Министерство образования Республики Беларусь

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Кафедра интеллектуальных информационных технологий

М. Д. СТЕПАНОВА

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

УЧЕБНО – МЕТОДИЧЕСКОЕ ПОСОБИЕ

по курсу "Статистические основы индуктивного вывода"
для студентов специальности "Искусственный интеллект"

Минск 2000

УДК 007:159.955: 519.23/.24 (075.8)
ББК 32.813 Я 73
С 79

Степанова М. Д.
С 79 Проверка статистических гипотез: Учебно-метод. пособие по курсу "Статистические основы индуктивного вывода"для студентов специальности "Искусственный интеллект". – Мн.: БГУИР, 2000. – 36 с. : ил. 1.
ISBN 985-444-087-7.

В пособии рассматриваются основные задачи и вероятностные модели проверки гипотез, приводится описание параметрических критериев проверки основных статистических гипотез. Рассмотрена задача проверки гипотез по экспертным данным и построение выводов с использованием байесовского правила. Приводятся примеры, иллюстрирующие применение статистических критериев.

Пособие предназначено для курса ″Статистические основы индуктивного вывода″.

УДК 007:159.955: 519.23/.24(075.8)

ББК 32.813

ISBN 985-444-087-7

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Проблемы, возникающие при принятии решений в различных областях деятельности, часто сводятся к оценке истинности одной или нескольких выдвигаемых гипотез на основе результатов анализа накопленной информации об изучаемом явлении.

В зависимости от специфики и свойств предметной области (ПрО), особенностей решаемых задач собранный материал может иметь своеобразную природу, структуру представления и хранения в памяти компьютера, а его обработка и анализ осуществляться специальными методами. В формализованных ПрО, где каждая проблема может быть описана аналитической моделью, а для ее решения используются строгие математические методы и алгоритмы, проверка гипотез опирается, как правило, на экспериментальные данные и осуществляется, в основном, статистическими методами с привлечением различных критериев, например, критериев значимости. Сущность такого рода задач состоит в проверке истинности некоторого предположительного утверждения (нулевой гипотезы) H_0 о характере некоторого массового явления по отношению к конкурирующей гипотезе H_1. В практике статистического анализа подобные задачи возникают на этапах принятия решений о выбираемых моделях данных и их зависимостях. Для этих целей используется достаточно широкий спектр методов и алгоритмических схем, соответствующих стандартным ситуациям.

В то же время возникает немало самостоятельных задач (особенно в тех областях, где случайная изменчивость является существенным элементом в модели), в основе решения которых лежит именно статистическая проверка гипотез (СПГ), приобретающая тем самым практически самостоятельный характер.

В слабо формализованных ПрО для принятия решений, кроме экспериментальных данных, могут использоваться также накопленные профессиональные знания (знания эксперта). Для получения обоснованного заключения об истинности той или иной гипотезы или их совокупности (выбора наиболее подходящей гипотезы) на основе анализа таких данных применяются специальные методы, имеющие необязательно строго статистическую основу. Примером такого рода проблем является часто возникающие на практике так называемые диагностические задачи, сущность которых состоит в определении достоверности целого набора выдвигаемых гипотез, исходя из имеющейся, как правило, неточной (неопределенной) информации об их признаках (свидетельствах). К наиболее известным методам решения задач такого типа (методам работы с неопределенностью) относятся метод коэффициентов уверенности, субъективный байесовский метод, интервальный метод Демпстера–Шеффера, методы, реализующие различные "возможностные" подходы, базирующиеся на теории нечетких множеств и выводов Заде.

Различия между этими методами весьма существенны и тесно связаны с видом представления информации (особенно нечеткой), с предпосылками, положенными в основу распространения информации, с учетом влияния свидетельств, поступающих из различных источников и др. Каждый из этих методов эффективен в определенных условиях. В качестве критериев выбора того или иного метода работы с неопределенностью могут выступать цели, которым отвечает конкретная задача (ее особенности, важность последствий ошибочного решения), а также степень знаний эксперта по отношению к априорной информации и предполагаемая компетенция пользователя системы.

На практике целесообразно решать одни и те же сложные задачи несколькими методами.

1. СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

Гипотеза – это утверждение, которое может или не может быть истинным. Практический пример ситуации, о которой можно сказать, что мы "проверяем гипотезу". Предположим, вы подъезжаете к дому вашей матери вечером и высказываете предположение (гипотезу), что она дома. Вы "проверяете гипотезу", определяя, горит ли свет в окнах. Свидетельство (отсутствие света), казалось бы, наводит на мысль о том, что ваша гипотеза ложная. Но это свидетельство не является достаточно полным, чтобы сделать определенный вывод (заключение). Ваш вывод станет более определенным, если вы убедитесь, что у дома нет автомобиля вашей матери. Чтобы быть абсолютно уверенным, что вашей матери нет дома, вы должны войти в дом и посмотреть, там она или нет. Но если вы сделаете это, то вы уже больше не "проверяете" гипотезу, а детерминированно устанавливаете, ложна она или нет. Некоторые считают, что гипотеза проверяется в обстоятельствах, прямая верификация которых не возможна (или практически не осуществима), но при этом доступна частичная информация, в противовес которой проверяется гипотеза.

Статистическая гипотеза – это гипотеза, которая допускает наблюдения статистической природы. Такие наблюдения могут возникать в различных областях деятельности человека. Вот некоторые примеры.

1) Эта игральная кость правильная.
2) Сведения нашего поставщика деталей о своей продукции ложны.
3) Новый метод обучения лучше, чем старый.
4) Требования стандарта на чистоту воздуха в нашем городе не выдерживаются.
5) Данные о занятости населения предполагают наличие дискриминационной политики при найме на работу.

Так как эти гипотезы являются статистическими, то они могут быть сформулированы более точно.

1) Вероятности выпадения каждой грани такой игральной кости равны. Это означает, что мы имеем равномерное распределение случайной переменной, которая представляет собой число точек на лицевой поверхности каждой грани данной кости.

2) Средняя длина детали, поступившей от нашего поставщика, больше, чем он заявлял (или меньше, или отличается от заявленной).

3) Средняя оценка стандартного тестирования у обучавшихся по новому методу выше, чем у обучавшихся по старому методу.

4) Значения параметров, характеризующих чистоту воздуха в городе, больше, чем установлено стандартом.

5) Для некоторых работодателей переменная "принятие на работу" не будет независимой от переменной "пол" (или "этническая принадлежность", "вероисповедание" и т. д.).

Из приведенных примеров следует, что *статистическая гипотеза* – это утверждение относительно характера или неизвестных параметров распределения случайных величин. Гипотеза называется *простой*, если она полностью определяет распределение случайной величины, Например, значение некоторого параметра Θ *в точности* равно заданной величине Θ_0. В других случаях гипотеза называется *сложной*.

Гипотеза (1) является гипотезой относительно абстрактной модели, вероятностном распределении случайной переменной, которая описывает игральную кость. Данная гипотеза может быть проверена с помощью наблюдений, например, 100 бросаний кости. Следующие три гипотезы касаются параметра. Они проверяются с помощью выборочных данных. Эти данные рассматриваются как выборочное распределение, необходимое для оценки параметра. Последняя гипотеза утверждает независимость двух качественных (нечисловых) переменных. Она проверяется путем сравнения наблюдаемых данных с истинными данными, ожидаемыми в случае независимости переменных. Отметим, что только три из пяти гипотез включает параметр, две другие представляют собой статистически утверждения другого рода.

Для каждого из этих примеров практически невозможно непосредственно определить истинность гипотезы. Например, вероятностное распределение для кости является моделью всех возможных бросаний, которые мы не можем наблюдать. Практически невозможно измерить длину каждой из сотен, а может быть и тысяч поступающих деталей. Для гипотезы (3) невозможно протестировать сегодня всех студентов, которые должны обучаться по новому методу в ближайшие 15 лет. Конечно, можно протестировать студентов через год после окончания обучения. Но оценку эффективности нового метода нужно делать до его реализации, а не после. А для гипотезы (4) можно ли проверить каждый кубический метр воздуха в городе? Наконец, что означает в (5) прямая верификация для каждого человека?

Из-за невозможности определить истинность гипотезы прямым путем, мы "проверяем" гипотезу, т.е. устанавливаем, не противоречит ли высказанная нами гипотеза имеющимся выборочным данным. Эта процедура носит название ***статистической проверки гипотез***.

