Curtin University, Semester 2, 2026

ECOM5005 Tutorial 10

Simple Linear Regression

Learning Outcomes

1. Conduct a simple regression and interpret the meaning of the regression coefficients.
2. Use regression analysis to predict the value of a dependent variable.
3. Make inferences about the slope.
4. Make inferences about the correlation coefficient.

PART 1 – Short answer questions

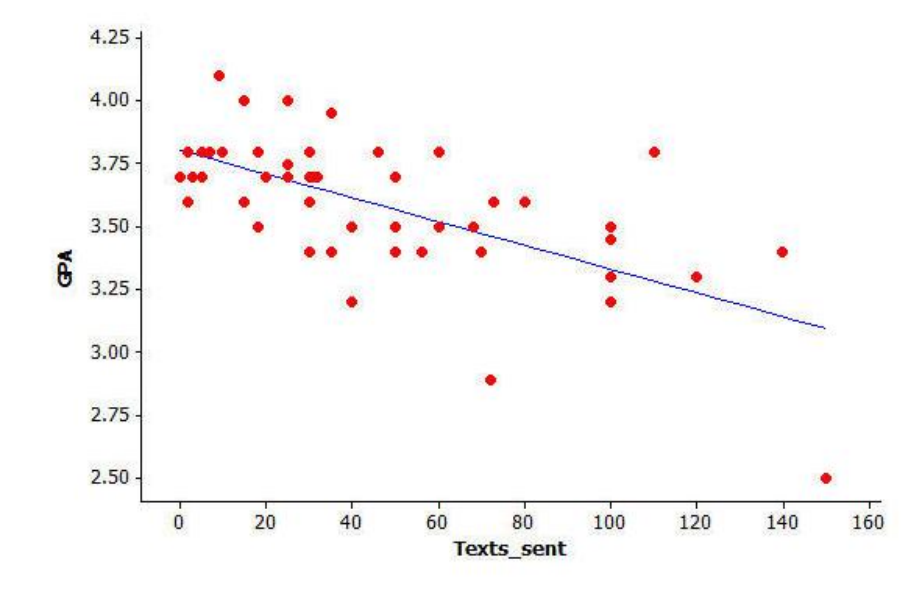
Question 1. High blood pressure

In a study of college freshmen, researcher found that students who watched TV for an hour or more on weeknights were significantly more likely to have blood pressure, compared to those students who watched less than an hour of TV on weeknights. Does this mean that watching more TV raises one's blood pressure? Explain your reasoning.

Question 2. Testing and Grades

A researcher suspects that there is a relationship between the number of text messages high school students send and their academic achievement. To explore this, she asks a random sample of 52 students at her school how many text messages they sent yesterday and what their grade point average (GPA) was during the most recent marking period. Her data are summarised in the scatter plot below. The line of best fit is also shown.

The equation of the line of best fit is GPA-hat = 3.8 - 0.005 (Texts sent). Interpret the quantities -0.005 and 3.8 in the context of these data.



Question 3. Used Subaru Foresters

Jane wants to sell her Subaru Forester and does research online to find other cars for sale in her area. She checks on carsales.com.au and finds 22 Subaru Foresters recently listed, along with their distance travelled (in km), age (in years), and listed price (in dollars).

She examines the scatterplot of price versus age and determines that a linear model is appropriate. She finds the equation of the least squares regression equation:

predicted price = 24,247.56 - 1482.06 age

a. What variable is the explanatory (independent) variable and what are the units it is measured in? What variable is the response (dependent) variable and what are the units it is measured in?

b. What is the slope of the least squares regression line and what are its units?

c. Interpret the slope of the least squares regression line in the context of the problem, discussing what the slope tells you about how price and age are related. Use appropriate units in your answer.

d. What is the y-intercept of the least squares regression line? Interpret the y-intercept in the context of the problem, including appropriate units.

Question 4. Digital advertising and weekly online sales

An online retailer analyses 24 regional advertising campaigns. Weekly digital advertising spend X is measured in $000s and weekly online sales Y is measured in $000s. The fitted model is:

predicted weekly sales = 58.4 + 4.72(advertising spend).

