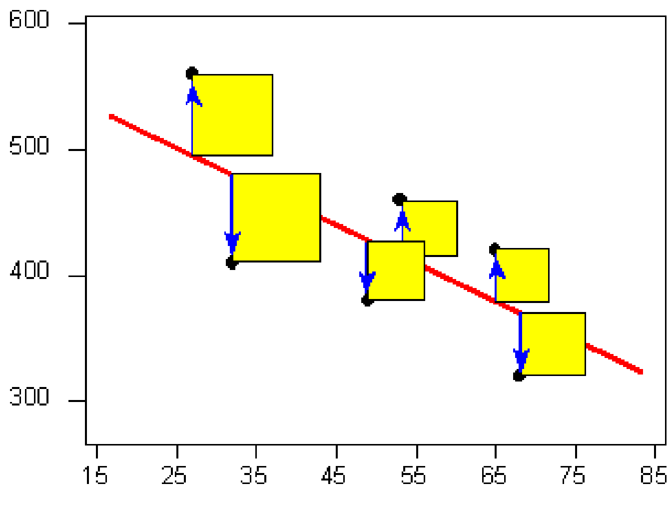
Case Q→Q: Summary

Learning Objective: In a given context, carry out the appropriate inferential method for comparing relationships and draw the appropriate conclusions.

So far, the researchers have observed linearity in the data, and based on a test concluded that this linear relationship between age and legibility distance can be generalized to the entire population of drivers.

Since that is the case, the researchers would now like to estimate the equation of the straight line that governs the linear relationship between age and legibility distance among drivers. As we commented earlier, this is done by finding the line that best fits the pattern of our observed data. Recall that this line is called the least squares regression line, which is the line that minimizes the sum of the squared vertical deviations:



In the Exploratory Data Analysis section, we presented the actual formulas for the slope and intercept of the line. We are not going to repeat those here; we will obtain those values from the output:

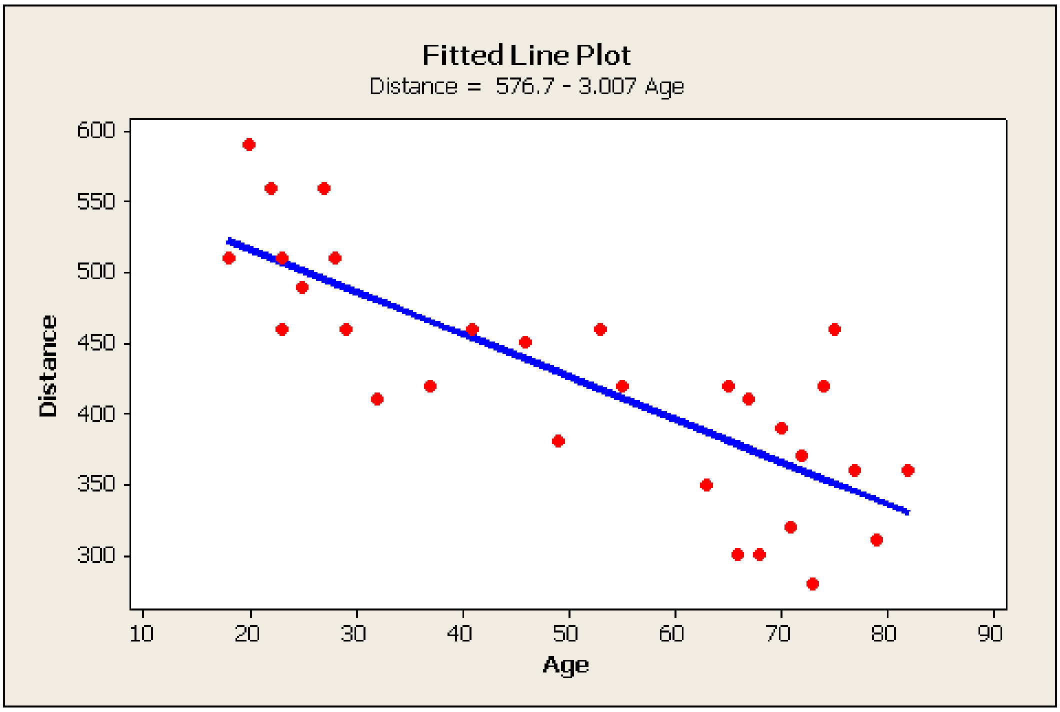
Regression Analysis: Distance versus Age

```
The regression equation is
Distance = 577 - 3.01 Age


Predictor     Coef  SE Coef      T      P
Constant    576.68    23.47  24.57  0.000
Age        -3.0068   0.4243  -7.09  0.000


S = 49.7616   R-Sq = 64.2%   R-Sq(adj) = 62.9%
```

