Tutorial Acitivity 5

Sean Leggett - BDA201 - Winter 2020
March 2, 2020

Question

Is there a linear relationship between the age at which a child first begins to speak and his or her mental ability later on? To answer this question a study was conducted in which the age (in months) at which a child first spoke and the child's score on an aptitude test as a teenager were recorded:

Age 14, 26, 10, 9, 15, 20, 17, 11, 8, 20

Score 96, 71, 83, 91, 99, 87, 93, 100, 104, 94

Answer the following questions using R (where applicable)

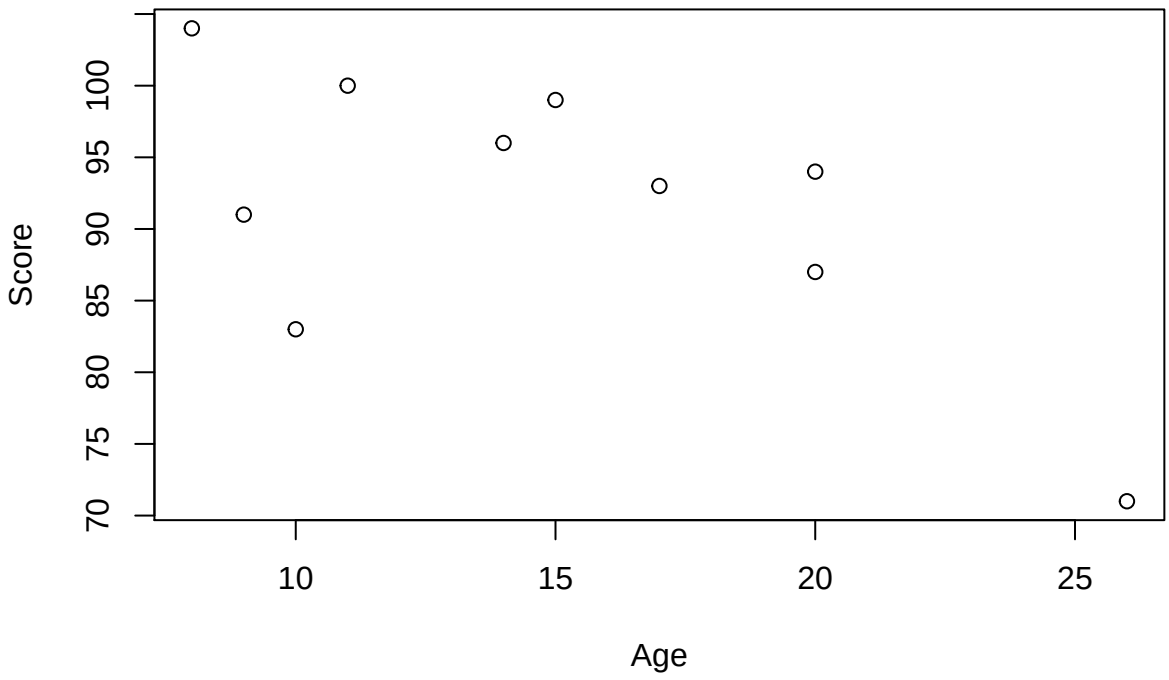
Draw a scatterplot and determine whether there appears to be a linear relationship between these two variables. What is the relationship, how would you describe it? Calculate and show covariance Calculate and show Pearson correlation coefficient r Determine and show what percentage of the variability in the aptitude score can be explained by the variability in the age at which a child begins speaking.

Answer

Scatterplot:

```
##create df
agescore <- data.frame('Age' = c(14, 26, 10, 9, 15, 20, 17, 11, 8, 20),
                       'Score' = c(96, 71, 83, 91, 99, 87, 93, 100, 104, 94)
)

agescoreplot <- plot(agescore)
```




```
agescoreplot
```

```
## NULL
```

There seems to be a general correlation between early speech and later scores on mental ability. There are a couple of outliers. This is a fairly strong, negative relationship on face value; the later you learn to speak the less likely you will achieve higher scores on acuity tests. Learning to speak after the age of 25 leaves us a little suspicious of this datapoint.

Calculate Covariance

```
age <- agescore$Age
score <- agescore$Score
agecov <- cov(score, age)
agecov
```

```
## [1] -34.77778
```

Calculate Pearson Correlation Coefficient

```
## we prefer cor.test to simply cor
coftest <- cor.test(age, score, method = "pearson")
coftest
```

```
##
##  Pearson's product-moment correlation
##
## data:  age and score
## t = -2.2755, df = 8, p-value = 0.05244
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.900902305  0.004604594
## sample estimates:
##       cor
## -0.626839
```

This supports the interpretation of strong negative correlation while not necessarily overwhelming.

Demonstrate what percentage of variability in aptitude scores can be explained by variability in age:

This could be demonstrated by assuming r = 0.626839, therefore r-squared = :

```
variability <- 0.626839 * 0.626839
variability
```

```
## [1] 0.3929271
```

39% of the variability in aptitude scores can be explained by variability in age.