Homework 07

Christina Lee

10/26/2021

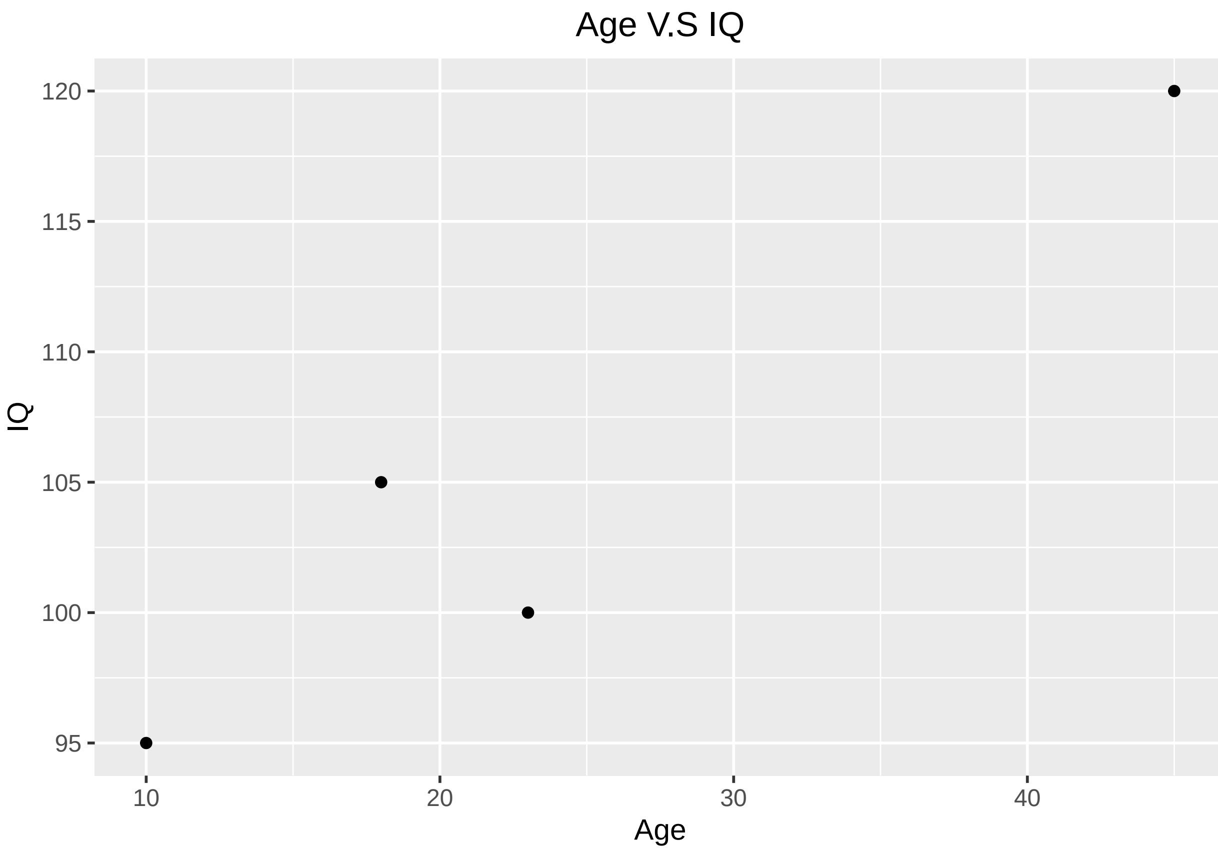
Is there an effect of Age on IQ? Please perform all calculations by hand using the equations in the lessons unless otherwise specified.

1. Plot these four points using R.

```r
# first create the table and save it as data frame.
ageIQ <- data.frame(Age = c(23,18,10,45), IQ = c(100,105,95,120))
ageIQ
```

```
##   Age  IQ
## 1  23 100
## 2  18 105
## 3  10  95
## 4  45 120
```

```r
# create the plot.
library(ggplot2)
ggplot(data=data.frame(ageIQ), mapping=aes(x=Age, y=IQ)) + geom_point() + ggtitle("Age V.S IQ") +
  theme(plot.title= element_text(hjust=0.5))
```



2. Calculate the covariance between age and IQ.

$$cov(x, y) = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$cov(x, y) = \frac{(23-24)(100-105) + (18-24)(105-105) + (10-24)(95-105) + (45-24)(120-105)}{3}$$

$$cov(x, y) = 153.\bar{3}$$

```r
x <- c(23,18,10,45)  # Age
y <- c(100,105,95,120) # IQ

# covariance calculated by R to verify hand calculation.
cov(x,y)
```

```
## [1] 153.3333
```

3. Calculate their correlation. What does the number you get indicate?

Sx = 14.98
Sy = 10.80

$$r = \frac{cov(x, y)}{S_x S_y}$$

$$r = \frac{153.3}{(14.98)(10.80)}$$

$$r = 0.95$$

```r
x <- c(23,18,10,45)  # Age
y <- c(100,105,95,120) # IQ

# correlation calculated by R to verify hand calculation.
cor(x,y)
```

```
## [1] 0.9470957
```

The correlation calculated by hand is 0.95 which is fairly close to 1. This indicates that when x (Age) is higher than its mean, y (IQ) is as well – Age and IQ are positively correlated.

