

Классические статистические тесты

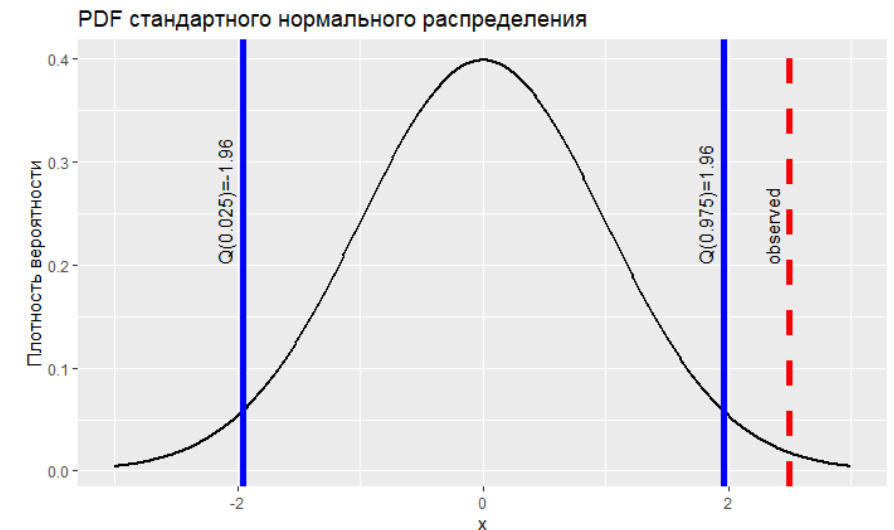
Арнаут Олег, д.м.н.,
магистр по наукам о данных,
врач анестезиолог-реаниматолог

Клиническая и статистическая гипотеза

- Клиническая (исследовательская) гипотеза
 1. препарат А вызывает фиброз миокарда
 2. курение сопряжено с повышенным риском сердечно-сосудистых заболеваний
 3. концентрация IL-1 является биологическим маркером гистологических изменений при бронхиальной астме
- Статистическая гипотеза - строго сформулированное утверждение касательно параметра (параметров) генеральной совокупности (но не выборки)

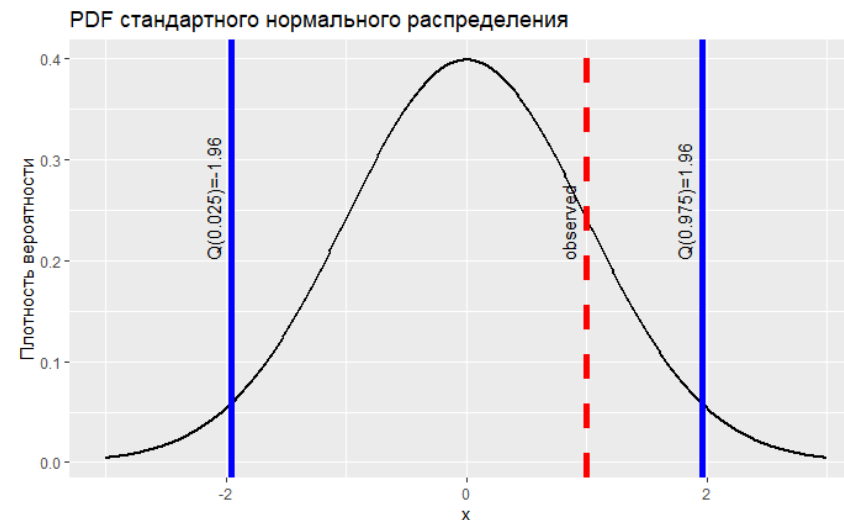
Алгоритм проверки статистических гипотез

- Формулируем нулевую и альтернативную гипотезы (формализация исследовательской гипотезы на языке математики)
- Фиксируем уровень значимости: $\alpha=0.05$
- Выбираем статистический тест (критерий) для проверки гипотезы
- Находим критическое значение статистики
- Находим наблюдаемое значение статистики
- Отвергаем или не отвергаем нулевую гипотезу в соответствии наблюдаемым и критическим значением статистики



Алгоритм проверки статистических гипотез

- Формулируем нулевую и альтернативную гипотезы (формализация исследовательской гипотезы на языке математики)
- Фиксируем уровень значимости: $\alpha=0.05$
- Выбираем статистический тест (критерий) для проверки гипотезы
- Находим критическое значение статистики
- Находим наблюдаемое значение статистики
- Отвергаем или не отвергаем нулевую гипотезу в соответствии наблюдаемым и критическим значением статистики



Если при проверке нулевая гипотеза не отвергается, нельзя считать её доказанной !!!

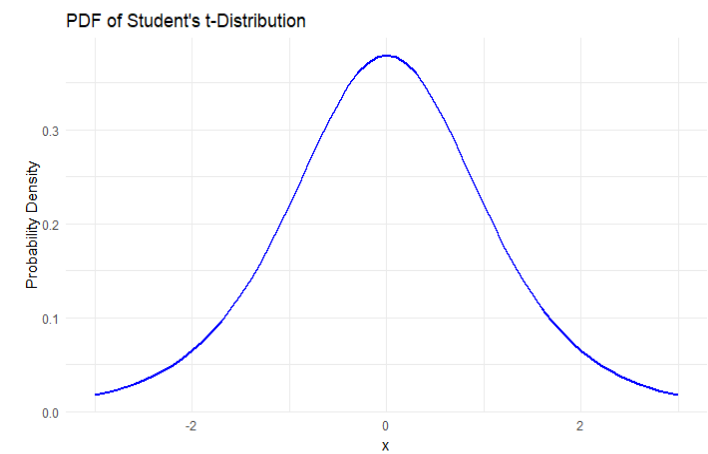
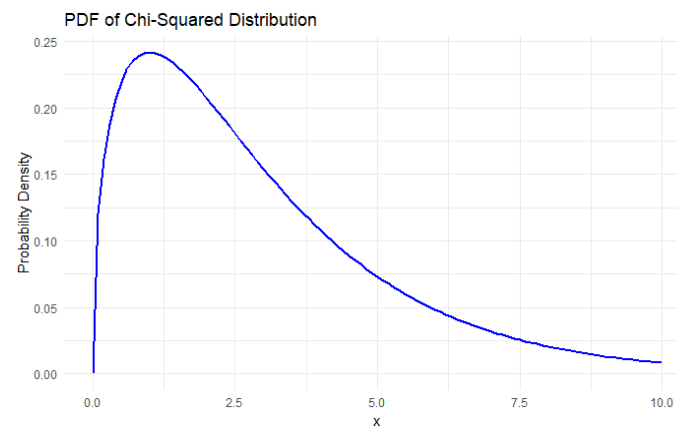
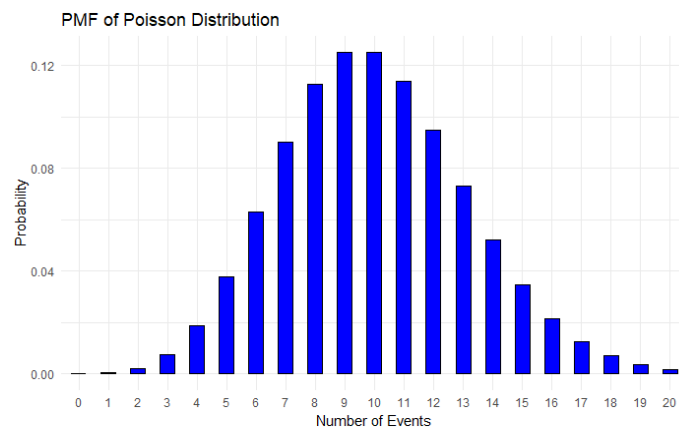
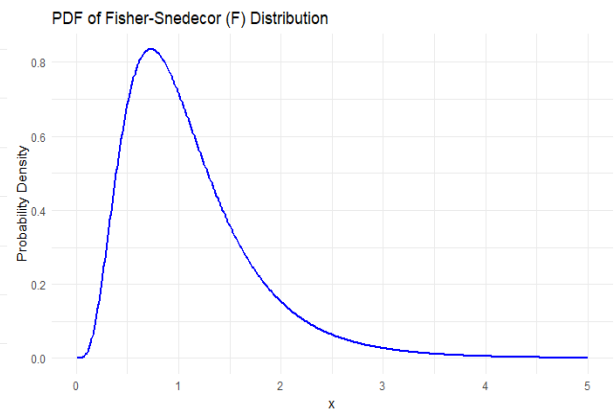
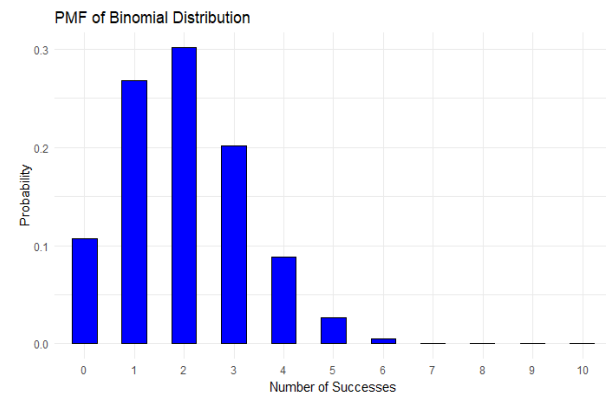
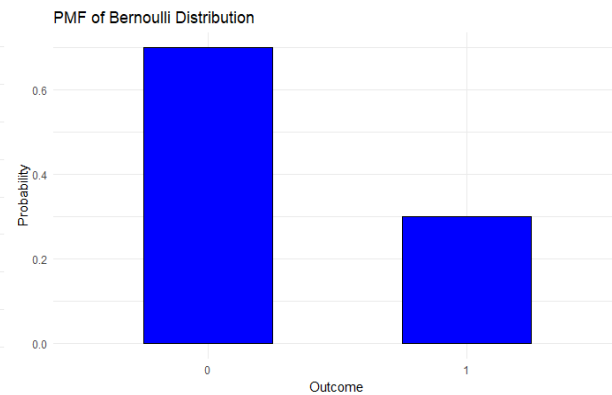
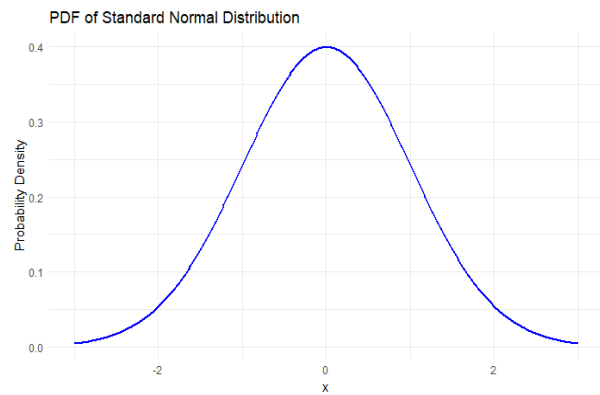
Презумпция нулевой гипотезы

