hw_6

Andrew Mikolinski

2023-11-16

Question 1a

```
d = data.frame(x <- c(110.5, 105.4, 118.1, 104.5, 93.6, 84.1, 77.8, 75.6),
               y <- c(5.755, 5.939, 6.010, 6.545, 6.730, 6.750, 6.899, 7.862))
fit_d <- lm(y ~ x, data=d)
coefficients <- coef(fit_d)
beta1_hat <- coefficients[2]
print(beta1_hat)
```

```
##           x
## -0.03717469
```

since the β value is negative, it means that as the plant height increases, the grain yield is going to decrease.

Question 1b

```
anova(fit_d)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## x          1 2.42357 2.42357  18.455 0.005116 **
## Residuals  6 0.78794 0.13132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
t.test(x, y, paired = TRUE, conf.level = .95)
```

```
##
##  Paired t-test
##
## data:  x and y
## t = 15.441, df = 7, p-value = 1.153e-06
```

```
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   75.91127 103.36623
## sample estimates:
## mean difference
##        89.63875
```

The ANOVA resulted in a very high F value and the t test yielded a low p value indicating that the results are significant

Question 1c

```r
qt(0.025, 6)
```

```
## [1] -2.446912
```

```r
summary(fit_d)
```

```
##
## Call:
## lm(formula = y ~ x, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34626 -0.27605 -0.09448  0.27023  0.53495
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.137455   0.842265  12.036    2e-05 ***
## x           -0.037175   0.008653  -4.296  0.00512 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3624 on 6 degrees of freedom
## Multiple R-squared:  0.7547, Adjusted R-squared:  0.7138
## F-statistic: 18.46 on 1 and 6 DF,  p-value: 0.005116
```

```r
confint(fit_d, alpha = 0.025)
```

```
##                   2.5 %      97.5 %
## (Intercept)  8.07650745 12.19840320
## x           -0.05834895 -0.01600043
```

The confidence intervals that I got by hand are (8.077, 12.198) and (-0.0586, -0.0158). The confidence interval for &]$beta\ 0$ are very accurate with the R output, the values calculated for &]$beta\ 1$ are not quite as accurate but this may be a rounding error on my part.The equations I used were: $\beta_0 \pm$ -2.447(0.842) where $\beta_0 =$ 10.13746 and $\beta_1 \pm$ -2.447(0.842) where $\beta_1 =$ -0.0372

Question 1d

```
anova(fit_d)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## x          1 2.42357 2.42357  18.455 0.005116 **
## Residuals  6 0.78794 0.13132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coef(fit_d)["x"]
```

```
##           x
## -0.03717469
```

```
coef(fit_d)[1]
```

```
## (Intercept)
##    10.13746
```

The raw residuals are 6 degrees of freedom, 0.78794 SS, and 0.13132 MS. The linear equation is $\hat{y} =$ 10.13746 - 0.03717X.

Question 1e

```
sigma_hat <- sigma(fit_d)
sigma_hat
```

```
## [1] 0.3623848
```

Question 1f

```
mu_hat <- predict(fit_d, newdata = data.frame(x = 100), interval = "confidence", level = 0.95)
mu_hat
```

```
##        fit      lwr      upr
## 1 6.419986 6.096321 6.743651
```

Question 1g

```r
prediction <- predict(fit_d, newdata = data.frame(x = 100), interval = "prediction", level = 0.95)
print(prediction)
```

```
##        fit      lwr      upr
## 1 6.419986 5.476038 7.363934
```

The fit is the same between, but the 95% interval is wider for the answer to question 1g

