

# Regression Tools

$$\bar{y} = 85.0$$

$$\bar{x} = 24.8$$

$$SS(x) = 1,743.64$$

$$SS(y) = 16,976.00$$

$$SS(xy) = 2,766.00$$

So we can calculate

$$\text{Slope} = b = \frac{SS(xy)}{SS(x)} = \frac{2,766.00}{1,743.64} = 1.59$$

$$\begin{aligned}\text{Intercept} = a &= \bar{y} - b\bar{x} \\ &= 85.0 + 1.59(24.8) \\ &= 45.54\end{aligned}$$

The straight line depicting the regression relationship of  $y$  on  $x$  is

$$\begin{aligned}\hat{y} &= a + bx \\ \hat{y} &= 45.54 + 1.59x\end{aligned}$$

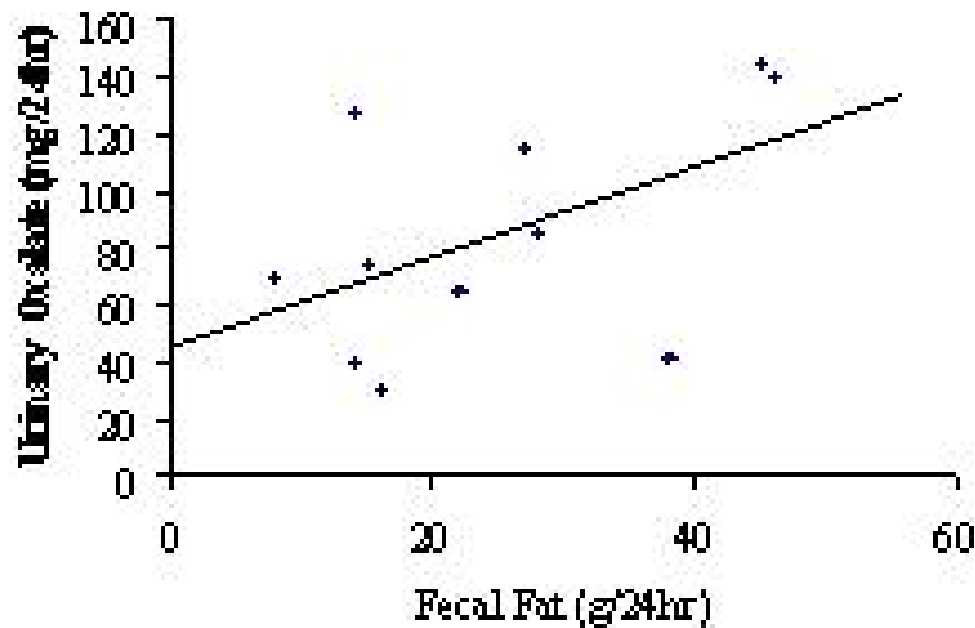
At  $x = 40$ , the regression estimate for  $y$  is:

$$\hat{y} = 45.54 + 1.59(40) = 109.14$$

With this information, we can add the regression line

$$\hat{y} = 45.54 + 1.59x$$

to the scatter plot, as shown below.



$$\hat{y} = 45.54 + 1.59(x = 45) = 117.09$$

## Hypothesis test for regression of urinary oxalate on fecal fat

1. *The Hypothesis:*  $H_0: \beta = 0$  vs  $H_1: \beta \neq 0$
2. *The  $\alpha$  level:*  $\alpha = 0.05$
3. *The assumptions:* Random normal samples for y-variable from populations defined by x-variable
4. *The test statistic:* ANOVA as specified by

ANOVA				
Source	df	SS	MS	F
Regression	1	bSS(xy)	SS(Reg)/1	MS(Reg)/MS(Res)
Residual	n-2	SS(Res) <sup>a</sup>	SS(Res)/(n-2)	
Total	n-1	SS(y)		

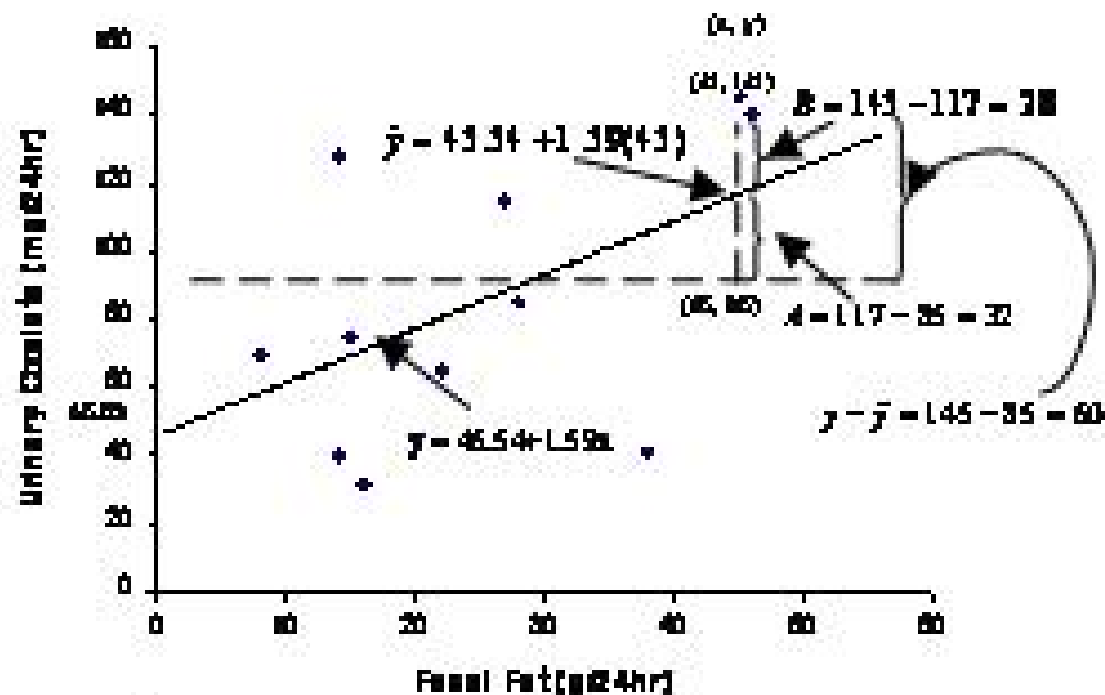
<sup>a</sup>SS(Residual) = SS(y) - SS(Regression)

5. *The critical region:* Reject  $H_0: \beta = 0$  if the value calculated for  $F$  is greater than  $F_{0.95}(1,9) = 5.12$

