1. Correlations and Measures of Association

References

Heiman (2014). Chapters 2 and 7.

Healey (2013). Chapters 11 and 12.

1.1. Level (Scale) of Measurement

• Four levels (Stevens, 1951)

• Three properties for classifying data:

- magnitude: whether the measured trait/attribute can be rank ordered in terms of its intensity

- equal intervals: whether each unit difference between two scale points reflects the same amount of trait/attribute difference

- absolute zero: whether "0" refers to the complete absence of the trait/attribute

1.1.1. Nominal Scale

- Data are classified into mutually exclusive categories
- Qualitative difference
- One-to-one transformation

1.1.2. Ordinal Scale

- Same as nominal, but the categories are rank-ordered
- Difference between categories has no numerical meaning
- Monotonic transformation

1.1.3. Interval Scale

- Same as ordinal, but difference between measurements is meaningful
- Arbitrary zero point
- Linear transformation

1.1.4. Ratio Scale

- Same as interval, but ratio between measurements is meaningful

- Absolute zero point

- Linear transformation through the origin (rescaling)

Scale	Mathematical Properties				Transformation
Nominal	$= \neq$				one-to-one
Ordinal	$= \neq$	$> <$			monotonic
Interval	$= \neq$	$> <$	$+ -$		linear
Ratio	$= \neq$	$> <$	$+ -$	$\times \div$	rescaling

- Different scales require the use of different statistical techniques

1.2. Correlation Analysis

- Direction of influence:

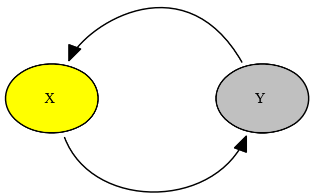
  1. symmetric

     $X \leftrightarrow Y$ (X is correlated with Y)

  2. asymmetric

     $X \rightarrow Y$ (X predicts Y)

  3. reciprocal



(X and Y are mutually reinforcing)

- Level of measurement
- An adequate method should show us
    - direction
    - strength
    - nature of relationship

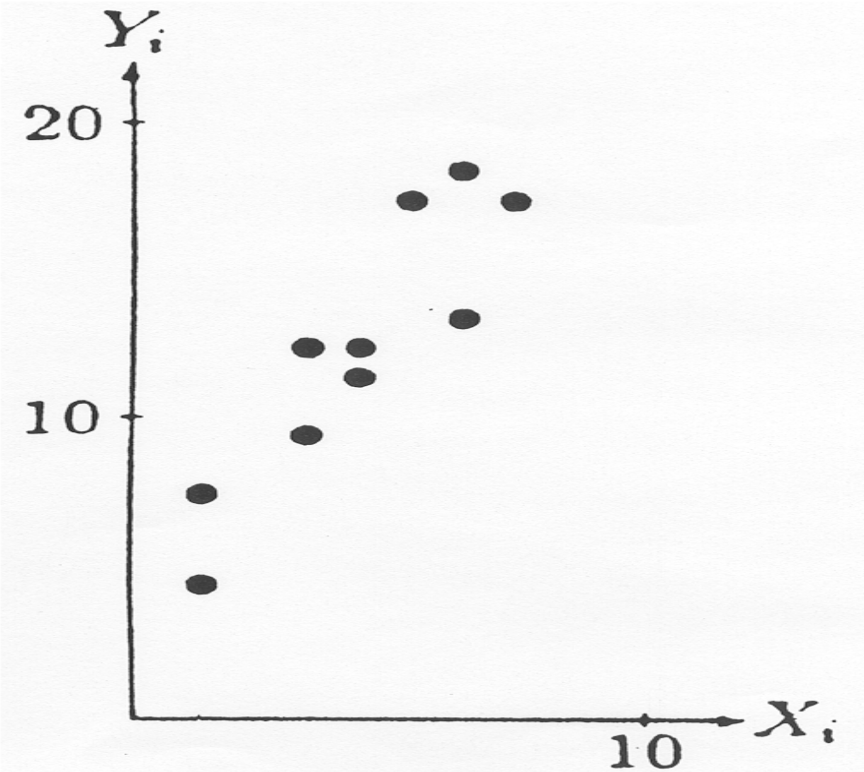
1.2.1. Scatter Plot and Indices

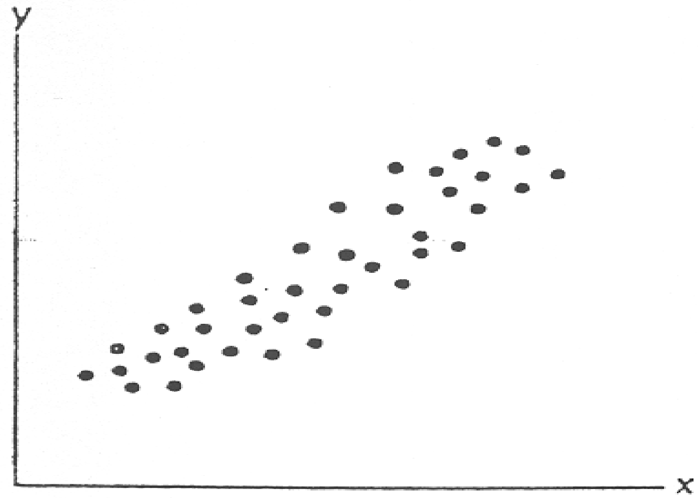
Example 1. A study about the relationship between job commitment (*X*) and job performance (*Y*) (data: example1_1.dat)

Ss	X	Y	$X-\overline{X}$	$Y-\overline{Y}$
1	1	4	-3	-8
2	1	7	-3	-5
3	3	9	-1	-3
4	3	12	-1	0
5	4	11	0	-1
6	4	12	0	0
7	5	17	1	5
8	6	13	2	1
9	6	18	2	6
10	7	17	3	5

