

Раздел 1. Введение в курс «Анализ данных в маркетинге».

Лекция 1. Место и задачи анализа данных в маркетинге.

Систематизация и представление статистических данных. (2 часа)

Анализ данных это понятие означает совокупность действий, осуществляемые исследователем в процессе изучения неких данных с целью формирования определенного представления о характере описываемого явления. Предметом является техника формирования определенного представления. Основная цель анализа данных - выявление (подтверждение, корректировка) каких-то интересующих исследователя статистических закономерностей; или, другими словами, - определенного рода сжатие, усреднение содержащейся в данных информации.

Задачу поиска закономерности иногда отождествляют с задачей *объяснения* интересующего исследователя явления (напомним, что главный смысл объяснения состоит в подведении объясняемого явления под какой-либо закон, явление – это не обязательно наша содержательная закономерность). Достижение соответствующей цели является актуальной в любом социологическом исследовании. Ее почти всегда можно считать основной целью анализа.

Наряду с объяснением изучаемого явления, представляется целесообразным всегда иметь в виду по крайней мере еще две цели: описание исходных данных и осуществляемое на основе выявленной закономерности предсказание того или иного явления. *Описание* - цель, достичь которую часто бывает необходимо прежде, чем непосредственно приступить к поиску основной интересующей исследователя закономерности. *Предсказание* тоже зачастую считается основной целью научного исследования и с этим трудно спорить (афоризм О.Конта «Знать, чтобы предвидеть»).

Описание требуется для того, чтобы исследователь мог хотя бы самым приблизительным образом сориентироваться в том "море" данных. Ведь

далеко не всегда социологу полностью ясно, каков вид закономерностей, "скрывающихся" за интересующими его данными, какими признаками эти закономерности должны описываться, возможно, ли подобрать соответствующие признаки и т.д. Описание может помочь дать ответы на такие вопросы, чтобы потом уже можно было более направленно искать интересующие социолога соотношения.

Прогноз тех или иных характеристик жизни общества по существу служит целью выявления любой закономерности.

Для понимания сути анализа данных важно отметить, что и при описании данных, и при прогнозе могут использоваться алгоритмы того же рода, что и при поиске основной закономерности. Границы между этими тремя целями часто бывают размыты. Кроме того, можно выделить и другие цели. Упомянем здесь лишь одну из них, лежащую в русле уже упомянутой нами гуманитарной парадигмы – *понимание* изучаемого явления.

Как известно, термин "понимание" как название одной из главных функций науки с конца XIX века является ключевым для социологии.

Уделяется огромное внимание анализу сходства и различия смыслов терминов "объяснение" и "понимание" как отражений соответствующих функций науки. Бытует точка зрения, в соответствии с которой главная познавательная функция наук о природе – объяснение (подведение единичного объекта под общий закон, в результате чего уничтожается неповторимость объекта), а наук о духе – понимание (т.е. изучение объекта в его неповторимости). Мы присоединяемся к другому мнению, в соответствии с которым любая наука (это особенно касается наук о человеке и, в частности, социологии) должна и объяснять, и понимать (свое "понимание" имеется, скажем, даже в математике; этого мы здесь не касаемся).

Нужно упомянуть о понимании как об одной из познавательных функций социологии в силу огромной важности достижения понимания изучаемого объекта (человека) в любом социологическом исследовании.

«Понимание» обычно достигается с помощью мягких методов исследования.

Основными задачами, которые решает анализ данных являются:

1. Классификация объектов – поиск однотипных групп объектов, создание типологии.
2. Сжатие информации:
 - Одномерный анализ – описательная статистика;
 - Многомерный анализ – связь между признаками;
 - Поиск латентных переменных.

Статистическая закономерность в социологии.

В науке принято выделять две основные формы закономерной связи явлений, отличающиеся по характеру вытекающих из них предсказаний: динамические и статистические закономерности. В законах динамического типа предсказание имеет точный, определенный однозначный вид; в статистических же законах предсказание носит не достоверный, а лишь вероятностный характер.

Нас интересуют в основном статистические закономерности, так называемые закономерности "в среднем".

