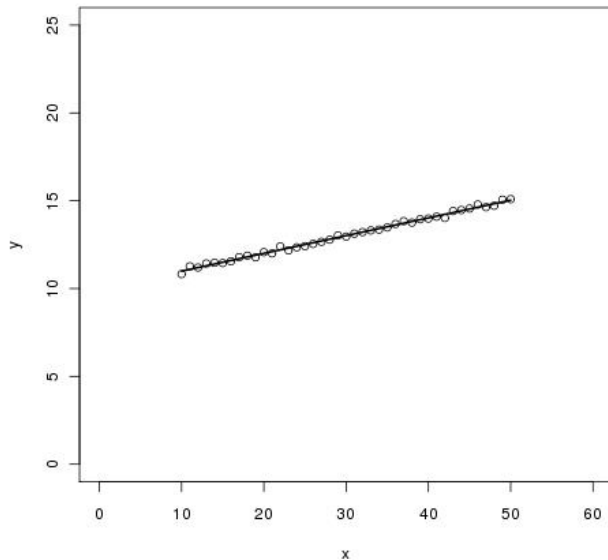
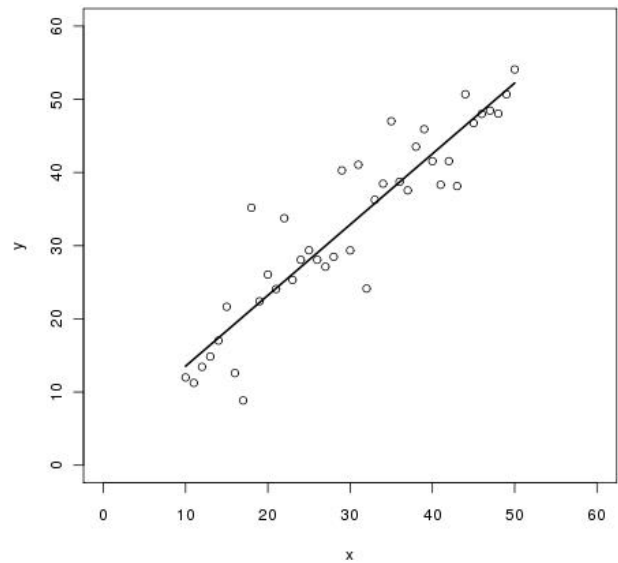
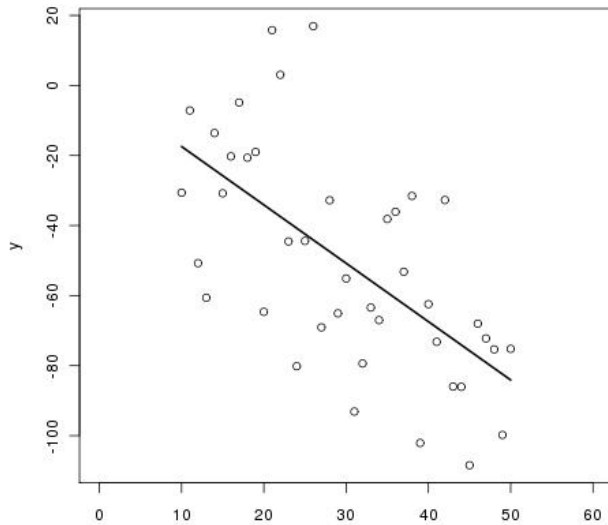


ACE2013: Statistics for Marketing and Management: Supplementary handout 2

Simple linear regression analyses were performed on three sets of paired observations (x, y) . The resulting scatterplots, with superimposed regression lines, are shown below.

Use the information in the table to write down:

1. the correlation coefficient;
2. the full regression equation;
3. the R^2 statistic, for each set of paired observations.



$\hat{\beta}_0$	2, -3, 10
$\hat{\beta}_1$	0.1, 1, -1.4
$\hat{\sigma}$	5, 0.1, 30
r	0.98, 0.8, -0.6
R^2	0.64, 0.96, 0.36

Exam-style question

A business analyst has been given the task of investigating factors that could influence a factory's productivity. The business analyst requests the factory's monthly productivity figures for the last year (Y , in thousands of units), as well as the monthly total number of overtime hours worked (X_1), the monthly total number of sickness days recorded (X_2) and staff morale (X_3 – an average score out of 10). In certain months of the year, the factory also has a “shut-down” period of a week or more – usually December (for Christmas), January (for New Year) and August (the “summer shut-down”). Thus, the business analyst defines a fourth variable (X_4) that takes the value 1 if a particular month has a “shut-down” period, and zero otherwise.

- Write down the general form of the full regression model that could be used to explain the factory's productivity. Which variable is an *indicator variable*?
- The business analyst uses Minitab to fit the model you specified in part (a); the resulting output, representing the “full model”, is shown below. Look at it, and then answer the questions that follow.

Regression Analysis: Productivity versus overtime hours, sickness day, ...

The regression equation is
 Productivity = 1.1 + 0.443 overtime hours - 1.39 sick days - 0.258 Morale
 - 1.38 Shut-down

Predictor	Coef	SE Coef	T	P
Constant	1.10	10.29	0.11	0.919
overtime hours	0.4433	0.2869	A	0.183
sick days	-1.3936	B	-2.03	0.098
Morale	-0.2584	0.7883	C	0.756
Shut-down	-1.381	D	-1.10	0.320

S = 1.15872 R-Sq = 95.3% R-Sq(adj) = 91.6%

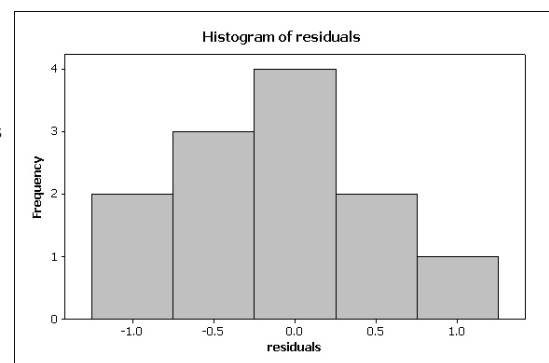
- Write down the full estimated regression equation.
 - Find **A**, **B**, **C** and **D**.
 - Which variable should be removed before the business analyst re-fits the multiple linear regression model? Explain your answer.
 - Comment on the R^2 statistic.
- (c) After re-analysing the productivity data, the business analyst obtains the following output from Minitab representing the “final model”:

Regression Analysis: Productivity versus overtime hours, sickness hours

The regression equation is
 Productivity = - 1.33 + 0.458overtime hours - 1.53sick days

Predictor	Coef	SE Coef	T	P
Constant	-1.333	9.084	-0.15	0.888
overtime hours	0.4582	0.1741	2.63	0.034
sick days	-1.5269	0.6316	-2.42	0.046

S = 1.09296 R-Sq = 94.2% R-Sq(adj) = 92.5%



- Explain what has happened between “full model” fit and the “final model” fit.
- Comment on the R^2 statistic in the “final model” and compare it to that in the “full model”.
- Comment on the histogram of residuals.
- What can we expect productivity to be when there is a total of 36 hours of overtime worked and 3 days sickness taken?