

Проверка статистических гипотез

Вопрос 1. Общая схема проверки статистических гипотез

Статистической гипотезой H_0 (нулевой гипотезой) называется любое предположение относительно закона распределения исследуемой случайной величины X.

Гипотезы бывают простые и сложные. **Простия гипотеза** содержит одно утверждение и полностью определяет закон распределения величины X в отличии от сложной, когда проверяется несколько утверждений. Гипотезы бывают **параметрическими**, если выдвигается предположение о параметрах распределения при известном законе и **непараметрическими**, если предположение выдвигается о самом виде закона распределения.

Наряду с выдвинутой гипотезой Н₀ рассматривают противоречащую ей гипотезу Н₁ (*альпернативную гипотезу*). Если выдвинутая гипотеза Н₀ будет отвергнута, то имеет место альтернативная гипотеза Н₁.

Критерием проверки статистической гипотезы называется некоторое правило, позволяюще принять ее или отвергнуть.

Причем критерии строятся с помощью случайной величины К (часто именно ее называют критерием), для которой известно распределение. **Наблюдаемым значением криперия** К_{набл} называют значение критерия, вынисленное по данным выборки.

Параметрические гипотезы проверяются с помощью критериев *значимости*, а непараметрические – с помощью критериев *согласия*.

В случае проверки гипотез возможны ошибки:

Ошибка 1-го рода состоит в том, что будет отвергнута правильная гипотеза. Вероятность ошибки первого рода α называется *уровнем значимости* критерия, по которому производится проверка.

Ошибка 2-го рода состоит в том, что будет принята неправильная гипотеза. Если β — вероятность ошибки второго рода, то величина 1- β называется **мощностью критерия**.

Критической областью называется совокупность значений критерия, при которых нулевую гипотезу отвергают.



Если уровень знач

Если уровень значимости α уже выбран и задан объем выборки, то критическую область следует строить так, чтобы мощность критерия была максимальной. Выполнение этого требования должно обеспечить минимальную ошибку 2-го рода, что более желательно.

Основной принцип проверки статистических гипотез: если $K_{\text{набл}}$ принадлежит критической области — гипотезу H_0 отвергают, если же $K_{\text{набл}}$ принадлежит области принятия гипотезы, то гипотезу H_0 принимают.

Вопрос 2. Проверка гипотез о нормальном законе распределения генеральной совокупности

После того, как получен эмпирический закон распределения выборки X и по данным этого вариационного ряда построен полигон относительных частот, делается вывод (выдвигается гипотеза H_0) о законе распределения:

H₀: генеральная совокупность распределена по нормальному закону. И выдвигается гипотеза H₁, противоречащая гипотезе H₂ или ее отвергающая.

Проверка гипотезы о предполагаемом законе неизвестного распределения

производится при помощи специально подобранной случайной величины—

критерия согласия.

Разработано несколько таких критериев: χ^2 -Пирсона, Колмогорова, Смирнова и др.

Рассмотрим критерий χ^2 -Пирсона, как классический пример применительно к проверке гипотезы о нормальном законе распределения Γ С.

Пусть нам задан уровень значимости α (γ — доверительная вероятность, то есть вероятность принять верную гипотезу; α — это вероятность отвергнуть

верную гипотезу, причем $\alpha + \gamma = 1$).

Для того, чтобы при заданном α проверить гипотезу о нормальном распределении ГС, надо вынислить теоретические вероятности. Плотность распределения для нормального закона есть функция:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-M[x])^2}{2D[x]}}$$

Тогда, пользуясь формулой нахождения вероятности попадания случайной величины в интервал: $P(a < x < b) = \int_a^b f(x) dx$, имеем $\forall j = \overline{1,k}$

$$p_j = P(a_{j-1} < x < a_j) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{\text{B}}} e^{-\frac{(\bar{x}_j - M_{\text{B}}[x])^2}{2D_{\text{B}}[x]}} \cdot h$$

0000 X 000 +

Где a_j $(j = \overline{0, k})$ – границы частных подынтервалов; \bar{x}_j – середина ј-го частичного подынтервала; h – длина частичного подынтервала. Далее составляется таблица:

X	$\overline{X_1}$	<u> </u>	•••	$\overline{\mathbf{x}_{k}}$	
μ	μ_1	μ_2	• • •	μ_k	Эмпирические частоты
p	p_1	p_2		p_k	Теоретические частоты

Оценка отклонения эмпирических вероятностей μ_j $(j=\overline{1,k})$ от теоретических вероятностей p_j $(j=\overline{1,k})$ производится с помощью критерия

Пирсона
$$\chi^2$$
: $\chi^2_{\text{набл}} = \sum_{j=1}^k \frac{(\mu_j - p_j)^2 \cdot N}{p_j}$

По таблице критических точек распределения χ^2 по заданному уровню значимости α и числу степеней свободы r=k-3 (k- количество подытервалов) находим критическое значение $\chi^2_{\rm кp}(\alpha,r)$ правосторонней критической области.