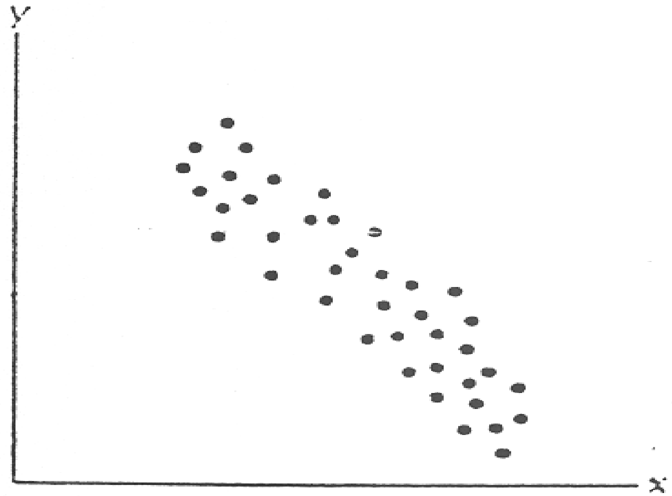
• Methods:

1. Scatter plot

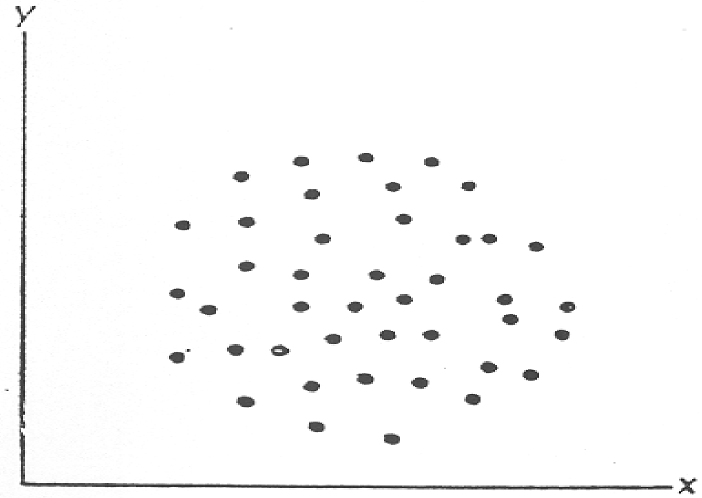




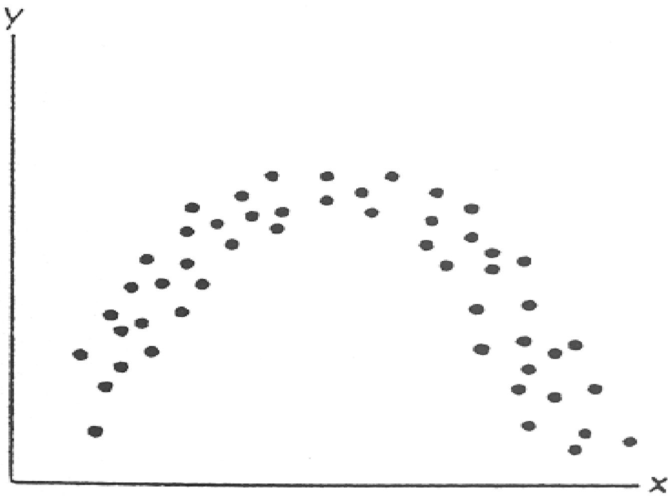
(a) $r > 0$



(b) $r < 0$

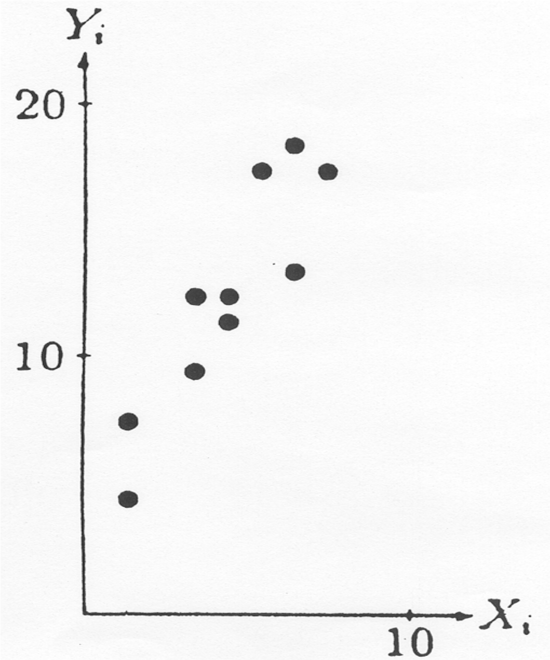


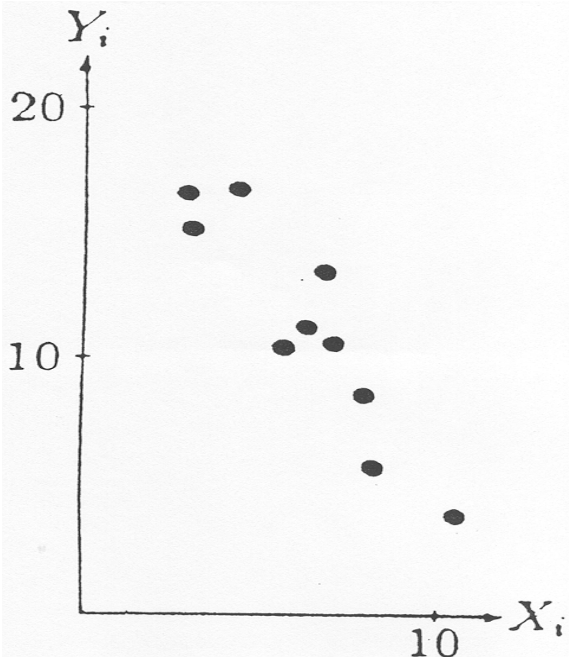
(c) $r \approx 0$



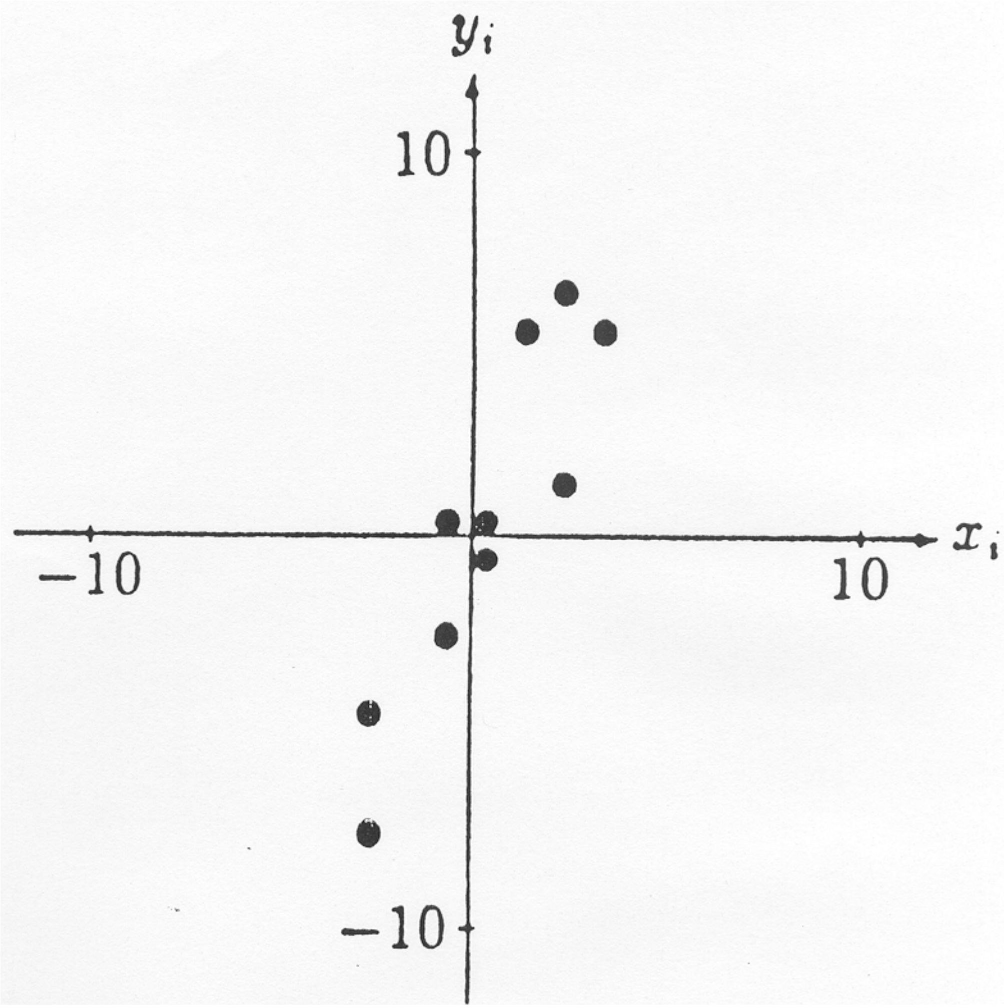
(d) $r \approx 0$

2. Cross-product of raw scores, $\sum XY$





3. Cross-product of centered scores, $\sum(X-\overline{X})(Y-\overline{Y})$



4. Cross-product of Z scores, $\sum \frac{(X-\overline{X})}{S_x} \frac{(Y-\overline{Y})}{S_y}$