Презумпция невиновности: человек считается невиновным, пока его вина не будет доказана в суде

Презумпция нулевой гипотезы: мы верим в нулевую гипотезу, пока данные не опровергнут её

Ошибки первого и второго рода неравнозначны: перед экспериментом мы фиксируем α , а β получается такой какой получается

Распределения*



R studio...

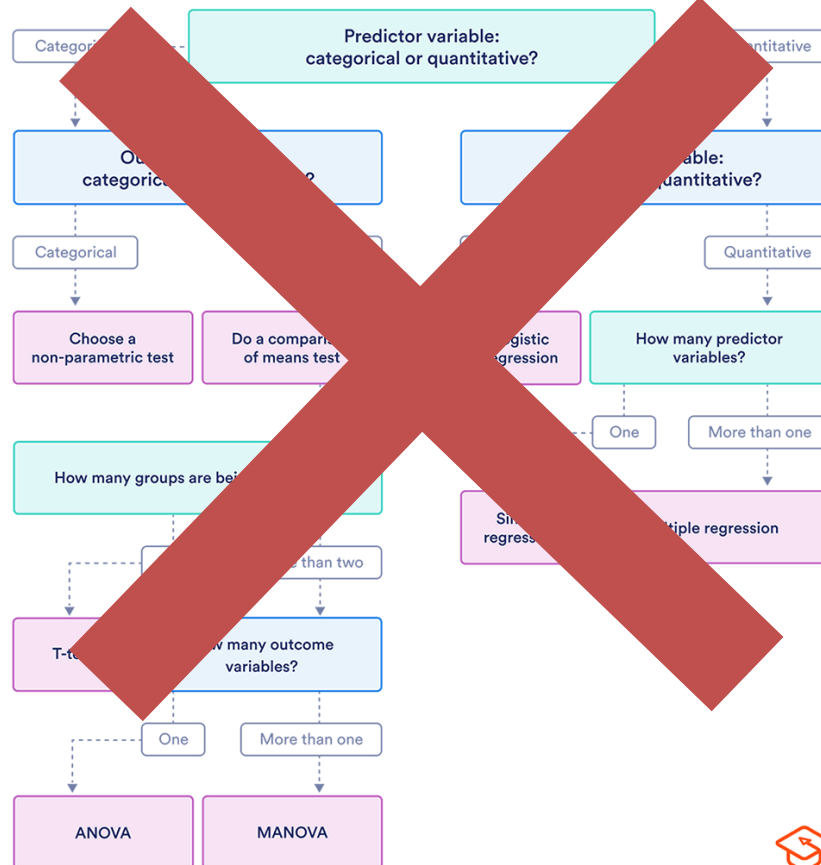
Критерии (статистические тесты)

- Параметрические - включают в себя расчёт параметров конкретного распределения
- Непараметрические - не завязаны на конкретное распределение
- Наиболее мощный критерий это критерий, который при фиксированном размере выборки и фиксированной ошибке первого рода будет давать наименьшую ошибку второго рода

Как выбрать статистический тест?

Choosing a statistical test

This flowchart helps you choose among parametric tests



Выбор статистического теста определяется статистической гипотезой !!!

Гипотезы¹ о средних

Асимптотические versus Точные тесты

- основываются на асимптотических свойствах
- подходят для больших выборок
- вычислительно более эффективны
- основываются на точных распределениях выборочных статистик
- не требуют больших выборок
- могут быть более надежными, так как они учитывают все возможные комбинации данных

Z-критерий для среднего (асимптотический)

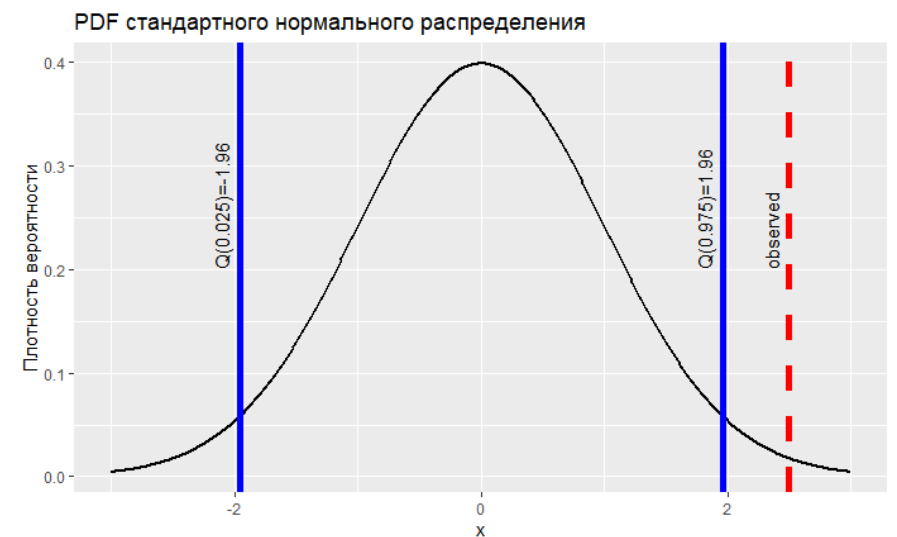
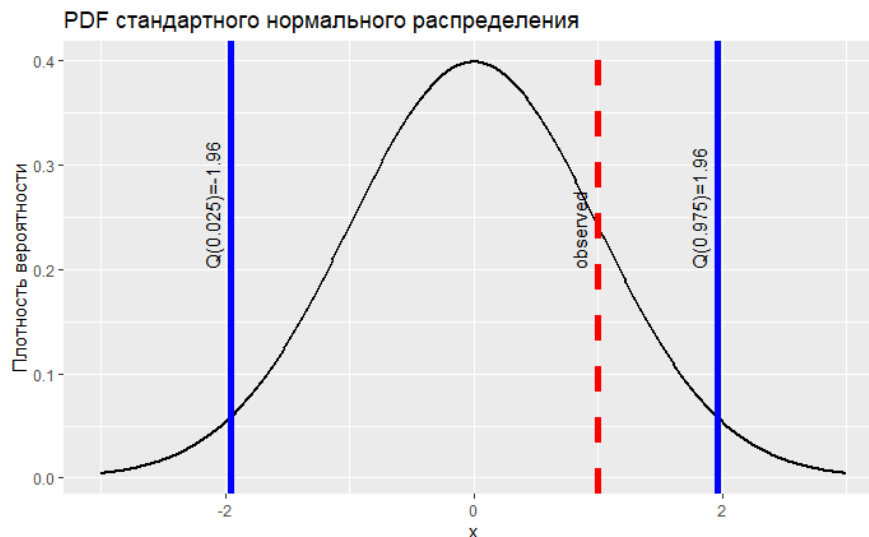
Критерий для проверки:

$$X_1, \dots, X_n \sim iid \text{ } \cancel{N}(\mu, \sigma^2)$$

$$z = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim N(0, 1)$$

$H_0: \mu = \mu_0$ (среднее в генеральной совокупности (популяции) равно μ_0)

$H_a: \mu < \neq > \mu_0$



Z-критерий для среднего (точный)

