

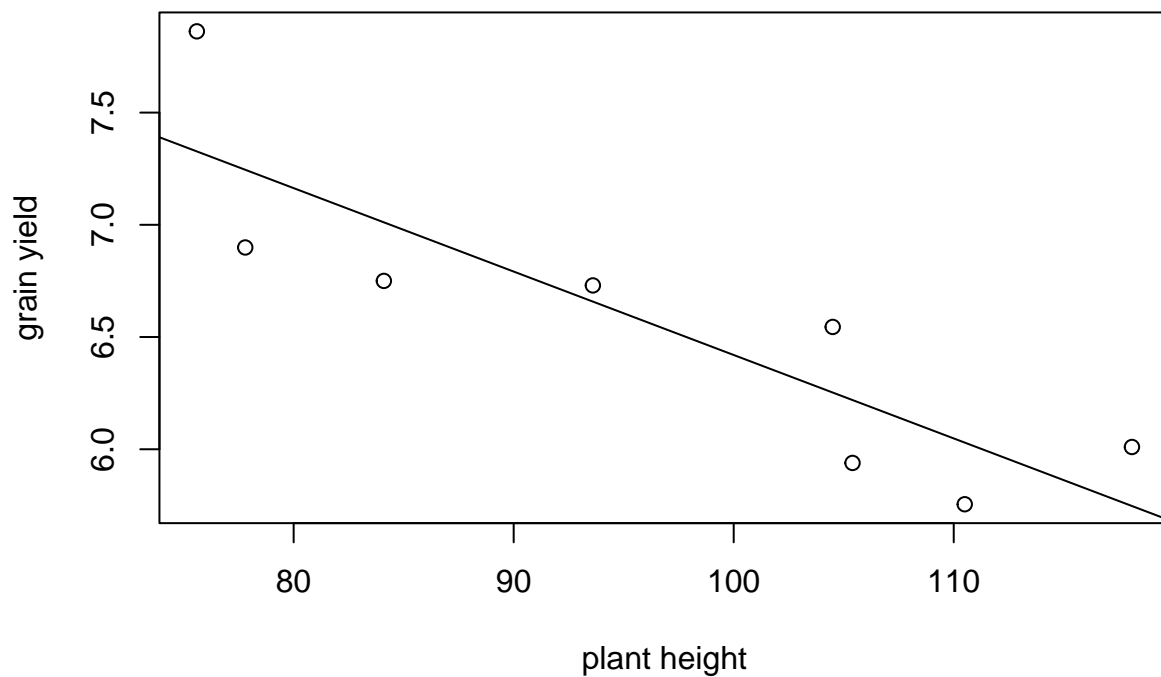
# Homework 6

Kirti Agrawal

14/11/2021

Answer 1 a

```
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))  
plot(d$x, d$y, xlab = "plant height", ylab = "grain yield")  
fit_d = lm(y ~ x, data = d)  
abline(fit_d)
```



“Using the abline command we got the least square regression line and from the plot we can identify the curve does not lie within the points of the graph. Hence, this is not a good fit.”

Answer1 b “ $H_0: B_1$  is equal to 0, using F test(ANOVA)”

```
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
```

“F statistics = 18.46 on 1 and 6 degree of freedom. The relationship between the plant height and grain yield is not zero.  $H_0: B_1$  is not equal to 0, using T test”

```
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
```

“T value = 12.036 & T statistics = -4.296 Same conclusion as the F test.”

