СТАТИСТИЧЕСКИЕ СПОСОБЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Выполнил студент
Группы пмпм-21
Юрек С.Г.

Методы статистической обработки результатов эксперимента.

Методами статистической обработки результатов эксперимента называются математические приемы, формулы, способы количественных расчетов, с помощью которых показатели, получаемые в ходе эксперимента, можно обобщать, приводить в систему, выявляя скрытые в них закономерности. Речь идет о таких закономерностях статистического характера, которые существуют между изучаемыми в эксперименте переменными величинами.

Некоторые из методов математико-статистического анализа позволяют вычислять так называемые элементарные математические статистики, характеризующие выборочное распределение данных, например выборочное среднее, выборочная дисперсия, мода, медиана и ряд других. Иные методы математической статистики, например дисперсионный анализ, регрессионный анализ, позволяют судить о динамике изменения отдельных статистик выборки. С помощью третьей группы методов, скажем, корреляционного анализа, факторного анализа, методов сравнения выборочных данных, можно достоверно судить о статистических связях, существующих между переменными величинами, которые исследуют в данном эксперименте.

Методы первичной статистической обработки результатов эксперимента

Все методы математико-статистического анализа условно делятся на первичные и вторичные. Первичными называют методы, с помощью которых можно получить показатели, непосредственно отражающие результаты производимых в эксперименте измерений. Соответственно под первичными статистическими показателями имеются в виду те, которые применяются в самих психодиагностических методиках и являются итогом начальной статистической обработки результатов психодиагностики. Вторичными называются методы статистической обработки, с помощью которых на базе первичных данных выявляют скрытые в них статистические закономерности.

К первичным методам статистической обработки относят, например, определение выборочной средней величины, выборочной дисперсии, выборочной моды и выборочной медианы. В число вторичных методов обычно включают корреляционный анализ, регрессионный анализ, методы сравнения первичных статистик у двух или нескольких выборок.

Мода

Числовой характеристикой выборки, как правило, не требующей вычислений, является так называемая мода. Модой называют количественное значение исследуемого признака, наиболее часто встречающееся в выборке. Для симметричных распределений признаков, в том числе для нормального распределения, значение моды совпадает со значениями среднего и медианы. Для других типов распределении, несимметричных, это не характерно. К примеру, в последовательности значений признаков 1, 2, 5, 2, 4, 2, 6, 7, 2 модой является значение 2, так как оно встречается чаще других значений - четыре раза.

Моду находят согласно следующим правилам:

1) В том случае, когда все значения в выборке встречаются одинаково часто, принято считать, что этот выборочный ряд не имеет моды. Например: 5, 5, 6, 6, 7, 7 - в этой выборке моды нет.

2) Когда два соседних (смежных) значения имеют одинаковую частоту и их частота больше частот любых других значений, мода вычисляется как среднее арифметическое этих двух значений. Например, в выборке 1, 2, 2, 2, 5, 5, 5, 6 частоты рядом расположенных значений 2 и 5 совпадают и равняются 3. Эта частота больше, чем частота других значений 1 и 6 (у которых она равна 1). Следовательно, модой этого ряда будет величина =3,5

3) Если два несмежных (не соседних) значения в выборке имеют равные частоты, которые больше частот любого другого значения, то выделяют две моды. Например, в ряду 10, 11, 11, 11, 12, 13, 14, 14, 14, 17 модами являются значения 11 и 14. В таком случае говорят, что выборка является бимодальной.

Могут существовать и так называемые мультимодальные распределения, имеющие более двух вершин (мод).

4) Если мода оценивается по множеству сгруппированных данных, то для нахождения моды необходимо определить группу с наибольшей частотой признака. Эта группа называется модальной группой.

Медиана

Медианой называется значение изучаемого признака, которое делит выборку, упорядоченную по величине данного признака, пополам. Справа и слева от медианы в упорядоченном ряду остается по одинаковому количеству признаков. Например, для выборки 2, 3, 4, 4, 5, 6, 8, 7, 9 медианой будет значение 5, так как слева и справа от него остается по четыре показателя. Если ряд включает в себя четное число признаков, то медианой будет среднее, взятое как полусумма величин двух центральных значений ряда. Для следующего ряда 0, 1, 1, 2, 3, 4, 5, 5, 6, 7 медиана будет равна 3,5.

