Проблемы проверки статистических гипотез

Цейтлин Н. А.

(Dr. Natan Tseitlin; tseitlin@gmx.net; фирма CuBe Matrix GbR, Гамбург, ФРГ, www.cubematrix.com)

Аннотация

Получил дальнейшее развитие «Альфа-метод проверки статистических гипотез»; определена роль аналитического статистика в решении проблем формулировки гипотез, выбора критического уровня значимости, расчёта оценок уровня значимости, мощности критерия, решения спорных случаев, выявления «порочной спирали» в теории проверки гипотез и её преодоления, а также – роль лица, принимающего решение, несущего ответственность за обоснованный вывод.

Реферат

Всегда легче заявлять о своих принципах, чем осуществлять их. (У. Черчилль)

В конце прошлого века автор собрал вместе и кратко описал различные методы проверки 23-х статистических гипотез относительно параметров нормально распределённых случайных величин – математических ожиданий и дисперсий. Удобная для практической работы модификация метода проверки гипотез, названная «Альфа-метод проверки гипотез», подробно, с рассуждениями и многочисленными примерами изложена в работах [1, 7, 8].

В настоящей статье автор делится с коллегами своим пониманием и накопленным опытом решения **проблем** формулировки гипотез, выбора критического уровня значимости, расчёта оценок уровня значимости, мощности критериев, решения спорных случаев, выявления «порочной спирали» в теории проверки гипотез и её преодоления, приводит ряд полезных замечаний и цитирует мнения коллег, в некоторых фрагментах дискуссий.

Формулируя статистические гипотезы, лицо, принимающее решения (ЛПР), заинтересовано в том, чтобы убедить читателя в справедливости предпочитаемой и впоследствии принятой им гипотезы. «Последнее слово» об отклонении или принятии проверяемой гипотезы отдаётся экспертной оценке уровня ответственности ЛПР за вывод.

38 с., 3 табл., 7 рис., 35 ссылок.

Ключевые слова

Прикладная математическая статистика; аналитический статистик; лицо, принимающее решение; пакет прикладных программ; вероятность; случайная величина; формулировка статистических гипотез; параметрическая гипотеза; нулевая гипотеза; альтернативная гипотеза; проверка гипотез; альфа-метод проверки статистических гипотез; доверительный интервал; метод доверительных интервалов; критический уровень значимости; оценка уровня значимости; численный метод проверки гипотез; дискуссии с коллегами.

Discussion of the problems of statistical hypothesis testing

(Dr. Tseitlin tseitlin@gmx.net; firm CuBe Matrix GbR, Hamburg, Germany, www.cubematrix.com)

Annotation

The author develops "An alpha method testing of statistical hypotheses" and discusses a function of the analytical statistician in the following problems solving: statement of hypotheses, choice of critical significance level, estimation of significance point, power of test, decision of questionable incident, overcoming of “circularity spiral” in the theory of testing hypotheses.

Diskussion der Probleme der statistischen Hypothesentests

(Dr. Tseitlin tseitlin@gmx.net; Firma CuBe Matrix GbR, Hamburg, Deutschland, www.cubematrix.com)

Abstract

At the end of the last century, the author has gathered and briefly describedvarious methods fortesting of23 types of statistical hypotheses on the parameters of normally distributed random variables - expectations and variances. A practically convenient modification of the hypotheses testing method, called "alpha-method of hypotheses testing", is presented in [1, 7, 8]in detail, with numerous examples and argumentations.
In this article the author shares with his colleagues his understanding and experience addressing the formulation of hypotheses, the choice of the critical level of significance, the estimation of the level of significance, power of the tests, the solutionsto questionable cases,theidentificationof "vicious spiral" in the theory of hypothesis testing and how to overcome it, brings a number of useful comments and quoted excerpts from discussions with colleagues expressing their views.
In formulating statistical hypotheses, the decision-maker, is interested in convincing the reader of the validity of the preferred and subsequently accepted hypothesis. "The finalsay" on the rejection or acceptance of the hypothesis being tested is given to an expert evaluation of the level of responsibility that the decision maker bears for the outcome.

3 tables, 4 figures, 21 pages, 32 references.

Keywords

Applied Mathematical Statistics; Analytical statistics; decision-maker; application package; probability; random variable; formulation of statistical hypotheses; parametric hypothesis; null hypothesis; alternative hypothesis; hypothesis testing; confidence interval; critical level of significance; assessment of the level of significance; numerical method of hypothesestesting; discussions with colleagues.

Содержание

«Если рассмотреть проблему достаточно внимательно, то вы увидите себя как часть этой проблемы». (Аксиома Дучарма)

Предисловие

Можно всё, что недоказуемо.

Важным **инструментом** научного исследования является **метод проверки статистических гипотез**. Удобная для практической работы **модификация** этого метода, названная «**Альфа-метод проверки гипотез**», подробно, с рассуждениями и многочисленными примерами изложена в работах [1, 7, 8].

Согласно **Альфа-методу**, в качестве статистической характеристики проверяемой гипотезы используется **точечная (однозначная) оценка уровня значимости.** Результат проверки гипотез зависит от соотношения между точечной **оценкой уровня значимости** и его **критическим значением**, априори задаваемым ЛПР в соответствии с избранным им уровнем ответственности за статистический вывод.

Многолетний опыт работы автора в роли аналитического статистика (АСа) совместно со специалистами в предметных областях показывает, что среди множества проблем, возникающих при проверке статистических гипотез, наиболее существенными являются проблемы формулировки гипотез и выбора критического уровня значимости для их проверки. Рассматривать эти проблемы приходится в совокупности с другими, возможно и менее существенными, но также – довольно значимыми проблемами.

Многие коллеги-предметники (биологи, медики, экологи, физхимики, химики-технологи, метеорологи, инженеры-строители, физики и др.) зачастую представляют результаты статистической обработки данных не вполне корректно. Например, модели регрессии представляют только точечными оценками (в виде средних значений оценок

коэффициентов или линий регрессии), а проверяя различные статистические гипотезы выбирают один и тот же критический уровень значимости 0,05. Тем не менее, суть построения статистических моделей – в их вероятностном представлении (например, с помощью эмпирических функций распределения или, хотя бы, в виде интервальных оценок), а выбор критического уровня значимости зависит от различных обстоятельств. В настоящей статье автор делится своим пониманием и опытом решения названных проблем.
Статья публикуется на одном из сайтов **Сети**, что создаёт определённые удобства для читателя: **определение любого термина читатель может мгновенно найти через поисковую машину, любую ссылку немедленно открыть.**
Автор желает читателю лёгкого чтения и глубокого понимания рассмотренных проблем. Критические замечания, предложения, советы и сведения коллег о применении изложенных в статье методов проверки гипотез на практике будут приняты с благодарностью.

Автор.

Некоторые обозначения: АС — аналитический статистик; ЛПР — лицо, принимающее решение; ППП – пакет прикладных программ; СВ — случайная величина; H_0 и H_1 – нулевая и альтернативная гипотезы; ДИ - доверительный интервал.

Тексты эпиграфов заимствованы из интернета.

1. Ключевые понятия [1, 4-8]

Сёма заглядывают в комнату, где папа с ножницами возится с книгами.
- Папа! Зачем ты уничтожаешь книги?
- Я не уничтожаю книги, а создаю свою!

Статистическими гипотезами назовём предположения о свойствах распределения генеральной совокупности, которые можно проверить, опираясь на данные случайной выборки из теоретически бесконечной генеральной совокупности. Генеральная совокупность есть идеализация действительной совокупности, из которой получена выборка. Обычно проверяемую гипотезу называют «нулевой гипотезой» и обозначают «H_0».

Определение. Нулевыми, «H_0», считаются гипотезы, утверждающие, что различие между сравниваемыми величинами отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями в выборках.

Остальные гипотезы, отличающиеся от H_0 и противопоставляемые им, называются альтернативными и обозначаются «H_1».

Смысл проверки статистических гипотез состоит в том, чтобы по данным случайной выборки принять или отклонить проверяемую гипотезу H_0 с минимальным риском ошибки.

Свод правил, указывающих, при каких результатах наблюдений гипотеза H_0 отклоняется, а когда - не отклоняется, называют «**статистическим критерием**».

2. Распределение вероятностей при испытании нулевой гипотезы

Выбор правильных решений приходит с опытом, а опыт приходит с накоплением неправильных решений.

Пусть $u \in W$ - некоторая функция наблюдений, составляющих выборку (проверочная статистика), а W - пространство всех возможных значений u. Разделим пространство W на критическую ω и допустимую W-ω области, такие, что если $u \in \omega$, то гипотеза H_0 неверна. Таким образом, выбор критерия проверки гипотезы H_0 сводится к выбору проверочной статистики u и критической области ω.

Если отклоняют гипотезу H_0, а она верна, совершают ошибку первого рода, вероятность которой (табл. 1)

Таблица 1. Распределение вероятностей (P) при испытании нулевой гипотезы (H_0) против альтернативной гипотезы (H_1)

Решение	Истина	
	H_0 верна	H_1 верна
Отклонить H_0 (принять H_1)	Ошибка первого рода, $P = \alpha$	Верно, $P = 1-\beta$
Отклонить H_1 (не отклонить H_0)	Верно, $P = 1-\alpha$	Ошибка второго рода, $P = \beta$

$$\alpha = P\,(u \in \omega \mid H_0) \tag{1}$$

называется «уровнем значимости».

Если не отклоняют гипотезу H_0, а она неверна, совершают ошибку второго рода с вероятностью

$$\beta = P[u \in (W - \omega) \mid H_1]. \tag{2}$$

Мерой полезности критерия или его способности отделить гипотезу H_0 от гипотезы H_1 служит **мощность критерия**

$$1 - \beta = P(u \in \omega \mid H_1). \tag{3}$$

Гипотеза называется **простой**, если она утверждает, что параметр θ может принять одно значение (например, $\theta = \theta_0$). **Сложной** считается гипотеза, утверждающая, что параметр θ может принять некоторые значения из заданного множества (например, $\theta > \theta_0$).

3. Виды параметрических гипотез

Есть и рассказывать следует медленно, стараясь тщательно разжёвывать продукт.

Рассмотрим **параметрические гипотезы** относительно $\theta \in \Theta$ - параметра распределения СВ Y:

-простую против простой альтернативы:

$$H_0: \theta = \theta_0; \qquad H_1: \theta = \theta_1; \tag{4}$$

(при $\theta_1 > \theta_0$ гипотеза H_1 является правосторонней альтернативой, и обозначается «$H_{1п}$»; при $\theta_1 < \theta_0$ гипотеза H_1 является левосторонней альтернативой, и обозначается «$H_{1л}$»);

- простую против сложной двусторонней альтернативы:

$$H_0: \theta = \theta_0; \qquad H_1: \theta \neq \theta_0; \tag{5}$$

(гипотеза H_1 называется двусторонней, потому что представляет собой две альтернативы: $H_{1л}$: $\theta < \theta_0$ - левостороннюю и $H_{1п}$: $\theta_1 > \theta_0$ – правостороннюю);

- сложную левостороннюю против сложной правосторонней альтернативы:

$$H_0: \theta \leq \theta_0; \quad H_{1п}: \theta > \theta_0; \tag{6}$$

- сложную правостороннюю против сложной левосторонней альтернативы:

$$H_0: \theta \geq \theta_0; \quad H_{1л}: \theta < \theta_0. \tag{7}$$

Критическая область двустороннего критерия для проверки гипотез (5) состоит из двух частей (рис. 1, поз. а, г).

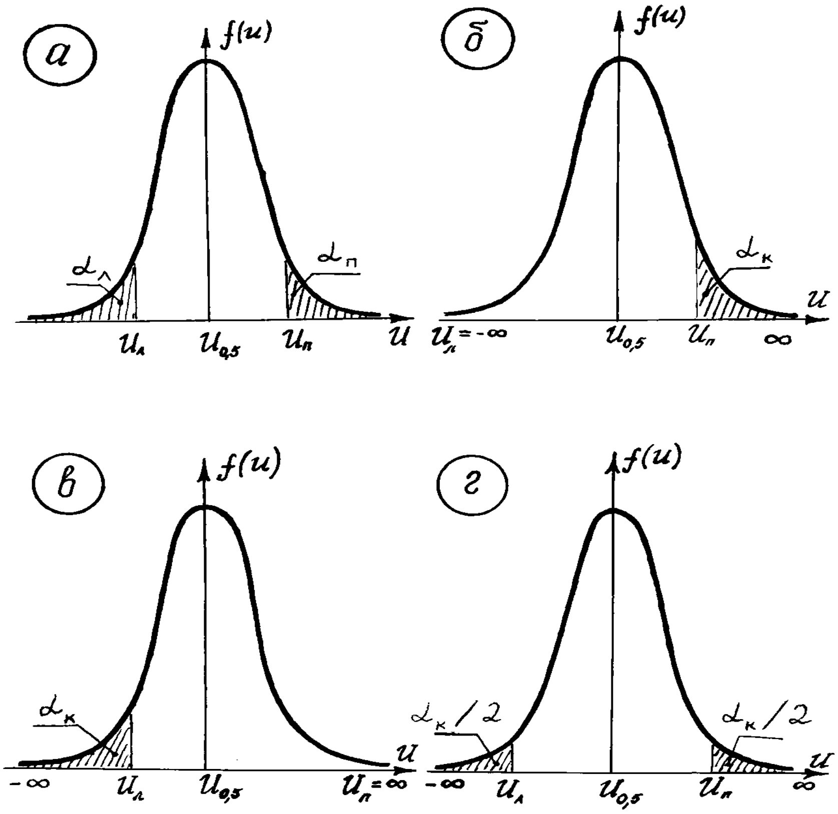

Рис. 1. Верхние α-пределы и критические области статистики u:

а – общий вид; б, в, г - изображения критических областей при право-, лево- и двусторонней альтернативах соответственно; $u_л$, $u_п$ - лево- и правосторонний верхний α-предел соответственно; α_k - критическое значение уровня значимости (площадь, заштрихованная под кривой f(u) - плотности распределения статистики u).

Обычно границы критической области выбираются так, чтобы вероятности попадания в левую и правую части были бы одинаковыми и равными $\alpha/2$. **Критическая область одностороннего (правостороннего или левостороннего) критерия** для проверки гипотез (4), (6) и (7) содержит всего одну зону, вероятность попадания в которую равна α (Рис. 1, поз. б, в).

При построении статистических критериев желательно свести к минимуму вероятности обоих типов ошибок - α и β. Однако уменьшение вероятности ошибки одного типа ведет к увеличению вероятности ошибки другого типа. Единственный способ уменьшить одновременно вероятностей ошибок обоих типов состоит в увеличении объёма выборки, что редко практикуется. Обычно выбирают определённый - критический уровень значимости α_k, и стремятся найти критерий, максимизирующий вероятность (1 - β) то есть, наболее мощный критерий.

4. Формулировка статистических гипотез

«Мы стремимся в большей степени к тому, чтобы отвести от себя страдания, нежели - к тому, чтобы получить удовольствие». (З. Фрейд)

Роли людей, участвующих в принятии решений, условно можно распределить так. **Эксперты** – специалисты в предметной области, среди которых может быть и аналитический статистик (АС) [1], занимаются подготовкой вариантов решений. **Лицо, принимающее решение (ЛПР),** несущее ответственность за ошибки, делает окончательный выбор. Иногда при решении важных задач ЛПР отвечает своими средствами, имуществом, свободой, и даже, головой за допущенные ошибки. Однако может быть, что при решении неважных задач, ЛПР, эксперты и АС - это одно лицо, **добросовестно** «играющее» названные роли.

В конечном итоге **ЛПР формулирует статистические гипотезы, выбирает критический уровень значимости, и принимает решение в пользу H_0 или H_1.**

Назовём научной гипотезой любое содержательное предположение о свойствах наблюдаемого объекта. Если свойства объекта могут быть измерены с помощью измерительных шкал, то появляется возможность сформулировать и логически соответствующие научным гипотезам - **статистические гипотезы** относительно параметров измеряемых свойств.

Рассмотрим **для простоты** две противоположные статистические гипотезы H_A и H_B. Одну из этих гипотез следует сформулировать в качестве проверяемой - нулевой (H_0), другую - в качестве альтернативной (H_1).

Существующая **техника выбора проверяемой нулевой гипотезы H_0** из двух противоположных гипотез H_A и H_B основана на **принципе логического «отрицания отрицания»**. Прежде чем сформулировать гипотезы как H_0, или H_1, ЛПР необходимо проанализировать, какие ошибки в принятии решений (см. табл. 1) приведут к более тяжёлым последствиям. Одну из гипотез (H_A или H_B), ошибочное отклонение которой приведёт к б**о**льшему ущербу (и поэтому наиболее нежелательно), ЛПР назовёт нулевой H_0 (если она не противоречит приведенному выше формальному определению понятия «нулевая гипотеза, H_0»), а другую, противопоставляемую ей гипотезу - альтернативной H_1. Такая формулировка позволяет ЛПР выбрать затем нужный критический уровень значимости α_k: чем **выше** его уровень ответственности Э и тяжелее ущерб от ошибочного отклонения проверяемой нулевой гипотезы H_0, тем **меньшее** значение α_k приходится принимать (подробнее см. ниже).

