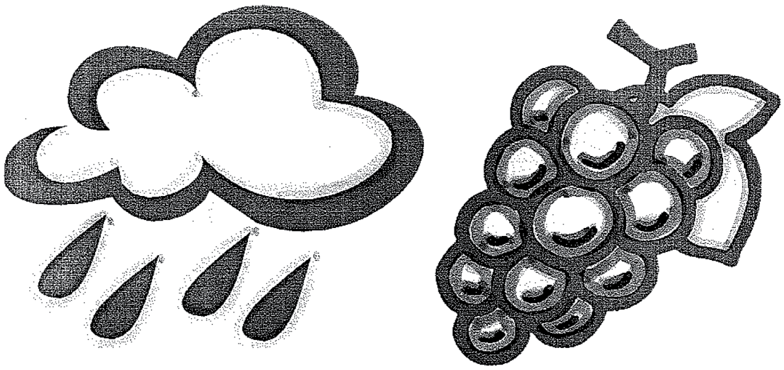
Maths Studies Project

Rainfall compared to Grape Vine Yield



Contents

	Page no.
A. Statement of Task	1
Plan	1
B. Data Collection and Presentation	2
Tables	2
Histograms	3
Scatter Diagrams	4 - 6
C. Analysis	7
Measures of Central Tendency	7 - 8
Measures of Dispersion	8 - 9
Correlation Coefficient	9
Linear Regression	9
D. Evaluation	10

A2

Maths Studies Project

A. Statement of Task

The aim of this project is to observe whether there is a correlation between rainfall and yield of specific cultivars by using statistics from the farm Constantia Uitsig in Cape Town, South Africa. The specific cultivars being: Sauvignon Blanc, Chardonnay and Cabernet Sauvignon. In order to understand this project one must realise that for rainfall the years range from 1994-2002 and for the yields they range from 1995-2003. This is because the rainy season is winter being June, July, August and this affects the yield for the following year when it is harvested in February and March. The Null-Hypothesis of this Project is: The more the rainfall in year n the more the cultivar's yield in year $n+1$.

Plan

The plan that this project will follow is:

- Collect data
- Organise data into comprehensive and clear tables
- Use tables to make a bar graph with the different information from the tables.
- Make use of different types of graphs, being: a Histogram or Bar Graph and a Scatter Diagram. Each graph showing yield compared to rainfall.
- Analyse this data by using the data to perform simple mathematical processes such as discovering the mean, median, mode and standard deviation.
- Use this data to perform more advanced, sophisticated and complex mathematical processes such as the line of best fit for the scatter diagram as well as the correlation coefficient (r) and line of best fit in order to see whether there is a correlation between these two factors and prove whether the null-hypothesis dependent or independent.
- Look at the graphs (as well as conclusions from mathematical processes) to decipher whether there is a correlation between these two factors or whether there is evidence to suggest that the amount of rainfall positively or negatively affects the Yields of different cultivars.

1

B2

B. Data Collection and Presentation

The data for this project was obtained by making an appointment with the manager of Constantia Uitsig farm from whom the data was obtained. Constantia Uitsig collects daily rainfall data and adds this together to get total monthly rainfall measurements (as seen in Diagram 1.1). Constantia Uitsig also collects the data for Tons/Hectare of each Cultivar's yield/year. With this information readily available I was able to make tables and subsequently graphs. This project will use selected data from these tables, which is represented in bold print.

Statistical data that has been collected is normally presented in the form of a table of values such as in Diagrams 1.1 and 1.2. In the table containing rainfall at Constantia Uitsig the monthly rainfall is included, although this project will be referring to the annual rainfall in most cases

Diagram 1.1
Table Showing Rainfall Data

Year:	Annual Rainfall	Mean Annual	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1994	**1041**	86.8	25	4	3	43	78	543	162	66	74	21	20	2
1995	**808**	67.3	9	19	8	23	82	115	220	134	36	108	22	32
1996	**1107**	92.3	1	49	40	37	63	224	144	150	196	87	56	60
1997	**724**	60.3	16	2	13	46	102	183	42	177	18	23	81	21
1998	**909**	75.8	12	-	19	66	296	104	172	79	59	11	66	25
1999	**1027**	85.6	2	4	-	108	65	180	220	242	147	3	42	14
2000	**784**	65.3	34	-	18	19	148	96	152	88	179	16	12	22
2001	**1356**	113.0	21	20	2	51	226	89	387	342	137	51	26	4
2002	**1035**	86.3	117	32	38	76	134	178	202	134	44	38	32	10

