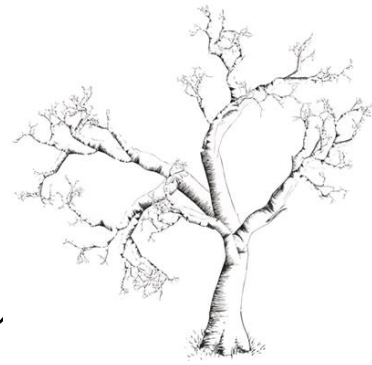


SAC REVISION



Tree Age, Diameter, Soil Quality and Growth Rate

The table below lists 24 eucalyptus trees planted on a site with their age(in years), diameter(in cm), soil quality(rated as A-Good or B-poor) and growth rate(rated as 1-fast, 2-medium or 3-slow).

1. For the numbers 1 – 100, write down the calculator steps and the numbers generated for:

- A random sample of 5

- A random sample of 20

Tree Age(yrs)	Diameter (cm)	Soil Quality	Growth Rate
4	2.03	A	1
34	20.32	B	2
5	2.03	A	1
8	2.54	B	2
40	19.05	A	2
8	5.08	A	3
10	7.62	B	2
30	15.24	A	1
12	8.89	A	1
13	12.45	B	3
16	6.35	B	1
20	11.68	B	2
22	13.97	A	1
25	14.73	B	2
28	16.51	B	3
30	15.24	A	2
14	8.89	B	2
33	17.78	B	1
14	8.89	B	2
35	16.51	A	3
38	17.78	B	3
10	5.08	B	3
42	19.05	A	1
18	11.43	A	3

Explain why more than the stated number of results may need to be generated.

Explain why a random sample is required.

For our analysis of this data we will use all the results in the supplied table.

2. For each of the four variables state the type of variable.

Tree age = _____

Diameter = _____

Soil Quality = _____

Growth rate = _____

3. State the type of graph required to allow a comparison of:

- age and growth rate

- soil quality and growth rate

- age and diameter

4. Construct a stemplot for the age of the 24 trees.

Remember to use the sort function in your spreadsheet page as it will make it easier to construct the stemplot.

DON'T FORGET A KEY

5. Construct a boxplot to allow comparison of soil quality and diameter.

What to show for a boxplot

- 5 number summary
- Outlier calculations and sentence
- Label axis and each boxplot

If asked to discuss the parallel boxplots then make comparisons between the shape, centre, spread and presence of outliers. Use words like higher or lower etc.

6. Construct a two-way frequency table to compare soil quality and growth rate. Prepare a parallel segmented bar chart to present this information. Interpret your results.

Two-way table

		EV	
RV			



Remember when calculating percentages work down each column and column total will be 100%.

When discussing if there is an association between the 2 variables make sure you refer to one row of %'s from the table.

