ОСНОВЫ ХИМИЧЕСКОЙ МЕТРОЛОГИИ
Лекция 3 Проверка статистических гипотез

лектор: Образовский Е. Г.

22 февраля 2015 г.

Ввиду принципиально статистического характера экспериментальных данных мы можем быть уверены в том, что ограниченная выборка обладает тем или иным свойством лишь с некоторой вероятностью.
Рассмотрим в общем виде принципы проверки статистических гипотез относительно выборки объема n экспериментальных результатов $x_1, x_2, ..., x_n$.

Правило, позволяющее отвергнуть некоторую гипотезу H (например, что распределение является нормальным) на основании данных выборки $x_1, \dots, x_n$, называется статистическим критерием. Критерий определяет некоторую область значений – критическую область. Гипотеза H отвергается, если рассчитанный по выборке некоторый параметр попадает в критическую область и не отвергается в противном случае.

Такое принятие или отбрасывание гипотезы не дает логического доказательства или опровержения этой гипотезы. Возможны четыре случая:

- **гипотеза H верна и принимается согласно критерию;**

Такое принятие или отбрасывание гипотезы не дает логического доказательства или опровержения этой гипотезы. Возможны четыре случая:

- **гипотеза H верна и принимается согласно критерию;**
- **гипотеза H неверна и не принимается согласно критерию;**

Такое принятие или отбрасывание гипотезы не дает логического доказательства или опровержения этой гипотезы. Возможны четыре случая:

- **гипотеза H верна и принимается согласно критерию;**
- **гипотеза H неверна и не принимается согласно критерию;**
- **гипотеза H верна, но отвергается согласно критерию(это ошибка первого рода);**

Такое принятие или отбрасывание гипотезы не дает логического доказательства или опровержения этой гипотезы. Возможны четыре случая:

- **гипотеза H верна и принимается согласно критерию;**
- **гипотеза H неверна и не принимается согласно критерию;**
- **гипотеза H верна, но отвергается согласно критерию(это ошибка первого рода);**
- **гипотеза H неверна, но принимается согласно критерию (это ошибка второго рода).**

Если гипотеза полностью фиксирует параметры распределения, то такую гипотезу называют простой или нулевой (сложная гипотеза ограничивает значения параметров некоторой областью значений). Конкурирующую с нулевой гипотезой (H_0) называют альтернативной (H_1). Проиллюстрируем эти рассуждения рис. 1.

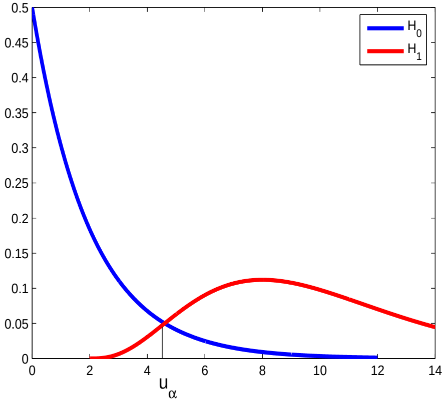


Рис.: Кривая H_0 – плотность вероятности при справедливости нулевой гипотезы; кривая H_1 – плотность вероятности для конкурирующей гипотезы; u_α – критическое значение для величины α вероятности ошибки первого рода

Выбор значения вероятности ошибки первого рода α является вопросом соглашения. В некоторых случаях достаточно взять $\alpha = 0,1$ (например, в случае экспрессного анализа для собственных нужд), в других случаях может быть недостаточно малым и $\alpha = 0,001$ (например, при анализе лекарств).

