

ДИСПЕРСИОННЫЙ АНАЛИЗ В ОЦЕНКЕ ПАРАМЕТРОВ РЫНОЧНОЙ КОНЪЮНКТУРЫ

В. Ю. Руденков

Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», Республика Беларусь

Научный руководитель А. М. Бондарева, канд. экон. наук, доцент

Дисперсионный анализ как метод статистического исследования, с помощью которого изучается влияние отдельных факторов на резульативный показатель, позволяет среди множества факторов выделить один и оценить его влияние на вариацию резульативного признака. Слово «дисперсионный» указывает на сопоставление процесса анализа компонент дисперсии изучаемой переменной. Общая изменчивость переменной раскладывается на составляющие: межгрупповую (факторную), обусловленную различием групп (средних значений), и внутригрупповую (остаточную), обусловленную случайными причинами. Чем больше частное от деления межгрупповой и внутригрупповой изменчивости, тем больше различаются средние значения сравниваемых выборок и тем выше статистическая значимость этого различия.

Дисперсионный анализ используется для обработки экспериментальных данных, результатов опроса, данных наблюдений. После первичной обработки данных возможно определить влияние фактора или группы факторов на резульативный показатель. Например, шанс получения травмы на производстве в зависимости от стажа работника, зависимость политических взглядов от социального положения, популярность определенных размеров одежды в зависимости от географии проживания и т. д.

В исследовании конъюнктуры рынка факторами независимой переменной чаще всего являются:

– социально-демографические, поведенческие и прочие характеристики потребителей (пол, возраст, социальное положение, доход, степень осведомленности о продукте и прочее);

– отдельные элементы комплекса маркетинга (цена, ассортимент, интенсивность рекламной кампании, интенсивность комплекса мер по стимулированию сбыта, варианты упаковки товара, виды рекламных роликов и прочее).

Зависимой переменной может выступать объем покупок, частота покупок, предпочтение товара или торговой марки, имидж фирмы, оценка потребительских намерений приобретения товара, вероятность покупки товара и другие переменные.

Количество независимых и зависимых переменных определяет вид и подвид дисперсионного анализа. Так с помощью *одномерного дисперсионного анализа* исследуется влияние одной или нескольких независимых переменных на одну зависимую переменную. Он включает в себя: однофакторный и многофакторный дисперсионный анализ. Примером постановки вопросов однофакторного дисперсионного анализа может быть: какой из трех (или более) рекламных роликов имеет лучшую запоминаемость? Влияет ли тип рекламы (плакаты, реклама по радио и др.) на число посетителей в кинотеатре? Примеры постановки вопросов двухфакторного дисперсионного анализа: влияет ли на выбор потребителя уровень образования (высшее, среднее, неполное среднее) и возраст? Как осведомленность о магазине (высокая, средняя, низкая) и представление о нем (позитивное, нейтральное, негативное) влияют на предпочтения потребителей? Пример постановки вопроса трехфакторного дисперсионного анализа: как меняется намерение потребителей купить товар при различных уровнях цен, каналах распределения и интенсивности рекламной кампании?

Главным преимуществом многофакторного дисперсионного анализа является возможность исследовать эффект взаимодействия факторов. Взаимодействие имеет место тогда, когда эффект одного фактора на зависимую переменную зависит от уровня других факторов. Например, многофакторный дисперсионный анализ позволяет ответить на вопрос: усиливают ли друг друга реклама по радио и мероприятия прямого маркетинга, проводимые в торговом центре?

С помощью *многомерного дисперсионного анализа* исследуется влияние нескольких независимых переменных на несколько зависимых переменных. Например: как интенсивность рекламы (высокая, средняя, низкая) и уровень цены (высокий, средний, низкий) одновременно влияют на объем продаж и имидж магазина?

В проведенной работе подробнее рассмотрим одномерный однофакторный дисперсионный анализ на примере выборки группы случайных людей, по которой определим самый востребованный размер одежды, а также зоны экономического риска, чтобы получить максимальную прибыль.

Допустим, что есть исследование, в котором 1000 случайных людей измеряли по росту. После чего все полученные результаты поделили на группы и посчитали, сколько человек в какую группу входит. Это и есть распределение людей по росту.

