205615894_stats101A_hw3

Takao

1/19/2022

1

```
football_data <- read.csv("football.csv")
head(football_data)
```

```
##            Team Games.Won Opponent.Rushing.Yards
## 1  Washington        10                   2205
## 2    Minnesota       11                   2096
## 3 New England       11                   1847
## 4      Oakland       13                   1903
## 5   Pittsburgh       10                   1457
## 6   Baltimore       11                   1848
```

(a)

The response variable is the number of games won or Games.Won The predictor variable is the number of yards gained rushing by opponents or Opponent.Rushing.Yards

(b)

```
slr.model1 <- lm(Games.Won ~ Opponent.Rushing.Yards, data = football_data)
print(slr.model1)
```

```
##
## Call:
## lm(formula = Games.Won ~ Opponent.Rushing.Yards, data = football_data)
##
## Coefficients:
##            (Intercept)  Opponent.Rushing.Yards
##              21.788251               -0.007025
```

```
b.coeff <- coefficients(slr.model1)

plot(football_data$Games.Won ~ football_data$Opponent.Rushing.Yards, xlab = "# of yards gained rushing

abline(a = b.coeff[1], b = b.coeff[2])
```
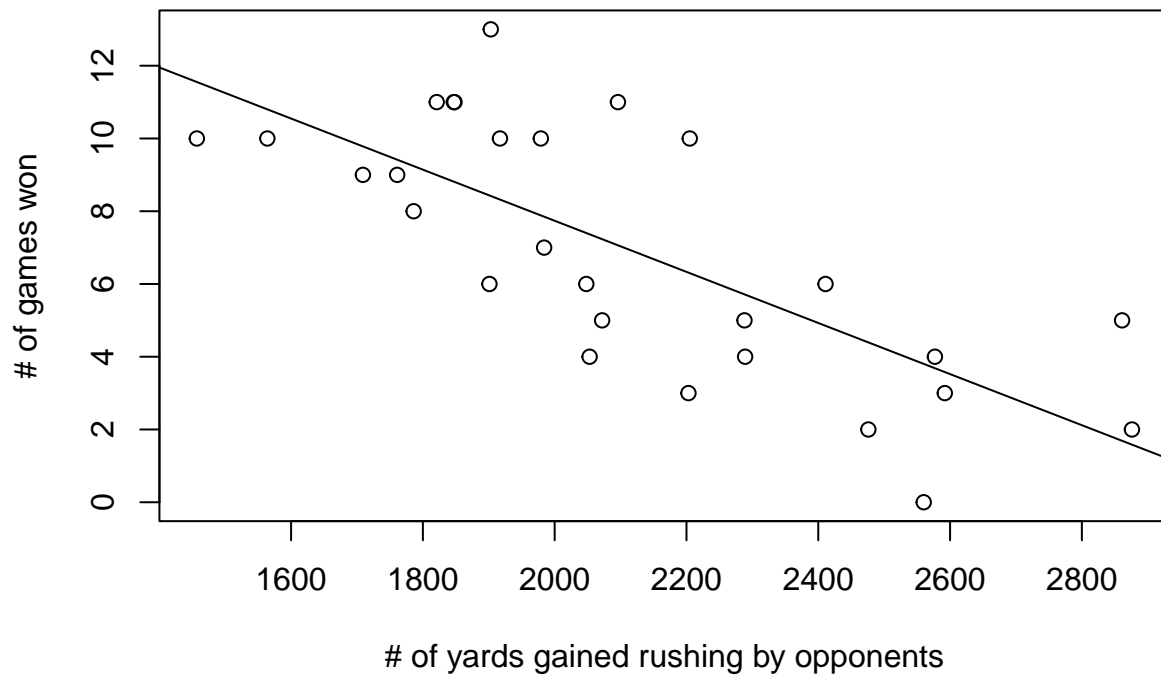
```
b.coeff
```

```
##           (Intercept) Opponent.Rushing.Yards
##            21.7882509             -0.0070251
```

Intercept = 21.788251 In the 1976 National Football League, we predict the number of games won (per 14-game season) by a team is 21.788251 games when the number of yards gained rushing by opponents is 0 yards.

Slope = -0.0070251 In the 1976 National Football League, we predict that the number of games won (per 14-game season) by a team will decrease by 0.0070251 games when the number of yards gained rushing by opponents increases by 1 yard.

(c)

```
anova(slr.model1)
```

```
## Analysis of Variance Table
##
## Response: Games.Won
##                       Df Sum Sq Mean Sq F value    Pr(>F)
## Opponent.Rushing.Yards  1 178.09 178.092  31.103 7.381e-06 ***
## Residuals              26 148.87   5.726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2

The ANOVA table shows a p-value of 0.000007381. Based on a significance level of $\alpha = 0.05$, we have significance evidence to reject the null hypothesis. since $0.000007381 < 0.05$. There is sufficient evidence that the number of yards gained rushing by opponents has an effect on the number of games won (per 14-game season) by a team.

(d)

```
confint(slr.model1, level = 0.95)
```

```
##                          2.5 %       97.5 %
## (Intercept)         16.246064040 27.330437725
## Opponent.Rushing.Yards -0.009614347 -0.004435854
```

(-0.009614347, -0.004435854)

If, in repeated random sampling, we construct a large number of confidence intervals, 95% of those intervals will contain the true parameter value β_1.

We are 95% confident that the interval (-0.009614347, -0.004435854) contains the true value of the slope β_1.

(e)

```
summary(slr.model1)
```

```
##
## Call:
## lm(formula = Games.Won ~ Opponent.Rushing.Yards, data = football_data)
##
## Residuals:
##     Min      1Q Median     3Q     Max
## -3.804 -1.591 -0.647  2.032   4.580
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            21.788251   2.696233   8.081 1.46e-08 ***
## Opponent.Rushing.Yards -0.007025   0.001260  -5.577 7.38e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.393 on 26 degrees of freedom
## Multiple R-squared:  0.5447, Adjusted R-squared:  0.5272
## F-statistic:  31.1 on 1 and 26 DF,  p-value: 7.381e-06
```

The value of R^2 is shown at the output "Multiple R-squared: 0.5447" This means that the model $Y = \beta_0 + \beta_1 x + \epsilon$ explains 54.47% of the total variability in the response number of games won (per 14-game season) by a team.

(f)

Using the jargon of the problem, we know that this is a mean response so this is a confidence interval, not a prediction interval

```
dataprediction <- data.frame(Opponent.Rushing.Yards = 2000)

predict(slr.model1, dataprediction, interval = "confidence", level = 0.95)
```

```
##       fit      lwr      upr
## 1 7.73805 6.765753 8.710348
```

(6.765753, 8.710348)

We are 95% confident that the interval (6.765753, 8.710348) contains the true mean number of games won if the opponents' yards rushing is 2000 yards.