

# ИНСТРУКЦИЯ ПО ВЕДЕНИЮ НАУЧНОГО ИССЛЕДОВАНИЯ

Логическая Статистика





Книга: Инструкция по ведению научного исследования

Лекция: Логическая Статистика

Автор лекции: Альмира Абилова

В то время как цель описательной статистики состоит в том, чтобы описать и обобщить характеристики вашей выборки, цель логической статистики – сделать выводы, которые выходят за пределы ваших непосредственных данных / выборки. Например, логическая статистика может использоваться для проверки различных гипотез о взаимосвязи между различными переменными или, что еще важнее, для оценки характеристик населения на основе данных выборки. Другими словами, логическая статистика позволяет обобщать результаты.

## Вопросы о количественных данных с использованием логической статистики

### *Как участники моего исследования сравниваются со всем населением?*

Эти типы вопросов сравнивают выборку с населением. Предположим, вы проводите исследование студентов, проходящих определенный университетский курс. Вы можете спросить, являются ли процентные доли мужчин или женщин в вашей выборке, или их средний возраст, или их интересы статистически сходными со студентами университетов по всей стране. Чтобы ответить на такие вопросы, вам потребуется доступ к данным о всех студентах страны.

### *Есть ли различия между двумя или более группами респондентов?*

Вопросы, в которых сравниваются две или более групп, очень распространены и часто называются «между субъектами». Если придерживаться темы студентов, вы можете спросить, могут ли студенты мужского и женского пола иметь одинаковые интересы; или могут ли студенты разных этнических групп проходить разные курсы; или имеют ли студенты, обучающиеся в разных кампусах, разные средние баллы.

### *Изменились ли мои респонденты с течением времени?*

Вопросы, связанные с данными до и после с одной и той же группой респондентов или респондентов со сходными характеристиками. Примером такого типа вопросов может быть «Изменились ли привычки студентов последнего года обучения по сравнению с первым годом обучения в университете?»

### *Есть ли связь между двумя или более переменными?*

Вопросов, используются для нахождения отношения и корреляции, или причины и следствия. Примерами вопросов взаимоотношений являются: «Есть ли связь между временем, потраченным на обучение, и удовлетворенностью университетом?» И «Существует ли корреляция между полом студентов и внешкольными занятиями, в которые они вовлечены?» Вопросы, ищущие причину и следствие, делятся на зависимые и независимые переменные. Например, «Зависит ли удовлетворенность от времени обучения?» И «Зависит ли стресс от пройденного курса?» Причинно-следственные связи могут также искать более одной независимой переменной для объяснения вариации в зависимой переменной (многомерный анализ). Например, «Зависит ли удовлетворение от обучения в университете от сочетания продолжительности обучения, возраста и карьерных устремлений?»

## Статистическая значимость

Как уже говорилось, цель логической статистики состоит в том, чтобы обобщить результаты вашей выборки на все население. Но поскольку данные все еще ограничены выборкой, невозможно сказать со 100% -ной уверенностью, что любое обобщение не имеет вероятности ошибки. Однако логическая статистика позволяет вам оценить вероятность того, что наблюдаемая разница больше, чем случайность или случайное обнаружение, и на самом деле является статистически значимой.

Статистическая значимость, как правило, относится к «р-значению», которое оценивает фактическую вероятность того, что ваши выводы более чем совпадение. Обычные значения  $p$  равны 0,05, 0,01 и 0,001, что говорит о том, что вероятность того, что ваши результаты произошли случайно, составляет 5 из 100, 1 из 100 или 1 из 1000 соответственно. По сути, чем ниже значение  $p$ , тем увереннее могут быть исследователи в том, что полученные данные являются подлинными. Имейте в виду, что исследователи обычно не принимают результаты, у которых значение  $p$  больше 0,05, поскольку вероятность того, что результаты совпадают или вызваны ошибкой выборки, слишком велика.

## Понимание и выбор правильного статистического теста



Существует множество статистических тестов, которые могут помочь вам ответить на типы вопросов, которые мы сегодня обсудили. И хотя важно понимать основную логику, которая управляет наиболее распространенными статистическими тестами, такие программы, как SPSS и SAS, способны запускать такие тесты без необходимости разбираться в технических особенностях их математических операций. Однако проблема определения того, какой тест подходит для вашей конкретной задачи, все еще остается. К счастью, вы можете обратиться к ряду тестовых селекторов, доступных в настоящее время в Интернете (см. Тестовый селектор Билла Трохима на сайте [www.socialresearchmethods.net/kb/index.htm](http://www.socialresearchmethods.net/kb/index.htm)) и с помощью таких программ, как MODSTAT и SPSS.

