ADS LAB – 1 ASSIGNMENT

Name: **MAHVISH ISHAQ** Date: **19/10/2025**

1)

Date: 18/10/25 ADS LAB Page No.: 1

ASSIGNMENT - 1

1. Given the following data of Temperature (°c) and Power consumption (kWh):

(a) Derive a regression equation $Y = a + bX$, using the least squares method and calculate a (intercept) and b (slope). Also compute the value of ΣX, ΣY, ΣXY.

(b) Using your predicted values ($\hat{Y}$), compute R^2.

Temperature °C (X)	Power Consumption (kWh) (Y)
10	300
12	310
14	320
16	330
18	345
20	360
22	370
24	390
26	420
28	450

(a) $\hat{Y} = a + bx$ → regression equation

$\Sigma X = 190$

$\Sigma Y = 3595$

$\Sigma X^2 = 3940$

$\Sigma XY = 70910$

Date : 18/10/25

	X	Y	XY	X^2	$\hat{Y}$	SS_{res}	SS_{tot}
	10	300	3000	100	288.5	132.25	3540.25
	12	310	3720	144	304.28	32.718	2450.25
	14	320	4480	196	320.06	0.0036	1560.25
	16	330	5280	256	335.84	34.105	870.25
	18	345	6210	324	351.62	43.82	210.25
	20	360	7200	400	367.4	54.76	0.25
	22	370	8140	484	383.18	173.71	110.25
	24	390	9360	576	398.96	80.28	930.25
	26	420	10920	676	414.74	27.66	3660.25
	28	450	12600	784	430.52	379.47	8190.25
Σ	190	3595	70910	3940		958.77	21522.5

Computing slope b

$$b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2}$$

$n = 10$, $\Sigma XY = 70910$

$\Sigma X = 190$, $\Sigma Y = 3595$, $\Sigma X^2 = 3940$

$(\Sigma X)^2 = 190^2 = 36100$

$$\therefore\ b = \frac{(10)(70910) - (190)(3595)}{(10)(3940) - 36100}$$

$$= \frac{709100 - 683050}{39400 - 36100} = \frac{521}{66} = 7.89$$

$\boxed{b = 7.89}$

Computing intercept a

$$a = \bar{Y} - b\bar{X}$$

$\bar{X} = \frac{\Sigma X}{n}$ $\quad \bar{X} = \frac{190}{10} = 19$, $\bar{Y} = \frac{3595}{10} = 359.5$

$\bar{Y} = \frac{\Sigma Y}{n}$

$\hat{Y}$
$= 209.6 + 7.89 \times 10$
$209.6 + 7.89 \times 12$
$209.6 + 7.89 \times 14$
$209.6 + 7.89 \times 16$
$209.6 + 7.89 \times 18$
$209.6 + 7.89 \times 20$
$209.6 + 7.89 \times 22$
$209.6 + 7.89 \times 24$
$209.6 + 7.89 \times 26$
$209.6 + 7.89 \times 28$

SS tot
$(300-359.5)^2$ … $(450-359.5)^2$

SS res
$(300-288.5)^2$, $(310-304.28)^2$, $(320-320.06)^2$, $(330-335.84)^2$, $(345-351.62)^2$, $(360-367.4)^2$, $(370-383.18)^2$, $(390-398.96)^2$, $(420-414.74)^2$, $(450-430.52)^2$

Date : 18/10/25 Page No. : 3

$a = 359.5 - (7.89)(19)$

$= 359.5 - 149.91 = 209.59$

$\boxed{a = 209.6}$

Regression Equation $\rightarrow$ $\hat{Y} = 209.6 + 7.89X$

(b) Computing R^2

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$SS_{residual} = \Sigma (Y - \hat{Y})^2$ $\hat{Y}$ = predicted value

$SS_{total} = \Sigma (Y - \bar{Y})^2$ $\bar{Y}$ = mean value.

See the values of $\hat{Y}$, SS_{res}, SS_{tot} from table.

$SS_{res} = 958.77$

$SS_{tot} = 21522.5$

$$R^2 = 1 - \frac{958.77}{21522.5}$$

$= 1 - 0.044$

$= 0.956$

$\therefore \boxed{R^2 = 0.95}$ (equivalent to 1)

→ 95% of variation in power is explained by temperature

→ Perfect Linear relationship.

- Model fits data very well.

→ 2^{nd} and 3^{rd} answer are attached in .ipynb format.

2) (a) USE PYTHON (statsmodels) TO FIT MODEL AND COMPARE.