5. Average cross-product of Z scores, $\frac{1}{n-1}\sum\frac{(X-\overline{X})}{S_x}\frac{(Y-\overline{Y})}{S_y}$

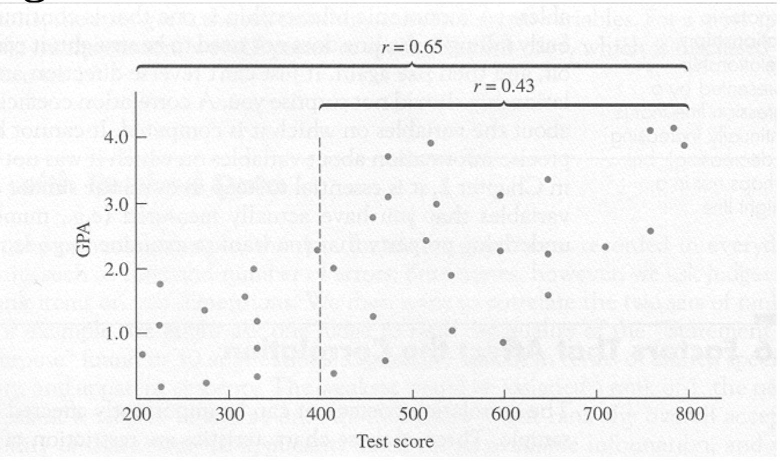
1.2.2. Pearson Product-Moment Correlation Coefficient

$$r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n} z_{x_i} z_{y_i} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}\sqrt{\sum_{i=1}^{n} Y_i^2 - \mathrm{n}\overline{Y}^2}}$$

- Symmetric
- Range: [-1, +1]
- Interval scale or above
- Linear relation
- Descriptive and inferential

1.2.3. Factors Affecting Correlation

- Range restriction



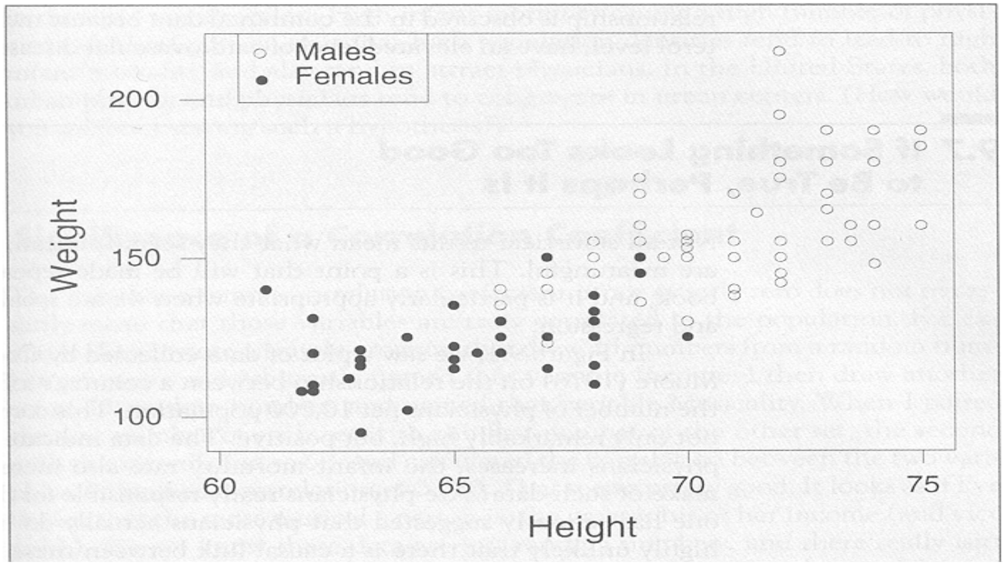
- Correlation corrected for range restriction (r_c):

