STA 2101/442 Assignment Twelve[1]

These questions are practice for the final exam, and are not to be handed in. Material like this may or may not be on the final.

1. Consider simple regression through the origin in which the explanatory variable values are random variables rather than fixed constants. In addition, the explanatory variable values cannot be observed directly. Instead, we observe X_i plus a piece of random noise. The model is this: Independently for $i = 1, \ldots, n$, let

$$
\begin{aligned}
Y_i &= X_i\beta + \epsilon_i \\
W_i &= X_i + e_i,
\end{aligned}
\tag{1}
$$

where

- X_i has expected value μ and variance σ_x^2,
- ϵ_i has expected value 0 and variance σ_ϵ^2, and
- e_i has expected value 0 and variance σ_e^2
- X_i, e_i and and ϵ_i are all independent.

Again, the X_i values are unavailable. All we can see are the pairs (W_i, Y_i) for $i = 1, \ldots, n$.

(a) Following common practice, we ignore the measurement error and apply the usual regression estimator with W_i in place of X_i. The parameter β is estimated by

$$
\widehat{\beta}_{(2)} = \frac{\sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i^2}
$$

Is $\widehat{\beta}_{(2)}$ a consistent estimator of β? Answer YES, NO or IMPOSSIBLE TO DETERMINE. Show your work.

item Now consider instead the estimator

$$
\widehat{\beta}_{(3)} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n W_i}.
$$

Is $\widehat{\beta}_{(3)}$ a consistent estimator of β? Answer YES, NO or IMPOSSIBLE TO DETERMINE. Show your work.

(b) If X_i, ϵ_i and e_i are normally distributed, Model (1) has five parameters: β, μ, σ_x^2 σ_ϵ^2 and σ_e^2. Perhaps surprisingly, it is possible to obtain the maximum likelihood estimates explicitly without differentiating anything. Do it if you can. It is acceptable to start the calculation and just indicate how you would finish it without giving formulas for all the MLEs. Is $\widehat{\beta}_{(3)}$ the MLE of β?
Hint: Begin by calculating the covariance matrix of $(W_i, Y_i)'$.

(c) Even without assuming normal distributions, a large-sample confidence interval for β is within reach. Indicate how you would derive it, not necessarily giving all the details. There is more than one way to get the standard error you need.

---

[1]Copyright information is at the end of the last page.

2. Steel is made by heating iron and adding some carbon. A steel company conducted an experiment in which knife blades were manufactured using two different amounts of carbon (Low and High), and three different temperatures (Low, Medium and High). Of course even the Low temperature was very hot. A sample of knife blades was manufactured at each combination of carbon and temperature levels, and then the breaking strength of each blade was measured by a specially designed machine. The response variable is breaking strength.

   (a) In a table with one row for each treatment combination, please make columns giving the coefficients of the contrast or contrasts you would use to test for main effects of Temperature.

   (b) In another table with one row for each treatment combination, please make columns giving the coefficients of the contrast or contrasts you would use to test the Temperature by Carbon Level interaction.

   (c) In one last table with one row for each treatment combination, please make columns showing how you would set up dummy variables for both independent variables, using *effect coding* (that's the scheme with the -1s).

   (d) Write $E(Y|\mathbf{X} = \mathbf{x})$ for the regression model, using the names from your table above. Include the interactions!

   (e) Using the β values from your answer to the preceding question, state the null hypothesis you'd use to test whether the effect of carbon level on breaking strength depends on the temperature.

3. Consider a two-factor analysis of variance in which each factor has two levels. Use this regression model for the problem:

$$Y_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + \beta_3 d_{i,1} d_{i,2} + \epsilon_i,$$

where $d_{i,1}$ and $d_{i,2}$ are dummy variables.

   (a) Make a two-by-two table showing the four treatment means in terms of β values. Use *effect coding* (the scheme with 0, 1, -1). In terms of the β values, state the null hypothesis you would use to test for

      i. Main effect of the first factor
      ii. Main effect of the second factor
      iii. Interaction

   (b) Make a two-by-two table showing the four treatment means in terms of β values. Use *indicator dummy variables* (zeros and ones). In terms of the β values, state the null hypothesis you would use to test for

      i. Main effect of the first factor
      ii. Main effect of the second factor
      iii. Interaction

   (c) Which dummy variable scheme do you like more?

