**ЛАБОРАТОРНАЯ РАБОТА №3**

# Доверительные интервалы
# Проверка статистических гипотез

- **Доверительный интервал** — это интервал, построенный с помощью случайной выборки из известного распределения, такой, что он содержит данный параметр с заданной вероятностью.
- **Интервальное оценивание** — построение интервала, в котором с некоторой вероятностью находится истинное значение оцениваемого параметра.

Зачем нужны доверительные интервалы?

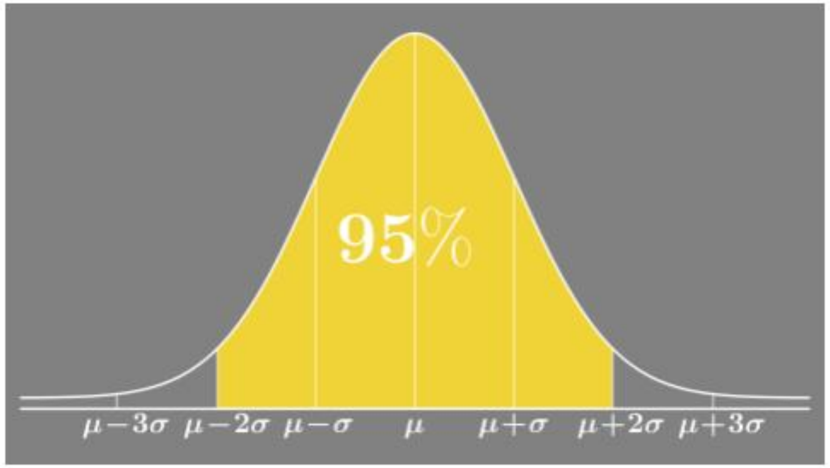
Все значения изучаемого признака, лежащие внутри доверительного интервала считаются приемлемыми, допустимыми. Если признак принимает значение лежащие вне доверительного интервала – это значение является *аномальным*.

*Выявление аномалий* – важная задача интеллектуального анализа данных, направленная на обнаружение в данных выбросов, шумов, отклонений и исключений.

# Правило двух сигм

Если случайная величина имеет нормальное распределение с математическим ожиданием $\mu$ и дисперсией $\sigma^2$ $(X \sim N(\mu, \sigma^2))$, то с вероятностью примерно 95% она принимает значение из интервала $\mu \pm \sigma^2$:

$$\mathbf{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95.$$




При решении статистических задач правила двух сигм недостаточно:
во-первых, эта оценка неточная, во-вторых, хочется строить такие оценки не только для вероятности 0.95, но и для любой другой.

# Уточнение правила двух сигм

Пусть задано число $\alpha \in (0,1)$.Тогда *квантилем* порядка $\alpha$ случайной величины $X$ называется такая величина $X_\alpha$, что:

$$\mathbf{P}(X \leq X_\alpha) \geq \alpha, \quad \mathbf{P}(X \geq X_\alpha) \geq 1 - \alpha.$$

Существуют другие эквивалентные определения квантиля. В частности, если случайная величина $X$ задана функцией распределения $F(x)$:

$$F(x) = \mathbf{P}(X \leq x),$$

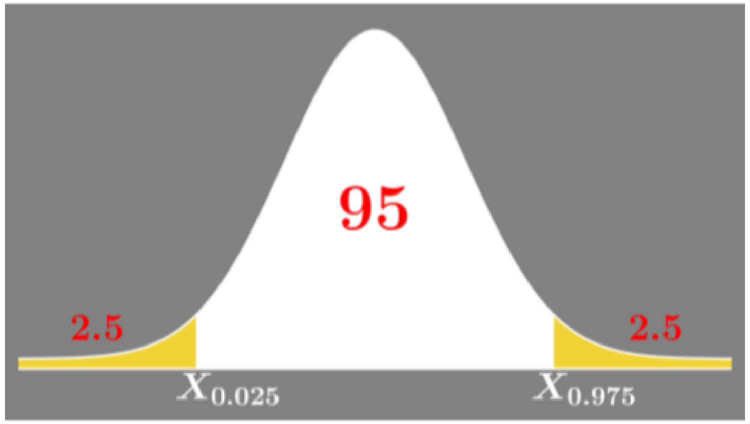
$$X_\alpha = F^{-1}(\alpha) = \inf\{x \colon F(x) \geq \alpha\},$$

то есть наименьшее $x$, для которого функция распределения $F(x) \geq \alpha$.

