ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ МЕДИЧНИЙ УНІВЕРСИТЕТ

Кафедра медичної фізики та інформаційних технологій

Методичні вказівки до виконання лабораторних робіт

з дисципліни

**МЕДИЧНА ІНФОРМАТИКА**

**(комп'ютерна обробка медико-статистичних даних)**

**для лікарів - інтернів**

**Укладачі: Богданова Т.Л.**

**Кльованик О.А.**

**Семенов А.С.**

**Краматорськ, 2018**

УДК 61:004.451.7 (076.5)

Б73

Донецький національний медичний університет, кафедра медичної фізики та інформаційних технологій №1.

Методичні вказівки до виконання лабораторних робіт з дисципліни МЕДИЧНА ІНФОРМАТИКА (комп'ютерна обробка медико-статистичних даних) для лікарів - інтернів, Краматорськ. - 2018. – 66с.

**Автори:** Богданова Т.Л., Кльованик О.А., Семенов А.С.

**Рецензенти:**

1. Гетьман І.А. – к.т.н., доцент кафедри комп’ютерних інформаційних технологій Донбаської державної машинобудівної академії.

2. Загребельний С.Л. – к.пед.н., доцент, зав. кафедрою інформатики і інженерної графіки Донбаської державної машинобудівної академії.

***Затверджено в якості методичних вказівок для лікарів - інтернів на циклі „Медична інформатика” на засіданні Вченої Ради Донецького Національного медичного університету***

***(протокол № 8 від 25червня 2018 р.)***

© Богданова Т.Л., 2018 р.

© Кльованик О.А., 2018 р.

© Семенов А.С., 2018 р.

# ЗМІСТ

ВСТУП………………………………………………………………………..…..4

ТЕСТОВІ ЗАВДАННЯ ДЛЯ ПЕРЕВІРКИ ВХІДНОГО РІВНЯ ЗНАНЬ…...…8

1 МЕТОДИ СТАТИСТИЧНОГО АНАЛІЗУ МЕДИЧНИХ ДАНИХ……....12

2 Оцінка ВІДМІННОСТЕЙ між групами…………………………….14

2.1. Кількісні ознаки…………………………......................……………..14

2.1.1. Нормальний розподіл………………………………………..…14

2.1.2. Ненормальний розподіл………...........................……………...15

2.2 Порядкові ознаки……………………………………….……………...16

2.3 Якісні ознаки……………………………………….…….…………...16

2.4 Аналіз залежності між ознаками…………………………………..…17

3 СУЧАСНЕ ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ СТАТИСТИЧНОЇ ОБРОБКИ БІОМЕДИЧНИХ ДОСЛІДЖЕНЬ…………………………….….19

# 4 СТИСЛІ ВІДОМОСТІ ПРО системУ Statistica 6…………………22

5 МетодичНІ рекомендацІЇ ДО вИКОНАННЯ практичНих заВДАНЬ………………………............................................................…...…....23

5.1 Лабораторна робота № 1. Дослідження ефективності щеплень…..23

5.2 Лабораторна робота № 2. Створення рандомізированої вибірки при організації медико - біологічних досліджень…………………………………34

5.3 Лабораторна робота № 3. Порівняння виживаності у двох і більше групах………………………………………………………………………..…..43

5.4. Лабораторна робота № 4. Перевірка розподілу на нормальність…48

5.5. Лабораторна робота № 5. Порівняння вибірок………………….….54

тестові завдання для досягнення конкретних цілей навчання ………………………………………………………………..…59

Список рекомендованої літератури…………………………..…64

ДОДАТОК А. Таблиця значень t Стьюдента для P=0,05; P=0,01………………………………………………………………………..…..65

ДОДАТОК Б. ВИБІР МЕТОДУ СТАТИСТИЧНОГО АНАЛІЗУ В ЗАЛЕЖНОСТІ ВІД УМОВ ЕКСПЕРИМЕНТУ……………………………..66

**ВСТУП**

Доказова медицина (evidence-based medicine) – це розділ медицини, заснований на доказах, що передбачає пошук, порівняння, узагальнення та широке поширення отриманих доказів для використання в інтересах хворих (Evidence Based Medicine Working Group, 1993).

Доказова медицина – це новий підхід, напрям або технологія збору, аналізу, узагальнення та інтерпретації наукової інформації. Доказова медицина передбачає добросовісне, з'ясовне і засноване на здоровому глузді використання найкращих сучасних досягнень для лікування кожного пацієнта (Sackett D.L. et al., 1996). Основна мета впровадження принципів доказової медицини в практику охорони здоров'я – оптимізація якості надання медичної допомоги з точки зору безпеки, ефективності, вартості та ін. значущих чинників.

Жоден практичний лікар не володіє достатнім досвідом, що дозволяє вільно орієнтуватися у всьому різноманітті клінічних ситуацій. Можна покладатися на думки експертів, авторитетні керівництва і довідники, проте це не завжди надійно через так званого ефекту запізнювання – перспективні терапевтичні методи впроваджуються в практику через значний час після отримання доказів їх ефективності (Antman E.T. et al., 1992). З іншого боку, інформація в підручниках, посібниках та довідниках часто застаріває ще до їх публікації, а вік провідного лікування досвідченого лікаря негативно корелює з ефективністю лікування (Sackett D.L. et al., 1991). Ці висновки були отримані за допомогою основного статистичного інструменту доказової медицини – мета аналізу (В.А. Горьков і співавт., 1998).

**Основні тенденції розвитку біомедичних наук визначають наступні чинники:**

* глобалізація інформаційних процесів;
* велика кількість біомедичних досліджень, що проводяться;
* широкий спектр лікарських засобів (ЛЗ) на фармацевтичних ринках;
* збільшення потоку медичної інформації (видається близько 40 000 біомедичних журналів, що публікують приблизно 2 млн статей щорічно) (Oxman A., Guyall G., 1988);
* гостро стоїть проблема раціонального витрачання коштів в системі охорони здоров'я.

**Ці основні тенденції визначають наступні потреби практичної медицини:**

* необхідність критичної оцінки інформації, призначеної для практичних лікарів і керівників охорони здоров'я;
* вибір системних підходів для прийняття рішень в медицині (лікувальних, діагностичних, управлінських та ін.).

Отже, існує необхідність узагальнення біомедичних знань і широкого інформування медичної громадськості про результати новітніх досліджень.

Потенційні можливості застосування принципів доказової медицини в практиці охорони здоров'я – значні. В першу чергу, їх застосування дозволяє використовувати об'єктивні критерії до всіх аспектів фармакотерапії. Принципи доказової медицини дають можливість з урахуванням новітньої і достовірної інформації оптимізувати вплив на прийняття рішення таких багато в чому суб'єктивних факторів, як інтуїція і кваліфікація лікаря, думки авторитетних експертів, рекомендації популярних посібників і довідників. Таким чином, доказова медицина припускає об'єднання індивідуального клінічного досвіду лікаря з найкращими доступними незалежними клінічними доказами з систематизованих досліджень (Юр'єв К.Л., Логановський К.М., 2000.).

При цьому принципи доказової медицини дозволяють розробляти найбільш ефективні, безпечні та економічні сучасні терапевтичні стратегії, які можуть бути реалізовані на державному, регіональному, популяційному, субпопуляційний і індивідуальному рівнях, сприяючи вибору оптимального варіанту в кожному конкретному клінічному випадку.

Зупинимося на деяких аспектах практичного застосування принципів доказової медицини. Перш за все вони можуть бути застосовані для підвищення якості надання медичної допомоги: це розробка клінічних рекомендацій для практичних лікарів і впровадження систем стандартизації в охорону здоров'я.

**Основні передумови виникнення доказової медицини:**

1) зростаюча потреба в критичній оцінці величезної кількості медичної інформації (з метою встановлення її надійності і достовірності), зростання числа альтернативних методів лікування і діагностики, необхідність їх грамотного вибору на основі надійних наукових даних;

2) обмежений обсяг фінансування на надання медичної допомоги (з огляду на цей фактор з усього різноманіття пропонованих методів повинні бути обрані лише ті, доказ ефективності яких не викликає сумнівів).

Поняття доказової медицини включає в себе вміння оптимального вибору лікарем найкращого способу лікування і діагностики для конкретного пацієнта, використовуючи як накопичений досвід колег, так і сучасні наукові досягнення в галузі медицини.

Доказова медицина розробляє наукові основи лікарської практики - звід правил для прийняття клінічних рішень. Головний постулат доказової медицини: кожне клінічне рішення має базуватися на чітко доведених наукових фактах.

**Як наука доказова медицина базується на двох основних напрямках: клінічної епідеміології та медичної статистики.**

**Клінічна епідеміологія −** це наука, яка розробляє методи клінічних досліджень, які мінімізують вплив систематичних і випадкових помилок. Мета клінічної епідеміології **−** розробка і застосування таких методів клінічного спостереження, які дають можливість робити справедливі висновки.

На відміну від фундаментальних біомедичних наук, клінічну медицину цікавлять питання, відповіді на які можуть дати дослідження тільки на живих людях, а не на експериментальних тварин, культурах тканин або клітинних мембранах. Клінічне дослідження важко віднести до "чистому експерименту". Тут об'єкт вивчення – пацієнт, який вільний сам визначати свої вчинки, а експериментатор - лікар з особистим професійним досвідом, схильностями і часом помилковими судженнями. Ось чому в клінічних дослідженнях завжди закладена небезпека *систематичних помилок* (упередженості), уникнути яких можна лише дотримуючись чітким науковим принципам.

**Медична статистика** – один з найважливіших інструментів доказової медицини.

Медична громадськість довго не бажала визнавати важливість статистики, почасти тому, що вона применшувала значення клінічного мислення. Подібний підхід ставив під сумнів компетентність лікарів, що спираються на постулати неповторності кожного хворого, і, отже, індивідуальності обраної терапії. Особливо це було помітно у Франції - країні, яка подарувала світові безліч дослідників, які вивчали проблеми ймовірності: П'єра де Ферма, П'єра-Симона Лапласа, Авраама де Муавра, Блеза Паскаля і Симеона Дениса Пуассона. У 1835 р уролог Ж. Сівіаль опублікував статтю, з якої випливало, що після безкровного видалення каменів сечового міхура виживають 97% хворих, а після 5175 традиційних операцій вижили тільки 78% хворих. Французька академія наук призначила комісію лікарів для того, щоб перевірити дані статті Ж. Сівіаля. У звіті цієї комісії було висловлено і обґрунтовано думку про недоцільність застосування статистичних методів в медицині: «Статистика насамперед відмовляється від конкретної людини і розглядає його в якості одиниці спостереження. Вона позбавляє його будь – якої індивідуальності для того, щоб виключити випадкові впливу цієї індивідуальності на досліджуваний процес або явище. У медицині такий підхід неприйнятний ». Однак, подальший розвиток медицини і біології показало, що в дійсності статистика є найпотужнішим інструментом цих наук.

До середини 19 століття «... вже були розроблені основні принципи статистики і відоме поняття ймовірності подій. У книзі «Загальні принципи медичної статистики» Жюль Гавар застосував їх до медицини. Ця книга чудова тим, що в ній вперше підкреслювалося, що висновок про перевагу одного методу лікування перед іншим повинен ґрунтуватися не тільки на умоглядному укладанні, але витікати з результатів, отриманих в процесі безпосереднього спостереження достатньої кількості хворих, які отримували лікування по порівнюваним методикам. Можна сказати, що Гавар фактично розробив статистичний підхід, на якому в наші дні ґрунтується доказова медицина.

Фактором, які зробили значний вплив на розвиток математичних методів статистики, стало відкриття закону великих чисел Яковом Бернуллі (1654-1705) і поява теорії ймовірності, основи якої розробив французький математик і астроном П'єр Симон Лаплас (1749-1827). Помітним етапом в низці цих подій для медичної статистики стала публікація робіт бельгійського вченого А. Кетле (1796-1874), що вперше застосував на практиці математико-статистичні методи дослідження. У своїй роботі «Про чоловіка та розвитку її здібностей» А. Кетле вивів тип середнього людини, наділеного, поряд із середніми показниками фізичного розвитку (ріст, вага), середніми розумовими здібностями і середніми моральними якостями. В цей же період часу в Росії виходить робота лікаря Бернуллі «Про щеплення проти віспи: про смерть і теорії ймовірності».

Медична статистика як точка докладання методів математичної статистики посідає особливе місце. Це особливе місце обумовлено великою роллю медицини у виникненні статистики як самостійної науки і істотним впливом науково-дослідних розробок медико-біологічних проблем на появу багатьох методів статистичного аналізу. В даний час, з метою підкреслити особливий статус медико - біологічної математичної статистики, для її позначення все частіше використовують термін **біометрія.**

Більшість методів статистичного аналізу є універсальними і можуть застосовуватися не тільки в різних галузях медичної статистики, але і в найрізноманітніших галузях людської діяльності. Наприклад, з точки зору формальної логіки статистичний прогноз інфекційної захворюваності і прогноз курсу долара - одна і та ж завдання. Більшість методів статистичного аналізу є універсальними і можуть застосовуватися не лише в різних галузях медичної статистики, але і в найрізноманітніших галузях людської діяльності. Наприклад, з точки зору формальної логіки статистичний прогноз інфекційної захворюваності і прогноз курсу долара – одне і те ж завдання.

**Методи медичної статистики можна розділити на наступні групи:**

* 1. Збір даних, який може бути пасивним (спостереження) або активним (експеримент).
  2. Описова статистика, яка займається описом і поданням даних.
  3. Порівняльна статистика, яка дозволяє проводити аналіз даних в досліджуваних групах і порівняння груп між собою з метою отримання певних висновків. Ці висновки можуть бути сформульовані у вигляді гіпотез або прогнозів.

**ТЕСТОВІ ЗАВДАННЯ ДЛЯ ПЕРЕВІРКИ ВХІДНОГО РІВНЯ ЗНАНЬ**

**1 ПРЕДМЕТАМИ ВИВЧЕННЯ МЕДИЧНОЇ СТАТИСТИКИ Є:**

1) здоров'я населення в цілому і окремих віково-статевих груп

2) виявлення і встановлення залежностей між рівнем здоров'я і чинниками довкілля

3) фінансові результати діяльності медичних установ

4) дані про мережу, діяльність, кадри установ охорони здоров'я

5) оцінка статистичної достовірності результатів медико-біологічних, клінічних і експериментальних досліджень

**2 СТАТИСТИЧНА ГІПОТЕЗА ЦЕ:**

1) складова частина об'єкту спостереження, що підлягає вивченню і реєстрації відповідно до програми дослідження

2) явище, яке підлягає детальному вивченню, і усі облікові ознаки якого можуть бути виміряні тільки кількісно

3) явище, яке підлягає детальному вивченню і його облікові ознаки повинні носити тільки якісний, описовий характер

4) формальне строго сформульоване припущення про істинність розподілу ознаки, який дослідник припускає перевірити за емпіричними даними

**3 СТАТИСТИЧНА СУКУПНІСТЬ – ЦЕ:**

1) група, що складається з великого числа відносно однорідних елементів (одиниць спостереження), узятих разом у відомих межах часу і простору

2) явище, яке підлягає детальному вивченню, і усі облікові ознаки якого можуть бути виміряні тільки кількісно

3) явище, яке підлягає детальному вивченню і його облікові ознаки повинні носити тільки якісний, описовий характер

4) місце або територія, де здійснюється статистичне дослідження

**4 ГЕНЕРАЛЬНА СТАТИСТИЧНА СУКУПНІСТЬ – ЦЕ:**

1) явище, яке підлягає детальному вивченню, і усі облікові ознаки якого можуть бути виміряні тільки кількісно

2) явище, яке підлягає детальному вивченню і його облікові ознаки повинні носити тільки якісний, описовий характер

3) місце або територія, де здійснюється статистичне дослідження

4) набір усіх можливих одиниць спостереження, які можуть бути до неї віднесені відповідно до мети дослідження

**5 ВИБІРКОВА СТАТИСТИЧНА СУКУПНІСТЬ – ЦЕ:**

1) явище, яке підлягає детальному вивченню, і усі облікові ознаки якого можуть бути виміряні тільки кількісно

2) частина генеральної сукупності, відібрана спеціальним методом і призначена для характеристики генеральної сукупності

3) місце або територія, де здійснюється статистичне дослідження

4) набір усіх можливих одиниць спостереження, які можуть бути до неї віднесені відповідно до мети дослідження

**6 ВИДАМИ СТАТИСТИЧНИХ ВЕЛИЧИН, ЯКІ ВИКОРИСТОВУЮТЬСЯ ПРИ ВИКОНАННІ МЕДИКО - БІОЛОГІЧНИХ ДОСЛІДЖЕНЬ, ЯВЛЯЮТЬСЯ:**

1) абсолютні

2) інтенсивні

3) екстенсивні

4) співвідношення

5) динамічні

6) прості

**7 ВАРІАЦІЙНИЙ РЯД – ЦЕ:**

1) ряд числових значень ознаки (варіант) отриманих в однорідній статистичній сукупності, окремі одиниці якої характеризують кількісні відмінності ознаки, що вивчається, або явища

2) ряд чисел, записаних в реєстраційний документ (анкета, бланк, карта та ін.), отриманих при вимірі ознаки, що вивчається, у одиниці спостереження відповідно до плану і програми дослідження

3) кількісне вираження ознаки однієї одиниці спостереження, що вивчається

**8 До ПОЗНАЧЕНЬ ЕЛЕМЕНТІВ ВАРІАЦІЙНОГО РЯДУ ВІДНОСЯТЬСЯ:**

1) V – варіанта

2) p – частота повторень варіанти

3) n – загальне число спостережень

4) m – помилка репрезентативності

5) σ – середнє квадратичне відхилення

**9 РОЗРІЗНЯЮТЬ НАСТУПНІ ВИДИ ВАРІАЦІЙНИХ РЯДІВ :**

1) простій

2) ранжируваний

3) згрупований

4) дискретний (переривчастий)

5) безперервний

6) інтервальний

7) складний

8) репрезентативний

**10 ЗАГАЛЬНИМИ ХАРАКТЕРИСТИКАМИ ЗНАЧЕНЬ ВАРІАНТ ВАРІАЦІЙНОГО РЯДУ ЯВЛЯЮТЬСЯ:**

1) середня арифметична

2) мода

3) медіана

4) амплітуда

5) коефіцієнт варіації

**11 ВЕЛИЧИНА, що НАЙБІЛЬШ ЧАСТО ЗУСТРІЧАЄТЬСЯ У ВАРІАЦІЙНОМУ РЯДУ, НАЗИВАЄТЬСЯ:**

1) медіаною (*Ме)*

2) середньою арифметичною (*М*)

3) модою (*Мо*)

4) середнім квадратичним відхиленням

**12 МЕДІАНА ВАРІАЦІЙНОГО РЯДУ (Ме) – ЦЕ:**

1) найбільша за значенням варіанту варіаційного ряду

2) варіанти, що зустрічається частіше за інших у варіаційному ряду

3) варіанти, розташована в середині ранжируваного варіаційного ряду

4) середня арифметична, вычесленная за способом моментів

**13 СПОСОБАМИ ОЦІНКИ ВИДУ РОЗПОДІЛУ ОЗНАКИ Є:**

1) графік розподілу ознаки у варіаційному ряду

2) коефіцієнт асиметрії (*β*)

3) коефіцієнт ексцесу (*γ*)

4) критерій Стьюдента (t)

5) коефіцієнт варіації (C)

**14 КОЕФІЦІЄНТ АСИМЕТРІЇ ХАРАКТЕРИЗУЄ:**

1) зміщення графіку розподілу ознаки в експерименті від кривої Гауса вліво або управо

2) відхилення форми графіку розподілу ознаки, що вивчається, від кривої Гауса

3) варіабельна ознаки

4) достовірність відмінності середніх величин

**15 СЕРЕДНЄ КВАДРАТИЧНЕ ВІДХИЛЕННЯ ХАРАКТЕРИЗУЄ:**

1) можливі відхилення параметрів генеральної сукупності в порівнянні з вибірковою сукупністю

2) різницю між вариантой і середньою арифметичною

3) середнє відхилення усіх варіант варіаційного ряду від середньої арифметичної

**16 "ПРАВИЛО ТРЬОХ СИГМ" – ЦЕ:**

1) правило відбору одиниць спостереження в статистичну сукупність

2) імовірнісна залежність між значенням середньої арифметичної, середнім квадратичним відхиленням і варіантами

3) відношення середньої величини до середнього квадратичного відхилення

**17 РЕПРЕЗЕНТАТИВНІСТЬ – ЦЕ:**

1) відповідність середньої арифметичної величини вибірковій сукупності аналогічному параметру генеральної сукупності

2) поняття, що характеризує зв'язок між ознаками

3) характеристика методики дослідження

**18 ДЛЯ ОЗНАК, що ПІДКОРЯЮТЬСЯ НОРМАЛЬНОМУ ЗАКОНУ РОЗПОДІЛУ, ДОСТОВІРНІСТЬ ВІДМІННОСТІ ДВОХ СЕРЕДНІХ ВЕЛИЧИН ВИЗНАЧАЄТЬСЯ З ДОПОМОГОЮ:**

1) помилки репрезентативності

2) коефіцієнти варіації

3) середньою арифметичною

4) критерію Стьюдента

**1 МЕТОДИ СТАТИСТИЧНОГО АНАЛІЗУ МЕДИЧНИХ ДАНИХ**

У медичних дослідженнях найбільш часто використовують дві групи методів:

1) Критерії, що дозволяють виявити відмінності між досліджуваними препаратами і досліджуваними групами хворих з яких - небудь параметрами.

2) Методи, що встановлюють наявність і ступінь взаємозв'язку між двома або кількома ознаками.

Рідше використовується третя група, до якої відносяться більш складні способи аналізу даних - багатовимірні методи дослідження (факторний аналіз, кластерний аналіз і т. ін.).

В першу чергу вибір критеріїв визначається тим, які параметри Ви збираєтеся аналізувати:

- Кількісні (наприклад: рівень артеріального тиску і т.д.), їх можна впорядкувати і щоб між ними чинити арифметичні дії.

- Порядкові (наприклад: ступінь тяжкості захворювання - легка, середня, важка) їх можна впорядкувати, але арифметичні дії над ними робити не можна.

