**Отчет**

**о результатах исследования эффекта синхронизации максимумов процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) с максимумами геомагнитной возмущенности (по данным Кр-индекса)**

Рассматриваются данные с 01.03.2017 по 30.04.2017 процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) трех групп пациентов без нагрузки: в каждой группе по 6 пациентов, которые кратко обозначаются *i\_j* (*i* – номер группы, *j* - номер пациента в группе)

В качестве количественной меры близости между рядом данных пациента и рядом значений Кр-индекса выбрано среднее значение расстояний (в днях) от каждого дня максимума данных пациента до ближайшего дня максимума Кр-индекса.

**Методика выявления эффекта синхронизации максимумов процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) с максимумами геомагнитной возмущенности (по данным Кр-индекса).**

1. По данным Кр-индекса с 01.03.2017 по 30.04.2017 определяются дни максимумов, которые отмечаются значениями 1, остальные дни отмечаются значениями 0.

Для этого рассматривается список Кр-значений с 28.02.2017 по 01.05.2017:

[15.0, 34.0, 33.0, 25.0, 27.0, 25.33, 28.0, 23.33, 18.67, 22.0, 17.0, 8.67, 13.33, 4.0, 8.0, 11.33, 8.67, 4.67, 3.0, 2.67, 2.0, 27.67, 30.0, 15.33, 10.67, 4.67, 6.0, 37.67, 30.33, 27.0, 26.33, 31.67, 23.67, 14.33, 7.67, 23.67, 15.33, 12.33, 16.0, 23.33, 20.67, 9.0, 18.33, 8.33, 7.33, 20.33, 10.33, 6.33, 6.33, 10.67, 20.67, 30.33, 18.67, 39.33, 35.67, 26.67, 19.0, 16.33, 12.33, 10.33, 11.67, 10.33] – всего 62 дня

В результате получен список дней максимумов Кр-значений с 01.03.2017 по 30.04.2017:

[1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1] – всего 60 дней

1. По данным коэффициента симметрии зубца Т электрокардиограммы пациента определяются дни максимумов со 02.03.2017 по 29.04.2017, которые отмечаются значениями 1, остальные дни отмечаются значениями 0.

Например, для списка значений коэффициента симметрии зубца Т электрокардиограммы первого пациента с 01.03.2017 по 30.04.2017:

[0.8, 0.6, 0.69, 1.2, 1.71, 0.65, 0.66, 0.73, 0.75, 0.78, 0.76, 0.55, 0.56, 0.61, 0.63, 0.82, 0.8, 0.9, 1.0, 0.8, 0.34, 0.47, 0.61, 0.74, 0.59, 0.6, 0.62, 0.52, 0.59, 0.57, 0.55, 0.53, 0.63, 0.72, 0.64, 0.55, 0.42, 0.54, 0.66, 0.66, 0.54, 0.59, 0.69, 0.61, 0.58, 0.63, 0.68, 0.8, 0.61, 0.63, 0.84, 0.84, 0.76, 0.67, 0.58, 0.68, 0.65, 0.63, 0.63, 0.91]

получен список максимумов значений коэффициента симметрии зубца Т электрокардиограммы первого пациента со 02.03.2017 по 29.04.2017 :

[0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0].

1. По двум рядам данных, полученных в пунктах 1,2 строится распределение расстояний (которые измеряются числом дней) от дней максимумов данных пациента до ближайшего дня максимума Кр-индекса:
2. объем выборки равен числу максимумов данных пациента,
3. значение варианта равно 0, если день максимума данных пациента совпадает с днем максимума Кр-индекса,
4. значение варианта больше 0, если день максимум данных пациента наступает позже дня максимума Кр-индекса,
5. значение варианта меньше 0, если день максимума данных пациента наступает раньше дня максимума Кр-индекса.

Например, для первого пациента получаем следующее распределение расстояний:

[1, -1, 1, 1, -3, 2, 0, 2, -2, -1, 0, 1, 1, 3, -3, 0, 1, -1, 3]

## Анализ полученных данных

## Для статистической оценки параметров анализируемых выборок определяем вид их распределения. При построении теоретической модели распределений вероятности полученных выборок используются следующие методы:

## *графические методы* проверки визуальной нормальности с помощью построения гистограмм и специальных диаграмм полученных данных, и качественной оценки того, что данные были получены из нормального распределения;

* *статистические методы* применения тестов, которые вычисляют статистику по данным и количественно определяют, насколько вероятно, что данные были получены из нормального распределения.