6. *The result:*  
 $SS(\text{Reg}) = bSS(xy) = 1.59(2,766.00) = 4,397.94$   
 $SS(\text{Total}) = SS(y) = 16,976.00$   
 $SS(\text{Res}) = 16,976.00 - 4,397.94 = 12,578.06$

ANOVA				
Source	df	SS	MS	F
Regression	1	4,397.94	4,397.94	3.15
Residual	9	12,578.06	1,397.56	
Total	10	16,976.00		

7. *The conclusion:* Accept  $H_0: \beta = 0$  since  $F < 5.12$



$$R^2 = 0.2985$$

$$\hat{y} = \text{regression estimate} = 45.54 + 1.59x = 117$$

$$C = A + B = y - \bar{y} \text{ Total deviation} \Rightarrow SS(y)$$

$$A = \hat{y} - \bar{y} \text{ Explained by line} \Rightarrow SS(\text{Reg})$$

$$B = y - \hat{y} \text{ Left over} \Rightarrow SS(\text{Residual})$$

## Correlation Tools

$$SS(x) = 1,743.64$$

$$SS(y) = 16,976.00$$

$$SS(xy) = 2,766.00$$

The estimate of the correlation coefficient is:

$$r_{xy} = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}}$$

$$r_{xy} = \frac{2,766.00}{\sqrt{[1,743.64][16,976.00]}} = 0.5084$$

# The Correlation Coefficient

$r$  measures linear association

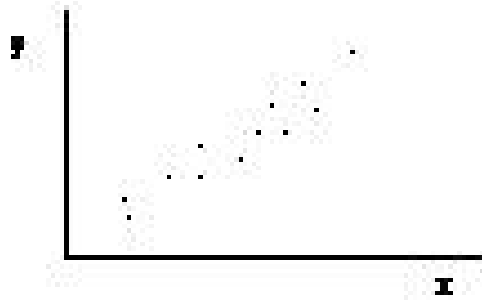
$r$  has values between  $-1 \leq r \leq +1$

$r \approx +1$  implies strong positive linear association

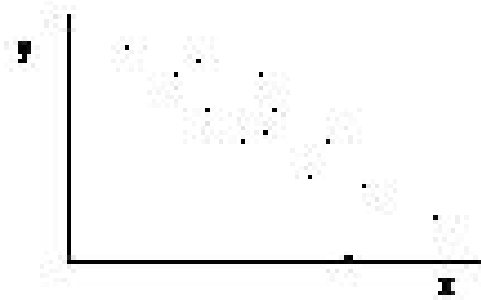
$r \approx -1$  implies strong negative linear association

$r \approx 0$  implies no linear association

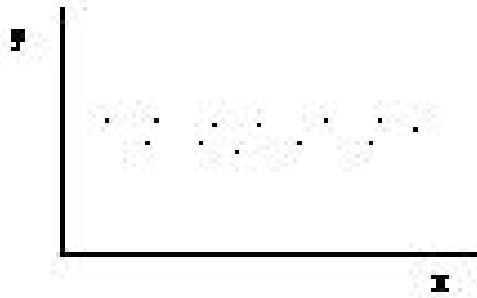




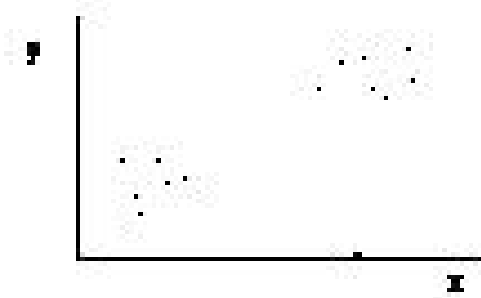
$r \approx +1$



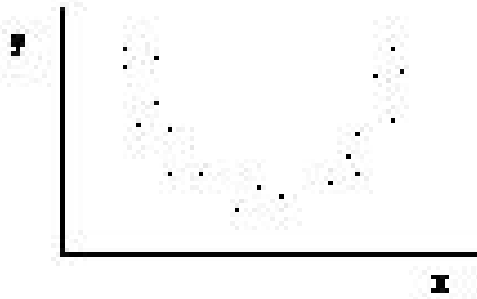
$r \approx -1$



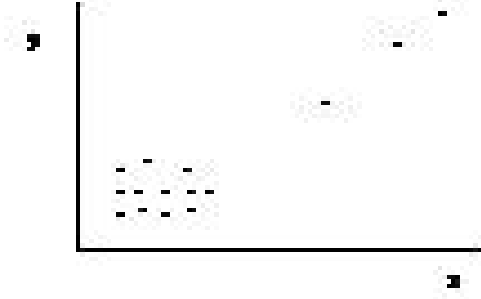
$r \approx 0$



$r \approx +1$



$r \approx 0$



$r \approx 0$

# Correlation Hypothesis Testing

The hypothesis of interest deals with whether there is linear association between  $x$  and  $y$ . If there is no such association, we would have  $\rho = 0$ . Hence, the hypotheses of interest are:

$$H_0: \rho = 0 \quad \text{vs} \quad H_1: \rho \neq 0$$

which we can test by using the test statistic:

$$t = r \left[ \frac{n-2}{1-r^2} \right]^{1/2} \approx t_{(n-2)}$$

Note that this calculation requires only the sample estimate  $r$  of the correlation coefficient  $\rho$  and the sample size  $n$  and that we need to use the  $t$  distribution with  $n - 2$  degrees of freedom.

## Test of correlation hypothesis for urinary oxalate and fecal fat, n = 11, r = 0.5084

1. *The Hypothesis:*  $H_0: \rho = 0$  vs  $H_1: \rho \neq 0$
2. *The  $\alpha$  level:*  $\alpha = 0.05$
3. *The assumptions:* Random sample from bivariate normal distribution
4. *The test statistic:*

$$t = r \left[ \frac{n-2}{1-r^2} \right]^{1/2} \approx t_{(n-2)}$$

5. *The critical region:* Reject  $H_0: \rho = 0$  if the value calculated for  $t$  is not between  $\pm t_{0.975}(9) = 2.262$

6. *The result:*  $r = 0.5084, n = 11$

$$\begin{aligned} t &= 0.5084 \left[ \frac{9}{1 - (0.5084)^2} \right]^{\frac{1}{2}} \\ &= 0.5084 \left[ \frac{9}{1 - 0.2585} \right]^{\frac{1}{2}} \\ &= 1.77 \end{aligned}$$

7. *The conclusion:* Accept  $H_0: \rho = 0$  since  $t = 1.77$  is between  $\pm t_{0.975}(9) = 2.262$

