Homework\_4

Sohail Shaikh | Tanmay Khairnar | Samarth Sathe | Ashutosh Lonkar | Aishwarya Jawalkar

2023-11-14

df = read.csv("E:\\Linder\_college\\Linear Regression\\dataset\\alumni.csv")  
  
X1 = df$percent\_of\_classes\_under\_20  
X2 = df$student\_faculty\_ratio  
X3 = df$private  
Y = df$alumni\_giving\_rate  
  
head(df)

## school percent\_of\_classes\_under\_20  
## 1 Boston College 39  
## 2 Brandeis University 68  
## 3 Brown University 60  
## 4 California Institute of Technology 65  
## 5 Carnegie Mellon University 67  
## 6 Case Western Reserve Univ. 52  
## student\_faculty\_ratio alumni\_giving\_rate private  
## 1 13 25 1  
## 2 8 33 1  
## 3 8 40 1  
## 4 3 46 1  
## 5 10 28 1  
## 6 8 31 1

## (a) What is the estimated coefficient for Private or Public School X3 ? Was this estimate significance at the α=0.05 level? Clearly write out the null and alternative hypotheses, observed t-statistic, and p-value

model = lm(Y ~ X1 + X2 + X3, data = df)  
  
summary(model)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.757 -6.320 -2.273 5.152 25.669   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.78364 13.67220 2.690 0.01005 \*   
## X1 0.07725 0.17873 0.432 0.66768   
## X2 -1.39835 0.51075 -2.738 0.00889 \*\*  
## X3 6.28534 5.35633 1.173 0.24693   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.06 on 44 degrees of freedom  
## Multiple R-squared: 0.5747, Adjusted R-squared: 0.5457   
## F-statistic: 19.81 on 3 and 44 DF, p-value: 2.818e-08

The estimated coefficient for Private or Public school X3 is 6.285. We get the p value to be 0.24693 and observed t statistic (t value) to be 1.173

Null Hypothesis: The coefficient B3 (Beta3) is zero. Alternate hypothesis: The coefficient B3(Beta 3)is not zero.

For coefficient Beta 3 (B3) we do not reject the Null hypothesis as its p value is 0.24693 which is greater than 0.05. Hence there is insufficient evidence to conclude that the predictor variable has a statistically significant relationship with the response variable.

## (b) Was this the model significance at the α=0.05 level? Clearly write out the null and alternative hypotheses of this model, observed F-statistic, and p-value.

summary(model)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.757 -6.320 -2.273 5.152 25.669   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.78364 13.67220 2.690 0.01005 \*   
## X1 0.07725 0.17873 0.432 0.66768   
## X2 -1.39835 0.51075 -2.738 0.00889 \*\*  
## X3 6.28534 5.35633 1.173 0.24693   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.06 on 44 degrees of freedom  
## Multiple R-squared: 0.5747, Adjusted R-squared: 0.5457   
## F-statistic: 19.81 on 3 and 44 DF, p-value: 2.818e-08

NUll Hypothesis: All the coefficent B1 (Beta 1) ,B2 (Beta 2) and B3 (Beta 3) are zero

Alternate hypothesis: Atleast one of the coefficient is not zero

We get the F-statistic value as 19.81 on 3 and 44 DF and p value = 2.818e-08

Since the p value is less than 0.05 we reject the null hypothesis at significance level alpha = 0.05. We conclude that at least one of the predictor variables in the model has a significant relationship with the response variable.

## (c) Repeat part a and b now. Add the interaction between Student/Faculty Ratio (X2) and Private or Public School(X3). What is the estimated slopes for X2 now? Interpret the results of with the interaction term.

model2 <- lm(Y ~ X1 + X2 + X3 + X2:X3, data = df)  
  
summary(model2)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3 + X2:X3, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.021 -6.275 -2.207 6.439 23.455   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 25.572987 15.388370 1.662 0.1038   
## X1 0.005511 0.182487 0.030 0.9760   
## X2 -0.584364 0.737736 -0.792 0.4326   
## X3 27.907119 15.265512 1.828 0.0745 .  
## X2:X3 -1.477699 0.978896 -1.510 0.1385   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.931 on 43 degrees of freedom  
## Multiple R-squared: 0.5961, Adjusted R-squared: 0.5585   
## F-statistic: 15.86 on 4 and 43 DF, p-value: 4.745e-08

## Repeat part a and b

# Part (a)

The estimated coefficient for Private or Public School X3 is 27.907.

