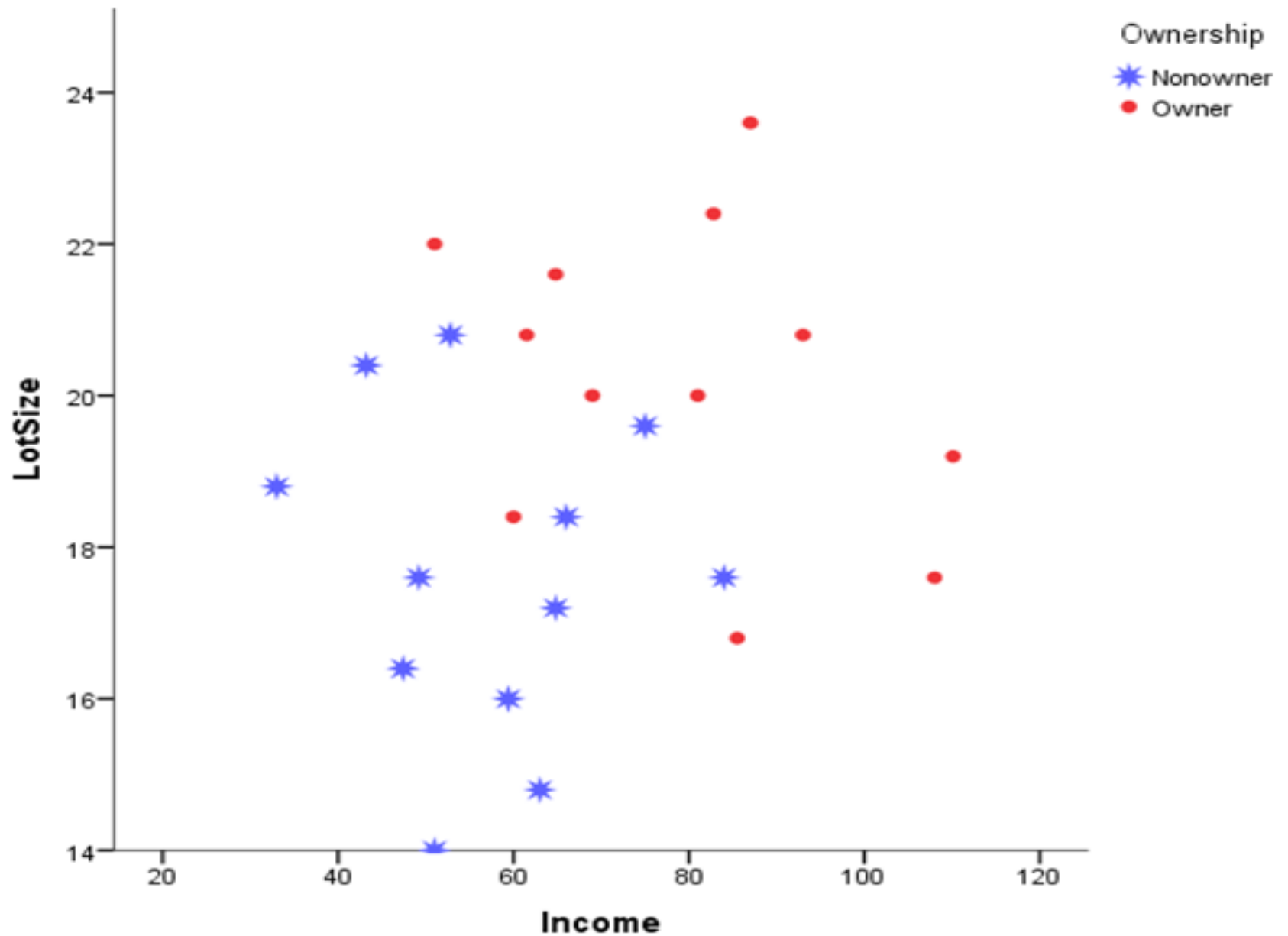
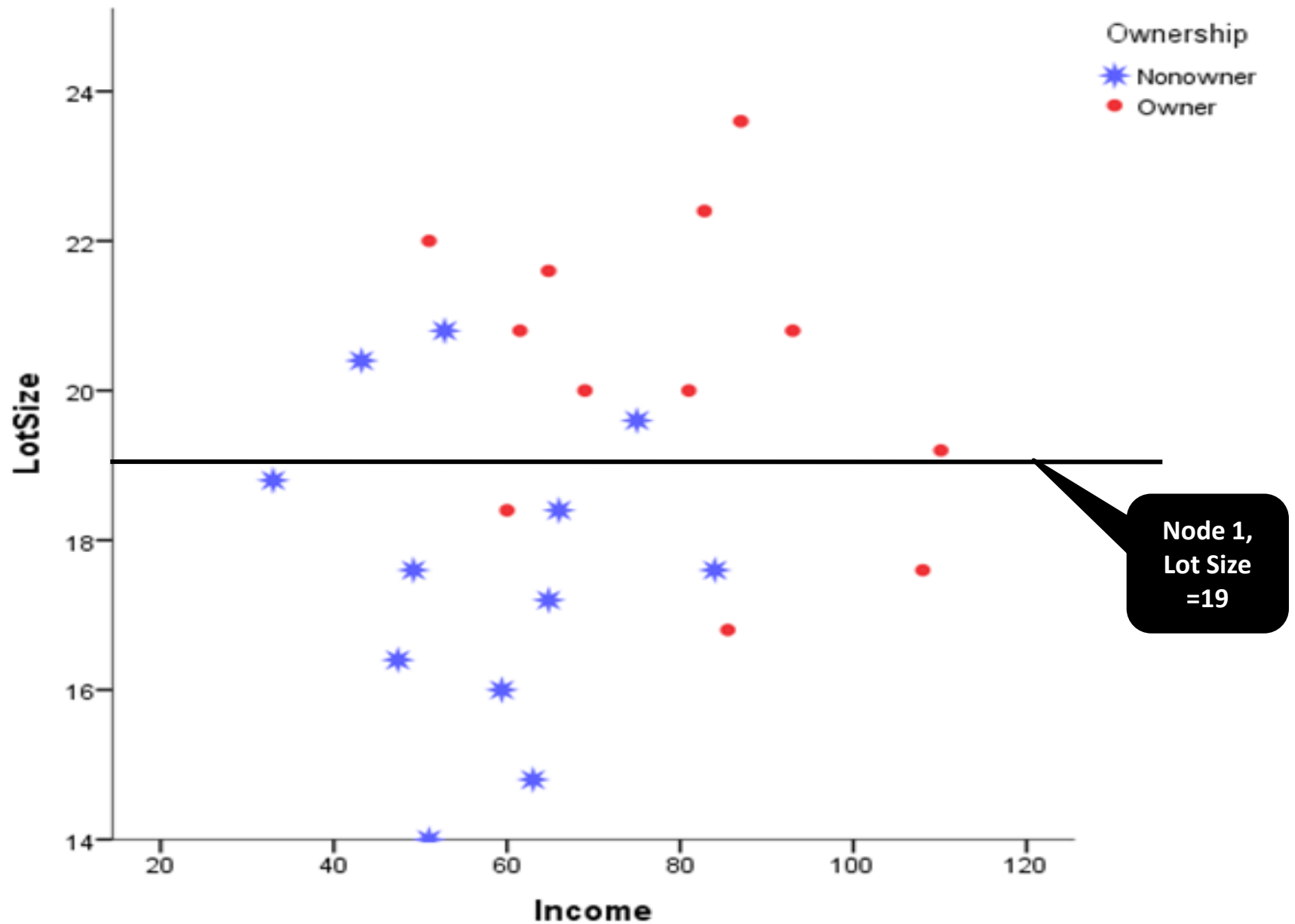


