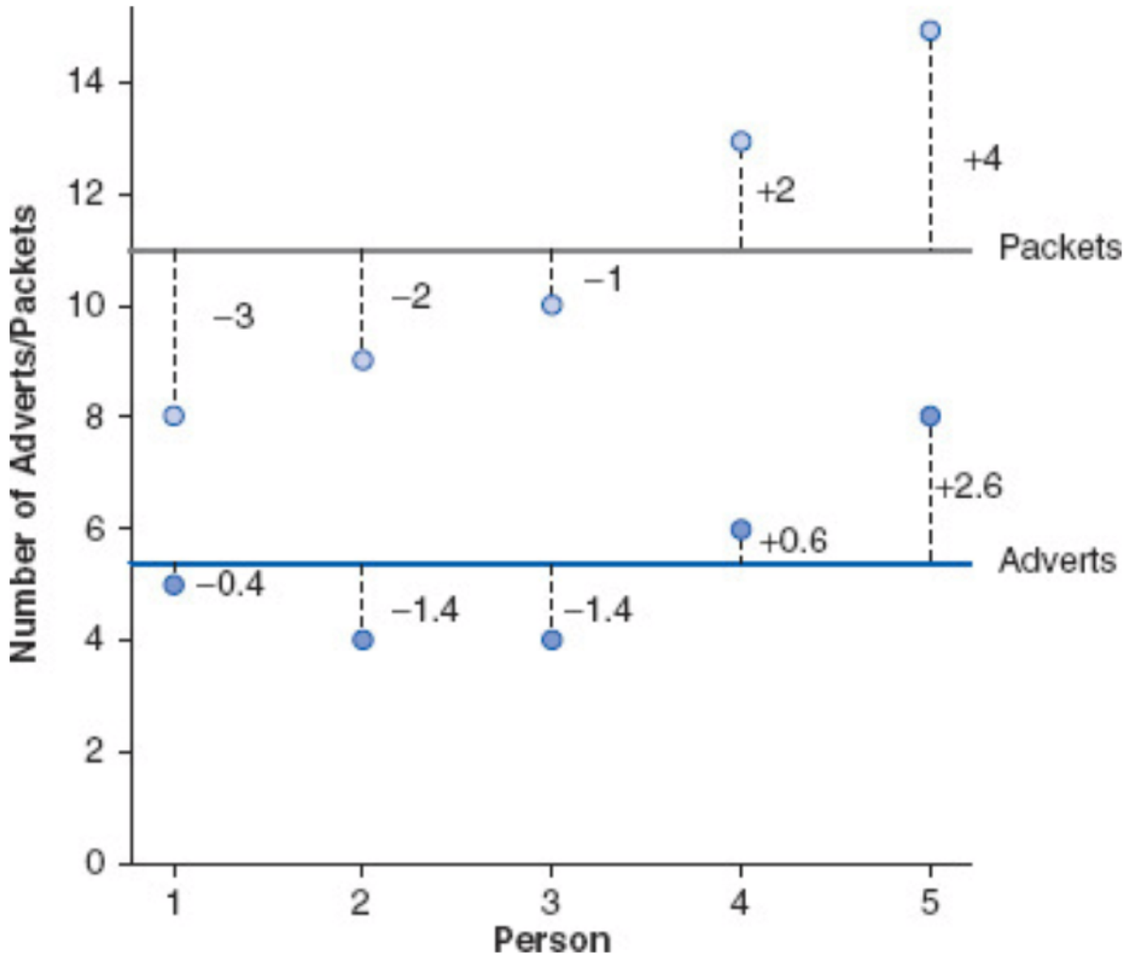
TAIPEI MEDICAL UNIVERSITY 1960
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi}\, dx$$

Two Variables

$$SS_{XY}$$

$$\mathrm{cov}(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$



- **Deviation**: differences between observed values and the average.
- **Variance**: SS in the variable divided by degree of freedom.
- **Covariance**: changes in **x** are met with similar changes in **y** (**paired**).
  - ➡ **Product-sum** is the approach for "similarity" (Ex: Fourier Transform).
- Variance-Covariance Matrix
  - ➡ If there are multiple variables, we can formulate var-cov matrix to remove auto-correlation.
- Covariance is dependent on the scale of measurement (e.g., units).

TAIPEI MEDICAL UNIVERSITY
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Pearson's Correlation $= \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x) \times \mathrm{var}(y)}}$

- **Standardization**: to overcome the dependence problem on scale, we convert covariance into a standard set of units, which is ***correlation***.

between -1 and +1

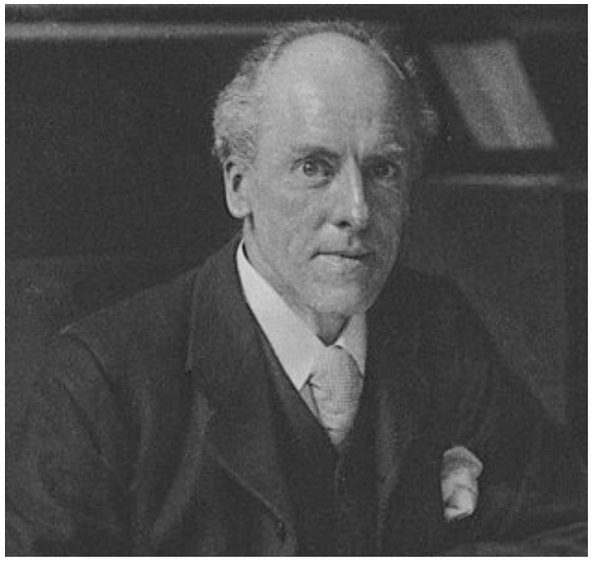
$$r = \frac{\mathrm{SS}_{XY}}{\sqrt{\mathrm{SS}_X \cdot \mathrm{SS}_Y}}$$

Sum of squares of paired deviation

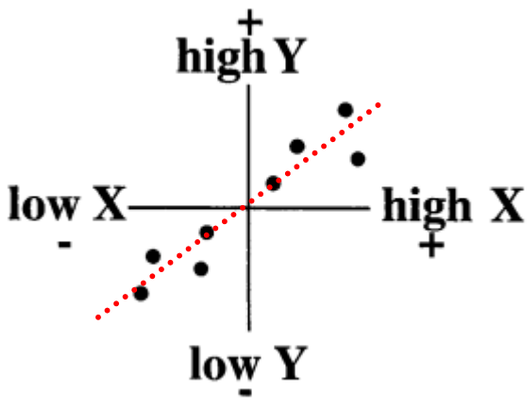
Correction factor on Amplitudes

Karl Pearson

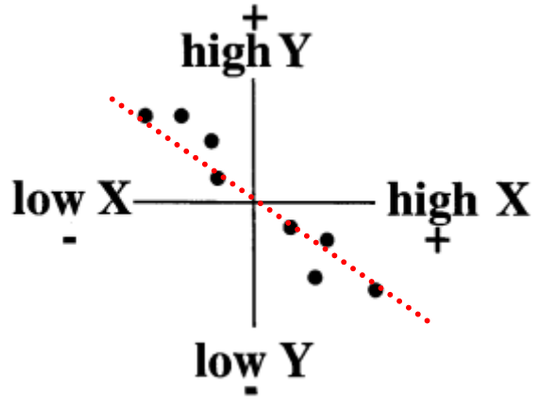
1857-1936



- **Positive** correlation
  - ➢ Most high X values are paired with high Y values
  - ➢ Most low X values are paired with low Y values



- **Negative** correlation
  - ➢ Most high X values are paired with low Y values
  - ➢ Most low X values are paired with high Y values



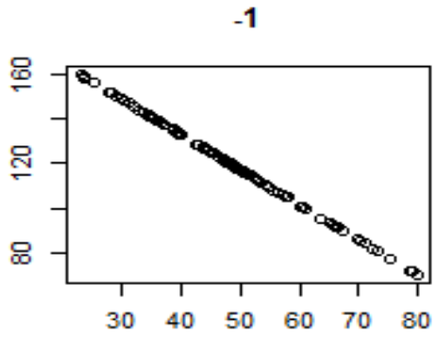
Pearson's Correlation

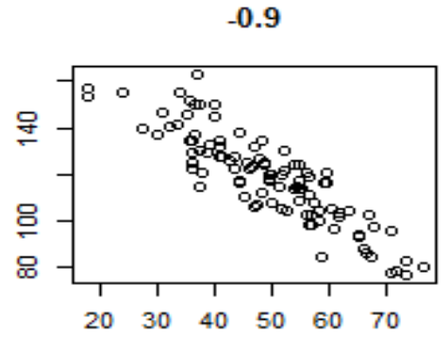
$$r = \frac{SS_{XY}}{\sqrt{SS_X \cdot SS_Y}}$$

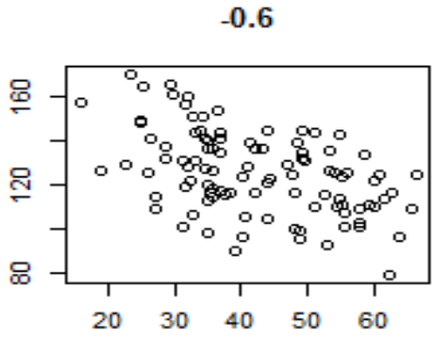
$$SSY = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

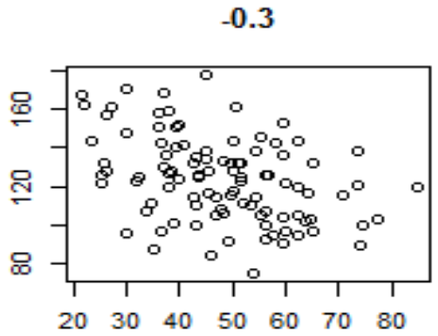
$$SSX = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

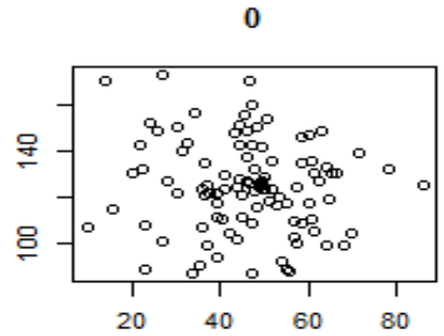
$$SSXY = \sum xy - \frac{\sum x \sum y}{n}$$

