SOLUTIONS FOR “TREE AGE AND DIAMETER”

1. This question requires a random sample to be taken so that bias is eliminated. The easiest way to do this is to go to StatList editor and put the cursor into the heading of list1. Then select F4, Probability, RandInt(1,100,5). Please note that the answers should contain no duplicates. So it sometimes easier to generate a few extra say RandInt(1,100,8) and take the first unique 5 numbers.
   To achieve a sample of 20 use RandInt(1,100,25)
   Random sample eliminates bias.

2.

Tree Age(yrs)	Diameter (cm)	Soil Quality	Growth Rate
Discrete numerical	Numeric continuous	Categorical nominal	Categorical ordinal

3.

- age and growth rate — parallel boxplots
- soil quality and growth rate — parallel segmented barcharts
- age and diameter — scatterplot

4.

Distribution of Ages for 24 trees

0	4	5	8	8				
1	0	0	2	3	4	4	6	8
2	0	2	5	8				
3	0	0	3	4	5	8		
4	0	2						

Key: 0|4 means 4 years

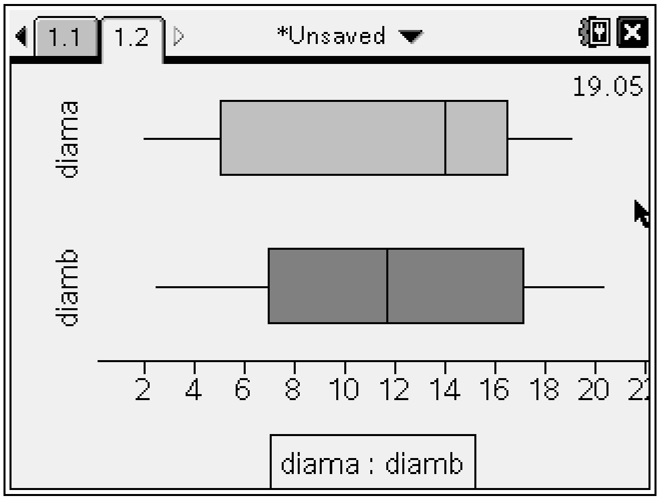
5. Soil quality is categorical and diameter is numerical. This then requires parallel boxplots. To prepare the data all results with a soil quality of A(good) could be entered in list1 and all other results, ie. soil quality B(poor) are entered in list2. Two boxplots are then defined in Plot Setup.
   Remember the following points:

   A clear even numerical scale is required below the two boxplots. You must not draw two separate scales as this means it makes it very hard to compare!

   The key points need not be labeled – but if your scale is good the reader should be able to interpret the values.

   Summary statistics:

Soil Quality A		Soil Quality B	
	min = 2.03		min = 2.54
	Q1 = 5.08		Q1 = 6.99
	Median = 13.97		Median = 11.68
	Q3 = 16.51		Q3 = 17.15
	Max = 19.05		Max =20.32



Soil Quality A outlier check
$5.08 - 1.5 \times 11.43 \le x \le 16.51 + 1.5 \times 11.43$
$0\ (-12.065) \le x \le 33.665$
All values lie within the range above therefore no outliers.

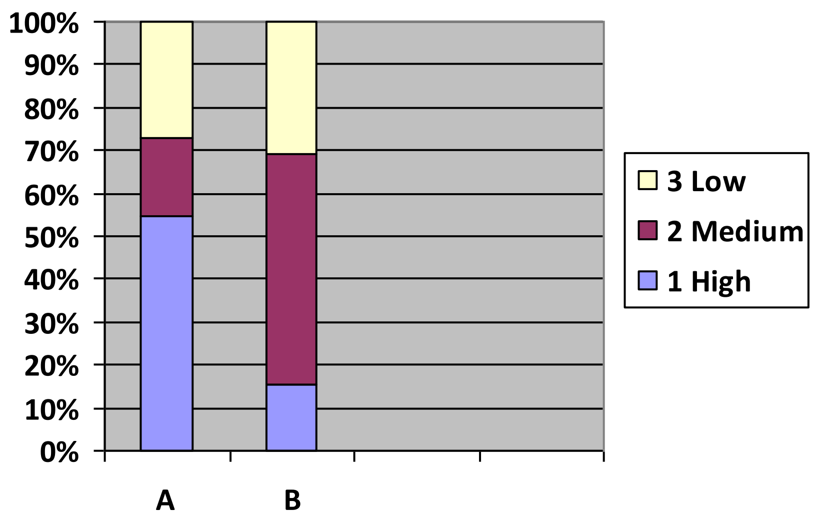
Soil Quality B outlier check
$6.99 - 1.5 \times 10.16 \le x \le 17.15 + 1.5 \times 10.16$
$0\ (-8.25) \le x \le 32.39$
All values lie within the range above therefore no outliers

Don't forget to label the axes with an appropriate name. And a title is good!

