

1. (a) **Case A:** outlier, low leverage, low/moderate influence
Case B: not an outlier, high leverage, low influence
Case C: outlier, high leverage, high influence
- (b) **Case A:** deleting A would result in small changes to $\hat{\beta}_0$ and $\hat{\beta}_1$, and a decrease in $\hat{\sigma}$
Case B: deleting B would result in little change in $\hat{\beta}_0$, $\hat{\beta}_1$, or $\hat{\sigma}$
Case C: deleting C would result in a decrease in $\hat{\beta}_0$, a big increase in $\hat{\beta}_1$, and a decrease in $\hat{\sigma}$
2. (a) If tensile strength is fixed, abrasion loss decreases in a (approximately) linear manner as hardness increases. The rate of decrease is about the same for all levels of tensile strength. If hardness is fixed, the trend between tensile strength and abrasion loss has a bent shape: abrasion loss decreases with increased tensile strength until the value of tensile strength reaches approximately 190 and then the trend flattens off. The shape of the relationship between tensile strength and abrasion loss is reasonably consistent for all levels of hardness.
- (b) The lowest values of abrasion loss occur when hardness is near its maximum value (just less than 90) and for values of tensile strength between 190 and 240.
3. (a) $\text{clearance} = \beta_0 + \beta_1 \text{conc} + \beta_2 \text{conc}^2 + \beta_3 \text{age} + \beta_4 \text{weight} + \epsilon$
- (b) i. $\hat{\beta}_2 = 15.42$, $\text{se}(\hat{\beta}_2) = 7.79$, if we test $H_0: \beta_2 = 0$, given that all the other terms are in the model, we get $t_0 = 1.98$ and a p-value of 0.0575 (some evidence that conc^2 should be kept in the model).
ii. the fitted regression surface explains 87.72% of the variability in the observed creatinine clearance measurements
- (c) For each increase of 1 year in the age of a patient, creatinine clearance decreases on average by 0.73 units, provided conc and weight are held constant
- (d) For $\text{conc} = 1$ the contribution due to conc is $-84.82 \times 1 + 15.43 \times 1^2 = -69.39$. For $\text{conc} = 2$, the contribution due to conc is $-84.82 \times 2 + 15.43 \times 2^2 = -107.92$. Creatinine clearance will be 38.5 units lower for $\text{conc} = 2$ than $\text{conc} = 1$.
4. (a) An unusual value of covariance ratio indicates that the observation has a large influence on the standard errors and covariances of the estimated coefficients. If an observation does not have a large effect it should have a covariance ratio that is approximately 1.
- (b) VIFs are used to detect multicollinearity (near linear relationships among the explanatory variables). VIFs measure how much the variance of each estimated coefficient is being inflated by multicollinearity. VIFs can take any value greater than 1. If a regressor is completely unrelated (linearly) to the other regressors then it will have a VIF of 1. The stronger the linear relationship between a regressor and the other regressors the larger the VIF. So any regressor with a large VIF is involved in a near linear relationship with some of the other regressors. VIFs greater than 10 indicate relationships that are very close to linear.