

Университет ИТМО

Практическая работа №2
по дисциплине «Визуализация и моделирование»

Автор: Никулина Ксения Григорьевна

Поток: ВИМ 1.1

Группа: К3221

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

1 Heart Failure Prediction

Сердечно-сосудистые заболевания являются причиной смерти номер 1 во всем мире, забирая в среднем 17.9 миллионов жизней в год, что составляет 31% от всех смертей по всему миру. Остановка сердца - результат сердечно-сосудистых заболеваний и данный датасет содержит 12 показателей, которые можно использовать для предсказания смертности от остановки сердца.

Название столбца	Описание данных	Тип данных	Шкала
Возраст	Возраст пациента	Integer	Интерв
Анемия	Снижение гемоглобина	Binary	Относит
Креатинкиназа	Уровень фермента	Integer	Интерв
Диабет	Наличие диабета	Binary	Относит
Фракция выброса	% крови при сокращении	Integer	Интерв
Высокое кровяное давление	Гипертония	Binary	Относит
Тромбоциты	Тромбоциты в крови	Float	Интерв
Креатинин	Уровень креатинина в крови	Float	Интерв
Натрий	Уровень натрия в крови	Integer	Интерв
Пол	Пол пациента	Binary	Относит
Курение	Курит ли пациент	Binary	Относит
Время	Кол-во дней наблюдения	Integer	Интерв
Смерть	Умер ли пациент	Binary	Относит

1. Визуализируем график распределения количества осматриваемых по каждому возрасту. Можем заметить, что минимальный возраст осматриваемого пациента равен 40 годам, а максимальный возраст - 95 лет. Наибольшее количество осматриваемых людей было в возрасте 60 лет

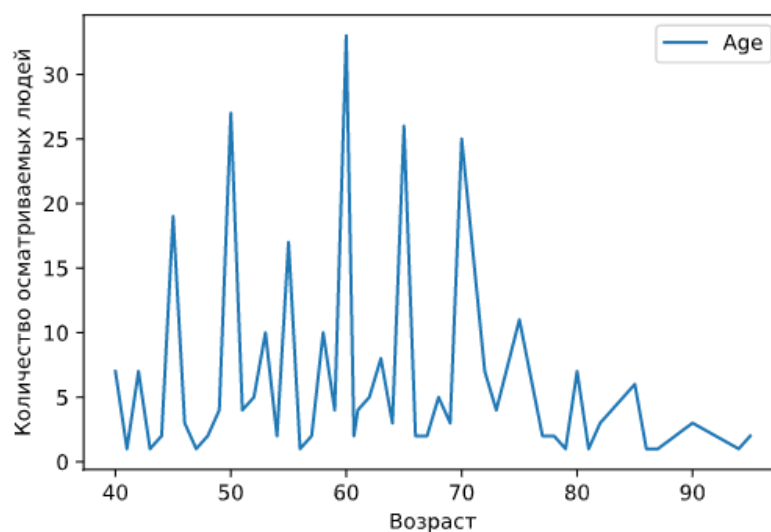


Рис. 1 график распределения

2. Визуализируем в виде графика количество исследуемых женщин и мужчин в сравнении с общим количеством осматриваемых пациентов. Можно заметить что мужчин и женщин примерно одинаковое количество. Из этого можно сделать вывод, что исследование не было направлено на какой то определенный пол.

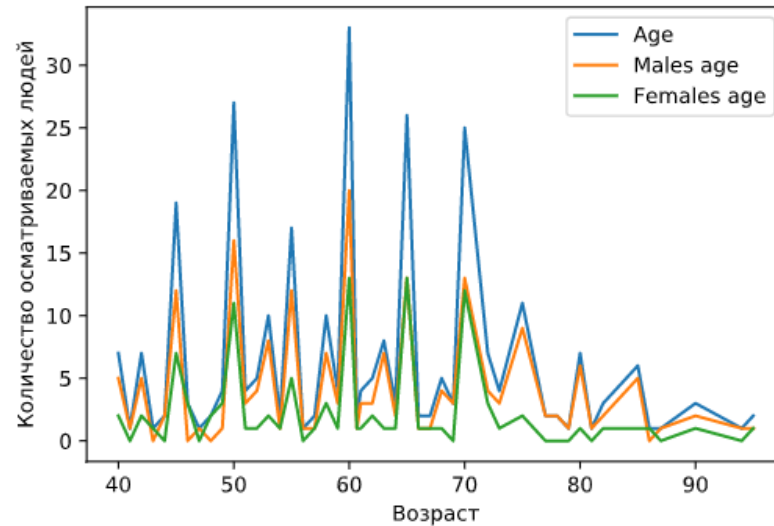


Рис. 2 график распределения

3. Подтверждаю то, что написала выше при помощи диаграммы. То, что женщин меньше можно объяснить недостатком данных и при большем датасете их количество бы сравнялось

