

# РАЗРАБОТКА WEB-ПРИЛОЖЕНИЯ ДЛЯ АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ РЕЗУЛЬТАТОВ МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ

Сластин А.П., Мазорчук М.С.

*Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ»,  
г. Харьков, Украина*

**Цель работы:** разработка веб-приложения, предназначенного для статистической обработки и визуализации медицинских наборов данных.

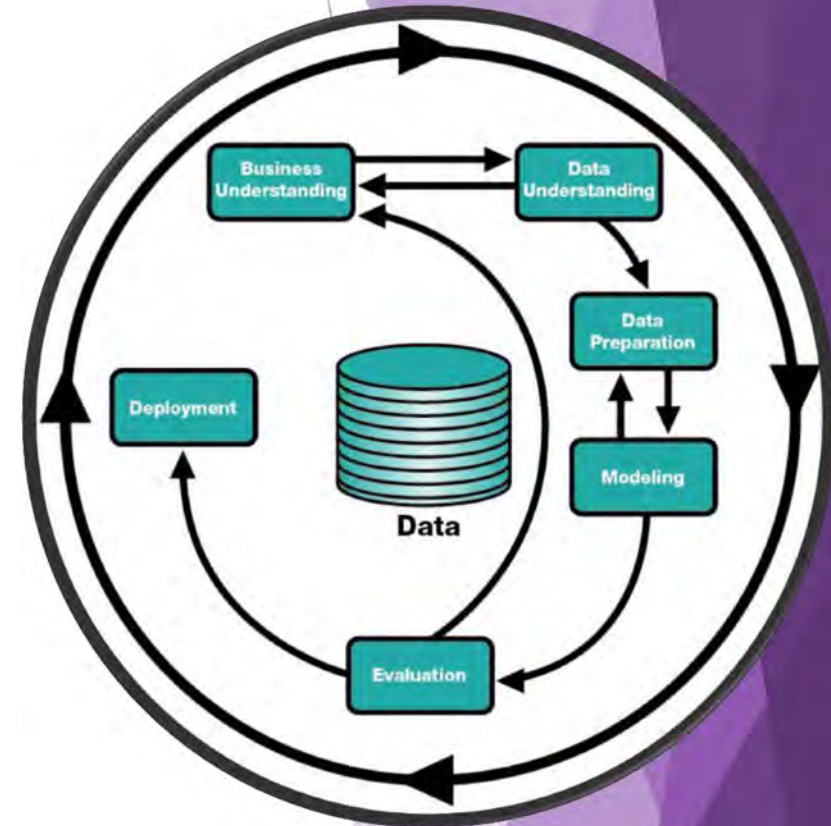
**Объект исследования:** процесс оценивания значимых факторов о пациентах, имеющих различные острые заболевания печени и перенесшие осложнения после оперативных вмешательств.

**Предмет исследования:** модели прогнозирования результатов лечения пациентов с учетом влияния значимых факторов на основе статистических методов.

# Постановка задачи

Для анализа врачами-исследователями были собраны данные по 32 показателям 127 пациентов, имеющих различные острые заболевания печени и перенесшим осложнения после оперативных вмешательств.

Необходимо построить модель прогнозирования осложнений пациентов при различных оперативных вмешательствах с учетом влияния множества факторов.



# Предварительная обработка данных

Medical research

The research data

Show 10 entries

	id	fn	dates	diagnosis	sex	age	level_bilia_block	time_yellow	diag_prehospital	weight	oper_risk	cholangitis	pathology	days_common	days_before
1	8	Пелешенко И.Г. №6006	12/12/11-16/01/13	3	муж	49	1	1	1	1	3	0		34	
2	28	Власенко А.А №4840	29/09/14-14/09/14	3	муж	67	1	1	0	0	3	2	0	15	
3	31	Дмитрусенко Т.И. №1690	21/03/14-22/04/14	3	жен	54	1	1	0	0	3	1	0	32	
4	73	Ильченко О.М.	09/02/12-07/03/12	1	жен	76	2	1	0	0	3	0	1	27	
5	75	Суббота Т.В.	19/05/14-09/07/14	1	жен	53	2	3	1	2	3		1	51	
6	76	Филова Н.Ф.	13/05/13-20/05/13	3	жен	68	1	3	0	2	3	0	0	7	
7	78	Мамонова П.И.	23/11/15-22/12/15	1	жен	86	2	2	0	2	3	0	1	29	
8	79	Губенко О.М.	16/03/15-15/04/15	2	жен	38	1	2	0	0		0	0	30	
9	108	Толстой И.Г. №4937	07/09/16-09/09/16	3	муж	70	1	3	0	2	3	1	1	2	
10	129	Чернышевич Т.В. №4774	01/09/16-13/09/16	1	жен	78	2	1	0	1	3	0	1	12	

Showing 1 to 10 of 127 entries

1 2 3 4 5 ... 13 Next

Search:

Пол: преобразование в двухфакторную переменную

Длительность желтухи: преобразование в трехфакторную переменную

Количество койко-дней общее: преобразование в числовую переменную и замена пропущенных значений средним

Патология: преобразование в двухфакторную переменную

# Методы и модели анализа данных

В зависимости от вида шкал измерения переменных для исследования связей между ними используют различные статистические методы, которые показаны на таблице ниже.

Шкала измерения влияющих переменных	Шкала измерения зависимых переменных	Применяемые методы
Интервалов или отношений	Интервалов или отношений	Регрессионный и корреляционный анализ
Время	Интервалов или отношений	Анализ временных рядов
Наименований или порядка	Интервалов или отношений	Дисперсионный анализ <sup>3</sup>
Смешанный	Интервалов или отношений	Ковариационный и регрессионный анализ
Наименования или порядка	Наименования или порядка	Анализ ранговых корреляций и таблиц сопряженности <sup>4</sup>
Наименования или порядка	Интервалов или отношений	Дискриминантный анализ; кластерный анализ; таксономия

→ Для исследования связи метрических признаков используется регрессионный анализ

→ Если число факторов больше двух, то для анализа значимости различий метрических признаков между группами используется дисперсионный анализ

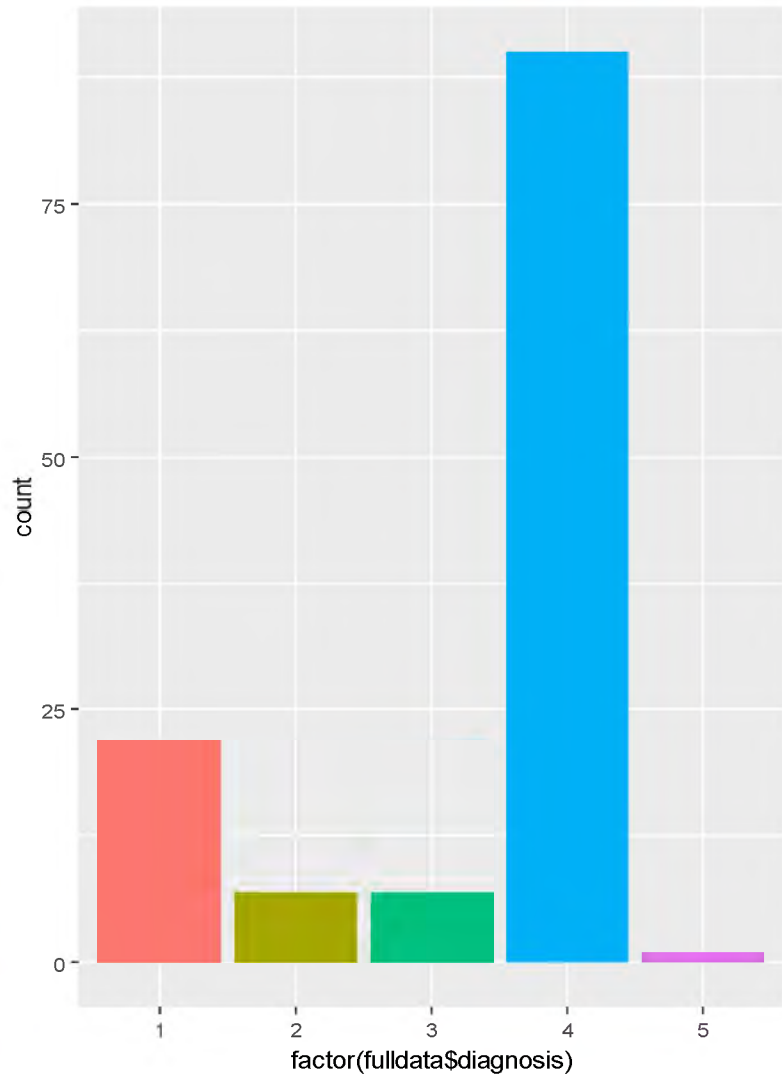
→ Для оценивания вероятности попадания признака в группу на основе анализа метрических факторов используется дискриминантный анализ