# Уточнение правила двух сигм

Определение квантиля можно использовать для уточнения правила двух сигм: требуется найти такие границы отрезка, что случайная величина $X$ лежит внутри него с вероятностью ровно 95%.

На рисунке показана *плотность* вероятности нормально распределённой случайной величины (*плотность* — это функция, интеграл от которой по всей числовой прямой равен 1, а по любому отрезку — вероятности попадания случайной величины в этот отрезок; интеграл — это площадь под кривой). У плотности можно выделить левый и правый «хвосты», так, чтобы их площади были равны 2.5%. Тогда площадь под центральной частью графика будет равна 95% (0.95). По определению, границы таких хвостов задаются квантилями $X_{0.025}$ и $X_{0.975}$. Искомый интервал найден: $\mathbf{P}(X_{0.025} \le X \le X_{0.975}) = 0.95.$



# Построение доверительных интервалов для среднего

Часто не достаточно построить точечную оценку среднего по выборке (выборочное среднее), и хочется понять, в каком диапазоне может меняться среднее. Именно в таких случаях используют доверительные интервалы для среднего. Рассмотрим *два способа построения доверительных интервалов*: с помощью $z$-интервала и $t$-интервала.

1) Для построения $z$-интервала необходимо знать дисперсию выборки или выдвинуть какое-то предположение о её значении: $\bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, где

$z_{1-\frac{\alpha}{2}}$ — значение верхнего квантиля уровня $\frac{\alpha}{2}$, $\frac{\sigma}{\sqrt{n}}$ — средняя ошибка выборки.

Случаи, когда известна дисперсия, очень редки, на практике ее значение практически никогда неизвестно. Пример случая, когда можно использовать $z$-интервал,— оценка работы некоторого прибора, в таких случаях обычно известна погрешность, а значит, и дисперсия.

2) В случаях, когда дисперсия неизвестна, лучше не делать ничем не подкреплённых предположений о её значении, а использовать $t$-интервал. Вместо гипотетической дисперсии в этом методе используется выборочная дисперсия $S^2$:

$$\bar{X}_n \pm t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}.$$

# Построение доверительных интервалов для доли

В таких случаях работа ведется с генеральной совокупностью, состоящей из бинарных событий. Это такие события, каждое из которых можно описать 0 или 1, или по-другому, связать с успехом или с неудачей. В жизни довольно много примеров таких событий: проигрыш или выигрыш в лотерею, покупка или не покупка товара, клик или не клик на рекомендацию. Доверительный интервал для доли можно строить на основе нормального распределения с использованием центральной предельной теоремы. Формула для такого интервала: $\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$.

Следующий метод, который очень часто используют, — это доверительный интервал Уилсона. Этот метод является некоторым улучшением предыдущего метода, которое позволяет получать качественные оценки в крайних случаях (то есть когда доля близка к 0 или 1). Формула для расчета:

$$\frac{1}{1+\frac{z^2}{n}} \left( \hat{p} + \frac{z^2}{2n} \pm z \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n} + \frac{z^2}{4n^2}} \right), \quad z \equiv z_{1-\frac{\alpha}{2}}.$$

$\hat{p}$ — частота встречаемости признака в выборке.

# Проверка статистических гипотез

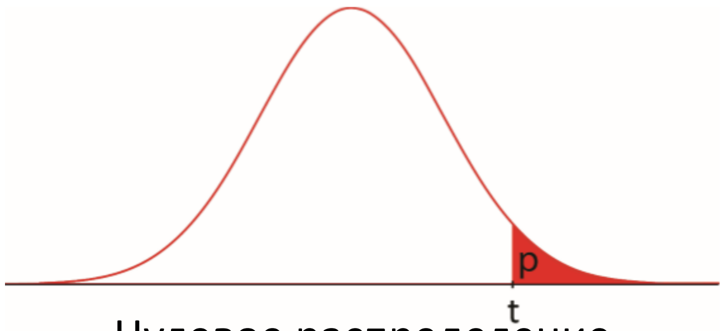
Проверка статистических гипотез — это важнейший инструмент, которым необходимо владеть в совершенстве, чтобы заниматься анализом данных.

