

# Проверка статистических гипотез

Понятие статистической гипотезы. Проверка гипотез о параметрах законов распределения.

# Понятие статистической гипотезы

- **Статистической гипотезой** (или просто **гипотезой**) называется любое предположение о генеральной совокупности, проверяемое по выборке.
- Если гипотеза содержит некоторое утверждение о параметрах распределения случайной величины (когда сам закон распределения считается известным), то она называется **параметрической**, и **непараметрической** – в иных случаях.
- Нулевой (основной) гипотезой  $H_0$  называется предположение, которое выдвигается изначально, пока наблюдения не заставят признать обратное.
- **Альтернативной (конкурирующей)** гипотезой  $H_1$  называется гипотеза, которая противоречит нулевой гипотезе  $H_0$  и которую принимают, если отвергнута основная гипотеза.
- Гипотезы бывают **простые** (содержащие только одно предположение) и **сложные** (состоящие из конечного или бесконечного числа простых гипотез).

# Задачи статистической проверки гипотез:

- Относительно некоторой генеральной совокупности высказывается та или иная гипотеза  $H$ .
- Из этой генеральной совокупности извлекается выборка.
- Необходимо указать правило, с помощью которого можно было по выборке ответить на вопрос о том, следует ли отклонить гипотезу  $H$  или принять её.
- Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость её проверки.

**!** Статистическими методами гипотезу можно только **опровергнуть или не опровергнуть**, но не доказать.

# Проверка статистических гипотез

- Имея две гипотезы  $H_0$  и  $H_1$ , необходимо на основе выборочных данных либо принять основную гипотезу  $H_0$ , либо конкурирующую  $H_1$ .
- Правило, по которому принимается решение принять или отклонить гипотезу  $H_0$  (или  $H_1$ ), называется **статистическим критерием** (или просто **критерием**) проверки гипотезы  $H_0$ .
- **Статистикой** (или **тестом**) критерия называют случайную величину  $T$ , которая служит для проверки статистических гипотез.

# Общая схема проверки статистических гипотез

1. Для основной гипотезы  $H_0$  формулируется альтернативная гипотеза  $H_1$ .
2. Выбирается уровень значимости проверки – малое число  $\alpha > 0$ .
3. Рассматриваются теоретические выборки значений случайных величин, о которых сформулирована гипотеза  $H_0$ , и выбирается (формируется) случайная величина  $t$ . Значения и распределение  $t$  полностью определяются по выборкам при предположении о верности гипотезы  $H_0$ . (обычно  $t$  выбирают из перечисленных ниже:

$U$  – нормальное распределение,

$\chi^2$  – распределение Пирсона,

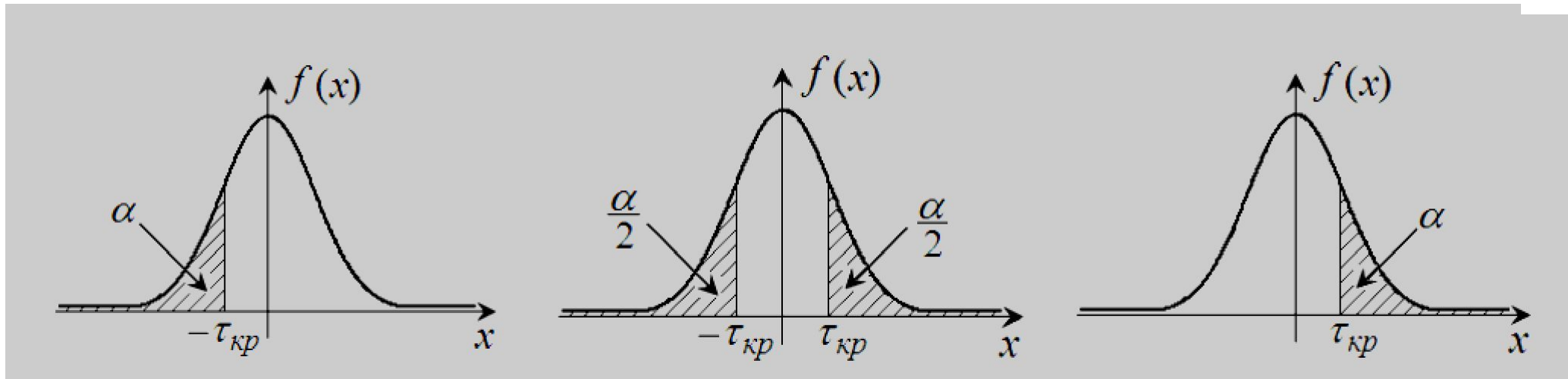
$T$  – Стьюдента,

$F$  – Фишера-Снедекора)

4. На числовой оси задают интервал  $D$ , такой, что вероятность попадания случайной величины  $t$  в этот интервал:  $P(t \in D) = 1 - \alpha$ .

Интервал  $D$  называется **областью принятия гипотезы  $H_0$** , а оставшаяся область числовой оси – **критической областью** (величина  $t = t_{кр}$  – **критическое значение теста проверки**).

Различают три типа критических областей. Критическая область определяется с учётом гипотез:



5. По реализациям анализируемых выборок вычисляется конкретное (наблюдаемое) значение теста  $T$  (обозначим его  $T = T_{\text{набл}}$ ) и проверяется выполнение условия  $P(T \in D) = 1 - \alpha$ :

а) если оно выполняется (например,  $T_{\text{набл}} < T_{\text{кр}}$  для правосторонней области), то гипотеза  $H_0$  принимается в том смысле, что она не противоречит опытным данным и нет оснований её отвергнуть;

б) если условие не выполняется ( $T_{\text{набл}} > T_{\text{кр}}$  для правосторонней области), то полагается, что гипотеза  $H_0$  неверна и её отвергают.

Для каждого критерия имеются соответствующие таблицы, по которым и находят критическое значение, удовлетворяющее приведённым выше соотношениям.

*Принятие гипотезы  $H_0$  следует расценивать не как раз и навсегда установленный, абсолютно верный содержащийся в ней факт, а лишь как достаточно правдоподобное, не противоречащее опыту утверждение.*

# Ошибки при проверке гипотез

<i>ошибка I рода</i>	<i>ошибка II рода</i>
Отвергается основная (нулевая) гипотеза, хотя она верна.	Отвергается конкурирующая гипотеза, хотя она верна.
Вероятность ошибки $P(H_1 H_0) = \alpha$ , <i><math>\alpha</math> – уровень значимости критерия</i> (обычно $\alpha = 0,05 ; 0,01; 0,005; 0,001$ ).	Вероятность ошибки $P(H_0 H_1) = \beta$ (величина $\beta$ , как правило, заранее неизвестна)
Вероятность принять верную (нулевую) гипотезу $P(H_0 H_0) = 1 - \alpha$ .	Вероятность принять верную (конкурирующую) гипотезу $P(H_1 H_1) = 1 - \beta,$ <i><math>(1 - \beta)</math> – мощность критерия.</i>



# Гипотеза о математическом ожидании нормального распределения при известной дисперсии генеральной совокупности

- Пусть генеральная совокупность  $X$  распределена по нормальному закону.
- Генеральная средняя  $a$  хотя и неизвестна, но есть основания предполагать, что она равна предполагаемому значению  $a_0$ .
- Необходимо проверить гипотезу  $H_0: a = a_0$  против альтернативной:  $H_1: a \neq a_0$ , или  $H_1: a < a_0$ , или  $H_1: a > a_0$ .