Из рассматриваемых нами гипотез (4) – (7) только из пары гипотез (5) гипотезу $\theta \neq \theta_0$ невозможно, к сожалению, назвать нулевой H_0 (двустороннюю гипотезу $\theta \neq \theta_0$ сформулировать в качестве нулевой нельзя в принципе, несмотря на то, что «её ошибочное отклонение может иметь более ущербные последствия», так как подобная формулировка будет противоречить приведенному выше определению понятия «нулевая гипотеза, H_0»). Поэтому в данном **исключительном** случае ЛПР придётся выбирать нужный критический уровень значимости α_k, пользуясь «обратной шкалой»: чем **выше** его уровень ответственности, тем **большее** значение α_k надо принять (подробнее см. ниже)!

Важно отметить, что подобные «исключительные случаи» нередко встречаются на практике, когда исследователь заинтересован в принятии нулевой гипотезы H_0: при проверке согласия распределений – эмпирического и теоретического (часто – нормального), при проверке гипотезы об адекватности эмпирической функции регрессии (о равенстве дисперсий отклика – остаточной и воспроизводимости), при проверке гипотезы об отсутствии мультиколлинеарности (корреляции между базисными функциями) в регрессионном анализе, проверке гипотезы об отсутствии разности между двумя центрами распределения СВ и в ряде других задач.

Часто при определении гипотезы ЛПР бывает **удобнее использовать более простой принцип «предпочтительности»**, основанный на сравнении преимуществ от верного принятия одной из двух гипотез (или, перефразируя З. Фрейда: «Некоторые люди стремятся в большей степени к тому, чтобы получить удовольствие, нежели - к тому, чтобы отвести от себя страдания»).

Прежде, чем сформулировать гипотезы H_0 или H_1, ЛПР может определить, какая правильно принятая гипотеза даст ему большие преимущества (см. табл. 1). После этого гипотезу (H_A или H_B), которую предпочитает принять ЛПР, следует назвать альтернативной H_1, а противопоставляемую ей гипотезу - нулевой, H_0 (если это не противоречит приведенному выше определению понятия «нулевая гипотеза H_0»).

Заметим, что критическая область критерия для сложной альтернативы может быть построена только, исходя из соотношения **равенства** (см. рис. 1). Поэтому в запись проверяемой нулевой гипотезы H_0 **всегда** дописывается знак равенства!

Пример 1. Если из двух гипотез $\theta < \theta_0$ и $\theta > \theta_0$ ошибочное отклонение гипотезы $\theta > \theta_0$ приведёт к большему ущербу (и поэтому нежелательно), то, согласно **принципу «отрицания отрицания»**, ЛПР назовёт нулевой гипотезу H_0: $\theta \geq \theta_0$ (это не противоречит приведенному выше определению понятия «нулевая гипотеза, H_0»), а противопоставляемую ей гипотезу $\theta < \theta_0$ - альтернативной, H_1: $\theta < \theta_0$.

Согласно **принципу «предпочтительности»**, если ЛПР предпочитает принять $\theta < \theta_0$, то её и следует сформулировать альтернативной H_1: $\theta < \theta_0$, и затем - H_0: $\theta \geq \theta_0$.

Проверяемой всегда является нулевая гипотеза H_0, которой противопоставляется альтернативная гипотеза H_1. Одна из этих гипотез является **предпочтительной** для исследователя. Выбирая предпочтительную гипотезу, ЛПР должен чётко представлять себе, чем он рискует, и какой должна быть критическая (пороговая) вероятность понесённого ущерба в случае ошибки. Поэтому рассмотрим подробно процедуру выбора критического уровня значимости.

5. Выбор критического уровня значимости

Право на глупость - одна из гарантий свободного развития личности.

Проблема выбора критического уровня значимости часто возникает перед ЛПР (или пусть – перед исследователем, но - в роли ЛПР). По традиции, сложившейся в «докомпьютерную эру» - в начале прошлого века, когда «рабочими инструментами» статистика были - логарифмическая линейка и бумажные статистические таблицы, прикладниками использовались «популярые» значения **критических уровней значимости** $\alpha_к \in$ (0,1, 0,05, 0,01, 0,001). Перечисленные значения $\alpha_к$ приводились только на входе таблиц. Так что достигаемые в расчётах уровни значимости α определять по этим таблицам можно было лишь путём интерполяции. А малые значения α, $\alpha < 0{,}001$ даже не публиковались!

С этой «традицией» давно следует покончить: ведь почти все статистические ППП для ЭВМ оканчиваются расчётом точечных оценок α уровней значимости, которые используются для статистических выводов.

Обратимся к рекомендациям классиков. Э. Леман **[5]** в фундаментальной монографии [**5,** с. 91] пишет: «Обычно выбор уровня значимости α_k до некоторой степени произволен, поскольку в большинстве ситуаций нет точной границы для «разрешенной» вероятности ошибки первого рода. Стало обычным выбирать для α_k одно из стандартных значений, таких, как 0,005; 0,01 и 0,05. Эта стандартизация имеет некоторые преимущества, так как она позволяет сократить объём используемых таблиц. Никакой другой специальной причины для выбора именно этих значений нет.

Важным обстоятельством, влияющим на выбор α_k, является отношение ЛПР к гипотезе H_0 до проведения эксперимента. Если ЛПР твёрдо верит в истинность гипотезы H_0, то потребуется убедительное свидетельство против неё для того, чтобы ЛПР отказалось от своей уверенности; соответственно уровень значимости α_k будет выбран весьма низким.

Н. А. Плохинский **[6]** вводит нематематическое, экспертное - для ЛПР - понятие «уровень ответственности за выводы» - об отклонении гипотезы H_0: чем он выше, тем меньше α_k. Автор работы **[1]** предложил при выборе α_k учитывать также и наименование предпочитаемой исследователем гипотезы: выбирая значения α_k, ответственность за выводы несёт ЛПР. Уровень ответственности одни ЛПР могут оценить в процентах, скажем, от 0% до 100% (табл. 2), другие (как писал Э. Леман **[5]**) –

Таблица 2. Ориентировочные нормы для выбора лицом, принимающим решения (ЛПР) критического значения уровня значимости α_{ki} [1].

Уровень **ответственности** ЛПР за вывод		Уровень значимости α_{ki}*, если для ЛПР предпочтительна гипотеза:	
название	экспертная оценка $Э_i$*	альтернативная, H_1	нулевая, H_0
нижний предел	0	1	0
малый	10	0,6	$5\cdot10^{-5}$
	20	0,4	10^{-3}
	30	0,2	0,006
обычный	40	0,1	0,02
	50	0,05	0,05
	60	0,02	0,1
большой	70	0,006	0,2
	80	10^{-3}	0,4
	90	$5\cdot10^{-5}$	0,6
верхний предел	100	0	1

*$i \in (0 \cup 1)$, то есть, индекс $i = 1$, если предпочтительна альтернативная гипотеза H_1 или $i = 0$, если предпочтительна (в порядке исключения) нулевая гипотеза H_0.

непосредственно в «субъективных вероятностях», в данном случае - в некотором, основанном на опыте или интуитивном представлении ЛПР - о пороге критической вероятности α_{ki} практически невозможного события.

Если исследователь предпочтитает альтернативную гипотезу H_1, то рекомендуется [7] определять

$$\alpha_{к1} = (1 - 0{,}01Э_{к1})^{4,3} \text{ [и - обратно: } Э_{к1} = 100(1 - \alpha_{к1}^{0,233})]; \quad (8)$$

если же предпочтительна нулевая гипотеза H_0, то

$$\alpha_{к0} = (0{,}01Э_{к0})^{4,3} \text{ [и - обратно: } Э_{к0} = 100\,\alpha_{к0}^{0,233}]. \quad (9)$$

Формулы (8) и (9) являются аналитической аппроксимацией взаимосвязи величин уровня значимости $\alpha_{кi}$ и экспертных оценок $Э_{кi}$, $i \in (0 \cup 1)$, данных авторитетными экспертами (см. табл. 2) (очевидно, что здесь $0 \neq \alpha_{кi} \neq 1$ и $0 \neq Э_{кi} \neq 100\%$).

Если окажется, что $\alpha_i < \alpha_{кi}$, то проверяемая гипотеза H_0 отклоняется в пользу альтернативы H_1, если же $\alpha_i \geq \alpha_{кi}$, то гипотеза H_0 не отклоняется.

Если эксперт оценивает уровень ответственности за вывод «средним баллом», и принимает $Э_i = 50\%$, то $\alpha_{к0}$ и $\alpha_{к1}$ приблизительно равны «общепринятому» [3 – 6] критическому значению уровня значимости: $\alpha_{к0} \approx \alpha_{к1} \approx 0{,}05$. Однако следует учесть, что при низком уровне ответственности ($Э_i < 50\%$) получаем $\alpha_{к0} < \alpha_{к1}$, а при высоком уровне ответственности ($Э_i > 50\%$) получаем $\alpha_{к0} > \alpha_{к1}$.

Таким образом, решение ЛПР (или исследователя в роли ЛПР) принять или отклонить гипотезу H_0 зависит от того критического уровня $Э_{кi}$ ответственности за вывод, который принимается **им** в данном, конкретном исследовении. Поэтому вместо того, чтобы сравнивать уровни значимости - оценку α_0 (или α_1) и её критическое значение $\alpha_{к0}$ (или $\alpha_{к1}$), проще сравнивать уровни ответственности за вывод – оценку $Э_i$ $[i \in (0 \cup 1)]$ и её критическое значение $Э_{кi}$ ($0\% < Э_{кi} < 100\%$) где значение величины $Э_1$ вычисляется по обратной формуле (8), когда предпочтительна альтернативная гипотеза H_1, а значение величины $Э_0$ - по обратной формуле (9), когда предпочтительна нулевая гипотеза H_0.

Теперь **если окажется, что оценка $Э_i$ уровня ответственности за вывод выше критического значения $Э_{кi}$, т. е., $Э_i \geq Э_{кi}$, то предпочитаемая исследователем гипотеза H_0 (или H_1) принимается, и – наоборот: если $Э_i < Э_{кi}$, то предпочитаемая исследователем гипотеза H_0 (или H_1) не принимается.**

Замечание 1. Э. Леман [5] пишет: «Иногда может оказаться желательным использование значений α_k (превосходящих «обычные»), например, 0,1 или 0,2».

Э. Леман, по-видимому, имел в виду случай, когда исследователь заинтересован продемонстрировать справедливость проверяемой гипотезы H_0 «с высоким уровнем ответственности $Э_0$» (см. табл. 2) или же - когда исследователь заинтересован продемонстрировать справедливость альтернативной гипотезы H_1, но - «с очень малым уровнем ответственности $Э_1$» (см. табл. 2).

Замечание 2. «Общепринятые» [4, 5, 6] критические значения уровней значимости $\alpha_{к0} \approx \alpha_{к1} \approx 0{,}05$ являются анахронизмом, данью устаревшей «традиции». Критические значения уровней ответственности за вывод – оценки $Э_{кi}$, в зависимости от предпочитаемой исследователем гипотезы H_0 (или H_1), рекомендуется задавать в диапазоне ($50\% < Э_{кi} < 100\%$), и как можно ближе к 100%. Тогда, например, по формулам (8) и (9) при $Э_{кi} = 80\%$ получим $\alpha_{к1} \approx 0{,}001$, а $\alpha_{к0} \approx 0{,}4$, а при $Э_{кi} = 90\%$ получим $\alpha_{к1} \approx 5\times10^{-5}$, а $\alpha_{к0} \approx 0{,}6$.

Принимая или отклоняя гипотезу H_0, ЛПР (или исследователь в роли ЛПР) может вообще не задаваться критическими значенияи $\alpha_{кi}$ и $Э_{кi}$, а просто приводить оценки достигнутых критериев α_i и $Э_i$ для того, чтобы уже читатель (в роли ЛПР) мог сам согласиться или не согласиться с выводом авторов. Такая аргументация выводов представляется наиболее приемлемой.

Очевидно, что делая вывод на основании статистических данных, исследователь должен избегать категорических формулировок («достоверно», «доказано» и т. п.), так как опыт не «доказывает», а лишь «не противоречит» принятым гипотезам. Действительно, значение $(1 - \alpha_i)$ характеризует вероятность, с которой не отклоняют гипотезу H_0, если она верна. Меры надежности «принять нулевую гипотезу H_0» не существует. Поэтому об отклоненной гипотезе H_0 говорят, что она отклонена с пренебрежимо малой вероятностью ошибиться ($\alpha_i < \alpha_{iк}$) и с соответствующим уровнем ответственности за вывод, превышающем критический ($Э_i > Э_{кi}$). Если же гипотеза H_0 не отклонена, то говорят, что «данные опытов ей не противоречат на уровне значимости α_i, где $\alpha_i > \alpha_{кi}$», или, иначе, что «найденный уровень значимости α_i слишком велик, больше критического значения ($\alpha_i > \alpha_{iк}$), и его недостаточно для того, чтобы отклонить проверяемую гипотезу H_0».

Отсюда также следует, что «индекс i для обозначения предпочитаемой исследователем гипотезы» H_i $[i \in (0 \cup 1)]$ следует всегда писать как при уровнях значимости α_i и $\alpha_{кi}$, так и при экспертных оценках **уровня ответственности ЛПР за вывод** $Э_i$ и $Э_{кi}$.

В своей исследовательской работе ЛПР не может быть беспристрастным. Если ЛПР, например, создал новый товар лучше старого, то тогда он предпочтёт альтернативную гипотезу H_1 о том, что параметры его нового товара лучше параметров прототипа; если же он создал математическую модель объекта исследования, и она должна быть адекватной описываемому объекту, то ЛПР предпочтёт нулевую гипотезу H_0 о равенстве параметров, рассчитанных с помощью математической модели, параметрам моделируемого объекта. Естественно, ЛПР - оппонент может попытаться опровергнуть достоинства нового товара или сомневаться в адекватности математической модели. В подобных случаях удобным подспорьем разным ЛПР и служит табл. 2, в которой формализованы существующие взгляды на выбор α_k.

Для удобства расчетов на ЭВМ данные табл. 2. аппроксимированы простыми формулами (8) и (9).

6. Вернёмся к формулировкам статистических гипотез

Ошибки - не лженаука. Лженаука - это непризнание ошибок.

Сравнивая две **простые** гипотезы H_A: $\theta = \theta_0$ и H_B: $\theta = \theta_1$ (где θ - некий статистический параметр, а θ_0 и θ_1 – его значения), ЛПР желает подтвердить одну гипотезу (например, H_A) и опровергнуть другую (например, H_B). В этом случае он должен сказать, что проверяет нулевую гипотезу $H_0 = H_B$ против альтернативной гипотезы $H_1 = H_A$. ЛПР - оппонент желает опровергнуть гипотезу H_A и подтвердить H_B. В этом случае он должен сказать, что проверяет нулевую гипотезу $H_0 = H_A$ против альтернативной гипотезы $H_1 = H_B$. Оба ЛПР пользуются графой «предпочтительна альтернативная гипотеза H_1» в табл. 2.

Однако если одна из гипотез (например, H_A: $\theta = \theta_0$) – **простая, а вторая** (например, H_B: $\theta \neq \theta_0$) – **сложная двусторонняя,** то гипотезу H_B нельзя принять в качестве нулевой, и если для ЛПР предпочтительной является гипотеза H_A и её, естественно, нельзя сформулировать в виде альтернативной H_1, то при выборе α_k из табл. 1. **приходится пользоваться графой «предпочтительна нулевая, H_0».**

Наименование предпочтительной гипотезы удобно отмечать в индексе при выбранном из табл. 2. значении критического уровня значимости. Например, уровень $\alpha_{k0} = 0{,}2$ выбран в соответствии с предпочтением нулевой гипотезы (большая ответственность за ошибку); тот же уровень $\alpha_{k1} = 0{,}2$, но предпочтительна H_1 (малая ответственность за ошибку)!

Замечание 3. Табл. 2. можно также использовать для выбора доверительной вероятности $1 - \alpha_k$, необходимой при интервальном оценивании параметров [1]. Статистические оценки параметров распределения СВ должны быть надёжными. Это требование соответствует предпочтению альтернативной гипотезы H_1. **С увеличением**

ответственности ЛПР за вывод интервальные оценки получаются более широкими.

