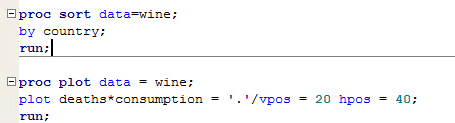
**Homework 4 STAT: 2010 Statistical Methods and Computing**

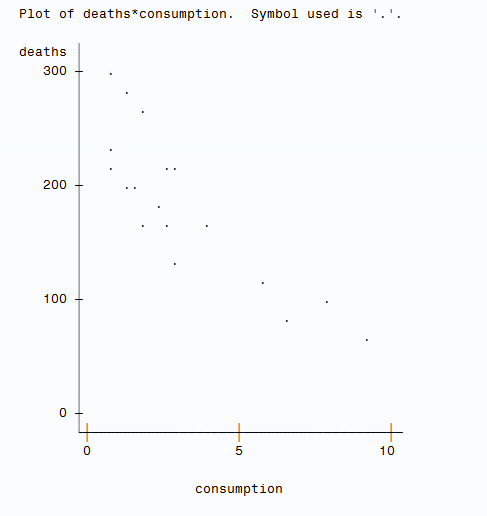
**Yubing Li**

**00808366**

**1. Winehrt**

**(a) Scatterplot**:



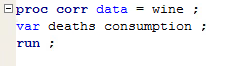


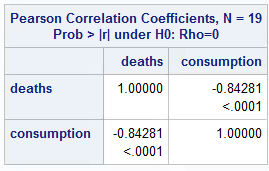
**(b)** There’s a linear pattern in the plot. We could tell from the scatterplot that the association between the consumption of wine and the deaths from heart disease is negative, roughly linear, and moderately strong.

**(c)** The direction is negative.

A negative linear relationship indicates that the more wine a people consume, the smaller possibility he has to die of heart disease.

I think the data give good evidence to show the relationship because the pattern is moderately strong, negative linear, with no unusual features.

**(d) Correlation coefficient:**

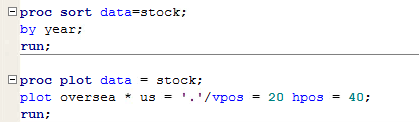


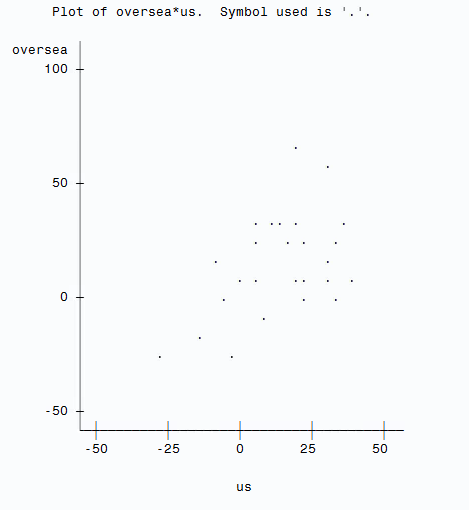
The correlation coefficient r is -0.8428.

**(e)** The correlation coefficient r will not change if the unit of the wine consumption changes.

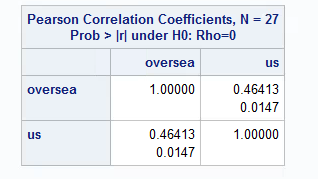
**2. Stocks**

**(a) Scatterplot:**





**(b) Correlation coefficient:**

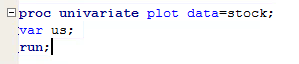


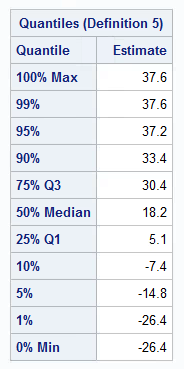
The correlation coefficient r is 0.46413.

The r value is greater than 0, which indicates there’s a positive relationship between U.S. and overseas returns. However, the value is not close to 1, so we say there’s not a very strong linear relationship.