NUll Hypothesis: The coefficent B3 (Beta 3) is zero.

Alternate hypothesis: The coefficent B3 (Beta 3) is not zero

We get the F-statistic value as 15.86 on 4 and 43 DF, p-value: 0.0745

Since the p value is greater than 0.05 we donot reject the null hypothesis at significance level alpha = 0.05. Hence there is insufficient evidence to conclude that the predictor variable X3 has a statistically significant relationship with the response variable

## Part (b)

NUll Hypothesis: All the coefficent B1 (Beta 1) ,B2 (Beta 2) ,B3 (Beta 3) and B4 (Beta 4) are zero

Alternate hypothesis: Atleast one of the coefficient is not zero

We get the F-statistic value as 15.86 on 4 and 43 DF, p-value: 4.745e-08

Since the p value is less than 0.05 we reject the null hypothesis at significance level alpha = 0.05. We conclude that at least one of the predictor variables in the model has a significant relationship with the response variable.

## (d) Compare the model with and without the interaction term. Which model would you choose based on the data? Please explain your choice.

#Model without interation term  
summary(model)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.757 -6.320 -2.273 5.152 25.669   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.78364 13.67220 2.690 0.01005 \*   
## X1 0.07725 0.17873 0.432 0.66768   
## X2 -1.39835 0.51075 -2.738 0.00889 \*\*  
## X3 6.28534 5.35633 1.173 0.24693   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.06 on 44 degrees of freedom  
## Multiple R-squared: 0.5747, Adjusted R-squared: 0.5457   
## F-statistic: 19.81 on 3 and 44 DF, p-value: 2.818e-08

#Model with interaction term  
summary(model2)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3 + X2:X3, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.021 -6.275 -2.207 6.439 23.455   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 25.572987 15.388370 1.662 0.1038   
## X1 0.005511 0.182487 0.030 0.9760   
## X2 -0.584364 0.737736 -0.792 0.4326   
## X3 27.907119 15.265512 1.828 0.0745 .  
## X2:X3 -1.477699 0.978896 -1.510 0.1385   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.931 on 43 degrees of freedom  
## Multiple R-squared: 0.5961, Adjusted R-squared: 0.5585   
## F-statistic: 15.86 on 4 and 43 DF, p-value: 4.745e-08

Inorder to pick a model we would check the statistical significance, model fit for both the cases.

Model without Interaction:

We have the Adjusted R squared to be 0.5467 and F statistice to be 19.81 on 3 and 44 DF, p-value: 2.818e-08

Model with Interaction:

We have the Adjusted R squared to be 0.5585 and F statistice to be 15.86 on 4 and 43 DF, p-value: 4.745e-08

From above results we see that model with interaction has slightly higher Adjusted R squared indicating that model with interaction explains slighltly higher proportion of variance in data. And both the models have lower p values (< 0.05) suggesting both models are statistically significant.

However In the model with interaction the interaction term (X2\*X3) is not statistically significant, suggesting that the interaction term may not be essential for explaining the variation in the response variable

Hence given that the model without interaction exhibits a similar adjusted R-squared, a notable influence of X2 on the response variable, and a simple structure, it could be considered a better choice as compared to the model with interaction.

## (e) What is the predicted alumni giving rate for an observation with (X1=40, X2=5, X3=1) ?

new\_data <- data.frame(X1 = c(40), X2 = c(5), X3 = c(1))  
predictions <- predict(model, newdata = new\_data)  
  
predictions

## 1   
## 39.16739

For observation X1 = 40, X2 = 5 and X3 = 1, the model predicts the alumni giving rate of 39.167 %.

**##Q.2) Simulation Study (Simple Linear Regression). Assume mean function E(Y|X)=10+5X−2X2**

**a. Generate data with X1∼N(μ=3,σ=0.5), sample size n=200, and error term ϵ∼N(μ=0,σ=0.5).**

set.seed(7052)  
n <- 200  
x1 <- rnorm(n, mean = 3, sd = 0.5)  
  
error <- rnorm(n, mean = 0, sd = 0.5) # e ~ N(0, sigma = 0.5)  
  
  
y1 <- 10 + (5\*x1) - (2\*(x1^2)) + error # equivalent

**b. Fit a simple linear regression using just X. What is the estimated regression equation? Please conduct model estimation, inference, and residual diagnostics. What do you conclude?**