DIFFERENCES EQUAL ASSOCIATION

7. Find the 68% confidence interval for the age of the trees.

Mean =

Std dev =

Lower limit =

Upper limit =

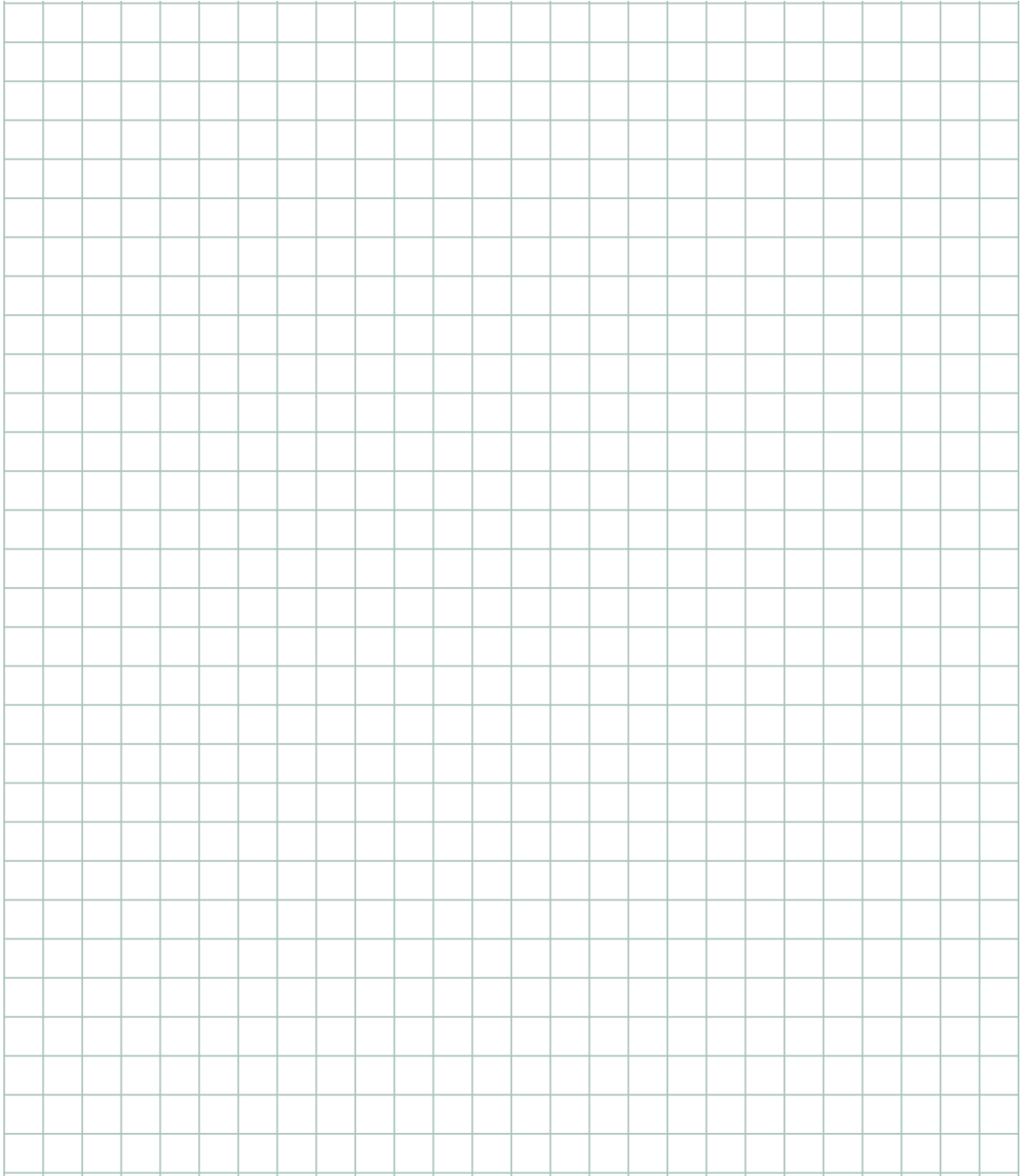
Therefore the 68% confidence interval is _____

If the lower limit is a negative number rewrite it as 0 as the data is all positive.

We would now like to determine how the size of the trees increases with age.....

8. Identify the explanatory variable, stating reasons.

9. Construct a scatterplot for this data, and comment on it.



Comment:

When commenting on a scatterplot you are to discuss form, direction, strength and the presence of outliers

10. Calculate the least squares regression model and write it in terms of the correct variables. You must draw the model on your scatterplot.

To draw the line on your scatterplot it is good to label the 2 sets of coordinates that you used.

