Homework 13

Put your name and student ID here

2021-05-27

Q1: Consider the linear model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2), i = 1, \ldots, n.$$

1. Derive the maximum likelihood estimators (MLE) for β_0, β_1. Are they consistent with the least square estimators (LSE)?
2. Derive the MLE for σ^2 and look at its unbiasedness. Is it better than the unbiased estimator $\hat{\sigma}^2 = S_e^2/(n-2)$ by taking account for MSE?
3. A very slippery point is whether to treat the x_i as fixed numbers or as random variables. In the class, we treated the predictors x_i as fixed numbers for sake of convenience. Now suppose that the predictors x_i are iid random variables (independent of ϵ_i) with density $f_X(x;\theta)$ for some parameter θ. Write down the likelihood function for all of our data $(x_i, y_i), i = 1, \ldots, n$. Derive the MLE for β_0, β_1 and see whether the MLE changes if we work with the setting of random predictors?

Q2: Consider the linear model without intercept

$$y_i = \beta x_i + \epsilon_i,\ i = 1, \ldots, n,$$

where ϵ_i are independent with $E[\epsilon_i] = 0$ and $Var[\epsilon_i] = \sigma^2$.

- Write down the least square estimator $\hat{\beta}$ for β, and derive an unbiased estimator for σ^2.
- For fixed x_0, let $\hat{y}_0 = \hat{\beta} x_0$. Work out $Var[\hat{y}_0]$.

Q3: Genetic variability is thought to be a key factor in the survival of a species, the idea being that "diverse" populations should have a better chance of coping with changing environments. Table below summarizes the results of a study designed to test that hypothesis experimentally. Two populations of fruit flies (Drosophila serrata)-one that was cross-bred (Strain A) and the other, in-bred (Strain B)-were put into sealed containers where food and space were kept to a minimum. Recorded every hundred days were the numbers of Drosophila alive in each population.

Date	Day number	Strain A	Strain B
Feb 2	0	100	100
May 13	100	250	203
Aug 21	200	304	214
Nov 29	300	403	295
Mar 8	400	446	330
Jun 16	500	482	324

- Plot day numbers versus population sizes for Strain A and Strain B, respectively. Does the plot look linear? If so, please use least squares to figure out the coefficients and their standard errors, and plot the two regression lines.
- Let β_1^A and β_1^B be the true slopes (i.e., growth rates) for Strain A and Strain B, respectively. Assume the population sizes for the two strains are independent. Under the same assumptions of $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ for both strains, do we have enough evidence here to reject the null hypothesis that $\beta_1^A \leq \beta_1^B$ (significance level $\alpha = 0.05$)? Or equivalently, do these data support the theory that genetically mixed populations have a better chance of survival in hostile environments. (提示：仿照方差相同的两个正态总体均值差的

假设检验，构造相应的 t 检验统计量)