В большинстве случаев рутинного анализа придерживаются следующего правила:

- **гипотеза принимается, когда ошибка первого рода возможна более чем в 5 % случаев ($P = 1 - \alpha \leqslant 0,95$), тогда различие считается незначимым;**

В большинстве случаев рутинного анализа придерживаются следующего правила:

- **гипотеза принимается, когда ошибка первого рода возможна более чем в 5 % случаев ($P = 1 - \alpha \leqslant 0,95$), тогда различие считается незначимым;**
- **гипотеза отбрасывается, если ошибка первого рода может появиться менее чем в 1 % случаев ($P = 1 - \alpha \geqslant 0,99$); различие считается значимым;**

В большинстве случаев рутинного анализа придерживаются следующего правила:

- **гипотеза принимается, когда ошибка первого рода возможна более чем в 5 % случаев ($P = 1 - \alpha \leqslant 0,95$), тогда различие считается незначимым;**
- **гипотеза отбрасывается, если ошибка первого рода может появиться менее чем в 1 % случаев ($P = 1 - \alpha \geqslant 0,99$); различие считается значимым;**
- **если $0,95 \leqslant P \leqslant 0,99$, то различие интерпретируется как спорное.**

Рис.: К. Пирсон

В качестве примера рассмотрим проверку различных гипотез с помощью χ^2-распределения (распределения Пирсона).

Пусть имеется выборка $x_1, \ldots, x_n$, значения которой подчиняются нормальному закону распределения.

Найдем закон распределения для величины $\chi^2 = \sum_{i=1}^{n} x_i^2$.

χ^2-Распределение

Если имеется одна переменная x_1, то для величины $u_1 = x_1^2$ плотность функции распределения имеет вид $f_1(u_1) \sim u_1^{-1/2} e^{-u_1/2}$. Для числа переменных $n = \nu$ функция распределения имеет вид

$$\mathbf{f}_\nu(\chi^2) = \frac{(\chi^2)^{\nu/2-1}\,\mathbf{e}^{-\chi^2/2}}{2^{\nu/2}\Gamma(\nu/2)}.$$

Это упоминавшееся выше гамма-распределения, а величина ν называется числом степеней свободы. На рис. 3 приведены типичные графики χ^2 распределения для некоторых значений числа степеней свободы.

χ^2-Распределение

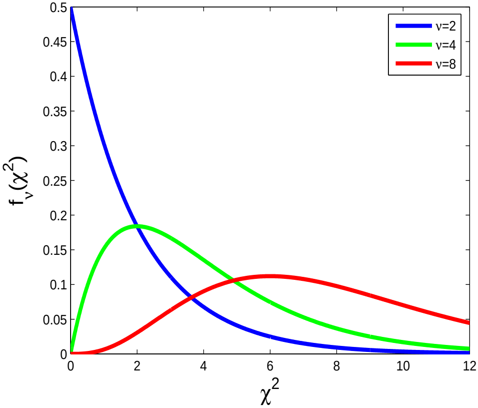
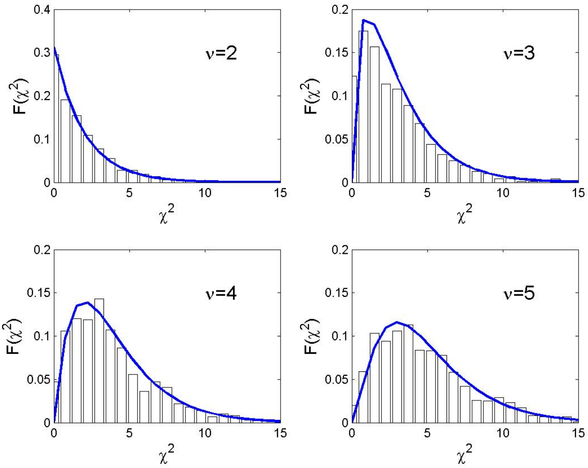


Рис.: χ^2-Распределение для $\nu = 2, 4, 8$

Численное моделирование χ^2-распределения



Как применять χ^2-распределение для проверки гипотезы, что выборка объемом n имеет нормальное распределение? Для этого вычисляют среднее $\bar{x}$ и стандартное отклонение s:

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{1}^{n} x_i, \quad s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

и приводят экспериментальные данные к стандартному виду $u_i = (x_i - \bar{x})/s$. Весь диапазон значений u_i в данной выборке разбивают на k неперекрывающихся интервалов, с расчетом, чтобы в каждый интервал попала не менее пяти значений (один из возможных вариантов приведен ниже).

