1. **ANOVA** one factor Data File Table 6.19-PlasmaPhosphate.xls Please refer to Dobson Exercise 6.5 and answer the following questions
   - (a) Test the hypothesis that there are no mean differences among the three groups.
   - (b) Assume independent groups and normality with common variance, obtain a 95% confidence interval for the difference in means between the hyperinsulinemic obese group and the nonhyperinsulinemic obese group.
   - (c) Using an appropriate model, plot the standardized residuals against the observation index. Also produce a normal probability plot for the standardized residuals.

2. **ANOVA** two factors with unbalanced data Data File Table 6.21-UnbalancedData.xls Please refer to Dobson Exercise 6.8 and answer the following questions.
   - (a) Test the hypothesis (at 5%) that there are no interaction effects
   - (b) Test the hypothesis (at 5%) that there is no effect due to Factor A by comparing the models

     $$\mathrm{E}\left[Y_{jkl}\right] = \mu + \alpha_j + \beta_k \quad \mathrm{E}\left[Y_{jkl}\right] = \mu + \beta_k$$

   - (c) Test the hypothesis (at 5%) that there is no effect due to Factor A by comparing the models

     $$\mathrm{E}\left[Y_{jkl}\right] = \mu + \alpha_j \quad \mathrm{E}\left[Y_{jkl}\right] = \mu$$

   - (d) Compare your conclusions for (b) and (c) and explain the difference.

3. **one-factor ANOVA** You are given the following information Table 1 and Table 2 below:

Group	Count	Average	Variance
Hypinsul&obese	11	4.19	0.7929
Non-hypinsul&obese	10	3.58	0.1707
Control	12	2.93	0.4584

Table 1: One Factor with 3 level Summary

Source of Variation	Sum sq	df	MS	f-stat	p-value
Between Groups	*	2	*	*	*
Within Groups	*	*	*		
Total	23.7218	32			

Table 2: One-factor ANOVA

   - (a) Please calculate the sum sq for Within Groups in Table 2 above. (*hint*: Within Groups Sum sq is the sum of the residual sum of squares over three groups. You could get the info from Table 1.)
   - (b) Please fill in the rest of the asterisks in Table 2 above.

(c) Please assert whether the three groups in this one factor ANOVA model have the same mean.

4. **National Life Expentancies**. Referred to Frees' Exercise 5.3 and 5.4.

   (a) Begin the data set from $n = 185$ countries that have valid (nonmissing) life expectancy (LIFEEXP). (Note that: Check the whether data valid for a variable x, sum(data,is.na(x))). Plot the LIFEEXP versus GDP (gross domestic product) and PRIVATEHEALTH (private expenditures on health). From these plots, describe why it is desirable to use logarithmic transforms, lnGDP and lnHEALTH, respectively. ALso plot LIFEEXP versus lnGDP and lnHEALTH to confirm your intuition.

   (b) Return to the full dataset of $n = 185$ countries and run a regression model using FERTILITY, FUBLICEDUCATION, and lnHEALTH as explanatory variables.

      (i) Provide plots of standardized residuals.

      (ii) Identify all outliers (outlier defined as the standardized residual greater then 3 in absolute value), (hint: which.is.max(x)).

      (iii) Provide plot of leverage.

      (iv) Identify all high leverages (high leverage defined as $h_{ii} > 3 \times \bar{h}$).

      (v) Identify a data point if any that is both outlier and high leverage. Then calculate the cook's distance for the point in b(iv), determine the decomposition of the distance, the attribution from outlier and high leverage respectively.

   (c) Variance Inflation Factors.

      (i) Brief explain the idea of of collinearity and a variance inflation factor.

      (ii) What constitutes a large variance inflation factor?

      (iii) Calculate the VIF for lnHEALTH by obtaining R^2_{lnHEALTH}.

      (iv) Calculate all three explanatory variables without rerun the regression using the relation of VIF to $se(b_j)$ that explained in class.