# Средства анализа данных и разработки приложений

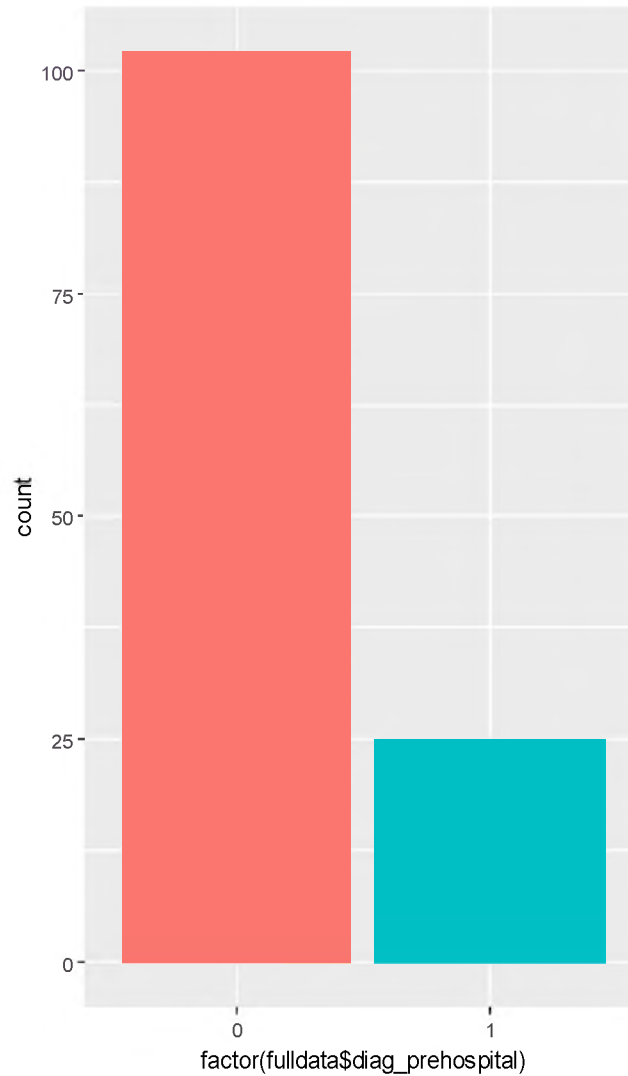
- ▶ R — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом.
- ▶ R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ.
- ▶ R поддерживает широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов. Пакеты представляют собой библиотеки для работы специфических функций или специальных областей применения.
- ▶ *Shiny* — это пакет (фреймворк) для быстрой разработки веб-приложений на базе R. Использование *Shiny* (как своего сервера или облачного сервера) позволит продемонстрировать результаты анализа в более удобной форме.