Least squares regression line is : _____

11. Interpret the values of the gradient and y-intercept in the context of the problem.

Gradient =

Y-intercept =

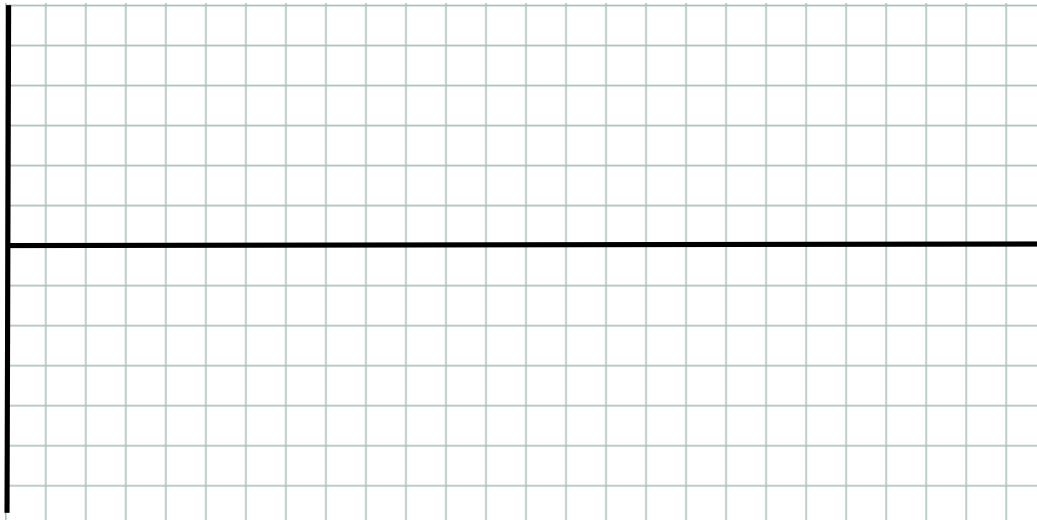
12. For the least squares regression model, find the values of the correlation coefficient and the coefficient of determination. Interpret both of these values in the context of the problem.

Correlation coefficient =

Coefficient of determination =

13. Using the least squares model calculate the residual value for the 4 year old tree. Show all steps of working.

14. Construct a suitably labelled residual plot. Comment on the suitability of the linear model.



Comment:

15. Two alternative models are tried which involve transformation of **one** of the variables:

$$y^2 \text{ vs } x$$

$$y \text{ vs } \log_{10} x$$

Construct a scatterplot and appropriate statistical results for each transformation to determine the most appropriate transformed model. You must provide an adequate explanation for the model you choose.

(Complete both of these on your CAS. After you have produced the Scatterplot and residual plot, complete the table below)

	Rule/Model/Equation	r	r^2	Residual shape
Original				
Diameter square				
log(age)				

Which model would you choose and why?

