TruongWileyHW08

Duy Truong

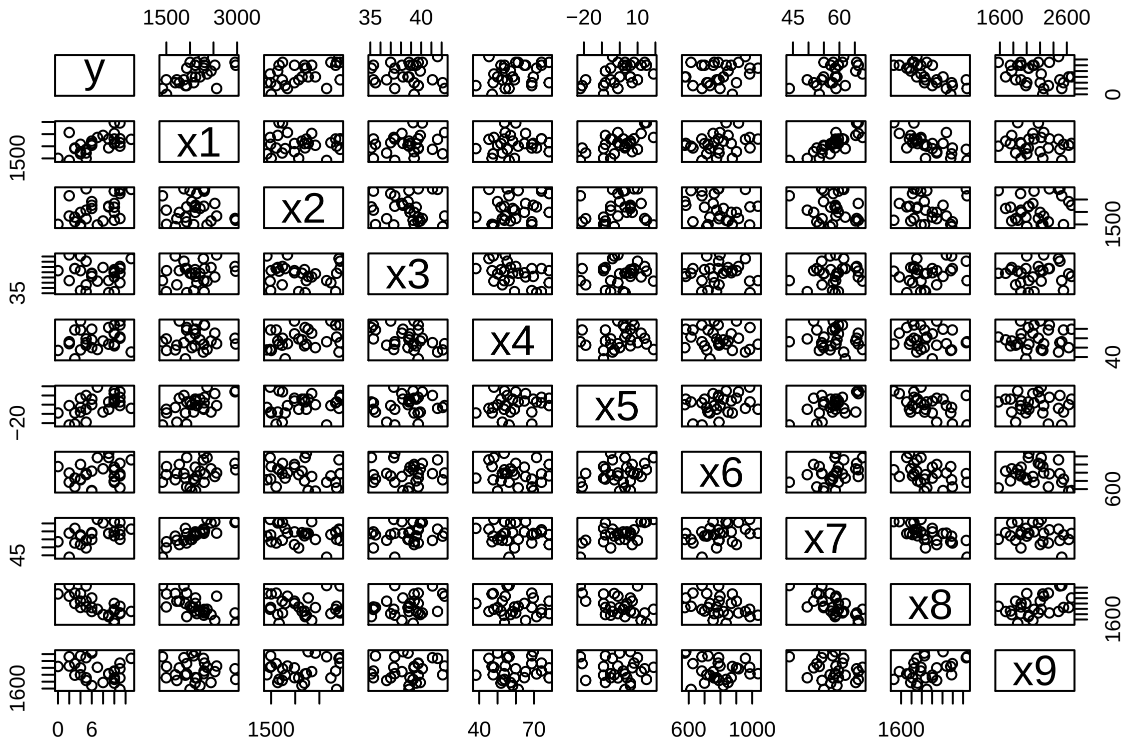
February 18, 2019

R Markdown

```
dat=read.csv("C:\\Users\\Nick\\Documents\\0_Spring 2019\\Applied Regression\\Labs_HW\\Data_Sets\\B1.csv
dat
```

```
##     y   x1   x2   x3   x4  x5   x6   x7   x8   x9
## 1  10 2113 1985 38.9 64.7   4  868 59.7 2205 1917
## 2  11 2003 2855 38.8 61.3   3  615 55.0 2096 1575
## 3  11 2957 1737 40.1 60.0  14  914 65.6 1847 2175
## 4  13 2285 2905 41.6 45.3  -4  957 61.4 1903 2476
## 5  10 2971 1666 39.2 53.8  15  836 66.1 1457 1866
## 6  11 2309 2927 39.7 74.1   8  786 61.0 1848 2339
## 7  10 2528 2341 38.1 65.4  12  754 66.1 1564 2092
## 8  11 2147 2737 37.0 78.3  -1  761 58.0 1821 1909
## 9   4 1689 1414 42.1 47.6  -3  714 57.0 2577 2001
## 10  2 2566 1838 42.3 54.2  -1  797 58.9 2476 2254
## 11  7 2363 1480 37.3 48.0  19  984 67.5 1984 2217
## 12 10 2109 2191 39.5 51.9   6  700 57.2 1917 1758
## 13  9 2295 2229 37.4 53.6  -5 1037 58.8 1761 2032
## 14  9 1932 2204 35.1 71.4   3  986 58.6 1709 2025
## 15  6 2213 2140 38.8 58.3   6  819 59.2 1901 1686
## 16  5 1722 1730 36.6 52.6 -19  791 54.4 2288 1835
## 17  5 1498 2072 35.3 59.3  -5  776 49.6 2072 1914
## 18  5 1873 2929 41.1 55.3  10  789 54.3 2861 2496
## 19  6 2118 2268 38.2 69.6   6  582 58.7 2411 2670
## 20  4 1775 1983 39.3 78.3   7  901 51.7 2289 2202
## 21  3 1904 1792 39.7 38.1  -9  734 61.9 2203 1988
## 22  3 1929 1606 39.7 68.8 -21  627 52.7 2592 2324
## 23  4 2080 1492 35.5 68.8  -8  722 57.8 2053 2550
## 24 10 2301 2835 35.3 74.1   2  683 59.7 1979 2110
## 25  6 2040 2416 38.7 50.0   0  576 54.9 2048 2628
## 26  8 2447 1638 39.9 57.1  -8  848 65.3 1786 1776
## 27  2 1416 2649 37.4 56.3 -22  684 43.8 2876 2524
## 28  0 1503 1503 39.3 47.0  -9  875 53.5 2560 2241
```

```
plot(dat)
```



Problem 3.1a Fit a multiple linear regression model relating the number of games won to the team ' s passing yardage (x2), the percentage of rushing plays (x7), and the opponents ' yards rushing (x8).