Answer1 c “95% confidence interval for the intercept  $B_0$   $B_0^{\wedge} = 10.137455$   $t = 2.447$   $s.e.(B_0^{\wedge}) = 0.842265$   
95% confidence interval for the intercept  $B_0 = 10.137455 + 2.447 * 0.842265 = 12.223745405$  95% confidence  
interval for the intercept  $B_0 = 10.137455 - 2.447 * 0.842265 = 8.051164595$ ”

```
confint(fit_d)
```

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

Answer1d

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

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

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

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

“linear regression line for this dataset is  $\hat{Y} = 10.13746 + (-0.03717469)X$  Raw residual = 0.3624”

Answer1 e “estimate  $\sigma^2$  of the error variance  $\sigma^2 = 0.13132$ ”

Answer 1 f

```
M = predict(fit_d, newdata = data.frame(x = 100), interval = "confidence")
print(M)
```

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

Answer 1g

```
N = predict(fit_d, newdata = data.frame(x = 100), interval = "prediction")
print(N)
```

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

“Prediction interval provided the wider range as compared to the confidence interval.”

Answer 1h

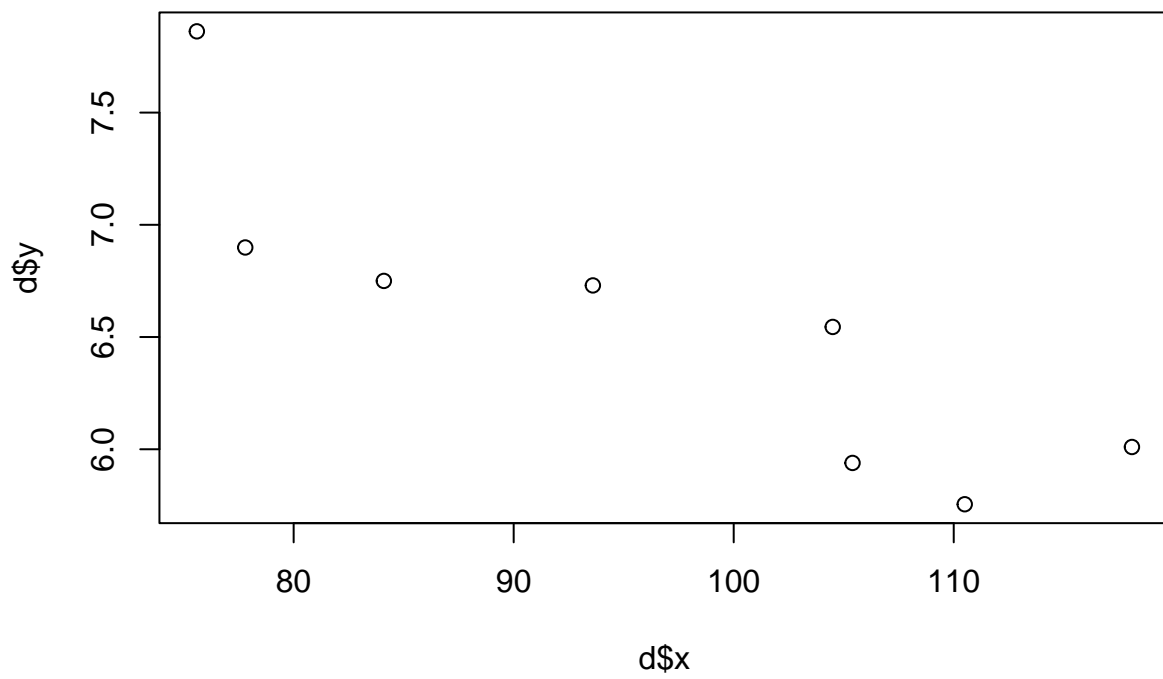
```
x = summary(fit_d)
print(x$r.squared)
```

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

“ $R^2$  is 0.75, it means 75% of the variation in the grain yield is explained by the plant height.”