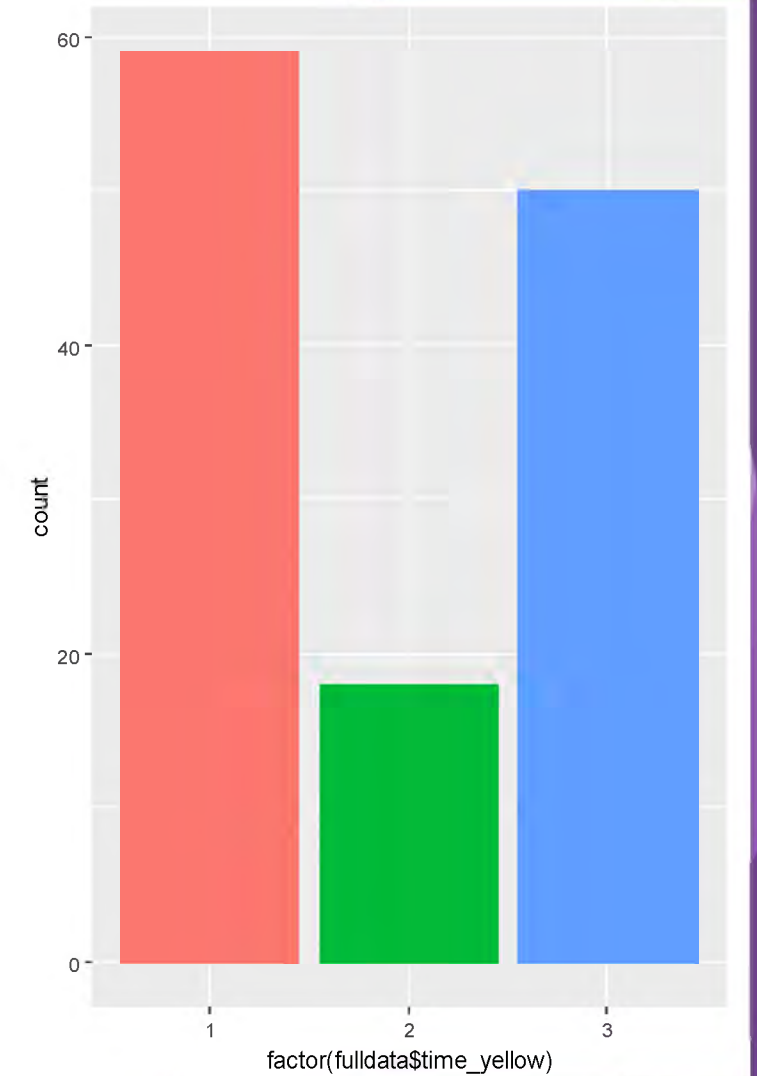
# Первичный анализ данных



Диагноз поступивших пациентов



Правильность диагноза на догоспитальном этапе

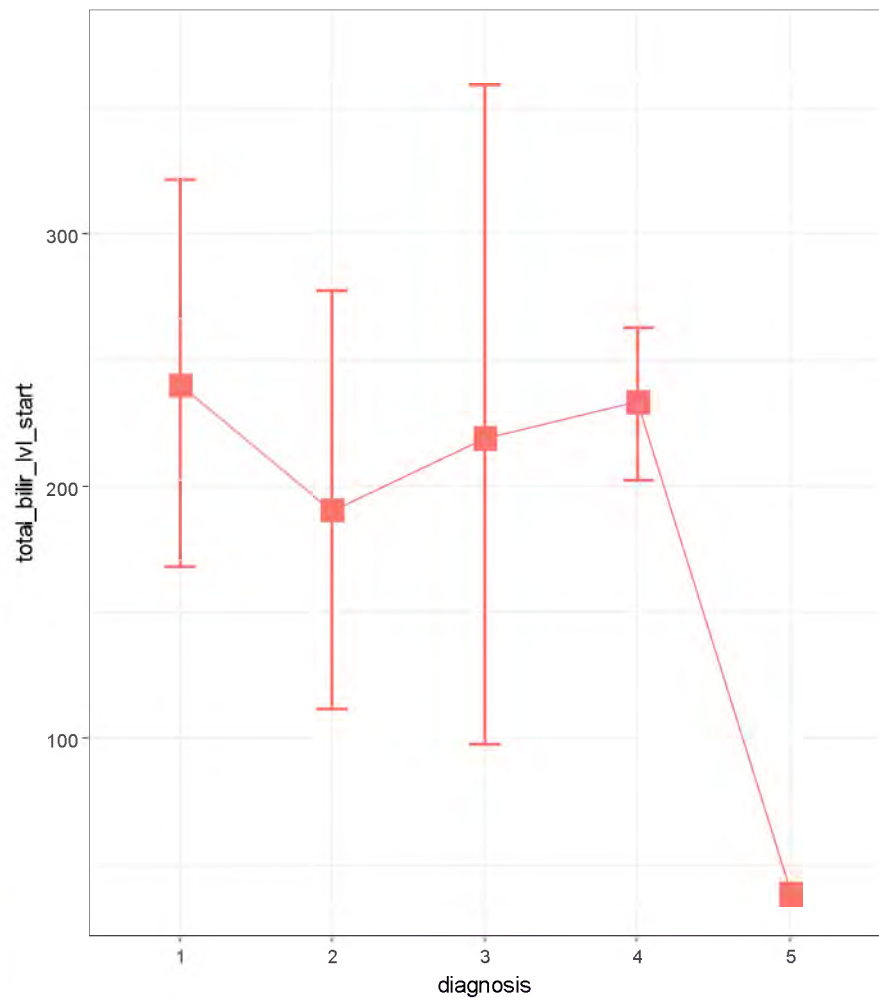


Длительность желтухи

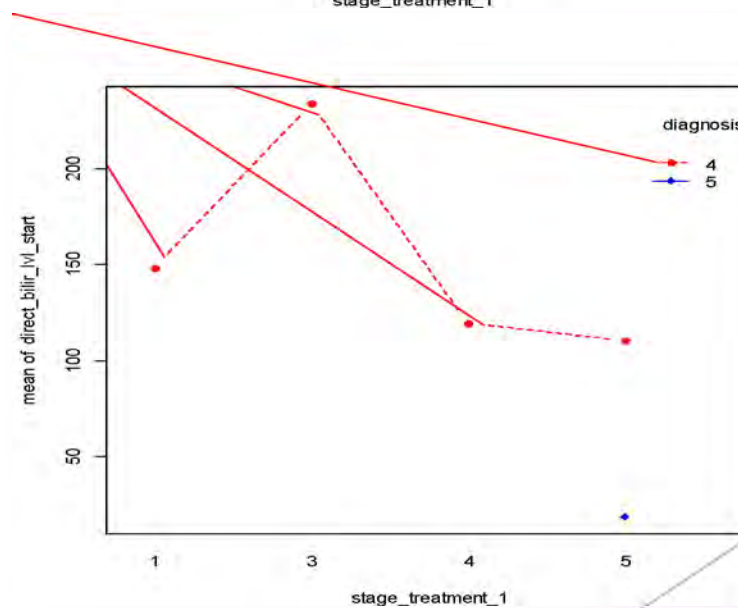
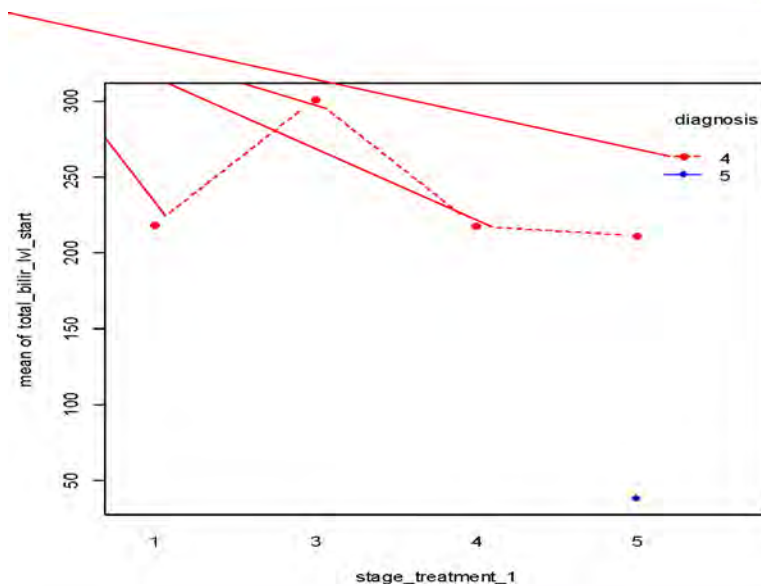


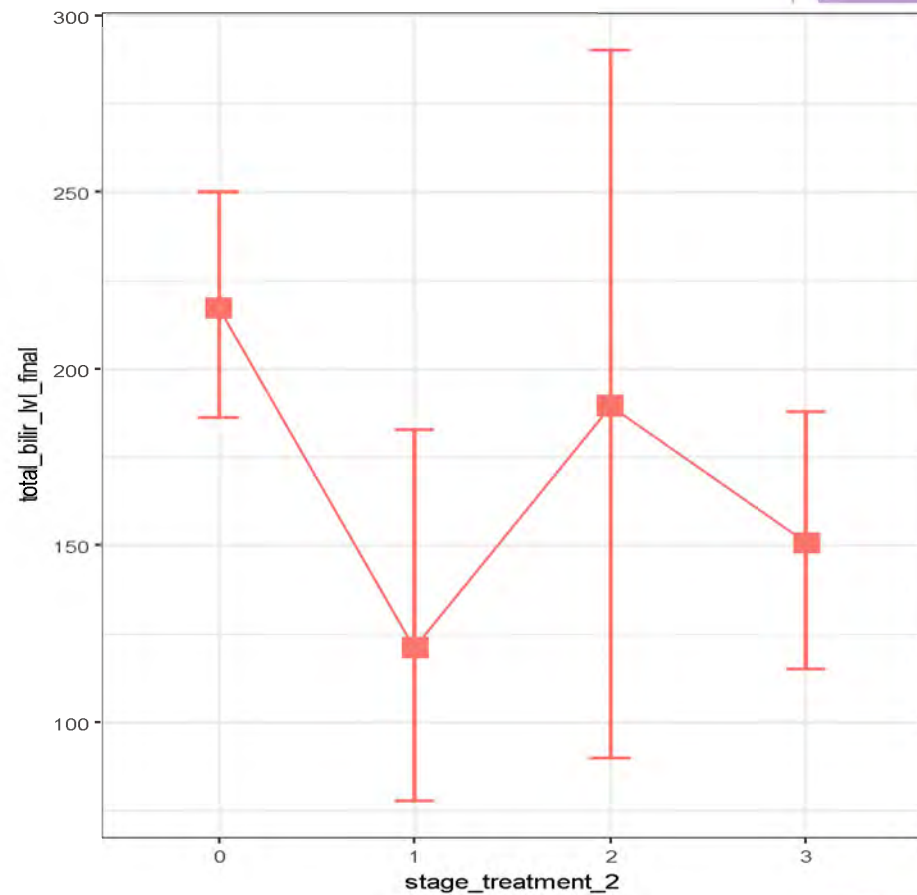
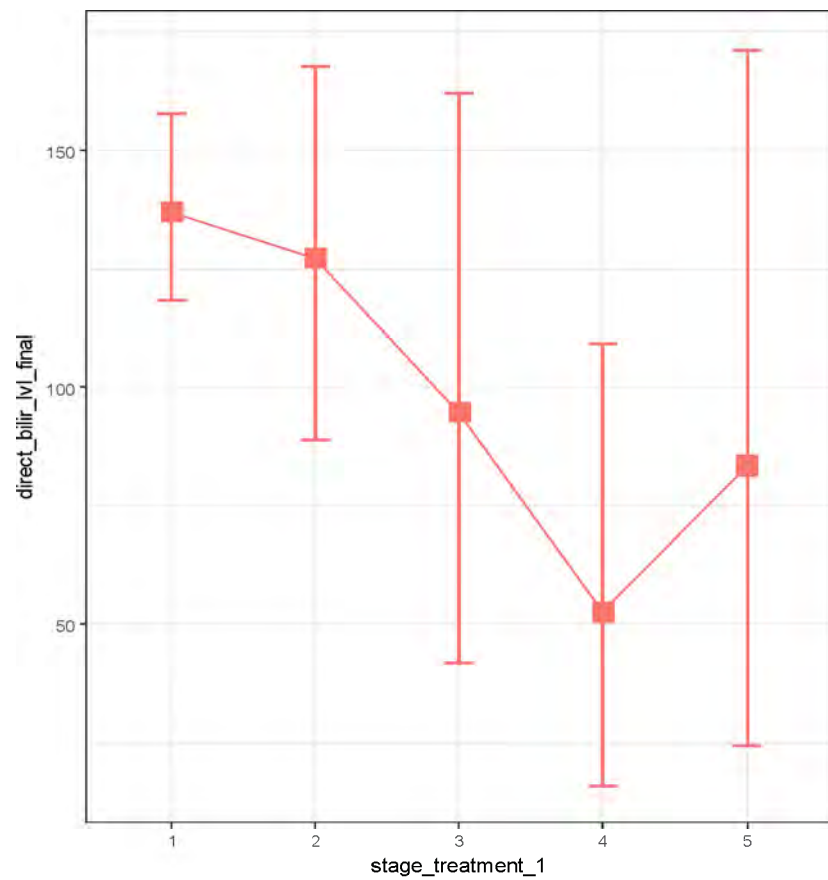