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$a = a_0$	$U = \frac{\bar{x} - a_0}{\sigma} \sqrt{n}$	$a \neq a_0$	$ U  < u_{кр},$ $\Phi(u_{кр}) = 0,5 - \frac{\alpha}{2}$
$\sigma^2 = \sigma_2^2$ известно		$a < a_0$	$U > -u_{кр},$ $\Phi(u_{кр}) = 0,5 - \alpha$
		$a > a_0$	$U < u_{кр},$ $\Phi(u_{кр}) = 0,5 - \alpha$

# Пример 1

- Из нормальной генеральной совокупности с известным средним квадратическим отклонением  $\sigma = 5$  извлечена выборка объёма  $n=100$ , и по ней найдено выборочное среднее 26,5. Требуется на уровне значимости  $\alpha=0,05$  проверить гипотезу  $H_0: a=a_0=25$  против альтернативной гипотезы  $H_1: a \neq a_0$ .
- Решение. Найдем значение статистики критерия

$$U = \frac{\bar{x} - a_0}{\sigma} \sqrt{n} = \frac{26,5 - 25}{5} \sqrt{100} = 3.$$

$$\Phi(u_{кр}) = 0,5 - \frac{0,05}{2} = 0,475$$

из соотн  $u_{кр} = 1,96$   
Лапласа

находим по таблице

Т. к.  $|U| > u_{кр}$ , то основная гипотеза отвергается.

# Гипотеза о математическом ожидании нормального распределения при неизвестной дисперсии генеральной совокупности

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$a = a_0$  $\sigma^2 = \sigma_2^2$ неизвестно	$T = \frac{\bar{x} - a_0}{s} \sqrt{n}$	$a \neq a_0$	$ T  < t_{кр}$ , $t_{кр} = t_{\alpha, n-1}$ для двусторонней области
		$a < a_0$	$T > -t_{кр}$ , $t_{кр} = t_{\alpha, n-1}$ для односторонней области
		$a > a_0$	$T < t_{кр}$ , $t_{кр} = t_{\alpha, n-1}$ для односторонней области

$s$  – исправленное среднее квадратическое отклонение.

Значение  $t_{кр}$  находим по таблице Стьюдента

## Пример 2

- По выборке объёма  $n = 16$ , извлечённой из нормальной генеральной совокупности, найдены  $\bar{x} = 12,4$  и  $s = 1,2$ . Требуется при уровне значимости  $0,05$  проверить нулевую гипотезу  $H_0: a = 11,8$  при конкурирующей гипотезе  $H_1: a \neq 11,8$ .
- Решение: Найдём наблюдаемое значение статистики критерия

$$T = \frac{\bar{x} - a_0}{s} \sqrt{n} = \frac{12,4 - 11,8}{1,2} \sqrt{16} = 2$$

Поскольку конкурирующая гипотеза имеет вид  $a \neq a_0$ , то искомая критическая область двусторонняя. Из таблицы критических точек распределения Стьюдента найдём по уровню значимости  $\alpha = 0,05$  и числу степеней свободы  $k = n - 1 = 15$  критическую точку  $t_{кр} = t_{кр}(0,05; 15) = 2,13$ .

Т. к.  $|T| < t_{кр}$ , то нет оснований отвергнуть нулевую гипотезу.

# Гипотеза о сравнении генеральных дисперсий нормального распределения

- **Гипотезы о дисперсиях** возникают достаточно часто, так как дисперсия характеризует такие исключительно важные показатели, как точность машин, приборов, технологических процессов, степень однородности совокупностей, риск, связанный с отклонением доходности активов от ожидаемого уровня, и т. д.

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$\sigma_x^2 = \sigma_y^2$  $a_x$ и $a_y$ неизвестны	$F = \frac{s_{\max}^2}{s_{\min}^2}$	$\sigma_x^2 \neq \sigma_y^2$	$F < F_{кр}$ , $F_{кр} = F_{\alpha/2, n_1-1, n_2-1}$
		$\sigma_x^2 > \sigma_y^2$	$F < F_{кр}$ , $F_{кр} = F_{\alpha, n_1-1, n_2-1}$

## Пример 3

Измерения одной и той же физической величины проведены двумя методами. Получены следующие результаты:

В первом случае:  $x_1=9,6$ ;  $x_2=9,8$ ;  $x_3=10,0$ ;  $x_4=10,2$ ;  $x_5=10,6$ .

Во втором случае:  $y_1=10,4$ ;  $y_2=9,7$ ;  $y_3=10,0$ ;  $y_4=10,3$ .

