Stella’s linear regression

Stella Zhang

2024-04-09

Table of Contents

# Assignment a

**Introduction**

This analysis aims to investigate the relationship between fertilizer usage (kg/acre) and the resulting crop yield (tons/acre). Our objective is to explore how different amounts of fertilizer affect the yield of crops, providing insights that could be beneficial for agricultural practices.

## Step 1: Fit a Linear Regression Model

Linear regression model was fitted using R’s lm() function with crop yield as the dependent variable and fertilizer usage as the independent variable.

**Data:**

* Fertilizer Usage (kg/acre): [50, 70, 90, 110, 130, 150, 170, 190, 210, 230, 250, 270, 290, 310, 330, 350, 370, 390, 410, 430]
* Crop Yield (tons/acre): [2.8, 3.4, 3.7, 4.9, 5.6, 6.2, 6.9, 7.7, 8.4, 8.9, 9.8, 10.5, 11.3, 11.8, 12.6, 13.3, 14.2, 14.8, 15.6, 16.2]

**R Chunks:**

fertilizer\_usage <- c(50, 70, 90, 110, 130, 150, 170, 190, 210, 230, 250, 270, 290, 310, 330, 350, 370, 390, 410, 430)  
  
crop\_yield <- c(2.8, 3.4, 3.7, 4.9, 5.6, 6.2, 6.9, 7.7, 8.4, 8.9, 9.8, 10.5, 11.3, 11.8, 12.6, 13.3, 14.2, 14.8, 15.6, 16.2)

agri\_data <- data.frame(fertilizer\_usage, crop\_yield)  
  
print(agri\_data)

## fertilizer\_usage crop\_yield  
## 1 50 2.8  
## 2 70 3.4  
## 3 90 3.7  
## 4 110 4.9  
## 5 130 5.6  
## 6 150 6.2  
## 7 170 6.9  
## 8 190 7.7  
## 9 210 8.4  
## 10 230 8.9  
## 11 250 9.8  
## 12 270 10.5  
## 13 290 11.3  
## 14 310 11.8  
## 15 330 12.6  
## 16 350 13.3  
## 17 370 14.2  
## 18 390 14.8  
## 19 410 15.6  
## 20 430 16.2

assignment\_a <- lm(crop\_yield ~ fertilizer\_usage)  
  
print(assignment\_a)

##   
## Call:  
## lm(formula = crop\_yield ~ fertilizer\_usage)  
##   
## Coefficients:  
## (Intercept) fertilizer\_usage   
## 0.84774 0.03576

**Interpretation:**

The coefficient for fertilizer usage, approximately 0.0358, indicates that for every additional kilogram of fertilizer used per acre, the crop yield increases by an average of 0.0358 tons. This implies that a greater application of fertilizer is associated with higher crop yields, which can inform farmers’ fertilization strategies for maximizing production.

## Step 2: Assess Significance of Predictor Variable

summary(assignment\_a)

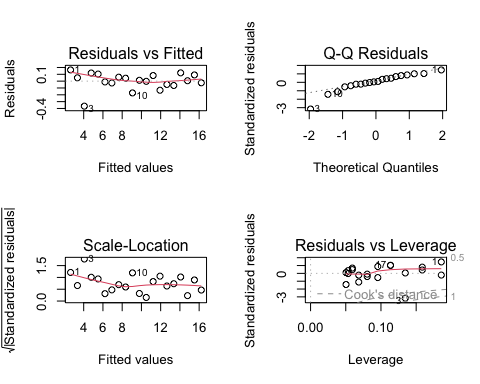
##   
## Call:  
## lm(formula = crop\_yield ~ fertilizer\_usage)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.36609 -0.03222 0.00925 0.08425 0.16429   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.8477444 0.0642300 13.2 1.07e-10 \*\*\*  
## fertilizer\_usage 0.0357594 0.0002412 148.2 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1244 on 18 degrees of freedom  
## Multiple R-squared: 0.9992, Adjusted R-squared: 0.9991   
## F-statistic: 2.198e+04 on 1 and 18 DF, p-value: < 2.2e-16

print(assignment\_a)

