Введение в экономико-математическое моделирование

Лекция 19. Проверка гипотез

канд. физ.-матем. наук, доцент Д. В. Чупраков

usr10381@vyatsu.ru

Структура лекции

Статистическая гипотеза

Определеине

Статистическая гипотеза это предположение

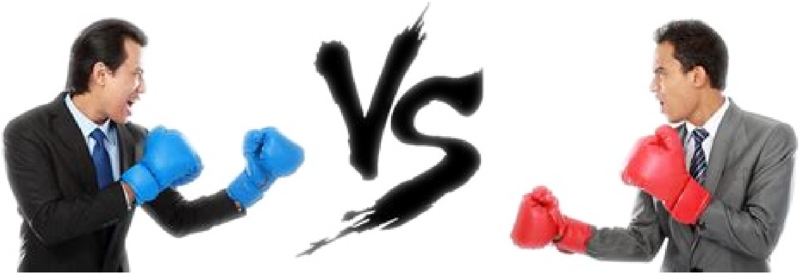
- о виде распределения генеральной совокупности или
- о величинах неизвестных параметров известного распределения генеральной совокупности,

которое может быть проверено на основании выборочных показателей.

По количеству предположений гипотезы делятся на:

- простые — это гипотезы, содержащие только одно предположение;
- сложные — гипотезы, состоящие из конечного или бесконечного числа простых гипотез.

Mortal Combat



Нулевая гипотеза H_0 — гипотеза, подлежащая проверке.

Конкурирующая (альтернативная) гипотеза H_1 — любое утверждение, которое противоречит нулевой гипотезе.

Нулевая гипотеза — утверждение, принимаемое по умолчанию.

Проверяя статистическую гипотезу исследователь пытается показать несостоятельность нулевой гипотезы, несогласованность её с имеющимися опытными данными, то есть отвергнуть гипотезу.

При этом подразумевается, что должна быть принята другая, альтернативная (конкурирующая), исключающая нулевую гипотезу.

Отвергнуть нулевую гипотезу — значит сделать вывод, что конкурирующая гипотеза H_1 лучше описывает реальность, чем нулевая гипотеза H_0

Презумпция невиновности

Для нулевой гипотезы действует своеобразная „презумпция невиновности“':

Нулевая гипотеза считается верной, пока не будет доказано обратное (нулевая гипотеза отвергнута) сверх необходимых сомнений (т. е. в статистически значимой степени).

Истинность нулевой гипотезы невозможно доказать, но можно показать, что в данный момент нет причин сомневаться в ней.

Статистический критерий

Статистический критерий — правило, которое позволяет на основе имеющихся данных отвергнуть нулевую гипотезу.

- **Параметрические** — критерии, которые служат для проверки гипотез о параметрах распределений генеральной совокупности (чаще всего нормального распределения).
- **Непараметрические** – критерии, которые для проверки гипотез не используют предположений о распределении генеральной совокупности. Эти критерии не требуют знания параметров распределений.
- **Критерии согласия** — служат для проверки гипотез о согласии распределения генеральной совокупности, из которой получена выборка, с ранее принятой теоретической моделью (чаще всего нормальным распределением).

Статистика критерия

В основе критерия лежит статистика критерия — искусственно сконструированная функция

$$T_n = T(X_1, X_2, \ldots, X_n)$$

от выборки $X_1, X_2, \ldots, X_n$.

- Статистика критерия является случайной величиной.
- Закон распределения статистики критерия должен быть известнен!

Обозначение статистики

В зависимости от закона распределения статистику обозначают через:

- U или Z, если она имеет нормальное распределение;
- F или v^2 — распределение Фишера;
- χ^2 — распределение «хи квадрат»;
- t — распределение Стьюдента.

Критическая область

Множество всех значений статистики критерия разбивается на два непересекающихся подмножества:

- Критическую область — включает значения статистики, появление которых при справедливости H_0 практически невозможно.
- Область допустимых значений (область принятия гипотезы) — значения которые может принимать статистика при условии справедливости нулевой гипотезы H_0;

- Статистика подбирается так, чтобы область допустимых значений и критическая область были интервалами.
- Вид критической области зависит от типа альтернативной гипотезы.