Предполагается, что результаты измерений распределены в выборке нормально и выборки независимы. Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости  $\alpha=0,1$ ?

# Решение примера 3

Будем судить о точности методов по величине дисперсии.

Основная гипотеза  $H_0: D(X)=D(Y)$

Альтернативная гипотеза  $H_1: D(X)\neq D(Y)$

Найдем исправленные выборочные дисперсии:  $s_X^2 = 0,148$ ;  $s_Y^2 = 0,1$

Находим статистику:  $F_{\text{набл}} = \frac{0,148}{0,1} = 1,48$

Критическая область двусторонняя, поэтому по уровню значимости  $\alpha/2=0,05$  и числам степеней свободы  $k_1=5-1=4$ ,  $k_2=4-1=3$  находим критическую точку  $F_{\text{кр}}(0,05; 4; 3)=9,12$ .

Т.к.  $F_{\text{набл}} < F_{\text{кр}}$ , то нет оснований отвергать нулевую гипотезу.  
Следовательно, оба метода обеспечивают одинаковую точность измерений.

# Проверка гипотез о равенстве двух средних нормальных генеральных совокупностей, дисперсии которых известны (большие независимые выборки)

Имеются две независимые выборки больших объемов ( $n_1 > 30, n_2 > 30$ ), по которым найдены выборочные средние. Генеральные дисперсии  $D(X), D(Y)$  известны.

Необходимо проверить на уровне значимости  $\alpha$  нулевую гипотезу  $H_0$ .

$M(X) =$

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$a_x = a_y$  $\sigma_x^2$ и $\sigma_y^2$ известны	$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}}$	$a_x \neq a_y$	$ U  < u_{кр},$ $\Phi(u_{кр}) = 0,5 - \frac{\alpha}{2}$
		$a_x < a_y$	$U > -u_{кр},$ $\Phi(u_{кр}) = 0,5 - \alpha$
		$a_x > a_y$	$U < u_{кр},$ $\Phi(u_{кр}) = 0,5 - \alpha$



## Пример 4

Для проверки эффективности новой технологии отобраны две группы рабочих: в первой группе численностью  $n_1=50$  чел., где применялась новая технология, выборочная средняя выработка составила  $\bar{x}=85$  (изделий), во второй группе численностью  $n_2=70$  чел. выборочная средняя  $\bar{y}=78$  (изделий). Предварительно установлено, что дисперсии выработки в группах равны соответственно  $\sigma_x^2=100$  и  $\sigma_y^2=74$ .

На уровне значимости  $\alpha=0,05$  выяснить влияние новой технологии на среднюю производительность.

# Решение примера 4

Проверяемая гипотеза  $H_0: a_x = a_y$ , т. е. средние выработки рабочих одинаковы по новой и старой технологиям.

В качестве конкурирующей гипотезы можно взять  $H_1: a_x > a_y$

Находим фактическое значение статистики критерия

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} = \frac{85 - 78}{\sqrt{\frac{100}{50} + \frac{74}{70}}} = 4.$$

При альтернативной гипотезе  $H_1$  по таблице Лапласа из

$$\Phi(u_{кр}) = 0,5 - 0,05 = 0,45$$

$$u_{кр} = 1,64$$

найдем критическое значение

Т.к.  $U > u_{кр}$ , то гипотеза  $H_0$  отвергается, можно сделать вывод, что новая технология позволяет повысить среднюю выработку рабочих.

Проверка гипотезы о равенстве двух средних нормальных генеральных совокупностей, дисперсии которых неизвестны и одинаковы (малые независимые выборки)

Имеются две независимые выборки малых объемов ( $n_1 < 30, n_2 < 30$ ), по которым найдены выборочные средние и исправленные выборочные дисперсии. Генеральные дисперсии  $D(X), D(Y)$  неизвестны, но предполагаются одинаковыми.

