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

\[ \rho = \frac{\sum (x-u_x)(y-u_y)}{\sqrt{\sum (x-u_x)^2 \sum (y-u_y)^2}} \]

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

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

\[ r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} \]

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

\[ = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}} \]
Fecal Fat (g/24 hr) and Urinary Oxalate (mg/24 hr) secreted by a random sample of n = 11 persons

<table>
<thead>
<tr>
<th>Patient</th>
<th>Fecal Fat (g/24hr)</th>
<th>Urinary Oxalate (mg/24hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>65</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>85</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>115</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>128</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>145</td>
</tr>
<tr>
<td>11</td>
<td>46</td>
<td>140</td>
</tr>
<tr>
<td>Sum</td>
<td>273</td>
<td>935</td>
</tr>
<tr>
<td>Mean</td>
<td>24.8</td>
<td>85.0</td>
</tr>
<tr>
<td>Person</td>
<td>Fecal Fat</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>x²</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>256</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>196</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>1,444</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>225</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>484</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>784</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>729</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>196</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>2,025</td>
</tr>
<tr>
<td>11</td>
<td>46</td>
<td>2,116</td>
</tr>
</tbody>
</table>

|        |          |    |                |    |
| Sum    | 273      | 8,519| 935           | 96,451|
| 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         |    |

20 - 9
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

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

\[
\text{Intercept } a = \bar{y} - b\bar{x} \\
= 85.0 + 1.59(24.8) \\
= 45.54
\]

The straight line depicting the regression relationship of \( y \) on \( x \) is

\[
\hat{y} = a + bx \\
\hat{y} = 45.54 + 1.59x
\]

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:**  \( H_0: \beta = 0 \text{ vs } H_1: \beta \neq 0 \)

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

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>( bSS(xy) )</td>
<td>( SS(\text{Reg})/1 )</td>
<td>( \frac{MS(\text{Reg})}{MS(\text{Res})} )</td>
</tr>
<tr>
<td>Residual</td>
<td>( n-2 )</td>
<td>( SS(\text{Res})^a )</td>
<td>( SS(\text{Res})/(n-2) )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( n-1 )</td>
<td>( SS(y) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^aSS(\text{Residual}) = SS(y) - SS(\text{Regression})\)
5. The critical region: Reject $H_0: \beta = 0$ if the value calculated for $F$ is greater than $F_{0.05}(1,9) = 5.12$

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

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>4,397.94</td>
<td>4,397.94</td>
<td>3.15</td>
</tr>
<tr>
<td>Residual</td>
<td>9</td>
<td>12,578.06</td>
<td>1,397.56</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>16,976.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. The conclusion: Accept $H_0: \beta = 0$ since $F < 5.12$
\[ R^2 = 0.2585 \]

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

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

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

\[ B = y - \hat{y} \quad \text{Left over} \Rightarrow SS(\text{Residual}) \]
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
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:

$$H_0: \rho = 0 \quad vs \quad H_1: \rho \neq 0$$

which we can test by using the test statistic:

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

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: $r = 0.5084$, $n = 11$

$$t = 0.5084 \left[ \frac{9}{1 - (0.5084)^2} \right]^{\frac{1}{2}}$$

$$= 0.5084 \left[ \frac{9}{1 - 0.2585} \right]^{\frac{1}{2}}$$

$$= 1.77$$

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 $\pm t_{0.075} (38) = 2.02$

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

Objectives: 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 nations with various health care systems, national health services, and social security systems.

Results: Health care expenditures were strongly correlated in potential countries, with a low level of development and infant mortality rates. Economic and demographic factors are strongly dependent on each country's level of development and position within the worldwide classification of health care services.

The socioeconomic level of development, as measured by per capita gross domestic product (GDP) and health care expenditures, has a similar level of gross domestic product (GDP). Conversely, the same level of GDP and health care expenditures in some countries has led to better health outcomes in some countries.

Related factors influence health care systems. Countries that have lower levels of infant mortality rates and higher levels of GDP and health care expenditures have more effective health care systems and better health outcomes.

