

К. А. Б Р А У Н Л И

СТАТИСТИЧЕСКИЕ
ИССЛЕДОВАНИЯ
В ПРОИЗВОДСТВЕ

И*Л

*Издательство
иностранной
литературы*

*

INDUSTRIAL
EXPERIMENTATION

by
K. A. BROWNLEE

1947

К. А. БРАУНЛИ

СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ В ПРОИЗВОДСТВЕ

ПЕРЕВОД С АНГЛИЙСКОГО

В. А. ГОВОРКОВА

ПОД РЕДАКЦИЕЙ АКАДЕМИКА

А. Н. КОЛМОГорова

1949

ИЗДАТЕЛЬСТВО

ИНОСТРАННОЙ ЛИТЕРАТУРЫ

Москва

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Небольшая книга Браунли дает хорошее представление о технике применения методов математической статистики при промышленных испытаниях в Англии. С нашей точки зрения книга эта чрезмерно рецептурна и догматична. Ценность ее по преимуществу в хорошо подобранных примерах, большинство которых вполне конкретны и реальны. На постепенно усложняющихся примерах автор дает возможность читателю, не привыкшему к языку общих формул, усвоить технику расчетов. Многие примеры удачны еще и в том отношении, что показывают пограничные случаи, когда значимость тех или иных различий действительно не очевидна с первого взгляда и тем не менее устанавливается вычислением, или, наоборот, когда какой-либо выбор кажется неискушенному читателю вполне очевидным, подсчет же дает для него лишь весьма скромный „уровень значимости“.

Наиболее полное теоретическое освещение всех приемов, рекомендуемых Браунли, содержится в книге Г. Крамера „Математические методы статистики“, русский перевод которой издан Государственным издательством иностранной литературы. Однако книга Крамера далеко не элементарна. Менее подготовленному читателю можно рекомендовать прочесть соответствующие разделы из „Элементарного курса математической статистики“ В. И. Романовского (Госпланиздат, 1938). Для того чтобы помочь читателю связать чисто рецептурное изложение Браунли с общими теоретическими курсами, к переводу приложены краткие комментарии редактора перевода.

Особенно большое внимание Браунли уделяет *дисперсионному анализу*. К сожалению, на русском языке нет достаточно подробного изложения этой важной теории, а достаточно полного *критического* ее изложения вообще не существует. Как

указано в комментариях редактора перевода, автор несколько увлекается этим методом, и из его изложения можно получить преувеличенное впечатление о его силе. Тем не менее, популярное изложение приемов дисперсионного анализа с большим числом факторов и учетом их взаимодействия, данное Браунли, весьма ценно, так как именно в таких сложных случаях для предварительной ориентировки и выделения наиболее существенных факторов и их взаимодействий дисперсионный анализ особенно полезен.

А. Колмогоров

ПРЕДИСЛОВИЕ

Настоящая монография основана на инструкции, выпущенной ранее Директоратом артиллерийских заводов взрывчатых веществ для лиц, связанных с испытаниями в опытном и общезаводском масштабе по химическим производственным процессам на этих заводах.

Опыт проведения многих испытаний доказал желательность проверки результатов с помощью критериев значимости. Однако не было описания соответствующих методов, предназначенного для тех, кто должен применять критерии на практике, не обладая достаточными познаниями в их теоретических обоснованиях. Настоящая книга и представляет собой попытку дать такого рода описание.

Очевидно, что для удобного и экономичного проведения проверок с помощью критериев значимости испытания должны планироваться надлежащим образом. Мы считаем, что излагаемые методы проверок должны стать таким же стандартным инструментом заводского экспериментатора, каким химические весы служат экспериментатору в лаборатории.

При проведении заводского испытания не ставится вопрос о выборе между применением статистического расчета с приложением надлежащих критериев значимости и обычными методами; речь идет о выборе правильного метода. Даже простейшее испытание требует проверки значимости его результатов.

Нельзя не осудить строго случаи, когда консультируются со статистиком только после окончания испытания и спрашивают его: „Что можно получить из результатов?“. Важно, чтобы испытание проводилось в удобной для анализа форме; этого можно достигнуть, вообще говоря, только путем проектирования испытания при консультации статистика или с должным учетом привлекаемых статистических принципов.

Настоящая монография выпускается поэтому как руководство по планированию испытаний в заводском масштабе и по объяснению их результатов; мы надеемся, что описанные здесь методы станут частью повседневной работы тех, кто проводит испытания.

