WINE CONSUMPTION AND HEART DISEASE

10. Brief Version of the Case Study

10.1 Problem Formulation

The goal of the case study is to examine the relationship between the average wine consumption and the mortality rate of ischemic heart disease for men aged 55 to 64 years old for 18 industrialized countries. The simple linear regression and SPSS will be used to describe the relationship between the two variables.

The data from the experiment are available in the SPSS file wine.sav located in the STAT 252 directory on the FTP server.

The following is a description of the variables in the data file:

Column	Name of Variable	Description of Variable
1	COUNTRY	Country Name
2	WINE	Average wine consumption (in liters per person)
3	MORTAL	Number of Ischemic Heart Disease Deaths (per 1,000 men aged 55 to 64 years old)

We will use SPSS to answer the following questions using the data:

1. Do these data suggest that the heart disease rate is associated with average wine consumption? If so, how can that relationship be described?

2. Is the simple linear regression model appropriate to describe the relationship between wine consumption and mortality of heart disease?

3. Do any countries have significantly higher or lower death rates than others with similar wine consumption rates?

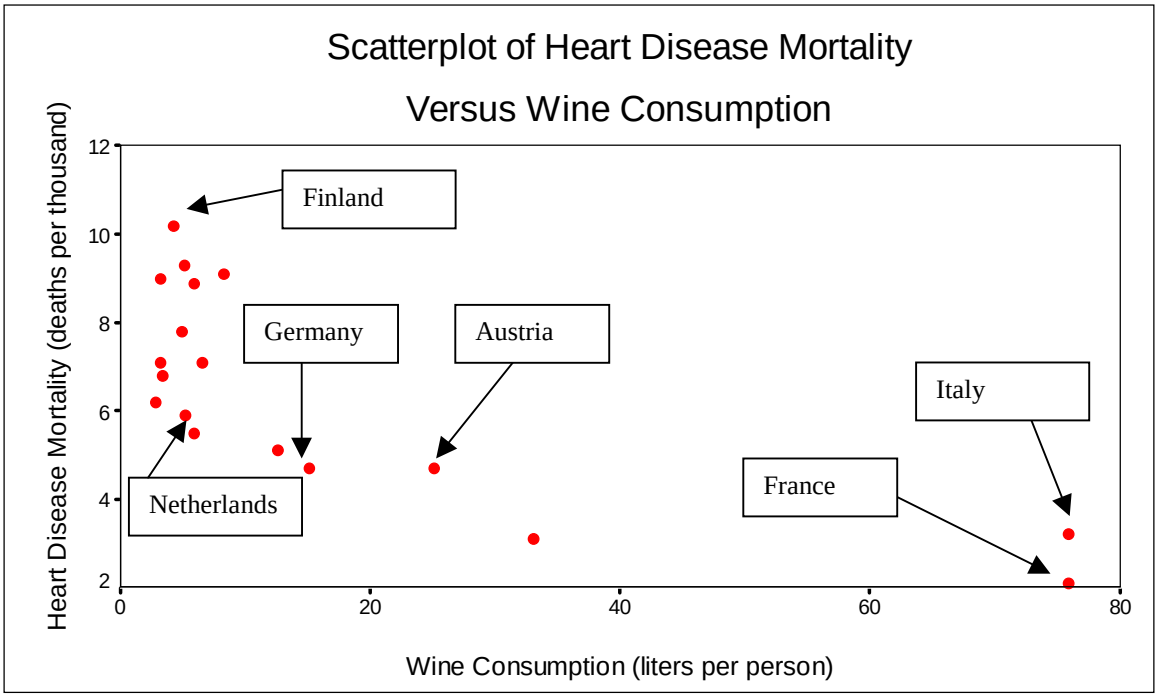
10.2 Study Design

The 18 industralized countries used in the study were not selected randomly from any well-defined population of countries. Therefore, the observed pattern cannot be inferred to hold in some general population, for example the population of all industralized countries, unless we assume that the countries are representative of the population.