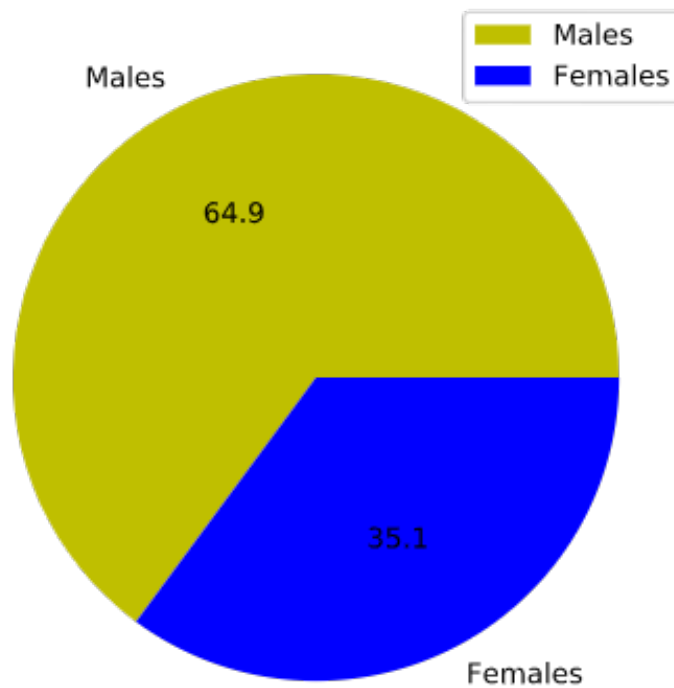


Рис. 3 диаграмма распределения

4. Визуализируем в виде диаграммы количество людей, попадавших в больницу с анемией и без нее. Анемия — это патологическое состояние, которое характеризуется пониженным уровнем эритроцитов и/или гемоглобина в крови, из-за чего ухудшается передача кислорода в ткани и возникает гипоксия, т. е. кислородное голодание тканей. Можно сделать вывод, что люди с анемией попадают в больницу с сердечной недостаточностью реже, чем без нее

```
anaemia_df_a = df.loc[df["anaemia"] == 1]
a_num_yes = anaemia_df_a.shape[0]

anaemia_df_a = df.loc[df["anaemia"] == 0]
a_num_no = anaemia_df_a.shape[0]

anaemia_pie = pd.DataFrame({"": [a_num_yes, a_num_no]},
                           index=["People with anaemia", "People with no anemia"])
anaemia_pie.plot.pie(y="",
                    colors=["pink", "olive"],
                    autopct="%.1f",
                    fontsize=10,
                    figsize=(5, 5));
```

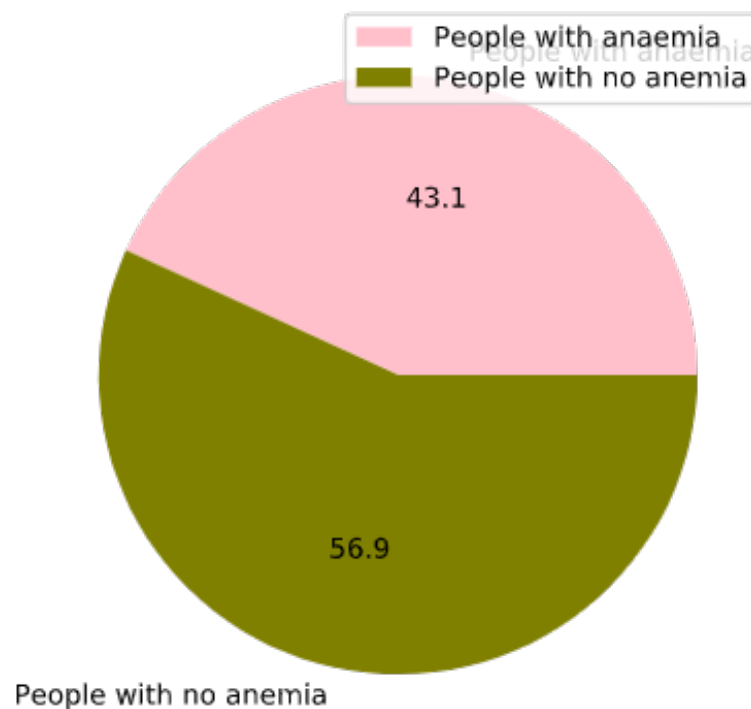


Рис. 4 диаграмма распределения

5. Посмотрим, как колеблется число респондентов в группе интерквартильного размаха - содержащей наиболее релевантные данные