Январь, 1946 г.

Р. Боуден,
директор заводов взрывчатых веществ
Министерства снабжения

ИЗ ПРЕДИСЛОВИЯ АВТОРА

Настоящая монография предназначена для лиц, ведущих исследовательскую работу, в качестве руководства по современным статистическим методам как в отношении применения критериев значимости для получения надежных выводов из данных испытаний, так и в отношении применения статистических расчетов в целях получения наибольшей точности с наименьшими затратами.

Предмет рассматривается исключительно с практической точки зрения. Теория сведена к минимуму, а математический аппарат — к простой арифметике. Каждый рассмотренный статистический метод подробно иллюстрирован практическими примерами, чтобы показать точно, в чем его суть. Некоторые более сложные методы могут показаться довольно трудно выполнимыми: к ним настоятельно рекомендуется подходить путем овладения более простыми приемами. К последним выработается тем самым привычка и доверие, и читатель увидит, что более сложные методы являются сравнительно простым развитием знакомых ему простых приемов.

Применение статистических методов требует определенного количества вычислений. В основном это — суммирование чисел и их квадратов, легко выполнимое с помощью таблиц квадратов (см., например, Барлоу „Таблицы квадратов и др.“) и простого арифмометра. Для проведения большого количества корреляционных расчетов желательна вычислительная машина, производящая умножение и деление, однако эти машины дороги, и при недостаточной загрузке их приобретение не оправдывается. Как бы то ни было, та или иная форма механизации вычислений существенно важна.

Теоретическим основанием этой книги служили методы, разработанные в основном профессором Р. А. Фишером в его трудах: „Статистические методы для исследовательских работников“ и „Проектирование испытаний“*.

Таблицы III—VI перепечатаны из книги проф. Р. А. Фишера и д-ра Фрэнка А. Ийтс „Статистические таблицы для исследований в биологии, сельском хозяйстве и медицине“**.

Некоторые факторы для диаграмм контроля качества перепечатаны из Британского стандарта 600 R „Диаграммы контроля качества“.

Все приведенные примеры являются результатами экспериментальных и исследовательских работ на заводах взрывчатых веществ.

* R. A. Fisher, „Statistical Methods for Research Workers“ и „The Design of Experiments“.

** R. A. Fisher & F. A. Yates, „Statistical Tables for Biological, Agricultural and Medical Research“.

ГЛАВА I

ВВЕДЕНИЕ

(А) ОШИБКИ ПРИ НАБЛЮДЕНИЯХ

Результаты экспериментальной исследовательской работы всегда содержат ошибки наблюдений. Ни один исследователь не будет всерьез это оспаривать. Тем более замечательно, что лишь в последнее время становится общепризнанной необходимость оценивать ошибки результатов каждого исследования для суждения об их достоверности; соответствующая операция известна в статистике под названием „проверка значимости“. Можно прямо сказать, что отсутствие этих проверок значимости является основной причиной неудовлетворительного качества многих исследований, произведенных в промышленности; их ошибки велики и притом больше, чем предполагал исследователь, который относил ряд обнаруженных эффектов за счет влияния варьированных им факторов, тогда как на самом деле они возникли исключительно из-за случайных колебаний, вызванных ошибками наблюдений. Если бы были произведены строгие проверки значимости, многие из сделанных утверждений оказались бы совершенно недостаточно обоснованными; некоторые оказались бы просто неверными. (Заметим, что из недостаточной обоснованности того или иного утверждения не вытекает, конечно, что само утверждение неверно; может оказаться, что автор был прав в своем рискованном утверждении. Однако директор предприятия, рассматривая рекомендацию своего консультанта, должен знать, основана ли она на научно доказанном факте или на субъективной догадке.)

(В) КЛАССИЧЕСКИЕ И ПРОИЗВОДСТВЕННЫЕ ИСПЫТАНИЯ

В вопросе оценки ошибки возникает одно из принципиальных различий между исследованиями в лаборатории и в производстве, т. е. в заводском масштабе. Исследователь, работающий в лаборатории, находится обычно в более благоприятном положении, будучи в состоянии осуществлять точный контроль над всеми независимыми переменными; его материалы могут быть наивысшего качества, а измерительная аппаратура — высшей точности; его ошибки обычно малы.