Western European countries have a more efficient and effective health care system than those in developing countries. The level of health care services in these countries can be classified into two basic categories: national health services and social security systems. National health services are based on egalitarian principles and are financed through public taxation. In contrast, social security systems, on the other hand, are financed mostly through mandatory payroll contributions and provide universal levels of care, and the health care services in social security systems are generally more accessible.

Introduction

The health of a population is influenced by many factors, including biology, environment, genetics, and health care systems. National health services and health care systems are strongly dependent on each country's level of development and position within the worldwide classification of health care services.

The socioeconomic level of development, as measured by per capita gross domestic product (GDP) and health care expenditures, has a similar level of gross domestic product (GDP). Conversely, the same level of GDP and health care expenditures in some countries has led to better health outcomes in some countries.

Related factors influence health care systems. Countries that have lower levels of infant mortality rates and higher levels of GDP and health care expenditures have more effective health care systems and better health outcomes.

Western European countries have a more efficient and effective health care system than those in developing countries. The level of health care services in these countries can be classified into two basic categories: national health services and social security systems. National health services are based on egalitarian principles and are financed through public taxation. In contrast, social security systems, on the other hand, are financed mostly through mandatory payroll contributions and provide universal levels of care, and the health care services in social security systems are generally more accessible.
TABLE 2—Correlation Matrix of Selected Health and Socioeconomic Indicators of Western European Countries

<table>
<thead>
<tr>
<th>Health- Care Expenditures per Capita</th>
<th>% Population Cured</th>
<th>Public Expenditures</th>
<th>Infant Mortality Rate</th>
<th>PYLL Males</th>
<th>PYLL Females</th>
<th>Life Expectancy Males</th>
<th>Life Expectancy Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP per Capita</td>
<td>-0.66</td>
<td>0.02</td>
<td>-0.33</td>
<td>-0.08</td>
<td>-0.10</td>
<td>-0.48</td>
<td>0.35</td>
</tr>
<tr>
<td>Health care expenditure per Capita</td>
<td>-12</td>
<td>0.14</td>
<td>0.85</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>% population educated</td>
<td>0.35</td>
<td>0.71</td>
<td>0.57</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Public health expenditure</td>
<td>0.19</td>
<td>0.35</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>% of GNP spent on health care</td>
<td>-12</td>
<td>-0.16</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td>Gini coefficient</td>
<td>-0.16</td>
<td>-0.15</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.12</td>
</tr>
<tr>
<td>Infant mortality rate</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Literacy rate</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Life expectancy</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>Life expectancy</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Note: P < 0.01 at the 95% level.

Results

Mean values of all variables for the year 1990 of the health care systems are shown in Table 2. When we compared the scores of socioeconomic indicators by GDP quintiles, only two variables, GDP and health care expenditures, reached statistical significance.

Pearson's correlation coefficients were statistically significant for health care expenditures and GDP and for both variables and infant mortality (Table 2).

In simple regression models, health care expenditures and GDP were associated with life expectancy for males and females associated with potential years of life (5 and 10 years). However, these were not associated with the same health indicators for males. Health care expenditures were statistically significant at 5% and 10% of the variability in potential years of life for both men and women (Table 2).

In the analysis, health care expenditures proved to be a better explanatory variable of infant mortality (M = 0.5) than GDP (M = 0.4). Type of health care system was included in models III and IV (Table 3). According to the results, infant mortality rates would be lower for national health systems, compared with social security systems at similar levels of GDP model III and health care expenditures (model IV). The magnitude of the differences, as estimated by the coefficients of 5%, would be of approximately 3% (model III) or 1% (model IV).

Table 1 shows the best model that estimates the rate of observed to predicted infant mortality ratio, according to GDP. This ratio decreases as health care expenditures increase. However, increases in health care expenditures would result in greater decreases in the infant mortality rate for countries with lower health expenditures. The model predicts that at average health care expenditures ($/P$), the infant mortality rate would be about 15% lower in national health systems. Figure 1 illustrates the marginal 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.

American Journal of Public Health, 1995
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 $\pm t_{0.025}(15) = 2.1315$.