Необходимо проверить на уровне значимости  $\alpha$  нулевую гипотезу  $H_0$ :

$M(X) = M(Y)$

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$a_x = a_y$ $\sigma_x^2$ и $\sigma_y^2$ неизвестны, но равны	$T = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $s = \sqrt{\frac{s_x^2 \cdot (n_1 - 1) + s_y^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}$	$a \neq a_0$	$ T  < t_{кр}$ , $t_{кр} = t_{\alpha, n_1 + n_2 - 2}$ для двусторонней области
		$a < a_0$	$T > -t_{кр}$ , $t_{кр} = t_{\alpha, n_1 + n_2 - 2}$ для односторонней области
		$a > a_0$	$T < t_{кр}$ , $t_{кр} = t_{\alpha, n_1 + n_2 - 2}$ для односторонней области

# Пример 5

Из двух партий изделий, изготовленных на двух одинаково настроенных станках, извлечены малые выборки, объемы которых  $n = 10$  и  $m = 12$ . Получены следующие результаты:

<i>Первый станок</i>					<i>Второй станок</i>			
Контролируемый размер изделий $x_i$	3,4	3,5	3,7	3,9	Контролируемый размер изделий $y_i$	3,2	3,4	3,6
Частота (число изделий) $n_i$	2	3	4	1	Частота (число изделий) $m_i$	2	2	8

При условии значимости 0,02 проверим гипотезу  $H_0: M(X) = M(Y)$  о равенстве средних размеров изделий при конкурирующей гипотезе  $H_1: M(X) \neq M(Y)$ . Предполагается, что случайные величины  $X$  и  $Y$  распределены нормально.

# Решение примера 5

Найдем выборочные средние и исправленные дисперсии для каждой выборки:

	3,6		0,0267
	3,5		0,0255

Для рассматриваемого критерия Стьюдента предполагается, что генеральные дисперсии одинаковы, поэтому надо сравнить дисперсии, используя критерий Фишера – Снедекора при  $H_1 : D(X) \neq D(Y)$ .

$F_{\text{набл}} = 0,0267 / 0,0255 = 1,05$ . По таблице имеем  $F_{\text{кр}}(0,01; 9; 11) = 4,63$ .

Так как  $F_{\text{набл}} < F_{\text{кр}}$ , дисперсии различаются незначимо.

Теперь вычислим наблюдаемое значение критерия Стьюдента.  $T_{\text{набл}} = 1,45$ .

По уровню значимости 0,02 и числу степеней свободы  $k = n + m - 2 = 10 + 12 - 2 = 20$  находим по таблице Стьюдента критическую точку  $t_{\text{дв.кр}}(0,02; 20) = 2,53$ .

Так как  $T_{\text{набл}} < t_{\text{дв.кр}}$ , нет оснований отвергнуть нулевую гипотезу; следовательно, средние размеры изделий существенно не различаются.

# Проверка гипотезы о равенстве двух средних нормальных генеральных совокупностей с неизвестными дисперсиями (зависимые выборки)

Пусть генеральные совокупности  $X$  и  $Y$  распределены нормально, причем их дисперсии неизвестны.

Из этих совокупностей извлечены зависимые выборки одинакового объема  $n$ , варианты которых соответственно равны  $X_i$  и  $Y_i$ .

Введем обозначения:

$d_i = X_i - Y_i$  - разности вариант с одинаковыми номерами;

$\bar{d} = \sum d_i / n$  - средняя разностей вариант с одинаковыми номерами;

$s_d = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2 / n}{n-1}}$  - «исправленное» среднее квадратическое отклонение.

Для того чтобы проверить нулевую гипотезу  $H_0: M(X) = M(Y)$  о равенстве двух средних нормальных совокупностей  $X$  и  $Y$  с неизвестными дисперсиями в случае двух зависимых выборок одинакового объема при конкурирующей гипотезе  $H_1: M(X) \neq M(Y)$ , надо:

- 1) вычислить наблюдаемое значение критерия  $T_{\text{набл}} = \bar{d} \frac{\sqrt{n}}{s_d}$
- 2) по таблице критических точек распределения Стьюдента по заданному уровню значимости  $\alpha$  и числу степеней свободы  $k = n - 1$  найти критическую точку  $t_{\text{дв.к}}(\alpha, k)$ :
  - если  $|T_{\text{набл}}| < t_{\text{дв.к}}(\alpha, k)$ , нет оснований отвергнуть нулевую гипотезу;
  - если  $|T_{\text{набл}}| > t_{\text{дв.к}}(\alpha, k)$ , нулевую гипотезу отвергают.