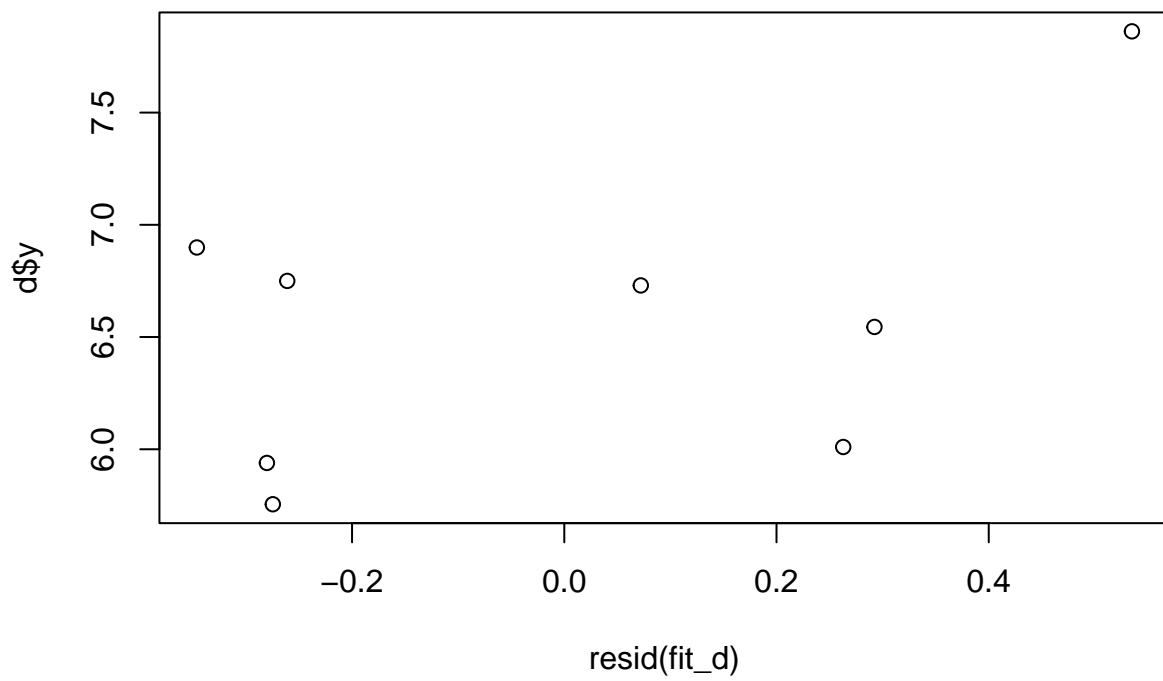
Answer2a

```
df = data.frame(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(d$x, d$y)
```



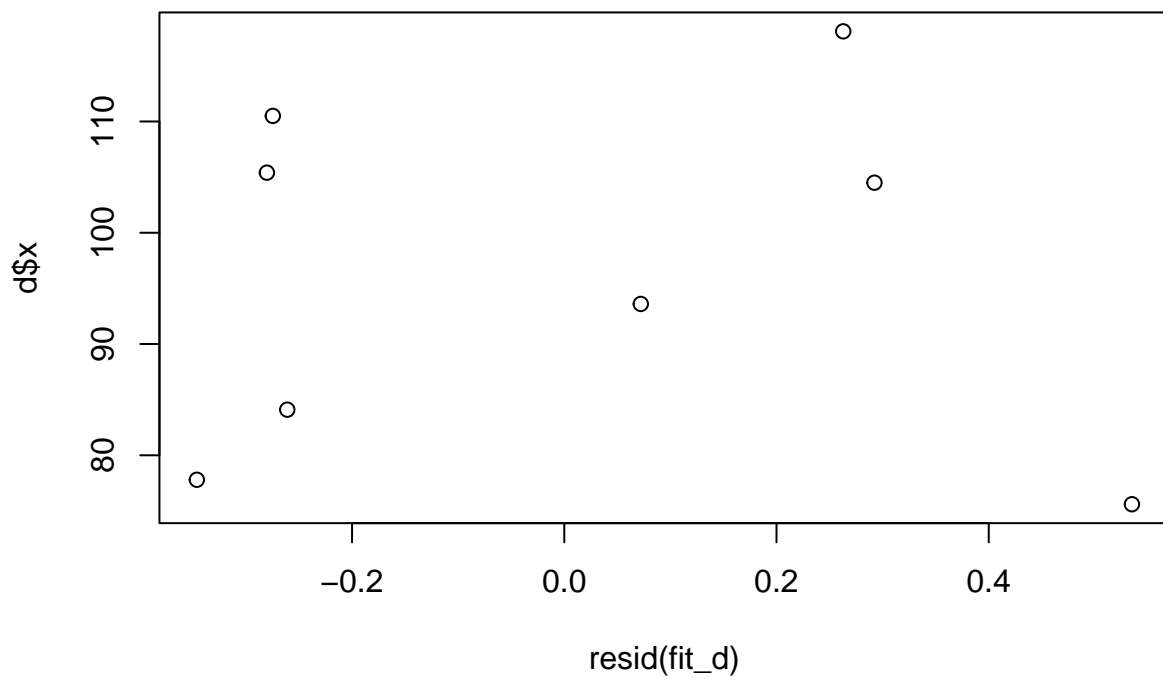
Answer2b

```
plot(resid(fit_d), d$y)
```