<u>Правило 1.</u> Если $\chi^2_{\rm набл} < \chi^2_{\rm кp}$, тогда нет оснований отвергать гипотезу H_0 о нормальном законе распределения генеральной совокупности (то есть

эмпирические и теоретические частоты различаются незначимо (случайно)).

<u>Правило 2.</u> Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, тогда гипотеза \mathcal{H}_0 отвергается.

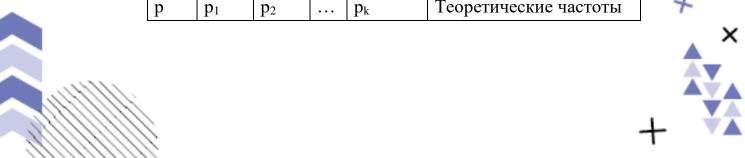
Вопрос 3. Проверка гипотез о других законах распределения генеральной совокупности

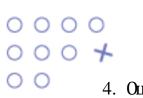
Для того, чтобы при заданном уровне значимости α , проверить гипотезу о показательном распределении ГС надо:

- 1. Вынислить $M_{_{\rm B}}[x]$ и принять в качестве оценки параметра $\pmb{\lambda}$ показательного распределения величину, обратную выборочной средней: $\lambda = \frac{1}{M_{_{\rm B}}[x]}$
- 2. Вынислить теоретические вероятности p_j $(j=\overline{1,k})$. Поскольку плотность распределения для показательного (экспоненциального) закона есть $f(x) = \begin{cases} 0, & x < 0 \\ \lambda \cdot e^{-\lambda x}, & x \geq 0 \end{cases}$ тогда $p_j = P(a_{j-1} < x < a_j) = e^{-\lambda \cdot a_{j-1}} e^{-\lambda \cdot a_j}$, где a_j $(j=\overline{0,k})$ границы частных подынтервалов.

3. Составляем сводную таблицу:

X	$\overline{X_1}$	$\overline{X_2}$	•••	$\overline{\mathbf{x}_{k}}$	
μ	μ_1	μ_2	• • •	μ_k	Эмпирические частоты
p	p_1	p_2	•••	p_k	Теоретические частоты





- 4. Оценку отклонения эмпирических вероятностей μ_j $(j=\overline{1,k})$ от теоретических вероятностей p_j $(j=\overline{1,k})$ производим с помощью критерия Пирсона χ^2 : $\chi^2_{\text{набл}} = \sum_{j=1}^k \frac{(\mu_j p_j)^2 \cdot N}{p_j}$
- 5. По таблице критических точек распределения χ^2 по заданному уровнюзначимости α и числу степеней свободыг=k-2 (k количество подытервалов) находим критическое значение $\chi^2_{\rm kp}(\alpha,r)$ правосторонней критической области.

Для того чтобы при заданном уровне значимости α, проверить гипотезу о равномерном распределении ГС надо:

- 1. Оценить параметрыа и с концы интервала, в котором наблюдались возможные значения X, по формулам (через a^* и c^* обозначены оценки параметров): $a^* = M_{\scriptscriptstyle B}[x] \sqrt{3} \cdot \sigma_{\scriptscriptstyle B}, \, c^* = M_{\scriptscriptstyle B}[x] + \sqrt{3} \cdot \sigma_{\scriptscriptstyle B}.$
- 2. Вынислить теоретические вероятности p_j $(j=\overline{1,k})$. Поскольку плотность распределения для показательного (экспоненциального) закона есть $f(x)=\frac{1}{c^*-a^*}$, тогда $p_j=P\big(a_{j-1}< x< a_j\big)=\frac{1}{c^*-a^*}\cdot h$, где a_j $(j=\overline{0,k})$ границы частных подынтервалов; h длина частичного подынтервала.

Получили, что все p_j ($j = \overline{1, k}$) равны одному числу $\frac{h}{c^* - a^*}$.

- 3. Составляем сводную таблицу на основе эмпирических вероятностей и рассчитанных теоретических вероятностей
- 4. Оценку отклонения эмпирических вероятностей μ_j $(j=\overline{1,k})$ от теоретических вероятностей p_j $(j=\overline{1,k})$ производим с помощью критерия Пирсона χ^2 : $\chi^2_{\text{набл}} = \sum_{j=1}^k \frac{(\mu_j p_j)^2 \cdot N}{p_j}$
- 5. По таблице критических точек распределения χ^2 по заданному уровнюзначимости α и числу степеней свободыг=k-3 (k количество подытервалов) находим критическое значение $\chi^2_{\rm кp}(\alpha,r)$ правосторонней критической области.