Результат сопоставления высказанной гипотезы с выборочным данными может быть либо ***отрицательным*** (данные наблюдения противоречат высказанной гипотезе, а поэтому гипотезу надо ***отклонить***), либо ***неотрицательным*** (данные наблюдения не противоречат высказанной гипотезе, а поэтому ее можно ***принять*** в качестве одного из возможных решений).

2. ОБЩАЯ СХЕМА СТАТИСТИЧЕСКОЙ ПРОВЕРКИ ГИПОТЕЗ

Задачи, сводящиеся к оценке истинности нулевой гипотезы H_0 по отношению к конкурирующей гипотезе H_1 могут быть решены с помощью различного рода статистических критериев (СК).

Несмотря на разнообразие самих гипотез и применяемых СК, их можно объединить в следующую общую логическую схему.

1. Выдвижение гипотез H_0, H_1.
2. Выбор ***уровня значимости*** α – вероятности ошибочного отклонения нулевой гипотезы. Эту величину называют также ***размером критерия*** (теста). Выбор величины уровня значимости α зависит от размера потерь, которые мы понесем в случае ошибочного решения. В большинстве практических задач пользуются стандартными значениями уровня значимости: α = 0.1; 0.05; 0.025; 0.01; 0.005; 0.001. Наиболее распространенной является величина уровня значимости α = 0.05 (в среднем в пяти случаях из 100 мы будем ошибочно отклонять высказанную гипотезу).
3. Выбор ***критической статистики*** (критерия) $\psi_k(x_1, x_2, ..., x_n)$ – некоторой функции от результатов наблюдений. Эта критическая статистика ψ_k сама является случайной величиной и в предположении справедливости нулевой гипотезы H_0 подчинена некоторому хорошо изученному закону распределения с плотностью $f_\psi(u)$. Критическая статистика строится на основании ***принципа отношения правдоподобия*** [1, 2].
4. Определение ***критической области*** W (множества значений критической статистики, при которых гипотеза отклоняется), исходя

из следующего условия: $P\,(\psi_k(x_1, x_2, ..., x_n) \in W \,\big|\, H_0\,) = \alpha$. Из таблиц распределения $f_\psi(u)$ находятся квантили уровня $\alpha/2$ и уровня $1-\alpha/2$, соответственно равные $\psi_{\alpha/2}$ и $\psi_{1-\alpha/2}$. Они разделяют всю область возможных значений случайной величины ψ_k на три части: 1 – ***область неправдоподобно малых*** $(-\infty, \psi_{\alpha/2}]$, 2 – ***правдоподобных*** $(\psi_{\alpha/2}, \psi_{1-\alpha/2})$, 3 – ***неправдоподобно больших*** $[\psi_{1-\alpha/2}, \infty)$ значений ***в условиях справедливости нулевой гипотезы H_0***. В тех случаях, когда опасными для нашего утверждения являются только односторонние отклонения, т. е. только "слишком маленькие" или только "слишком большие" значения критической статистики ψ_k, находят лишь одну квантиль: либо $\psi_{\alpha/2}$, которая будет разделять весь диапазон значений ψ_k на две части: область неправдоподобно малых и область правдоподобных значений; либо $\psi_{1-\alpha/2}$; она будет разделять весь диапазон значений ψ_k на область неправдоподобно больших и область правдоподобных значений.

5. Определение на основе выборочных данных $x_1, x_2, ..., x_n$ численной величины статистики $\hat{\psi}_k$.

6. Выработка решения. Если $\hat{\psi}_k \in W$, то гипотезу H_0 рекомендуется отклонить, в противном случае ее можно принять, так как имеющиеся данные не противоречат высказанной гипотезе. Однако для большей уверенности, прежде чем принять гипотезу, ее желательно подвергнуть проверке с помощью других критериев или повторить эксперимент, увеличив объем выборки.

Решение, принимаемое на основе статистического критерия, может оказаться ошибочным в двух случаях: когда ошибочно отклоняется гипотеза H_0 (с вероятностью α), и когда ошибочно принимается гипотеза H_0 (с вероятностью β), где вероятности ошибочных решений α и β – ошибки первого и второго рода, а величина $1 - \beta$ – мощность соответствующего критерия [1, 2].

3. ОШИБКИ ПЕРВОГО И ВТОРОГО РОДА

Возможные ошибки. В контексте реальной жизни мы обнаруживаем массу возможных ошибок: человеческие ошибки, машинные ошибки, ошибки из-за невнимательности, намеренные ошибки (саботаж), ошибки пользователя, ошибки спецификации, ошибки тривиальные, ошибки серьезные, ошибки распознаваемые, ошибки нераспознаваемые, ошибки фатальные, ошибки смешные и т. д.

С точки зрения статистической проверки гипотез существуют только два вида ошибок, называемые ***ошибкой I рода*** и ***ошибкой II рода***.

Ошибка I рода – это неправильное действие в соответствии с H_1: действовать в соответствии с H_1, если справедлива H_0. Ошибка II рода – это неправильное действие в соответствии с H_0: действовать в соответствии с H_0, если справедлива H_1. Вероятность ошибки интерпретируется как условная вероятность. Условные вероятности этих двух типов ошибок обозначаются соответственно α и β:

$\alpha =$ **P** (ошибка I рода) = **P** (действие в соответствии с $H_1 | H_0$ истинна);

$\beta =$ **P** (ошибка II рода) = **P** (действие в соответствии с $H_0 | H_1$ истинна).

В табл. 1 показаны возможности принятия решения и ошибки двух типов по отношению к гипотезе H_0. Отметим, что если гипотеза H_0 справедлива и она принимается, то в таблице указано, что решение принято правильно. Если справедлива гипотеза H_1, а принимается H_0, то при решении допущена ошибка II рода. Если справедлива гипотеза H_0, а принимается гипотеза H_1, то при решении допущена ошибка I рода.

Таблица 1

Решения и ошибки при статистической проверке гипотез

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Состояние реального мира (неизвестное нам)</td></tr>
<tr><td>H_1 ложна (H_0 истинна)</td><td>H_1 истинна</td></tr>
<tr><td rowspan="2">Наше решение, основанное на данных</td><td>Действие в соответствии с H_0</td><td>Правильное решение</td><td>Ошибка II рода</td></tr>
<tr><td>Действие в соответствии с H_1</td><td>Ошибка I рода</td><td>Правильное решение</td></tr>
</table>

Пусть $f(x; \theta_0)$, $f(x; \theta_1)$ – плотности распределения критической статистики соответственно при справедливости нулевой гипотезы H_0 и альтернативной гипотезы H_1, θ_0, θ_1 – параметры распределения при H_0 и H_1. Тогда ошибки I и II рода определяются выражениями

$$\alpha = \int_{x_k}^{\infty} f(x;\theta_0)dx \tag{1}$$

и

$$\beta = \int_{-\infty}^{x_k} f(x;\theta_1)dx\,, \tag{2}$$

где x_k – граница критической области W.

Существуют два возможных вывода при проверке гипотезы: либо мы отклоняем нулевую гипотезу ("отклонить H_0)", либо мы отказываемся отклонить нулевую гипотезу ("отказ отклонить H_0)". Рассмотрим смысл каждого из этих выводов и соответствующие ошибки на примере, связанном с контролем качества. Пусть p – доля дефектных изделий, выпускаемых производственной линией. Сформулируем две гипотезы:

H_0: $p = 0.01$;

H_1: $p > 0.01$.

Предполагая 5 %-ный уровень значимости, рассмотрим возможные выводы и их интерпретацию.

Отклонить H_0: Свидетельства предполагают, что контроль качества ослаб, необходимо остановить производство и выполнить корректирующие действия.

Отказ отклонить H_0: Тест не дает выводов, так как отсутствуют свидетельства того, что контроль качества ослаб, производство остается без изменений.

Ошибка I рода: С 5 %-ным риском ошибки мы останавливаем производство без необходимости. Доля дефектных изделий не превышает заданный уровень, даже если мы основываемся на неверных свидетельствах.

Ошибка II рода: Мы продолжаем производство с неизвестным риском ошибки, когда число дефектных изделий среди выпускаемых больше, чем допускают требования контроля качества.

Формальные выводы ведут к реальным действиям. "Отклонить H_0" означает "действовать в соответствии с H_1", "Отказ отклонить H_0" означает "действовать в соответствии с H_0".

Рассмотрим пример проверки гипотезы о среднем значении μ диаметра поставляемых деталей:

H_0: $\mu = 3.15$;

H_1: $\mu < 3.15$.

Отклонить H_0: Это означает, детали машины, которые мы получаем от поставщика, имеют недопустимо малый диаметр (среднее значение диаметра меньше, чем 3.15 мм). Ошибка в данном случае составляет 5 %.