6. Start by sorting the data into A1, A2, A3, B1, B2 and B3. Note "soil quality" is the heading of the columns as this is the independent variable.

	Soil Quality	
Growth Rate	**A(Good)**	**B(Poor)**
1 (Fast)	6 (54.5%)	2 (15.4%)
2 (Medium)	2 (18.2%)	7 (53.8%)
3 (Slow)	3 (27.3%)	4(30.8%)
	11 (100%)	13 (100%)

Create two columns one for soil quality A and one for soil quality B. Each bar should go up to 100% on the frequency axis. Divide each column according to the percentages in the table above.



Make sure you have a title, label the legend and label each column.
Growth rate may be associated to soil quality as 54.5% of trees show a fast growth rate in good soil compared to only 15.45% of trees in poor soil.

7. Calculate the mean and standard deviation for the age of the trees using CAS 1 variable statistics.
mean = 21.21 standard deviation = 11.85

The 68% confidence interval is found by calculating two endpoints – the mean plus or minus one standard deviation.

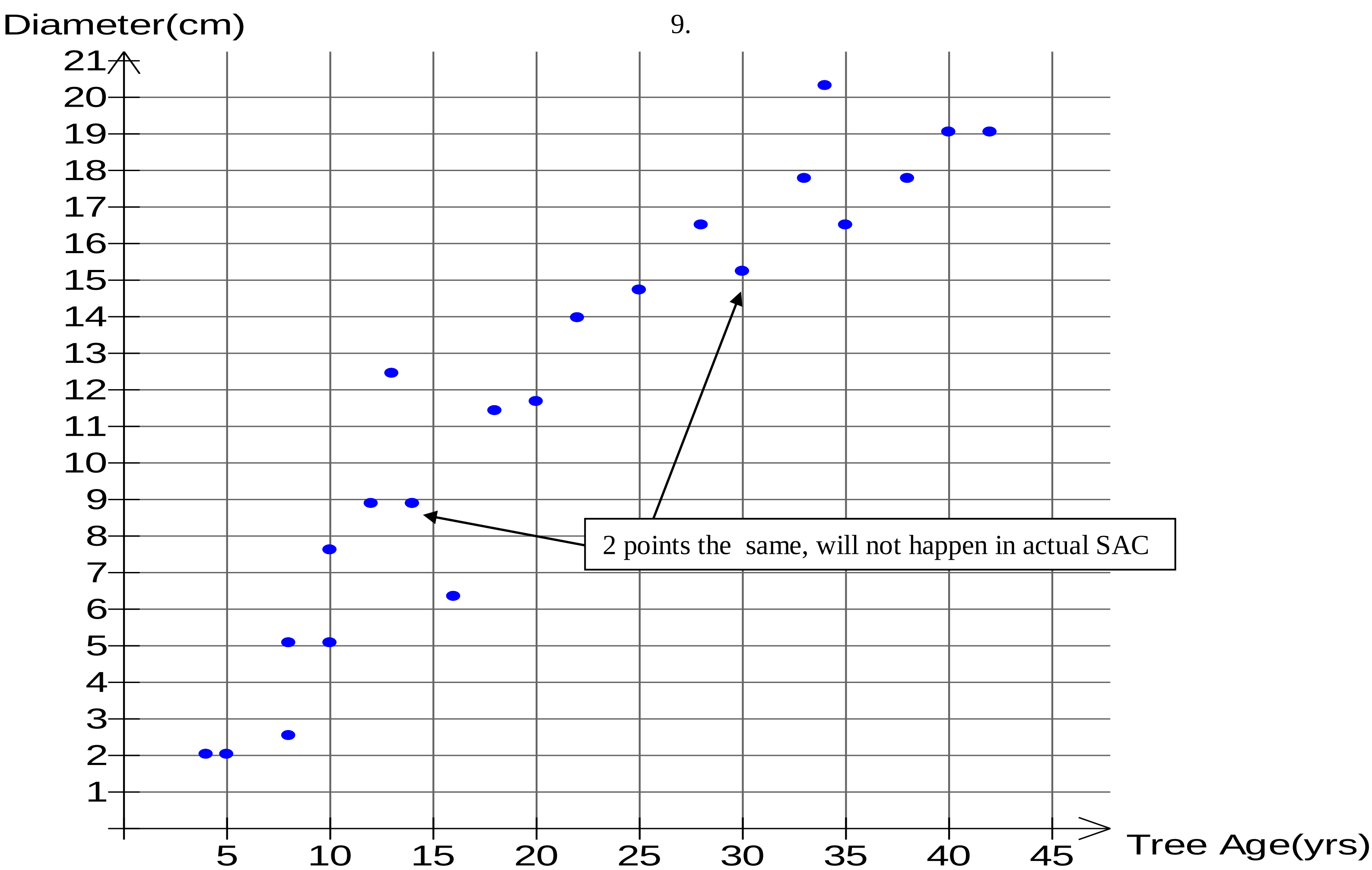
Lower limit 21.21-1 x 11.85 = 9.36
Upper limit 21.21+1 x 11.85 = 33.06