# Дисперсионный анализ

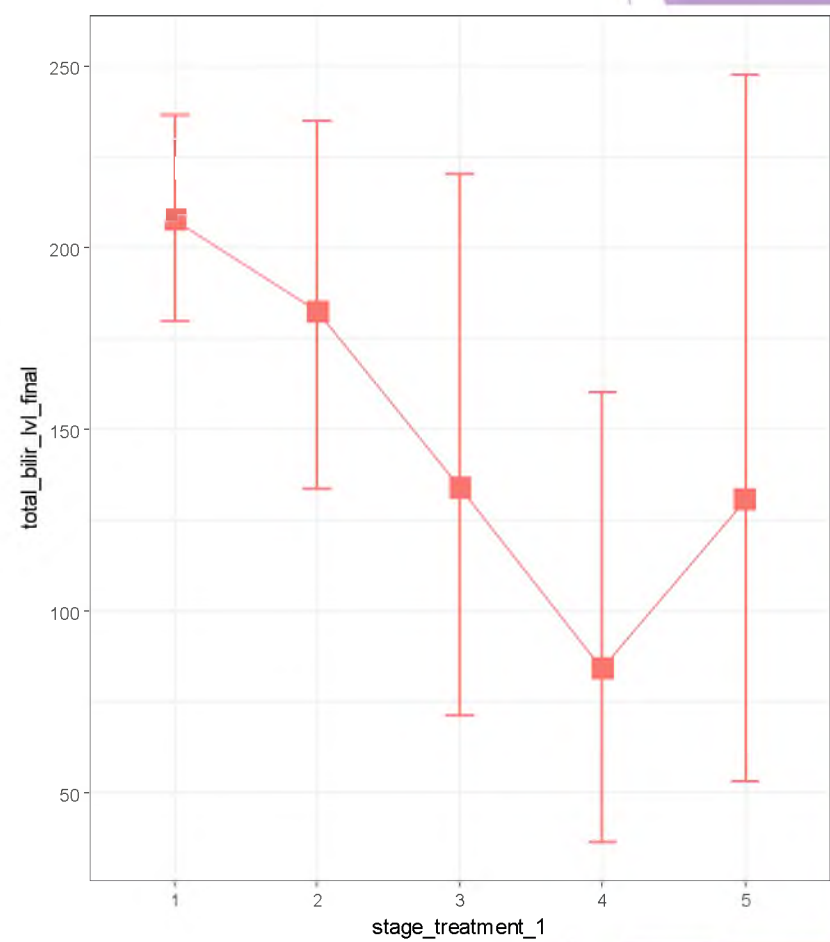
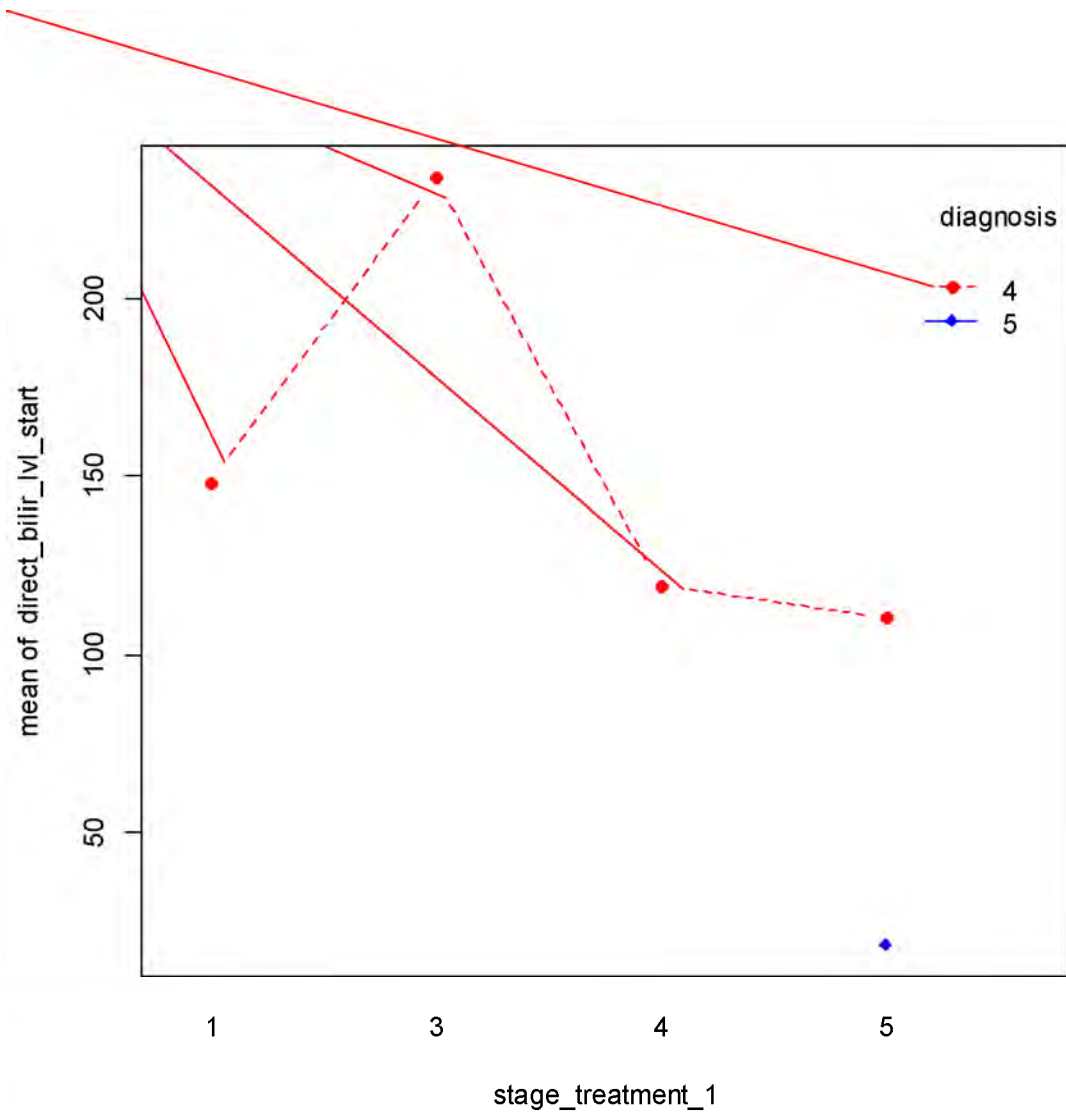


Начальный уровень билирубина на первом этапе лечения

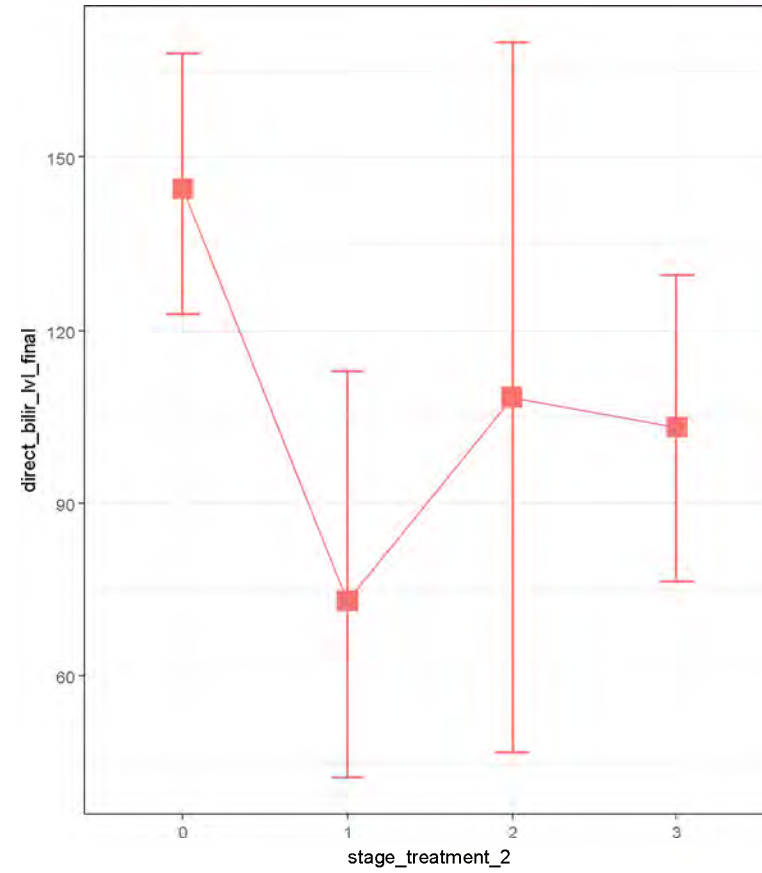
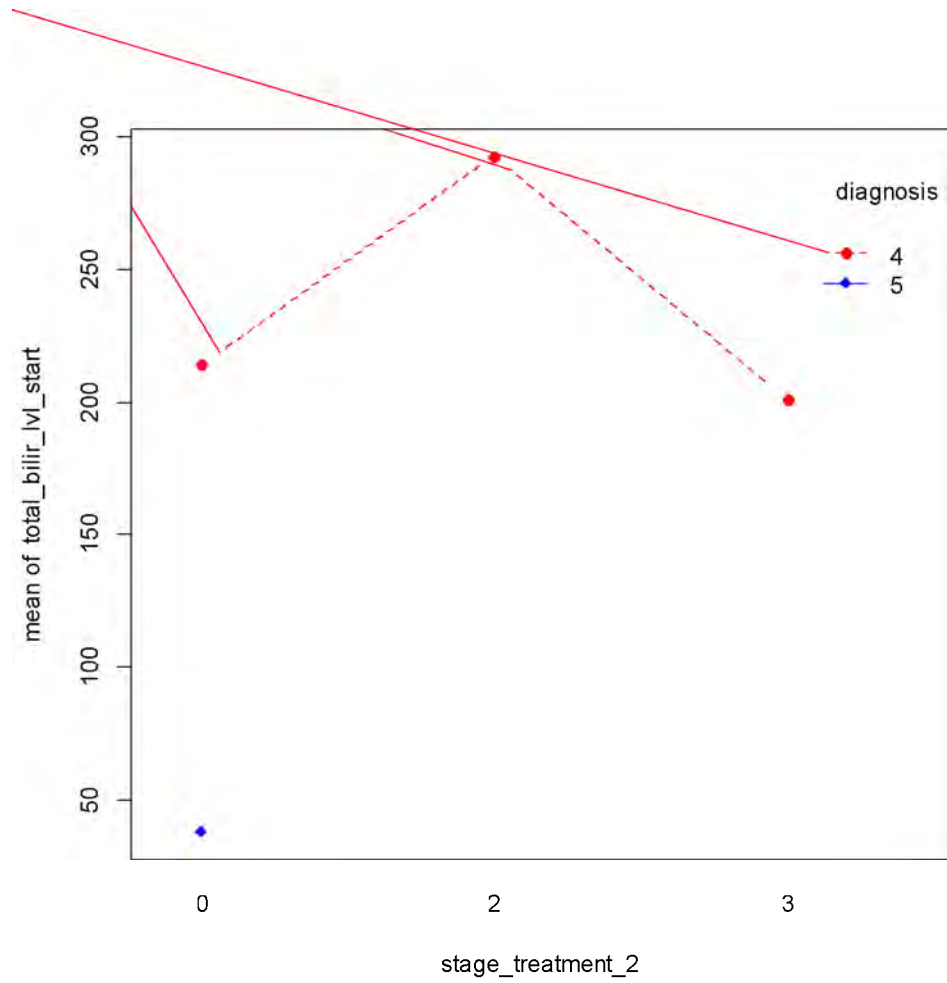




