**§5. Дисперсионный анализ**

В биологических исследованиях часто приходится выяснять, как действуют те или иные внутренние или внешние факторы на свойства, параметры биологических объектов. Для этого обычно берут выборки из различных генеральных совокупностей (определяемых различными условиями, например, количеством удобрений, внесенных в почву, среднегодовой температурой, количеством осадков и так далее) и сравнивают средние значения этих выборок. Если какая-либо пара выборок по t-критерию Стьюдента статистически достоверно отличаются друг от друга, то делают вывод о влиянии данного фактора на среднее значение признака. Короче говоря, влияние фактора сказывается на сред­нем значении случайной величины. Изменение среднего значения необходимо выявить статистическими методами.

Но при большом числе выборок вероятность того, что различия у двух из них окажутся такими, что фактическое значение статистического критерия при условии правильности нулевой гипотезы попадёт в критическую область, становится большой, даже если все эти выборки принадлежат одной генеральной совокупности. Следовательно, попарное сравнение средних в данном случае может дать неверный ответ с большей вероятностью. Чтобы обойти эту трудность, Фишер разработал метод одновременного сравнения сразу нескольких средних. Этот метод основан на изучении дисперсии признака, и поэтому называется дисперсионным анализом (ANOVA - analysis of variance). Планирование эксперимента для выяснения того, влияет ли данный фактор на среднее значение исследуемого признака, состоит в следующем: эксперименты проводят в нескольких условиях, отличающихся значениями данного фактора (например, это может быть температура, при которой культивировали некий микроорганизм); экспериментальные результаты, полученные при разных значениях фактора, рассматриваются как выборки из соответствующих генеральных совокупностей. Если фактор влияет на исследуемый признак, то выборочные средние должны статистически достоверно отличаться друг от друга. Выявление этих отличий осуществляется методом дисперсионного анализа. Все данные, полученные по всем выборкам, представляют единый статистиче­ский комплекс, но данные сгруппированы по градациям данного фактора.

Суть дисперсионного анализа состоит в разложении общей дисперсии признака (по всему статистическому комплексу) на две части: факториальную дисперсию, обусловленную вариацией исследуемого фактора, и остаточную дисперсию, обусловленную действием всех других неучитываемых факторов – и последующем сравнении этих вкладов в общую дисперсию с помощью критерия Фишера. Если факториальная дисперсия значимо превосходит остаточную, то делают вывод о существовании влиянии данного фактора на исследуемый признак.

Если генеральная совокупность разбита на группы, то имеет место следующее соотношение:

$$D\_{общ}=D\_{внгр}+D\_{межгр,}$$

где Dобщ – общая дисперсия, вычисляется по формуле:

$$D\_{общ}=\frac{\sum\_{}^{}\left(x\_{ij}-\overbar{x}\right)^{2}}{N-1}$$

$D\_{общ}$ – внутригрупповая дисперсия, вычисляемая по формуле:

$$D\_{внгр}=\frac{\sum\_{j}^{}\left(x\_{ij}-\overbar{x\_{i}}\right)^{2}}{N\_{i}}$$

$D\_{межгр}$ – межгрупповая дисперсия, вычисляемая по формуле:

$$D\_{межгр}=\frac{\sum\_{j}^{}N\_{j}\left(\overbar{x}\_{j}-\overbar{x}\right)^{2}}{N}$$

Для проведения дисперсионного анализа необходимо выполнение следующих условий: выборки должны принадлежать нормально распределенным генеральным совокупностям, у которых математические ожидания могут быть разными, но дисперсии должны быть одинаковыми. Следовательно, перед проведением дисперсионного анализа необходимо проверить выборки на нормальность распределения и на равенство дисперсий. Если условия выполнены, то можно проводить анализ, если нет – то выявлять различие средних обычным методом парных сравнений, но уровень доверительной вероятности следует поднять, а к выводам относиться с осторожностью.

Рассмотрим пример применения дисперсионного анализа. Пусть необходимо сравнить *к* сортов пшеницы по урожайности.

Простейшим способом сравнения является следующий эксперимент. Опытное поле разбивают на *к* участков. Каждый участок подразделяют на делянки. На данный участок высевают пшеницу одного сорта. Подсчитывают урожай на каждой делянке. Результаты сводят в таблицу.