Однофакторный анализ – статистический анализ одной переменной за раз. Он состоит из таких мер, как центральная тенденция, дисперсия и распределение. Хотя одномерный анализ не рассматривает корреляцию, причину и следствие или моделирование, этот тип анализа является краеугольным камнем описательного исследования и является важным предварительным этапом для всех типов и уровней статистического анализа.

Двусторонний анализ – статистический анализ взаимосвязи между двумя переменными. Двусторонний анализ оценивает отношения между двумя переменными; например, существует ли связь между уровнем образования и привычками просмотра телевидения, или между полом и доходом. Диапазон тестов, которые используются для изучения таких взаимосвязей, довольно широк и варьируется в зависимости от типа переменной. Наиболее распространенными тестами являются кросс-табуляции (Хи-квадрат, используется для двух номинальных переменных), ANOVA (используется для одной номинальной и одной переменной отношения) и корреляции (используется для двух переменных отношения).

Многофакторный анализ – статистический анализ, который исследует взаимосвязь между тремя и более переменными и позволяет исследователям искать причину и следствие, строить модели и проверять теории. С помощью многомерного анализа исследователи могут не только исследовать, зависит ли зависимая переменная от двух или более независимых переменных (т. е. Доход зависит как от пола, так и от уровня образования), но также признать связь между зависимыми переменными (т. е. Связь между полом и образованием достижение). Некоторыми из методов, используемых в многомерном анализе, являются факторный анализ, разработка, моделирование структурных уравнений, MANOVA, множественная регрессия, каноническая корреляция и анализ пути.

## Презентация количественных данных

Когда дело доходит до презентации количественных данных, может возникнуть реальное искушение сделать графики, диаграммы и таблицы для каждой переменной в вашем исследовании. Но важно противостоять этому искушению и определить, что является самым важным в вашей работе. Ваши результаты должны рассказать историю, связанную с вашими целями, задачами и вопросами исследования.

Теперь, когда речь заходит о том, как ваши данные должны быть представлены, я думаю, что есть одно золотое правило: ваша презентация не должна быть сложной работой для читателя. Когда дело касается статистики, ваши данные не должны быть трудными для расшифровки. Вам не нужно быть статистиком, чтобы понять это. Ваша задача состоит в том, чтобы представить ваши данные в графической и устной форме, чтобы их значения были понятны. Любые графики и таблицы, которые вы представляете, должны облегчить задачу для читателя. Таким образом, хотя вам необходимо включить адекватную информацию, вы не хотите вдаваться в информационную перегрузку.

## Использование графиков

Как говорится, картинка стоит тысячи слов, поэтому хороший график имеет большое значение в презентации ваших результатов. Для того, чтобы ознакомить вас с наиболее часто используемыми графиками в количественном анализе, обратите внимание на этот набор данных с четырьмя переменными в SPSS, используя гипотетический пример данных опроса 60 учеников (30 мужчин и 30 женщин), которые собираются окончить среднюю школу.



Пол  
1 = женский  
2 = мужской

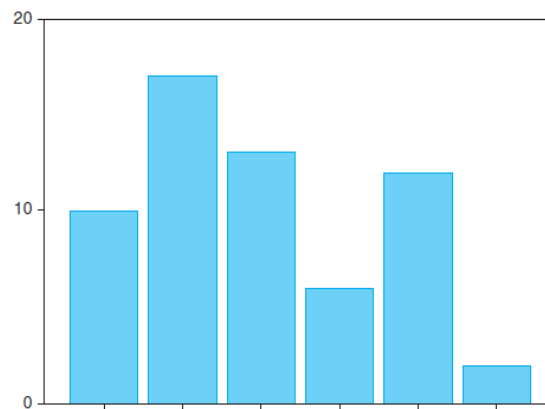
Средний балл (из 4)

Планы после  
выпускного  
1 = профессионально-  
техническое обучение  
2 = университет  
3 = рабочая сила  
4 = путешествие  
5 = не определились  
6 = другое