Имеется некоторая выборка $X$ объема $n$ из неизвестного распределения $P$.

Проверяется нулевая гипотеза $H_0$ о распределении $P$ против общей альтернативы $H_1$. Это делается с помощью статистики $T$, которая является функцией от выборки. Для этой статистики известно нулевое распределение $F(x)$, то есть распределение при справедливости нулевой гипотезы.

По нулевому распределению, по его хвостам (разным, в зависимости от типа альтернативы) вычисляется достигаемый *уровень значимости – вероятность получить такое же значение статистики, какое было получено в эксперименте, или ещё более экстремальное:* $p = \mathbf{P}(T \geq t \,|\, H_0)$.

Чем ниже достигаемый уровень значимости, тем сильнее данные свидетельствуют против нулевой гипотезы в пользу альтернативы.



Нулевое распределение

Существенная особенность механизма проверки гипотез — его несимметричность относительно пары нулевая гипотеза – альтернатива. Эта особенность тесно связана с понятиями *ошибок первого и второго рода*. Нулевая гипотеза может быть либо верна, либо неверна. В результате проверки гипотезы ее можно либо принять, либо отвергнуть.

| | $H_0$ верна | $H_0$ неверна |
|---|---|---|
| $H_0$ принимается | $H_0$ верно принята | Ошибка II рода |
| $H_0$ отвергается | Ошибка I рода | $H_0$ верно отвергнута |

На главной диагонали находятся верные решения: либо принимается верная нулевая гипотеза, либо отвергается неверная нулевая гипотеза. А вот на побочной диагонали располагаются ошибки. Совершить ошибку первого рода — значит отвергнуть верную нулевую гипотезу. Если же принимается неверная нулевая гипотеза, то это ошибка второго рода.

В механизме проверки гипотез ошибки первого и второго рода неравнозначны. Ошибка первого рода критичнее, вероятность отвержение нулевой гипотезы в случае, когда она верна, жестко ограничивается. Если нулевая гипотеза отвергается при значении уровня значимости $p \leq \alpha$, то вероятность ошибки первого рода получается ограниченной сверху:

$$\mathbf{P}(H_0 \text{ отвергнута} \mid H_0 \text{ верна}) = \mathbf{P}(p \leq \alpha \mid H_0) \leq \alpha.$$

Что касается ошибки второго рода, то она минимизируется по остаточному принципу. Понятие ошибки второго рода связано с понятием мощности статистического критерия. *Мощность* — это вероятность отвергнуть неверную нулевую гипотезу:

$$\text{pow} = \mathbf{P}(\text{отвергаем } H_0 | H_1) = 1 - \mathbf{P}(\text{принимаем } H_0 | H_1).$$

Чтобы найти идеальный критерий для проверки пары нулевая гипотеза – альтернатива, нужно среди всех корректных критериев выбрать критерий с максимальной мощностью.

# Пример 1. Статистически значимо, практически незначимо

В рамках исследования на протяжении трех лет у большой выборки женщин измеряли вес, а также оценивали, насколько активно они занимаются спортом. По итогам исследования выяснилось, что женщины, которые в течение этого времени упражнялись не меньше часа в день, набрали значительно меньше веса, чем женщины, которые упражнялись менее 20 минут в день. Статистическая значимость этого результата достаточно высока: $p < 0.001$. Проблема в размере эффекта: разница в набранном весе между двумя исследуемыми группами женщин составила всего 150 граммов. 150 граммов за 3 года —это не очень много. Крайне сомнительно, что этот эффект имеет какую-то практическую значимость.