# Пример 6

Двумя приборами в одном и том же порядке измерены шесть деталей и получены следующие результаты измерений (мкм):

$x_i$	2	3	5	6	8	10
$y_i$	10	3	6	1	7	4

При уровне значимости 0,05 проверим нулевую гипотезу о равенстве результатов измерений в предположении, что они распределены нормально.

Решение:  $d_1 = -8$ ,  $d_2 = 0$ ,  $d_3 = -1$ ,  $d_4 = 5$ ,  $d_5 = 1$ ,  $d_6 = 6$ .

Выборочная средняя  $\bar{d} = 0,5$ , «исправленное» среднее квадратическое отклонение,  $s_d = 5,01$

$T_{\text{набл}} = 0,24$ .

По таблице находим критическую точку  $t_{\text{дв.кр}}(0,05; 5) = 2,57$ .



# Сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности

Пусть  $n$  - объем выборки, по которой найдена исправленная дисперсия  $s^2$ .

При заданном уровне значимости  $\alpha$  проверить нулевую гипотезу  $H_0: \sigma^2 = \sigma_0^2$  о равенстве неизвестной генеральной дисперсии  $\sigma^2$  гипотетическому (предполагаемому) значению  $\sigma_0^2$ .

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$\sigma^2 = \sigma_0^2$ $a$ неизвестно	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 \neq \sigma_0^2$	$\chi_{1-\frac{\alpha}{2}; n-1}^2 < \chi^2 < \chi_{\frac{\alpha}{2}; n-1}^2$
		$\sigma^2 < \sigma_0^2$	$\chi^2 > \chi_{1-\alpha; n-1}^2$
		$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi_{\alpha; n-1}^2$

Если число степеней свободы  $k > 30$ , то критическую точку  $\chi_{\text{кр}}^2(\alpha, k)$  можно найти из равенства Уилсона — Гильферти:

$$\chi_{\text{кр}}^2(\alpha, k) = k \left[ 1 - \frac{2}{9k} + z_{\alpha} \sqrt{\frac{2}{9k}} \right]^3,$$

где  $z_{\alpha}$  находят, используя функцию Лапласа, из равенства  $\Phi(z_{\alpha}) = \frac{1 - 2\alpha}{2}$ .

# Пример 7

Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать  $\sigma_0^2 = 0,1$ . Взята проба из 25 случайно отобранных изделий. Получены следующие результаты измерений:

Контролируемый размер изделий пробы $x_i$	3,0	3,5	3,8	4,4	4,5
Частота $n_i$	2	6	9	7	1

При уровне значимости 0,05 проверим, обеспечивает ли станок требуемую точность.

# Решение примера 7

$$H_0: \sigma^2 = 0,1.$$

$$H_1: \sigma^2 > 0,1.$$

Найдем исправленную выборочную дисперсию.

$$s^2 = 0,1975$$

Найдем наблюдаемое значение критерия:  $\chi^2_{\text{набл}} = 48$

Найдем по таблице критическую точку:  $\chi^2_{\text{кр}}(0,05; 24) = 36,4.$

Имеем  $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$ , следовательно, нулевую гипотезу отвергаем, т.е. станок не обеспечивает необходимую точность и требует наладки.

# Сравнение наблюдаемой относительной частоты с гипотетической вероятностью появления события

Пусть по достаточно большому числу  $n$  независимых испытаний, в каждом из которых вероятность  $p$  появления события постоянна, но неизвестна, найдена относительная частота  $m/n$  (где  $m$  - число появлений события).