(b) INTERPRET RESULTS (POSITIVE/NEGATIVE SLOPE, ACCURACY)

(a)

```python
import statsmodels.api as sm

# X = Temperature , Y = Power Consumption
X = [10, 12, 14, 16, 18, 20, 22, 24, 26, 28]
Y = [300, 310, 320, 330, 345, 360, 370, 390, 420, 450]

# constant term for intercept
X = sm.add_constant(X)

# Create and fit model and compare
model = sm.OLS(Y, X).fit()

# Display regression summary
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.955
Model:                            OLS   Adj. R-squared:                  0.950
Method:                 Least Squares   F-statistic:                     171.6
Date:                Sun, 19 Oct 2025   Prob (F-statistic):           1.10e-06
Time:                        17:45:52   Log-Likelihood:                -37.005
No. Observations:                  10   AIC:                             78.01
Df Residuals:                       8   BIC:                             78.61
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        209.5152     11.962     17.515      0.000     181.931     237.100
x1             7.8939      0.603     13.099      0.000       6.504       9.284
==============================================================================
Omnibus:                        1.026   Durbin-Watson:                   0.581
Prob(Omnibus):                  0.599   Jarque-Bera (JB):                0.781
Skew:                           0.568   Prob(JB):                        0.677
Kurtosis:                       2.236   Cond. No.                         68.7
==============================================================================
```

(b) INTERPRETING RESULTS

-->SLOPE (b = 7.89):

* The slope is POSITIVE, meaning that as temperature increases, power consumption also increases.

-->INTERCEPT (a = 209.5):

* When the temperature is 0°C, the predicted power consumption is approximately 209.5 kWh.

-->R^2 = 0.95:

*The model explains 95% of the variation in power consumption. This indicates a very strong or PERFECT LINEAR RELATIONSHIP between temperature and power consumption.

3) Using Python, Perform Linear Regression on the dataset attached in excel format.

In [15]:

```
# ------------------------------------------------------------
# Linear Regression with Confidence Intervals
# ------------------------------------------------------------

# Step 1: Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import statsmodels.api as sm
from google.colab import files

# Step 2: Upload Excel file
print(" Please upload your Experience_Salary.xlsx file")
uploaded = files.upload()

# Step 3: Load dataset into DataFrame
data = pd.read_excel("Experience_Salary.xlsx")

# Step 4: Display all rows
pd.set_option('display.max_rows', None)
print("\n Full Dataset Preview:")
print(data)

# Step 5: Define independent and dependent variables
X = data[['Experience_Years']]
y = data['Salary_USD']

# Step 6: Linear Regression using sklearn
model = LinearRegression()
model.fit(X, y)
y_pred = model.predict(X)

# Step 7: Compute R²
r2 = r2_score(y, y_pred)
```

```python
# Step 8: Confidence Interval using statsmodels
X_sm = sm.add_constant(X)  # Add constant for intercept
sm_model = sm.OLS(y, X_sm).fit()
pred_summary = sm_model.get_prediction(X_sm).summary_frame(alpha=0.05)  # 95% CI

# Step 9: Combine results
result = data.copy()
result['Predicted'] = y_pred
result['Lower_CI'] = pred_summary['obs_ci_lower']  # Lower bound of CI
result['Upper_CI'] = pred_summary['obs_ci_upper']  # Upper bound of CI
print("\n Result with Predictions and Confidence Intervals:")
print(result.head())

# Step 10: Display model parameters
print("\n Regression Equation:")
print(f"Intercept (a): {model.intercept_:.2f}")
print(f"Slope (b): {model.coef_[0]:.2f}")
print(f"R² Score: {r2:.3f}")

# Step 11: Plot regression with confidence intervals
plt.figure(figsize=(8,5))
plt.scatter(X, y, color='blue', label='Actual Data')
plt.plot(X, y_pred, color='red', linewidth=2, label='Regression Line')
plt.fill_between(X['Experience_Years'], result['Lower_CI'], result['Upper_CI'], color='gray', alpha=0.5, label='95% Confide
plt.xlabel('Experience (Years)')
plt.ylabel('Salary (USD)')
plt.title('Experience vs Salary - Linear Regression')
plt.legend()
plt.show()

# Step 12: Interpretation
print("\n INTERPRETATION:")
print(f"- The SLOPE is POSITIVE ({model.coef_[0]:.2f}), meaning salary increases with experience.")
print(f"- R² of {r2:.3f} indicates the model explains about {r2*100:.1f}% of salary variation.")
print("- The shaded area in the plot shows the 95% CONFIDENCE INTERVAL of predicted salaries.")
```

```
 Please upload your Experience_Salary.xlsx file
```