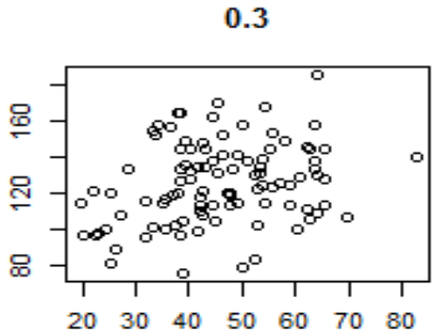


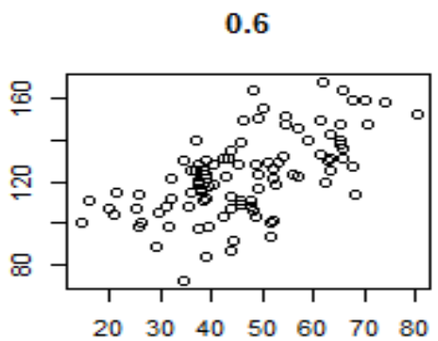


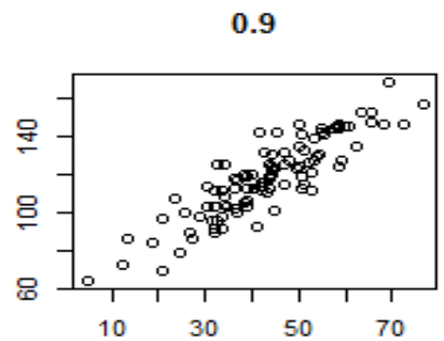


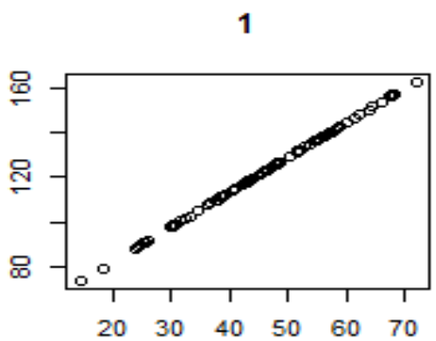








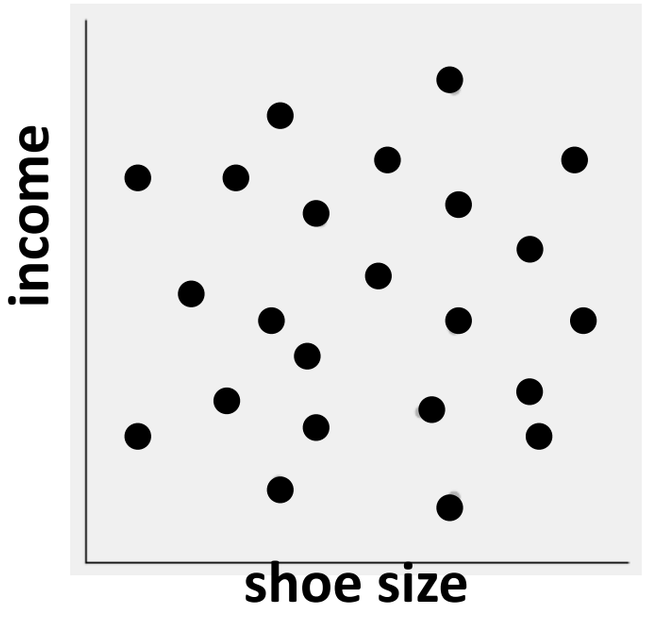




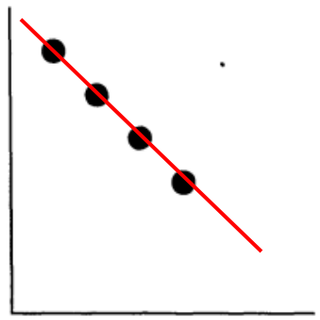
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

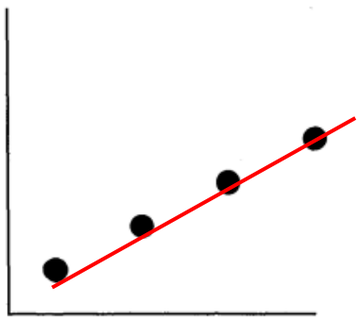
Issues of Correlation

- **No correlation**
  - ➢without linear trend

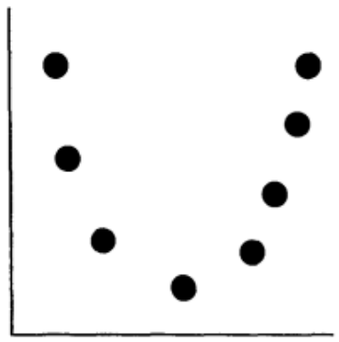


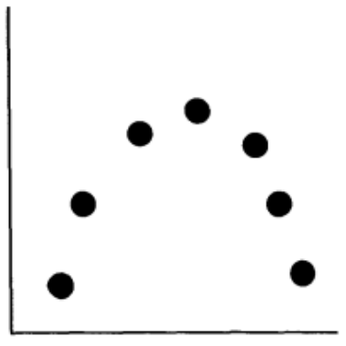
- **Linear correlation**
  - ➢ Roughly following a straight line





- **Curvilinear correlation**
  - ➢ Systematic pattern that is not a straight line



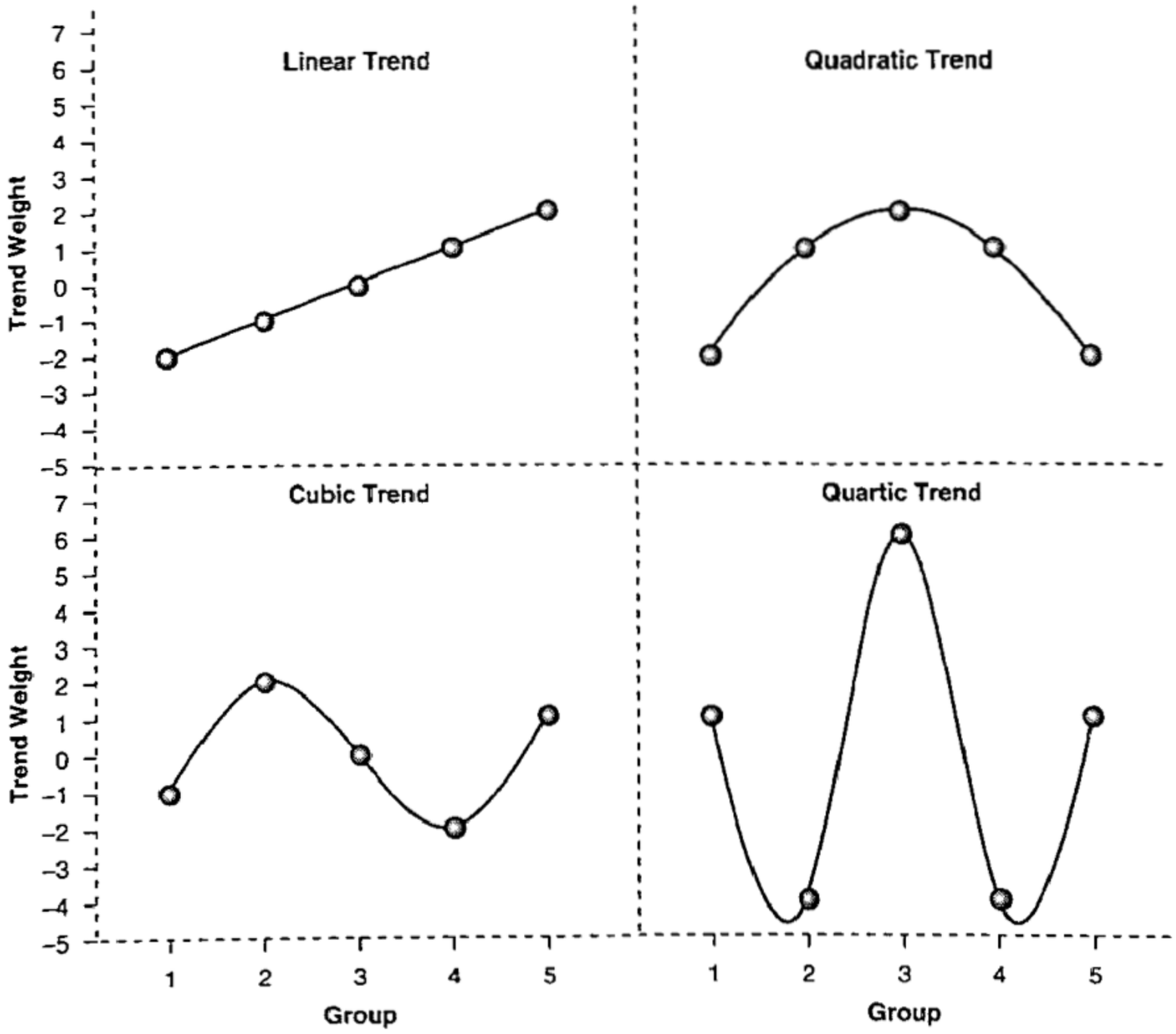


CHI 2017.

- Pearson's ***r*** indicates a linear relationship only.
- If relationship is curvilinear:
  - ➢do not use Pearson's ***r***
  - ➢Use Spearman's **ρ**

TAIPEI MEDICAL UNIVERSITY 1960
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

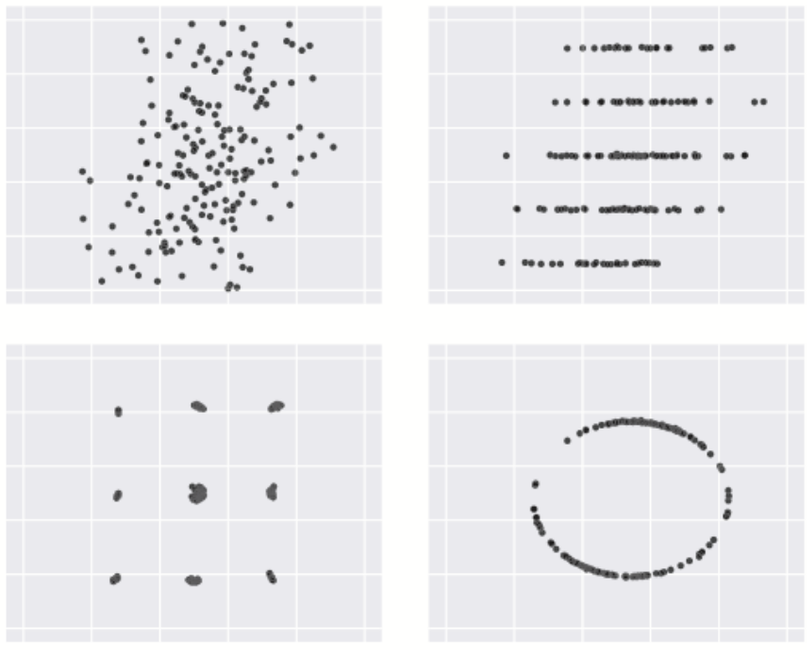
Issues of Correlation

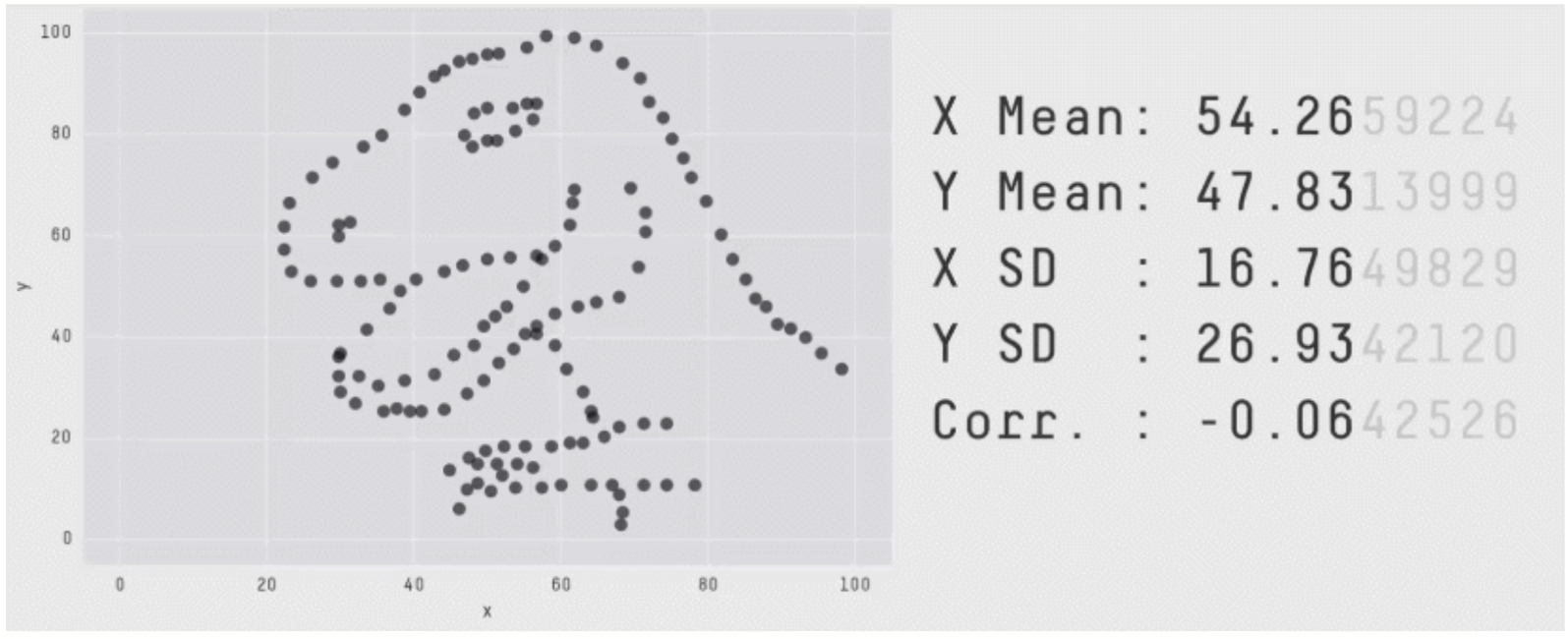


- Pearson's ***r*** indicates a **linear** relationship only.

Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing

Justin Matejka and George Fitzmaurice
Autodesk Research, Toronto Ontario Canada
{*first.last*}@autodesk.com





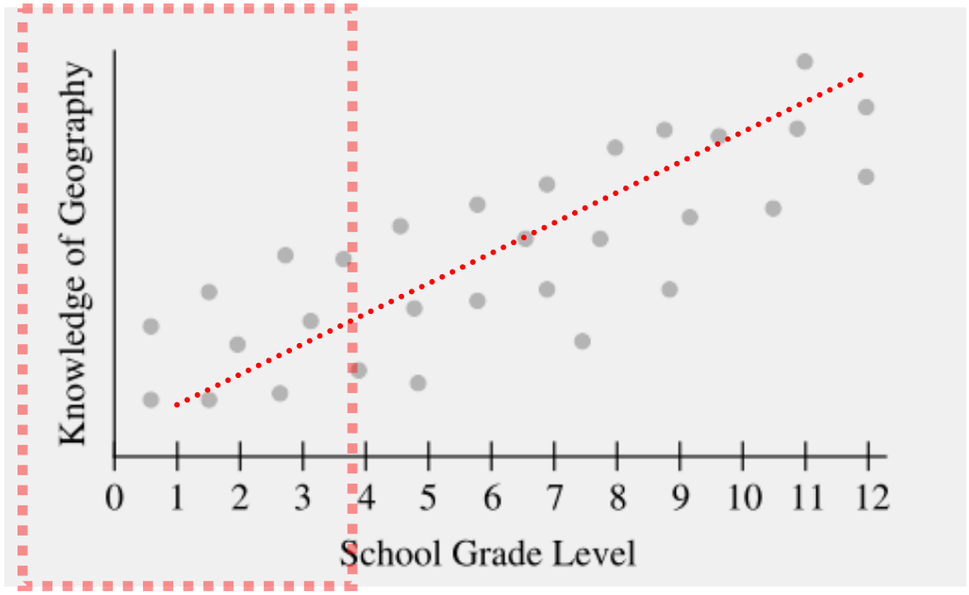
Make sure reporting the correlation with scatter plots!

https://www.autodeskresearch.com/publications/samestats

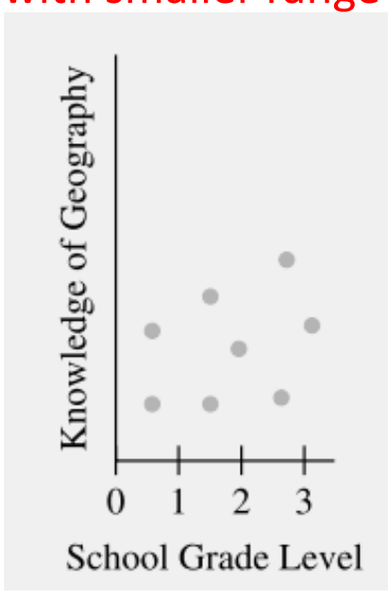
Issues of Correlation

- **Range**: limited range may restrict the emergence of correlation.

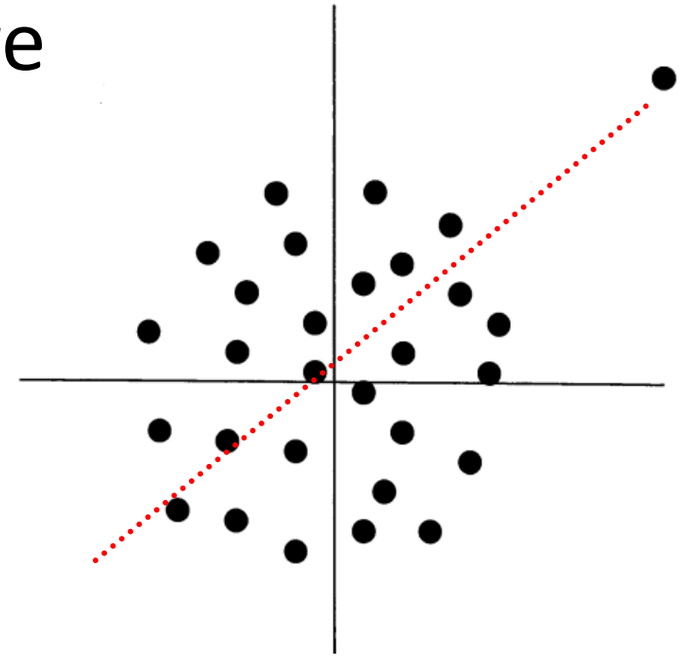
Positive correlation with knowledge of geography and school grade level



No correlation with smaller range

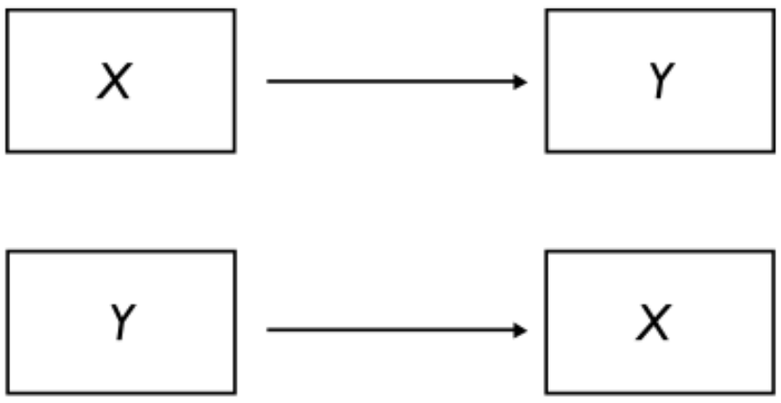


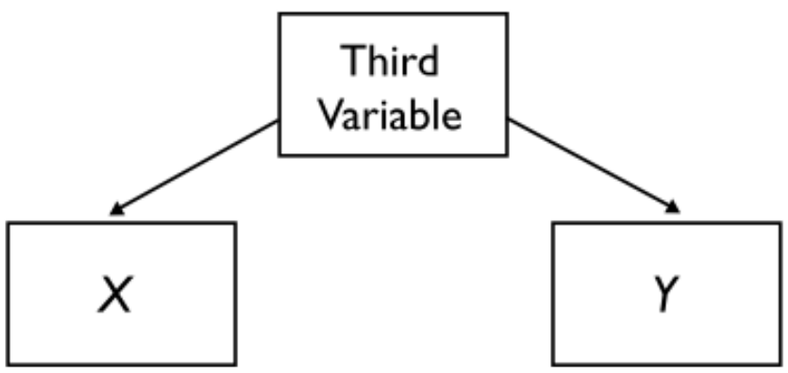
- **Outlier**: extreme score(s), have especially large influences in the correlation.

