**Темы дипломных работ 2023-2024**

**Доцент кафедры ТСиВС, к.ф.-м.н. Лукинов Виталий Леонидович**

**e-mail:** [vitaliy.lukinov@sscc.ru](mailto:vitaliy.lukinov@sscc.ru)

**Темы бакалаврских работ**

[1. Разработка онлайн инструментов для аналитической обработки данных на маркетплейсах](#_aoep9fn1ui76)

[2. Разработка онлайн калькулятора для прогнозирования риска тромбоза](#_s4kuitt8to09)

[3. Разработка онлайн сервиса анализа совместимости лекарств](#_akw1h7lpelag)

[4. Разработка онлайн сервиса обработки рентгеновских снимков](#_jdfftnpirldy)

[5. Имитационное моделирование аварий вычислительных блоков в большой системе](#_tejknbo89tsz)

[6. Разработка сервиса по расчету размера групп (выборки) в клинических исследованиях](#_yf5bmbv1ph24)

[7. Разработка онлайн сервиса для планирования набора в клинических исследованиях](#_7byl8axcg7zp)

[8. Разработка онлайн сервиса ведения клинических исследованиях](#_dk9v7nc8k5u1)

[9. Исследование алгоритмов поиска экстремумов](#_wxcmdpywzrmm)

[10. Разработка параллельного генератора псевдослучайных чисел](#_blxcs76dqvxt)

[11. Разработка системы автоматической балансировки производительности функций из BLAS](#_fp0azvcgly2u)

[12. Разработка пакета статистического сравнительного анализа данных](#_85l60wn2yzln)

[13. Разработка пакета автоматизированного графического анализа данных](#_ccz3ea58jmzx)

[14. Разработка библиотеки для визуализации результатов мета-анализов](#_9ktnqef43i29)

[15. Моделирования COVID в Новосибирске.](#_b35ccg1x006r)

[16. Создание библиотеки расчета многофакторных моделей регрессий.](#_kdzmsuwd2256)

[17. Создание библиотеки для выявления предикторов риска негативных событий.](#_wwixotiz8mws)

### Разработка онлайн инструментов для аналитической обработки данных на маркетплейсах

Задачи по теме «Вайлдберриз» (<https://openapi.wildberries.ru/> )

1. Автоответы. Получение данных по апи продавца, автоматический ответ покупателю на отзыв в зависимости от окраски отзыва и кол-во звезд, при появление вопросов будет отправлено письмо(сообщение) менеджеру для ответа.

2. Фин аналитика. Сбор и хранение данных продаж по апи продавца, вывод в Excel. Дашборды.

3. Сбор данных со страницы ВБ по определенному запросу по определенным параметрам/предварительная фильтрация данных для дальнейшей работы (размерная сетка, список брендов конкурентов), вывод данных в ексель в определенной форме.

Задачи по теме «Озон»

( https://docs.ozon.ru/api/seller/#section/Upravlyajte-zakazami-FBO-FBS-i-rFBS/Shema-FBS-Standart )

1. Контроль остатков, есть такой вариант решения <https://habr.com/ru/articles/672194/> , хотелось бы получить программно, ответов файл в Excel для дальнейшей работы.

2. Фин аналитика. Сбор и хранение данных по апи продавца, вывод в Excel. Дашборды.

(<https://docs.ozon.ru/api/seller/#tag/FinanceAPI> )

***Средства разработки*** – по желанию

***Количество студентов – до 4***

1. Разработка сервиса голосового рецепта

Создание сервиса для голосового распознавания назначаемых лекарств и процедур в соответствии с текущей базой и автоматического составления рецепта для пациента

***Средства разработки*** – по желанию

***Количество студентов – 1***

1. Разработка сервиса голосового описания снимков МРТ

Создание сервиса для голосового распознавания описания снимков МРТ и составления текстового описания с привлечением базы описательных шаблонов

***Средства разработки*** – по желанию

***Количество студентов – 1***

### Разработка онлайн калькулятора для прогнозирования риска тромбоза

Необходимо создать онлайн калькулятор в котором по введенным показателям пациента рассчитывается риск возникновения тромбоза, а также происходит сбор данных для контроля и дальнейшего обучения моделей прогноза. При вычислении риска используются готовые (созданные) модели логистической регрессии и деревьев градиентного бустинга. Калькулятор планируется внедрить в практику врачей.

***Средства разработки*** – R или Python

***Количество студентов – до 2***

### Разработка онлайн сервиса анализа совместимости лекарств

Нужно разработать сервис для анализа совместимости принимаемых лекарственных средств. Пользователь должен иметь возможность удобного быстрого ввода лекарств и получить анализ по совместимости и рисках их применения. Данные о рисках и совместимости получаются с помощью нейросетей.