Diagram 1.2
Table Showing Cultivar Yield in Tons/Hectare

Cultivar	1995	1996	1997	1998	1999	2000	2001	2002	2003	Mean Yield
Sauvignon Blanc 1	**5.02**	**8.05**	**8.44**	**6.8**	**6.45**	**9.5**	**6.9**	**6**	**5.4**	7.0
Sauvignon Blanc 2	5.7	6.6	8.5	7	5.39	7.76	6.4	6.8	4.6	6.5
Chardonnay 1	**6.2**	**4**	**8.8**	**13.6**	**2.5**	**9.2**	**3.2**	**4.4**	**8.1**	6.7
Chardonnay 2	6.3	3.8	8.3	11	3.86	8.8	5.5	4.5	8.6	6.7
Cabernet Sauvignon 1	**5.9**	**6.7**	**11.7**	**6**	**5.1**	**4**	**4.82**	**5.64**	**7**	6.3
Cabernet Sauvignon 2	6.1	7.6	9.6	6.45	3.85	9.7	7.64	3.12	7	6.8
Semillon	**0**	**10.9**	**9.9**	**11.3**	**7.1**	**8,9**	**9.78**	**11**	**10.7**	8.8
Merlot	**8.7**	**7.02**	**16.16**	**9.9**	**10.3**	**13.97**	**9.4**	**9.8**	**12**	10.8

The following step is to present the data in a more meaningful way by drawing graphs. Most people find it easier and quicker to interpret data in picture form rather than in a table. This is clearly evident when reading newspapers and financial magazines, in this project the data will be represented by means of a Histogram or Bar Graph and Scatter Diagrams.

2

C2

Histogram

Diagram 2.1

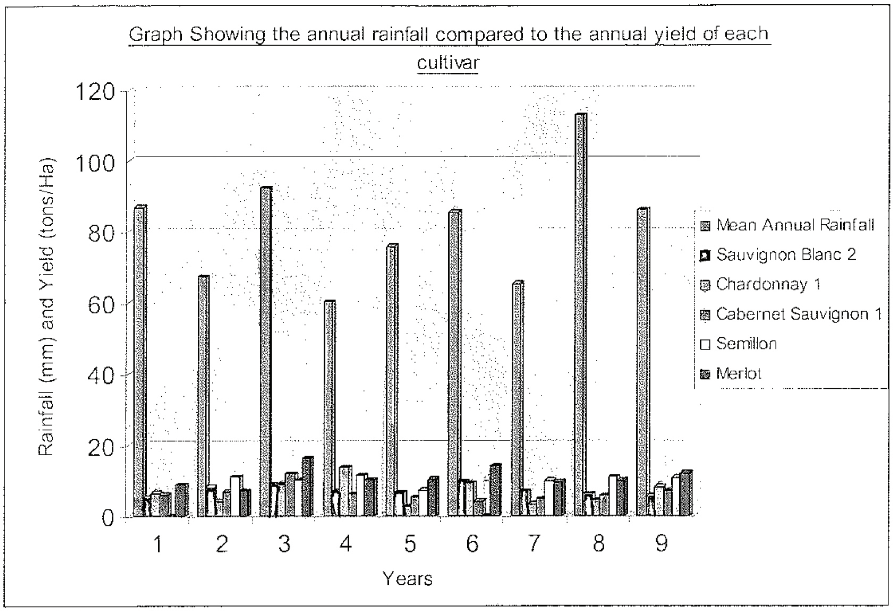
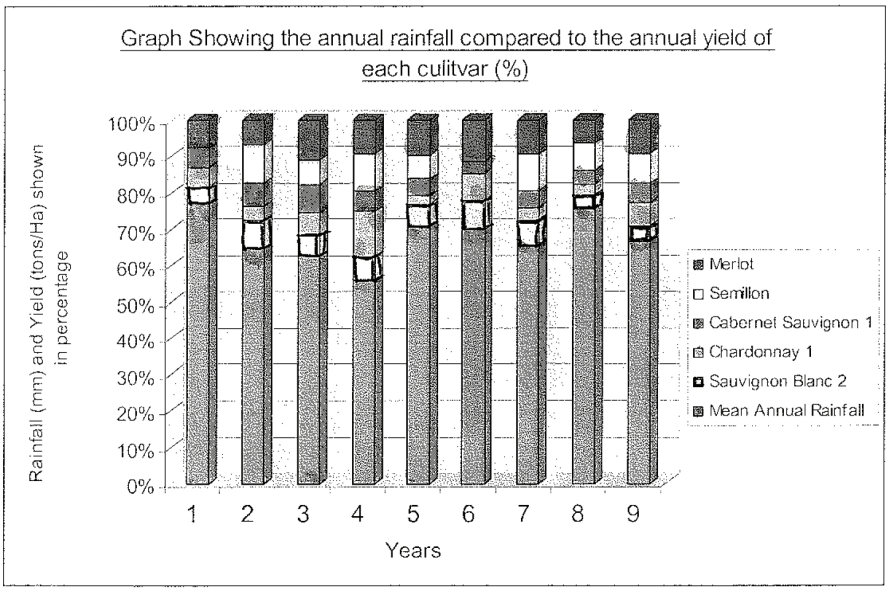