Знание медианы полезно для того, чтобы установить, является ли распределение частных значений изученного признака симметричным и приближающимся к так называемому нормальному распределению. Средняя и медиана для нормального распределения обычно совпадают или очень мало отличаются друг от друга. Если выборочное распределение признаков нормально, то к нему можно применять методы вторичных статистических расчетов, основанные на нормальном распределении данных. В противном случае этого делать нельзя, так как в расчеты могут вкрасться серьезные ошибки.

Выборочное среднее

Выборочное среднее (среднее арифметическое) значение как статистический показатель представляет собой среднюю оценку изучаемого в эксперименте психологического качества. Эта оценка характеризует степень его развития в целом у той группы испытуемых, которая была подвергнута психодиагностическому обследованию. Сравнивая непосредственно средние значения двух или нескольких выборок, мы можем судить об относительной степени развития у людей, составляющих эти выборки, оцениваемого качества.

Выборочное среднее определяется при помощи следующей формулы:

$$\overline{X} = \frac{(X_1 + X_2 + \ldots + X_n)}{n} = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} X_i \right)$$

где x - выборочная средняя величина или среднее арифметическое значение по выборке; n - количество испытуемых в выборке или частных психодиагностических показателей, на основе которых вычисляется средняя величина; x_k - частные значения показателей у отдельных испытуемых. Всего таких показателей n, поэтому индекс k данной переменной принимает значения от 1 до n; ∑ - принятый в математике знак суммирования величин тех переменных, которые находятся справа от этого знака. Выражение соответственно означает сумму всех x с индексом k, от 1до n.

Разброс выборки

- Разброс (иногда эту величину называют размахом) выборки обозначается буквой R. Это самый простой показатель, который можно получить для выборки - разность между максимальной и минимальной величинами данного конкретного вариационного ряда, т.е.
- $R = x_{max} - x_{min}$
- Понятно, что чем сильнее варьирует измеряемый признак, тем больше величина R, и наоборот. Однако может случиться так, что у двух выборочных рядов и средние, и размах совпадают, однако характер варьирования этих рядов будет различный. Например, даны две выборки:
- X = 10 15 20 25 30 35 40 45 50X = 30 R = 40
- Y = 10 28 28 30 30 30 32 32 50 Y=30 R = 40
- При равенстве средних и разбросов для этих двух выборочных рядов характер их варьирования различен. Для того чтобы более четко представлять характер варьирования выборок, следует обратиться к их распределениям

Дисперсия

Дисперсия - это среднее арифметическое квадратов отклонений значений переменной от её среднего значения.

Дисперсия как статистическая величина характеризует, насколько частные значения отклоняются от средней величины в данной выборке. Чем больше дисперсия, тем больше отклонения или разброс данных.

$$D = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$$

где 5 - выборочная дисперсия, или просто дисперсия;

(……) - выражение, означающее, что для всех x, от первого до последнего в данной выборке необходимо вычислить разности между частными и средними значениями, возвести эти разности в квадрат и просуммировать;

n - количество испытуемых в выборке или первичных значений, по которым вычисляется дисперсия. Однако сама дисперсия, как характеристика отклонения от среднего, часто неудобна для интерпретации. Для того, чтобы приблизить размерность дисперсии к размерности измеряемого признака применяют операцию извлечения квадратного корня из дисперсии. Полученную величину называют стандартным отклонением.

Из суммы квадратов, делённых на число членв ряда извлекаеся квадратный корень.

$$S_x = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Иногда исходных частных первичных данных, которые подлежат статистической обработке, бывает довольно много, и они требуют проведения огромного количества элементарных арифметических операций. Для того чтобы сократить их число и вместе с тем сохранить нужную точность расчетов, иногда прибегают к замене исходной выборки частных эмпирических данных на интервалы. Интервалом называется группа упорядоченных по величине значений признака, заменяемая в процессе расчетов средним значением

Методы вторичной статистической обработки результатов эксперимента

- С помощью вторичных методов статистической обработки экспериментальных данных непосредственно проверяются, доказываются или опровергаются гипотезы, связанные с экспериментом. Эти методы, как правило, сложнее, чем методы первичной статистической обработки, и требуют от исследователя хорошей подготовки в области элементарной математики и статистики.