χ^2-Распределение

Для каждого интервала определяют число h_k попавших в него экспериментальных значений и сравнивают с теоретически ожидаемым h_k^t в предположении верности проверяемой гипотезы и вычисляют величину

$$\chi^2 = \sum_{\mathbf{k}} \frac{(\mathbf{h_k} - \mathbf{h_k^t})^2}{\mathbf{h_k^t}}.$$

Это значение сравнивают с критическим значением $\chi_t^2(P;\nu)$, которое теоретически определяется из вида функции распределения. Если вычисленное по экспериментальным данным χ^2 меньше, чем χ_t^2, то проверяемую гипотезу принимают.

Пример. Имеется $n = 200$ результатов анализа и нужно проверить нормальность распределения. Все результаты разбили на 8 интервалов:

в интервал $x_i \leqslant \bar{x} - 1,5 \cdot \sigma$ попало $h_1 = 14$ результатов;

в интервал $\bar{x} - 1,5 \cdot \sigma < x_i \leqslant \bar{x} - 1,0 \cdot \sigma$ попало $h_2 = 29$ результатов;

в интервал $\bar{x} - 1,0 \cdot \sigma < x_i \leqslant \bar{x} - 0,5 \cdot \sigma$ попало $h_3 = 30$ результатов;

в интервал $\bar{x} - 0,5 \cdot \sigma < x_i \leqslant \bar{x}$ попало $h_4 = 27$ результатов;

в интервал $\bar{x} < x_i \leqslant \bar{x} + 0,5 \cdot \sigma$ попало $h_5 = 28$ результатов;
в интервал $\bar{x} + 0,5 \cdot \sigma < x_i \leqslant \bar{x} + 1,0 \cdot \sigma$ попало $h_6 = 31$ результатов;
в интервал $\bar{x} + 1,0 \cdot \sigma < x_i \leqslant \bar{x} + 1,5 \cdot \sigma$ попало $h_7 = 28$ результатов;
в интервал $x_i \geqslant \bar{x} + 1,5 \cdot \sigma$ попало $h_8 = 13$ результатов.
Эти результаты представлены на рис. 4

χ^2-Распределение

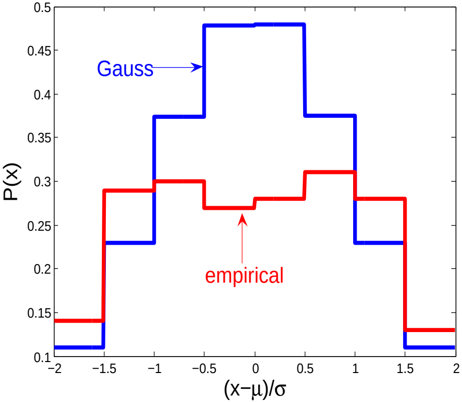


Рис.: Сравнение экспериментальной частоты с теоретической гауссовой

Рассчитывается ожидаемое число результатов h_{it} в каждом интервале i в предположении нормальности распределения. Например:

$$\mathbf{h_{1t} = n \cdot \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-1,5} e^{-x^2/2} dx, \quad h_{2t} = n \cdot \frac{1}{\sqrt{2\pi}} \int\limits_{-1,5}^{-1,0} e^{-x^2/2} dx, \ldots.}$$

Обычно табулированы значения только для положительных величин верхнего предела интегрирования. Для определения значений интеграла с отрицательными верхними пределами следует использовать свойство

$$\mathbf{F = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-u} e^{-x^2/2} dx = 1 - \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+u} e^{-x^2/2} dx.}$$

χ^2-Распределение

Вычисляем

$$\chi^2 = \sum_i \frac{(h_i - h_i^t)^2}{h_i^t} = 17,6$$

и сравниваем с критическим значением $\chi_t^2(P = 0,99; \nu = 8 - 3) = 15,1$ Поскольку экспериментальное значение χ^2 превышает теоретически ожидаемое, гипотезу о нормальности распределения отвергаем.

