

ДИСПЕРСИОННЫЙ АНАЛИЗ

1. Сущность дисперсионного анализа.
2. Дисперсионный анализ многофакторного опыта

1. Сущность дисперсионного анализа Дисперсионный анализ был разработан и введен в практику сельскохозяйственных исследований английским ученым Фишером, который открыл закон распределения средних квадратов, или дисперсий. При дисперсионном анализе одновременно обрабатываются данные нескольких выборок (вариантов), составляющих единый статистический комплекс, структура которого определяется схемой и методикой эксперимента.

Сущностью дисперсионного анализа является расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части, соответствующие структуре эксперимента, и оценка действия и взаимодействия изучаемых факторов по критерию Фишера. С помощью дисперсионного анализа проверяется нулевая гипотеза, которая формулируется следующим образом: между средними по вариантам опыта нет существенных различий.

Модель дисперсионного анализа определяется количеством изучаемых факторов, типом сравнительного эксперимента, а в полевом опыте также и методом размещения вариантов по делянкам. Для применения дисперсионного анализа опыт должен проводиться в нескольких повторностях.

Наиболее простой является схема однофакторного лабораторного и вегетационного опыта, а также полевого опыта с модельными растениями, где нет территориально организованных повторений. В этом случае общее число квадратов отклонений, т.е. общее варьирование, расчленяется на 2 компонента: варьирование вариантов и варьирование случайных факторов:

$$C_y = C_v + C_z$$

где C_y – общее варьирование, т.е. общая изменчивость изучаемого признака;

C_v - варьирование вариантов, которое определяется влиянием изучаемого фактора и определяет варьирование между выборками;

C_z – варьирование случайных факторов, или остаточное варьирование, которое определяется влияние изучаемых факторов и определяет варьирование внутри выборки.

Общее число степеней свободы также делится на 2 компонента: число степеней свободы для вариантов и число степеней свободы для случайного варьирования:

$$v_y = v_v + v_z$$

где v_y – общее число степеней свободы, которое определяется по формуле $N-1$, где N – общее число значений признака (общее число наблюдений);

v_v - число степеней свободы вариантов, которое определяется по формуле $l-1$, где l – число вариантов опыта;

v_z – число степеней свободы случайных факторов, или остатка, которое рассчитывается по формуле $v_y - v_v$.

При проведении однофакторного полевого опыта методом рендомизированных повторений общее варьирование S_y делится на 3 компонента: варьирование вариантов S_v , варьирование повторений S_p и случайное варьирование, или варьирование ошибки S_z .

Варьирование повторений обусловлено влиянием почвенного плодородия и определяет варьирование признака между повторениями.

Общее число степеней свободы v_y также разлагается на 3 части: степени свободы для вариантов v_v , степени свободы для повторений v_p и для случайного варьирования v_z . Где

v_p - определяется по формуле $n-1$, где n – число повторений в опыте;

v_z – рассчитывается по формуле $v_y - v_v - v_p$.

Варьирование вариантов и варьирование случайных факторов делят на соответствующие им степени свободы и получают две дисперсии: дисперсию вариантов и дисперсию ошибки.

Отношение дисперсии вариантов к дисперсии ошибки дает фактическое значение критерия Фишера F .

Для оценки действия изучаемых факторов фактическое значение критерия Фишера сравнивают с теоретическим, которое берут из статистической таблицы с учетом принятого уровня значимости и числа степеней свободы вариантов и ошибки.

Если $F_{\text{факт.}} < F_{\text{теор.}}$, нулевая гипотеза принимается, следовательно между вариантами нет существенных различий.

Если $F_{\text{факт.}} \geq F_{\text{теор.}}$, нулевая гипотеза отвергается, следовательно между вариантами есть существенные различия.

Критерий Фишера лишь показывает, что между изучаемыми вариантами есть существенные различия, но не указывает между какими средними имеются эти различия.

Для оценки существенности различий между средними рассчитывают **наименьшую существенную разность (НСР)**, которая указывает предельную величину ошибки для разности двух выборочных средних d . Если разность двух выборочных средних больше или равна НСР, то она существенна, если она меньше НСР – несущественна.