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4555 -1.5350 -0.0247  1.1613  3.5937
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.2029899  8.6811423  -0.023 0.981538
## x1           0.0058707  0.0011320   5.186  2.6e-05 ***
## x2           0.0034436  0.0008569   4.019 0.000502 ***
## x3          -0.3247098  0.2205033  -1.473 0.153857
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.216 on 24 degrees of freedom
## Multiple R-squared:  0.6394, Adjusted R-squared:  0.5943
## F-statistic: 14.19 on 3 and 24 DF,  p-value: 1.59e-05
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```
##  (Intercept)           x1           x2           x3
## -0.202989943  0.005870742  0.003443616 -0.324709827
```

For our coefficients we have x1 = 0.0059, x2 = 0.0034, and x3 = -0.3247. x1 represents the team's passing yardage, for every increase in the yardage holding x2 and x3 constant, there is a increase of the number of games won by 0.0059. x2 represents the percentage of a team's rushing plays, every increase of a teams rushing plays, holding x1 and x3 constant, increases the number of games won by 0.0034. x3 represents the opponents' rushing yards, holding x1 and x2 constant, the number of games won decreasesby 0.3247 for each yard rushed by the opponent.

3.1 Calculate/Extract the hat matrix, coefficient and var-covariance matrix estimates.

```r
#hat matrix
Hat_Matrix <- X %*% solve(t(X) %*% X) %*% t(X)
# Coefficient
solve(t(X) %*% X) %*% t(X)%*%y
```

```
##            [,1]
##    -1.808372059
## x1  0.003598070
## x2  0.193960210
## x3 -0.004815494
```

```r
coef(fit)
```

```
##  (Intercept)           x1           x2           x3
## -0.202989943  0.005870742  0.003443616 -0.324709827
```

```r
#Var-Covariance Matrix Estimaes
vcov( fit )
```

```
##               (Intercept)            x1            x2            x3
## (Intercept) 75.3622324412 -7.199137e-04 -2.064387e-03 -1.792896e+00
## x1          -0.0007199137  1.281444e-06  2.201546e-08 -5.256411e-05
## x2          -0.0020643872  2.201546e-08  7.342127e-07  1.180410e-05
## x3          -1.7928959641 -5.256411e-05  1.180410e-05  4.862170e-02
```

3.2 Using the results of Problem 3.1, show numerically that the square of the simple correlation coeffi cient between the observed values yi and the fitted values y^i equals R2.

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4555 -1.5350 -0.0247  1.1613  3.5937
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.2029899  8.6811423  -0.023 0.981538
## x1           0.0058707  0.0011320   5.186  2.6e-05 ***
## x2           0.0034436  0.0008569   4.019 0.000502 ***
## x3          -0.3247098  0.2205033  -1.473 0.153857
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.216 on 24 degrees of freedom
```

```
## Multiple R-squared:  0.6394, Adjusted R-squared:  0.5943
## F-statistic: 14.19 on 3 and 24 DF,  p-value: 1.59e-05
```

```
## [1] 0.638401
```

```
## [1] 0.6394093
```

$$r_{y,\hat{y}}^2 = \left(\frac{Cov(y,\hat{y})}{\sqrt[2]{Var(y)Var(\hat{y})}} \right)^2$$

$$r_{y,\hat{y}}^2 = \frac{Cov(y,\hat{y})}{\sqrt[2]{Var(y)Var(\hat{y})}} \frac{Cov(y,\hat{y})}{\sqrt[2]{Var(y)Var(\hat{y})}}$$

$$r_{y,\hat{y}}^2 = \frac{Cov(y,\hat{y})Cov(y,\hat{y})}{Var(y)Var(\hat{y})}$$

$$r_{y,\hat{y}}^2 = \frac{Cov(\hat{y}+e,\hat{y})Cov(\hat{y}+e,\hat{y})}{Var(y)Var(\hat{y})}$$

$$r_{y,\hat{y}}^2 = \frac{\left(Cov(\hat{y},\hat{y}) + Cov(\hat{y},e)\right)\left(Cov(\hat{y},\hat{y}) + Cov(\hat{y},e)\right)}{Var(y)Var(\hat{y})}$$

$$r_{y,\hat{y}}^2 = \frac{Cov(\hat{y},\hat{y})Cov(\hat{y},\hat{y})}{Var(y)Var(\hat{y})}$$

$$r_{y,\hat{y}}^2 = \frac{Var(\hat{y})Var(\hat{y})}{Var(y)Var(\hat{y})}$$

$$r_{y,\hat{y}}^2 = \frac{Var(\hat{y})}{Var(y)} = \frac{\frac{1}{n}\sum_i^N(\hat{y}_i - \bar{\hat{y}})^2}{\frac{1}{n}\sum_i^N(y_i - \bar{y})^2} = \frac{\sum_i^N(\hat{y}_i - \bar{\hat{y}})^2}{\sum_i^N(y_i - \bar{y})^2} = \frac{ESS}{TSS} = R^2$$

$$r_{y,\hat{y}}^2 = R^2$$

$$E(\hat{\beta})] = E[(X^TX)^{-1}X^T(X\beta+\varepsilon)] = \beta + E[(X^TX)^{-1}X^T\varepsilon)] = \beta + E[(X^TX)^{-1}X^T\varepsilon|X)] = \beta + E[(X^TX)^{-1}X^TE[\varepsilon|X])] = \beta$$

$$E(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] = E[((X^TX)^{-1}X^T\varepsilon)((X^TX)^{-1}X^T\varepsilon)^T] = E[(X^TX)^{-1}X^T\varepsilon\varepsilon\, X(X^TX)^{-1}] = E[(X^TX)^{-1}X^T\sigma^2X(X^TX)^{-1}$$