**(c) 5-number summary & boxplot:**

**U.S.**





Minimum is -26.4

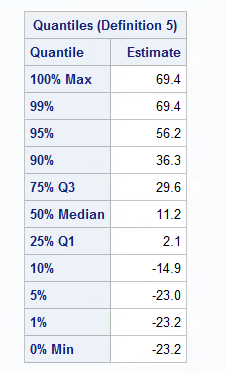
1st quartile is 5.1

Median is 18.2

3rd quartile is 30.4

Maximum is 37.6

**Overseas**



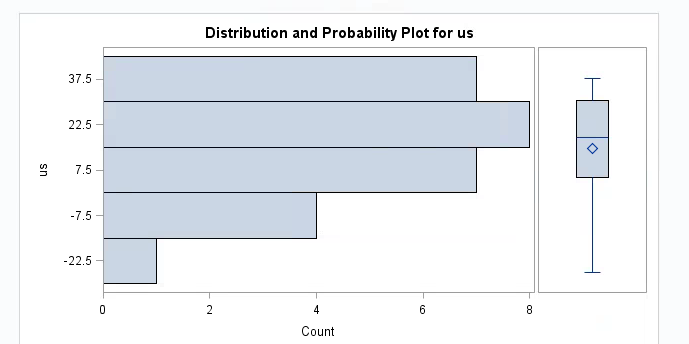
Minimum is -23.2

1st quartile is 2.1

Median is 11.2

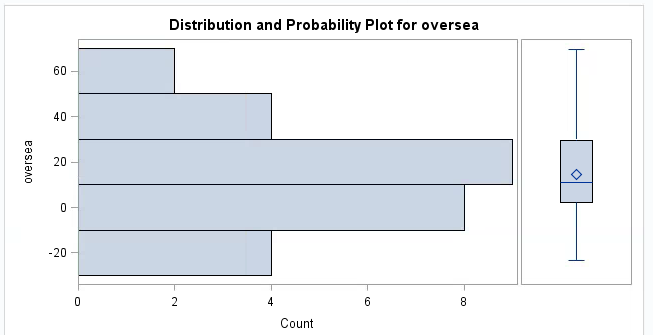
3rd quartile is 29.6

Maximum is 69.4



Boxplot for U.S. returns

Boxplot for overseas returns

**(d)** The returns are generally higher in U.S. in this period.

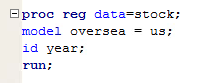
The 5-number summary shows that Q1, median and Q3 are higher in U.S. than overseas. In the boxplots, the values are concentrated in the interval, approximately from 20 to 40 for U.S. returns while the values are concentrated in the interval, approximately from 0 to 20 for overseas returns.

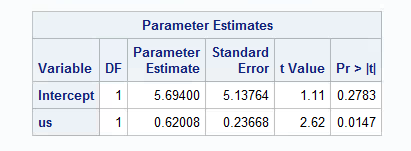
**(e)** The returns are more volatile in overseas in this period.

The 5-number summary suggests that the minimum is lower in overseas and the maximun is higher in overseas, which indicates the values are more spreed out in overseas.

The boxplots show that there are more values in overseas than in U.S. are off the higher bound and lower bound.

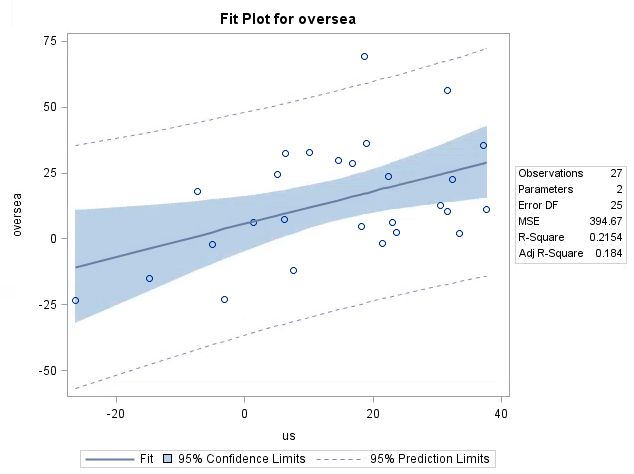
**(f) Regression:**



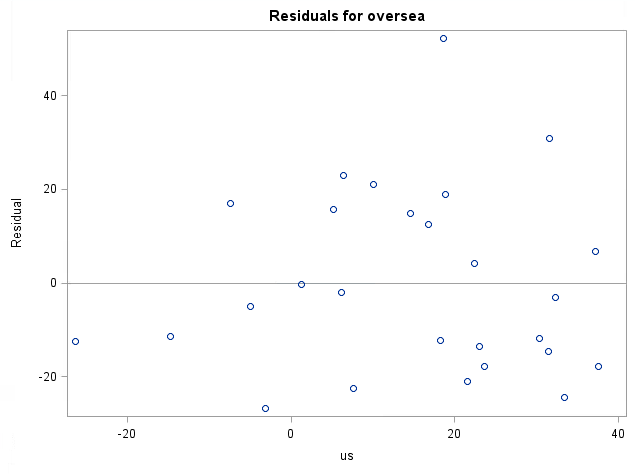


The least-squares regression model is

= 5.694 + 0.62008 × (U.S. returns)

**(g) Scatterplot w/ regression:**

“Fit line” is the least square regression line

**(h) Residual plot:**

**(i)** The predicted overseas stocks is 5.694 + 0.62008 × (33.4%) = 26.4047%

The predictions using the regression is 24.3047% higher than the actual value. Therefore, we are not confident to use this regression model to predict overseas stocks from U.S. stocks.

**(j) See the circles in (a), (g) and (h)**

The year is 1986.

**(k)** The point we circle in **(j)** is very influential, it’s an outlier and if we remove it, the regression model will be very different.

The points in dash circles are less influential than the one we circle in **(j)**.