Группировка вариантов по НСР. Опытные варианты можно распределить по группам в зависимости от величины отклонения средних значений опытных вариантов от контроля или стандарта. Если отклонение от стандарта превышает НСР, по вариант относится к I группе, находится в пределах \pm НСР – ко II группе, если отклонение с отрицательным знаком по модулю больше НСР вариант относится к III. Варианты, входящие в I группу существенно превосходят стандарт, во II группу - находятся на уровне стандарта, в III группу – значительно уступают стандарту.

2. Дисперсионный анализ многофакторного опыта.

Многофакторный дисперсионный комплекс – это совокупность исходных наблюдений, позволяющих статистически оценить действие и взаимодействие нескольких изучаемых факторов на изменчивость результативного признака. Например, если мы изучаем в многофакторном опыте одновременно влияние удобрений и орошения на урожайность какой-либо культуры, то мы должны общее варьирование вариантов, вызванное применением этих агротехнических приёмов разложить на варьирование урожайности, происходящее под действием орошения, варьирование урожайности, обусловленное действием удобрений и варьирование урожайности, обусловленное взаимодействием этих двух факторов. В полевом опыте эффект от совместного применения изучаемых факторов может быть больше (синергизм) или меньше (антагонизм) суммы эффектов от отдельного применения каждого из них. В первом случае существует положительное взаимодействие факторов, во втором – отрицательное. Когда факторы не взаимодействуют, прибавка от их совместного применения равна сумме прибавок от отдельного применения (аддитивизм).

В многофакторном опыте источников варьирования больше, чем в однофакторных.

Дисперсионный анализ многофакторного опыта проводится в 2 этапа. На первом этапе данные обрабатываются как в однофакторном опыте. Ко второму этапу переходят в том случае, когда фактическое значение критерия Фишера больше теоретического и, следовательно, между вариантами есть существенные различия.

Схема проведения второго этапа дисперсионного анализа многофакторного опыта зависит от метода размещения вариантов по делянкам полевого опыта. Обычно многофакторные полевые опыты закладываются методом рендомизированных повторений или методом расщепленных делянок.

На втором этапе при закладке опытов методом рендомизированных повторений сумма квадратов вариантов разлагается на компоненты, соответствующие источникам варьирования – главные эффекты изучаемых факторов и их взаимодействия, на соответствующие компоненты разлагается также и число степеней свободы вариантов.

Сумма квадратов отклонений для вариантов C_v и число степеней свободы вариантов расчленяется на компоненты в зависимости от структуры опыта: в двухфакторном опыте - на 3, а в трехфакторном – на 7 компонентов.

$$C_v = C_A + C_B + C_{AB} \quad (7)$$

$$C_v = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC} \quad (8)$$

При обработке результатов двухфакторного опыта рассчитывают 3 значения $НСР_{05}$: $НСР_{05A}$ – для оценки действия фактора А, $НСР_{05B}$ – для оценки действия фактора В и $НСР_{05AB}$ – для оценки взаимодействия факторов.

В опытах с расщеплёнными делянками имеется большие делянки первого порядка, которые расщепляются (делятся) на более мелкие делянки второго порядка и т.д. Статистическая обработка данных таких опытов имеет свою специфику, так как ошибки на больших и малых делянках неодинаковы по величине и поэтому оценка существенности разниц по одному значению $НСР$ не может быть правильной. Первоначальную обработку данных этих опытов проводят в той же последовательности, что и обработку данных многофакторного опыта поставленного методом рендомизированных повторений. Отличием здесь является разложение суммы квадратов случайных факторов C_z на вариабельность делянок первого (ошибка I), второго (ошибка II) и т.д. порядков.

При обработке результатов двухфакторного опыта, заложенного методом расщепленных делянок, рассчитывают 4 значения $НСР_{05}$: $НСР_{05A}$ –

для оценки действия фактора А, $НСР_{05B}$ – для оценки действия фактора В, и $НСР_{05}'$ – для оценки взаимодействия факторов на делянках I порядка, $НСР_{05}''$ – для оценки взаимодействия факторов на делянках II порядка.