Пример 2. Предполагается продажа большой партии пива. Допустим, некий признак пива 1-го сорта есть H_A: $\theta = \theta_0$, а 2-го сорта - H_B: $\theta = \theta_1$. Цена пива 1-го сорта высокая, а 2-го сорта - низкая. Выбор между гипотезами H_A и H_B осуществляют два ЛПР - продавец (производитель) и покупатель (потребитель) продукции. Необходимо сформулировать гипотезы H_0 и H_1 с позиций продавца и покупателя, а также задать критические уровни значимости α_k.

Рассмотрим два принципа формулировки статистических гипотез.

1) Принцип «отрицания отрицания». Стратегия **покупателя**: ошибочное отклонение H_B (второй сорт) в пользу H_A (первый сорт) принесет больший ущерб, чем ошибочное отклонение H_A (первый сорт) в пользу H_B (второй сорт). Поэтому покупатель принимает $H_0 = H_B$ (тогда $H_1 = H_A$).

2) Принцип «предпочтительности» [1]. **Покупателю** выгодна гипотеза H_A (получать дорогую продукцию 1-го сорта, если она может быть дешевой, 2-го сорта), поэтому он принимает, что $H_1 = H_A$ (тогда $H_0 = H_B$). **Продавцу** выгодна гипотеза H_B (сбывать дешевую продукцию 2-го сорта, если она может быть дорогой, 1-го сорта); поэтому он принимает $H_1 = H_B$ (тогда $H_0 = H_A$).

Пример 3. Задание критического уровня значимости

Покупатель и продавец принимают большой уровень ответственности за вывод (Э = 70%) и из табл. 2. задают $\alpha_k = 0{,}006$. Требуется решить пример 2 при условиях, когда признак пива 1-го сорта H_A: $\theta = \theta_0$, а 2-го сорта - H_B: $\theta \neq \theta_0$.

Решение. Аналогично решению в примере 2, продавец принимает, что $H_1 = H_B$: $\theta \neq \theta_0$; $H_0 = H_A$: $\theta = \theta_0$ и задаёт $\alpha_{k1} = 0{,}006$; покупатель **пытается** сформулировать гипотезы «наоборот»: $H_1 = H_A$: $\theta = \theta_0$, $H_0 = H_B$: $\theta \neq \theta_0$. Однако такая формулировка невозможна, так как противоречит определению нулевой гипотезы H_0 (нулевой - H_0 может быть только гипотеза о равенстве)!

В этом исключительном случае покупателю приходится принять также, как и продавцу, что $H_0 = H_A$: $\theta = \theta_0$ и $H_1 = H_B$: $\theta \neq \theta_0$. Различие же в интересах продавца и покупателя при одном и том же высоком уровне ответственности (Э = 70%) сказывается при выборе критического уровня значимости из Табл. 2: покупатель принимает $\alpha_{k0} = 0{,}2$.

Пример 4. Комиссия тестирует кандидата в эксперты качества напитков по сенсорной чувствительности вкуса с помощью N пар проб. Претендент должен уверенно отличать на вкус воду от слабого водного раствора соли. Это значит, что вероятность π правильного ответа должна быть высокой, например, H_A: $\pi > 0{,}9$, а при неуверенном ответе H_B: $\pi < 0{,}9$. Необходимо сформулировать гипотезы H_0 и H_1 с позиции комиссии.

Решение. Комиссия заинтересована принять кандидата в эксперты, обладающего высокой сенсорной чувствительностью. Поэтому она принимает, что $H_1 = H_A$: $\pi > 0{,}9$ и $H_0 = H_B$: $\pi \leq 0{,}9$.

Заметим, что в решении этого примера к неравенству H_B добавлен знак равенства. Это позволяет выбрать критическую область критерия для сложной правосторонней альтернативы, исходя из равенства $\pi = 0{,}9$.

Таким образом, использование экспертной оценки $Э_i$ (%) уровня ответственности за вывод расширяет возможности ЛПР (или исследователя «в роли» ЛПР) сознательно выбирать значения критического уровня значимости α_{ki}.

7. Расчёт уровня значимости

Все мы - невежды, только - в разных областях. (У. Роджерс)

«Рабочий» диапазон значений уровня значимости α лежит, **в основном**, в пределах $0 < \alpha < 0,5$ (см. табл. 2). Поэтому для **практической** работы удобно ввести следующие **обозначения: u_P - верхний Р - предел СВ u [7, с. 39], когда (см. рис.1)**

$$P(u > u_P \mid H_0) = P\ ;\ 0 < P < 1. \tag{10}$$

Тогда $u_{0,5}$ будет медианой.

Назовём $u_л < u_{0,5}$ - левосторонним (л-) (нижним) Р-пределом, когда

$$P(u < u_л \mid H_0) = \alpha_л \ \text{и}\ \alpha_л < 0,5 \tag{11}$$

и $u_п > u_{0,5}$ - правосторонним (п-) (верхним) Р-пределом, когда

$$P(u > u_п \mid H_0) = \alpha_п \ \text{и}\ \alpha_п < 0,5, \tag{12}$$

где $\alpha_л$ и $\alpha_п$ - вероятности попадания статистики u в области, расположенные слева или справа от пределов $u_л$ и $u_п$ соответственно.

Замечание 4. Когда вместо соотношения (10) используют выражение $P(u < u_P|H_0) = P$ [12, с. 122], то u_P называют «Р-квантилью»; $u_л$ и $u_п$ - соответственно «нижней и верхней Р-квантилью». В наших обозначениях (11) левосторонний (л-) (нижний) Р-предел $\mathbf{u_л}$ есть «нижняя Р-квантиль». Для «верхней (1 - Р)-квантили» справедливо выражение $P(u < u_{1-P} \mid H_0) = 1 - P$.

На это различие обозначений и названий следует обратить внимание читателя, чтобы не было путаницы при употреблений введенного Шеффе термина «верхний Р-предел» - в настоящей работе и «Р-квантиль» - в других источниках!

Замечание 5. Погрешность **задаваемых ЛПР** экспертных оценок «$Э_k$» (см. табл. 2) в области уровней ответственности ($Э_k > 50\%$), обычно не превышает 5% размаха шкалы. Погрешность же соответствующего критического уровня значимости α_k столь велика, что результаты вычислений по формулам (10) - (12) достаточно записывать всего с одной значащей цифрой, например, $\alpha_л = 0,4; 0,04; 0,004; 0,0004; 0,00004$ и т. п.

Пример 5. Вряд ли ЛПР сможет «почувствовать» различие даже в 10 раз двух «достигнутых» в результате вычислений по формулам (10) - (12) оценок $\hat{\alpha}$, равных 10^{-6} и 10^{-7}, так как экспертные оценки «Э» уровней ответственности ЛПР за вывод различаются всего на 1,5 пункта: 96% и 97,5% соответственно.

Рассмотрим в качестве статистики критерия сам уровень значимости α [3]**.** Пусть $\Omega \subseteq W$ - область, на которой оценка уровня значимости

$$\hat{\alpha} = \int_\Omega f(u)du = P(u \in \Omega \mid H_0). \tag{13}$$

В зависимости от выбора альтернативной гипотезы область Ω принимает вид (см. рис. 1): $\Omega = (-\infty, \hat{u}]$ при л-альтернативе (4), в случае, когда $\theta_1 < \theta_0$ для гипотез (7); $\Omega = [\hat{u}, \infty)$ при п-альтернативе (4), в случае, когда $\theta_1 > \theta_0$ для гипотез (6);

$\Omega = (-\infty, \hat{u}_л] \cup [\hat{u}_п, \infty)$ (где Y-символ дизъюнкции «или») при д-альтернативе (5), где $u_л$ и $u_п$ таковы, что $\alpha_л = P(u < \hat{u}_л) = P(u > \hat{u}_п) = \alpha_п$. Отсюда следует, что $\hat{\alpha}$ является монотонной функцией $\hat{u}$ (рис. 2) и поэтому также может служить проверочной статистикой критерия.

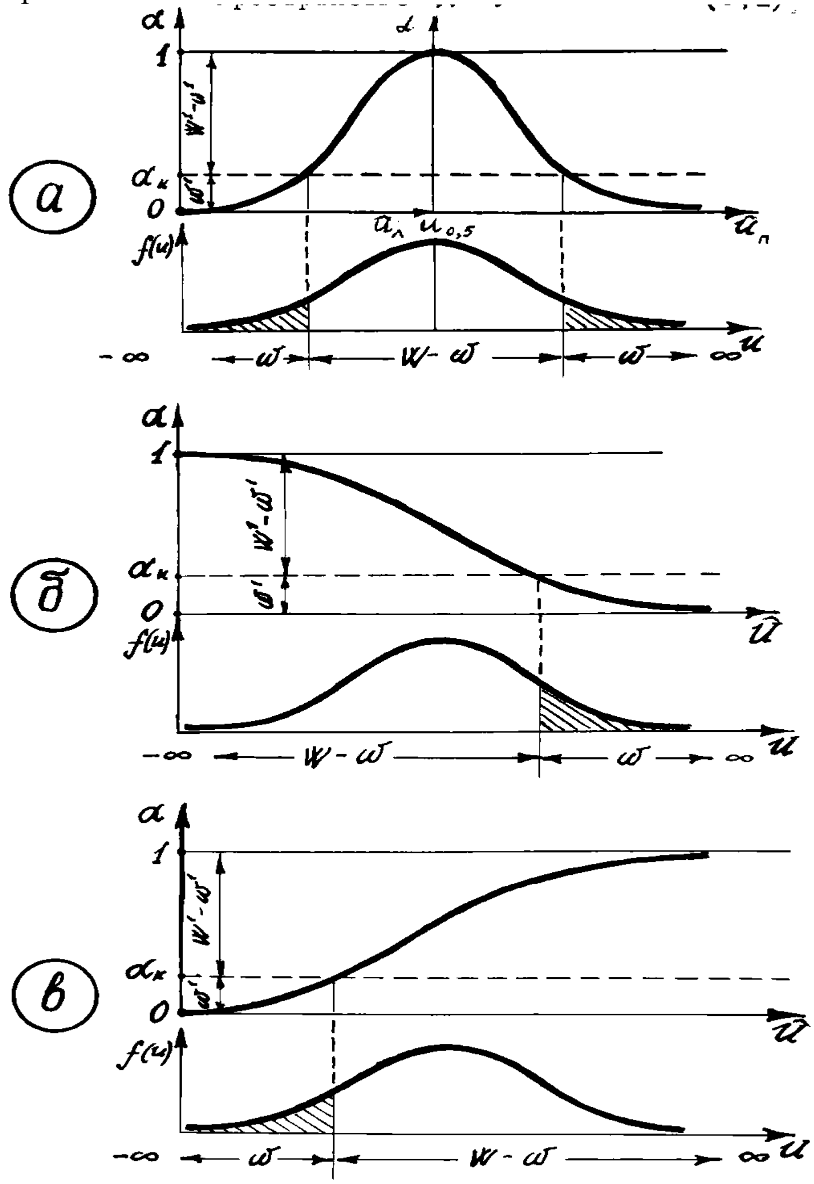

Рис. 2. Отображение критической области ω статистики u-критерия из пространства w в пространство $W' = \{\alpha : 0 < \alpha < 1\}$:

а – при двусторонней альтернативе H_1; б – при правосторонней; в – при левосторонней; ω, ω' - критические области u- и α-критериев; $W - \omega$, $W' - \omega'$ - допустимые области u- и α-критериев; $f(u)$ - плотность распределения проверочной статистики критерия u; α_k - критическое значение уровня значимости.

Пространство W' всех возможных значений α есть

$$W' = \left\{\alpha: \quad 0 < \alpha < 1\right\}; \tag{14}$$

критическая ω' и допустимая $W' - \omega'$ области имеют вид

$$\omega' = \{\alpha \in W' : 0 < \alpha \le \alpha_k\}; \tag{15}$$

$$W' - \omega' = \{\alpha \in W' : \alpha_k < \alpha < 1\}, \tag{16}$$

где α_k - заданный ЛПР критический уровень значимости.

Формулу (11) теперь можно записать следующим образом:

$$\hat{\alpha} = \begin{cases} n\hat{\alpha}_л;\ \hat{\alpha}_л = p(u < \hat{u}_л \mid H_0), \text{если } \hat{u} = \hat{u}_л; \\ n\hat{\alpha}_п;\ \hat{\alpha}_п = p(u > \hat{u}_п \mid H_0), \text{если } \hat{u} = \hat{u}_п, \end{cases} \tag{17}$$

где коэффициент n = 1 для односторонней (л- или п-) альтернативы; n = 2 – для двусторонней (д-) альтернативы.

Действительно, согласно формуле (15) при двусторонней (а), правосторонней (б) и левосторонней (в) альтернативам H_1 имеем:

- формулу **двустороннего** (д-) критерия для проверки гипотезы H_0 (5) (см. рис. 2 а): $\hat{\alpha} = 2\hat{\alpha}_л$, если $\hat{u} = \hat{u}_л$, или $\hat{\alpha} = 2\hat{\alpha}_п$, если $\hat{u} = \hat{u}_п$;

- для проверки гипотез H_0 (4) при $\theta_1 > \theta_0$ и (6) получаем формулу **правостороннего** (п-) критерия (см. рис. 2 б): $\hat{\alpha} = \hat{\alpha}_п$;

- формулу **левостороннего** (л-) критерия для проверки гипотез (4) при $\theta_1 < \theta_0$ и (7) (см. рис. 2 в): $\hat{\alpha} = \hat{\alpha}_л$.

Поскольку $\hat{\alpha}$ - вероятность ошибочного отклонения гипотезы H_0, если она верна и $\hat{\alpha} > 0$, то условие отклонения H_0 (при попадании в критическую область ω') имеет вид

$$\hat{\alpha} < \alpha_k. \tag{18}$$

Интерпретация α-критерия: в каждых ста выборках объема N элементов в среднем только $100\hat{\alpha}$ раз будет отклоняться H_0, если она верна [1].

Замечание 6. Зоотехники [6], медики и биологи [9] (в отличии, например, от представителей технических наук [14, 17]) по традиции, принятой в **биометрике** (математической статистике, адаптированной к решению задач биологических наук), приводят оценку уровня значимости (в виде неравенства, обозначая оценку буквой Р, а не α, как это принято в математико-статистической литературе) после каждого утверждения о справедливости альтернативной гипотезы.

Используя «альфа-метод», ЛПР может дать более точную, чем неравенство, оценку уровня значимости. Например, вместо грубого неравенства $\hat{\alpha} < 0{,}001$ записывать более определённый результат - равенство $\hat{\alpha} = 0{,}000007$ (есть разница?!).

Замечание 7. Описание большинства методов проверки гипотез относительно параметров нормального распределения СВ приведены в работе [1]. Однако современный ЛПР может теперь особенно не заботиться о конкретных методах и формулах расчёта оценок $\hat{\alpha}$ для решения тех или иных конкретных задач проверки гипотез, потому что существующие ППП охватывают большинство необходимых методов проверки гипотез, и все они оканчиваются расчётом «достигнутых» оценок $\hat{\alpha}_i$. ЛПР остаётся лишь с помощью табл. 2 или формул (8) и (9) сравнить соответствующие уровни ответственности за выводы $Э_i$ с априори заданными $Э_{ki}$ и сделать вывод относительно проверяемой гипотезы H_{0i}.

8. Мощность критерия

Проверку статистических гипотез всегда необходимо доводить до конца, чтобы узнать, с чего её нужно было начинать.

При проверке статистических гипотез бывает недостаточно оценить вероятность $\hat{\alpha}_i$ ошибки первого рода (1). Следует также определить мощность 1 - β (3) критерия (или вероятность β (2) ошибки второго рода) при фиксированном уровне значимости α_{ik}. Для этого необходимо задать значение параметра нецентральности θ_λ [19] - величину отклонения истинного значения параметра θ от проверяемого θ_0.