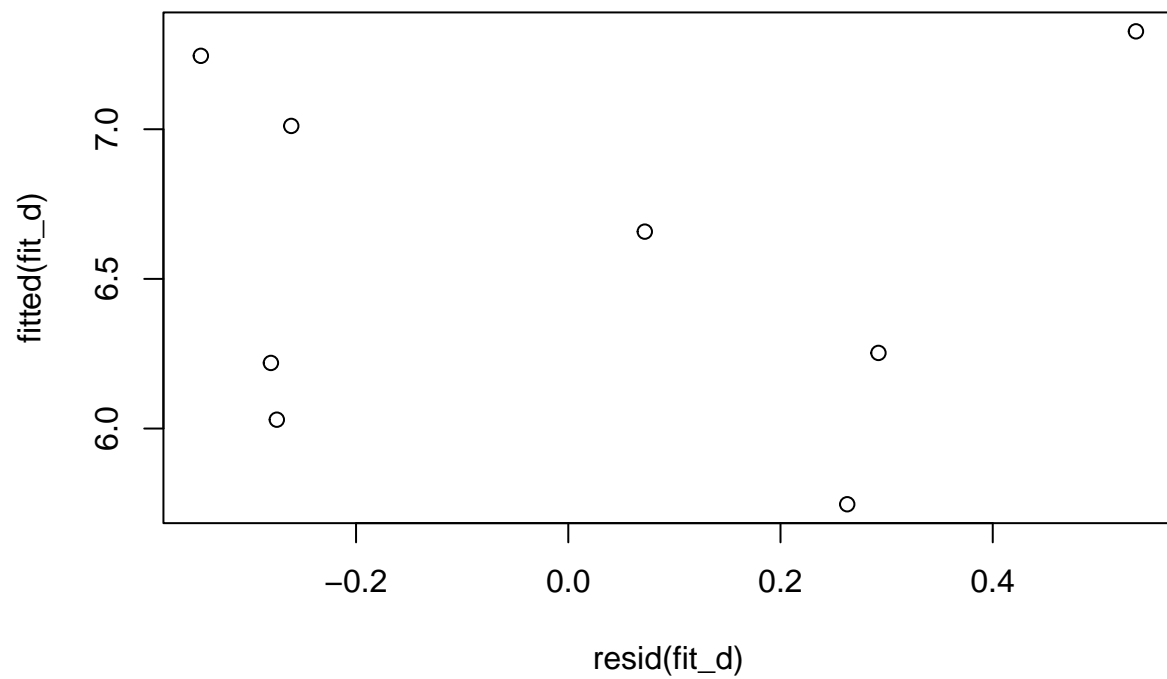
Answer2c

```
plot(resid(fit_d), d$x)
```



Answer2d

```
plot(resid(fit_d), fitted(fit_d))
```



Answer2e “c & d is a meaningful fit, as the values are scattered randomly and no clusters of values found. And b & d have similar patterns and are clustered”