

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

An Investigation of the Methods for Estimating Usual Dietary Intake Distributions

A Thesis Presented in Partial Fulfillment of the Requirements for the Degree of

Master of Applied Statistics

at Massey University, Albany, New Zealand

Stefan Kremenov Stoyanov

Student ID # 00038377

2008

Abstract

The estimation of the distribution of usual intake of nutrients is important for developing nutrition policies as well as for etiological research and educational purposes. In most nutrition surveys only a small number of repeated intake observations per individual are collected. Of main interest is the long-term usual intake which is defined as long-term daily average intake of a dietary component. However, dietary intake on a single day is a poor estimate of the individual's long-term usual intake. Furthermore, the distribution of individual intake means is also a poor estimator of the distribution of usual intake since usually there is large within-individual compared to between-individual variability in the dietary intake data. Hence, the variance of the mean intakes is larger than the variance of the usual intake distribution. Essentially, the estimation of the distribution of long-term intake is equivalent to the estimation of a distribution of a random variable observed with measurement error.

Some of the methods for estimating the distributions of usual dietary intake are reviewed in detail and applied to nutrient intake data in order to evaluate their properties. The results indicate that there are a number of robust methods which could be used to derive the distribution of long-term dietary intake. The methods share a common framework but differ in terms of complexity and assumptions about the properties of the dietary consumption data. Hence, the choice of the most appropriate method depends on the specific characteristics of the data, research purposes as well as availability of analytical tools and statistical expertise.

Acknowledgements

I would like to thank my supervisor Dr. Barry McDonald for sparking my interest in measurement error models and for his support and encouragement during my work on the thesis.

I am grateful to Dr. Hoffmann and Dr. Tooze for the useful comments and computer code which were of great help in completing this research.

Table of Contents

1.	INTRODUCTION	7
2.	LITERATURE REVIEW	9
2.1	DODD ET AL (2006).....	9
2.2	NATIONAL CANCER INSTITUTE METHOD (2006)	17
2.3	THE NATIONAL ACADEMY OF SCIENCES (2003)	19
2.4	HOFFMANN ET AL (2002A, 2002B).....	23
2.5	NUSSER ET AL (1996).....	31
2.6	WALLACE AND WILLIAMS (2005).....	35
2.7	GAY (2000).....	37
2.8	SLOB (1993)	38
2.9	BUCK ET AL (1995).....	40
2.10	SLATER ET AL (2004)	42
2.11	CHANG ET AL (2001)	45
2.12	CARROLL ET AL (1995, 2006)	49
3.	METHODS.....	55
3.1	THE DATA.....	55
3.2	DATA PREPARATION.....	56
3.3	PRELIMINARY ANALYSIS	59
4.	RESULTS.....	67
4.1	HOFFMAN ET AL METHOD (2002)	68
4.2	ISU METHOD (1996).....	73
4.3	CHANG ET AL METHOD (2001)	76
4.4	NATIONAL CANCER INSTITUTE METHOD (2006)	80
5.	DISCUSSION AND SUGGESTIONS FOR FUTURE RESEARCH.....	86
5.1	SIMPLIFIED POWER METHOD	87
5.2	EMPIRICAL COMPARISON OF THE METHODS FOR ESTIMATING USUAL INTAKE DISTRIBUTIONS	95
6.	CONCLUSION	102
7.	BIBLIOGRAPHY.....	104

List of Tables and Graphs

Tables:

TABLE 1: MAIN NUTRIENTS.....	7
TABLE 2 : DODD ET AL (2006). DATA TRANSFORMATIONS AND BIAS ADJUSTMENTS	14
TABLE 3 : DODD ET AL (2006). COMPARISON OF THE MODEL STEPS	15
TABLE 4: DODD ET AL (2006). STRENGTHS AND WEAKNESSES OF THE MODELS	16
TABLE 5: ANOVA TABLE: ONE-WAY CLASSIFICATION WITH BALANCED DATA	21
TABLE 6: HOFFMANN ET AL (2002A). COMPARISON OF METHODS FOR ESTIMATING USUAL INTAKE DISTRIBUTION.....	24
TABLE 7: SLATER ET AL (2004). EXPECTATIONS OF THE COMPONENTS OF VARIANCE	44
TABLE 8: CARROLL ET AL (2006). COMPARISON BETWEEN ORDINARY OLS AND MEASUREMENT ERROR MODEL ESTIMATORS	53
TABLE 9: KARKECK (1987). NUMBER OF REPEATED MEASUREMENTS REQUIRED	57
TABLE 10: SAMPLE COUNTS FOR DAY OF THE WEEK AND MONTH OF PREGNANCY. EIGHT REPEATS PER INDIVIDUAL.....	59
TABLE 11: DIETARY COMPONENTS CORRELATIONS.....	63
TABLE 12: REGRESSION RESULTS: RELATIONSHIP BETWEEN MEANS AND STANDARD DEVIATIONS. FOUR REPEATS PER INDIVIDUAL.....	64
TABLE 13: REGRESSION RESULTS: RELATIONSHIP BETWEEN MEANS AND STANDARD DEVIATIONS. SIX REPEATS PER INDIVIDUAL	64
TABLE 14: REGRESSION RESULTS: RELATIONSHIP BETWEEN MEANS AND STANDARD DEVIATIONS. EIGHT REPEATS PER INDIVIDUAL	64
TABLE 15: T-VALUES FOR TESTS OF HETEROGENOUS WITHIN-PERSON VARIANCE	65
TABLE 16: F-VALUES FOR DAY OF THE WEEK EFFECT	65
TABLE 17: F-VALUES FOR WEEKEND EFFECT	65
TABLE 18: F-VALUES FOR SEASONAL EFFECT.....	66
TABLE 19: F-VALUES FOR INTERVIEW SEQUENCE EFFECT	66
TABLE 20: HOFFANN ET AL METHOD (2002). TRANSFORMATION PARAMETERS AND OUTPUT STATISTICS FOR A SET OF NUTRIENT INTAKES.....	68
TABLE 21: HOFFMANN ET AL METHOD (2002). EXPECTED VALUE IN THE ORIGINAL SCALE	69
TABLE 22: HOFFMANN ET AL. METHOD (2002). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL PROTEIN INTAKE.....	71
TABLE 23: HOFFMANN ET AL. METHOD (2002). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL CARBOHYDRATES INTAKE.....	72
TABLE 24: HOFFMANN ET AL. METHOD (2002). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL VITAMIN E INTAKE.....	72
TABLE 25: ISU METHOD (1996). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL PROTEIN INTAKE	74
TABLE 26: ISU METHOD (1996). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL CARBOHYDRATES INTAKE	75
TABLE 27: ISU METHOD (1996). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL VITAMIN E INTAKE	75
TABLE 28: CHANG ET AL METHOD (2001). ESTIMATED RATIOS OF WITHIN TO BETWEEN-INDIVIDUAL VARIANCE.....	77
TABLE 29: CHANG ET AL METHOD (2001). ANDERSON-DARLING GOODNES OF FIT TEST. GAMMA DISTRIBUTION.....	77
TABLE 30: CHANG ET AL. METHOD (2001). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL PROTEIN INTAKE.....	78
TABLE 31: CHANG ET AL METHOD (2001). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL CARBOHYDRATES INTAKE.....	78
TABLE 32: CHANG ET AL METHOD(2001). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL VITAMIN E INTAKE	79
TABLE 33: NCI METHOD (2006). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL PROTEIN INTAKE	81
TABLE 34: NCI METHOD(2006). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL CARBOHYDRATES INTAKE.....	82
TABLE 35: NCI METHOD (2006). COMPARISON BETWEEN OBSERVED AND ESTIMATED DISTRIBUTION OF USUAL VITAMIN E INTAKE.....	82
TABLE 36: USUAL PROTEIN INTAKE COEFFICIENTS OF VARIATION	83
TABLE 37: COMPARISON OF THE MODELING PROCESS STEPS	84
TABLE 38: COMPARISON OF THE STRENGTHS AND WEAKNESSES OF THE MODELS.....	85
TABLE 39: SIMPLE POWER METHOD. ANDERSON-DARLING TEST OF NORMALITY	88
TABLE 40 : COMPARISON BETWEEN THE PROTEIN USUAL INTAKE DISTRIBUTION ESTIMATED BY DIFFERENT STATISTICAL METHODS AND THE OBSERVED DISTRIBUTION OF INDIVIDUAL 2-DAY MEANS	95
TABLE 41 : COMPARISON BETWEEN THE CARBOHYDRATES USUAL INTAKE DISTRIBUTION ESTIMATED BY DIFFERENT STATISTICAL METHODS AND THE OBSERVED DISTRIBUTION OF INDIVIDUAL 2-DAY MEANS	95
TABLE 42: COMPARISON BETWEEN THE VITAMIN E USUAL INTAKE DISTRIBUTION ESTIMATED BY DIFFERENT STATISTICAL METHODS AND THE OBSERVED DISTRIBUTION OF INDIVIDUAL 2-DAY MEANS	95

Graphs:

GRAPH 1: OBSERVED DAILY PROTEIN INTAKE.....	60
GRAPH 2: OBSERVED DAILY CARBOHYDRATES INTAKE.....	60
GRAPH 3: OBSERVED DAILY VITAMIN E INTAKE.....	61
GRAPH 4: MEANS AND STANDARD DEVIATIONS OF THE TRANSFORMED PROTEIN INTAKE.....	62
GRAPH 5: MEANS AND STANDARD DEVIATIONS OF THE TRANSFORMED CARBOHYDRATES INTAKE.....	62
GRAPH 6: MEANS AND STANDARD DEVIATIONS OF THE TRANSFORMED VITAMIN E INTAKE.....	63
GRAPH 7: HOFFMANN ET AL. METHOD (2002). PROTEIN USUAL INTAKE DISTRIBUTION.....	70
GRAPH 8: ISU METHOD (1996). PROTEIN USUAL INTAKE DISTRIBUTION.	73
GRAPH 9: CHANG ET AL METHOD (2001). PROTEIN USUAL INTAKE DISTRIBUTION.	77
GRAPH 10: NCI METHOD (2006). PROTEIN USUAL INTAKE DISTRIBUTION.	81
GRAPH 11: BOX-COX TRANSFORMATION OF PROTEIN INTAKE	88
GRAPH 12: NATURAL CUBIC SPLINE FIT TO THE POWER TRANSFORMED PROTEIN INTAKE.....	92
GRAPH 13: NATURAL CUBIC SPLINE FIT TO THE LOG TRANSFORMED VITAMIN E INTAKE	93
GRAPH 14: PROTEIN USUAL INTAKE DISTRIBUTION. AVERAGE ESTIMATED BIAS	96
GRAPH 15: CARBOHYDRATES USUAL INTAKE DISTRIBUTION. AVERAGE ESTIMATED BIAS	97
GRAPH 16: VITAMIN E USUAL INTAKE DISTRIBUTION. AVERAGE ESTIMATED BIAS	98