Основные понятия теории вероятностей и математической статистики взяты из работ [1],[2]. Все расчеты получены с помощью компьютерных программ на языке программирования Python.

## Проверка визуальной нормальности

Для качественной проверки нормальности анализируемых выборок рассматривались такие графики данных, как гистограммы, непараметрический метод ядерной оценки плотности (KDE или Kernel Density Estimation) и графики квантилей ([Q-Q plots](https://en.wikipedia.org/wiki/Q-Q_plot) или Quantile-Quantile plots). На таких графиках изображаются квантили двух распределений - эмпирического (т.е. построенного по анализируемым данным) и теоретически ожидаемого стандартного нормального распределения. При нормальном распределении проверяемой выборки точки на графике квантилей должны выстраиваться в прямую линию, исходящую под улом 45 градусов из левого нижнего угла графика.

Результаты такого визуального анализа в нашем случае показывают, что все рассматриваемые распределения достаточно хорошо аппроксимируются графиком нормального распределения.

Например, гистограммы, ядерная оценка плотности и графики квантилей распределения данных первого пациента представлены на рис. 1,2,3.

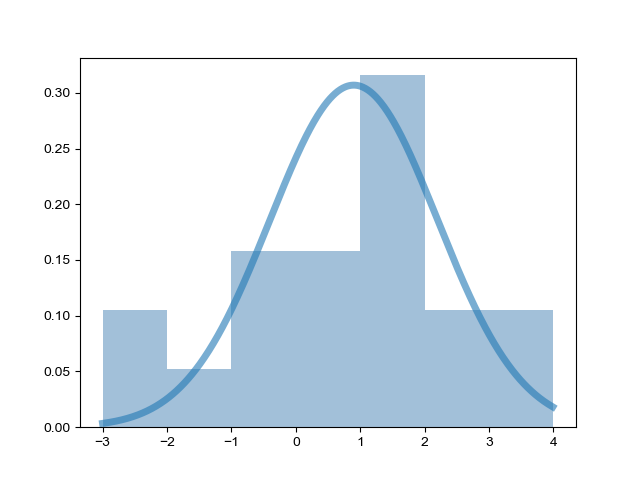
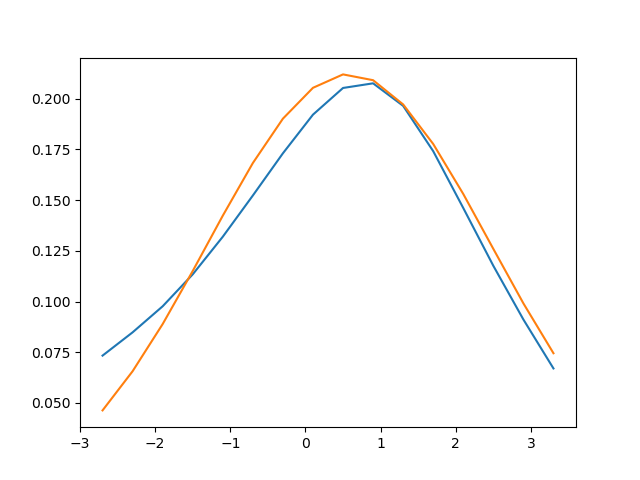
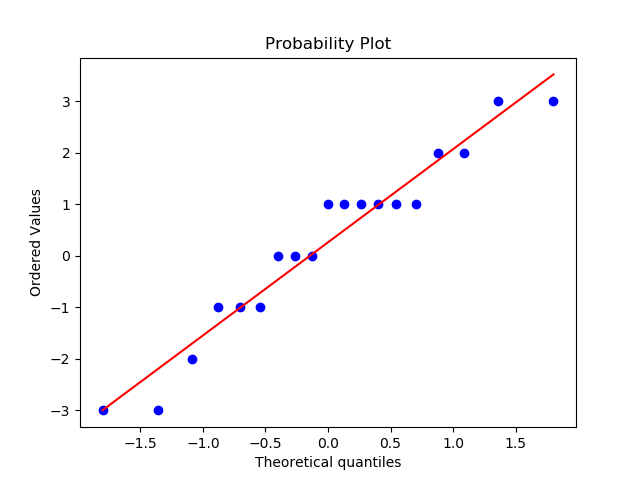


