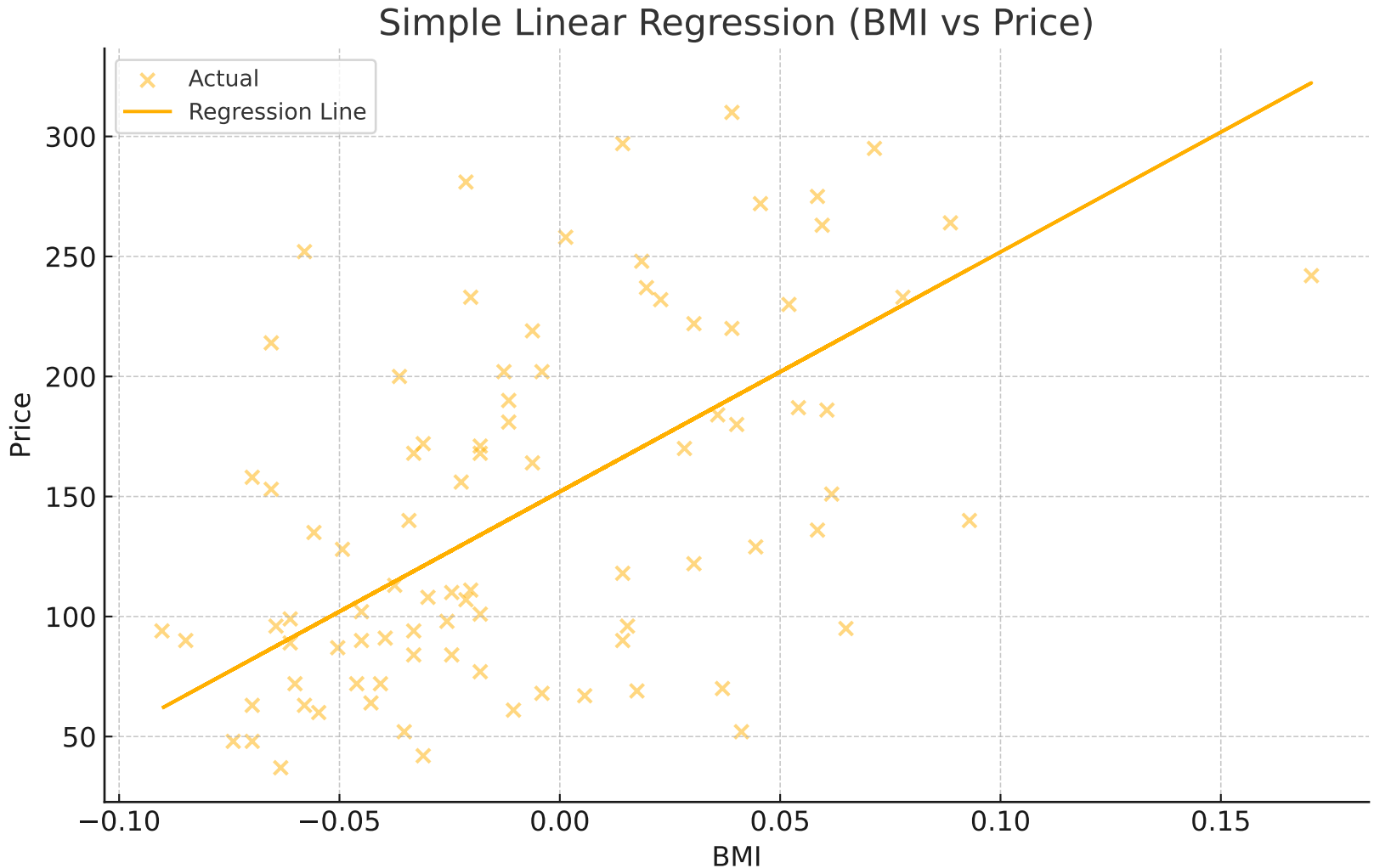
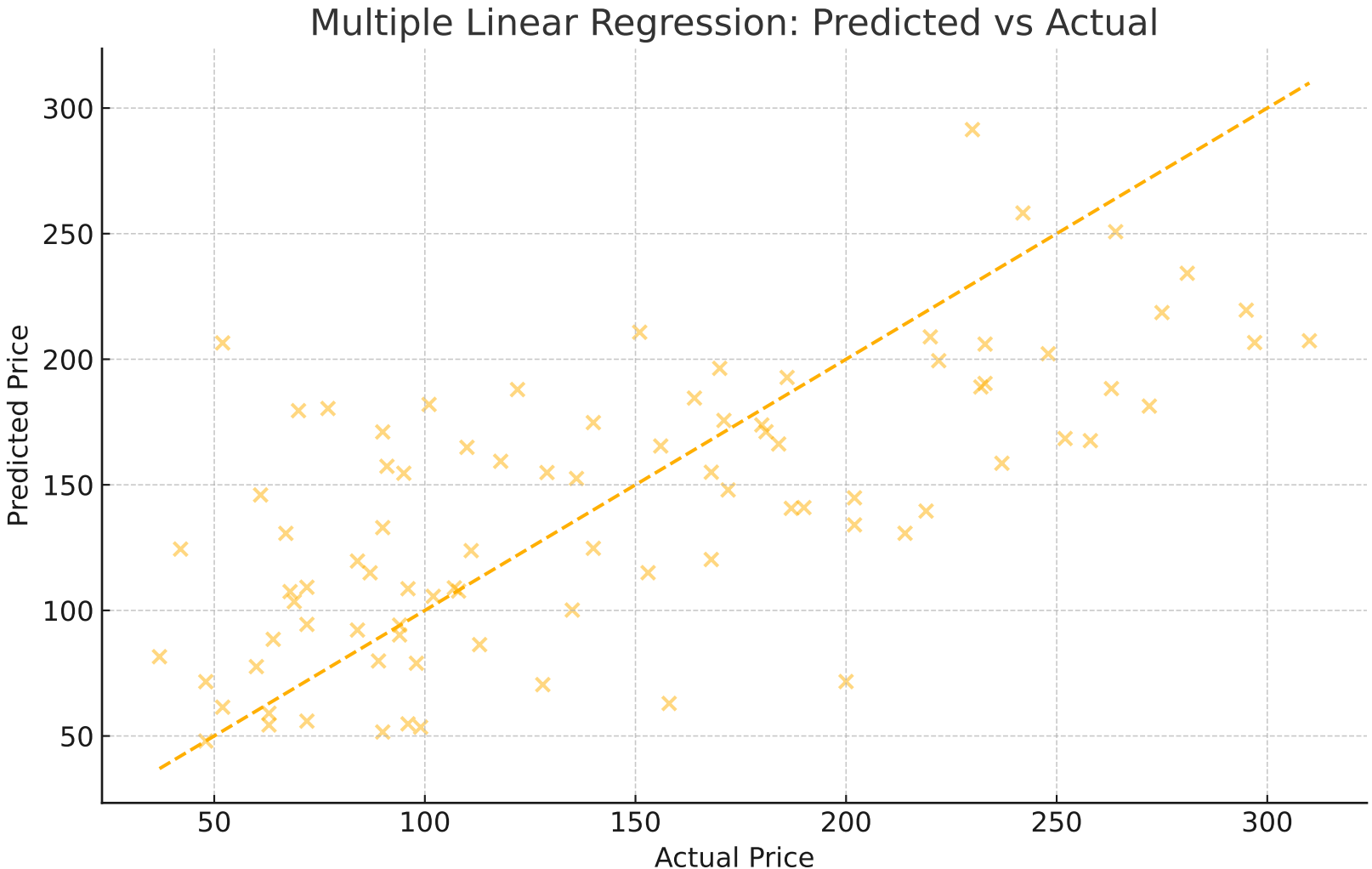
Linear Regression Solution