Therefore the 68% confidence interval is $9.36 \leq x \leq 33.06$

8. Explanatory variable is age as it is being used to predict the diameter.

Note the description in the question says that "we would like to determine how their size increases with age".

9.



From the scatterplot we can see that there is a strong positive linear association with no clear outliers.

10. The least squares regression line is found using CAS and is given by

$$Diameter = 1.847 + 0.461 \times age$$

When sketching least squares regression line onto scatterplot please calculate 2 sets of coordinates.

11. Gradient = 0.461

On average for every 1 year increase in age the tree's diameter increases by 0.461 cm

Y-int = 1.847

The tree would be 1.847cm in diameter at 0 years of age (applicable in this case?)

12. $r = 0.945$

This indicates a strong positive association

$r^2 = 0.893$

This indicates that 89% of the variation in diameter can be explained by the variation in age.

13. To calculate a residual you must first find the predicted value using the least squares model.

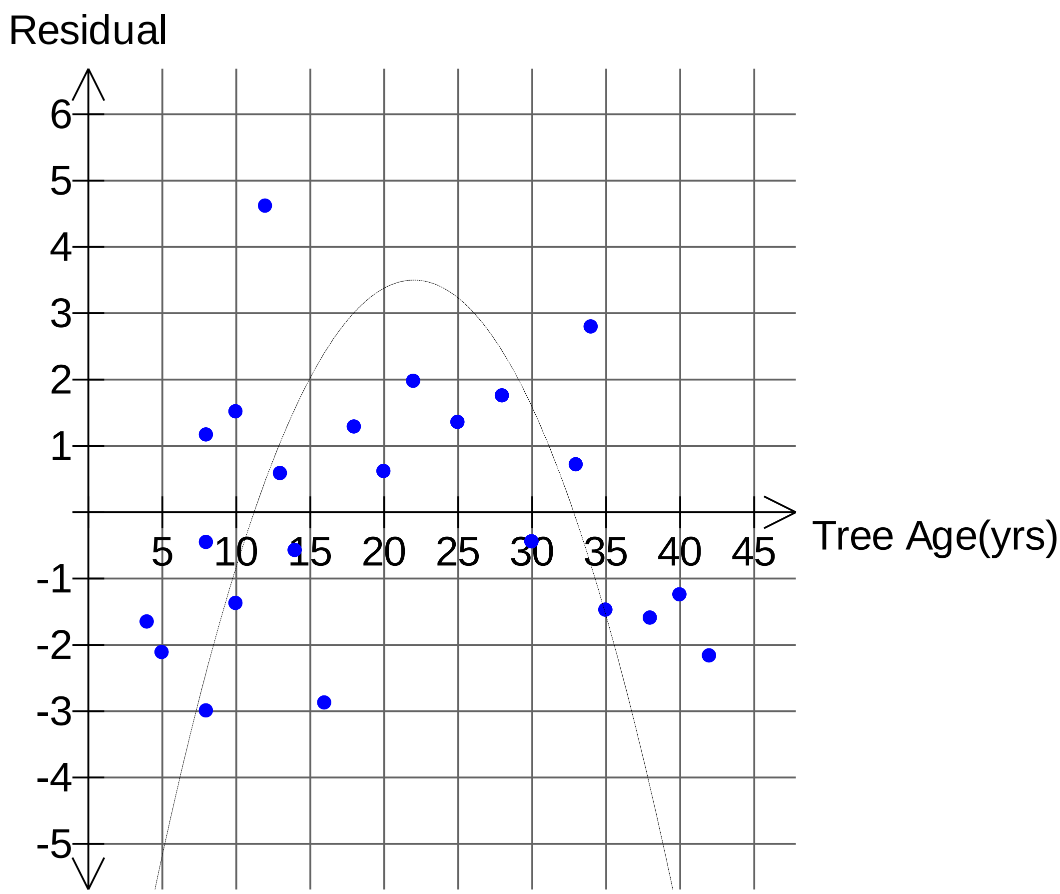
$$predicted = 1.847 + 0.461 \times 4 = 3.61$$