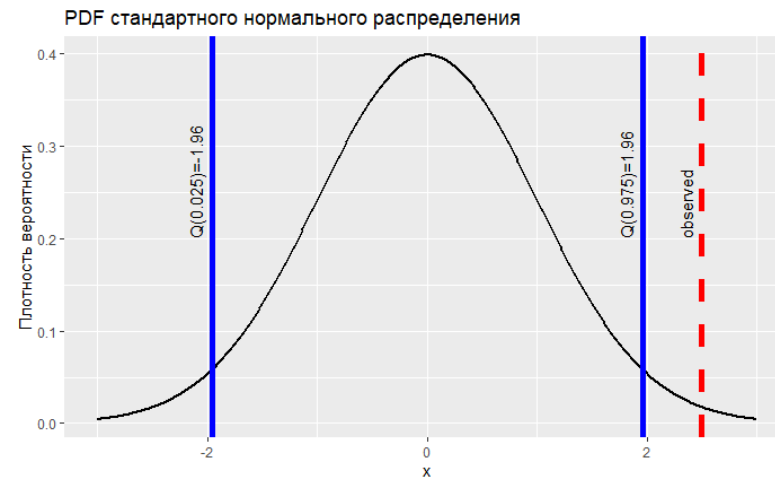
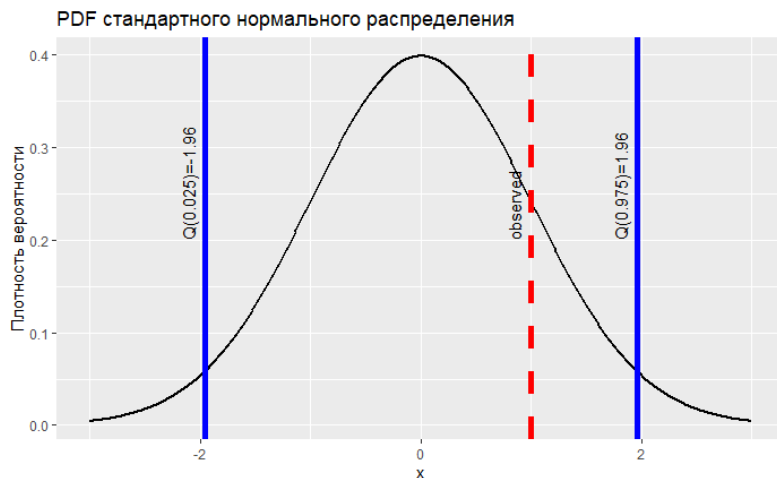
Критерий для проверки:

$$X_1, \dots, X_n \sim iid N(\mu, \sigma^2)$$

$$z = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

$H_0: \mu = \mu_0$ (среднее в генеральной совокупности (популяции) равно μ_0)

$H_a: \mu < \neq > \mu_0$ σ^2 — известна



t-критерий для среднего (точный)

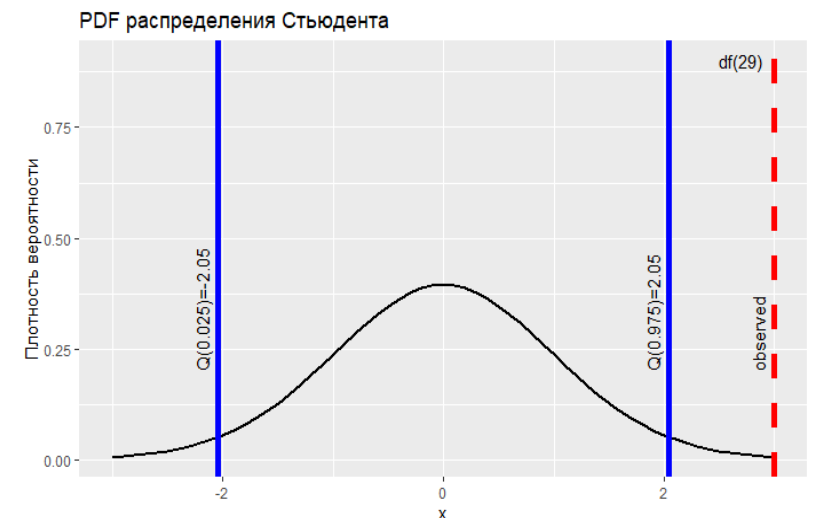
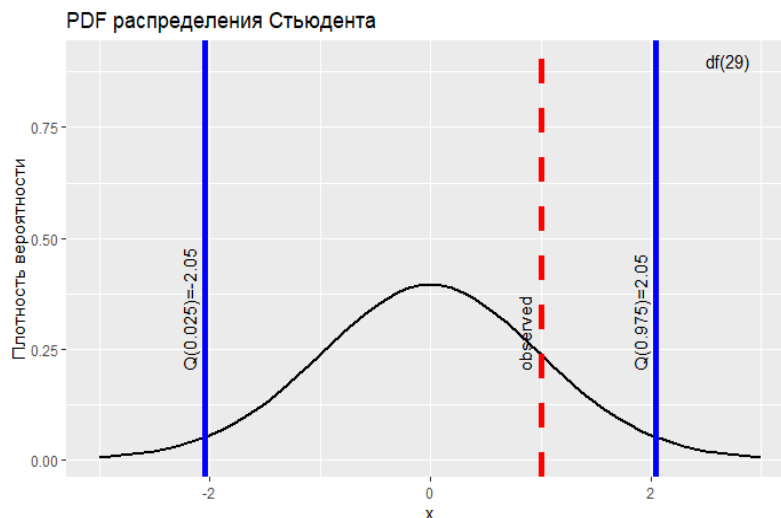
$$X_1, \dots, X_n \sim iid N(\mu, \sigma^2)$$

Критерий для проверки:

$$t = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim t(n-1)$$

$H_0: \mu = \mu_0$ (среднее в генеральной совокупности (популяции) равно μ_0)

$H_a: \mu < \neq > \mu_0$ σ^2 — НЕизвестна



R studio...

Пример

Предположим, мы проводим экспериментальное исследование, в котором хотим проверить, уровень гликолизированного гемоглобина (HbA1c) у пациентов с сахарным диабетом II типа после применения нового метода лечения.

Нам известно, что в общей популяции пациентов с сахарным диабетом II типа уровень HbA1c составляет 10%. Мы хотим узнать, отличается ли уровень HbA1c в целевой группе после применения нового метода.

t-критерий для среднего (точный). Пример

$$X_1, \dots, X_n \sim iid N(\mu, \sigma^2)$$

$H_0: \mu = 10$ средний уровень HbA1c равен 10

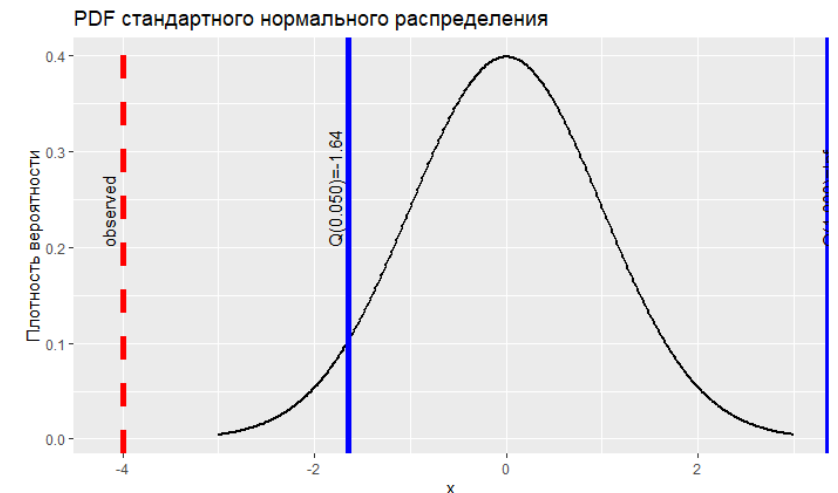
$H_a: \mu < 10$ средний уровень HbA1c меньше 10

$$n = 100$$

$$\hat{\mu} = 8$$

$$\hat{\sigma}^2 = 25$$

$$t = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{-2}{\sqrt{\frac{25}{100}}} = -4$$



t-критерий для среднего (точный). Пример

$$X_1, \dots, X_n \sim iid N(\mu, \sigma^2)$$

$H_0: \mu = 10$ средний уровень HbA1c равен 10

$H_a: \mu < 10$ средний уровень HbA1c меньше 10

$$n = 100$$

$$\hat{\mu} = 8$$

$$\hat{\sigma}^2 = 25$$

$$t = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

Z-критерий для разности средних (асимптотический). Выборки независимые

$$X_1, \dots, X_n \sim iid (\mu_1, \sigma_1^2)$$

$$Y_1, \dots, Y_m \sim iid (\mu_2, \sigma_2^2)$$

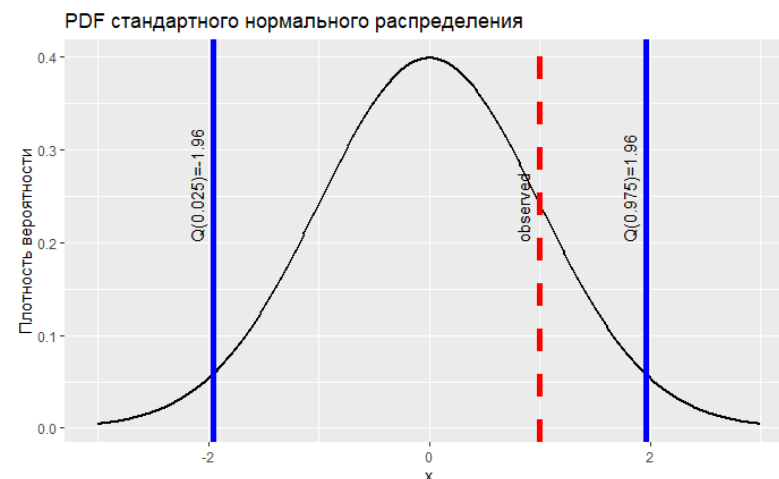
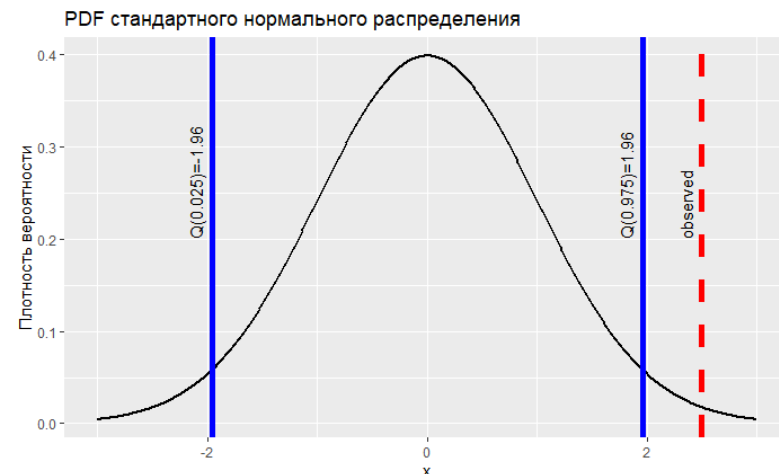
Выборки независимые

