STAT 425 — Spring 2016 — Sections 1UG, 1GR

Homework 1

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

1. The *simple regression through the origin* model is like a simple linear regression model, but without the intercept:

   $$Y_i \ = \ \beta_1 x_i \ + \ e_i, \qquad i = 1, 2, \ldots, n$$

   with $\mathrm{E}(e_i) = 0$, $\mathrm{Var}(e_i) = \sigma^2 > 0$, and $\mathrm{Cov}(e_i, e_j) = 0$ if $i \neq j$.

   The *ordinary least squares* estimate minimizes the *residual sum of squares*

   $$RSS(\beta_1) \ = \ \sum_{i=1}^{n} \left(y_i - \beta_1 x_i\right)^2$$

   (a) [2 pts] Take the derivative of RSS, and set the resulting expression equal to zero. (This is sometimes called the called the *normal equation*.)

   (b) [2 pts] Solve the equation of the previous part.

   (c) [2 pts] To find your solution in part (b), you made an assumption about the values of $x_1, x_2, \ldots, x_n$. What is that assumption, and why is it needed?

   (d) [2 pts] Show that the expression you found in part (b) really is a minimizer of $RSS(\beta_1)$. (Hint: Take the second derivative.)

2. For a constant matrix $\boldsymbol{A}$ and a random vector $\boldsymbol{Z}$,

   $$\mathrm{E}(\boldsymbol{AZ}) \ = \ \boldsymbol{A}\,\mathrm{E}(\boldsymbol{Z}) \qquad \mathrm{Var}(\boldsymbol{AZ}) \ = \ \boldsymbol{A}\,\mathrm{Var}(\boldsymbol{Z})\boldsymbol{A}^T$$

   (assuming expectations and variances all exist).

   Consider the linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ under the Gauss-Markov conditions. For each of the following random vectors, determine the mean vector and the variance-covariance matrix (in terms of $\boldsymbol{X}$, $\boldsymbol{\beta}$, and σ^2). Simplify, if possible.

   (a) [2 pts] $\boldsymbol{e}$

   (b) [2 pts] $\boldsymbol{Y}$

   (c) [2 pts] $\hat{\boldsymbol{\beta}}$

   (d) [2 pts] $\hat{\boldsymbol{Y}}$ (the random vector for which the realization is the computed vector $\hat{\boldsymbol{y}}$ of fitted values)

3. The data set `ais` (in package `alr4`) provides data on athletes. Use `help(ais)` for information about the variables. Fit a regression model with weight (kg) as the response, and sex, height, sum of skin folds, and percent body fat as predictors.

   (a) [2 pts] Present a summary of your fitted model. (Use the R `summary` function.)

   (b) [2 pts] Give the least squares estimates of all coefficients.

(c) [2 pts] What is the *name* for the proportion of variation in the response explained by the predictors? What is its *value*, for the model you fit?

(d) [2 pts] Which observation (case number) has the largest (positive) residual? Also, what is its fitted value?

(e) [2 pts] Supposing all other predictors are held constant, what would be the difference in weight (kg) for a male compared to a female, according to this model?

(f) [2 pts] Which independent variables are statistically significant at the 5% (0.05) level?

(g) [2 pts] Compute individual 95% confidence intervals for all of the regression coefficients.

(h) [2 pts] Predict the weight (kg) of a 170 cm tall female with sum of skin folds equal to 60 and 12% body fat. Also, give a 95% *prediction* interval.

(i) [2 pts] Fit a model with only sex and height as the predictors, and use an F-test to compare it with the full model.

4. Using the `fuel2001` data set (in package `alr4`), fit a regression model with `FuelC` as the response, and `Income`, `Pop`, and `Tax` as predictors.

   (a) [2 pts] Present a summary of your fitted model. (Use the R `summary` function.)

   (b) [2 pts] Using the summary, test the (null) hypothesis that $\beta_{\texttt{Income}} = 0$.

   (c) [2 pts] Using the summary, test the (null) hypothesis that $\beta_{\texttt{Income}} = \beta_{\texttt{Pop}} = \beta_{\texttt{Tax}} = 0$.

   (d) [2 pts] Add `Drivers` as another predictor, and present a summary of your fitted model.

   (e) [2 pts] Use an F-test to test whether $\beta_{\texttt{Drivers}} = 0$.

   (f) [2 pts] Compare your results in the previous part with the results of a t-test for $\beta_{\texttt{Drivers}} = 0$. Are they the same?

5. [GRADUATE SECTION ONLY]

   [4 pts] Weisberg (Fourth Edition), Exercise 2.12.

Some reminders:

- Unless otherwise stated, all data sets can be found in either the `alr4` package or the `faraway` package in R.
- Unless otherwise stated, use a 5% level ($\alpha = 0.05$) in all tests.