

Статистической гипотезой называется любое утверждение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин.

Проверка статистической гипотезы состоит в том, чтобы сформулировать такое правило, которое позволило бы по результатам проведенных наблюдений принять или отклонить гипотезу. Правило, согласно которому гипотеза принимается или отвергается, называется *критерием* проверки статистической гипотезы.

Гипотезу, которую мы проверяем, будем называть *основной* или *нулевой* гипотезой, и будем всегда обозначать H_0 . Альтернативные или *конкурирующие* гипотезы будем обозначать $H_1, H_2, \dots, H_m$.

С проверкой статистических гипотез связывают ошибки двух типов. *Ошибкой первого рода* называют событие, когда верная проверяемая гипотеза отвергается критерием. *Ошибкой второго рода* называют событие, когда неверная проверяемая гипотеза принимается критерием. Вероятности ошибок первого и второго рода обозначают α и β , соответственно. Вероятность ошибки второго рода зависит от выдвигаемой конкурирующей гипотезы. Вероятность отклонения ложной проверяемой гипотезы, т.е. принятия правильного решения в пользу конкурирующей, называется *мощностью*, и она равна $1 - \beta$. Вероятность ошибки первого рода также называют *уровнем значимости* критерия.

5.1. Виды статистических гипотез

Обычно статистические гипотезы делят на следующие виды: *однородности*, если имеется две или более выборок случайных величин; *независимости*, если имеется выборка многомерной случайной величины; *случайности*, если есть предположения о независимости и одинаковом распределении наблюдений в выборке; *о виде распределения*, если есть предположения о законе распределения случайной величины (рисунок 5.1).

