



## РЕЦЕНЗИЯ

От: *доц. д-р Калоян Валентинов Хараламчиев*  
*СУ „Св. Климент Охридски“*

*Научна специалност: Статистика и демография*

Относно: дисертационен труд за присъждане на образователна и научна степен **„доктор“** по научна специалност *„Статистика и демография“* в УНСС.

**Основание** за представяне на рецензията: участие в състава на научното жури по защита на дисертационния труд съгласно Заповед №792/28.03.2022 на Ректора на УНСС.

Автор на дисертационния труд: *Цвета Цонкова Цонкова*

Тема на дисертационния труд: *Статистическо съчетаване на данни от извадкови изследвания за анализа в социално-икономическата област*

### **1. Информация за дисертанта**

Дисертантът се е обучавал по докторска програма към *катедра „Статистика и иконометрия“*, факултет *„Приложна информатика и статистика“* на УНСС по научна специалност *„Статистика и демография“* съгласно Заповед на Зам.-ректора по НИД на УНСС 838/03.04.2015. Обучението е осъществено в *задочна* форма през периода 2015-2018 г.

### **2. Обща характеристика на представения дисертационен труд**

Дисертационният труд е в обем от 188 страници и включва увод, три глави, заключение, списък на използваната литература и три приложения.

Темата на дисертационния труд – съчетаване на данни от различни извадкови изследвания, е изключително актуална. Актуалността произтича от два фактора. Първо, статистическото съчетаване на данни позволява да се правят изследвания с по-малко изучавани признаци, което не натоварва излишно респондентите и поевтинява самите изследвания, но след съчетаването се получават синтетични бази данни, в които се съдържат признаци от всички изследвания, което обогатява

възможностите за анализ. Второ, темата е практически непозната в България, за което свидетелства и списъкът на използваната литература, в който няма нито едно заглавие на български език.

Целта на дисертационния труд е дефинирана на стр. 4 и включва две части – разработване на технология за статистическо съчетаване на данни и апробация на разработената технология. Дефинираната цел е адекватна на темата на дисертационния труд, реалистична е и е изпълнима.

В дисертационния труд са дефинирани осем задачи (стр. 4-5). Задачите са добре дефинирани. Последователността на задачите е правилна, като тяхното изпълнение закономерно ще доведе до изпълнение на целта.

Обектът и предметът са дефинирани правилно (стр. 5).

Основната теза на дисертационния труд (стр. 4) е убедителна и е доказана в текста на дисертацията.

Списъкът с използваната научна литература съдържа 56 заглавия и отделно има приложен списък с три интернет страници. Всичките 56 заглавия са на английски език и обхващат периода 1968-2017 г. Това свидетелства, че докторантът познава както класически текстове от самото начало на възникване на този изследователски проблем, така и съвременни текстове по темата.

### **3. Оценка на получените научни и научно-приложни резултати**

В първа глава са разгледани историческото развитие, същността, значението, методите и възможностите за статистическото съчетаване на данни от различни изследвания. Важен принос на докторанта е разработването на авторска класификация на методите за статистическо съчетаване на данни на базата на шест класификационни критерия.

Във втора глава е описана технологията на статистическото съчетаване на данни. Тази технология съдържа осем етапа, които в началото на главата са представени вербално (стр. 44-46) и схематично (стр. 47). След това в останалата част от главата всеки от етапите е разгледан в детайли.

В трета глава е описана методика за статистическо съчетаване на данни от две конкретни извадкови изследвания, провеждани в България. По същество това е конкретно приложение на технологията, описана във втора глава. Прави впечатление изключително педантичното следване на технологията с необходимото внимание към детайлите на всеки етап. Изрично искам да обърна внимание на направената проверка на качеството на получения синтетичен файл.

Също така изрично искам да подчертая нещо, което за мен е голям плюс на дисертационния текст, а именно това, че в приложения са представени всички изчислителни алгоритми, реализиран в IBM SPSS Statistics и в R. Това дава възможност за проверимост на получените резултати, както и дава възможност други изследователи да проведат същото статистическо съчетаване в бъдеще, от което да се провери доколко получените резултати са устойчиви във времето и повторяеми.

В заключение, може да се каже, че поставените в увода задачи са изпълнени, което логично довежда до постигане на целта на изследването.

#### **4. Оценка на научните и научно-приложни приноси**

В автореферата са посочени пет научни и научно-приложни приноса. Първите два могат да се класифицират като научни, а последните три – като научно-приложни. Съгласен съм с така формулираните приноси и смятам, че те са реални и наистина присъстват в текста.

Изрично искам да открия три от тях:

- Създаването на авторска класификация на методите за статистическо съчетаване на данни (Принос №2);
- Разработването на технология за статистическо съчетаване на данни (Принос №4);
- Апробирането на разработената технология по данни от две български извадкови изследвания (Принос №5).

#### **5. Оценка на публикациите по дисертацията**

Докторантът има две публикации по темата на дисертацията си. Едната е публикуван доклад от международна научна конференция. Втората е статия в списание „Статистика“.

Тези публикации представят съществени части на дисертационното изследване и дават възможност на широката научна публика да се запознае с резултатите от него.

#### **6. Оценка на автореферата**

Авторефератът отразява точно структурата и съдържанието на дисертационния труд.