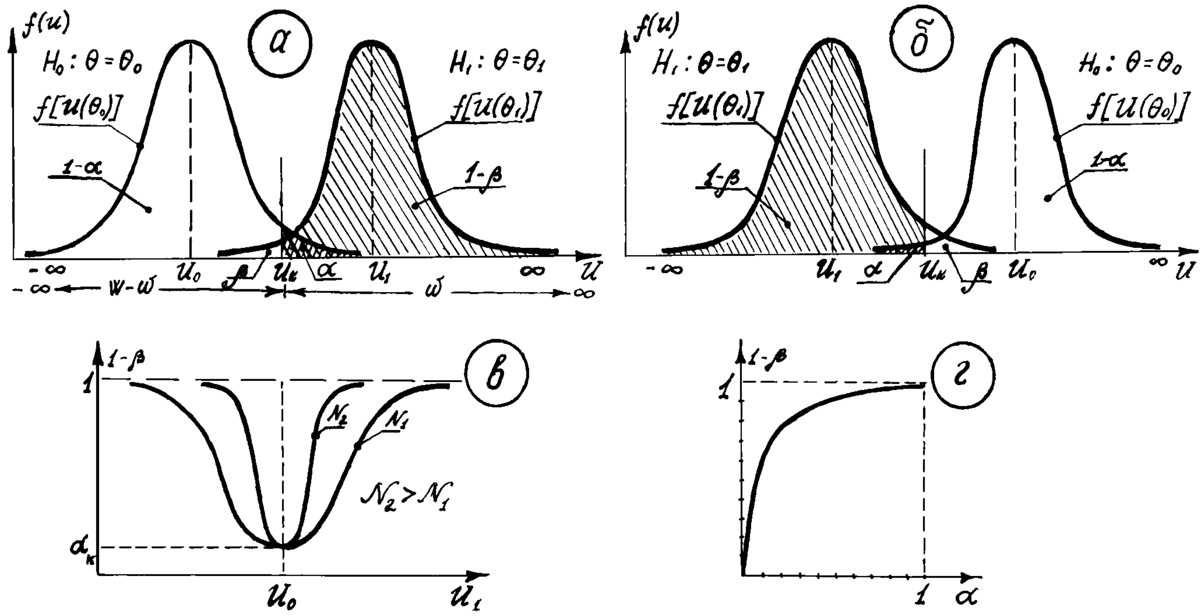

Рис. 3. Мощность критерия (1-β) в случае простой право- (а) и левосторонней (б) альтернативы:
в - кривые функции мощности критерия для различных условий при двусторонней альтернативе; г - зависимость мощности критерия 1- β от уровня значимости α.

Пусть, например, $\theta = \theta_1 = \theta_0 + \theta_\lambda$ и, следовательно, нулевая гипотеза H_0: $\theta = \theta_0$ неверна. Критическая область строится, как обычно, на основе плотности вероятности центрального распределения $f(u) = f[u, u(\theta_0)]$, группирующегося вокруг точки $u_0 = u(\theta_0)$, в предположении, что гипотеза H_0 справедлива. На самом же деле СВ u будет иметь иную, нецентральную плотность распределения вероятностей $f_1(u) = f_1[u, u(\theta_1)]$, группирующуюся вокруг точки $u(\theta_1) = u(\theta_0 + \theta_\lambda) = u_1$ (см. рис. 3, поз. а, б). Вероятность ошибки второго рода β при попадании статистики $\hat{\alpha}$ в допустимую область $w' - \omega'$ характеризуется функцией $f_1(u)$ и при выбранном значении α_k зависит от θ_λ. Если $\theta_\lambda \to \infty$, то $\beta \to 0$; если $\theta_\lambda \to 0$, то $\beta \to 1 - \alpha$ [20].

Зависимость 1 - β от θ_λ, согласно определению (3), характеризуется функцией мощности

$$1-\beta = \int_\omega f_1(u)\,du = 1 - P(u_{kл} < u < u_{kп} \mid H_1) = P(u < u_{kл} \mid H_1) + P(u > u_{kп} \mid H_1) \qquad (19)$$

(здесь $u_{kл}$, $u_{kп}$ - критические значения левого и правого α-пределов u-распределения) и числом сторон альтернативной гипотезы (см. рис. 3, в).

Вероятность 1 - β зависит также от объёма выборки N: если $N \to \infty$, то при заданных $\alpha_k > 0$ и $\theta_\lambda > 0$ значение $\beta \to 0$ [1] (см. рис. 3, в).

Зависимость 1 - β от α (см. рис. 3, г) называется «оперативной характеристикой критерия»; если $\alpha \to 0$, то при заданных $\theta_\lambda > 0$ и N значение $(1 - \beta) \to 0$.

Если функция плотности нецентрального распределения $f_1(u)$ может быть получена простым смещением функции центрального распределения $f_0(u)$ на $u_\lambda = u_1 - u_0$ (см. рис. 3), то при заданном α_k уравнение (8) можно привести к виду

$$1-\beta = \begin{cases} P\left(u > u_{kn} - u_\lambda \mid H_0\right) \ при \ u\left(\theta_1\right) > u_{0,5}\left(\theta_0\right), \quad \theta_\lambda > 0; \\ P\left(u < u_{kл} - u_\lambda \mid H_0\right) \ при \ u\left(\theta_1\right) < u_{0,5}\left(\theta_0\right), \quad \theta_\lambda < 0. \end{cases} \qquad (20)$$

9. Спорный случай

Не понял сам, - дай понять другому.

Результаты проверки гипотез зависят не только от природы изучаемого явления, результатов эксперимента, математической методики проверки гипотез, но и от субъективных для ЛПР факторов, связанных с его выбором предпочтительной гипотезы и заданного им критического уровня значимости.

Поэтому нередко результаты одного и того же опыта могут привести разных ЛПР к противоположным выводам. Рассмотрим эту ситуацию подробнее.

В зависимости от ошибочного решения в пользу H_1 или H_0 обычно терпят ущерб различные заинтересованные стороны, назовём их А и В. Такими сторонами могут быть, например, поставщик и потребитель товара [18], институт-разработчик новшества и предприятие или банк, финансирующие внедрение новшества, истец и ответчик (в судебной практике) и т. п. В такой обстановке вполне может возникнуть и **спорный случай**.

Пример 6. Две стороны А и В проверяют пары гипотез вида (4), (6) и (7). Для определённости рассмотрим гипотезы $(\theta > \theta_0)$ и $(\theta < \theta_0)$. Если сторона А сформулирует одну из этих гипотез в виде (6) (см. рис. 3, а), то сторона В сформулирует их в виде (7) (см. рис. 3, б). Стороны А и В выберут свои критические значения α_{kA} и $\alpha_{kБ}$ (которые могут и совпадать - это не важно). Важно то, что значения α_{kA} и $\alpha_{kБ}$ делят пространство W проверочной статистики u критерия на три части: область принятия сторонами А и В гипотезы $(\theta < \theta_0)$, когда $u < u_A$, область принятия сторонами А и В гипотезы $(\theta > \theta_0)$, когда $u \geq u_B$ и «спорную область» $u_A \leq u < u_B$ (где u_A и u_B - соответствующие критические значения u-критериев, определяемые сторонами А и В).

Пример 7. Две стороны А и В проверяют гипотезы вида (5) H_0: $\theta = \theta_0$ и H_1: $\theta \neq \theta_0$. Для стороны А предпочтительна гипотеза H_1, для стороны В - H_0 (напомним: сторона В в этом «исключительном» случае не может сформулировать гипотезу $\theta \neq \theta_0$ как нулевую). Сторона А с большим уровнем ответственности за выводы задаст уровень значимости $\alpha_{k1} = 0{,}006$ (см. табл. 2), найдет, скажем, $\hat{\alpha} = 0{,}02 > \alpha_{k1}$ и примет решение в пользу H_0. Проверяя эти же гипотезы, сторона В, руководствуясь столь же высокой ответственностью за вывод, задаст другой критический уровень значимости $\alpha_{k0} = 0{,}2$ (см. табл. 2). Поскольку теперь $\hat{\alpha} = 0{,}02$ будет меньше α_{k0}, сторона В примет решение в пользу гипотезы H_1: $\theta = \theta_0$. Такой случай также является спорным.

Для решения спорного случая (когда это важно) привлекают третью сторону - С. Это, например, арбитр или судья. В спорных случаях сторона С неформально, с учётом всех обстоятельств «дела» выбирает одно из трех решений - принять H_0 или H_1 или продолжить исследование. В последнем случае, если сравниваются простые гипотезы (4), критические значения α_k и β_k могут быть выбраны сторонами А и В из табл. 2 и использованы для планирования объёма будущей выборки N_δ [1].

К. Доерфель [23] предлагает искусственно ввести в рассмотрение область «спорного случая» в окрестности значения «традиционного» критического уровня значимости $\alpha_k = 0{,}05$. Сопоставление предложения Доерфеля с требованием описанного выше п. 4. порядка действий для проверки гипотез позволяет представить ***область спорного случая «по Доерфелю»*** следующим образом.

Пространство w всех возможных значений α делится на критическую $v_k = \left\{0 < \alpha \le \alpha_k^-\right\}$, «спорную» $v_c = \left\{\alpha_k^- < \alpha \le \alpha_k^+\right\}$ и допустимую $v_д = \left\{\alpha_k^+ < \alpha < 1\right\}$ области. Спорной областью v_c может быть один из интервалов для уровня значимости α_k в табл. 2 (причём, совсем не обязательно, чтобы по устаревшей «традиции» $\alpha_k = 0{,}05$). При попадании статистики критерия $\hat{\alpha}$ в «спорную» область v_c никакого решения относительно проверяемой гипотезы H_0 не принимается.

На наш взгляд, название области v_c «спорной» неудачно, так как автор не указал субъектов «спора». В нашем примере такие субъекты были указаны - это стороны А и В, а «спорные области» определены совсем иначе. Лучше, по-видимому, тогда уж рассматривать зону v_c как «область неопределенности по Доерфелю».

Особенностью подхода Доерфеля является акцентирование внимания на нечёткости (размытости) выбираемой исследователем границы между критической $v_к$ и допустимой $v_д$ областями.

К сожалению, теория проверки статистических гипотез не касается «рецептов» выбора критического значения уровня значимости α_{ik}. Это уже - задача экспертного оценивания [1, с. 730].

Из приведенных рассуждений следует, что в решении проблем проверки гипотез имеется некоторая доля субъективизма ЛПР по принципу: «***если нельзя, но очень хочется, то можно***». Пусть, например, ЛПР стороны А предпочтет некоторую гипотезу H_1, выберет некоторый уровень значимости α_{ik} и получит оценку $\hat{\alpha}_i$, ненамного превышающую α_{ik}. Возможно, «с меньшей ответственностью», вновь выбрав другое значение α_{ik} из табл. 2, большее, чем $\hat{\alpha}_i$, ЛПР стороны А все же примет «предпочтительную гипотезу» H_1.

Однако, другой стороне (В) часто не так важно мнение стороны А, как та аргументация, на которой основано это мнение (например, что найденного стороной А в исследовании уровня значимости $\hat{\alpha}_i$ не достаточно для того, чтобы отклонить H_0, так как, по мнению ЛПР стороны В, $\hat{\alpha}_i > \alpha_{ik}$). Важнейшим из аргументов здесь является оценка $\hat{\alpha}_i$ уровня значимости критерия. Естественно, что автор работы (сторона А) в своём отчёте (научной публикации) должен представить материалы, позволяющие в случае необходимости стороне В перепроверить корректность расчёта уровня значимости $\hat{\alpha}$, оценить мощность (или функцию мощности) критерия и объём будущей выборки.

О «надёжности» $1-\hat{\beta}$ принятой альтернативной гипотезы H_1 исследователи обычно совсем не говорят, потому что ни её, ни функцию мощности не оценивают (то ли - по неведению, то ли из-за сложности вычислений).

Отметим также, что некоторые исследователи неверно называют вероятность (1 - $\hat{\alpha}_i$) «надёжностью нулевой гипотезы H_0» (чем эта вероятность больше, тем гипотеза H_0 «надёжней»), хотя на самом деле всё - с точностью наоборот: чем вероятность (1 - $\hat{\alpha}_i$) больше (ближе к 1), тем точечная оценка $\hat{\alpha}_i$ уровня значимости меньше, и, следовательно, гипотеза H_0 - менее надёжная!

10. Действия исследователя, занятого проверкой параметрических гипотез [1]

Если ничто другое не помогает, прочтите, наконец, инструкцию! (Аксиома Кана)
Чем тщательнее проработан план, тем глупее причина, по которой он рухнет.

Опыт ряда авторов [18, с. 199], [19, с. 33], [21], [22, с. 246] позволяет рекомендовать исследователю следующую **последовательность действий, предпринимаемых для проверки параметрических гипотез**.

1) Формулируется научная гипотеза.

2) Осуществляется математико-статистическая постановка задачи.

2.1) Обобщается априорная информация о параметрах распределения величин, обосновываются или постулируются законы распределения величин.

2.2) Выбирается статистическая характеристика гипотезы u закон распределения которой $f(u)$ известен.

3) Путём содержательного анализа предпочтений ЛПР и предполагаемых рисков противоположных гипотез (см. п. 4), формулируются две статистические гипотезы - нулевая H_0 и альтернативная H_1.

4) Задаётся критический уровень значимости α_{ki} в зависимости от уровня ответственности ЛПР $Э_i$ за выводы и наименования предпочтительной гипотезы ($i = 0$ – предпочтительна H_0; $i = 1$ – предпочтительна H_1).

5) Выполняются действия (эксперименты), имеющие целью получить представительную (рандомизированную и достаточно объёмную) выборку.

6) Вычисляется оценка $\hat{\alpha}_i$ «альфа-критерия».

6.1) По данным выборки рассчитываются точечные оценки параметров распределения величин.

6.2) В зависимости от сформулированных гипотез определяются правая $u_п$ или левая $u_л$ оценки статистических характеристик критерия.

6.3) Вычисляется, наконец, точечная оценка $\hat{\alpha}_i$ критерия.

7) Отклоняется или не отклоняется нулевая гипотеза H_0.

7.1) Если $\alpha_i < \alpha_{ki}$, то гипотеза H_0 отклоняется и принимается H_1.

7.2) Если $\alpha_i \geq \alpha_{ki}$, то гипотеза H_0 не отклоняется, а H_1 отклоняется.

8) Определяется (при необходимости) мощность критерия $1 - \beta$.

8.1) Если проверяются простые или сложные гипотезы, то вычисляется значение $1 - \beta$ или строится функция мощности критерия соответственно.

8.2) Иногда рассчитывается оперативная характеристика критерия [1].

9) Для планируемых в будущем исследований определяется объём выборки $N_б$, обеспечивающий заданные уровень значимости α_k и мощность критерия $1 - \beta_k$.

Прежде всего, необходимо констатировать факт, что практически во всех известных нам исследованиях **описанный выше порядок действий не соблюдался**! Статистические гипотезы формулировались, как правило, не до получения данных, а - после. Использовался приём «разведочного анализа» [17, с. 214]: данные наносили на графики, рисовали кривые и определяли, какие гипотезы можно сформулировать. Затем формулировали гипотезы и проверяли их. Мощность статистических критериев и, тем более, функцию мощности, никто никогда не оценивал.

Выход из создавшегося положения видится в том, чтобы после завершения такого «некорректного» цикла повторить экспериментальное исследование в порядке действий, предписанном статистической теорией.

11. Параметрический и численный методы проверки гипотез

Чем дальше вы убежите от проблем, тем дольше вам придётся возвращаться назад, чтобы их решить.

Статистические **оценки параметров распределения СВ** получают в результате следующих действий [1].

Постулируется или обосновывается некоторый теоретический закон распределения рассматриваемой **СВ** X с теоретической (генеральной) функцией распределения вероятностей $F_T(\varphi)$. Путем специально поставленных экспериментов получают случайную представительную выборку $\{x_i\}$ ($i = \overline{1,N}$) из генерального (обычно-бесконечного) множества значений СВ X. Подбирается нужная статистика $\varphi\{x_i\}$, закон распределения которой заранее хорошо изучен (то есть, существуют таблицы, номограммы или аналитические аппроксимации функции $P = F_1(\varphi)$ распределения статистики).

Посредством масштабирования функции $F_1(\varphi)$ получают функцию распределения искомой статистики $\hat{\varphi}$. Поскольку функция эта довольно сложная, для отчёта о проделанной работе её принято определенным образом табулировать. Обычно для этого задают две вероятности – точную 0,5 и большую доверительную вероятность В (1 > В > 0,5). Решая три уравнения $0{,}5 = F(\hat{\varphi})$, $(1\text{-}B)/2 = F(\hat{\varphi}_n)$ и $(1\text{+}B)/2 = F(\hat{\varphi}_b)$ относительно $\hat{\varphi}$, вычисляют точечную, или однозначную оценку $\hat{\varphi}$, нижнюю $\hat{\varphi}_n$, и верхнюю $\hat{\varphi}_b$ границы **доверительного интервала (ДИ**), такого, чтобы

$$P(\hat{\varphi}_n \le \varphi \le \hat{\varphi}_b) = B.$$

Из многих недостатков приложения этой методики (потеря информации в отчёте о функции распределения статистики $\varphi(x_i)$, несоблюдение постулата о случайности выборки [2, 10] и др.) рассмотрим один устранимый недостаток. Это - сомнительность **гипотезы** (или постулата) о том, что выборка $\{x_i\}$ получена из генеральной совокупности с заранее известным (с точностью до параметров) законом распределения $F_T(X)$.

Мощности статистических критериев (особенно при малых выборках) обычно не хватает для того, чтобы отклонить эту **гипотезу**. Поэтому в большинстве практических задач это просто - постулат. Если же объём выборки достаточно велик и гипотезу эту отклоняют, возникает трудная задача подбора нормализующего преобразования или другой заранее изученной теоретической функции $F_T'(X)$ [1, раздел 2].

