**ЛЕКЦИЯ 6 ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ**

*1 Сущность и метод дисперсионного анализа*

*2 Однофакторные статистические комплексы*

## 1 Сущность и метод дисперсионного анализа

Сущность дисперсионного анализа заключается в изучении статистического влияния одного или нескольких факторов на результативный признак.

### 1.1 Результативный признак

**Результативный признак** (*Y*) – это элементарное качество или свойство объектов, изучаемое как результат влияния факторов: организованных в исследовании (*X*) и всех остальных, неорганизованных в данном исследовании (*Z*).

Результативными признаками могут быть:

* *точно измеряемые особенности объектов* (длина, ширина, рост, сила, содержание гемоглобина, артериальное давление и т. д.);
* *неточно измеряемые признаки* (густота раствора в баллах, умственные способности и т. д.);
* *комбинированные признаки* (отношение размеров тела, индексы продуктивности, средние из нескольких данных для одного объекта и т. д.);
* *качественные признаки* (масть, цвет глаз, болезнь, выздоровление, смерть и т. д.).

### 1.2 Фактор

***Фактор*** – это любое влияние, воздействие или состояние, разнообразие которых может, так или иначе, отражаться в разнообразии результативного признака.

Факторами могут быть:

* *физические влияния* (температура, влажность, радиационное излучение);
* *химические влияния* (питание, стимуляторы, мутагены, алкоголь);
* *биологические влияния* (здоровье и болезни, биостимуляторы, наследственность, талантливость, идиотизм);
* *признаки объектов* (возраст, пол, сорт, порода, национальность, ареал обитания, условия жизни);
* отдельные признаки, принимаемые за аргумент при изучении других признаков – функций (длина ног рысака как один из факторов, определяющих его резвость).

### 1.3 Градации факторов

***Градации факторов*** – это степень их действия (нулевое действие в контрольной группе), или состояние объектов изучения (пол, возраст, обученность и т. д.).

Градациями факторов могут быть:

* разная температура, влажность, разные дозы облучения, разная продолжительность физических, воздействий;
* разные питательность и состав корма, разные дозы стимуляторов и химических мутагенов, разные стадии опьянения;
* разные периоды болезни, степени таланта, разные отцы или классы отцов, разные матери или классы матерей;
* разные возраст, пол, сорт, порода;
* разные ареалы и условия жизни;
* разная величина признака, принятого за аргумент.

### 1.4 Градации комплекса

***Градации комплекса*** – это опытные группы исследования. Каждая градация комплекса соответствует одной градации фактора и включает те объекты (с их данными), которые подвергались одной степени действия фактора или находились в одном из изучаемых состояний.

Организация градаций комплекса может осуществляться разными способами: подбор опытных и контрольных групп, привлечение материалов ранее проведенных наблюдений и опытов, систематизация записей производственной отчетности.

Подбор объектов в градации из ранее намеченных генеральных совокупностей (соответствующих каждой градации) производится рандомизированно, т. е. по принципу случайной выборки, без учета развития изучаемого признака (перед отбором).

### 1.5 Дисперсионный комплекс

***Дисперсионный комплекс*** – это совокупность градаций с привлеченными для исследования данными и средними из данных по каждой градации (частные средние) и по всему комплексу (общая средняя).

Если изучается действие одного фактора, комплекс называется однофакторным, двух факторов – двухфакторным, трех и более факторов – многофакторным.

Если во все градации подбирается одинаковое число данных, комплекс называется **равномерным**, неодинаковое число данных – **неравномерным**. Если градации двух и многофакторных комплексов заполнены неодинаковым числом данных, но так, что данные по градациям одного фактора находятся в одинаковом отношении в градациях всех остальных факторов, такой комплекс называется **пропорциональным**.

Комплексы, составленные по принципу случайной выборки, называются **рандомизированными**.

При изучении количественных признаков в градации комплекса заносятся данные – числовые результаты измерения изучаемого признака у каждого отдельного объекта.

### 1.6 Статистические влияния

***Статистическое влияние*** – это отражение в разнообразии результативного признака того разнообразия фактора (его градаций), которое организовано в исследовании.

Для оценки влияния фактора необходимо выявить разнообразие его действия, т. е. установить, насколько различно действуют его градации на результативный признак.

