

Значение и интерпретация ошибки среднего в клиническом исследовании и эксперименте

Г. П. Тихова

ООО «ИнтелТек Лаб», Петрозаводск

Importance and interpretation of standard error of mean in clinical study and trial

G. P. Tikhova

IntelTeck Lab Ltd, Petrozavodsk

В статье рассмотрен содержательный смысл и различие между стандартной ошибкой среднего и среднеквадратическим отклонением, изложены факторы, влияющие на величину этих статистических параметров. *Ключевые слова:* обработка клинических данных, ошибка среднего, стандартное отклонение.

This paper highlighted the meaning and difference between standard error of mean and standard deviation as well as describes the factors that affect the values of this statistical parameters. *Key words:* clinical data processing, standard error of mean, standard deviation.

Прежде чем погрузиться в тонкости различных методов статистической обработки данных, зададимся вопросом: что мы собственно считаем с их помощью? Как рассчитанные на нашей выборке статистические параметры связаны с описанием исследуемой нами популяции?

Для примера рассмотрим числовую случайную величину с нормальным законом распределения – ударный объем у здоровых беременных женщин со сроком гестации не менее 30 недель.

Итак, во-первых, мы четко определяем интересующую нас популяцию – здоровые беременные женщины со сроком гестации 30 недель и более. Представим мысленно, что мы можем измерить ударный объем у всех без исключения членов этой популяции. (Конечно, на самом деле это сделать нереально!) Регистрируем ударный объем у 1-й беременной и получаем, допустим, 60 мл, затем у 2-й – 55 мл, у 3-й – 78 мл и т. д. В конечном итоге мы получаем бесконечное множество реализаций случайной величины под названием ударный объем. Можем ли мы посчитать среднее значение всех полученных измерений? Конечно, можем. Что же это будет за величина? Это будет точное среднее значение ударного объема, присущее популяции беременных со сроком от 30 недель и более. Подчеркну, именно точное среднее значение, без всякой ошибки, потому что по условию нашего мысленного эксперимента мы измерили ударный объем у всех без исключения членов популяции. Поскольку мы перебрали всех членов популяции, то откуда же взяться ошибке? Мы

не оставили никакой неопределенности, включив в расчет всю популяцию, поэтому и смогли посчитать точное среднее значение без всякой ошибки. Но из этого следует, что в рассматриваемой популяции среднее значение – это объективная реальность, а не придуманная математиками величина. Причем это среднее значение равно вполне конкретному числу, присущему именно этой популяции. В другой популяции оно, вообще говоря, может быть другим, и если это так, то по показателю ударного объема эти две популяции различаются, и данный показатель может служить признаком отличия этих популяций друг от друга. Точно также, имея полученные измерения, мы можем подсчитать точное значение стандартного отклонения ударного объема в этой популяции. Это же касается и медианы, и моды, и всех остальных статистик, которые обычно рассчитываются для числовой случайной величины. Что же получается? Все эти интегральные параметры случайной величины – не искусственное изобретение, а вполне реальные свойства, имеющие точные значения в конкретной популяции. Можно сказать, точные значения этих параметров определяются особенностями конкретной популяции и в свою очередь характеризуют ее, как отличительные признаки, присущие только ей. Когда мы проводим реальное исследование, то вынуждены довольствоваться выборкой, т. е. лишь частью (и, как правило, далеко не большей частью) интересующей нас популяции. Однако мы все-таки рассчитываем и среднее значение, и стандартное отклонение, и другие статистики в условиях,

когда подавляющее большинство членов популяции оказываются за рамками нашего исследования. Что же мы делаем и что получаем в итоге? Поскольку мы включаем в наш расчет лишь часть популяции, привносим некоторую неопределенность, неточность в расчет среднего и других статистик и в итоге получаем не точные значения, а лишь ОЦЕНКИ этих параметров. Такие оценки называют выборочными оценками, потому что они рассчитаны по конкретной выборке и в некоторой степени зависят от ее особенностей. Если мы повторим исследование, наберем другую выборку из этой же популяции, и снова рассчитаем среднее, то это новое рассчитанное значение, скорее всего, будет отличаться от первого, хотя, по сути, мы рассчитали то же самое. Почему нам следуем ожидать этого несовпадения? Потому что на выборке мы получаем лишь оценку точного значения, а не само популяционное значение статистики, и эта оценка, полученная на выборке, всегда, подчеркну, всегда без исключения имеет ошибку. Ошибка эта называется статистической, потому что она обусловлена случайными факторами. В отличие от коварной и вредоносной ошибки смещения, статистическая ошибка, можно сказать, безобидна, т. к. ее всегда можно рассчитать и сделать на нее поправку. Она всегда явная и точно известна по величине. Но если ошибки смещения можно попытаться избежать, то статистическая ошибка – это неотъемлемая часть выборочной оценки. При расчете любого параметра на выборке, статистическая ошибка будет иметь место и от нее никуда не деться.