**Test of correlation hypothesis for Tono-Pen vs Goldman intraocular pressure, n = 40, r = 0.6574**

1. *The hypothesis:*  $H_0: \rho = 0$  vs  $H_1: \rho \neq 0$

2. *The assumptions:* Random sample  
bivariate normal distribution

3. *The  $\alpha$ -level:*  $\alpha = 0.05$

4. *The test statistic:* 
$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

5. *The rejection region:* Reject  $H_0: \rho = 0$ , if  $t$  is not between  $\pm t_{0.975}(38) = 2.02$

6. *The result:*  $n = 40$ ,  $r = 0.6574$ ,  $r^2 = 0.44$ ,

$$t = 0.6574 \sqrt{\frac{38}{1 - 0.6574^2}} = 0.66 \sqrt{\frac{38}{0.56}} = 5.44$$

7. *The conclusion:* Reject  $H_0: \rho = 0$

Since  $t = 5.44$  is not between  $\pm 2.02$

# Health Indicators and the Organization of Health Care Systems in Western Europe

## ABSTRACT

**Objective.** This study investigated the association between health care systems and health indicators in developed countries.

**Methods.** Cross-national comparisons were conducted with regression analysis between 17 Western European countries with two types of health care systems: national health services and social security systems.

**Results.** Health care expenditure rates were strongly correlated to potential years of life lost to ischemic and to infant mortality rates; they were positively correlated to life expectancy for females. Regression models predicted that countries with national health services systems would have lower infant mortality rates at similar levels of gross domestic product (GDP) and health care expenditure. Finally, increases in health care expenditure would decrease the risk of ischemic and potential infant mortality rates according to GDP; this decrease would be greater in countries with national health services than in those with social security systems. This model predicted this difference to be about 15% on average levels of health expenditure.

**Conclusions.** National health services seem to be more efficient in providing lower infant mortality rates than social security systems in Western European countries. (*J Gen Intern Med 1998;13:20-24*)

Josée Ahoja, MD, Alberto Goyens, MD, MPH, and  
Luisa Navarro, MD, PhD

## Introduction

The health of a population is reflected by many factors: longevity, infant mortality, life expectancy, and the health care system.<sup>1</sup> Development and life expectancy strongly depend on each country's level of development<sup>2</sup> and position within the worldwide division of labor.<sup>3</sup>

The socioeconomic level of development, as measured by per capita gross domestic product (GDP), is an important predictor of the health of populations.<sup>4</sup> However, some "regulation" variables have achieved greater improvements in the health of their populations than less regulation ones with similar levels of GDP.<sup>5-7</sup> Cross-national comparisons between developed countries have found better health indicators in countries with a more egalitarian distribution of income.<sup>8,9</sup>

Political factors influence health care systems organization. Kopp<sup>10</sup> and others<sup>11-13</sup> have related the strength of military movements and the power of social democratic parties to the development of universal systems of social protection, such as national health services, in Western developed countries.

Western European countries currently involved in a process of political and economic integration share as a basic value the guarantee of access to health care services for all citizens.<sup>14</sup> Despite important differences,<sup>15</sup> the health care systems of these countries can be classified as two basic categories: national health services and social security systems. National health services are based on egalitarian principles and are financed through general taxation, and the health care services are, in general, publicly owned and managed. Social security systems are the other form: the national

health service (financed by social payments and private insurance funds of workers) and the health care services are, in general, private. In countries with social security systems, the share of private health expenditure in total health expenditure tends to be higher.<sup>16-17</sup>

Health system organization and the quality of capital are an interesting characteristic of a nation<sup>18</sup> and therefore of the health care systems in Western Europe. Therefore, a common goal of these countries is to improve their health care systems efficiency.<sup>19</sup> Our health care system may be considered more efficient than another if, for the same level of health care expenditure (costs), it produces better health indicators of the same people, especially health service coverage.<sup>20,21</sup>

Several political and economic variables were related to a higher level of the right of all citizens to have access to health care services through the health care system: the relative performance of the two types of health care systems in Western Europe. The aim of this study is to analyze these two health care systems in terms of their efficiency—that is, the relationship between health care expenditure and health indicators.

*In the course of this study, the authors were aided by Department of Health Policy and Planning Unit, School of Health and Public Health, The Johns Hopkins University, Baltimore, Md; Dr Antonio Goyens, who is with the Ministerio de Sanidad y Consumo, Madrid, and Antonio Navarro, as well as the Dirección Asistencial de Salud Pública, Granada, Spain.*

Received for review 10/19/97; accepted for publication 1/16/98.

This study was supported from C. Goyens, Ministerio Sanidad y Consumo, Granada, and Dirección Asistencial de Salud Pública, Granada, Spain.

TABLE 3—Correlation Matrix of Selected Health and Socioeconomic Indicators of Western European Countries

	Health Care Expenditures per Capita	% Population Covered	Public Health Expenditures, 1964	Infant Mortality Rate	WV.L. Aging	WV.L. Population	Life Expectancy, Males	Life Expectancy, Females	
Change observed (1960-64) per capita	1.00	-.06	.02	-.20	-.04	-.10	-.40	.20	
Health care expenditure: total per capita		-.12	-.14	-.40	-.04	-.07	-.30	-.23	
% population covered			.65	.04	.00	.00	-.10	-.14	
Public health expenditures				-.15	-.10	-.00	.07	-.11	
Infant mortality rate					.20	.00	-.00	-.14	
WV.L. Aging						.27	-.00	-.40	
WV.L. Population							-.00	-.00	
Life expectancy, males								.20	
Life expectancy, females									.20

Note:  $R^2$  for  $X$  indicates value of the best linear regression equation of  $X$  on  $Y$ .

WV.L.P. were obtained. Because of the high correlations, the ranking of these variables in the order of the selected infant mortality rate, the projected infant mortality rate on the basis of a given level of WV.L.P. This ratio is expressed as a percentage and provides an estimate of infant mortality not explained by WV.L.P. in a given country.

This ratio of observed to expected infant mortality rates according to WV.L.P. cases in both 1960 and 1964, and infant mortality rate indicated by the WV.L.P. was further analyzed the results are shown in Table 4 and in Figure 1.

### Results

Mean values of all variables for the two types of health care systems are shown in Table 5. When we compared the scores of socioeconomic indicators between the two groups, only two variables, WV.L.P. and health care expenditures, showed statistical significance.

Regression correlations coefficients were statistically significant for health care expenditures and WV.L.P. and for health care expenditures and infant mortality (Table 6).

