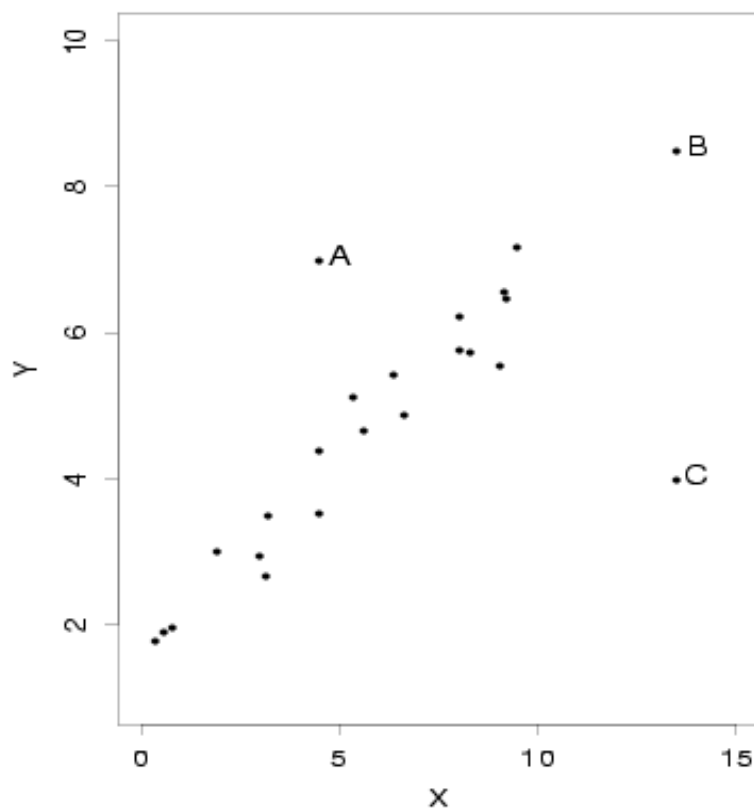


Total marks = 30

Notes: Short answers are preferable to long answers (answers in “point form” are acceptable).



1. Consider a linear regression model that involves just one explanatory variable

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad \epsilon \sim N(0, \sigma)$$

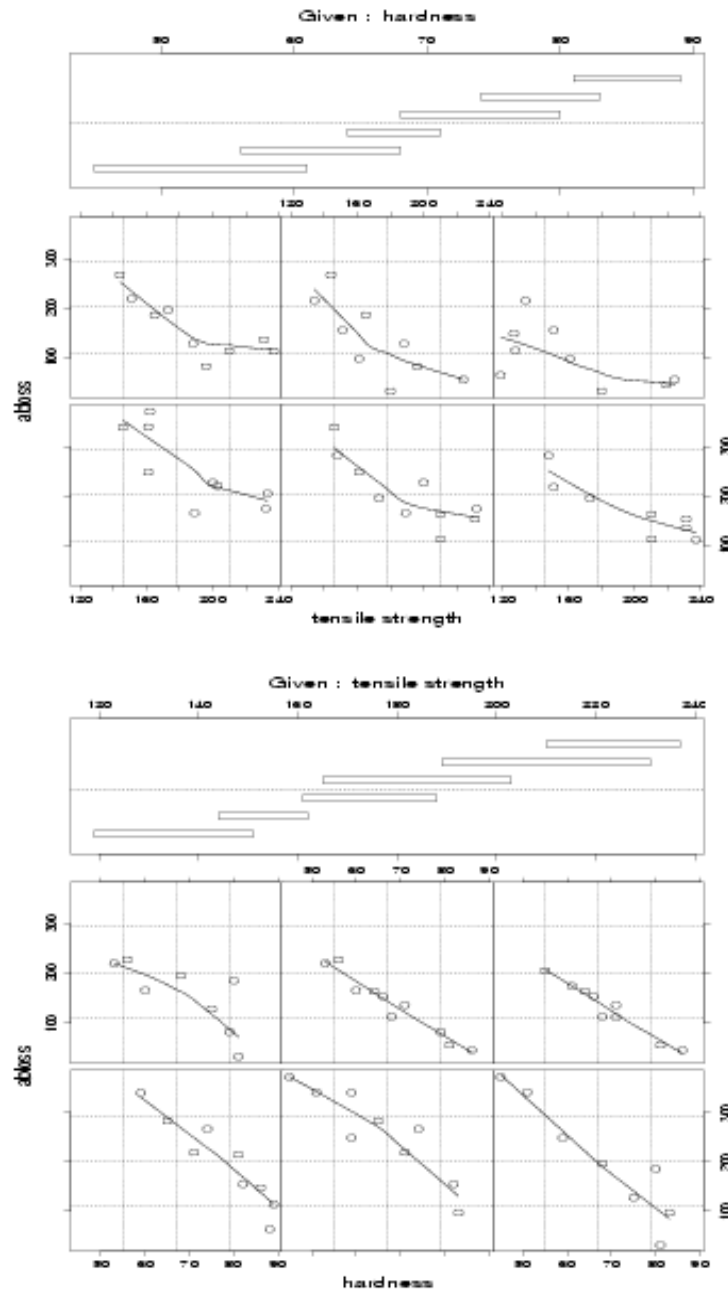
and suppose that the data contains a single unusual observation. The points labelled A, B and C in the above figure above illustrate three possible ways an observation could be unusual. Now consider three cases: for each case the data set consists of the unlabelled points and **one** of the labelled points. For each of these three cases:

