



## УНИВЕРСИТЕТ ЗА НАЦИОНАЛНО И СВЕТОВНО СТОПАНСТВО

### СТ А Н О В И Щ Е

От: доц. д-р Пламен Иванов Петков,  
Стопанска академия „Д. А. Ценов“ – Свищов,  
Катедра „Статистика и приложна математика“,  
Научна специалност „Статистика и демография“

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

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

#### **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 на НИД на УНСС № 838/03.04.2015. Обучението е осъществено в задочна форма през периода 10.03.2015 – 10.3.2022 година.

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

Представеният дисертационен труд на тема “Статистическо съчетаване на данни от извадкови изследвания за анализа в социално-икономическата област” с автор Цвета Цонкова Цонкова е в обем от 188 страници, от които 120 страници чист текст, 5 страници библиографска справка и 63 страници с приложения, в които са представени резултатите

от анализа на емпиричния материал с помощта на SPSS, използваните кодове в среда R, с които се извършва съчетаването на данни от две сложни по характер реални извадкови изследвания, осъществявани от Националния статистически институт, както и първите 100 записа, съответстващи на приложения в дисертацията синтетичен набор от данни. Разработката се характеризира с логическа обвързаност между отделните части, постигнат е оптимален баланс и е осъществен плавен преход между отделните компоненти. В структурно отношение тя се състои от *списък на използваните съкращения, съдържание, увод, три глави, заключение, списък с цитираната литература и приложения.*

Заглавието е коректно формулирано и очертава проблемите, които ще бъдат разисквани в разработката – какво представлява съчетаването на данни, какви са неговите възможности за приложение при данни, получени от извадкови изследвания, какви подходи могат да бъдат апробирани за нуждите на анализа в социално-икономическата област. Актуалността на разглежданата проблематика е във всякакво съмнение. Всеизвестно е, че официалната социална статистика е насочена към изследвания на доходи, потребление, здравеопазване, образование, пазар на труда, използване на информационно-комуникационни технологии, социално включване и други, но поради ред причини, сред които водещи са финансовите, отделните изследвания не могат да обхванат самостоятелно всички възможни аспекти. От тази гледна точка съчетаването на данни от две и повече изследвания, било то изчерпателни или извадкови, представлява необходимия и надежден инструментариум, чрез който може да се осъществи коректен синтез и получаване на обща статистическа информация.

В увода са изпълнени всички формални изисквания по отношение на задължителните атрибути като е обоснована актуалността и значимостта на разглежданата в дисертационния труд проблематика, извършен е кратък, но пълноценен обзор на статистическата теория и практика, свързана с отделни аспекти на статистическото съчетаване на данни, дефинирани са изследователска теза, обект, предмет, времеви обхват, цел, задачи и две изследователски хипотези на научното изследване, посочени са основните информационни източници, които са използвани за осъществяване на емпиричната част от дисертацията, софтуерните продукти, с които е осъществена обработката и анализа на информацията и са изведени ограничителните условия, които съпътстват настоящата разработка. Обект на изследване са данните от две извадкови изследвания, провеждани ежегодно от Националния статистически институт – използване на информационни комуникационни технологии (ИКТ) от домакинствата и

изследването на доходите и условията на живот (EU-SILC), а предметът е насочен към предлагане на методология, с която ефективно да се съчетаят данните от тези две извадкови изследвания. Целта на дисертацията е да се оцени възможността, да се разработи и апробира с реални данни технология за прилагане на методи за статистическо съчетаване на данни от две различни извадкови изследвания. За конкретизиране на целта са поставени 8 изследователски задачи, от които 5 са с теоретико-методологичен, а 3 – с практико приложен характер. С теоретико-методологична насоченост са задачите, свързани с изясняване същността на понятието „статистическо съчетаване на данни“, преглед на съществуващи подходи и методи за статистическо съчетаване на данни от различни източници, проучване на опита за съчетаване на данни от различни източници в световен мащаб, анализ на възможностите за приложение на статистическото съчетаване на данни от различни източници в българската статистическа практика и разработване на методология за статистическо съчетаване на данни от две извадкови изследвания с обзор на различните подходи за хармонизиране и обединяване на масивите от данни. Практико-приложните задачи се свързват с апробиране на разработената методология за съчетаване на статистически данни изследването „Използване на ИКТ от домакинствата“ и изследването на доходите и условията на живот (EUSILC) като се акцентира върху качеството и ефективността на използваните методи, извършване на описателен анализ на данните от получения интегриран масив от данни и анализ на резултатите от експеримента с формулиране на изводи и препоръки. Основната теза се свързва със съществуването на обективна необходимост от разработване на технология за статистическо съчетаване на данни от различни източници за задълбочаване на анализа в социално-икономическата сфера в България. Коректно са изброени ограничителни условия, които са присъщи за настоящото изследване.