Отвержение и принятие гипотезы

Условие отвержения гипотезы

Если значение статистики попадает в критическую область, то гипотеза H_0 отвергается в пользу альтернативной.

Условие согласия гипотезы

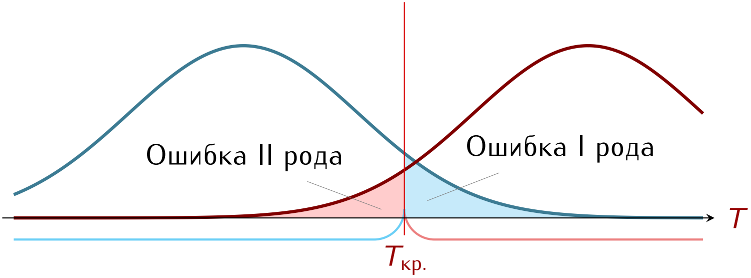
Если значение статистики попадает в область допустимых значений, то гипотеза H_0 не противоречит наблюдаемым значениям, поэтому нет оснований отвергать ее.

Нулевая гипотеза принимается только волевым решением исследователя.

Матрица ошибок

	H_0 верна	H_0 неверна
H_0 принята	Верное решение	Ошибка II рода
H_0 отвергнута	Ошибка I рода	Верное решение

- Ошибка первого рода — отвержение верной гипотезы H_0.
- Ошибка второго рода — принятие ошибочной гипотезы H_0.



Уровень значимости

Определеине

Вероятность ошибки первого рода называется уровнем значимости α.

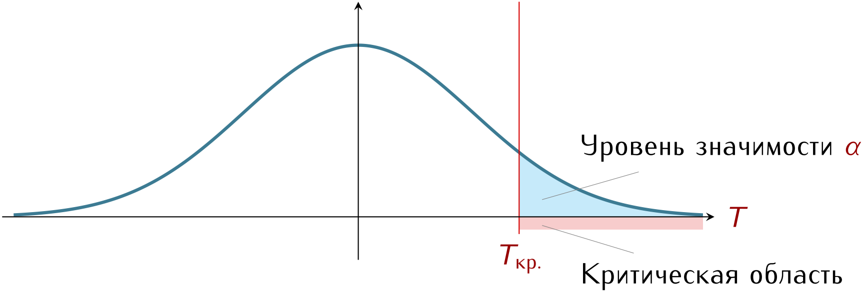
Уровень значимости α устанавливается из значений следующего ряда:

$$0.05, \quad 0.01, \quad 0.005, \quad \ldots$$

события с такими вероятностями считаются практически невозможными.
Допустимая величина уровня значимости определяется теми последствиями, которые наступают после совершения ошибки.

Критическое значение

Так как область допустимых значений и критическая область являются интервалами, то существует граничная точка, разделяющая их

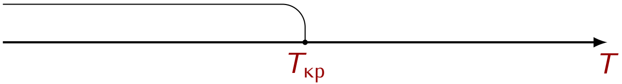


Критическое значение статистики — граница области допустимых значений статистики, при условии, что нулевая гипотеза H_0 верна.

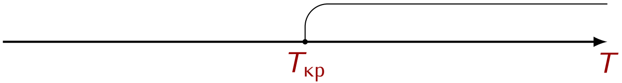
Типы критической области

Односторонняя:

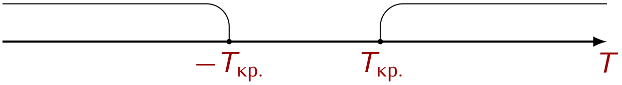
- Левосторонняя — определяется $P(T < T_{кр}) = \alpha$



- Правосторонняя — определяется $P(T > T_{кр}) = \alpha$



Двухсторонняя — определяется $P(T > |T_{кр}|) = \dfrac{\alpha}{2}$



Мощность критерия

Определеине

Мощность критерия — вероятность попадания критерия в критическую область при условии, что верна конкурирующая гипотеза.

Если β — вероятность ошибки второго рода, то мощность критерия равна $1-\beta$.

Чем больше мощность критерия, тем меньше вероятность совершить ошибку второго рода.

После выбора уровня значимости α следует строить критическую область так, чтобы мощность критерия была максимальной.