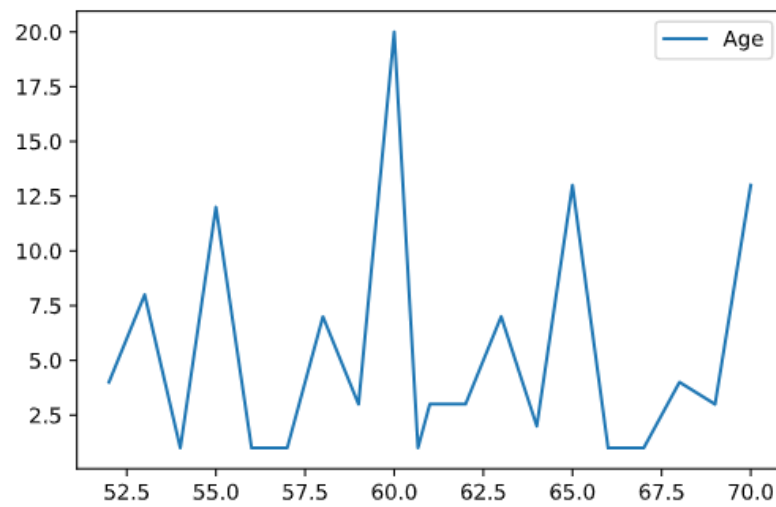


Рис. 5 график распределения

6. Визуализируем в виде солбчатой диаграммы количество живых и умерших мужчин в зависимости о возраста. Можно заметить, что после 70 при попадании в больницу шанс на выживания 50 на 50, а до 70 лет при попадании в больницу выжили большее количество пациентов

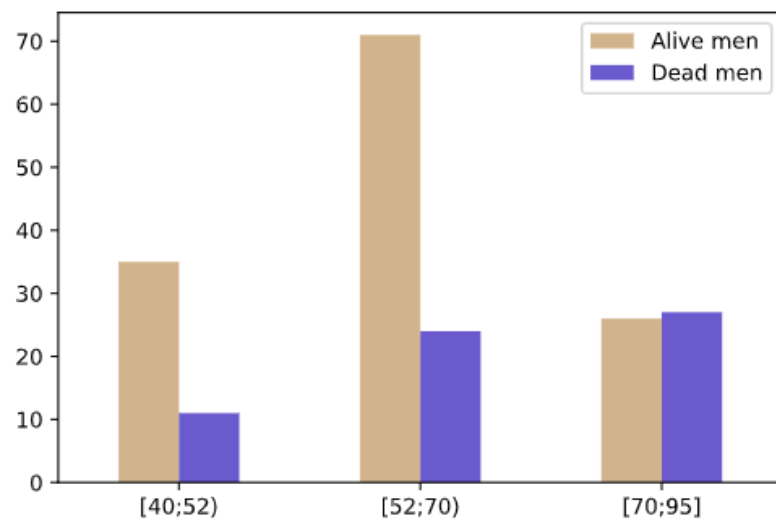


Рис. 6 диаграмма распределения

7. Визуализируем в виде солбчатой диаграммы количество живых и умерших женщин в зависимости о возраста. Можно заметить, что после 70 при попадании в больницу шанс на выживания 50 на 50, а до 70 лет при попадании в больницу выжили большее количество пациенток

```
falive = [female_df.query("age < @q1 & DEATH_EVENT == 0").shape[0],  
          female_df.query("@q1 <= age < @q3 & DEATH_EVENT == 0").shape[0],  
          female_df.query("age >= @q3 & DEATH_EVENT == 0").shape[0]]  
  
fdead = [female_df.query("age < @q1 & DEATH_EVENT == 1").shape[0],  
         female_df.query("@q1 <= age < @q3 & DEATH_EVENT == 1").shape[0],  
         female_df.query("age >= @q3 & DEATH_EVENT == 1").shape[0]]  
  
dead_or_alive = pd.DataFrame({"Alive women": falive, "Dead women": fdead}, index=index)  
dead_or_alive.plot.bar(color=["teal", "silver"],  
                        rot=0);
```