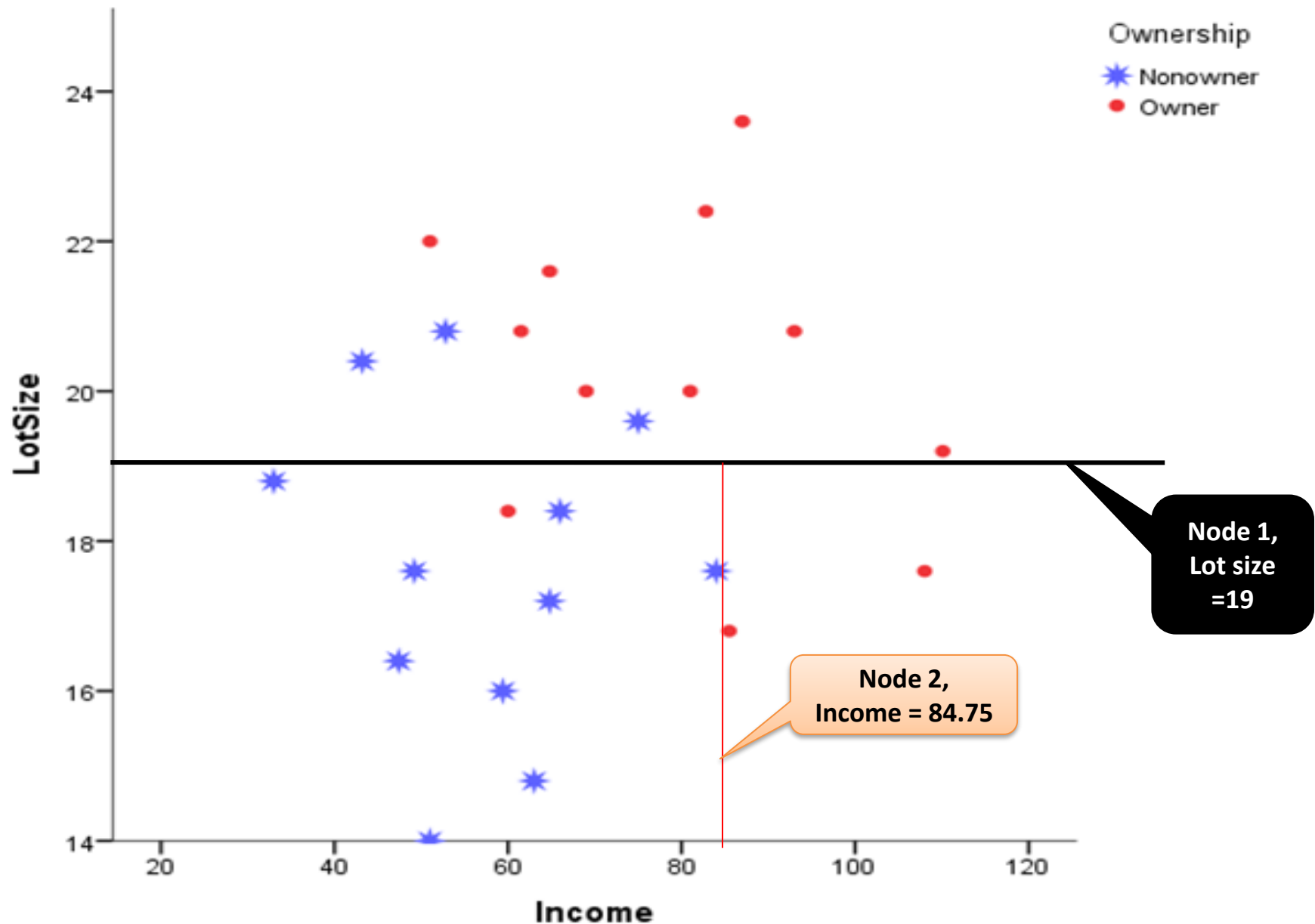
Decision Trees

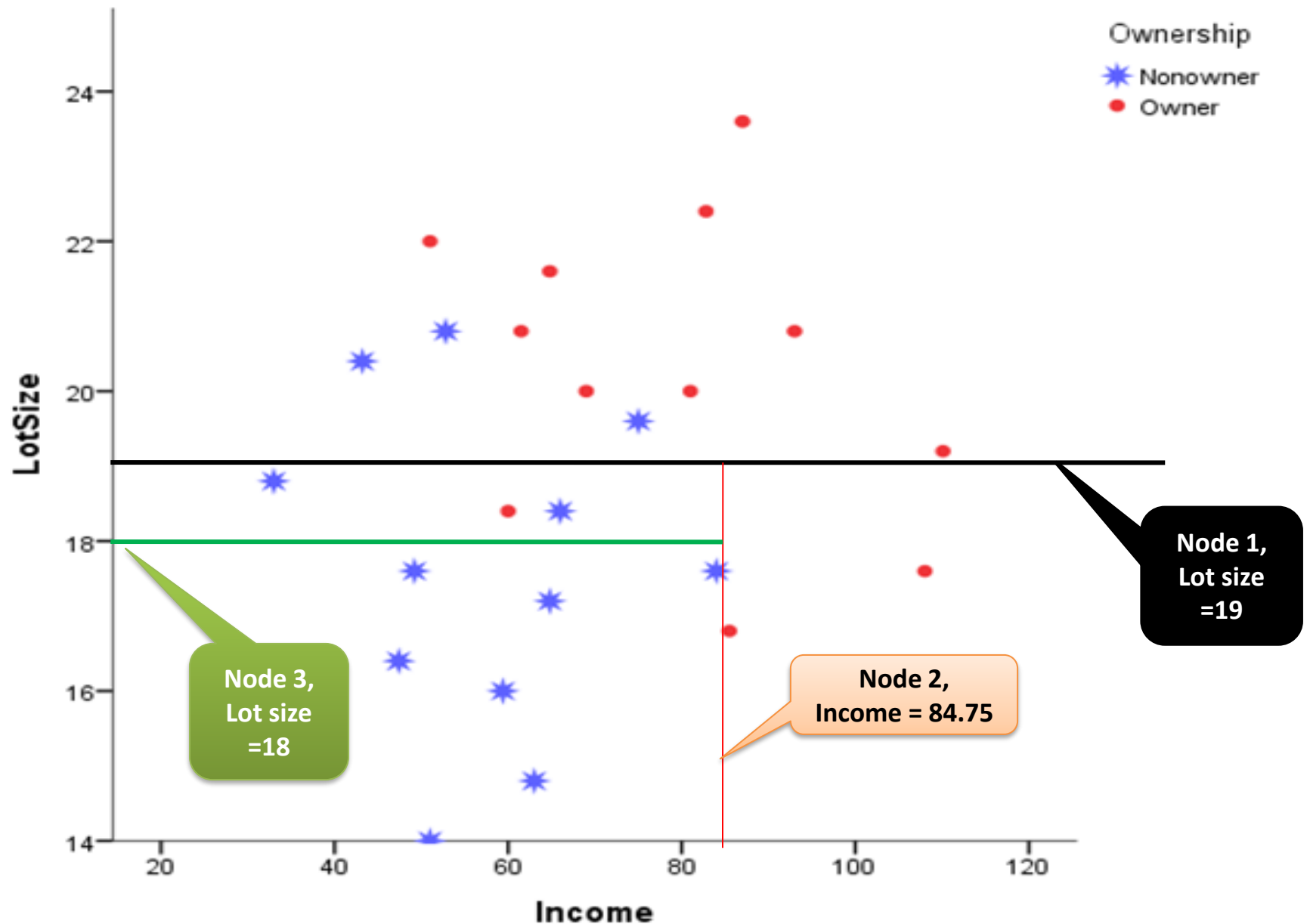
Basic Concepts

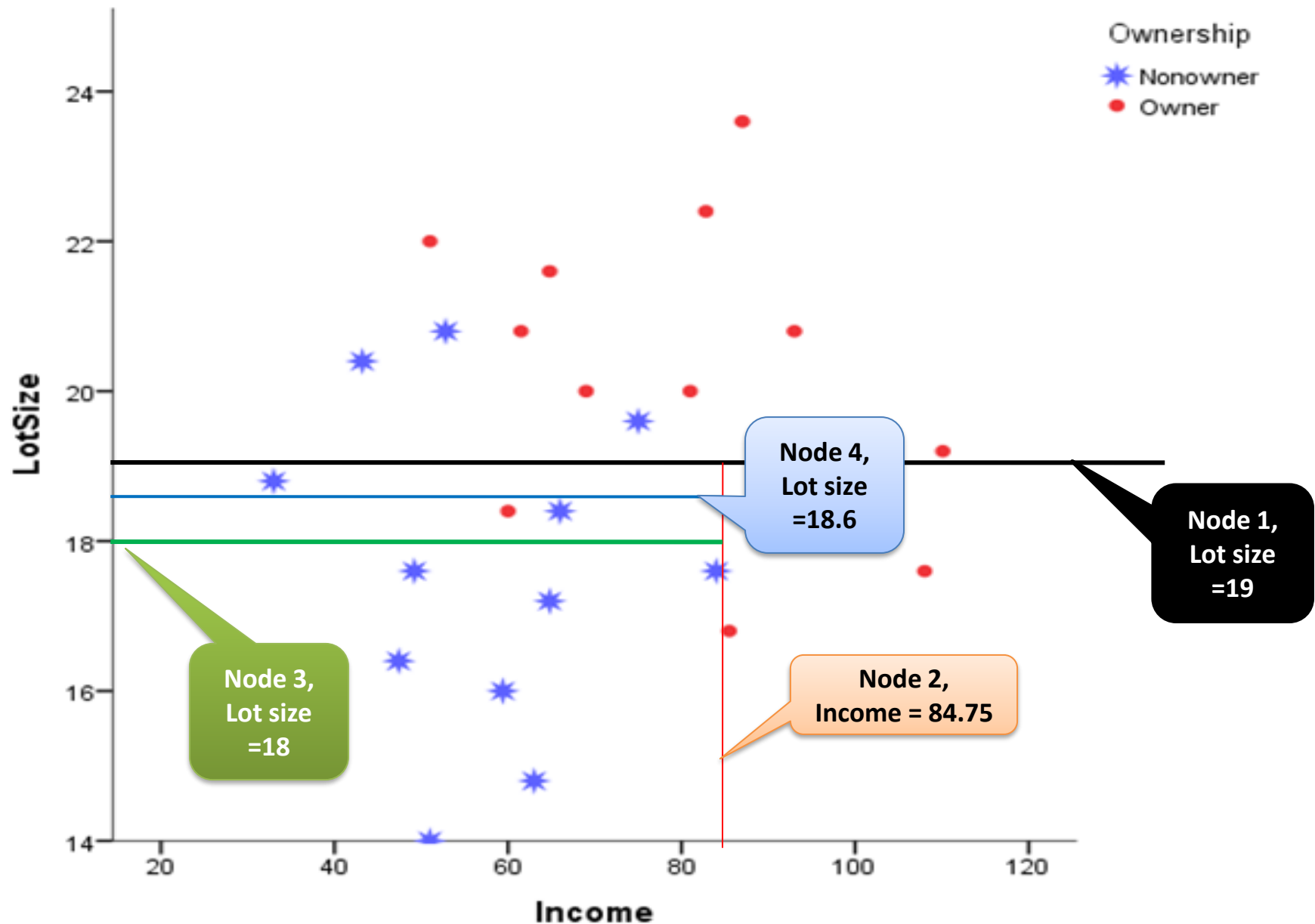
Household Number	Income	Lot Size	Ownership
1	60	18.4	Owner
2	85.5	16.8	Owner
3	64.8	21.6	Owner
4	61.5	20.8	Owner
5	87	23.6	Owner
6	110.1	19.2	Owner
7	108	17.6	Owner
8	82.8	22.4	Owner
9	69	20	Owner
10	93	20.8	Owner
11	51	22	Owner
12	81	20	Owner
13	75	19.6	Nonowner
14	52.8	20.8	Nonowner
15	64.8	17.2	Nonowner
16	43.2	20.4	Nonowner
17	84	17.6	Nonowner
18	49.2	17.6	Nonowner
19	59.4	16	Nonowner
20	66	18.4	Nonowner
21	47.4	16.4	Nonowner
22	33	18.8	Nonowner
23	51	14	Nonowner
24	63	14.8	Nonowner