Assuming identical slopes for both restricted and unrestricted samples:

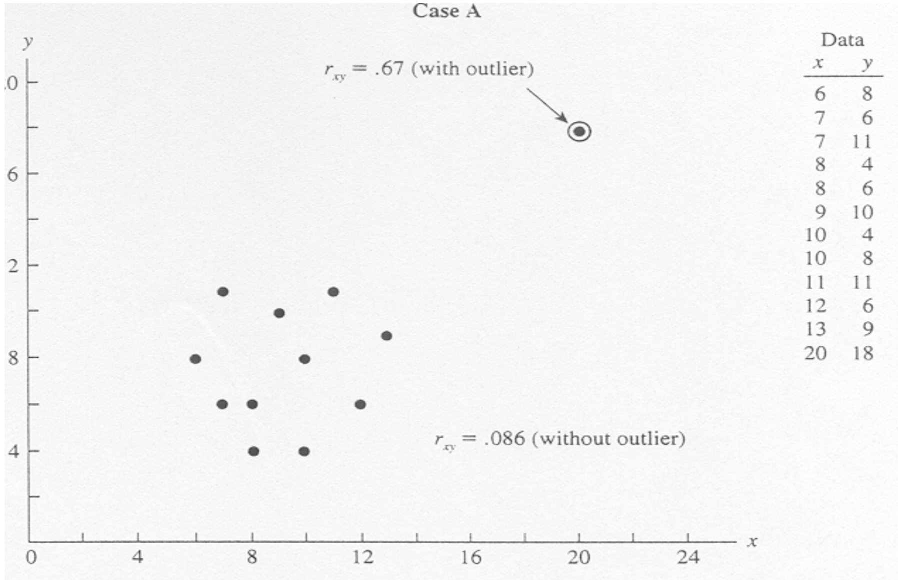
$$r_c = \frac{r}{\sqrt{r^2+(1-r^2)\frac{s_x^2}{s_c^2}}} = \frac{0.43}{\sqrt{.43^2+(1-.43^2)(0.31)}} = 0.65$$

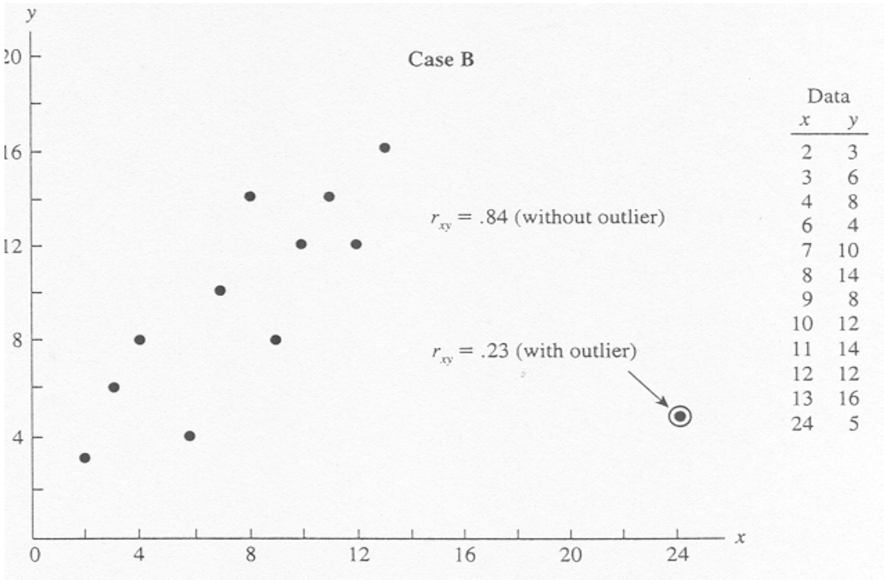
$r =$ correlation between X and Y on restricted sample
$s_x^2 =$ variance of X on restricted sample
$s_c^2 =$ variance of X on unrestricted sample

• Heterogeneous subsamples



• Outliers


Case A
y
x
r_{xy} = .67 (with outlier)
r_{xy} = .086 (without outlier)
0
4
8
12
16
20
24
Data
x y
6 8
7 6
7 11
8 4
8 6
9 10
10 4
10 8
11 11
12 6
13 9
20 18


Case B
y
x
r_{xy} = .84 (without outlier)
r_{xy} = .23 (with outlier)
4
8
12
16
0
4
8
12
16
20
24
Data
x y
2 3
3 6
4 8
6 4
7 10
8 14
9 8
10 12
11 14
12 12
13 16
24 5

1.2.4. Hypothesis Testing

• Case 1. Standard Test

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$ (2-tailed)
$\rho > 0$ or $\rho < 0$ (1-tailed)

assumption: bivariate normal distribution

test statistic: $\mathrm{t} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(\mathrm{df}=n-2)}$

decision: reject H_0 at α level of significance if

$|\mathrm{t}| > t_{(n\text{-}2,\frac{\alpha}{2})}$ (2-tailed)
$\mathrm{t} > t_{(n\text{-}2,\alpha)}$ or $\mathrm{t} < -t_{(n\text{-}2,\alpha)}$ (1-tailed)

drawback: can only test $\rho = 0$

• Example 1 (cont.): example1_1.R

```
# Example 1: Job Commitment and Performance

# set work directory
setwd("c:/users/wchan/google drive/stat6108/data")

# load library Hmisc
library(Hmisc)

# import data
mydata <- read.table("example1_1.dat", header=TRUE)

sink("example1_1.out", split=TRUE)
list(mydata)

# Scatter plot
attach(mydata)
plot(commitment, performance, main="Scatter plot of performance against commitment", xlab="job
commitment", ylab="job performance")

# Person correlation coefficient
cat("\n compute Pearson correlation and its test \n")
rcorr(as.matrix(mydata), type="pearson")
sink()
```