Численное моделирование проверки нормальности распределения

Численное моделирование проверки нормальности распределения для большего числа интервлов

Сравнение результатов проверки гипотезы о нормальности распределения для данных имеющих нормальное и треугольное распределения

Это распределение применяется для проверки гипотезы о совпадении двух различных оценок стандартного отклонения. В качестве критерия используется величина

$$F_{\nu_1,\nu_2} = \frac{S_1^2}{S_2^2} \quad \text{при условии } S_1 > S_2,$$

где

$$S_1^2 = \frac{1}{\nu_1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2, \quad \nu_1 = n_1 - 1,$$

$$S_2^2 = \frac{1}{\nu_2}\sum_{i=1}^{n_2}(y_i - \bar{y})^2, \quad \nu_2 = n_2 - 1.$$

Рис.: Р. Фишер

F-Распределение Фишера

Плотность распределения значений F дается формулой

$$\mathbf{f}(\mathbf{F};\nu_1,\nu_2) = \mathbf{C}\frac{\mathbf{F}^{(\nu_1/2)-1}}{(\nu_1\mathbf{F}+\nu_2)^{(\nu_1+\nu_2)/2}},$$

где нормировочный множитель C равен

$$\mathbf{C} = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}(\nu_1)^{\nu_1/2}(\nu_2)^{\nu_2/2}.$$

Максимальное значение плотности распределения Фишера достигает при

$$\mathbf{F_0} = \frac{\nu_2(\nu_1-2)}{\nu_1(\nu_2+2)}.$$

Пример распределения Фишера представлен на рис. 6 для случая $\nu_1 = 4$, $\nu_2 = 6$.

F-Распределение Фишера

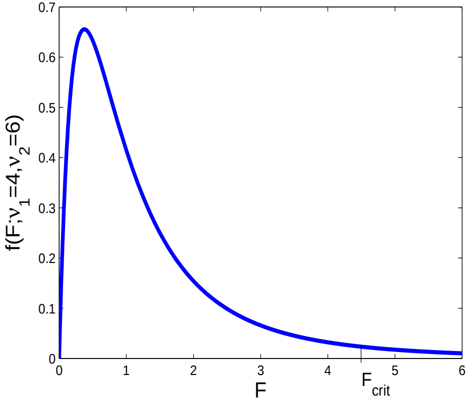


Рис.: F-Распределение Фишера для $\nu_1 = 4, \nu_2 = 6$

F-Распределение Фишера

Критическое значение величины $F_{кр}$ для заданного значения доверительной вероятности $P = 1 - \alpha$ находится из условия

$$\int\limits_{\mathbf{F}_{кр}}^{\infty} \mathbf{f}(\mathbf{F}; \nu_1, \nu_2)\mathbf{dF} = \alpha.$$

Эти величины приведены в виде таблиц. Для случая $\nu_1 = 2$ можно получить аналитическое выражение для $F_{кр}$:

$$\mathbf{F}_{кр} = \frac{\nu_2}{2}\left[\left(\frac{1}{\alpha}\right)^{2/\nu_2} - 1\right].$$

Пример[1]. Имеется две выборки: $x_1 = 0,49$, $x_2 = 0,45$, $x_3 = 0,45$ и $y_1 = 0,52$, $y_2 = 0,55$, $y_3 = 0,50$, $y_4 = 0,52$. Стандартные отклонения для этих выборок равны $S_1 = 0,023$, $S_2 = 0,021$. Проверка по критерию Фишера

$$\mathbf{F = \frac{S_1^2}{S_2^2} = \frac{0,023^2}{0,021^2} = 1,20 < F_{tab}(P = 0,95; \nu_1 = 2, \nu_2 = 3) = 9,55}$$

показывает, что различие незначимо.