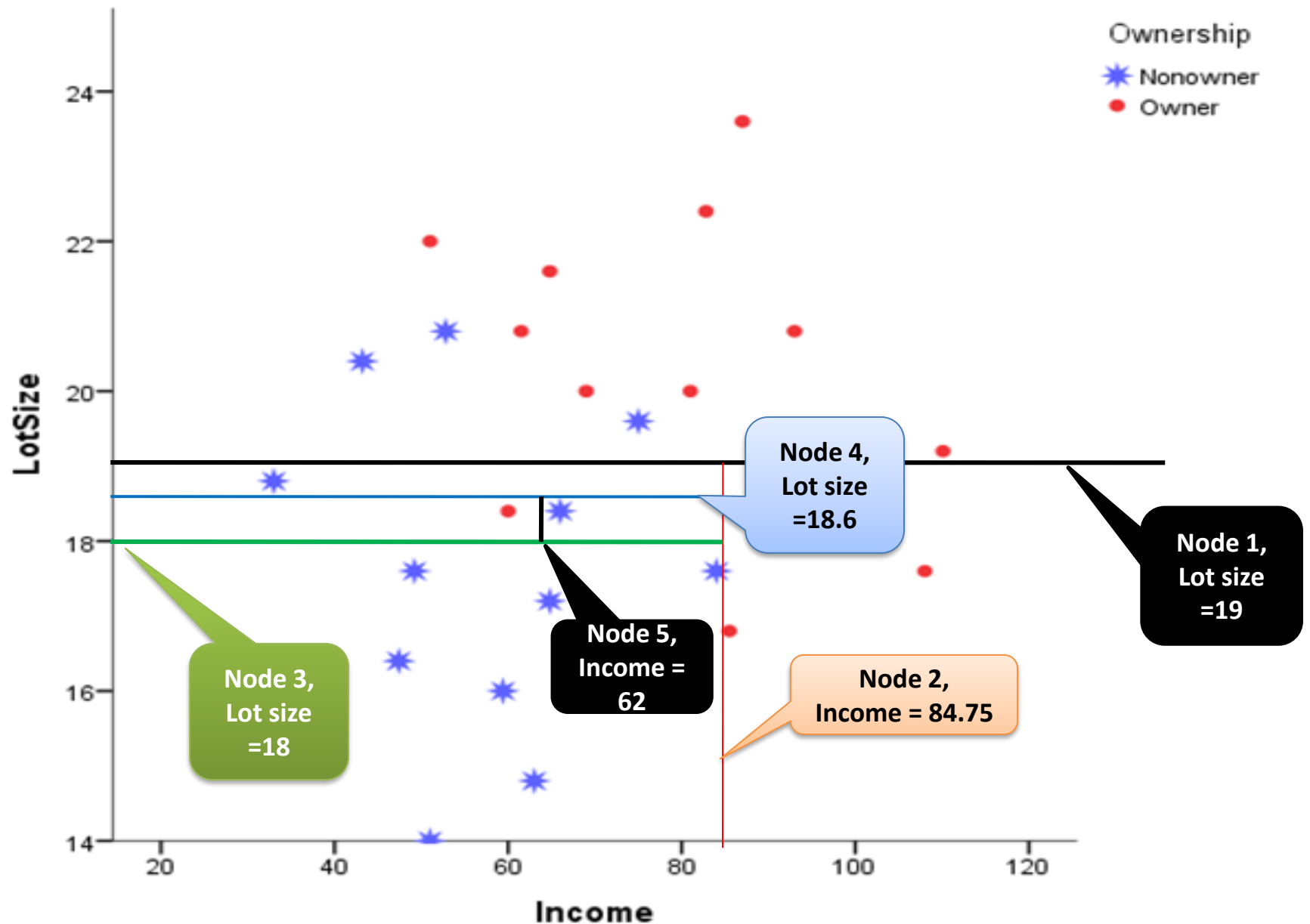


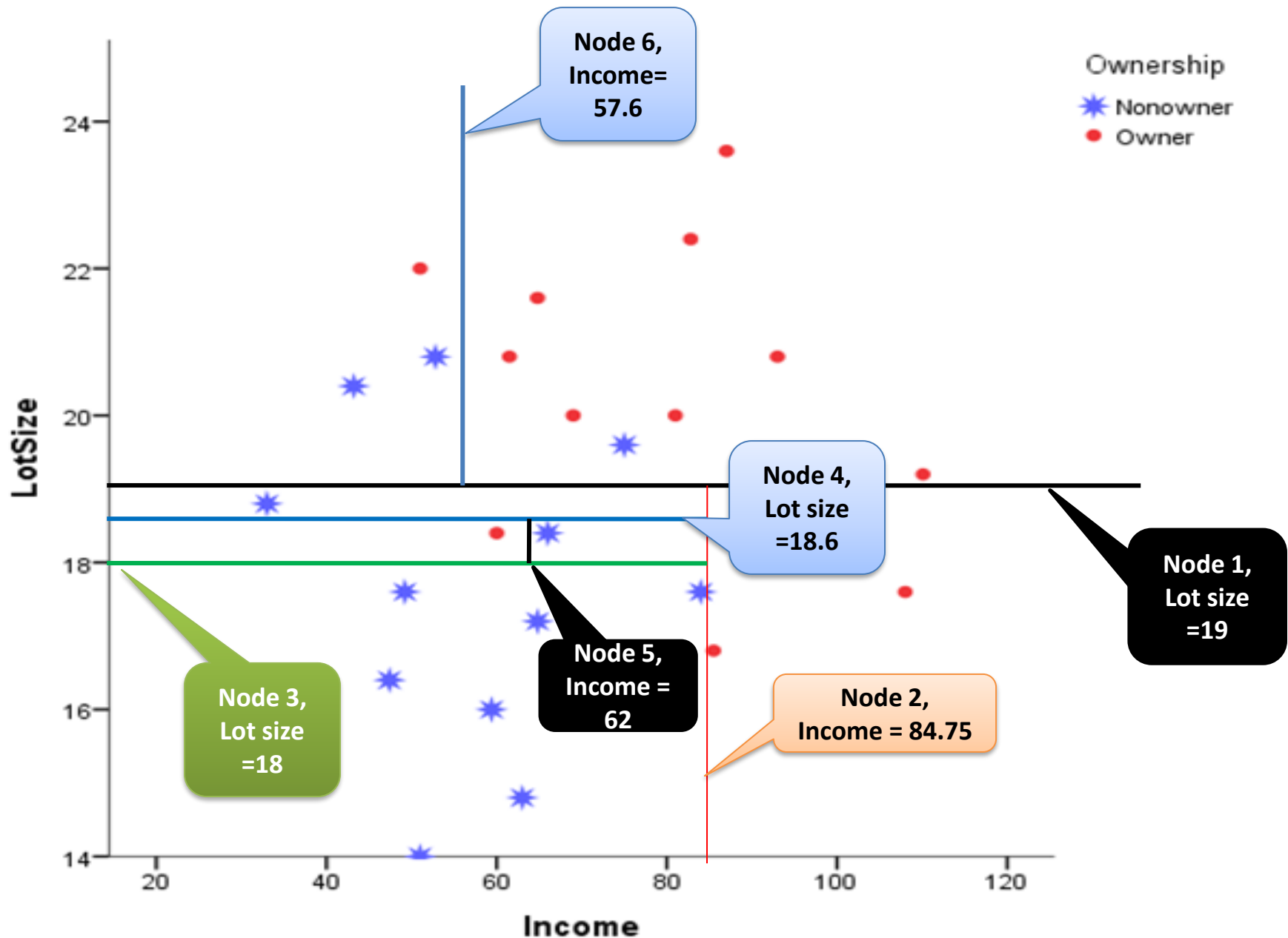