***Аналог*** <https://www.drugs.com/newdrugs.html>

***Средства разработки*** – по желанию

***Количество студентов – до 2***

### Разработка онлайн сервиса обработки рентгеновских снимков

Необходимо разработать нейросеть для распознавания позвоночника человека и автоматического вычисления углов для расчета параметров сагиттального баланса. Сервис востребован для выбора оптимального лечения пациентов с заболеваниями позвоночника. Результаты работы могут быть внедрены в общероссийскую систему здравоохранения.

***Средства разработки*** – Python

***Количество студентов – до 2***

### Имитационное моделирование аварий вычислительных блоков в большой системе

В суперкомпьютерах с сотнями тысяч вычислительных элементов гарантированно случаются аварии отдельных элементов. Например, в последнем релизе суперкомпьютера [Sunway TaihuLight](https://www.top500.org/system/178764) с гордостью сообщили о 13 часовой безаварийной работе с момента запуска. Актуальна задача об исследовании пределов возможности расширения (добавления новых вычислительных элементов) больших систем с повышением производительности.

Необходимо запрограммировать и смоделировать поведение большой однородной системы вычислительных элементов в которой с заданной интенсивностью может происходить авария вычислительного элемента, с заданным временем происходит восстановление, есть пул запасных элементов и т.п. Стоит задача нахождения пределов расширения вычислительной системы при котором увеличивается прирост производительности с учетом аварий.

Имитационное моделирование будет необходимо сделать на основе известной легко программируемой и распараллеливаемой модели пуассоновских потоков.

***Средства разработки*** – по желанию, предпочтительно С/С++

***Количество студентов – 1***

### Разработка сервиса по расчету размера групп (выборки) в клинических исследованиях

Для планирования клинических исследований очень важно рассчитать необходимое количество пациентов в исследовании. Необходимо создать онлайн сервис для расчета размера групп на основе предположений (гипотез) относительно целевых показателей, сервис помимо расчета размеров выборки должен предоставлять графики и формировать полноценный текстовый отчет по расчетам в виде файла .docx, включающий справочную информацию, таблицы со моделируемыми сценариями, основную гипотезу и графики. Результаты работы очень востребованы и будут использоваться в реальных научных исследованиях.

**Аналоги:** <http://powerandsamplesize.com/Calculators/>

<https://www.calculator.net/sample-size-calculator.html>

<https://www.surveymonkey.com/mp/sample-size-calculator/>

***Средства разработки*** – по желанию Python или R

***Количество студентов – до 2 (back/front).***

### Разработка онлайн сервиса для планирования набора в клинических исследованиях

Требуется продолжить разработку сервиса по рандомизированному и стратифицированному набору в клинических исследованиях. Для чистоты экспериментов в клинических исследованиях пациентам случайным образом назначают вид лечения. Задачей сервиса является создание предварительного плана случайных распределений лечений среди пациентов в исследовании таким образом, чтобы этот план можно было потом проверить, нельзя было предугадать распределение, можно было выслать подтверждающую ссылку на план любому заинтересованному лицу после набора пациентов, была возможность проводить большие многоцентровые исследования (несколько планов для разных медицинских центров) с единовременным набором в рамках одного проекта.

***Средства разработки*** – Python.

***Количество студентов – до 2.***

### Разработка онлайн сервиса ведения клинических исследованиях

Требуется продолжить разработку сервиса по ведению клинических исследований.

Онлайн сервис должен помогать создавать и вести базу данных пациентов в клинических исследованиях.

**Аналоги:** https://redcap.icts.uiowa.edu/redcap/index.php?action=training

https://clinicaltrials.gov/

***Средства разработки*** – Python, по желанию.

***Количество студентов – до 3.***

### Исследование алгоритмов поиска экстремумов

Необходимо разработать тестовую систему (сборка, запуск и набор тестов) и исследовать вычислительные свойства различных современных алгоритмов детерминированного и стохастического поиска экстремума функции и их производительность. Будут использоваться готовые реализации алгоритмов («рой частиц». «стохастического градиента» и др.) из библиотек и собственные реализации алгоритмов.

***Средства разработки*** – по желанию python, С/С++

***Количество студентов – 1.***

### Разработка параллельного генератора псевдослучайных чисел

Нужно разработать и реализовать параллельный алгоритм генерации псевдослучайных чисел, создать тестовую систему, сравнить производительность и качества генератора с другими генераторами.