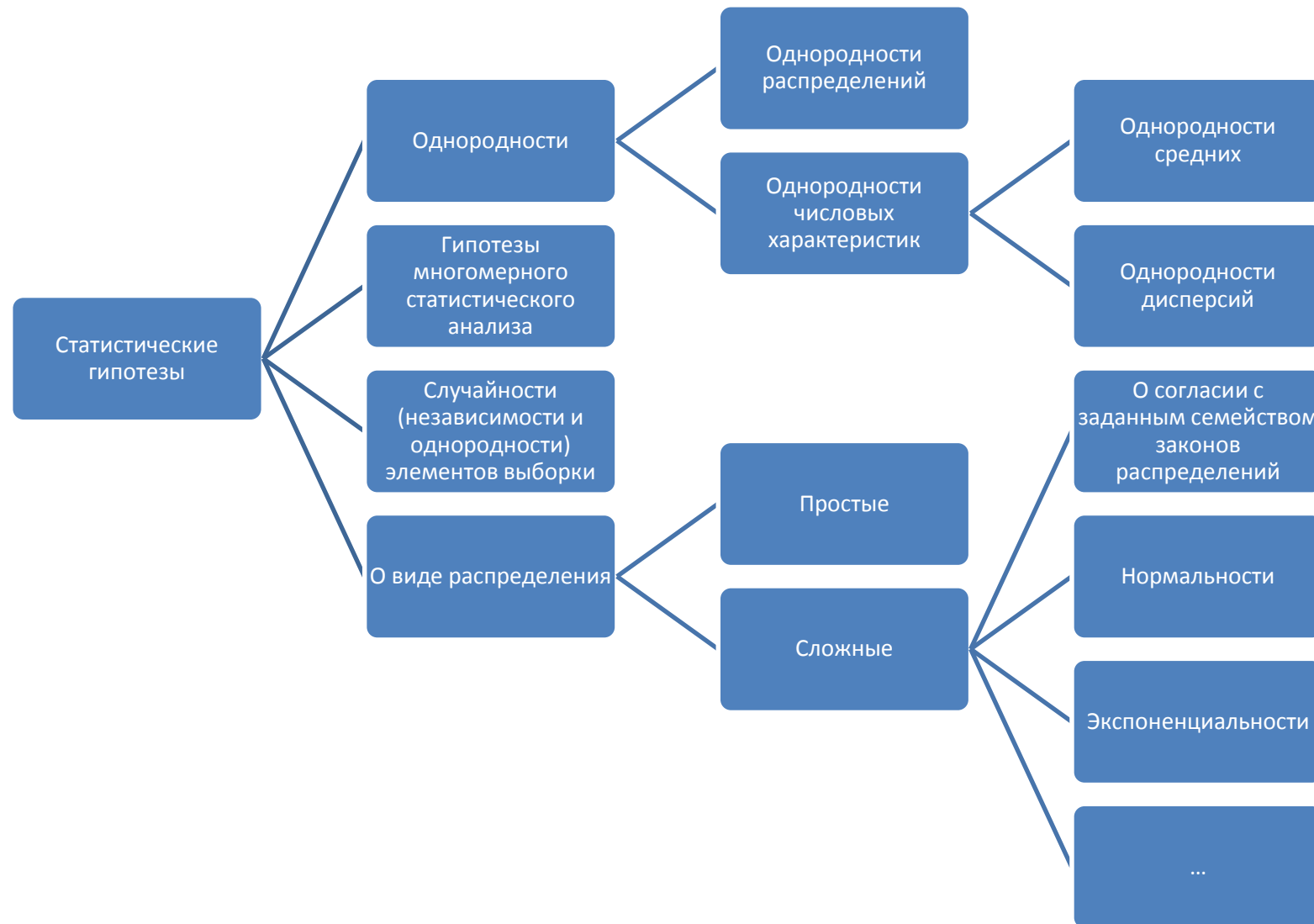


Рисунок 5.1 – Классификация статистических гипотез

1. Гипотеза о виде распределения

Пусть имеется выборка $X_n = \{X_1, \dots, X_n\}$ наблюдаемой случайной величины с функцией распределения $F_{\xi}(x)$.

а) *Простой* гипотезой является утверждение $H_0: F_{\xi}(x) = F(x)$, где $F(x)$ полностью задана.

б) *Сложной* гипотезой является утверждение $H_0: F_{\xi}(x) \in F = \{F(x, \theta), \theta \in \Theta\}$.

2. Гипотеза однородности

Пусть произведено k серий независимых наблюдений $\mathbb{X}_{n_1} = \{X_1^1, \dots, X_n^1\}$, $\mathbb{X}_{n_2} = \{X_1^2, \dots, X_n^2\} \dots \mathbb{X}_{n_k} = \{X_1^n, \dots, X_n^n\}$ и пусть $F_i(x)$ – функция распределения i -й серии. Чтобы проверить менялось ли распределение от серии к серии можно сформулировать гипотезу *однородности*

$$H_0: F_1(x) = F_2(x) = \dots = F_k(x),$$

при этом само распределение $F(x)$ может быть неизвестным.

3. Гипотеза независимости

В эксперименте наблюдается двумерная случайная величина $\xi = (\xi_1, \xi_2)$ с неизвестной функцией распределения $F_\xi(x, y)$ и есть основания предполагать, что ξ_1 и ξ_2 независимы. В этом случае нужно проверить гипотезу *независимости*:

$$H_0: F_\xi(x, y) = F_{\xi_1}(x)F_{\xi_2}(y),$$

где $F_{\xi_1}(x)$ и $F_{\xi_2}(y)$ некоторые одномерные функции распределения.

4. Гипотеза случайности

Результат эксперимента описывается n -мерной случайной величиной $\bar{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ с некоторой функцией распределения $F_{\bar{\xi}}(x_1, \dots, x_n)$. Чтобы проверить утверждение о том, что все наблюдения независимы и одинаково распределены, можно сформулировать гипотезу *случайности*:

$$H_0: F_{\bar{\xi}}(x_1, \dots, x_n) = F_{\xi}(x_1)F_{\xi}(x_2) \dots F_{\xi}(x_n)$$

Любой критерий проверки статистической гипотезы разбивает выборочное пространство на доверительную область $\mathcal{X}_0$ и критическую область $\mathcal{X}_1$. При попадании выборки в критическую область гипотеза отвергается, а при попадании в доверительную область – принимается. Чаще всего такое разбиение производится с помощью одномерной *статистики* – функции от выборки, поэтому критическая и доверительная область формулируются уже как подмножества множества вещественных чисел.

Доверительная область включает такие значения статистики критерия, при которых гипотеза принимается, а критическая область – значения, при которых гипотеза отвергается. Кроме того, вероятность попадания выборки (статистики критерия) в критическую область, когда гипотеза верна, по определению равна вероятности ошибки первого рода, а вероятность попадания выборки (статистики критерия) в доверительную область, когда гипотеза не верна, равна вероятности ошибки второго рода.

