

Homework V

STAT 4355 Applied Linear Models

Due date: March 29th, 2022

Total points possible: 80

Instruction: Clearly mark the problem number. Be sure to clearly write the answer and its derivation.

1. (Chapter 3, pp. 65-67) Consider the multiple linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon.$$

For the general linear hypothesis approach, find the appropriate **D** and **d**. (15 points)

- (a) $H_0 : \beta_1 + 2\beta_2 = 3$
 - (b) $H_0 : \frac{\beta_1 + \beta_2}{2} = \beta_3$
 - (c) $H_0 : \beta_1 = \beta_2, \beta_3 = \beta_4$
 - (d) $H_0 : \beta_1 - 2\beta_2 = 4\beta_3, \beta_1 + 2\beta_2 = 0$
 - (e) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$
2. (R programming) The National Football League data are in the attached CSV file. (65 points)

y	Games won (per 14-game season)
x_1	Rushing yards (season)
x_2	Passing yards (season)
x_3	Punting average (yards/punt)
x_4	Field goal percentage (FGs made/FGs attempted 2season)
x_5	Turnover differential (turnovers acquired–turnovers lost)
x_6	Penalty yards (season)
x_7	Percent rushing (rushing plays/total plays)
x_8	Opponents' rushing yards (season)
x_9	Opponents' passing yards (season)

Table 1: Variable description

- (a) (Lecture Notes Ch 3 pp. 16-18) Draw nine scatterplots for the number of games won (y) against the nine variables: x_1, \dots, x_9 . (5 points)

From the plots, the number of games won (y) seem to be associated with the team's rushing yardage (x_1), passing yardage (x_2), turnover differential (x_5), percentage of rushing plays (x_7), and opponents' yards rushing (x_8). We fit a multiple linear regression model relating y to x_1, x_2, x_5, x_7 , and x_8 . The corresponding parameters are denoted by $\beta_1, \beta_2, \beta_5, \beta_7, \beta_8$.

- (b) Write down the fitted linear model, and report (i) the variance estimate ($\hat{\sigma}^2$) (ii) R^2 and (iii) *adjusted* R^2 . (5 points)
- (c) Interpret the six parameter estimates: $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_5, \hat{\beta}_7, \hat{\beta}_8$, and $\hat{\sigma}^2$. (5 points)

- (d) (Lecture Notes Chapter 3, pp.45-46) Obtain the `anova` output using the code on p.46, and construct ANOVA table as in p.45 based on the output. (10 points)
- (e) (Lecture Notes Chapter 3, pp.45-47) Obtain the `anova` output using the code on p.47 and construct ANOVA table as in p.45 based on the output. (10 points)
- (f) Confirm that the two ANOVA tables from (d)-(e) are the same and perform test for significance of regression at significance level of 0.05. (5 points)
- (g) Test the value of passing yardage given all the other 4 predictors at significance level of 0.05. (5 points)
 - i. Write down the null and alternative hypotheses.
 - ii. Specify the distribution of the test statistics under the null hypothesis.
 - iii. Find the observed test statistic and p-value, and draw a conclusion.

In the model with the 5 predictors, it is not significant that the number of games won (y) is associated with the team's rushing yardage (x_1), turnover differential (x_5), and percentage of rushing plays (x_7). Therefore, we investigate the contribution of the 3 predictors (x_1, x_5, x_7) to the model.

- (h) Test an appropriate subset of coefficients at significance level of 0.05. (10 points)
 - i. Write down the null and alternative hypotheses.
 - ii. Specify the distribution of the test statistics under the null hypothesis.
 - iii. Find the observed test statistic and p-value, and draw a conclusion.
- (i) Which model do you prefer to consider based on the test performed in (h)? (5 points)
- (j) Attach the R code and console output you used for (a)-(h). (5 points)