Question 1h

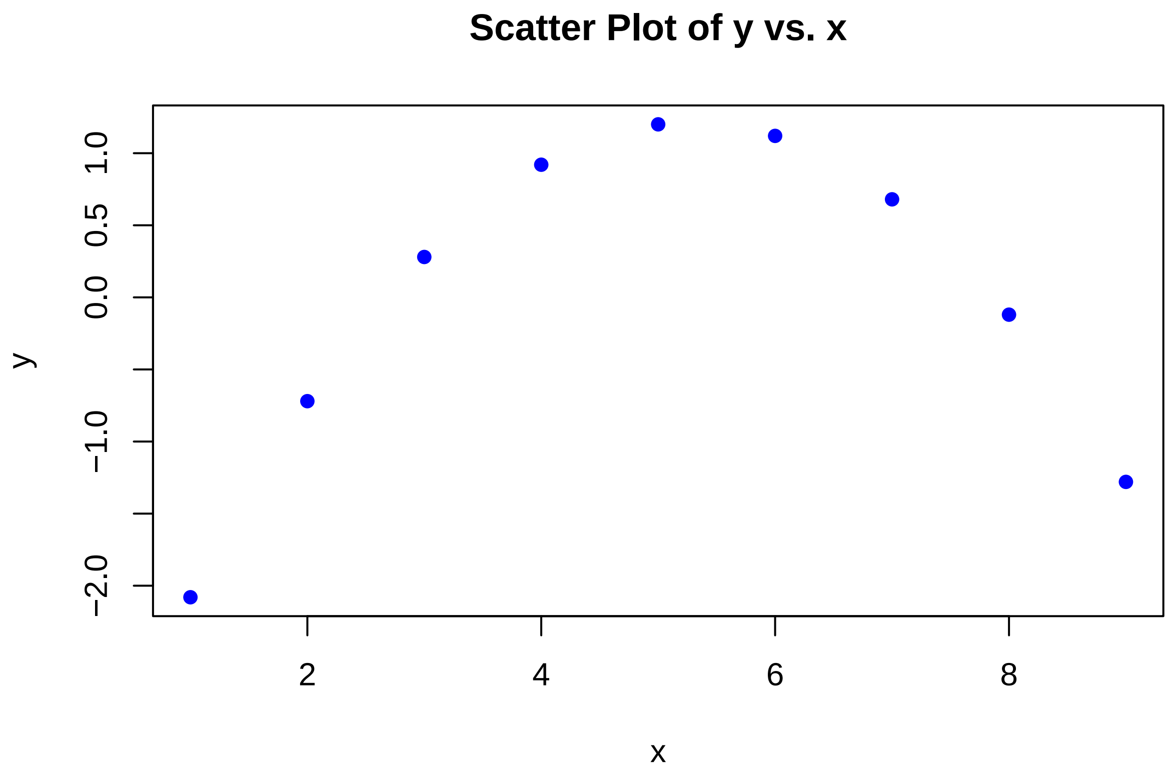
```r
summary_stats <- summary(fit_d)
r_squared <- summary_stats$r.squared
print(r_squared)
```

```
## [1] 0.7546518
```

This value for the coefficient of determination R^2 means that the model explains most of the variance in the response data but not all of it