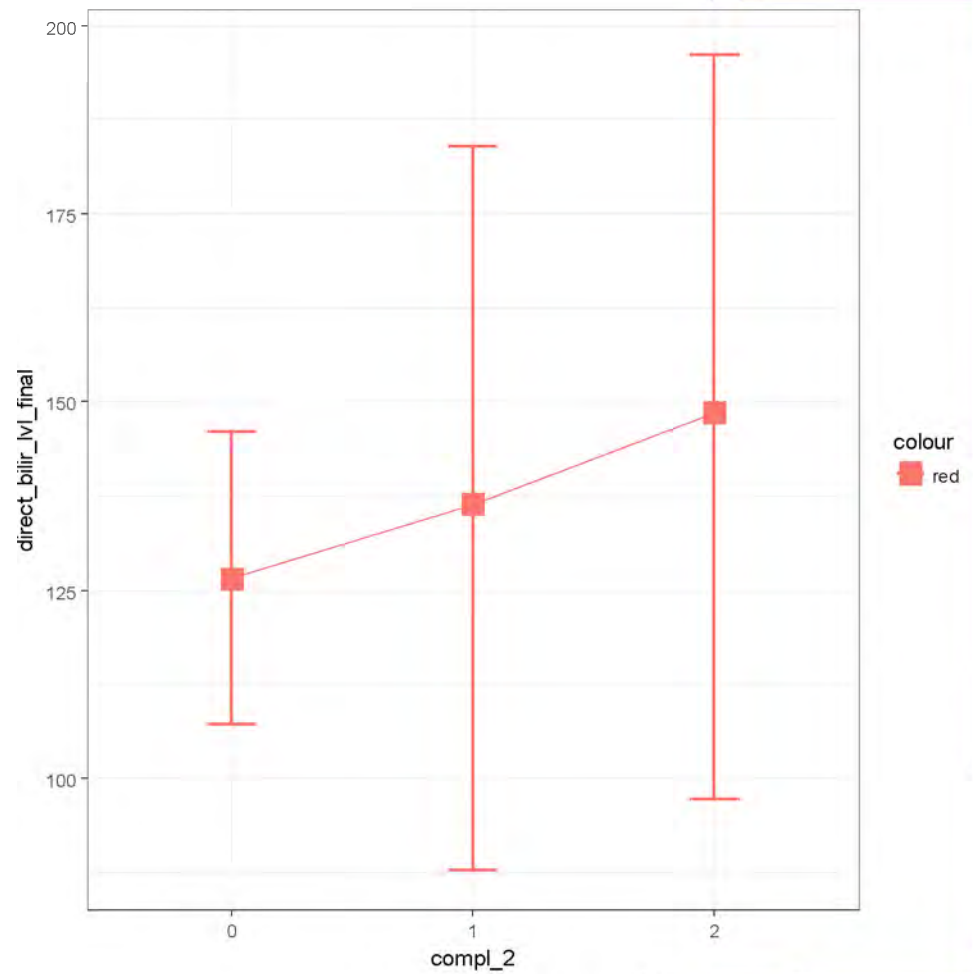
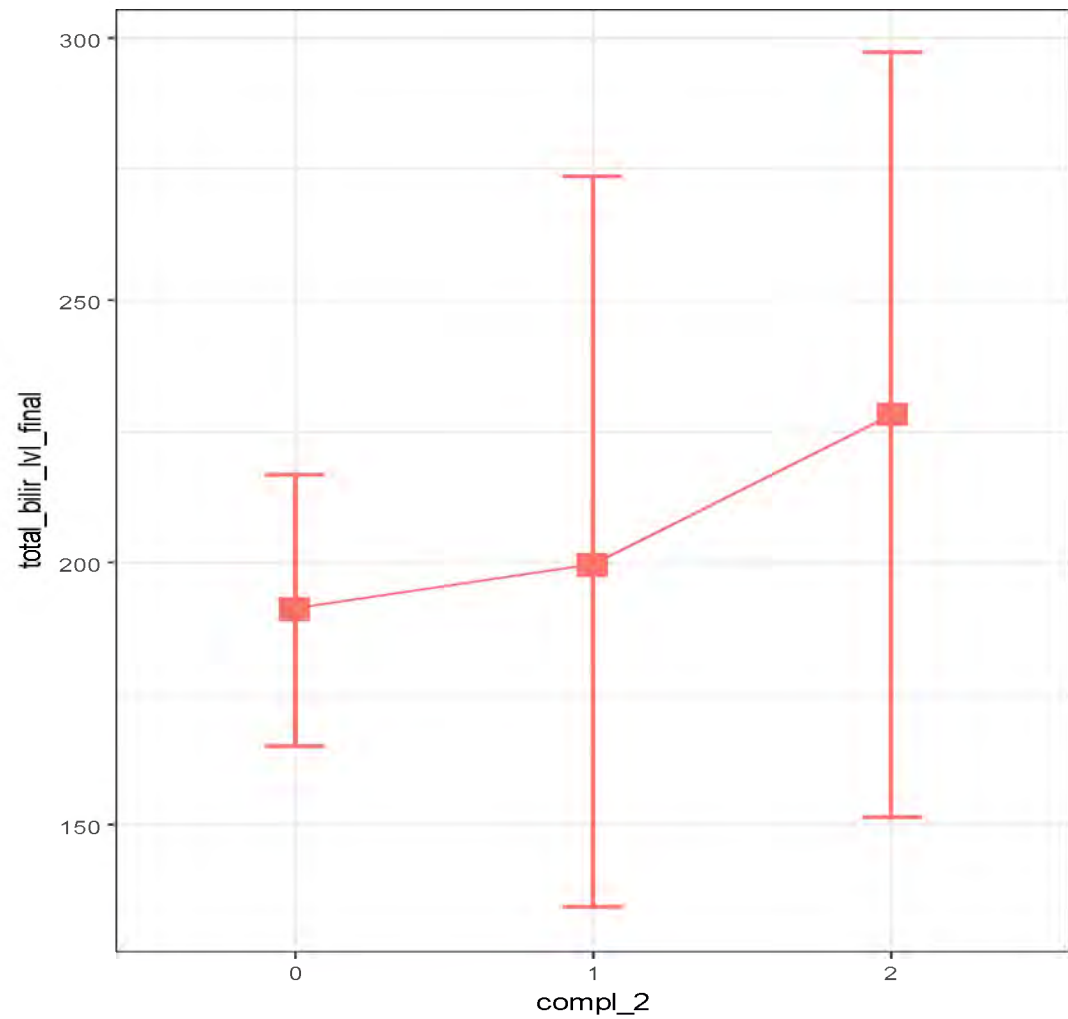
Уровень билирубина после лечения на различных этапах