##   
## Call:  
## lm(formula = crop\_yield ~ fertilizer\_usage)  
##   
## Coefficients:  
## (Intercept) fertilizer\_usage   
## 0.84774 0.03576

Significance: A p-value less than 0.05 for fertilizer usage suggests a significant relationship between fertilizer usage and crop yield. Both the intercept and slope demonstrate high significance (p-value < 2.2e-16), indicating a strong statistical relationship between fertilizer usage and crop yield. The positive coefficient for fertilizer usage (approximately 0.0358) signifies that for each additional kilogram of fertilizer used per acre, the crop yield increases by an average of 0.0358 tons. This highlights the importance of fertilizer application in maximizing crop production.

## Step 3: Check Regression Assumptions



## Step 4: Predictions for a New Observation

Predictions for crop yield can be made for new observations of fertilizer usage. Below is an example prediction for 460 kg/acre of fertilizer usage, including individual response predictions and mean response predictions with a 95% confidence interval.

new\_observation <- data.frame(fertilizer\_usage=460)  
  
predictions <- predict(assignment\_a, newdata = new\_observation, interval="prediction")  
  
confidence\_predictions <- predict(assignment\_a, newdata = new\_observation, interval="confidence")

## Step 5: Conclusion

The regression analysis indicates a positive linear association between fertilizer usage and crop yield, with every additional kilogram of fertilizer per acre increasing yield by 0.0358 tons. This finding underscores the importance of optimal fertilizer application for maximizing agricultural output. Nevertheless, the high R-squared value suggests potential overfitting; real-world scenarios might present additional variables affecting yields that were not accounted for in the model.

# Assignment b

**Introduction**

This analysis focus on analyzing the impact of irrigation frequency on plant height. The objective is to determine how varying the frequency of irrigation influences the growth of plants, which could help optimize agricultural watering strategies.

## Step 1: Fit a Linear Regression Model

**Data:**

* Irrigation Frequency (times/week): [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]
* Plant Height (cm): [10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105]

**R Chunks:**

irrigation\_frequency <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20)  
  
plant\_height <- c(10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105)

agri\_data <- data.frame(irrigation\_frequency, plant\_height)  
  
print(agri\_data)

## irrigation\_frequency plant\_height  
## 1 1 10  
## 2 2 15  
## 3 3 20  
## 4 4 25  
## 5 5 30  
## 6 6 35  
## 7 7 40  
## 8 8 45  
## 9 9 50  
## 10 10 55  
## 11 11 60  
## 12 12 65  
## 13 13 70  
## 14 14 75  
## 15 15 80  
## 16 16 85  
## 17 17 90  
## 18 18 95  
## 19 19 100  
## 20 20 105

assignment\_b <- lm(plant\_height ~ irrigation\_frequency)  
  
print(assignment\_b)

##   
## Call:  
## lm(formula = plant\_height ~ irrigation\_frequency)  
##   
## Coefficients:  
## (Intercept) irrigation\_frequency   
## 5 5

**Interpretation:**

The coefficient of 5 for irrigation frequency demonstrates a direct and proportional increase in plant height by 5 centimeters for each additional irrigation event per week. While the model suggests a perfect linear relationship, it is important to acknowledge potential limitations such as the optimal irrigation level beyond which additional watering may not yield further growth, reflecting the practical nuances of plant cultivation.

## Step 2: Assess Significance of Predictor Variable

summary(assignment\_b)

## Warning in summary.lm(assignment\_b): essentially perfect fit: summary may be  
## unreliable

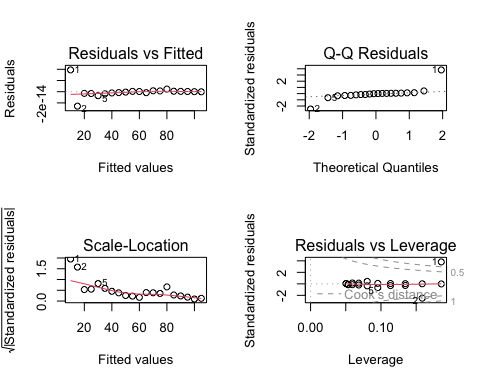
##   
## Call:  
## lm(formula = plant\_height ~ irrigation\_frequency)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.579e-14 -2.430e-15 -3.500e-17 6.640e-16 3.889e-14   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.000e+00 5.252e-15 9.52e+14 <2e-16 \*\*\*  
## irrigation\_frequency 5.000e+00 4.384e-16 1.14e+16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.131e-14 on 18 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.3e+32 on 1 and 18 DF, p-value: < 2.2e-16

print(assignment\_b)