Напротив, экспериментатор на производстве часто не может добиться полного контроля над всеми переменными, так как

масштаб испытания зачастую столь обширен, что для такого контроля понадобилась бы армия инспекторов. Достижение хорошего контроля может оказаться также и технически трудным; так, например, в хорошо оборудованной лаборатории легко поддерживать температуры в весьма узких пределах с помощью обычной термостатической аппаратуры. При испытаниях же в заводском масштабе это обычно совершенно неосуществимо. В любых лабораторных условиях достаточно легко задержать процесс; с другой стороны, задержка работы поглотительной башни высотой 15 м (в производстве серной кислоты) обошлась бы абсурдно дорого.

Кроме того, для экспериментатора на производстве важно проводить испытания так, чтобы возможно меньше нарушать нормальный ход производства. Может оказаться, что для осуществления контроля над независимыми переменными потребовалось бы столько времени и внимания к каждой партии продукции, что производительность завода заметно сократилась бы. Или же для выяснения влияния какой-нибудь независимой переменной может оказаться желательным менять эту переменную в широких пределах, чтобы получить достаточно далекие друг от друга показания; это могло бы повести к выпуску большого количества продукции, не удовлетворяющей стандарту, т. е. брака. В таком случае приходится работать в узком диапазоне изменения независимой переменной; это значит, что эффект, подлежащий наблюдению, будет намного уменьшен и может оказаться размытым погрешностями наблюдений.

Поэтому результаты испытаний в производстве по сравнению с лабораторными обычно содержат гораздо большие погрешности. При этих условиях бывает трудно решить, является ли результат данного индивидуального испытания истинным или ошибочным: возникает необходимость применять статистические критерии значимости. Но для получения значимого результата (т. е. такого, в котором наблюдаемый эффект значительно больше величины ошибки в его определении), если не применять специальных статистических приемов, требуется большое количество испытаний, а это в заводских условиях крайне нежелательно из-за дороговизны.

Наконец, быть может, уместно заметить, что, по опыту автора, даже в лабораторных условиях после объективного подсчета с помощью точных статистических методов часто оказывается, что результаты содержат гораздо большие ошибки, чем думает и утверждает экспериментатор.

Из сказанного выше следует, что экспериментатору на производстве приходится преодолевать ряд дополнительных трудностей, с которыми не сталкивается экспериментатор в лаборатории; поэтому экспериментатор на производстве нуждается

в особых дополнительных приемах исследования. Эти приемы разработаны в последние годы применительно к сельскому хозяйству рядом исследователей (в частности, Р. Фишером и его коллегами на Ротэмстэдской экспериментальной станции), которым приходилось считаться с погодой и изменениями плодородности почв среди целого ряда факторов, находившихся вне их контроля.

(С) ПОВТОРЕНИЯ ИСПЫТАНИЯ

Вообще, существуют два альтернативных метода уменьшения слишком большой ошибки в оценке какого-либо эффекта. Первый метод заключается в усовершенствовании техники эксперимента, например, в применении более прецизионного термостатического оборудования, во взвешивании на лучших весах, в принятии более строгих мер предосторожности против испарения и т. п. Другой метод, который в производственных условиях часто является единственно возможным, состоит в многократном повторении эксперимента и осреднении результатов. К сожалению, этот метод не очень продуктивен, так как ошибка среднего значения обратно пропорциональна квадратному корню из числа наблюдений. Беря среднее четырех наблюдений, получаем ошибку среднего в два раза меньше ошибки одиночного наблюдения; среднее шестнадцати наблюдений имеет ошибку в четыре раза меньшую, чем ошибка одиночного наблюдения.

Эксперименты в лабораторных условиях обходятся сравнительно дешево; поэтому бывает возможным получить гораздо большее число наблюдений, чем строго необходимо для вывода заключения с заданным уровнем значимости. Благодаря этому, точной проверки значимости не требуется. Работа же в производственном масштабе обходится столь дорого и требует столь больших затрат сырья, энергии и т. п., что проведение числа наблюдений, большего чем требуется для поставленной цели — вывода того или иного заключения с заданным уровнем значимости, было бы непростительной расточительностью.

(D) ПРОЕКТИРОВАНИЕ ЭКСПЕРИМЕНТОВ. СЛУЧАЙНЫЕ БЛОКИ

До сих пор теория статистики была упомянута лишь в связи с теорией ошибок, для которой она, конечно, является основой. Однако теория статистики может оказать большую помощь и в проектировании порядка проведения экспериментов, имеющих целью получить неискаженную оценку ошибки и свести ошибку к минимуму.