# Part b: Fit Simple Linear Regression  
fit <- lm(y1 ~ x1)

plot(fit)

A graph with black dots and red line

Description automatically generatedA graph of a graph with numbers and a line

Description automatically generatedA graph of a number of dots

Description automatically generated with medium confidenceA diagram of a number of black dots

Description automatically generated with medium confidence

# Print summary of the linear regression model  
summary(fit)

##   
## Call:  
## lm(formula = y1 ~ x1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7767 -0.4085 0.1226 0.5522 1.3576   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 28.9036 0.3686 78.42 <0.0000000000000002 \*\*\*  
## x1 -7.4517 0.1179 -63.22 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.839 on 198 degrees of freedom  
## Multiple R-squared: 0.9528, Adjusted R-squared: 0.9526   
## F-statistic: 3997 on 1 and 198 DF, p-value: < 0.00000000000000022

# Extract coefficients and standard errors  
coefficients <- coef(fit)  
std\_errors <- summary(fit)$coef[, "Std. Error"]

# Observed t-statistic and p-value for the slope (β1)  
t\_stat\_x1 <- as.numeric(coefficients["x1"] / std\_errors["x1"])  
p\_value\_x1 <- 2 \* (1 - pt(abs(t\_stat\_x1), df = n - 2))  
  
  
# Interpretation of the estimates and test results  
cat("Estimated Prediction Equation: Y = ", round(coefficients["(Intercept)"], 2), " + ", round(coefficients["x1"], 2), " \* X1\n")

## Estimated Prediction Equation: Y = 28.9 + -7.45 \* X1

cat("Estimated Coefficients:\n", coefficients, "\n")

## Estimated Coefficients:  
## 28.90365 -7.451686

cat("Standard Errors:\n", std\_errors, "\n")

## Standard Errors:  
## 0.3685676 0.1178638

cat("t-statistic X1 (β1): ", round(t\_stat\_x1, 2), "\n")

## t-statistic X1 (β1): -63.22

cat("p-value X1 (β1): ", format(p\_value\_x1, scientific = TRUE, digits = 2), "\n")

## p-value X1 (β1): 0e+00

# adjusted R^2 value extraction  
r\_sqrd1 <- round(summary(fit)$adj.r.squared\*100,2)

Since the p-value is much smaller than the typical significance level of 0.05, we can confidently conclude that the estimated coefficient for X is highly significant. This indicates that the predictor variable X has a significant effect on the response variable Y.

The R-squared value of 95.26% accuracy, indicates that approximately 95.26% of the variance in Y can be explained by its linear relationship with X.

Therefore, we can rely on the precision of the estimate provided by the standard error. This model can be considered robust to make predictions based on the given data. The negative coefficient for X suggests that as X increases, Y tends to decrease.

**c. Update the model from part b) by adding a quadratic term. Conduct model estimation, inference, and residual diagnostics. What do you conclude? Does this model seem to fit the data better? Please explain.**

fit2 <- lm(y1 ~ x1 + I(x1^2))

plot(fit2)

A graph with black dots and red line

Description automatically generatedA graph of a line

Description automatically generatedA diagram of a number of values

Description automatically generatedA graph of a number of individuals

Description automatically generated with medium confidence

# Print summary of the linear regression model  
summary(fit2)

##   
## Call:  
## lm(formula = y1 ~ x1 + I(x1^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.3707 -0.3234 0.0025 0.3365 1.1679   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.89785 0.88912 12.257 < 0.0000000000000002 \*\*\*  
## x1 4.61948 0.58351 7.917 0.000000000000174 \*\*\*  
## I(x1^2) -1.96853 0.09455 -20.821 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4701 on 197 degrees of freedom  
## Multiple R-squared: 0.9853, Adjusted R-squared: 0.9851   
## F-statistic: 6581 on 2 and 197 DF, p-value: < 0.00000000000000022

# Extract coefficients and standard errors  
coefficients2 <- coef(fit2)  
std\_errors2 <- summary(fit2)$coef[, "Std. Error"]

# Observed t-statistic and p-value for the slope (β1)  
t\_stat\_x2 <- as.numeric(coefficients2["x1"] / std\_errors2["x1"])  
p\_value\_x2 <- 2 \* (1 - pt(abs(t\_stat\_x2), df = n - 2))  
  
t\_stat\_x3 <- as.numeric(coefficients2["I(x1^2)"] / std\_errors2["I(x1^2)"])  
p\_value\_x3 <- 2 \* (1 - pt(abs(t\_stat\_x3), df = n - 2))  
  