Например, для изучения влияния радиационного облучения на мутационный процесс требуется организовать минимум две градации фактора (облучения нет, облучение есть) и установить, велико ли различие в частоте мутаций при разных градациях облучения.

Если разнообразие результативного признака большое, т. е. если частота мутаций при облучении значительно отличается от частоты без облучения, значит влияние данного облучения велико; если такого различия нет или оно мало, значит, изучаемый фактор действует слабо, конечно, при данных дозах и условиях.

**Оценка влияния фактора по разнообразию действия его градаций есть основной принцип дисперсионного анализа**.

Если нет разнообразия результативного признака по градациям фактора, нет и статистического влияния, хотя физиологическое влияние этого фактора может быть сильным.

Например, стимулятор мог сильно повышать плодовитость, но одинаково при всех дозах, организованных в данном исследовании (сильный эффект при любой не нулевой дозе). Тогда при сильном физиологическом влиянии этот стимулятор показал бы очень малое статистическое влияние, конечно, только при организации комплекса без нулевой градации.

При проведении дисперсионного анализа изучаются три основных вида статистических влияний:

### 1) Факториальное влияние – это простое или комбинированное статистическое влияние изучаемых факторов.

В однофакторных комплексах изучается простое влияние одного фактора при определенных организованных в опыте градациях и при определенных общих условиях.

Этого не следует забывать при интерпретации результатов анализа. Например, если действие температуры оказалось сильным при градациях 10°–15°–20°, это не значит, что столь же сильное статистическое влияние проявится и при любых других градациях фактора, например при 20°–25°–30°.

Точно так же, если влияние температуры изучалось при нормальной влажности и естественном освещении, нельзя ожидать такой же степени влияния температуры при повышенной (пониженной) влажности и при искусственном освещении.

При анализе двухфакторных дисперсионных комплексов изучаются 4 факториальных влияния:

- влияние первого фактора при усредненном влиянии второго;

- влияние второго фактора при усредненном влиянии первого;

- влияние сочетания градаций обоих факторов;

- суммарное действие обоих организованных факторов.

При анализе влияний *первого* и *второго факторов* действуют те же ограничения, какие необходимо иметь в виду при анализе однофакторных комплексов: выявляется степень влияния только при данных градациях каждого фактора и при данных условиях.

Кроме того, следует помнить, что действие каждого фактора в двухфакторном комплексе изучается при усредненном влиянии другого фактора и усредненном действии всех остальных, неорганизованных в данном комплексе факторов. Например, если изучается степень проявления в данных условиях влияния отцов (первый фактор) и матерей (второй фактор), то сила отцовского влияния (при данном составе матерей и при определенных условиях) изучается при усредненной реализации влияния матерей, а сила материнского влияния – при усредненном влиянии отцов, участвовавших в проведенных скрещиваниях.

Третье влияние, или *влияние сочетания градаций обоих факторов*, возникает вследствие того, что второй фактор часто действует различно при разных градациях первого. То же можно наблюдать и в отношении первого фактора: его действие часто проявляется неодинаково при различных градациях второго фактора. Например, если изучается действие стимулятора линьки (две градации второго фактора – контроль, опыт) на самцов и самок (две градации первого фактора), то может случиться так, что введение стимулятора даст большой эффект только для самок, а для самцов – незначительный.

Такое разнообразие действий одного фактора при разных градациях другого создает дополнительное статистическое влияние (сверх изолированных влияний каждого фактора), которое учитывается как особый вид факториальных влияний.

*Суммарное действие факторов* (четвертое влияние) включает в себя изолированные влияния каждого из факторов и влияние сочетаний их градаций. Это суммарный представитель всех факториальных влияний в двухфакторном дисперсионном комплексе.

### 2) Случайное влияние – это действие тех многих факторов, которые не организованы в изучаемом дисперсионном комплексе и составляют общий фон, на котором действуют организованные факторы. Так как неорганизованных факторов много и действуют они в разных направлениях, их влияние рассматривается как случайное, т. е. не вытекающее из закономерности действия организованных факторов.

Во всех дисперсионных комплексах случайные влияния выявляются и измеряются единообразно и для одной и той же очень важной цели: для определения той базы, с которой, как с эталоном, сравниваются факториальные влияния при определении их достоверности.

**Чем больше факториальное влияние отличается от случайного, тем большая достоверность приписывается этому факториальному влиянию**.

