STOR 664: FALL 2022
Midterm Exam, October 6, 2022

Open book in-class exam: time limit 75 minutes.

This is a single multi-part question but each part will be graded independently of the other parts. You are allowed to consult course notes (printed or e-read), homework assignments and any personal notes you have made during the course. Other outside materials are not permitted. Computers or ipads may be used *only* for the purpose of accessing pre-stored course notes; they are not to be used for computations during the exam. A hand-held calculator is permitted. Answers should preferably be written in a university examination book ("blue book"). You may consult the teaching assistant (in class) or the instructor (email, text or phone) if the wording is unclear or if you think there might be an error, but the teaching assistant or instructor will not give hints how to solve the exam. The university Honor Code is in effect at all times.

Consider the linear regression model

$$y_i \quad = \quad \beta_0 + x_i^2\beta_1 + x_i\beta_2 + x_i^3\beta_3 + \epsilon_i, i = 1, \ldots, n$$

where the ϵ_i are independent normally distributed random variables with mean 0 and a common unknown variance σ^2. (Please note the order of covariates: x_i^2, x_i and x_i^3, in that order.) Defining $S_k = \sum_{i=1}^n x_i^k$, $T_k = \sum_{i=1}^n y_i x_i^k$ for any $k \geq 0$, we assume that the x_i's are symmetric about 0 to guarantee that $S_k = 0$ for all odd values of k.

(a) Find explicit expressions for the least squares estimators $\hat{\beta}_0, \cdots, \hat{\beta}_3$, and their variances. You should express the answer in terms of the values of S_k and T_k for $k \geq 0$, or any expressions derived from them. [**25 points**]

(b) We would like to test the hypotheses $H_0 : \ \beta_2 = \beta_3 = 0$ versus the alternative H_1 that at least one of β_2 or β_3 is not 0. How would the estimates in (a), and their variances, change under the assumption that H_0 is true? [**10 points**]

(c) Now suppose we are setting up the formal F test of H_0 against H_1. Write SSE_0 and SSE_1 for the residual sum of squares under H_0 and H_1 respectively. Show that

$$SSE_0 - SSE_1 \quad = \quad A\hat{\beta}_2^2 + B\hat{\beta}_2\hat{\beta}_3 + C\hat{\beta}_3^2$$

where A, B and C are constants that you should identify (functions of n, S_2, S_4, etc.). Hence write down the formal test of H_0 against H_1 and define the rejection region for a test of significance level 0.01. (You are not expected to make an explicit numerical calculation but describe how to calculate it; for example, you may use R notation to define the needed percentage point of the F distribution, which will be one component of your answer.) [**25 points**]

Turn the page for the last two parts of the question.

(d) Now consider the case where x_i goes from –3 to +3 in steps of 0.5 (so $n = 13$). You can assume (no need to check this) $S_2 = 45.5$, $S_4 = 284.375$, $S_6 = 2099.094$, the last to three decimal places. Also assume $\sigma^2 = 40$, $\beta_2 = 1$, $\beta_3 = 0.5$. What, in that case, will be the power of the test in part (c)? [**20 points**]

(You should give detailed numerical calculations as far as you are able to go, but the final answer will depend on the non-central F distribution and you should give the formula for calculating that as an R function or any equivalent notation that makes clear how to do the numerical calculation. If you cannot do explicit numerical calculations, at least state the formulas on which they may be based.)

(e) Now suppose that the real purpose of the experiment is to determine a 95% prediction interval for a new observation taken at a new value $x = x^*$. Show (i) how to calculate such a prediction interval under the assumption H_1, (ii) how the calculations in (i) would change if the experimenter did indeed assume H_0 to be true. [**20 points**]

(Since it's not possible for you to give a numerical answer here, you should describe precisely the seqeence of steps, including any formulas for percentage points of relevant probability distributions. For (ii), you do not need to repeat the full calculation of (i), but indicate at which steps the calculation would change.)