Simple Regression:
  MAE: 52.2600
  MSE: 4061.8259
  R2: 0.2334

Multiple Regression:
  MAE: 42.7941
  MSE: 2900.1936
  R2: 0.4526

Coefficients for Multiple Regression:
  age: 37.9040
  sex: -241.9644
  bmi: 542.4288
  bp: 347.7038
  s1: -931.4888
  s2: 518.0623
  s3: 163.4200
  s4: 275.3179
  s5: 736.1989
  s6: 48.6707


Simple Linear Regression (BMI vs Price)
Actual
Regression Line
Price
BMI
300
250
200
150
100
50
−0.10
−0.05
0.00
0.05
0.10
0.15


Multiple Linear Regression: Predicted vs Actual
Predicted Price
Actual Price
50
100
150
200
250
300

Linear Regression Interview Questions and Solutions

1. What assumptions does linear regression make?
   - Linearity: The relationship between dependent and independent variables is linear.
   - Independence: Observations are independent of each other.
   - Homoscedasticity: Constant variance of residuals across all levels of predictors.
   - Normality: Residuals are normally distributed for inference.
   - No multicollinearity: Predictors are not highly correlated.
   - No autocorrelation: Residuals are not correlated (esp. in time series).

2. How do you interpret the coefficients?
   - A coefficient β_i represents the expected change in the dependent variable for a one-unit increase in predictor X_i, holding other predictors constant.

3. What is R^2 score and its significance?
   - $R^2 = 1 - (SS_{res} / SS_{tot})$. It measures the proportion of variance in the dependent variable explained by the model (0 to 1 range). Higher is better.

4. When would you prefer MSE over MAE?
   - MSE (Mean Squared Error) penalizes larger errors more, is differentiable, and useful in gradient-based optimization. Use when you want to emphasize larger deviations.

5. How do you detect multicollinearity?
   - Compute Variance Inflation Factor (VIF): VIF > 5 or 10 indicates issues.
   - Inspect correlation matrix of predictors.
   - Check condition number of the design matrix.

6. What is the difference between simple and multiple regression?
   - Simple regression: one predictor variable.
   - Multiple regression: two or more predictor variables.

7. Can linear regression be used for classification?
   - Not directly; it predicts continuous values. For classification, use logistic regression or other classifiers.

8. What happens if you violate regression assumptions?
   - Estimates may be biased, inefficient, or inconsistent.
   - Hypothesis tests and confidence intervals become invalid.
   - Predictions may be unreliable.