Еще один пример связан с клиническими испытаниями гормонального препарата «Премарин», который облегчает симптомы менопаузы. В 2002 году эти испытания были прерваны досрочно, поскольку было обнаружено, что прием препарата ведет к значимому увеличению риска развития рака груди (на 0.08%), инсульта (на 0.08%) и инфаркта (на 0.07%). Этот эффект статистически значим; при этом на первый взгляд кажется, что размеры эффектов ничтожны. Например, если кому-то сказать, что его любимые конфеты повышают риск возникновения инфаркта на 0.07%, вряд ли это заставит человека отказаться от этих конфет. Тем не менее, если пересчитать размеры эффектов на всю популяцию людей, которым этот препарат может быть потенциально приписан, результатом будут тысячи дополнительных смертей. Разработчики препарата не могут взять на себя эту ответственность, поэтому такой препарат немедленно запрещают и снимают с рынка. Этот пример показывает, что практическую значимость результата нельзя определить на глаз.

# Пример 2. Статистически незначимо, практически значимо

Еще один пример — это испытание лекарства, которое замедляет ослабление интеллекта у людей, страдающих болезнью Альцгеймера. В этом исследовании очень сложно измерить размер эффекта. В течение эксперимента одна часть испытуемых должна принимать лекарство, а другая — плацебо. Только по прошествии нескольких лет можно будет сравнить эти две группы. Поэтому такое исследование длится долгое время и дорого стоит. Если при испытании оказывается, что разница между снижением IQ в контрольной группе, где люди принимали плацебо, и тестовой группе, где люди принимали препарат, составляет 13 пунктов, это различие очень большое, и на практике этот эффект крайне значим. При этом может оказаться, что статистическая значимость не была достигнута, то есть $p > \alpha$, и формально нулевую гипотезу об отсутствии эффекта лекарства нельзя отвергнуть. Если предмет исследования очень важен, то, оказавшись в подобной ситуации, возможно, стоит продолжать исследования: набрать еще выборку, уменьшить дисперсию оценки размера эффекта и убедиться в том, что важное открытие не упущено.

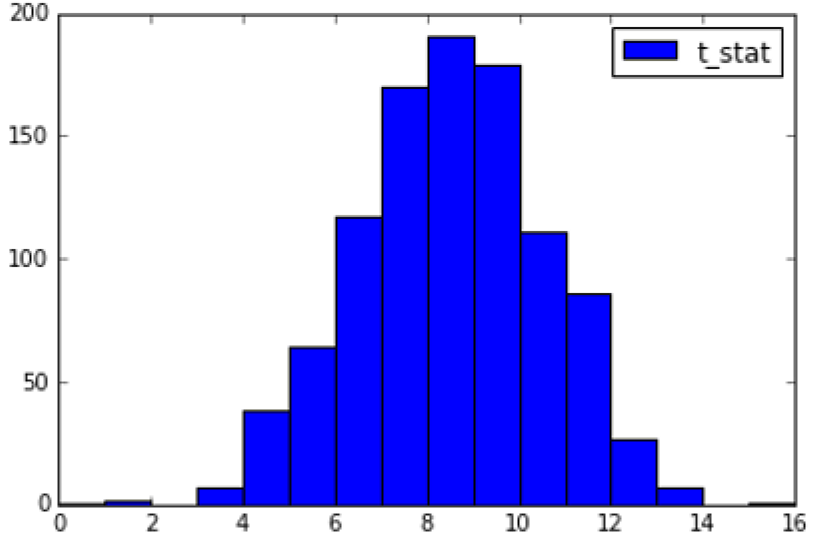
# Биномиальный критерий для доли

Джеймс Бонд утверждает, что он предпочитает пить мартини взболтанным, но не смешанным. Чтобы проверить это на практике, можно предложить Джеймсу Бонду пройти так называемое «слепое тестирование»: завязать ему глаза, несколько раз предложить на выбор взболтанный и смешанный мартини, а после этого спросить, какой напиток он предпочитает. В данном случае, если бы Джеймс Бонд выбирал взболтанный напиток, это считалось бы *успехом*, потому что его выбор соответствует его утверждению. В противном случае считалось бы, что произошла *неудача*, так как выбор утверждению не соответствует.

