**MATH 8050**

# Homework Assignment 5
(Due on 6/17 9:45am in class)

1. An analyst wanted to fit the regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \quad i = 1, \ldots, n$$

by the method of least squares when it is known that $\beta_3 = -2$. How can the analyst obtain the desired fit by using a multiple regression computer program?

2. A hospital administrator wished to study the relation between patient satisfaction ($Y$) and patient's age ($x_1$), severity of illness ($x_2$), and anxiety ($x_3$). The data (46 patients) are available in `PatientSatisfaction.txt`.

   (a) Obtain a scatterplot matrix and correlation matrix. What information do these provide?

   (b) Fit a first-order regression model for three predictor variables to the data and state the estimated regression function. How is $\hat{\beta}_2$ interpreted here?

   (c) Test whether there is a regression relation, using $\alpha = 0.10$. State the hypotheses, decision rule, p-value, and conclusion. What does your test imply about $\beta_1$, $\beta_2$, and $\beta_3$?

   (d) Find the coefficient of multiple determination $R^2$ and interpret it here.

   (e) Obtain a 90% confidence interval for the mean satisfaction for a 35-year-old patient with severity index 45 and anxiety index 2.2.

3. An endocrinologist explored the relationship between steroid level ($Y$) and age ($x$) in 27 healthy females (age 8–25). Data: `SteroidLevels.txt`.

   (a) Fit a model: $y = \beta_0 + \beta_1 x_i' + \beta_{11}(x_i')^2 + \varepsilon_i$, where $x_i' = x_i - \bar{x}$. Plot the data and fitted model. Comment on fit. Compute $R^2$.

   (b) Test whether or not there is a regression relation ($\alpha = 0.01$). State hypotheses, decision rule, p-value, and conclusion.

   (c) Predict the steroid level of a female at age 15 with 99% prediction interval. Interpret your interval.

   (d) Test if the quadratic term can be dropped from the model ($\alpha = 0.01$). State hypotheses, decision rule, and conclusion.

   (e) Express the fitted function obtained in (a) in terms of the original variable $x$.

4. Refer to the patient satisfaction data in 2.

   (a) Obtain the ANOVA table that decomposes regression sum of squares into extra sums of squares for: $x_2$; $x_1$ given $x_2$; $x_3$ given $x_2$ and $x_1$.

(b) Test whether $x_3$ can be dropped given that $x_1$ and $x_2$ are retained (use $F^*$, $\alpha = 0.025$).

(c) Test whether $x_2$ and $x_3$ can be dropped given that $x_1$ is retained ($\alpha = 0.025$). Give p-value.

(d) Test whether $\beta_1 = -1.0$ and $\beta_2 = 0$ ($\alpha = 0.025$). State hypotheses, full and reduced models, decision rule, and conclusion.

5. Suppose a dataset with sample size $n = 25$ and 3 potential regressors (predictors). Use the adjusted $R^2$ to perform variable selection.

   (a) Find adjusted $R^2$ for the null and full models. Given: Residual standard error = 40, $R^2$ of full model = 0.6.

   (b) Use forward selection to choose the final model.

   (c) Did this procedure fail to find the true best model? If yes, explain.

   (d) What is the AIC of the chosen model from question 2?

Adjusted $R^2$ for Candidate Models

Table 1: *

| Model | Adjusted $R^2$ |
|---|---|
| $E(Y) = \beta_0$ | — |
| $E(Y\|X_1) = \beta_0 + \beta_1 X_1$ | 0.488 |
| $E(Y\|X_2) = \beta_0 + \beta_2 X_2$ | 0.331 |
| $E(Y\|X_3) = \beta_0 + \beta_3 X_3$ | 0.491 |
| $E(Y\|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ | 0.653 |
| $E(Y\|X_1, X_3) = \beta_0 + \beta_1 X_1 + \beta_3 X_3$ | 0.452 |
| $E(Y\|X_2, X_3) = \beta_0 + \beta_2 X_2 + \beta_3 X_3$ | 0.427 |
| $E(Y\|X_1, X_2, X_3)$ | — |