Простой выход из сложившейся ситуации предложил Эфрон [11]. Необходимо строить по выборке **эмпирическую функцию распределения СВ (ЭФР СВ),** $F_Э\{x_i\}$, считать её «**теоретической»**, а функцию $F(\varphi)$ распределения статистики $\varphi(X_i)$ строить (и табулировать) численным методом статистических испытаний (Монте-Карло) с помощью ЭВМ. Метод получил название «**бутстреп**». Становятся ненужными многие статистические таблицы; бутстреп-**метод приобретает черты универсальности**: построение функции распределения статистики φ, традиционное интервальное оценивание параметров и проверка статистических гипотез выполняются одними и теми же приёмами, требующими, правда, применения ЭВМ.

В качестве аппроксимации теоретической функции распределения Эфрон по традиции использовал довольно грубую кусочно-постоянную функцию (ступенчатую, имеющую разрывы непрерывности первого рода).

Пусть, согласно идее Фишера, генеральная функция распределения $F_T(X)$ существует. Можно предположить, что эта функция - непрерывная, такая, что для её описания можно использовать менее грубую, чем кусочно-постоянная и более удобную для расчётов непрерывную функцию, например, кусочно-линейную (особенно - для малых выборок) или непрерывно дифференцируемую (особенно - для больших выборок) [2, 12-15].

В работе [12] иллюстрируются, по крайней мере, два преимущества кусочно-линейной аппроксимации эмпирической функции распределения $F_Э\{x_i\}$ перед кусочно-постоянной аппроксимацией, особенно при малом [16] объёме N выборки (до 10 элементов):

1) возможность имитации с помощью кусочно-линейной аппроксимации заданного количества N **любых значений** измеряемой СВ X (а - не только наблюдаемых в эксперименте, обеспечиваемое кусочно-постоянной аппроксимацией);

2) если считать кусочно-линейную аппроксимацию «теоретической», то она ближе, а при $N \to \infty$ ещё и быстрей [13] сходится к генеральному распределению F(x) СВ X, чем кусочно-постоянная аппроксимация.

Таким образом, мы видим два разных метода проверки гипотез, основанных на расчёте статистической оценки уровня значимости: **традиционный аналитический метод** в предположении о «теоретическом» нормальном законе распределения СВ и численный бутсреп-метод Эфрона, который базируется на аппроксимации наблюдаемой в каждом эксперименте эмпирической функции распределения $F_Э\{x\}$ СВ.

Отметим следующее отличие бутсреп-метода проверки гипотез от аналитического метода. Теоретически область определения нормального закона распределения СВ – вся числовая ось. Это обстоятельство даёт возможность при проверке параметрических гипотез получать уровни значимости α сколь угодно малыми (например, $\alpha = 10^{-10}$), но отличными от 0. В то же время, кусочно-линейная аппроксимация эмпирической функции распределения $F_Э\{x\}$ СВ является финитной функцией, ограниченной некоторым конечным интервалом области определения. Следовательно, получать уровни значимости α сколь угодно малые, но отличные от 0 можно не всегда: иногда будем получать и значения $\alpha = 0$.

Однако, если вместо кусочно-линейной аппроксимации $F_Э\{x\}$ СВ использовать адекватную аппроксимацию в виде взвешенной суммы гауссиан, например,

$$F_Э\{x\} = \sum_{i=1}^{c} q_i F_i(\nu_i, \sigma_i, x); \qquad \sum_{i=1}^{c} q_i, = 1; \quad q_i > 0, \tag{21}$$

где с – количество слагаемых, q_i – весовые коэффициенты; ν_i, σ_i – коэффициенты, имеющие смысл центров и стандартных отклонений гауссиан $F_i(\nu_i, \sigma_i, x)$. Тогда вполне возможно получать и довольно малые значения α, отличные от 0 (при $\alpha > 0$) и численным бутсреп-методом.

Гипотезу об адекватности аппроксимации (21) можно проверять любым известным непараметрическим методом [1, 14]. Важно, что часто при малом объёме выборки эта аппроксимация оказывается адекватной данным уже на первом шаге (при количестве слагаемых $c = 1$). Это позволяет немедленно переходить к хорошо освоенному и обеспеченному ППП параметрическому методу проверки гипотез.

12. «Порочная спираль» в теории проверки гипотез

Вы когда-нибудь видели незамкнутый круг? (У. Черчилль)

Рассмотрим условие $\alpha_i < \alpha_{кi}$ отклонения нулевой гипотезы H_0. Оценка α_i является СВ, потому что определяется через другие СВ; величина $\alpha_{кi}$, вообще говоря, также является СВ, потому что определяется как обобщённая экспертная оценка [1, с. 730], вычисляемая, в свою очередь, как некое осреднение экспертных оценок отдельных экспертов. Получается так, что условие $\alpha_i < \alpha_{кi}$ также следует рассматривать как новую гипотезу, например, в виде $M\{\alpha_i\} < M\{\alpha_{кi}\}$ против другой гипотезы $M\{\alpha_i\} > M\{\alpha_{кi}\}$, где

M{●} - оператор «математическое ожидание». Если учесть, что законы распределения величин α_i и $\alpha_{кi}$ явно - не нормальные, то проверку последней пары гипотез, казалось бы, можно осуществлять численным методом (см., например, [1, 11, 12]). Однако в этом случае нам пришлось бы определять «новые» уровни значимости $\alpha^1{}_i$ и $\alpha^1{}_{кi}$, а для их сравнения опять формулировать подобную пару гипотез и т. д. Таким образом, процедура проверки гипотез может войти в «порочную спираль», разрыв которой на первом же «витке», по-видимому, и лежит в сравнении только точечных оценок α_i и $\alpha_{кi}$. **Возможно,** придётся решать эту задачу за две итерации: следующая пара уровней значимости $\alpha^1{}_i$ и $\alpha^1{}_{кi}$ будет менее «важна» для принятия решения, и на этом итерации «порочной спирали» можно закончить. Эта идея требует дальнейшего исследования.

Решая проблему выбора критического уровня значимости (по умолчанию - в предпочнтении альтернативной гипотезы H_1), автор работы [9] предлагает рассматривать не только точечную оценку α_i уровня значимости альфа-критерия, но и всё его распределение как СВ, и, в частности, верхний альфа-предел $\alpha^1{}_i$ правостороннего доверительного интервала, или, так сказать, «оценку уровня значимости $\alpha^1{}_i$ уровня значимости α_i альфа-критерия». На примере (объём выборки N = 10), при критическом значении точечной оценки, равном $\alpha_{iк} = 0{,}05$, верхний альфа-предел $\alpha^1{}_i$ оценки уровня значимости альфа-критерия α_i оказался столь великим ($\alpha^1{}_i = 0{,}3$), что признать «значимость» такого критерия не представлялось возможным.

Действительно, большое значение величины $\alpha^1{}_i$ ($\alpha^1{}_i = 0{,}3$) получено по малому объёму N выборки (N = 10). С ростом объёма N выборки верхний альфа-предел $\alpha^1{}_i$ правостороннего доверительного интервала должен приближаться к точечному значению уровня значимости α_i. Например, при N = 20 получим $\alpha^1{}_i = 0{,}2$, а при N = 50 получим $\alpha^1{}_i = 0{,}06$ (более точные числа могут быть получены – либо теоретически, либо – численным бутстреп-методом).

С другой стороны, если же ЛПР предпочтёт нулевую гипотезу H_0 (индекс i = 0), то его уже будет интересовать не верхний альфа-предел $\alpha^н{}_i$, а другая «оценка уровня значимости $\alpha^н{}_i$ уровня значимости α_i альфа-критерия», то есть, нижний альфа-предел $\alpha_i{}^н$ левостороннего доверительного интервала уровня значимости α_i. Тогда значение величины $\alpha^н{}_i$, полученное по малому объёму N выборки (N = 10) будет слишком мало ($\alpha^н{}_i = 0{,}004$) для принятия гипотезы H_0. С ростом объёма N выборки нижний альфа-предел $\alpha^н{}_i$ левостороннего доверительного интервала должен приближаться к точечному значению уровня значимости α_i «снизу». Например, при N = 20 получим $\alpha^н{}_i = 0{,}02$, а при N = 50 получим $\alpha^н{}_i = 0{,}08$ (более точные числа также могут быть получены – либо теоретически, либо – численным бутстреп-методом).

Поэтому автор статьи [9] предлагает не спешить с выводом, а «для большей надёжности» ужесточить точечную оценку критического значения $\alpha_{iк}$ альфа-критерия до меньшего значения, например, - до $\alpha_{iк} = 0{,}001$. И вот тогда-то $\alpha^1{}_i$ и станет приемлемым: ($\alpha^1{}_i = 0{,}02$ при N = 10). Правда, и в этом случае «уровень значимости $\alpha^2{}_i$ уровня значимости $\alpha^1{}_i$ уровня значимости α_i альфа-критерия» также будет СВ и, следовательно, верхний альфа-предел оценки $\alpha^2{}_i$ получится опять-таки, довольно большим…

По-видимому, продолжать раскручивать эту «порочную спираль» нет смысла. Придётся ЛПР, руководствуясь высоким уровнем ответственности за вывод $Э_i$, большим 50% (см. табл. 2), останавливаться на «первом витке спирали», оставаясь в рамках «классического» сравнения только точечной оценки α-критерия α_1 с априори заданным ЛПР критическим значением $\alpha_{1к}$, выбирая его достаточно малым - не более 0,001, когда предпочтительна альтернативная гипотеза H_1 или - в рамках «классического» сравнения только точечной оценки α-критерия α_0 с априори заданным ЛПР критическим значением $\alpha_{0к}$, выбирая его достаточно большим, более 0,2 (как, кстати, рекомендовал Э. Леман), когда предпочтительна нулевая гипотеза H_0.

13. Ещё некоторые замечания

Если ты - умный, то поймешь, что ничего не понимаешь. (М. Жванецкий)

Замечание 8. Пара слов – о «Р-значении”. Термины «Р-значение» или «Р-уровень» ввёл в биометрику К. А. Браунли (K. A. Brownlee, 1960) [4]. Эти термины кажутся неудачным, потому что они - синонимы давно (до 1960 г.) известного термина «уровень значимости». А после 1960 г. термин «Р-значение» стал широко использоваться прикладниками (в основном, биологами, медиками и экологами) наряду с другими неудачными терминами – «достоверная корреляция при Р < 0,05», «доказанная разность при Р < 0,05» и т. п. [9]. Принято говорить не о «достоверности», а об экспериментально обнаруженной закономерности.

Замечание 9. В ряде работ в последние годы используется термин «достигнутый уровень значимости», и пишут, например, «р = 0,0757» или «р = 0,000».

Слово «достигнутый» - лишнее. По этой логике тогда надо везде писать: «достигнутое среднее», или «достигнутая оценка дисперсии» и т. п. Лучше просто писать «получена точечная оценка уровня значимости…».

Замечание 10. Имеет место разнобой между принятым в математической статистике обозначением уровня значимости «α» и «компьютерным» - «р» (потому что на большинстве алфавитно-цифровых печатающих устройсты старых ЭВМ 50-х годов прошлого века греческая буква «α» отсутствовала). В отчётах следует писать обозначение «α», а - не «р».

Замечание 11. С выбранным критическим значением $\alpha_{iк}$ «достигнутый уровень значимости» α_i сравнивают редко. В отчёте следует приводить критическое значение $\alpha_{iк}$, а также - как именно, из каких соображений оно выбрано.

Замечание 12. Не ясно, зачем используется большая точность в записи числа α_i (например, «α_i = 0,0757» - с тремя значащими цифрами)!? Где используется «хвост» этого числа - 0,0057? Достаточно писать округлённо - до одной значащей цифры (например, α_i = 0,08). Ведь используется это число только лишь для сравнения его с выбранным ЛПР (зачастую, предвзято) критическим значением $\alpha_{iк}$. При этом отброшенный «хвост» (здесь - 0,0057) нигде не используется.

Замечание 13. Значение «α = 0,000» часто приводят, копируя результат записи числа на дисплее ЭВМ с малым числом нулей после запятой. Это - неправильно. Не бывает «α = 0,000». Обычно $\alpha > 0$, и поэтому это число следует писать точней, с одной значащей цифрой, например, вместо «α = 0,000» писать «$\alpha = 2\times10^{-8}$».

Замечание 14. Часто статистические обобщения делаются неправильно - только с помощью средних величин. Поясним это на простом примере. Допустим, в фирме работает n = 99 человек и все получают одинаковую зарплату У (пусть для ровного счёта У = 1000 евро в месяц), а один из них - директор - получает в сто раз больше, то есть 100У = 100000 евро. Тогда на одного человека приходится зарплаты «в среднем» (nУ + 100У)/100 = 1,99У = 1990 евро. Получается, что работники фирмы получают «в среднем» не так уж и плохо! Как можно интерпретировать такую статистику? А - никак.

Давайте посмотрим, что пишут в средствах массовой информации о продолжительности жизни населения разных стран: «в Германии средняя продолжительность жизни 78,2 года, а в Японии – 81,5». Верно ли теперь, что мы получили представление о соотношении продолжительности жизни населения в упомянутых странах? Неверно! Может, просто в Японии больше очень старых людей или смертность новорожденных меньше, чем в Германии!? Из средних значений это не следует.

Очевидно, что подобные «обобщения» результатов наблюдений приводит к поверхностным суждениям. Для науки и практики (и, даже, - для просвящённого читателя) следует сравнивать непосредственно распределения вероятностей сопоставляемых СВ!

Пример 8. Рассмотрим кривую распределения возраста населения Германии (Рис. 4). Всё, что можно сделать с изображёнными распределениями – это сравнивать

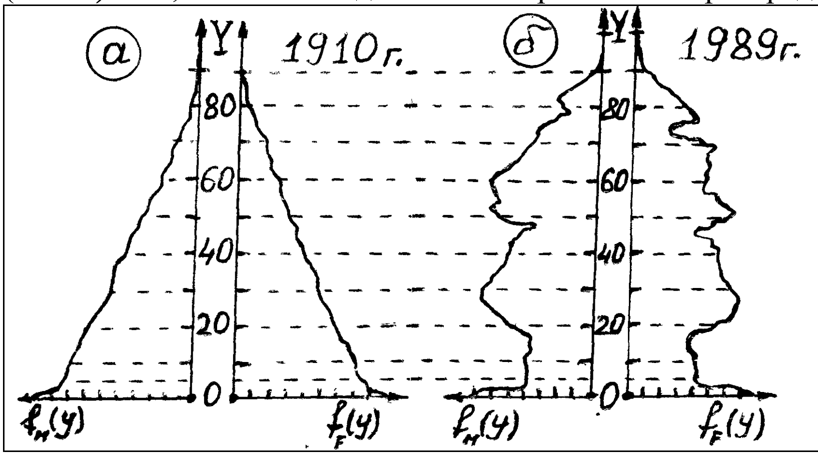

Рис. 4. Распределение возраста жителей Германии в разные годы:

а – 1910 г.; б – 1989 г.; Y – возраст, лет; $f_м(y)$ и $f_F(y)$ – эмпирические функции плотности вероятностей распределения возраста для мужчин и женщин, соответственно.

их друг с другом, так сказать, «непосредственно» по частям - по усмотрению специалистов. Тогда мы можем сопоставить, например, возраст групп людей в определённых (заинтересованными специалистами – акушерами, педиатрами, педагогами, военными, врачами, геронтологами, историками, социологами, страхователями жизни, пенсии и др.) временных интервалах (новорожденных, малышей и т. д., вплоть до престарелых), разных полов, исторических периодов и т. п. Сопоставление приведенных кривых может дать много полезной информации вдумчивому читателю [1, с. 229]. А средняя величина – повторим - это всего лишь абстрактная точка, одна из нескольких математических характеристик положения всего распределения! Здесь уместно заметить, что упомянутыми «характеристиками положения» могут быть и такие «средние» как средняя геометрическая, средняя гармоничная, средняя квадратичная, а также - медиана, мода и др. В каких случаях какое «среднее» значение используют – см. в [1, с. 246].

В процессе статистической обработки **одномерной выборки** из множества значений СВ возникает ряд проблем, решение которых не настолько формализованы, чтобы эту работу можно было полностью доверить ЭВМ. Часто исследователю приходится задумываться над каждым этапом расчётов и принимать решение с учётом специфических особенностей конкретного исследования.

Важно подчеркнуть, что при сравнении нескольких (М) СВ Y_i $\left(i=\overline{1,M}\right)$ следует сравнивать между собой непосредственно их распределения - функции плотности вероятности $f_i(y)$ [или, лучше, - функции распределения $F_i(y)$]. Если сравниваемые функции распределения описываются одинаковыми параметрическими функциями $F_i\ \left(y,\underline{\theta}_i\right),\ i=\overline{1,M}$, где $\underline{\theta}_i$ - вектор параметров i-того распределения, то в этом случае сравнивать можно и векторы параметров $\underline{\theta}_i$, имея в виду, все же, главную задачу: сравнение вероятностных распределений непосредственно!