Статистическая закономерность возникает как результат взаимодействия большого числа элементов, составляющих совокупность, и характеризуют не столько поведение отдельного элемента совокупности, сколько всю совокупность в целом. Проявляющаяся в статистических закономерностях "необходимость" возникает вследствие взаимной компенсации и уравнивания множества случайных факторов, "пробивает" себе дорогу через массу случайностей, контрпримеров, отступлений от нее.

Статистические закономерности вполне адекватно описывают массовые явления случайного характера, а именно такого рода явления и изучает обычно социолог.

Чаще всего, говоря о статистичности социальных закономерностей, исследователи имеют в виду законы развития больших социальных групп и общества в целом. При этом подобная статистичность обычно рассматривается в контексте анализа известной дилеммы о соотношении общих закономерностей развития общества и свободой воли отдельного человека.

Проблема уникального и среднего.

Нахождение разного рода статистических закономерностей является привычным делом каждого социолога, проводящего эмпирическое исследование. Но нам представляется некорректным, когда статистический подход связывается только с большими группами или обществом в целом.

Статистические модели могут использоваться и при попытке "понять" отдельного человека, и при изучении разного рода групп людей, в том числе общества в целом.

Адекватно отражающими суть статистического подхода при изучении отдельного человека являются то, что актуальность для социологии изучения статистических закономерностей аргументируется посредством рассмотрения детерминированной и стохастической (вероятностной) составляющей в психологии человека, анализа механизма выполнения эмоциональными формами психологической деятельности человека роли стохастических регуляторов поведения.

Анализ данных позволяет находить статистические закономерности. Этим определяется его важность для социолога, но социолога не могут не интересовать и многие такие явления, которые не носят статистического характера по несколько иным причинам.

Интересуясь лишь статистическими закономерностями, мы игнорируем "аномалии", отступления от средней зависимости, что вряд ли можно считать допустимым. Анализ "аномалий" предусматривается грамотным использованием традиционных статистических методов: любой статистический пакет предусматривает выдачу пользователю т.н. резко

отклоняющихся наблюдений. Но такие специфичные объекты могут интересовать социолога не только как некоторые "огрехи" найденной статистической закономерности. Вряд ли следует бездумно выбрасывать соответствующие объекты из дальнейшего рассмотрения. Иногда анализ мнения такого "отклоняющегося" респондента может дать больше, чем выявление многих статистических закономерностей.

Изучение фактов, не укладывающихся ни в какие статистические закономерности, анализ случайных флуктуаций, не выражающихся в статистически значимых характеристиках, может стать целью исследования. Исследователь может стремиться найти такие "возмущения" в общественной жизни, такие ее "переломные" точки системы, которые свидетельствуют либо о ее разрушении, либо о зарождении новой системы.

Естественно, что при такой постановке задачи методы, направленные на поиск "средних" закономерностей, скрывающихся за наблюдаемыми фактами, т.е. статистические методы, перестают играть главенствующую роль.

Поиск уникальных точек вообще может не ассоциироваться с поиском закономерности.

Социология "между двух стульев". Все выявляемые закономерности слишком приблизительно отражают то, что интересует исследователя. Поэтому для социолога очень остро стоит вопрос о постоянном неформальном изучении отдельных объектов. В частности, актуальным является поиск объектов, не похожих на других, уникальных в своем роде. И существуют методы, позволяющие это делать.

Вероятно, для социолога наиболее важными методами, позволяющими находить и изучать уникальные точки рассматриваемой системы, являются т.н. мягкие методы общения с респондентами и анализа полученной от них информации. В литературе нет установившейся традиции по поводу трактовки этого термина. К числу мягких методов

опроса относят, биографический метод, разные виды не формализованного интервью – глубинное, фокусированное

Ясно, что мягкие методы действительно дают возможность изучать уникальные настроения отдельных респондентов и тем самым подводить исследователя к обнаружению "точек перелома" системы. Однако следует заметить, что они же могут использоваться и для прямо противоположных целей - для поиска закономерностей "в среднем".

Математическая статистика как основа анализа социологических данных: признак, частота встречаемости его значения, частотное распределение, статистическая закономерность в социологии.