Требуется при заданном уровне значимости  $\alpha$  проверить нулевую гипотезу, состоящую в том, что неизвестная вероятность  $p$  равна гипотетической вероятности  $p_0$ , т.е.  $H_0: p = p_0$

$H_0$	Статистика критерия	$H_1$	Область принятия $H_0$
$p = p_0$ достаточно большие $n$ , $np_0 > 5$ , $nq_0 > 5$ , $q_0 = 1 - p_0$	$U = \frac{p^* - p_0}{\sqrt{p_0 q_0}} \sqrt{n},$ $p^* = \frac{m}{n}$	$p \neq p_0$	$ U  < u_{кр},$ $\Phi(u_{кр}) = 0,5 - \frac{\alpha}{2}$
		$p < p_0$	$U > -u_{кр},$ $\Phi(u_{кр}) = 0,5 - \alpha$
		$p > p_0$	$U < u_{кр},$ $\Phi(u_{кр}) = 0,5 - \alpha$

## Пример 8

Партия изделий принимается, если вероятность того, что изделие окажется бракованным, не превышает 0,02. Среди случайно отобранных 480 изделий оказалось 12 дефектных. Можно ли принять партию?

Решить самостоятельно.

# Проверка гипотезы о равенстве дисперсий нескольких нормальных генеральных совокупностей по выборкам одинакового объема (критерий Кочрена)

Пусть генеральные совокупности  $X_1, X_2, \dots, X_m$  распределены нормально.

Из этих совокупностей извлечены  $m$  независимых выборок одинакового объема  $n$  и по ним найдены исправленные выборочные дисперсии  $s_1^2, s_2^2, \dots, s_m^2$ , все с одинаковым числом степеней свободы  $k = n - 1$ .

Требуется на уровне значимости  $\alpha$  проверить нулевую гипотезу об однородности дисперсий, т.е. гипотезу  $H_0: D(X_1) = D(X_2) = \dots = D(X_m)$  о равенстве между собой генеральных дисперсий.

В качестве критерия проверки нулевой гипотезы примем **критерий Кочрена (Кохрена)** - отношение максимальной исправленной дисперсии к сумме всех исправленных дисперсий.

Для этого надо:

1) вычислить *наблюдаемое значение критерия*

$$G_{\text{набл}} = \frac{s_{\text{max}}^2}{s_1^2 + s_2^2 + \dots + s_l^2};$$

2) по таблице критических точек распределения Кочрена найти *критическую точку*  $G_{\text{кр}}(\alpha, k, l)$ :

- если  $G_{\text{набл}} < G_{\text{кр}}$ , нет оснований отвергать нулевую гипотезу;
- если  $G_{\text{набл}} > G_{\text{кр}}$ , нулевую гипотезу отвергают.

При условии однородности дисперсий независимых выборок одинакового объема в качестве оценки генеральной дисперсии принимают **среднюю арифметическую исправленных дисперсий**.



# Проверка гипотезы о равенстве дисперсий нескольких нормальных генеральных совокупностей по выборкам различного объема (критерий Бартлетта)

Пусть генеральные совокупности  $X_1, X_2, \dots, X_m$  распределены нормально.

Из этих совокупностей извлечены  $m$  независимых выборок, вообще говоря, различных объемов  $n_i$  (некоторые объемы могут быть одинаковыми; если все выборки имеют одинаковый объем, используется критерий Кочрена).

По выборкам найдены исправленные выборочные дисперсии  $s_1^2, s_2^2, \dots, s_m^2$ .

Требуется на уровне значимости  $\alpha$  проверить нулевую гипотезу об однородности дисперсий, т.е. гипотезу  $H_0: D(X_1) = D(X_2) = \dots = D(X_m)$  о равенстве между собой генеральных дисперсий.