  
  
# Interpretation of the estimates and test results  
cat("Estimated Prediction Equation: Y = ", round(coefficients2["(Intercept)"], 2), " + ", round(coefficients2["x1"], 2), " \* X", " + ", round(coefficients2["I(x1^2)"], 2), " \* (X^2)\n")

## Estimated Prediction Equation: Y = 10.9 + 4.62 \* X + -1.97 \* (X^2)

cat("Estimated Coefficients:\n", coefficients2, "\n")

## Estimated Coefficients:  
## 10.89785 4.619479 -1.968531

cat("Standard Errors:\n", std\_errors2, "\n")

## Standard Errors:  
## 0.8891173 0.5835127 0.09454601

cat("t-statistic X1 (β1): ", round(t\_stat\_x2, 2), "\n")

## t-statistic X1 (β1): 7.92

cat("t-statistic X2 (β2): ", round(t\_stat\_x3, 2), "\n")

## t-statistic X2 (β2): -20.82

cat("p-value X1 (β1): ", format(p\_value\_x2, scientific = TRUE, digits = 2), "\n")

## p-value X1 (β1): 1.7e-13

cat("p-value X2 (β2): ", format(p\_value\_x3, scientific = TRUE, digits = 2), "\n")

## p-value X2 (β2): 0e+00

# adjusted R^2 value extraction  
r\_sqrd2 <- round(summary(fit2)$adj.r.squared\*100,2)

Since the p-value is much smaller than the typical significance level of 0.05, we can confidently conclude that the estimated coefficient for X and X^2 is highly significant. This indicates that the predictor variables X and X^2 has a significant effect on the response variable Y.

The R-squared value of 98.51% accuracy, indicates that approximately 98.51% of the variance in Y can be explained by its linear as well as quadratic relationship with X. This accuracy is higher than the original model and gives a more accurate representation of the underlying data and its patterns.

Therefore, we can rely on the precision of the estimate provided by the standard error. This updated model with a quadratic term is very robust to make predictions based on the given data. The positive coefficient for X suggests that as X increases, Y also tends to increase. However, the quadratic term has a negative coefficient.

**d. Generate Y data with using E(√Y|X)=10+5X−2X2. Fit data with the model in part c with the quadratic term, i.e., lm(y ~ x + I(x^2)). Does the assumption of constant error variance appear to be violated?**

set.seed(7052)  
sqrt\_y <- 10 + (5\*x1) - (2\*(x1^2)) + error  
  
y2 <- sqrt\_y^2  
  
fit3 <- lm(y2 ~ x1 + I(x1^2))

plot(fit3)

A graph of a graph with numbers and a line

Description automatically generated with medium confidenceA graph with a line

Description automatically generatedA graph of a line graph

Description automatically generated with medium confidenceA graph of a number of numbers and a line

Description automatically generated with medium confidence

# Print summary of the linear regression model  
summary(fit3)

##   
## Call:  
## lm(formula = y2 ~ x1 + I(x1^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.957 -5.824 -1.701 4.415 59.325   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 572.499 19.975 28.66 <0.0000000000000002 \*\*\*  
## x1 -268.942 13.109 -20.52 <0.0000000000000002 \*\*\*  
## I(x1^2) 31.413 2.124 14.79 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.56 on 197 degrees of freedom  
## Multiple R-squared: 0.9356, Adjusted R-squared: 0.935   
## F-statistic: 1432 on 2 and 197 DF, p-value: < 0.00000000000000022

# Extract coefficients and standard errors  
coefficients3 <- coef(fit3)  
std\_errors3 <- summary(fit3)$coef[, "Std. Error"]

# plot  
plot(fit3$fitted.values, resid(fit3),  
 xlab = "Fitted Values", ylab = "Residuals",  
 main = "Residuals vs. Fitted Values")  
abline(h = 0, col = "red", lty = 2)

A graph with dots and lines

Description automatically generated

# Observed t-statistic and p-value for the slope (β1)  
t\_stat\_x4 <- as.numeric(coefficients3["x1"] / std\_errors3["x1"])  
p\_value\_x4 <- 2 \* (1 - pt(abs(t\_stat\_x4), df = n - 2))  
  
t\_stat\_x5 <- as.numeric(coefficients3["I(x1^2)"] / std\_errors3["I(x1^2)"])  
p\_value\_x5 <- 2 \* (1 - pt(abs(t\_stat\_x5), df = n - 2))  
  