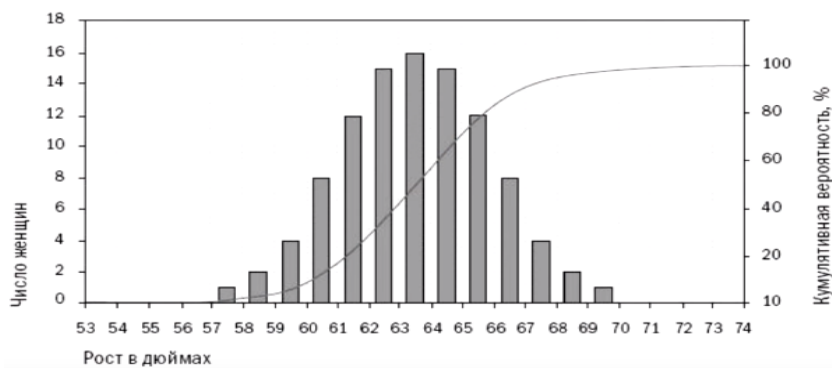


Рис. 1. График распределения людей в группы по росту

На графике виден максимум, в котором больше всего людей определенного роста, если набор исходных данных случайный, то этот максимум называется *математическим ожиданием* или ожидаемым значением. То есть если взять из толпы случайного человека, то его рост будет близок к ожидаемому значению.

Также есть производитель одежды и в его планах всю произведенную одежду продать. Одна из главных причин, почему его продукцию не купят, заключается в том, что имеется риск произвести одежду с неправильным соотношением размеров, вследствие чего будут затрачены ресурсы, а готовая продукция не раскупится. Чтобы этого избежать, необходимо воспользоваться математической статистикой для дисперсионного анализа возможных рисков.

Имеется распределение с исследования 1000 человек. По нему можно понять, какой размер будет востребован больше всего (математическое ожидание), какие размеры должны быть максимальными и минимальными, но непонятно, какие должны быть точные соотношения размера одежды. То есть отходя от самого востребованного размера на один шаг размера непонятно, на сколько надо сделать одежды меньше.

С помощью математической статистики можно по имеющемуся массиву данных посчитать следующие значения: μ – математическое ожидание (ожидаемое значение 178); σ^2 – дисперсия (12,25); σ – среднееквадратичное отклонение (3,5), где значения в скобках были взяты для исследования.

Дисперсия – это значение, которое характеризует, насколько распределение случайных величин размытое, то есть насколько пологи боковины линии распределения.



Рис. 2. Влияние дисперсии на плотность вероятности распределения

Теперь распределение случайной величины можно записать в виде формулы и построить не просто столбчатую диаграмму, а непрерывный график (рис. 3).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} - \text{функция плотности вероятности нормального распределения, где } \sigma \text{ определяется из массива данных по формуле } \sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}.$$

где σ определяется из массива данных по формуле $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$.

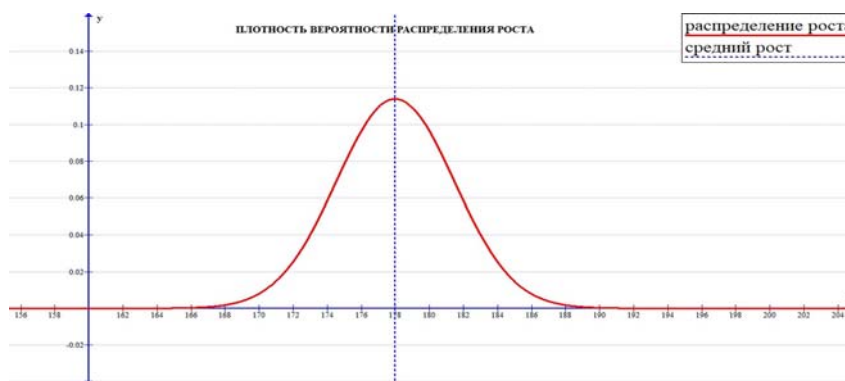


Рис. 3. График плотности вероятности распределения роста

Далее можно узнать доли и вероятность попадания величины роста в любом диапазоне, что будет очень просто. Самое главное свойство распределения – это то, что оно по всей своей площади дает 100 % (рис. 4) или же единицу, если построено в долях. Из этого следует, что чем выше линия распределения роста или чем ближе исследуемый промежуток к значению математического ожидания, тем больше становится площадь исследуемого промежутка и тем больше вероятность попадания случайного человека в этот промежуток.