(4,2.03) $actual = 2.03(\,from\ table)$

$$residual = 2.03 - 3.691 - 1.66$$

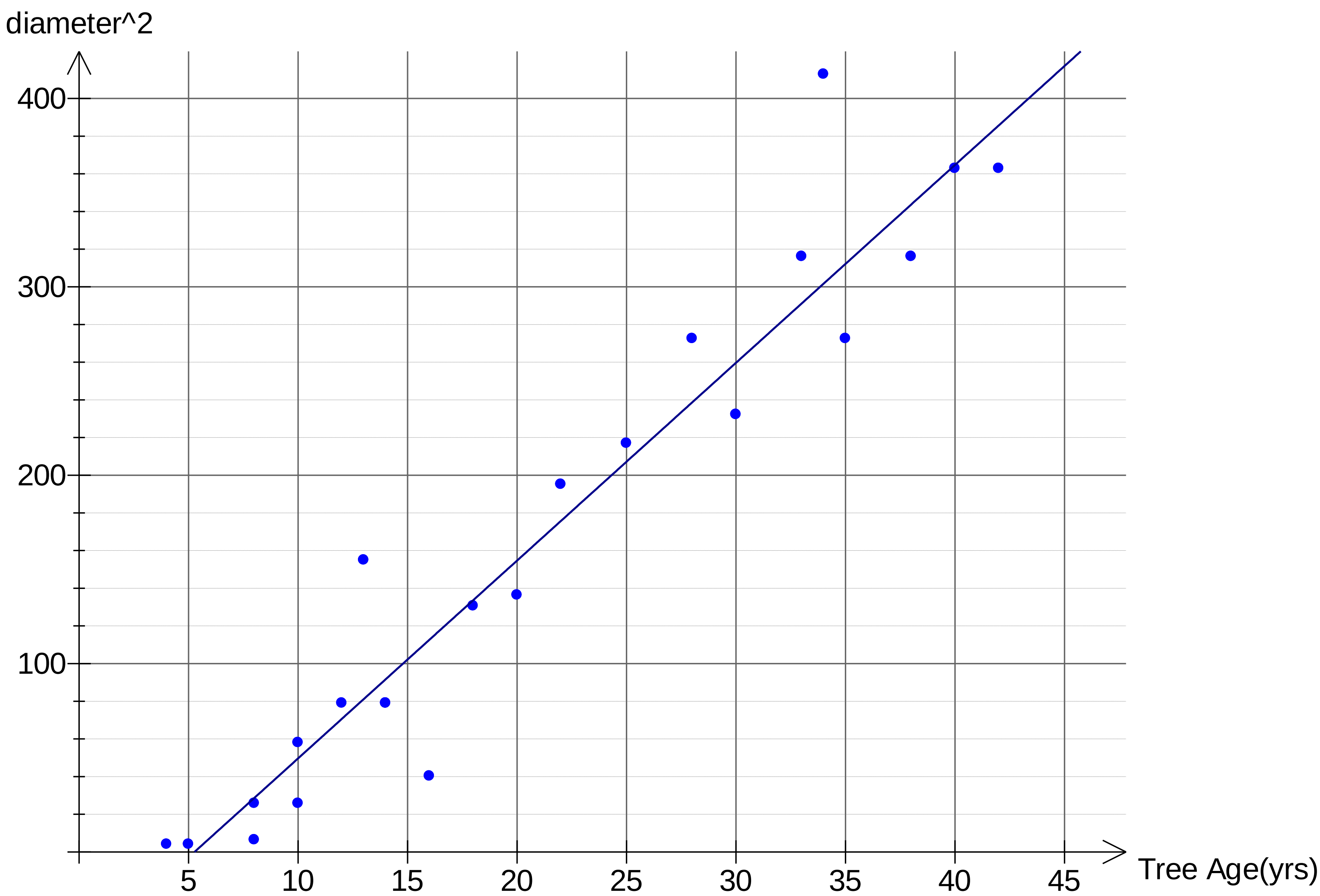
Remember the values on your CAS will be more accurate than ones produced using a rounded model (rule). The most important thing is that you understand what the residual represents.