Таблица

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № сорта | Наблюденияурожая наотдельныхделянках | Числонаблюдений | Суммарныйурожай | Среднийуpoжай |
| 1 | х11, х12, *…* $х\_{1n\_{1}}$ | n1 | T1 | $$\overbar{X\_{1}}$$ |
| 2 | Х21, х22, *…* $х\_{2n\_{2}}$ | n2 | T2 | $$\overbar{X\_{2}}$$ |
| 3 | Х31, х32, *…* $х\_{3n\_{3}}$ | n3 | T3 | $$\overbar{X\_{3}}$$ |
| 4 | . . . . . . . . . . . | . . . . | . . .  | . . . |
| К | Хk1, хk2, *…* $х\_{kn\_{k}}$ | nk | Tk | $$\overbar{X\_{k}}$$ |
| Итого: | $N=\sum\_{}^{}n\_{i}$  | $G=\sum\_{}^{}T\_{i}$  | $\overbar{X}=\frac{G}{N}$  |

Здесь $\overbar{X\_{i}}$ – средние урожаи i-того сорта, $\overbar{X}$ – общий средний урожай.

Нулевая гипотеза дисперсионного анализа заключается в предположении, что урожайности сортов одинаковы, то есть, что математические ожидания урожайностей для разных сортов равны друг другу:

Н0 : М1 = М2 = ...=МК

$H\_{1}$ : не все средние равны

Конкурирующая гипотеза заключается в предположении, что хо­тя бы для одной пары средних равенство не выполняется.

Чтобы проверить нулевую гипотезу необходимо вычислить зна­чение статистического критерия:

$$F\_{ф}=\frac{S\_{ф}^{2}}{S\_{ост}^{2}}$$

где $S\_{ф}^{2}$ - оценка факториальной дисперсии; $S\_{ост}^{2}$ - оценка остаточной дисперсии.

 Оценки дисперсий вычисляют так: сначала вычисляют соответствующие суммы квадратов отклонений, а за тем эти суммы делят на соответствующие числа степеней свободы. Это видно из таблицы:

|  |  |  |  |
| --- | --- | --- | --- |
| Компонентадисперсии | Суммаквадратов | Сте-пенисвобо-ды | Оценкидисперсий |
| Факто-риальная | $\sum\_{}^{}\left(\overbar{x}\_{i}-\overbar{x}\right)^{2}$  | k - 1 | $$S\_{ф}=\frac{\sum\_{}^{}\left(\overbar{x}\_{i}-\overbar{x}\right)^{2}}{k-1}$$ |
| Общая | $\sum\_{i,j}^{}\left(x\_{ij}-\overbar{x}\right)^{2}$  | N - 1 |  |
| Оста-точная  | $\sum\_{}^{}\left(x\_{ij}-\overbar{x}\right)^{2}-\sum\_{}^{}\left(x\_{i}-\overbar{x}\right)^{2}$  | n - k |  |

Рабочие формулы для вычисления сумм квадратов отклонений таковы: сначала вычисляют общую сумму квадратов отклонений по формуле:

$\sum\_{}^{}\frac{\left(T\_{i}\right)^{2}}{n\_{i}}-\frac{\left(\sum\_{}^{}T\_{i}\right)^{2}}{N}$

Затем факториальную сумму квадратов (т.е. между выборками) по формуле:

$\sum\_{}^{}\frac{\left(T\_{i}\right)^{2}}{n\_{i}}-\frac{\left(\sum\_{}^{}T\_{i}\right)^{2}}{N}$

После этого из общей суммы вычитают факториальную и получают остаточную. Как и обычно при сравнении дисперсий по таблице Фишера находят критическую точку для правостoронней критической области. Если $F\_{∅ }$**<**$F\_{кр'} $то нулевая гипотеза отвергается, т.е. не все урожайности сортов одинаковы.

Если $F\_{∅ }$**<**$F\_{кр'} $то урожайности сортов различаются недостоверно.

**Пример.** Пусть имеется три вида червей. Биолог берет случайную выборку из каждого вида и измеряет длину особей. Данные приведены в таблице.

 Длина червей трех видов (см):

|  |  |  |
| --- | --- | --- |
| 1 | 2 | 3 |
| 10.28.28.98.08.38.0 | 12.210.69.913.08.110.811.5 | 9.210.59.28.79.0 |
| $Т\_{1}$=51.6 | $T\_{2}$=76.1 | $T\_{3}$46.6 |

$\sum\_{}^{}T\_{i}=174,3$ $N=18$

Проверить гипотезу о равенстве средних, то есть

$H\_{0 }$: $M\_{1}$=$M\_{2}$=$M\_{3}$

$H\_{1}$: не все средние равны.