##   
## Call:  
## lm(formula = plant\_height ~ irrigation\_frequency)  
##   
## Coefficients:  
## (Intercept) irrigation\_frequency   
## 5 5

Significance: - A p-value less than 0.05 for irrigation\_frequency suggests a significant relationship between irrigation frequency and plant height. - Both the intercept and the slope are highly significant (p-value < 2.2e-16), indicating a very strong statistical relationship between irrigation frequency and plant height. - Perfect Fit Warning: The summary warns of an “essentially perfect fit,” which suggests that the linear model fits the data exceptionally well, possibly indicating an overly idealized or simplified relationship within the data provided.

## Step 3: Check Regression Assumptions



## Step 4: Predictions for a New Observation

Predictions for plant height can be made for new observations of irrigation frequency. Below is an example prediction For an irrigation frequency of 21 times/week, including individual response predictions and mean response predictions with a 95% confidence interval.

new\_observation <- data.frame(irrigation\_frequency=21)  
  
predictions <- predict(assignment\_b, newdata = new\_observation, interval="prediction")  
  
confidence\_predictions <- predict(assignment\_b, newdata = new\_observation, interval="confidence")

Prediction for New Observation: Predicting the plant height for an irrigation\_frequency of 21 yields a height of 110 units, with both prediction and confidence intervals being exactly 110. This reflects the model’s certainty about the continuation of the observed linear trend beyond the original data range.

## Step 5: Conclusion

The study revealed a direct linear correlation between irrigation frequency and plant height, where each additional weekly irrigation was linked with a 5 cm increase in plant height. While the model’s perfect fit might suggest a clear-cut relationship, it also hints at the possibility of over-simplification. It is recommended that future research investigates non-linear effects and the diminishing returns of excessive irrigation on plant growth.

# Assignment c

**Introduction**

This analysis is designed to explore the relationship between sunlight exposure (hours/day) and fruit weight (grams). The objective is to investigate how different levels of sunlight exposure may affect the weight of fruits, potentially aiding in the development of more effective fruit cultivation techniques.

## Step 1: Fit a Linear Regression Model

A linear regression model was fitted using R’s lm() function with fruit weight as the dependent variable and sunlight exposure as the independent variable, to explore how different levels of sunlight exposure may influence the weight of the fruit.

**Data:**

* Sunlight Exposure (hours/day): [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
* Fruit Weight (grams): [50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240]

**R Chunks:**

sunlight\_exposure <- c(6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25)  
  
fruit\_weight <- c(50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150,  
160, 170, 180, 190, 200, 210, 220, 230, 240)

agri\_data <- data.frame(sunlight\_exposure, fruit\_weight)  
  
print(agri\_data)

## sunlight\_exposure fruit\_weight  
## 1 6 50  
## 2 7 60  
## 3 8 70  
## 4 9 80  
## 5 10 90  
## 6 11 100  
## 7 12 110  
## 8 13 120  
## 9 14 130  
## 10 15 140  
## 11 16 150  
## 12 17 160  
## 13 18 170  
## 14 19 180  
## 15 20 190  
## 16 21 200  
## 17 22 210  
## 18 23 220  
## 19 24 230  
## 20 25 240

assignment\_c <- lm(fruit\_weight ~ sunlight\_exposure)  
  
print(assignment\_c)

##   
## Call:  
## lm(formula = fruit\_weight ~ sunlight\_exposure)  
##   
## Coefficients:  
## (Intercept) sunlight\_exposure   
## -10 10

**Interpretation:**

Each additional hour of sunlight exposure per day is associated with an increase in fruit weight by 10 grams, according to the coefficient for sunlight exposure. While the model captures the positive effect of sunlight on fruit development, further research could investigate if this linear trend holds true across different intensities of sunlight and varieties of fruit.