Diagram 2.2



3

C2

Scatter Diagram: This will later be used to find the **correlation coefficient (r)** to show the strength of the linear relationship. The line of best fit is also indicated.

Diagram 3.1

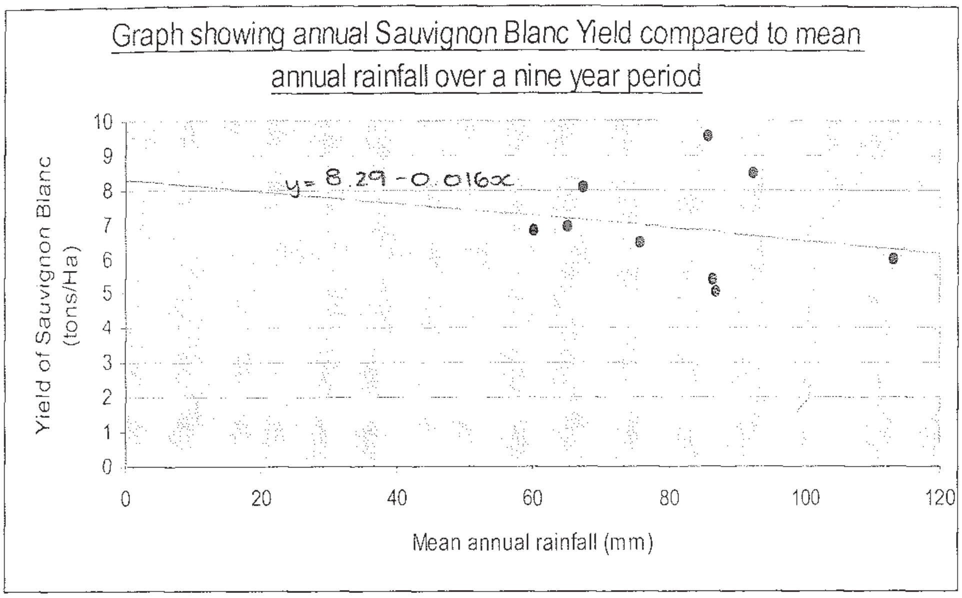
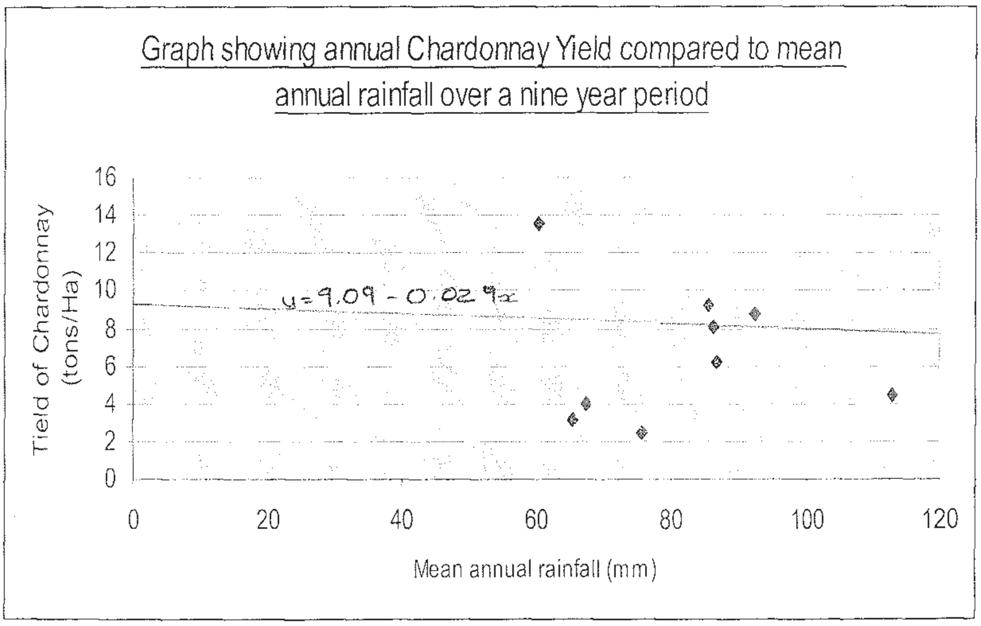