- Якісні (наприклад: забарвлення бактерій за Грамом - червона або синя), їх не можна впорядкувати і над ними не можна проводити арифметичні дії, їх можна описати, підрахувавши число спостережень, що мають одне і теж значення, а потім оцінити (порівняти) групи за кількістю спостережень, мають різні значення.

Другим фактором, що має більше значення для кількісних ознак, є характер розподілу, які спостерігаються Вами об'єктів. Виходячи з меркантильних міркувань слід розглянути два види розподілів:

1) Нормальний.

2) Ненормальний (яке включає ряд конкретних розподілів).

Існують кілька способів оцінки приналежності, отриманого Вами розподілу до нормального:

1) Побудова гістограми частотного розподілу ознаки і визначення основних статистичних параметрів сукупності. Якщо Ви отримаєте форму гістограми близьку до тієї, що зображена на рис. 1, то Ви маєте справу з нормальним розподілом. При цьому медіана близька або дорівнює середньому, інтервал включає не менше 95% спостережень, асиметрія близька до нуля.

2) Використання нормального імовірнісного графіка (функція реалізована в комп'ютерних статистичних програмах). Якщо при цьому, отримані значення ознаки розташовуються майже на одній прямій, то розподіл близький до нормального.

3) Використання різних критеріїв (Колмогорова-Смирнова), які оцінюють ступінь відхилення Ваших даних від теоретичних, мають нормальний розподіл з тієї ж середньої і дисперсією, що і у Вашій групі. Якщо відмінності невеликі, то Ваші дані мають розподіл близьке до нормального. Графічне зображення нормального розподілу (рис. 1):

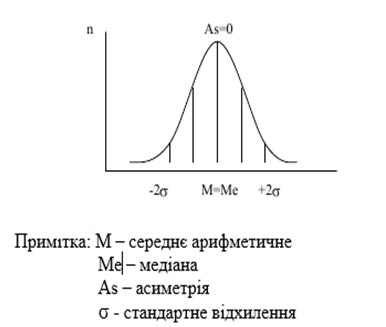


Рис. 1

Оцінка характеру розподілу повинна проводитися обов'язково, якщо, звичайно, Вас цікавить достовірність дослідження.

До числа інших факторів, що впливають на вибір конкретного критерію належать:

- кількість аналізованих груп,

- порівняння залежних або незалежних груп.

У разі нормального (або близького до нього) розподілу Ваших даних, для їх опису слід використовувати середнє і дисперсію. Для порівняння і оцінки відмінностей між групами слід використовувати параметричні критерії, тобто ті, які для виявлення відмінностей оперують параметрами розподілу (середнім і дисперсією).

У разі нормального (або близького до нього) розподілу Ваших даних, для їх опису слід використовувати середнє і дисперсію. Для порівняння і оцінки відмінностей між групами слід використовувати параметричні критерії, тобто ті, які для виявлення відмінностей оперують параметрами розподілу (середнім і дисперсією).

Якщо Ви сумніваєтеся в характері отриманого Вами розподілу, використовуйте непараметричні методи. Однак, слід пам'ятати, що часто непараметричні методи мають обмеження за обсягом аналізованої вибірки (як правило не більше 30 спостережень)

**2 Оцінка ВІДМІННОСТЕЙ між групами**

**2.1. Кількісні ознаки**

***2.1.1. Нормальний розподіл***

1. Порівняння незалежних груп.

а) Порівняння двох груп

ПРИКЛАД 1. Оцінити відмінності між чоловіками (перша група) і жінками (друга група) за рівнем артеріального тиску

МЕТОД - Звичайний двосторонній критерій Стьюдента (t-критерій) оцінює відмінності середніх.

б) Порівняння трьох і більше груп.

ПРИКЛАД 2. Оцінити відмінності за рівнем артеріального тиску в групах з різним віковим цензом: 30-39 років, 40-49 років, 50-59 років (відповідно 1-а, 2-а, 3-я групи).

МЕТОД - Дисперсійний аналіз оцінює відмінності середніх і дисперсії; встановлює наявність відмінностей за всіма групами (в ряду оцінюваних груп) .Щоб встановити між якими двома з трьох груп існують відмінності, а між якими відсутні, слід використовувати критерій Стьюдента з поправкою Бонферроні (для виключення помилки пов'язаної з множинним порівнянням).

Пояснимо викладене. Число попарних порівнянь дорівнює 3 (1 і 2 групи, 2 і 3 групи, 1 і 3 групи). Припустимо, що рівень значущості в двох порівняннях дорівнював 0,02, в третьому - 0,01 (без застосування поправки Бонферроні). Помилкове виявлення відмінностей зростає пропорційно числу порівнянь. Таким чином, справжній рівень значущості буде дорівнює добутку отриманої достовірності на число порівнянь, тобто 0,02х3 = 0,06 і 0,01х3 = 0,03. Отже при рівні значущості прийнятому в медицині (р> 0,05), достовірні відмінності виявлені тільки в одному порівнянні.

При числі порівнянь більше 8 (тобто при числі порівнюваних груп більше 4) слід використовувати критерій Ньюмена-Кейлса.

в) Порівняння декількох груп з контрольною.

ПРИКЛАД 3. Порівняти масу міокарда лівого шлуночка у здорових осіб (контрольна група) з масою міокарда у хворих I, II, III ступенем артеріальної гіпертонії (1, 2, 3 групи відповідно).

МЕТОД - Критерій Стьюдента з поправкою Бонферроні при кількості груп не більше 4. При більшій кількості груп – критерій Даннета.

2. Порівняння залежних груп (порівняння повторних вимірів)

а) Порівняння двох груп.

ПРИКЛАД 4. Оцінити достовірно чи змінився ударний обсяг у хворих на серцеву недостатність після лікування дигоксином.

МЕТОД - Парний критерій Стьюдента.

Оцінює відмінності середніх різниць показника. Вимагає нормального розподілу різниць.

б) Порівняння трьох і більше груп.

ПРИКЛАД 5. Оцінити достовірність зміни рівня холестерину при лікуванні правастатином в різні терміни спостереження - через 1, 3, 6,9 місяців.

МЕТОД - Дисперсійний аналіз повторних вимірювань. Виявляє відмінності в усій послідовності порівнянь. Для визначення між якими двома групами достовірні відмінності, а між якими - ні, використовують критерій Стьюдента або Ньюмена-Кейлса для повторних вимірів (відмінністю від аналогічних критеріїв для незалежних груп є те, що в якості оцінки дисперсії беруть залишкову варіацію, а при знаходженні критичної позначки використовують число ступенів свободи залишкової варіації).

***2.1.2. Ненормальний розподіл***

1 Порівняння незалежних груп

Приклади завдань є аналогічними при нормальному розподілі

а) Порівняння двох груп

МЕТОД - Критерій Манна-Уїтні. Альтернативними методами можуть бути: критерії Уайта, Ван дер Вадена, Колмогорова-Смирнова.

б) Порівняння трьох і більше груп.

МЕТОД - Критерій Круськала-Уолліса. Виявляє відмінності в ряду досліджуваних груп. Для оцінки відмінностей між двома конкретними групами проводять попарне їх порівняння за допомогою непараметричних варіантів критеріїв Ньюмена-Кейлса або Даннет (при однаковій кількості спостережень в групах); критерій Данна (при неоднаковій кількості спостережень в групах).

2 Порівняння залежних груп (повторні вимірювання)

а) Порівняння двох груп.

МЕТОД - Критерій Уілконсона (Вілконсона). Альтернативний метод - критерій знаків.

б) Порівняння трьох і більше груп.

МЕТОД - Критерій Фрідмана. Виявляє відмінності між усіма групами. Для оцінки відмінностей між конкретними двома групами проводять попарне порівняння за допомогою адаптованих (непараметричних) варіантів критеріїв Ньюмена-Кейлса або Даннета.

**2.2 Порядкові ознаки**

Алгоритм вибору методу аналізу той же самий, що і при ненормальному розподілі кількісних ознак.

**2.3 Якісні ознаки**

1 Порівняння незалежних груп

а) Порівняння двох груп

ПРИКЛАД 6. Чи впливає призначення стрептокінази на реканалізацію коронарних артерій у хворих на гострий інфаркт міокарда. Плацебо отримували 40 хворих, реканалізація зареєстрована у 8 чоловік. Стрептокіназу вводили 50 хворим, реканалізація відзначена у 20 чоловік.

МЕТОД - Критерій Z з поправкою Йетсена на безперервність. Аналогічний з математичного виразу критерієм Стьюдента, але використовуються часткові [p і (1-p)] значення ознаки. Табличний варіант прикладу 6 виглядає наступним чином:

|  |  |  |
| --- | --- | --- |
| Препарат | Є ефект | Немає ефекту |
| Плацебо | 0,2 | 0,8 |
| Стрепкокіназа | 0,4 | 0,6 |

Критерій Z застосуємо, якщо твір числа спостережень на значення частки більше 5 в кожній з клітин таблиці. При недотриманні останнього правила слід використовувати точний критерій Фішера (в цьому випадку поле таблиці вводять абсолютне число спостережень).

б) Порівняння трьох і більше груп.

ПРИКЛАД 7. Умова та ж сама, що і в прикладі 6, але доданий третій варіант лікування - стрептокіназа + гепарин, при якому з 45 тих, що лікувалися, реканалізація спостерігалася у 25 осіб.

МЕТОД - Таблиця спряженості з оцінкою відмінностей за допомогою критерію χ2. В поле таблиці вводять абсолютні числові значення ознаки. Критерій χ2 має високу надійність, якщо у всіх полях таблиці немає значень ознаки менше 5. При цьому оцінюється наявність відмінностей між усіма групами. Для виявлення між якими двома групами достовірні відмінності, а між якими - ні, слід будувати кілька таблиць з чотирьома полями (що містять результати двох з трьох методів лікування) з оцінкою кожної з них за критерієм χ2 з поправкою Бонферроні.

2. Порівняння залежних груп (повторні виміри)

а) Порівняння двох груп.

ПРИКЛАД 8. При лікуванні 40 хворих АГ анаприліном ефект спостерігався у 25 осіб, а при лікуванні цих хворих гипотиазидом ефект був відзначений у 16 чоловік.

МЕТОД - Критерій Мак-Німара.

Критерій також може бути використаний, але в цьому випадку він має меншу чутливістю.

б) Порівняння трьох і більше груп

ПРИКЛАД 9. Аналогічний прикладу 8, але додається третій препарат празозин - при лікуванні яких ефект спостерігався у 18 хворих. Оцінити достовірні відмінності між препаратами за кількістю хворих з наявністю ефекту.

МЕТОД - Критерій Кокрена. Альтернативним методом може служити таблиця спряженості з оцінкою відмінностей по .

Для виявлення конкретних двох груп між якими відмінності за кількістю хворих з ефектом достовірні, загальну таблицю слід розбити на кілька таблиць з чотирьома полями і порівняти їх послідовно за допомогою критерію .

**2.4 Аналіз залежності між ознаками**

1. Кількісні ознаки, нормальний розподіл

ПРИКЛАД 10. Вивчити наявність і характер взаємозв’язку між ростом і вагою дорослих чоловіків.

МЕТОД - Регресійний аналіз.

Дозволяє оцінити характер зв’язку (лінійна, нелінійна) між досліджуваними ознаками; побудувати рівняння регресії, використовуючи який можна передбачити ступінь зміни однієї з досліджуваних ознак (наприклад, ваги) при зміні другої ознаки (росту) на певну величину.

МЕТОД - Кореляційний аналіз

Щоб оцінити силу зв’язку між вагою і ростом слід використовувати кореляційний аналіз, а саме визначити коефіцієнт кореляції Пірсона.

Цей метод крім ступеню взаємозв’язку відображає її напрям (прямий чи зворотній).

2. Кількісні ознаки, ненормальний розподіл або порядкові ознаки

ПРИКЛАД 11. Визначити, чи є залежність між степенем тяжкості пневмонії і числом ускладнень цього захворювання.

МЕТОД - Коефіцієнт кореляції Спірмена.

При аналізі кількох груп використовують коефіцієнт кореляції Кенделла.

3. Якісні ознаки.

ПРИКЛАД 12. Оцінити, чи є залежність між результатом захворювання (видужав, помер) і схемою лікування (традиційна, традиційна + новий препарат).

МЕТОД - Коефіцієнт спряженості.

При таблице з чотирьома полями може використовуватися коефіцієнт асоціації Юла.

**3 СУЧАСНЕ ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ СТАТИСТИЧНОЇ ОБРОБКИ БІОМЕДИЧНИХ ДОСЛІДЖЕНЬ**

18

Всі програми статистичної обробки даних можна розділити на професійні, напівпрофесійні (популярні) і спеціалізовані. Статистичні програми відносяться до наукомісткого програмного забезпечення, ціна їх часто недоступна індивідуальному користувачеві. Професійні пакети мають велику кількість методів аналізу, популярні пакети - кількість функцій, достатню для універсального застосування. Спеціалізовані ж пакети орієнтовані на якусь вузьку область аналізу даних. Творці програмних статистичних пакетів заявляють, що їх продукт перевершує аналоги. Відсутність у більшості дослідників часу для освоєння декількох програм, робить непростим її вибір. Далі наведена базова інформація про присутніх на ринку основних напівпрофесійних програмних пакетах придатних для статистичної обробки біомедичних даних.

**MS Excel**. Найчастіше згадуваною (і використовуваною) у вітчизняних статтях є додаток MS Excel з пакету офісних програм компанії Microsoft MS Office. Причини цього криються в широкому поширенні цього програмного забезпечення, наявності російськомовної версії, тісною інтеграцією з MS Word і PowerPoint. Однак, MS Excel - це електронна таблиця з досить потужними математичними можливостями, де деякі статистичні функції є просто додатковими вбудованими формулами. Розрахунки зроблені при її допомоги не визнаються авторитетними біомедичними журналами. Також в MS Excel неможливо побудувати якісні наукові графіки. Безумовно, MS Excel добре підходить для накопичення даних, проміжного перетворення, попередніх статистичних прикидок, для побудови деяких видів діаграм. Однак остаточний статистичний аналіз необхідно робити в програмах, які спеціально створені для цих цілей.

**STADIA**. Програма вітчизняної розробки з 16-и річною історією. Включає в себе всі необхідні статистичні функції. Вона прекрасно справляється зі своїм завданням - статистичним аналізом. Але. Програма зовні практично не змінюється з 1996 року. Графіки і діаграми, побудовані за допомогою STADIA, виглядають в сучасних презентаціях архаїчно. Колірна гамма програми (червоний шрифт на зеленому) дуже стомлює в роботі. До позитивних якостей програми можна віднести російськомовний інтерфейс і наявність книг описують роботу. Наприклад, [[[1]](#footnote-1)]

**SPSS** (Statistical Package for Social Science). Найчастіше використовуваний пакет статистичної обробки даних з більш ніж 30-и річною історією Відрізняється гнучкістю, потужністю можна використовувати для всіх видів статистичних розрахунків застосовуються в біомедицині. Нещодавно вийшла 13-а англомовна версія. Існує російськомовне представництво компанії яке пропонує повністю русифіковану версію SPSS 12.0.2 для Windows. З'явився підручник російською мовою, що дозволяє крок за кроком освоїти можливості SPSS, репетитор по статистиці російською мовою, що допомагає у виборі потрібної статистичної або графічної процедури для конкретних даних і завдань, а також довідка з SPSS Base і SPSS Tables. Російський офіс SPSS регулярно проводить навчальні курси з аналізу даних за допомогою програмного забезпечення SPSS. На російську мову переведена книга по SPSS, яка вийшла в світ у 2002 році в Київському видавництві «Діасофт» під назвою «SPSS 10: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей».

**STATA**. Професійний статистичний програмний пакет з data-management system, який може застосовуватися для біомедичних цілей. Один з найпопулярніших в освітніх і наукових установах США поряд з SPSS. Програма добре документована, видається спеціальний журнал для користувачів системи. Однак можливості попереднього ознайомлення з демо-версією немає.

**STATISTICA**. Виробником програми є фірма StatSoft Inc. (США) яка випускає статистичні додатки, починаючи з 1985 року. STATISTICA включає велику кількість методів статистичного аналізу (більше 250 вбудованих функцій) об'єднаних наступними спеціалізованими статистичними модулями: Основні статистики і таблиці, Непараметрична статистика, Дисперсійний аналіз, множинна регресія, Нелінійне оцінювання, Аналіз часових рядів і прогнозування, Кластерний аналіз, Аналіз, Дискримінантний функціональний аналіз, Аналіз тривалості життя, Канонічна кореляція, Багатомірне шкалювання, Моделювання структурними рівняннями і ін. Нескладний в освоєнні цей статистичний пакет може бути рекомендований для біомедичних досліджень будь-якої складності.

**JMR**. Один зі світових лідерів в аналізі даних. Розвиває цей статистичний пакет SAS Institute який викупив в кінці 2002 року відому статистичну програму StatView. Однак особливих переваг для медико-біологічної статистики цей програмний продукт не має.

**SYSTAT** Статистична система для персональних комп'ютерів Остання 11 версія має непоганий інтуїтивно зрозумілим інтерфейсом. Компанія Systat Software також розробляє популярні у вітчизняних дослідників SigmaStat і SigmaPlot, які є відповідно, програмою статистичної обробки та програмою побудови діаграм. При спільній роботі стають єдиним пакетом для статистичної обробки і візуалізації даних.

**NCSS**. Програма розвивається з 1981 року і розрахована на непрофесіоналів в області статистичної обробки. Інтерфейс системи багатовіконний і як наслідок цього явища - трохи незвичний у використанні. Всі дії користувача супроводжуються підказками. Зараз доступна версія 2004 р.

**MINITAB 14**. Статистичний пакет MINITAB в даний час випускається у версії 14. З сайту виробника можна взяти повнофункціональний пробний варіант програми, яка працює 30 днів. Це досить зручний в роботі програмний пакет, який має хороший інтерфейс користувача, хороші можливості по візуалізації результатів роботи. Має докладну довідку.

**STATGRAPHICS PLUS**. Досить потужна статистична програма. Містить більше 250 статистичних функцій, генерує зрозумілі, що настроюються звіти.

**PRISM**. Ця програма створювалася спеціально для біомедичних цілей. Інтуїтивно зрозумілий інтерфейс дозволяє в лічені хвилини проаналізувати дані і побудувати якісні графіки. Програма містить основні часто застосовуються статистичні функції, яких в більшості досліджень буде досить. Однак, як відзначають самі розробники, програма не може повністю замінити серйозних статистичних пакетів.

На якій програмі зупинити свій вибір? Безумовно, дорожнеча програм не дозволяє їх міняти. Тому має сенс подивитися демо-версії, розібратися з роботою і потім робити остаточний висновок. Російськомовні версії (з документацією) мають тільки SPSS та STATISTICA.

Що стосується можливих рекомендацій, то вони наступні:

- Якщо потрібен потужний, загальновизнаний пакет з простим і зрозумілим навіть початківцям користувачам інтерфейсом, то краще скористатися SPSS.

- Для початківців і професіоналів, яким потрібна підказка і розвинена документація російською мовою, можна рекомендувати STATISTICA. Це потужна програма з професійними можливостями.

- Для невибагливих користувачів, які обмежуються в своїх дослідженнях стандартними статистичними методами можна рекомендувати англомовну програму Prism.

# 4 СТИСЛІ ВІДОМОСТІ ПРО системУ Statistica 6

1 Структура пакету Statistica 6.

Пакет Statistica 6складається з декількох незалежно працюючих модулів. У кожному модулі зібрані логічні пов’язані між собою статистичні процедури. Завантажити можна відразу декілька модулів, але активним може бути лише один.

2 Створення нової таблиці даних.

Обрати пункт меню *Файл – Создать – Ok (File – New – Spreadsheet - ОК)*. Буде створена пуста електронна таблиця розміром 10х10. У стовпчиках розміщені перемінні (Vars), у рядках – спостереження (Cases).

3 Видалення і додавання нових перемінних і спостережень.

Робота з перемінними і спостереженнями виконується за допомогою кнопок панелі інструментів *Переменные (Variables)* і *Наблюдения (Cases)*. Команди *Добавить (Add)* і У*далить (Delete)* видаляють виділенні або додають після виділенего рядка або стовпчика нові. При додаванні потрібно вказати кількість елементів, що додаються і вказати після якого елементу буде виконано додавання – *Ok*.

4 Корегування таблиць

Корегування зводиться до змін імен перемінних, вмісту стовпчиків повністю та окремих комірок. Дляцього подвійний клік по назві стовпчика відкриє вікно, у якому можна задати нове ім’я стовпчику у рядку *Имя (Name)* і нові значення випадкам за допомогою формули, а також можна змінити вигляд числа (автоматично – 8 позицій, після коми – 3).

Корегування числових значень в окремих комірках виконується як в EXCEL.

5 Обчислення характеристик для значень перемінних

Наприклад, максимальне, мінімальне, середнє значення, дисперсія і т. ін. Активувати таблицю даних, потім активувати пункти меню *Анализ – Основные статистики и таблицы - Описательные статистики (Statistics – Basic statistics – Descriptive statistics), обрати* зі списку потрібні статистичні характеристики (наприклад, min або mean), *Ok*, виділити перемінні, для яких шукаються характеристики *Ok, Ok*. На екран буде виведена таблиця з потрібними характеристиками.

6 Отримання графіку та рівняння лінійної регресії

Петивувати таблицю. обрати пункт меню *Графика – Диаграммы рассеяния (Graphs - Scatterplots)* обрати перемінні (для аргументу – Х і функції – Y), *Ok, на вкладці Дополнительно (Advanced)* обрати *Простой (Regular) – Линейная (Linear)* – *Ok.* З’явиться графік лінійної регресії, над яким записано рівняння регресії.

7 Типи файлів в Statistica

Разом 4 типи файлів:

\*.sta - вихідні дані;

\*.sсr - результати статичної обробки даних;

\*.stg - графік;

\*.rtf– звіт.

В одному сеансі роботи з пакетом може бути створено або відчинено скільки завгодно файлов \*.sсr, \*.stg, \*.rtf, але лише один файл з вихідними даними \*.sta.