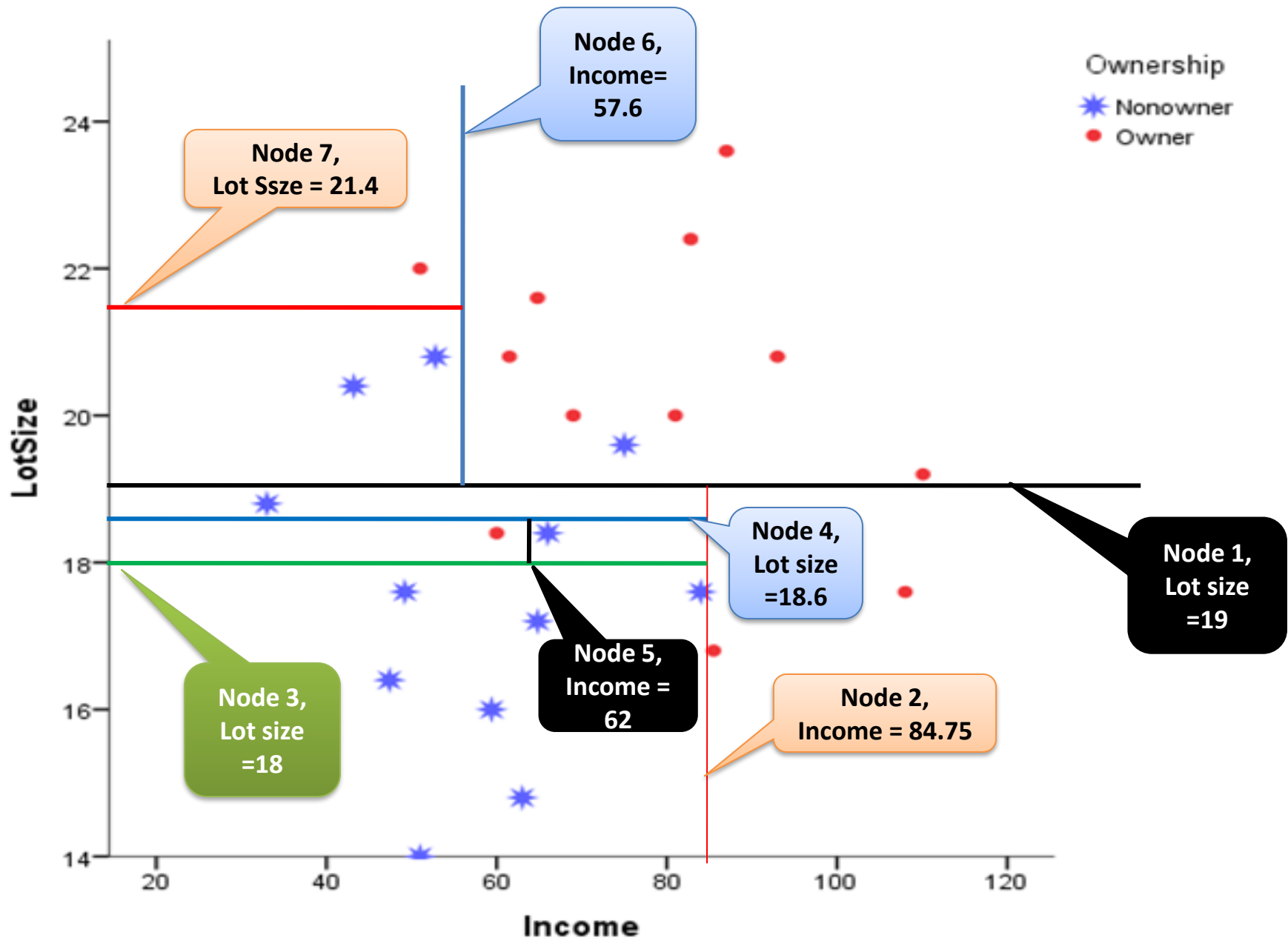


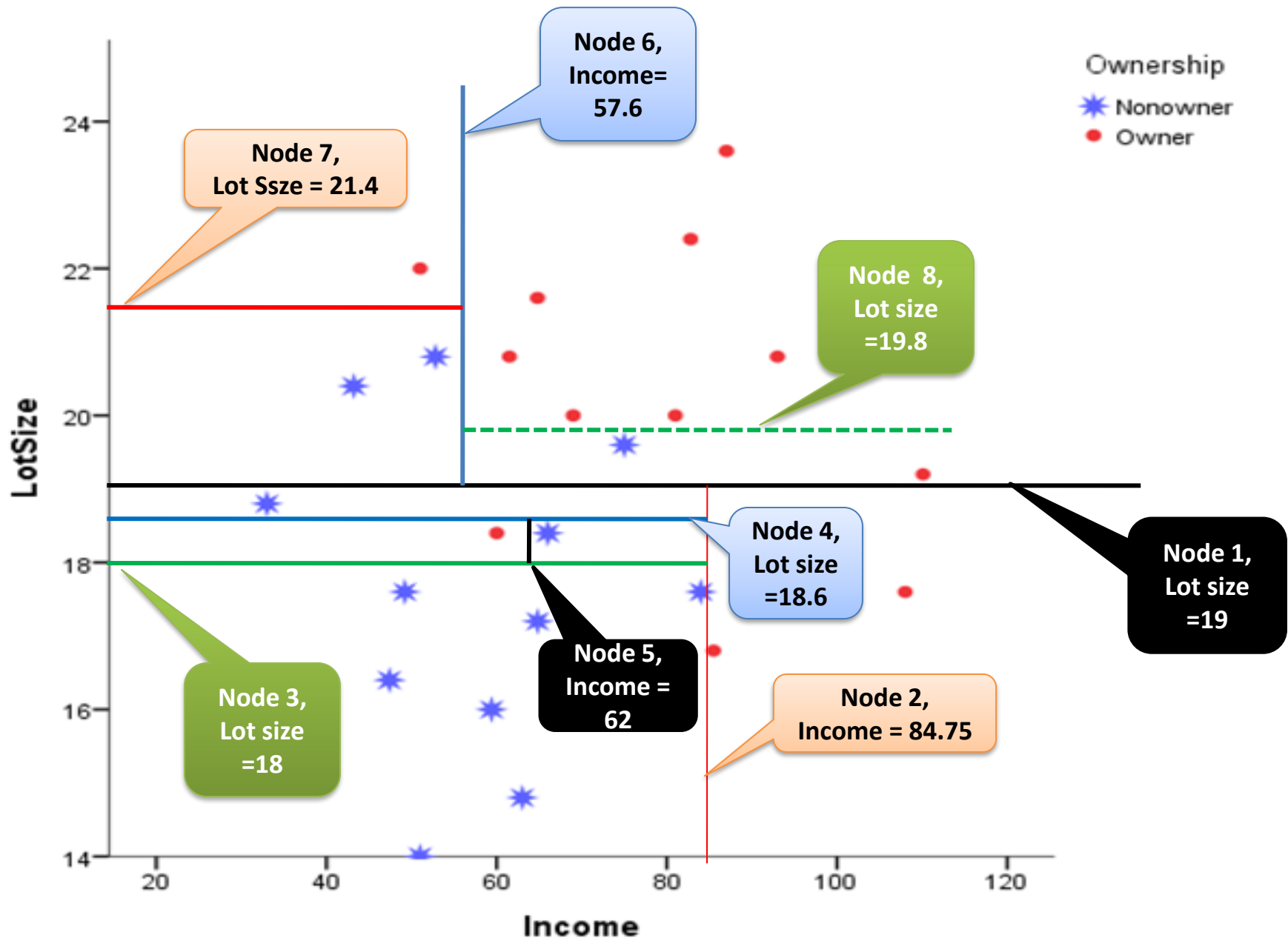