This is an observational study, so no causation can be inferred. Though, there is a strong negative association between death rates of heart disease and wine consumption, one cannot claim that drinking wine causes a reduction in heart disease deaths. One cannot rule out the possibility that confounding variables are responsible for the observed differences among the death rates in the 18 countries. The 18 countries differ in many aspects, social, environmental, and cultural. These differences, not the level of wine consumption might be responsible for the observed differences in death rates of heart disease.

10.3 Displaying and Describing the Relationship between Heart Disease Mortality and Wine Consumption

SPSS produces the following scatterplot of heart disease mortality versus wine consumption on the original scales of measurement:



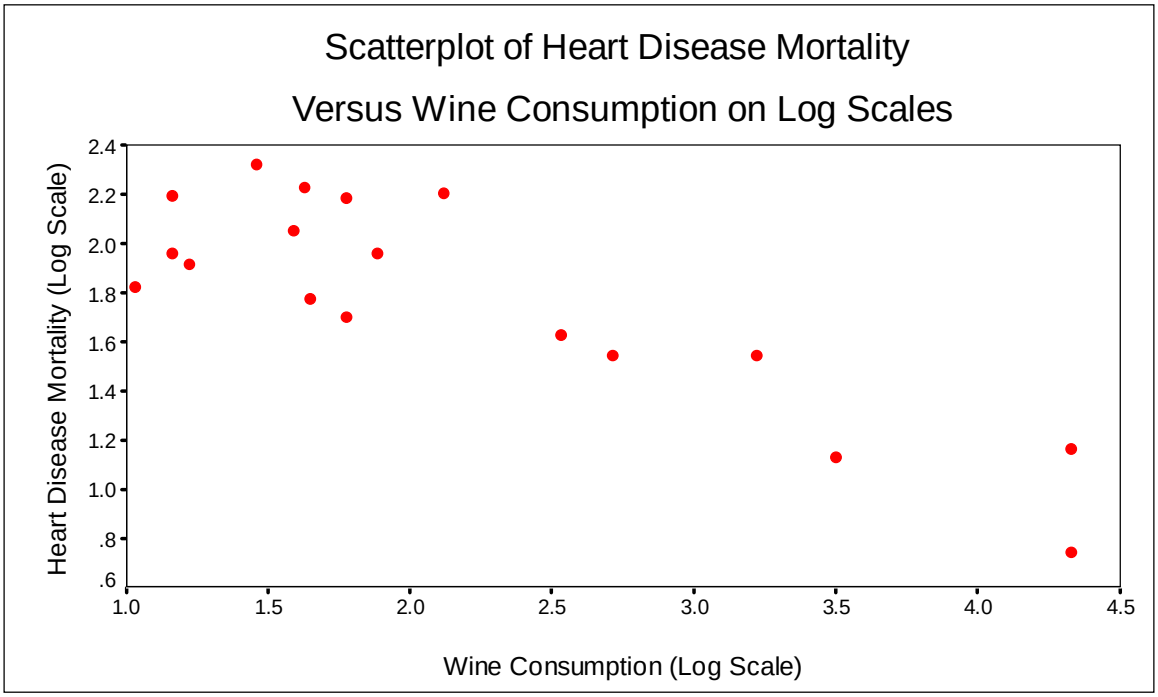
The pattern in the data shows that countries with high wine consumption have very low death rates and countries with low wine consumption have high death rates. It looks there is a strong negative association between heart disease mortality and wine consumption.

France (the lowest point in the lower right part of the plot) with the wine consumption of 75.9 litres per person has the lowest heart disease mortality (2.1 deaths per thousand). However, Italy has the same average wine consumption but much higher (3.2) mortality

rate. On the other hand, Germany and Austria have identical heart disease mortality but the wine consumption in Austria (25.1) is much higher than in Germany (15.1). Finland and Netherlands have very similar wine consumption (4.3 and 5.2, respectively) but very different heart disease mortality (10.2 and 5.9, respectively). Therefore, even if the wine consumption indeed has an effect on the mortality rate, there still exist other factors that make the rates different in countries with similar wine consumption.

