HW10-R-ajmn100

Pg 621 # 11.73 based on 11.68

> ## Pg621#11.73 based on 11.68

> x1<-c(-3,-2,-1,0,1,2,3);

> x2<-x1^2;

> X<-cbind(1,x1,x2);

> Y<-c(1,0,0,-1,-1,0,0);

> B <- solve(crossprod(X))%\*%crossprod(X,Y);B;

[,1]

-0.7142857

x1 -0.1428571

x2 0.1428571

> SSE <- crossprod(Y)-crossprod((X%\*%B),Y);SSE;

[,1]

[1,] 0.5714286

> SS <- SSE/(length(x1)-2-1);SS;

[,1]

[1,] 0.1428571

> a <- c(0,1,2);a;

[1] 0 1 2

> SXX <- crossprod(a,solve(crossprod(X)))%\*%a;SXX;

[,1]

[1,] 0.08333333

> TS <- (B[2]+2\*B[3])/(SS\*SXX)^0.5;TS;

[,1]

[1,] 1.309307

Pg 622 # 11.76

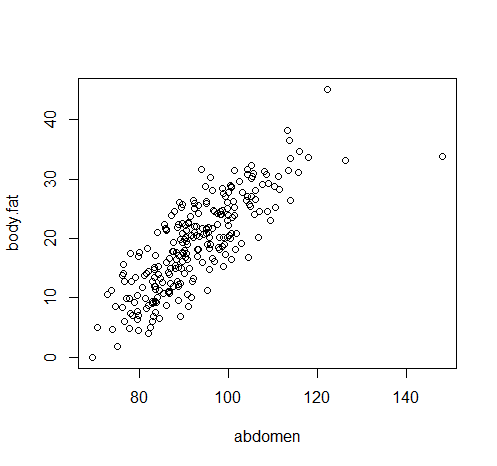
|  |
| --- |
| > ## Pg622#11.76  > XX <- rbind(c(151401.8, 2.6, 100.5, -28082.9),  + c( 2.6, 1.0, 0.0, 0.4),  + c( 100.5, 0.0, 8.1, 5.2),  + c(-28082.9, 0.4, 5.2, 6038.2));  > XX <- XX\*10^-4  > B <- c(38.83, - 0.0092, -0.92, 11.56)  > a <- c(1,914,65,6)  >  > atXXa <- crossprod(a,XX)%\*%a;atXXa  [,1]  [1,] 92.76617  > atB <- crossprod(a,B);atB  [,1]  [1,] 39.9812 |
|  |
| |  | | --- | |  | |

This is in regard to the file fat.R which consists of a data set called fat. In order to load this data set, just open this script, select all of the text, and then hit Ctrl+R. This data set consists of several body measurements that can be used to predict body-fat percentage (the column body.fat in the data frame). These measurements are taken on 252 adult men using a scale with a height measurement apparatus and a measuring tape (measuring in cm), in hopes of being able to come up with a low-cost alternative to the special apparatuses needed to truly measure body fat percentage.

1. Make a plot of body.fat vs abdomen. What kind of relationship is there between these two variables?

> attach(fat)

> plot(abdomen, body.fat)



**We can observe a positive linear association between the body fat and abdomen. Also, there is positive correlation.**

1. Fit a least squares regression line using body.fat as the response (y) and abdomen as the only predictor (x). What is the equation for the best fitting line?

> reg2 <- lm(body.fat ~ abdomen); reg2;

Call:

lm(formula = body.fat ~ abdomen)

Coefficients:

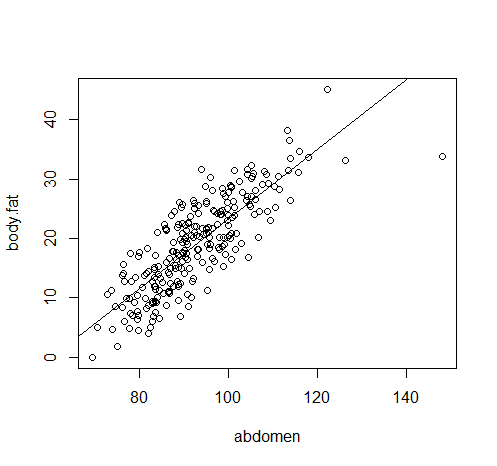
(Intercept) abdomen

-35.1966 0.5849

**The fitted line is: y = -35.1966 + 0.5849 x**

**y := body.fat; x := abdomen.**

> abline(a = reg2$coef[1], b = reg2$coef[2])



1. In a test to determine if there is no linear association between body.fat and abdomen, what would value of the test statistic be? Would you reject the null of no linear association in this case?