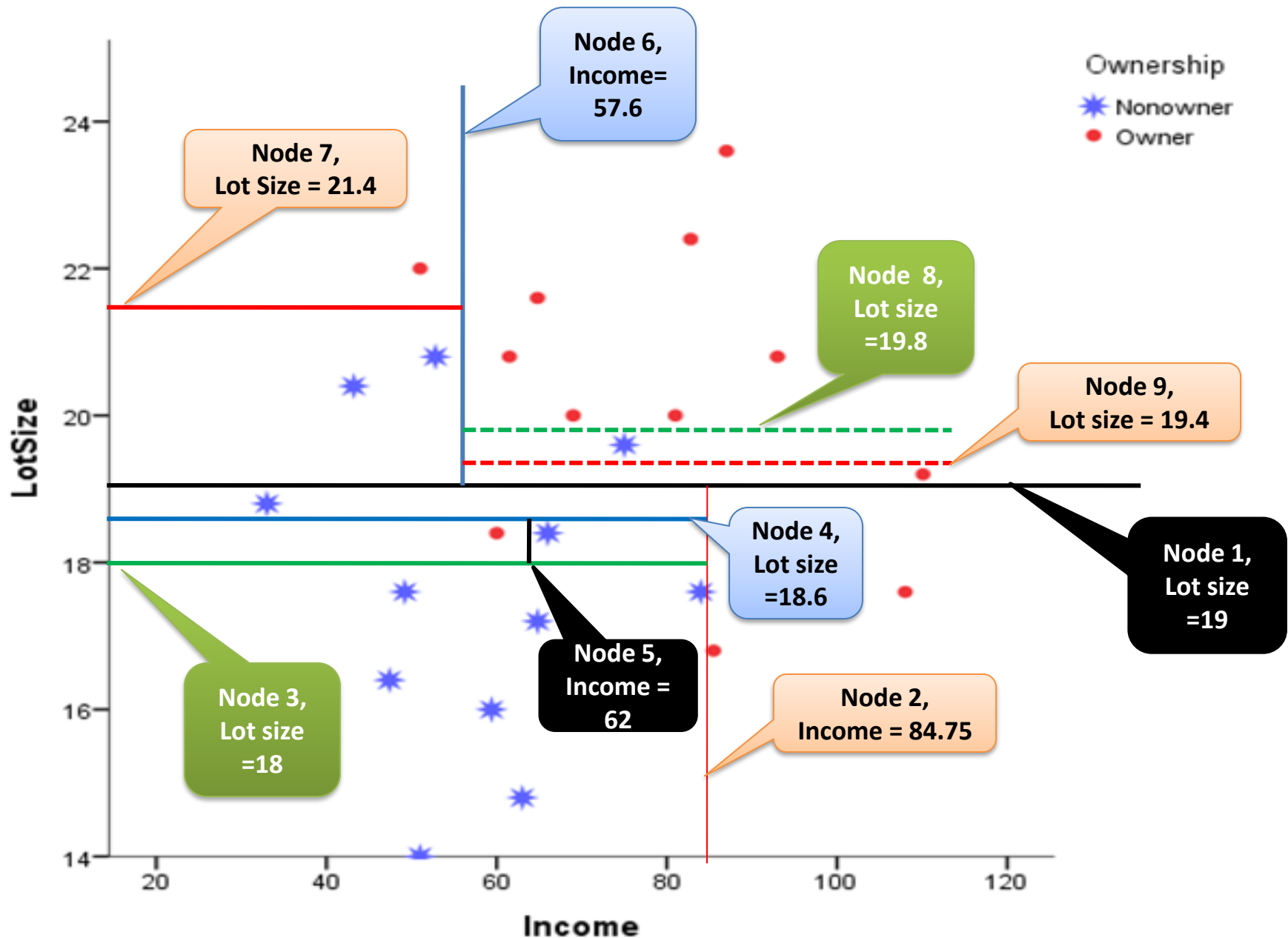