A strong relationship exists between mortality and wine consumption, but a straight line does not provide an adequate description of the data. More precisely, mortality rate decreases exponentially as wine consumption increases. This pattern suggests the need for transformation. France and Italy are clearly separated from the rest of the data, and therefore they can be treated as outliers.

The natural logarithm transformation applied both to the response variable (mortality) and explanatory variable (wine consumption) produces the following plot:



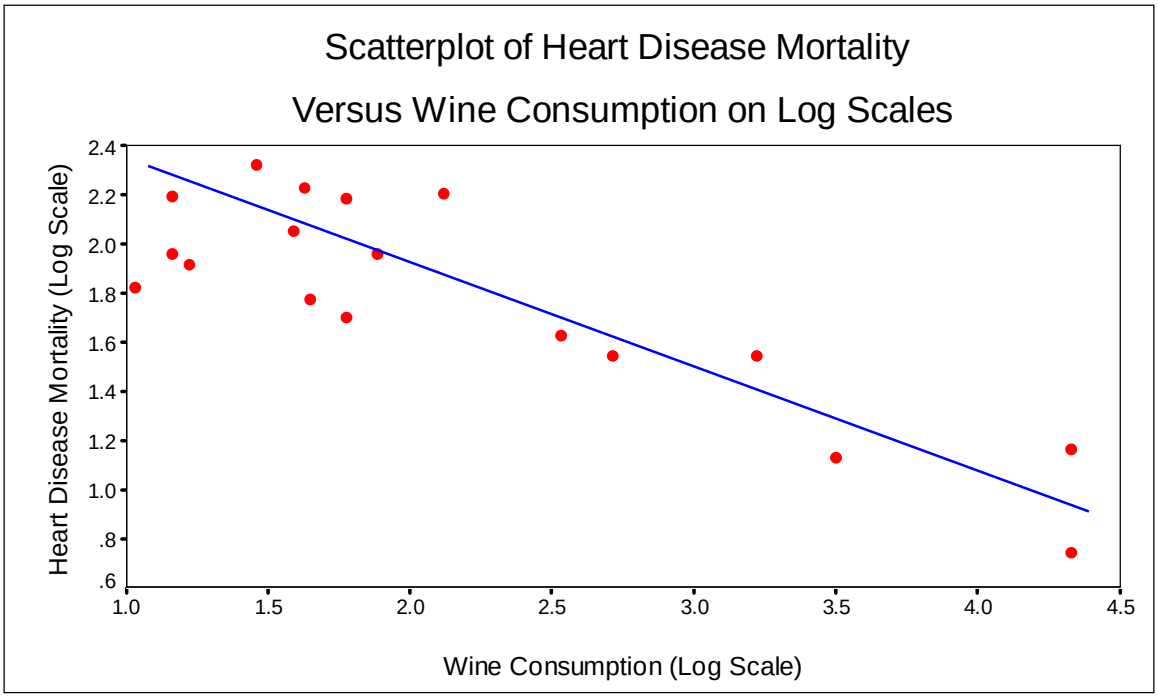
As you can see, the logarithm transformation applied to both variables made the relationship between the two variables approximately linear. Thus, the Pearson correlation coefficient can be used to measure the strength and direction of the relationship between the log-transformed variables:

PEARSON CORRELATION COEFFICIENTS		
	LNMORTAL	**LNWINE**
LNMORTAL	1.00	-0.8593
LNWINE	-0.8593	1.00

The above matrix shows that there is a high negative correlation between the transformed observations (on log scale).

10.4 Simple Linear Regression Model

In the previous section we obtained a scatterplot of heart disease mortality on the log scale versus wine consumption on the log scale. The plot revealed a linear pattern between the two variables.



Thus the following simple regression models is suitable:

$$LNMORTAL = \beta_0 + \beta_1 * LNWINE + ERROR.$$

Here LNMORTAL is the natural logarithm of MORTAL, and LNWINE is the natural logarithm of WINE. The random variable *ERROR* is assumed to follow a normal distribution with the mean of zero and an unknown standard deviation σ. The standard deviation is constant at all levels of *LNWINE*. The variable *ERROR* follows a normal distribution at each level of LNWINE.