14. Фрагменты дискуссий с коллегами

Мы не всегда свободны от ошибок, по поводу которых критикуем других.

Приведенные фрагменты дискуссий демонстрируют различие взглядов коллег на проблемы, возникающие при проверке гипотез.

14.1. Фрагменты дискуссии с профессором А. И. Орловым [7, с. 274]

Критик – врач: усыпляет хлороформом похвал, а потом оперирует. (К. Ижиковский)

А. И. Орлов (О): …К сожалению, параметрические семейства, распределения Стьюдента и Фишера существуют лишь в головах авторов учебников по математической статистике. В реальной жизни их нет. Поэтому прикладная статистика использует, в основном, непараметрические методы, в которых распределения результатов наблюдений могут иметь произвольный вид. В моей книге [14] на примере нормального распределения показана невозможность практического использования параметрических семейств для описания распределений конкретных данных.

Ц: В математике используется ряд абстрактных допущений (о безразмерной точке, об идеальной прямой и др., в том числе и о нормальном распределении), которые также «существуют только в головах математиков, а не в реальном мире». Ни одно из подобных допущений в природе не встречается. Однако эти допущения позволяют использовать математические, в том числе и параметрические методы.

В прикладной статистике существует большое количество нечётких экспертных решений (прежде всего – предметных: о важности задачи, о верности исходных данных, об отутствии выбросов в данных, о корректности математической модели, о погрешности измерений, о выборе критического уровня значимости при проверке гипотез и т. п.), куда более существенно влияющих на исход обработки данных, чем «некорректное» решение о выборе типа распределения. Что же делать?

На наш взгляд, аналитический статистик (АС), вооружённый современным ППП, должен статистическое оценивание и проверку гипотез выполнять и параметрическими, и непараметрическими методами, и только после этого приступать к содержательному анализу результатов обработки данных.

О: Распространенные ППП по статистике (SAS, SPSS и др.) соответствуют уровню научных исследований 50-х - 70-х годов XX-го века и, следовательно, отстают от переднего края исследований на 30 – 50 лет. Так что у каждого есть выбор – пользоваться стандартными, но устаревшими ППП, либо потратить ресурсы на современные ППП, чтобы войти в мир современной статистики.

Часто ли распределение результатов ваших наблюдений является нормальным?

Ц: Обычно определенное параметрическое распределение (чаще всего – нормальное) по выборкам малого объёма ($2 < N < 10$) мы просто постулируем. Даже по данным выборок «среднего» объёма ($9 < N < 100$) бывает нетрудно подогнать гипотезу о нормальном распределении [1] (бывает достаточно исправить или отбросить 1 - 2 выброса, если они есть). А если научная теория того требует, то можно принять параметрическое распределение на основании теории.

Другая проблема – получить большую выборку значений исследуемой величины (по Вашим данным [14] – более 2500 точек!) и проверить согласие с определенным параметрическим распределением. Такие выборки «дороги» и редко встречаются. Для обработки таких данных достаточно пользоваться распространенными ППП.

14.2. Фрагменты отзыва д. т. н., проф. Г. В. Гореловой

(г. Таганрог, Россия, g.v.gorelova@gmail.com)

Хорошие люди принесут вам счастье, плохие наградят опытом, худшие дадут урок, а лучшие подарят воспоминания. Цените каждого!

Г: Хочу через пару лекций изложить моим студентам «гипотезы по Цейтлину». Ваш материал вдохновил меня дать и этот раздел в курс лекций, хотя по госстандарту, который неизвестно, кем писан, этот раздел исключён для бакалавров нашей специальности.

Ц: Большое спасибо за такую честь. Буду считать это наградой мне за труд!

14.3. Фрагменты дискуссии с к. б. н., доц. М. В. Комаровой

(г. Самара, Россия; marinakom@yandex.ru)

В корне ошибаются только математики.

К: На кого Ваш труд ориентирован?

Ц: Статья ориентирована на коллег – аналитических статистиков (АСов), а также – исследователей-предметников, желающих познакомиться с проблемами проверки статистических гипотез.

К: Врачи в массе своей любят чёткую инструкцию, рецепт, руководство к действию (для такой-то проблемы - такой метод анализа данных и принятия решений).

Ц: Для врачей и других специалистов в предметных областях исследований, впервые обращающихся к решению задач проверки статистических гипотез, можно рекомендовать книги [4, 6, 19 и 23] с «чёткими инструкциями», а затем - более обстоятельные книги [1, 7, 14, 17, 21] и настоящую статью.

К: В медицинских статьях и диссертациях сейчас указывают не просто результаты проверки статистических гипотез, а - точные Р-значения и сами по себе значения признаков. Медицинское сообщество всегда в первую очередь смотрит на собственно числа и их **клиническую значимость**. При желании каждый читатель может делать свой вывод.

Ц: В связи с этим я хочу обратить внимание исследователей на то, что среднее значение случайной величины – это всего лишь абстрактная, содержательно не интерпретируемая точка, одна из мер положения всего распределения величины! Поэтому, например, указание только на простое, пусть, даже, высоко статистически значимое различие средних значений приводит к поверхностным суждениям. **Клиническую значимость** имеет не просто сравнение средних значений (см. Замечание 14), а сравнение наблюдаемых вероятностных распределений изучаемых величин (см. Пример 8)!

К: Термин «достоверный» - нравится нам или нет, - но используется в медицине как синоним «статистически значимый». Я вообще большой проблемы в этом не вижу.

Ц: Согласен: проблема - небольшая. Это – как обычная стилистическая ошибка. Так и зацикливаться на ней не надо, а просто - писать грамотно (см., например, [24]).

14.4. Фрагменты дискуссии с доктором В. К. Шитиковым [7, с. 222].

Дискуссия математиков закончилась дракой: чего-то не поделили.

В. К. Шитиков (Ш): В прикладных областях (экологии и проч.) принят хороший термин «достоверность отличий» («достоверный» в словаре С. И. Ожегова определяется как «надёжный, не вызывающий сомнений»). «**Сдвиг распределений**

достоверен при таком-то уровне значимости» – звучит уже вполне по-русски. Но «крупные специалисты» в области лингвистики русского языка (Е. С. Вентцель, В. П. Леонов) неожиданно приписали термину «достоверность» исключительную идентичность событию с вероятностью, равной 1. Однако поскольку многовековая русская литература, по самым «**достоверным** слухам», этого толкования не придерживается, я делаю *вывод*, что в научных текстах замена термина «значимость» на «достоверность» вполне допустима.

Н. А. Цейтлин (Ц): Термин «достоверный» (англ.: «reliable», «trustworthy») [Толковый словарь русского языка Ушакова: http://slovari.yandex.ru/dict/ushakov/] по-русски означает «подлинный, несомненный, не вызывающий сомнений». Термин «достоверность» действительно получил широкое хождение в русскоязычных публикациях: «достоверная связь», «достоверное влияние», «достоверность закономерности», «достоверность различия между средними арифметическими» и т. п. **перлы**, коллекция которых собрана в специальной «Кунсткамере» [24]. Подмена термина «значимый» термином «достоверный» магически меняет в глазах читателя смысл статистического вывода: одно дело какая-то там случайная «значимость», а совсем другое – подлинная «достоверность»! Причины таких ошибок – две: нарушение принципов формальной логики и непонимание основ прикладной статистики.

Вероятность α случайных событий может принимать значение в интервале от 0 до 1 **включительно** ($0 \le \alpha \le 1$). Однако событие с вероятностью 0 называют «невозможным», а с вероятностью 1 – «достоверным»; событие же с вероятностью ислючительно внутри этого интервала ($0 < \alpha < 1$) называют «случайным». Часто встречаемая в публикациях замена термина «значимость» на «достоверность» или аналогичная более грубая ошибка – замена термина «проверка непротиворечивости статистической гипотезы» на «доказательство статистической гипотезы» – говорит о том, что существует, возможно, не осознаваемое стремление авторов отождествлять неслучайность различий неких статистических оценок параметров с достоверностью или доказанностью полученных результатов, декларируя тем самым обязательность и неизбежность именно этих выводов исследования. Иными словами, декларируется обязательное взаимно-однозначное соответствие изучаемого комплекса условий и описываемого результата, как следствия этих условий.

В **математике** доказа́тельством называется цепочка логических умозаключений, показывающая, что при каком-то наборе **аксиом** и **правил вывода** верно некоторое утверждение. В отличие же от других наук, в математике недопустимы эмпирические доказательства: все утверждения доказываются исключительно логическими способами. Короче: **доказывать можно теоремы, а не статистические гипотезы!**

Отсюда следует, что в научных текстах замена терминов «значимость» на «достоверность», а также «проверка непротиворечивости статистической гипотезы» на «доказательство статистической гипотезы» недопустима.

Ш: Есть такое понятие «терминологическое поле» – это искусственно очерченная и специально охраняемая от посторонних проникновений область существования терминов и их дефиниций, внутри которого они отличаются от обычных слов. Вся русская художественная и техническая литература никогда не воспринимала «достоверность» события как некую магическую 100%-ю подлинность. И если господа Зорин и Леонов вытаптывают себе «поле» вокруг этого слова, наделяя его мистическими свойствами, то это – их проблемы. Вы же пишете книгу для широкого читателя, а, как указывает «Словарь маркетинговых терминов» (http://www.glossostav.ru/word/2349), достоверность – это показатель вероятности того, что **истинное значение изучаемого параметра генеральной совокупности попадет в доверительный интервал**. А уж странные манипуляции с двойными английскими переводами – это просто от отсутствия иных аргументов.

Ц: Приведенный пример из самодеятельного «Словаря маркетинговых терминов» может также занять достойное место в Кунсткамере [24]: в учебниках и монографиях по математической статистике - как в зарубежных (в работе [8] приведены 9 ссылок), так и в отечественных (11 ссылок - в работе [8]) вероятность накрытия доверительным интервалом параметра генеральной совокупности принято называть просто **доверительной вероятностью**. По этой причине мы не можем рекомендовать широкому читателю употребление доморощенного термина «достоверность».

В заключение необходимо отметить следующее. В настоящей статье воспроизводятся давно высказанные идеи автора [1]. Однако нынче «воз и ныне – там»: никто, кроме автора не проверяет так гипотезы, как описано в книге [1], возможно, потому что мало кто из коллег её прчитал, а если кто и прчитал, то не понял, а если кто и понял, то - не так, как надо; многие (особенно, не статистики) считают, что «имеющиеся ППП по МС перекрывают все потребности статистики, которые только и могут возникнуть». Вот автор настоящей статьи и пытается найти коллег, которые подхватили бы описанные ростки идей, повторно изложенных выше, развили их дальше, и посадили бы их в почву современных ППП по статистике.

14.5. Фрагменты дискуссии с к. б. н. Л. А. Животовским (med-gen@mail.ru)

Не стой на месте: иди, ищи новые тупики!

Ж: Вашу статью я прочёл и обратил внимание на центральную в статье табл. 2. Мне Ваш подход нравится, так как ориентирован на то, чтобы заставить ЛПР задумываться, а не формально выбирать "пороговый" уровень значимости ($P < 0.05$ или $P < 0.01$). Это важно - убрать веру в абсолютность уровней значимости.

Ц: Спасибо за понимание. Многочисленные «черти» сидят в деталях всей концепции проверки гипотез, и в одной статье, даже, столь обширной, трудно было вывести всех «чертей» на белый свет. Для этого приведены ссылки - как на классические источники, так и на предыдущие работы автора, в предположении того, что читатель, желающий глубже проникнуть в концепцию, обратится к ссылкам.

Авторы современных учебников по статистике описывают предложение Неймана и Пирсона, согласно которому в качестве статистики критерия используется по-разному интерпретируемые функции выборки, кроме однозначно интерпретируемой – уровня значимости α_i. В «альфа-методе» предлагается подняться на вершину этой «пирамиды функций», и использовать наболее информативную функцию - уровень $Э_i$ ответственности ЛПР за вывод!

В принятии решения ЛПР присутствуют два компонента – формальный – математический - теория проверки гипотез, оканчивающаяся расчётом оценки уровня значимости α_i, и неформальный – экспертный, оканчивающийся сравнением уровней ответственности ЛПР за вывод $Э_i$. Вот тут-то - в глубинах предпочтений ЛПР - собака и зарыта. Как формально сопоставить оценку α_i уровня значимости с оценкой $Э_i$ уровня ответственности ЛПР? Автор монографии [1], предложил сопоставлять колонки α_i и $Э_i$ в табл. 2, дополнив их индексами i предпочтения гипотез. Затем были кратко изложены алгоритмы расчёта уровней значимости как функции вариантов статистик Фишера, используемых для проверки 23-х наиболее распространённых гипотез относительно параметров нормального распределения СВ - математических ожиданий и дисперсий (см. табл. 3 и сопроводительный текст в книге [1]).

Аналогично адаптируются методики проверки **не**параметрических гипотез, которые также оканчиваются оценками α_i для табл. 2.

Вот теперь вся эта совокупность текстов, табл. 2 и, в качестве примеров - методики проверки параметрических и непараметрических гипотез и есть «Альфа-метод». И вот теперь можно обо всём этом сказать в двух словах приблизительно так:

«Альфа-методом» является такой метод проверки статистических гипотез, который включает два компонента - формальный (математический расчёт оценки уровня значимости статистического критерия) и неформальный - экспертный, позволяющий лицу, принимающему решение, в зависимости от степени предпочтений, сделать вывод об отклонении или принятии проверяемой гипотезы.

14.6. Фрагменты дискуссии с к.б.н. Л. Н. Пороховником (perevesty@mail.ru)

Для тех, кого поднимают на щит, главное - не оторваться от земли.

П: В тексте имеются фрагменты, целесообразность которых сомнительна, например, многочисленные шутливые эпиграфы: их обилие меня раздражало. Видимо, я привык к более академическому стилю.

Ц: Это – «мой персональный стиль», включающий эпиграфы и диспуты. Многие читатели (в основном, студенты) жаловались на скучную академичность моих предыдущих работ. Пришлось пойти навстречу молодым читателям.

П: Насколько я смог понять, для математика содержание - банально.

Ц: В том-то и дело, что банально, но - недостаточно. Для того-то я и привожу дискуссии с коллегами-математиками, чтобы это показать.

П: Возможно, для слабо подкованной в математике аудитории было бы неплохо закрепить материал примерами.

Ц: Я же даю в работе многочисленные ссылки на конкретные примеры, накопленные за многие годы.

П: А я и в пределах написанного не всё понимаю. Например, про «порочную спираль» Вы пишете, что уменьшение уровня значимости, на котором принимается гипотеза, будет раскручивать какую-то «спираль». Объясните, пожалуйста: какую?

Ц: Широко распространено понимание оценки параметра, получаемой в виде функции случайных величин как **числа, то есть, точечной, или однозначной величины.** Например, запись Р-значения есть точечная оценка уровня значимости. Сама же оценка уровня значимости как функция случайных величин, есть случайная величина с каким-то своим «законом» распределения. Для него можно также построить (например, методом Монте-Карло) доверительный интервал, который будет зависеть от числа степеней свободы проверочной статистики. Поэтому сравнение точечной оценки уровня значимости с его критическим значением – также – гипотеза, которую надо проверять. Значит, опять необходимо найти новый уровень значимости и сравнить его с заново задаваемым критическим значением. И так далее, и т. д. Это и есть открытая в работе (если никто не обратил на это внимание раньше) «**порочная спираль**».

П: Я понимаю только то, что уменьшение порогового Р-значения от 0,05 до 0,001 чревато тем, что мы пропустим многие ценные работы только из-за того, что в них было, скажем, $P = 0{,}002$, хотя поток публикаций, конечно, сократится. Причём, для обеспечения высокой значимости придётся даже в очевидных работах увеличивать выборку так, что никаких грантов не хватит. Следовательно, упадёт объём всех публикаций, а не только - ложных ("ложное срабатывание сигнализации от ветерка"), что сильно замедлит научный прогресс. Правильно ли я понимаю?

Ц: Вопросы пропуска многих ценных работ, увеличения выборок, объёма всех публикаций и замедления научного прогресса не имеет отношения к математическим проблемам проверки гипотез. Они имеют отношение к теории экспертного оценивания (ТЭО), и в рамках ТЭО решаются. Подробнее см., например, гл. 10 в книге [1].

14.7. Фрагменты дискуссии с проф. С. А. Кондратовым (kondratov@rune.lg.ua)

Чтоб дело мастера боялось, он знает много страшных слов.