Первый этап лечения



Второй этап лечения



Осложнения второго этапа лечения

# Ответ желтухи на дренирование

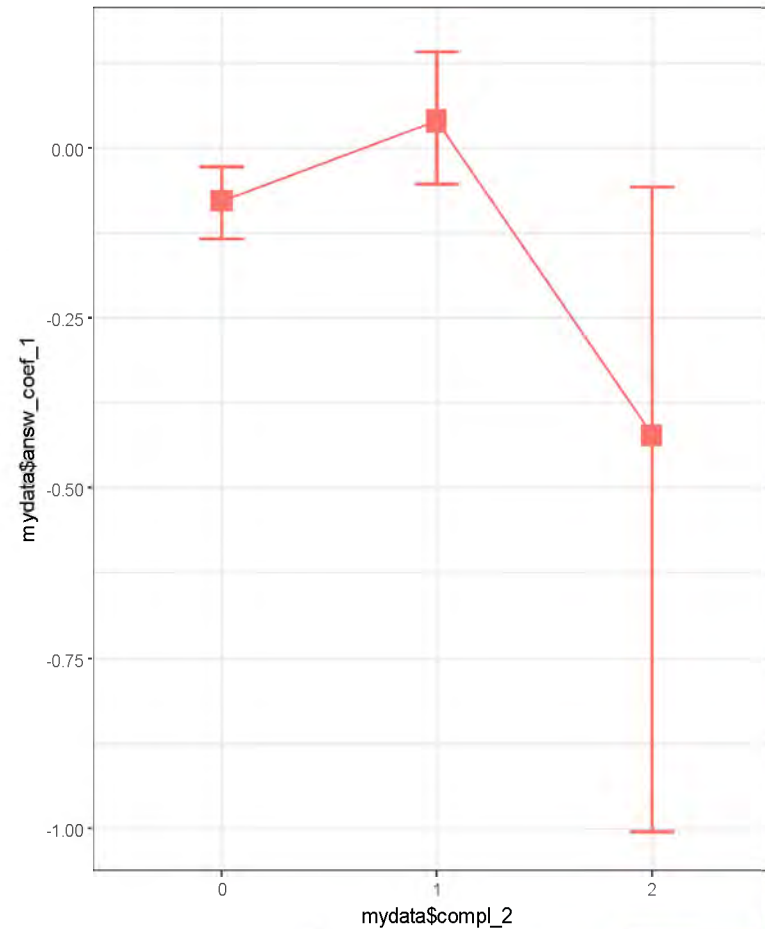
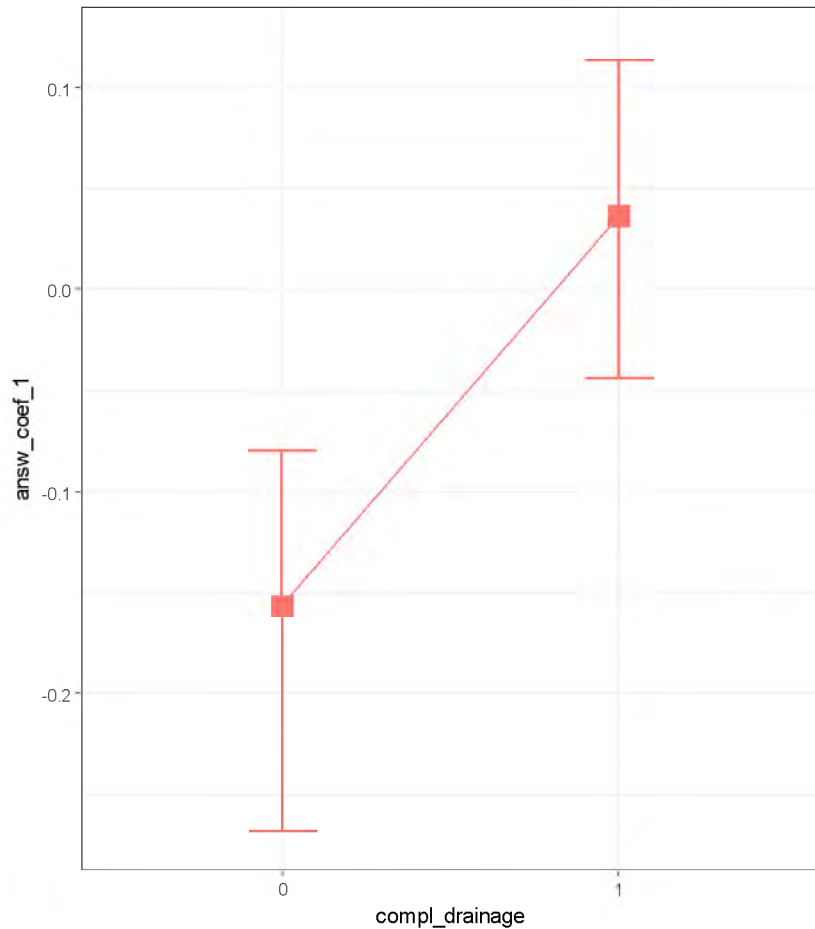
- ▶ Для определения темпа билиарной декомпрессии использована формула ответа желтухи на дренирование, предложенная в 1978 г. Т. Shimizu и К. Yoshida, которые обнаружили, что график снижения общего билирубина сыворотки крови соответствует не линейной функции, а функции натурального логарифма.



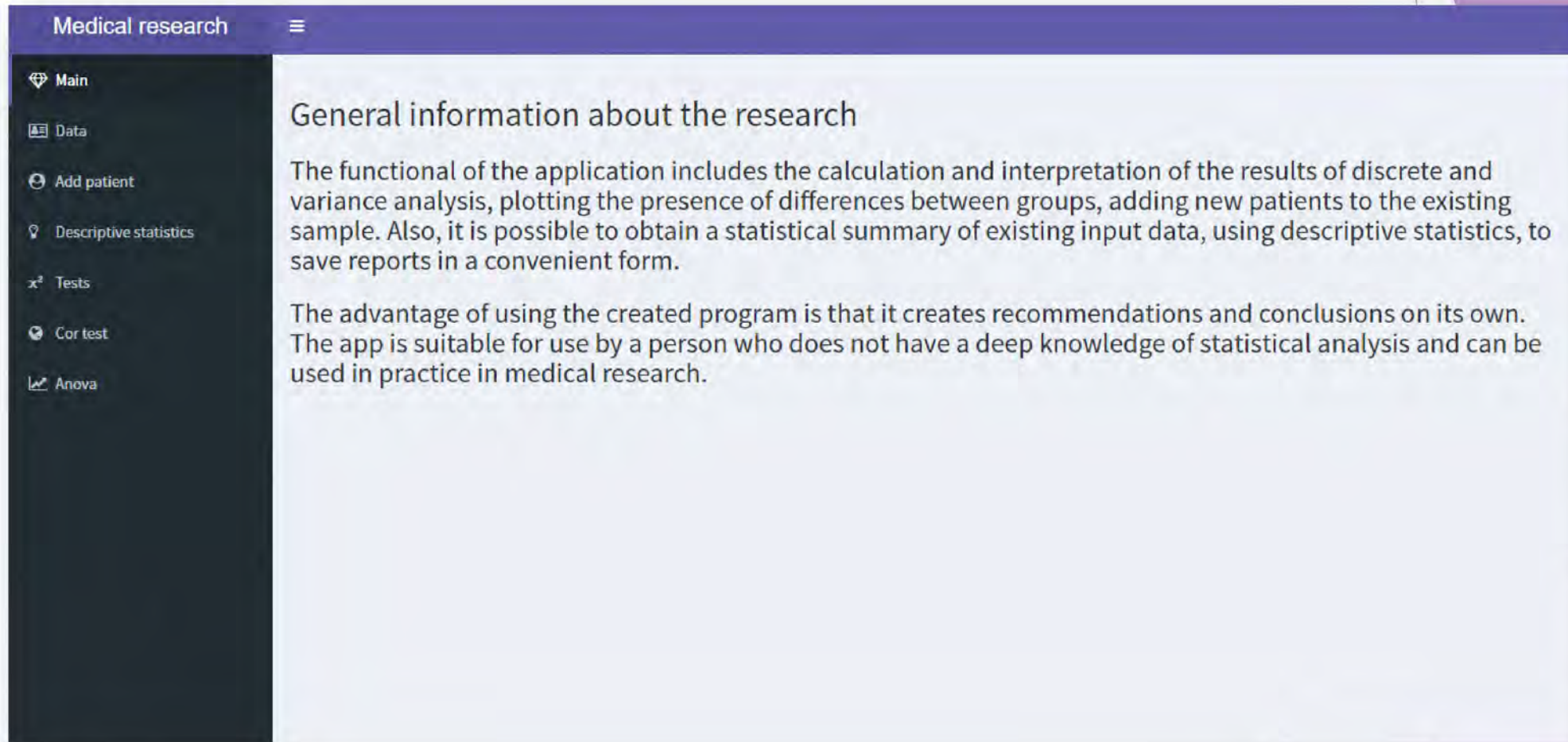
- 
- ▶ где  $B$  - ответ желтухи на декомпрессию,  $A$  - исходный уровень общего билирубина сыворотки крови,  $y$  - уровень общего билирубина сыворотки крови после декомпрессии,  $x$  - количество дней после декомпрессии.
  - ▶ Этот параметр имеет отрицательное значение и считается быстрым, если  $B < -0,09$ , средним при  $-0,09 < B < -0,05$ , затяжным (дозированным) при  $-0,05 < B < -0,025$ , рефрактерным при  $B > -0,025$ .