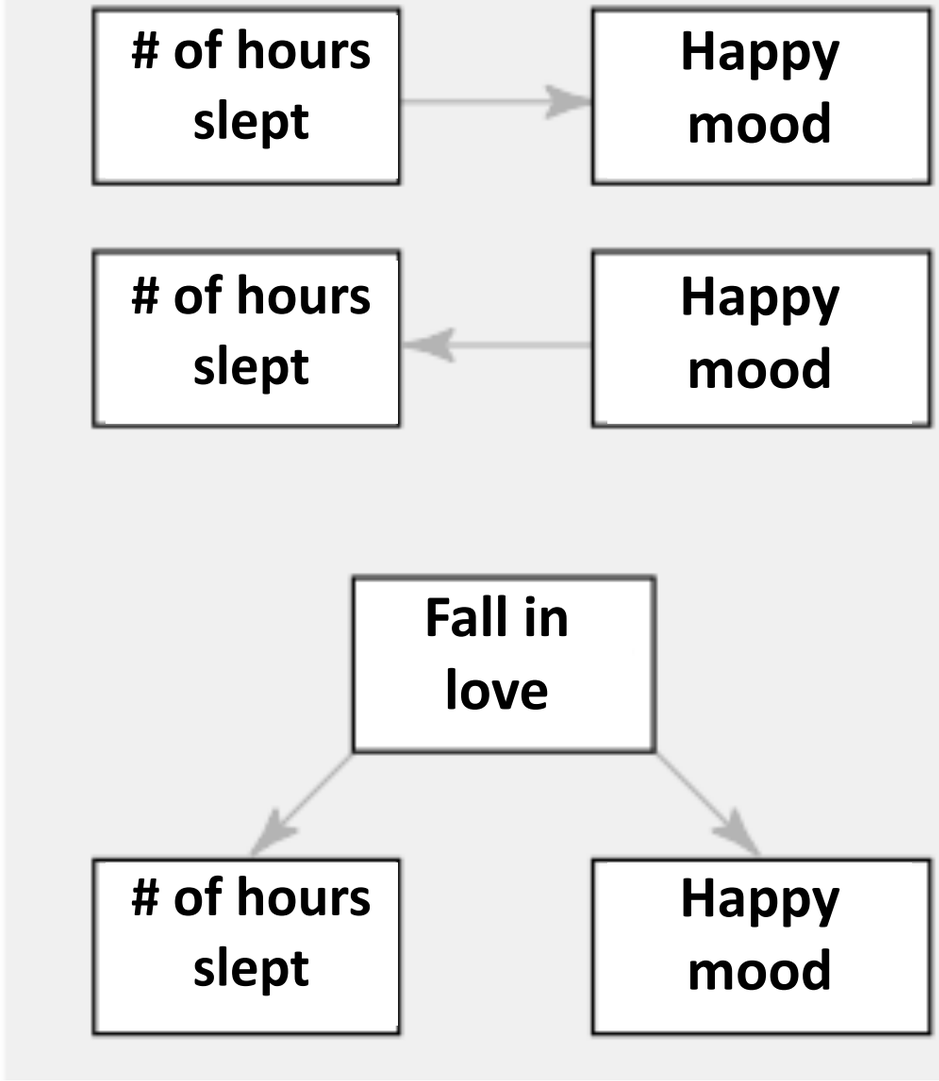


No correlation
vs.
positive correlation

Issues of Correlation
Correlation ≠ Causality







- **No indication of the direction of Causality.**

Correlation coefficients

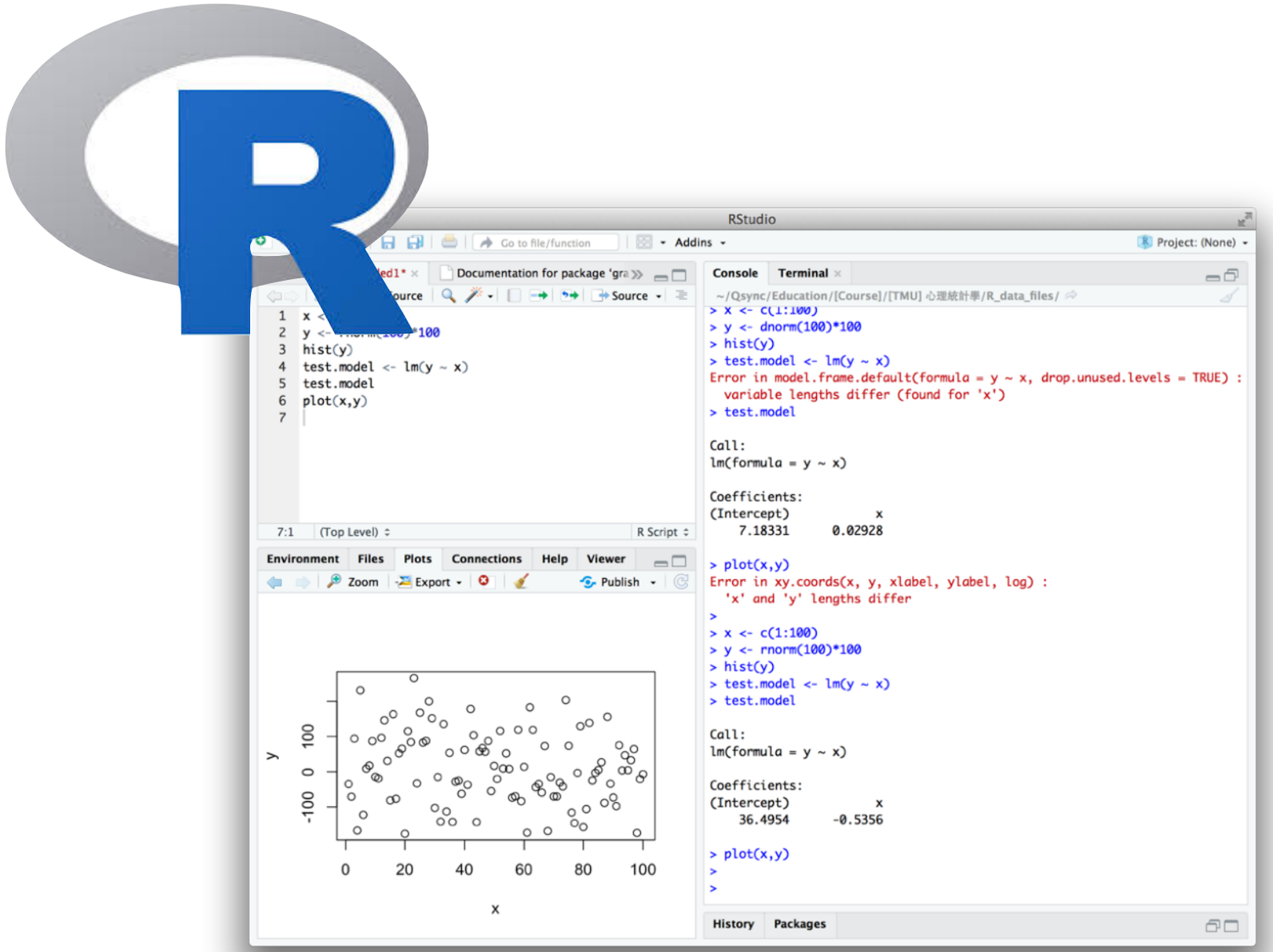
- **Measure relationship between two variables.**
  - Correlation coefficients lie between -1 and +1.
  - Plot the figure to make sure the valid linear correlation.

- **Pearson's** correlation coefficient, ***r***, is a parametric statistic and requires interval data for both variables.
- **Spearman's** correlation coefficient, **ρ**, is a non-parametric statistic and requires only ordinal data for both variables.
- **Kendall's** correlation coefficient, **τ**, is a non-parametric statistic but probably better for small samples.

- A ***partial correlation*** quantifies the relationship between two variables while controlling the effects of a third variable on *both* variables in the original correlation; whereas a ***part correlation*** controls the effects of a third variable on *only one* variable in the original correlation.

Pearson’s correlation coefficients

① PARAMETRIC CORRELATION



TAIPEI MEDICAL UNIVERSITY
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Assumption of Pearson's Correlation

- **The data are with continuous intervals.**
  ➡ for it to be a measure of the linear relationship between two variables.
- For the test statistic to be valid, **the sampling distribution has to be normally distributed.**

	Pearson	*Spearman*	*Kendall*	*p-value*	*C.I.*	*Multiple Correlation*
cor	✓	✓	✓			✓
cor.test	✓	✓	✓	✓	✓	
cor_test	✓	✓	✓	✓		✓

DEMO

Exam vs. Anxiety

- **Background**: A psychologist was interested in the effects of exam stress and revision on exam performance. She had devised a questionnaire to assess state anxiety relating to exams. This scale produced a measure of anxiety scored out of 100.
- ***Anxiety*** was measured before an exam, and the ***percentage mark*** of each student on the exam was used to assess exam performances.
- She also measured the *number of hours spent revising*.

Measure 1: Anxiety before exams

Measure 2: Percentage mark of each student

- **Set up Hypothesis**:

① Hypothesis

H_0: Anxiety has nothing to do with Exam mark.

H_a: Anxiety is linearly correlated with Exam mark.

DEMO

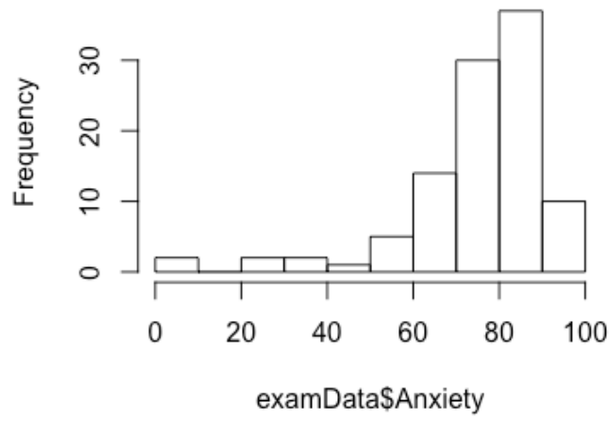
Exam vs. Anxiety

- **Hypothesis**:
  - $H_0: r = 0$
  - $H_a: r \neq 0$
- **Data import**:
  - *examData = read.delim("ExamAnxiety.dat", header = TRUE)*

Load:
ExamAnxiety.dat

② Assumption

- **Assumption check:**




```
> pastecs::stat.desc(examData$Exam, basic=F, norm=T)
      median         mean      SE.mean CI.mean.0.95          var
60.000000000 56.572815534  2.556001435  5.069816727 672.913763564
     std.dev      coef.var     skewness      skew.2SE      kurtosis
25.940581404  0.458534389 -0.362455926 -0.761660046 -0.910222445
    kurt.2SE     normtest.W     normtest.p
-0.964980721  0.955157617  0.001521502
>
> pastecs::stat.desc(examData$Anxiety, basic=F, norm=T)
      median         mean      SE.mean CI.mean.0.95          var
7.904400e+01 7.434367e+01 1.692979e+00 3.358015e+00 2.952163e+02
     std.dev      coef.var     skewness      skew.2SE      kurtosis
1.718186e+01 2.311139e-01 -1.953528e+00 -4.105118e+00 4.732992e+00
    kurt.2SE     normtest.W     normtest.p
5.017725e+00 8.224243e-01 8.650105e-10
```

臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

Let's skip the violation first, use Spearman later.

DEMO

Exam vs. Anxiety

③ Testing

General form:

- ***cor(x, y, use="string", method="correlation type")***
- ***cor_test(x, method="correlation type")***

Input 1 variable: one variable against ALL
Input 0 variable: all possible pairwise correlation

- **Parameters**:
  - *x, y*: numeric variable or dataframe.
  - *use=(**everything** / all.obs / complete.obs / pairwise.complete.obs)*
    - everything: report an NA when there is missing values.
    - all.obs: report an error when there is missing values.
    - **complete.obs**: missing values are handled by **casewise** deletion.
    - **pairwise.complete.obs**: when variable numbers are more than 2, then choose pairwise for deletion & only works with '**pearson**'.
  - *method=(**pearson** / spearman / kendall)*: choosing method types.

DEMO

Exam vs. Anxiety

③ Testing

General form:

- ***cor(x, y, use="string", method="correlation type")***
- ***cor_test(x, method="correlation type")***

Input 1 variable: one variable against ALL
Input 0 variable: all possible pairwise correlation

- **Practice**:
  - ***cor(examData$Exam, examData$Anxiety, use = "complete.obs", method = "pearson")***
  - input as a data frame
    - **multiple correlation ←**

```
              Exam    Anxiety     Revise
Exam     1.0000000 -0.4409934  0.3967207
Anxiety -0.4409934  1.0000000 -0.7092493
Revise   0.3967207 -0.7092493  1.0000000
```

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

DEMO

Exam vs. Anxiety

③ Testing

General form:

- ***cor.test(x, y, alternative="string", method="correlation type", conf.level=0.95)***

- **Parameters**:
  - *x, y*: numeric variable or dataframe.
  - *alternative=(**two.sided** / less / greater)*
    - less: when you predict that correlation is less than zero.
    - greater: when predicting that correlation is greater than zero.
  - *method=(**pearson** / spearman / kendall)*: choosing method types
  - *conf.level*: 0.95 by default, change by your expectation.

- **Practice:**
  - ***cor.test(examData$Exam, examData$Anxiety, method="pearson")***

DEMO

Exam vs. Anxiety

③ Testing

General form:

- ***cor.test(x, y, alternative="string", method="correlation type", conf.level=0.95)***

```
        Pearson's product-moment correlation

data:  examData$Exam and examData$Anxiety
t = -4.938, df = 101, p-value = 3.128e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5846244 -0.2705591
sample estimates:
       cor
-0.4409934
```

$$t = \frac{r - 0}{s_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}, \qquad \text{with } \nu = n - 2.$$

s_r = standard error of ***r***

- **How to calculate C.I.?**

confidence interval = t-value × standard error

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學

Fisher's Z Transform

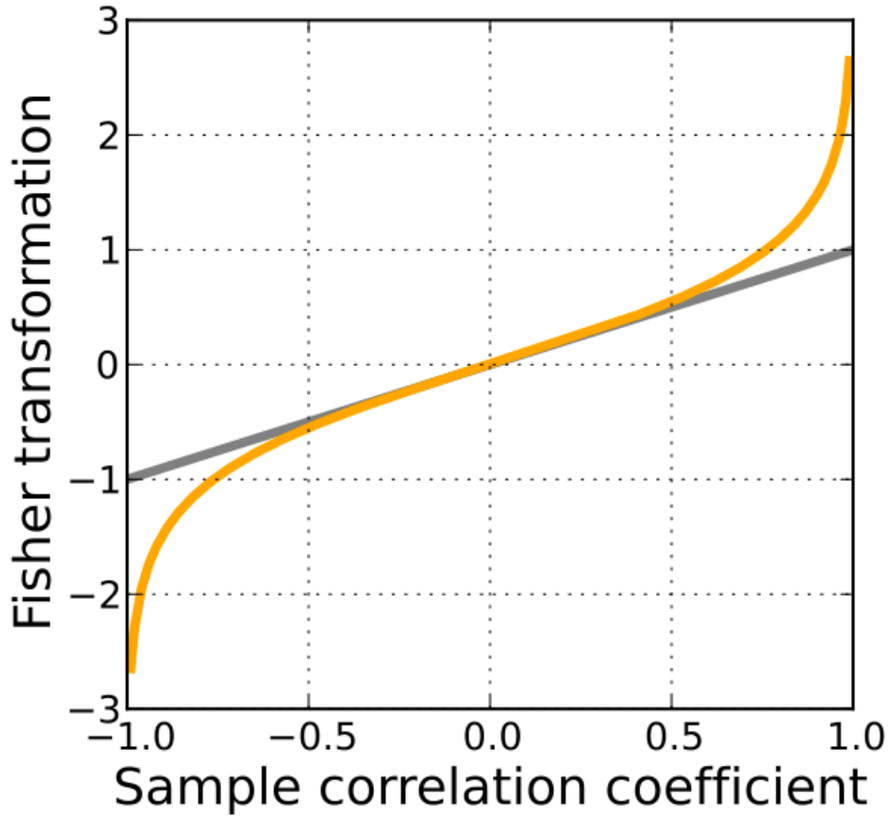
- **An inverse hyperbolic tangent of *r*** (Fisher, 1921)
  - Alternative approach to calculate <u>confidence interval</u>, when ***r*** values have been deemed significantly different from 0 via a ***t*** test.
  - For the next parametric testings on the ***r*** value. (e.g., connectivity)