Алгоритм проверки гипотез

1. Формулируются гипотезы H_0 и H_1.
2. По виду гипотезы выбирается статистический критерий T;
3. Выбирается уровень значимости критерия α. Он равен вероятности допустить ошибку первого рода.
4. По выборочным данным вычисляется вычисляется наблюдаемое значение сатистики $T_{\text{набл.}}$
5. По уровню значимости α вычисляется критическое значение $T_{\text{кр}}$, разделяющее критическую область и область допустимых значений.
6. Определяется неравенство, задающее критическую область.

- Если $T_{\text{набл.}}$ попадает в критическую область, то нулевая гипотеза отвергается.
- Если $T_{\text{набл.}}$ попадает в область допустимых значений, то нулевая гипотеза не противоречит наблюдаемым данным.

Задача о костях

Бросаются две игральные кости. Задайте случайную величину „Число выпавших очков“.

s_i	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Сыграем в кости?

Пусть было совершено 144 броска двух игральных костей: И получен следующий статистический ряд числа очков

s_i	2	3	4	5	6	7	8	9	10	11	12
n_i	2	4	10	12	22	29	21	15	14	9	6

Как Вы считаете, кости утяжелены или нет?

Сыграем в кости?

Подсчитаем Каково было ожидаемое число выпадений каждого количества очков при $n = 144$ бросках

Значение s_i	2	3	4	5	6	7	8	9	10	11	12
Вероятность p_i	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$
Ожид. число np_i	4	8	12	16	20	24	20	116	12	8	4
Наблюд. число n_i	2	4	10	12	22	29	21	15	14	9	6

Насколько вероятно, что кости утяжелены?

Критерий согласия I

Оценим, как велико суммарное отклонение наблюдаемых значений от теоретических.

$$V = (n_1 - np_1) + (n_2 - np_2) + \ldots + (n_{11} - np_{11}).$$

Подставим значения и получим $V = 0$.
Отклонения скомпенсировали друг друга!

Возьмем квадраты отклонений:

$$V = (n_1 - np_1)^2 + (n_2 - np_2)^2 + \ldots + (n_{11} - np_{11})^2$$

Подставим значения и получим $V =$.
Плохие кости приведедут к невероятно большому значению V.

Проблема: Отклонения в разности $(n_7 - np_7)^2$ будут встречаться чаще, чем в разности $(n_1 - np_1)^2$ так как вероятность выпадения 7 очков в шесть раз больше, чем выпапение 2 очков.

Скомпенсируем вклад каждого слагаемого:

$$\chi^2 = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} + \ldots + \frac{(n_{11} - np_{11})^2}{np_{11}}$$

Подсчитаем:

$$\chi^2 = \frac{(2-4)^2}{4} + \frac{(4-8)^2}{8} + \ldots + \frac{(6-4)^2}{4} = 7\frac{7}{48}$$

Теперь вопрос: $7\frac{7}{48}$ — это „невероятно большое число“ или все же нет?

Критерий согласия χ^2 Пирсона

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

- n — объем выборки;
- k — число интервалов разбиения выборки;
- n_i — число значений выборки, попавших в i–й интервал;
- np_i — теоретическая частота попадания значений случайной величины X в i–й интервал.

Замечание

При использовании критерия согласия χ^2 достаточно большими должны быть как общее число опытов $n > 100$, так и значения в отдельных интервалах $n_i \geqslant 5$.
Если для некоторых интервалов условие $n_i \geqslant 5$ нарушается, то соседние интервалы объединяются в один.

Применение критерия χ^2 Пирсона

Для проверки гипотезы о законе распределения генеральной совокупности

1. По гистограмме выбирается наиболее подходящее гипотетическое распределение.
2. Формулируется гипотеза H_0 — „генеральная совокупность подчинена выбранному закону распределения“.
3. Находятся теоретические вероятности наблюдаемых значений.
4. По формуле $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ вычисляют наблюдаемое значение критерия $\chi^2_{\text{набл.}}$
5. По известному уровню значимости α и числу степеней свободы k в таблице значений χ^2 находят теоретическое значение $\chi_{\alpha,k}$.
6. Если $\chi^2_{\text{набл.}} > \chi_{\alpha,k}$, то гипотеза H_0 отвергается.