Отказ отклонить H_0: Тест не дает выводов. Отсутствуют свидетельства нарушений условий поставки партии деталей, поставляемая партия принимается.

Ошибка I рода: Существует 5 %-ный риск того, что, основываясь на ошибочных свидетельствах, мы отклоним партию деталей, хотя среднее значение диаметра не будет меньше, чем 3.15 мм.

Ошибка II рода: Так как у нас отсутствовали необходимые свидетельства, мы принимаем поставляемую партию деталей, хотя их размеры не соответствуют требованиям.

4. ОБЛАСТЬ ПРИНЯТИЯ И ОТКЛОНЕНИЯ ГИПОТЕЗЫ

Область отклонения гипотезы. Задача отклонения или принятия гипотезы H_0 решается на основании данных. Рассмотрим ее на примере данных контроля качества (p – доля дефектных изделий в выпускаемой продукции):

$H_0 : p = 0.01,$

$H_1 : p > 0.01.$

Вероятность ошибки 1-го рода α представляет собой уровень значимости теста. Предположим, что $\alpha = 0.05$:

$$\alpha = 0.05 = P\ (\text{отклонить } H_0 \mid H_0 \text{ верна}).$$

Значение $\hat{p}$, которое разделяет область "отказ отклонить H_0" и область "отклонить H_0", называется критическим значением $\hat{p}$ и обозначается p_c. Если $\hat{p} > p_c$, мы отклоняем H_0, в противном случае отказываемся отклонить H_0. Значения $\hat{p}$, при которых происходит отклонение H_0, называется ***областью отклонения гипотезы***:

$$\textit{область отклонения}: \{\hat{p} \mid \hat{p} > p_c\}.$$

Критическое значение определяется из выражения (1).

Пример. Определить значение p_c для $\alpha = 0.05$. Полагая, что p имеет нормальное распределение, найдем значение 5%-ной отрезной точки $Z = 1.645$. Тогда $1.645 = (p_c - 0.01)/\textit{с.к.о.}$.

Среднее квадратическое отклонение (*с.к.о.*) вычисляется в предположении справедливости нулевой гипотезы H_0. Пусть объем выборки $n = 700$. Решив приведенное выше уравнение, найдем $p_c = 0.0162$ и область отклонения гипотезы:

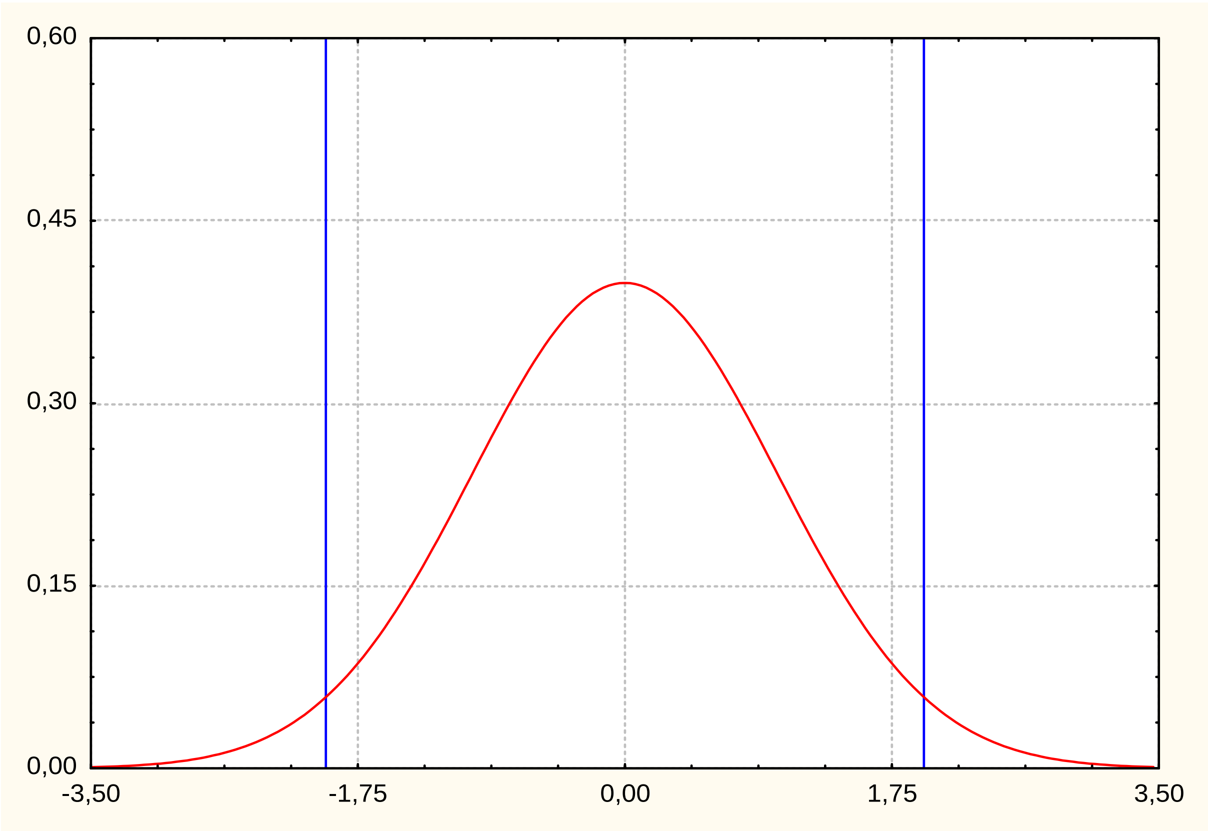
$$\textit{область отклонения}: \{\hat{p} \mid \hat{p} > 0.0162\}.$$

Граница критической области W определяется путем решения уравнения (1) относительно x_k при фиксированном уровне значимости α (рис. 1)

Пример. Проверяется гипотеза о значении математического ожидания: H_0 : $\mu = \mu_0 = 20$, H_1: $\mu > \mu_0$.

Допустим, что величина σ^2 известна и равна 16, объем выборки $n = 9$, имеет место нормальное распределение. При $\alpha = 0.05$ решаем уравнение (1) относительно x_k :

$$\int_{x_k}^{\infty} f(x;\mu,\sigma)dx = \alpha,$$



←Отклонить H_0 | ←Принять H_0→ | Отклонить H_0→

Рис. 1 Область отклонения и принятия гипотезы H_0
при двусторонней альтернативе

где f (x; μ. σ) – плотность нормального распределения, μ = 20, σ = 4/3. Значение x_k равно 22.2.

5. РЕШАЮЩЕЕ ПРАВИЛО И СТАТИСТИКА КРИТЕРИЯ

В общем виде решающее правило формулируется следующим образом:

Если значение статистики критерия, вычисленное по выборке, попадает в область отклонения гипотезы, то следует действовать в соответствии с H_1.

Для рассматриваемого примера с контролем качества решающее правило будет таким:

Если в контролируемой партии продукции число дефектных изделий превысит 11 (1.62% от 700), то производство необходимо остановить и принять меры для устранения причин брака.

Значение статистики критерия вычисляется по выборке. В примере с контролем качества статистика критерия представляет собой значение $Z = (p_c - 0.01)/с.к.о.$

При проверке гипотезы о значении математического ожидания статистики критерия имеют вид

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \quad \text{или} \quad t = \frac{\overline{X} - \mu_0}{s / \sqrt{n}} .$$

В этих формулах μ_0– это значение математического ожидания μ, определяемое нулевой гипотезой. При проверке гипотезы о значении дисперсии статистика критерия имеет вид $\chi^2 = (n-1)\, s^2 / \sigma_0^2$, где σ_0^2 – значение дисперсии, задаваемое нулевой гипотезой.

С помощью статистики критерия можно выразить область отклонения гипотезы и связанное с ней решающее правило. Например, если используется статистика критерия t, то

область отклонения = $\{t \mid t > t_c\}$,

где t_c – квантиль уровня α t–распределения с n – 1 степенью свободы. Если $\alpha = 0.05$ и $n = 10$, то $t_c = 1.8331$. Тогда решающее правило будет следующим:

Если вычисленное по выборке значение t больше, чем 1.8331, то гипотеза H_0 *отклоняется.*

Запишем решающее правило принятия или отклонения нулевой гипотезы, основанное на критическом значении статистики критерия ψ_k:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если} \quad \psi_k < \Delta(\alpha), \\ H_1, & \text{если} \quad \psi_k \geq \Delta(\alpha), \end{cases}$$

где $\Delta(\alpha)$ – порог теста (критическое значение статистики критерия) размера α.