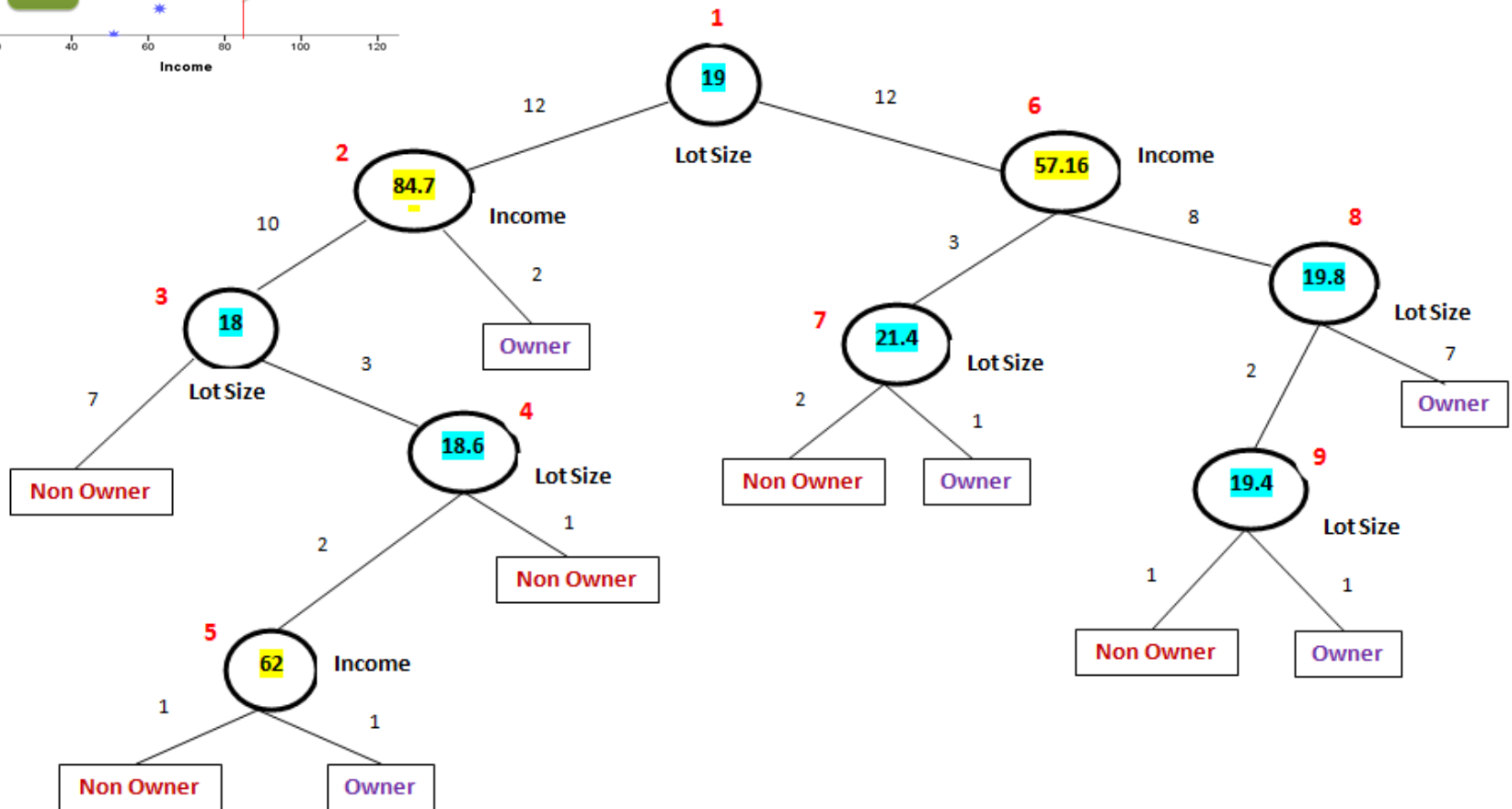
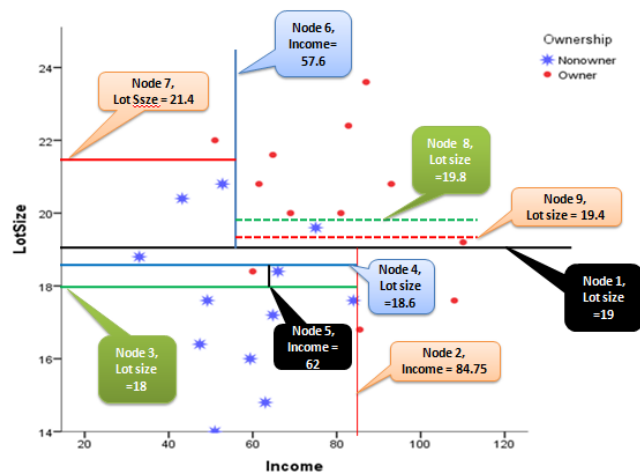