К: Статья Ваша, безусловно, интересна методически и хорошо написана. Но найдёт ли она своего читателя на русскоязычном сайте – большой вопрос.

Надо сказать, что уровень «вероятностно-статистического» образования у нас - крайне низкий. И такая ситуация - повсеместно. Посему о проверке гипотез выпускники вузов имеют очень слабые понятия. В результате выросло несколько поколений инженеров, которые не умеют обрабатывать данные и не понимают, для чего это нужно. С другой стороны, в тех отраслях, которые сейчас интенсивно развиваются, статистику используют для решения серьёзных задач, но - на уровне **под**становки данных в сертифицированные программы, определяемые соответствующими стандартами, где «шаг вправо - шаг влево – стреляют».

Основной идеей Вашей работы является отказ от традиционного уровня значимости 0,05 (замечу, что «технари» считают, что этот уровень выводится математикой (sic!)) и использование вместо этого гибкой ситуативной величины, определяемой ЛПР в каждом конкретном случае. Мне эта идея импонирует, но, к сожалению, на практике, возникает вопрос: **где брать экспертов для оценки критической вероятности и ЛПР, которые эту оценку определяют**?

Ц: Вопрос уместный. Действительно, этому надо, наверно, как-то обучать экспертов и ЛПР. Но - как? Может, приводить наглядные примеры на каких-то жизненных ситуациях? Давайте подумаем. Пока что я ограничился своими «рекомендациями», основанными на, опять же, рекомендациях «опытных статистиков».

К: Может быть, в Германии с этим полегче?

Ц: В Германии, по моим сведениям, работают специалисты в предметных областях, непосредственно - с дорогим (и честно приобретённым) программным обеспечением (ПО). Инструкции к ПО написаны достаточно подробно, так что после года практики эти специалисты довольно успешно решают свои задачи. О критическом уровне значимости 0,05 они особенно не задумываются. Дело в том, что хорошие результаты проверки гипотез (например, в регрессионном анализе) дают оценки уровней значимости гораздо меньшие 0,05. Так что и задумываться особенно не о чём.

К: Хочу обратить Ваше внимание на вопрос, связанный с мощностью критерия, то есть, вероятностью ошибки второго рода. Его считают редко, хотя есть практические задачи, в том числе – гостированные, где это необходимо делать. Однако, если вопросы проверки гипотез с уровнем значимости рассматриваются детально, и во многих случаях, то вопрос расчёта мощности критерия мало где рассматривается. При этом в литературе часто рассуждают о мощности, например, критерия Стьюдента и Вилкоксона, говорят, что первый мощнее второго (см. работы А. И. Орлова), не указывая, как мощность зависит от базы сравнения.

Ц: Как правило, одинаковая мощность непараметрического критерия достигается большим числом степеней свободы по сравнению с параметрическим критерием. Но А. И. Орлов утверждает, что разность эта невелика.

К: Так что, не всё чисто, тут тоже надо бы поработать!

Ц: Потомкам останется ещё немало работы!

К: В целом, Ваша статья мне понравилась. Она носит выраженный методический характер, посему все определения понятий очень уместны. Может быть, имеет смысл сделать некий научно-методический сайт, который мог бы быть площадкой для дискуссий?

Ц: Так я же начал работу в этом направлении (см. мой блог "Прикладная математическая статистика").

К: Но, опять-таки, его лучше сделать либо англоязычным, либо двуязычным. Также неплохо было бы поместить некие численные примеры.
Ц: Они, естественно, накоплены и приведены в ссылках.

В заключение выражаю уважаемым коллегам благодарность за колоссальный труд на прочтение, конструктивную критику и предложения, внесенные в ходе дискуссии.

15. Четыре меры качества статистического вывода

Будь внимателен, прежде чем ошибиться.

Диалектический «закон возникновения нового качества вследствие накопления количественных изменений» гласит: «развитие осуществляется путём накопления количественных изменений в предмете, что приводит к выходу за пределы меры и скачкообразному переходу к новому качеству».

Выделим четыре меры качества статистического вывода: «не значимо», «значимо», «практически достоверно», «достоверно» (табл. 3).

Таблица 3. Меры качества статистического вывода

Альтернатива H_1	Оценка α_i
не значима	$\alpha_{кi} \le \alpha_i < 1$
значима	$\alpha_{нi} \le \alpha_i < \alpha_{кi}$
практически достоверна	$0 < \alpha_i < \alpha_{нi}$
достоверна	$\alpha_i = 0$

α_i - оценка уровня значимости;
$\alpha_{кi}$ - критическое значение;
$\alpha_{нi}$ – ничтожное значение.

Очевидно, что при значении уровня значимости $\alpha_i = 0$ (точно) вполне допустимо говорить о **достоверном** отклонении проверяемой гипотезы H_0 и соответствующем **достоверном** признании альтернативной гипотезы H_1.

Такие случаи, как показано выше, могут встретиться, например, в процессе проверки гипотез бутстреп-методом.

Рассмотрим такое статистическое понятие как «**практически достоверное**» событие, которое встречается в процессе проверки гипотез и параметрическим методом, когда значения уровня значимости α_i ничтожно мало отличается от 0 (то есть, $\alpha_i \approx 0$ приблизительно). Обозначим ничтожно малую вероятность $\alpha_{нi}$. Если получен уровень значимости $0 < \alpha_i < \alpha_{нi}$, например, при $\alpha_{нi} = 10^{-10}$ (по А. Н. Колмогорову [9]), то с уровнем $Э_i$ ответственности ЛПР за вывод в пределах $99{,}5\% < Э_i < 100\%$ можно уже говорить о «практически достоверном» отклонении гипотезы H_0 и признании гипотезы H_1; если же $\alpha_{нi} \le \alpha_i < \alpha_{кi}$, то должно говорить о высокой значимости альтернотивной гипотезы H_1 и, наконец, при $\alpha_{кi} \le \alpha_i < 1$ – приходится говорить о незначимости альтернотивной гипотезы H_1.

Таким образом, возможно, что табл. 3 позволит прийти экспертам к компромиссу в дискуссии о качестве статистического вывода.

16. Метод доверительных интервалов для парного сравнения статистических параметров и функций распределения

Если не знаешь, как поступить, то поступай, как знаешь.

Обозначения

ДИ – доверительный(-ные) интервал(-лы);
МДИ - метод доверительных интервалов;
СВ - случайная(-ные) величина(-ны);
МО – математическое(-кие) ожидание(-ния);
СО - среднеквадратичное(-ные) отклонение(-ния);

ФР – функция(-ции) распределения(ний) СВ;
ГФР - генеральная(-ные) функция(-ции) распределения СВ;
ЭФР – эмпирическая(-кие) функция(-ции) распределения СВ.

Иногда бывает полезно представлять читателю результаты статистической обработки данных наглядным образом с помощью МДИ. Задача парного сравнения статистических параметров (МО, СО, долей, коэффициентов корреляции или регрессии и т. п.) и функций распределения ряда СВ решается «графически» - в виде нескольких специально рассчитанных ДИ. Этот приём позволяет читателям-предметникам наглядно представлять убедительные аргументы в пользу соответствующих статистических гипотез. Алгоритмы расчётов и ряд примеров применения МДИ изложены в книге [1].

МДИ был предложен [34, 1] для проверки гипотез о равенстве пар однотипных параметров нормально распределённых СВ - МО и СО. Дальнейшее развитие МДИ получил в работе [33]: было показано, что МДИ можно воспользоваться и для проверки гипотез о согласии пар ГФР СВ.

Даны m независимых выборок $\{y_{ji}\}_{i=1}^{N_j}$ значений y_{ji} СВ Y_j (j = 1, 2, …, m) равных объёмов - по $N_j = N$ элементов – каждая, распределенных нормально, $Y_j \sim N(\nu_{Yj}, \sigma_{Yj})$, имеющих непрерывные ГФР $F_i(y) = P(Y_i \le y)$, где: ν_{Yj} – МО; σ_{Yj} – СО.

Необходимо с помощью МДИ [34, 1] проверить $U = C_m^2 = m(m - 1)/2$ нулевых гипотез H_{0nq} о равенстве пар **параметров** ν_{Yn} и ν_{Yq} или σ_{Yn} и σ_{Yq}, или - о согласии пар ГФР $F_n(y)$ и $F_q(y)$:

$$H_{0nq}: u_n(y) = u_q(y), \ (1 \le n < q \le U) \qquad (22)$$

против U альтернативных гипотез H_{1nq}: $u_n(y) \ne u_q(y)$, где $u_n(y)$, $u_q(y)$ – **параметры** или ГФР.

МДИ заключается в сравнении на графике двусторонних ДИ, построенных в виде пар плеч $\delta_{jP/2}$ (отрезков прямых с двух сторон от точечных оценок параметров) с такой доверительной вероятностью (1-P), чтобы на априори заданном (см. табл. 2) [1, 7] уровне значимости $\alpha_к$ читатель мог визуально определить значимость различия сравниваемых пар параметров. Если ДИ перекрываются, то соответствующая гипотеза H_{0nq} не отклоняется, если же - не перекрываются, то - отклоняется в пользу H_{1nq}.

Критическим событием является равенство суммы плеч $\delta_{nP/2} + \delta_{qP/2}$ и критического расстояния Δ_λ между сравниваемыми оценками параметров, т. е.

$$\delta_{nP/2} + \delta_{qP/2} = \Delta_\lambda; \ (1 \le n < q \le U), \qquad (23)$$

где $\lambda = \lambda_к$ – критическая вероятность, и $\lambda_к = \lambda(\alpha_к)$.
Для проверки гипотез (22) относительно МО в виде H_{0nq}: $\nu_{Yn} = \nu_{Yq}$ при равных СО ($\sigma_{Yn} = \sigma_{Yq} = \sigma_Y$) Р. Фишер, основываясь на теореме Бонферрони, предложил [22] использовать верхний $\lambda_к$-предел статистики Стьюдента $t_{\lambda;mf}$, где $\lambda = \lambda_к = \alpha_к/(2U)$; mf - число степеней свободы. Это предложение и формула (23) позволили [34] для построения 100(1-P)-процентных ДИ использовать соотношение

$$t_{P/2;mf} = 2^{-0,5} t_{\lambda;mf}; \ \lambda = \lambda_к = \alpha_к/(2U), \qquad (24)$$

где $t_{P/2,mf}$ - верхний P/2-предел распределения Стьюдента (см. пример на рис. 5).

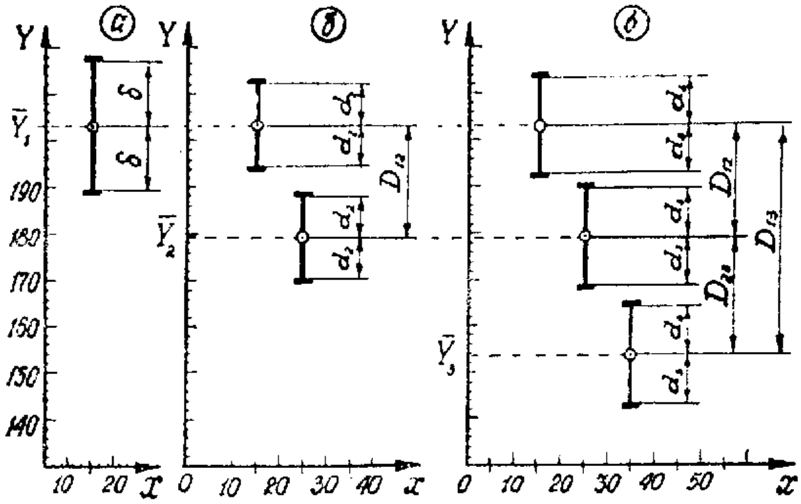

Рис. 5. Интервальное оценивание и проверка гипотез методом ДИ относительно перепадов давления завихрителя вихревого пылеулавливающего аппарата при различных углах установки лопаток [34, 1]:
а) 95%-ный доверительный интервал для ν_{Y_1}; *б)* два 83%-ных доверительных интервала для проверки гипотезы (22) о равенстве центров $\nu_{Y_1} = \nu_{Y_2}$; *в)* три 98%-ных доверительных интервала для проверки гипотез (22) о равенстве центров $\nu_{Y_1} = \nu_{Y_2}; \nu_{Y_1} = \nu_{Y_3}; \nu_{Y_2} = \nu_{Y_3}$.
X - угол установки лопаток завихрителя, град, *Y* - перепад давлений завихрителя.

При N → ∞ (а практически - при N ≥ 25) зависимость (24) принимает вид [1, с. 225]:

$$Z_{P/2} = 2^{-0,5} Z_\lambda;\ \lambda = \lambda_к = \alpha_к/(2U), \qquad (25)$$

где $Z_{P/2}$ и Z_λ - верхние P/2- и λ-пределы распределения Гаусса. Аналогичное соотношение для построения ДИ СО в МДИ было получено при проверке гипотез (22) относительно СО в виде H_{0nq}: $\sigma_{Yn} = \sigma_{Yq}$ [1] (см. пример на рис. 6).

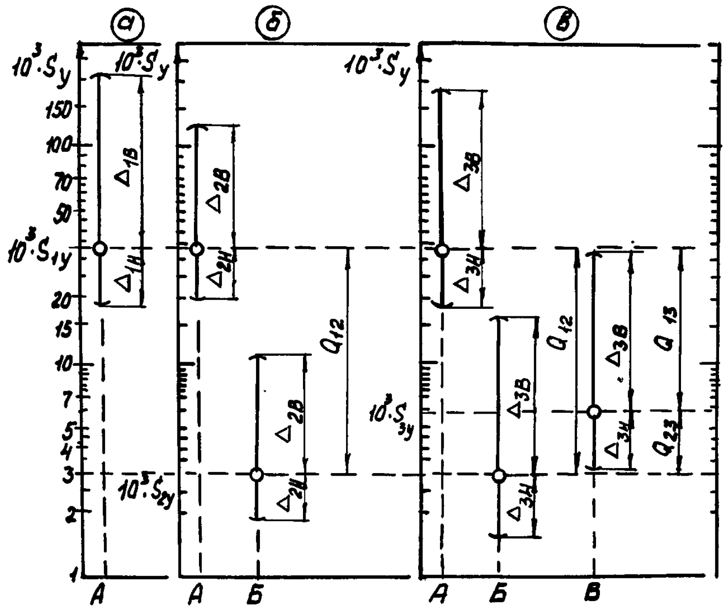

Рис. 1. Статистические оценивание и проверка гипотез относительно среднеквадратичных отклонений (СО) случайных погрешностей измерения массовой концентрации магния в растворе бикарбоната магния методом доверительных интервалов (ДИ) [1]:

а) 95%-ный ДИ для σ_{1Y}; б) два 87%-ных ДИ для проверки гипотезы (22) о равенстве СО $\sigma_{1Y} = \sigma_{2Y}$; в) три 93,3%-ных ДИ для проверки гипотез (22) о равенстве СО $\sigma_{1Y} = \sigma_{2Y}$, $\sigma_{1Y} = \sigma_{3Y}$, $\sigma_{2Y} = \sigma_{3Y}$; S_Y - СО, г/л; S_{1Y}, S_{2Y}, S_{3Y} – СО, полученные по результатам наблюдений, выполненных лаборантами А, Б и В соответственно; Δ_{iH}, Δ_{iB} - нижние и верхние плечи ДИ; Q_{ij} - расстояния между значениями S_{iY} и S_{jY}.

Для проверки гипотез (22) в виде H_{0nq}: $F_n(y) = F_q(y)$ о согласии пар ГФР автор работы [33] предложил воспользоваться соотношением, аналогичным формуле (25):

$$D_{P/2;N} = 2^{-0,5}D_{\lambda;N};\ \lambda = \lambda_к = \alpha_к/(2U);\ \lambda < 0,2, \tag{26}$$

где $D_{P/2;N}$ и $D_{\lambda;N}$ - верхние P/2- и λ-пределы распределения Колмогорова-Смирнова. Подставив зависимость $D_{\lambda;N} = ((-0,5\ln(\lambda/2))/N)^{0,5}$, справедливую [35, с. 372] для непрерывной ГФР $F(y)$, в (26), получим

$$D_{P/2;N} = ((-\ln(\alpha_к/2m(m-1))/4N)^{0,5};\ N > 70. \tag{27}$$

Данные для **примера** (рис. 7) заимствованы из маркетингового исследования [7];

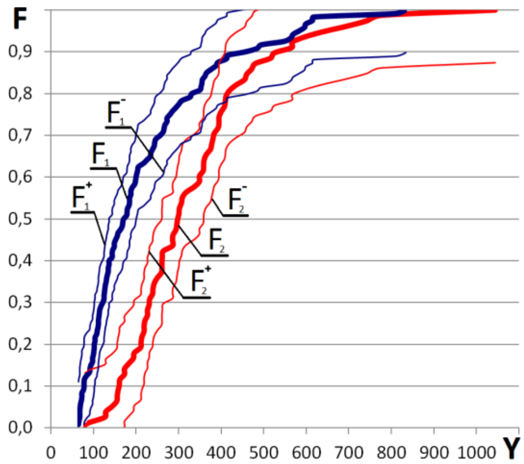

Рис. 7. Сравнение распределений времени пребывания мужчин F_1 и женщин F_2 в павильоне напитков:
Y – время, с; F_i – эмпирические ФР (ЭФР); F_i^- и F_i^+ – нижние и верхние границы соответствующих ДИ; m = 2; $\alpha_к$ = 0,05.