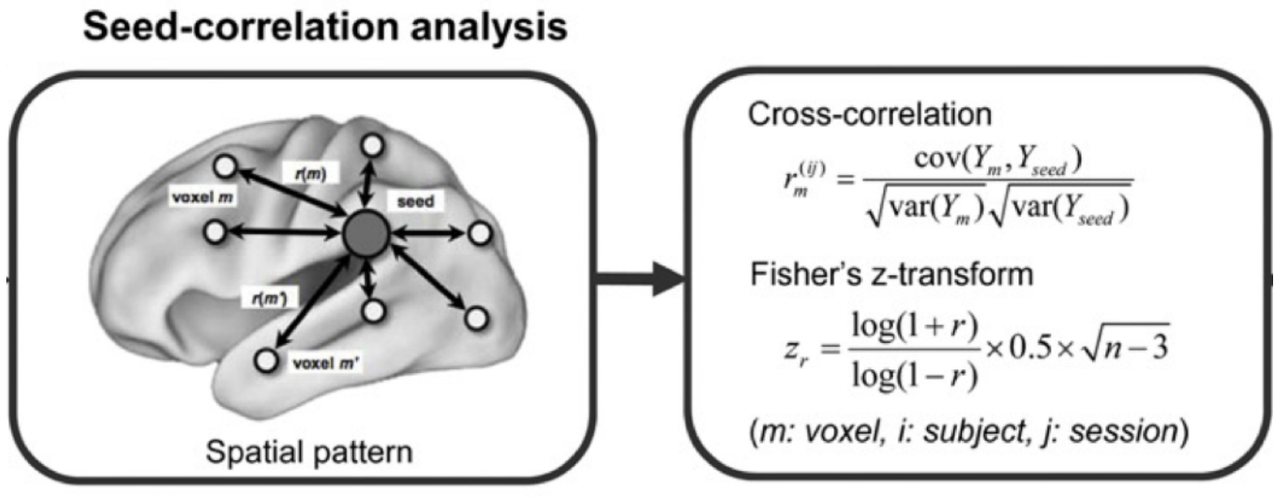
$$Z = \tanh^{-1} r = 0.5 \ln\left(\frac{1+r}{1-r}\right)$$

The resulting z_r has a standard error of:

$$SE_{z_r} = \frac{1}{\sqrt{N-3}}$$

- **L1 = z + 1.96 x SE_{zr}**
- **L2 = z - 1.96 x SE_{zr}**





DEMO

Exam vs. Anxiety

④ Effect Size

- **Effect size**: correlation coefficient itself is the effect size!

$$r^2 = \frac{SS_R}{SS_{Total}} = \frac{(SS_{XY})^2}{SS_X \cdot SS_Y}$$

- You may calculate ***R²*** as the coefficient of determination.

```
> R2
             Exam    Anxiety    Revise
Exam     100.00000  19.44752  15.73873
Anxiety   19.44752 100.00000  50.30345
Revise    15.73873  50.30345 100.00000
```

⑤ Decision

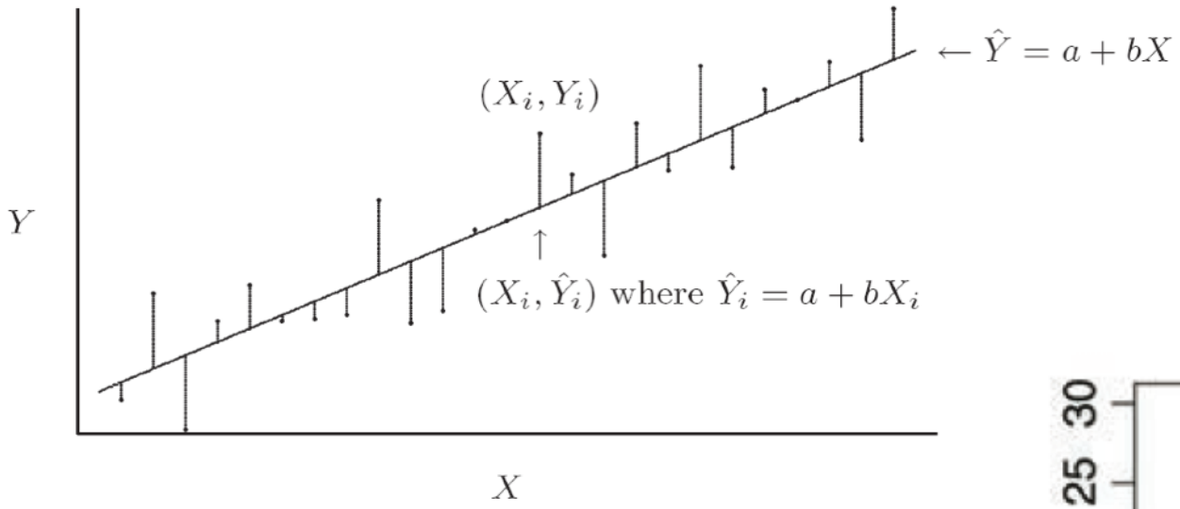
- **Reporting decision:**

  Exam performance was significantly correlated with exam anxiety, *r* = -.44, and time spent revising, *r* = .40; the time spent revising was also correlated with exam anxiety, *r* = -.71 (all *p* < 0.001).

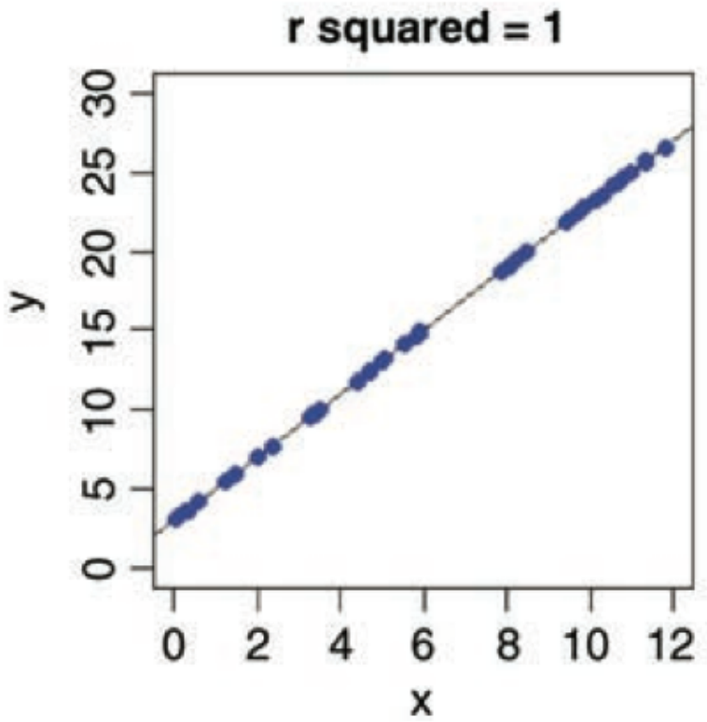
```
        Spearman's rank correlation rho

data:  examData$Anxiety and examData$Exam
S = 255790, p-value = 2.245e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.4046141
```

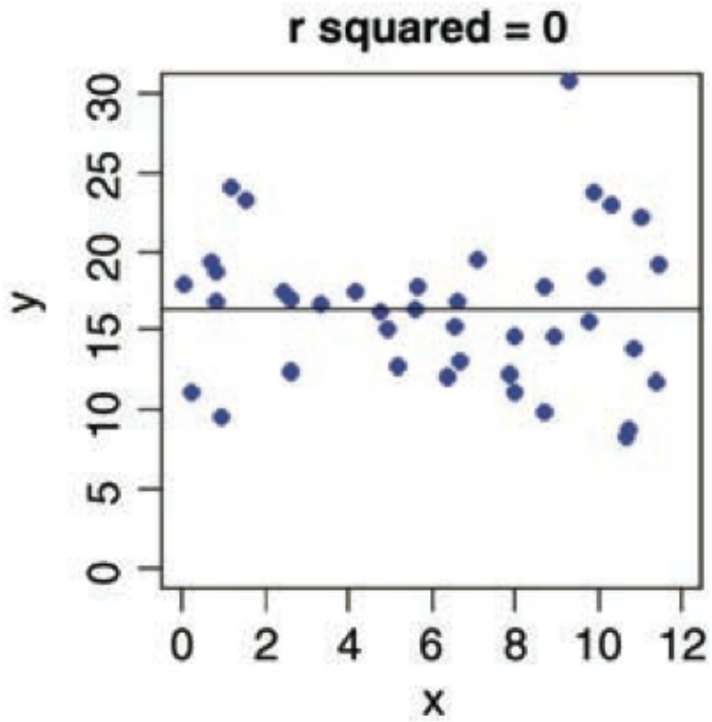
Coefficient of Determination

- **R^2**
  - R^2 lies between 0 and +1.
  - The proportion of the variance in the dependent variable ($\mathbf{Y}$) that is predictable from (explained by) the independent variable ($\mathbf{X}$).



$$r^2 = \frac{\mathrm{SS_R}}{\mathrm{SS_{Total}}} = \frac{(\mathrm{SS}_{XY})^2}{\mathrm{SS}_X \cdot \mathrm{SS}_Y}$$



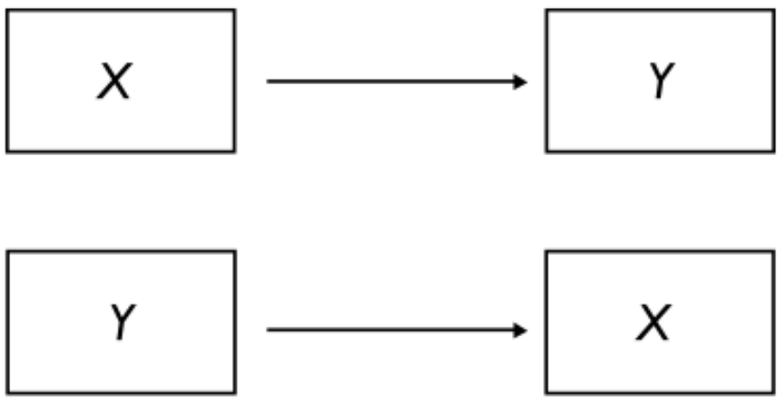


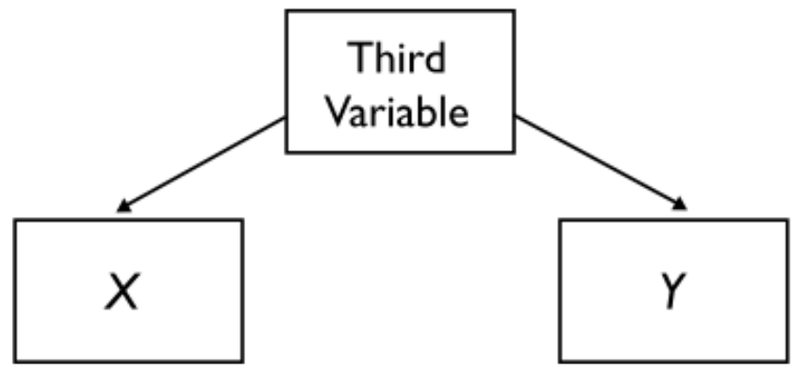
TAIPEI MEDICAL UNIVERSITY 臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