Начнем с одного из простейших понятий, с так называемого «случайного блока».

Пусть желают сравнить каких-нибудь четыре способа „обработки“. Слово „обработка“ имеет здесь обобщенный смысл; например, могут рассматриваться четыре рода удобрений для посева пшеницы, или четыре различных значения температуры в процессе нитрации, или четыре значения концентрации пластификатора в пластмассе и т. п. Предположим, затем, что материал поступает на обработку партиями такого размера, что возможно провести только четыре эксперимента по каждой партии. Предположим, что поступающие друг за другом партии материала могут заметно различаться по качеству. Назовем отдельные партии „блоками“.

В упомянутом примере с посевом пшеницы можно рассматривать как „блок“ четыре соседних опытных участка земли; они, вероятно, более сходны друг с другом по урожайности, чем другие участки, принадлежащие к другим более удаленным блокам, находящимся, возможно, в другом поле. В примере процесса нитрации понятие „блока“ можно отнести к баку, наполненному хорошо размешанной кислотой, и партии материала, подлежащего нитрации, тщательно размешанного до достижения однородности. Затем предположим, что каждое наблюдение будет производиться четыре раза.

Обозначим способы обработки буквами *A*, *B*, *C* и *D* и блоки цифрами 1, 2, 3 и 4. Эксперимент можно провести, например, в порядке, приведенном в таблице 1.1.

ТАБЛИЦА 1.1.

Блок	Обработка			
1	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
2	<i>B</i>	<i>B</i>	<i>B</i>	<i>B</i>
3	<i>C</i>	<i>C</i>	<i>C</i>	<i>C</i>
4	<i>D</i>	<i>D</i>	<i>D</i>	<i>D</i>

Достаточно взглянуть на таблицу 1.1, чтобы убедиться, что такой порядок проведения эксперимента был бы нелепым, так как любые различия в результатах могут быть отнесены в равной степени как за счет различий между блоками, так и за счет различий между способами обработки.

Несколько лучше было бы распределить способы обработки в чисто случайном порядке. Можно заготовить 16 карточек,

написав на четырех из них *A*, на других четырех *B* и т. д., положить их в шляпу и затем вынимать по одной и раскладывать по порядку, слева направо по блокам, идя сверху вниз. Результат одной из таких операций показан в таблице 1.2.

ТАБЛИЦА 1.2.

Блок	Обработка			
1	<i>D</i>	<i>B</i>	<i>B</i>	<i>D</i>
2	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>
3	<i>B</i>	<i>A</i>	<i>C</i>	<i>A</i>
4	<i>C</i>	<i>C</i>	<i>D</i>	<i>A</i>

Если карточки были перемешаны как следует, средние четырех результатов по каждому способу обработки будут правильны со статистической точки зрения, но величина ошибки в каждом среднем будет преувеличена, поскольку она будет охватывать и различия между блоками. Поэтому эксперимент будет не так точен, как он мог бы быть.

Очевидно, будет лучше распределить способы обработки по одному на каждый блок, как показано в таблице 1.3.

ТАБЛИЦА 1.3

Блок	Обработка			
1	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
2	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
3	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
4	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>

Различие между средними по четырем наблюдениям *A* и по четырем наблюдениям *B* будет теперь совершенно независимо от различий между блоками.

Более удовлетворительно, однако, было бы наметить порядок применения способов обработки по каждому блоку различным, в разбивку, а не систематическим, в порядке *A, B, C, D*, как указано в таблице 1.3. В самом деле, если, например, существует тенденция получения высокого результата в первом эксперименте каждого блока, то получится высокий итоговый результат по способу обработки *A*, что неверно.

Этого можно избежать, распределив порядок проведения экспериментов, как показано в таблице 1.4.

ТАБЛИЦА 1.4.

Блок	Обработка			
	<i>D</i>	<i>B</i>	<i>A</i>	<i>C</i>
1	<i>D</i>	<i>B</i>	<i>A</i>	<i>C</i>
2	<i>B</i>	<i>C</i>	<i>A</i>	<i>D</i>
3	<i>C</i>	<i>B</i>	<i>A</i>	<i>D</i>
4	<i>A</i>	<i>C</i>	<i>D</i>	<i>B</i>

Мы пришли теперь к понятию „случайного блока“. Все подлежащие сравнению способы обработки распределены по блокам, по равному числу в каждом, причем блок достаточно велик, чтобы охватить по крайней мере по одному разу каждый способ обработки. Повторения исследований осуществляются по различным блокам, порядок проведения везде сделан в разбивку. Разумеется, если объем блока таков, что в пределах отдельного блока можно сделать несколько повторений, весь эксперимент и следует провести именно так; при повторениях же в различных блоках ничего не выигрывается*.