and ask the software to plot it for us on the scatterplot so we can see how well it fits the data.

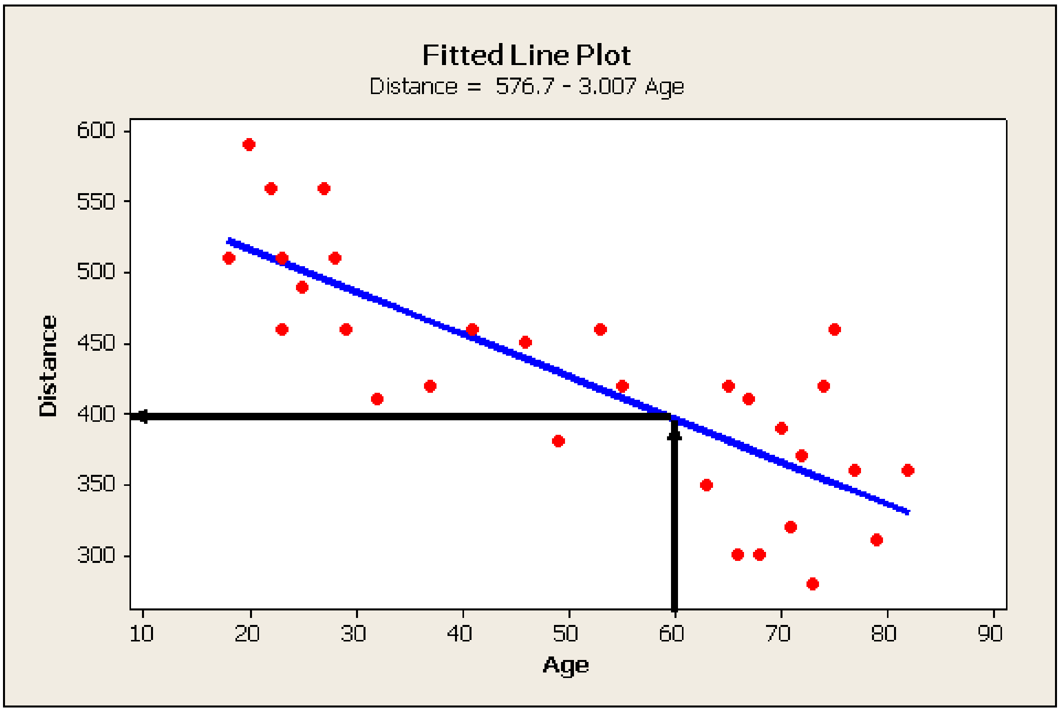


Based on the observed data, the researchers conclude that the linear relationship between age and legibility distance among drivers can be summarized with the line:

DISTANCE = 576.7 - 3.007 * AGE

In particular, the slope of the line is roughly -3, which means that for every year that a driver gets older (1 unit increase in X), the maximum legibility distance is reduced, on average, by 3 feet (Y changes by the value of the slope).