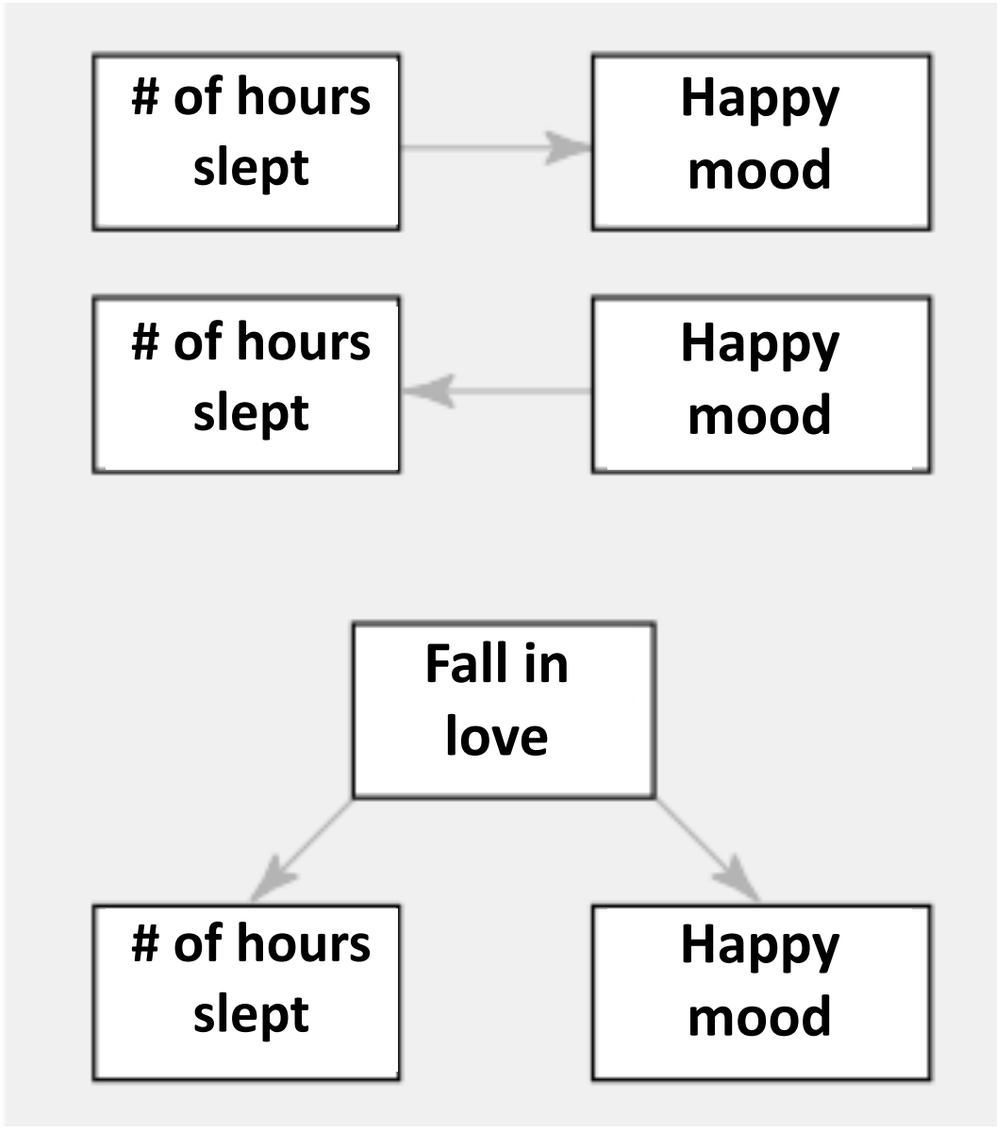
An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

② CONCERNS OF CORRELATION

Correlation ≠ Causality







- **No indication of the direction of Causality.**

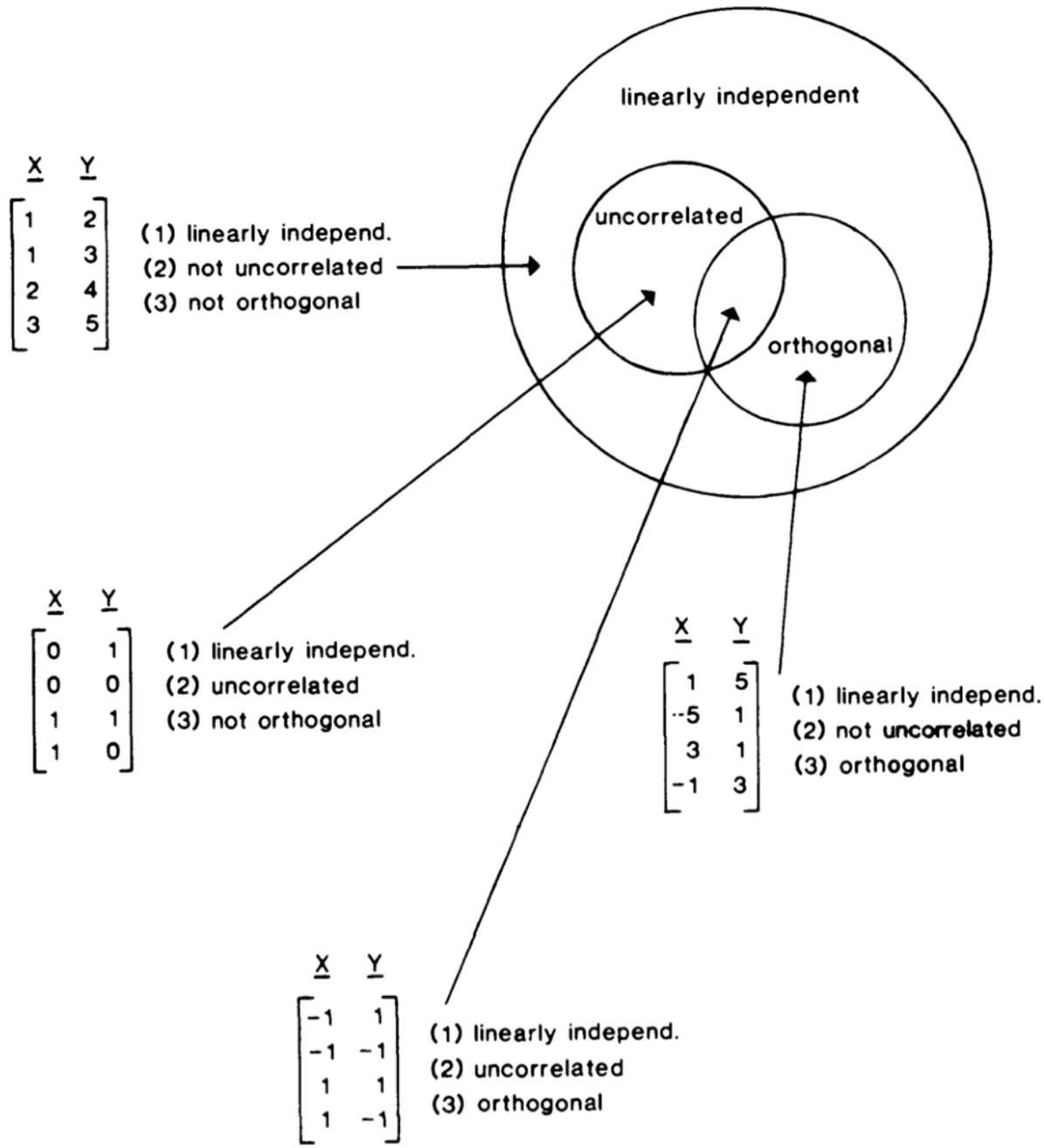
Issues of Correlation
Independence / Uncorrelated

Let **X** and **Y** be vector observations of the variables X and Y. Then

1. **X** and **Y** are linearly independent iff there exists no constant a such that $a\mathbf{X} - \mathbf{Y} = 0$ (**X** and **Y** nonnull vectors).
2. **X** and **Y** are orthogonal iff $\mathbf{X}'\mathbf{Y} = 0$.
3. **X** and **Y** are uncorrelated iff $(\mathbf{X} - \bar{X}\mathbf{1})'(\mathbf{Y} - \bar{Y}\mathbf{1}) = 0$, where $\bar{X}$ and $\bar{Y}$ are the means of **X** and **Y**, respectively, and **1** is a vector of ones.

To these definitions:

1. **Uncorrelated** random variables are not necessarily independent.
2. Pearson correlation measures the strength of **linear relationship** between two variables.
3. Statistical independence requires no specific relationship of any kind (e.g., linearity).

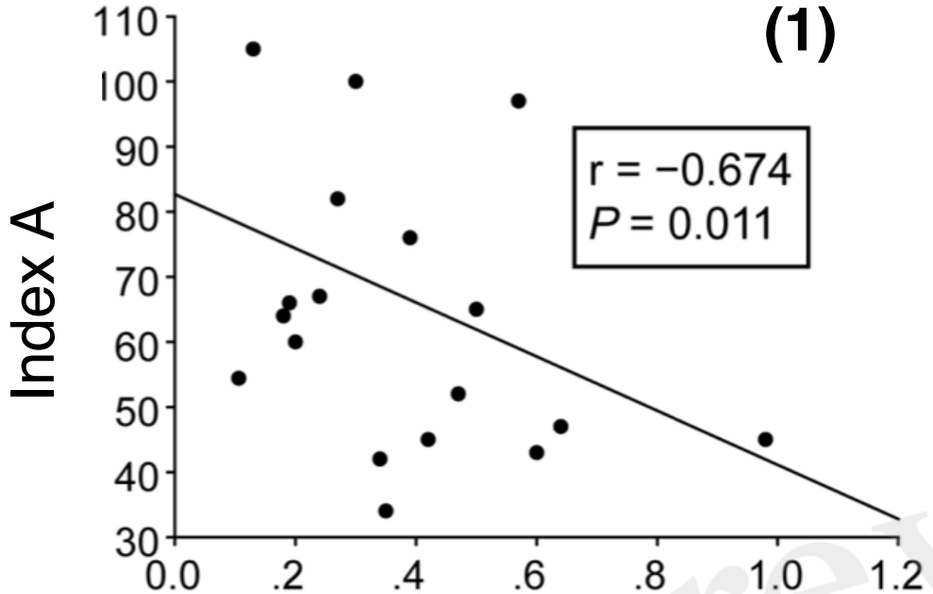


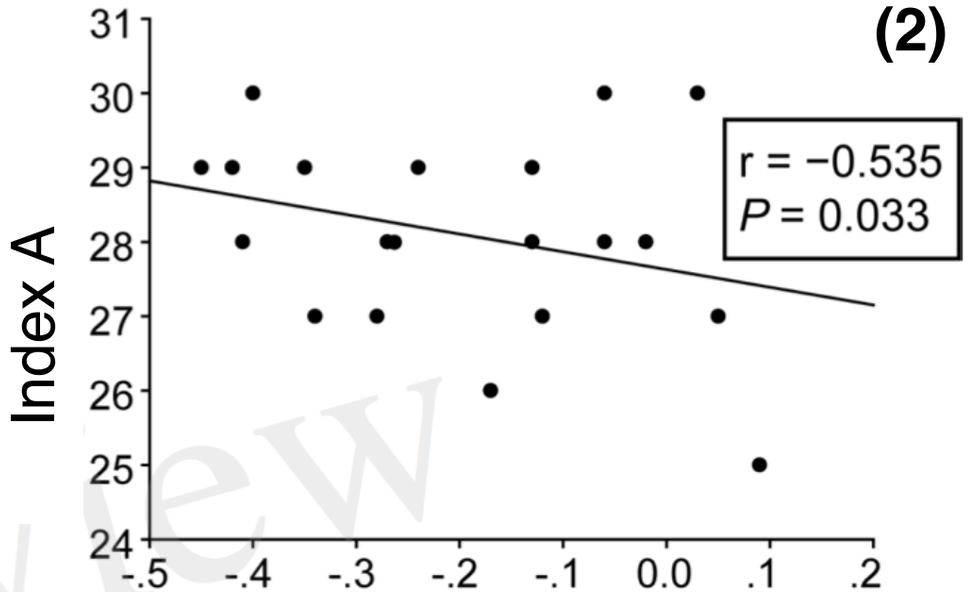
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

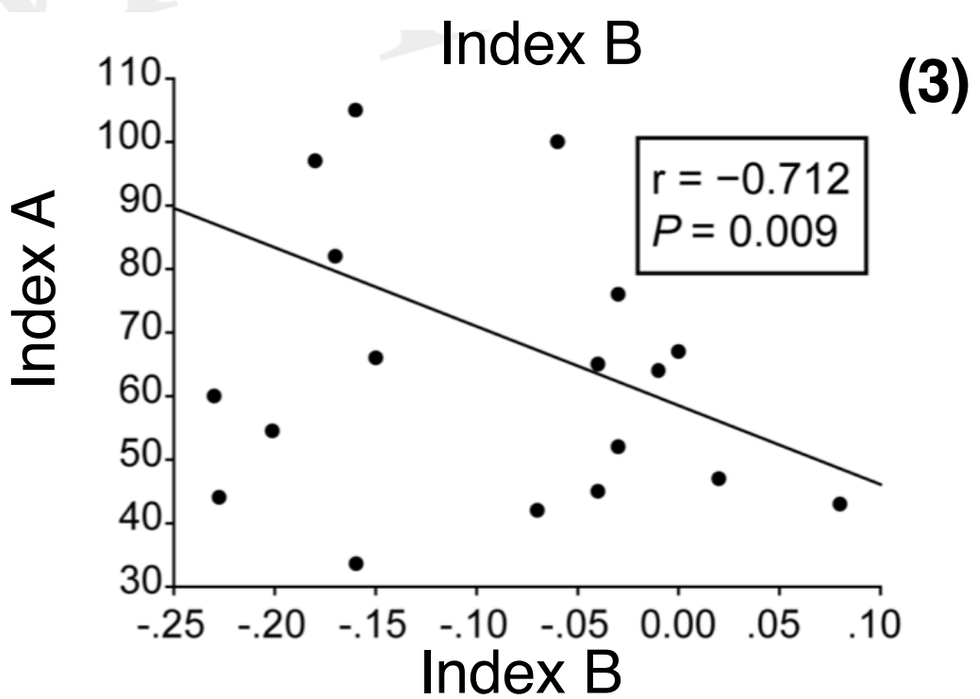
The American Statistician, Vol. 38, No. 2. (May, 1984), pp. 133-134.