Требуются реализации для генератора на многоядерной рабочей станции и на графической карте. Один студент выбирает одну реализацию. В идеале поддержка мульти платформенной работы.

***Средства разработки*** – OpenMP, CUDA, C/C++.

***Количество студентов – до 3.***

### Разработка системы автоматической балансировки производительности функций из BLAS

Необходимо реализовать часть параллельных функций для базовых операций линейной алгебры BLAS (математические операции с векторами и матрицами – умножения и т.п.) с помощью декомпозиции данных для параллельных расчетов, провести параметризацию декомпозиции, после чего написать скрипты для выявления набора параметров предоставляющих наилучшую производительность на тестовых наборах матриц и векторов.

Средства разработки – C/C++, по желанию.

***Количество студентов – до 2.***

### Разработка пакета статистического сравнительного анализа данных

Необходимо разработать пакет R (библиотеку), создающую таблицы с результатами статистического сравнения данных в разных группах и записывающую результаты работы в файл формата .docx с автоматическим созданием выводов из результатов сравнений.

***Средства разработки*** – R, пакет officer для записи в .docx.

***Количество студентов – до 2.***

### Разработка пакета автоматизированного графического анализа данных

Необходимо разработать пакет R (библиотеку), создающую графики с результатами статистического сравнения данных в разных группах и записывающую результаты работы в файл формата .docx с автоматическим созданием выводов из результатов сравнений.

***Средства разработки*** – R, пакет ggplot2 для создания графиков, пакет officer для записи в .docx.

***Количество студентов – до 2.***

### Разработка библиотеки для визуализации результатов мета-анализов

Мета-анализы – это медицинские обзорные статьи, в которых анализируются и сравниваются разные подходы для лечения одного и того же заболевания. С помощью статистических методов происходит обработка числовых результатов из разных статей для понимания следующих пунктов:

* какой способ лечения лучший и какой худший;
* какой средний эффект лечения среди всех подходов;
* насколько устойчив эффект – если ли устойчивая тенденция в лечении
* есть ли зависимость эффекта от определенных показателей до начала лечения и т.п.

Предлагается сделать пакет (библиотеку) языка R с помощью пакета metafor для автоматического формирования графиков и таблиц в мета-анализе.

