Final-Assignment–Question-2.R

snmus

2022-11-21

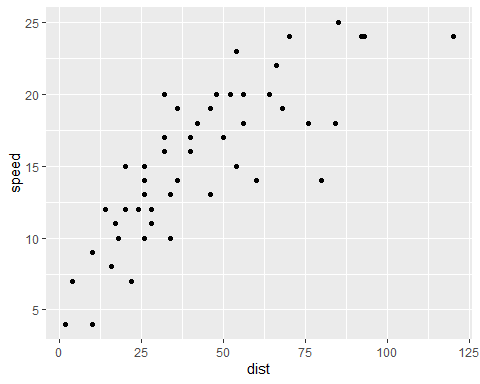
#Question2   
cars

## speed dist  
## 1 4 2  
## 2 4 10  
## 3 7 4  
## 4 7 22  
## 5 8 16  
## 6 9 10  
## 7 10 18  
## 8 10 26  
## 9 10 34  
## 10 11 17  
## 11 11 28  
## 12 12 14  
## 13 12 20  
## 14 12 24  
## 15 12 28  
## 16 13 26  
## 17 13 34  
## 18 13 34  
## 19 13 46  
## 20 14 26  
## 21 14 36  
## 22 14 60  
## 23 14 80  
## 24 15 20  
## 25 15 26  
## 26 15 54  
## 27 16 32  
## 28 16 40  
## 29 17 32  
## 30 17 40  
## 31 17 50  
## 32 18 42  
## 33 18 56  
## 34 18 76  
## 35 18 84  
## 36 19 36  
## 37 19 46  
## 38 19 68  
## 39 20 32  
## 40 20 48  
## 41 20 52  
## 42 20 56  
## 43 20 64  
## 44 22 66  
## 45 23 54  
## 46 24 70  
## 47 24 92  
## 48 24 93  
## 49 24 120  
## 50 25 85

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.3

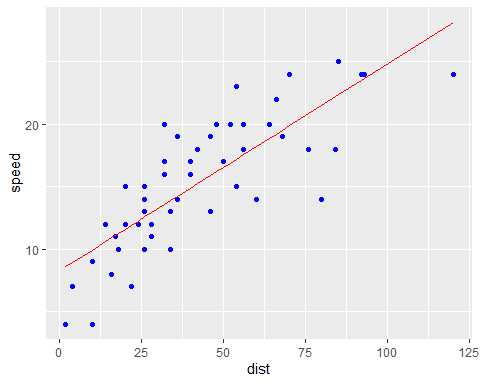
#(i) scatter plot of speed against distance.   
#y axis=speed and x axis =distance.  
#create ggplot variable   
#assign type of plot to data.  
cars\_plot<-ggplot(data=cars,mapping=aes(x=dist,y=speed))  
cars\_plot +  
 geom\_point()



#(ii)correlation  
corrcoeff\_cars<-cor(cars$dist,cars$speed,use="everything",method=c("spearman"))  
corrcoeff\_cars

## [1] 0.8303568

#A Pearson's correlation coeff of 0.8068949, Kendall's of 0.6689901 and Spearman's of 0.8303568, show speed and distance have a strong positive correlation.  
  
#(iii) line of best fit  
#creating simple linear regression model   
lm\_cars<- lm(speed~dist,data=cars)  
  
#store predictions as well as variable we want to plot against in dataframe for plotting of line of best fit  
predicted\_cars<-data.frame(speed\_pred=predict(lm\_cars,cars),dist=cars$dist)  
  
#creating plot  
ggplot(data =cars, aes(dist, speed)) +   
 geom\_point(color='blue') +  
 geom\_line(color='red',data =predicted\_cars, aes(dist, speed\_pred))



#(iv) interpretation of results and pvalue  
summary(lm\_cars)

##   
## Call:  
## lm(formula = speed ~ dist, data = cars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.5293 -2.1550 0.3615 2.4377 6.4179   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.28391 0.87438 9.474 1.44e-12 \*\*\*  
## dist 0.16557 0.01749 9.464 1.49e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.156 on 48 degrees of freedom  
## Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438   
## F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

#lm\_cars ; Speed=B1+B2dist +u  
#H0:B2=0, Ha:B2=!0  
#Pr|>t|>Pvalue=0. Therefore,we reject the null hypothesis