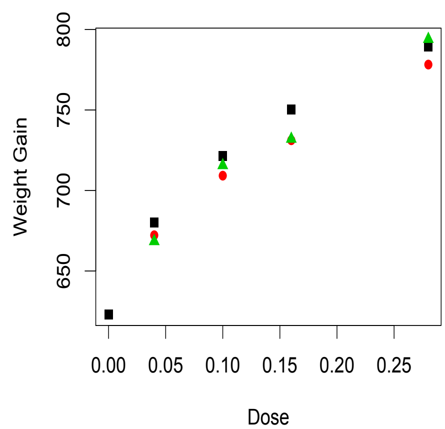
Homework 6 (Written Section)

1. Show that $\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

2. Cook and Weisberg (1999) describe an experiment with turkey growth. Methionine is an amino acid essential for normal growth in turkeys; if they have too little, the birds can be malnourished, but if they have too much, it could be toxic. To find the optimal level of methionine for the birds, an experiment was conducted. A total of 60 pens of turkeys were randomly assigned to one of three commercial sources of methionine and one of four doses of methionine, ranging from 0.04% to 0.28% of the total diet. (Five pens were assigned to each treatment combination.)

   An additional 10 pens were assigned to a control group receiving no methionine at all. A plot of the relationship between dosage and average weight gain for the birds from each of the groups of five pens, along with the average from the 10 control pens, is shown below. (Black squares are the averages for Source 1, red circles for Source 2, and green triangles for Source 3.)



   (a) Would you suggest weighting this model? Why or why not? If so, what weights would you use?

   (b) Would you suggest using a polynomial model? Why or why not? If so, what order polynomial would you use?

   (c) Would you suggest using separate lines or polynomials for the three groups? Would you use separate intercepts? Would you use interactions? Explain.

   (d) Write down your suggested model, using variables x = dosage, $iSource2$ = indicator for being methionine from Source 2, and $iSource3$ = indicator for being methionine from Source 3.

3. In a one-way ANOVA model with k = 3 groups and 4 observations per group:

   (a) Use the F-statistic in Model Reduction Method 2 to derive a statistic for testing whether the average of the means of the first two groups is the same for the mean of the third group. That is, create the F-statistic for testing $H_0 : (\mu_1+\mu_2)/2 = \mu_3$. (Hint: Don't fit a model with a y-intercept. It makes everything easier.)

   (b) Where $\hat{\mu}_1 = 5.6, \hat{\mu}_2 = 7.9, \hat{\mu}_3 = 6.1$, and SSE = 12.8, test your hypothesis. Use $\alpha = 0.05$.

4. (Old Qualifying Exam Question) A randomized trial was conducted to investigate the relationship between a continuous response y and four treatments A, B, C, and D. The sample size was $n = 200$, with 50 observations in each of the four treatment groups. Let y be the 200×1 vector of response values, ordered so that the first 50 entries are for treatment group A, the next 50 for B, then C, and finally D. The regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ was fit, where $\mathbf{X}$ is the 200×4 design matrix given by:

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

and where each entry is a column vector of length 50. The estimated regression coefficients were $\hat{\boldsymbol{\beta}}' = [37.5, -11.5, 1.0, -27.7]$, with standard errors 2.75, 3.89, 3.89, 3.89, and residual standard deviation $\hat{\sigma} = 19.45$. Also:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.02 & -0.02 & -0.02 & -0.02 \\ -0.02 & 0.04 & 0.02 & 0.02 \\ -0.02 & 0.02 & 0.04 & 0.02 \\ -0.02 & 0.02 & 0.02 & 0.04 \end{bmatrix}$$

   (a) Interpret each of the four regression parameters.

   (b) What assumptions are required for the regression model?

   (c) What is an approximate 95% confidence interval for the mean difference in response between treatment groups B and A (so, the difference $\mu_B - \mu_A$)?

   (d) What is an approximate 95% confidence interval for the mean response in treatment group B?