

DEPARTMENT OF EPIDEMIOLOGY  
SCHOOL OF PUBLIC HEALTH  
UNIVERSITY OF NORTH CAROLINA

EPID 160

PRINCIPLES OF EPIDEMIOLOGY

Fall Semester

OBJECTIVE: To Illustrate Some Principles of the Scientific Method:  
The Interrelationship of Variables

I. Associations Between Two Variables:

For purposes of this course an association is said to exist between two variables when changes in one are accompanied by changes in the other.

Example (1)

Impairments by Age: (U. S., 1959)

Age (in years)	Impairments: Rate per 1000 Population
Under 25	52.9
25 - 44	130.6
45 - 64	212.4
65 - 74	376.6
75 and over	615.0

Source: U. S. National Health Survey, 1959.

In this example the variables age and impairments are associated in that increasing age is accompanied by an increasing rate of impairment.

Example (2)

Impairments by Sex (U. S., 1959)

Sex	Impairments: Rate per 1000 Population
Both Sexes	141.4
Male	160.8
Female	123.1

Source: U. S. National Health Survey, 1959.

In this example impairments are associated with sex in that the impairment rate varies when the sex varies.

Exercise 1.

Number of Persons Injured per 1000 Persons  
Per Year by Age (U.S. July-December 1957)

Age (in years)	Number Injured Per 1000 Persons Per Year
Under 5	244
5 - 14	370
15 - 24	387 *
25 - 44	255
45 - 64	274 *
65+	265
All Ages	298

*Bimodal*

Source: U. S. National Health Survey, May, 1958.

What is the nature of the association demonstrated in the above table?

Exercise 2.

Infant Mortality from Bronchitis and Pneumonia by Family Size

Number of Children Mother Has Borne	Number of Infant Deaths (28 days-1 Year) Per 1000 Live Births Per Year
1	3.2
2	4.6
3	7.6
4	9.4
5 and over	13.8

Source: Morris, J. N., Uses of Epidemiology, E. & S. Livingstone, London, 1957, p. 79.

What is the nature of the association demonstrated in the above table?

Exercise 3.

Number of Duodenal Ulcers Per 100,000 Population Per Year  
by Age for Each Sex. (City of York 1952 - 1957)

MEN	
Age (in years)	Number of Duodenal Ulcers Per 100,000 Population Per Year
15 - 24	143
25 - 34	268
35 - 44	237
45 - 54	220
55 - 64	247
65 and over	148
All Ages	215

WOMEN	
Age (in years)	Number of Duodenal Ulcers Per 100,000 Population Per Year
15 - 24	37
25 - 34	55
35 - 44	87
45 - 54	71
55 - 64	40
65 and over	38
All Ages	57

What is the nature of the associations demonstrated in the above tables?

II. Interrelationships of More Than Two Variables:

In epidemiological studies some of the most valuable clues are obtained from analysis of the relationships between 3 or more variables.

The following examples are presented to illustrate the various types of interrelationships that may exist between 3 variables.

Example (3)

Some Variables Associated With Birth Weight of Male Infants\*

Table 3A. Birth Weight by Race

Race	Mean Birth Weight
Negro	3102 grams
White	3457 grams

What conclusions can be drawn from these data?

Table 3B. Birth Weight by Income Level

Income	Mean Birth Weight
Low income groups	3300 grams
High income groups	3500 grams

*only one variable mentioned*

What conclusions can be drawn from the data presented in Table 3B?

Taking the information from Tables A and B together we can now say that both race and income are associated with birth weight. Specifically, the association is that Negro race and low income groups are associated with a low birth weight. White race and high income groups are associated with a high birth weight.

This might be diagrammed as follows:



The question that now must be asked is whether race is associated with birth weight because of some purely racial factor or whether race is associated with birth weight only because there is an undue concentration of one of the income groups in one race.

To put this in other words: Do Negro babies tend to have low birth weights because of some racial (genetic) factor common to Negroes, or is it because most Negroes are in the low income groups, and it is the low income that forms the curcial association with low birth weight?

Exactly the same question could be asked about the association of low income groups and low birth weight. Is this association brought about by virtue of some of the things in the way of life of low income groups that differs from the way of life of high income groups, or does this association exist because most of the low income groups from which these figures were drawn happen to be Negroes?

---

\* The figures in Tables 3A, B, and C are approximations derived from several different studies.