Структурата се характеризира със завършен вид и коректно отразява представената информация чрез приложение на принципа от „общото към частното“. В първа глава се обхващат теоретичните основи при статистическото съчетаване на данни. Първо се описва същността на статистическото съчетаване на данни, след това акцентът се пренася върху съчетаването на данни от сложни извадкови изследвания, представени са методите, намиращи приложение в тази насока и е осъществена тяхната класификация. Извършен е кратък исторически преглед на процеса по осъществяване на статистическо съчетаване на данни, отразено е нарастващото значение на този тип анализ и са маркирани основните възможности за съчетаване на данни от различни източници в българската статистическа практика. Втората глава е от методологичен характер и в нея

се представя технологията при реализирането на статистическо съчетаване на данни от две извадкови изследвания. Отражени са основните проблеми, съпътстващи подготовката на процеса по съчетаването на данни и са посочени основните фактори за постигане на краен успех при технологията за статистическото съчетаване на данни от различни източници – съгласуване на концепциите и определенията, съгласуване на изместени източници на данни, съгласуване на несъвместими променливи, избор на съчетаващи променливи и оценка на качеството на резултатите от съчетаването на данните. В трета глава е приложена методика за съчетаване на данни за 2020 г. от изследването на доходите и условията на живот (EU-SILC) и изследването „Използване на информационни и комуникационни технологии (ИКТ)”. Коректно е осъществен анализ на масивите от данни, които се набавят с помощта на посочените извадкови изследвания, предприети са стъпки по тяхното хармонизиране и е съставена рамка за статистическото им съчетаване. Аргументирано е осъществен подбор на съчетаващите променливи, описани са приложените методи за съчетаване на данните и е извършен надежден анализ на резултатите от реализираното съчетаване на данните. В библиографската справка са включени общо 56 литературни източници и 3 интернет сайта като всички са на английски език.

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

Като цяло дисертационният труд се характеризира със сериозна задълбоченост на обхвата на изследването, както от теоретично, така и от практико-приложно и методологично естество. Докторантът е положил значителни усилия по отношение на изследваната от него проблематика, разкрил е особеностите и специфичните черти на обекта на изследване, анализиран е в детайли научните достижения в световен аспект и успешно и достигнал до открояване на нерешените проблеми в социално-икономическата сфера с помощта на методологията за съчетаване на данни от две независими сложни извадкови изследвания. По отношение на стила и езика на работата се наблюдава коректно поднасяне на информацията, която в голяма степен е стилово и смислово изчистена, обмислена в необходимата дълбочина и поднесена по начин, който подобава на едно завършено научно произведение.

За проверка на дефинираните работни хипотези, изследователската теза и поставената цел е използван подходящ от методологична точка инструментариум, разчитащ на теоретичен анализ и синтез, както и приложение на съвременни математико-статистически методи. В групата се отнасят статистически методи, с които може да се установят връзките

между променливите в етапа на хармонизирането на дефиниции и класификации – логистична регресия, дискриминантен анализ, коефициенти на рангова корелация и статистически инструментариум за проверка на хипотези, оценка на синтетичния и донорския набор от данни, основаваща се на различни критерии за измерване на сходства и различия в етапа на проверката на качеството на резултатите от съчетаването на данните: индекс на различие, коефициент на припокриване на две разпределения, коефициент на Bhattacharya, разстояние на Хелингер, приложение на регресионно-корелационния анализ при идентифициране на общите променливи, оказващи съществено от статистическа гледна точка влияние върху определените на предходен етап целеви променливи, методи за съчетаване на данните – случаен хот дек метод, метод на разстоянието до най-близкия съсед и рангов хот дек метод, графичен и табличен подход при представяне на статистическа информация.

