

Statistics 101 – Homework 5

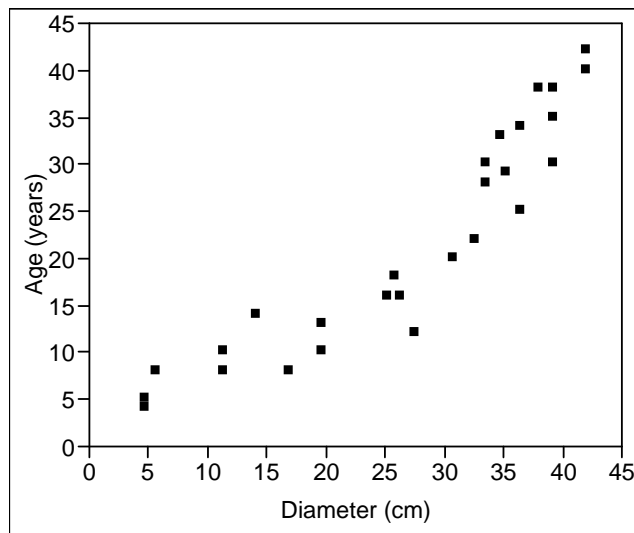
Solution

Assignment:

1. How old is a tree? Cutting a tree down and counting the number of growth rings can determine the age of a tree. It would be nice to have a non-destructive method for determining the age of a tree. A forester measured 27 trees of the same species that had been cut down and determined the age of each tree by counting the growth rings. Below are the data.

Diameter, cm	Age, years	Diameter, cm	Age, years	Diameter, cm	Age, years
4.6	4	14.0	14	33.5	30
4.6	5	25.1	16	39.1	30
5.6	8	25.7	18	34.7	33
11.2	8	30.7	20	36.3	34
16.8	8	32.5	22	39.1	35
11.2	10	26.2	16	37.9	38
19.6	10	36.3	25	39.1	38
27.4	12	33.5	28	41.9	40
19.6	13	35.1	29	41.9	42

- a) Plot the data. Use Age as your Y, Response and Diameter as your X, Factor. Describe the association between diameter and age. Be sure to include form, direction, strength and comment on any potential outliers.



The form of the relationship is basically linear but with a slight curve upward. The direction is positive. The strength is fairly strong. There are no potential outliers.

- b) Find the correlation between age and diameter. What does this indicate about the strength of the linear association?

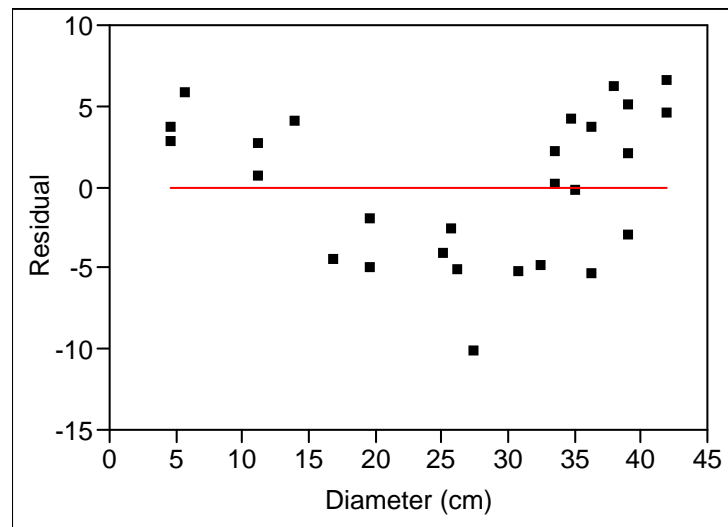
The correlation is the square root of R^2 with the same sign as the direction of the relationship. $r = \sqrt{R^2} = \sqrt{0.855655} = +0.925$. This indicates a fairly strong positive linear association between age and diameter.

- c) Fit the least squares regression line relating age to diameter. Report the prediction equation and give an interpretation of the estimated slope coefficient.

$$\text{Predicted Age} = -2.872 + 0.9175 * \text{Diameter}$$

For each additional 1 cm of diameter, the age of the tree increases, 0.9175 years, on average.