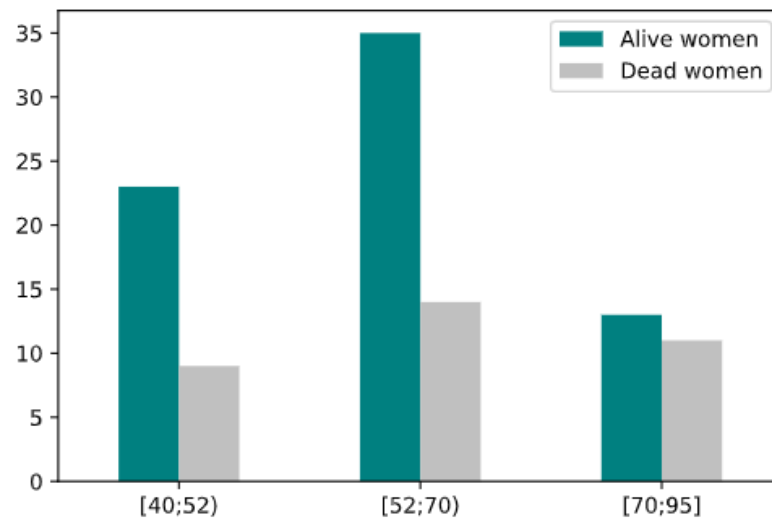


Рис. 7 диаграмма распределения

8. В виде солбчатой диаграммы можно наглядно увидеть количество мужчин, которые курили/не курили и у которых есть/нет анемия с тем количеством дней, сколько они были под наблюдением. Количество мужчин, которые не курили и у которых нет анемии превышает количество других при долгом наблюдении. Пациенты без вредных привычек и без анемии проводят дольше времени под наблюдением и это может быть результатом того, что другая половина мужчин умирает.

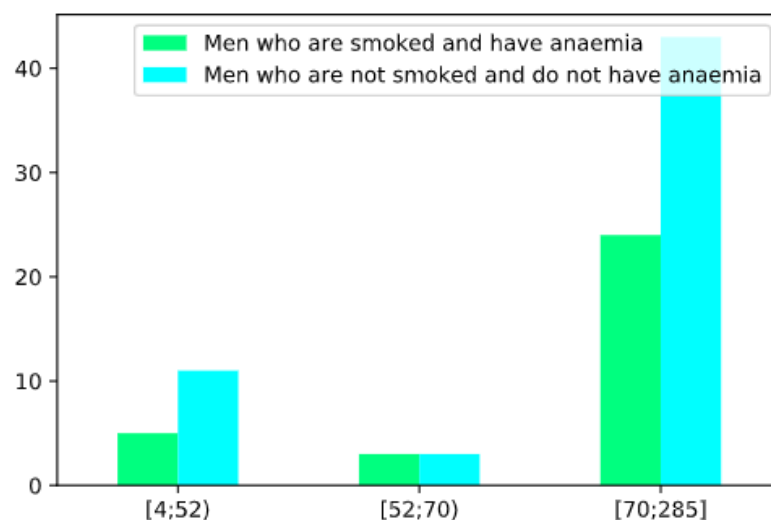


Рис. 8 даиграмма распределения

9. В виде солбчатой диаграммы можно наглядно увидеть количество женщин, которые курили/не курили и у которых есть/нет анемия с тем количеством дней, сколько они были под наблюдением. Количество женщин, которые не курили и у которых нет анемии превышает количество других при долгом наблюдении (скорее всего это связано с тем, что другие пциентки умерли во время наблюдения). Из этой диаграммы можно сделать вывод, что у женщины без вредных привычек и без анемии шанс выжить намного больше, чем у курящей женщины с анемией.