The simple linear regression model can be stated equivalently as follows:

$$\mu\{LNMORTAL \mid LNWINE)\} = \beta_0 + \beta_1 * LNWINE.$$

From the above equation, we can obtain

$$MEDIAN\{LNMORTAL \mid LNWINE\} = \exp(\beta_0)(WINE)^{\beta_1}.$$

The SPSS simple linear regression model output for the problem has the following form:

LINEAR REGRESSION

Multiple R	.85932
R Square	.73843
Adjusted R Square	.72209
Standard Error	.22854

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	2.35918	2.35918
Residual	16	.83566	0.05223

F = 45.16980 Signif F = .0000

According to the output, the value of the correlation coefficient between the logarithm of heart disease mortality and the logarithm of wine consumption is 0.85932. The value of R^2 (0.73843) says that 73.843% of the variation in the log-mortality was explained by the linear regression on log-wine consumption.

We analyze the ANOVA table associated with the simple regression. The sum of squares due to the regression model is reported as 2.35918, and the sum of squares due to error (residual sum of squares) is 0.83566. The residual mean square is an estimate of the variance σ^2 and is equal to 0.05223. Thus an estimate of the residual standard deviation is $\sqrt{0.05223} = 0.228539$.

The value of the F statistic is equal to 45.16980 with the corresponding p-value of 0 provides very strong evidence of the utility of the model.

Now we analyze the part of the output providing the estimates of the regression parameters.

---------- Variables in the Equation ----------

Variable	B	SE B	95% Confidence Interval B		Beta
LNWINE	-.355596	.052909	-0.467759	-0.243433	-0.859321
(Constant)	2.555552	.126897	2.286542	2.824561	

Variable	T	Sig T
LNWINE	-6.721	.0000
(Constant)	20.139	.0000

According to the output, the estimated regression line of heart disease mortality (log scale) on wine consumption (log scale) is

$$\mu\{LNMORTAL \mid LNWINE)\} = -0.355596 * LNWINE + 2.555552$$

The association between mortality (log scale) and wine consumption (log scale) is negative and significant (estimate of the slope is -0.355596 with reported p-value of zero).

The estimated regression line was obtained when both the response and explanatory variable were logged. The log transformation was necessary to fit the data to a straight line model and to make the assumptions of the simple linear regression model satisfied. The validity of the simple linear regression assumptions for the data is discussed in **Section 6**.

Based on the estimated regression line equation, we have

$$MEDIAN\{LNMORTAL \mid LNWINE\} = \exp(2.555552)(WINE)^{-0.355596},$$

or

$$MEDIAN\{LNMORTAL \mid LNWINE\} = 12.87841 * (WINE)^{-0.355596},$$

Thus, a doubling of wine consumption is associated with a $2^{-0.355596} = 0.781568$ fold change in the median of heart disease mortality (22% drop in heart disease mortality).

10.5 Summary

As we stressed it earlier, this is an observational study, so no causation can be inferred. We cannot claim that wine has a protective effect against heart disease. We can only state that there is strong evidence of a negative association between heart disease mortality and wine consumption. One cannot rule out the possibility that confounding variables are responsible for the observed differences among the death rates in the 18 countries. The 18 countries differ in many aspects, social, environmental, and cultural. These differences, not the level of wine consumption might be responsible for the observed differences in death rates of heart disease. For example, as wine drinking is said to be related to a relaxed way of living, it is possible that this attitude, very common in some of these countries, reduces the risk of heart disease.

Further, the countries were not randomly selected. The statistical results apply only to the participating countries. Any extrapolation of the pattern to other countries comes from the assumption that the relationship between wine consumption and heart disease is similar for others. The best wording of results would emphasize that the association could not have arisen from a random assignment of mortality numbers to wine consumption values.

The study was conducted only on men, so we cannot say for sure whether a similar association holds for women.