14.



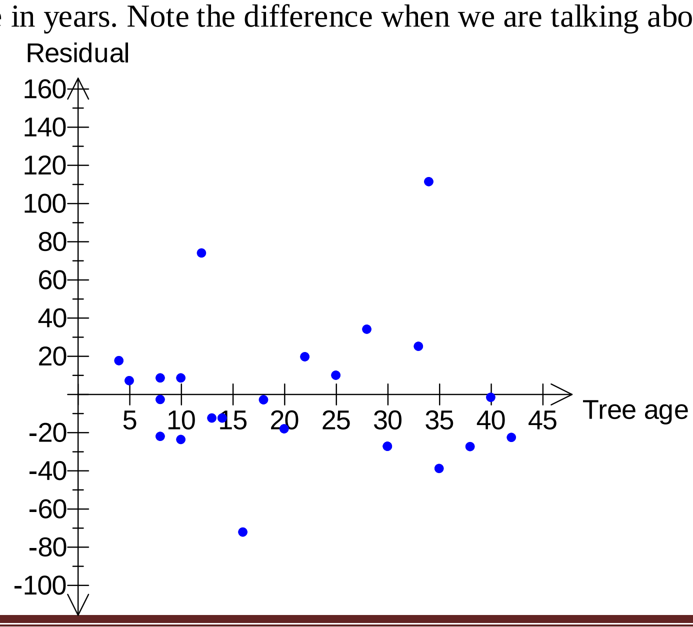
As can be seen above the residual shows a clear pattern, this indicates that the linear model is not appropriate because the original data was probably non-linear.

15. **TRANSFORMATIONS** y^2 **transformation (be prepared to explain the nature of this transformation – the y-values are stretched)**