Thoughts

Review of an Article





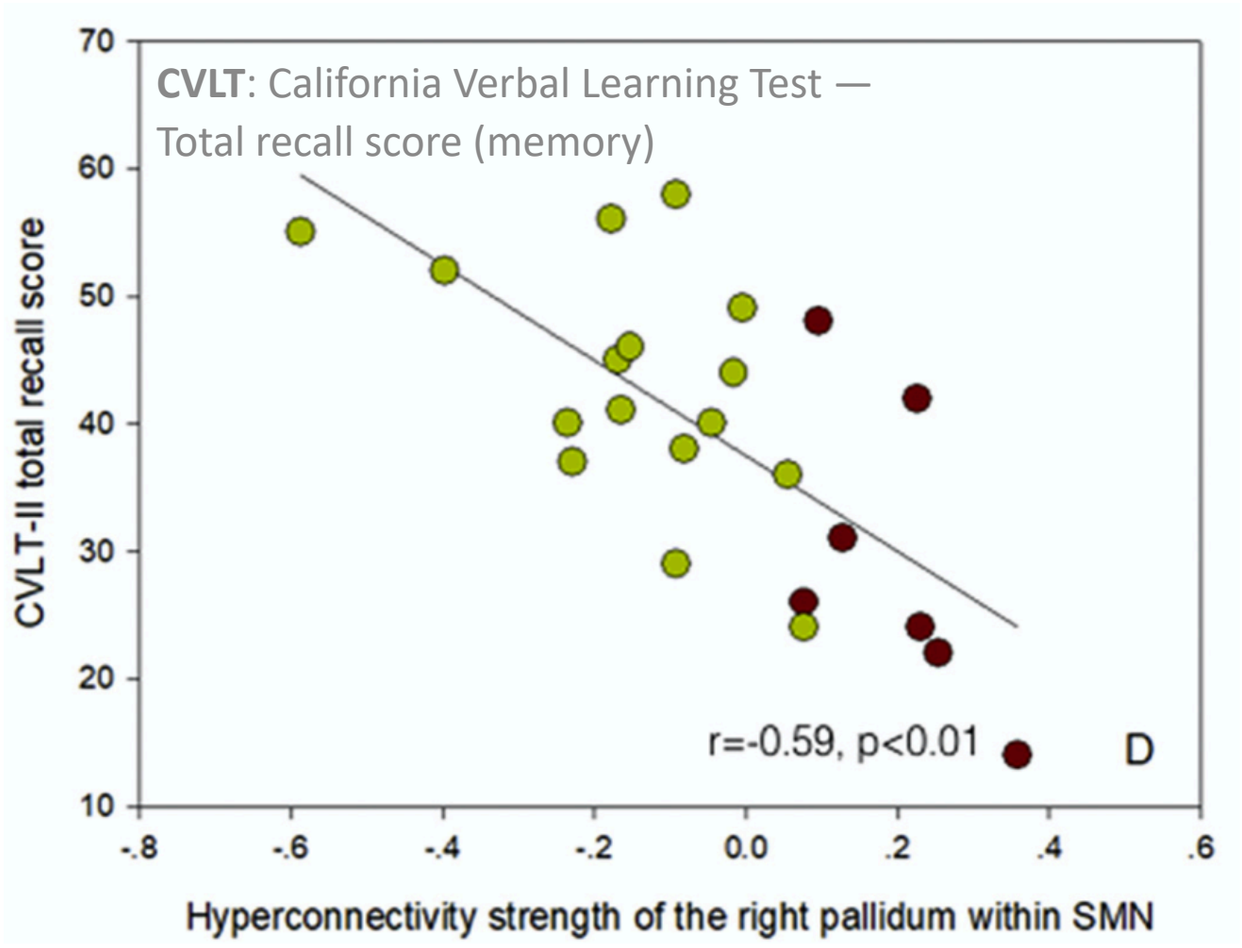


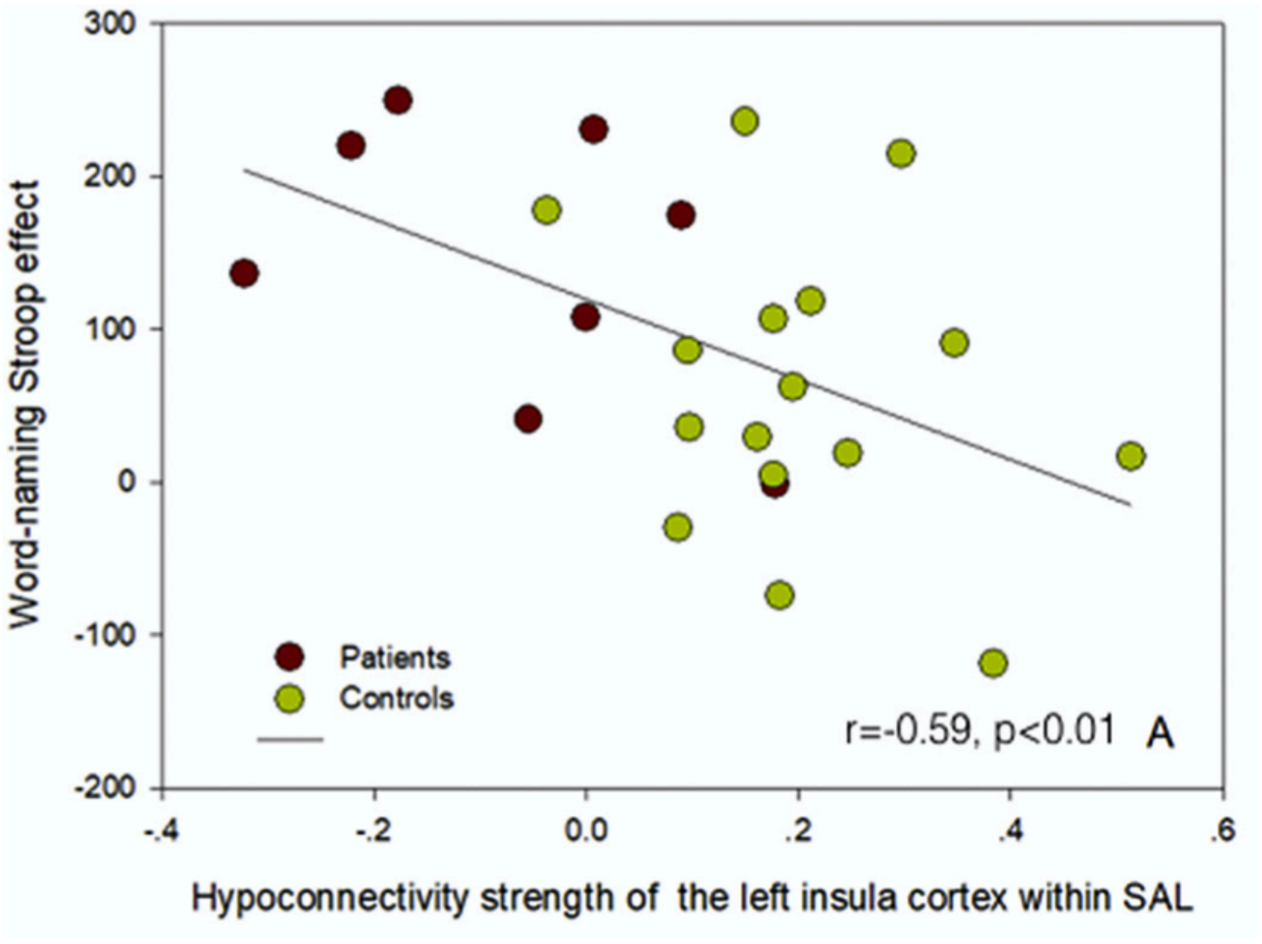
- **This is a figure in one manuscript under review.**

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Thoughts

Correlation btw Groups





- **The reporting figure shows the negative correlation between brain connectivity and the questionnaire.**
- **Is the correlation valid within group?**

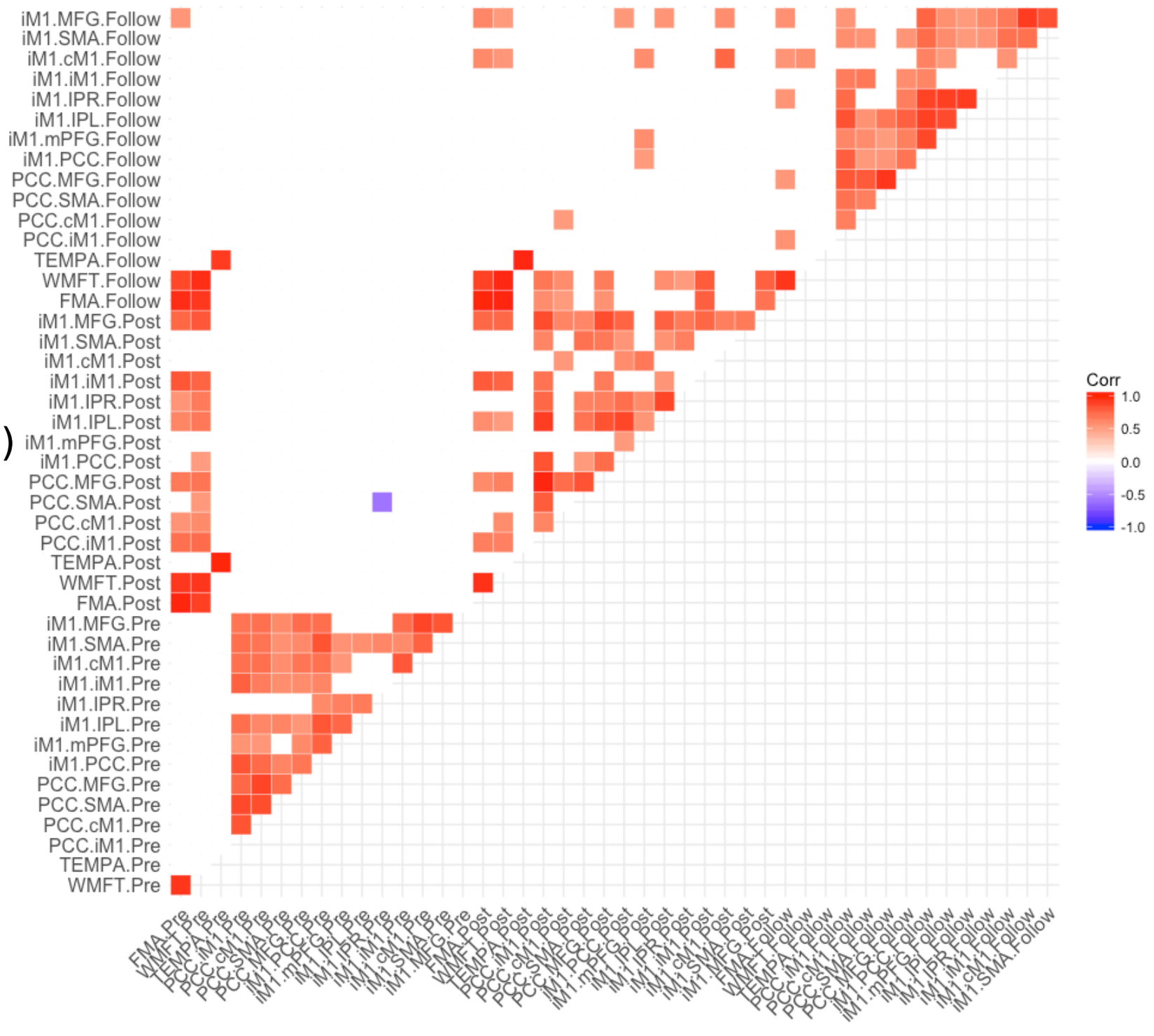
Huang et al. ***Neuroimage: Clinical*** 2018