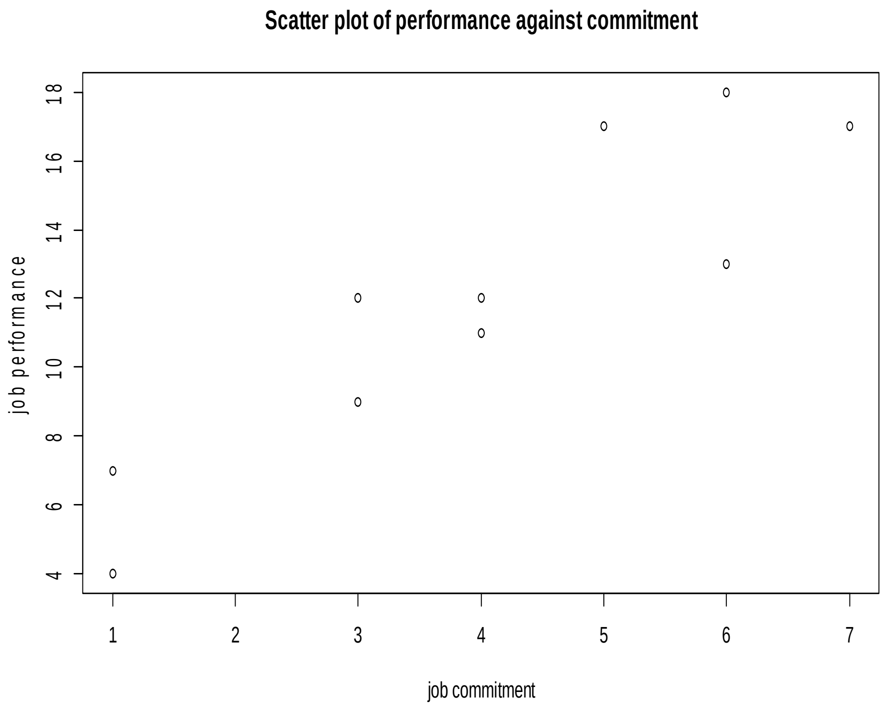
• Example 1 (cont.): example1_1.out

```
[[1]]
   commitment performance
1           1           4
2           1           7
3           3           9
4           3          12
5           4          11
6           4          12
7           5          17
8           6          13
9           6          18
10          7          17


 compute Pearson correlation and its test
            commitment performance
commitment         1.0         0.9
performance        0.9         1.0

n= 10


P
            commitment performance
commitment             3e-04
performance 3e-04
```


Scatter plot of performance against commitment
job performance
4
6
8
10
12
14
16
18
1
2
3
4
5
6
7
job commitment

- **Case 2. Testing Nonzero ρ**

H_o: $\rho = \rho_o$
H_1: $\rho \neq \rho_o$ (2-tailed)
$\rho > \rho_o$ or $\rho < \rho_o$ (1-tailed)

define: $g(\rho) = \frac{1}{2}\ln(\frac{1+\rho}{1-\rho})$ *Fisher transformation*

assumptions: large sample size, bivariate normal distribution

test statistic: $z = \sqrt{n-3}\,(g(r) - g(\rho_o)) \sim N(0,1)$

decision: reject H_o at α level of significance if

$|z| > Z_{(\frac{\alpha}{2})}$ (two-tailed)
$z > Z_\alpha$ or $z < -Z_\alpha$ (one-tailed)

• Example 2. The correlation between motivation and income for a sample of 30 females is 0.3.

Test: $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$

Method 1: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

Conclusion?

Method 2: $z = \sqrt{n-3}\,(g(r) - g(0))$

Conclusion?

- **Case 3. Comparing ρ from Two Independent Samples**

H_o: $\rho_1 = \rho_2$
H_1: $\rho_1 \neq \rho_2$ (two-tailed)
$\rho_1 > \rho_2$ or $\rho_1 < \rho_2$ (one-tailed)

assumptions: large samples and normal distributions

test statistic: $z = \frac{(g(r_1) - g(r_2)) - (g(\rho_1) - g(\rho_2))}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \sim N(0,1)$

sample 1: $(r_1; n_1)$ sample 2: $(r_2; n_2)$

decision: reject H_o at α level of significance if

$|z| > Z_{(\frac{\alpha}{2})}$ (two-tailed)
$z > Z_\alpha$ or $z <$ -Z_α (one-tailed)

• Example 2 (cont.). From a sample of 50 males, the correlation between motivation and income is 0.7. Is the difference in correlation between male and female significant?

Case 4. Testing Different Correlations within a Sample

• Example 3. Let $I =$ income, $J =$ job performance (extrinsic variables);
$K =$ motivation, $L =$ job commitment (intrinsic variables).

Which pair of variables has a stronger relationship: ρ_{ij} or ρ_{kl}?

H_o: $\rho_{ij} = \rho_{kl}$
H_1: $\rho_{ij} \neq \rho_{kl}$

assumptions: large sample size, multivariate normal distribution

test statistic: $\frac{1}{\sigma}\left[(r_{ij} - r_{kl}) - (\rho_{ij} - \rho_{kl})\right] \sim N(0, 1)$

- Formulas (Olkin & Finn, 1995; *Psychological Bulletin*) :

$$\sigma^2 = \text{var}(r_{ij} - r_{kl}) = \text{var}(r_{ij}) + \text{var}(r_{kl}) - 2\text{cov}(r_{ij}, r_{kl})$$

$$\text{var}(r_{ij}) = (1 - \rho_{ij}^2)^2/n$$

$$\begin{aligned}\text{cov}(r_{ij}, r_{kl}) = [&\tfrac{1}{2}\rho_{ij}\rho_{kl}(\rho_{ik}^2 + \rho_{il}^2 + \rho_{jk}^2 + \rho_{jl}^2) + \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk} \\ &- (\rho_{ij}\rho_{ik}\rho_{il} + \rho_{ji}\rho_{jk}\rho_{jl} + \rho_{ki}\rho_{kj}\rho_{kl} + \rho_{li}\rho_{lj}\rho_{lk})]/n\end{aligned}$$

