# Advanced Statistics Assignment#2

Abdul Saboor
20L-1113
BDS-4A

**Question # 1:**

Formulas being used:

$\hat{y} = a + b\,x$ where
$b = [N\,\Sigma(xy) - \Sigma x\,\Sigma y] / [N\,\Sigma(x^2) - (\Sigma x)^2]$
$a = \bar{y} - b\,\bar{x}$

(a)

| **Time (X)** | **Temperature (Y)** | **X^2** | **XY** |
|---|---|---|---|
| 0 | 82 | 0 | 0 |
| 1 | 71 | 1 | 71 |
| 2 | 59 | 4 | 118 |
| 3 | 52 | 9 | 156 |
| 4 | 43 | 16 | 172 |
| 5 | 35 | 25 | 175 |
| Total = 15 | Total = 35 | Total = 55 | Total = 692 |
| $\bar{x}$ = 15 / 6 = 2.582 | $\bar{y}$ = 342 / 6 = 57 | | |

Putting values in formulae:
$b = [6 * 692 - 15 * 342] / [6 * 55 - 15^2]$
b = -978 / 105 = -9.3143

a = 57 - 2.5 (-9.3143) = 80.2857

**$\hat{y}$ = 80.28 - 9.31 x (Least Square Regression Line)**

(b)

x = 6
$\hat{y}$ = 80.28 - 9.31x
$\hat{y}$ = 80.28 - 9.31(6)
**$\hat{y}$ = 24.42**

The temperature of the coffee after 6 minutes would be 24.42 Fahrenheit.

**Question # 2:**

**(a) Final regression line:**

```python
import pandas as pd
import numpy as np
data=pd.read_csv('Chem_Process.csv')

x=np.array(data['Temperature'])
y=np.array(data['Yield'])
x=np.vstack([x,np.ones(len(x))]).T
m,c=np.linalg.lstsq(x,y)[0]
print(f'Regression line is: yhat = ({m})x + {c}')
```

✓ 0.2s

```
Regression line is: yhat = (1.9951684696867256)x + 17.001591728645355
```

**(b) The yield of chemical process at 35 degrees Celsius:**

```python
x=35
print(f'The yield of chemical process at 35 degrees Celsius: {m*x+c}')
```

✓ 0.2s

```
The yield of chemical process at 35 degrees Celsius: 86.83248816768075
```

**Question # 3:**

Homoscedasticity:

This is the term referred to a situation where our residuals or error is almost constant across the values of independent variables.

*What kind of insights can we draw out of homoscedasticity?*

If the variance in error is homoscedastic, we can say that our model is well defined. For large variance in error the model is not fitting well to the given data.

Heteroscedasticity:

This is the term referred to situation where there is large variance in the error or residuals across all the values of independent variables.

*What kind of insights can we draw out of homoscedasticity?*

Heteroscedasticity in a prediction may not cause a bias in the coefficient estimates but it does tend to make the values less precise. This lower precision tends to deviate the estimates further away from the actual values.

**Question # 4:**

Formulas being used:

ŷ = b0 + b1*x1 + b2*x2

b0 = y – b1X1 – b2X2

b1 = [(Σx2^2)(Σx1y) – (Σx1x2)(Σx2y)] / [(Σx1^2) (Σx2^2) – (Σx1x2)^2]

b2 = [(Σx1^2)(Σx2y) – (Σx1x2)(Σx1y)] / [(Σx1^2) (Σx2^2) – (Σx1x2)^2]

(a)

| $y$ | $x1$ | $x2$ | $x1^2$ | $x2^2$ | $x1y$ | $x2y$ | $x1x2$ |
|---|---|---|---|---|---|---|---|
| 140 | 60 | 22 | 3600 | 484 | 8400 | 3080 | 1320 |
| 155 | 62 | 25 | 3844 | 625 | 9610 | 3875 | 1550 |
| 159 | 67 | 24 | 4489 | 576 | 10653 | 3816 | 1608 |
| 179 | 70 | 20 | 4900 | 400 | 12530 | 3580 | 1400 |
| 192 | 71 | 15 | 5041 | 225 | 13632 | 2880 | 1065 |
| 200 | 72 | 14 | 5184 | 196 | 14400 | 2800 | 1008 |
| 212 | 75 | 14 | 5625 | 196 | 15900 | 2968 | 1050 |
| 215 | 78 | 11 | 6084 | 121 | 16770 | 2365 | 858 |
| Total = 1542 | Total = 555 | Total = 145 | Total = 38767 | Total = 2823 | Total = 101895 | Total 25364 | Total = 9859 |
| ȳ = 1452/8 =181.5 | x̄1 = 555/8 = 69.375 | x̄2 = 145/8 = 18.125 | | | | | |

$\Sigma x1^2$ = $\Sigma x1^2$ – $(\Sigma x1)^{\mathbf{2}}$ / n

$\Sigma x1^2$ = 38767 – $(555)^2$/ 8

$\Sigma x1^2$ = 263.875

$\Sigma x2^2$ = $\Sigma x2^2$ – $(\Sigma x2)^{\mathbf{2}}$ / n

$\Sigma x2^2$ = 38767 – $(145)^2$/ 8

$\Sigma x2^2$ = 194.875

Σx1y = ΣX1y – (ΣX1Σy) / n

Σx1y = 101895 – (555*1452) / 8

Σx1y = 1162.5

Σx2y = ΣX2y – (ΣX2Σy) / n

Σx2y = 25364 – (145*1452) / 8

Σx2y = -953.5

Σx1x2 = ΣX1X2 – (ΣX1ΣX2) / n

Σx1x2 = 9859 – (555*145) / 8

Σx1x2 = -200.375

Putting values in formulas:
$b_1 = [(194.875)(1162.5) – (-200.375)(-953.5)] / [(263.875) (194.875) – (-200.375)2]$
$b_1 = 3.148$
$b_2 = [(263.875)(-953.5) – (-200.375)(1152.5)] / [(263.875) (194.875) – (-200.375)2]$
$b_2 = -1.656$
$b_0 = 181.5 – 3.148(69.375) – (-1.656)(18.125)$
$b_0 = -6.867$

**$\hat{y} = -6.867 + 3.148x_1 – 1.656x_2$ (Multiple Linear Regression Line)**

(b)

$x_1 = 76$
$x_2 = 13$
$\hat{y} = -6.867 + 3.148x_1 – 1.656x_2$
$\hat{y} = -6.867 + 3.148(76) – 1.656(13)$
**$\hat{y} = 210.853$**