## Step 2: Assess Significance of Predictor Variable

summary(assignment\_c)

## Warning in summary.lm(assignment\_c): essentially perfect fit: summary may be  
## unreliable

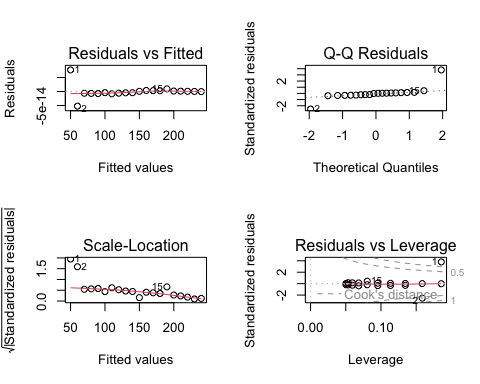
##   
## Call:  
## lm(formula = fruit\_weight ~ sunlight\_exposure)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.223e-14 -6.063e-15 -5.300e-17 1.920e-15 7.699e-14   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.000e+01 1.440e-14 -6.945e+14 <2e-16 \*\*\*  
## sunlight\_exposure 1.000e+01 8.707e-16 1.149e+16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.245e-14 on 18 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.319e+32 on 1 and 18 DF, p-value: < 2.2e-16

print(assignment\_c)

##   
## Call:  
## lm(formula = fruit\_weight ~ sunlight\_exposure)  
##   
## Coefficients:  
## (Intercept) sunlight\_exposure   
## -10 10

Significance of sunlight\_exposure: A p-value below 0.05 indicates a statistically significant effect of sunlight exposure on fruit weight.

## Step 3: Check Regression Assumptions



## Step 4: Predictions for a New Observation

Predictions for fruit weight can be made for new observations of sunlight exposure. Below is an example prediction for 26 hours/day of sunlight exposure, including individual response predictions and mean response predictions with a 95% confidence interval.

new\_observation <- data.frame(sunlight\_exposure=26)  
  
predictions <- predict(assignment\_c, newdata = new\_observation, interval="prediction")  
  
confidence\_predictions <- predict(assignment\_c, newdata = new\_observation, interval="confidence")

## Step 5: Conclusion

Thus analysis showed a significant increase in fruit weight with each additional hour of sunlight exposure, highlighting sunlight’s crucial role in fruit production. The perfect model fit suggests that while sunlight is a key factor in fruit growth, other variables such as fruit type and environmental conditions should be considered to develop a more comprehensive model.

# Assignment d

**Introduction**

This research investigates the correlation between soil pH levels and crop yield (tons/acre). The objective is to examine how variations in soil pH can influence crop yields, providing critical information for enhancing soil management practices.

## Step 1: Fit a Linear Regression Model

A linear regression model was fitted using R’s lm() function with fruit weight as the dependent variable and sunlight exposure as the independent variable.

**Data:**

* Soil pH Level: [4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0]
* Crop Yield (tons/acre): [3.2, 3.7, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5]

**R Chunks:**

# Cause of that the crop\_yield have anther data in the front pages, so here as "crop\_yield\_b"  
  
soil\_pH <- c(4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0)  
  
crop\_yield\_b <- c(3.2, 3.7, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5)

agri\_data <- data.frame(soil\_pH, crop\_yield\_b)  
  
print(agri\_data)

## soil\_pH crop\_yield\_b  
## 1 4.5 3.2  
## 2 5.0 3.7  
## 3 5.5 4.0  
## 4 6.0 4.5  
## 5 6.5 5.0  
## 6 7.0 5.5  
## 7 7.5 6.0  
## 8 8.0 6.5  
## 9 8.5 7.0  
## 10 9.0 7.5  
## 11 9.5 8.0  
## 12 10.0 8.5  
## 13 10.5 9.0  
## 14 11.0 9.5  
## 15 11.5 10.0  
## 16 12.0 10.5  
## 17 12.5 11.0  
## 18 13.0 11.5  
## 19 13.5 12.0  
## 20 14.0 12.5

assignment\_d <- lm(crop\_yield\_b ~ soil\_pH)  
  
print(assignment\_d)