Рис.1. Гистограмма распределения данных первого пациента и график плотности нормального распределения (синий график)

Рис.2. Ядерная оценка плотности распределения данных первого пациента (синий график) и график плотности нормального распределения (красный график)



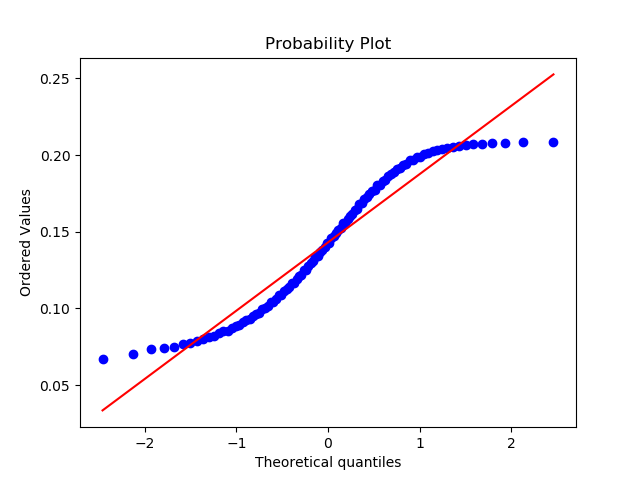


Рис.3. Графики квантилей распределения данных первого пациента и ядерной оценки плотности распределения этих данных

**Тесты статистической нормальности**

Для проверки нормальности рассматриваемых выборок применялись известные статистические тесты, специально разработанные для проверки нормальности распределения данных (см. [2],[4]). В общем виде проверяемая при помощи таких тестов нулевая гипотеза H0 формулируется следующим образом: "Анализируемая выборка происходит из генеральной совокупности, имеющей нормальное распределение". Если получаемая при помощи того или иного теста вероятность ошибки *p* оказывается меньше некоторого, заранее принятого уровня значимости (в нашем случае, ), нулевая гипотеза отклоняется.

Интерпретация теста.

Для анализируемой выборки каждый тест вычисляет следующие значения:

* *Статистика* - это значение, вычисленное по тесту, которое может быть интерпретировано в контексте теста путем сравнения его с критическими значениями из распределения тестовой статистики.
* *p-значение* - это значение, которое используется для интерпретации теста с целью проверки, была ли выборка получена из нормального распределения.

Каждый тест вычисляет специфическую для теста статистику, с помощью которой интерпретируется результат проведенного анализа. С другой стороны, *p*-значение используется для быстрой и точной интерпретации статистики в практических приложениях.

В тестах предполагается, что образец был выведен из нормального распределения. Технически это называется нулевой гипотезой, или гипотезой H0. Пороговый уровень выбирается как уровень значимости (в нашем случае, 5% или 0,05), который используется для интерпретации значения *р*.

При реализации этих тестов значение *p* интерпретируется следующим образом:

* если , то гипотеза H0 отклоняется, т.е. делается вывод, что распределение не нормальное,
* если , то гипотеза H0 принимается, т.е. делается вывод, что распределение нормальное.

В нашем случае для статистического анализа рассматриваемых выборок применялись такие известные статистические тесты нормальности, как тест Шапиро-Вилка, тест K2 Д'Агостино, тест Андерсона-Дарлинга и тест Колмогорова-Смирнова (см. [2], [4]).

Результаты такого тестирования статистической нормальности показывают, что почти для всех рассматриваемых распределений (16 распределений из 18) по всем тестам принимается гипотеза H0 о нормальности распределения и только для двух распределений гипотеза H0 по двум тестам принимается, а по двум тестам отклоняется.