Пример

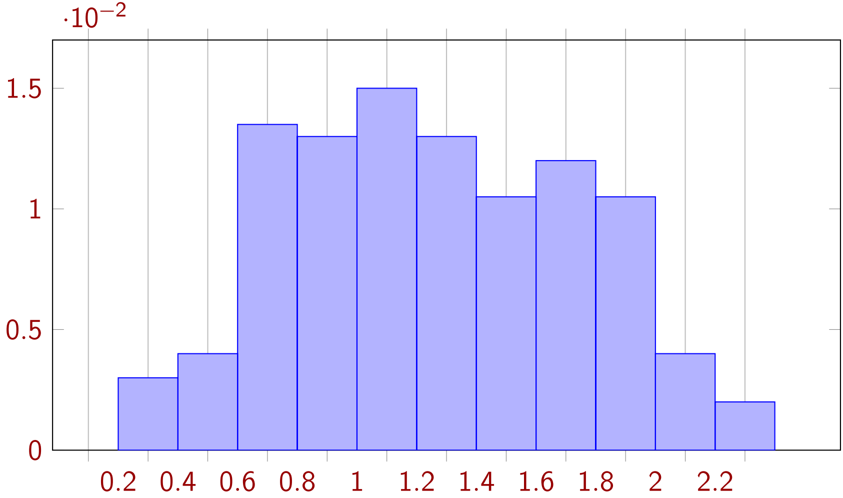
По данным выборки выдвинуть гипотезу о распределении генеральной совокупности и проверить ее при уровне значимости 0.05.

x_i	n_i
0.2–0.4	6
0.4–0.6	8
0.6–0.8	27
0.8–1.0	26
1.0–1.2	30
1.2–1.4	26
1.4–1.6	21
1.6–1.8	24
1.8–2.0	21
2.0–2.2	8
2.2–2.4	4

Для определения типа закона распределения построим гистограмму: $\Delta x_i = 0.1$

x_i	x_i^*	n_i	w_i	$\frac{w_i}{\delta x_i}$
0.2–0.4	0.3	6	0.03	0.003
0.4–0.6	0.5	8	0.04	0.004
0.6–0.8	0.7	27	0.135	0.0135
0.8–1.0	0.9	26	0.13	0.013
1.0–1.2	1.1	30	0.15	0.015
1.2–1.4	1.3	26	0.13	0.013
1.4–1.6	1.5	21	0.105	0.0105
1.6–1.8	1.7	24	0.12	0.012
1.8–2.0	1.9	21	0.105	0.0105
2.0–2.2	2.1	8	0.04	0.004
2.2–2.4	2.3	4	0.02	0.002

Решение. Выдвижение гипотезы II



Предположим, что закон распределения нормальный.

Для проверки гипотезы склеим два последних столюца, так как в последнем число измерений меньше 5.
Вычислим среднее выборочное и выборочную дисперсию

x_i	x_i^*	n_i	w_i	$x_i^* w_i$	$(x_i^*)^2 w_i$
0.2–0.4	0.3	6	0.03	0.009	0.0027
0.4–0.6	0.5	8	0.04	0.02	0.01
0.6–0.8	0.7	27	0.135	0.0945	0.06615
0.8–1.0	0.9	26	0.13	0.117	0.1053
1.0–1.2	1.1	30	0.15	0.165	0.1815
1.2–1.4	1.3	26	0.13	0.169	0.2197
1.4–1.6	1.5	21	0.105	0.1575	0.23625
1.6–1.8	1.7	24	0.12	0.204	0.3468
1.8–2.0	1.9	20	0.10	0.19	0.361
2.0–2.4	2.2	12	0.06	0.132	0.2904
Σ		200	1	1.258	1.8198

Решение. Параметры распределения II

- Выборочное среднее $\bar{x} = 1.258$
- Выборочная дисперсия $s^2 = 1.8198 - 1.258^2 \approx 0.237$
- Исправленная выборочная дисперсия $\hat{s}^2 = \frac{n}{n-1} s^2 = \frac{200}{199} 0.237 = 0.238$
- Исправленное квадратичное отклонение $\hat{s} = \sqrt{\hat{s}^2} = 0.488$