$H_0: \mu_1 = \mu_2$ среднее в популяции X (μ_1) равно среднему в популяции Y (μ_2)

$$H_a: \mu_1 < \neq > \mu_2$$

Критерий для проверки:

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}} \sim N(0, 1)$$



Точные критерии для разности средних. Выборки независимые

Критерий для проверки:

дисперсии
известны

$$z = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

Нормальное
распределение

дисперсии
неизвестны,
но равны

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}} \sim t(n + m - 2)$$

Распределение
Стьюдента

дисперсии
неизвестны

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}} \sim t(\nu)$$

Распределение
Стьюдента
Метод Уэлча

$$\nu = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1 - 1)} + \frac{\sigma_2^4}{n_2^2(n_2 - 1)}}$$

Разности средних. Независимые выборки. Пример

- Мы исследуем влияние нового препарата на основной метаболизм крыс. У вас есть две группы крыс: контрольная, которая получает плацебо и экспериментальная группа, которая получает препарат.
- Наша задача - определить, меняется ли основной метаболизм крыс (ОМК) под воздействием препарата. Известно что размер выборки $n_1 = n_2 = 10$. Среднее значение ОМК в контрольной группе составила 8.2 ккал/ч/кг, в экспериментальной группе ОМК составила 10 ккал/ч/кг ($SD_1 = 6$, $SD_2 = 4$)
- Выберите статистический тест. Аргументируйте свой выбор. Сформулируйте нулевую и альтернативную гипотезы. Расчитайте соответствующий критерий и проинтерпретируйте полученный результат

Разности средних. Независимые выборки. Пример

$$X_1, \dots, X_n \sim iid (\mu, \sigma^2)$$

$$Y_1, \dots, Y_m \sim iid (\mu, \sigma^2)$$

$H_0: \mu_1 = \mu_2$ средний уровень ОМК в группах не отличается

$H_a: \mu_1 \neq \mu_2$ средний уровень ОМК в группах отличается

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}}$$

Разности средних. Независимые выборки. Пример

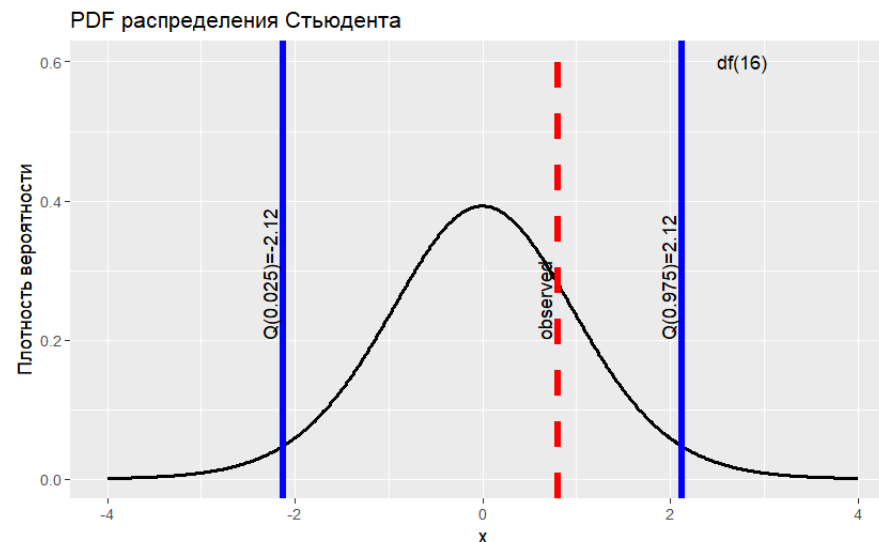
$$X_1, \dots, X_n \sim iid (\mu, \sigma^2)$$

$$Y_1, \dots, Y_m \sim iid (\mu, \sigma^2)$$

$H_0: \mu_1 = \mu_2$ средний уровень ОМК в группах не отличается

$H_a: \mu_1 \neq \mu_2$ средний уровень ОМК в группах отличается

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}} = \frac{10 - 8.2}{\sqrt{\frac{36}{10} + \frac{16}{10}}} = 0.79$$



Разность средних (зависимые выборки)

Выборки зависят друг от друга:

$$X_1, \dots, X_n \sim iid N(\mu_1, \sigma_1^2) \quad Y_1, \dots, Y_n \sim iid N(\mu_2, \sigma_2^2)$$

- Измерения делаются на одних и тех же объектах
- Можем посмотреть прирост на отдельных объектах

$$d_i = X_i - Y_i$$

Критерий для проверки:

$$t = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim t(n-1)$$

- Используем распределение Стьюдента, дисперсию считаем по формуле:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$$

Разность средних (зависимые выборки). Пример

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

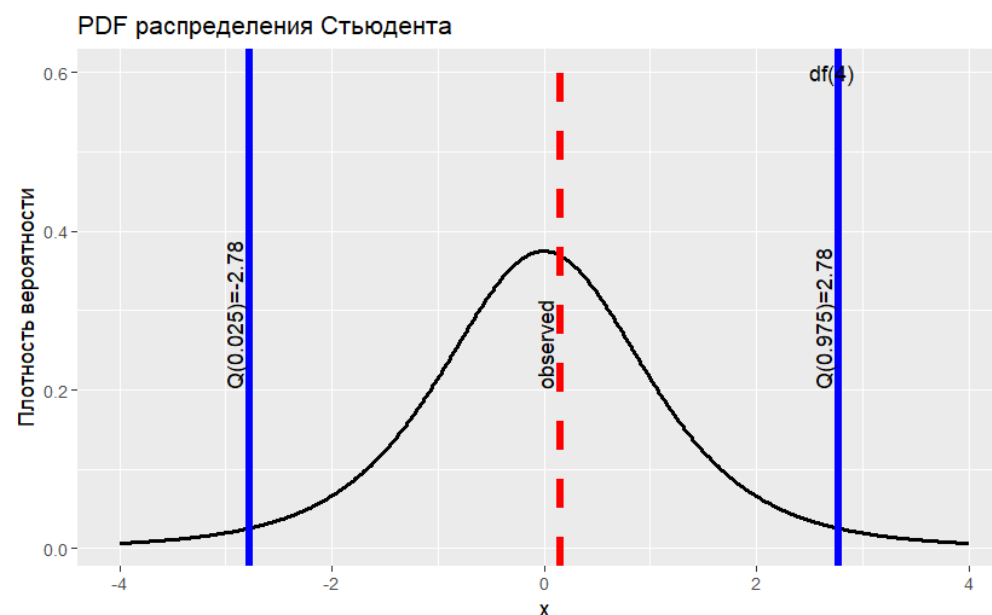
среднее в популяции X (μ_1) равно среднему в популяции Y (μ_2), где обе выборки получены на одних и тех же пациентах до и после воздействия (pre-post trials)

$$\hat{d} = \frac{1}{5} \sum_{i=1}^5 d_i = -6$$

$$\sigma^2 = \frac{1}{5-1} \sum_{i=1}^5 (d_i - \hat{d})^2 = 92.5$$

$$t = \frac{\hat{d} - 0}{\sqrt{\frac{\sigma^2}{n}}} = \frac{-6 - 0}{\sqrt{\frac{92.5}{5}}} = -1.39$$

Препарат 1	50	40	45	45	35
Препарат 2	60	30	30	35	30
d_i	10	-10	-15	-10	-5



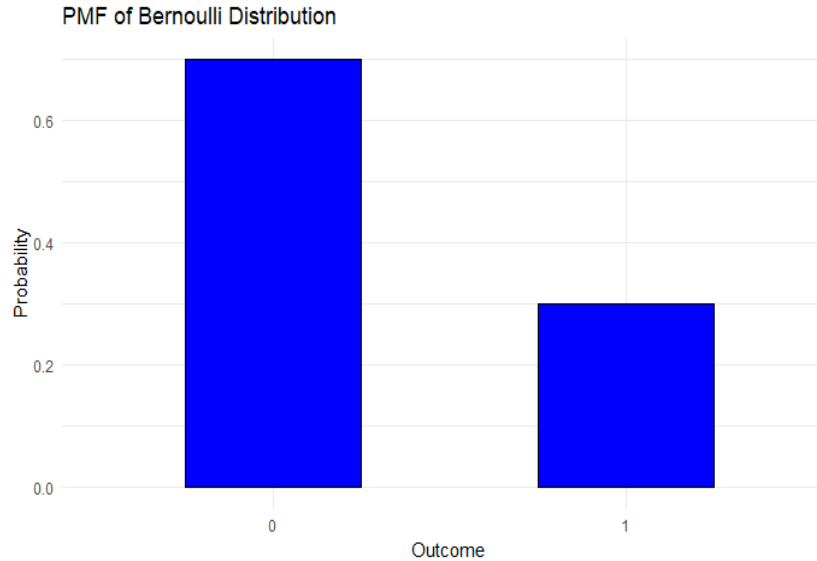
Резюме

- Благодаря свойствам ЦПТ, критерии основанные на ней широко распространены
- Для маленьких выборок лучше использовать точные критерии
- Если предполагается равенство дисперсий, это тоже не мешает проверить, но может возникнуть проблема с тестом на равенство дисперсии
- Для разных тестов критерии считаются в соответствии с логикой нулевой гипотезы