8 Створення автозвіту.

Автозвіт необхідно створювати при кожному сеансі роботи з пакетом для того, щоб усі результати роботи (таблиці і графіки) потрапляли в автозвіт.

Активувати таблицю \*.sta. Далі вказати шлях для створення автозвіту: *Файл – Добавить в Отчет – Новый Отчет. (File – Add to report – Add to new report)* З’явиться пусте вікно автозвіту. Його можна звернути. Усі розрахункові таблиці і графіки з’являться у звіті. Файл звіту має розширення \*.rtf, його можна редагувати у WORD.

**5 МетодичНІ рекомендацІЇ ДО вИКОНАННЯ практичНих заВДАНЬ**

**5.1 Лабораторна робота № 1. Дослідження ефективності щеплень**

*Стислі теоретичні відомості*

Багато задач медицини можуть бути вирішені за допомогою статистики і всі модулі системи STATISTICA, так чи інакше, використовуються в медицині. Перш за все, в медицині статистика використовується в задачах, пов'язаних з вибірковими обстеженнями, з перевіркою ефективності різних доз різних ліків, діагностика захворювань на підставі проведених медичних аналізів, виявлення сезонних чинників і прихованих періодичностней (наприклад, визначення того, як народжуваність залежить від місяця або дня тижня), оцінка інтенсивності викликів швидкої допомоги в залежності від часу доби, прогнозування одужання хворих, оцінка залежностей між різними змінними, наприклад, як стан зубів дітей пов'язано зі способом годування (годування грудьми або штучне годування) і т. ін.

Ситуація ускладняється тим, що часто дослідник має неповними даними (спостереження можуть бути цензурувати, наприклад, пацієнт переведений в інший госпіталь або виписаний і зв'язок з ним втрачено). Для аналізу неповних спостережень застосовуються спеціальні статистичні методи, реалізовані в модулі *Аналіз виживаності*.

Крім того, дані можуть містити багато пропусків. Методи обробки пропущених значень доступні в будь-якому модулі системи.

Якщо аналізуються дані невеликого обсягу і немає ніяких обґрунтованих припущень про форму розподілу спостережуваних величин, застосовуються методи, реалізовані в модулі *Непараметрична статистика*.

Якщо потрібно провести класифікацію хворих, описуваних декількома ознаками, то слід використовувати модуль *Дискримінантний аналіз*. Однак методи дискримінантного аналізу "працюють", якщо перемінні виміряні в досить багатій шкалою, наприклад, інтервального. Якщо предиктори виміряні в номінальною або порядкової шкалою, то використовують методи модуля *Дерева класифікації*.

Для задач прогнозування одужання пацієнтів застосовуються методи *множинної регресії*. Однак у багатьох ситуаціях доводиться мати справу з кусочно - лінійними моделями, тому що до певного критичного моменту (наприклад, через 23 дні після операції) адекватна одна лінійна модель і хороший прогноз дає одне підмножина предикторов. Після критичного моменту, модель змінюється і змінюється безліч предикторов, що дозволяють будувати прогноз. Методи модуля *Нелінійне оцінювання* містять необхідний інструментарій для аналізу таких задач.

Спеціальні методи аналізу багатовхідних таблиць спряженості реалізовані в модулях *логлінійний аналіз* і *Аналіз відповідності*.

Широке коло завдань може бути вирішене в модулі *Основні статистики і таблиці*.

*Приклад виконання*

Розглянемо наступний приклад «**Дослідження ефективності щеплень» за допомогою** аналізу в модулі *Основные статистики и таблицы (Basic statistics)*.

Припустимо, Вам потрібно дослідити, як пов’язані щеплення проти деякої хвороби і захворювання цією хворобою. Випадковим чином Ви обираєте історії хвороб з архіву і створюєте файл даних (рис. 2):

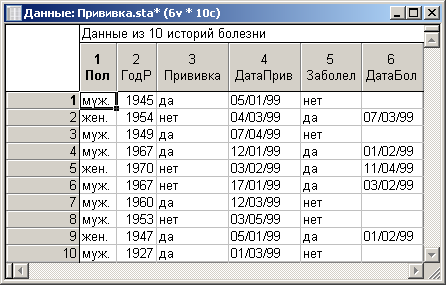


Рис. 2

У файлі міститься додаткова інформація: стать, рік народження, дата щеплення і, якщо обстежуваний захворів, то також наводиться дата захворювання. Ця інформація корисна для подальшого, поглибленого аналізу даних, а в ході виконання цього завдання вона не використовується. Ви хочете встановити, як пов'язані дві ознаки (і чи пов'язані вони взагалі): щеплення і захворюваність.

Досліджувані перемінні: щеплення і захворюваність є *категоріальним*, тому що приймають тільки 2 значення: так чи ні (суб'єкт потрапляє в одну з двох категорій). Такі змінні сильно відрізняються, наприклад, від перемінних, що вимірюють температуру або тиск, рівень холестерину, які виміряні в багатшою шкалою - інтервального шкалою (грубо кажучи, є безперервними).

Для вивчення зв'язків або залежності між категоріальним змінними розроблений спеціальний апарат - *таблиці спряженості*, до побудови яких в системі STATISTICA ми зараз перейдемо.

У даному прикладі буде побудована таблица 2 на 2 чи, як її ще иноді називають, чотирьохкомірчата таблиця, оскільки в ній є лише 4 комбінаціх (так, так), (так, ні), (ні, ні), (ні, так) значень перемінних: ПРИВИВКА, ЗАБОЛЕЛ. Така таблиця зветься таблицею спряженості. Для цього потрібно обрати пункт меню *Анализ – Основные статистики и таблицы (Statistics – Basic statistics)* (рис.3).

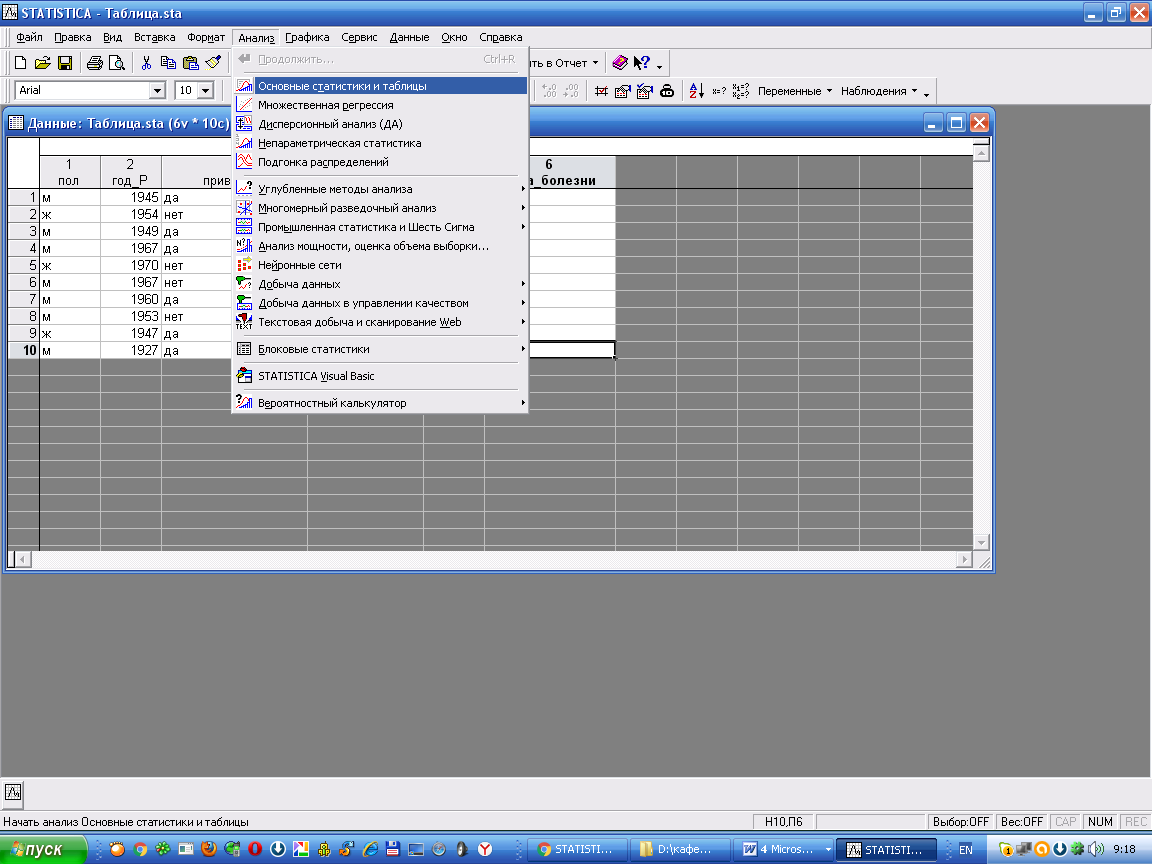


Рис. 3

У стартовій панелі модуля *Основные статистики и таблицы* *(Basic statistics)* потрібно обрати процедуру *Таблицы сопряженности, флагов и заголовков (Tables and banners)* (рис.4).

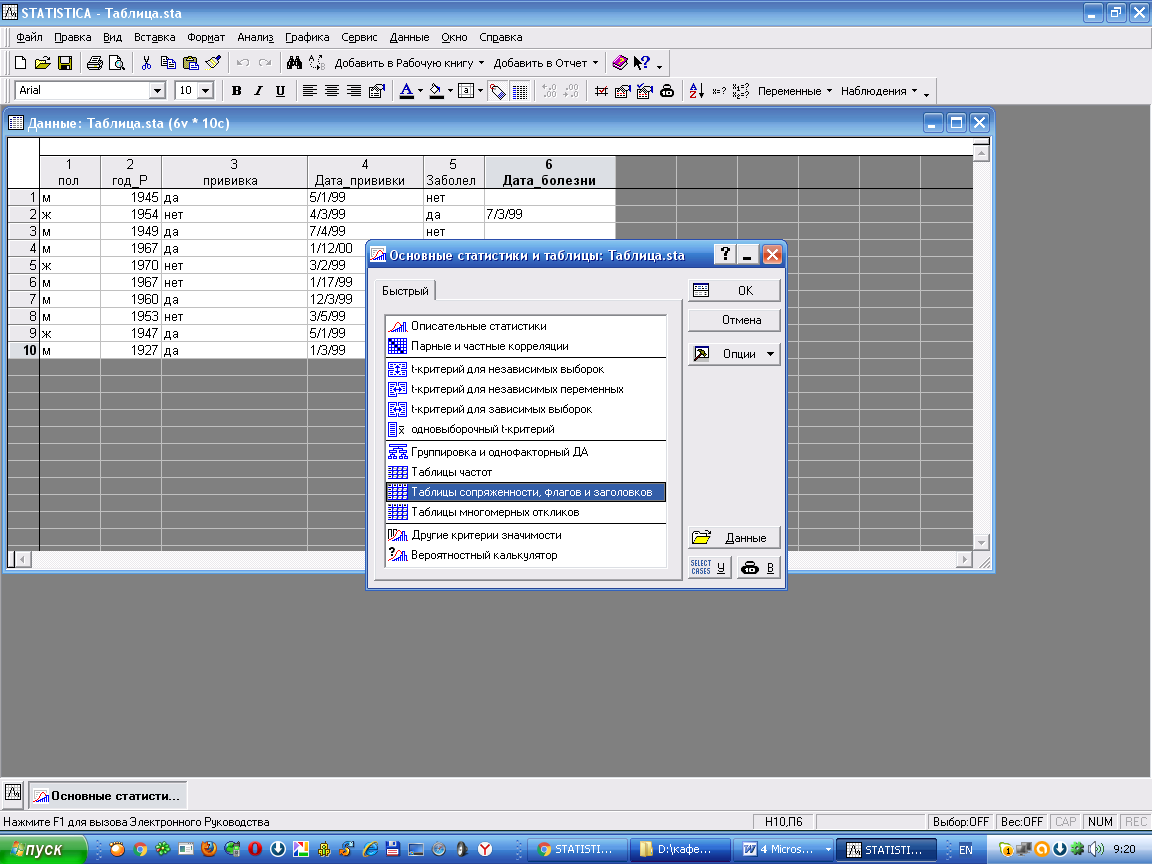


Рис. 4

Нажмите кнопку *ОК*, а потім у діалоговому вікні *Задайте таблицы (Crosstabulation tables)* нажмите кнопку *Задайте таблицы (Specify tables (select variables)* (рис. 5).

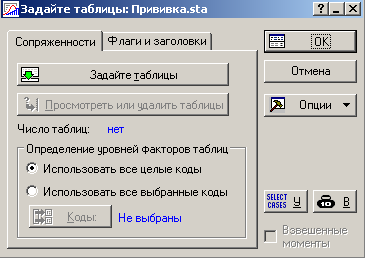


Рис. 5

Оберіть перемінні, як вказано нижче (рис. 6):

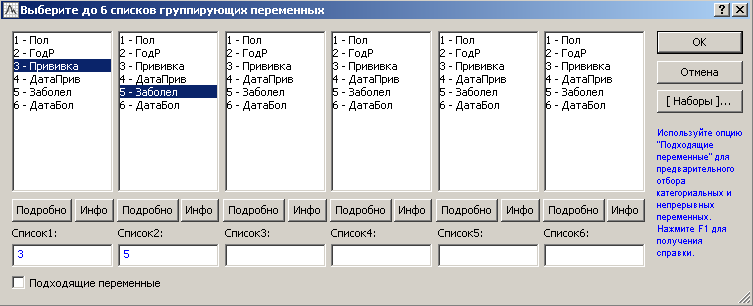


Рис. 6

Разом можна обрати до 6 списків перемінних шрупування (тобто тих перемінних, які задають розділ пацієнтів на групи), нам достатньо обрати тільки 2 перемінні. Натисніть кнопку ОК. Ви знову повернетесь до диалогового вікна задания таблицы, де слід натиснути кнопку *Использовать все выбранные коды (Use selected grouping codes only)* (рис. 7):

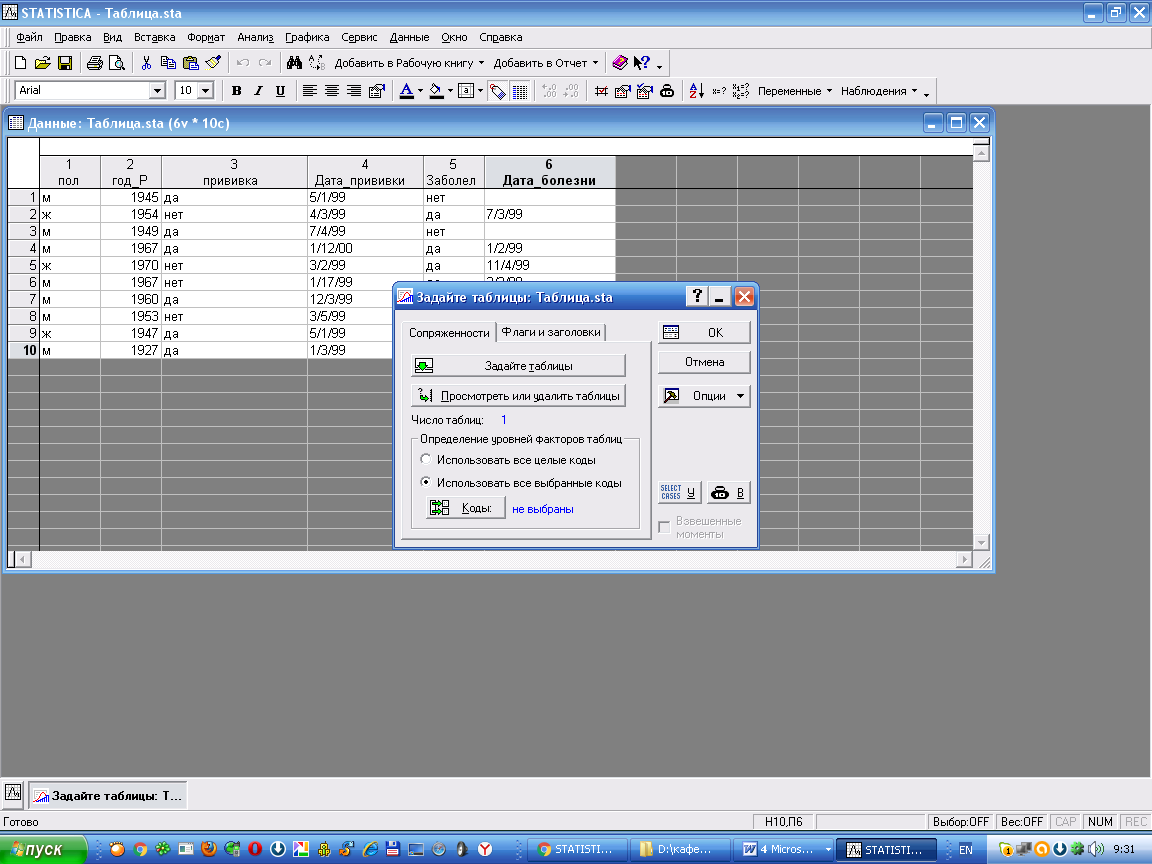


Рис. 7

Далі натисніть *Коды – Все – Все – ОК (Codes – All – All - OK)* (рис.8):

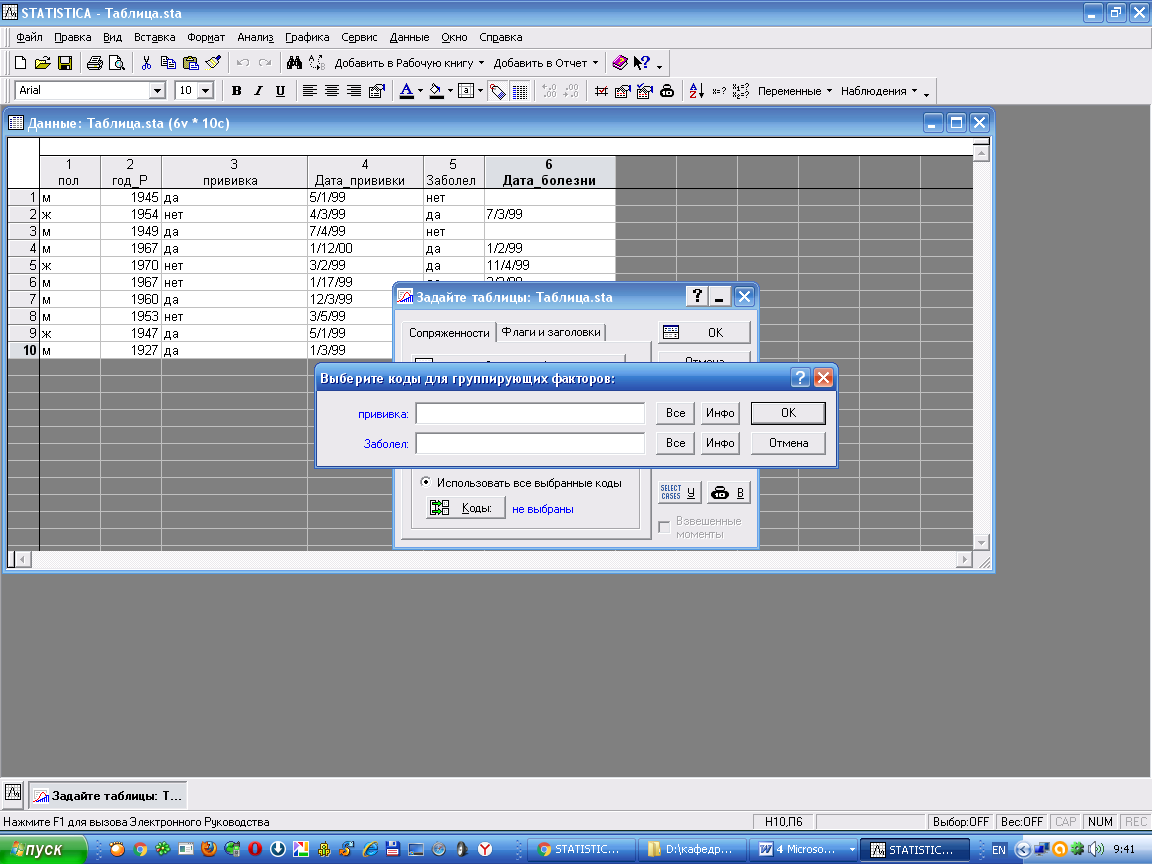


Рис. 8

Вікно *Задайте таблицы* тепер виглядає таким чином (рис. 9):

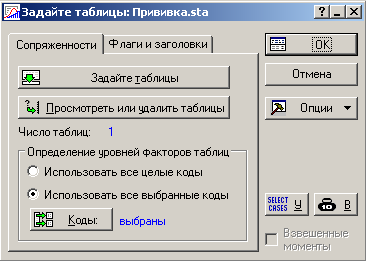


Рис. 9

Натисніть кнопку *ОК*. Тепер, у відкрившимуся діалоговому вікні *Результаты кросстабуляции (Crosstabulation tables results),* потрібно обрати вкладку *Дополнительно (Advanced)* (рис. 10):

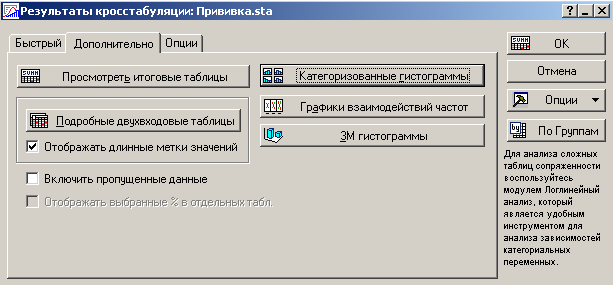


Рис. 10

Натисніть кнопку *Просмотреть итоговые таблицы*  (*Summary: Review summery tables*) і Ви побачите потрібну таблицю спряженості. Ця таблиця містить у стислому вигляді всю інформацію, що дозволяє оцінити заледність між категоріальними перемінними: ПРИВИВКА, ЗАБОЛЕЛ (рис. 11):

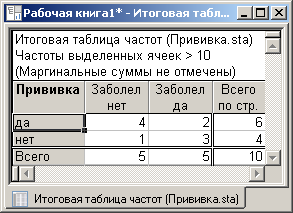


Рис. 11

З отриманої таблиці можна побачити, що 1 людина з числа досліджених – не мав щеплення і не захворів (ліва верхня комірка), 3 людини не мали щеплення і захворіли (наступна комірка у першому рядку), 4 людини мали щеплення і не захворіли (перша комірка у другому рядку), 2 людини мали щеплення і захворіли (друга комірка у другому рядку).