[1] Здесь (и далее) в промежуточных вычислениях количество значащих цифр берется на одну больше, чем в исходных данных, чтобы уменьшить ошибки округления.

F-Распределение Фишера

t-Распределение Стьюдента

Рис.: У. Госсет ("Стьюдент")

Применяется для проверки гипотезы о совпадении средних значений двух выборок. По двум наборам значений $x_1, \ldots, x_{n_1}$ и $y_1, \ldots, y_{n_2}$ вычисляют средние значения и дисперсии:

$$\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i, \quad S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2,$$

$$\bar{y} = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(y_j - \bar{y})^2$$

и проверяют совпадение оценок S_1 и S_2 по критерию Фишера.

Если $F = S_1^2/S_2^2 < F_{tab}(\nu_1 = n_1 - 1, \nu_2 = n_2 - 1)$, то по закону сложения ошибок находят

$$\mathbf{S_{\bar{x}-\bar{y}}} = \sqrt{\frac{\mathbf{S_1^2}}{\mathbf{n_1}} + \frac{\mathbf{S_2^2}}{\mathbf{n_2}}}$$

и вычисляют величину

$$\mathbf{t} = \frac{\bar{\mathbf{x}} - \bar{\mathbf{y}}}{\mathbf{S_{\bar{x}-\bar{y}}}}$$

Теоретическое распределение имеет вид

$$\mathbf{f(t,\nu)} = \frac{\mathbf{C_t}}{(\mathbf{1} + \mathbf{t^2}/\nu)^{(\nu+\mathbf{1})/\mathbf{2}}},$$

где нормировочный коэффициент C_t равен

$$\mathbf{C_t} = \frac{\mathbf{1}}{\sqrt{\pi\nu}} \frac{\mathbf{\Gamma}\left(\frac{\nu+\mathbf{1}}{\mathbf{2}}\right)}{\mathbf{\Gamma}\left(\frac{\nu}{\mathbf{2}}\right)},$$

а число степеней свободы $\nu = n_1 + n_2 - 2$ и показано на рис. 8 для случая $\nu = 4$.

t-Распределение Стьюдента

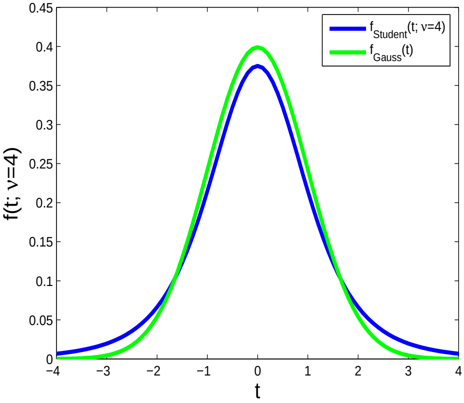


Рис.: t-Распределение Стьюдента для $\nu = 4$. Для сравнения приведено также нормальное распределение

t-Распределение Стьюдента

В случае значимого различия оценок стандартных отклонений вычисляют средневзвешенную величину

$$\mathbf{t} = \frac{\bar{\mathbf{x}} - \bar{\mathbf{y}}}{\sqrt{\frac{\mathbf{S}_1^2}{\mathbf{n}_1} + \frac{\mathbf{S}_2^2}{\mathbf{n}_2}}}$$

и сравнивают ее со средневзвешенным критическим значением

$$\mathbf{t}' = \frac{\mathbf{w}_1 \mathbf{t}_1 + \mathbf{w}_2 \mathbf{t}_2}{\mathbf{w}_1 + \mathbf{w}_2},$$

где

$$\mathbf{t}_1 = \mathbf{t}(\mathbf{n}_1; \mathbf{P} = 0,95), \quad \mathbf{w}_1 = \frac{\mathbf{S}_1^2}{\mathbf{n}_1}; \qquad \mathbf{t}_2 = \mathbf{t}(\mathbf{n}_2; \mathbf{P} = 0,95), \quad \mathbf{w}_2 = \frac{\mathbf{S}_2^2}{\mathbf{n}_2}$$

Пример. Необходимо установить, значимо ли различие среднего значения выборки из $n = 10$ результатов анализа, приведенных в таблице.