$Diameter^2 = -55.344 + 10.501 \times age$ $r = 0.958$ $r^2 = 0.918$

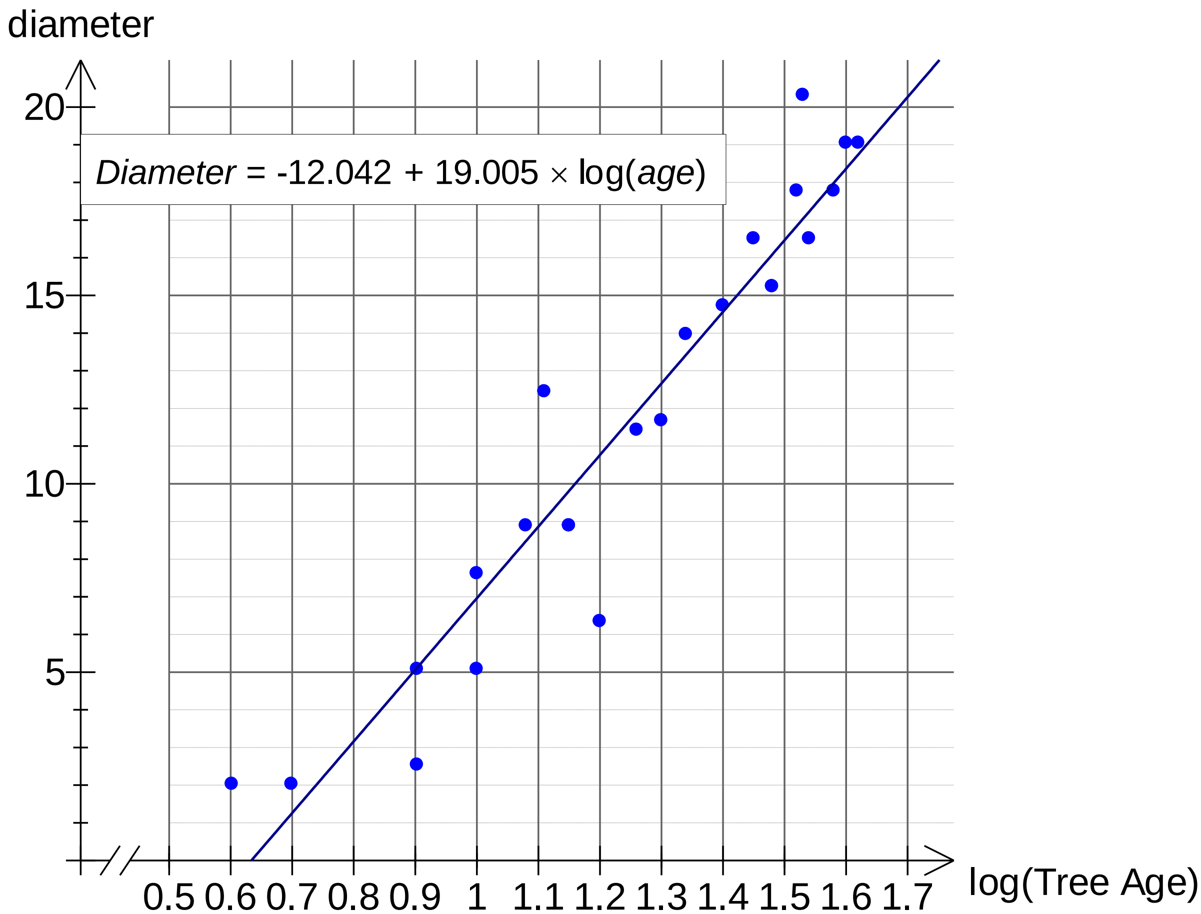
ie. 91.8% of the variation in the square of the diameter can be explained by the variation in tree age in years. Note the difference when we are talking about a data transformation.



Remember to label the axes with the correct terms.

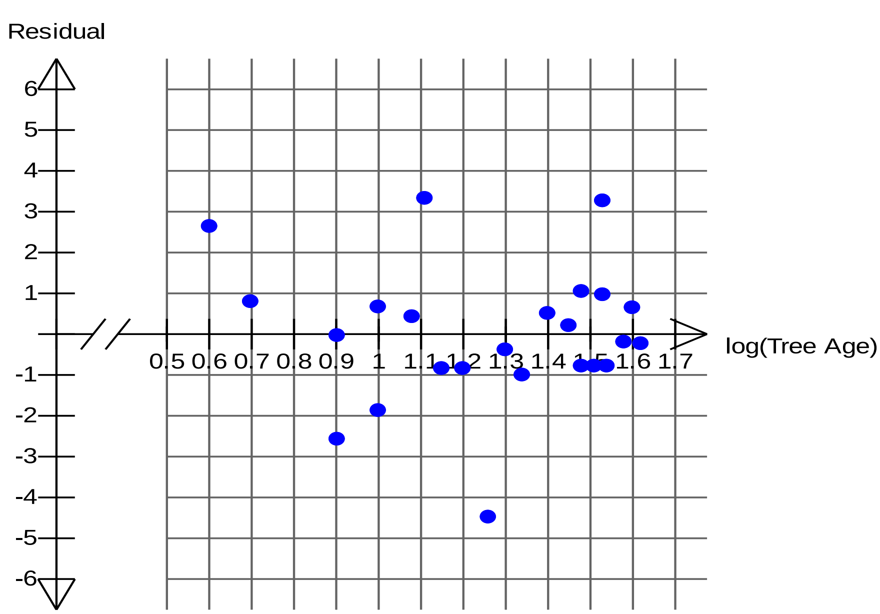
The residuals appear to have improved as there no longer seems to be a clear pattern. Both the correlation coefficient and the coefficient of determination have improved(closer to 100%).