Compare residual plots and coefficient of determination

16. For the non-linear model you choose, find its rule. Use the model to predict the diameter of a tree whose age is:

27.5 years

45 years

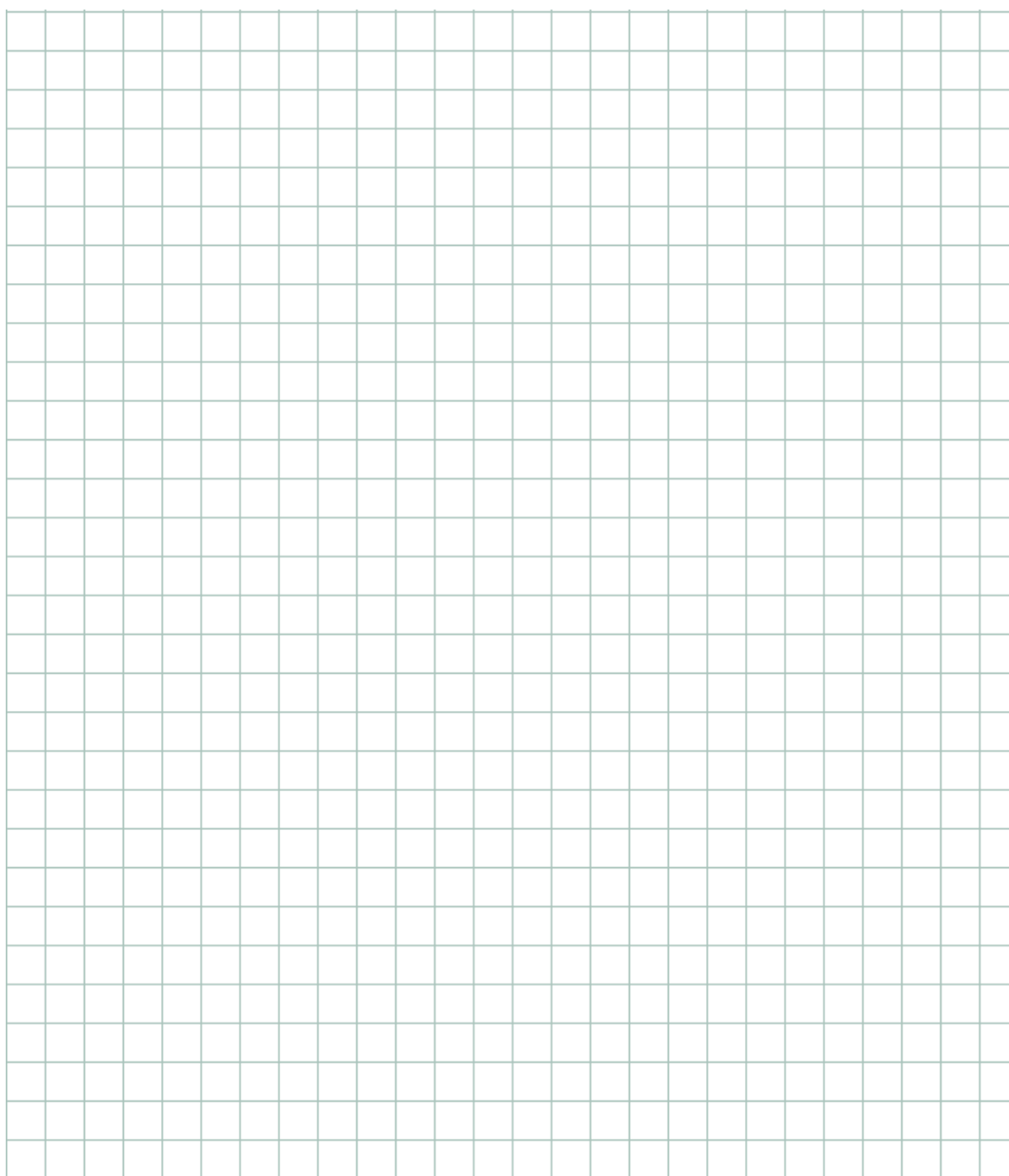
100 years

Comment on the reasonableness of these predictions in the context of the problem.

17. The data in the table below shows the rainfall (mm) in the area where the trees are growing. Create a time series plot ensuring that you include an appropriate scale, labels and titles.

YEAR 1	RAINFALL	YEAR 2	RAINFALL
January	100	January	100
February	215	February	155
March	30	March	55
April	85	April	70
May	80	May	95
June	40	June	100
July	50	July	60
August	30	August	45
September	80	September	45
October	80	October	25
November	130	November	85
December	50	December	40

The data represents two years of recorded rainfall and when you construct the time series plot it is one continuous graph so the January of Year 2 comes directly after December of Year 1 on the x axis



18. Describe the general pattern and variation for the above time series plot.

19. Use five-median smoothing to smooth the time series. Plot the smoothed time series on the plot above. Mark each smoothed data point with a cross (×).

20. In another part of the state, the rainfall was found to be seasonal and modelled using the seasonal indices/least squares regression method. The equation found was:

$$\text{Deseasonalised Rainfall} = 91.193 - 0.464 \times \text{month}$$

where *month* = 1 is January – Year 1

The seasonal index for February was 0.84. What does this tell us about rainfall in February?

The correct wording should not use seasonal average but refer specifically to the type of season, for this example it would be monthly average

21. Using the model given, predict the **actual** rainfall for February, Year 3.

Predicting actual rainfall but equation is for deseasonalised rainfall therefore it is a 2 step process to find the actual rainfall

22. The seasonal index for July is 1.20, by what percentage does the actual rainfall need to be adjusted to allow for seasonality.