Гипотезы о долях

Z-критерий для доли (асимптотический)

$$X_1, \dots, X_n \sim iid Bern(p)$$

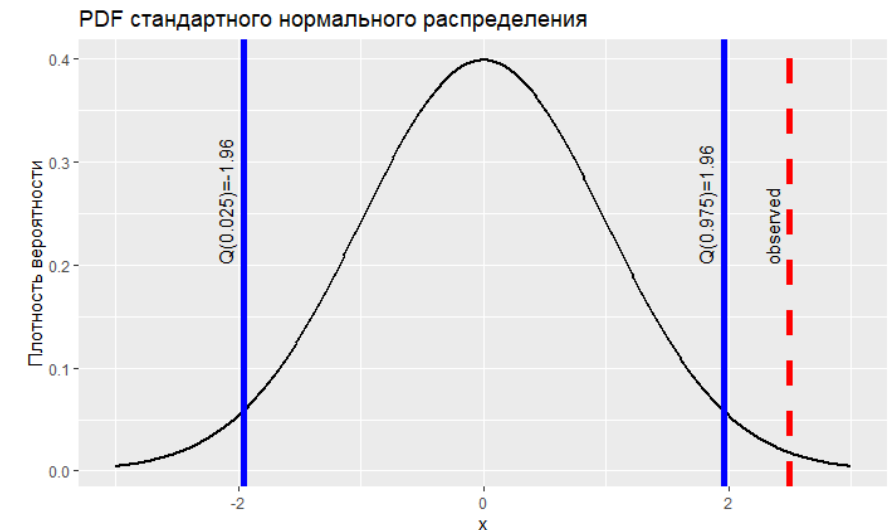
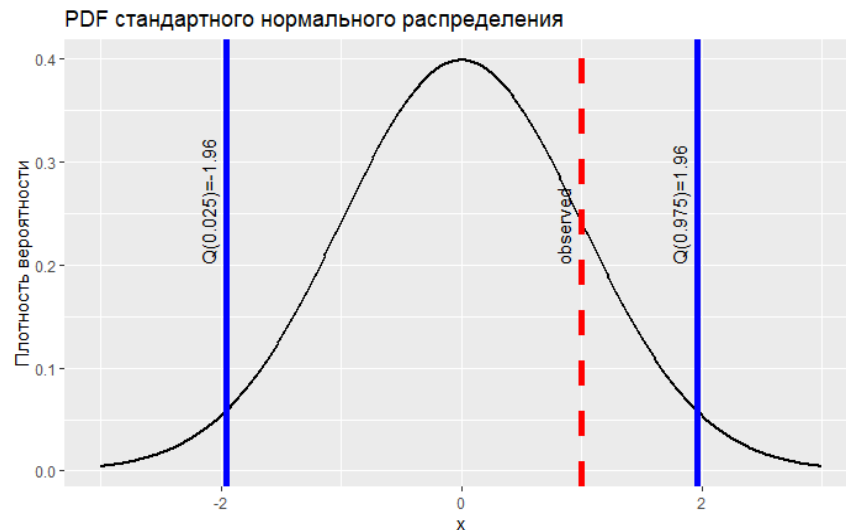


Критерий для проверки:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$$

$H_0: p = p_0$ доля 'успеха' в генеральной совокупности (популяции) равна p_0

$H_a: p \neq p_0$



Z-критерий для доли (асимптотический). Пример

Исследование эффективности нового метода обучения для улучшения памяти студентов.

Группе студентов предложили новый метод обучения для улучшения памяти. После проведения обучения и тестирования, студентов делят на две категории: тех, кто показал улучшение памяти (успех), и тех, кто не показал улучшения (неуспех).

Исследователь хочет проверить, если доля студентов с улучшенной памятью после нового метода обучения больше целевой доли (p_0) в 50%.

Z-критерий для доли (асимптотический). Пример

$$X_1, \dots, X_n \sim iid Bern(p)$$

$H_0: p = 0.5$ Доля студентов с улучшенной памятью после нового метода обучения равна 50%.

$H_a: p > 0.5$ Доля студентов с улучшенной памятью после нового метода обучения больше 50%.

$$p = 0.7$$

$$p_0 = 0.5$$

$$n = 100$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 (1 - p_0)}{n}}}$$

Z-критерий для доли (асимптотический). Пример

$$X_1, \dots, X_n \sim iid Bern(p)$$

$H_0: p = 0.5$ Доля студентов с улучшенной памятью после нового метода обучения равна 50%.

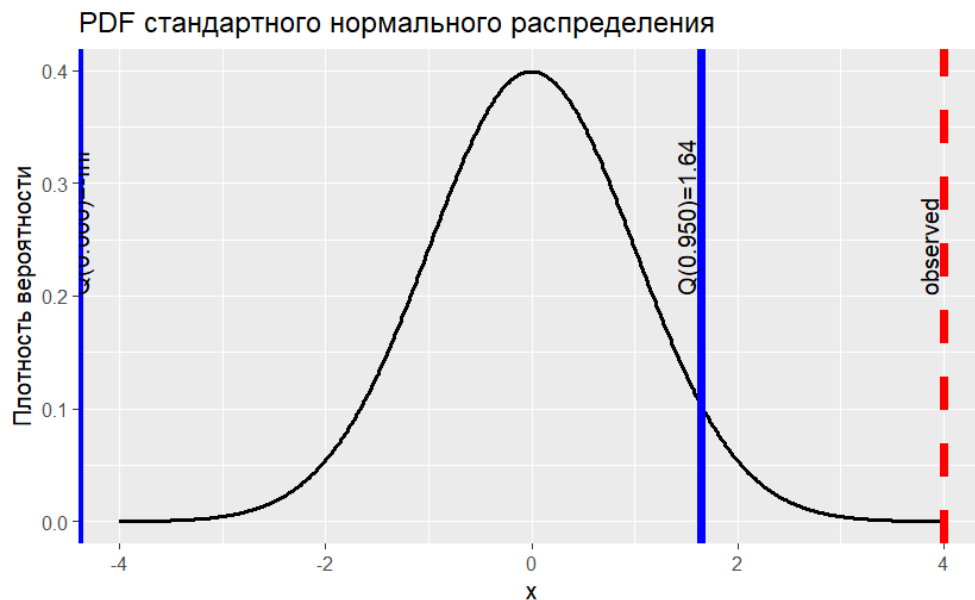
$H_a: p > 0.5$ Доля студентов с улучшенной памятью после нового метода обучения больше 50%.

$$p = 0.7$$

$$p_0 = 0.5$$

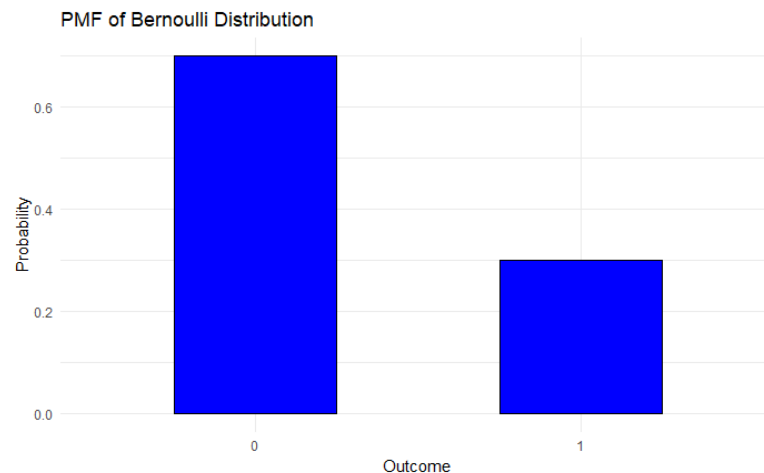
$$n = 100$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.7 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = \frac{0.2}{\sqrt{\frac{0.25}{100}}} = \frac{0.2 \times 10}{0.5} = 4$$



Точный критерий для доли

$$X_1, \dots, X_n \sim iid Bern(p)$$



Критерий для проверки:

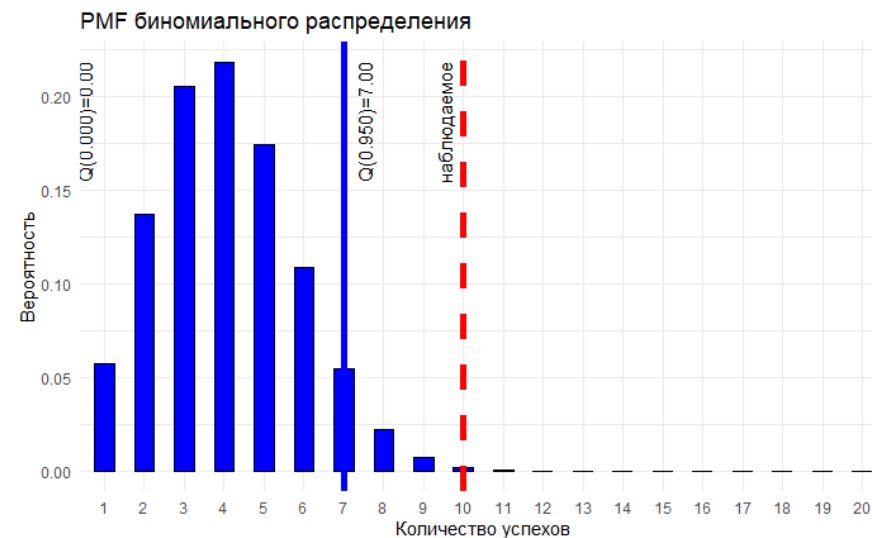
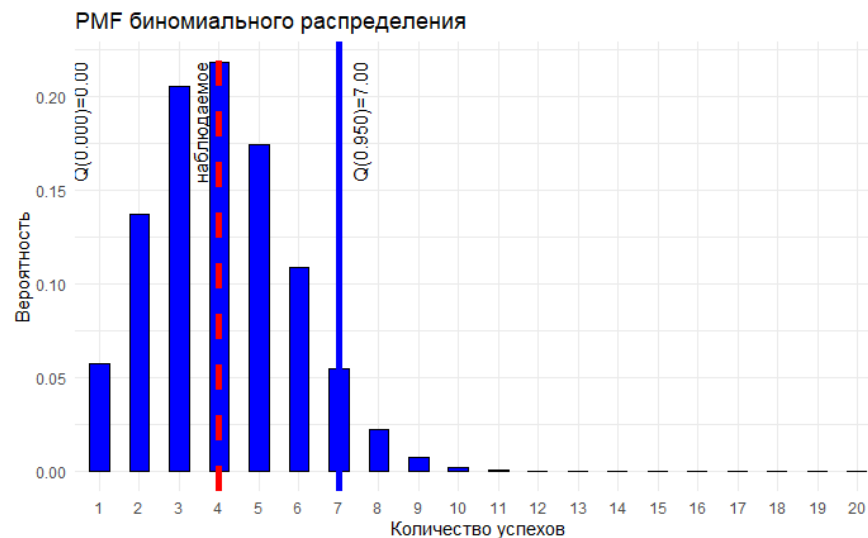
$$X \sim Binom(n, p)$$

X – количество успехов

$$H_0: p = p_0$$

доля 'успеха' в генеральной совокупности (популяции) равна p_0

$$H_a: p < \neq > p_0$$



Точный критерий для доли. Пример

$$X_1, \dots, X_n \sim iid Bern(p)$$

$H_0: p = 0.5$ Доля студентов с улучшенной памятью после нового метода обучения равна 50%.

$H_a: p > 0.5$ Доля студентов с улучшенной памятью после нового метода обучения больше 50%.

Критерий для проверки:

$$p = 0.7$$

$$p_0 = 0.5$$

$$n = 100$$

$$X \sim Binom(n, p)$$

Точный критерий для доли. Пример

$$X_1, \dots, X_n \sim iid Bern(p)$$

$H_0: p = 0.5$ Доля студентов с улучшенной памятью после нового метода обучения равна 50%.

$H_a: p > 0.5$ Доля студентов с улучшенной памятью после нового метода обучения больше 50%.

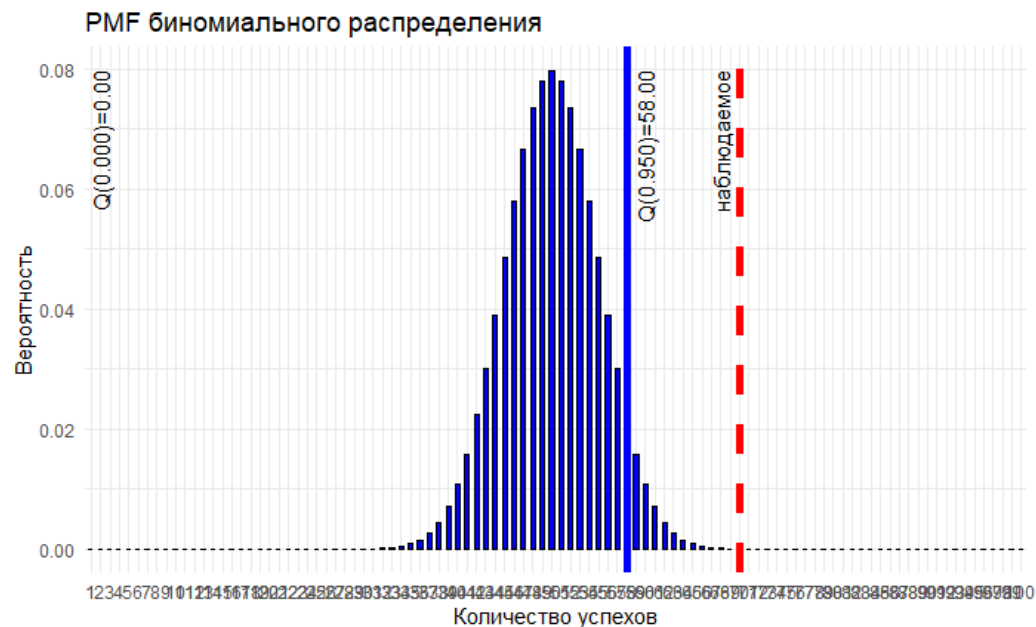
Критерий для проверки:

$$p = 0.7$$

$$p_0 = 0.5$$

$$n = 100$$

$$X \sim Binom(n, p)$$



Z-критерий для разности долей для независимых выборок (асимптотический)

$$X_1, \dots, X_n \sim iid Bern(p)$$

$$Y_1, \dots, Y_{n_y} \sim iid Bern(p_y)$$

Критерий для проверки:

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{P(1-P) \cdot \left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} \sim N(0, 1)$$

$$P = \frac{m_x + m_y}{n_x + n_y}$$

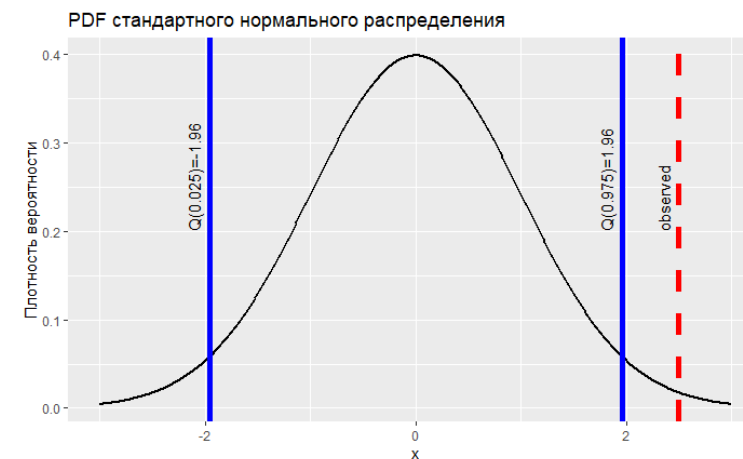
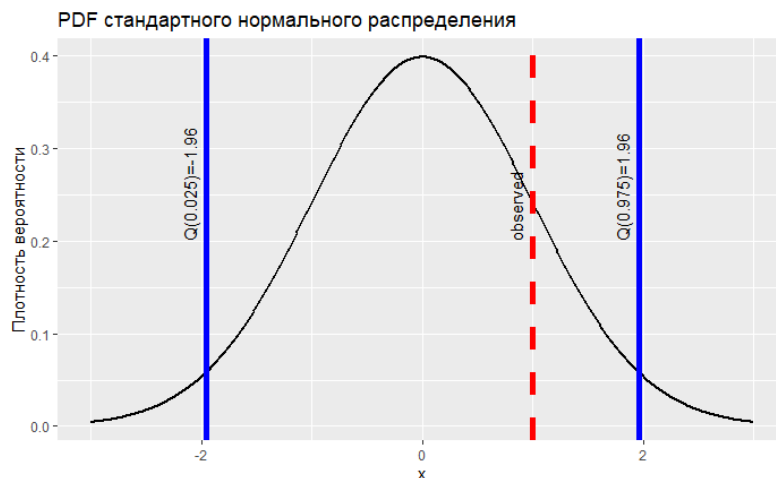
m_x, m_y – количество успехов

n_x, n_y – количество наблюдений

$$H_0: p_x = p_y$$

$$H_a: p_x < \neq > p_y$$

доли 'успеха' в генеральной совокупности (популяции) X
и генеральной совокупности (популяции) Y равны



Z-критерий для разности долей для независимых выборок (асимптотический). Пример

Исследование эффективности нового метода обучения для улучшения памяти студентов.

Студентов случайно разделили на 2 группы. Контрольная группа занималась по традиционной методике, Экспериментальная группа работала по новой методике обучения для улучшения памяти. После проведения обучения и тестирования, студентов из исследуемых групп делят на две категории: тех, кто показал улучшение памяти (успех), и тех, кто не показал улучшения (неуспех).

Исследователь хочет проверить, отличается ли доля студентов с улучшенной памятью в группах.