TRANSFORMATIONS **log(*x*)**



$Diameter = -12.042 + 19.005 \times \log(age)$ $r = 0.955$ $r^2 = 0.912$

ie. 91.2% of the variation in diameter can be explained by the variation in the log of tree age.



The residuals show a slight pattern. Both the correlation coefficient and the coefficient of determination have also improved compared to the linear model.

Summary of transformation findings

	r	r^2	Residuals
Original	0.945	0.893	Pattern
Diameter square	0.958	0.918	Random
log(age)	0.955	0.912	Pattern

The best model is then age against diameter^2 as it has r^2 closest to 100% and the residual plot was randomly scattered indicating the data was probably linear.

$$Diameter^2 = -55.344 + 10.501 \times age$$

Don't forget the data transformation!
Not diameter but diameter squared

Most importantly make sure your answer makes SENSE!
If you forget the diameter squared for the prediction of 100 you get a tree diameter of 994 cm ..... metres in diameter – this would be an incredibly large tree.

16.

a. 27.5 years

$$Diameter^2 = -55.344 + 10.501 \times 27.5$$
$$diameter^2 = 233.434$$
$$diameter = \sqrt{233.434}$$
$$diameter = 15.28\,cm$$

b. 45 years

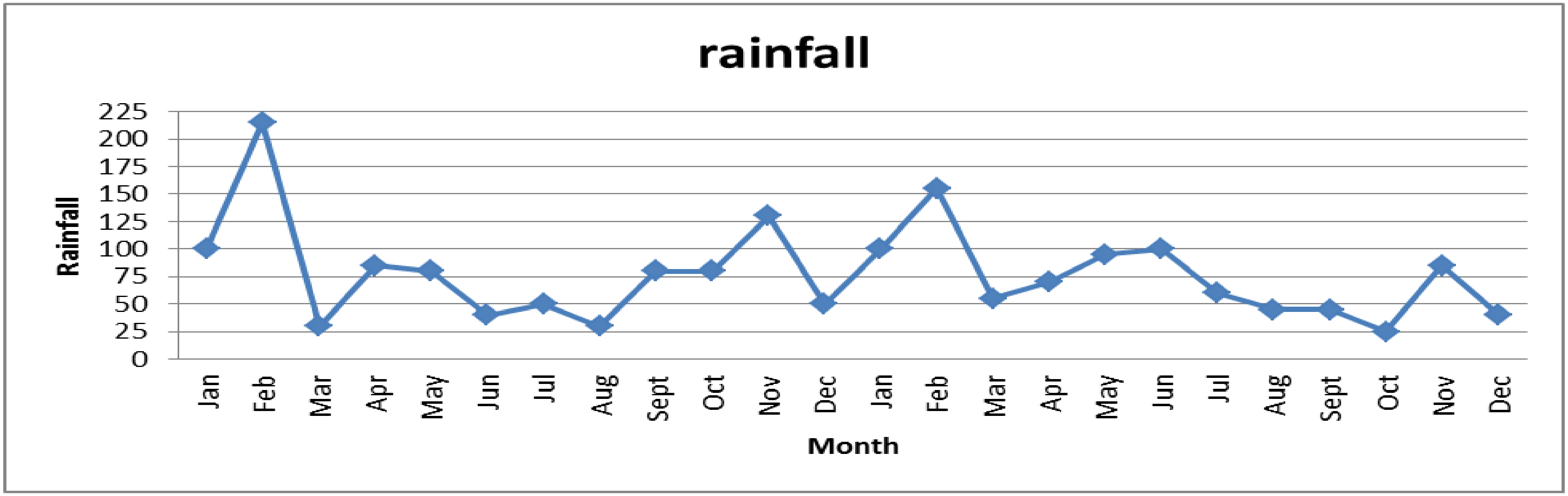
$$Diameter^2 = -55.344 + 10.501 \times 45$$
$$diameter^2 = 417.201$$
$$diameter = \sqrt{417.201}$$
$$diameter = 20.43\,cm$$

c. 100 years

$$Diameter^2 = -55.344 + 10.501 \times 100$$
$$diameter^2 = 994.756$$
$$diameter = \sqrt{994.756}$$
$$diameter = 31.54\,cm$$