#### **7. Критични бележки, препоръки и въпроси**

Имам няколко бележки и препоръки, които са по-скоро редакционни, и не променят цялостното ми добро впечатление от дисертационния труд.

1. Според мен, точката за историческото развитие на статистическото съчетаване на данни от различни източници се появява доста късно в текста. За мен правилното ѝ място е след точка 1.2, но преди описанието на методите.
2. На много места в текста (например на стр. 5, 6, 13, 15, 16, 17 и т.н.) се говори за „сложни извадкови изследвания“ без обаче да е дефинирано какво означава това. Предполагам, че става въпрос за превод на английския термин *complex samples*, но това също никъде не е указано. Смятам, че когато се използва едно понятие, то не трябва да бъде оставено като саморазбиращо се, а трябва да бъде дефинирано. Също така на стр. 19 се говори за „теоретични извадки“, отново без дефиниция какво означава това. Предполагам, че се има предвид планираната извадка, но това не става ясно от текста.
3. На няколко места в текста (например на стр. 18, 23, 57 и т.н.) във формулите е използвана главната латинска буква I, без да е пояснено какво се означава с нея, което затруднява четенето на формулите.
4. Когато в текста се споменават цитираните автори (извън скобите) в едни случаи имената са изписани на кирилица, а в други на латиница. Добре е да има единство и да са само на кирилица или само на латиница.
5. На стр. 53 в Таблица 8 Хи-квадрат е посочен като корелационен коефициент. Той обаче е статистически критерий за проверка на хипотези, а не е корелационен коефициент, макар че на базата на Хи-квадрат са конструирани производни корелационни коефициенти, като например посоченият в същата таблица коефициент на Крамер.
6. На стр. 54 е написано „корелационното отношение  $\eta^2$ “. Обаче корелационното отношение е първата степен –  $\eta$ . Втората степен е коефициент на определеност.
7. На стр. 59 има формула (30), за която не е обяснено значението на нито един от използваните символи.
8. На стр. 59 във формула (31) и на стр. 61 във формула (36) са използвани две различни означения за едно и също нещо – индекса на различие. Също така на стр. 61 във формули (35) и (36) и на стр. 71 са използвани три различни означения за едно и също нещо – разстоянието на Хелингер. Смятам, че еднаквите неща трябва да се означават с еднакви символи.

9. На стр. 61 е написано, че разстоянието на Хелингер приема стойности между 0 и  $\sqrt{2}$ . Обаче разстоянието на Хелингер е корен от единица минус коефициента на Бхатчаря, за който пък от своя е написано, че приема стойности от 0 до 1 (стр. 60). Това е противоречие. Ако е вярно, че коефициентът на Бхатчаря приема стойности от 0 до 1, тогава и разстоянието на Хелингер ще приема стойност от 0 до 1. Ако е вярно, че разстоянието на Хелингер приема стойности между 0 и  $\sqrt{2}$ , тогава коефициентът на Бхатчаря трябва да приема стойности между -1 и 1. Кое от двете е вярно?
10. На стр. 67 при номерирането на категориите на образованието е изпуснат номер 5.
11. На стр. 71-72 е направена проверка за съгласуваност на разпределенията, след което за тези признаци, чиито разпределения не се съгласуват добре, са окрупнени категории. След повторната проверка разпределенията вече се съгласуват добре. Но това е изкуствено произведен ефект, тъй като всяко окрупняване на категории заличава част от различията, тъй като изчезва междугруповото различие и остава само вътрешногруповото различие. По този начин, заличавайки различията, съгласуваността изкуствено се увеличава.
12. На стр. 77 са посочени две причини логистичната регресия да бъде предпочетена пред дискриминантния анализ. Втората причина е съществената и само тя е достатъчна. Първата причина не е добре обоснована и според мен е напълно излишна.
13. На стр. 79 и 85 на практика ръчно е реализирана стъпкова регресия. Моята препоръка е стъпковата регресия да не се прави ръчно, а да се използва вградената функция за стъпкова регресия в IBM SPSS Statistics.
14. На стр. 80 коефициентите на Кокс и Снел и на Нейджълкърк са интерпретирани по един и същи начин, което е проблематично, тъй като имат различни теоретични граници. Коефициентът на Нейджълкърк е нормиран в границите от 0 до 1, докато коефициентът на Кокс и Снел има плаваща горна граница. В този смисъл интерпретацията на коефициента на Нейджълкърк е правилна, но интерпретацията на коефициента на Кокс и Снел е проблематична.
15. На стр. 107 в Таблица 28 са посочени процентите по квинтилни групи. На теория те би трябвало да са по 20%, обаче процентите в

таблицата варират от 14,0 до 27,9. На какво се дължи това разминаване спрямо теоретичните 20%?

16. На стр. 114 във Фигура 10 се създава впечатлението, че има голямо различие между сравняваните проценти. Това обаче е изкуствено произведен ефект чрез свиване на мащаба на скалата. Моята препоръка е мащабът на скалата да се направи между 0% и 100% и тогава различията може би няма да се окажат чак толкова големи.

## **8. Заключение**

*Предлагам на членовете на уважаемото научно жури да присъдят ОНС „доктор“ на Цвета Цонкова Цонкова по научно направление 3.8. Икономика, научна специалност „Статистика и демография“.*

29.04.2022/София

Подпис: .....