Diagram 3.2



4

Diagram 3.3

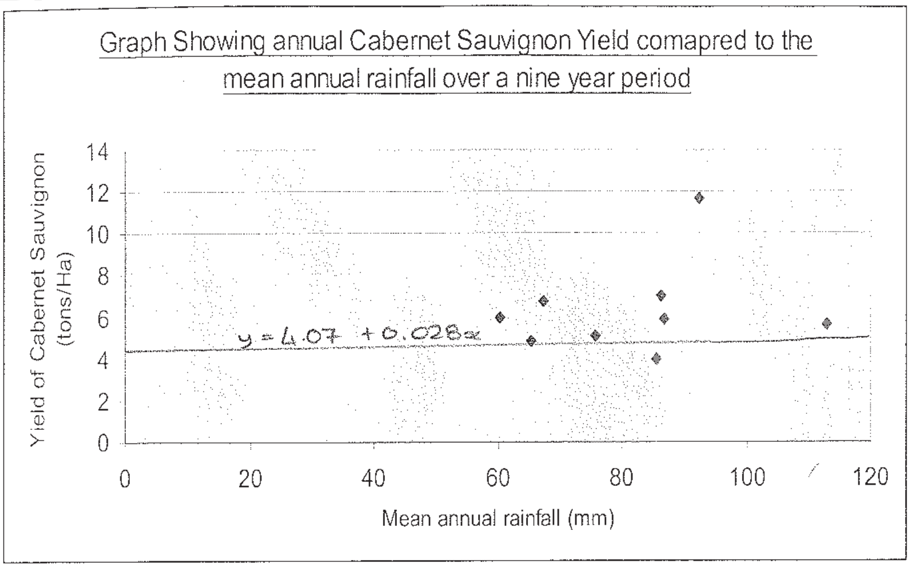
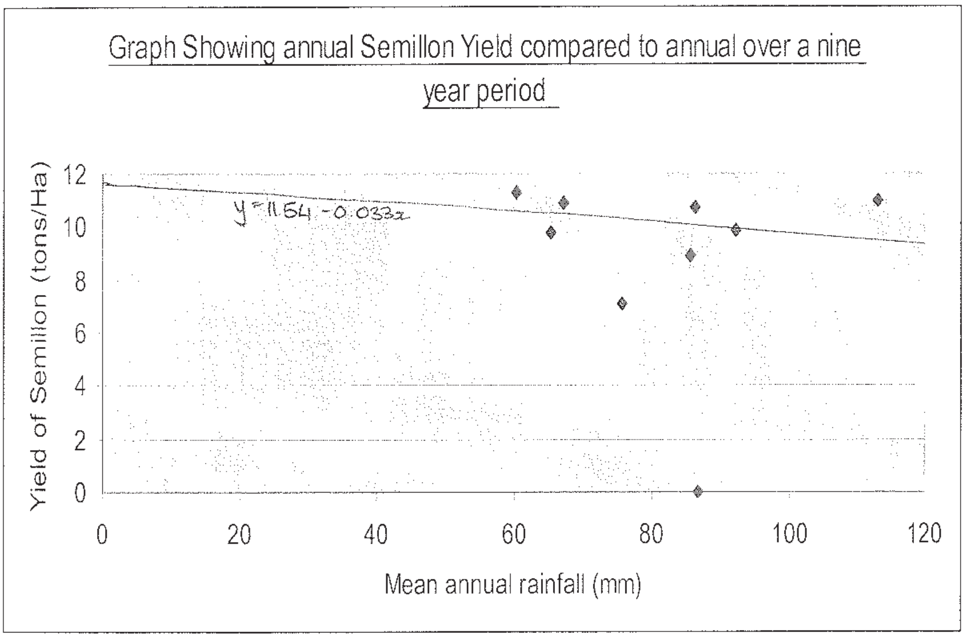
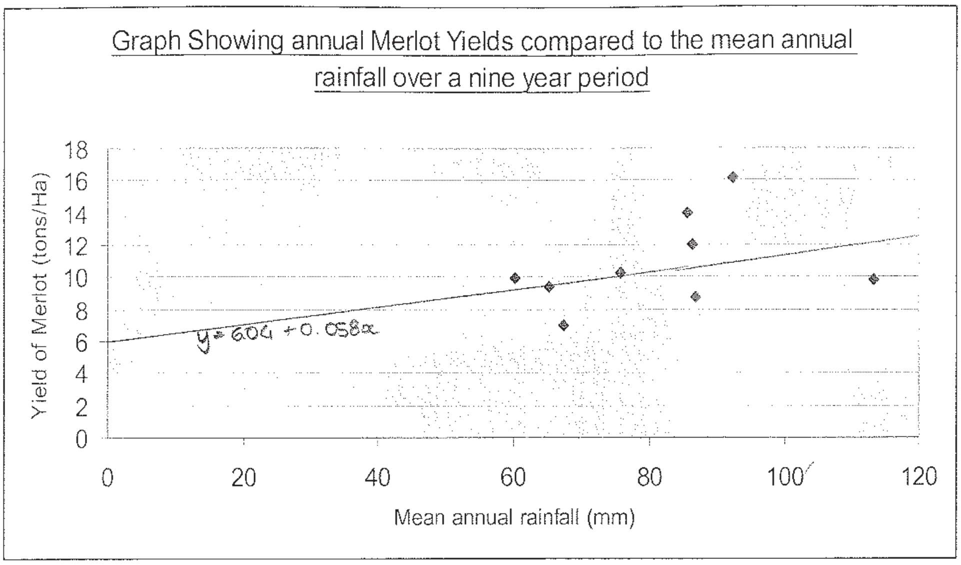


Diagram 3.4



5

Diagram 3.5


Graph Showing annual Merlot Yields compared to the mean annual rainfall over a nine year period
Yield of Merlot (tons/Ha)
18
16
14
12
10
8
6
4
2
0
0
20
40
60
80
100
120
Mean annual rainfall (mm)
y= 6.04 +0.058x

6

C2

C. Analysis

Measures of central tendency
In statistics one speaks of three different "averages", these can also be called "measures of central tendency". These averages are known as:

Mean: This is often referred to as the average. To obtain the mean one must add up all the scores and divide by the total number of scores. This is the method used to calculate the **mean annual rainfall** in Diagram 1.1 and the **mean yield** in Diagram 1.2.

Mode: This is the number or score which occurs most frequently. One can have more than one mode. E.g. Refer to Diagram 1.1: The mode for the annual rainfall is in 2001 with 1356mm. Refer to Diagram 1.2: The modes for cultivars in bold print are as follows:
Table Showing the Mode for different Cultivars
Diagram 4.1

Cultivar	Mode	Year
Sauvignon Blanc	9.5	2000
Chardonnay	13.6	1998
Cabernet Sauvignon	11.7	1997
Semillon	11.3	1998
Merlot	16.16	1997

Median: This is the middle-most score. The data must be arranged in ascending of descending order. If there is no middle score, one must take the mean of the two numbers in the middle. E.g.
Diagram 4.2

Year:	Annual Rainfall
1994	1041
1995	808
1996	1107
1997	724
1998	909
1999	**1027**
2000	784
2001	1356
2002	1035

This is the data in ascending order:
724;784;808;909;1027;1035;1041;1107;1356
1027 is the middle-most score; this is the median for this data.

