

Take Home

June 4, 2018

Show all your work. You need to submit your report electronically to rguhaniy@ucsc.edu by 6/10/2018 11:59 PM. The report is limited to 8 pages. Attach your code and it will not be counted within the 8 page limit.

1. Consider the linear regression model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2).$$

Use the following data. Please show your work. Do not use R package to run linear

y:	82	79	74	83	80	81	84	81
x_1	10	9	9	11	11	10	10	12
x_2	15	14	13	15	14	14	16	13

regression. Using R for simple algebra is okay.

- (a) Provide the least square estimates of β_1 , β_2 and σ^2 . (5 points)
 - (b) Provide 95% confidence intervals for β_1 and $2 * \beta_1 + \beta_2$. (10 points)
 - (c) Perform a $\alpha = 0.01$ level test for $H_0 : \beta_2 = 3$. (5 points)
 - (d) Find p-value for the test $H_0 : \beta_1 = \beta_2$. (5 points)
2. Consider the setting and the dataset in the previous question. Use the R package to run linear regression. Provide
 - (a) p-value for testing $\beta_2 = 0$. (5 points)
 - (b) Draw the joint confidence set for (β_1, β_2) . (10 points)
 - (c) Add an intercept to the model and check if predictor coefficients are significant. (10 points).
 3. Consider a linear regression model given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

where $\mathbf{X} = [\mathbf{1} : \mathbf{x}_1 : \mathbf{x}_2 : \cdots : \mathbf{x}_p]$. Show that the model fitting statistic R^2 for this model is simply the square of the correlations between observed and predicted values of y . (20 points)

4. Christensen presents mathematics ineptitude scores (Score y_{ijk}) for a group of $N = 35$ students categorized by
- Major i (1 = Economics, 2 = Anthropology, and 3 = Sociology);
 - High school background (“BG”) j (1 = Rural and 2 = Urban).

The output from fitting a 2-way ANOVA model with interaction is on the last page. The model is

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \epsilon_{ijk}$$

Also, you do not need to read the Section 7.2 (“2-way ANOVA with interaction”), but it might help just getting familiar with the model. While fitting the model we use the constraint $\alpha_1 = \eta_1 = 0$. Also $\gamma_{ij} = 0$ if $i = 1$ or $j = 1$.

- (a) Which group of students has the lowest average score? (What is it?) Which group of students has the highest average score? (What is it?) (10 points)
- (b) In the `summary(.)` output there is an F-statistic, $F = 2.553$ with 5 and 29 degrees of freedom.
- (i) What are the null and alternative hypotheses being tested? (5 points)
- (ii) What conclusion would you make? (Please state in general terms that relate to the groups rather than parameters). (10 points)