Activity3_2

Maria Sandate

2025-01-23

Question 1.

Data on structurally deficient highway bridges is compiled by the Federal Highway Administration (FHWA) and reported in the National Bridge Inventory (NBI). For each state, the NBI lists the number of structurally deficient bridges and the total area (thousands of square feet) of the deficient bridges. The data for the 50 states (plus the District of Columbia and Puerto Rico) are saved in the FHWABRIDGE file. For future planning and budgeting, the FHWA wants to estimate the total area of structurally deficient bridges in a state based on the number of deficient bridges.

```
data <- read.delim("C:/Users/sanda/Documents/Langara College/DANA-4810-001/Chapter 3/FHWABRIDGE.txt")

x<- data$NumberSD
y<- data$SDArea
```

a.Fit the least squares prediction equation.

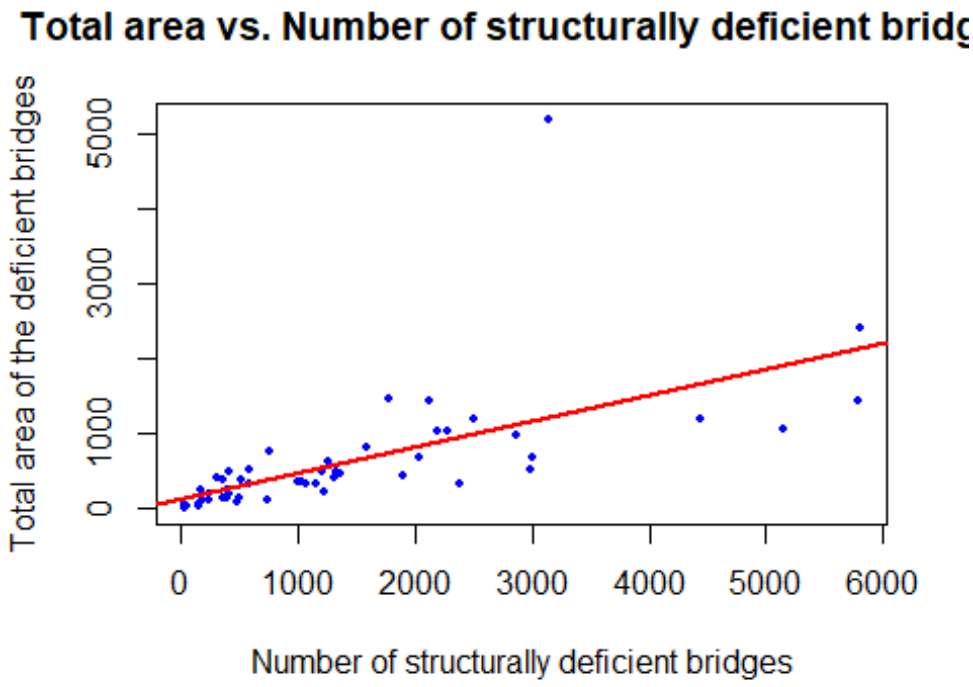
```
model=lm(y~x)
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -831.0 -146.3 -107.2  104.7 3972.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 119.86392  123.02005   0.974    0.335
## x             0.34560    0.06158   5.613 8.69e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 635.2 on 50 degrees of freedom
## Multiple R-squared:  0.3865, Adjusted R-squared:  0.3743
## F-statistic:  31.5 on 1 and 50 DF,  p-value: 8.695e-07
```

b. Plot the data and graph the least squares line as a check on your calculations.

```
# Scatterplot or Scatter Diagram
# Note: For 2 variables, X-var goes first in plot()
plot(x, # X-variable
     y, # Y-variable
     main = "Total area vs. Number of structurally deficient bridges",
     ylab = "Total area of the deficient bridges",
     xlab = "Number of structurally deficient bridges",
     cex = 0.8, # size of the dot
     pch = 20, # style of the dot, default is 1
     col = "blue")

abline(a = 119.86392, b = 0.34560, col = "red", lwd = 2)
```



c. List the assumptions required for the regression analysis.

-The mean of the probability distribution is epsilon is 0 -The variance of the probability distribution of epsilon is constant for all settings of the independent variable x -The probability distribution of epsilon is normal. -The values of epsilon associated with any two observed values of y are independent.

d. Locate the estimated standard error of the regression model, s, on the printout.

```
y_length = length(y)

s=sqrt(sum(model$residuals^2)/(y_length-2))
s

## [1] 635.186
```

e. Use the value of s to find a range where most (about 95%) of the errors of prediction will fall.

```
rang_95=2*s
rang_95

## [1] 1270.372
```

The range where most of the errors of prediction will fall is y_hat +-1,270.4 (y_hat +-2*s)

Question 2

Refer to the data on sale prices and total appraised values of 76 residential properties in an upscale Tampa, Florida, neighborhood, (Activity chapter3_1).

```
load ("C:/Users/sanda/Documents/Langara College/DANA-4810-001/Chapter 3/TAMPALMS.Rdata")

y <- TAMPALMS$Sale_Price
x <- TAMPALMS$Market_Val

model=lm(y~x)
summary(model)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -282.171  -24.829    1.807   29.791  188.792
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.35868   13.76817   0.099    0.922
## x            1.40827    0.03693  38.132   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68.76 on 74 degrees of freedom
## Multiple R-squared:  0.9516, Adjusted R-squared:  0.9509
## F-statistic:  1454 on 1 and 74 DF,  p-value: < 2.2e-16
```

a.Use the output to determine whether there is a positive linear relationship between appraised property value x and sale price y for residential properties sold in this neighborhood. That is, determine if there is sufficient evidence (at α = .01) to indicate that β1, the slope of the straight-line model, is positive.

Ho: beta1=0 Ha: beta1>0

beta1=1.40827

Because p_value is <2.2e-16/2 (less than α = .01), we reject Ho, so beta1 is positive

b. Find a 95% confidence interval for the slope,β1, on the printout. Interpret the result practically.

```
confint(model, level=0.95)
```

```
##                   2.5 %    97.5 %
## (Intercept) -26.075006 28.792368
## x             1.334683  1.481858
```

We are 95% confident that the sale price will increase between $1335 and $1482 for every $1000 increase in market value.

c.What can be done to obtain a narrower confidence interval in part b?

Increase the sample size because the confidence interval decreases as the sample size increases.