In simple regression models, health care expenditures and WV.L.P. were associated with life expectancy for females and males, respectively, with projected rates of life span for males. However, they were not associated with the same health indicators for males. Health care expenditures explained 20% of the variability in projected rates of life span for females,  $R^2 = 74\%$ ,  $F = 402$  and 37% of life expectancy for females,  $F = 3.85$

TABLE 4—Regression Models for Infant Mortality Rate as a Function of Infant Expenditure, Health Care Expenditures, and Type of Health Care System, Western Europe

	R-SQUARE			
	Model I	Model II	Model III	Model IV
Change observed (1960-64) per capita	-.0.26 (0.112)		-.0.40 (0.117)	
Health care expenditure per capita 1964		-.0.06 (0.070)		-.0.06 (0.070)
Health care expenditure: infant mortality ratio (1960-64) per capita 1964			-.0.13 (0.100)	-.0.13 (0.100)
$R^2$	.00	.00	.01	.00
	.00	.00	.01	.01

Note: All variables significantly explanatory.

$F = 205$ , WV.L.P. explained 20% of the variability in projected life span for the females,  $R^2 = 5.7\%$ ,  $F = 105$  and 20% of life expectancy for females,  $R^2 = 4.9\%$ ,  $F = 105$ . No other socioeconomic indicator was explained significantly and their correlation to these two health indicators.

In the analysis, health care expenditures were provided in the a linear regression equation of infant mortality,  $R^2 = 4\%$  (for WV.L.P.  $R^2 = 74\%$ ). Type of health care system was included in models III and IV (Table 6); however, no more, infant mortality rates varied the model for the health care systems compared with model health care expenditures and WV.L.P. model III and health care expenditures provided 1% of the variability of life expectancy, as indicated by the coefficient

$R^2 = 1\%$ , which is an approximately 1% increase life span for female IV's.

Table 8 shows the best model that explained the rates of observed to projected infant mortality rates according to WV.L.P. This ratio distribution on health care expenditures (observed to expected) showed increases in health care expenditures would result in greater decreases in the infant mortality rate for countries with normal health systems. The model predicts that an average health care expenditure of \$100 per year, and infant mortality rate would be about 1.5% lower in normal health systems. Figure 1 illustrates the limiting effect of this model on the ratio for the countries included in the study as health care expenditures increase, the decreasing variability of infant mortality



# Test of correlation hypothesis for infant mortality rate and gross domestic product, $n = 17$ , $r = -0.64$

## example

1. *The hypothesis:*  $H_0: \rho = 0$  vs  $H_1: \rho \neq 0$

2. *The assumptions:* Random sample  
bivariate normal distribution

3. *The  $\alpha$ -level :*  $\alpha = 0.05$

4. *The test statistic:* 
$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

5. *The rejection region:* Reject  $H_0: \rho = 0$ , if  $t$  is not between  $\pm t_{0.975}(15) = 2.13$

6. *The result:*  $n = 17, r = -0.64,$

$$t = -0.64 \sqrt{\frac{15}{1 - (-0.64)^2}} = -0.64 \sqrt{\frac{15}{0.59}} = -3.23$$

7. *The conclusion:* Reject  $H_0: \rho = 0$

Since  $t = -3.23$  is not between  $\pm 2.1315$

## Test of correlation hypothesis for life expectancy for males and females, $n = 17$ , $r = 0.67$

1. *The hypothesis:*  $H_0: \rho = 0$  vs  $H_1: \rho \neq 0$
2. *The assumptions:* Random sample  
bivariate normal distribution
3. *The  $\alpha$ -level :*  $\alpha = 0.05$
4. *The test statistic:* 
$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

5. *The rejection region:* Reject  $H_0: \rho = 0$ , if  $t$  is not between  $\pm t_{0.975}(15) = 2.1315$

6. *The result:*

$$n = 17, r = 0.67,$$

$$t = 0.67 \sqrt{\frac{15}{1 - 0.67^2}} = 0.67 \sqrt{\frac{15}{1 - 0.45}}$$

$$t = 0.67 \sqrt{\frac{15}{0.55}} = 0.67 \sqrt{27.27}$$

$$t = 0.67(5.22) \quad t = 3.49$$

7. *The conclusion:*

Reject  $H_0: \rho = 0$

Since  $t = 3.49$  is not between  $\pm 2.1315$

# Social Capital, Income Inequality, and Mortality

## ABSTRACT

**Objectives:** Recent studies have demonstrated that income inequality is related to mortality rates. It was hypothesized, in this study, that income inequality is related to reduction in social cohesion and that displacement in social capital is in turn associated with increased mortality.

**Methods:** In this cross-national ecological study based on data from 29 states, social capital was measured by weighted regression on two items from the General Social Survey: per capita density of membership in voluntary groups in each state and level of social trust, as gauged by the proportion of residents in each state who believed that people could be trusted. Age-standardized total and cause-specific mortality rates in 1990 were obtained for each state.

**Results:** Income inequality was strongly associated with both per capita group membership ( $r = -.51$ ) and lack of social trust ( $r = .73$ ), in turn, both social trust and group membership were associated with total mortality, as well as rates of death from coronary heart disease, malignant neoplasms, and firearm mortality.

**Conclusions:** These data support the notion that income inequality leads to increased mortality via displacement in social capital. *Am J Public Health*. 1997;87:1894-1898.

*Author Keywords:* PHA, Income P, Kennedy, KMM, Kimberly, Lawrence, AM, and Deborah Froehner-Nishi, MD

### Introduction

A number of cross-national studies have indicated that the degree of income inequality in a given country is strongly related to the country's level of mortality.<sup>1-5</sup> In one investigation of nine nations included in the Human Development Study,<sup>6</sup> a correlation of .86 was reported between average life expectancy and percentage of people allowed to the 70th of the population at the lowest income levels. Two recent UK studies independently demonstrated an association between income inequality and mortality.<sup>7,8</sup> Kondo et al.<sup>7</sup> examined the relationship between degree of household income inequality and state-level variation in all-cause and cause-specific mortality. The degree of income inequality in each state was measured by the Ratio Head Index, which is equivalent to the proportion of aggregate income that goes to households from households above the mean and transformed to show before the mean in order to achieve perfect equality in the distribution of household incomes.<sup>9</sup> The higher the Ratio Head Index, the more unequal the distribution of income. The overall correlation of the Ratio Head Index to all-cause mortality in 1990 was .61 ( $p < .0001$ ). After adjustment for poverty, a 1% rise in the Ratio Head Index was associated with an increase in age-adjusted total mortality rate of .17 deaths per 100,000 (95% confidence interval [CI] = .06, .26).<sup>7</sup> The Ratio Head Index was also associated with deaths from specific causes, including coronary heart disease, cancer, and infant mortality.