TAIPEI MEDICAL UNIVERSITY
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Practice

Correlation Map

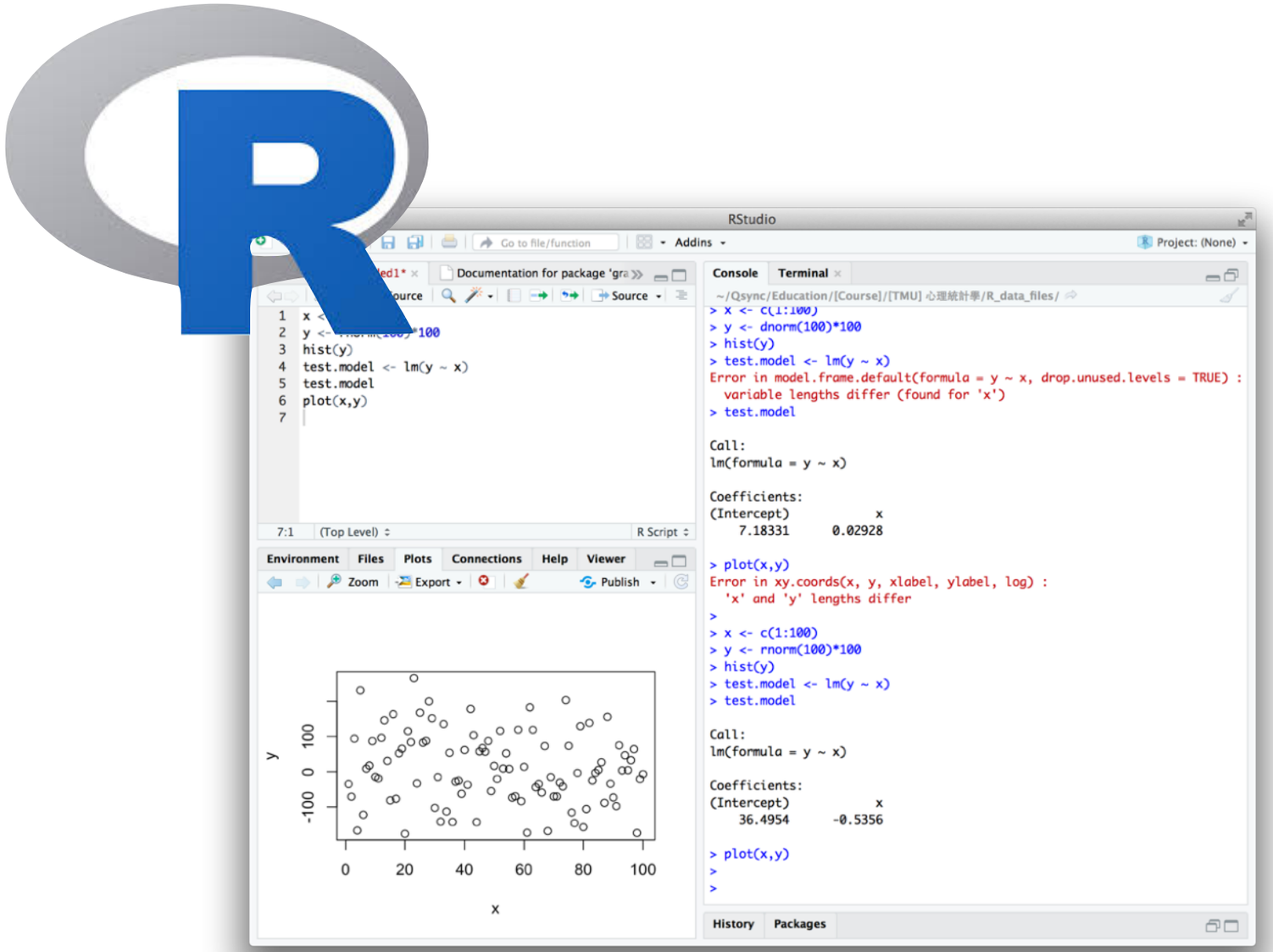
```
library(ggcorrplot)

Rtable<-cor(RehabIDX)
Rsqr<-
round(cor(RehabIDX)^2,2)
Pval <-
round(cor_pmat(RehabIDX),3)
ggcorrplot(Rtable,hc.order
= FALSE, outline.col =
"white",type
="upper",p.mat=Pval, insig
= "blank")
```



TAIPEI MEDICAL UNIVERSITY
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Spearman’s ρ & Kendall’s τ

③ NON-PARAMETRIC CORRELATION



TAIPEI MEDICAL UNIVERSITY
臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Correlation Coefficient

Parametric vs. Nonparametric

- If the data are **non-normal**, use non-parametric methods.
- **Comparison between correlation methods**
  - Pearson's ***r*** and Spearman's $\boldsymbol{\rho}$ are generally similar sizes.
  - Kendall's $\boldsymbol{\tau}$ is about 66-75% smaller than both Spearman's ρ and Pearson's *r*. (Strahan, 1982)
- $\boldsymbol{\rho^2}$ is usually a good approximation of $\boldsymbol{r^2}$ (esp. no ties & nearly normal distribution).
  - Kendall's $\boldsymbol{\tau^2}$ does not tell us about the proportion of variance shared by two variables.
  - so $\boldsymbol{\tau}$ should not be squared.

Spearman's Correlation Coefficient ρ

H_0: $\rho_s = 0$

H_a: $\rho_s \neq 0$,

FORMULA 10.12. Spearman's rank correlation coefficient (assuming no ties) is

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

where $d_i = r_{X_i} - r_{Y_i}$ is the difference in the rank of X_i and Y_i.

- Spearman's test works by first ranking the data, and then applying Pearson's equation to those ranks, rather than to the original data.
- If there are no ties among Xs and no ties among Ys, ***ρ*** can be computed much more simply than Pearson's ***r***.
- Pearson's ***r*** and Spearman's ***ρ*** are generally similar sizes.

Spearman's ρ

EXAMPLE 10.8. A study by Musch and Hay examined relative age effects in soccer in various northern and southern hemisphere countries. For each country, a sample consisting of all players in the highest professional soccer league was investigated. Their data for Germany are given below. A cut-off date of August 1 applies in Germany. Since participation cut-off dates vary by country, foreign players were excluded from their analysis. For each country, the distribution of professional players' birthdays was computed by month. These birthday distributions were then compared with that of the general population of that country.

Month	Actual players	Expected players	Difference
1	37	28.27	8.73
2	33	27.38	5.62
3	40	26.26	13.74
4	25	27.60	−2.60
5	29	29.16	−0.16
6	33	30.05	2.95
7	28	31.38	−3.38
8	25	31.83	−6.83
9	25	31.16	−6.16
10	23	30.71	−7.71
11	30	30.93	−0.93
12	27	30.27	−3.27

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Spearman's ρ

SOLUTION. To calculate r_s:

1. Separately rank each group of variables as shown in the following table. The data pairs have been ordered using the months in the first column as X's and the differences in the second column as Y's. Their ranks are then listed in the third and fourth columns.

X: month	Y: difference	r_X	r_Y	d_i
1	8.73	1	11	−10
2	5.62	2	10	−8
3	13.74	3	12	−9
4	−2.60	4	6	−2
5	−0.16	5	8	−3
6	2.95	6	9	−3
7	−3.38	7	4	3
8	−6.83	8	2	6
9	−6.16	9	3	6
10	−7.71	10	1	9
11	−0.93	11	7	4
12	−3.27	12	5	7

3. Square these differences and sum them to obtain $\sum_i d_i^2$.
4. Use the Formula 10.12 to obtain r_s,

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

where n is the number of pairs of variables.

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學 TAIPEI MEDICAL UNIVERSITY

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Spearman's ρ

$$
\begin{aligned}
r_s &= 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \\
&= 1 - \frac{6\left[(-10)^2+(-8)^2+(-9)^2+(-2)^2+(-3)^2+(-3)^2+3^2+6^2+6^2+9^2+4^2+7^2\right]}{12(12^2-1)} \\
&= -0.727.
\end{aligned}
$$

with $n = 12$. Let $\alpha = 0.05$. From Table C.13, the critical value for the Spearman rank correlation is 0.587. The table lists only positive critical values. We need to compare the magnitude (absolute value) of r_s to the critical value. Since $0.727 > 0.587$, reject H_0 and accept H_a. The negative Spearman rank correlation ($r_s = -0.727$) is significant and indicates that there is an excess in the number of "goliath" players (those born early in the competition year) and a lack of players born late in the competition year among professional soccer players in Germany.

- If **ρ < 0.306**, there will be no stat. significance.

TAIPEI MEDICAL UNIVERSITY 臺北醫學大學

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Kendall's Correlation Coefficient

τ

The usual test performed is two-sided with the hypotheses being:

- H_0: $\tau = 0$ (or "X and Y are independent")
- H_a: $\tau \neq 0$ (or "X and Y are not independent").

FORMULA 10.11. The Kendall correlation coefficient is defined in terms of the difference between C and D divided by the total number of comparisons,

$$\tau = \frac{C - D}{\frac{n(n-1)}{2}} = \frac{2(C - D)}{n(n-1)}.$$

- If all $\binom{n}{2} = \frac{n(n-1)}{2}$ comparisons are concordant (a "perfect" positive correlation), then $C = \frac{n(n-1)}{2}$ and $D = 0$, so $\tau = +1$, as it should.
- Similarly, if all $\frac{n(n-1)}{2}$ comparisons are discordant (a "perfect" negative correlation), then $D = \frac{n(n-1)}{2}$ and $C = 0$, so $\tau = -1$.
- In all other cases, $-1 < \tau < +1$.
- Finally, ties are not counted in τ since they do not constitute evidence for either a positive or negative correlation.

TAIPEI MEDICAL UNIVERSITY 1960

Kendall's τ

A study by DeMeis and Stearns examined relative age effects on academic and social performance in Geneva. The data in the following table come from one part of their study, which examined the number of students in grades K through 4 evaluated for the district's Gifted and Talented Student Program.

The first and second columns give the month after the cut-off date in which the student was born. The third column lists the number of students in each month cut-off category in grades K through 4 evaluated for the district's Gifted and Talented Student Program.

Remove the ties (equal pairs)

Concordant: both larger
Ex: (190, 186) vs. (182, 185)

Discordant: interaction
Ex: (180, 188) vs. (182, 185)

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Birth month	Month after cut-off	Students evaluated
December	1	53
January	2	47
February	3	32
March	4	42
April	5	35
May	6	32
June	7	37
July	8	38
August	9	27
September	10	24
October	11	29
November	12	27

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Kendall's τ

X_i	Y_i	Rank X_i	Rank Y_i	Concordant pairs below (X_i, Y_i)	Discordant pairs below (X_i, Y_i)	Tied pairs below (X_i, Y_i)
1	53	1	12	0	11	0
2	47	2	11	0	10	0
3	32	3	5.5	4	4	1
4	42	4	10	0	8	0
5	35	5	7	2	5	0
6	32	6	5.5	2	4	0
7	37	7	8	1	4	0
8	38	8	9	0	4	0
9	27	9	2.5	1	1	1
10	24	10	1	2	0	0
11	29	11	4	0	1	0
12	27	12	2.5	0	0	0

$$C = 0 + 0 + 4 + 0 + 2 + 2 + 1 + 0 + 1 + 2 + 0 + 0 = 12.$$

Similarly, $D = 52$, and $E = 2$. So

$$\tau = \frac{2(C - D)}{n(n-1)} = \frac{2(12 - 52)}{12(12-1)} = -0.606.$$

Kendall's τ is interpreted in the same way that Pearson's r is. It appears that there is a moderately strong negative correlation between month after cut-off date and referrals for gifted evaluation.

An Introduction to Biostatistics 2008, 2nd Ed. By Thomas Glover and Kevin Mitchell. Ting Lung Book Co.

Discussion

1. **Parametric correlation**
   - Pearson correlation coefficients

2. **Concerns of correlation**
   - Considerations from the scatter plot

3. **Non-parametric correlation**
   - Spearman & Kendall