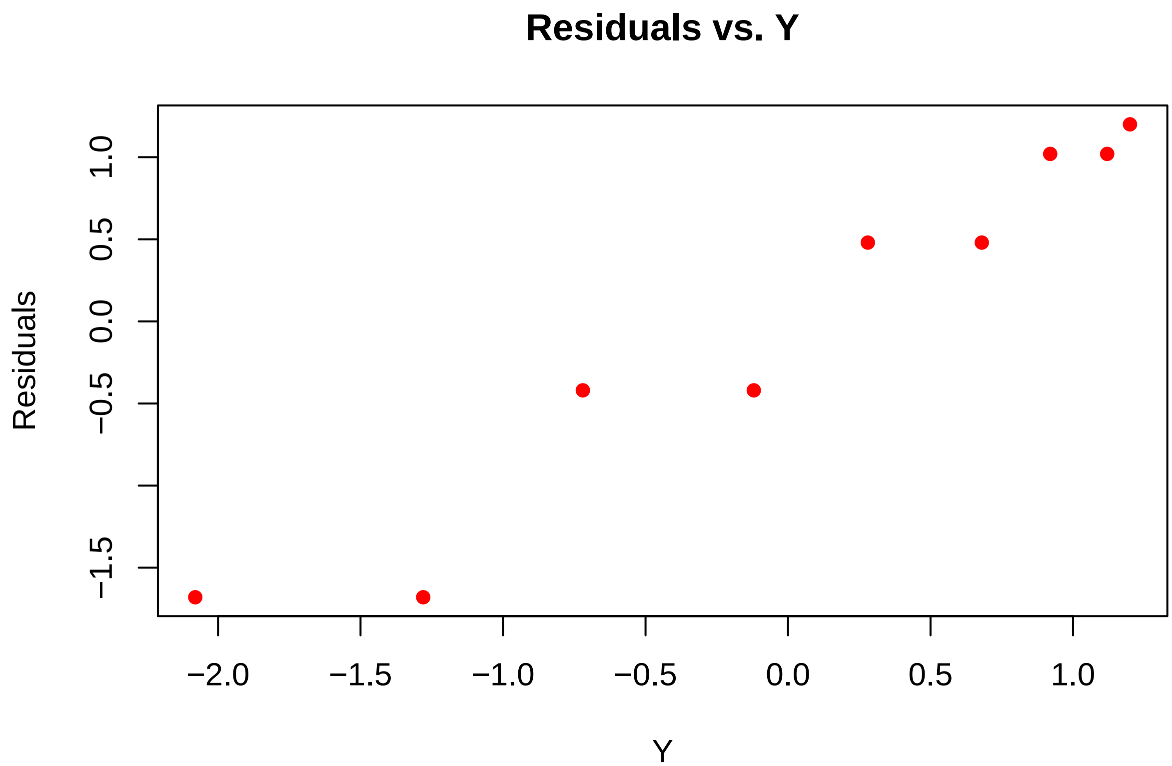
#Question 2a

```r
X = c(1, 2, 3, 4, 5, 6, 7, 8, 9)
Y = c(-2.08, -0.72, 0.28, 0.92, 1.20, 1.12, 0.68, -0.12, -1.28)
plot(X, Y, pch = 16, col = "blue", main = "Scatter Plot of y vs. x", xlab = "x",
     ylab = "y")
```



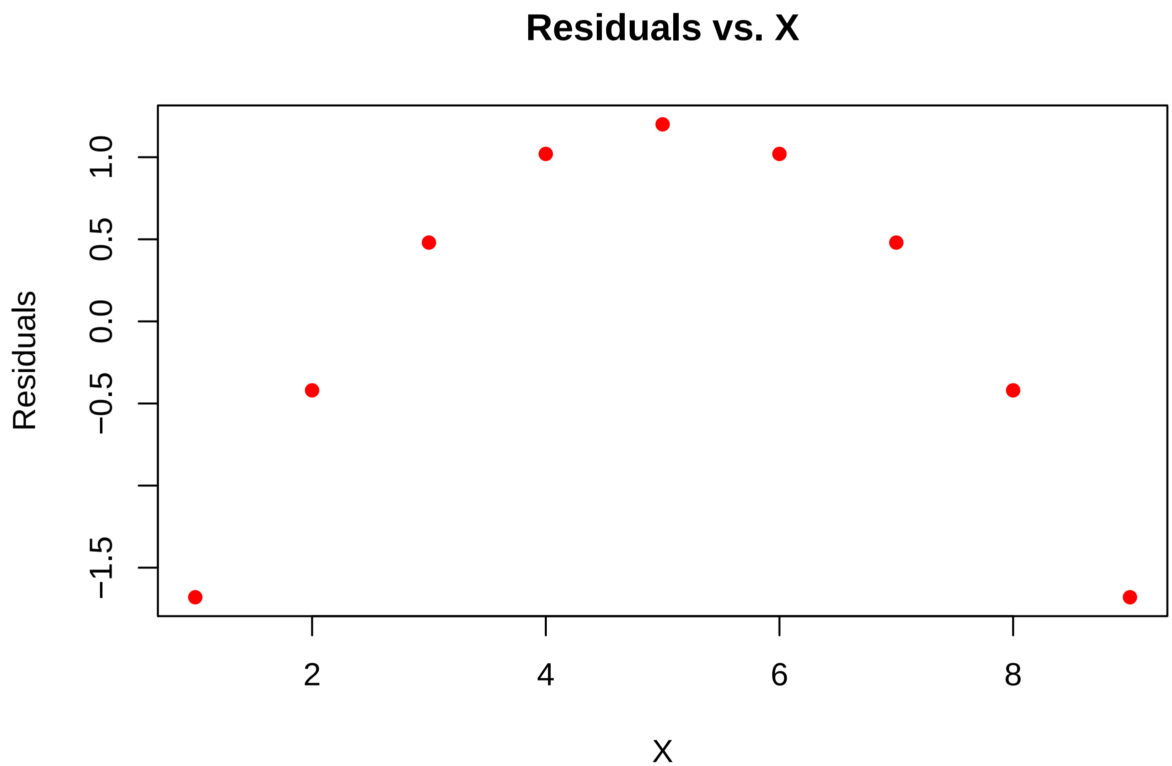
Question 2b

```
model <- lm(Y ~ X)
residuals1 <- residuals(model)
plot(Y, residuals1, pch = 16, col = "red", main = "Residuals vs. Y", xlab = "Y",
     ylab = "Residuals")
```