4. Calculate the regression coefficients B0 and B1 and write out the equation of the best-fit line relating age and IQ.

First solve for B1:

$$r = \frac{cov(x,y)}{S_x S_y} = \beta_1 \frac{S_x}{S_y}$$
$$\frac{r}{\frac{S_x}{S_y}} = \beta_1$$
$$\beta_1 = \frac{0.95}{\frac{14.98}{10.80}}$$
$$\beta_1 = 0.68$$

Once we solved for B1, we can now solve for B0:

$$\beta_0 = \hat{y} - \beta_1 \hat{x}$$

$$\beta_0 = 105 - 0.68 * 24$$
$$\beta_0 = 88.68$$

Therefore, the best-fit line equation is then:

$$y = 88.68 + 0.68x$$

5. Calculate the predicted yi for each xi.

$$104.32 = 88.68 + 0.68(23)$$
$$\hat{y_i} = 104.32$$

$$100.92 = 88.68 + 0.68(18)$$
$$\hat{y_i} = 100.92$$

$$95.48 = 88.68 + 0.68(10)$$
$$\hat{y_i} = 95.48$$

$$119.28 = 88.68 + 0.68(45)$$

$$\hat{y_i} = 119.28$$

6. Calculate R^2 from the TSS/SSE equation. How does it relate to the correlation? What does the number you get indicate?

$$r^2 = \frac{TSS - SSE}{TSS}$$

$$TSS = \sum_i (y_i - \bar{y})^2$$

$$TSS = \sum_i (100 - 105)^2 + (105 - 105)^2 + (95 - 105)^2 + (120 - 105)^2$$

$$TSS = 350$$

$$SSE = \sum_i (y_i - \hat{y_i})$$

$$SSE = \sum_i (100 - 104.32)^2 + (105 - 100.92)^2 + (95 - 95.48)^2 + (120 - 119.28)^2$$

$$SSE = 36.05$$

$$r^2 = \frac{350 - 36.05}{350}$$

$$r^2 = 0.897$$

r^2 is a measure of the proportion of the total variation in y explained by x –how strong/weak are the variables correlated with each other. r^2 is also sometimes known as the proportional reduction in error so, we can then say we have reduced the error by 89.7%, or 89.7% of the variation in y is explained by the model –x and y are very much correlated.

7. Calculate the standard error of B1, and use that to test (using the t test) whether B1 is significant.

Let's first calculate the se for y hat.

$$se_{\hat{y}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2}}$$

$$se_{\hat{y}} = \sqrt{\frac{SSE}{n-2}}$$

$$se_{\hat{y}} = \sqrt{\frac{36.05}{4-2}}$$

$$se_{\hat{y}} = 4.24$$

Once we've calculated the se for y hat, solve for se of B1:

$$se_{\beta_1} = se_{\hat{y}} \sqrt{\frac{1}{\sum (x_i - \bar{x})^2}}$$

Solve for se of B1:

$$se_{\beta_1} = 4.24\frac{1}{\sqrt{674}}$$
$$se_{\beta_1} = 0.16$$

Now, conduct a t test to see whether B1 is significant:

$$H_a : \beta_1 \neq 0$$

$$H_o : \beta_1 = 0$$

$$Test\ Statistic = \frac{\beta_1 - \mu_0}{se_{\beta_1}}$$
$$Test\ Statistics = \frac{0.68 - 0}{0.16}$$

$$Test\ Statistics = 4.25$$

$$df = n - k - 1$$

$$df = 4 - 1 - 1$$

$$df = 2$$

Now, calculate the threshold value for two-tailed test:

```
qt(0.975,2) # upper threshold
```

```
## [1] 4.302653
```

```
qt(0.025,2) # lower threshold
```

```
## [1] -4.302653
```

$$T_{critical} = \pm 4.30$$

Results:
Since our t statistic value = 4.25 is not greater than 4.30, nor less than -4.30 the threshold value we fail to reject the null. This indicates that the positive correlation between age and IQ is not statistically significant –the change in age does not really produce a change in IQ and is just a statistical illusion.

8. Calculate the p-value for B1 and interpret it.

```
2*pt(4.25,2,lower.tail=F)
```

```
## [1] 0.05115253
```

Since p-value 0.05115253 is > 0.05 we again fail to reject the null hypothesis –the effect of age on IQ is not statistically significant.

9. Calculate the 95% CI for B1 and interpret it.

$$95\%\, CI = \beta_1 = 0.68 \pm 4.30 * 0.16 = [-0.008, 1.368]$$

We are 95% certain that as age increases per year, it will result in an increase in IQ between -0.008 and 1.368, with our best guess being 0.68 increase in IQ as age increases each year.