# Ответ желтухи на дренирование

- ▶ В результате обработки данных, было определено что ответ желтухи на дренирование является быстрым у 17 пациентов, средним - у 13, затяжным - у 9 и рефрактерным у 88.
- ▶ На рисунках изображены графики взаимосвязи коэффициента и осложнениями после этапов лечения.



# Разработка web-приложения





# Ввод исходных данных (обучающая выборка)

Medical research

Main

Data

Add patient

Descriptive statistics

Tests

Cor test

Anova

## The research data

Show 10 entries

Search:

	id	fio	dates	diagnosis	sex	age	level_bilial_block	time_yellow	diag_prehospital	weight	oper_risk	cholera
1	8	Пелешенко И.Г. №6006	12/12/11-16/01/13	3	муж	49	1	1	1	1	3	0
2	28	Власенко А.А №4840	29/09/14-14/09/14	3	муж	67	1	1	0	0	3	2
3	31	Дмитрусенко Т.И. №1690	21/03/14-22/04/14	3	жен	54	1	1	0	0	3	1
4	73	Ильченко О.М.	09/02/12-07/03/12	1	жен	76	2	1	0	0	3	0
5	75	Суббота Т.В.	19/05/14-09/07/14	1	жен	53	2	3	1	2	3	2
6	76	Филова Н.Ф.	13/05/13-20/05/13	3	жен	68	1	3	0	2	3	0
7	78	Мамонова П.И.	23/11/15-22/12/15	1	жен	86	2	2	0	2	3	0
8	79	Губенко О.М.	16/03/15-15/04/15	2	жен	38	1	2	0	0	3	0

# Добавление новых пациентов

Medical research

Adding information about a new patient

**Diagnosis:**  
1 - Holedocholitis

**Sex:**  
1 - Male

**Age:**  
50

**Bilial block level:**  
1 - Proximal

**Duration of jaundice:**  
1 - Acute

**Correct diagnosis at the prehospital stage:**  
0 - Yes

**Body weight:**  
0 - Normo

Main  
Data  
Add patient  
Descriptive statistics  
Tests  
Cor test  
Anova

# Оценивание связи между факторами

Medical research

- Main
- Data
- Add patient
- Descriptive statistics
- Tests
- Cor test
- Anova

On this page you can see results of dependence between numeric values

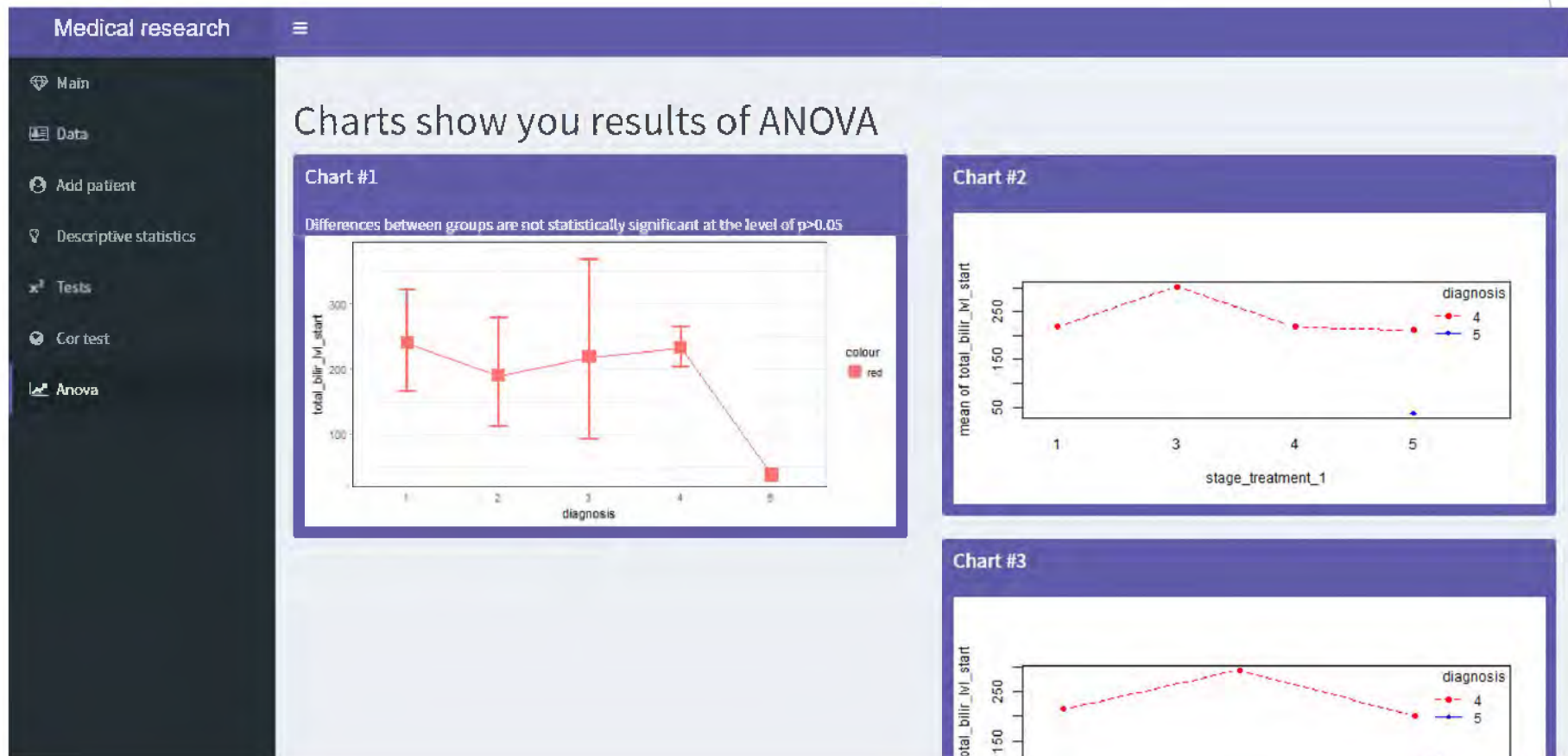
## Table of dispersion and covariation

Show 10 entries Search:

	mydata_num	days_common	days_before_1	drainage_dur	days_after_2	total_bilir_lv_start	direct_bilir_lv_start
mydata_num	134.687539057618	-51.5816772903387	-10.2844019497563	-16.3338957630296	-22.2421092452729	127.455843019623	80.99274348059
days_common	-51.5816772903387	296.540057492813	23.4667541557305	66.5777402824647	194.293935914261	-371.564954380702	-310.3132020997
days_before_1	-10.2844019497563	23.4667541557305	21.4028246469191	9.82395950506187	-5.90788651418573	-114.467166604174	-94.85204687649
drainage_dur	-16.3338957630296	66.5777402824647	9.82395950506187	61.1287339082615	-4.85055997464603	19.7126859142607	-26.48726335678
days_after_2	-22.2421092452729	194.293935914261	-5.90788651418573	-4.85055997464603	216.141682554748	-310.051551312782	-204.1742051380
total_bilir_lv_start	127.455843019623	-371.564954380702	-114.467166604174	19.7126859142607	-310.051551312782	23451.8942744657	15275.43958342
direct_bilir_lv_start	80.9927434805943	-310.313202099738	-94.8520468764934	-26.4872633567863	-204.174205138092	15275.4395834271	11386.07130210
coefficient_1	0.654759082985775	-7.26960659162257	-1.3356668061323	-1.53332646882025	-3.95353575472169	80.7915132606928	50.27264205361
total_bilir_lv_final	107.367747708007	-473.362733408324	-93.0668814927546	-168.555017975694	-271.617679978159	17541.4228629068	11055.70099767
direct_bilir_lv_final	34.1808217722785	-319.281029871266	-64.7919636148423	-126.10404173743	-172.030032150327	11209.2781238742	8173.985659229