$N_1 = 78$; $N_2 = 120$; по формуле (27): $D_{P/2;78} = ((-\ln(0,05/2\times2(2-1))/4\times78)^{0,5} = 0,119$; $D_{P/2;120} = 0,0955$. Нижние F_i^- и верхние F_i^+ границы ДИ построены путём сдвига ЭФР F_i вниз и вверх на величины $D_{P/2;78} = 0,119$ и $D_{P/2;120} = 0,0955$ соответственно; медианы распределений (при F = 0,5) $y_{м1} = 175$ с и $y_{м2} = 308$ с различаются значимо (ДИ не пересекаются). Однако различия времени пребывания в павильоне напитков менее 110 с и более 316 с между 15% и 30% мужчин и женщин соответственно, не значимы.

Метод ДИ обладает рядом преимуществ по сравнению с традиционными методами статистической обработки данных.

1. Зачастую при традиционном способе сравнения СВ ограничиваются сравнением лишь центров распределений, игнорируя различия в формах функций распределения.

2. Традиционный способ наглядного представления выборки данных заключается в построении **гистограмм** распределения СВ. Поскольку гистограмма является дифференциальной функцией распределения СВ, то для её построения приходится разбивать выборку данных на группы произвольного объёма. В то время как кумулятивная функция распределения не зависит от произвола эксперта-статистика и даёт однозначное и наглядное представление о распределении СВ.

3. Значимое отличие выборочного распределения от нормального (например, при наличии выбросов данных) не является критическим при проверки гипотез, потому что для этого используются устойчивые к выбросам непараметрические методы. Что же касается наличия выбросов, то решение этой проблемы переносится в область компетенции экспериментатора: если он посчитает, что выбросы не являются грубыми ошибками, то в общей выборке их можно сохранить.

4. Методы проверки гипотез с помощью гистограмм распределения нам не известны, а - с помощью кумулятивных функций методом ДИ – изложены в работе [33].

5. Известно, что сглаженная эмпирическая дифференциальная функция распределения СВ даёт легко интерпретируемую картину распределения СВ. Такое сглаживание возможно в пределах ДИ гистограммы распределения СВ с помощью подходящей непрерывно дифференцируемой функции (см, например, аппроксимацию (21) в работе [1]). Для расчёта оценок коэффициентов q_i, ν_i и σ_i этой аппроксимации достаточно было использовать критерий минимума максимума расстояния Колмогорова-Смирнова D с ограничением $D \le D_{\alpha,N}$, где $D_{\alpha,N}$ – половина ДИ.

17. Теория практических невероятностей

Всё, что нельзя исправить, следует считать совершенным.

В своём шутливом эссе «**Теория практических невероятностей**» (ТПН) [1, с. 115] автор пишет: «Введём обозначение: α_k – критический уровень **незначимости** и пусть $0 < \alpha_k < 1$. Возможно, лучше было бы сохранить привычное для статистиков название α_k - «критический уровень значимости». Однако, у авторов этого термина что-то неладно с логикой. Нормальный человек никак не может понять: почему повышение значимости (обычное - для предпочитаемой альтернативной гипотезы H_1) соответствует... уменьшению уровня значимости? В нашей ТПН с этим безобразием покончено: повышение **незначимости** соответствует повышению **уровня незначимости** α_k и - наоборот!

Пусть $\boldsymbol{\alpha}$ - вероятность некоторого события и $1 > \boldsymbol{\alpha} > 0$. Тогда «практически невероятным» будем называть всякое событие, вероятность которого меньше критической $\boldsymbol{\alpha}_k$, то есть при $0 < \alpha < \alpha_k < 1$. Вся проблема заключается только в выборе критического уровня незначимости $\boldsymbol{\alpha}_k$.

По этому вопросу сотни лет высказывались все, кто мало-мальски считал себя практичным человеком, умеющим оценивать риск. Мы остановились на таблице 2, обобщающей мнение статистиков. Мнения специалистов в предметных областях ещё ждут своего обобщения.

18. Из недавнего опыта применения «Альфа-метода проверки гипотез»

Итак, вы написали трактат и хотите опубликовать его. Поступите следующим образом. Если вы понимаете, что написали и можете доказать это - пошлите трактат в математический журнал. Если вы понимаете, что написали, но не можете доказать это - пошлите в журнал по физике. Если вы не понимаете, что написали, но можете доказать это - пошлите в журнал по экономике. Если вы ничего не понимаете, что написали и не можете доказать это – поместите свой трактат на своём сайте в интернете.

Для решения рутинных статистических задач авторы обычно используют «традиционные» приёмы решения, которые, тем не менее, могут быть модернизированы, в частности, путём применения «Альфа-метода проверки гипотез».

Педагогика (2011 - 2012 г.) В ряде работ [12, 25-27] рассмотрены различные **метрологические аспекты оценки успеваемости учащихся**.

Трейдинг (2013 г.) В работах [28, 29] приведены **способы описания рядов цен акций**, чтобы, основываясь на их динамике, можно было пользоваться современным программным обеспечением по **прогнозированию временных рядов**.

Маркетинговые исследования (2014 г.) В работах [8, 30-32] приведены **алгоритмы статистического анализа быстрой реакции пользователей на телевизионную рекламу интернет-магазина**.

19. Задачи дальнейших исследований

Надежней сразу самому не сделать, чем предложить другому сделать и ждать, пока он не сделает.

Как видно из сказанного выше, проблемы проверки гипотез невозможно полностью формализовать. Без содержательных экспертных оценок и мнений ЛПР - как на начальном этапе (формулировки гипотез), так и на конечном (принятии решения) - обойтись невозможно. Тем не менее, для математиков-статистиков, занятых совершенствованием формальных методов, остаётся ещё поле для исследований.

1. Дальнейшего изучения требует решение проблемы выбора между параметрическими и непараметрическими или численными методами проверки гипотез.

2. Дальнейшего исследования требует решение задачи сравнения оценки уровня значимости и его критического значения за две итерации (см. раздел 12).

3. Открытым остаётся вопрос о границах мер качества статистического вывода.

Выводы

Уже поздно возвращаться назад, чтобы все правильно начать, но ещё не поздно устремиться вперёд, чтобы все правильно закончить.

Формулируя статистические гипотезы, лицо, принимающее решение (ЛПР или исследователь «в роли» ЛПР) должен убедить читателя в справедливости предпочитаемой и впоследствии принятой им гипотезы.

«Общепринятые» критические значения уровней значимости (0,05) являются анахронизмом, данью устаревшей «традиции».

Критические значения уровней ответственности ЛПР (или исследователя «в роли» ЛПР) за вывод (см. табл. 2) – оценки $Э_{кi}$, в зависимости от предпочитаемой ЛПР гипотезы – нулевой H_0 (i = 0) или альтернативной H_1 (i = 1) - рекомендуется задавать в диапазоне ($50\% < Э_{кi} < 100\%$), как можно ближе к 100%.

«Последнее слово» об отклонении или принятии проверяемой гипотезы отдаётся экспертной оценке $Э_{кi}$ уровня ответственности ЛПР за вывод.

Литература

Гении - это такие умники, которые все придумали и сделали до тебя.

[1] Цейтлин Н. А. Из опыта аналитического статистика. - М.: Солар, 2007. - 906 с. www.cubematrix.com/oldsite/anlagen/as.pdf.

[2] Эльясберг П. Е. Измерительная информация. Сколько её нужно? Как её обрабатывать? – М: Наука. - 2011 г. – 208 с.

[3] Цейтлин Н. А. Применение методов математической теории эксперимента в содовой промышленности. Обзорная информация. Серия «Содовая промышленность». - М.: НИИТЭХИМ, 1984. - 48 с.

[4] Brownlee K. A. (1960). Statistical Theory and Methodology in Science and Engineering, John Wiley & Sons, Inc., New York, p. 236. (Браунли К. А. Статистическая теория и методология в науке и технике. - М.: Наука, 1977. - 408 с. http://www.twirpx.com/file/547505/ и http://knigi.tor2.net/index.php?id=102268)

[5] Леман Э. Проверка статистических гипотез. - М.: Наука, 1964. - 498 с.

[6] Плохинский Н. А. Руководство по биометрии для зоотехников. - М.: Колос. - 256 с.

[7] Горбач А. Н., Цейтлин Н. А. Покупательское поведение: анализ спонтанных последовательностей и регрессионных моделей в маркетинговых исследованиях. - Киев: Освіта України, 2011. - 298 с. http://www.cubematrix.com/oldsite/anlagen/asp.pdf.

[8] Цейтлин Н. А., Горбач А. Н. Маркетинговое исследование телевизионной рекламной кампании интернет-магазина. CuBe Matrix, Inh. A. Gorbach. Hamburg – 2015. - 111 с. www.cubematrix.com/oldsite/book/tv_ad_analysis.pdf.

[9] Хромов-Борисов Н. Н. Синдром статистической снисходительности и доверчивости или значение и назначение Р-значения. (Санкт-Петербургский Государственный Медицинский Университет им. акад. И. П. Павлова). 16.02.2011 г. http://tele-conf.ru/aktualnyie-problemyi-tehnologicheskih-izyiskaniy/sindrom-statisticheskoy-snishoditelnosti-ili-znachenie-i-naznachenie-p-znacheniya.html.

[10] Алимов Ю. И. О практической ценности теории оценок. - Автоматика, 1981, № 2, с. 84-94.

[11] Эфрон Б. Нетрадиционные методы многомерного статистического анализа. Сб. статей. - М.: ФиС, 1988.- 263 с. http://www.biznesbooks.com/2010-01-07-17-48-08/2367-jefron-b-netradicionnye-metody-mnogomernogo-statisticheskogo-analiza

[12] Цейтлин Н. А. Метрологические аспекты оценки успеваемости учащихся с. 124 / Современные проблемы математики и её приложения в естественных науках и информационных технологиях. - Сб. тез. докл. междунар-й конф. «Тараповские чтения - 2012». Харьков: "Апостроф", 2012. – 136 с. http://tarapov-conf.univer.kharkov.ua/reports/39.pdf

[13] Z. Ciesielski, R. Zieliński. POLYNOMIAL AND SPLINE ESTIMATORS OF THE DISTRIBUTION FUNCTION WITH PRESCRIBED ACCURACY, APPLICAT. MATHEMATICAE 36, 1 (2009), pp. 1–12.

[14] Орлов А. И. Эконометрика: учебник для вузов. – Ростов н/Д: Феникс, 2009. – 572 с. http://orlovs.pp.ru/econ.php#ek1.

[15] Устойчивые статистические методы оценки данных / пер. с англ. Ю. И. Малахова; под ред. Н. Г. Волкова. – М.: Машиностроение, 1984. – 232 с.

[16] Гаскаров Д. В., Шаповалов В. И. Малая выборка. - М.: Статистика, 1978. – 248 с.

[17] Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Классификация и снижение размерности: Справочное издание. - М.: Финансы и статистика, 1989. - 608 с.

[18] Справочник по надежности / Пер. с англ. Ю. Г. Епишина и Б. А. Смиренина; Под ред. Б. Р. Левина. Т. 1. - М.: Мир, 1969. – 340 с.

[19] Головач А. В., Ерина А. М., Трофимов В. П. Критерии математической статистики в экономических исследованиях. – М.: Статистика, 1973. - 136 с.

[20] Оуэн Д. Сборник статистических таблиц. М.: АН СССР, 1966. – 568 с.

[21] Статистические методы в экспериментальной физике / Идье В., Драйард, Джеймс Ф. и др.; Пер. с англ.; Под ред. А. А. Тяпкина. – М.: Атомиздат, 1976. – 336 с.

[22] Fisher R. M. The design of experiments. - London: Oliver and Boyd, 1935.- 360 p.

[23] Доерфель К. Статистика в аналитической химии. - М.: Мир, 1969.-248 с.

[24] Зорин Н. А. О неправильном употреблении термина «достоверность»: http://www.biometrica.tomsk.ru/let1.htm.

[25] Сословский В. Г., Цейтлин Н. А. Рациональные оценки успеваемости учащихся с поправками на угадывание правильных ответов при рутинном компьютерном тестировании. (с. 86 - 95). –Харьковский национальный автомобильно-дорожный университет \ Сб. научных трудов. -Харьков: «Міськдрук», 2011. - 399 с. http://www.youtube.com/watch?v=DtOU_qsai8o&list=PLC097436338E0A811&index=17&feature=plpp_video

[26] Цейтлин Н. А. Статистический подход к оцениванию знаний учащихся. Тез. докл. 3-й международной научно-практической конф. «Компьютерное моделирование в химии и технологиях и устойчивое развитие». Киев – Рубежное, 10-12 мая 2012 г. 284 с. / Сб. научных трудов. - Киев: «КПИ», 2012. – с. 254 – 256. http://www.cmct.org.ua/; http://pedsovet.org/component/option,com_mtree/task,viewlink/link_id,74598/Itemid,6/;

[27] Цейтлин Н. А. Компьютерное тестирование и статистическая обработка оценок, с.154 - 158. / Теория и практика измерения и мониторинга компетенций и других латентных переменных в образовании: материалы XVII (01 – 02 февраля 2012 г.), XVIII (25 – 26 июня 2012 г.) всероссийских научно-практ. конф-й / отв. ред. А. А. Маслак; филиал ФГБОУ ВПО «КубГУ» в г. Славянске-на-Кубани. – Славянск-на-Кубани: Изд. центр филиала ФГБОУ ВПО «КубГУ» в г. Славянске-на-Кубани, 2012. – 173 с.

[28] Цейтлин Н. А. Описание временного ряда цен акций. 1. Редукция временного ряда, С. 119-120. Сб. тез. докл. междунар. школы-конф. ТАРАПОВСКИЕ ЧТЕНИЯ-2013 "Современные проблемы математики, механики, информатики". Харьков, 29 сентября - 4 октября 2013 г. http://theormech.univer.kharkov.ua/abstr013.html

[29] Цейтлин Н.А. Описание временного ряда цен акций. 2. Анализ временного ряда. - Там же, с. 120-121. http://theormech.univer.kharkov.ua/abstr013.html

[30] Цейтлин Н. А., Горбач А. Н. Измерение быстрой реакции пользователей на телевизионную рекламу интернет-магазина. - Сб. трудов XVIII Международной научно-практич. конф. «Системный анализ в проектировании и управлении» 1 – 3 июля 2014 г. ч. 1. – СПб., изд. Политех. Ун-та. - Сообщение 1. Алгоритмы расчётов. - с. 68 – 75. - Сообщение 2. Погрешности методов измерения. - с. 76 – 83.

[31] Цейтлин Н. А., Горбач А. Н. Измерение быстрой реакции пользователей на телевизионную рекламу интернет-магазина. - Дев'ята міжнародна науково-практична конференція «Математичне та імітаційне моделювання систем. МОДС 2014". Тези доповідей. – Жукин. – 2014. – 23-27 червня 2014р. – 426 с. - Сообщение 1. Описание методов измерения. - с. 99 - 102. - Сообщение 2. Исследование погрешности методов измерения с. 103 - 105.

[32] Gorbach A., Tseitlin N. Tv or not tv. Modeling-Ansatz untersucht Werbewirkung. - Research & Results, www.research-results.de, 7, 2014. - S. 54, 55. www.cubematrix.com/oldsite/download/tv_or_not_tv_modeling-ansatz_artikel_r&r.pdf

(Перевод: Горбач А., Цейтлин Н. ТВ или – не ТВ: исследование эффективности телевизионной рекламы. – «Научные исследования и результаты», 7, 2014. - С. 54, 55.).

[33] Цейтлин Н. А. Парное сравнение статистических параметров и функций распределения методом доверительных интервалов. Сборник тезисов докладов международной школы-конференции «Тараповские чтения-2013». «Современные проблемы математики, механики, информатики». Харьков, 1-15 марта 2016 г. - с. 120. http://theormech.univer.kharkov.ua/tar-conf/presents/059.pdf

[34] Цейтлин Н. А. Проверка гипотез методом доверительных интервалов. В кн. Методы математической статистики в основной химии. Труды, Т. 55, НИОХИМ, Харьков, 1981, с. 82-89.

[35] Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Справочное издание. Основы моделирования и первичная обработка данных. - М.: Финансы и статистика, 1983. - 471 с.