Predicting Student Performance in Final Exam

Objective: Predict final exam score (dependent variable)
Independent Variables:

- x1: Hours studied per week
- x2: Attendance percentage
- x3: Assignment completion rate (%)
- x4: Sleep hours before exam day
- x5: Use of online learning tools (binary: 0 = No, 1 = Yes)

S.#	Hours Studied	Attendance %	Assignment %	Sleep Hours	Online Tools (0/1)	Final Exam Score
1	10.99	70.85	91.79	5.76	1	128.94
2	9.72	80.79	92.8	6.16	0	115.72
3	11.3	81.57	95.42	8.12	0	126.23
4	13.05	76.98	95.27	7.92	1	131.76
5	9.53	83.39	83.11	6.97	1	134.69
6	9.53	89.04	85.31	7.18	1	127.67
7	13.16	103.86	92.58	8.92	1	137.78
8	11.53	86.75	92.57	6.11	1	122.46
9	9.06	87.58	92.58	7.82	1	116.21
10	11.09	84.26	109.26	6.7	1	133.87
11	9.07	65.81	92.85	6.67	1	112.92
12	9.07	84.73	95.68	8.65	0	113.37
13	10.48	85.6	94.77	8.24	1	124.56
14	6.17	109.63	93.26	8.22	0	115.35
15	6.55	83.08	88.42	8.96	1	125.79
16	8.88	88.02	93.79	7.03	1	128.11
17	7.97	84.65	86.14	8.02	0	113.78
18	10.63	73.31	88.82	6.53	1	127.71
19	8.18	96.43	87.57	7.49	1	123.2
20	7.18	92.52	90.41	6.8	0	110.77

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8030
R Square	0.6448
Adjusted R Square	0.5180
Standard Error	5.6472
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	810.52	162.10	5.08	0.0073
Residual	14	446.47	31.89		
Total	19	1256.99			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	82.68	26.41	3.13	0.0074	26.03	139.33
Hours Studied	2.25	0.81	2.79	0.0144	0.52	3.98
Attendance %	0.17	0.15	1.14	0.2722	-0.15	0.49
Assignment %	-0.07	0.26	-0.28	0.7846	-0.63	0.49
Sleep Hours	0.73	1.53	0.47	0.6428	-2.56	4.01
Online Tools (0/1)	8.55	3.06	2.79	0.0144	1.98	15.13

SUMMARY OUTPUT INTERPRETATION

Regression Statistics

Metric	Value	Interpretation
Multiple R	0.803	Strong correlation between actual and predicted exam scores.
R Square (R^2)	0.645	~64.5% of the variance in final exam scores is explained by the model.
Adjusted R Square	0.518	Accounts for number of predictors; still a decent model fit for 5 predictors with 20 observations.
Standard Error	5.65	On average, predicted scores deviate from actual scores by ~5.65 points.
Observations	20	Number of students (data points).

ANOVA Table

Source	df	SS	MS	F	Significance F
Regression	5	810.52	162.10	5.083	**0.0073**
Residual	14	446.47	31.89		
Total	19	1256.99			

- **F-statistic = 5.08** with **p-value = 0.0073** → The model is statistically significant overall (at $\alpha = 0.05$).
- This means that at least one predictor meaningfully contributes to predicting the final exam score.

🧮 Regression Equation

$$\text{Score} = 82.68 + 2.25x_1 + 0.17x_2 - 0.07x_3 + 0.73x_4 + 8.55x_5$$

Where:

- x_1: Hours Studied
- x_2: Attendance %
- x_3: Assignment Completion %
- x_4: Sleep Hours
- x_5: Use of Online Tools (0 or 1)

Detailed Coefficient Interpretation

Variable	Coefficient	P-value	Significance?	Interpretation
Intercept	82.68	0.007	✅ Yes	Predicted score for a student with all independent variables at 0. Mostly theoretical.
Hours Studied	**2.25**	**0.014**	✅ Yes	Each additional hour studied per week increases score by ~2.25 points, holding other variables constant.
Attendance %	0.17	0.272	❌ No	Positive but not statistically significant.
Assignment %	-0.07	0.785	❌ No	Slight negative and not significant. Possibly due to multicollinearity or overfitting.
Sleep Hours	0.73	0.643	❌ No	Positive, but effect not statistically significant.
Online Tools	**8.55**	**0.014**	✅ Yes	Students using online tools score ~8.55 points higher on average. Very impactful variable.

Conclusion

- **Significant Predictors**:
    - Hours Studied
    - Use of Online Tools
- **Model Quality**:
    - Good fit for a small sample (R^2 = 64.5%)
    - Room for improvement with more data and possibly interaction terms.
- **Recommendations**:
    - Encourage students to use online tools and invest more time studying.
    - Consider gathering more observations to improve estimate precision.