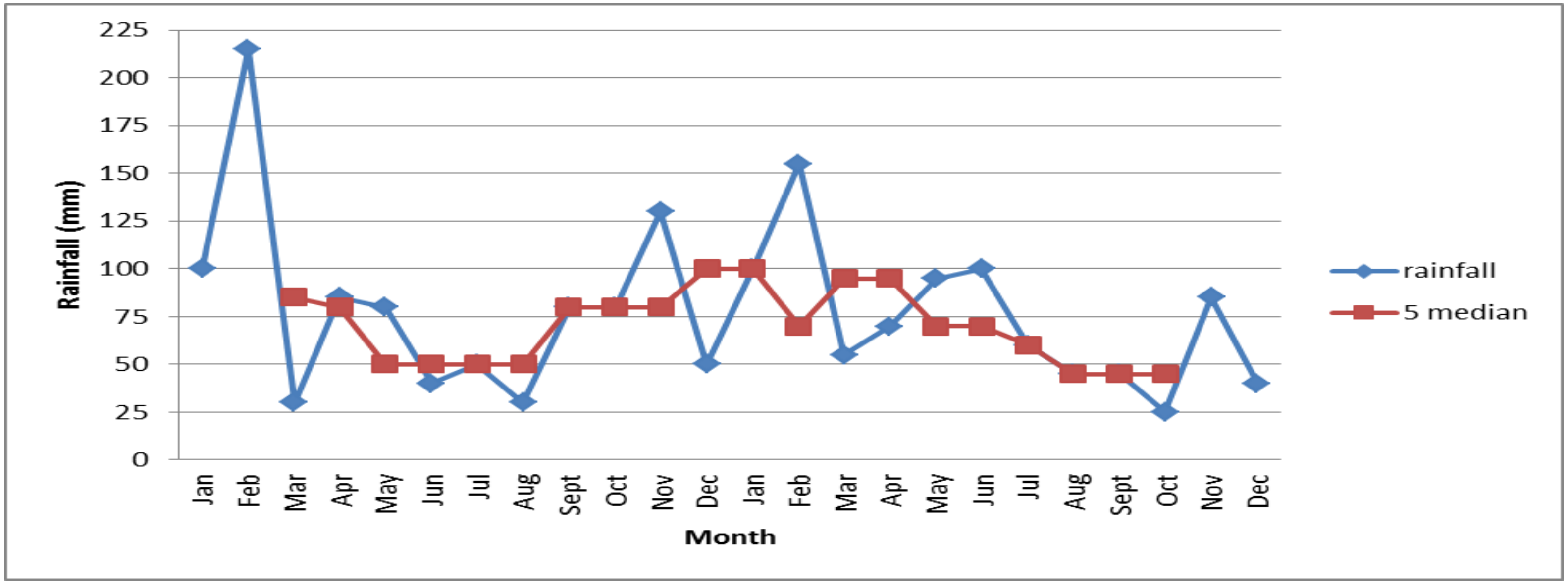
Given that the data set contains values between 4 and 42 years of age we would expect that the predictions for a. and possibly b.(45 years is just outside the data set) to be reasonably accurate. However the value of 100 years requires extrapolation beyond the known data set and would need to be treated with some uncertainty.

17.



18. Students should mention common peaks for Feb and Nov indicating the data has seasonality.

19.



20. A seasonal index of 0.84, tells us:

On average, February rainfall is 16% below the monthly average.

21. $\textbf{\textit{Deseasonalised Rainfall}} = \mathbf{91.193} - \mathbf{0.464} \times \textbf{\textit{month}}$

where $\textbf{\textit{month}} = \mathbf{1}\ \textbf{\textit{is January}} - \textbf{\textit{Year}}\ \mathbf{1}$

February Year 3 = 26

Predicted Deseasonalised Rainfall = 91.193 – 0.464 × 26 = 79.129
Predicted Actual Rainfall = 79.129 × 0.84 = 66.4684
Predicted rainfall for February Year 3 is 66mm

22. July SI = 1.20

0.2/1.2*100 = 16.7%

Actual rainfall must be adjusted (decreased) by approximately 16.7% to allow for seasonality.