Z-критерий для разности долей для независимых выборок (асимптотический). Пример

$$X_1, \dots, X_n \sim iid Bern(p_x)$$

$$Y_1, \dots, Y_{n_y} \sim iid Bern(p_y)$$

$H_0: p_x = p_y$ Доли студентов с улучшенной памятью в группах равны

$H_a: p_x \neq p_y$ Доли студентов с улучшенной памятью в группах не равны

$$p_x = 0.2$$

$$p_y = 0.4$$

$$n_x = n_y = 100$$

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{P(1 - P) \cdot \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$$

Z-критерий для разности долей для независимых выборок (асимптотический). Пример

$$X_1, \dots, X_n \sim iid Bern(px)$$

$$Y_1, \dots, Y_{n_y} \sim iid Bern(p_y)$$

$H_0: p_x = p_y$ Доли студентов с улучшенной памятью в группах равны

$H_a: p_x \neq p_y$ Доли студентов с улучшенной памятью в группах не равны

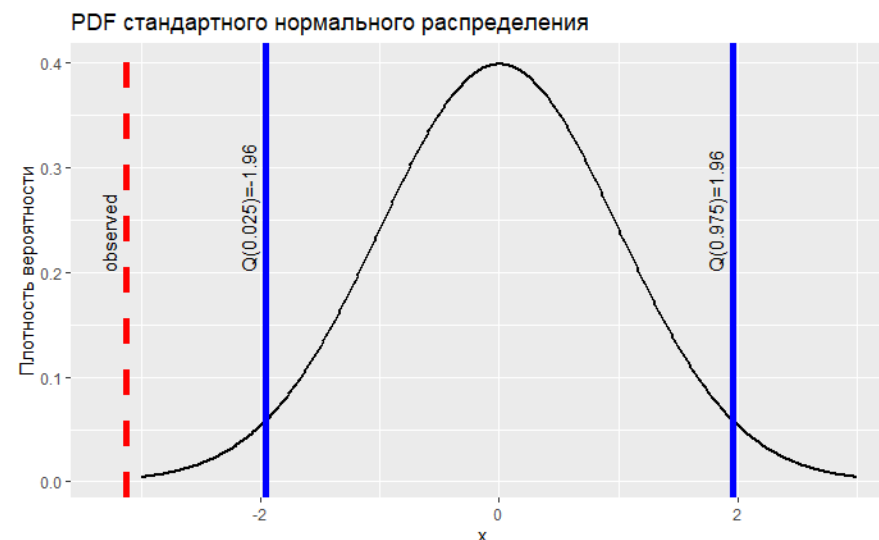
$$p_x = 0.2$$

$$p_y = 0.4$$

$$n_x = n_y = 100$$

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{P(1 - P) \cdot \left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} = -3.09$$

$$P = \frac{20+40}{200} = 0.3$$



Z-критерий для разности долей для зависимых выборок (асимптотический).

$$X_1, \dots, X_n \sim iid Bern(p_x)$$

$$Y_1, \dots, Y_n \sim iid Bern(p_y)$$

$$H_0: p_x = p_y$$

$$H_a: p_x < \neq > p_y$$

Выборки зависимые !!!

доля в популяции X (p_x) равна доле в популяции Y (p_y), где обе выборки получены на одних и тех же пациентах до и после воздействия (pre-post trials)

Критерий для проверки:

$$z = \frac{c - b}{\sqrt{c + b - \frac{(c - b)^2}{n}}} \sim N(0, 1)$$

		До (X)	
		0	1
После (Y)	0	a	<i>b</i>
	1	<i>c</i>	d

Z-критерий для разности долей для зависимых выборок (асимптотический). Пример

Исследование эффективности новой реабилитационной программы для пациентов после инсульта.

Группе пациентов после инсульта предложили новый метод реабилитации для улучшения способность пациентов самостоятельно передвигаться.

Для начала оценивают статус пациента (самостоятельно передвигается или самостоятельно не передвигается), проводят реабилитацию, опять оценивают статус, сопоставляя результаты до и после реабилитации. Соответственно пациентов делят на 3 категории: улучшение способность пациентов самостоятельно передвигаться, ухудшение таковой и тех у кого изменений нет.

Исследователь планирует проверить, эффективен ли новый метод.

Z-критерий для разности долей для зависимых выборок (асимптотический).

$$X_1, \dots, X_n \sim iid Bern(p_x)$$

$$Y_1, \dots, Y_n \sim iid Bern(p_y)$$

$H_0: p_x = p_y$ Доля пациентов которые передвигаются не изменилась

$H_a: p_x < p_y$ Доля пациентов которые передвигаются увеличилась

		До (X)	
		0	1
После (Y)	0	30	10
	1	20	40

Выборки зависимые !!!

$$z = \frac{c - b}{\sqrt{c + b - \frac{(c - b)^2}{n}}}$$

Z-критерий для разности долей для зависимых выборок (асимптотический).

$$X_1, \dots, X_n \sim iid Bern(p_x)$$

$$Y_1, \dots, Y_n \sim iid Bern(p_y)$$

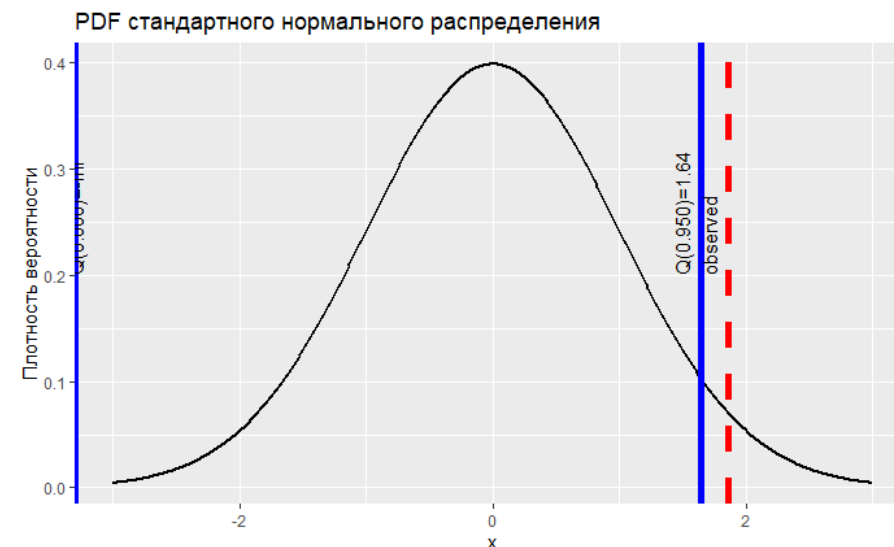
$H_0: p_x = p_y$ Доля пациентов которые передвигаются не изменилась

$H_a: p_x < p_y$ Доля пациентов которые передвигаются увеличилась

		До (X)	
		0	1
После (Y)	0	30	10
	1	20	40

Выборки зависимые !!!

$$z = \frac{c - b}{\sqrt{c + b - \frac{(c - b)^2}{n}}} = 1.86$$



Резюме

- Для проверки гипотезы о долях используется как z-тест, основанный на ЦПТ, так и точные, основанные на распределении t.
- В случае независимых и зависимых выборок статистика считается немного по-разному, для зависимых мы акцентируем внимание только на изменениях

Гипотезы о дисперсии

χ^2 – критерий для дисперсии (точный)

Критерий для проверки:

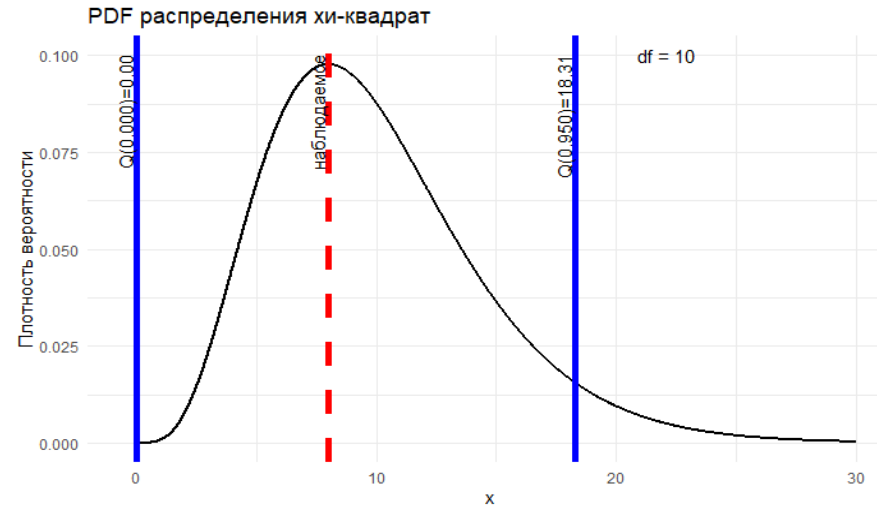
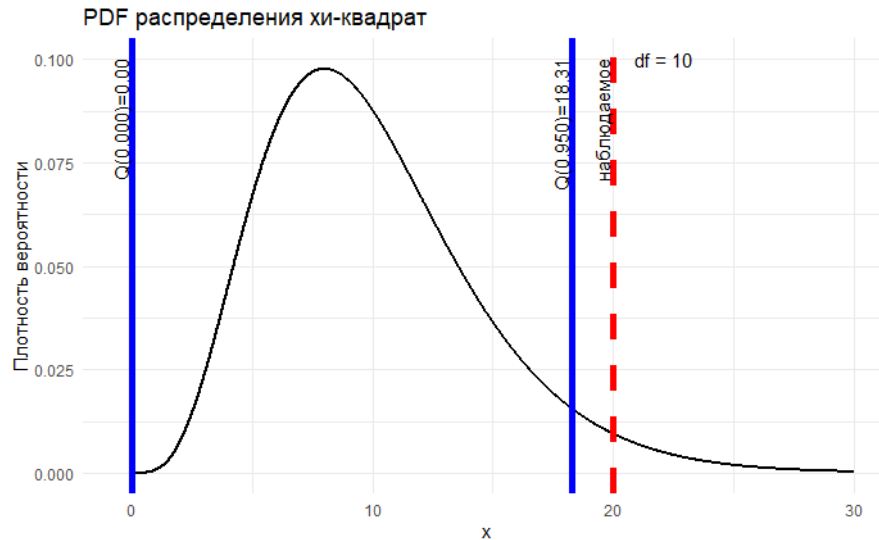
$$X_1, \dots, X_n \sim iid N(\mu, \sigma^2)$$