Таким образом, рассчитывая на нашей выборке среднее значение, стандартное отклонение и другие параметры числовой случайной величины, мы на самом деле проводим оценивание их реальных точных значений в популяции, из которой эта выборка была взята. Оценка, как мы уже упоминали, может быть более или менее точной, но необходимо смириться с тем, что на выборке мы никогда не получим абсолютно точной величины, которая имеет место в данной исследуемой популяции. А это значит, что, вообще говоря, каждую рассчитанную на выборке статистику мы должны сопровождать ее ошибкой, чтобы можно было понять, с какой точностью оценен тот или иной параметр. На практике однако принято указывать ошибки не всех рассчитываемых статистических параметров, а лишь наиболее важных для сравнения выборок. Так, например, как правило, указывают ошибку среднего. О ней и поговорим поподробнее.

Поскольку ошибка среднего рассчитывается по очень простой формуле непосредственно из стандартного отклонения, то ее часто путают с этим последним. В научных статьях принято

указывать рядом со средним значением либо ошибку среднего, либо стандартное отклонение, но это только потому, что, зная одно, мы легко можем вычислить другое при известном объеме выборки. Из этого однако не следует, что ошибка среднего и стандартное отклонение – это одно и то же. Указанные два параметра описывают совершенно разные свойства исследуемого показателя. Ошибка среднего характеризует точность выборочной оценки среднего значения исследуемого показателя в заданной популяции и зависит от варибельности, объективно присущей этому показателю в условиях данной популяции, и репрезентативности выборки:

- чем чище и больше наша выборка, тем меньше ошибка;
- чем меньше варибельность исследуемого показателя, тем меньше ошибка.

Иными словами, ошибка среднего дает нам понять, насколько сильно мы отклоняемся в своих расчетах от точного значения среднего в популяции. Если бы наши расчеты охватили всю заданную популяцию, то ошибка среднего была бы равна нулю, т. е. отсутствовала бы полностью.

Что касается стандартного отклонения, то по своему смыслу к точности оценки среднего оно не имеет никакого отношения. Как мы уже разбирали ранее, стандартное отклонение характеризует объективно существующую варибельность исследуемого показателя, и поскольку мы здесь обсуждаем только случайные величины с нормальным вероятностным распределением, эта статистика не может обращаться в ноль ни при каких обстоятельствах, даже в случае охвата всей популяции. Стандартное отклонение, рассчитанное по выборке, само является приближенной оценкой точного популяционного значения стандартного отклонения и, соответственно, тоже имеет статистическую ошибку, величина которой определяется теми же условиями, что и ошибка среднего. Просто эту ошибку, как правило, не указывают. В критериях сравнения она не играет такой важной роли, как ошибка среднего.

Таким образом, стандартное отклонение характеризует объективную природу данных, с которыми мы имеем дело в исследовании, а ошибка среднего отражает точность и адекватность нашего исследования и полученных результатов. С увеличением объема выборки ошибка среднего стремится к нулю, а стандартное отклонение – к определенному числу, являющемуся точным значением стандартного отклонения этого показателя в популяции.

Иными словами, стандартное отклонение определяется объективной природой исследуемых данных, а ошибка среднего зависит:

- от объективной природы данных (исследуемой случайной величины), а именно ее вариабельности, выражаемой в стандартном отклонении;
- субъективной особенности конкретного исследования – объема выборки.