- d) Describe the plot of residuals. Explain why a linear model is not an adequate model for the relationship between age and diameter.



The residuals are above zero for small diameters (< 15 cm), below zero for medium diameters (15 to 35 cm) and primarily above zero for large diameters (> 35 cm). This under, over, under prediction is an indication of a curved relationship and that the linear model is not adequate. We could do better with a model that accounts for the curvature.

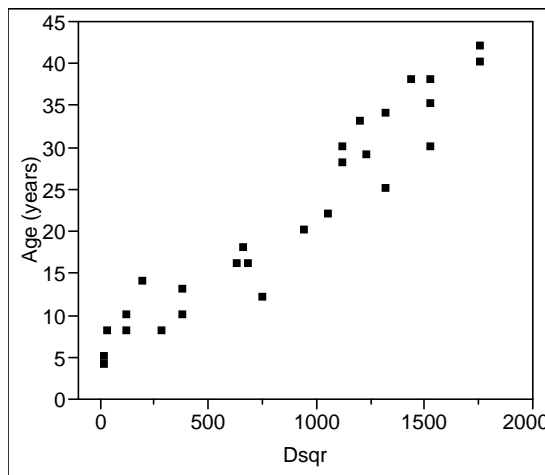
- e) If you used the linear model, would you generally over-predict or under-predict the age of trees with large diameters, say greater than 45 cm? Explain Briefly

This linear model would tend to under-predict the age of trees with large diameters. Virtually all of the ages for diameters greater than 35 cm are under-predicted, so for larger diameters one would expect under-predicted ages.

2. Cross-sectional area might be a better predictor of age. Cross-sectional area is proportional to the square of the diameter. Re-express the data by squaring the diameter, i.e. create a new variable $Dsqr=(Diameter)^2$.

a) Plot Age as the Y, Response and Dsqr as the X, Factor. Does this scatter plot look better than the scatter plot in 1? Explain briefly.

The scatter plot appears to have less curvature, i.e. it appears more linear.



b) Fit a least squares regression line relating Age and Dsqr. Report the prediction equation and give an interpretation of the estimated slope coefficient.

$$\text{Predicted Age} = 4.650 + 0.019892 * \text{Diameter}^2$$

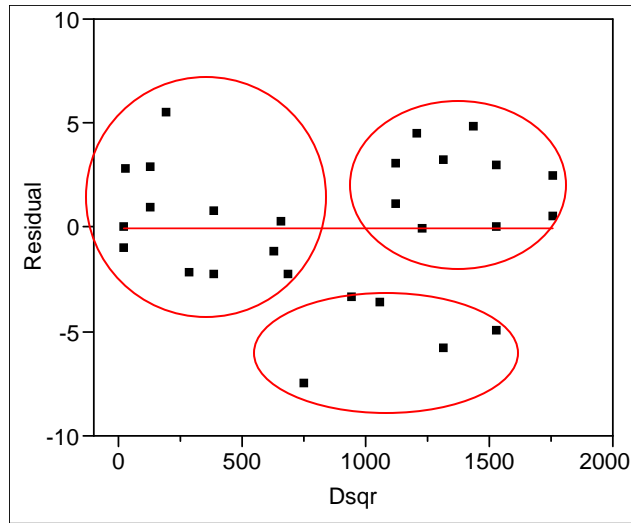
For each additional 1 cm² of Diameter², the age of the tree increases, 0.019892 years, on average.

c) Use the equation in b) to predict the age of a tree that has a 45 cm diameter.

$$\text{Predicted Age} = 4.650 + 0.019892 * (45)^2 = 4.650 + 40.28 = 44.93 \text{ years}$$