6. The result:

\[
\begin{align*}
    n &= 17, r = 0.67, \\
    t &= 0.67 \frac{15}{\sqrt{1 - 0.67^2}} = 0.67 \frac{15}{\sqrt{1 - 0.45}} \\
    t &= 0.67 \frac{15}{\sqrt{0.55}} = 0.67 \sqrt{27.27} \\
\end{align*}
\]

\[t = 0.67(5.22) \quad t = 3.49\]

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

Objective: 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 deprivation in social capital is in turn associated with increased mortality.

Methods: In this cross-sectional study, data from 39 states' social capital was measured using weighted responses to two items from the General Social Survey: percent of respondents identifying as members in voluntary groups in each state and level of social trust. An index of income inequality was calculated for each state using data from the Luxembourg Income Study.

Results: Income inequality was strongly correlated with both social trust (r = -.46, p < .001) and social cohesion (r = .39, p < .01). After adjustment for income inequality was still associated with lower social trust (β = -0.2, p < .001). The Robins Index was associated with a 1.5% increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7. The Robins Index was associated with 5% increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7. The Robins Index was associated with a 1.5% increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7. The Robins Index was associated with a 1.5% increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7.

Conclusion: These data support the notion that income inequality leads to increased mortality via disinvestment in social capital. Am J Public Health. 1997;87:1491-1498.

Introduction

A number of cross-sectional studies have demonstrated that the degree of income inequality in a given society is strongly related to the society's level of mortality. The measure of income inequality used in the present study was the average deviation of income for the top 10% of the population relative to the bottom 50% of households. Income inequality was defined as the degree of income inequality in each state as measured by the Bohn-Hatt Index, which is equivalent to the proportion of aggregate income that was received from households above the mean who had 20% or more of the mean income in order to achieve perfect equality in the distribution of household income. The higher the Bohn-Hatt Index, the more unequal the distribution of income. The overall correlation of the Bohn-Hatt Index to all-cause mortality in 1990 was 0.34 (p < 0.001). After adjusting for income inequality, the Bohn-Hatt Index was associated with an increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7. The Robins Index was associated with a 5% increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7. The Robins Index was associated with a 1.5% increase in adjusted total mortality rate of 21.7 deaths per 100,000 persons, confidence interval 6.5, 56.7.
poverty index is a proxy for income level and does not reflect other sources of income such as noncash benefits from food stamps, Medicaid, and public housing. Poverty thresholds are updated annually to reflect changes in the Consumer Price Index. The poverty variable used in the analysis represents the percentage of households in a given state below the federal poverty level (FPL). In 1983, an income of less than $15,550 for households with two family members.

Measurement of Social Capital

The age-adjusted mortality rates for each state in 1983 were obtained from the National Center for Health Statistics of the US Department of Health and Human Services. These data were obtained from the CDC's database on CHAMPS-WH (Death Rates by Race and Hispanic Origin by State, 1988-1990).

All mortality rates were directly age standardized to the US population and expressed as the number of deaths per 100,000 person-years in the state. For each state with poverty ratio, we used the percentage of deaths as the number of deaths per 100,000 person-years in the state. We also examined the percentage of deaths as the number of deaths per 100,000 person-years in the state.

Data Analysis

p values were used to examine the relationship between social capital and mortality rates. Two sets of models were examined for each outcome of interest. In one set of models, we examined the relationship between social capital and mortality rates. In the second set of models, we examined the relationship between social capital and mortality rates and examined the impact of social capital on mortality through its impact on social capital.

Results

Relationships among Social Capital, Measures of Income, and Mortality

The four indicators of social capital—extent of participation in civic associations and the weighted proportions of respondents who agreed that "most people would not be likely to cheat you if they got a chance"—were highly correlated with each other Table 1. Since these variables may not represent an exhaustive list of "social capital" indices.
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 Chlamydia and gonorrhea in Winnipeg, Manitoba, Canada, and to assess implications for control programs.

Methods. Chlamydial and gonococcal surveillance data collected from two large geographically distinct census areas were analyzed. Data from the Winnipeg, Manitoba, census area were compared with those from the Central Hudson Valley, New York, census area, and with the results of a recent study of Chlamydia transmission dynamics in the United States.