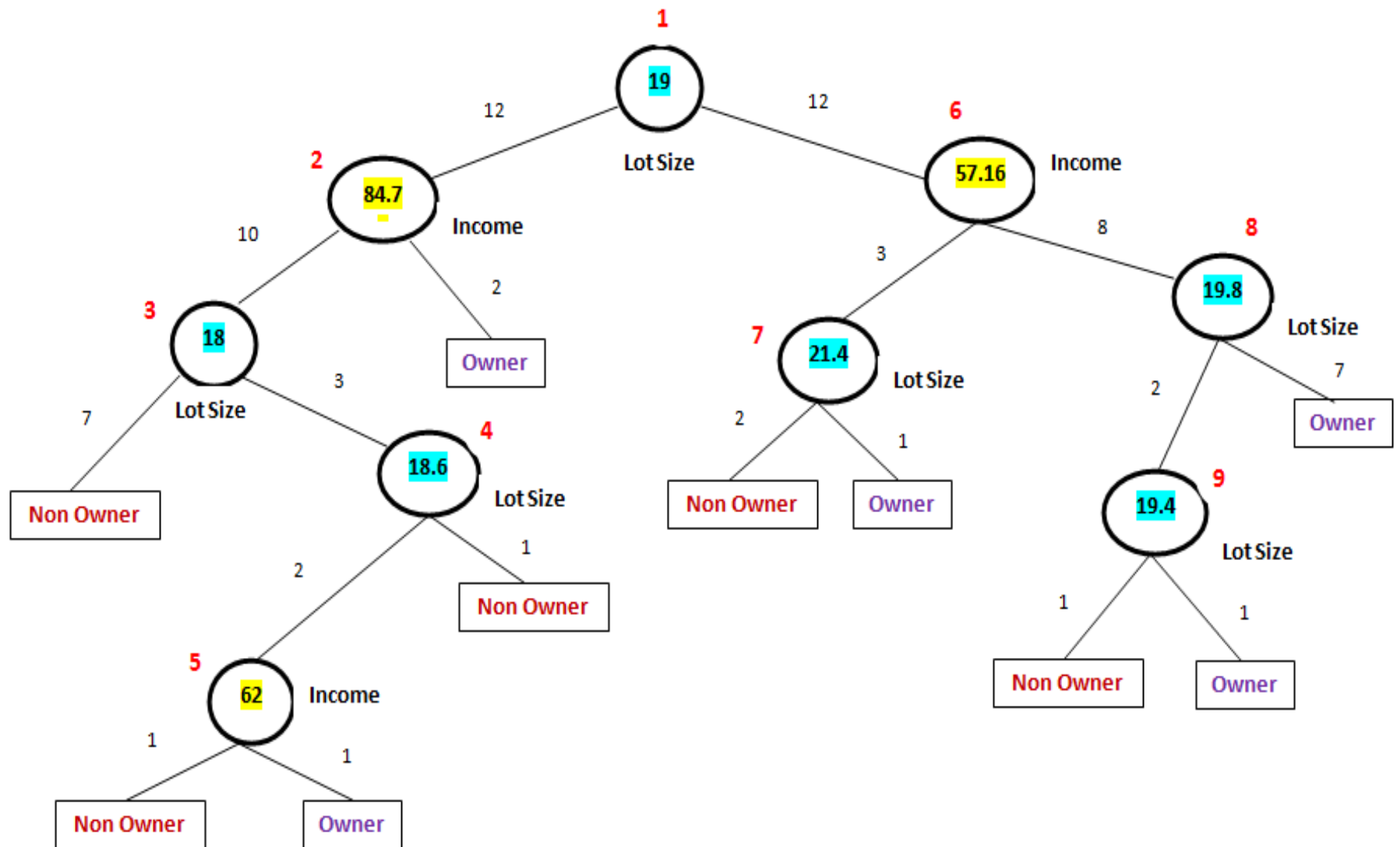












Impurity Gini Index

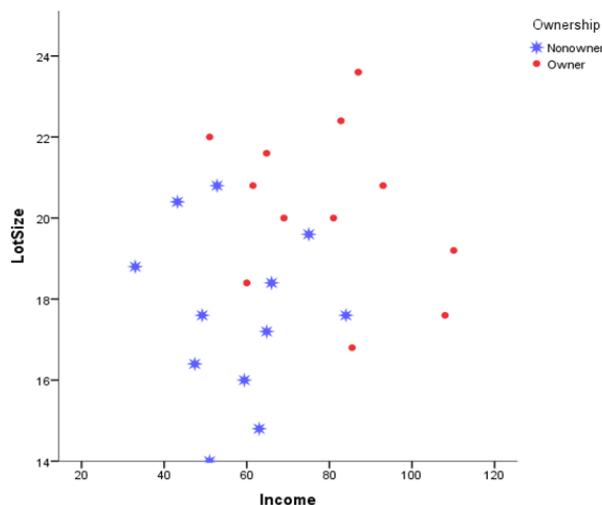
$$\text{Gini Index for rectangle } A, I(A) = 1 - \sum_{k=1}^m p_k^2$$

where, k is particular class (owner, non – owner); m is number of classes

$$\text{Gini Index for rectangle } A, I(A) = 1 - \left\{ \left(\frac{12}{24} \right)^2 + \left(\frac{12}{24} \right)^2 \right\}$$

$$\text{Gini Index for rectangle } A, I(A) = 1 - \{(0.5)^2 + (0.5)^2\}$$

$$\text{Gini Index for rectangle } A, I(A) = 1 - \{0.25 + 0.25\} = 0.50$$



Gini Index before split

For Upper Rectangle, there were 9 owners and 3 non-owners

$$\text{Gini Index for rectangle A, } I(A) = 1 - \left\{ \left(\frac{9}{12} \right)^2 + \left(\frac{3}{12} \right)^2 \right\}$$

$$\text{Gini Index for rectangle A, } I(A) = 1 - \{(0.75)^2 + (0.25)^2\}$$

$$\text{Gini Index for rectangle A, } I(A) = 1 - \{0.5625 + 0.0625\} = \mathbf{0.375}$$

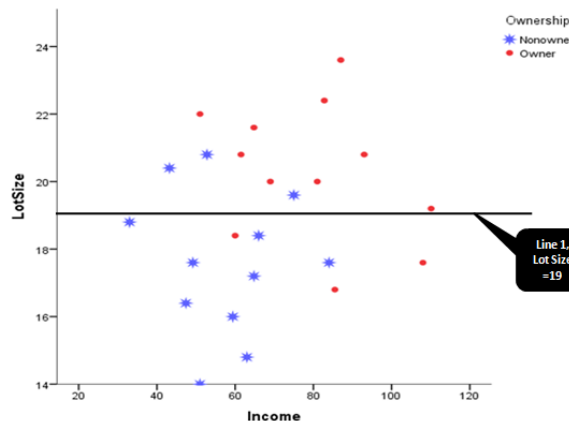
**Total
Impurity**

For Lower Rectangle, there were 3 owners and 9 non-owners

$$\text{Gini Index for rectangle A, } I(A) = 1 - \left\{ \left(\frac{3}{12} \right)^2 + \left(\frac{9}{12} \right)^2 \right\}$$

$$\text{Gini Index for rectangle A, } I(A) = 1 - \{(0.25)^2 + (0.75)^2\}$$

$$\text{Gini Index for rectangle A, } I(A) = 1 - \{0.0625 + 0.5625\} = \mathbf{0.375}$$



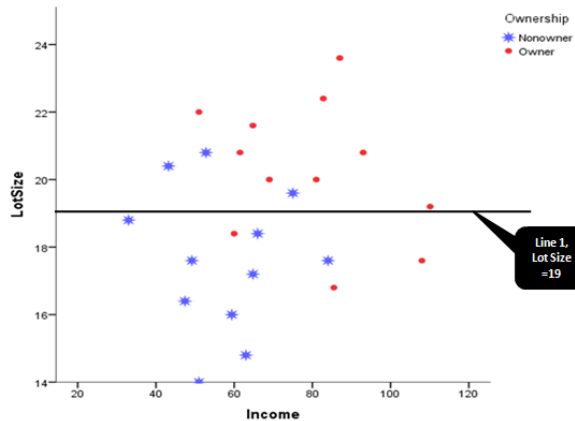
Total Impurity

$$TI = w_{Upper\ Rectangle} \times GI_{Upper\ Rectangle} + w_{Lower\ Rectangle} \times GI_{Lower\ Rectangle}$$

$$TI = 0.50 \times \mathbf{0.375} + 0.5 \times \mathbf{0.375}$$

$$\mathbf{TI = 0.375}$$

Entropy Measure



TOP RECTANGLE	Proportion of Owner in Top Rectangle	Proportion of non-Owner in Top Rectangle
p_k	-0.75	-0.25
$\log_2(p_k)$	-0.415037499	-2
$p_k * \log_2(p_k)$	0.311278124	0.5
Total =		0.811278124

Entropy measure before split

$$entropy(A) = - \sum_{k=1}^m p_k \log_2(p_k)$$

$$entropy(A) = -0.5 \times \log_2(0.5) + -0.5 \times \log_2(0.5)$$

$$entropy(A) = -0.5 \times (-1) + -0.5 \times (-1)$$

$$entropy(A) = 0.5 + 0.5 = \mathbf{1}$$

BOTTOM RECTANGLE	Proportion of Owner in Bottom Rectangle	Proportion of non-Owner in Bottom Rectangle
p_k	-0.25	-0.75
$\log_2(p_k)$	-2	-0.415037499
$p_k * \log_2(p_k)$	0.5	0.311278124
Total =		0.811278124
Weighted Sum	$= 0.5 * 0.811 + 0.5 * 0.811 = \mathbf{0.811}$	

Decision Trees

Regression & Classification

Data: student

Decision Tree

**“tell me and
i’ll forget.
show me
and i may
remember.
involve me
and i learn.
- Benjamin Franklin**