STAT 509 Section E01 Homework 6

Due in class on April 21

1. In an article in Statistics and Computing Carlin and Gelfand investigated the age (x) and length (y) of 27 captured dugongs (sea cows).

x =1.0, 1.5, 1.5, 1.5, 2.5, 4.0, 5.0, 5.0, 7.0, 8.0, 8.5, 9.0, 9.5, 9.5, 10.0, 12.0, 12.0, 13.0, 13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5

y =1.80, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47, 2.19, 2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43, 2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57

1. Find the least squares estimates of the slope and the intercept in the simple linear regression model. Find an estimate of σ^2 .

Hint: Use R to load the data and the regression function lm to find the least squres estimates and an estimate of σ^2 .

R code is list below:

x=c(1.0, 1.5, 1.5, 1.5, 2.5, 4.0, 5.0, 5.0, 7.0, 8.0, 8.5, 9.0, 9.5, 9.5, 10.0, 12.0, 12.0, 13.0, 13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5)

y=c(1.80, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47, 2.19,2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43, 2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57)

fit=lm(y~x)

summary(fit)

> x=c(1.0, 1.5, 1.5, 2.5, 4.0, 5.0, 5.0, 7.0, 8.0, 8.5, 9.0, 9.5, 10.0, 12.0, 12.0, 13.0, 13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5)

> y=c(1.88, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47, 2.19, 2.26, 2.40, 2.396, 2.41, 2.50, 2.32, 2.32, 2.43, 2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57)

> fit=lm(y~x)

Error in model.frame.default(formula = y ~ x, drop.unused.levels = TRUE) : variable lengths differ (found for 'x')

> fit Error: object 'fit' not found

> fit=lm(y,x) Error in formula.default(object, env = baseenv()) : invalid formula

> fit=lm(y ~ x) Error in model.frame.default(formula = y ~ x, drop.unused.levels = TRUE) : variable lengths differ (found for 'x')

> clear Error: object 'clear' not found

> x=c(1.0, 1.5, 1.5, 1.5, 2.5, 4.0, 5.0, 5.0, 7.0,

+ 8.0, 8.5, 9.0, 9.5, 9.5, 10.0, 12.0, 12.0, 13.0,

+ 13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5)

> y=c(1.80, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47,

+ 2.19,2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43, 2.47,

+ 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57)

> fit=lm(y~x)

> summary(fit)

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-0.35439 -0.06554 0.03407 0.11959 0.24920

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.019769 0.053128 38.02 < 2e-16 \*\*\*

x 0.028718 0.003966 7.24 1.38e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1592 on 25 degrees of freedom

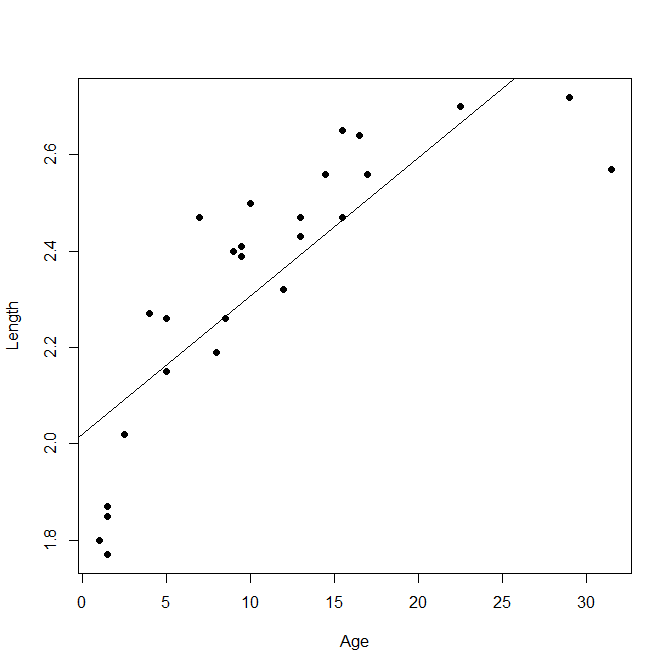
Multiple R-squared: 0.6771, Adjusted R-squared: 0.6642

F-statistic: 52.42 on 1 and 25 DF, p-value: 1.378e-07p

1. Estimate the mean length of dugongs at age 11.

> plot(x,y,xlab= "Age", ylab= "Length",pch=16)

> abline(fit)



Rough estimate of 2.3 based on line

1. Based on the R outputs, test the hypotheses H0 : β1 = 0 versus H1 : β1 =/ 0.

Based on the pvalue of p-value: 1.378e-07 we reject H0

1. What is the coefficient of determination, R2 ? (hint: in R, it is called “Multiple R-squared”)

Multiple R-squared: 0.6771,

1. What is the correlation between the response (y) and predictor (x) variables? R code: cor(x,y)

> cor(x,y)

[1] 0.8228604