Номер **i**	**Содержание** **x_i, %**	**Номер** **i**	**Содержание** **x_i, %**
1	1,60	6	1,76
2	1,60	7	1,78
3	1,67	8	1,78
4	1,70	9	1,81
5	1,73	10	1,81

и аттестованным значением $\mu = 1,67$ %.

По экспериментальным данным находим среднее $\bar{x} = 1,72$ % и стандартное отклонение $S = 0,079$ %. Поскольку

$$\mathbf{t} = \frac{|1,72 - 1,67|}{0,079/\sqrt{10}} = 2,15 < \mathbf{t}(\mathbf{P} = 0,95; \nu = 9) = 2,26,$$

можно сделать заключение о незначимости отличия среднего значения от аттестованного содержания.

t-Распределение Стьюдента

Критерий Колмогорова–Смирнова.

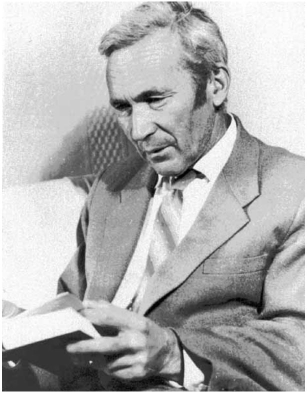


Рис.: А. Н. Колмогоров

Для проверки гипотез относительно свойств небольших выборок применяют критерий Колмогорова-Смирнова. Проверим являются ли нормально распределенными следующие результаты измерений:

$x_1 = 4.71$, $x_2 = 4.46$, $x_3 = 5.59$,
$x_4 = 4.68$, $x_5 = 5.24$, $x_6 = 4.52$,
$x_7 = 5.11$, $x_8 = 4.80$, $x_9 = 4.26$.

Критерий Колмогорова–Смирнова.

Приведя эти данные к стандартному виду $u_i = (x_i - \bar{x})/s$, получим $u_1 = -1.33$, $u_2 = -0.85$, $u_3 = -0.71$, $u_4 = -0.33$, $u_5 = -0.26$, $u_6 = -0.04$, $u_7 = 0.69$, $u_8 = 1.00$, $u_9 = 1.83$.
Эмпирическое распределение имеет ступенчатый характер с увеличением вероятности на величину $1/n$ (где n–число измерений) при $u = u_i$, как показано на рис. 10.

Критерий Колмогорова–Смирнова.

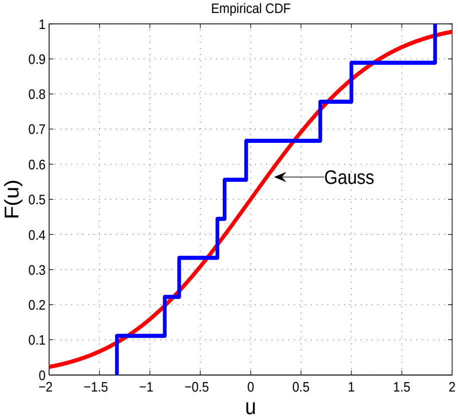


Рис.: Эмпирическое распределение для $n = 9$ измерений (ступенчатая кривая) и нормальное распределение (гладкая кривая).

Критерий Колмогорова–Смирнова.

Разность абсолютных значений эмпирическим и теоретическим распределениями не должна превышать значений, указанных в таблице

число измерений n	**Критическая разность $d(n, P = 0.95)$**	**число измерений n**	**критическая разность $d(n, P = 0.95)$**
3	0,376	12	0,242
4	0,375	13	0,234
5	0,343	14	0,226
6	0,323	15	0,219
7	0,304	16	0,213
8	0,288	17	0,207
9	0,274	18	0,202
10	0,261	19	0,197
11	0,251	20	0,192

Критерий Колмогорова–Смирнова.