- Describe how the labelled observation is unusual (is it an outlier? a high leverage point? an influential point? all three?). [3 marks]
- Explain what effect deleting the labelled point would have on the fitted model (How would $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}$ be affected?). [4 marks]

2. Example 3 in your course notes contains some tyre abrasion data. The idea was to relate the amount of abrasion in a standard abrasion test to the hardness and tensile strength of the rubber. Two trellis plots are given below for this data set.

(a) Use these plots to briefly explain how abrasion loss (abloss) is related to hardness and tensile strength. [4 marks]

(b) It is desirable to have low abrasion loss. Clearly, indicate what combination of values for hardness and tensile strength result in low values of abloss. [2 marks]



3. Creatinine clearance (`clearance`) is an important measure of kidney function but is difficult to obtain because it requires a 24 hour urine collection. Data was collected by a kidney specialist to determine whether creatinine clearance can be predicted using measurements of creatinine concentration in mg per decilitre (`conc`), age in years (`age`), and weight in kg (`weight`) since these measurements are much easier to collect.

Summary statistics were generated for the data using S-plus:

```
> summary(kidney.df)
      clearance      conc      age      weight
Min.   : 30.00  Min.   :0.68  Min.   :16.00  Min.   : 50.00
1st Qu.: 59.00  1st Qu.:0.94  1st Qu.:42.00  1st Qu.: 65.00
Median : 85.00  Median :1.13  Median :63.00  Median : 71.00
Mean   : 85.21  Mean   :1.25  Mean   :55.97  Mean   : 72.55
3rd Qu.:110.00 3rd Qu.:1.50  3rd Qu.:69.00  3rd Qu.: 80.00
Max.   :140.00  Max.   :2.52  Max.   :78.00  Max.   :107.00
```

The following model was used.

```
>kidney.fit3<-lm(clearance ~ conc + I(conc^2) + age + weight, data = kidney.df)
> summary(kidney.fit3)
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	151.2033	21.1118	7.1620	0.0000
conc	-84.8245	23.2816	-3.6434	0.0011
I(conc^2)	15.4265	7.7885	1.9807	0.0575
age	-0.7324	0.1348	-5.4341	0.0000
weight	0.7431	0.1647	4.5125	0.0001

Residual standard error: 11.87 on 28 degrees of freedom

Multiple R-Squared: 0.8727

F-statistic: 47.97 on 4 and 28 degrees of freedom, the p-value is 3.896e-12

- (a) Write down the theoretical form (i.e. $\text{clearance} = \beta_0 + \dots$) for this model. [2 marks]
- (b) What does each of the following lines from the output indicate about the fitted model? [2 marks each]
- I(conc^2) 15.4265 7.7885 1.9807 0.0575
 - Multiple R-Squared: 0.8727
- (c) Describe how the fitted model relates age to creatinine clearance (be more specific than "creatinine clearance increases/decreases as age increases"). [2 marks]
- (d) How much difference in creatinine concentration does the fitted model predict between a patient for has `conc = 1` and `conc = 2` (assuming age and weight are the same)? [3 marks]

4. (a) What does an unusual value of the *covariance ratio* statistic for an observation indicate? Approximately, what value of covariance ratio do you expect for an observation that does not have a large effect on the regression analysis? [3 marks]
- (b) What problem are variance inflation factors (VIFs) used to detect? Explain how VIF's are used diagnose this problem (What do the VIFs actually measure? What range of values can the VIFs take? What values indicate that a problem is present?) [3 marks]