В данном случае необходимо проверить нулевую гипотезу $H_0$: Джеймс Бонд не различает два вида напитков и выбирает наугад, против некоторой альтернативы. Но альтернатива могла бы быть разной. С одной стороны, можно рассматривать двустороннюю альтернативу (Джеймс Бонд отличает два вида напитков, и у него есть некоторые предпочтения) или одну из односторонних (Джеймс Бонд предпочитает взболтанный мартини, так, как он утверждает, или Джеймс Бонд предпочитает смешанный). Такой эксперимент нужно провести $n$ раз и в качестве $T$-статистики использовать количество единиц выборки или сумму элементов выборки. Если нулевая гипотеза справедлива, то есть Джеймс Бонд выбирает напиток наугад, то можно было бы равновероятно получить любую комбинацию из нулей и единиц.

Таких комбинаций ровно $2^n$, поэтому для того, чтобы получить нулевое распределение, можно было бы сгенерировать все эти наборы данных, на каждом посчитать значение статистики и таким образом получить распределение. На самом деле, в данном случае этот шаг можно пропустить, потому что исследуемая выборка состоит из нулей и единиц и взята из распределения Бернулли с вероятностью успеха $p$. В данном случае вероятность успеха $p = 0.5$, потому что если нулевая гипотеза справедлива, то успех и неудачи происходят равновероятно. Соответственно выборка представляет из себя сумму $n$ независимых одинаково распределенных величин из распределения Бернулли. Значит, нулевое распределение статистики — это биномиальное распределение с параметрами $n$ (количество экспериментов) и $p$ (вероятность успеха).

Нулевое распределение при параметрах $n = 16$ и $p = 0.5$ показано на рисунке . Оно выглядит так, как и ожидалось: пик находится в центре.




Биномиальное распределение с параметрами $n = 16$ и $p = 0.5$

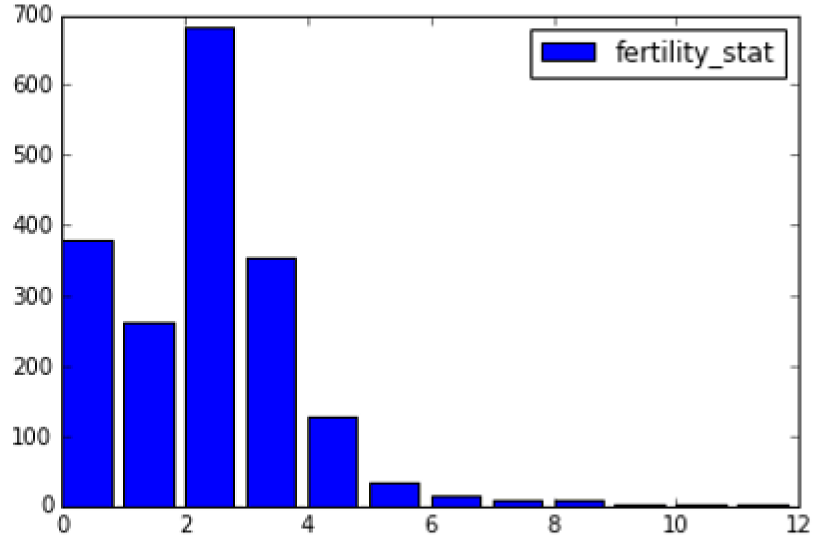
Итак, сначала можно протестировать нулевую гипотезу против односторонней альтернативы $H_1$: Джеймс Бонд предпочитает взболтанный мартини. При такой альтернативе более вероятно попасть в правый конец распределения, то есть получить много единиц в выборке. Пусть было проведено 16 испытаний, и при этом в 12 из них Джеймс Бонд выбрал взболтанный мартини, то есть произошел успех. Если построить соответствующее нулевое распределение, то в данном случае $T$-статистика была бы равна 12 и интерес представлял бы правый «хвост» распределения. В данном случае требуется просуммировать высоту столбцов, начиная со столбца, соответствующего 12, и правее, то есть правый «хвост» распределения, полученное значение — это достигаемый уровень значимости. В данном случае получается $p$-value 0.038, это говорит о том, что на уровне значимости 0.05 нулевая гипотеза отвергается. То есть если успех происходит 12 раз из 16, то можно сделать вывод, что Джеймс Бонд предпочитает взболтанный мартини.