Это правило в дисперсионном анализе не имеет исключений. При определении достоверности факториальных влияний за базу сравнения можно принимать только случайные влияния.

### 3) Общее влияние – это влияние всех организованных и неорганизованных факторов, определивших такое развитие признака, которое наблюдалось в дисперсионном комплексе. Общее влияние служит базой для определения доли влияний – факториальных и случайных.

##

## 2 Однофакторный дисперсионный комплекс

Дисперсионный анализ разработан и введен в практику сельскохозяйственных и биологических исследований английским ученым Р.А. Фишером, который открыл закон распределения отношения средних квадратов (дисперсий).

Дисперсионный анализ широко используется для планирования эксперимента и статистической обработки его данных. Если в недалеком прошлом считали, что роль математика состоит лишь в анализе экспериментальных данных, то работы Р.А. Фишера коренным образом изменили эту точку зрения, и в настоящее время статистическое планирование опыта в соответствии с требованиями дисперсионного анализа и математическая интерпретация результатов: непременные условия успешного получения ответов на вопросы, интересующие экспериментатора. Статистически обоснованный план эксперимента определяет и метод математического анализа результатов. Поэтому современный эксперимент нельзя правильно спланировать, не зная основ дисперсионного анализа.

**При дисперсионном анализе одновременно обрабатывают данные нескольких выборок** (вариантов), составляющих единый статистический комплекс, оформленный в виде специальной рабочей таблицы. Структура статистического комплекса и его последующий анализ определяются схемой и методикой эксперимента.

**Сущностью дисперсионного анализа является расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части – компоненты, соответствующие структуре эксперимента, и оценка значимости действия и взаимодействия изучаемых факторов по *F*–критерию**.

Если обрабатывают однофакторные статистические комплексы, состоящие из нескольких независимых выборок, например *l*–вариантов в вегетационном опыте, то общая изменчивость результативного признака, измеряемая общей суммой квадратов *СY*, расчленяется на два компонента: варьирование между выборками (вариантами) *СV* и внутри выборок *Cz*, Следовательно, в общей форме изменчивость признака может быть представлена выражением:



Здесь вариация между выборками (вариантами) представляет ту часть общей дисперсии, которая обусловлена действием изучаемых факторов, а дисперсия внутри выборок характеризует случайное варьирование изучаемого признака, т. е. ошибку эксперимента.

Общее число степеней свободы (*N–1*) также расчленяется на две части – степени свободы для вариантов (*l–1*) и для случайного варьирования (*N–l*);

Суммы квадратов отклонений по данным эксперимента – статистического комплекса с *l*-вариантами – обычно находят в такой последовательности. При общем числе наблюдений *N* определяются суммы по вариантам V и общая сумма всех наблюдений *ΣХ*. Затем вычисляются:

1. корректирующий фактор (поправка) 
2. общая сумма квадратов 
3. сумму квадратов для вариантов 
4. сумму квадратов для ошибки 

Две последние суммы квадратов *СY* и *CZ* делятся на соответствующие им степени свободы, т. е. приводятся к сравниваемому виду – одной степени свободы вариации. В результате получают два средних квадрата:

вариантов  и ошибки 

Эти средние квадраты и используют в дисперсионном анализе для оценки значимости действия изучаемых факторов. Оценка проводится путем сравнения дисперсии вариантов  с дисперсией ошибки  по критерию Фишера:

.

Таким образом, за базу – единицу сравнения принимают средний квадрат случайной дисперсии, которая определяет случайную ошибку эксперимента. При этом проверяемой нулевой гипотезой служит предположение: все выборочные средние являются оценками одной генеральной средней и, следовательно, различия между ними несущественны.

Если *Fфакт < Fтеop*, то нулевая гипотеза *Но: d*=0 не отвергается; между всеми выборочными средними нет существенных различий и на этом проверка заканчивается. Нулевая гипотеза отвергается, когда *Fфакт ≥ Fтеop*.

В этом случае дополнительно проводят оценку, между какими средними имеются значимые различия.

Теоретическое значение критерия F для принятого в исследовании уровня значимости находят по таблицам приложений с учетом числа степеней свободы для дисперсии вариантов и случайной дисперсии. В большинстве случаев избирают 5%–ный, а при более строгом подходе 1%–ный или даже 0,1%–ный уровень значимости.