Итак, гипотеза:

$$H_0 : X \sim N(1.258, 0.488)$$

Решение. Проверка гипотезы I

По критерию согласия χ^2 Пирсона проверим гипотезу:

$$H_0 \colon X \sim N(1.258, 0.488)$$

Расчитаем теоретические вероятности попадения в промежуток

$$P(x_i \leqslant x \leqslant x_{i-1}) = \Phi\left(\frac{x_{i+1} - \bar{x}}{s}\right) - \Phi\left(\frac{x_i - \bar{x}}{s}\right)$$

Вычисление функции Φ проведем с помощью функции Excel: `=ГАУСС()`.

x_i	x_i^*	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
0.2–0.4	0.3	6	0.024	0.270
0.4–0.6	0.5	8	0.049	0.359
0.6–0.8	0.7	27	0.085	5.817
0.8–1.0	0.9	26	0.125	0.048
1.0–1.2	1.1	30	0.154	0.023
1.2–1.4	1.3	26	0.162	1.248
1.4–1.6	1.5	21	0.144	2.096
1.6–1.8	1.7	24	0.108	0.251
1.8–2.0	1.9	20	0.069	2.749
2.0–2.4	2.2	12	0.055	0.109
Σ		200	0.975	12.969

Вычислим наблюдаемое значение χ^2

Решение. Проверка гипотезы III

$$\chi^2_{\text{набл.}} = \sum_{i=1}^{10} \frac{(n_i - np_i)^2}{np_i} =$$

$$= 0.270 + 0.359 + 5.817 + 0.048 + 0.023 +$$

$$+ 1.248 + 2.096 + 0.251 + 2.749 + 0.109 =$$

$$= 12.969$$

Вычислим критическое значение χ^2 по уровню значимости $\alpha = 0.05$ и числу степеней свободы $k = 10 - 1 = 9$
Воспользуемся функцией Excel: =`ХИ2ОБР(0,05; 9)`.

$$\chi^2\text{кр.}_{=}\chi^2_{0.05;9} = 16.919.$$

Решение. Проверка гипотезы IV

Сравним значения

$$\chi^2\text{набл.} = 12.969 < 16.919 = \chi^2\text{кр.}$$

Таким образом, наблюдаемые данные не противоречат гипотезе о нормальном распределении $N(1.258, 0.488)$ генеральной совокупности.

Гипотеза об однородности выборки

Пусть есть две независимые выборки с неизвестными функциями распределения $F_1(x)$ и $F_2(x)$.
Гипотеза H_0 — эти выборки из одной и той же генеральной совокупности:

$$H_0 \colon F_1(x) = F_2(x).$$

Альтернативная гипотеза H_1 — эти выборки из разных генеральных совокупностей:

$$H_1 \colon F_1(x) \neq F_2(x).$$

Критерий Колмогорова—Смирнова

1. Наблюдаемое значение вычисляется по формуле

$$\lambda'_{\text{набл.}} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \max \left| F^*_{n_1}(x) - F^*_{n_2}(x) \right|$$

   где
   - n_1, n_2 — объемы выборок
   - $F_{n_1}(x)$, $F_{n_2}(x)$ выборочные функции

2. Критическое значение λ'_α вычисляется на основе уровня значимости α
   - по таблицам Колмогорова—Смирнова при $n_1, n_2 \leqslant 20$;
   - по таблицам распределения Колмогорова при $n_1, n_2 \geqslant 50$.

3. Гипотеза отвергается, если $\lambda'_{\text{набл.}} > \lambda_\alpha$.

Пример

В течение месяца выборочно осуществлялась проверка торговых точек города по продаже овощей. Результаты двух проверок по недовесам покупателям одного вида овощей приведены в таблице:

	0; 10	10; 20	20; 30	30; 40	40; 50	50; 60	60; 70	70; 80	80; 90
n_1	3	10	15	20	12	5	25	15	5
n_2	5	12	8	25	10	8	20	7	5

Можно ли считать при уровне значимости 0,05, что недовесы овощей являются устойчивым и закономерным процессом при продаже овощей в данном городе (т.е. описываются одной и той же функцией распределения)?