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

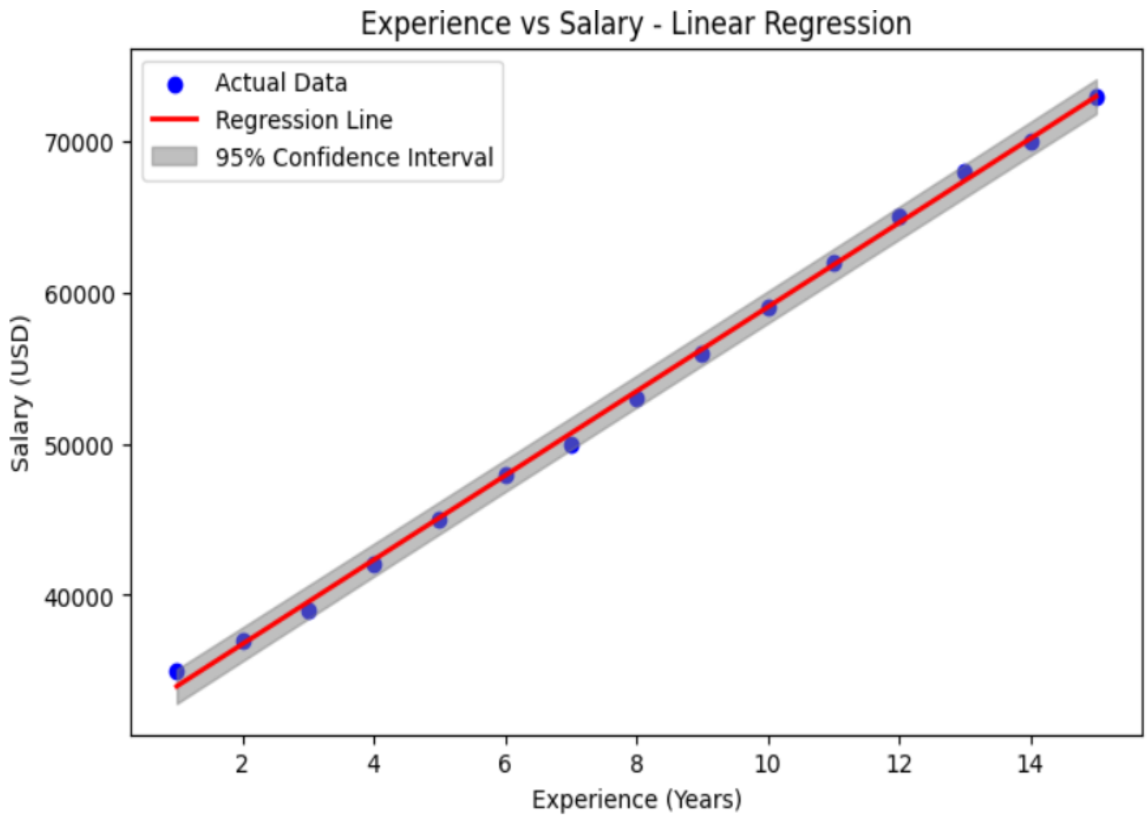
```
Saving Experience_Salary.xlsx to Experience_Salary (13).xlsx

 Full Dataset Preview:
    Experience_Years  Salary_USD
0                  1       35000
1                  2       37000
2                  3       39000
3                  4       42000
4                  5       45000
5                  6       48000
6                  7       50000
7                  8       53000
8                  9       56000
9                 10       59000
10                11       62000
11                12       65000
12                13       68000
13                14       70000
14                15       73000

 Result with Predictions and Confidence Intervals:
   Experience_Years  Salary_USD     Predicted      Lower_CI      Upper_CI
0                 1       35000  33966.666667  32834.350978  35098.982355
1                 2       37000  36752.380952  35641.436801  37863.325104
2                 3       39000  39538.095238  38445.561177  40630.629299
3                 4       42000  42323.809524  41246.572261  43401.046786
4                 5       45000  45109.523810  44044.335918  46174.711701

 Regression Equation:
Intercept (a): 31180.95
Slope (b): 2785.71
R² Score: 0.999
```

•




```
INTERPRETATION:
- The SLOPE is POSITIVE (2785.71), meaning salary increases with experience.
- R² of 0.999 indicates the model explains about 99.9% of salary variation.
- The shaded area in the plot shows the 95% CONFIDENCE INTERVAL of predicted salaries.
```