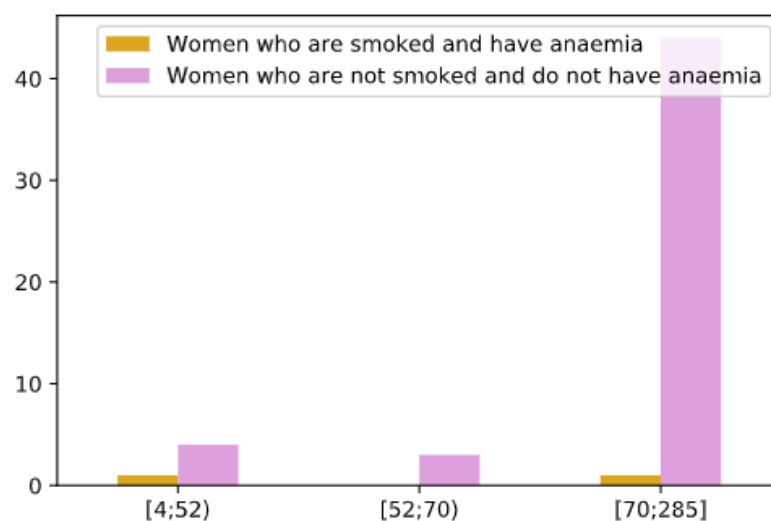


Рис. 9 даиграмма распределения

10. Визуализировали связь между количеством дней наблюдения, курящими/не курящими женщинами и с их выживаемостью. При малом количестве дней наблюдения курящих и живых женщин практически нет из этого можно сделать вывод, что они либо оказались в столбце "курящие и мертвые" или, дожили до 70+ дня наблюдения, или умирали сразу. Количество умерших не курящих женщин в каждом диапазоне примерно одинаковое и скорее всего это связано с тем, что у них могли быть и другие заболевания/привычки, которые вызывают сердечную недостаточность. При большом количестве дней наблюдений количество курящих/не курящих пациенток сильно разнится, скорее всего это связано с тем, что мало курящих женщин дожили до акого количества дней наблюдения и умерали от сердечной недостаточности раньше либо сразу.

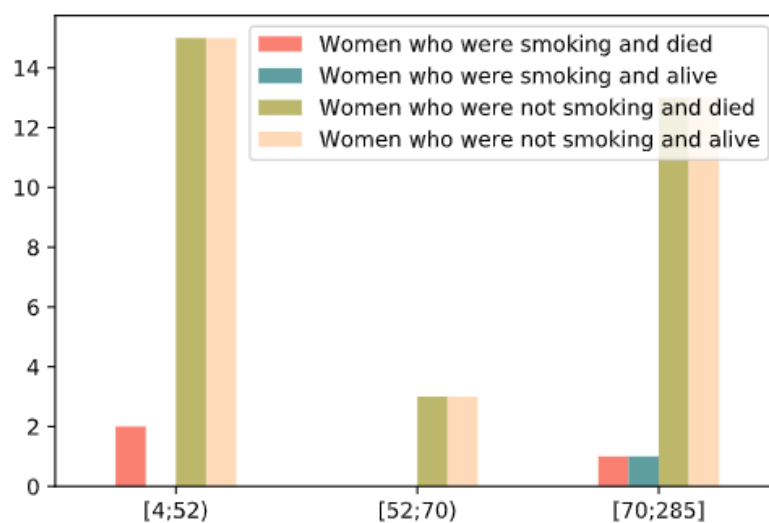


Рис. 10 даиграмма распределения

11. Визуализировали связь между количеством дней наблюдения, курящими/не курящими мужчинами и с их выживаемостью. При малом количестве дней можно заметить, что мужчин, которые курили и выжили совсем нет из этого можно сделать вывод, что они скорее всего оказались в графе "курящие и мертвые" либо они были под наблюдением больше, чем 51 день. Удивительно, но при большем количестве дней наблюдения больше всего курящих и живых мужчин. Очень мало курящих и умерших (можно сделать вывод, что мало пациентов дожило до такого количества дней наблюдения). Примерно одинаково количество мужчин, которые не курили во всех промежутках времени.

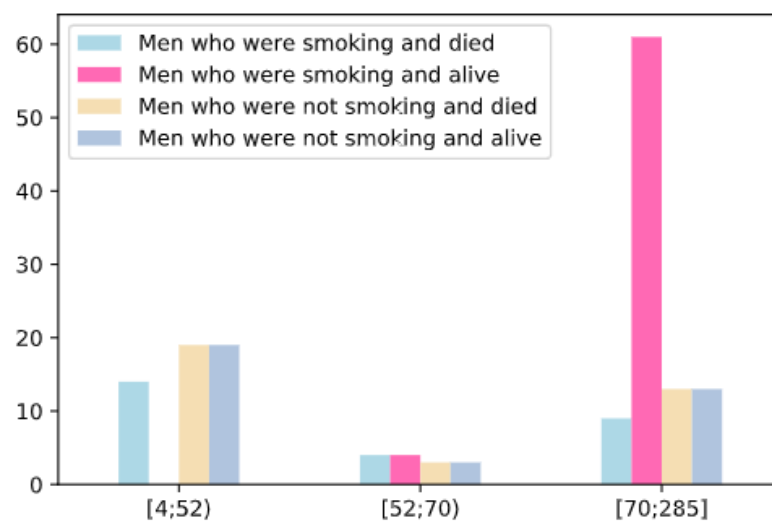


Рис. 11 даиграмма распределения