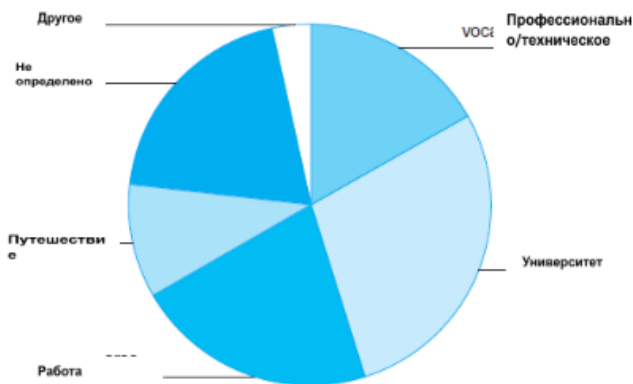
Важность университета  
1 = совсем неважно, до  
7 = необходимо

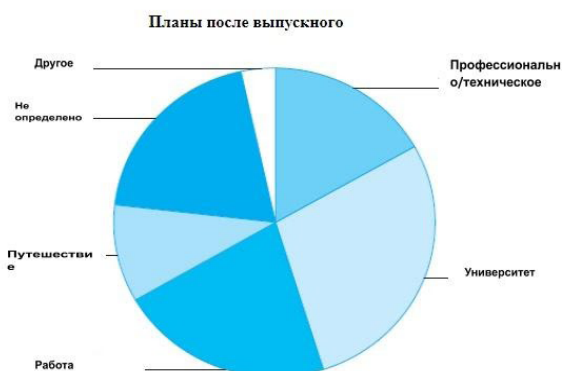
1	3.10	2	7
2	2.60	3	2
1	1.30	3	2
2	3.60	5	6
2	1.50	1	1

«Планы после окончания» – это номинальная переменная, поэтому гистограммы и круговые диаграммы

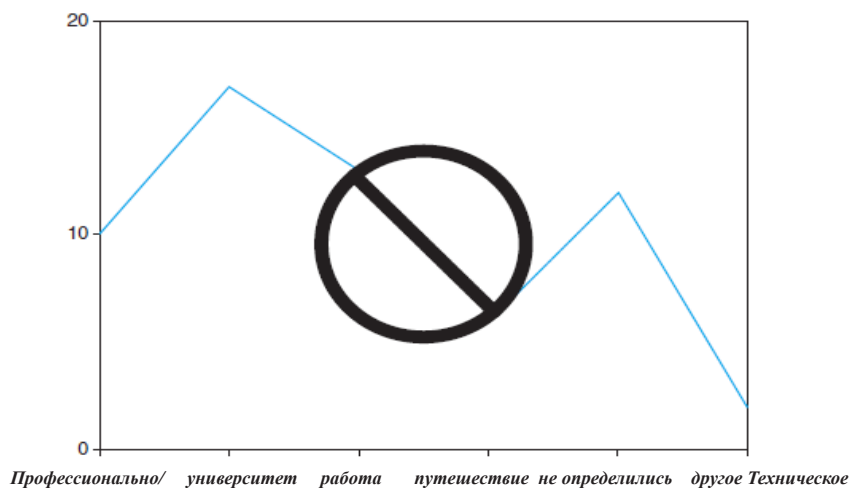


### Планы после выпускного





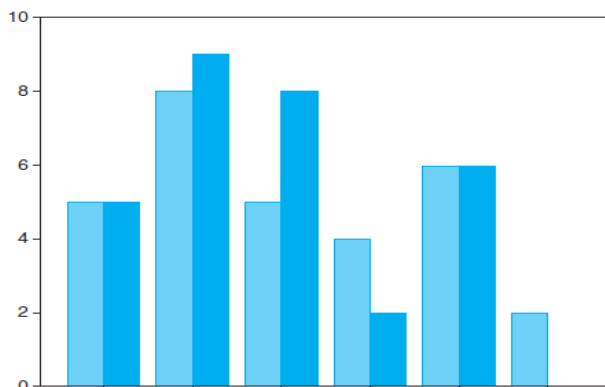
(Рисунок 1 и 2), как правило, работают хорошо. Однако линейный график не работает (Рисунок 3),



потому что он лучше подходит для отображения изменений во времени – это не то, что мы пытаемся изучить.

Линейный график лучше подходит для данных, которые показывают что-то вроде изменения среднего балла учащегося (GPA) в течение их школьной карьеры (GPA будет находиться на одной оси, а год обучения – на другой).