  
  
# Interpretation of the estimates and test results  
cat("Estimated Prediction Equation: Y = ", round(coefficients3["(Intercept)"], 2), " + ", round(coefficients3["x1"], 2), " \* X", " + ", round(coefficients3["I(x1^2)"], 2), " \* (X^2)\n")

## Estimated Prediction Equation: Y = 572.5 + -268.94 \* X + 31.41 \* (X^2)

cat("Estimated Coefficients:\n", coefficients3, "\n")

## Estimated Coefficients:  
## 572.4988 -268.9424 31.41328

cat("Standard Errors:\n", std\_errors3, "\n")

## Standard Errors:  
## 19.97466 13.10903 2.124044

cat("t-statistic X1 (β1): ", round(t\_stat\_x4, 2), "\n")

## t-statistic X1 (β1): -20.52

cat("t-statistic X2 (β2): ", round(t\_stat\_x5, 2), "\n")

## t-statistic X2 (β2): 14.79

cat("p-value X1 (β1): ", format(p\_value\_x4, scientific = TRUE, digits = 2), "\n")

## p-value X1 (β1): 0e+00

cat("p-value X2 (β2): ", format(p\_value\_x5, scientific = TRUE, digits = 2), "\n")

## p-value X2 (β2): 0e+00

# adjusted R^2 value extraction  
r\_sqrd3 <- round(summary(fit3)$adj.r.squared\*100,2)  
r\_sqrd3

## [1] 93.5

Since the p-value is much smaller than the typical significance level of 0.05, we can confidently conclude that the estimated coefficient for X and X^2 is highly significant. This indicates that the predictor variables X and X^2 has a significant effect on the response variable √Y.

The R-squared value of 93.5% accuracy, indicates that approximately 95.52% of the variance in Y can be explained by its linear as well as quadratic relationship with X. This accuracy is lower than the original quadratic model using Y.

The positive coefficient for X suggests that as X increases, Y also tends to increase. However, the quadratic term has a negative coefficient.

**e. Now repeat part d but transform Y with MASS::boxcox . What is the lambda value you choose? Does this model seem to fit the data better?**

set.seed(7052)  
  
# finding lambda  
bc <- MASS::boxcox(y2 ~ x1 + I(x1^2))

A graph with numbers and lines

Description automatically generated

(lambda <- bc$x[which.max(bc$y)])

## [1] 0.8686869

The lambda value is 0.8686869.

#fitted model  
fit4 <- lm(y2 ~ x1 + I(x1^2))  
summary(fit4)

##   
## Call:  
## lm(formula = y2 ~ x1 + I(x1^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.957 -5.824 -1.701 4.415 59.325   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 572.499 19.975 28.66 <0.0000000000000002 \*\*\*  
## x1 -268.942 13.109 -20.52 <0.0000000000000002 \*\*\*  
## I(x1^2) 31.413 2.124 14.79 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.56 on 197 degrees of freedom  
## Multiple R-squared: 0.9356, Adjusted R-squared: 0.935   
## F-statistic: 1432 on 2 and 197 DF, p-value: < 0.00000000000000022

plot(fit4$fitted.values, resid(fit4),  
 xlab = "Fitted Values", ylab = "Residuals",  
 main = "Residuals vs. Fitted Values")  
abline(h = 0, col = "red", lty = 2)

A graph with dots and lines

Description automatically generated

# adjusted R^2 value extraction  
r\_sqrd4 <- round(summary(fit4)$adj.r.squared\*100,2)  
r\_sqrd4

## [1] 93.5

Since the p-value is much smaller than the typical significance level of 0.05, we can confidently conclude that the estimated coefficient for X and X^2 is highly significant. This indicates that the predictor variables X and X^2 has a significant effect on the response variable Y.

The R-squared value of 93.5% accuracy, indicates that approximately 93.5% of the variance in Y can be explained by its linear as well as quadratic relationship with X. This accuracy is the same as that seen in part d).

The negative coefficient for X suggests that as X increases, Y tends to decrease. However, the quadratic term has a positive coefficient.

A notebook with writing on it

Description automatically generated

A piece of paper with writing on it

Description automatically generated