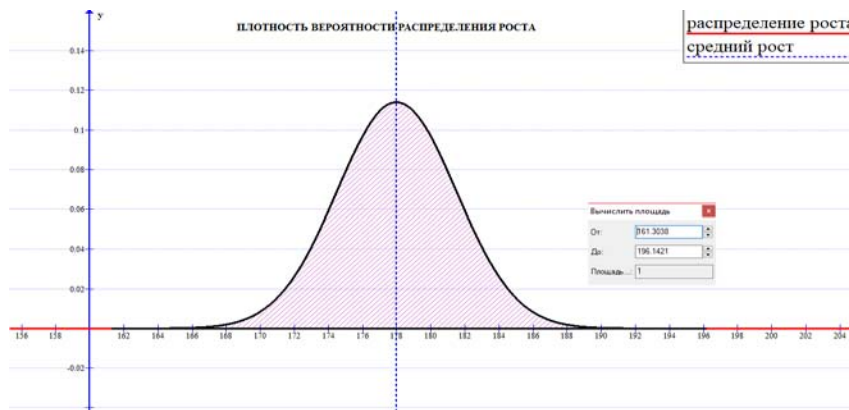


Рис. 4. Плотность вероятности распределения роста

Если взять кусочек графика, который нам нужен для вычисления доли, например, для производства штанов N -го размера, то есть возможность по нужным измерениям произвести просчет площади плотности вероятности в определенном диапазоне. Примером послужит интервал в 2 см (см. рис. 4).

После расчета видно, с какой вероятностью случайный человек окажется в этом диапазоне роста (в нашем случае это вероятность равна 3,2 %) (рис. 5). Тем самым появляется возможность спрогнозировать соотношение спроса на разные размеры одежды и не складировать ткань в виде готовой продукции, которую невозможно будет реализовать, что сводит риски к минимуму.



Рис. 5. График плотности вероятности распределения роста (фрагмент)

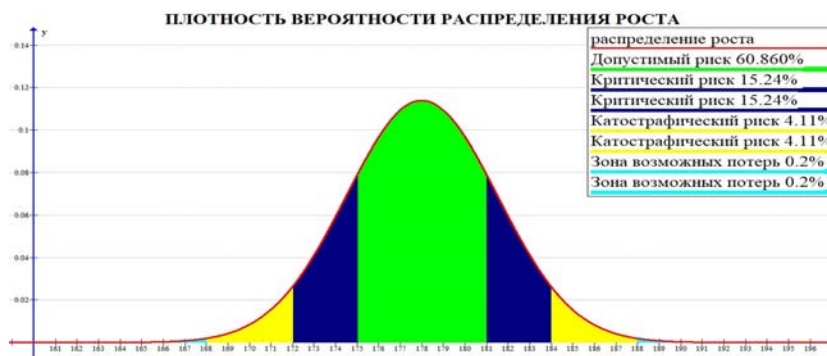


Рис. 6. Плотность вероятности распределения роста людей с делением на зоны риска

На рис. 6 обозначены зоны риска, где проценты обозначают вероятность попадания случайного человека в одну из зон. Из этого можно сделать вывод, что больше прибыли мы получим в зоне допустимого риска, где вероятность попадания человека в эту зона равна 60,86 %. Но чтобы получить максимальную прибыль, необходимо на каждую из зон затратить определенный процент от имеющихся ресурсов, т. е. на ту же зону допустимого риска необходимо потратить 60,89 % от имеющихся ресурсов, на зону критического риска – 30,48 %, на зону катастрофического риска – 8,11 % и на зону возможных потерь – 0,4 %.

Таким образом, в ходе данной работы было выявлено, что для получения максимальной прибыли необходимо правильно использовать статистические данные и уметь их правильно проанализировать. Причем данные не всегда должны быть больших объемов, что приводит к небольшим неточностям по сравнению с большим массивом данных, но экономит ресурсы.