ST590-602 Homework #2

Instructions: please answer the questions below. You are welcome to use output from statistical software to substantiate your answers.

1. Reconsider the cornyields and rainfall data. Consider linear regressions of yield (Y) on the following sets of predictor variables (x_1 is rainfall, $x_2 = x_1^2$ and x_3 is year):

   - Model 1: Simple linear regression on rainfall,

   $$E(Y|x_1) = \beta_0 + \beta_1 x_1$$

   - Model 2: Quadratic regression on rainfall, (a model with x_1 and $x_2 = x_1^2$)

   $$E(Y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

   - Model 3: Regression model in which rainfall, x_1, and year (x_3) effects are *additive*

   $$E(Y|x_1, x_3) = \beta_0 + \beta_1 x_1 + \beta_3 x_3$$

   - Model 4: Quadratic regression on rainfall, x_1, with non-additive year effects.

   $$E(Y|x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3$$

   - Model 5: A model selected with the assistance of variable selection software:

   $$E(Y|x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3$$

   - Model 6: A model with non-additive linear rain and year effects

   $$E(Y|x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \boldsymbol{\beta_3 x_3} + \beta_4 x_1 x_3$$

   (a) Obtain the least squares estimates of the regression coefficients for each model.

   (b) Conduct F-tests comparing the following pairs of models. For each comparison, state the implicit null hypothesis (H_0) being tested and conduct the test at level $\alpha = .05$. Additionally, report the p-value associated with the test/comparison. Using a policy that adopts the reduced/nested model unless there is "significant" evidence against H_0, speficy the model you'd choose for each comparison.

      i. Model 1 versus Model 2
      ii. Model 3 versus Model 4
      iii. Model 4 versus Model 5
      iv. Model 1 versus Model 3
      v. Model 2 versus Model 4
      vi. Model 6 versus Model 4

   (c) Is Model 2 nested in Model 3?

   (d) Using Models 1 and 6, estimate the increase in mean yield when rainfall increases by one inch in the year 1900.

2. A random sample leads to $n = 11$ bivariate measurements $(y_1, x_1), \ldots, (y_{11}, x_{11})$ with sample means and sample variances

$$\bar{x} = 80, \ \bar{y} = 85, s_x^2 = 1, \ s_y^2 = 20.$$

(a) Complete the ANOVA table.

Source	Sum of Squares	d.f.	Mean Square	F-ratio	p-value
Regression	160	1	160	?	
Error	?	?	?		
Corrected Total	200	10			

(b) Determine the uncorrected total sum of squares, $\sum y_i^2$.

(c) The sample correlation coefficient was $r_{xy} = -.894$. Report the slope of the least squares regression line.

(d) Ignoring x, obtain 95% confidence limits for $E(Y)$.

(e) Consider the subpopulation with $x = 80$, obtain 95% confidence limits for $E(Y|X = 80)$.

(f) Consider another observation, Y_{12}.

i. Ignoring x, obtain 95% prediction limits for Y_{12}. (Is the interval wider than the one in part (d).)

ii. Supposing $x_{12} = 80$, obtain 95% prediction limits for $E(Y_{12}|X = 80)$. (Is the interval wider than the one in part (e).)