- Cheung & Chan (2004; *ORM*) used SEM technique to test dependent correlations

1.2.5. Conditional Relationships

• Three intercorrelated variables X, Y, and Z with correlation matrix

$$\begin{pmatrix} 1.00 & & \\ r_{yx} & 1.00 & \\ r_{xz} & r_{yz} & 1.00 \end{pmatrix}$$

• Suppose Z has an effect that is in some sense prior to X and Y, so it influences both these variables but is not influenced by them

$$r_{yx} = r_{\text{due to } z} + r_{\text{unique } yx}$$

• *Partial correlation* ($r_{yx.z}$) measures the unique (linear) relationship between Y and X given the effect of Z has been removed from both Y and X. That is,

$$r_{yx.z} = \frac{r_{yx} - r_{yz} r_{xz}}{\sqrt{1 - r_{yz}^2}\sqrt{1 - r_{xz}^2}}$$

- Example 4. Detecting spurious relationship

$X =$ income $\qquad Y =$ health $\qquad Z =$ age

$r_{yx} =$ -.35 $\qquad r_{xz} =$ .60 $\qquad r_{yz} =$ -.50

$r_{yx.z} =$

- Example 5. Discovering hidden relationship

$X =$ income $\qquad Y =$ job satisfaction $\qquad Z =$ working hours

$r_{yx} =$ .00 $\qquad r_{xz} =$ .60 $\qquad r_{yz} =$ -.50

$r_{yx.z} =$

1.3. Other Measures and Tests of Association

	Interval/ Ratio	Ordinal	Nominal
Interval/ Ratio	Pearson's r		
Ordinal		Spearman's r_s Gamma, G	
Nominal			Cramer's V Lambda, L

1.3.1. Spearman's Rank Order Correlation (ρ_s)

• When two variables (X and Y) are ordinal, we can use Spearman's rank-order correlation (ρ_s) to measure their relationship

X	Y	rank(X)	rank(Y)	d
1	1	1	1	0
2	4	2	2	0
3	9	3	3	0
4	16	4	4	0

• If there are no ties in ranks for both X and Y, the sample estimate of ρ_s is

$$r_{\mathrm{s}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where $d_i = \text{rank}(x_i) - \text{rank}(y_i)$.

• r_{s} can be understood as the Pearson r for ranks

• When ties exist, the formula will *inflate* the value of r_s

• Correction of ties:

$$r_\mathrm{s} = \frac{(n^3-n)-6\Sigma d^2-(T_x+T_y)/2}{\sqrt{(n^3-n)^2-(T_x+T_y)(n^3-n)+T_xT_y}}$$

such that $T_x = \sum_{i=1}^{g}(t_i^3 - t_i)$, g is the number of groupings of different tied ranks and t_i is the number of tied ranks in the ith grouping.

• Small effect of ties when g and/or t_i is small

1.3.2. Statistical Test

H_o: $\rho_s = 0$
H_1: $\rho_s \neq 0$ (2-tailed)
$\rho_s > 0$ or $\rho_s < 0$ (1-tailed)

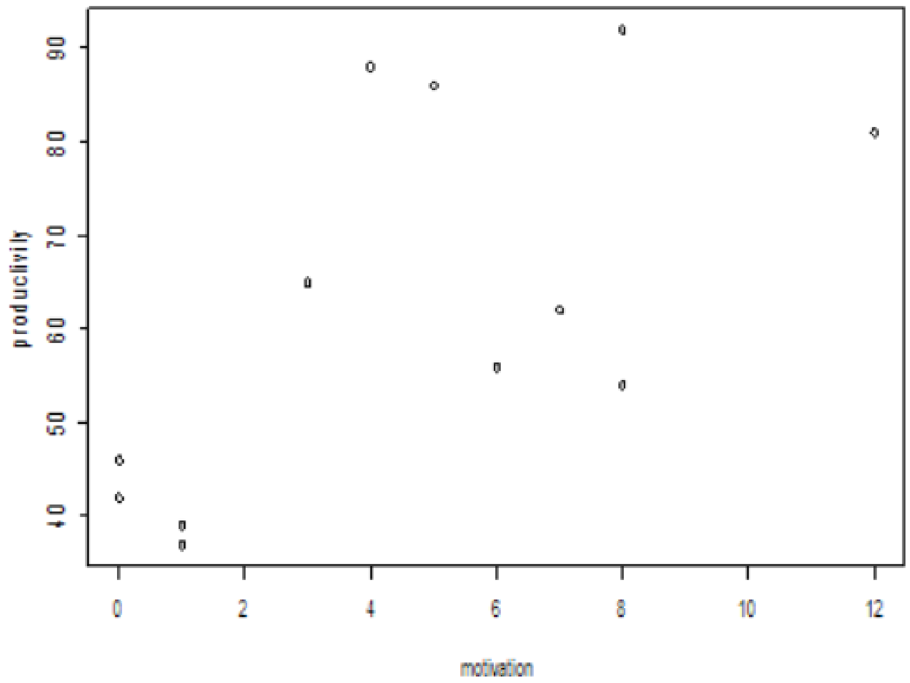
test statistic: $\mathrm{t} = r_s\sqrt{\frac{n-2}{1-r_s^2}} \sim t_{(\mathrm{df}=n-2)}$

decision: reject H_0 at α level of significance if

$|\mathrm{t}| > t_{(n\text{-}2,\frac{\alpha}{2})}$ (2-tailed)
$\mathrm{t} > t_{(n-2,\alpha)}$ (+ve association)
$\mathrm{t} < -t_{(n-2,\alpha)}$ (-ve association)

- Example 6. Motivation and Productivity (data: example1_6.dat)