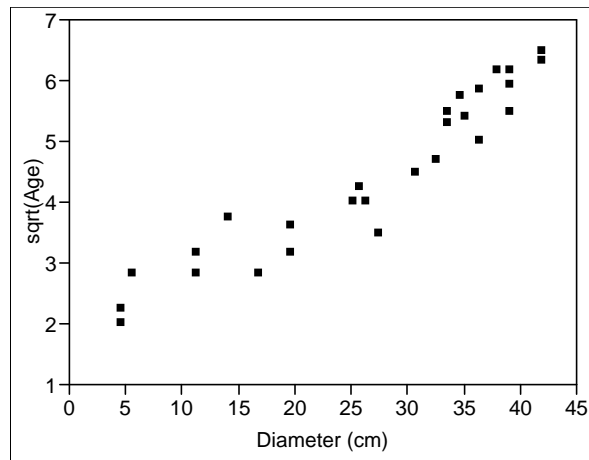
d) Describe the plot of residuals. Is this a more appropriate model than the one you fit in 1? Explain briefly.

The plot of residuals shows residuals above and below the zero line for all values of Diameter². There is not a strong pattern, although there may be some grouping of residuals, and so this is a more appropriate model than the simple linear model in 1.



3. Instead of re-expressing the diameter, re-express age by taking the square root of age, $\text{sqrt}(\text{Age})$.
- a) Plot $\text{sqrt}(\text{Age})$ as the Y, Response and Diameter as the X, Factor. Does this scatter plot look better than the one in 1? Explain briefly.

The plot looks a little better in that it is more linear than the plot in 1. There is still a little bit of curvature.



- b) Fit a least squares regression line relating $\text{sqrt}(\text{Age})$ and Diameter. Report the prediction equation and give an interpretation of the estimated slope coefficient.

$$\text{Predicted } \text{sqrt}(\text{Age}) = 1.5935 + 0.107155 * \text{Diameter}$$

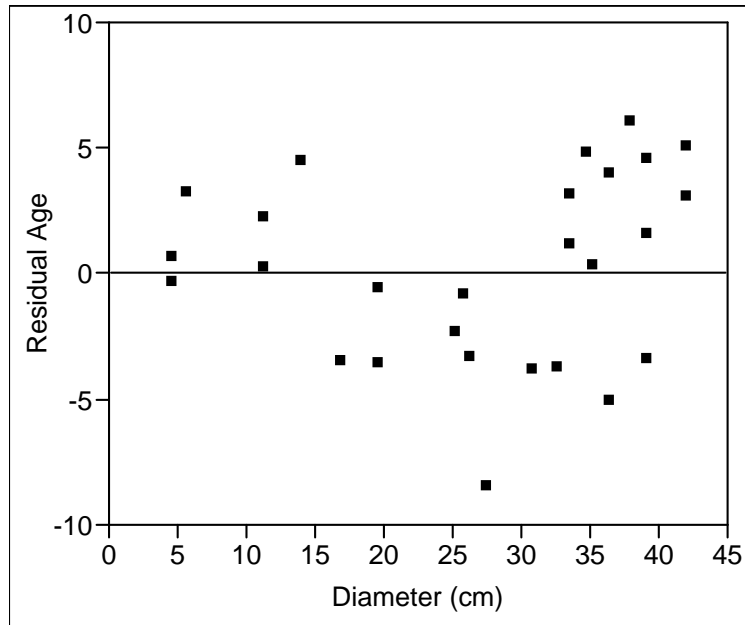
For each additional 1 cm of diameter the $\text{sqrt}(\text{Age})$ increases, on average, by 0.1070155 years.

c) Use the equation in b) to predict the age of a tree that has a 45 cm diameter.

$$\text{Predicted sqrt(Age)} = 1.5935 + 0.107155*(45) = 6.415475$$

$$\text{Predicted Age} = (6.415475)^2 = 41.16 \text{ years}$$

d) Below is a plot of residual Age (using the equation in b)) versus Diameter. Describe the plot of residuals. Is this a more appropriate model than the one you fit in 1? Is this a more appropriate model than the one you fit in 2? Explain briefly.



For small diameters (< 15 cm), the residual age tends to be positive for medium diameters (15 to 35 cm) the residual age tends to be negative and for larger diameters (> 35 cm) the residual age tends to be positive. This curvature is not as pronounced as with the simple linear model in 1, but more pronounced than with the model in 2. This is a more appropriate fit than in 1, but may not be as appropriate as the fit in 2.