学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　陳　一凡

学　位　論　文　題　名

Study of unobserved factors in fatty acids with exploratory data analysis
（探索的データ解析による脂肪酸内在因子に関する研究）

Fatty acid is a fundamental component of lipids in plants, animals, and microorganisms. It plays an important role in human health and fat-related diseases. A comprehensive measurement of diverse fatty acids in biological samples straightforwardly generates to a multi-variate and complex dataset. Therefore, we can hardly extract the characteristics of fatty acid directly from the dataset, and there may exist many unobserved factors; unobserved factors give us useful information but cannot be directly derived from the data without appropriate analysis methods. Thus, exploratory data analysis, which are statistic techniques to find out unobserved factors of dataset for suggestive information, is considered to be applied on fatty acid datasets. In this study, exploratory data analysis was applied to human serum fatty acid dataset and cow milk fatty acid dataset.

Our study consists of 6 chapters.

Chapter 1 is about the background, purpose and structure of our study.

In chapter 2, we outline fatty acid and vitamin D, such as their structure, their circulation, their function, to explain the complexity of fatty acid and vitamin D. Fatty acid is complicated because of its composites. Fatty acid, usually refers to total fatty acid, consists of free fatty acid and esterified fatty acid. Esterified fatty acid is further divided into cholesterol ester (CE), triglyceride, and phospholipid. Fatty acids undergo various metabolic processes including dietary intake, enteric absorption, plasma lipoprotein metabolism, enzymatic modification, and secretion as triglyceride in the milk. Therefore, analyzing unobserved factors in fatty acid dataset is quite challenging

In chapter 3, we introduced methods of exploratory data analysis including dimension reduction methods and functional data analysis. Functional data analysis is a statistical approach to the datasets which are represented by functions. Dimension-reduction methods are considered as an effective way to extract the common unobserved factors of high-dimension dataset.

In chapter 4, exploratory data analysis, dimension reduction methods, were applied to serum fatty acids dataset. The serum datasets are consisting of total fatty acid dataset, free fatty acid dataset, and CE dataset. Besides total and free fatty acid datasets containing also the dataset of 25-hydroxyvitamin D3 (vitamin D). The three datasets are determined by LC/MS/MS from a general Japanese population (n=545; men:women=245:300) in Suttu town, Hokkaido. Firstly, three basic dimension reduction methods (factor analysis, principal component analysis, and independent component analysis) were conducted to variables related to total and free fatty acid. These analyses successfully characterized the fatty acid datasets, reflecting their physicochemical natures, metabolisms, and food sources. Factor analysis and principal component demonstrated the association of omega-3 fatty acids (20:5 and 22:6) with vitamin D, suggesting fish oil as the common source of vitamin D and we were told that it is the

first time to find out vitamin D and omega-3 fatty acids (20:5 and 22:6) are in a positive relation in the aspect of clinic. Secondly, two basic dimension reduction methods (principal component analysis and factor analysis) were conducted to variables related to CE. They both reflected size (concentration), food source, fat solubility, and biological aspect of CE species. In comparison between PCA (PC4) and factor analysis (factor 4), the latter was found more suggestive for a biological aspect of n-6 FAs. Cholesteryl docosahexaenoate (DHA) was found unique by factor analysis, possibly relevant to the unique accumulation of DHA in the brain. Informatics approach, especially factor analysis, might be useful for analysis of complicated metabolism of CE species in the serum. However, applying dimension reduction methods separately to the part of variables is not enough to find the general unobserved factor of the dataset. Therefore, we analyzed the dataset by common principal component analysis (CPCA)to find out the common unobserved factors of the three sets of variables in the dataset; total fatty acid, free fatty acid and CE. In the result of CPCA, we found that there were four components explainable: first one was considered as size factor, and more importantly the rest three were considered as plant intake, animal fat intake and fish oil intake correspondingly. Therefore, we found food intake information of fatty acid successively, which can reduce the cost in clinical research by reducing the cost of questionnaire about food intake.

In chapter 5, functional data analysis (functional regression analysis and functional clustering analysis) were conducted to find the unobserved factors of milk fatty acid concentration and the corresponding environment datasets (daily sum of rainfall, daily average wind speed, and daily average temperature). The dataset was obtained from 18 areas every two months and continued to one year in Hokkaido, Japan (n=18, p=6). With the characteristics of functionalized dataset and the derivatives of dataset, we can analyze not only the data themselves but also the changing trend from other environment datasets. Firstly, functional clustering analysis was conducted to the total fatty acid dataset of milk samples and three environment datasets to reveal the unobserved relationship between milk fatty acid concentration and other variables. We obtained eight dendrograms in functional clustering analysis. Cophenetic correlation coefficient was used to measure the similarities among the dendrograms. Multidimensional scaling was used to clearly visualize the dissimilarities of the dendrograms. As a result, we found that milk fatty acid concentration has relationships with the average wind speed and the average temperature, and also that the trend of milk concentration has relation with the trends of daily sum of rainfall and average wind speed. This result can benefit farmers and even milk company to find a proper location in Hokkaido for economical milk production.

In chapter 6, we concluded that dimension reductions and functional data analysis can serve as a useful tool to extract valuable information from complex datasets of fatty acids.