Scott Kobos

Stat318 HW10

4)

a)

```
> baseball.data= read.delim("clipboard", header = T)
> attach(baseball.data)
> baseball.mod= lm(BA44 ~ BA43)
> summary(baseball.mod)

Call:
lm(formula = BA44 ~ BA43)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.052177 -0.017221  0.006301  0.027441  0.039027 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  0.10368    0.06636   1.562   0.1332  
BA43         0.67517    0.25128   2.687   0.0138 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02968 on 21 degrees of freedom
Multiple R-squared:  0.2558,	Adjusted R-squared:  0.2204 
F-statistic:  7.22 on 1 and 21 DF,  p-value: 0.0138
```

b)

```
> anova(baseball.mod)
Analysis of Variance Table

Response: BA44
          Df    Sum Sq   Mean Sq F value Pr(>F)  
BA43       1 0.0063595 0.0063595  7.2196 0.0138 *
Residuals 21 0.0184984 0.0008809                 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

c)

Variability explained by the line is the SSR= .0063595

d)

Variability not explained by the line is the SSE=.0184984 e)

The residual standard error provides an estimate for σ^2, as the RSE is equivalent to the square root of the MSE, which is just σ, so squaring the RSE gives us σ^2. RSE=.02968, $.02968^2$ = 7.279204E-4= σ^2

f)

R^2= .2558. This low of an R^2 value tells us the line is a bad fit.

Interpretation: 25.58% of the variability in 1944 batting averages is explained by the line with 1943 batting averages.