	Motivation		Productivity			
Subject	Data	Rank	Data	Rank	d_i	d_i^2
A	0	1.5	42	3	−1.5	2.25
B	0	1.5	46	4	−2.5	6.25
C	1	3.5	39	2	1.5	2.25
D	1	3.5	37	1	2.5	6.25
E	3	5	65	8	−3.0	9.00
F	4	6	88	11	−5.0	25.00
G	5	7	86	10	−3.0	9.00
H	6	8	56	6	2.0	4.00
I	7	9	62	7	2.0	4.00
J	8	10.5	92	12	−1.5	2.25
K	8	10.5	54	5	−5.5	30.25
L	12	12	81	9	3.0	9.00
						$\Sigma d_i^2 = 109.50$



• Example 6 (cont.): example1_6.R

```
# Example 6: Motivation and Productivity

# set work directory
setwd("c:/users/wchan/google drive/stat6108/data")

# load library Hmisc
library(Hmisc)

# import data
mydata <- read.table("example1_6.dat", header=TRUE)

sink("example1_6.out", split=TRUE)
list(mydata)
cat("\n compute Spearman correlation and its test \n")
rcorr(as.matrix(mydata), type="spearman")
sink()
```

• Example 6 (cont.): example1_6.out

```
   motivation productivity
1           0           42
2           0           46
3           1           39
4           1           37
5           3           65
6           4           88
7           5           86
8           6           56
9           7           62
10          8           92
11          8           54
12         12           81
```

```
 compute Spearman correlation and its test
             motivation productivity
motivation         1.00         0.62
productivity       0.62         1.00

n= 12


P
             motivation productivity
motivation              0.0333
productivity 0.0333
```

1.3.3. Gamma (γ)

- Spearman's r_s becomes less and less useful when there are too many ties

- Example 7. Productivity and Company Image (data: example1_7.dat)

Productivity (X)	Company Image (Y)
low	low
⋮	⋮
low	moderate
⋮	⋮
low	high
⋮	⋮
moderate	low
⋮	⋮
moderate	moderate
⋮	⋮
moderate	high
⋮	⋮
high	low
⋮	⋮
high	moderate
⋮	⋮
high	high
⋮	⋮

Company Image

Productivity	Low	Moderate	High
Low	10	5	2
Moderate	8	9	7
High	2	6	8

- γ measures the relationship between the row variable (X) and the column variable (Y) in a contingency table when both of them are in ordinal scales.

- Sample estimate of γ is

$$G = \frac{\text{\# of concordant pairs } (C) - \text{\# of discordant pairs } (D)}{\text{\# of concordant pairs } (C) + \text{\# of discordant pairs } (D)}$$

- $G \in [-1, +1]$

- How to count C and D?

B	f_B	A	f_A	No. of Pairs
(L,L)	10	(M,M)	9	90
		(M,H)	7	70
		(H,M)	6	60
		(H,H)	8	80
(L,M)	5	(M,H)	7	35
		(H,H)	8	40
(M,L)	8	(H,M)	6	48
		(H,H)	8	64
(M,M)	9	(H,H)	8	72
				C=559

B	f_B	A	f_A	No. of Pairs
(L,H)	2	(M,L)	8	16
		(M,M)	9	18
		(H,L)	2	4
		(H,M)	6	12
(L,M)	5	(M,L)	8	40
		(H,L)	2	10
(M,H)	7	(H,L)	2	14
		(H,M)	6	42
(M,M)	9	(H,L)	2	18
				D=174

1.3.4. Statistical Test

H_o: $\gamma = \gamma_o$
H_1: $\gamma \neq \gamma_o$ (2-tailed)
$\gamma > \gamma_o$ or $\gamma < \gamma_o$ (1-tailed)

assumption: large sample size

test statistic: $z = \frac{G-\gamma_o}{\sqrt{\mathrm{var}(G)}} \sim N(0,1)$

where var(G) is the asymptotic variance of G (Goodman & Kruskal, 1963)

decision: reject H_o at α level of significance if

$|z| > Z_{(\frac{\alpha}{2})}$ (H_1: $\gamma \neq \gamma_o$)
$z > Z_{\alpha}$ (H_1: $\gamma > \gamma_o$)
$z < -Z_{\alpha}$ (H_1: $\gamma < \gamma_o$)

• Example 7 (cont.): example1_7.R

productivity	image	count
1_Low	1_Low	10
1_Low	2_Moderate	5
1_Low	3_High	2
2_Moderate	1_Low	8
2_Moderate	2_Moderate	9
2_Moderate	3_High	7
3_High	1_Low	2
3_High	2_Moderate	6
3_High	3_High	8

```
# Example 7: Productivity and Company Image

# set work directory
setwd("c:/users/wchan/google drive/stat6108/data")

# load library vcdExtra
library(vcdExtra)

# import data
mydata <- read.table("example1_7.dat", header=TRUE)

# How to create cross classification table?

# Method 1: From Raw Data set
table1 <- table(mydata$productivity, mydata$image)

# Method 2: From Data Set Weighted by Frequency
table2 <- xtabs(count ~ productivity+image, data=mydata)

# Method 3: Create a table directly
table3 <- matrix(c(10, 5, 2, 8, 9, 7, 2, 6, 8), nrow=3, ncol=3, byrow=TRUE)
colnames(table3) <- c("Low","Moderate","High")
rownames(table3) <- c("Low","Moderate","High")
table3 <- as.table(table3)

sink("example1_7.out", split=TRUE)
list(table1,table2,table3)
cat("\n Goodman and Kruskal gamma coefficient \n")
GKgamma(table2, level = 0.95)
sink()
```

• Example 7 (cont.): example1_7.out

```
[[1]]

             1_Low 2_Moderate 3_High
  1_Low          1          1      1
  2_Moderate     1          1      1
  3_High         1          1      1

[[2]]
            image
productivity 1_Low 2_Moderate 3_High
  1_Low         10          5      2
  2_Moderate     8          9      7
  3_High         2          6      8

[[3]]
         Low Moderate High
Low       10        5    2
Moderate   8        9    7
High       2        6    8


 Goodman and Kruskal gamma coefficient
gamma        : 0.525
std. error   : 0.137
CI           : 0.257 0.794
```