μ – НЕизвестно

$$H_0: \sigma^2 = \sigma_0^2 \quad (\text{дисперсия в генеральной совокупности (популяции) равна } \sigma_0^2)$$

$$H_a: \sigma^2 > \sigma_0^2$$

$$\frac{(n-1) \cdot s^2}{\sigma_0^2} \sim \chi_{n-1}^2$$



χ^2 – критерий для дисперсии (точный). Пример

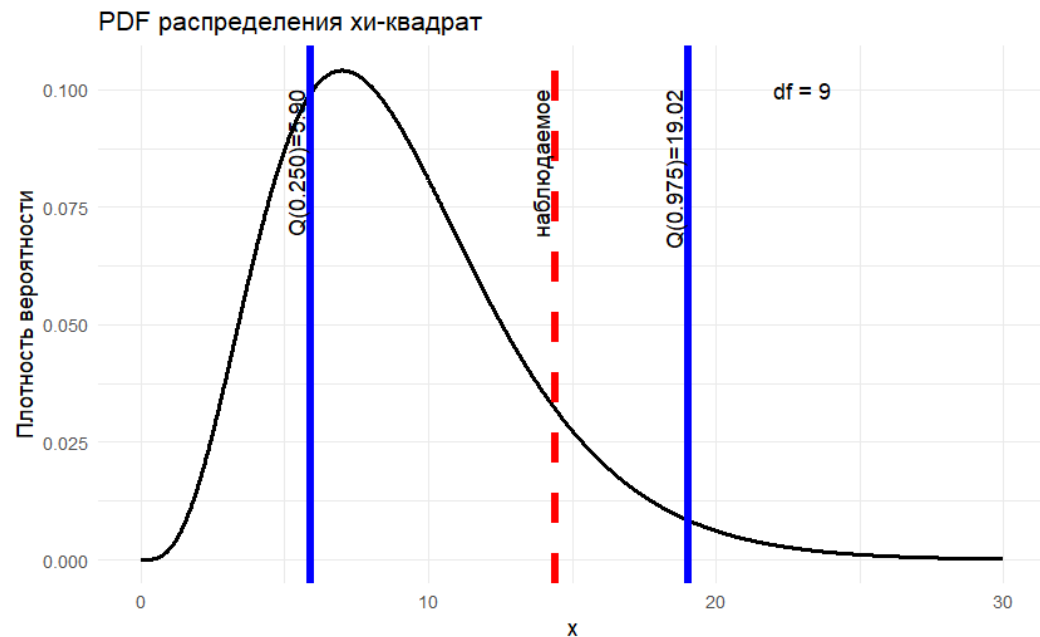
H_0 : дисперсия биомаркера = 0.5 $H_0: \sigma^2 = 0.5$ $n = 10$
 H_a : дисперсия биомаркера > 0.5 $H_a: \sigma^2 > 0.5$ $s^2 = 0.8$

Предположение:

$$X_1, \dots, X_n \sim iid N(\mu, \sigma^2)$$

$$\chi_{obs}^2 = \frac{(10 - 1) \cdot 0.8}{0.5}$$

$$\chi_9^2(0.95) = 14.4$$



Тест Фишера для отношения дисперсий (точный)

$$X_1, \dots, X_n \sim iid N(\mu_x, \sigma_x^2)$$

$$Y_1, \dots, Y_m \sim iid N(\mu_y, \sigma_y^2)$$

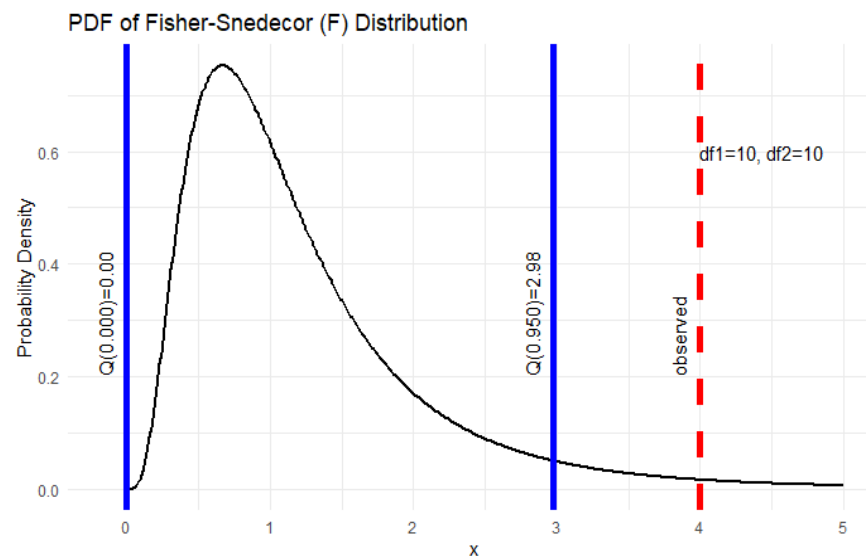
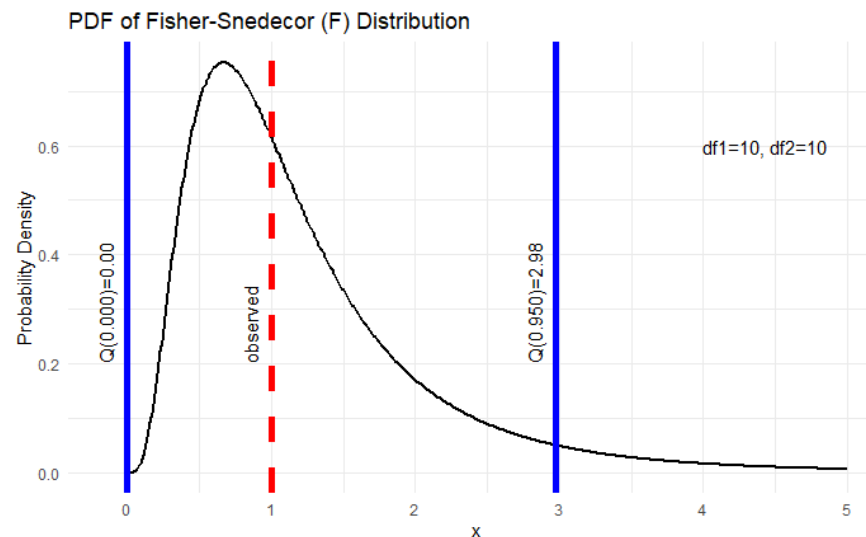
Выборки независимые

$$H_0: \sigma_x = \sigma_y$$

$$H_a: \sigma_x < \neq > \sigma_y$$

Критерий для проверки:

$$\frac{S_x^2}{S_y^2} \sim F_{n-1, m-1}$$



Тест Фишера для отношения дисперсий. Пример

Предположим, мы проводим исследование, в котором измеряем уровень ALT двух разных групп животных: группе А и группе Б.

Нас интересует, есть ли разница в дисперсиях этих измерений между группами.

Тест Фишера для отношения дисперсий. Пример

H_0 : дисперсии в группах одинаковы

H_a : дисперсия в группе В больше чем в группе А

Предположение:

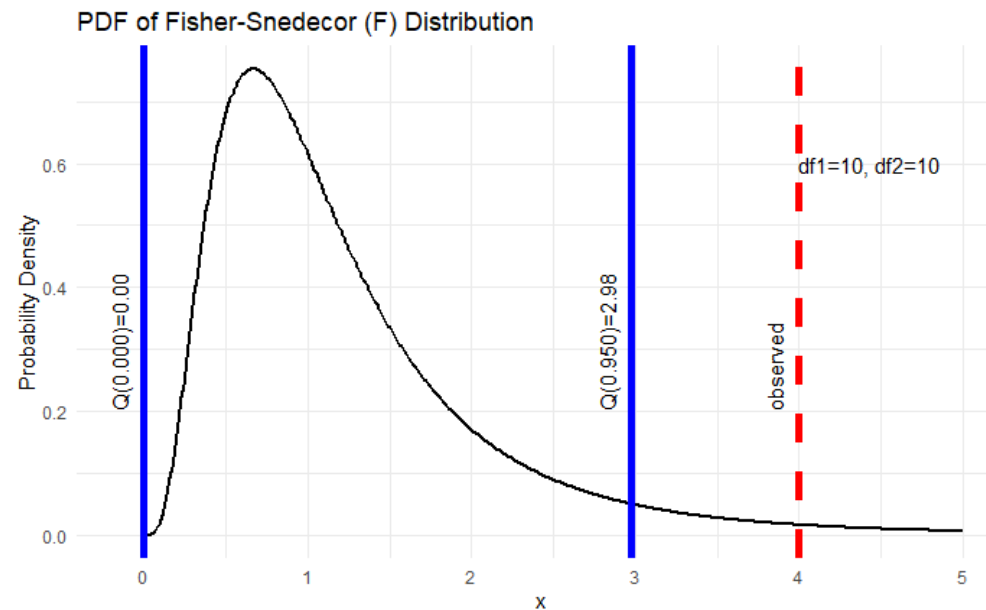
$$X_1, \dots, X_n \sim iid N(\mu_x, \sigma_x^2)$$

$$Y_1, \dots, Y_m \sim iid N(\mu_y, \sigma_y^2)$$

Выборки независимые

$$F_{obs} = \frac{44}{11}, n = 11$$

$$F_{10,10}(0.95) = 2.98$$



Резюме

- Выбор статистического теста определяется статистической гипотезой.
- Параметрические тесты бывают асимптотическими и точными.
- Для разных тестов критерии оцениваются в соответствии с логикой нулевой гипотезы.
- С помощью рассмотренных статистик можно строить доверительные интервалы.

Спасибо за внимание !!!