Обсуждаемую группу методов можно разделить на несколько подгрупп:

1. Регрессионное исчисление.
2. Методы сравнения между собой двух или нескольких элементарных статистик (средних, дисперсий и т.п.), относящихся к разным выборкам.
3. Методы установления статистических взаимосвязей между переменными, например их корреляции друг с другом.
4. Методы выявления внутренней статистической структуры эмпирических данных (например, факторный анализ). Рассмотрим каждую из выделенных подгрупп методов вторичной статистической обработки на примерах.

Регрессионное исчисление

- Регрессионное исчисление - это метод математической статистики, позволяющий свести частные, разрозненные данные к некоторому линейному графику, приблизительно отражающему их внутреннюю взаимосвязь, и получить возможность по значению одной из переменных приблизительно оценивать вероятное значение другой переменной (7).
- Графическое выражение регрессионного уравнения называют линией регрессии. Линия регрессии выражает наилучшие предсказания зависимой переменой (Y) по независимым переменным (X).

Регрессию выражают с помощью двух уравнений регрессии, которые в самом прямом случае выглядят, как уравнения прямой.

$Y = a_0 + a_1 * X$ (1)

$X = b_0 + b_1 * Y$ (2)

В уравнении (1) Y - зависимая переменная, X - независимая переменная, a_0 - свободный член, a_1 - коэффициент регрессии, или угловой коэффициент, определяющий наклон линии регрессии по отношению к осям координат.

В уравнении (2) X - зависимая переменная, Y - независимая переменная, b_0 - свободный член, b_1 - коэффициент регрессии, или угловой коэффициент, определяющий наклон линии регрессии по отношению к осям координат.

Количественное представление связи (зависимости) между X и Y (между Y и X) называется регрессионным анализом. Главная задача регрессионного анализа заключается в нахождении коэффициентов a 0, b 0, a1 и b 1 и определении уровня значимости полученных аналитических выражений, связывающих между собой переменные X и У.

При этом коэффициенты регрессии a 1 и b 1 показывают, насколько в среднем величина одной переменной изменяется при изменении на единицу меры другой. Коэффициент регрессии a 1 в уравнении можно подсчитать по формуле:

$$a1 = r_{xy} \cdot \frac{S_y}{S_x}$$

- а коэффициент b 1 в уравнении по формуле

$$b1 = r_{yx} \cdot \frac{S_x}{S_y}$$

- где r_{yx} - коэффициент корреляции между переменными X и Y;
- S_x - среднеквадратическое отклонение, подсчитанное для переменной X;
- S_y - среднеквадратическое отклонение, подсчитанное для переменной У/

Для применения метода линейного регрессионного анализа необходимо соблюдать следующие условия:

1. Сравниваемые переменные X и Y должны быть измерены в шкале интервалов или отношений.

2. Предполагается, что переменные X и Y имеют нормальный закон распределения.

3. Число варьирующих признаков в сравниваемых переменных должно быть одинаковым.

Заключение

Если данные, полученные в эксперименте, качественного характера, то правильность делаемых на основе их выводов полностью зависит от интуиции, эрудиции и профессионализма исследователя, а также от логики его рассуждений. Если же эти данные количественного типа, то сначала проводят их первичную, а затем вторичную статистическую обработку. Первичная статистическая обработка заключается в определении необходимого числа элементарных математических статистик. Такая обработка почти всегда предполагает как минимум определение выборочного среднего значения. В тех случаях, когда информативным показателем для экспериментальной проверки предложенных гипотез является разброс данных относительного среднего, вычисляется дисперсия или квадратическое отклонение. Значение медианы рекомендуется вычислять тогда, когда предполагается использовать методы вторичной статистической обработки, рассчитанные на нормальное распределение, Для такого рода распределения выборочных данных медиана, а также мода совпадают или достаточно близки к средней величине. Этим критерием можно воспользоваться для того, чтобы приблизительно судить о характере полученного распределения первичных данных.

Вторичная статистическая обработка (сравнение средних, дисперсий, распределений данных, регрессионный анализ, корреляционный анализ, факторный анализ и др.) проводится в том случае, если для решения задач или доказательства предложенных гипотез необходимо определить статистические закономерности, скрытые в первичных экспериментальных данных. Приступая к вторичной статистической обработке, исследователь прежде всего должен решить, какие из различных вторичных статистик ему следует применить для обработки первичных экспериментальных данных. Решение принимается на основе учета характера проверяемой гипотезы и природы первичного материала, полученного в результате проведения эксперимента.