In an independent study, Kaplan et al.<sup>8</sup> examined the association between income inequality—as measured by the share of aggregate income earned by the

bottom 20% of households—and state-level variations in total mortality. A strong association was found between their measure of income inequality and age-adjusted total mortality rates in 1990 ( $r = -.56$ ,  $P < .001$ ). Moreover, the degree of income inequality in each state in 1990 was a powerful predictor of levels of total mortality 10 years later.

The pathways and mechanisms mediating the association between income inequality and mortality levels remain to be established.<sup>10</sup> One hypothesis is that rising income inequality results in increased levels of frustration, which may have deleterious behavioral and health consequences.<sup>11</sup> Another, that greater large disparities in income to development also tend to be the case than underdevelopment in human capital (e.g., education, health care, and other factors that promote health).<sup>12</sup> Recently, it has been hypothesized that the growing gap between the rich and the poor has led to declining levels of social cohesion and trust, or displacement in "social capital."<sup>13-15</sup> Social capital has been defined as the "stocks of social capitalization, such as civic participation, norms of reciprocity, and trust in others, that facilitate cooperation for mutual benefit."<sup>16</sup> Social capital is thus a community-level "resource" variable whose counterpart at the individual level is represented by a person's

The authors are with the Harvard School of Public Health, Boston, Mass.; Robert Kennedy and Hillary Clinton School and the Department of Health and Social Behavior, Bruce P. Kennedy and Hillary Clinton School of Public Health, Boston; and the Center for Health Policy and Management.

Requests for reprints should be sent to Robert Kennedy, PhD, Department of Health and Social Behavior, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115.

This paper was accepted November 8, 1996.

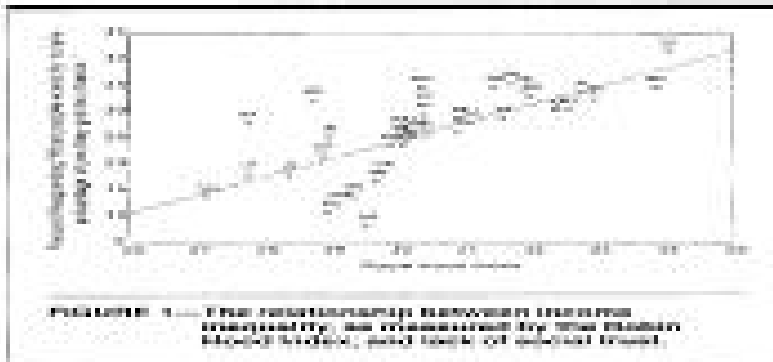


FIGURE 1.—The relationship between percent elderly, as measured by the United States Census, and both of social trust.

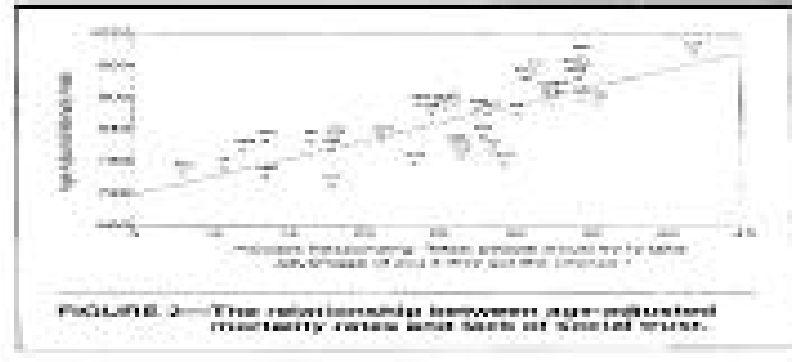


FIGURE 2.—The relationship between age-adjusted mortality rates and both of social trust.

percent) tends to greatly over-represent rural and often not urban cities, whereas, for income such as personal income from food stamps, Medicaid, and public housing, elderly households are equally likely to reflect changes in the retirement phase transition. The poverty variable used in the analysis represents the percentage of households in a given state before the federal poverty level in 1994, as reported by the U.S. Census Bureau (1994) with their own methodology.<sup>12</sup>

**Measurement of Social Capital**  
 (See Appendix)

The age-adjusted mortality rates for each state in 1996 were obtained from the Surveillance Mortality Files compiled by the National Center for Health Statistics of the Centers for Disease Control and Prevention (CDC). The data were obtained from the CDC's database via URL: <http://www.cdc.gov>.

All mortality rates were directly age standardized to the US population and expressed as the number of deaths per 100,000 persons (except in the case of infant mortality, for which death rates were expressed per 1,000 live births). In addition to all-cause mortality, we derived the following major causes of death categories from the US Surveillance Mortality Database by ICD9: 1st causes (1001-14) and 2nd through 4th, respiratory (1001-14) and 2nd through 4th, malignant neoplasms (150-209) and 2nd through 4th, cardiovascular disease (210-259) and 2nd through 4th, and unintentional injuries (800-899) and 2nd through 4th.

TABLE 1.—Correlations among indicators of Social Capital, Poverty, Income, Mortality, and Mobility

	Poverty	Mortality	Model-based Trust	Group Mobility	Systemic Trust	Total
Mortality	.07*					
Model-based Trust	.14*	.09*				
Group-based Mobility	.08*	.04*	.08*			
Systemic Trust	.08*	.04*	.08*	.04*		
Personal Income	.08*	.04*	.08*	.04*	.08*	.04*
Percentage of the population aged 65 and over	.08*	.04*	.08*	.04*	.08*	.04*

\*Measured by the percentage regression. \*\*Model-based trust by the best regression of age 65 and over. \*\*\*Measured by the percentage regression. \*\*\*\*Age 65 and over, mobility, mobility with group mobility by the percentage regression.

**Model Analysis**

A binary, least squares regression was used to estimate the relationship of social capital indicators to mortality rates. Two sets of models were estimated: the first set of models, in the form of a model, was estimated the weighted social capital indicators (e.g., weighted average group membership and weighted average social trust) against all-cause and cause-specific mortality rates. In the second set of models, we adjusted the regression models for state variables in population of persons. To estimate the effects of mobility on mortality, the Model-based Trust and Group-based Mobility variables were carried out a path analysis<sup>13</sup> based on a causal model in which

mortality affects mobility through an unobserved causal agent.