Results. High incidence rates of both chlamydia and gonorrhea characterized the two large geographically distinct areas. A decline in the number of reported cases of chlamydia (37%) and gonorrhea (48%) occurred between 1988 and 1993. For chlamydia, the decline was more prominent in non-white areas, whereas for gonorrhea it was similar in both races.

Conclusion. Chlamydia and gonorrhea continue to be evolving through different epidemics, primarily, with chlamydial transmission in response to a newly introduced control program, becoming more concentrated and gonococcal transmission becoming more widespread in the face of a sustained control effort. Improved understanding of the transmission dynamics of chlamydia and gonorrhea is needed to develop an effective strategy to control these diseases. The basic reproductive number is the most important parameter in the spread of STIs. The current number of infected individuals in a highly susceptible population is not sufficient to maintain the spread of STIs. Instead, a combination of interventions targeting specific high-risk groups is necessary to achieve a sustainable reduction in the spread of STIs. This approach is critical for the success of control programs.

James F. Blanchard, MD, PhD, Stephen Major, MD, MPH, Charles Caron, PhD, and Robert C. Bridger, MD, FRCP

Bacterial sexually transmitted diseases (STDs) remain an important public health problem worldwide. Most frequent STDs are highly prevalent among certain racial minorities and are therefore becoming the focus of major disease control efforts. This study aimed to assess the transmission dynamics of STIs in a geographically distinct population and to compare these findings with previous studies from other areas. The results indicate that the transmission dynamics of STIs in Winnipeg are similar to those in other urban areas, but there are differences in the demographic and sociocultural factors that influence the spread of these infections. Improved understanding of the transmission dynamics of STIs will be crucial for the development of effective control programs.
Methods

Descriptive Epidemiology

Manitoba is a province in Canada with a population of approximately 1.2 million. Winnipeg, with a population of close to 600,000, is the only major city in the province. All cases of chlamydia and gonorrhea are notifiable to the provincial health department by both physicians and laboratories under the Public Health Act. Since bacteriuria in the general population is a high proportion of positive diagnostic tests for chlamydia and gonorrhea, the laboratory-confirmed cases are reported to Winnipeg, a special part of the provincial health surveillance system is conducted for most persons infected with chlamydia or gonorrhea. Sexual contacts of those with chlamydia and gonorrhea are also notifiable to the provincial health department. As a measure of comparing the incidence of chlamydia and gonorrhea by geographic location, the province was divided into 322 geographic sites defined by pre-existing administrative boundaries (280 rural municipalities for non-Winnipeg residents and the 24 postal areas for Winnipeg residents). The average annual incidence of chlamydia and gonorrhea per 100,000 population was computed on the basis of all reported cases. Incidence rates were directly standardized for age and gender to the total 1991 Manitoba population. Pearson's moment correlation coefficients were used to assess the geographic relationship between chlamydia and gonorrhea rates. A separate 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.

Winnipeg Data Source

Winnipeg was divided into 3 geographic risk areas ("core," "special," and "periphery") separately for chlamydia and gonorrhea. This was done empirically by aggregating postal areas based on their average annual incidence rates between 1991 and 1995 and visually selecting the break points from the distribution curve (see Figure 3). The Census of Canada was used to estimate unemployment rates, mean household income, language usually spoken at home, mobility status, and population density for the Winnipeg postal areas (Statistics Canada, 1996, Manitoba 1301). The incidence of other selected communicable diseases was derived from the provincial notifiable disease registry.

Results

STD Incidence Rates by Geographic Site

Figure 1 shows the incidence of chlamydia and gonorrhea by geographic location, the province was divided into 322 geographic sites defined by pre-existing administrative boundaries (280 rural municipalities for non-Winnipeg residents and the 24 postal areas for Winnipeg residents). The average annual incidence of chlamydia and gonorrhea per 100,000 population is differen...
Hypothesis test for correlation between gonorrhea rate and chlamydia rate

1. The hypothesis: \[ H_0: \rho = 0 \text{ vs } H_1: \rho \neq 0 \]

2. The assumptions: random sample, bivariate normal distribution

3. The \( \alpha \)-level: \( \alpha = 0.05 \)

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