Showing 1 to 10 of 11 entries Previous 1 2 Next

# Визуализация результатов анализа



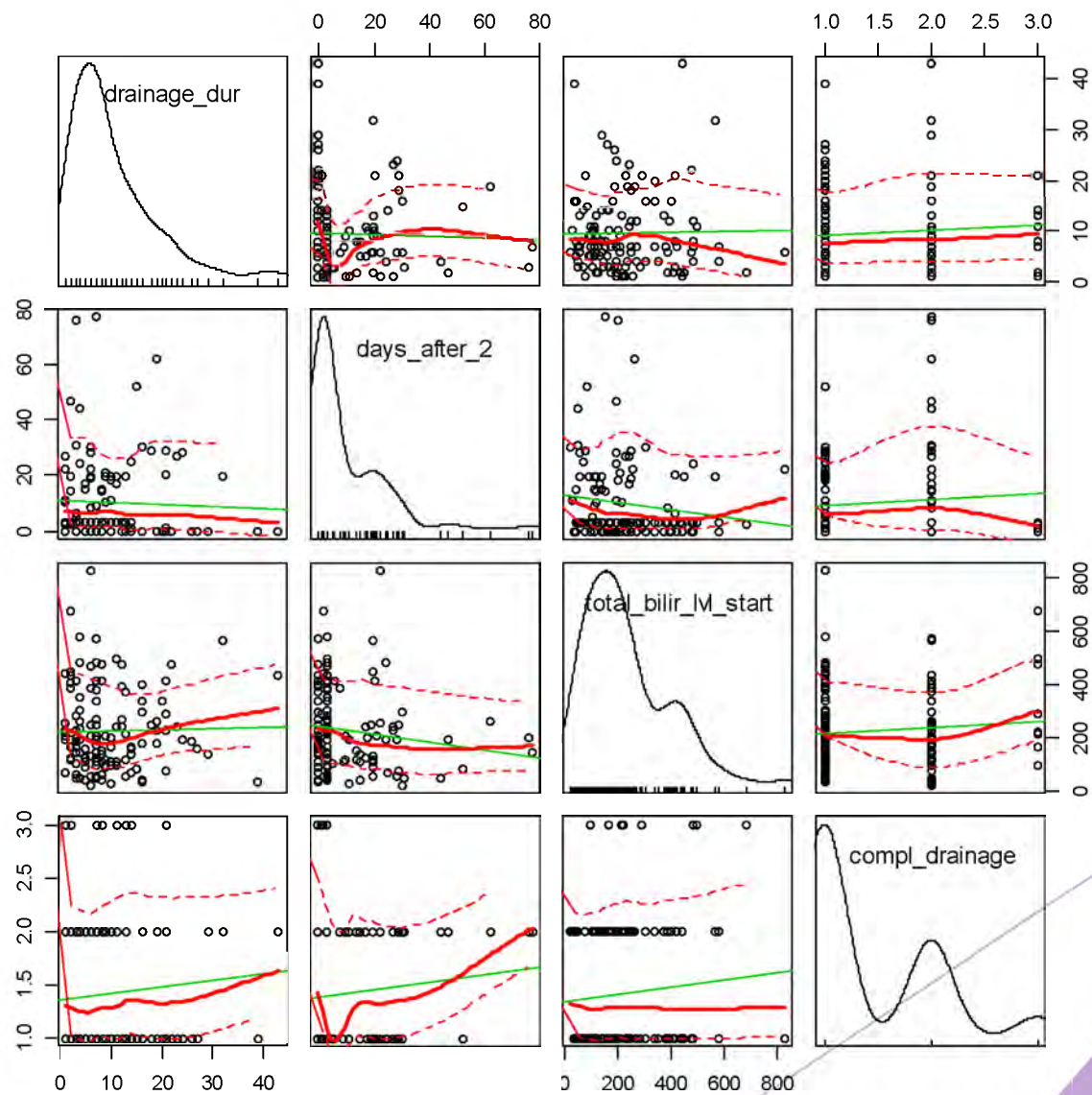
# Функция классификации пациентов на основе дискриминантного анализа

## ► Коэффициенты классификации

	LD1	LD2	LD3	LD4
Уровень билварного блока	1.083725229	0.4467749662	0.090667381	-0.287467697
Операционный риск	1.593731070	0.6974110821	0.711058023	-0.583224756
Количество койко-дней общее	0.011961383	0.1035938549	-0.055797108	-0.024428639
Количество койко-дней после 1 этапа	-0.051125210	0.0012365458	0.054974058	-0.003790320
Длительность дренирования	-0.085908779	-0.0369808973	0.093797767	0.073024578
Количество койко-дней перед 2 этапом	0.017228188	-0.1033551957	0.020450821	0.049304588
Общий уровень исходного билирубина	-0.002897815	0.0003142019	-0.004314542	-0.009477009
Осложнения после дренирования 1	1.444581540	0.4028703723	1.187926669	0.346382933
Осложнения после дренирования 2	-0.030785452	0.9215661451	-2.078132841	2.678499535
Общий уровень конечного билирубина	0.003131592	-0.0024784509	0.006333952	0.010501991
Осложнения после 2 этапа лечения 1	0.762438714	0.5174654282	-0.386788329	-0.388711640
Осложнения после 2 этапа лечения 2	1.058399563	-0.2623565511	0.409915886	-0.442516392
Ответ желтухи на дренирования	-2.389471320	-0.6153943934	-0.6153943934	-4.183214503

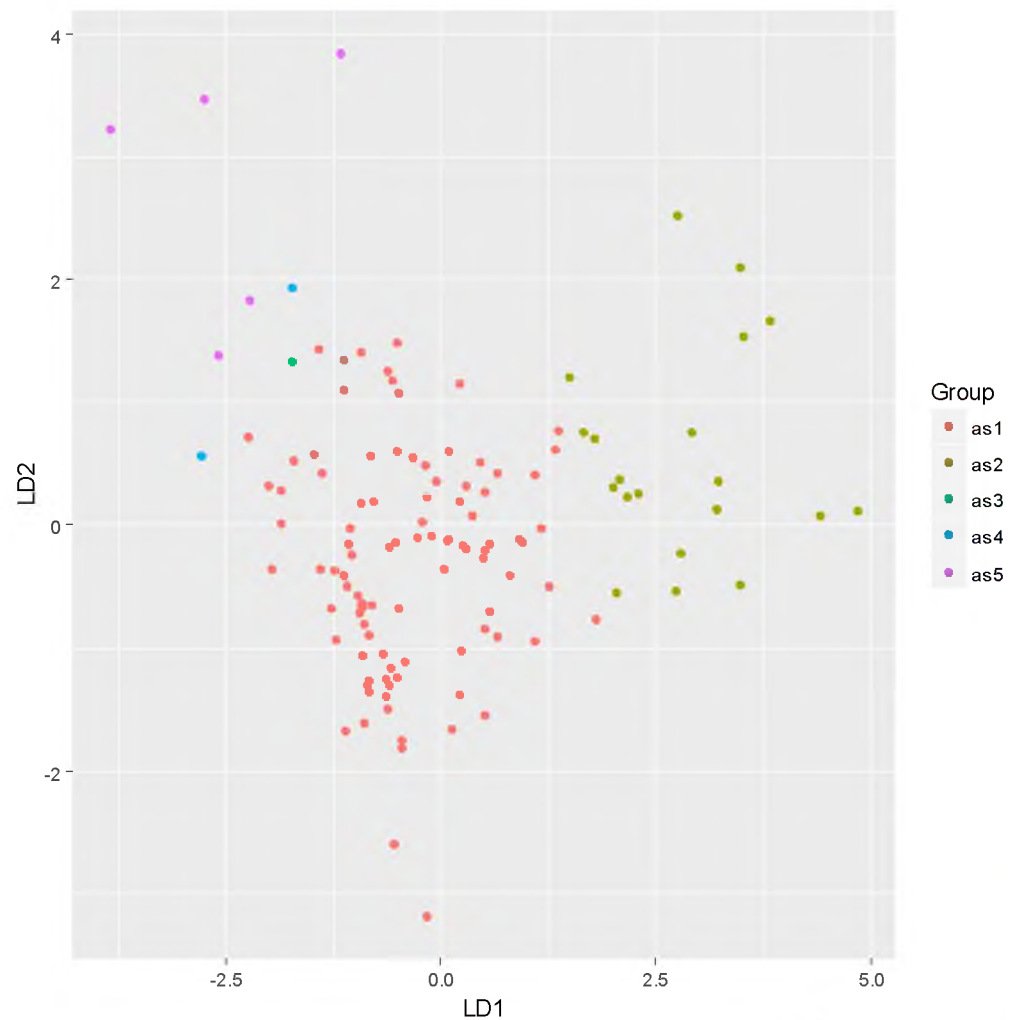
# Функция классификации пациентов на основе дискриминантного анализа

► График классификации



# Функция классификации пациентов на основе дискриминантного анализа

- ▶ График рассеивания



# Функция классификации пациентов на основе дискриминантного анализа

Точность модели выражена двумерной таблицей, где отображено сколько наблюдений получилось точно классифицировать, а сколько - нет.

	1	2	3	4	5
1	89	3	2	2	2
2	3	16	0	0	0
3	1	0	1	0	0
4	0	0	0	2	0
5	1	0	0	0	4

Как мы видим, по результатам не все наблюдения были классифицированы верно, но ошибка составляет всего 11%. То есть можно сказать, что дискриминантные функции хорошо разделяют классы.



# Результаты классификации новых пациентов

- ▶ Точность модели выражена двумерной таблицей, где отображено сколько наблюдений получилось точно классифицировать, а сколько - нет.

# Выводы

- ▶ Преимуществом использования созданной программы является то, что она самостоятельно создает рекомендации и выводы.
- ▶ Приложение пригодно для использования человеком, который не имеет глубоких познаний в области статистического анализа и может быть использовано на практике в медицинских исследованиях.