The answer to these questions are obtained by controlling each variable. This means comparing the birth weights of Negro babies in low income groups with the birth weight of white babies in low income groups, then comparing Negro birth weights in high income groups with white birth weights in high income groups. In this way the income group variable has been held constant or controlled. The same data can be examined to find out the effect of holding race constant.

Table 3C. Mean Birth Weight by Race and Income Groups

Race	Low Income Group	High Income Group
Negro	3300 grams	3482 grams
White	3319 grams	3469 grams

You will notice in Table 3C that 3 variables - Race, Income Group, and Birth weight - are all included.

What conclusions can be drawn from Table C?

Example (4)

Some Variables Associated with Prematurity

Table 4A. Incidence of Prematurity in Relation to Work During Pregnancy

	Number	Percentage of Births Which Were Premature
Mothers of single first births	1318	6.8
Mothers not gainfully employed during pregnancy	780	4.7
Mothers gainfully employed for less than 28 weeks of pregnancy	285	8.4
Mothers gainfully employed for more than 28 weeks of pregnancy	253	11.1

1307.46  
743.74  
262.06  
28.08 224.92

Source: Stewart A., "A Note on the Obstetric Effects of Work During Pregnancy." Br. J. Prev. and Soc. Med., 9: p. 159, July, 1955.

What association is demonstrated by Table 4A?

Table 4B. Incidence of Prematurity According to Mothers' Social Class

Social Class	Percentage of Live Births Born Prematurely
Highest social class	5.5
Lowest social class	7.4

Source: Rider, Rowland V., et. al., "Associations Between Premature Births and Socio-Economic Status." Am. J. P. H., 45: p. 1022: 1955.

What association is demonstrated by Table 4B?

As in the previous example (Tables 3A, B, and C) we now have interrelationships between 3 variables. Both work during pregnancy and social class are seen to be associated with prematurity. Specifically, mothers who work during pregnancy and low social class mothers have the highest prematurity rates.

Once again the questions must be asked as to whether mothers who work during pregnancy have a higher rate of premature births than do housewives because of some aspect of working, or because working mothers are more likely to be of low social class. We therefore have to control for the variables social class and working mothers.

Table 4C. Incidence of Prematurity According to Mothers' Social Class for Working and Non-Working Mothers

Social Class	Percentage Premature Births		
	Mothers not gainfully employed	Mothers employed for less than 28 weeks of pregnancy	Mothers employed for more than 28 weeks of pregnancy
Highest	2.6	3.3	5.9
Middle	4.0	8.5	10.4
Lowest	7.8	10.4	13.7

*Both variables working*

Source: Stewart, A., "A Note on the Obstetric Effects of Work During Pregnancy." Br. J. Prev. Med., 9: p. 159, 1955.

What conclusions can be drawn from Table 4C?

How do the relationships between the 3 variables - Working Mothers, Social Class, and Prematurity - differ from the relationships between the three variables - Race, Social Class, and Low Birth Weight (from Example 3)?

Example (5)

Some Variables Associated with Rheumatoid Arthritis

*In ER notes*

Table 5A. Prevalence of Rheumatoid Arthritis by Education

Education Level	Rheumatoid Arthritis per 100 Population
Less than 5th grade	15.4
5th - 8th grade	6.3
9th grade and over	4.1

Describe the association demonstrated in Table 5A.

Table 5B. Prevalence of Rheumatoid Arthritis by Income Level

Income Level	Rheumatoid Arthritis per 100 Population
Less than \$3000 per year	7.4
\$3000 - \$4499	5.1
\$4500 and over	3.2

Describe the association demonstrated in Table 5B.

From the accumulated data of Tables 5A and 5B, what is the next question that must be asked? How can this questions be answered?

Table 5C. Prevalence of Rheumatoid Arthritis by Education and Income

Income	Education		
	Less than 5th grade	5th-8th grade	9th grade & over
Less than \$3000	6.5	4.5	10.0
More than \$3000	38.6	4.6	2.0

Source: (Tables 5A, B, and C) King, Stanley H., and Sidney Cobb. "Psychosocial Factors in the Epidemiology of Rheumatoid Arthritis." J. Chronic Dis., 7: p. 466, 1958.

What are the conclusions that can be drawn from Table 5C?

How do the relationships between the three variables - Education, Income, and Rheumatoid Arthritis - differ from the relationships between the three variables of Example (3) and the three variables of Example (4)?