**Results**

**Relationships among Social Capital, Mortality, and Mobility**

The main indicators of social capital—rates of participation in civic activities and the weighted proportion of respondents who reported that "most people would be helpful and loyal if you let them get on their own"<sup>14</sup>—were significantly correlated with mobility, or "people really talk and are interested"<sup>15</sup>, were highly correlated with each other (Table 1). Since these variables may not represent an exhaustive list of "social capital" influ-

**For Table 1, Correlation between Mortality and Social Mistrust, n = 39, r = 0.79**

1. *The hypothesis:*  $H_0: \rho = 0$  vs  $H_1: \rho \neq 0$

2. *The assumptions:* Random sample  
bivariate normal distribution

3. *The  $\alpha$  level:*  $\alpha = 0.05$

4. *The test statistic:* 
$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

5. *The rejection region:*

Reject  $H_0: \rho = 0$ , if  $t$  is not  
between  $\pm t_{0.975}(37) = 2.02$

6. *The result:*

$n = 39, r = 0.79, r^2 = 0.6241,$

$$t = 0.79 \sqrt{\frac{37}{1 - 0.79^2}} = 0.79 \sqrt{\frac{37}{0.3759}} = 7.8$$

7. *The conclusion:*

Reject  $H_0: \rho = 0$   
Since  $t = 7.8$  is not between  
 $\pm 2.02$



# The Evolving Epidemiology of Chlamydial and Gonococcal Infections in Response to Control Programs in Winnipeg, Canada

## ABSTRACT

**Objective.** The purpose of this study was to describe and compare the transmission dynamics of chlamydial and gonorrheal infections in Winnipeg, Manitoba, Canada, and to assess implications for control programs.

**Methods.** Chlamydia and gonorrhea serosurvey data reports (1988 through 1995) and sentinel-testing reports (1988 through 1995) were obtained.

**Results.** High incidence rates of both chlamydia and gonorrhea characterized geographic core areas characterized by low socioeconomic status. A decline in the number of reported cases of chlamydia (60%) and gonorrhea (50%) occurred between 1988 and 1995. For chlamydia, the decline was most pronounced in men, core areas, and, while less pronounced, it was similar in core and noncore areas.

**Conclusions.** Chlamydia and gonorrheal appear to be evolving through defined epidemic phases, with chlamydia transmission, in response to a newly introduced control program, becoming more core dependent and gonorrhea transmission becoming more sporadic in the face of a sustained control effort. Focused control programs, based on an understanding of the transmission dynamics of chlamydial and gonorrheal, may make their contribution a significant one. *J Public Health. 1998;20: 240-246.*

Authors: P. Brumwell, PhD, PhD, Stephen Miles, PhD, MPH, C. Colwell, PhD, MPH, PhD, P. Frank, PhD, MPH, MPH, J. Wu, MD, Chris M. Armstrong, MD, PhD, and P. Martin, C. Macdonald, PhD, DPH

Sexually acquired transmitted disease (STD) control is an important public health problem worldwide.<sup>1,2</sup> Most bacterial STDs are curable with relatively short courses of antimicrobials and hence they have become the focus of major disease control efforts.<sup>3</sup>

STD control programs should be based on a sound understanding of the population-level transmission dynamics of STD pathogens. Previous theoretical and empirical findings have provided valuable insight into the transmission dynamics of STDs.<sup>4-7</sup> These findings have identified the importance of core groups in the spread of STDs within populations. The theoretical underpinning of the core group concept is derived from the basic reproductive number, which defines the average number of new infections derived by the average number of secondary infections arising from infected individuals in a fully susceptible population. The basic reproductive number is the product of 2 population parameters: transmission of the disease, contact rate between infectious and susceptible individuals, and duration of infectivity. One when the basic reproductive number is greater than 1, then an infectious agent successfully spreads in a population. Distinct core groups, however, have demonstrated that the average rate of partner change (contact rate) is much higher within the population in core subgroups than outside STDs.<sup>8</sup> Instead, subpopulations with higher rates of partner change have greater and sustained for the spread and transmission of STDs in the entire population.

Various approaches have been used to identify STD core groups. At the aggregate level, members of a core group can be defined according to risk indicators such as rate of partner change.<sup>9-11</sup> However, identifying individuals within core groups can be difficult for public health practitioners and

initially challenging for individuals. Therefore, recent methods have been used to define geographic "core areas" within which STD transmission rates are highest.<sup>12</sup> These geographic core areas are hypothesized to contain a higher proportion of the core group members and thus provide a natural target for focusing control efforts.<sup>13-15</sup> However, Woodhouse and colleagues<sup>16</sup> note that the size, location, and rates of core groups in STD transmission studies under the influence of control programs,<sup>17</sup> furthermore, they suggest that these changes result in a somewhat dynamic pattern and thus control programs should be responsive to such changes.

In this paper, we describe the epidemiology of Chlamydia trachomatis, Neisseria gonorrhoea, and Mycoplasma genitalium (genitally transmitted infections in Winnipeg, Canada. By examining the role of geography, core area populations in the spread of these 3 STDs over time, we provide evidence suggesting that the epidemiology of chlamydia and gonorrhea has evolved in response to control programs, as proposed by Woodhouse and Aitai.<sup>18</sup>

Dr. Brumwell is with the Epidemiology Unit and Public Health Branch, Manitoba Health, and the Department of Medical Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada; Dr. Miles is with the Department of Medical Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada; Dr. Colwell is with the Department of Medical Microbiology and Tropical Medicine, University of Manitoba, Winnipeg, Manitoba, Canada; Dr. Frank is with the Public Health Branch, Manitoba Health.

Dr. Martin is currently with the Centre for Disease Control, 4115 York, University Ave, Winnipeg, Manitoba, R3S 0V8, Canada; Dr. Wu is with the Public Health Branch, Manitoba Health.

This paper was accepted for publication

**Abstract**

*Changyuan Dingshenkang*

Statistics in a province in China with a population of approximately 4.4 million, Wenzhou, with a population of about 1.2 million, is the only major city in the province. All cases of tuberculosis and gonorrhoea are notified to the provincial health department by both physicians and laboratories under the Public Health Act. Tissue bacteriologic in the government routinely report all positive diagnostic tests for tuberculosis and gonorrhoea. Ten laboratory-confirmed cases are investigated in Wenzhou, specially trained public health nurses interview and conduct contact tracing for these persons infected with tuberculosis or gonorrhoea. Sexual contacts of these with tuberculosis and gonorrhoea are also notified to the provincial health department. As a means of comparing the incidence of tuberculosis and gonorrhoea by geographic location, the province was divided into 222 geographical locations defined by postalizing administrative boundaries (288 total municipalities for non-Wenzhou residents and the 24 postal

areas for Wenzhou residents). The average annual incidence of tuberculosis and gonorrhoea per 100 000 population was compared on the basis of all reported cases. Incidence rates were directly standardized for age and gender to the total 1991 Statistics population. Post- and non-postal municipalities' geographical areas were used to assess the geographic relationship between tuberculosis and gonorrhoea rates. A regression test was used to assess the statistical significance of differences in proportions. SAS (Version 6.12; SAS Institute Inc, Cary, NC) was used in performing all statistical analyses.

