

1.10 Simple Linear Regression

Question 1

An investor receives a research report that analyzes two securities through a simple linear regression. The study discloses the coefficient of determination (R^2) and the sum of squares error (SSE). To estimate the accuracy of the model using an absolute measure of the errors, the investor will *most likely* need to know the:

- A. sample size.
- B. regression coefficients.
- C. correlation between the securities.

Question 2

An economist runs the following regression models between two macroeconomic datasets with variable transformations:

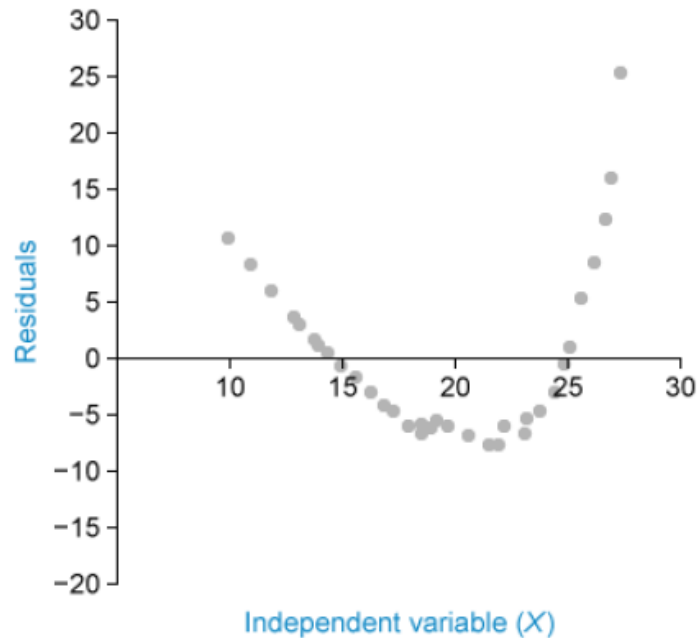
Regression	M1	M2
Model	$\ln Y = b_0 + b_1 X + \varepsilon$	$\ln Y = b_0 + b_1 \ln X + \varepsilon$
Intercept	5.723	5.338
Slope	0.233	1.184
Coefficient of determination (R^2)	0.830	0.573
Standard error of the estimate (s_e)	0.415	0.658
F-statistic	137.001	37.624

Based only on this data, which of the following functional forms is the *most appropriate* fit for the sample data?

- A. Lin-log
- B. Log-lin
- C. Log-log

Question 3

An economist runs a linear regression and analyzes the regression's residual plot:



The *most appropriate* conclusion is that the pattern of the residual plot indicates the presence of:

- A. outliers.
- B. nonlinearity.
- C. heteroskedasticity.

Question 4

A linear regression analysis generates the following ANOVA (analysis of variance) table:

ANOVA Table				
Source	Degrees of freedom	Sum of squares	Mean squares	F-statistic
Regression	1	9.5231	9.5231	163.4274
Residual	78	4.5452	0.0583	
Total	79	14.0683		

Based on this data, the standard error of the estimate (s_e) is *closest* to:

- A. 0.241
- B. 0.323
- C. 0.677

Question 5

An analyst runs a linear regression on crude oil prices (in USD per barrel) and a company's gross margin (in %), with the following results:

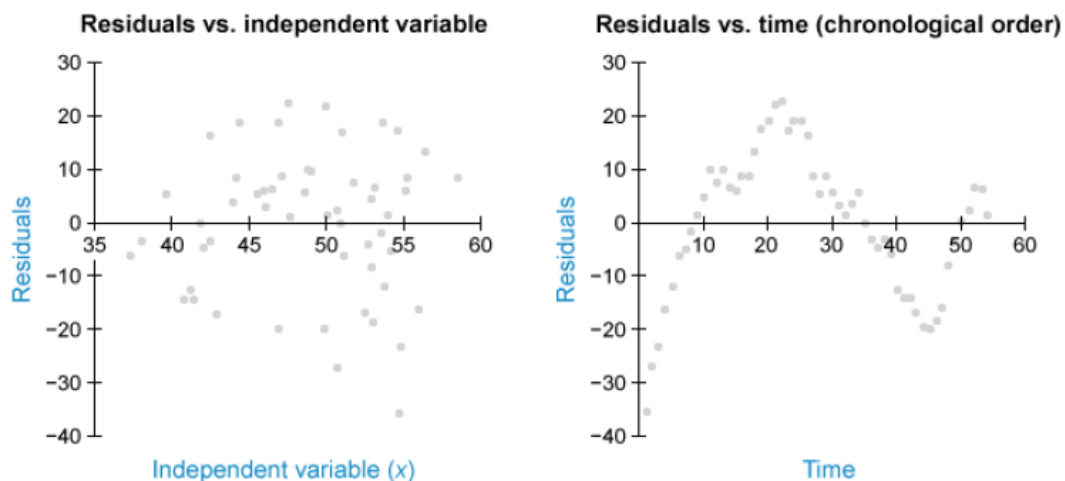
Regression statistics	
Coefficient of determination (R^2)	0.541
Standard error (S_e)	1.841
Observations (quarters) (n)	12
Coefficients	
Intercept	9.358
Crude oil	0.561
Descriptive statistics (Crude oil)	
Mean (\bar{X})	55.408
Standard deviation (S_x)	3.397