У нижньому рядку записані суми значень по стовпчиках. У крайньому правому стовпчику – суми значень по рядках. Ці значення іноді називають маргінальними або крайніми, оскільки вони знаходяться по вінцях таблиці.

### Далі розглянемо графічні зображення таблиць спряженості.

Побудуємо 3-вимірну гістограму частот. Для цього потрібно обрати пункт меню *Графика – 3М последовательные графики – Гистограммы двух переменных (Graphs – 3D Seq. – Bivariate histograms)* (рис. 12, 13):

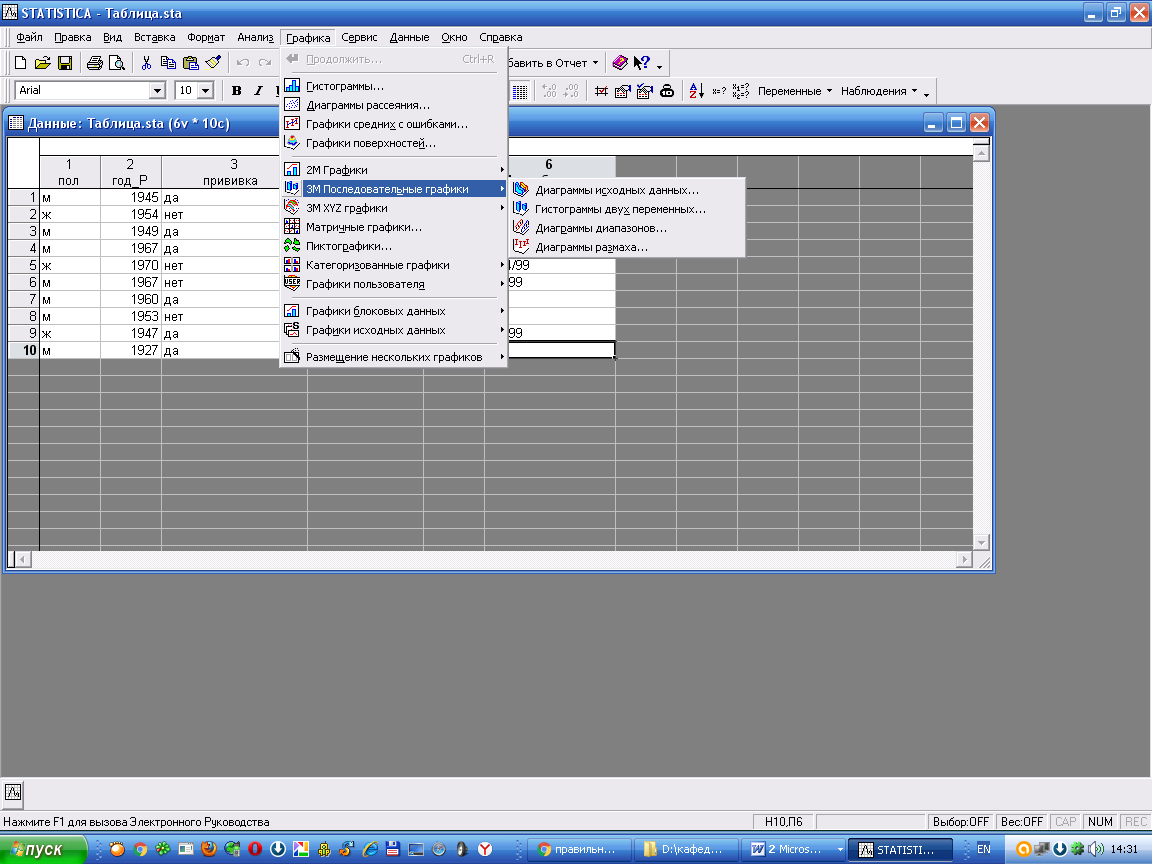


Рис. 12

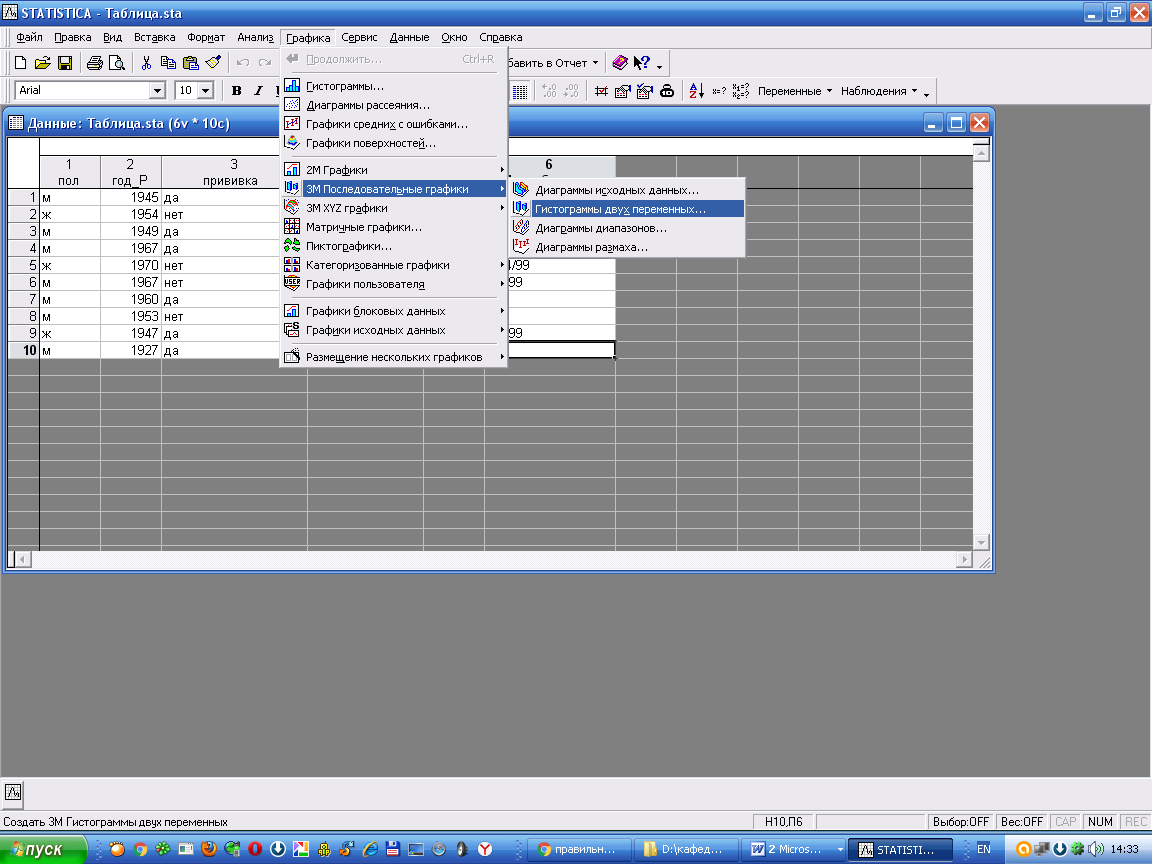


Рис. 13

В вікні, що з’явилося, *3М Гистограммы двух переменных (3D Bivariate histograms)* оберіть перемінні, як вказано нижче (рис. 14) і натисніть *ОК – ОК*:

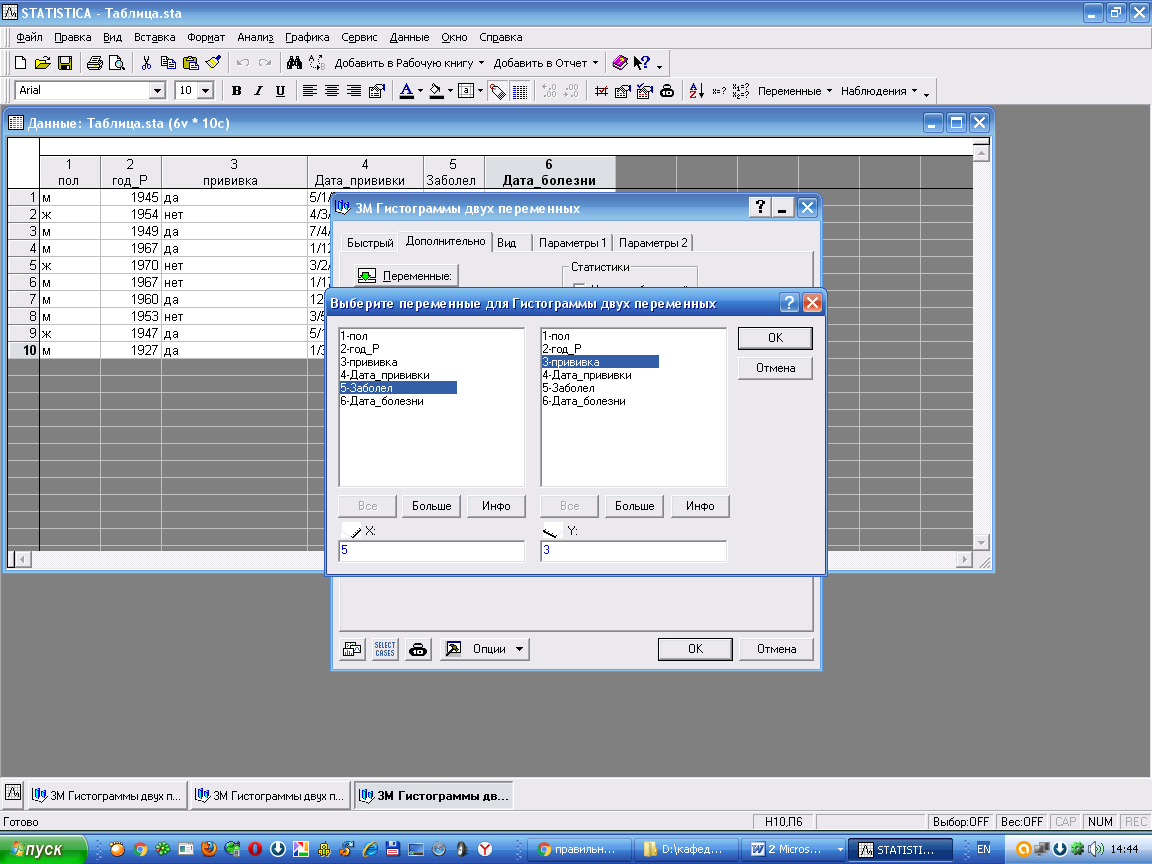


Рис. 14

Отримаємо *гістограму* частот. Це 3-вимірна гістограма частот (рис.15):

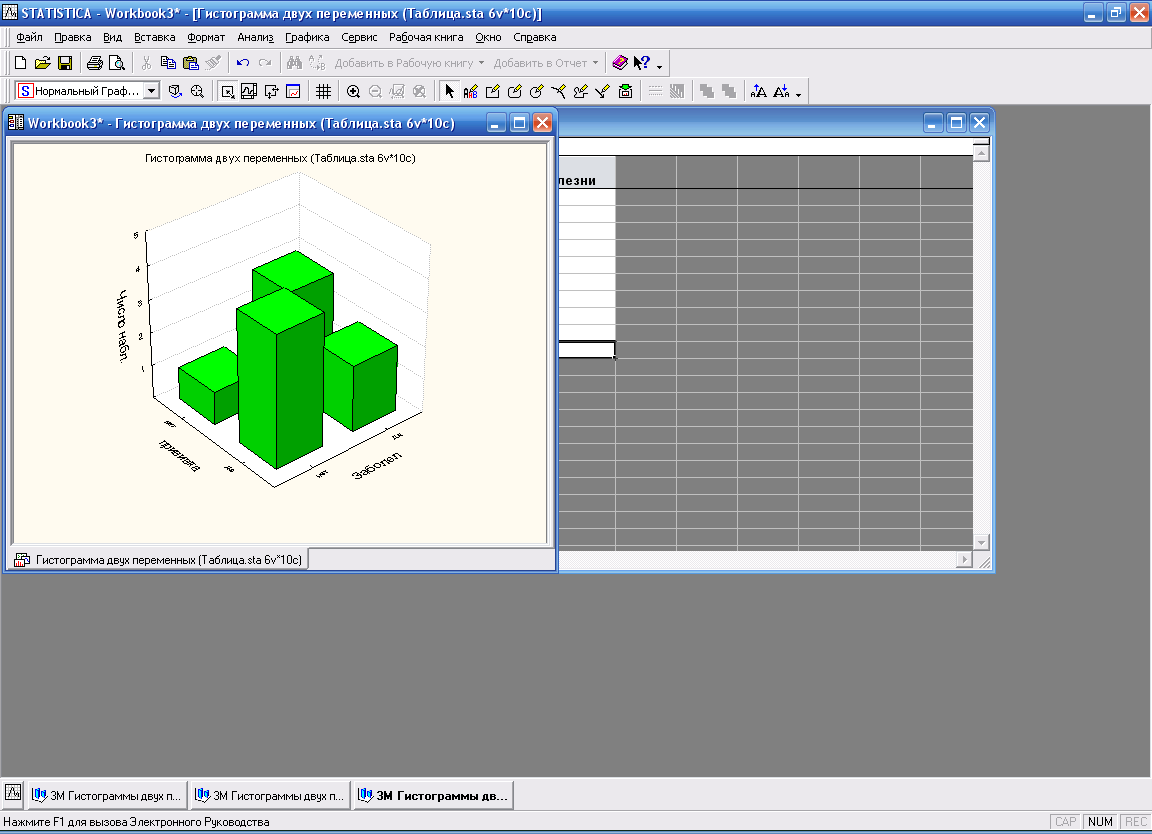


Рис. 15

Побудуємо *Категоризовані гістограми*. Для цього потрібно обрати пункт меню *Категоризованные графики – Гистограммы (2D – Histograms - Categorized)* (рис.16, 17):

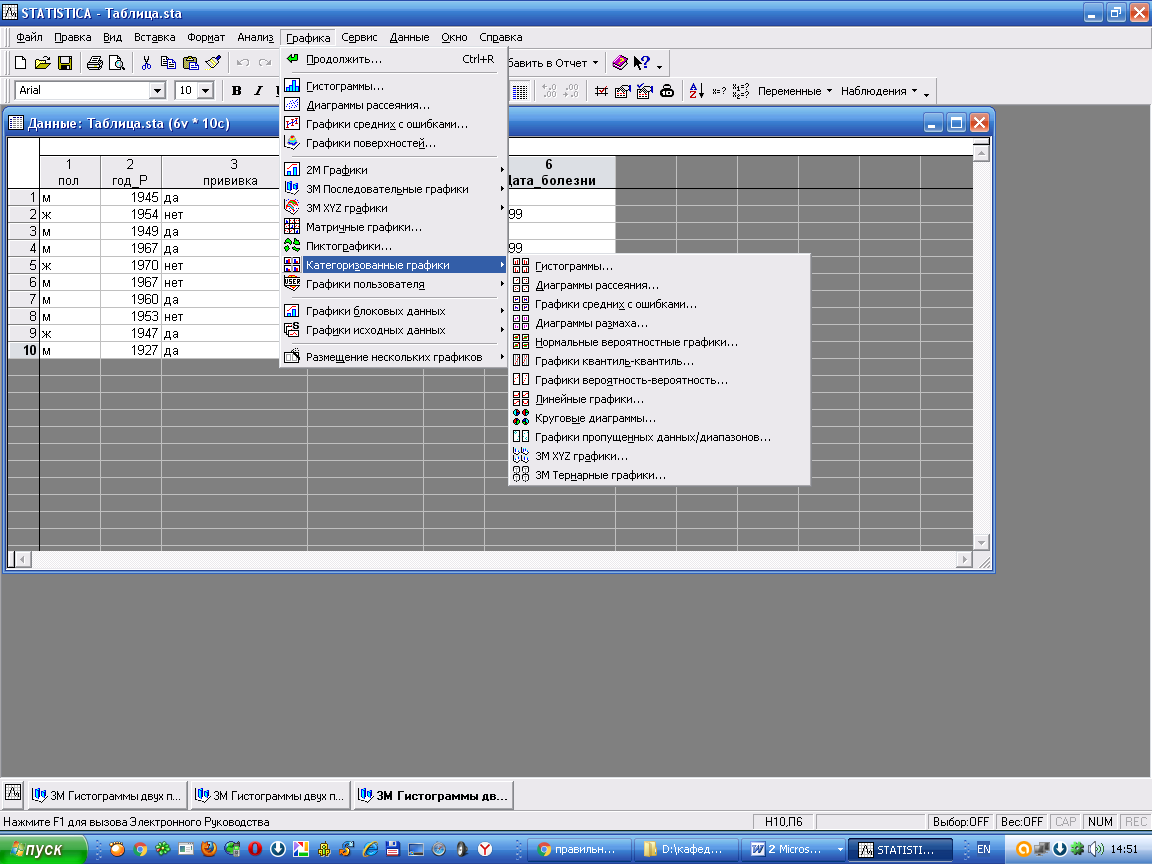


Рис. 16

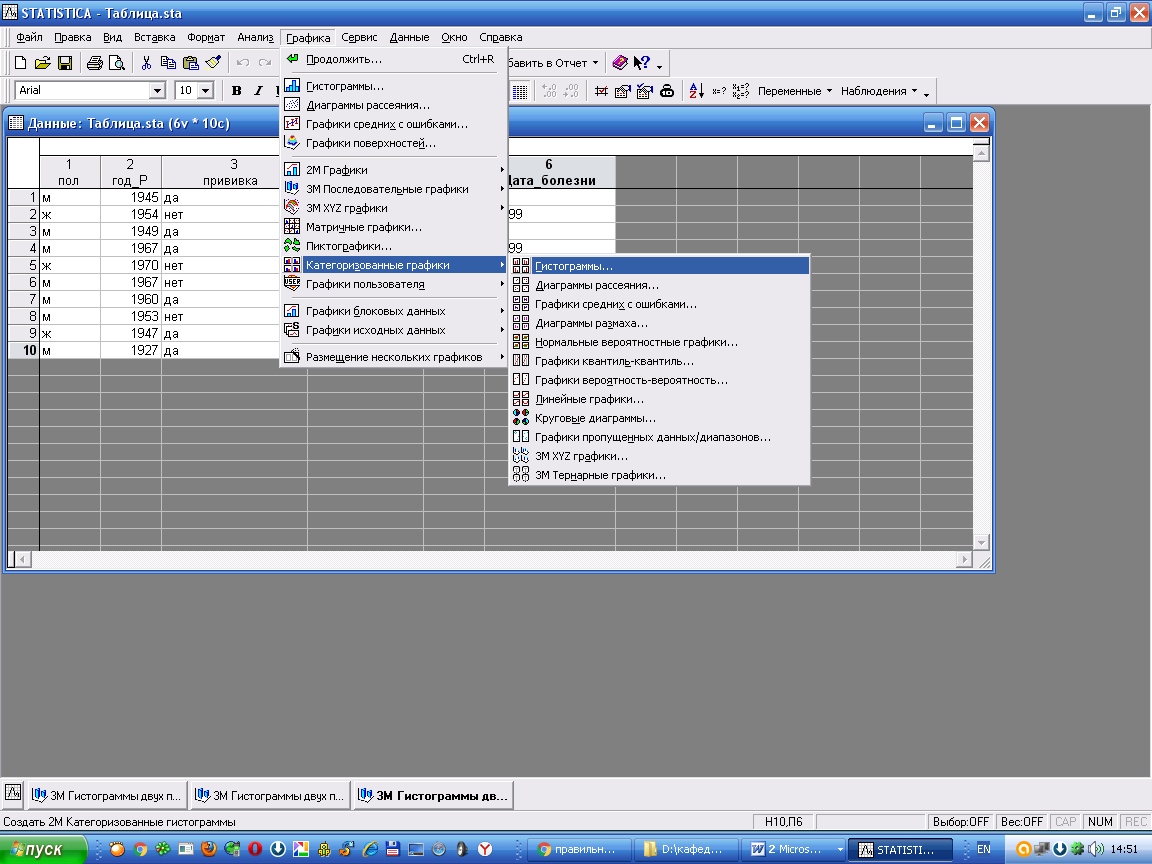


Рис. 17

У вікні, що з’явилося, 2*М Категоризованные гистограммы (2D categorized histograms)* потрібно обрати перемінні, як вказано нижче (рис.18) і натиснути *ОК – ОК*:

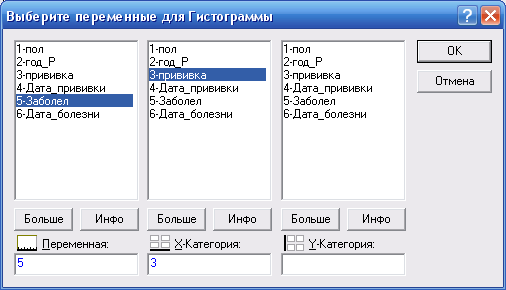


Рис. 18

Отримаємо наступні категоризовані гістограми (рис. 19):

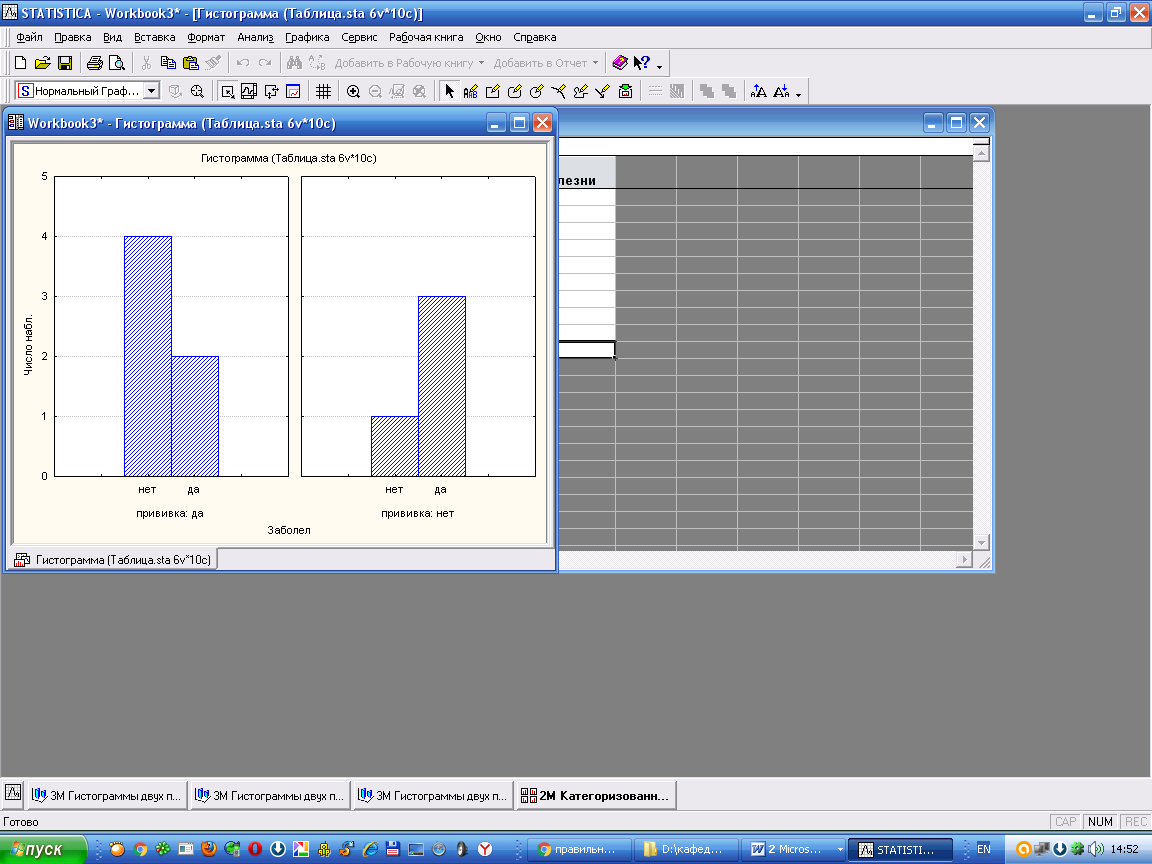


Рис. 19

Кожен з цих графіків по-своєму зручний і корисний. На другому, категорізованому графіку, як можна побачити, всі дослідження розбиті на дві категорії (групи): ПРИВИВКА - нет, ПРИВИВКА – да. З графіка видно, що число тих, хто захворів, у групі, без щеплень, більше числа хворих у групі, зі щепленнями. Якщо щеплення не зроблене, то кількість тих, хто не захворів, – 1, а хто захворів – 3. Якщо щеплення зроблене, то число тих, хто не захворів, – 4, число хворих – 2.