### Кластерная гистограмма

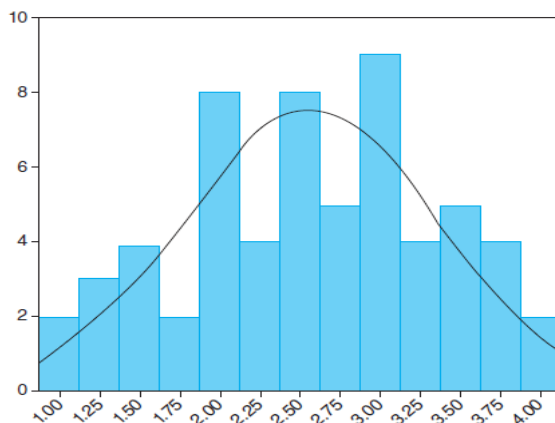




(Рисунок 4) позволяет сравнивать распределение номинальной переменной для двух или более групп. В этом примере «планы после выпуска» сравниваются по полу.

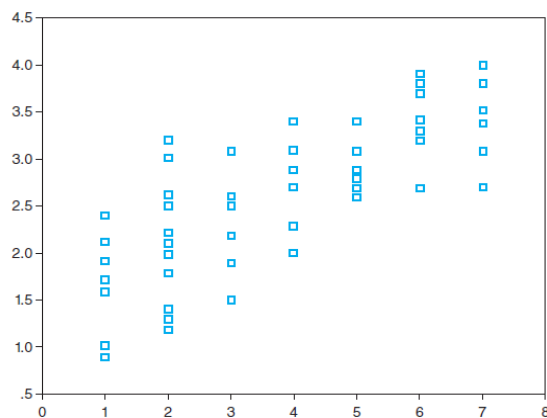
### *Изучение Среднего балла*

## Гистограмма



(Рисунок 5) подходит для отображения распределения интервальных и относительных данных. Здесь мы можем увидеть распределение Средний балл (относительная переменная), полное среднее значение, стандартное отклонение и график нормальной кривой.

### *Изучение взаимосвязи между Средним баллом и «Важностью университета»*



В этом примере график рассеяния используется для построения относительной переменной (Средний балл) по отношению к порядковой переменной (важность университета). График показывает положительную корреляцию между Средний балл и воспринимаемой важностью университета: то есть, хотя отношения не идеальны, так как GPA растет, так же, как и воспринимаемая важность. Чтобы выяснить точный уровень ассоциации, вы можете запустить корреляцию Пирсона, которая показывает корреляцию 0,773, значительную на уровне 0,01 (вы на 99% уверены, что модель не случайна).

## Создание графиков

Как же нужно создавать эти типы графиков? Есть два способа сделать это, и в обоих нужно будет посидеть перед компьютером. Во-первых, попросите кого-нибудь из знающих показать вам, что делать, а затем попрактикуйтесь. Второй – сесть самому и поиграть со статистическими программами. В большинстве из них есть отличные интерактивные онлайн-учебники, которые помогут вам пройти процесс от А до Я. Хотя чтение поможет вам познакомиться с программами, лучший способ научиться делать графики – это их сделать.



Книга: Инструкция по ведению научного исследования

Лекция: Логическая Статистика

Автор лекции: Альмира Абилова

## Создание таблиц

Идея в том, чтобы представить сложность с простотой. Таблицы, как правило, более сложны, чем графики, потому что они пытаются обобщить взаимосвязь между рядом переменных (многомерный анализ). А поскольку эта сводка может включать несколько статистических операций, программы статистики редко выкладывают таблицы в готовом виде. Теперь вы можете использовать автоматический генератор таблиц (вы можете найти их в Интернете – одним из примеров является TableMaker по адресу [www.bagism.com/tablemaker](http://www.bagism.com/tablemaker)), и это, безусловно, может помочь, но есть вероятность, что вам все равно нужно будет выполнить некоторые манипуляции самому. Задача здесь – найти баланс между достаточным и слишком большим количеством информации. Ваши таблицы не должны:

1. Давать вашим читателям головную боль;
2. Усыплять их; или
3. Запутать их еще больше.

Несмотря на то, что ваши таблицы не будут стоять в одиночестве (вам нужно будет разъяснить их значение для ваших читателей), они должны быть чем-то, с чем неспециалист может поработать и извлечь уроки. Ваши таблицы будут наиболее эффективными, если: вы сможете предоставить четкую и адекватную информацию; вы не предполагаете, что ваш читатель обладает всеми необходимыми знаниями; и вы их составляете максимально просто.