1. By Computing the Multiple regression of time on Cases and Distance, I found the bellows:

**The fitted model:**

 y = 2.341231 + 1.615907\*x1 + 0.014385\*x2

**The estimated residual standard deviation:**

3.259 on 22 degrees of freedom

**P-values for the overall model:**

4.687e-16

**P-values for each of the two predictors:**

P values for Cases: 3.25e-09

P values for Distance: 0.000631

Text, letter

Description automatically generated

Fig4: Multiple regression summary chart

1. **Residual plots and the histogram of the residuals with comments:**

The measured residuals are:

**Graphical user interface, text

Description automatically generated**

Fig:5 Measured Residuals

The density plot of the residuals:

**Chart, histogram

Description automatically generated**

Fig6: Density Plot of the residuals

From the density plot of residuals, I can see the assumption of normality is not followed.

Residual vs Fitted, Normal Q-Q, Scale Location and Residual vs Leverage:

Chart

Description automatically generated

Fig7:Residual vs Fitted, Normal Q-Q, Scale Location and Residual vs Leverage

The Residual vs Fitted values plot shows the constant variance within the intervals.

The Normal Q-Q plot shows the trend of residuals is not totally aligned with the line, hence the normality was not followed by the data distribution.

The scale-Location plot showing root square of standardized residuals along with Fitted values.

The Residuals vs Leverage plot is showing some influential points(9 and 22) along with cook distance plot.

Histogram of residuals:

**Chart, histogram

Description automatically generated**

Fig 8:Histogram of residuals:

From the histogram of residuals plot, I can see the trends pattern is not following normal distribution.

Normal Q-Q Plot of the residuals:

**Chart, scatter chart

Description automatically generated**

Fig9:Normal Q-Q Plot of the residuals

From the above Normal Q-Q plot of the residuals I can see the assumption

of normality is not followed.

**(c)**

According to Cook’s distance there is an observation in this data set which is extremely influential. This observation is number 9 data point.

**Cooks Distance plot:**

Chart, scatter chart

Description automatically generated

Fig 10: Cook’s Distance Plot

The next most influential observation data point is 22.

*The full source code is provided in* ***Appendix C.***

**Appendix :**

data **=** read**.**csv**(**"softdrink.csv"**)**

# In[27]:

ls**(**data**)**

# In[28]:

data$Cases

# In[29]:

data$Distance

# In[30]:

data$Time

# In[31]:

attach**(**data**)**

# In[32]:

Cases

# In[33]:

ls**(**data**)**

# In[34]:

Distance

# In[35]:

Time

# In[36]:

linm**=**lm**(**Time**~**Cases**+**Distance**)**

# In[37]:

summary**(**linm**)**

# In[50]:

anova**(**linm**)**

# In[38]:

resid**(**linm**)**

# In[39]:

plot**(**density**(**resid**(**linm**)))**

# In[55]:

par**(**mfrow**=**c**(**2**,**2**))**

plot**(**linm**,**pch**=**16**)**

# In[56]:

hist**(**residuals**(**linm**))**

# In[57]:

qqnorm**(**residuals**(**linm**))**

qqline**(**residuals**(**linm**))**