Table showing the yield of selected cultivars in ascending order
Diagram 4.3

S. Blanc	5.02	5.4	6	6.45	**6.8**	6.9	8.05	8.44	9.5
Chard. 1	2.5	3.2	4	4.4	**6.2**	8.1	8.8	9.2	13.6
Cab. Sauv 1	4	4.82	5.1	5.64	**6.9**	6	6.7	7	11.7
Semillon	0	7.1	8.9	9.78	9.9	10.7	10.9	11	11.3
Merlot	7.02	8.7	9.4	9.8	**9.9**	10.3	12	13.97	16.16

7

The median for these selected cultivars is in bold print, except for Semillon, which has two middle scores. Therefore the median for Semillon is (9.9 + 10.7) ÷ 2 = 10.3

Measures of dispersion
This project will use and explore two different measure of dispersion, these being the range and the standard deviation.

Range: is important statistical measure of dispersion.

Finding the difference between the highest and the lowest scores in any set of data locate the "range" or "spread" of the scores.

With reference to Diagram 1.1 the range of the annual rainfall figures is 1356 - 724 = 632mm.

With reference to Diagram 1.2 the range for cultivars in bold print are as follows:

Table showing the range of selected cultivars
Diagram 5.1

Cultivars	Range
Sauvignon Blanc	4.48
Chardonnay	11.1
Cabernet Sauvignon	7.7
Semillon	4.2
Merlot	9.14

Standard Deviation: this measure of dispersion is considered to be better than the above mentioned as it takes all of the scores into account.

Calculating the Standard Deviation takes several steps, these will be indicated below by using the rainfall figures.

Step 1.

Year:	Total Annual	$(score - mean)^2$
1994	1041	4121.64
1995	808	28493.44
1996	1107	16952.04
1997	724	63907.84
1998	909	4596.84
1999	1027	2520.04
2000	784	37171.84
2001	1356	143792.64
2002	1035	3387.24

Step 2: Add all the numbers in the third column and divide by the number of scores = 33882.6

C2

8

C2

Step 3: Take the square root of the answer in step 2 = 184
Therefore the standard deviation for this data is 184mm from the mean.

Correlation Coefficient:
Also called the **Pearson correlation coefficient**, this more advanced mathematical process indicates the strength of the relationship between two factors. In this case, by putting rainfall on the x-axis and the yield (each separately) on the y-axis one can determine whether rainfall has a strong effect on the amount of yield produced. Rainfall, however, is independent of yield as the size of the yield does not affect the rainfall.

In Diagrams 3.1 to 3.5 the correlation coefficient is represented by r. Two forms of the formulae to compute this value are:
($\bar{x}$ and $\bar{y}$ are the mean and x and y are the scores)

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$

$$r = \frac{S_{xy}}{S_x S_y} \quad \text{where } S_{xy} = \frac{1}{n}\left(\sum xy - n\bar{x}\bar{y}\right)$$

These are the correlation coefficients for the scatter diagrams in this project. These results were obtained using the TI-83 Plus calculator:
Diagram 3.1: $r = -0.184$
Diagram 3.2: $r = -0.136$
Diagram 3.3: $r = 0.204$
Diagram 3.4: $r = -0.152$
Diagram 3.5: $r = 0$.