Мы очень подробно рассмотрели выборочную оценку среднего и его ошибку, но другие статистики, рассчитанные на выборке, также обладают некоторой статистической ошибкой в силу тех же причин, что и выборочное среднее значение. Из этого следует, что, указывая полученное в ходе расчетов значение статистического параметра, мы должны указать и его статистическую ошибку. Все в точности так же, как это принято для выборочного среднего значения. Это касается всех рассчитываемых статистик независимо от типа исследуемого показателя. По популярности расчета следом за средним значением идет оценка частот или долей для качественных, а иногда и порядковых показателей, поэтому на одном таком примере мы и рассмотрим, как правильно описывать такие результаты.

В исследовании, посвященном спинальной анестезии у беременных, необходимо было определить долю различных вариантов интраоперационной тошноты и рвоты (ИОТР), а именно:

- сколько процентов пациенток не имели никаких осложнений,
- у какой доли беременных отмечалась гипотония во время операции,
- сколько процентов пациенток жаловались на тошноту,
- у какой доли наблюдаемых развилась интраоперационная рвота.

Исследуемая популяция определена как беременные со сроком гестации более 35 недель (беременность без патологий), спинальная анестезия проводилась по поводу планового кесарева сечения. Понятно, что в этой популяции мы исследуем качественный признак ИОТР, который является случайной величиной, принимающей в каждом конкретном случае одно из перечисленных значений, и во всей популяции доля каждого из вариантов вполне определена и имеет точное значение. Всю популяцию мы охватить в своем исследовании не можем, поэтому довольствуемся выборкой и, следовательно, полученные процентные доли (или частоты осложнений) будут лишь оценками точных популяционных величин. Что из этого следует? То, что все полученные числа, выборочные значения долей, будут иметь статистические ошибки, которые мы также должны рассчитать

и указать. В качестве примера приведена табл., взятая из работы А. М. Погодина и Е. М. Шифмана, напечатанной в 1-м номере нашего журнала за 2009 г. [1].

Частота эпизодов ИОТР в исследуемых группах

Группа*	Частота ИОТР (доля \pm ошибка доли, %)
I (n = 35)	17,1 \pm 2,4
II (n = 33)	18,2 \pm 2,6
III (n = 36)	11,1 \pm 1,6
Всего (n = 104)	15,4 \pm 1,3

* Стратификация пациентов на 3 группы производилась в соответствии с антиэметиком, применявшимся в составе премедикации.

Если мы выделим из той же популяции выборку большего объема, то выборочные значения долей у нас будут ближе к реальным, а ошибки – меньше. Независимо от того, что и на каких данных вы считаете, поведение всех расчетных статистических параметров будет одинаково:

- если параметр отражает объективную реальность, то его выборочные оценки с увеличением объема выборки будут приближаться к конкретным числам – точным значениям в популяции;
- если параметр характеризует точность и адекватность исследования, то с увеличением объема выборки его значение будет неуклонно уменьшаться и стремиться к нулю.

Все статистические ошибки с ростом объема выборки снижаются и стремятся к нулю.

Процентные доли не являются исключением, поэтому при указании результатов в процентных долях необходимо записывать полученное выборочное значение с его статистической ошибкой (ошибкой доли), поскольку сравнение долей требует такой же математической строгости, как и сравнение средних значений. Нельзя утверждать, что одна доля больше другой лишь на том основании, что мы получили два числа, одно больше другого. Их разность может укладываться в статистическую ошибку одного из них (или обоих) и они могут быть просто неточными, приближенными, оценками одного и того же точного популяционного значения. Например, в приведенной табл. в группе I частота ИОТР составила 17,1%, а в группе II – 18,2%, однако мы не можем утверждать, что в I группе частота меньше, чем во II, поскольку ошибки этих долей покрывают оба этих значения, указывая на то, что они с точки зрения статистики равны, а их различие – просто случайная неточность оценки, которую привносит данная, не очень большая, выборка и природная вариабельность данных.

Вопросы, которые мы здесь разобрали, порой кажутся второстепенными и мелкими, но на самом деле они требуют пристального внимания

исследователя при интерпретации результатов, полученных в ходе статистической обработки клинических данных. Глубокое понимание природы и источников статистической ошибки, а также строгий контроль ее величины позволят избежать неправомερных выводов: обнаружить эффект там, где его нет, или влияние фактора, когда в реальности оно отсутствует.

Литература

1. *Погодин А. М., Шифман Е. М.* Профилактика тошноты и рвоты при спинномозговой анестезии во время операции кесарева сечения // Регионарная анестезия и лечения острой боли. 2009; 1: 11–14.