1.3.5. Cramer's V

- When two variables, S and T, are nominal, we use Cramer's V to measure the association between them

- Example 8. Soda Preference (data: example1_8.dat)

Soda Preference

Gender	Coke	Pepsi	Coke Light
Male	60	20	30
Female	10	10	70

- $V = \sqrt{\frac{X^2}{n(k-1)}}$ where $k = \min(r, c)$

$X^2 = \sum_{\text{cells}} \frac{(f_o - f_e)^2}{f_e}$ where f_o=observed frequency

f_e=expected frequency

- $V \in [0, 1]$

1.3.6. Statistical Test

H_o: S and T are independent
H_1: S and T are not independent

test statistic: 1. Pearson chi-square:

$$X^2 = \sum_{\text{cells}} \frac{(f_o - f_e)^2}{f_e} \sim \chi^2(\text{df}=(r-1)(c-1))$$

2. Likelihood ratio G^2:

$$G^2 = 2 \sum_{\text{cells}} f_o \ln(\frac{f_o}{f_e}) \sim \chi^2(\text{df}=(r-1)(c-1))$$

assumption: large sample size

decision: reject H_o at α level of significance if

$X^2 > \chi^2_\alpha(\text{df})$ Pearson chi-sq. test
$G^2 > \chi^2_\alpha(\text{df})$ Likelihood ratio test

remarks: when n is large, $X^2 \simeq G^2$
require $f_e > 5$ for each cell

• Example 8 (cont.): example1_8.R

```
# Example 8: Soda Preference

# set work directory
setwd("c:/users/wchan/google drive/stat6108/data")

# load library vcdExtra
library(vcdExtra)

# Create a table directly
table <- matrix(c(60, 20, 30, 10, 10, 70), nrow=2, byrow=TRUE)
rownames(table) <- c("Male", "Female")
colnames(table) <- c("Coke","Pepsi","Coke Light")
table <- as.table(table)

sink("example1_8.out", split=TRUE)
writeLines("\n Print table \n")
table
writeLines("\n Row Totals \n")
margin.table(table,1)
writeLines("\n Column Totals \n")
margin.table(table,2)
writeLines("\n Cramer's V coefficient \n")
assocstats(table)
sink()
```

• Example 8 (cont.): example1_8.out

```
 Print table

       Coke Pepsi Coke Light
Male     60    20         30
Female   10    10         70

 Row Totals

  Male Female
   110     90

 Column Totals

      Coke      Pepsi Coke Light
        70         30        100

 Cramer's V coefficient

                    X^2 df   P(> X^2)
Likelihood Ratio 57.476  2 3.3062e-13
Pearson          53.583  2 2.3147e-12

Phi-Coefficient   : NA
Contingency Coeff.: 0.46
Cramer's V        : 0.518
```

1.3.7. Lambda (λ)

• An asymmetric type of association

• Data consist of antecedent-consequent pairs

• Example 9. Salary and Transportation (data: example1_9.dat)

	Salary (X)			
Transportation (Y)	decrease	no change	increase	Total
walking	10	1	4	15
bus	5	3	6	14
minibus	3	12	2	17
taxi	3	3	8	14
Total	21	19	20	60

- Treating Y as outcome, sample estimate of λ is

$$L_Y = \frac{\mathrm{E}_Y - \mathrm{E}_{Y\backslash X}}{\mathrm{E}_Y}$$

where E_Y = no. of errors in Y
$\mathrm{E}_{Y\backslash X}$ = no. of errors in Y given X

- Treating X as outcome, sample estimate of λ is

$$L_X = \frac{\mathrm{E}_X - \mathrm{E}_{X\backslash Y}}{\mathrm{E}_X}$$

where E_X = no. of errors in X
$\mathrm{E}_{X\backslash Y}$ = no. of errors in X given Y

- Use symmetric λ if one cannot identify the outcome variable,

$$L_{sym} = \frac{(\mathrm{E}_Y + \mathrm{E}_X) - (\mathrm{E}_{Y\backslash X} + \mathrm{E}_{X\backslash Y})}{\mathrm{E}_Y + \mathrm{E}_X}$$

- λ is a PRE (proportional reduction in error) measure

• Example 9 (cont.): example1_9.R

```
# Example 9: Salary and Transportation

# set work directory
setwd("c:/users/wchan/google drive/stat6108/data")

# load library DescTools
library(DescTools)

# import data
mydata <- read.table("example1_9.dat", header=TRUE)

# Create a table
mytable <- xtabs(count ~ transportation+salary, data=mydata)
rownames(mytable) <- c("walking", "bus", "minibus", "taxi")
colnames(mytable) <- c("decrease","no change","increase")

sink("example1_9.out", split=TRUE)
writeLines("\n Print Data Set \n")
mydata
writeLines("\n Print table \n")
mytable
writeLines("\n Lambda coefficient: Row variable (transportation) as outcome \n")
Lambda(mytable, direction="row", conf.level=.95)
writeLines("\n Lambda coefficient: Column variable (salary) as outcome \n")
Lambda(mytable, direction="column", conf.level=.95)
writeLines("\n Symmetric Lambda coefficient \n")
Lambda(mytable, direction="symmetric", conf.level=.95)
sink()
```

• Example 9 (cont.): example1_9.out

```
Print Data Set
   transportation salary count
1               1      1    10
2               1      2     1
3               1      3     4
4               2      1     5
5               2      2     3
6               2      3     6
7               3      1     3
8               3      2    12
9               3      3     2
10              4      1     3
11              4      2     3
12              4      3     8

 Print table
              salary
transportation decrease no change increase
       walking       10         1        4
       bus            5         3        6
       minibus        3        12        2
       taxi           3         3        8

 Lambda coefficient: Row variable (transportation) as outcome
   lambda    lwr.ci    upr.ci
0.3023256 0.1197385 0.4849126

 Lambda coefficient: Column variable (salary) as outcome
   lambda    lwr.ci    upr.ci
0.3846154 0.1448107 0.6244201

 Symmetric Lambda coefficient
   lambda    lwr.ci    upr.ci
0.3414634 0.1576402 0.5252866
```