Порог теста определяется для двусторонней альтернативы как квантиль уровня $1 - \alpha/2$ распределения статистики критерия, для односторонних альтернатив как квантиль уровня α (левосторонняя альтернатива) или как квантиль уровня $1 - \alpha/2$ (правосторонняя альтернатива).

6. ПОНЯТИЕ *P*–ЗНАЧЕНИЯ

Стандартная процедура проверки нулевой гипотезы состоит в том, что мы наблюдаем реализацию случайной величины X и смотрим, попадает x в область принятия или отклонения нулевой гипотезы. Однако часто более удобно выполнять другую процедуру, являющуюся в некотором смысле обратной к описанной. Вместо того, чтобы работать с фиксированным уровнем значимости α и только соглашаться с принятием или отклонением H_0, можно найти α, соответствующее реализации случайной величины X. Число α, полученное таким образом, называется ***P*–значением**.

Сформулируем решающее правило принятия или отклонения нулевой гипотезы, основанное на P–значении.

Пусть X – результат наблюдения, $Y = Y(X)$ – статистика критерия (скалярная случайная величина), с помощью которой проверяется нулевая гипотеза H_0, $\Delta(\alpha)$ – критическое значение (порог теста) уровня α, т. е. $\mathbf{P}\{Y > \Delta(\alpha) \mid H_0\} = \alpha$, $F(Y)$ – функция распределения статистики Y при нулевой гипотезе H_0. Возможны следующие три ситуации.

1. Проверяется нулевая гипотеза H_0: $\xi = a$ против правосторонней альтернативы $H_1 : \xi > a$. В этом случае критическая область имеет вид $R = \{X : F(Y(X)) > F(\Delta(\alpha) = 1- \alpha\}$. P–значение есть величина $P = 1 - F(Y(X))$, которая получается в результате подстановки значения статистики критерия в ее функцию распределения. Решающее правило:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если } P > \alpha, \\ H_1, & \text{если } P \le \alpha. \end{cases}$$

2. Проверяется нулевая гипотеза H_0: $\xi = a$ против левосторонней альтернативы $H_1 : \xi < a$. В этом случае критическая область имеет вид $R = \{X : F(Y(X)) < F(\Delta(\alpha) = \alpha\}$. P–значение есть величина $P' = F(Y(X))$. Решающее правило имеет вид:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если } P' > \alpha, \\ H_1, & \text{если } P' \le \alpha. \end{cases}$$

3. Проверяется нулевая гипотеза H_0: $\xi = a$ против двусторонней альтернативы $H_1 : \xi \neq a$. В этом случае критическая область имеет вид $R = \{X : F(Y(X)) > F(\Delta(\alpha) = \alpha/2$ и $F(Y(X)) < F(\Delta(\alpha) = \alpha/2\}$. P–значение есть величина P'', определяемая из уравнения $1 - P''/2 = F(Y(X))$. Решающее правило имеет вид:

$$\text{принимается гипотез } \begin{cases} H_0, & \text{если } P'' > \alpha, \\ H_1, & \text{если } P'' \le \alpha. \end{cases}$$

Пример. Нормальная случайная величина X, дисперсия D которой равна 100, приняла значение 125. Необходимо проверить гипотезу $\xi = 100$ против альтернативы $\xi \neq 100$ при уровне значимости $\alpha = 0.05$.

Определим значение статистики критерия $Z = |X - \xi|/\sqrt{D} = |125 - 100|/10 = 2.5$.

Решаем уравнение $1 - P/2 = \Phi(2.5)$, где $\Phi(\cdot)$ – функция стандартного нормального распределения. Получаем $1 - P/2 = 0.994$, откуда $P = 0.012$. Таким образом, гипотеза H_0 отклоняется.

Малое P–значение предполагает два объяснения для имеющихся данных:

- Гипотеза может быть истинной, а данные "плохими" из-за ошибки выборки. Например, выпадение всех гербов при 100 бросаниях монеты возможно, но маловероятно.
- Гипотеза может быть неверной, а данные "неплохими". Допустим, монета имеет смещенный центр тяжести (60% появлений герба). Тогда выпадение 58 гербов при 100 бросаниях – хорошие данные.

При выработке решения можно воспользоваться традиционной интерпретацией P-значений:

1) $P > 0.1$– имеем хорошее согласие с H_0;

2) $P = 0.05$ – есть некоторые сомнения в истинности H_0;

3) $P = 0.02$ – довольно сильный довод против H_0;

4) $P \leq 0.01$ – гипотеза почти наверняка не подтверждается.

7. МОЩНОСТЬ КРИТЕРИЯ

Мощность критерия π – это вероятность различения альтернативной гипотезы H_1, если она верна (или вероятность правильного отклонения нулевой гипотезы H_0):

$$\pi = 1 - \beta = \mathbf{P}\{\text{действия в соответствии с } H_1 \mid H_1 \text{ верна}\}.$$

$$\pi = \int_{x_k}^{\infty} f(x;\theta_1)dx \quad , \qquad (3)$$

где x_k – граница критической области W.

Это понятие имеет важное теоретическое и практическое значение. С теоретической точки зрения мощность критерия является одной из главных характеристик, на основании которой выбирается критерий проверки гипотезы среди возможных. В математической статистике существует развитая теория оптимальных критериев, с помощью которой идентифицируются свойства различных критериев.

Практическое использование мощности критерия рассмотрим на примере данных контроля качества продукции: $H_0 : p = 0.01$, $H_1 : p > 0.01$.

В этом случае мощность критерия представляет собой вероятность того, что мы правильно отследим плохую ситуацию:

$$\mathbf{P}\{\text{действия в соответствии с } H_1 \mid p > 0.01\}.$$

Это вероятность того, что мы остановим производство и примем меры для устранения причин брака, если в контролируемой партии продукции число дефектных изделий слишком велико. В данном примере область отклонения гипотезы при уровне значимости α определяется множеством значений $\{\hat{p} \mid \hat{p} > 0.0162\}$, а мощность критерия $\pi = \mathbf{P}\{\hat{p} > 0.0162 \mid p > 0.01\}$.

Пример. Рассмотрим гипотезы: 1) $H_0 : \mu = \mu_0 = 22$; 2) $H_0 : \mu = \mu_0 = 22$; и альтернативную гипотезу H_1: $\mu > \mu_0$. Как и в примере из раздела 5, примем, что величина σ^2 известна и равна 16, объем выборки n = 9, имеет место нормальное распределение. Граница критической области x_k равна 22.2. Определим мощность критерия относительно каждой из этих гипотез согласно (3):

$$\pi_1 = \mathbf{P}\{\overline{X} > 22.2 \mid \mu = 22\}, \qquad \pi_2 = \mathbf{P}\{\overline{X} > 22.2 \mid \mu = 24\}.$$

Эти вероятности можно записать в виде интегралов (3):

$$\pi_1 = \int_{22.2}^{\infty} f_{\overline{X}}(\bar{x}; 22; 4/3)\, d\bar{x},$$

$$\pi_2 = \int_{22.2}^{\infty} f_{\overline{X}}(\bar{x}; 24; 4/3)\, d\bar{x}.$$

Решив уравнения, получим мощности, соответствующие каждой из гипотез: $\pi_1 = 0.44038$, $\pi_2 = 0.91149$.

8. ОСНОВНЫЕ ТИПЫ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

По схеме, приведенной в разделе 2, можно решить большое количество практических задач, применяя различные статистические критерии. Высказываемые в ходе решения задач гипотезы можно подразделить на следующие типы:

об общем виде закона распределения исследуемой случайной величины;

об однородности двух или нескольких выборок;

о числовых значениях характеристик исследуемого явления или процесса;

об общем виде зависимости, существующей между компонентами исследуемого многомерного признака;

о независимости и стационарности ряда наблюдений.

8.1. Гипотезы о типе закона распределения исследуемой случайной величины

Задача состоит в подборе некоторой модельной функции распределения $F_{\text{mod}}(X)$, с помощью которой можно описать исследуемую функцию распределения $F_\xi(X)$ случайной величины ξ. Проверяемая гипотеза имеет вид

$$H:\ F_\xi(X) \equiv F_{\text{mod}}(X),$$

где гипотетическая модельная функция может быть как заданной *однозначно* (тогда $F_\xi(X) = F_0(X)$, где $F_0(X)$ – полностью известная функция), так и *заданной с точностью до принадлежности к некоторому параметрическому семейству* (тогда $F_{\text{mod}}(X) = F(X; \Theta)$, где Θ – некоторый параметр, значения которого неизвестны, но могут быть оценены по выборке).