**5.2 Лабораторна робота № 2. Створення рандомізированої вибірки при організації медико-біологічних досліджень**

*Стислі теоретичні відомості*

Обов’язковою умовою добре спланованого експериментального дослідження є проведення рандомізації. Дослівний переклад англійського "random" означає "створений або обраний навмання, випадковий, безладний". За деякими даними, сьогодення біля 20% статей, що публікуються у провідних медичних журналах світу, мають результати рандомізованих досліджень. Під рандомізацією розуміють процедуру, що забезпечує випадковий розподіл хворих до експерементальної і контрольної групи. Слід особливо підкреслити, що рандомізацію проводять вже після того, як хворого було включено до експерименту згідно з протоколом клінічного дослідження. Фахівці, які займаються цією проблемою, підкреслюють, що випадковий, або рандомізований, розподіл не є синонімом безладного, при якому процес розподілу не піддається математичному опису. Рандомізація вважається погано організованою при розподілі хворих на групи по номеру історії хвороби, страхового полісу чи даті народження. Краще за все використовувати таблицю випадкових чисел, методом конвертів чи шляхом централізованого комп’ютерного розподілу варіантів лікування. Нажаль, згадування процесу рандомізації не тотожне правильному його проведенню. Дуже часто у статтях не вказаний спосіб рандомізації, що ставить під сумнів добре організоване дослідження.

Деякі дослідники підкреслюють перед початком випробування розподіл пацієнтів на підгрупи з однаковим прогнозом і тільки потім рандомізувати їх окремо в кожній підгрупі (стратифікаційна рандомізація). Коректність стратифікаційної рандомізації визнається далеко не всіма.

*Приклад виконання*

Маємо дані про пацієнтів ( Відкрийте файл даних *Heart.sta* за допомогою меню Файл – Открыть (File - Open)), що поступили до лікарень для трансплантації серця. Таблиця з даними має вигляд:

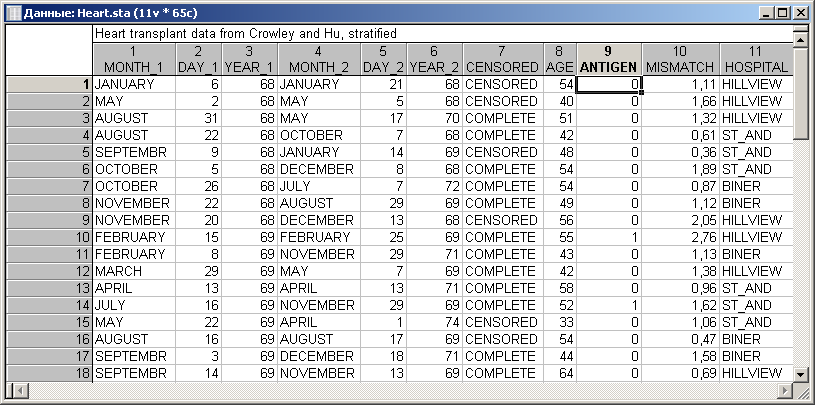


Рис. 20

Перші три стовпчики у цієї таблиці є датами трансплантації серця (у наступній послідовності: місяць-день-рік), 4, 5 и 6 стовпчики - дати, коли відповідний пацієнт або помер, либо був виключений з дослідження (іншими словами, цензурований, наприклад, з пацієнтом був утрачений зв’язок).

Перемінна Цензуровано - Censored є індикатором цензурування з кодом, який показує, чи є відповідний догляд завершеним або цензурованим (0 - завершено, тобто пацієнт помер; 1 - цензурований). Перемінна Age – вік пацієнта, Antigen – показник несумісності антигенів, Mismatch – ступінь несумісності тканин.

Перемінна Hospital є фіктивною перемінною для групування, яка відображає, до якої з трьох лікарень відноситься пацієнт.

## Рандомізація

Рандомізація - формування випадкової вибірки по вихідній таблиці даних.З цих даних будемо відокремлювати випадкові підмножини різними методами.

### **Підмножина**

У меню *Данные (Data)* обрати команду *Подмножество (Subset)*.

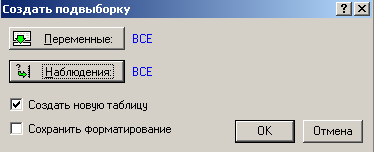


Рис. 21

*Умови вибору перемінних.* Обрати перемінні та визначити умови вибору спостережень в активному файлі даних.

Припустимо, що нам потрібні дані по хворим, що знаходяться тілько в лікарні ST\_AND. Натиснувши кнопку Наблюдения (Cases), вказуємо заданні нами умови і натиснемо OK.

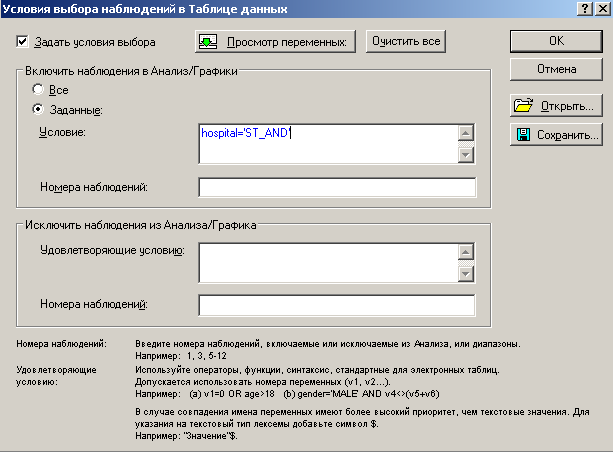


Рис. 22

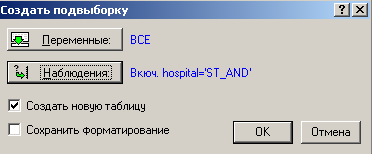


Рис. 23

В результаті, отримаємо дані, що відносяться тільки до потрібної нам лікарні.

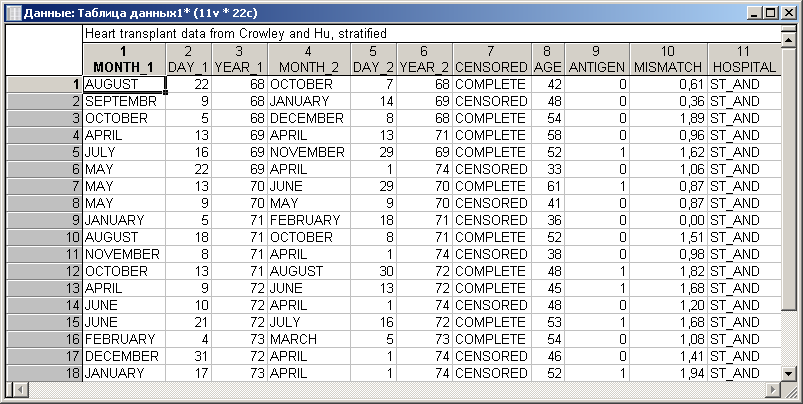


Рис. 24

### **Випадковий вибір**

В меню Данные (Data) потрібно обрати команду *Случайный выбор (Sampling)*.

У цьому діалозі знаходяться три вкладки: *Простой выбор (Simple Sampling)*, *Стратифицированный (Stratified Sampling)* і *Параметры (Options)*.

### Вкладка **Простой выбор (Simple Sampling)**

1) Простая случайная выборка (Simple Random Sampling). При виборі цього правила, дані будуть обиратися випадково.

Існують два способи вибору підмножини у загальній множині: вказати відсоток спостережень або вказати наближене число спостережень (цей параметр встановлюється в меню Опции (Options)). Припустимо, що нам потрібно проаналізувати 40% пацієнтів.

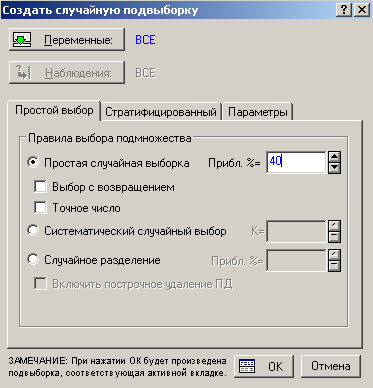


Рис. 25

Після натискання кнопки OK, отримаємо наступні результати:

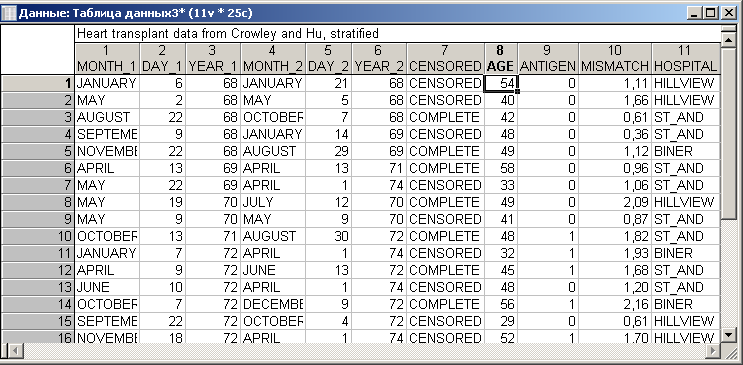


Рис. 26

Поставивши гачечок у полі *Выбор с возвращением (With replacement)*, отримаємо наступний результат: при включенні спостереження до підмножини, це спостереження знов потрапляє до вихідної множини спостережень (таким чином, одне спостереження може траплятися декілька разів у підсумковій підмножині).

2) Систематический случайный выбор (Systematic random sapling). Використовуючи цей метод, підмножина буде скоалатися за допомогою систематичного випадкового вибору.

Наприклад, якщо ввести число 5 в поле K=, то серед перших п’яти спостережень буде випадково обрано одне, а потім STATISTICA буде обирате кожне п’яте спостереження у вихідній множині даних.

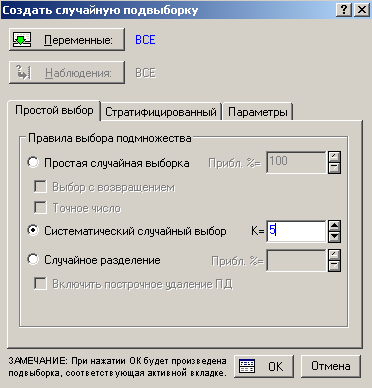


Рис. 27

3) Разделенный случайный выбор (Split node random sampling). При виборі цього методу, всі спостереження будуть випадково розділені на два файли даних. Необхідно вказати відсоток спостережень або наближене число спостережень (цей параметр установлюється в меню ***Опции (Options***).

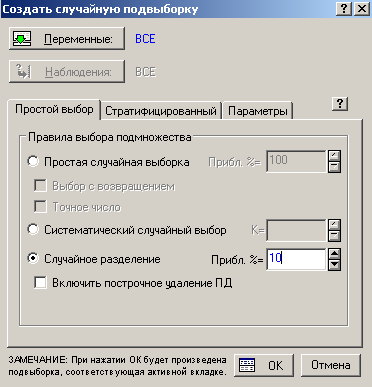


Рис. 28

Вкладка ***Стратифицированный (Stratified Sampling)***

За допомогою даної опції буде створена стратифікована випадкова вибірка на основі поточного файлу даних. Можна вказати кілька перемінних стратифікації, які містять цілі кодові значення, що визначають окремі групи (страти).

Стратифікована вибірка буде побудована на основі комбінації всіх кодів у всіх перемінних стратифікації.

Наприклад, обравши у якості перемінної стратифікації CENSORED або HOSPITAL, будуть створені різні долі вибірки для комбінацій COMPLETE-HILLVIEW, COMPLETE- ST\_AND, COMPLETE- BINER, CENSORED-HILLVIEW, CENSORED-ST\_AND, CENSORED-BINER.

Опція *Равные вероятности (Uniform probability)* вказується для вибору однієї й тієї ж долі спостережень у всіх групах. Припустимо, загальний (для всіх груп) відсоток спостережень у вибірках складає 50% (можливо також вказати наближене число спостережень).

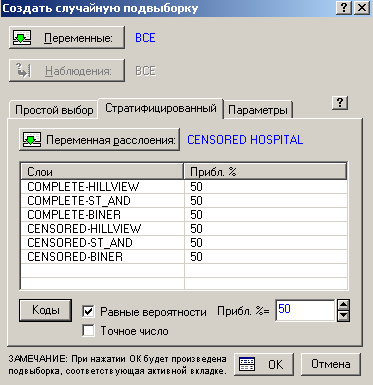


Рис. 29

Таким чином буде створена стратифікована випадкова вибірка.

### Розглянемо деякі опції вкладки *Параметры (Options):*

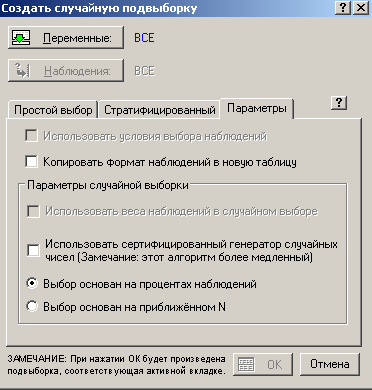


Рис. 30

1) ***Использовать условия выбора наблюдений (Use case selection condition expressions).*** Якщо ви обираєте цю опцію, то умови вибору спостережень, задаються при натисканні кнопки *Наблюдения (Cases)* в діалозі будуть використовуватися до створення підмножини.

Щоб ігнорувати умови вибору спостережень, слід убрати відмітку цієї опції

2) ***Копировать формат наблюдений в новую таблицу (Copy formatting to new spreadsheet cases)***. Обравши цю опцію, формат комірок буде копіюватися в нову підмножину.

3) ***Использовать веса наблюдений в случайном выборе***. Якщо у поточній Таблиці даних задані ваги спостережень, можна інтерпретувати ці ваги як множники спостережнь. У цьому випадку, відповідна дрібна вибірка буде ефективно підігнана до ваг спостережень.

4) ***Использовать сертифицированный генератор случайных чисел (Use Diehard-certified random number generation)***. ***STATISTICA*** використовує дуже точний і якісний генератор випадкових чисел у тих випадках, коли необхідно провести деякі процедури.

Однак у більшості процедур випадкового вибору або стратифікованого випадкового вибору можна використовувати більш прості і швидкі методи випадкового вибору спостережень. Зокрема, при роботі з дуже великими файлами даних можна прибрати позначку цієї опції для прискорення роботи.

5) ***Выбор основан на процентах наблюдений (Calculate based on percentage of cases)***. При виборі цієї опції файл даних розподіляється на основі відсотка спостережень.

6) ***Выбор основан на приближенном N (Calculate based on count of cases)***. При виборі цієї опції файл даних розподіляється на основі числа спостережень

**Завдання для самостійної роботи**

1. З представленої вибірки створити вибірку даних, що містить 40% пацієнтів, що належать лікарням

BINER.

HILLVIEW

ST\_AND

2. Створити дві випадкові вибірки , що містять по 20% (30, 40 %) пацієнтів загальної вибірки.

**5.3 Лабораторна робота № 3. Порівняння виживаності у двох і більше групах**

*Приклад виконання*

Дані про виживаність пацієнтів після операції по трансплантації серця містяться у файлі *Heart.sta (файл з даними з попередньої лабораторної роботи)*. Відкрийте цей файл даних за допомогою меню Файл – Открыть (File - Open).

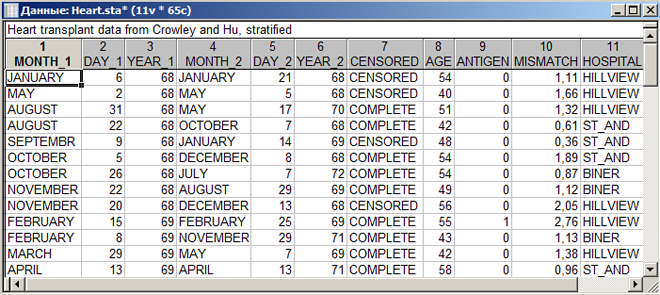


Рис. 31

Перші шість перемінних у цій множині даних – дати, точніше, дати трансплантації серця (у послідовності: місяць – день - рік), і дати, коли відповідний пацієнт або помер, або був вилучений з-під нагляду (іншими словами, цензурований, наприклад, з пацієнтом був утрачений зв’язок).

Перемінна *Цензурировано – Censored* є індикатором цензурування з кодом, який показує, чи є відповідний нагляд завершеним або цензурованим (*0 - завершено*; *1 - цензуровано*). Перемінная *Больница – Hospital* являє собою фіктивну перемінну групування, яка відображає, до якої з трьох лікарень відноситься пацієнт.

Замість записів даних у шести перемінних (місяць, день і рік початку нагляду і місяць, день і рік закінчення нагляду) ми можемо ввести тривалість життя за допомогою однієї перемінної або ввести дати початку та закінчення як значення двох різних перемінних.

У цьому досліджені представлені три лікарні: *Hillview* (код *1*), *St\_AND* (код *2*), і *Biner* (код*3*). На основі порівняння функцій виживання у цих трьох лікарнях з використанням деяких критеріїв х двома виборами з модулю *Анализ выживаемости (Survival)* можна зробити висновки як про виживаність пацієнтів, так і о можливості цих тестів.

**Специфікація аналізу.**

Відкрийте файл даних *Heart.sta*.

Потім потрібно обрати *Анализ выживаемости (Survival)* з меню *Анализ – Углубленные методы анализа* (*Statistics – Advanced models*) для відображення стартової панелі модуля *Анализ выживаемости (Survival and Failure Time Analysis),* двічі потрібно клацнути на рядку *Сравнение нескольких выборок (Comparing multiple samples)*, щоб відкрити діалогове вікно *Сравнение выживаемости в нескольких группах (Comparing survival in multiple groups)*.

Натисніть кнопку *Переменные (Variables (survival times, censoring indicator, grouping variable)*, щоб з’явилося стандартне вікно *Выберите переменные (Select survival times/dates, cens. indicator, grouping variable)*. Оберіть перші 6 перемінних як *Времен жизни (1, 2 или 6) (Survival)*. *STATISTICA* буде інтерпретувати першу і четверту перемінні у цьому списку як місяці, другу і п’яту - як дні, а третю і шосту - як рік.

Далі необхідно визначити перемінну *Censored* как *Перемінну* ц*ензурування (Cansoriring var.)* і перемінну *Hospital* как *Перемінну групування (Grouping variable)*.

Натисніть кнопку *OK* для повернення в діалоговое вікно *Сравнение выживаемости в нескольких группах (Comparing survival in multiple groups)*. Двічі клацніть по полю вводу *Код для полных наблюдений (Code for complete responses)*, щоб з’явилося діалогове вікно Переменная 7 (Variable 7). Тут потрібно обрати *Complete* і натиснути *OK*. Потім двічі клацнути по полю *Код для цензурированных наблюдений (Code for censored responses)* і обрати *Censored*.

Щоб ввести коди для перемінної групування, натисніть кнопку *Коды (для групп) (Codes (for groups)* і потім в діалоговому вікні спочатку натиснути кнопку *Все (All)* (щоб відобразити усі коди), а потім клацнути *OK* (щоб ввести ці коди і завершити цей діалог).