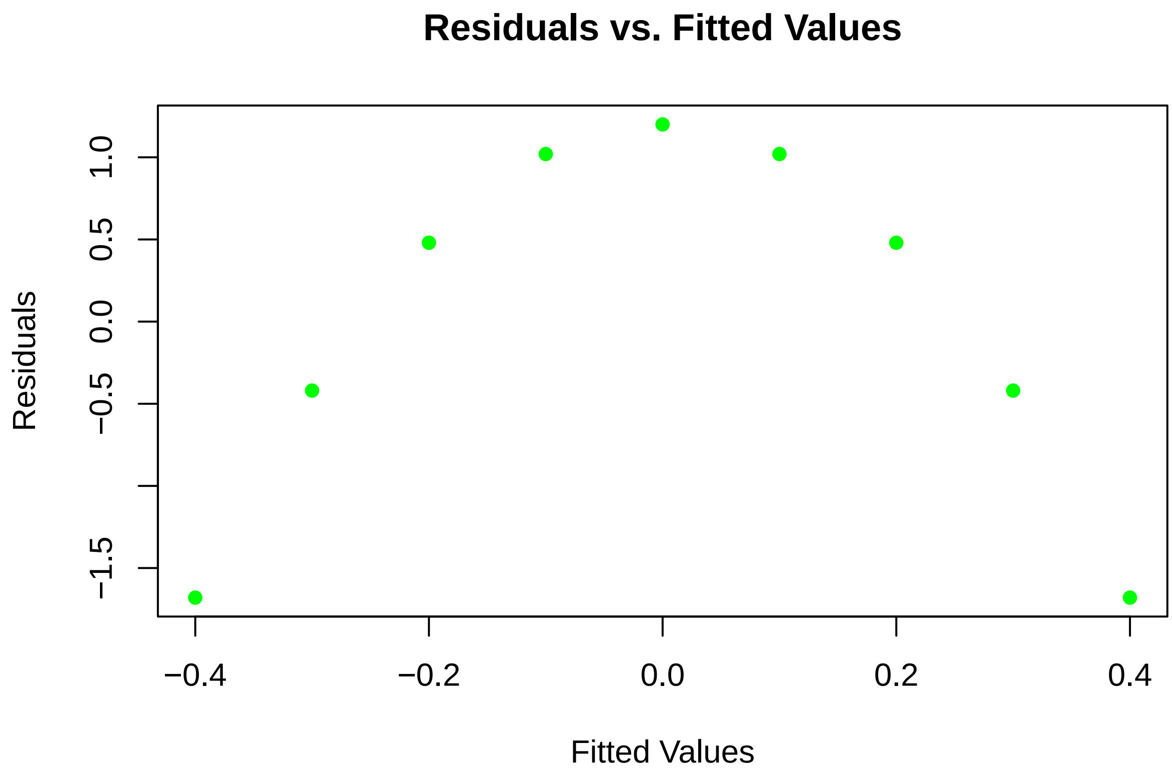
Question 2c

```
plot(X, residuals1, pch = 16, col = "red", main = "Residuals vs. X",
     xlab = "X", ylab = "Residuals")
```



Question 2d

```
fitted_values <- fitted(model)
plot(fitted_values, residuals1, pch = 16, col = "green",
     main = "Residuals vs. Fitted Values",
     xlab = "Fitted Values", ylab = "Residuals")
```



The only real difference in the data between $\hat{Y}$ and X is the scale of the data but the overall pattern is the same.The plot of Y vs residuals is a better indicator of the lack of fit because any pattern in the residuals is more closely related to the actual observed values.