Проверка гипотез об общем виде закона распределения осуществляется с помощью *критериев согласия*.

8.2. Гипотезы об однородности двух или нескольких выборок или некоторых характеристик анализируемых совокупностей

Пусть мы имеем l выборок. Обозначим вероятностный закон распределения, которому подчиняются наблюдения j-й выборки, средние значения и дисперсии соответственно $F_j\,(X)$, a_j , σ_j^2. Тогда основные гипотезы однородности можно записать в виде:

H_F: $F_1\,(X) \equiv F_2\,(X) \equiv \ldots \equiv F_l\,(X)$;

H_a **:** $a_1 = a_2 = \ldots = a_l$ **;**

H_σ **:** $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_l^2$ **.**

8.3. Гипотезы о числовых значениях параметров исследуемой совокупности

Задачи первого типа возникают, например, когда нужно проверить точность некоторого прибора, инструмента, устойчивость определенного технологического процесса и т. п. Они сводятся к проверке гипотезы о параметрах исследуемой генеральной совокупности. Формально гипотезы подобного рода приобретают вид:

H_0: $\Theta = \Delta_0$,

где Θ – некоторый параметр исследуемого распределения, а Δ_0 – область его конкретных значений. В качестве таких параметров часто рассматриваются наиболее информативные характеристики распределения: среднее значение (медиана в непараметрическом случае) и дисперсия.

8.4. Гипотезы о типе зависимости между компонентами исследуемого многомерного признака

При исследовании статистической зависимости, например, компоненты $x^{(2)}$ от компоненты $x^{(1)}$ анализируемого двумерного признака $X = (x^{(1)}, x^{(2)})$, бывает необходимо проверить *гипотезу об общем виде этой зависимости*. Например, гипотезу о том, что $x^{(2)}$ и $x^{(1)}$ связаны линейной регрессионной связью:

H: $\mathbf{E}\,(x^{(2)} | x^{(1)}) = b_0 + b_1 x$,

где b_0 и b_1 – некоторые неизвестные параметры модели.

Статистические критерии, с помощью которых проверяются гипотезы данного типа, называют *критериями адекватности*.

8.5. Гипотезы независимости и стационарности ряда наблюдений

Принятие решения о независимости или стационарности наблюдений (сохранения закона распределения) основывается на результатах проверки соответствующих гипотез, например:

$H:\ \mathbf{E}\,x_i = a = \text{const},\ i = 1, 2, \ldots, n;$

$H:\ r\,(x_i,\ x_{i+1}) = 0,\ i = 1, 2, \ldots, n-1,$

где $r\ (x_i,\ x_{i+1})$ – коэффициент корреляции, построенный по совокупности двумерных наблюдений $X = (x^{(1)}, x^{(2)})$.

9. ПРОВЕРКА СООТВЕТСТВИЯ ВЫБРАННОЙ МОДЕЛИ РАСПРЕДЕЛЕНИЯ ИСХОДНЫМ ДАННЫМ (КРИТЕРИИ СОГЛАСИЯ)

Пусть нами высказано предположение, что ряд наблюдений $x_1, x_2, \ldots, x_n$ образует случайную выборку из распределения с некоторой модельной функцией $F_0(X;\theta^{(1)},\ldots,\theta^{(s)})$, где общий вид функции F_0 (т. е. тип модели) считается известным, а параметры могут быть как известными, так и неизвестными.

Критерии согласия предназначены для проверки гипотезы

$$H_0:\ F_\xi(x) = F_0(x;\theta^{(1)},\ldots,\theta^{(s)}) \qquad (4)$$

и основаны на использовании различных мер расстояний между анализируемой эмпирической функцией распределения, определяемой по выборке, и гипотетической модельной $F_0(x;\theta^{(1)},\ldots,\theta^{(s)})$.

9.1. Критерий χ^2 Пирсона

Критерий χ^2 Пирсона позволяет проверить гипотезу (4), когда значения параметров $\theta^{(1)},\ldots,\theta^{(s)}$ неизвестны. Процедура проверки гипотезы состоит из следующих шагов.

1) Диапазон значений исследуемой случайной величины ξ разбивается на k взаимно исключающих и непересекающихся интервалов $I_1,...,I_k$.

2) Подсчитывается число наблюдений n_i, попадающих в интервал I_i, $i = 1,\ldots, k$.

3) Вычисляется ожидаемое число наблюдений ν_i в интервале I_i при условии справедливости гипотезы H_0.

4) Вычисляется статистика

$$\chi^2 = \sum_{i=1}^{k}(n_i - \nu_i)^2 / \nu_i,$$

которая при верной H_0 имеет χ^2–распределение с $f = k - s - 1$ степенями свободы.

5) Гипотеза о том, что исследуемая случайная величина ξ подчиняется закону распределения F_0, принимается на уровне значимости α, если

$$\Delta(\alpha/2) \le \chi^2 < \Delta(1 - \alpha/2),$$

где Δ(ε) – квантиль уровня ε имеет χ^2–распределение с $f = k - s - 1$ степенями свободы; если $\chi^2 \ge \Delta\ (1 - \alpha/2)$, гипотеза H_0 отклоняется.

Пример. Проверялся размер n = 200 деталей, изготовленных на токарном станке–автомате. В табл. 2 приведены границы интервалов I_i отклонения от номинального размера, n_i– число деталей со значением отклонения, попадающим в данный интервал, p_i – теоретическая вероятность попадания отклонений в интервалы, $\nu_i = np_i$ – теоретическая частота попадания отклонений в интервалы. Оценить с помощью критерия χ^2 согласие выборочного распределения с нормальным распределением при уровне значимости α = 0.05.

Вычислим теоретические вероятности p_i попадания отклонений в интервалы (y_i, y_{i+1}) по формуле

$$p_i = \Phi\ (z_{i+1}) - \Phi\ (z_i),$$

где Φ (·) – функция стандартного нормального распределения, $z_i = (y_i - \bar{y})/s$, $\bar{y}$ и $\bar{\sigma}$ – выборочные оценки среднего и среднего квадратического отклонения, $\bar{y} = 4.30$, $\bar{\sigma} = 9.71$. Находим значение $\chi^2 = 7.09$. Число степеней свободы $f = k - s - 1 = 9 - 2 - 1 = 6$. Критическое значение статистики χ^2 равно 14.45, следовательно, гипотеза о нормальности отклонений от номинального размера не противоречит наблюдениям.

Таблица 2

Вычисление статистики критерия χ^2

Границы интервалов (y_i, y_{i+1})	n_i	p_i	ν_i
(–20, –15)	7	0.0239	4.78
(–15, –10)	11	0.0469	9.38
(–10, –5)	15	0.0977	19.54
(–5, 0)	24	0.1615	32.30
(0, 5)	40	0.1979	39.58
(5, 10)	41	0.1945	38.90
(10, 15)	26	0.1419	28.38
(15, 20)	17	0/0831	16.62
(20, 30)	10	0.0526	10.52

9.2. Критерий Колмогорова

Определим расстояние Колмогорова между эмпирической и теоретической функциями распределения:

$$D = \max_x | F_э(x) - F_0(x) | \ .$$

Решающее правило:

$$\text{принимается гипотеза} \begin{cases} H_0, & \text{если} \quad n^{1/2}D < \Delta(\alpha), \\ H_1, & \text{если} \quad n^{1/2}D \geq \Delta(\alpha). \end{cases}$$

Если верна H_0 и $n \geq 20$, то независимо от F_0 (x) случайная величина $n^{1/2}$ D имеет распределение Колмогорова [1, 3]. Это позволяет определить $\Delta(\alpha)$. Значения $\Delta(\alpha)$ приведены в табл. 3.

Таблица 3

Значения квантилей распределения Колмогорова

α	0.01	0.05	0.1	0.2
Δ	1.63	1.36	1.22	1.07

Критерий Колмогорова рекомендуется применять, если теоретическая функция распределения известна полностью.

Пример. Эмпирическая функция распределения была определена по 20 значениям случайной величины. При этом наибольшая разность D между эмпирической и теоретической функциями распределения оказалась равной 0.33. Находим $n^{1/2} D = 20^{1/2}\, 0.33 = 1.48$. При уровне значимости $\alpha = 0.05$ нулевая гипотеза (результаты наблюдения подчиняются выбранному распределению) отклоняется, так как вычисленное значение статистики критерия $1.48 > \Delta(\alpha) = 1.36$.