Основными объектами изучения для математической статистики являются случайные величины. Случайными величинами в социологии служат признаки.

Для каждой совокупности значений случайной величины должна быть определена вероятность того, что, обследуя респондентов, социолог встретит значение из этой совокупности.

Напомним, что вероятностью события называют некоторую числовую характеристику степени возможности его появления в определенных, могущих повторяться неограниченное число раз, условиях. Выше в качестве события указывался выбор респондента. О вероятности этого события говорить не будем (поскольку такая вероятность связана с правилами построения выборки, которые мы не затрагиваем). В интересующем нас случае тот факт, что случайная величина приобретает некоторое значение, сам рассматривается как случайное событие. И именно задание соответствующих вероятностей сопрягается с определением случайной величины. Условия же реализации нашего случайного события – это условия, определяющие отбор респондента.

Совокупность вероятностей встречаемости значений рассматриваемой случайной величины называется отвечающим ей распределением

вероятностей, или просто ее распределением. Функция, задающая для определенных наборов значений случайной величины отвечающую им вероятность, называется функцией распределения этой случайной величины. Задать случайную величину, по существу, и означает задать соответствующее вероятностное распределение.

На практике часто функция плотности вероятности, определяющая, грубо говоря, вероятность встречаемости каждого значения случайной величины.

Саму вероятность исследователь никогда не наблюдает, в принципе не может измерить. Это – продукт нашего мышления, абстракция. Вместо вероятности исследователь обычно имеет дело с ее выборочной оценкой – относительной частотой встречаемости соответствующего события.

Математическая статистика позволяет находить широкий круг статистических закономерностей. Любая из них является некоторым набором параметров вероятностных распределений рассматриваемых случайных величин. Такого рода характеристиками являются, к примеру, разные меры средней тенденции, разброса значений случайных величин, связи между признаками и т.д.

Так как исследователь изначально имеет дело лишь с частотами, а не с соответствующими вероятностями, то фактически исходные случайные величины предстают перед ним в весьма приближенном виде. Вид закономерности, найденной для выборки, вообще говоря, будет отличаться от вида ее для генеральной совокупности. Важную роль должна играть оценка подобного различия, поскольку нас, вообще говоря, интересуют закономерности, свойственные генеральной совокупности, хотя на практике мы и имеем дело лишь с выборкой. Именно такую оценку мы и сможем сделать, пользуясь положениями математической статистики.

Фактически социолог пользуется случайными величинами, не употребляя, правда, соответствующего термина. В своей специфичной ситуации исследователь использует термин "признак" или "совокупность

признаков". Обращение социолога к математической статистике по существу начинается со статистической трактовки значений используемых признаков.

В качестве случайной величины перед социологом выступает признак (набор признаков), вместо вероятностей значений случайной величины исследователь имеет дело с относительной частотой встречаемости значений признака, вместо вероятностного распределения – с частотным, вместо параметров распределения – с отвечающими им статистиками. Рассчитав интересующие его статистики, он стремится обобщить результаты на генеральную совокупность.

Итак, основной объект, изучаемый математической статистикой, – случайная величина – является основным объектом изучения и для эмпирической социологии. Основные задачи, решаемые математической статистикой служат таковыми и для социолога, занимающегося изучением собранных эмпирических данных. Но, к сожалению, оказывается, что непосредственное применение математической статистики в социологии, зачастую, бывает весьма проблематично. Условия, предполагаемые строгими теоремами математической статистики, отнюдь не всегда выполняются на практике. И тогда вместо строгой математической статистики на сцену выступает не совсем строгое ее "приближение" - анализ данных.

При использовании математической статистики могут возникать следующие сложности:

1. Формирование выборки (иначе нельзя будет экстраполировать на генеральную совокупность);
2. Не все методы математической статистики можно использовать;
3. Отсутствие строгих алгоритмов решения многих практических задач.

Дескриптивная и индуктивная задачи анализа данных в социологии:

1. *Дескриптивная* (описательная) – поиск статистических закономерностей для выборки, описание и поиск взаимосвязей;

2. *Индуктивная* – обобщение полученных результатов на генеральной совокупности: статистическая оценка параметров; проверка статистических гипотез.