В случае двусторонней альтернативы гипотеза $H_1$ переформулируется следующим образом: Джеймс Бонд предпочитает какой-то один определенный вид мартини, не требуется выбирать, какой именно. При такой альтернативе будут очень вероятны либо большие значения статистики, либо очень маленькие. При расчете достигаемого уровня значимости будут учитываться как правый, так и левый концы распределения. Если предположить, что произошло 12 успехов, то есть 12 раз Джеймс Бонд выбрал взболтанный мартини, то необходимо просуммировать тот же самый правый конец, но теперь к нему добавляется и левый. Значение достигаемого уровня значимости $p = 0.077$, это больше, чем при проверке нулевой гипотезы против односторонней альтернативы. Соответственно, в данном случае нельзя отвергнуть гипотезу на уровне значимости 0.05, однако можно отвергнуть нулевую гипотезу на уровне значимости 0.1. Можно посмотреть, достаточно ли 13 успехов, чтобы отвергнуть нулевую гипотезу на уровне 0.05. В данном случае $p$-value $p = 0.021$, отвергнуть нулевую гипотезу на уровне значимости 0.05 можно.

# Критерий согласия Пирсона (Хи-квадрат)

Критерий согласия Пирсона (или критерий хи-квадрат) используется для проверки того, что некоторая наблюдаемая случайная величина подчиняется тому или иному теоретическому закону распределения.

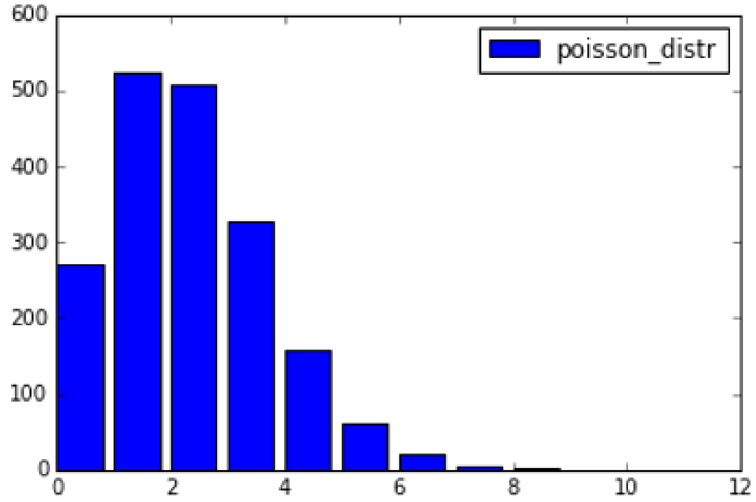
В качестве примера используются данные об исчерпанной рождаемости. Этот признак связан с количеством детей, родившихся у женщины на момент окончания репродуктивного возраста (приблизительно 45 лет). Для 1878 женщин старше 45, участвующих в социологическом опросе жителей Швейцарии, известно количество детей. Этот признак – типичный счётчик, поэтому его можно попробовать оценить с помощью распределения Пуассона. В данном случае выборка — это целочисленный вектор длины $n$ (в данном случае $n$ = 1878),где каждая компонента вектора —это количество детей, рожденных у женщины. В данном случае гипотеза $H_0$ : наблюдаемая величина имеет распределение Пуассона. Из распределения видно, что количество детей меняется от 0 до 11. Чаще всего у женщины не более четырёх детей. Наиболее часто встречающееся количество детей — это два ребёнка.