10. КРИТЕРИИ ОДНОРОДНОСТИ НОРМАЛЬНЫХ СОВОКУПНОСТЕЙ

В данном разделе рассматриваются задачи проверки однородности числовых характеристик (средних значений и дисперсий двух анализируемых совокупностей.

Наблюдаются две независимые случайные выборки $X^{(1)}$ и $X^{(2)}$ объемов n_1 и n_2 соответственно. $X^{(i)} = (x_{i1},...,x_{in_i})$ – выборка из нормального распределения с неизвестными математическим ожиданием μ_i и дисперсией σ_i^2 (i = 1, 2). Гипотезы равенства (однородности) математических ожиданий H_1 и дисперсий H_2 и их альтернативы $\overline{H}_1$ и $\overline{H}_2$ можно записать в виде

$$H_1:\ \mu_1 = \mu_2,\ \overline{H}_1:\ \mu_1 \neq \mu_2; \tag{5}$$

$$H_2:\ \sigma_1^2 = \sigma_2^2,\ \overline{H}_2:\ \sigma_1^2 \neq \sigma_2^2. \tag{6}$$

10.1. Однородность средних значений

Для проверки гипотезы равенства математических ожиданий в двух выборках в предположении нормальности распределений с равными дисперсиями ($\sigma_1^2 = \sigma_2^2$) используется критерий Стьюдента [1-3]:

$$\text{принимается гипотеза } \begin{cases} H_1, & \text{если} \quad |t| \le \Delta(\varepsilon), \\ \overline{H}_1, & \text{если} \quad |t| > \Delta(\varepsilon), \end{cases}$$

где

$$t = (\overline{X}_1 - \overline{X}_2) / (s^2 \ (1/n_1 + 1/n_2))^{1/2} \ - \qquad (7)$$

двухвыборочная статистика Стьюдента;

$\overline{X}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$ – выборочное среднее для выборки $X^{(i)}$;

$$s^2 = \sum_{i=1}^{2} (n_i - 1) s_i^2 / (\sum_{i=1}^{2} n_i - 2) \text{ – объединенная выборочная дисперсия;} \qquad (8)$$

$$s_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \overline{X}_i)^2 / (n_i - 1) \ - \qquad (9)$$

выборочная дисперсия для выборки $X^{(i)}$ $(i = 1, 2)$;

$\Delta(\varepsilon)$ – квантиль уровня $1–\alpha/2$ t–распределения Стьюдента с $n_1 + n_2 – 2$ степенями свободы.

Этот критерий оказывается чувствительным к отклонениям распределений от нормальности, различиям дисперсий σ_1^2 и σ_2^2 и неравенству объемов выборок n_1 и n_2.

Для обеспечения устойчивости решений к нарушению нормальности распределения и равенства дисперсий, разработаны устойчивые критерии, например, t–критерий Уэлча [1].

Для проверки гипотез

$H_\delta: \ \mu_1 - \mu_2 = \delta, \qquad \overline{H}_\delta: \ \mu_1 - \mu_2 \ne \delta$;

(значение δ фиксировано, в случае гипотезы равенства средних $\delta = 0$) при возможных нарушениях условия равенства дисперсий $(\sigma_1^2 \ne \sigma_2^2)$ используется тест вида

$$\text{принимается гипотеза } \begin{cases} H_\delta, & \text{если} \quad |t_1| \le \Delta(\varepsilon), \\ \overline{H}_\delta, & \text{если} \quad |t_1| > \Delta(\varepsilon). \end{cases}$$

Здесь

$t_1 = (\overline{X}_1 - \overline{X}_2 - \delta) / (s_1^2 / n_1 + s_2^2 / n_2)^{1/2}$ – статистика Уэлча,

$\Delta(\varepsilon)$ – квантиль уровня $1–\alpha/2$ t–распределения Стьюдента с ν_1 степенями свободы,

$$\nu_1 = \left(\frac{c_1^2}{n_1 - 1} + \frac{(1-c_1^2)^2}{n_2 - 1} \right)^{-1}, \quad c_1 = \frac{s_1^2 / n_1}{s_1^2 / n_1 + s_2^2 / n_2}.$$

Пример. Испытывались на растяжение образцы, у которых обработка поверхностей производилась двумя различными методами. Результаты испытаний приведены в табл. 4.

Таблица 4

Результаты испытаний образцов

Метод 1	16	14	19	20	15	18	18	19	17	18
Метод 2	13	19	14	14	15	10	17	21	13	15

Требуется решить при уровне значимости α = 0.05, могут ли данные испытаний принадлежать нормально распределенным совокупностям с одинаковыми средними значениями. Необходимо проверить гипотезы (5)

$$H_1: \ \mu_1 = \mu_2, \quad \overline{H}_1: \ \mu_1 \neq \mu_2;$$

в предположении независимости результатов испытаний, а также равенства дисперсий. Поскольку величина дисперсий неизвестна, оценим ее по обеим выборкам, используя формулу (8): s^2 = 6.85. Вычислим оценки $\overline{X}_1$ =17.4 и $\overline{X}_2$= 15.1. Подставив эти значения в (7), получим t = 1.97. Следует принять нулевую гипотезу H_1, так как порог двустороннего теста t (10; 0.975) = 2.101. Нет оснований утверждать, что один тип образцов прочнее другого.

10.2. Однородность дисперсий

Если необходимо выяснить, за счет чего обнаружилась неоднородность рассматриваемых выборок, то следует дополнительно произвести проверку однородности дисперсий. Статистический критерий однородности двух выборочных дисперсий основан на статистике

$$F = \frac{\sum_{i=1}^{n_1} (x_{1i} - \overline{X}_1)^2 /(n_1 - 1)}{\sum_{i=1}^{n_2} (x_{2i} - \overline{X}_2)^2 /(n_2 - 1)},$$

которая в условиях справедливости гипотезы H_2 (6) имеет F–распределение Фишера с числами степеней свободы числителя и знаменателя, равными соответственно $n_1 - 1$ и $n_2 - 1$.

Тест уровня значимости α для проверки (6) записывается следующим образом:

$$\text{принимается гипотеза } \begin{cases} H_2, & \text{если} \quad \Delta(1-\alpha/2) \le F \le \Delta(\alpha/2), \\ \overline{H}_2, & \text{если} \quad F > \Delta(\alpha/2) \text{ или } F < \Delta(1-\alpha/2), \end{cases}$$

где $\Delta(\varepsilon)$ – порог теста, определяемый при верной H_2 (6) как квантиль уровня ε F–распределения Фишера с $n_1 - 1$ и $n_2 - 1$ степенями свободы.

Если рассматривается односторонняя альтернатива $\sigma_1^2 > \sigma_2^2$, то H_2 отклоняется при $F \ge \Delta(1-\alpha)$, а для альтернативы $\sigma_1^2 < \sigma_2^2$ H_2 отклоняется при $F \ge \Delta(\alpha)$.

Пример. Необходимо установить, изменяется ли от одного дня к другому величина рассеяния температуры прибора. Для этого в первый день было проведено $n_1 = 12$ измерений, а во второй – $n_2 = 10$. Среднее квадратическое отклонение в первый день s_1 составило 23 градуса, а во второй – $s_2 = 30$ градусов. Известно, что температура прибора – это случайная величина, подчиняющаяся нормальному распределению.

При рассмотренных предположениях и $\alpha = 0.05$ критическое значение F–критерия с 9 и 11 степенями свободы $F(9;\ 11;\ 0.975) = 3.59$. Можно сделать вывод о том, что различий в дисперсиях, измеренных в разные дни, не установлено (принимается гипотеза H_2 (6)), так как вычисленное значение статистики $F = 30^2/23^2 = 1.70$, что меньше критического значения.