> summary(reg2)

Call:

lm(formula = body.fat ~ abdomen)

Residuals:

Min 1Q Median 3Q Max

-17.6257 -3.4672 0.0111 3.1415 11.9754

Coefficients:

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

(Intercept) -35.19661 2.46229 -14.29 <2e-16 \*\*\*

abdomen 0.58489 0.02643 22.13 <2e-16 \*\*\*

---

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

Residual standard error: 4.514 on 250 degrees of freedom

Multiple R-squared: 0.6621, Adjusted R-squared: 0.6608

F-statistic: 489.9 on 1 and 250 DF, p-value: < 2.2e-16

**Test: Ho : B1 = 0 Ha : B1 <> 0**

**From the data obtained on R: TS = t = 0.02643**

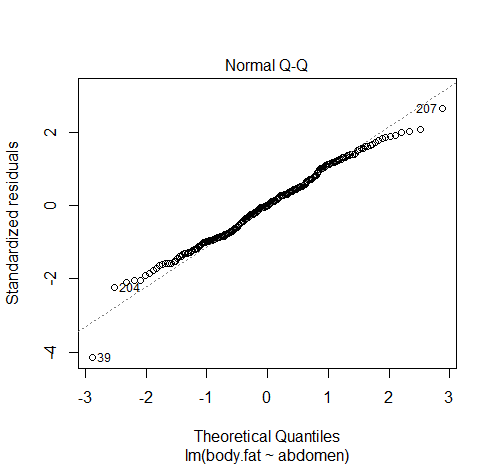
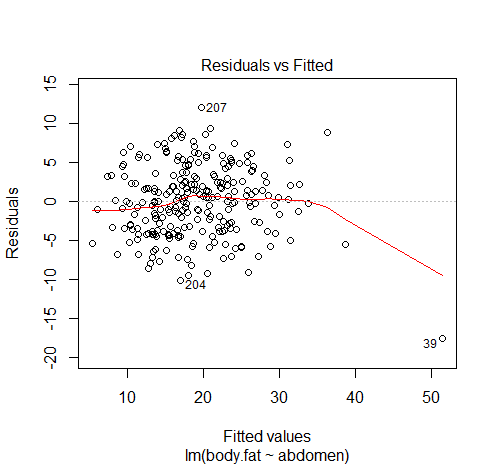
**With a p-value less than 2e-16**

**In this case we reject Ho.**

**Interpretation: At a significance level of 0.05 I’m there’s not enough to suggest that there’s no linear association between the body.fat and abdomen.**

1. Display the residual vs. fitted plot and the normal q-q plot of the residuals. Comment on whether you believe that the model assumptions are appropriate in this case.

> plot(reg2)



**The assumptions that the errors are iid are appropriate in the case because we can observe independency by looking at the random distribution of the residuals on the residual vs. fitted plot (due to no clear pattern). And, the q-q plot portrays a normality of the residuals with little deviations at the tails.**

1. Consider fitting the model where body.fat is the response and abdomen and age are the predictors.
   1. Find the equation for the least squares regression line.

> reg5 <- lm(body.fat ~ abdomen + age);reg5

Call:

lm(formula = body.fat ~ abdomen + age)

Coefficients:

(Intercept) abdomen age

-36.51507 0.56710 0.06605

**The equation is: y = -36.51507 + 0.56710\*x1 + 0.06605\*x2**

**y := body.fat; x1 := abdomen; x2 := age**

* 1. Find a 95% confidence interval for the average body fat percentage of all 60-year old men who have a 105cm abdomen in diameter.

> predict(reg5, newdata = data.frame(abdomen = 105, age = 60), level = 0.95, interval = "confidence")

fit lwr upr

1 26.99392 25.99695 27.99088

**Interpretation: I’m 95% confident that the mean body fat percentage of all 60-year-old men who have a 150cm abdomen in diameter is between 25.99% and 27.99%.**

* 1. Suppose a 50-year old man has a 90 cm abdomen in diameter. Find a 95% prediction interval for his body fat percentage.

predict(reg5, newdata = data.frame(abdomen = 90, age = 50), level = 0.95, interval = "prediction")

fit lwr upr

1 17.82685 9.040651 26.61304

**Interpretation: With 95% confidence I predict that the body fat percentage for a 50-year-old man who have a 90cm abdomen in diameter is between 9.04% and 26.61%.**