##   
## Call:  
## lm(formula = crop\_yield\_b ~ soil\_pH)  
##   
## Coefficients:  
## (Intercept) soil\_pH   
## -1.3798 0.9892

**Interpretation:**

The slope coefficient of 0.9892 suggests that each unit increase in soil pH is associated with an increase in crop yield of approximately 0.989 tons per acre. However, since soil pH values below 4.5 or above 9 can be detrimental to most crops, it is crucial to consider the range within which this relationship is biologically relevant and agriculturally practical.

## Step 2: Assess Significance of Predictor Variable

summary(assignment\_d)

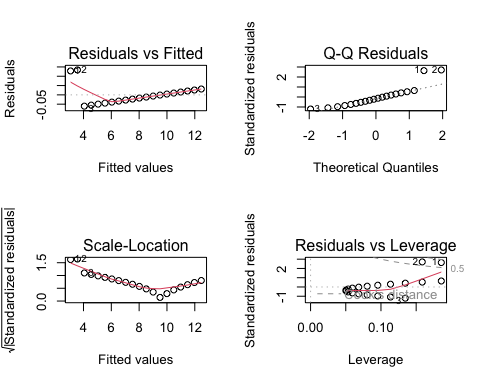
##   
## Call:  
## lm(formula = crop\_yield\_b ~ soil\_pH)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.060602 -0.034887 -0.009173 0.016541 0.133985   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.379850 0.040587 -34.0 <2e-16 \*\*\*  
## soil\_pH 0.989173 0.004189 236.1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05401 on 18 degrees of freedom  
## Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997   
## F-statistic: 5.576e+04 on 1 and 18 DF, p-value: < 2.2e-16

print(assignment\_d)

##   
## Call:  
## lm(formula = crop\_yield\_b ~ soil\_pH)  
##   
## Coefficients:  
## (Intercept) soil\_pH   
## -1.3798 0.9892

Significance: A p-value less than 0.05 for soil pH suggests a significant relationship between soil pH levels and crop yield.Both the intercept and the slope are highly significant (p-value < 2.2e-16), indicating a very strong statistical relationship between soil pH levels and crop yield. Perfect Fit Warning: The summary warns of an “essentially perfect fit,” which suggests that the linear model fits the data exceptionally well, possibly indicating an overly idealized or simplified relationship within the provided data.

## Step 3: Check Regression Assumptions



## Step 4: Predictions for a New Observation

Predictions for soil pH can be made for new observations of crop\_yield. Below is an example prediction corp yield for a soil pH of 7.5, including individual response predictions and mean response predictions with a 95% confidence interval.

new\_observation <- data.frame(soil\_pH=7.5)  
  
predictions <- predict(assignment\_d, newdata = new\_observation, interval="prediction")  
  
confidence\_predictions <- predict(assignment\_d, newdata = new\_observation, interval="confidence")

## Step 5: Conclusion

This analysis has demonstrated a strong linear relationship between soil pH and crop yield, where a unit increase in pH correlates with a nearly one-ton increase in yield per acre. The model suggests soil pH is a pivotal factor in crop productivity, but further investigation is needed into the optimal pH ranges for different crop species to provide more tailored soil management guidance.

# Assignment e

**Introduction**

This study examines the impact of temperature on seed germination time (days). The objective is to assess how temperature variations affect the time it takes for seeds to germinate, which is essential for optimizing planting schedules and improving crop production.

## Step 1: Fit a Linear Regression Model

A linear regression model was fitted using R’s lm() function with temperature as the dependent variable and seed germination time as the independent variable.