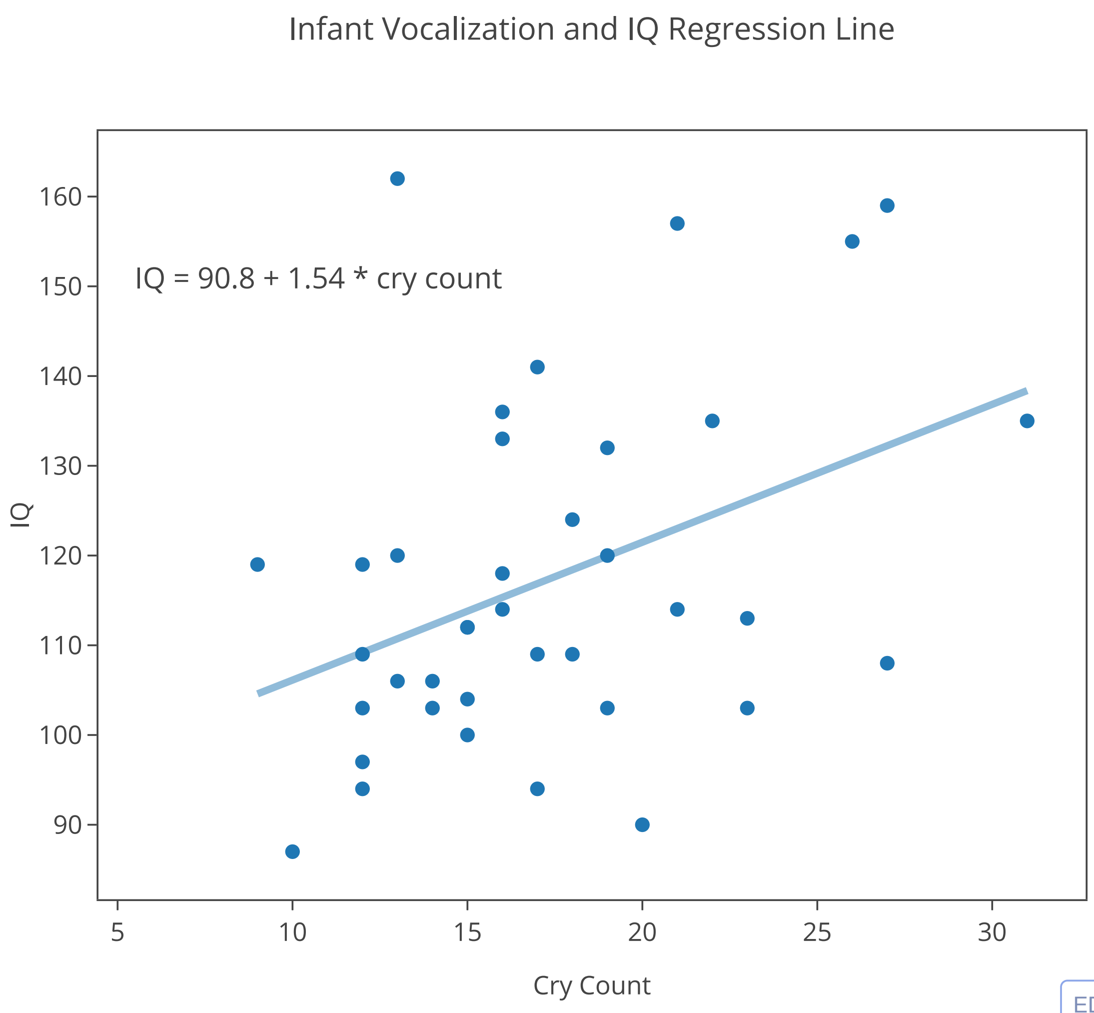
The researchers can also use this line to make predictions, remembering to beware of extrapolations (predictions for X values that are outside of the range of the original data). For example, using the equation of the line, we predict that the maximum legibility distance of a 60-year-old driver is: **distance = 576.7 - 3.007(60) = 396.28**. The following figure illustrates this prediction.



Scenario: Infant Vocalization and IQ

The purpose of this activity is to complete our discussion about our example that examines the relationship between vocalization soon after birth and IQ at age three. So far we explored the data using a scatterplot supplemented with the correlation r and discovered that the data display a moderately weak positive linear relationship. In addition, when we carried out the t-test for assessing the significance of this linear relationship and we concluded (based on the small p-value of 0.012) that the data provide fairly strong evidence of a moderately weak linear relationship between cry count soon after birth and IQ and age 3.

We would now like to consider using the least squares regression line for predicting IQ at age 3 based on cry count soon after birth. Below is the least squared regression line plotted on the scatterplot.