10.3. Равенство параметров двух биномиальных распределений

Имеются две выборки из биномиального распределения с параметрами p_j, n_j, $j = 1,2$. Необходимо проверить гипотезу о том, эти две выборки взяты из совокупностей, подчиняющихся биномиальному распределению с одинаковыми параметрами. Эту гипотезу можно записать как $H_0: \ p_1 = p_2$, или $H_0: \ p_1 - p_2 = 0$. Альтернативная гипотеза $H_1: \ p_1 \ne p_2$. Данная задача во многих прикладных областях называется задачей "сравнения двух долей". При достаточно большом объеме выборки можно использовать аппроксимацию нормальным распределением. Если нулевая гипотеза справедлива, то оценки p_j, относящиеся к каждой из выборок, являются несмещенными оценками одного и того же параметра. Следовательно, эти оценки можно объединить следующим образом:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2},$$

где X_j – число событий в выборке объема n_j, j =1,2. Величина $\hat{p}$ является оценкой максимального правдоподобия для параметра *p*. Вычислим среднее квадратическое отклонение разности выборочных оценок $\hat{p}_1 - \hat{p}_2$:

$$S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1 - \hat{p})\,(1/n_1 + 1/n_2)}\ .$$

Статистика критерия имеет вид

$$U = (\hat{p}_1 - \hat{p}_2)/S_{\hat{p}_1 - \hat{p}_2} = (X_1/n_1 - X_2/n_2)/S_{\hat{p}_1 - \hat{p}_2}\,.$$

При $p_1 = p_2$ распределение U почти совпадает со стандартным нормальным распределением $N(0,\ 1)$. Тест уровня значимости α использует статистику U и записывается следующим образом:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если} \quad |U| \le \Delta(\varepsilon), \\ H_1, & \text{если} \quad |U| > \Delta(\varepsilon), \end{cases}$$

где $\Delta(\varepsilon)$ – квантиль уровня $1-\alpha/2$ стандартного нормального распределения $u_{1-\alpha/2}$.

11. ГИПОТЕЗЫ О ЧИСЛОВЫХ ЗНАЧЕНИЯХ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК

11.1. Гипотезы о значении математического ожидания

Случай 1: дисперсия σ^2 известна. Пусть $x_1, x_2, ..., x_n$ – независимые одинаково распределенные случайные величины, имеющие нормальное распределение $N(\mu, \sigma^2)$ с неизвестным математическим ожиданием μ и известной дисперсией σ^2. Необходимо проверить гипотезу о числовом значении математического ожидания $H_0 : \mu = \mu_0$, против альтернативы $H_1 : \mu \neq \mu_0$. Решающее правило использует статистику критерия Z и имеет вид:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если} \quad |Z| \le \Delta(\varepsilon), \\ H_1, & \text{если} \quad |Z| > \Delta(\varepsilon), \end{cases}$$

где $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n})$, (10)

$\Delta(\varepsilon)$ – квантиль уровня $1–\alpha/2$ стандартного нормального распределения $u_{1-\alpha/2}$.

Если рассматривается односторонняя альтернатива $\mu > \mu_0$, то H_1 отклоняется при $Z \ge \Delta(1–\alpha)$, а для альтернативы $\mu < \mu_0$ H_1 отклоняется при $Z \le \Delta(\alpha)$.

Случай 2: дисперсия σ^2 неизвестна. В качестве выборочной оценки σ^2 используется s^2 (9). Статистика

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

имеет t–распределение Стьюдента с $\nu_1 = n - 1$ степенями свободы. Решающее правило имеет вид:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если} \quad |t| \le \Delta(\varepsilon), \\ H_1, & \text{если} \quad |t| > \Delta(\varepsilon), \end{cases}$$

$\Delta(\varepsilon)$ – квантиль уровня $1–\alpha/2$ t–распределения Стьюдента с $n - 1$ степенями свободы.

Если рассматривается односторонняя альтернатива $\mu > \mu_0$, то H_1 отклоняется при $t > \Delta(1–\alpha)$, а для альтернативы $\mu < \mu_0$ H_1 отклоняется при $t < \Delta(\alpha)$.

Пример. На заводе выпускаются стандартные трубы с внутренним диаметром (ВД), равным 2.40 дюйма, причем дисперсия этой величины σ^2 составляет 0.0004 дюйм^2. Для выборки объема $n = 25$ произведены измерения ВД. Среднее значение ВД, вычисленной по этой выборке, составляет 2.41 дюйма. Можно ли утверждать, что выборка взята из нормальной совокупности с указанными параметрами?

Вычислим значение статистики критерия Z по формуле (10):

$$Z = (2.41–2.40) / (0.02/\sqrt{25}) = 2.50.$$

При уровне значимости α = 0.05 гипотеза о равенстве среднего значения 2.40 дюйма отклоняется, так как Z =2.50 > $u_{1-\alpha/2}$= 1.96.

Если считать, что дисперсия σ^2 неизвестна, а ее выборочная оценка s^2 составляет 0.0004 дюйм^2, тогда для проверки гипотезы следует воспользоваться статистикой *t*. Вычисленное значение *t* сравнивается с квантилью уровня 1 – *α/2* = 0.975 *t*–распределения Стьюдента с *n* – 1 =24 степенями свободы, которая равна 2.064.

11.2. Гипотезы о значении дисперсии

Пусть $x_1, x_2, ..., x_n$ – независимые одинаково распределенные случайные величины, имеющие нормальное распределение $N(\mu, \sigma^2)$ с неизвестными математическим ожиданием μ и дисперсией σ^2. Необходимо проверить гипотезу о числовом значении дисперсии $H_0: \sigma^2 = \sigma_0^2$, против альтернативы $H_1: \sigma^2 \neq \sigma_0^2$. Статистика критерия χ^2 вычисляется по следующей формуле:

$$\chi^2 = (n-1)\ s^2/\sigma_0^2,$$

где s^2 – выборочная оценка дисперсии (9).

При уровне значимости α решающее правило имеет вид:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если} \quad \Delta(\alpha) < \chi^2 < \Delta(1-\alpha/2), \\ H_1, & \text{если} \quad \chi^2 \le \Delta(\alpha) \text{ или } \chi^2 \ge \Delta(1-\alpha/2), \end{cases}$$

$\Delta(\varepsilon)$ – квантиль уровня ε χ^2–распределения с $n-1$ степенями свободы.

Если рассматривается односторонняя альтернатива $\sigma^2 > \sigma_0^2$, то H_1 отклоняется при $\chi^2 > \Delta(1-\alpha)$, а для альтернативы $\sigma^2 < \sigma_0^2$ H_1 отклоняется при $\chi^2 < \Delta(\alpha)$.

Пример. В результате проведения стандартной процедуры проверки коэффициента упругости образцов резины установлено, что среднее квадратическое отклонение при измерении этого коэффициента составляет 18.0 единиц. Взята выборка объема $n = 20$ и получено $s = 23.2$ единицы. Обосновано ли предположение о нестабильности стандартной процедуры проверки коэффициента упругости?

Гипотезы запишутся следующим образом: H_0: $\sigma^2 = \sigma_0^2 = 324$, H_1: $\sigma^2 \neq \sigma_0^2$. Пусть уровень значимости $\alpha = 0.05$. Вычисленное значение статистики критерия χ^2 равно $19 \cdot (23.2)^2/(18.0)^2 = 31.56$. Критические значения для двустороннего критерия равны $\chi^2(19;\ 0.025) = 8.907$ и $\chi^2(19;\ 0.975) = 32.85$. Следовательно, нет оснований отклонить H_0. Критическое значение для одностороннего критерия (также при $\alpha = 0.05$ равно либо $\chi^2(19;\, 0.05) = 10.12$, либо $\chi^2(19;\, 0.95) = 30.14$.

11.3. Гипотезы о значении параметра биномиального распределения

Пусть $x_1, x_2, ..., x_n$ – независимые одинаково распределенные случайные величины, имеющие биномиальное распределение с параметрами p и n :

$$\mathbf{P}\{X = x\} = C_n^x p^x (1-p)^{n-x}.$$

Необходимо проверить гипотезу (при уровне значимости α) о числовом значении p_0 параметра p биномиального распределения $H_0 : p = p_0$, против альтернативы H_1: $p \neq p_0$. Критическое значение статистики при достаточно большом n вычисляется с применением нормальной аппроксимации по формуле

$$Z = \frac{\hat{p} - np_0}{\sqrt{p_0(1-p_0)n}},$$

где $\hat{p}$ – частота появления интересующего нас значения случайной биномиальной величины в последовательности из n испытаний. Решающее правило использует статистику критерия Z и имеет вид:

$$\text{принимается гипотеза } \begin{cases} H_0, & \text{если} \quad |Z| \le \Delta(\varepsilon), \\ H_1, & \text{если} \quad |Z| > \Delta(\varepsilon), \end{cases}$$

где $\Delta(\varepsilon)$ – квантиль уровня $1-\alpha/2$ стандартного нормального распределения $u_{1-\alpha/2}$.

Если рассматривается односторонняя альтернатива $p > p_0$, то H_1 отклоняется при $Z \ge \Delta(1-\alpha)$, а для альтернативы $p < p_0$ H_1 отклоняется при $Z \le \Delta(\alpha)$.