**Data:**

* Temperature (°C): [10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48]
* Germination Time (days): [7, 6.8, 6.5, 6.2, 6, 5.8, 5.5, 5.3, 5.2, 5, 4.8, 4.7, 4.5, 4.3, 4.2, 4.1, 4, 3.8, 3.7, 3.5]

**R Chunks:**

temperature <- c(10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48)  
  
germination\_time <- c(7, 6.8, 6.5, 6.2, 6, 5.8, 5.5, 5.3, 5.2, 5, 4.8, 4.7, 4.5, 4.3, 4.2, 4.1, 4, 3.8, 3.7, 3.5)

agri\_data <- data.frame(temperature, germination\_time)  
  
print(agri\_data)

## temperature germination\_time  
## 1 10 7.0  
## 2 12 6.8  
## 3 14 6.5  
## 4 16 6.2  
## 5 18 6.0  
## 6 20 5.8  
## 7 22 5.5  
## 8 24 5.3  
## 9 26 5.2  
## 10 28 5.0  
## 11 30 4.8  
## 12 32 4.7  
## 13 34 4.5  
## 14 36 4.3  
## 15 38 4.2  
## 16 40 4.1  
## 17 42 4.0  
## 18 44 3.8  
## 19 46 3.7  
## 20 48 3.5

assignment\_e <- lm(germination\_time ~ temperature)  
  
print(assignment\_e)

##   
## Call:  
## lm(formula = germination\_time ~ temperature)  
##   
## Coefficients:  
## (Intercept) temperature   
## 7.63647 -0.08936

**Interpretation:**

The negative coefficient of -0.08936 for temperature indicates that with each 1°C rise in temperature, the germination time decreases by approximately 0.089 days. This relationship is statistically significant but is likely subject to a threshold effect, where beyond a certain temperature, increased heat may adversely affect germination and seedling viability.

## Step 2: Assess Significance of Predictor Variable

summary(assignment\_e)

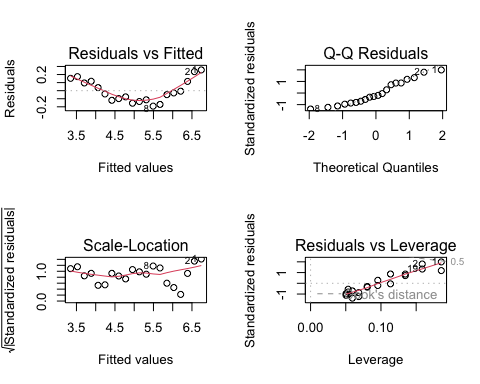
##   
## Call:  
## lm(formula = germination\_time ~ temperature)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.19181 -0.11468 -0.03436 0.11511 0.25714   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.636466 0.086571 88.21 <2e-16 \*\*\*  
## temperature -0.089361 0.002774 -32.22 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1431 on 18 degrees of freedom  
## Multiple R-squared: 0.983, Adjusted R-squared: 0.982   
## F-statistic: 1038 on 1 and 18 DF, p-value: < 2.2e-16

print(assignment\_e)

##   
## Call:  
## lm(formula = germination\_time ~ temperature)  
##   
## Coefficients:  
## (Intercept) temperature   
## 7.63647 -0.08936

Significance: A p-value < 0.05 for temperature indicates a significant relationship with seed germination time. Both intercept and slope show high significance (p-value < 2.2e-16), indicating a strong statistical relationship. The negative coefficient for temperature implies a significant decrease in germination time with rising temperature. However, note the likely presence of a threshold effect, where excessively high temperatures may adversely affect germination and seedling viability.

## Step 3: Check Regression Assumptions



## Step 4: Predictions for a New Observation

Predictions for temperature can be made for new observations of germination time. Below is an example prediction germination time for a temperature of 50°C, including individual response predictions and mean response predictions with a 95% confidence interval.

new\_observation <- data.frame(temperature=50)  
  
predictions <- predict(assignment\_e, newdata = new\_observation, interval="prediction")  
  
confidence\_predictions <- predict(assignment\_e, newdata = new\_observation, interval="confidence")

## Step 5: Conclusion

Temperature was found to inversely affect seed germination time significantly, with rising temperatures reducing the time needed for seeds to sprout. The model’s high R-squared value indicates a strong predictive power within the temperature range studied. However, the potential adverse effects of extreme temperatures on seed viability point to the need for further research on temperature thresholds for various crop species.