*Wenzhou Work Areas*

Wenzhou was divided into 3 geographic risk areas ("core", "interior", and "peripheral") separately for tuberculosis and gonorrhoea. This was done separately by aggregating postal areas based on their average annual incidence rates between 1991 and 1995 and visually connecting the best points from the distribution curves (see Figure 2a, 2b). Since the 1995 Census of China was used to estimate unemployment rates, which

represented income, language, ethnic patterns of labor, mobility status, and population density for the Wenzhou postal areas (Statistics Bureau, Statistics Bureau 1995). The incidence of other selected communicable diseases was identical from the provincial health administration registry.

**Results**

*SIR Incidence Rates by Geographic Site*

Figure 1 shows that tuberculosis and gonorrhoea incidence rates were highly correlated in both the provincial and urban levels (Wenzhou,  $r = 0.7$ ,  $P < 0.001$ ; Wenzhou,  $r = 0.82$ ,  $P < 0.001$ ). Directly standardized and gonorrhoea rates were also highly correlated, not just the data and analyzed the spatial distribution of incidence in the province and in the city of Wenzhou. Figure 2 shows the remarkably geographic variation in tuberculosis and gonorrhoea incidence rates. From 1991 through 1995, the average annual incidence of tuberculosis and gonorrhoea ranged from 0 to about 5000 per 100 000 population in differ-

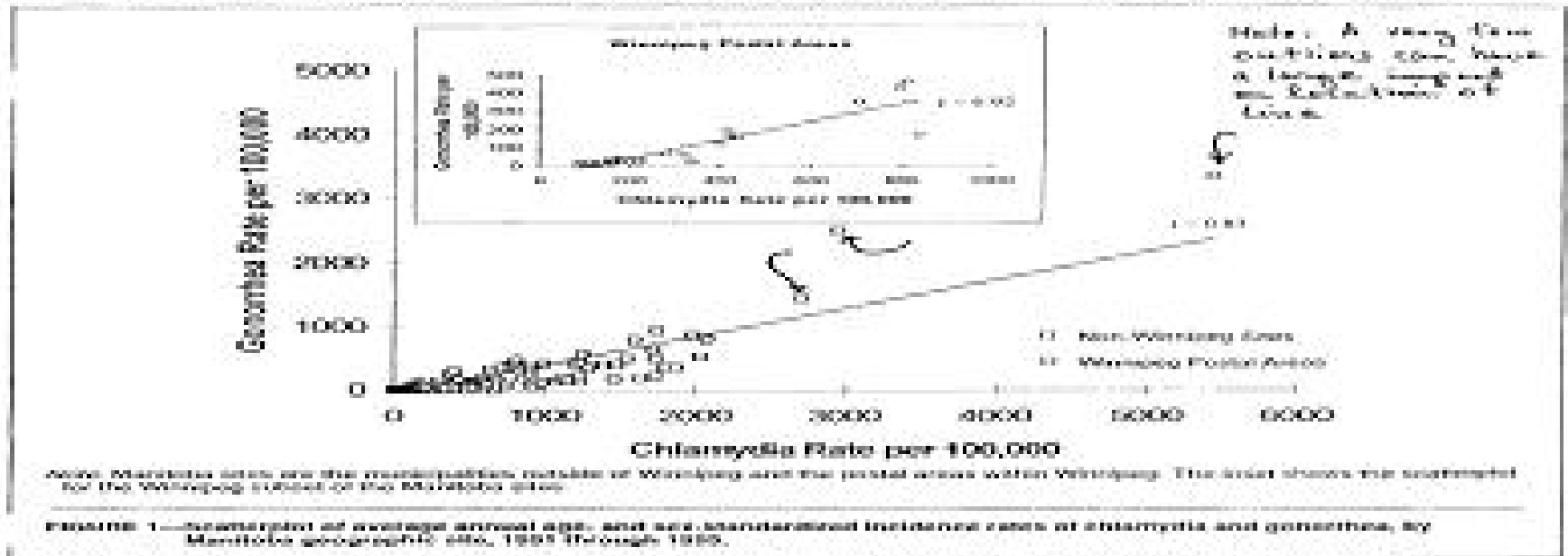


Figure 1—Scatterplot of average annual age- and sex-standardized incidence rates of tuberculosis and gonorrhoea, by Maritime geographic site, 1991 through 1995.

# Hypothesis test for correlation between gonorrhea rate and chlamydia rate

1. *The hypothesis:*  $H_0: \rho = 0$  vs  $H_1: \rho \neq 0$

2. *The assumptions:* random sample  
bivariate normal distribution

3. *The  $\alpha$ -level:*  $\alpha = 0.05$

4. *The test statistic:* 
$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

5. *The rejection region:* Reject  $H_0: \rho = 0$ , if  $t$  is not between  $\pm t_{0.975}(320) \approx 2.00$

6. *The result:*

$$n = 332, r = 0.83,$$

$$t = 0.83 \sqrt{\frac{320}{1 - 0.83^2}} = 0.83 \sqrt{\frac{320}{1 - 0.69}}$$

$$t = 0.83 \sqrt{\frac{320}{0.31}} = 0.83 \sqrt{1032.26}$$

$$t = 0.83(32.13) = 26.67$$

7. *The conclusion:*

Reject  $H_0: \rho = 0$

~~Since~~  $t = 26.67$  is not between

# Module 20: Correlation

This module focuses on the calculating, interpreting and testing hypotheses about the Pearson Product Moment Correlation Coefficient.

# Correlation

In module 19, we examined how two variables,  $x$  and  $y$ , relate to each other by using the simple linear regression tool. In that context,  $x$  is the independent variable and  $y$  is the dependent variable. Typical examples for the independent variable include measures of time, including age; whereas, typical examples for the dependent variable are continuous measurements such as blood cholesterol level. The general assumption is that there are separate normal distributions of the dependent variable  $y$  for each value of the independent variable  $x$ . Further, we need to assume that these separate normal distributions for the dependent variable all have the same population variance.

Clearly these assumptions are quite restrictive in that we are often interested in the relationship between two variables,  $x$  and  $y$ , where it is not at all clear which should be labeled the independent variable and which the dependent one. An example is the relationship between blood cholesterol level and blood pressure level.