***Примеры Графиков****:*

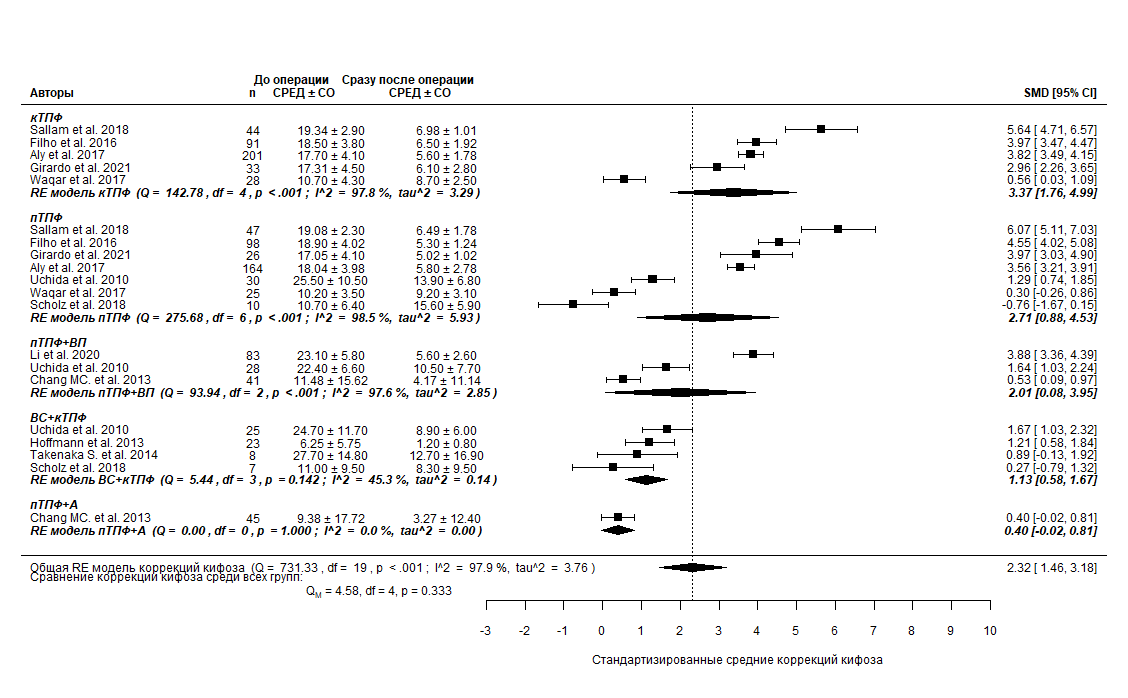
**

Рисунок коррекции кифоза разными хирургическими вмешательствами

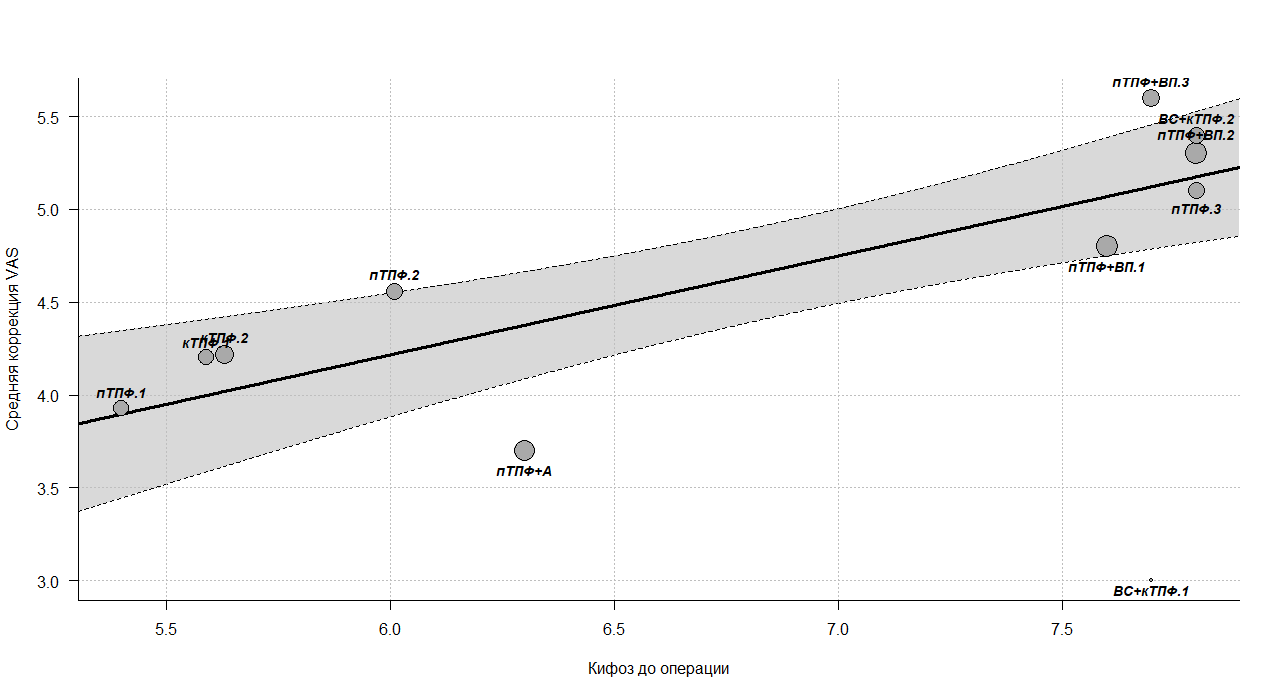
**

Рисунок зависимость снижения боли от исходного кифоза

***cсылки:***

[*https://www.metafor-project.org/doku.php/metafor*](https://www.metafor-project.org/doku.php/metafor)

<https://github.com/wviechtb/metafor>

***Средства разработки:*** *язык R*

***Количество студентов – 1.***

### Моделирования COVID в Новосибирске.

Требуется на основе пакета covasim смоделировать распространение эпидемии COVID-19 в Новосибирской области и сравнить результаты моделирования с другой программой на основе пуассоновских потоков.

Необходима разработка серверной части для выполнения расчетов, и веб интерфейса для ввода исходных данных и демонстрации результатов расчетов.

**cсылки:**

[*https://docs.idmod.org/projects/covasim/en/latest/index.html*](https://docs.idmod.org/projects/covasim/en/latest/index.html)