Кажется, что такие данные должны хорошо описываться распределением Пуассона. Лучшая оценка на параметр $\lambda$ для распределения Пуассона, — это просто выборочное среднее. Если его вычислить, получается $\lambda$ = 1.937. Необходимо проверить следующую гипотезу $H_0$ : наблюдаемая случайная величина имеет распределение Пуассона с параметром $\lambda$ = 2. Это можно делать с помощью критерия согласия Пирсона. Для этого нужно подготовить данные. Первое, что представляет интерес, — это наблюдаемые частоты. Известно, сколько раз встретилось каждое количество детей, так что интересующую величину несложно получить. Тогда элемент результирующего вектора  говорит о том, сколько раз в нашей выборке встретилось количество детей, равное 0 (в данном случае это 379), и последний 11 элемент означает, что 11 детей встретилось всего лишь 1 раз. Это наблюдаемые частоты.

Теперь нужно построить так называемые ожидаемые частоты. Это те частоты, которые бы наблюдались, если бы данные имели распределение Пуассона с параметром $\lambda$ = 2, и размер выборки был бы таким же. Результирующие ожидаемые частоты показаны на рисунке. Видно, что наблюдаемые частоты отличаются от ожидаемых. Строго оценить это различие можно с помощью критерия хи-квадрат. Статистика этого критерия:

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - np_i)^2}{np_i}.$$



Ожидаемые частоты для распределения Пуассона с параметром $\lambda = 2$

При справедливости нулевой гипотезы статистика имеет распределение хи-квадрат с числом степеней свободы $K - 1 - m$, где $m$ — число параметров распределения, оцененных по выборке. Достигаемый уровень значимости, полученный с помощью критерия хи-квадрат: $p$ = 1.77×10$^{-86}$. Это значение очень близко к нулю, значит, можно смело отвергнуть нулевую гипотезу о том, что данные имеют распределение Пуассона с параметром $\lambda$ = 2.

# Связь между проверкой гипотез и доверительными интервалами

Пусть требуется оценить качество предсказаний бинарного классификатора на тестовой выборке из 100 объектов. Этот классификатор верно предсказывает метку класса на 60 из 100 объектов. С одной стороны кажется, что 60 из 100 — это не очень много. С другой стороны, может быть, эта задача достаточно сложная, и предсказать лучше нельзя. Чтобы определить качество работы классификатора, его нужно сравнить с самым бесполезным классификатором — генератором случайных чисел. Если классы в задаче сбалансированы, то генератор случайных чисел в среднем будет угадывать метку у 50 объектов из 100, и вероятность угадать составит 0.5. Можно ли считать, что классификатор, который угадывает классы 60 из 100 объектов, лучше, чем генератор случайных чисел?

Чтобы ответить на этот вопрос, можно построить доверительный интервал для доли верно предсказанных меток. Результат при уровне доверия 0.95: [0.504, 0.696]. Соответствующая генератору случайных чисел вероятность 0.5 не содержится в этом интервале. Из этого следует, что исследуемый классификатор значимо лучше, чем генератор случайных чисел.

# Проверка гипотез и построение доверительных интервалов при сравнении двух классификаторов

Пусть теперь помимо описанного ранее бинарного классификатора имеется второй классификатор, который на той же самой тестовой выборке верно предсказывает метки для 75 объектов из 100. Требуется определить, какой из двух классификаторов лучше. С одной стороны, 75 больше, чем 60. Но с другой стороны, выборка из 100 объектов не очень большая, и такая разница может возникнуть и случайно. Учесть влияние случайности можно с помощью построения доверительных интервалов. Для первого классификатора доверительный интервал Уилсона для доли верных предсказаний: [0.502, 0.691]. Для второго классификатора такой же доверительный интервал: [0.657, 0.825]. Эти доверительные интервалы пересекаются по отрезку [0.657, 0.691]. Но пересечение доверительных интервалов не означает, что классификаторы нельзя различить по качеству. В данном случае выдвинута точечная нулевая гипотеза относительно двух параметров, $\theta_1$ и $\theta_2$, и необходимо проверить её против двусторонней альтернативы:

$$H_0\colon \theta_1 = \theta_2, \quad H_1\colon \theta_1 \neq \theta_2.$$

Правильным решением будет построить доверительный интервал для разности параметров $\theta_1$ и $\theta_2$, именно она полностью соответствует выдвинутой нулевой гипотезе (если $\theta_1 = \theta_2$, значит, их разность равна нулю). 95% доверительный интервал для разности долей в данной задаче: [0.022,0.278]. Этот доверительный интервал не содержит ноль, значит, можно утверждать, что второй классификатор значимо лучше. Если инвертировать этот доверительный интервал описанным ранее способом и выбрать наибольшее значение $\alpha$, при котором ноль попадает в доверительный интервал, получится уровень значимости $p = 0.022$. То есть, на уровне значимости 0.05 отвергается нулевая гипотеза о том, что два классификатора по качеству одинаковы.

Ранее не было учтено, что качество классификаторов определяется на одной и той же обучающей выборке, а значит, выборки в этой задаче — связанные. В такой ситуации доверительный интервал правильнее строить другим методом.

Для этого используется таблица сопряжённости и учитывается не количество ошибок каждого классификатора отдельно, а количество объектов, на которых классификаторы дали разные ответы (20 и 5 в этой таблице).

| I \ II | + | - | $\sum$ |
|---|---|---|---|
| + | 55 | 5 | 60 |
| - | 20 | 20 | 40 |
| $\sum$ | 75 | 25 | 100 |

Полученный 95% доверительный интервал для разности долей в связанных выборках равен [0.06,0.243]. Обратите внимание, что этот доверительный интервал уже, и его левая граница дальше отстоит от нуля, то есть, при учёте связанности увеличивается уверенность в том, что классификаторы отличаются. Для данного интервала достигаемый уровень значимости $p = 0.002$. Он почти в десять раз меньше, чем при построении доверительного интервала без учёта связанности выборок.

# Практическое задание

В файле water.txt представлено 61 наблюдение. Каждое наблюдение – город в Англии и Уэльсе. Города дополнительно поделены на северные и южные. Для каждого города известны средняя годовая смертность на 100000 населения (по данным 1958–1964) и концентрация кальция в питьевой воде (в частях на миллион). Чем выше концентрация кальция, тем жёстче вода.

1) Используя метод .describe() вычислите описательные статистики для северных и южных городов. Сравните средние значения смертности в северных и южных городах и значения концентрации кальция в питьевой воде.
2) Постройте 95% доверительные интервалы для средней годовой смертности по всем южным и северным городам. Отличаются ли границы интервалов?
3) Постройте 95% доверительные интервалы для средней концентрации кальция в питьевой воде для южных и северных городов. Отличаются ли границы интервалов?
4) **Какие можно сделать выводы?**

## ✓ Загрузка данных

```python
import pandas as pd
df = pd.read_csv('water.txt', sep = '\t')
df.head()
```

| | location | town | mortality | hardness |
|---|---|---|---|---|
| **0** | South | Bath | 1247 | 105 |
| **1** | North | Birkenhead | 1668 | 17 |

## ✓ Описательные статистики

```python
df.loc[df.location == 'South'].describe()
```

| | mortality | hardness |
|---|---|---|
| **count** | 26.000000 | 26.000000 |
| **mean** | 1376.807692 | 69.769231 |
| **std** | 140.269175 | 40.360682 |
| **min** | 1096.000000 | 5.000000 |
| **25%** | 1259.250000 | 40.250000 |

## ✓ Доверительный интервал

```python
from statsmodels.stats.weightstats import _tconfint_generic
import numpy as np

mort_mean = df.mortality.mean()

mort_mean_std = df.mortality.std() / np.sqrt(df.mortality.shape[0])

print('95% interval:', _tconfint_generic(mort_mean, mort_mean_std, df.mortality.shape[0] - 1, 0.05, 'two-sided'))
```

```
95% interval: (1476.0833413552848, 1572.2117406119285)
```

Функция для построения t-интервала

Среднее значение

Среднеквадратичное отклонение

Метод .shape возвращает размерность таблицы .shape[0] возвещает количество наблюдений $n$ =61

```python
df.shape
```

```
(61, 4)
```

Количество степеней свободы

$\alpha$

Альтернатива