For this situation, we have another tool we can use to measure and test hypotheses about the relationship between these two variables. The tool is called correlation and we focus here only on what is usually called the Pearson Product Moment Correlation Coefficient. There are other measures of correlation which we will not discuss here. There are restrictions for the use of this correlation tool as well, which include the basic assumption that the  $x$  and  $y$  variables together have a joint frequency distribution which is called the bivariate normal distribution. This distribution looks like a three-dimensional bell in a manner similar to the way a normal distribution for one variable looks like a cross section of a bell.



The degree of association or correlation between two variables is measured by the *correlation coefficient*. This is done in a manner similar to that for other population parameters and estimates of these parameters obtained by calculating statistics from samples. That is, there is a value for the population parameter for this coefficient which is estimated by selecting a random sample and calculating the appropriate coefficient using the data from this sample. We can also use the information from the sample to test hypotheses about the population.

The population parameter for the Pearson Product Moment Correlation Coefficient is defined as

$$\rho = \frac{\sum (x - \mu_x)(y - \mu_y)}{\sqrt{\sum (x - \mu_x)^2 \sum (y - \mu_y)^2}}$$

which is typically called rho, for the Greek letter it represents.

The estimate of  $\rho$  calculated from the sample data is the statistic

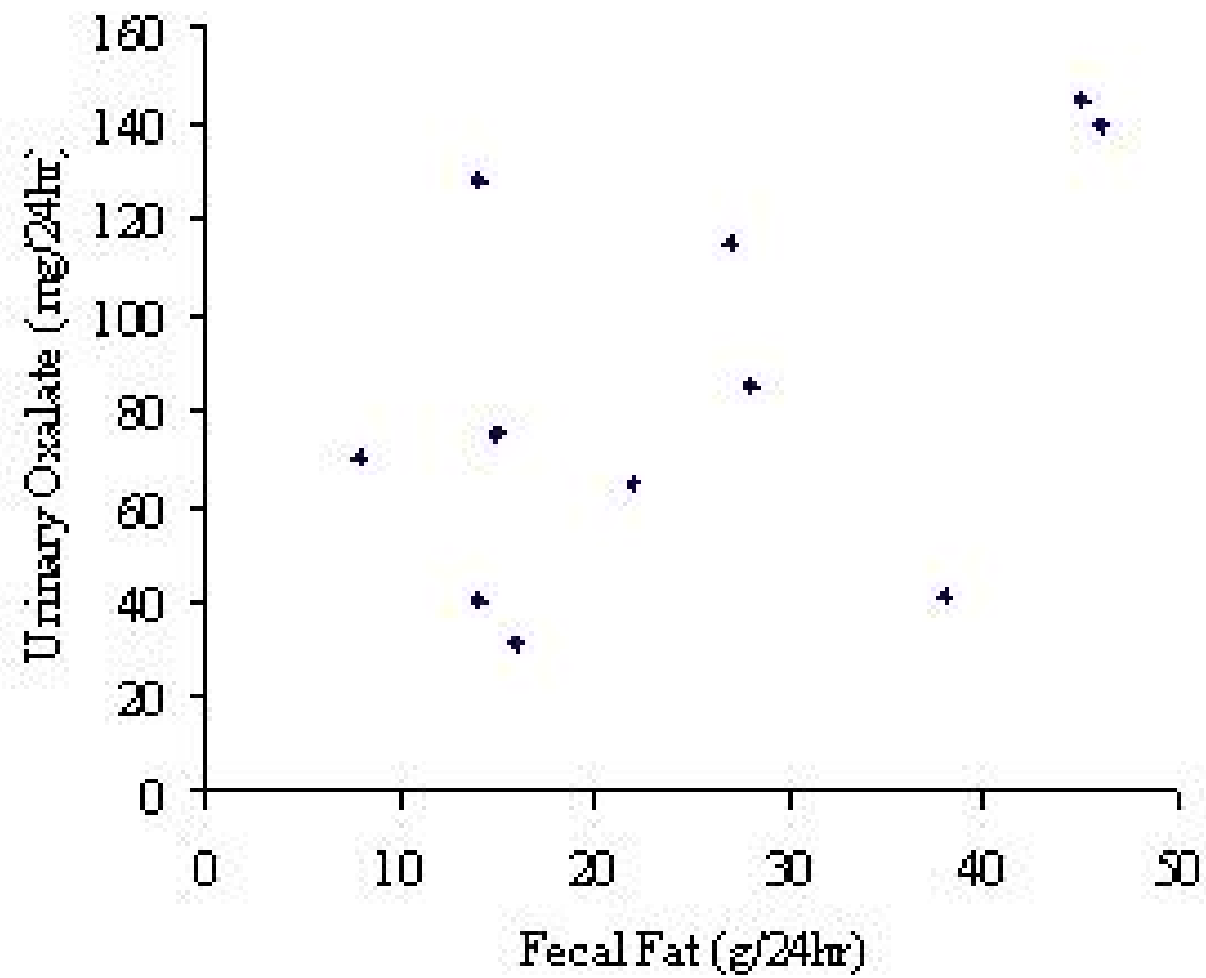
$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

$$r_{xy} = \frac{\sum xy - (\sum x)(\sum y) / n}{\sqrt{[\sum x^2 - (\sum x)^2 / n][\sum y^2 - (\sum y)^2 / n]}}$$

$$= \frac{SS(xy)}{\sqrt{SS(x)SS(y)}}$$

Fecal Fat (g/24 hr) and Urinary Oxalate (mg/24 hr)  
secreted by a random sample of  $n = 11$  persons

Patient	Fecal Fat (g/24hr)	Urinary Oxalate (mg/24hr)
	$x$	$y$
1	16	31
2	14	40
3	38	41
4	8	70
5	15	75
6	22	65
7	28	85
8	27	115
9	14	128
10	45	145
11	46	140
Sum	273	935
Mean	24.8	85.0



Person	Fecal Fat		Urinary Oxalate		
	x	x <sup>2</sup>	y	y <sup>2</sup>	xy
1	16	256	31	961	496
2	14	196	40	1,600	560
3	38	1,444	41	1,681	1,558
4	8	64	70	4,900	560
5	15	225	75	5,625	1,125
6	22	484	65	4,225	1,430
7	28	784	85	7,225	2,380
8	27	729	115	13,225	3,105
9	14	196	128	16,384	1,792
10	45	2,025	145	21,025	6,525
11	46	2,116	140	19,600	6,440
Sum	273	8,519	935	96,451	25,971
Mean	24.8		85.0		
Sum /n	6,775.36		79,475.00		
SS	1,743.64		16,976.00		2,766.00
Variance	174.36		1,697.60		
SD	13.2		41.2		