In a study of agricultural productivity, small apple farms are randomly assigned to use one of three Pesticides (Type A, B or C) and one of three Fertilizers (Type 1, 2 or 3). The dependent variable is total crop yield in kilograms, and there are two covariates: number of trees on the farm, and crop yield last year.

(a) In the table below, fill in the definitions of the dummy variables for Pesticide (p_1 and p_2), and the dummy variables for Fertilizer (f_1 and f_2). Use *effect coding* (the scheme with $0, \ 1, \ -1$).

Pesticide	Fertilizer	p_1	p_2	f_1	f_2
A	1				
A	2				
A	3				
B	1				
B	2				
B	3				
C	1				
C	2				
C	3				

(b) Write $E[Y|\mathbf{X}]$ for a model that includes a possible Pesticide by Fertilizer interaction as well as their main effects. Denote the covariates by X_1 and X_2. Of course the vector $\mathbf{X}$ includes p_1, p_2 and so on as well as X_1 and X_2. There are no interactions between the two covariates, or between covariates and factors.

(c) Give the null hypothesis you would test to answer each question below. The answers are in terms of the β parameters in your model. Some of the answers are the same. Except for the last one, assume that each question begins with "Controlling for number of trees and crop yield last year ..." .

Question	Null Hypothesis
Does type of pesticide affect average crop yield?	
Does the effect of fertilizer type on crop yield depend on the type of pesticide used?	
Does the effect of pesticide type on crop yield depend on the type of fertilizer used?	
Does fertilizer type affect average crop yield?	
Is there a main effect for pesticide type?	
Is there a main effect for fertilizer type?	
Is there an interaction between fertilizer type and pesticide type?	
Test both main effects and the interaction, all at the same time.	
Test both covariates simultaneously, controlling for the main effects and the interaction.	

4. The Ontario government selects a random sample of q grade schools. From each school, a random sample of k students is selected and given a reading test. School is the explanatory variable. Because the values of this variable represent a random sample from a larger populations, a *random effects* model is appropriate. A standard version that applies to this situation is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

where

μ is an unknown constant parameter.

$\tau_i \sim N(0, \sigma_\tau^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

τ_i and ϵ_{ij} are all independent, $i = 1, \ldots q$ and $j = 1, \ldots, k$.

(a) What is the distribution of Y_{ij}? Just write down the answer. You need not show any work.

(b) Are the Y_{ij} all independent? Consider two cases.

(c) What is the distribution of $\overline{Y}_i = \frac{1}{k} \sum_{j=1}^{k} Y_{ij}$? State your answer; the only work you need to show is your calculation of the variance.

(d) Find $Cov(\overline{Y}_i, Y_{ij} - \overline{Y}_i)$. Show your work.

(e) Define $SSTR = k \sum_{i=1}^{q} (\overline{Y}_i - \overline{Y}_.)^2$, where $\overline{Y}_. = \frac{1}{q} \sum_{i=1}^{q} \overline{Y}_i$. Find the distribution of $\frac{SSTR}{\sigma^2 + k\sigma_\tau^2}$. Hint: You can make this a very easy problem by using something from the formula sheet.

(f) Define $SSE = \sum_{i=1}^{q} \sum_{j=1}^{k} (Y_{ij} - \overline{Y}_i)^2$. Find the distribution of $\frac{SSE}{\sigma^2}$. Again you may use a well-known fact to make the problem easier, but *do not forget your answer to 4b*.

(g) The proportion of variance in an observation Y_{ij} that is explained by School is $\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma^2}$. Give a reasonable estimator for this quantity; show some work.

(h) What null hypothesis would you use to test for the effect of school on students' reading scores? State the null hypothesis in symbolic form; that is, it's a statement in terms of Greek letters.

(i) An exact (not large-sample) test is available for this hypothesis. Give a formula for the test statistic. Also state its distribution under H_0, including the degrees of freedom. Briefly indicate why it has the distribution you claim.