*https://github.com/institutefordiseasemodeling/covasim*

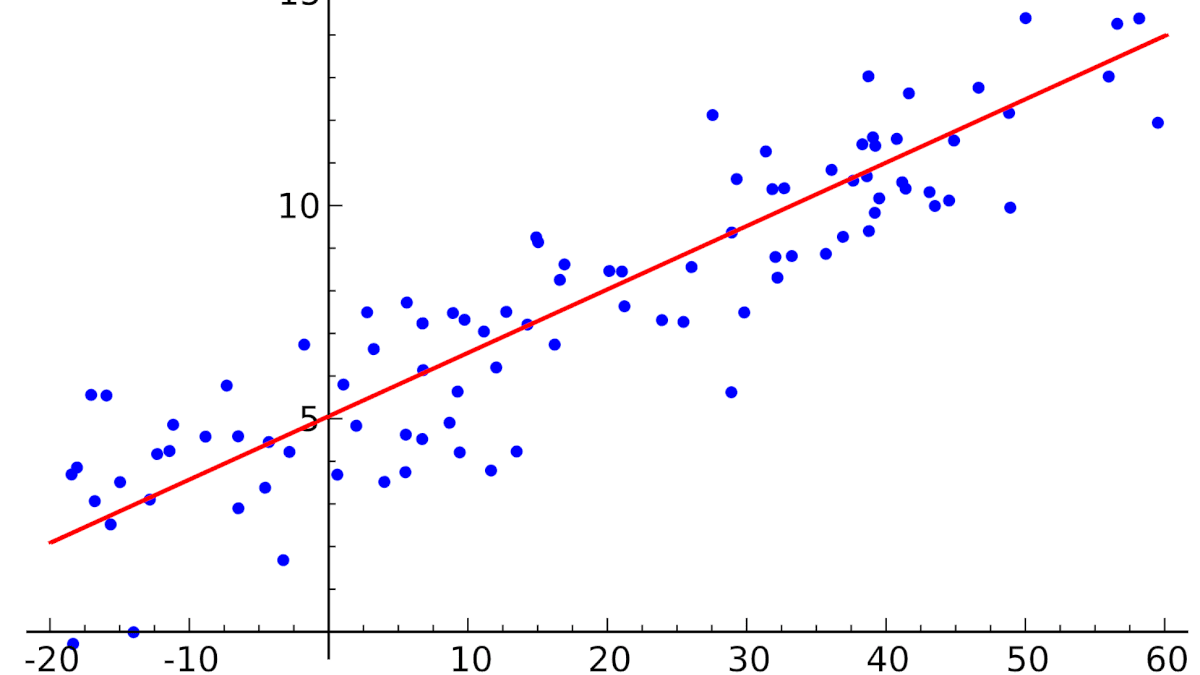
**Средства разработки: язык С/С++, python**

**Количество студентов – до 3.**

### Создание библиотеки расчета многофакторных моделей регрессий.

Требуется приложение, в котором пользователь может загрузить свои данные и получит в качестве результата модель (формулу линейной регрессии и прогноз) для предсказания значений одной переменной в зависимости от значений других переменных. Кроме того, после получения новых данных, система должна пересчитать и уточнить рассчитанную ранее модель.

Пример однофакторной модели (у зависит от одной переменной x)



***cсылки:***

[*https://temofeev.ru/info/articles/lineynaya-regressiya-i-metody-eye-vosstanovleniya/*](https://temofeev.ru/info/articles/lineynaya-regressiya-i-metody-eye-vosstanovleniya/)

*https://habr.com/ru/post/514818/*

***Средства разработки:*** *язык R*

***Количество студентов – до 2.***

### Создание библиотеки для выявления предикторов риска негативных событий.

Необходимо разработать алгоритмы поиска предикторов риска с помощью моделей пропорциональных рисков Кокса и логистической регрессии. Основной задачей является разработка визуализации качества созданных моделей для выбора наиболее подходящих предикторов. Отдельной задачей является создание веб приложения на основе созданной библиотеки полного перебора.

***Средства разработки:*** язык R для построение моделей, web клиент на любом языке

***Количество студентов – до 3.***

**Темы магистерских работ (больше года)**

1. **Разработка параллельного генератора случайных чисел**

*Нужно разработать и реализовать параллельный алгоритм генерации псевдослучайных чисел, создать тестовую систему, сравнить производительность и качества генератора с другими генераторами.*

*Требуются реализации для генератора на многоядерной рабочей станции, на графической карте, на гетерогенном вычислительном кластере. В идеале с поддержкой мульти платформенной работы.*

1. **Создание вычислительных алгоритмов непрерывной статистической обработки данных.**

*Необходимо разработать новые параллельные алгоритмы построения и оценки качества регрессий для непрерывно обновляющегося потока данных.*

***Количество студентов – до 2.***

1. **Разработка численных параллельных алгоритмов для моделирования физических процессов на суперкомпьютерах.**

*Разработка новых параллельных алгоритмов для численного анализа решений уравнений математической физики со случайными шумами.*

***Количество студентов – до 3.***

1. **Моделирование развития эпидемий.**

*Необходимо разработать и реализовать модификации известных алгоритмов для прогнозирования развития эпидемий*

***Количество студентов – до 2.***