The analyst forecasts an average oil price of USD 66 in the following quarter. Considering t -values of ± 2.228 , the 95% prediction interval for the company's gross margin is *closest* to:

- A. 40.6% to 52.1%.
- B. 42.3% to 50.5%.
- C. 44.2% to 48.6%.

Question 6

An analyst runs a simple linear regression (SLR) and plots the resulting residuals:



The regression's residuals are presented on the vertical axis of each graph. In the left graph, residuals are plotted against the regression's independent variable, and in the right graph, they are plotted in chronological order. Based only on the graphs, the model most likely violates which of the following linear regression assumptions?

- A. Linearity
- B. Independence
- C. Homoskedasticity

Question 7

The following table summarizes the relationship between two exchange rates, USD/GBP and USD/EUR for the last 12 months:

Statistic	Data
Standard deviation of USD/GBP	1.389
Standard deviation of USD/EUR	1.208
Correlation of USD/GBP and USD/EUR	0.781

An analyst runs a linear regression to test if USD/GBP explains changes in USD/EUR. The slope of this regression model is *closest* to:

- A. 0.405
- B. 0.679
- C. 0.898

Question 8

An economist builds a linear regression model (Model 1) to investigate if the unemployment rate can be used to anticipate trends in a stock index. Then, the economist builds a second model (Model 2) to test if the stock index is a good predictor of trends in GDP. The stock index is the dependent variable:

- A. in both models.
- B. only in Model 1.
- C. only in Model 2.

Question 9

A portfolio manager uses a regression model to evaluate how a company's debt-to-equity ratio influences its beta. The analysis of variance (ANOVA) for the regression model is:

ANOVA Table				
Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-statistic
Regression	1	0.1426	0.1426	?
Residual	8	0.2986	0.0373	
Total	9	0.4413		

The manager formulates the null hypothesis (H_0) that the slope coefficient is zero. If the critical F -value at a 5% significance is 5.32, the manager's most appropriate decision is to:

- A. reject H_0 since the F -statistic is less than 5.32
- B. reject H_0 since the F -statistic is greater than 5.32.
- C. fail to reject H_0 since the F -statistic is less than 5.32.

Question 10

An analyst runs a regression to determine whether the value of Stock Index A is a good predictor of the value of Stock Index B. Selected data are presented below:

	Coefficient	Standard Error
Intercept	31.930	91.950
Index A	3.587	0.198

The analyst expects Index A to reach a value of 509 next month and estimates the standard error of the forecast (s_f) to be 42.893. If the critical t -values at a 10% significance level are 1.701 (two-sided) and 1.313 (one-sided), the prediction interval for Index B is *closest* to:

- A. 1,785 to 1,931.
- B. 1,801 to 1,914.
- C. 1,815 to 1,901.

Question 11

An analyst evaluates the relationship between two securities and runs a linear regression based on 12 observations. Selected data from the ANOVA (analysis of variance) table is presented below:

Source	Sum of squares
Regression	20.601
Residual	8.326
Total	28.927

Based on this information, the F -statistic is *closest* to:

- A. 24.74
- B. 27.22
- C. 29.69

Question 12

A fixed income analyst runs a linear regression to evaluate the relationship between the price of two bonds. Selected data is presented below:

ANOVA Table			
Source	Degrees of freedom	Sum of squares	Mean squares
Regression	1	20.032	20.032
Residual	28	159.371	5.692
Total	29	179.404	

At a 5% significance level, the critical F -value is 4.196. Based only on the data, the *most appropriate* conclusion is that the:

- A. coefficient of determination (R^2) is 12.6%.
- B. standard error of the estimate (s_e) is 2.386.
- C. model is statistically significant since the F -statistic is 3.52.

Question 13

An analyst runs a simple linear regression using a sample of 10 observations. To assess the validity of the regression statistics, the analyst randomly draws 20 observations from the same data. The results of both regression analyses are:

	Sample size	Sum of squares total (SST)	Sum of squares errors (SSE)
Regression 1	10	167.00	47.75
Regression 2	20	334.00	95.50

Based on this information, the analyst should *most likely* conclude that:

- A. Regression 1 has a larger F -statistic.
- B. both regressions have the same F -statistic.
- C. Regression 2 has a larger F -statistic.