Например, тестирование статистической нормальности распределения данных первого пациента показывает, что для всех тестов принимается гипотеза H0 о нормальности этого распределения, что подтверждается соответствующим выводом компьютерной программы:

Shapiro-Wilk Test

Statistics=0.940, p=0.260

Sample looks Gaussian (fail to reject H0)

D'Agostino and Pearson's Test

Statistics=0.633, p=0.729

Sample looks Gaussian (fail to reject H0)

Anderson-Darling Test

Statistic: 0.467

15.000: 0.505, data looks normal (fail to reject H0)

10.000: 0.575, data looks normal (fail to reject H0)

5.000: 0.690, data looks normal (fail to reject H0)

2.500: 0.804, data looks normal (fail to reject H0)

1.000: 0.957, data looks normal (fail to reject H0)

Kolmogorov-Smirnov Test

Kolmogorov test for normality: 0/1000 = 0.00 rejects at rejection level 0.05

**Оценка параметров анализируемых распределений данных**

Согласно [1],[2], состоятельной, несмещенной и эффективной оценкой математического ожидания анализируемого распределения является в*ыборочная средняя*

,

состоятельной, несмещенной оценкой дисперсии анализируемого распределения является *исправленная выборочная дисперсия*

где

Значение называется *средним квадратическим отклонением* (или *стандартным отклонением*) анализируемого распределения.

Согласно теории интервальных оценок (см. [1],[2]), в нашем случае построение доверительного интервала для генеральной средней по малой выборке осуществляется с помощью *t*-распределения Стьюдента с степенями свободы по следующей формуле доверительной вероятности

где , – значение критерия Стьюдента (табулировано) и - *доверительная вероятность* или *надежность оценки* (в нашем случае ).

Таким образом, генеральная средняя распределения расстояний от дней максимумов пациента до ближайших дней максимумов *Kp*-значений с вероятностью 0,95 покрывается доверительным интервалом

Например, для первого пациента получены следующие точечные оценки генеральной средней и дисперсии распределения расстояний от максимумов пациента до ближайших максимумов *Kp*-значений:

выборочное среднее = 0.2631578947368421,

стандартное отклонение = 1.7119427333990098,

и генеральная средняя с вероятностью 0,95 покрывается доверительным интервалом

(-0.5842105263157897, 1.1105263157894738) радиуса 0.8473684210526318.

1. Результаты статистического анализа полученных распределений для всех 18 пациентов приведены в сводных таблицах 1,2:
2. в таблице 1 приведены результаты тестирования анализируемых выборок на нормальность распределений;
3. в таблице 2 для каждого пациента приведены распределение расстояний от дней максимумов процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) пациентов до ближайших дней максимумов геомагнитной возмущенности (по данным Кр-индекса), выборочное среднее и стандартное отклонение этого распределения расстояний, а также доверительный интервал оценки генеральной средней этого распределения с доверительной вероятностью 0,95.

**Заключение**

Полученные результаты статистического анализа данных показывают, что среднее значение расстояний от дней максимумов процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) пациентов до ближайших дней максимумов геомагнитной возмущенности (по данным Кр-индекса) незначительно отличается от значения 0, соответствующее совпадению дней максимумов процессов реполяризации миокарда желудочков пациентов с днями максимумов Кр-индекса:

1. разброс средних значений этих распределений от -0.9230769230769231 до 0.46153846153846156 свидетельствует о том, что дни максимумов процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) пациентов в среднем отличаются не более, чем на один день от дней ближайших максимумов геомагнитной возмущенности (по данным Кр-индекса);
2. разброс стандартных отклонений этих распределений от 1.166596222161777 до 1.730767331432956 свидетельствует о том, что дни максимумов процессов реполяризации миокарда желудочков (по данным коэффициента симметрии зубца Т электрокардиограммы) пациентов отклоняются от их средних значений не более, чем на два дня;
3. доверительные интервалы с вероятностью 0,95 покрывают генеральные средние этих распределений и содержат значение ноль, показывающее совпадение дней максимумов процессов реполяризации миокарда желудочков пациентов с днями максимумов Кр-индекса.

Список цитируемой литературы

1. Кремер Н.Ш. Теория вероятностей и математическая статистика. – М.: ЮНИТИ-ДАНА, 2006.
2. Кобзарь А.И. Прикладная математическая статистика. – М.: ФИЗМАТЛИТ, 2006.
3. Хардле В. Прикладная непараметрическая регрессия. – М.: МИР, 1993.
4. Лемешко Б.Ю., Лемешко С.Б. Сравнительный анализ критериев проверки отклонения распределения от нормального закона \\ Метрология. 2005. № 2. С.3-23.