

## Homework 3

(Due at 10 am on Mon. Oct. 08)

Please submit your assignment *on paper*, following the Guidelines for Homework posted at course website. (Even if correct, answers might not receive credit if they are too difficult to read.) Remember to include relevant computer output.

1. Show the following:

(a)  $\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i$

(b)  $E(\hat{\sigma}^2) = \sigma^2$

(c) If the errors are normally distributed, show that  $\hat{\beta}$  and  $\hat{\sigma}^2$  are independent.

(d) If the errors are normally distributed, the statistic

$$T := \frac{\hat{\beta}_1}{se(\hat{\beta}_1)},$$

has a  $t$  distribution when  $\beta_1 = 0$ .

2. Consider multiple linear regression with  $Y = (1, -2, 0, 1, -1, 2)$  and  $X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{pmatrix}$ .

We would like to estimate the parameters (i)  $\beta_0 + \beta_1 - \beta_2$ , and (ii)  $\beta_1 + \beta_2$ .

Derive the following quantities using the formulas and derivations in the class notes and not using any software (such as R, etc). You can use a calculator for basic computations such as matrix multiplication, inverse, etc. Show all your work.

- (a) The best linear unbiased estimator (BLUE) values for both these parameters.
  - (b) The standard errors for the estimators in the previous part.
  - (c) The correlation between the BLUEs of (i)  $\beta_0 + \beta_1 - \beta_2$  and (ii)  $\beta_2$ .
3. Using the `sat` data (from faraway package), perform the following analysis.
    - (a) Fit a model with total `sat score` as the response and `takers`, `ratio` and `salary` as predictors. Comment on the goodness of fit.
    - (b) Suppose you wish to claim that teachers' salary has a positive effect on the SAT scores. State the appropriate null and alternative hypothesis, the test statistic, the  $p$ -value, and your conclusion at significance level  $\alpha = 0.01$ .
    - (c) Test the hypothesis that the variable `ratio` has an effect on the SAT scores in the model from part (a).
    - (d) Test the hypothesis  $\beta_{takers} = \beta_{ratio} = \beta_{salary} = 0$ . Explain in words what this hypothesis means.

- (e) Compute the 95% and 99% CIs for the parameter associated with **salary**. Using just these intervals, what can we deduce about the  $p$ -value for **salary** in the regression summary?
  - (f) Compute and display 95% CIs for the parameters associated with **ratio** and **salary**.
  - (g) Now add **expend** (current expenditure per pupil) to the model and comment on the coefficients, their significance and the goodness of fit as compared to the model in part 3a.
  - (h) In the model of part 3g, test the hypothesis  $\beta_{\text{salary}} = \beta_{\text{expend}} = \beta_{\text{ratio}} = 0$ . Based on your entire analysis, do you feel any of these predictors have an effect on the response?
4. Using the **happy** data set (from faraway package), fit a model with **happy** as the response and the two variables (money and work) as predictors. Perform regression diagnostics on this model to answer the following questions. Display any plots that are relevant. Do not provide any plots about which you have nothing to say. Suggest possible improvements or corrections to the model where appropriate.
- Check the constant variance assumption for the errors.
  - Check the normality assumption.
  - Check for large leverage points.
  - Check for outliers.
  - Check for influential points.
  - Check the structure of the relationship between the predictors and the response.

Do the same diagnostic checks for the model with **happy** as the response and all of the other four variables (money, sex, love, and work) as predictors. What do you observe in comparison to the previous model diagnostics?