Question 14

A foreign exchange trader runs a linear regression with 24 observations in which BRL/USD is the independent variable and MXN/USD is the dependent variable and compiles the following results:

	$\frac{BRL}{USD}$	$\frac{MXN}{USD}$	Total sum of squares	Residual sum of squares
	X_i	Y_i	$\sum_{i=1}^n (Y_i - \bar{Y})^2$	$\sum_{i=1}^n (Y_i - \hat{Y})^2$
Mean	4.993	21.033		
Sum			4.932	4.190

The standard error of the estimate is *closest* to:

- A. 0.418
- B. 0.436
- C. 0.473

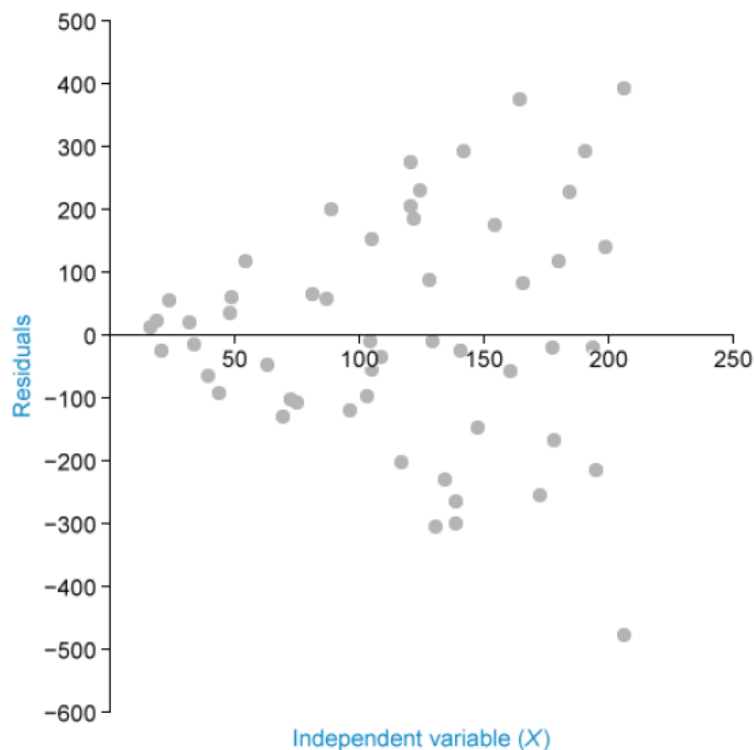
Question 15

Given that the goal of a linear regression is to explain the variation of the dependent variable Y, the sum of squares regression (SSR) is *best described* as the:

- A. total variation of Y.
- B. explained variation of Y.
- C. unexplained variation of Y.

Question 16

After running a simple linear regression, a trader analyzes the resulting residual plot:



Based on only the residual plot, this regression model *most likely* violated which of the following linear regression assumptions?

- A. Linearity
- B. Independence
- C. Homoskedasticity

Question 17

An analyst runs a linear regression between two variables but infers that they have a nonlinear relationship. Which of the following approaches would be *most appropriate* to address this issue?

- A. Transform one or both variables
- B. Use an indicator (dummy) variable
- C. Increase the number of observations

Question 18

A linear regression will most likely provide valid conclusions even if:

- A. the independent variable is not normally distributed.
- B. there is a strong seasonal pattern in a time series data.
- C. the dispersion of the dependent variable changes as the independent variable increases.

Question 19

The following table summarizes performance data for a mutual fund and its benchmark for the last 50 months:

	Fund Return (Y)	Benchmark Return (X)
Mean	1.48	1.20
Variance	5.60	5.07
Covariance (Y,X)		4.78

Based on this data, if an investor expects a 0% return for the benchmark next month, the predicted return of the mutual fund will be *closest* to:

- A. 0.35%
- B. 0.94%
- C. 1.48%

Question 20

An equity analyst evaluates how the percentage change in cotton prices affects the percentage change in the stock price of an apparel company. The analyst prepares a linear regression, resulting in an intercept of 0.3 and a slope of -0.7 . According to this regression, the predicted change in stock price given a 2% increase in cotton prices is *closest* to:

- A. -1.4%
- B. -1.1%
- C. -0.1%

Question 21

A trader runs a linear regression model and calculates a prediction interval. The trader reviews the results, then runs a second model based on the same data set and calculates a significantly narrower prediction interval. Based only on this information, the trader *most likely*:

- A. decreased the sample size.
- B. increased the level of significance.
- C. assumed the forecasted independent variable is farther from the sample mean.