Linear Regression:
In Diagrams 3.1 to 3.5 the **line of best fit** is shown. "Simple linear regression" is used to determine the straight line that best represents the way the data is dispersed on the scatter diagram. The method that was used in this project to determine the line of best fit is called "The method of least squares" by which the "line of regression of y on x" ($y = mx + c$) is found.
These are the lines of best fit for the scatter diagrams in this project. These results were obtained using the TI-83 Plus calculator:
Diagram 3.1: $y = 8.29 - 0.016x$
Diagram 3.2: $y = 9.09 - 0.029x$
Diagram 3.3: $y = 4.07 + 0.028x$
Diagram 3.4: $y = 11.54 - 0.033x$
Diagram 3.5: $y = 6.04 + 0.058x$

9

D2

D. Evaluation

The Null-Hypothesis: The more the rainfall in year n the more the cultivar's yield in year $n+1$. This is rejected as this is not the case with every cultivar.

One can tell by looking at the Histograms/Bar Graphs Diagram 2.1 and 2.2 how the annual yield of each cultivar is affected by the mean annual rainfall. The units and therefore scales of these two variables are different making it difficult to compare with accuracy the rise and fall of rainfall compared to that of the yield. These graphs are helpful, however, to make an educated estimation of each different yield's reactions after a year's rainfall. By looking at these Graphs this project estimates that Sauvignon Blanc and Cabernet Sauvignon are not greatly affected by variations in the annual rainfall whereas Chardonnay's yield increases after a low rainfall year and decreases after a high rainfall year. Merlot seems to increase after a year with high rainfall and Semillon seems to follow no continuous pattern.

The Scatter Diagrams with mean annual rainfall on the x-axis and annual yield of each different cultivar on the y-axis are a more effective way of deciphering how strong the two variable's linear relationship is. By looking at the equation and angle of the line of best fit and the correlation coefficient (r) one is able to see the extent of the linear relationship between annual rainfall and yield. The conclusion drawn from this data is that there is generally a weak linear relationship between these two variables with some cultivars having stronger dependencies on rainfall than others (e.g. Merlot has a stronger linear relationship than the others). By looking at the Correlation coefficient (r) one can see how strong or weak the linear relationship is. By looking at these results this project came to these conclusions:

Diagram 3.1: weak
Diagram 3.2: weak
Diagram 3.3: bit stronger but weak
Diagram 3.4: weak
Diagram 3.5: stronger but weak

The positive and negative values of (r) indicate whether the cultivar increases or decreases respectively when the rainfall increases. Therefore one can say that Sauvignon Blanc, Chardonnay and Semillon decrease with too much rain where as Cabernet Sauvignon and Merlot increase. These increases and decreases are small however they are prevalent and therefore an adequate conclusion.

E0

When finding the **line of regression of x on y** one could go on to make predictions of the data through **interpolation** and **extrapolation.** This is not possible with this data as there is not a strong enough dependence.

The simple mathematical processes to gauge the Measures of Central Tendency and Dispersion do not tell us anything about the dependence of these two variables, they merely help us to understand and work with the data available. One can say though that for **Standard Deviation** the more widely dispersed the data the larger the Standard Deviation.

10

To finally bring all the information together one can make a simple graph to see the effects of heavy rainfall on each different yield. One could look more closely at each individual cultivar's reaction to heavy rainfall to gauge whether different types of cultivars prefer more or less rainfall.

D. Table showing individual affects on different yields after select years of heavy rainfall.

Year	Rainfall	Sauvignon Blanc	Chardonnay	Cabernet Sauvignon	Merlot	Semillon
1996	1106	Stayed Same	x 2 +	x 2	x 2 +	- 1
1999	1027	Slightly More	x 3 +	Same	+ 3	+ 1
2001	1356	Same	+ 1	+ 1	+ 2	+ 2
2002	1035	Same	x 2	+ 2	+ 2	- 1

It appears that Sauvignon Blanc yields do not increase or decrease dramatically after heavy rains. The Chardonnay yields increased dramatically after a year with a high annual rainfall. The Cabernet Sauvignon yields do increase after a year with a high annual rainfall but not always significantly. The Merlot yields increase steadily after a year with a high annual rainfall. The Semillon yields either decrease or increase marginally after a year with heavy rainfall.

To conclude there is not a dramatically significant dependence of yield on rainfall. In order to make a more accurate description one would possibly need to use a greater number of years and more data from different blocks of the same cultivar. One can also not make an accurate alliance between these two variables as there are many other factors which affect the yield. These being: strength of wind, wind direction, temperature, direction of slope, soil, diseases, insects, time of rainfall and others.

11