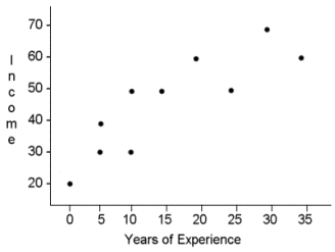
2.2 The "Least-Squares Line"

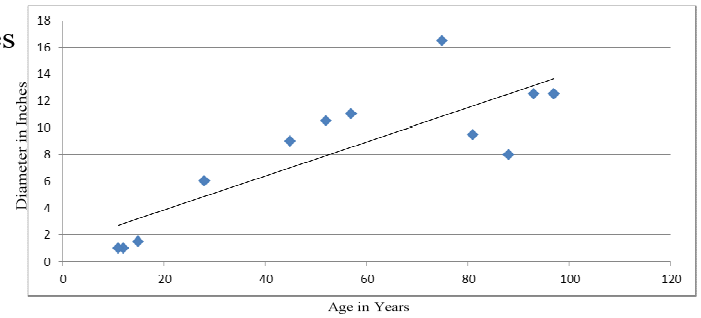


Example: You would like to determine whether there is a linear relationship between the diameter of a tree and its age.

Age (years)	Diameter (inches)
97	12.5
93	12.5
88	8.0
81	9.5
75	16.5
57	11.0
52	10.5
45	9.0
28	6.0
15	1.5
12	1.0
11	1.0

x = age of tree, in years
y = tree diameter, in inches

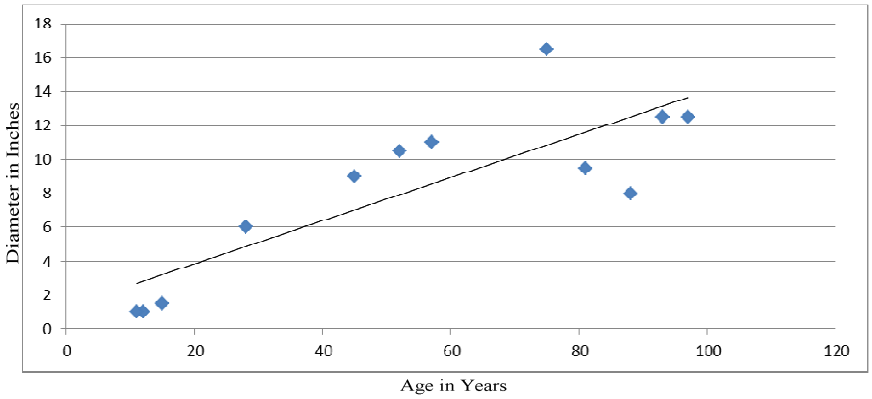
$y = 1.29 + 0.128x$



Predict the diameter of a tree that is 70 years old.

If two trees differ in age by 20 years, by how much would you predict their diameters differ?

How were the intercept and slope determined for the line?



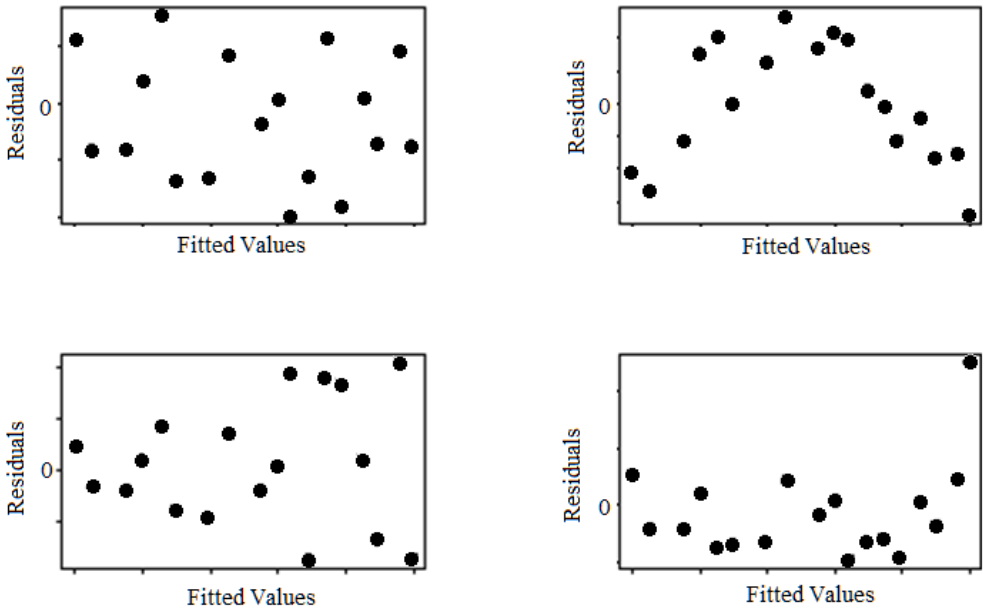
Use of the Least Squares Method assumes a linear model is appropriate and measurement errors are

- random and independent
- mean of zero
- spread is consistent across the graph of Fitted Values vs. Residuals
- normally distributed (but we don't know what this means yet)

The best diagnostic tool to check these assumptions is the residual plot.

- Fitted Values along horizontal axis
- Residuals along vertical axis

Residual Plots



Let's go back to our tree data:

Age x_i	Diameter y_i	Fitted Value $\hat{y}_i$	Residual = $y_i - \hat{y}_i$
11	1		
12	1		
15	1.5		
28	6		
45	9		
52	10.5		
57	11		
75	16.5		
81	9.5		
88	8		
93	12.5		
97	12.5		

Residual Plot for Tree Data