Діалогово вікно *Сравнение выживаемости в нескольких группах (Comparing survival in multiple groups)* тепер виглядає наступним чином:

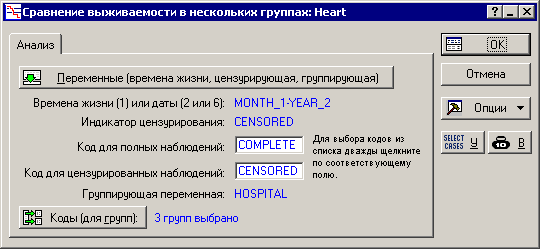


Рис. 32

Тепер, щоб продовжити, достатньо клацнути *OK*, і відчиниться діалогове вікно *Результаты сравнения выживаемости в нескольких группах (Comparing survival in multiple groups, Results)*.

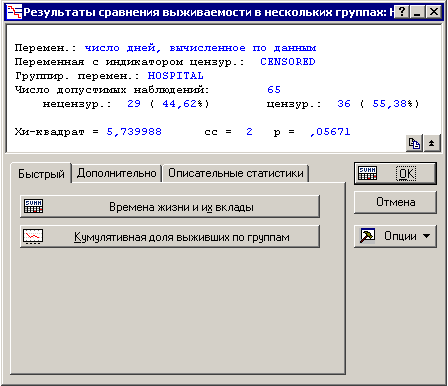


Рис. 33

Критерій *хі-квадрат* дає майже значуще відхилення у цьому прикладі (*p*<0,06). Тому ми можемо зробити попередній висновок, що є певні відмінності між лікарнями.

Щоб побачити ці відмінності, ми можемо побудувати графіки функций виживаності для різних груп, для чого достатньо натиснути кнопку График функции выживаемости по группам (Cumulative proportion surviving (Kaplan-Meier) by group) і отримаємо графік – *Кумулятивная доля выживших (*Cumulative proportion surviving (Kaplan-Meier*):*

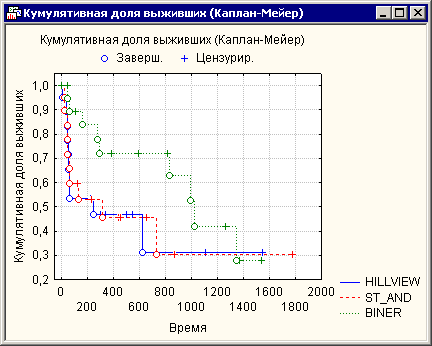


Рис. 34

Можна побачити, що функція виживаності у лікарні міста *Бінер* зменшується повільніше на початковому відрізку часу у порівняння з іншими лікарнями.

Звідси ми можемо зробити висновок, що з якоїсь причини пацієнти з трансплантованим серцем у лікарні *Бінера* мають більше шансів вижити, зокрема протягом критичних перших 100 днів після трансплантації.

Поверніться до діалогового вікна *Результаты сравнения выживаемости в нескольких группах (Comparing survival in multiple groups, Results)* і натисніть на вкладку *Дополнительно (Advanced)*. Щоб продивитися таблиці часу життя для цих трьох лікарень, натисніть кнопку *Процент выживших по группам (percent surviving by group)*. З’являться окремі електронні таблиці для кожної лікарні.

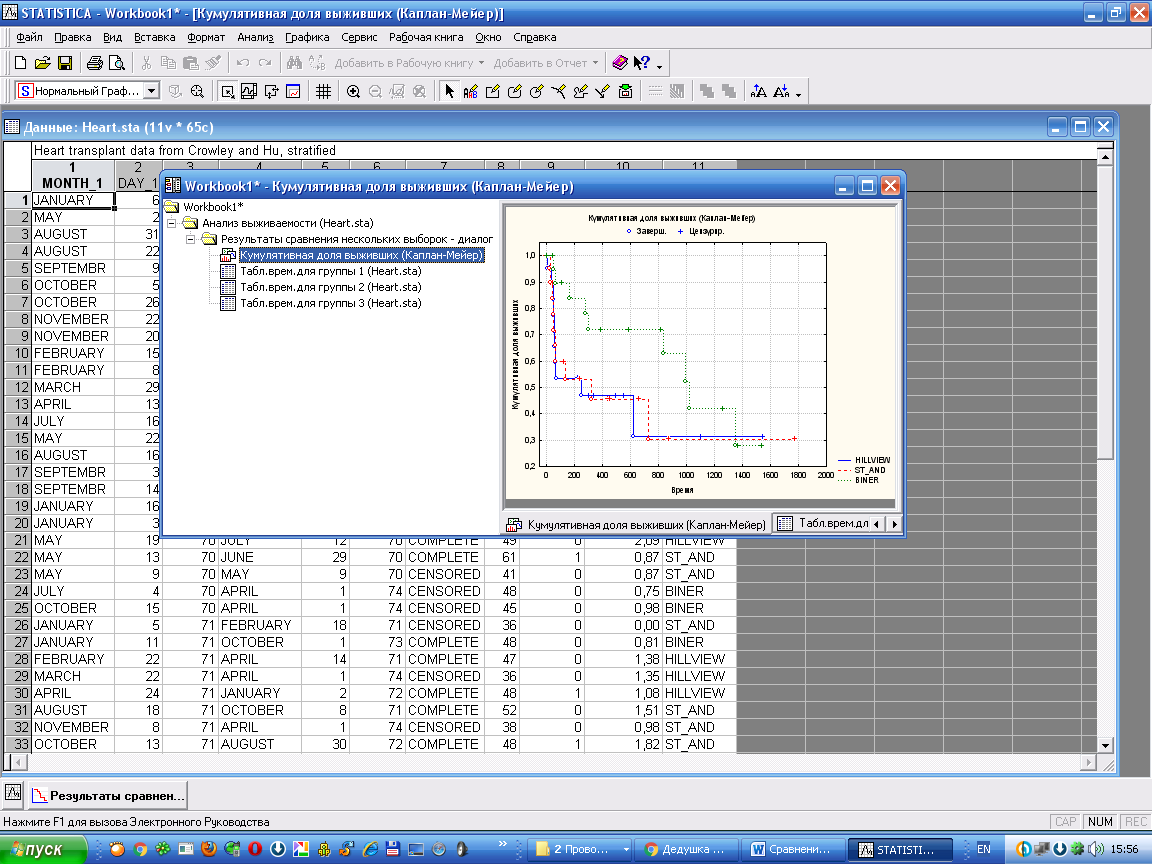


Рис. 35

Дані таблиці підтверджують приведені вище висновки.

**Порівняння виживаності у двох групах.**

Модуль *Анализ выживаемости (Survival)* містить різні методи для порівняння двох вибірок.

Натиснути кнопку *Отмена (Cancel)* в діалоговому вікні *Результаты сравнения выживаемости в нескольких группах (Comparing survival in multiple groups, Results)* і в діалоговому вікні *Сравнение выживаемости в нескольких группах (Comparing survival in multiple groups)* для повернення на стартову панель.

Потрібно обрати опцію *Сравнение двух выборок (Comparing two samles)* на стартовій панелі, щоб з’явилося вікно Сравнение выживаемости в двух группах (Comparing survival in two groups). Натисніть кнопку *Переменные (Variables (survival times, censoring indicator, grouping variable)* для того, аби вказати всі перемінні, як і раніше (див. вище). Також укажіть *Код для полных наблюдений (Code for complete responses)* і *Код для цензурированных наблюдений (Code for censored responses)*, як і раніше.

Щоб порівняти групи, двічі клацніть на полі ввода *Код для первой группы (Code for first group)* для відображення діалогового вікна Переменная 1 (Variable 1). Оберіть тут першу групу для порівняння (в нашому прикладі обрана *Hillview*) і натисніть *OK*.

Потім двічі клацніть на полі вводу *Код для второй группы (Code for second group)* і оберіть *Biner*. Діалоговое вікно *Сравнение выживаемости в двух группах (Comparing two samles)* тепер виглядяє наступним чином:

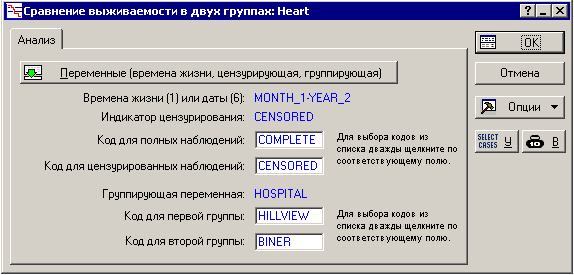


Рис. 36

Тепер нажмите кнопку *OK*, чтобы перейти до діалогового вікна *Двовибіркові критерії (Two samles tests results)*.

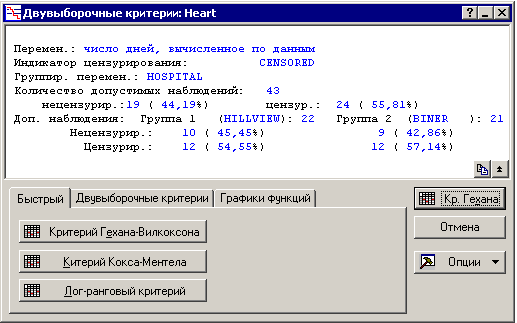


Рис. 37

У цьому прикладі деякі з критеріїв демонструють рівень значимості *p* біля 0,05, в той час як інші – незначні.

Тому з деякою долею впевненості ми можемо підсумувати, що виживаність пацієнтів з трансплантованим серцем у цих двох лікарнях значно відрізняються, при тому в лікарні *Бінера* виживаність вище.

**5.4. Лабораторна робота № 4. Перевірка розподілу на нормальність**

*Приклад виконання*

Нехай вихідними будуть дані 10-й, 50-й і 95-й процентилі біпаріетального і лобно-потиличного розміру в залежності від терміну вагітності, що відображені у таблиці:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Строк вагітності | Біпаріетальний розмір мм | | | Лобно-потиличний розмір, мм | | |
| 10 | 50 | 95 | 10 | 50 | 95 |
| 11 тижнів | 13 | 17 | 21 | - | - | - |
| 12 тижнів | 18 | 21 | 24 | - | - | - |
| 13 тижнів | 20 | 24 | 28 | - | - | - |
| 14 тижнів | 23 | 27 | 31 | - | - | - |
| 15 тижнів | - | 31 | - | - | - | - |
| 16 тижнів | 31 | 34 | 37 | 41 | 45 | 49 |
| 17 тижнів | 34 | 38 | 42 | 46 | 50 | 54 |
| 18 тижнів | 37 | 42 | 47 | 49 | 54 | 59 |
| 19 тижнів | 41 | 45 | 49 | 53 | 58 | 63 |
| 20 тижнів | 43 | 48 | 53 | 56 | 62 | 68 |
| 21 тижнів | 46 | 51 | 56 | 60 | 66 | 72 |
| 22 тижнів | 48 | 54 | 60 | 64 | 70 | 76 |
| 23 тижнів | 52 | 58 | 64 | 67 | 74 | 81 |
| 24 тижнів | 55 | 61 | 67 | 71 | 78 | 85 |
| 25 тижнів | 58 | 64 | 70 | 73 | 81 | 89 |
| 26 тижнів | 61 | 67 | 73 | 77 | 85 | 93 |
| 27 тижнів | 64 | 70 | 76 | 80 | 88 | 96 |
| 28 тижнів | 67 | 73 | 79 | 83 | 91 | 99 |
| 29 тижнів | 70 | 76 | 82 | 86 | 94 | 102 |
| 30 тижнів | 71 | 78 | 85 | 89 | 97 | 105 |
| 31 тижнів | 73 | 80 | 87 | 93 | 101 | 109 |
| 32 тижнів | 75 | 82 | 89 | 95 | 104 | 113 |
| 33 тижнів | 77 | 84 | 91 | 98 | 107 | 116 |
| 34 тижнів | 79 | 86 | 93 | 101 | 110 | 119 |
| 35 тижнів | 81 | 88 | 95 | 103 | 112 | 121 |
| 36 тижнів | 83 | 90 | 97 | 104 | 114 | 124 |
| 37 тижнів | 85 | 92 | 98 | 106 | 116 | 126 |
| 38 тижнів | 86 | 94 | 100 | 108 | 118 | 128 |
| 39 тижнів | 88 | 95 | 102 | 109 | 119 | 129 |
| 40 тижнів | 89 | 96 | 103 | 110 | 120 | 130 |

Занесемо вихідну таблицю в програму *STATISTICA*. *Вид даних в програмі зображено на рис. 38.*

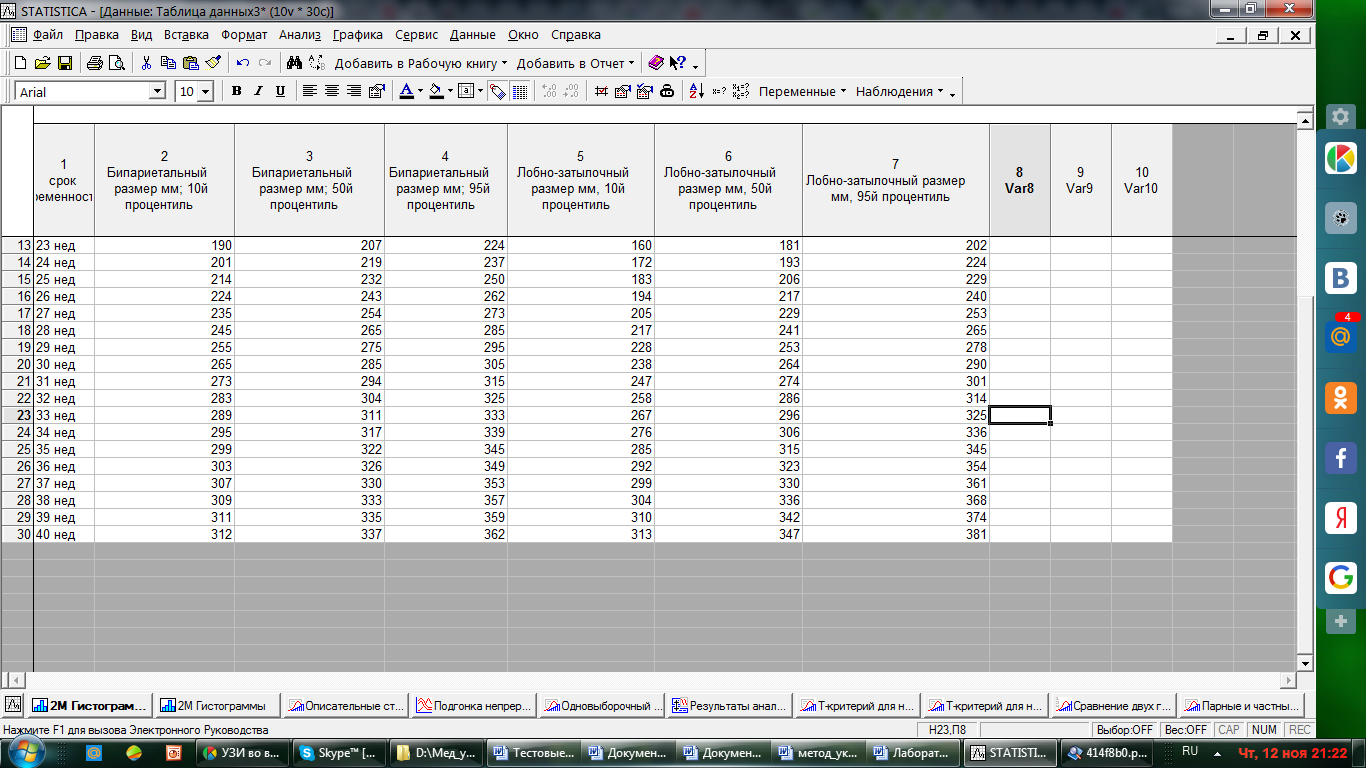


Рис. 38

**Перевірка гіпотези про нормальність візуальними методами**

Побудуємо гістограму за перемінною «Біпаріетальний розмір мм, 10-й процентіль». Для цього потрібно обрати з меню пункт Графика (Graphs)/Гистограммы (Histogram).

На вкладці Дополнительно (Advanced) вкажемо: Распределение (Fit type) = Нормальное (Normal), количество категорий (Categories) – 10, обираємо перемінну – «Біпаріетальный розмір мм, 10-й процентіль».

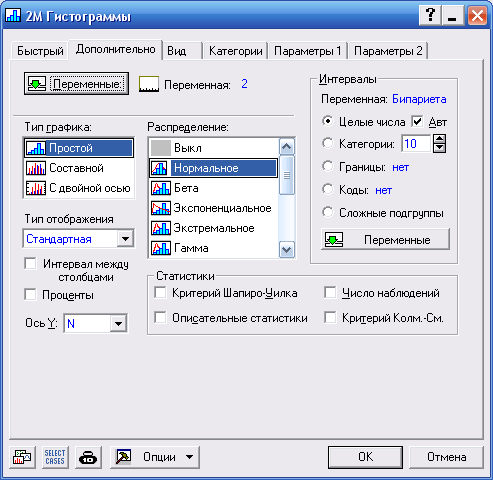


Рис. 39

Результат буде виглядати наступним чином:



Рис. 40

По цьому графіку зробити висновок у такому випадку доволі складно.

Побудуємо нормальний імовірнісний графік. Для цього потрібно обрати вкладки: *Графика (Graphs) – Графики исходных данных (Input data graphs) – Вероятностный график (Probability plot AGE) – Нормальный вероятностный (Normal probability)*.

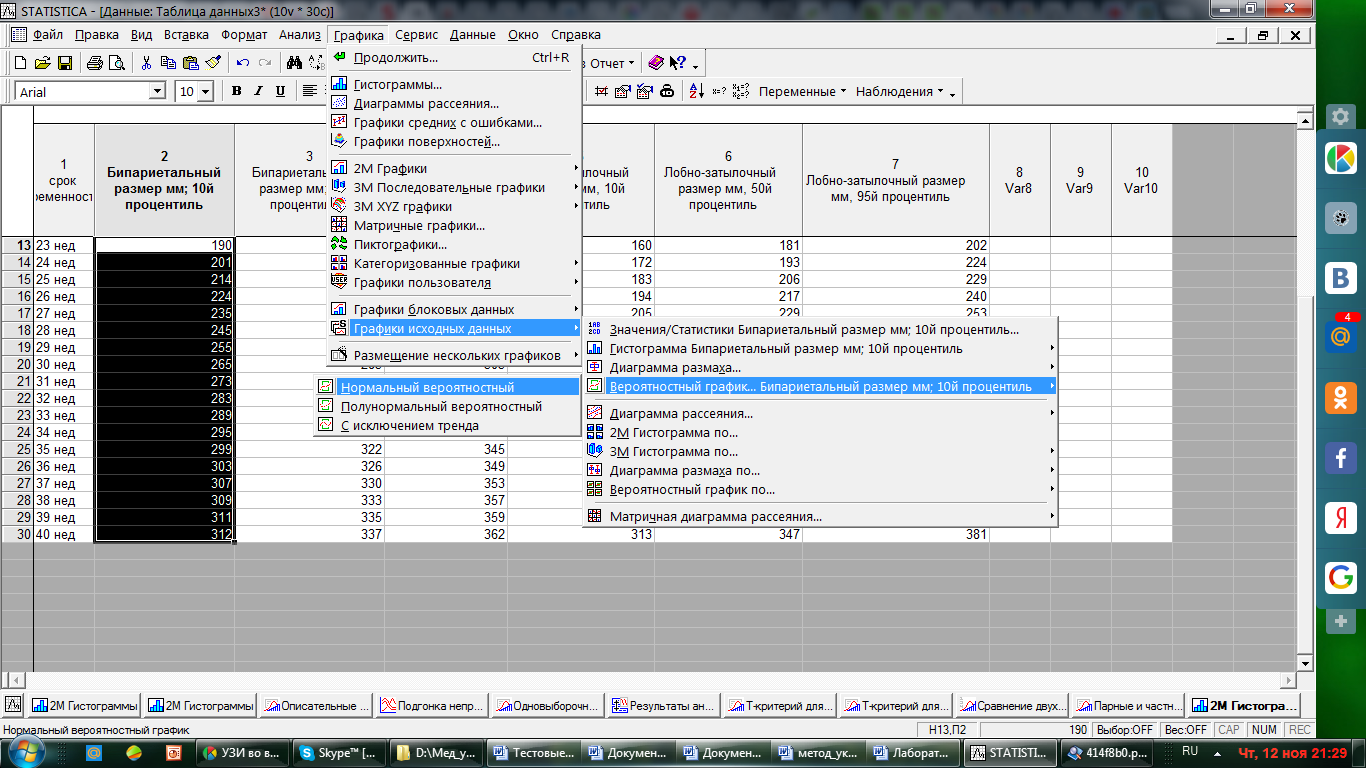


Рис. 41

Результат буде виглядати наступним чином:

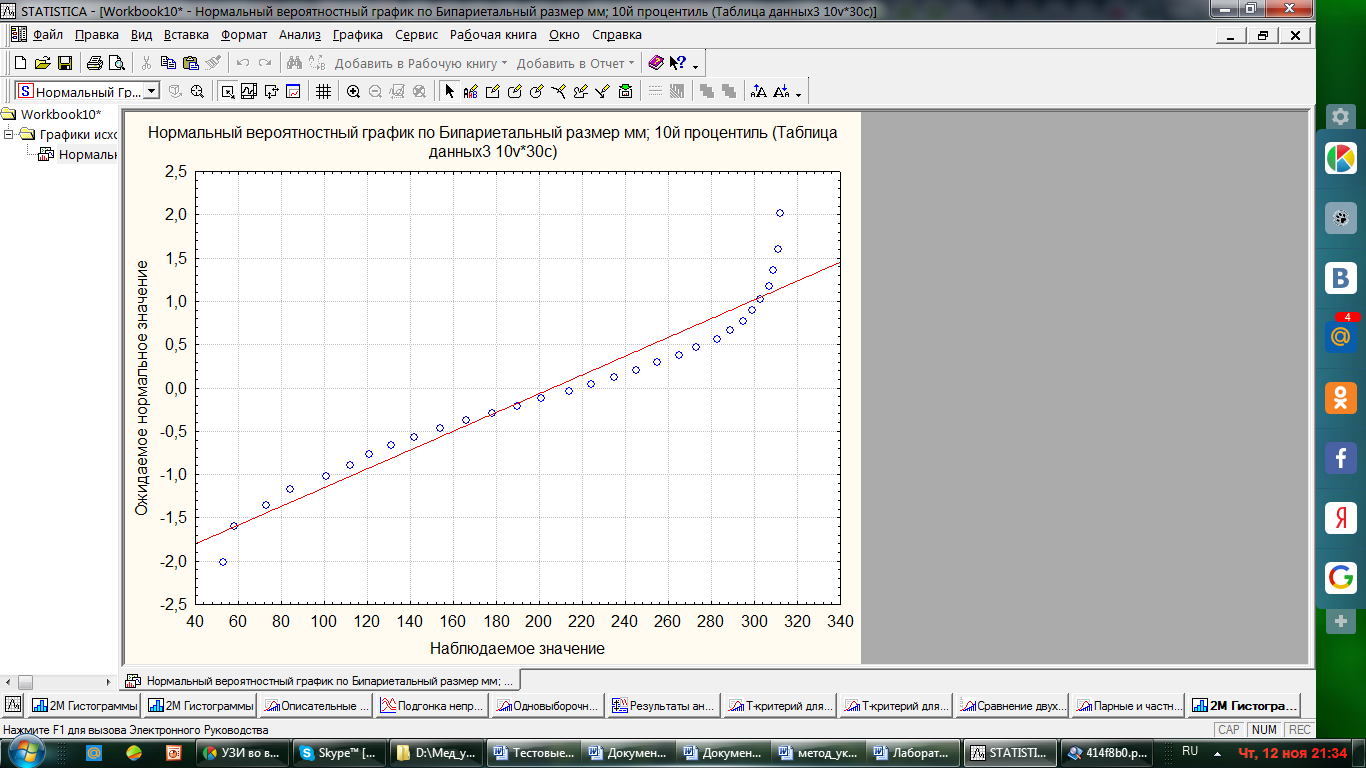


Рис. 42

Якщо значення, що спостерігаються, (відкладаємо їх по осі X) були б розподілені нормально, то всі значення на графіку повинні були попасти на прямую лінію. Однак цього не спостерігається.