Анализ результатов такого эксперимента, с проверкой уровня значимости результатов, рассмотрен в главе VII, (E).

(E) ЛАТИНСКИЙ КВАДРАТ

Рассмотренный в предыдущем разделе гипотетический эксперимент спроектирован достаточно строго в том отношении, что он не приведет к ложным заключениям; однако в некоторых случаях его можно еще усовершенствовать.

* См., впрочем, гл. XII, (E).

В качестве частного примера предположим, что в некотором многоступенном процессе партии материала по одной тонне делятся на четыре подпартии, которые обрабатываются приблизительно в одно и то же время в четырех стабилизаторах; требуется сравнить эффекты четырех различных способов обработки на этой последней ступени. При четырех способах обработки и четырех подпартиях в каждой партии применим прием „случайных блоков“, с распределением четырех способов обработки по четырем стабилизаторам в случайном порядке. Можно, однако, сделать так, что каждый из способов обработки придется на каждый стабилизатор один и только один раз, так что средние результаты будут совершенно независимы от возможных различий между стабилизаторами.

Такой порядок показан в таблице 1.5, где буквенные обозначения относятся к способам обработки.

ТАБЛИЦА 1.5

Партия	Стабилизатор 1	Стабилизатор 2	Стабилизатор 3	Стабилизатор 4
1	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
2	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>
3	<i>B</i>	<i>A</i>	<i>D</i>	<i>C</i>
4	<i>D</i>	<i>C</i>	<i>B</i>	<i>A</i>

Из таблицы 1.5 видно, что четыре подпартии, относящиеся к одной и той же партии, подвергаются всем четырем способам обработки. В то же время каждый из четырех способов обработки и каждая из партий приходится на каждый из стабилизаторов только по одному разу. Возможные ошибки из-за различий между партиями и между стабилизаторами, таким образом, исключаются из средних и из величины ошибки; этим достигается получение неискаженных оценок и минимальной величины ошибки.

Подсчет результатов латинского квадрата рассматривается в главе XII, (G).

(F) СБАЛАНСИРОВАННЫЕ НЕПОЛНЫЕ БЛОКИ

Объем обычного случайного блока должен быть достаточен для охвата одного полного комплекта сравниваемых способов

обработки. Однако может случиться, что число сравниваемых способов обработки больше размера блока.

Допустим, что сырье поступает партиями такой величины, что по каждой партии можно сделать 4 эксперимента; между партиями ожидается значительная дисперсия. Если требуется сравнить 2, 3 или 4 способа обработки, эта дисперсия между партиями не вызовет беспокойства у исследователя, поскольку каждый комплект из 2, 3 или 4 способов обработки можно будет провести по одной и той же партии. Если же требуется сравнить 5 или более способов обработки, то один или более из них придется провести по другой партии, а не по той, где проводилось испытание первых четырех способов; результат сравнения будет обесценен из-за влияния дисперсии между партиями.

Обычный метод обхода этого затруднения заключается в том, что один из способов обработки выбирается в качестве „стандартного“ и включается в каждую партию. Все способы обработки сравниваются затем непосредственно со стандартным; остальные сравнения способов между собой проводятся либо прямо — для способов, попавших в один и тот же блок, либо косвенным путем, например, из сравнения способа M со стандартным в его блоке, и из сравнения способа N со стандартным в блоке, в который входит N .

Хотя этот порядок испытаний и пригоден, но ни в какой мере не является достаточно удовлетворительным, так как здесь возникают три различные ошибки в трех различных типах сравнений. Кроме того, он и не наиболее продуктивен.

Рассмотрим альтернативный порядок проведения испытаний, показанный в таблице 1.6; здесь сравниваются пять способов обработки с четырехкратными повторениями, по пяти партиям.

ТАБЛИЦА 1.6

Партия №				
1	2	3	4	5
A	A	A	A	B
B	B	B	C	C
C	C	D	D	D
D	E	E	E	E

Можно сделать ряд замечаний относительно этого порядка проведения исследований:

(а) Каждый способ обработки попадает один и только один раз в четыре партии из пяти.