Пример. Мы наблюдали 60 раз герб при 100 бросаниях монеты. Необходимо при уровне значимости $\alpha = 0.05$ проверить гипотезу о "правильности" монеты, т.е. $H_0 : p = 0.5$ против альтернативы, что монета асимметрична, из-за чего чаще выпадает герб: H_1: $p = p_1 > p_0$. Вычисляем значение статистики критерия:

$$Z = \frac{60 - 0.5 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = 1.900.$$

По таблицам нормального распределения находим, что значение функции распределения равно 0.97128, тогда P–значение равно 0.02872. Следовательно, нулевая гипотеза о симметрии монеты должна быть отклонена при уровне значимости $\alpha = 0.05$.

12. ПРОВЕРКА ГИПОТЕЗ ПО ЭКСПЕРТНЫМ ДАННЫМ И ПОСТРОЕНИЕ ВЫВОДОВ

Предположим, что задано множество гипотез $H = \{h_i\}$, $i = 1,\ldots, N,$ и множество свидетельств (симптомов) $E = \{e_j\}$, $j = 1,\ldots, M$, относительно гипотез. На основании информации о свидетельствах и их взаимосвязях с гипотезами необходимо принять решение о достоверности гипотез. Такая постановка определяет так называемые диагностические задачи, в частности задачи медицинской и технической диагностики, разведки полезных ископаемых и др. В качестве априорной информации о задачах выступают знания о гипотезах и их взаимосвязях с симптомами, приобретенные экспертом из предыдущего опыта. Знания о фактическом состоянии симптомов задаются пользователем в процессе решения задачи. Решение конкретной практической задач может производиться с помощью различных механизмов вывода (МВ), опирающихся на тот или иной метод работы с неопределенностью. Наиболее часто при решении практических диагностических задач используются байесовские и близкие к ним МВ, основанные на коэффициентах уверенности и Демпстера-Шафера. Такие МВ имеют сходную структуру, которую можно представить в виде одного обобщенного алгоритма вывода.

Предположим, что связь между гипотезами и свидетельствами выражается в виде правил продукций "Если e, то h", каждому из которых ставится в соответствие значение функции $\delta : V \to R^n$, $n \geq 1$ (сила правила), где $V \subseteq HxE$ – множество всех пар взаимосвязанных гипотез и свидетельств. В общем случае сила правила – это уверенность в гипотезе в предположении, что свидетельство известно с определенностью. Значения силы правил задаются экспертом и хранятся в базе знаний.

Пусть $D(e)$ – достоверность события e, $D_0(h)$ – априорная достоверность гипотезы h; $C_d(e)$ – функция, задающая порядок использования свидетельства e согласно дисциплине d ; $R_s(x)$ – функция сканирования опытных данных в шкале x; $G(x)$ – функция преобразования сканированных данных в значение достоверности; $Q(x, y, z)$ – функция вычисления апостериорной оценки достоверности гипотез по x, y, z; E_h – множество свидетельств $e \in E$, связанных с h.

Итак, пусть заданы $D_0(h)$ и $\delta(e, h)$ $\forall h \in H$ и $\forall e \in E_h$.

Вычислить $D(h)$ $\forall\, h \in H$.

В принятых обозначениях опишем схему алгоритма вывода.

1. Начальная установка $D(h) = D_0(h)$ $\forall\, h \in H$.
2. Выбор свидетельства e согласно критерию $C_d(e)$ $\forall\, e \in E_h \subseteq E$.
3. Вычисление $D(e) = G(R_s(e))$.
4. Вычисление $D(h) = Q(D(h), \delta(e, h), D(e))$ $\forall\, h \in H$ $(e \in E_h)$. Выход.

При реализации функции $C_d(e)$ предусматривается возможность вернуться на несколько шагов назад.

Байесовский механизм вывода. Эти МВ основаны на использовании известной теоремы Байеса [4]. Для них достоверность $D(e) = P(e)$, где $P(e)$ – вероятность события e. Сила правила $\bar{\delta}(e, h) = (P(e|h, P(e)|\bar{h}))^T$. Вход $D_0(h)$ и $\bar{\delta}(e, h)$ $\forall\, h \in H$ и $\forall\, e \in E_h$.

Обозначим $P(h|e^{'})$ оценку достоверности гипотезы h , полученную после привлечения свидетельства e. Причем $P(h|e^{'}) = Q(x, \bar{y}, z)$, где z представляет опытную достоверность свидетельства e. Для практического вычисления $P(h|e^{'})$ можно воспользоваться следующей функцией $Q_1(P(h), \bar{\delta}(e, h), R_s(e))$:

$$Q_1(P(h), \bar{\delta}(e, h), R_s(e)) = \begin{cases} P(h) * (1 + \dfrac{R_s(e)}{5}) - \dfrac{R_s(e) * P(h \mid \bar{e})}{5}, & R_s(e) \le 0, \\ P(h) * (1 - \dfrac{R_s(e)}{5}) + \dfrac{R_s(e) * P(h \mid e)}{5}, & R_s(e) \ge 0. \end{cases}$$

Вероятности $P(h|e)$ и $P(h|\bar{e})$ определяются по формуле Байеса:

$$P(h|e) = \frac{P(h) * P(e \mid h)}{P(h) * P(e \mid h) + (1 - P(h)) * P(e \mid \bar{h})}$$

Модификацией байесовских механизмов вывода является интервальный байесовский МВ, при реализации которого все вероятности задаются в виде интервалов из [0, 1]. Использование интервальных оценок имеет смысл, так как эксперту часто трудно указать точечные значения оценок вероятностей, особенно $P(e)|\bar{h})$. При таком подходе, учитывая найденные на каждом шаге интервальные оценки текущих вероятностей, получаются интервальные оценки вероятностей гипотез. Следует подчеркнуть, что когда начальные вероятности задаются точечными значениями, то интервальный МВ совпадает с обычным байесовским.

ЛИТЕРАТУРА

1. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное изд.. / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. М.: Финансы и статистика, 1983.
2. Браунли К. А. Статистическая теория и методология в науке и технике. М.: Наука, 1977.
3. Вентцель Е. С. Теория вероятностей. М.: Наука, 1964.
4. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. М.: Финансы и статистика, 1982.

ПРИЛОЖЕНИЕ
ОДНОВЫБОРОЧНЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗ

Нулевая гипотеза	Допущения	Альтернативная гипотеза	Статистика критерия	Область отклонения
$\mu = \mu_0$	n большое, σ_0^2 известна или нормальность, σ_0^2 известна	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$	$Z \geq z_\alpha$ $Z \leq -z_\alpha$ $\lvert Z \rvert \geq z_{\alpha/2}$
$\mu = \mu_0$	n большое, σ_0^2 неизвестна	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$	$Z \geq z_\alpha$ $Z \leq -z_\alpha$ $\lvert Z \rvert \geq z_{\alpha/2}$
$\mu = \mu_0$	нормальность, n малое, σ_0^2 неизвестна	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$T = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}}$	$T \geq t_{\alpha,\, n-1}$ $T \leq -t_{\alpha,\, n-1}$ $\lvert T \rvert \geq t_{\alpha/2,\, n-1}$
$p = p_0$	биномиальные испытания, n большое	$p > p_0$ $p < p_0$ $p \neq p_0$	$Z = \dfrac{\hat{p} - np_0}{\sqrt{p_0(1 - p_0)n}}$	$Z \geq z_\alpha$ $Z \leq -z_\alpha$ $\lvert Z \rvert \geq z_{\alpha/2}$
$\sigma^2 = \sigma_0^2$	нормальность	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 = (n - 1)\ \dfrac{s^2}{\sigma_0^2}$	$\chi^2 \geq \chi^2_{\alpha,\, n-1}$ $\chi^2 \leq \chi^2_{1-\alpha,\, n-1}$ $\chi^2 \leq \chi^2_{1-\alpha/2,\, n-1}$ или $\chi^2 \geq \chi^2_{\alpha/2,\, n-1}$

Св. план 2000, поз. 19

Учебное издание

Степанова Маргарита Дмитриевна

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

УЧЕБНО – МЕТОДИЧЕСКОЕ ПОСОБИЕ

по курсу "Статистические основы индуктивного вывода"
для студентов специальности "Искусственный интеллект"

Редактор Т. Н. Крюкова

Корректор

Подписано в печать . . . Формат 60х84 1/16.

Бумага Печать офсетная. Усл.печ.л.

Уч.-изд. л. 2,0 Тираж 150 экз. Заказ

Белорусский государственный университет информатики и радиоэлектроники

Отпечатано в БГУИР. Лицензия ЛП № 156. 220027, Минск, П. Бровки, 6