Statistics Software Lab Report - 10 (Outputs file)

Name of the Student: Shatansh Patnaik
Roll No: 20MA20067

IIT Kharagpur
Statistics Software Lab

Problem-1

The carbonation level of a soft drink beverage is affected by the temperature of the product (X_1) and the filler operating pressure (X_2). Twelve observations were obtained and the resulting data are shown below:

Table 1: Data for Carbonation Level, Temperature, and Filler Operating Pressure

Carbonation Level (y)	Temperature (X_1)	Filler Operating Pressure (X_2)
2.6	21	31
2.4	21	31
17.32	24	31.5
15.6	24	31.5
16.12	24	31.5
5.36	22	30.5
6.19	22	31.5
10.17	23	30.5
2.62	21.5	31
2.98	21.5	30.5
6.92	22.5	31
7.06	22.5	30.5

a) Fit a second-order polynomial in two variables X_1 and X_2.
b) Find confidence intervals for the coefficients of the model.
c) Test for the significance of regression coefficients.
d) Find the coefficient of determination.

Solution for Problem-1

```
# Solution  for Exercise-1:
# a) Fitting a second order polynomial in two variables x1 and x2:
"The coefficient 1 is 3025.31866957059"
"The coefficient 2 is -194.272896029518"
"The coefficient 3 is -6.05066788960215"
"The coefficient 4 is 3.62587454971225"
"The coefficient 5 is 1.15424977741555"
"The coefficient 6 is -1.33171009507802"

# b) Confidence Intervals for the coefficients of the model:
"The 95% confidence interval for the coefficient 1 is -1980.44247708771 to
    8031.07981622889"
"The 95% confidence interval for the coefficient 2 is -517.422544763653 to
    128.876752704617"
"The 95% confidence interval for the coefficient 3 is -56.472352801759 to
    44.3710170225547"
"The 95% confidence interval for the coefficient 4 is -1.78126938018931 to
    9.0330184796138"
"The 95% confidence interval for the coefficient 5 is 0.362113626219123 to
    1.94638592861197"
```

```
"The 95% confidence interval for the coefficient 6 is -3.52461748507398 to
    0.861197294917941"

# c) Significance Testing for the Coefficients:
"The coefficient 1 is not significant"
"The coefficient 2 is not significant"
"The coefficient 3 is not significant"
"The coefficient 4 is not significant"
"The coefficient 5 is significant"
"The coefficient 6 is not significant"

# d) Coefficient of determination:
The R-squared value is 0.9932723
```

Problem-2

The data on 13 subjects of approximately the same height is taken on systolic blood pressure (Y) in relation to weight (X_1) and age (X_2).

Table 2: Data on Systolic Blood Pressure, Weight, and Age

X_1	X_2	Y
152	50	120
250	125	183
201	141	170
170	40	132
124	53	171
201	43	155
126	40	158
165	30	119
164	40	132
150	50	129
185	20	147
149	60	123

(a) Fit a multiple regression model.
(b) Test for the significance of all parameters of the model.
(c) Set up 95% confidence intervals for all regression coefficients.
(d) Calculate the coefficient of determination.

Solution for Problem-2

```
# Solution  for Exercise-2:
# a) Fitting a multiple regression model in two variables x1 and x2:
"The coefficient 1 is -65.0996781277957"
"The coefficient 2 is 1.07710145109936"
"The coefficient 3 is 0.425413431164021"
```

```

# b) Significance Testing for the Coefficients:
"The coefficient 1 is significant"
"The coefficient 2 is significant"
"The coefficient 3 is significant"

# c) Confidence Intervals for the coefficients of the model:
   Coefficient Lower Bound (95% CI) Upper Bound (95% CI)
   -65.0996781           -98.3982673           -31.8010889
X1   1.0771015             0.9053739             1.2488290
X2   0.4254134             0.2624199             0.5884069

# d) Coefficient of determination:
The R-squared value is 0.9576725
```

Problem-3

The following table presents the data concerning the percentage of conversion of n-heptane to acetylene and three explanatory variables. Fit full quadratic model in three variables to this using least squares. Examine the correlation matrix and detect multicollinearity. What statements can you make about the multicollinearity in the data? Resolve multicollinearity and fit the regression model in reduced number of variables.

Table 3: Data on Conversion of n-Heptane to Acetylene

Observation Contact Time	Conversion (%)	Temperature (°C)	Mole ratio
1	49.0	130	0.0120
2	50.2	130	0.0120
3	50.5	130	0.0115
4	48.5	130	0.0130
5	47.5	130	0.0135
6	44.5	130	0.0120
7	28.0	120	0.0400
8	31.5	120	0.0380
9	34.5	120	0.0320
10	35.0	120	0.0260
11	38.0	120	0.0340
12	38.5	120	0.0410
13	15.0	105	0.0840
14	17.0	105	0.0980
15	20.5	100	0.0920
16	29.5	100	0.0860