Как правило, встречаются три вида критических областей для статистики критерия:

- правосторонняя критическая область (t_α, ∞) ;
- левосторонняя $(-\infty, t_\alpha)$;
- двусторонняя $(-\infty, t_{\alpha_1}) \cup (t_{\alpha_2}, \infty)$, $\alpha_1 + \alpha_2 = \alpha$.

5.2. Вычисление достигаемого уровня значимости

Достижимый уровень значимости (p-value) определяется как вероятность попадания статистики критерия:

- в область $(S(X_n), \infty)$, если критическая область правосторонняя;
- в область $(-\infty, S(X_n))$, если критическая область левосторонняя;

где $S(X_n)$ - вычисленное значение статистики по реализации выборки.

Гипотеза отвергается, если достигаемый уровень значимости оказывается меньше заданной вероятности ошибки первого рода.

Если критическая область двусторонняя, то однозначного способа вычисления достигаемого уровня значимости нет. Например, можно вычислять p -value как $2 \cdot \min(p, 1 - p)$, где $p = P\{S \in (S(X_n), \infty) | H_0\}$.

Достижимый уровень значимости является случайной величиной, определенной на интервале $[0, 1]$, на основании которой делается статистический вывод о принятии гипотезы. Чем ближе значение p -value к 1, тем больше оснований для принятия гипотезы, чем ближе значение p -value к 0, тем больше оснований для отвержения гипотезы. Однако следует помнить о следующем важном замечании относительно p -value.

5.3. Выбор критерия проверки статистической гипотезы

Для проверки одной и той же гипотезы, как правило, существует несколько различных статистических критериев $T_1, T_2, \dots, T_k$. Выбор подходящего статистического критерия, вообще говоря, не является тривиальной задачей. Можно сформулировать следующие принципы подбора статистических критериев.

1. Должны выполняться “стандартные” предположения, обуславливающие возможность применения рассматриваемого критерия (например, о виде распределения случайной величины и о наблюдаемых данных). Так, например, нельзя применять критерий Колмогорова по группированным данным, или по наблюдениям дискретной случайной величины.

2. Критерий должен быть *состоятельным*, т.е. его мощность должна стремиться к единице с ростом объема выборки.
3. Критерий должен быть *несмещенным*, т.е. мощность должна быть больше, чем вероятность ошибки первого рода.
4. Критерий должен обладать наибольшей мощностью при заданном объеме выборки и заданном уровне значимости критерия.

Добиться выполнения последнего принципа на практике не представляется возможным, потому что построить наиболее мощный критерий удастся только в очень редких случаях, например, когда основная и конкурирующая гипотезы являются *простыми*. Чаще всего, для разных конкурирующих гипотез, для разных уровней значимости, для разных объемов выборки, более мощными оказываются разные критерии.

В этой ситуации для выбора оптимального критерия можно применить классическую теорию *принятия решений в условиях неопределенности*. Критерии являются стратегиями, конкурирующие гипотезы – состояниями среды, функция полезности $u(T_i, H_j)$ – это мощность критерия (таблица 5.1). Другим способом определения функции полезности при выборе критерия может быть стоимость проведения эксперимента по различению основной и конкурирующей гипотез с заданными вероятностями ошибок первого и второго рода.

Таблица 5.1 – Матрица полезности выбора критерия T_i при конкурирующих гипотезах H_j

$T_i \setminus H_j$	H_1	H_2	...	H_m
T_1	$u(T_1, H_1)$	$u(T_1, H_2)$	...	$u(T_1, H_m)$
T_2	$u(T_2, H_1)$	$u(T_2, H_2)$	...	$u(T_2, H_m)$
...	...	...	...	...
T_k	$u(T_k, H_1)$	$u(T_k, H_2)$	...	$u(T_k, H_m)$

Существуют разные подходы к выбору оптимальной стратегии при принятии решения в условиях неопределенности. В случае, когда нет никакой информации о том, какая конкурирующая гипотеза может быть верна, рациональным выглядит выбор критерия по правилу Вальда (известны также такие названия как «критерий крайнего пессимиста» или «критерий осторожного наблюдателя»):

$$T^* = \arg \max_{T_i} \min_{H_j} u(T_i, H_j).$$

Критерий, выбранный по правилу Вальда, максимизирует полезность против самой «неудобной» конкурирующей гипотезы.