The regression output reports R-squared = 0.64 and the standard error of the slope is 0.7547. The observed advertising-spend range is $5,000 to $25,000.

a. Identify the explanatory and response variables and interpret the slope and intercept in context.

b. For a region that spent $18,000 on digital advertising, calculate the predicted weekly sales. If actual sales were $151,000, calculate and interpret the residual.

c. Test H0: beta1 = 0 against H1: beta1 != 0 at the 5% significance level. State the degrees of freedom and conclusion.

d. Calculate the sample correlation coefficient r and interpret its sign and magnitude.

e. Should the model be used to predict sales for an advertising spend of $50,000? Explain.

Question 5. Emergency-department staffing and patient waiting time

A hospital studies 32 evening shifts to examine whether staffing is associated with mean patient waiting time. Let X be the number of nurses on duty and Y be the average waiting time in minutes. The fitted regression is:

predicted wait = 96.4 - 3.85(nurses).

The standard error of the slope is 0.95 and R-squared = 0.354. The observed staffing range is 7 to 15 nurses. On one evening with 12 nurses, the actual mean waiting time was 61 minutes.

a. Interpret the slope and intercept. Is the intercept practically meaningful?

b. Predict the waiting time with 12 nurses and calculate the residual for the observed 61-minute waiting time.

c. Test whether there is a statistically significant linear relationship at alpha = 0.05.

d. A residual plot has a clear fan shape: the spread of residuals increases at higher staffing levels, and there is one unusually large positive residual near 14 nurses. Which regression assumptions should concern you?

e. Can the hospital conclude that adding nurses causes shorter waiting times? Explain.

Question 6. Streaming promotions, subscriptions and autocorrelation

A streaming service regresses monthly new subscriptions Y (000s) on monthly promotional spending X ($000s) for 36 consecutive months. The fitted line is:

predicted new subscriptions = 12.6 + 1.85(promotional spending).

The regression reports R-squared = 0.49 and a Durbin-Watson statistic of D = 0.72. A time plot of the residuals shows long runs of positive residuals followed by long runs of negative residuals.

a. Interpret the slope and calculate the predicted number of new subscriptions when promotional spending is $20,000.

b. Use R-squared and the positive slope to determine the sign and approximate value of r.

c. What does D = 0.72 suggest about the independence-of-errors assumption?

d. Why should the analyst be cautious about the usual t tests, p-values and confidence intervals from this regression?

e. Suggest one sensible next modelling step before using this relationship for business decisions.

PART 2 – Data exercise: Engine size and highway fuel economy

The supplied Excel workbook contains 234 observations for popular vehicle models from model years 1999 and 2008. The data are derived from U.S. EPA fuel-economy records and distributed in the mpg teaching dataset. In this exercise, investigate the simple linear relationship between engine displacement (litres) and highway fuel economy (MPG).

(a) Identify the explanatory variable X and response variable Y. Create a scatter plot of highway MPG against engine displacement. Describe the direction, form and strength of the relationship.

(b) Use Excel Data Analysis ToolPak > Regression to estimate the simple linear regression model. Report the fitted equation.

(c) Interpret the slope and intercept in context. Comment on whether the intercept has a useful practical interpretation.

(d) Report and interpret R-squared and the standard error of the estimate.

(e) Test H0: beta1 = 0 against H1: beta1 != 0 at alpha = 0.05 using the regression output. Also calculate the correlation coefficient using Excel CORREL.

(f) Use the fitted equation to predict highway MPG for a vehicle with a 3.0-litre engine. For the first observation in the dataset (Audi A4, 1.8 L, highway MPG = 29), calculate its predicted MPG and residual.

(g) In the Regression dialog, request Residuals, Residual Plots, Line Fit Plots and Normal Probability Plots. Use these outputs to assess the LINE assumptions. State which assumptions appear reasonable and which deserve caution.

(h) Explain why the regression should not be interpreted as proving that increasing engine displacement causes a particular change in fuel economy.

PART 3 – Pearson MyLab

1. Please ensure that you have completed Quiz 5 will be due on this Friday (23-Oct-2026) at 5:pm AWST.

Note: If you are having issues with obtaining access to MyLab, please re-read the instructions (available in the welcome topic). If you are still unsuccessful, please send an email to the teaching team explaining the issue. It is important that we complete MyLab tasks on a weekly basis since the final e-test in this unit will be completed using MyLab.