Введем обозначения:

$k_i = n_i - 1$  — число степеней свободы дисперсии  $s_i^2$ ;

$k = \sum_{i=1}^l k_i$  — сумма чисел степеней свободы;

$(\bar{s})^2 = \frac{1}{k} \sum_{i=1}^l k_i s_i^2$  — средняя арифметическая исправлен-

ных дисперсий, взвешенная по числам степеней свободы;

$$V = 2,303 \left( k \lg(\bar{s})^2 - \sum_{i=1}^l k_i \lg s_i^2 \right);$$

$$C = 1 + \frac{1}{3(l-1)} \left( \sum_{i=1}^l \frac{1}{k_i} - \frac{1}{k} \right).$$

Для того чтобы при заданном уровне значимости  $\alpha$  проверить нулевую гипотезу об однородности дисперсий нормальных совокупностей, надо:

1) вычислить наблюдаемое значение критерия Бартлетта

$$B_{\text{набл}} = V/C;$$

2) по таблице критических точек распределения  $\chi^2$  по уровню значимости  $\alpha$  и числу степеней свободы  $m - 1$  (где  $m$  - число выборок) найти критическую точку  $\chi^2_{\text{кр}}(\alpha, m-1)$  правосторонней критической области:

- если  $B_{\text{набл}} < \chi^2_{\text{кр}}$ , нет оснований отвергать нулевую гипотезу;
- если  $B_{\text{набл}} > \chi^2_{\text{кр}}$ , нулевую гипотезу отвергают.

При условии однородности дисперсий в качестве оценки генеральной дисперсии принимают среднюю арифметическую исправленных дисперсий, взвешенную по числам степеней свободы:

$$(\bar{s})^2 = \frac{1}{k} \sum_{i=1}^m k_i s_i^2.$$

# Пример 9

По четырем независимым выборкам объемом  $n_1 = 17$ ,  $n_2 = 20$ ,  $n_3 = 15$ ,  $n_4 = 16$ , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии, соответственно равные 2,5; 3,6; 4, 1; 5,8. При уровне значимости 0,05 проверить гипотезу об однородности дисперсий и оценить генеральную дисперсию.

Решить самостоятельно

# Сравнение двух вероятностей биномиальных распределений

Пусть в двух генеральных совокупностях проводятся независимые испытания: в результате каждого испытания событие  $A$  может появиться в первой совокупности с неизвестной вероятностью  $p_1$  а во второй - с неизвестной вероятностью  $p_2$ .

По выборкам, извлеченным из первой и второй совокупностей, найдены соответствующие частоты  $w_1(A) = m_1/n_1$  и  $w_2(A) = m_2/n_2$ , где  $m_1, m_2$  - числа появлений события  $A$ ;  $n_1, n_2$  - количество испытаний.

В качестве оценок неизвестных вероятностей примем относительные частоты:  $p_1 = w_1, p_2 = w_2$ .

При заданном уровне значимости  $\alpha$  проверить нулевую гипотезу  $H_0$ :

$p_1 = p_2 = p$  о равенстве вероятностей появления события в двух генеральных совокупностях, имеющих биномиальные распределения,

наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}};$$

при конкурирующей гипотезе  $H_1: p_1 \neq p_2$  по таблице функции Лапласа находим критическую точку  $u_{\text{кр}}$  из равенства  $\Phi(u_{\text{кр}}) = \frac{1 - \alpha}{2}$

- если  $|U_{\text{набл}}| < u_{\text{кр}}$ , нет оснований отвергать нулевую гипотезу;
- если  $|U_{\text{набл}}| > u_{\text{кр}}$ , нулевую гипотезу отвергают.

При конкурирующей гипотезе  $H_1: p_1 > p_2$  критическую точку правосторонней критической области находят из равенства  $\Phi(u_{\text{кр}}) = \frac{1 - 2\alpha}{2}$

Та же формула используется и для левосторонней области.

# Пример 10

За смену отказали 15 элементов устройства А, состоящего из 800 элементов, и 25 элементов устройства В, состоящего из 1000 элементов. При уровне значимости  $\alpha = 0,05$  проверить гипотезу  $H_0: p_1 = p_2 = p$  о равенстве вероятностей отказа элементов обоих устройств при конкурирующей гипотезе  $H_1: p_1 \neq p_2$

Решить самостоятельно.