Тепер обчислимо деяку описову статистику для перемінної «Біпаріетальний розмір, мм, 10-й процентіль».

На вкладці Дополнительно (Advanced) діалогу Описательные статистики (Descriptive statistics) поставимо гачечки в полях Асимметрия (Skewness),стандартная ошибка асимметрии (Std. err., Skewness), Эксцесс (Kurtosis), Стандартная ошибка эксцесса (Std. err., Kurtosis).

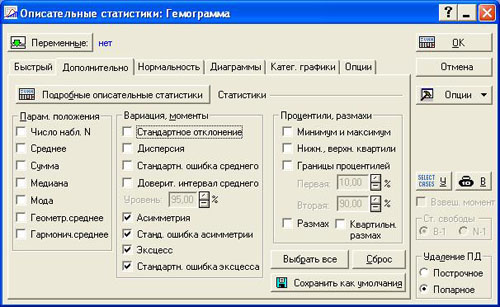


Рис. 43

Асиметрія - це показник симетричності / скошеності кривої розподілу, а ексцес визначає гостроту його вершини. При лівосторонній асиметрії її показник є позитивним і в розподілі переважають більш низькі значення ознаки. При правосторонній - показник позитивний і переважають більш високі значення. У всіх симетричних розподілах (в тому числі і у нормального розподілі) величина асиметрії дорівнює нулю.

Якщо в розподілі переважають значення близькі до середнього арифметичного, то формується розподіл з гострим верхом. В цьому випадку показник ексцесу прагне до позитивної величини. У нормального розподілу ексцес дорівнює нулю. Якщо у розподілу 2 вершини (бімодальне розподіл), то тоді ексцес прагне до негативної величини.

Розподіл оцінюється як імовірно близький до нормального, якщо встановлено, що від 50 до 80% всіх значень розташовуються в межах одного стандартного відхилення від середнього арифметичного, і коефіцієнт ексцесу по абсолютній величині не перевищує значення рівного 2.

Розподіл вважається достовірно нормальним якщо абсолютна величина показників асиметрії і ексцесу менше їх помилок репрезентативності в 3 і більше разів.

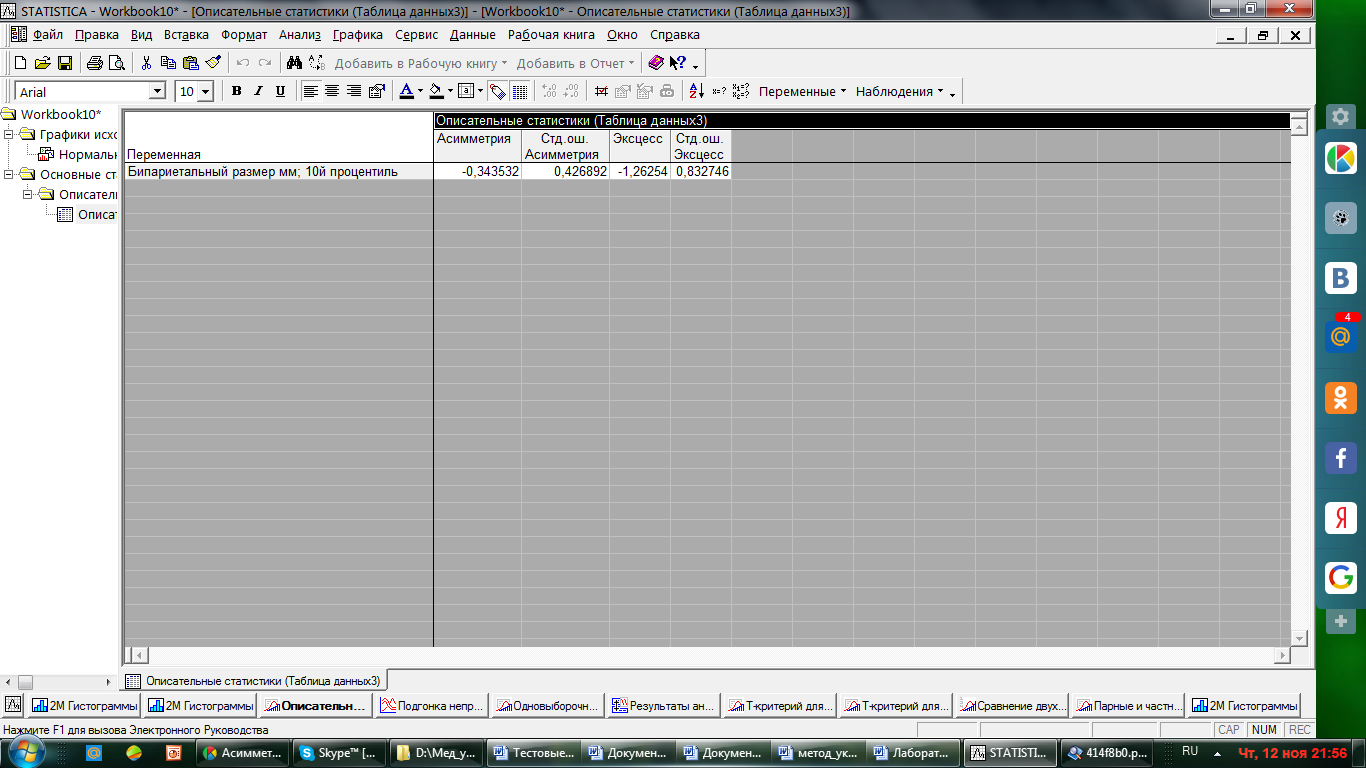


Рис. 44

Судячи зі значення ***Асиметрії*** (рис. 44), розподіл перемінної можна вважати нормальним (0 «майже» є в інтервалі ***Асиметрія ±Стандартна помилка Асиметрії***).

Але судячи по значенню ***Ексцесу***, гіпотезу про нормальність слід відкинути. Як правило, якщо знайдена хоча б одна істотна «невідповідність», гіпотезу сміливо відхиляють, у той час як відповідність навість усім відомим критеріям ще не говорить про справедливість гіпотези.

Таким чином, остаточний висновок: Розподіл перемінної ***«***Біпаріетальний розмір, мм, 10-й процентиль» істотно відрізняється від нормального розподілу

**Завдання для самостійної роботи:**

перевірити нормальність розподілу по рядам даних:

* Лобно-потиличний розмір, мм, 10-й процентиль;
* Лобно-потиличний розмір, мм, 50-й процентиль;
* Лобно-потиличний розмір, мм, 95-й процентиль.

**5.5. Лабораторна робота № 5. Порівняння вибірок**

*Стислі теоретичні відомості*

Найбільш часто зустрічається і досить складним математико-статистичним завданням є порівняння даних, отриманих в процесі ***спостережень*** або ***експериментів, в вибіркових сукупностях.*** Дослідник намагається описати результати спостереження кількісними методами і «на виході» отримує числовий масив тих чи інших доступних йому вимірювань - варіаційний ряд. Однак, як правило, змістовна інформація, міститься в результатах вимірювань, має набагато більшу цінність при порівнянні її з аналогічною інформацією, але отриманої деяким іншим чином. Наприклад, це може бути ситуація порівняння дослідних даних (коли ми якось вплинули на досліджуваний об'єкт або явище) з ***контрольною групою, в якій ніякого впливу на об’єкт спостережень не було.*** Можливо і порівняння двох варіантів дослідів. Можливо порівняння двох серій спостережень, розділених в просторі і часі і т.п.

Припустимо, що вдається помітити будь-які чисельні відмінності в характеристиках порівнюваних рядів. Насамперед виникає питання: наскільки ймовірним є, що ці відмінності невипадкові і будуть систематично повторюватися в подальшому при відтворенні умов експерименту або спостережень, тобто виявлені відмінності є **статистично значущими.**

**Вибір відповідного методу порівняння вибіркових сукупностей визначається декількома факторами: характером порівнюваних ознак (якісні або кількісні), числом груп, що порівнюються, залежністю або незалежністю вибірок, а також видом розподілу за ознакою.**

Вибірки є **незалежними**, якщо набір об'єктів дослідження в кожній з груп здійснювався незалежно від того, які об'єкти дослідження включені в іншу групу. Так, зокрема, відбувається при рандомізації, коли розподіл об'єктів відбувається випадковим чином. Прикладом порівняння незалежних вибірок може служити зіставлення даних аналізу крові в групі пацієнтів з аналогічними показниками в групі здорових.

Групи є **залежними (пов'язаними)** в динамічних дослідженнях, коли вивчаються одні й ті ж об'єкти в різні моменти часу. Наприклад, показники аналізу крові у одних і тих же пацієнтів до і після лікування.

Від виду розподілу і типу досліджуваної ознаки залежить вибір відповідного математико-статистичного критерію. Критерії діляться на два типи - **параметричні** і **непараметричні.**

**Параметричні критерії** – критерії, засновані на оцінці параметрів розподілу, до яких відносяться середнє арифметичне, середнє квадратичне відхилення, дисперсія. **Вони застосовні тільки в тому випадку, якщо чисельні дані підкоряються нормальному розподілу.** Якщо розподіл відрізняється від нормального, то слід користуватися так званими непараметричними критеріями.

**Непараметричні критерії** не засновані на оцінці параметрів розподілу і взагалі **не вимагають, щоб дані підпорядковувалися якомусь певному типу розподілу.** Непараметричні критерії дають більш грубі оцінки, ніж параметричні, але є більш універсальними. А параметричні методи більш точні, але лише в разі, якщо правильно визначено розподіл сукупності.

Перед тим як перейти до розгляду статистичних критеріїв, введемо поняття **нульовий** та **альтернативної гіпотез**, які нам будуть потрібні надалі.

На кожному кроці процесу аналізу даних висуваються **дві гіпотези.** Перша позначається  і називається **нульовою гіпотезою**. Друга гіпотеза позначається  і носить назву **альтернативної**, тобто протилежної за смислом. Під «нульовою гіпотезою» розуміють припущення про відсутність тої чи іншої цікавої досліднику події, яви або ефекту, а під «альтернативною» – про її наявність. Обидві гіпотези, як би вони не були сформульовані, обов’язково повинні мати **взаємовиключний** зміст.

Нульова гіпотеза **не може бути відкинута**, якщо її ймовірність буде вищою за який-небудь заздалегідь заданий рівень α, достатньо близький до 0, тобто . Ця величина α носить назву **рівня значимості нульової гіпотези**.

Альтернативна гіпотеза **може буть прийнята** лиши в том випадку, якщо її ймовірність досягне будь-якого наперед завданого рівня β або перевершить його, тобто . Ця величина β – **рівень довірчої ймовірності**. Він відповідає «рівням безпомилкових прогнозів», тобто ймовірностям 0,95, 0,99 і 0,999 (**область практичних достовірних подій**). Відповідно, α окреслює область **практично неможливих подій** з порогами ймовірності 0,05, 0,01 і 0,001.

Оскільки  і  – альтернативні гіпотези, то їх сумарна ймовірність рівняється одиниці. Отже, зростання ймовірності однієї з гіпотез автоматично призводить до зниження ймовірності іншої. Наприклад, якщо , це значить, що буде виконуватися умова . І в цьому випадку нульова гіпотеза може бути підкинута як подія майже неможлива, а альтернативна повинна бути прийнята як подія майже достовірна. Якщо ж , то . І у цій ситуації нульова гіпотеза не може бути відкинута, а альтернативна не може бути прийнята.

Наприклад, в процесі дослідження ставиться задача довести наявність статистично значущих відмінностей між результатами спостережень у дослідній і контрольній групах. Це означає, що дані, отримані при застосуванні того чи іншого статистичного критерію, повинні дозволити відкинути нульову гіпотезу про відсутність зазначених відмінностей.

*Приклад виконання*

Нехай вихідними будуть дані 10-й, 50-й і 95-й процентилі біпаріетального і лобно-потиличного розміру в залежності від терміну вагітності, які були використані в попередній лабораторній роботі.

**Програмна реалізація методу**

З точки зору прикладної статистики, задача зводиться до порівняння значень перемінних «Біпаріетальний розмір, 10-й процентиль» і «Біпаріетальний розмір, 95-й процентиль» (співставлення залежних вибірок).

Як ми вже дізналися, розподіл відрізняється від нормального, тому ми будемо користуватися методами непараметричної статистики

**Крок 1.** Запустимо модуль Непараметрическая статистика (Nonparametrics) (рис. 45).

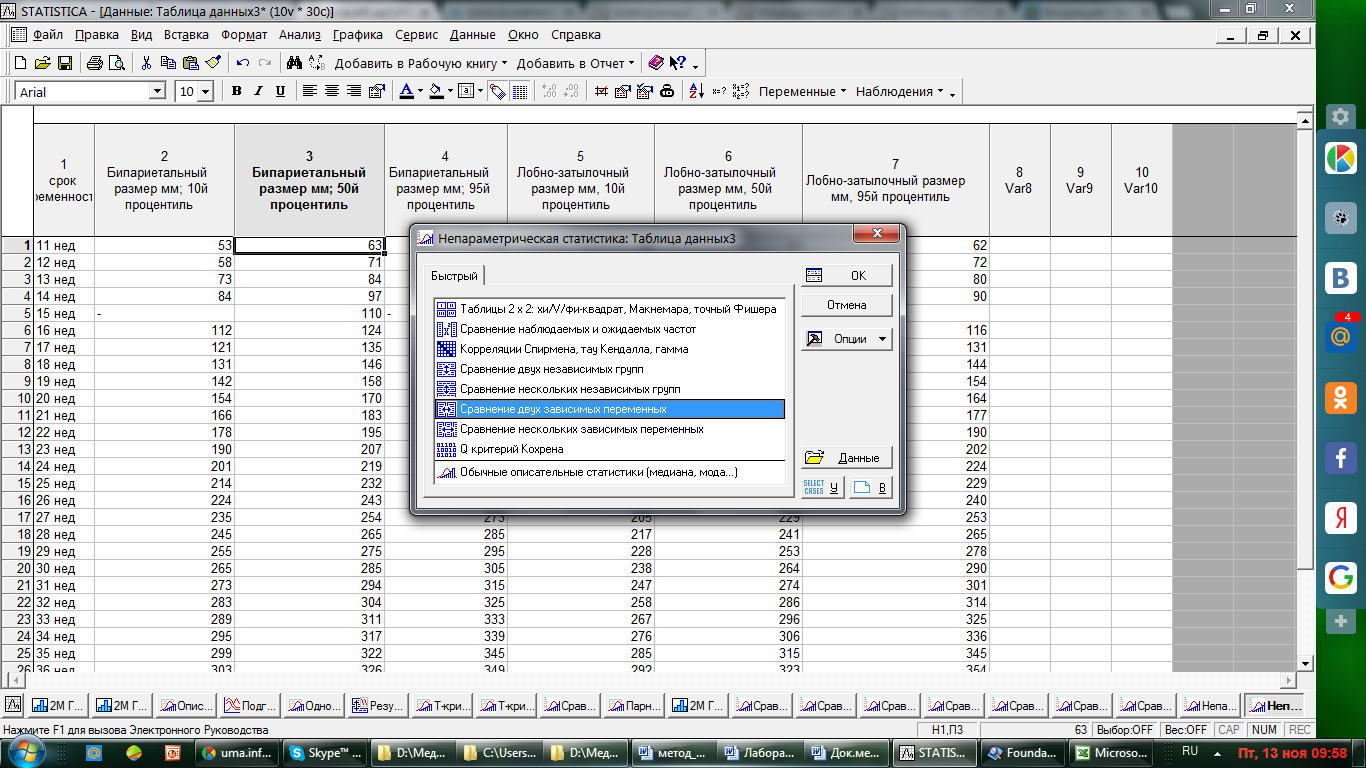


Рис. 45

Оберемо пункт ***Сравнение двух зависимых переменных (Comparing two independent samples (groups))***

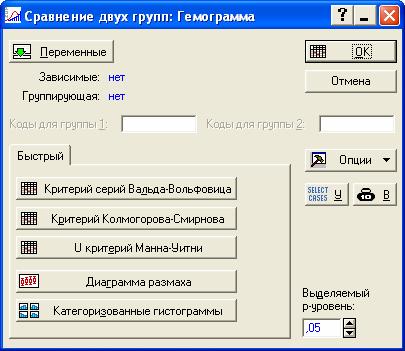


Рис. 46

**Крок 2.** Позначимо перемінні (рис. 47)

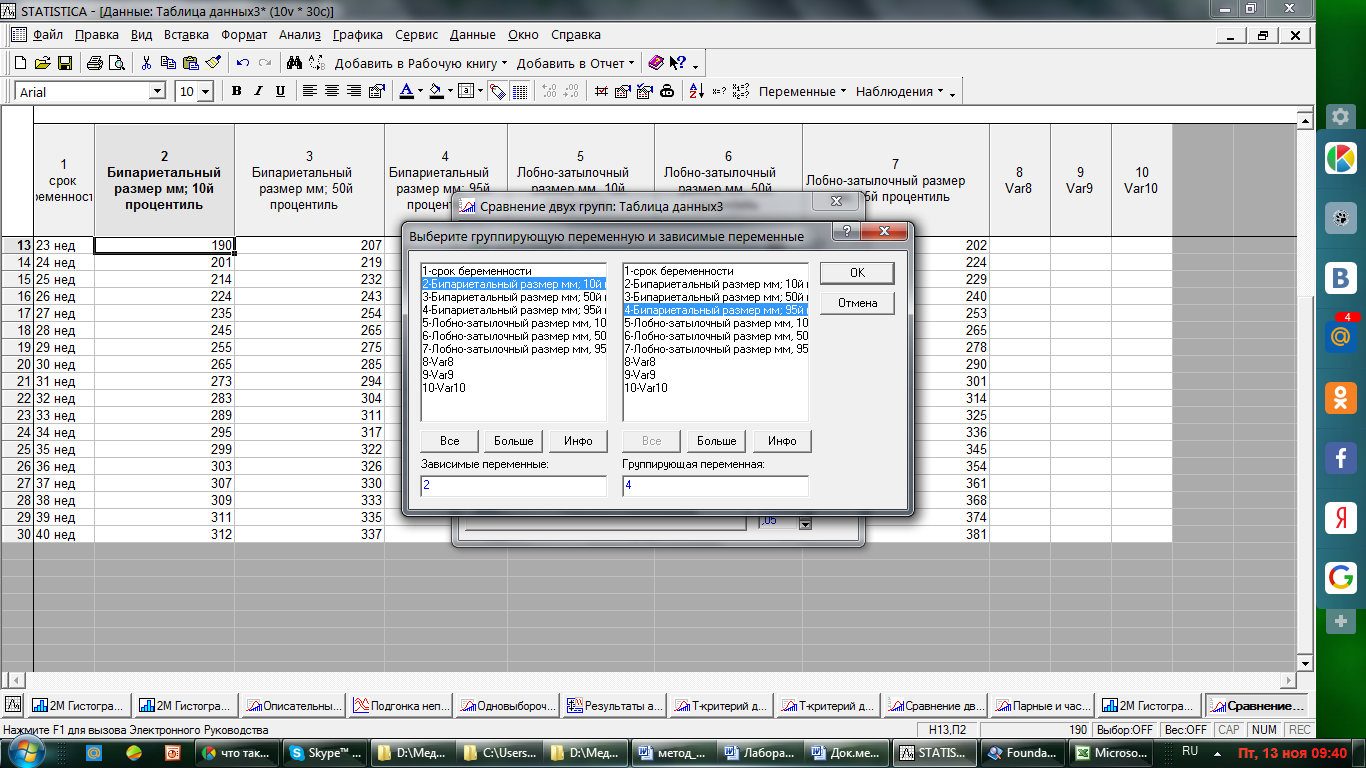


Рис. 47

Зверніть увагу: коди для перемінних (2 і 4) автоматично з’являться у відповідних вікнах.