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

Дисертационният труд съдържа научни и научно-приложни приноси, които надграждат теоретичните и емпиричните изследвания в областта на съвместното използване на резултати от няколко сложни извадкови изследвания. Същите са коректно дефинирани от докторанта в автореферата и отговарят на постигнатите резултати. Техният брой е сведен до 5 като 4 от тях носят теоретико-методологичен характер, свързан най-вече с идентификацията и структурирането на проблемите при съчетаване на данни от сложни извадкови изследвания, класификация на подходящите методи за съчетаване на данни, предложения за възможно приложение на методите за статистическо съчетаване на данни в българската статистическа практика и дефиниране на етапите, които следва да се изпълняват в процеса по съчетаване на данните. Приносът с практико-приложен етап е дефиниран като един, но в действителност той може да се представи като съвкупност от два независими приноса. От една страна е разработена и предложена собствена методика за съчетаване на данни от сложни извадкови изследвания, каквито са изследването за доходите и условията на живот (EU-SILC) и изследването „Използване на информационни и комуникационни технологии от домакинствата“ (ИКТ), а от друга - тази методика е апробирана с реални данни отнасящи се за 2020 г. Освен тези пет приноса, според мен има и още един, при това доста съществен, който докторантката скромно е пропуснала. Става дума за поместените в Приложение 2 тридесет и четири кода в среда R за провеждане на предложената от автора на настоящата разработка методика за съчетаването на данни от две сложни по характер реални извадкови

изследвания. Тези кодове могат да намерят успешна реализация както в статистическата практика на Националния статистически институт, така и в сродни изследвания, осъществявани от колеги от научните среди, а не на последно място биха могли да бъдат включени в учебните програми по дисциплини, свързани с приложение на извадкови изследвания, преподавани във висшите училища от България.

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

Авторефератът е в обем от 34 страници и отразява точно дисертационния труд като дава пълна представа за проведеното научно изследване. Написан е в стегнат и издържан стил, откъдето проличават уменията на докторанта да представи основните етапи и резултати от изследването си в синтезиран, балансиран и компактен вид. Към дисертационния труд са приложени две самостоятелни публикации - една статия, която е под печат и един доклад от участие в организирана от УНСС конференция. Публикациите са пряко свързани с темата, обекта, предмета и основната теза на дисертационния труд като отразяват основните идеи, представени в него.

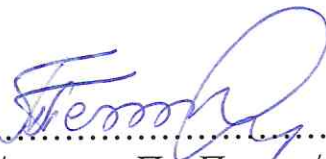
### **6. Критични бележки, препоръки и въпроси**

Към докторантката нямам въпроси, но имам няколко принципни препоръки и бележки. Дисертацията се базира на задоволителен литературен анализ и свидетелства за значителна запознатост с изследваната проблематика, но според мен от полза би било разширяването на обхвата на проучването. В списъка с използвана литература са посочени източници само на английски език. Липсва коментар по отношение на опита на други автори в българската литература, работили в сродни на изследваната в настоящата разработка област. Вярно е, че тук акцентът е върху съчетаването на данни от различни източници и настоящото изследване се явява пилотно, но би могло да се цитират български автори, които в своите разработки са прилагали използваните в настоящата разработка математико-статистически методи като регресионен и корелационен анализ, логистична регресия, дискриминантен анализ, а не на последно място и изследвания в областта на големите данни (big data). Литературните източници са използвани и цитирани добросъвестно и коректно, но са малко на брой позоваванията в текста. Добре би било използваните в изследването инструментариуми и методологии да се цитират с препратки към техния първоизточник – коефициент на Крамер, рангов корелационен коефициент на Спирмън, критерий на Валд, коефициент на Bhattacharya, разстояние на Хелингер и пр.

### Заклучение

Отчитайки завършения характер на представената разработка, вземайки предвид достойнствата и положителни страни на дисертацията, независимо от направените критични бележки, смятам, че докторантът притежава теоретични знания и статистически умения за практическо прилагане на статистическия инструментариум за нуждите на статистическо съчетаване на данни от независими изследвания, демонстрирани са способности за аргументирано отстояване на собствени виждания и за извършване на критичен анализ на теорията и практиката. Изхождайки от всичко посочено по-горе, както и от факта, че авторът има необходимия брой публикации, давам **положителна оценка** на дисертационния труд и ще гласувам **за присъждане на образователната и научна степен „доктор“** по научна специалност „Статистика и демография“ на Цвета Цонкова Цонкова.

27.04.2022 г.  
Свищов

Подпис:   
/доц. д-р Пл. Петков/