Modelli Statistici Avanzato

Homework n. 4

With reference to ***Section 3.2*** of the textbook of Fahrmeir, Kneib, Lang & Marx, briefly answer the following questions. Notation: OLS = Ordinary Least Squares, ML = Maximum Likelihood

1. What are the advantages of estimating the parameters using OLS instead of least absolute deviations?
2. Write the normal equations and derive the expression of the OLS estimator; declare what is the condition ensuring that the least squares solution is unique and make an example of violation of such condition.
3. Why ML estimation requires additional assumptions with respect to OLS? In which case the two estimation principles yield the same estimator?
4. Write the expression of the hat matrix and prove that it is symmetric and idempotent.
5. Write the ML estimator of the residual variance, discuss its bias, and write an unbiased estimator.
6. Prove that the residuals are orthogonal to the columns of the design matrix and exploit this result to show that the sum of the residuals is zero (as long as the model has the intercept).
7. Write the coefficient of determination; discuss the situation where it is zero and why a zero value does not imply that the response is unrelated with the explanatory variable(s).
8. Suppose we have a response y and a covariate x and fit the following models: M1 (y on x), M2 (y on x and x^2), M3 (y on ln(x)). Consider the coefficient of determination: if it is 0.15 in model M1, what can be said about its value in models M2 and M3?
9. Write down the three conditions for using the coefficient of determination as a tool for model comparison.
10. Write in a formal way the four standard assumptions on the model errors (zero mean, homoscedasticity, no correlation, normality).
11. Prove that the OLS estimator is unbiased; list the standard assumptions of the model errors that are needed to get this result.
12. Derive the covariance matrix of the OLS estimator; list the standard assumptions of the model errors that are needed to get this result.
13. Write the optimal prediction of y for a future observation with covariates equal to $\boldsymbol{x}_0$. In what sense this prediction is optimal?
14. Justify why if the errors are normally distributed then the OLS estimator is normally distributed.
15. Under the four standard assumptions on the model errors, write a statistic with chi-squared distribution which is based on the distance between the OLS estimates and the true parameter values.
16. Write a further assumption allowing to derive the asymptotic properties of OLS estimator without normality and mention a typical situation when such further assumption is true.
17. Derive the expectation and covariance matrix of the residuals.
18. Explain why the residuals cannot be directly used to judge the homoscedasticity assumption and how this problem can be circumvented.