Stats 203V: Introduction to Regression Models and ANOVA **Summer 2020**

Midterm

Due Date : July 18, 9:59 AM (PT)

For full credit, solutions (.pdf file, either typed or handwritten and scanned) must be uploaded to `Gradescope` by **9:59 am PT on July 18**. Any corrections to the problems will be posted on `Canvas` and will notified via a Canvas announcement.

This is an open-everything exam. The only rule is that you can not discuss about the exam with anyone. I do not think you will require any computing devises, but you are welcome if you want to use one.

If you have any doubt about any question of the exam, feel free to contact us via e-mail or Piazza. Remember to keep your message private if you use Piazza.

1. (3 points.) Write down the assumptions of the "usual" full rank linear regression model with n observations and p co-variates, including the intercept if present in the model (i.e. as discussed in class), in a form that assumes independent errors, but not assuming normal distributions.

2. (7 points) Consider a simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ (i = 1, \ldots, n)$, where ϵ_i are uncorrelated with mean zero and constant variance σ^2. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ are OLS estimates. Find a necessary and sufficient condition for $\hat{\beta}_0$ and $\hat{\beta}_1$ being uncorrelated.

3. (5 pts.) True or false, and explain.
   (a) If you fit a regression equation to data, the sum of the residuals is 0.
   (b) If the equation has an intercept, the sum of the residuals is 0.
   (c) In the regression model, $E(\hat{Y}|X) = X\hat{\beta}$.
   (d) In the regression model, $E(\hat{Y}|X) = X\beta$.
   (e) In the regression model, $E(Y|X) = X\beta$.

4. (10 points) Consider the linear regression model as in Question 1. $\hat{Y}_i$ be the fitted value for i-th observation obtained using OLS estimation. Show that

$$\sum_{i=1}^{n} \text{Var}(\hat{Y}_i) = p\sigma^2.$$

   [Hint : Try to write down the expression in terms of trace of a matrix and then use the formula for $\text{Var}(\hat{\beta})$.]

5. (3+5+2 points) Suppose Alice and Bob have observations of n individuals on three variables, namely (Y_i, x_i, z_i), $i = 1, \ldots, n$. They want to explore the effect of x and z-variable on Y by linear modelling. Alice went on performing a linear regression of Y on (x, z), including an intercept term. Bob tried to do something different and did a variable transformation creating new variable,s $u_i = x_i - z_i$ and $v_i = x_i + z_i$, and then ran a linear regression of Y on (u, v), including an intercept term.
   (a) Explicitly state the models they fit. You can assume full rank type assumptions, but clearly state those.

(b) Did they perform fundamentally different analysis? If not, then determine in which sense their analysis are equivalent.

(c) Provided you have no further information, which model do you prefer more, Alice's or Bob's?

6. (4+6+5 points) In order to estimate to parameters θ and ϕ, it is possible to make observations of three types: (a) the first type have expectation θ, (b) the second type have expectation $\theta + \phi$, and (c) the third type have expectation $\theta - 2\phi$. All observations are uncorrelated and have constant variance. Suppose that m observations of type (a), m observations of type (b) and n observations of type (c) are made. Write down the model for the observed data in multiple linear regression form and explicitly state the design matrix. Find the least squares estimates $\hat{\theta}$ and $\hat{\phi}$. Prove that this estimates are uncorrelated if $m = 2n$.

7. (10 points) Imon uses OLS to fit a regression equation with an intercept, and computes R^2. Elif wants to test the null hypothesis that all the coefficients are 0, except for the intercept. Can Elif compute F from R^2, n, and p? If so, what is the formula? If not, why not?

8. (7+3 points) Let $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, \quad i = 1, \ldots, n$, where the ϵ_i are independent $N(0, \sigma^2)$. Prove that the F-statistic for testing the hypothesis $H_0 : \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$ $(0 < q \leq p-1)$ is unchanged if a constant, c, say, is subtracted from each Y_i. Will the answer remain same if the original model didn't include any intercept term?