Используем критерий Колмогорова—Смирнова для проверки гипотезы $H_0: F_1(x) = F_2(x)$

	0; 10	10; 20	20; 30	30; 40	40; 50	50; 60	60; 70	70; 80	80; 90	Σ
n_1	3	10	15	20	12	5	25	15	5	110
n_2	5	12	8	25	10	8	20	7	5	100
w_1	$\frac{3}{110}$	$\frac{10}{110}$	$\frac{15}{110}$	$\frac{20}{110}$	$\frac{12}{110}$	$\frac{5}{110}$	$\frac{25}{110}$	$\frac{15}{110}$	$\frac{5}{110}$	1
w_2	$\frac{5}{100}$	$\frac{12}{100}$	$\frac{8}{100}$	$\frac{25}{100}$	$\frac{10}{100}$	$\frac{8}{100}$	$\frac{20}{100}$	$\frac{7}{100}$	$\frac{5}{100}$	1
F^*_{110}	$\frac{3}{110}$	$\frac{13}{110}$	$\frac{28}{110}$	$\frac{48}{110}$	$\frac{60}{110}$	$\frac{65}{110}$	$\frac{90}{110}$	$\frac{105}{110}$	$\frac{110}{110}$	
F^*_{100}	$\frac{5}{100}$	$\frac{17}{100}$	$\frac{25}{100}$	$\frac{50}{100}$	$\frac{60}{100}$	$\frac{68}{100}$	$\frac{88}{100}$	$\frac{95}{100}$	$\frac{100}{100}$	

- $\lambda'_{\text{набл.}} = \sqrt{\frac{n_1 n_2}{n_1+n_2}} \max|F_{n_1}(x) - F_{n_2}(x)| = \sqrt{\frac{11000}{210}} 0.089 \approx 0.644.$
- $\lambda'_{\text{кр.}} = \lambda_{0.05} = 1.36$ так как $\alpha = 0.05$, $n_1, n_2 \geqslant 50$.
- Так как $\lambda'_{\text{набл.}} < \lambda_{0.05}$, то наблюдаемые данные не противоречат гипотезе.

Ответ: нет оснований считать недовесы овощей системными.

Исключение грубых ошибок наблюдения

Грубые ошибки могут возникнуть из-за ошибок показаний измерительных приборов, ошибок регистрации, случайного сдвига запятой в десятичной записи числа и т. д.

Пусть $x^*, x_1, x_2, \ldots, x_n$ совокупность наблюдений, причем x^* резко выделяется.

Необходимо решить вопрос принадлежности резко выделяющегося значения x^* к остальным, т.е. проверить гипотезу

$$H_0 : x^* = \bar{x}.$$

В качестве конкурирующей гипотезы берется

- $H_1 : x^* < \bar{x}$ если значение x^* слишком маленькое.
- $H_2 : x^* > \bar{x}$ если x^* слишком большое.

Если гипотеза $H_0\colon x^* = \bar{x}$ справедлива, то статистика

$$t = \frac{\bar{x} - x^*}{\hat{s}}$$

имеет t-распределение Стьюдента с $n-1$ степенью свободы.

Алгоритм:

1. Наблюдаемое значение: $t_{\text{набл.}} = \frac{\bar{x}-x^*}{\hat{s}}$
2. Критическое значение $t_{\alpha,n-1}$ берется из таблицы Стьюдента по уровню значимости α и $n-1$ степеням свободы.
3. Гипотеза $H_0\colon x^* = \bar{x}$ отвергается, если $|t_{\text{набл.}}| > t_{\alpha,n-1}$

Минимальный уровень значимости

- В случае, когда в задаче не дан уровень значимости, возникает естественный вопрос. Какой уровень значимости всё-таки лучше 0.01, 0.02, 0.05 или 0.1? А может другой?
- Кроме того, получается, что ответ зависит от того, какой уровень значимости взяли.
- Допустим, мы не отвергли гипотезу при 0.05 уровне значимости. Но нам хочется знать, с какой вероятностью ошибки первого рода мы её можем отвергнуть. Ошибка в 0.06 может быть вполне допустимой, а ошибка в 0.25 это уж слишком много.
- Нужна величина, которая позволит указать пороговое значение уровня значимости.

Определение

Минимальный уровень значимости p-value – это минимальное значение α, при котором основная гипотеза ещё отвергается.