**Крок 3.** Натиснемо на кнопку критерий Вилкоксона (Wilcoxon matched pair test) (рис. 48).

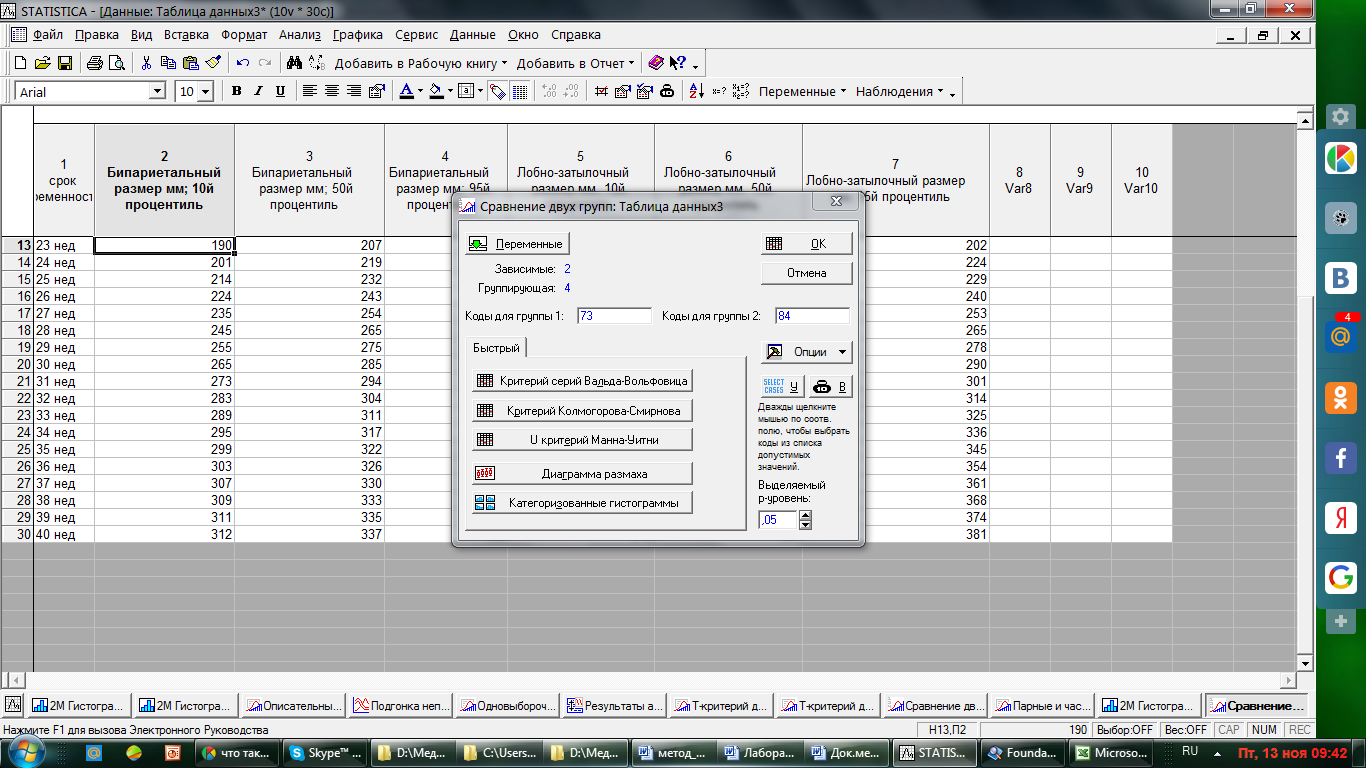


Рис. 48

Отримаємо наступний результат:

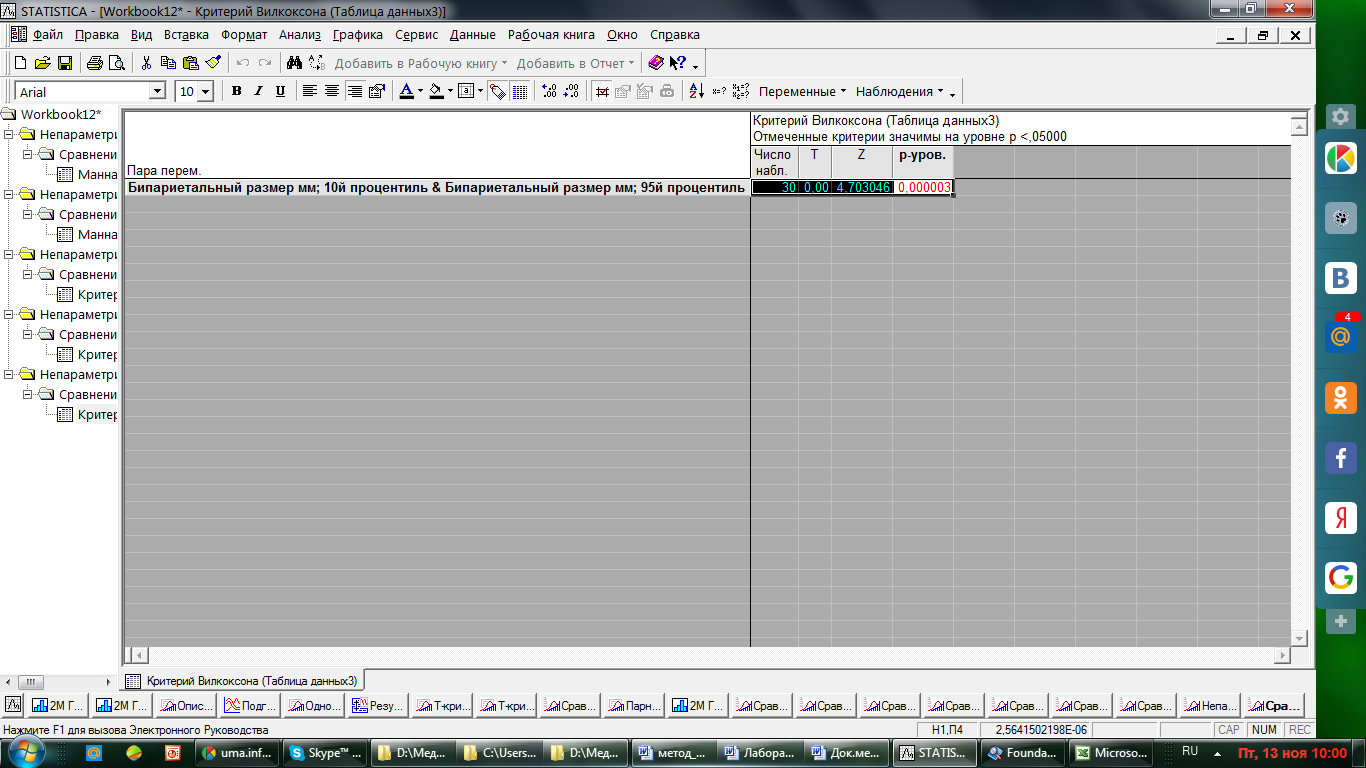


Рис. 49

Зверніть увагу на p-уровень (p-value): 0,000003. Гіпотезу про рівність функцій розподілу можна відхилити, різницю в значеннях можна вважати істотною.

**Завдання для самостійної роботи:**

Порівняти данні по вибіркам:

- Лобно- потиличний розмір 10-й процентиль і 50-й процентиль;

- Лобно-потиличний розмір 50-й процентиль і 95-й процентиль;

- Лобно-потиличний розмір 10-й процентиль і 95-й процентиль;

**тестові завдання для досягнення конкретних цілей навчання**

**1 ДОВІРЧІ МЕЖІ - ЦЕ:**

1) імовірнісна оцінка можливих відхилень, в межах яких може коливатися шукана середня величина ознаки при повторних дослідженнях

2) різниця між максимальною і мінімальною варіантами

3) різниця між середнім квадратичним відхиленням і помилкою репрезентативності

4) відношення середньої величини до середнього квадратичного відхилення

**2 ПОДІЯ В МЕДИКО - БІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ ВВАЖАЄТЬСЯ НЕ ДОСТОВІРНОЮ, ЯКЩО ЙОГО ВІРОГІДНІСТЬ РІВНА:**

1) 68,3%

2) 95,5%

3) 99,7%

4) 50,0%

**3 КРИТЕРІЙ ВІДПОВІДНОСТІ** **ВИКОРИСТОВУЄТЬСЯ У ВИПАДКУ:**

1) у разі зіставлення неоднорідних совокупностей

2) коли немає необхідності знати середні або відносні величини для оцінки достовірності відмінностей варіаційних рядів

3) коли необхідно оцінити міру впливу чинників на результуючу ознаку

4) пошуку взаємозв'язку між ознаками

**4 МЕТОД В СТАТИСТИЧНІЙ МАТЕМАТИЦІ, СПРЯМОВАНИЙ НА ПОШУК ЗАЛЕЖНОСТЕЙ В ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ ШЛЯХОМ ДОСЛІДЖЕННЯ ЗНАЧУЩОСТІ ВІДМІННОСТЕЙ ВАРІАБЕЛЬНОЇ ОЗНАКИ В ДОСЛІДЖУВАНІЙ СУКУПНОСТІ НАЗИВАЄТЬСЯ:**

1) кореляційний аналіз

2) дисперсійний аналіз

3) метод стандартизації

4) метод вирівнювання динамічних рядів

**5 У ОСНОВІ ДИСПЕРСІЙНОГО АНАЛІЗУ ЛЕЖИТЬ ПРИПУЩЕННЯ, ЩО:**

1) порівнювані статистичні сукупності непорівнянні (неоднорідні)

2) є взаємна зміна значень варіант двох варіаційних рядів

3) існує функціональна залежність двох ознак між собою

4) одні змінні можуть розглядатися як причини (чинники, незалежні змінні): f1, ., fk, а інші як наслідки (залежні змінні)

**6 СУТЬ ДИСПЕРСІЙНОГО АНАЛІЗУ ПОЛЯГАЄ В:**

1) розчленовуванні загальної дисперсії (D або SS) ознаки, що вивчається, на окремі компоненты, обумовлені впливом конкретних чинників, і перевірці гіпотез про значущість впливу цих чинників на досліджувану ознаку

2) вирівнюванні динамічного ряду за допомогою ковзаючої середньої

3) виявлення функціональної залежності двох ознак між собою

4) обчисленні загальних стандартизованных показників в порівнюваних совокупностях, які урівноважені завдяки вибраному стандарту

**7 У ДИСПЕРСІЙНОМУ АНАЛІЗІ ЧИННИКОМ НАЗИВАЄТЬСЯ:**

1) елементарна властивість об'єктів, що вивчається як результат впливу чинників

2) вплив, дія або стан, який відбивається на розмірах і різноманітності результативної ознаки

3) міра дії чинника (у тому числі нульове в контрольній групі) або стан об'єктів вивчення (пів, вікова група та ін.)

4) сукупність градацій даних (груп об'єктів спостереження, вибраних дослідником), що вивчаються, з вичисленими значеннями відносних або середніх величин по кожній градації

**8 У ДИСПЕРСІЙНОМУ АНАЛІЗІ РЕЗУЛЬТАТИВНОЮ ОЗНАКОЮ НАЗИВАЄТЬСЯ:**

1) елементарна властивість об'єктів, що вивчається як результат впливу чинників

2) вплив, дія або стан, який відбивається на розмірах і різноманітності результативної ознаки

3) міра дії чинника (у тому числі нульове в контрольній групі) або стан об'єктів вивчення (пів, вікова група та ін.)

4) сукупність градацій даних (груп об'єктів спостереження, вибраних дослідником), що вивчаються, з вичисленими значеннями відносних або середніх величин по кожній градації

**9 У ДИСПЕРСІЙНОМУ АНАЛІЗІ ГРАДАЦІЄЮ ЧИННИКА НАЗИВАЄТЬСЯ:**

1) елементарна властивість об'єктів, що вивчається як результат впливу чинників

2) вплив, дія або стан, який відбивається на розмірах і різноманітності результативної ознаки

3) міра дії чинника (у тому числі нульове в контрольній групі) або стан об'єктів вивчення (пів, вікова група та ін.)

4) сукупність градацій даних (груп об'єктів спостереження, вибраних дослідником), що вивчаються, з вичисленими значеннями відносних або середніх величин по кожній градації

**10 У ДИСПЕРСІЙНОМУ АНАЛІЗІ ФАКТОРІАЛЬНИМ ВПЛИВОМ НАЗИВАЄТЬСЯ:**

1) дія тих чинників, які не враховані в дисперсійному комплексі і складають загальний фон, на якому діють чинники, що враховуються

2) вплив, дія або стан, який відбивається на розмірах і різноманітності результативної ознаки

3) міра дії чинника (у тому числі нульове в контрольній групі) або стан об'єктів вивчення (пів, вікова група та ін.)

4) простий або комбінований статистичний вплив чинників, що вивчаються (врахованих)

**11 У ДИСПЕРСІЙНОМУ АНАЛІЗІ ВИПАДКОВИМ ВПЛИВОМ НАЗИВАЄТЬСЯ:**

1) дія тих чинників, які не враховані в дисперсійному комплексі і складають загальний фон, на якому діють чинники, що враховуються

2) вплив, дія або стан, який відбивається на розмірах і різноманітності результативної ознаки

3) міра дії чинника (у тому числі нульове в контрольній групі) або стан об'єктів вивчення (пів, вікова група та ін.)

4) простий або комбінований статистичний вплив чинників, що вивчаються (врахованих)

**12 ПРИ ВИКОНАННІ ДИСПЕРСІЙНОГО АНАЛІЗУ ВИВЧАЄТЬСЯ РІЗНОМАНІТНІСТЬ (ДИСПЕРСІЯ) ОЗНАК:**

1) міжгрупове

2) істинне

3) внутрішньогрупове

4) загальне

5) помилкове

**13 У ДИСПЕРСІЙНОМУ АНАЛІЗІ ВВАЖАЄТЬСЯ, ЩО МІЖГРУПОВА РІЗНОМАНІТНІСТЬ ЗАЛЕЖИТЬ ВІД:**

1) впливу досліджуваного чинника

2) розмірність ознаки

3) впливи якихось неврахованих випадкових чинників

4) числа груп дисперсійного комплексу

**14 У ДИСПЕРСІЙНОМУ АНАЛІЗІ ВВАЖАЄТЬСЯ, ЩО ВНУТРІШНЬОГРУПОВА РІЗНОМАНІТНІСТЬ ЗАЛЕЖИТЬ ВІД:**

1) впливу досліджуваного чинника

2) розмірність ознаки

3) впливи якихось неврахованих випадкових чинників

4) числа груп дисперсійного комплексу

**15 ТИПИ ВЗАЄМОЗВ'ЯЗКУ МІЖ ЯВИЩАМИ, ЯКІ МОЖНА ВСТАНОВИТИ МАТЕМАТИЧНИМИ МЕТОДАМИ:**

1) кореляційна

2) механічна

3) функціональна

4) непряма

**16 ЗАЛЕЖНІСТЬ МІЖ ОЗНАКАМИ, ПРИ ЯКІЙ КОЖНОМУ ЗНАЧЕННЮ ОДНІЄЇ ОЗНАКИ ВІДПОВІДАЄ СТРОГО ПЕВНЕ ЗНАЧЕННЯ ІНШОЇ ОЗНАКИ, НАЗИВАЄТЬСЯ:**

1) кореляційною

2) стандартизованной

3) функціональна

4) регресійна

**17 СТАТИСТИЧНИЙ АНАЛІЗ, ЯКИЙ ВИКОРИСТОВУЄТЬСЯ ДЛЯ ВИЯВЛЕННЯ ІМОВІРНІСНОГО ВЗАЄМОЗВ'ЯЗКУ МІЖ ОЗНАКАМИ В СОЦІАЛЬНО-ГІГІЄНІЧНИХ І МЕДИКО-БІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ, НАЗИВАЄТЬСЯ:**

1) кореляція

2) апроксимація

3) стандартизація

4) регресія

**18 КОРЕЛЯЦІЙНИЙ ВЗАЄМОЗВ'ЯЗОК МІЖ ОЗНАКАМИ, що ВИВЧАЮТЬСЯ, МОЖЕ БУТИ ВИЯВЛЕНИЙ:**

1) за допомогою статистичної таблиці

2) за допомогою графіку

3) розрахунком коефіцієнта кореляції

4) застосуванням методу стандартизації

**19 Про НАЯВНІСТЬ СИЛЬНОЇ ПРЯМОЇ КОРЕЛЯЦІЙНОЇ ЗАЛЕЖНОСТІ МОЖНА ГОВОРИТИ ПРИ ЗНАЧЕННІ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ:**

1) більше або рівно - 0,6

2) менше або рівне 0,1

3) більше 0,7

4) менше або рівне 0,4

**20 Про НАЯВНІСТЬ СИЛЬНОГО ЗВОРОТНОГО КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ МІЖ ОЗНАКАМИ МОЖНА ГОВОРИТИ ПРИ ЗНАЧЕННІ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ:**

1) від 0 до 0,3

2) від 0,3 до 0,7

3) від 0,7 до 1,0

4) від 0 до - 0,3

5) від - 0,3 до - 0,7

6) від - 0,7 до - 1,0

**21 ЗНАЧЕННЯ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ДЛЯ ПРЯМОГО ВЗАЄМОЗВ'ЯЗКУ МІЖ ОЗНАКАМИ:**

1) позитивні

2) негативні

3) будь-хто

4) менше або рівні - 1

**22 ОБЧИСЛЕННЯ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ЗА МЕТОДОМ КВАДРАТІВ (ПІРСОНУ) ВИКОНУЄТЬСЯ за ФОРМУЛОЮ:**

1)

2)

3)

4)

**23 ОБЧИСЛЕННЯ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ ЗА МЕТОДОМ РАНГІВ (СПРИМЕНА) ВИКОНУЄТЬСЯ за ФОРМУЛОЮ:**

1)

2)

3)

4)

**Список рекомендованої літератури**

1. Сергиенко В.И., Бондарева И.Б. Математическая статистика в клинических исследованиях.- М: ГЭОТАР МЕД., 2000.- 256 с.
2. Гланц С. Медико-биологическая статистика. – М.: Практика, 1999.– 459с.
3. Петри А., Сэбин К. Наглядная статистика в медицине / Пер. с англ. В.П. Леонова. – М.: ГЭОТАР МЕД., 2003. – 144с.
4. Гринхальх Т. Основы доказательной медицины: Пер. с англ.– М.: ГЭОТАР МЕД, 2006.–240 с.
5. Флэтчер Р., Флэтчер C., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины: Пер. с англ.– М.: Медиа сфера, 1998.– 349 с.
6. Страус Ш.Е. и др. Медицина, основанная на доказательствах. – М., ГЭОТАР.- 2010.-320с.

**ДОДАТОК А**

**Таблиця значень t Стьюдента для P=0,05; P= 0,01**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Кількість ступенів свободи** | **Р=0,05** | **Р=0,01** | **Кількість ступенів свободи** | **Р=0,05** | **Р=0,01** |
| 1 | 12,706 | 63,655 | 22 | 2,074 | 2,818 |
| 2 | 4,303 | 9,925 | 23 | 2,069 | 2,807 |
| 3 | 3,182 | 5,841 | 24 | 2,064 | 2,796 |
| 4 | 2,776 | 4,604 | 25 | 2,060 | 2,787 |
| 5 | 2,571 | 4,032 | 26 | 2,056 | 2,778 |
| 6 | 2,447 | 3,707 | 27 | 2,052 | 2,771 |
| 7 | 2,365 | 3,499 | 28 | 2,048 | 2,764 |
| 8 | 2,306 | 3,355 | 29 | 2,045 | 2,757 |
| 9 | 2,262 | 3,250 | 30 | 2,042 | 2,750 |
| 10 | 2,228 | 3,169 | 32 | 2,037 | 2,739 |
| 11 | 2,201 | 3,106 | 34 | 2,032 | 2,728 |
| 12 | 2,179 | 3,055 | 36 | 2,027 | 2,718 |
| 13 | 2,160 | 3,12 | 38 | 2,025 | 2,711 |
| 14 | 2,145 | 2,977 | 40 | 2,021 | 2,704 |
| 15 | 2,131 | 2,947 | 42 | 2,017 | 2,696 |
| 16 | 2,120 | 2,921 | 44 | 2,015 | 2,691 |
| 17 | 2,110 | 2,898 | 46 | 2,012 | 2,685 |
| 18 | 2,101 | 2,878 | 48 | 2,010 | 2,681 |
| 19 | 2,093 | 2,861 | 50 | 2,007 | 2,678 |
| 20 | 2,086 | 2,845 | 55 | 2,005 | 2,668 |
| 21 | 2,080 | 2,831 | 60 | 2,000 | 2,660 |
|  |  |  | ~ | 1,960 | 2,576 |

**ДОДАТОК Б**

**ВИБІР МЕТОДУ СТАТИСТИЧНОГО АНАЛІЗУ В ЗАЛЕЖНОСТІ ВІД УМОВ ЕКСПЕРИМЕНТУ**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Взаємозалежність ознак | Визначення відмінностей достовірності відмінностей між ознаками | | | | Взаємозв’язок ознак |
| Незалежні групи | | Залежні групи (повторні виміри) | |
| Дві групи | Більш двох груп | Одна група до і після лікування | Одна група, декілька  видів лікування |
| Кількісна,  нормальний  розподіл | Критерій Стьюдента | Дисперсійний аналіз. Потім критерій Стьюдента з поправкою Бонферроні або критерій Ньюмена-Кейлса | Парний критерій Стьюдента | Дисперсійний аналіз повторних вимірів, потім критерій Стьюдента або критерій Ньюмена-Кейлса для повторних вимірів | Регресійний аналіз.  Коефіцієнт кореляції Пірсона. |
| Кількісна,  ненормальний розподіл або порядкова ознака | Критерії Манна-Уітні, критерій Уайта, Ван дер Вадена, Колмогорова-Смирнова | Критерій Крускала-Уолліса, потім непараметричні варіанти критеріїв Ньюмена-Кейлса, Даннета або Данна. | Критерій Уілкоксона Критерій знаків | Критерій Фрідмана, потім непараметричні варіанти критеріїв  Ньюмена-Кейлса або  Даннета | Коефіцієнт рангової кореляції Спірмена, коефіцієнт кореляції Кендалла |
| Якісна ознака | Критерій , критерій Z, точний  критерій Фішера | Критерій , потім  таблиці з чотирьома полями з оцінкою розбіжностей по | Критерій Мак-Німара | Критерій Кокрена, потім таблиці з чотирьома полями з оцінкою розбіжностей по | Коефіцієнт пов’язаності. Коефіцієнт асоціації Юла |

1. Кулаичев А.П. Методы и средства анализа данных в среде Windows. - М: ИнКо, 2002. - 341 с. [↑](#footnote-ref-1)