10. Confirm your results by regressing IQ on Age using R

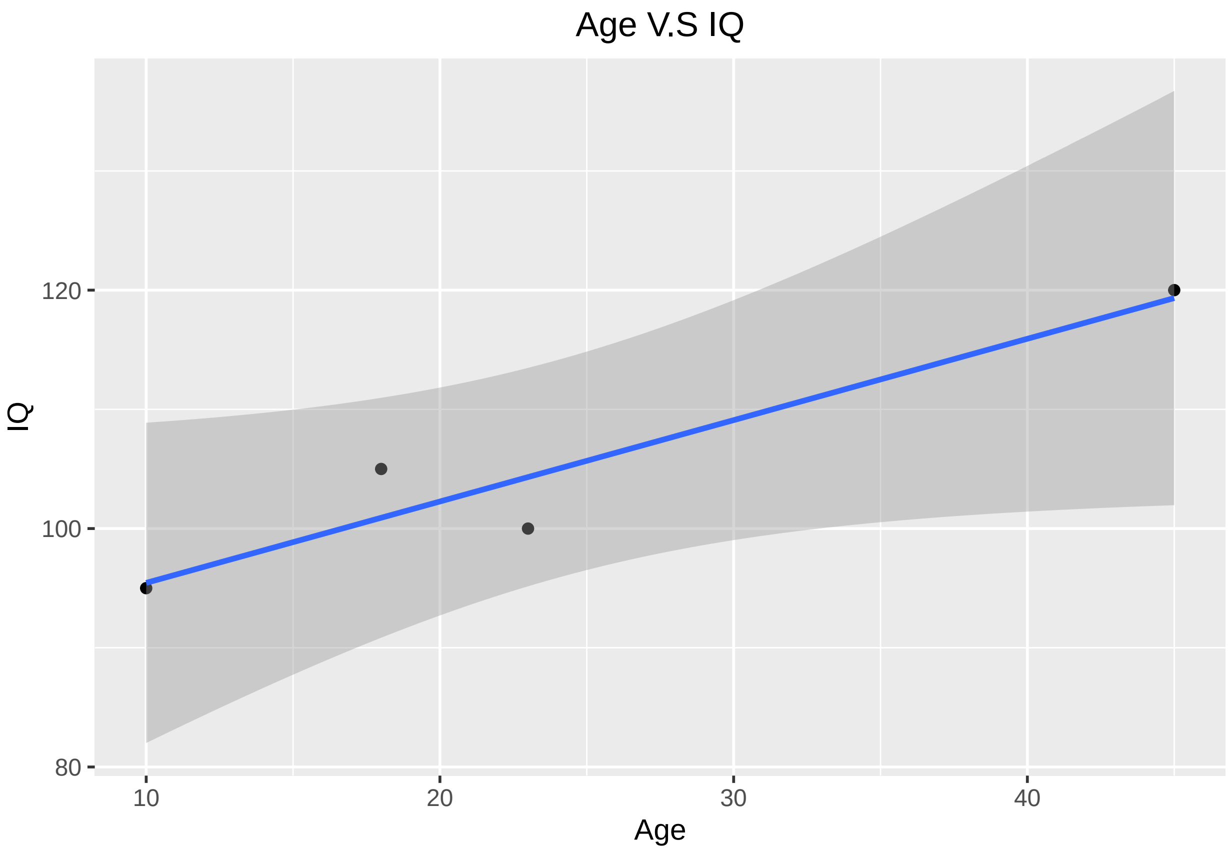
Results do match hand calculations above.

```
biv_model <- (lm(IQ ~Age, data =ageIQ))
summary(biv_model)
```

```
##
## Call:
## lm(formula = IQ ~ Age, data = ageIQ)
##
## Residuals:
##       1       2       3       4
## -4.3175  4.0950 -0.4451  0.6677
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  88.6202     4.4623  19.860  0.00253 **
## Age           0.6825     0.1635   4.173  0.05290 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.246 on 2 degrees of freedom
## Multiple R-squared:  0.897,  Adjusted R-squared:  0.8455
## F-statistic: 17.42 on 1 and 2 DF,  p-value: 0.0529
```

11. Plot your points again using R, including the linear fit line with its standard error.

```
# create the plot with best-fit line and standard error.
library(ggplot2)
ggplot(data=data.frame(ageIQ), mapping=aes(x=Age, y=IQ)) + geom_point() +
  geom_smooth(formula = y ~ x, method = "lm") + ggtitle("Age V.S IQ") +
  theme(plot.title= element_text(hjust=0.5))
```



12. What are you final conclusions about the relationship between age and IQ?

According to our results, r = 0.95 shows that age and IQ are positively correlated. With r^2 = 0.897, it suggests that we have reduced the error by 89.7% and 89.7% of the variation in y can be explained by the model. However, the t-test that we have conducted and with the p-value and t-statistics tells us that the positive correlation between age and IQ that we have witnessed is not, in fact, statistically significant –the change in age does not really produce a change in IQ and it is just a statistical illusion.