EDIT CHART

Learn By Doing

1/1 point (graded)

True or false? The least squares regression line fits the data well.

- True
- False ✔

Answer

Correct:

The least squares regression line does not fit the data very well due to the moderately weak linearity in the data. Visually, we see that the data points do not lay close to the line, resulting in a relatively poor fit of the line to the data.

Submit

Simple linear regression results:

Dependent Variable: IQ

Independent Variable: cry count

IQ = 90.75499 + 1.5363518 * cry count

Parameter estimates

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	90.75499	10.47342	≠ 0	36	8.665267	< 0.0001
Slope	1.5363518	0.5835417	≠ 0	36	2.6328056	0.0124

Learn By Doing

1/1 point (graded)

Use the least squares regression line (with intercept = 90.8 and slope = 1.54) to predict the IQ at age 3 of a newborn whose cry count is 19. Round your answer to ONE decimal place.

120.1

120.1

Answer

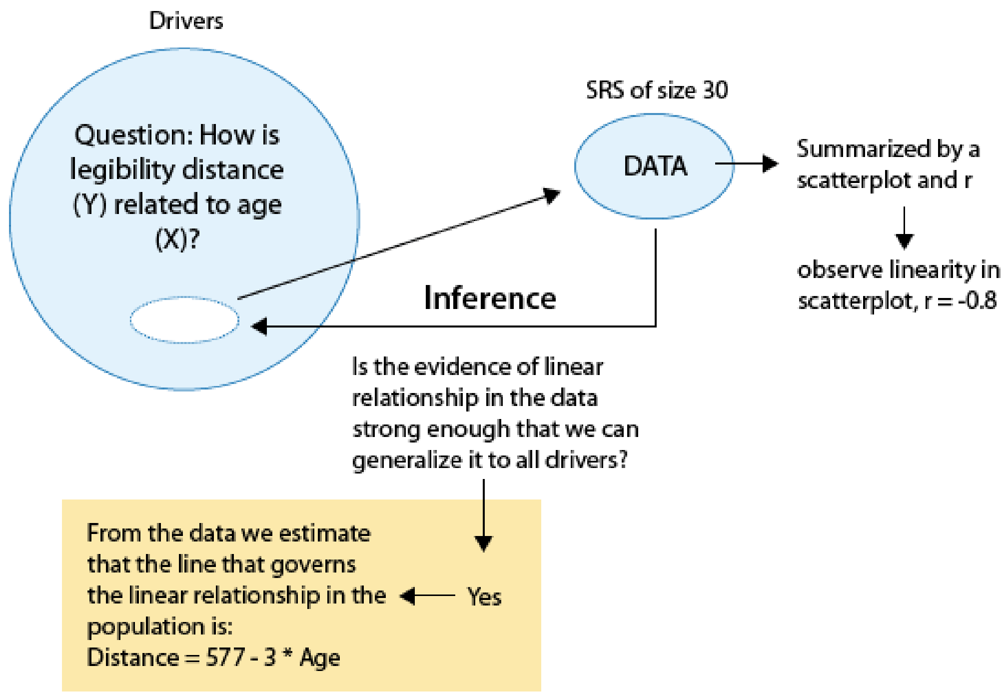
Correct:

The predicted IQ is obtained by plugging in a cry count of 19 into the regression line. We therefore obtain: Predicted IQ = 90.8 + 1.54 * 19 = 120.1.

Submit

Let's Summarize

Let's summarize in a figure all that the researchers have done:



- In the t-test for the significance of the linear relationship between two quantitative variables X and Y, we are testing
  - H_o: No linear relationship exists between X and Y.
  - H_a: A linear relationship exists between X and Y.
- The test assesses the strength of evidence provided by the data (as seen in the scatterplot and measured by the correlation r) and reports a p-value. The p-value is the probability of getting data such as that observed assuming that, in reality, no linear relationship exists between X and Y in the population.
- Based on the p-value, we draw our conclusions. A small p-value will indicate that we reject H_o and conclude that the data provide enough evidence of a real linear relationship between X and Y in the population.