Solution for Problem-3

```
# Solution  for Exercise-3:
Call:
lm(formula = Conversion ~ ., data = as.data.frame(X))

Residuals:
    Min      1Q  Median      3Q     Max
-4.0536 -1.9109 -0.5033  2.1621  4.0080

Coefficients: (1 not defined because of singularities)
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               4.164e+02  7.545e+02   0.552   0.5945
V1                               NA         NA      NA       NA
Reactor_Temperature      -8.442e-01  1.217e+00  -0.694   0.5054
Reactor_Temperature_sq    4.175e-04  4.917e-04   0.849   0.4178
Ratio_H2_to_n_Heptane     2.620e+00  9.212e-01   2.844   0.0193 *
Ratio_H2_to_n_Heptane_sq -8.129e-02  3.254e-02  -2.498   0.0340 *
Contact_Time              4.990e+02  5.085e+02   0.981   0.3521
Contact_Time_sq          -4.188e+03  3.716e+03  -1.127   0.2889
---

Residual standard error: 3.228 on 9 degrees of freedom
Multiple R-squared:  0.9559,
Adjusted R-squared:  0.9264
F-statistic: 32.48 on 6 and 9 DF,
p-value: 1.325e-05

The Correlation Matrix is as follows:
                      Reactor_Temperature Ratio_H2_to_n_Heptane Contact_Time
Reactor_Temperature             1.0000000             0.2236278   -0.9582041
Ratio_H2_to_n_Heptane           0.2236278             1.0000000   -0.2402310
Contact_Time                   -0.9582041            -0.2402310    1.0000000
```

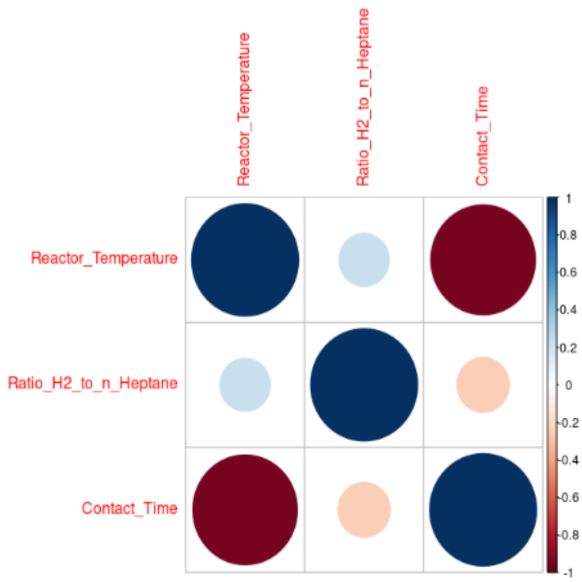

Figure 1: Correlation Plot

From the correlation matrix provided, we can draw several inferences and conclusions:

1. **Reactor Temperature and Contact Time**:
   - There is a strong negative correlation (-0.958) between Reactor Temperature and Contact Time. This indicates that as the Reactor Temperature increases, the Contact Time decreases, and vice versa.
   - This negative correlation suggests that there might be a relationship between these two variables, where changes in one variable affect the other variable in the opposite direction.
2. **Reactor Temperature and Ratio of H2 to n-Heptane**:
   - There is a weak positive correlation (0.224) between Reactor Temperature and the Ratio of H_2 to n-Heptane.
   - This suggests that there might be a slight tendency for Reactor Temperature to increase with the Ratio of H_2 to n-Heptane, but the correlation is not strong enough to make a definitive conclusion.
3. **Ratio of H2 to n-Heptane and Contact Time**:
   - There is a weak negative correlation (-0.240) between the Ratio of H_2 to n-Heptane and Contact Time.
   - This suggests that there might be a slight tendency for Contact Time to decrease as the Ratio of H_2 to n-Heptane increases, but again, the correlation is not very strong.
4. **Overall Interpretation**:
   - The strongest correlation exists between Reactor Temperature and Contact Time, indicating a potential significant relationship between these two variables.
   - The weak correlations between Reactor Temperature and the other two variables, as well as between the Ratio of H_2 to n-Heptane and Contact Time, suggest that these relationships may be less important or less clear.
   - Further analysis, such as regression modeling, hypothesis testing, or additional data collection, may be needed to better understand the relationships between these variables and draw more conclusive inferences.

```
# Solution  for Exercise-3 (Continued...):
Call:
lm(formula = Conversion ~ Reactor_Temperature + Contact_Time,
    data = newData)

Residuals:
    Min      1Q  Median      3Q     Max
-6.5572 -2.8978  0.1532  2.1379  9.2159

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -113.76373   61.10148  -1.862   0.0854 .
Reactor_Temperature    0.12489    0.04659   2.681   0.0189 *
Contact_Time         -38.77668  118.72437  -0.327   0.7492
---

Residual standard error: 4.162 on 13 degrees of freedom
Multiple R-squared:  0.894,
Adjusted R-squared:  0.8777
F-statistic:  54.8 on 2 and 13 DF,
p-value: 4.628e-07
```

Problem 4

Consider the data on space shuttle Challenger O-ring failures, which have been linked to temperature. The following table gives the data on temperatures at take-off and whether or not an O-ring failed.

Table 4: Data on Space Shuttle Challenger O-ring Failures

Failure (Y)	**Temperature** (X)
1	53
1	57
1	58
1	63
0	66
0	67
1	67
0	67
0	68
0	69
0	70
0	70
1	70
0	70
1	72
1	73
0	75
0	75
0	76
0	76
0	78
0	79
0	81

```
# Solution for Exercise-4:
intercept    0.5979680
temperature -0.2745997

# Significance Testing for Exercise-4:
"The coefficient 1 is not significant"
"The coefficient 2 is significant"
```