**Решение.** Нужно найти сначала общую сумму квадратов.

$\sum\_{}^{}x\_{ij}^{2}=\left(10.2\right)^{2}+\left(8.2\right)^{2}+\left(8.7\right)^{2}+\left(9.0\right)^{2}=1726.31$

$\left(ΣT\_{i}\right)^{2}$/N =$(174,3)^{2}$/18=1687,8

$\left(Σx\_{ij}^{2}\right)$=$\frac{\left(ΣT\_{i}\right)^{2}}{N}$= 38,505

Теперь найдем факториальную сумму квадратов:

$\sum\_{}^{}\left(\overbar{x}\_{i}-\overbar{x}\right)^{2}=\frac{\left(51.6\right)^{2}}{6}+\frac{\left(76.1\right)^{2}}{7}+\frac{\left(46.1\right)^{2}}{5}+\frac{\left(174.3\right)^{2}}{18}=17.583$

Отсюда находится остаточная сумма квадратов:

38,505-17,587=20,922.

Найдем теперь$ S\_{Ø}^{2}$ и $S\_{ост}^{2}$. Имеем:

$$S\_{∅}^{2}=\frac{17.583}{2}=8.792$$

$$S\_{ост}^{2}=\frac{20.922}{15}=1ю395$$

Теперь вычислим:

$$F\_{∅}=\frac{S\_{∅}^{2}}{S\_{ост}^{2}}=\frac{8.792}{1.335}$$

По таблице Фишера найдем $F\_{st}$ для $К\_{1}$=2 и $К\_{2}$=15. Для доверительной вероятности β=0,95 $F\_{st}$= 3,68. Значит, нулевую гипотезу нужно отклонить, то есть следует признать, что средняя длина не у всех этих трех видов червей одинакова.

Мы рассмотрели так называемый однофакторный дисперсионный комплекс. В данном случае результаты наблюдения соответствуют варьированию одного фактора, - сорта пшеницы. Но в научных исследованиях приходится выявлять влияние двух или более факторов. Иногда каждый фактор порознь не влияет на признак, а совместное их действие оказывается существенным. В таких случаях также может быть применен дисперсионный анализ, который позволит выявить наличие совместного влияния двух (и более) факторов.

**Контрольные вопросы и задания**

1. В каких случаях применяется дисперсионный анализ?
2. В чем суть дисперсионного анализа?
3. Каковы необходимые условия для проведения дисперсионного анализа?
4. На четырех разновозрастных группах мужчин измерялась скорость кровотока в сосудах (см/сек). Результаты приведены в таблице.

|  |  |  |
| --- | --- | --- |
| Возрастные группы мужчин | Наблюдения | Среднее |
| ПерваяВтораяТретьяЧетвертая | 7, 10, 129, 7, 1411, 16, 2015, 18, 17 | 9,6710,0016,6716,67 |

Методом дисперсионного анализа определите достоверность различия средних в этих группах.

1. Полярографическая активность фильтрата сыворотки крови при разной длительности заболевания(в отн. ед.) А:

А1  98 156 168 143 128 162 92 157 162 184

А2 90 79 101 86 171 142 82 104 108 64

А3 69 71 116 102 110 64 120 141 92 71

А4 51 51 71 94 64 66 72 135 75 124

Методом дисперсионного анализа проверьте гипотезу о равенстве средних.

1. Тридцать студентов-первокурсников были разделены на три группы по специально-экономическим признакам, и каждой из групп был предложен тест, позволяющий определить степень усвоения материала во время лекции:

Результаты приведены в таблице:

|  |  |  |
| --- | --- | --- |
| Группа А | Группа Б | Группа В |
| 21181717171615131287 | 202019181717141497 | 242323222020181716 |

Определите методом дисперсионного анализа, влияют ли социальные условия на усвоение материала лекции.

1. В сельскохозяйственном эксперименте 35 участников земли, имеющих примерно одинаковое плодородие, были засеяны семью сортами зерна. На каждый сорт отведено по 6 участков. Участки под сорта отводились случайно. В таблице приведены значения урожаев, полученных на всех участках (в бушелях/акр).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| I | II | III | IV | V | VI | VII |
| 1013111512 | 151417918 | 91917914 | 1715191113 | 121691315 | 1612151117 | 1310141419 |

Имеются ли существенные различия в урожайности этих сортов?

1. В таблице приведены оценки пяти студентов по трем дисциплинам (по 100-бальной шкале).

|  |  |  |
| --- | --- | --- |
| Статистика | Психология | Музыка |
| 8173679060 | 7391948783 | 6587849389 |

Проверить гипотезу об одинаковой успеваемости по этим трем дисциплинам.

1. Наблюдатели, расположенные в трех различных пунктах на поверхности Земли, проводили расшифровку радиосигналов, полученных из открытого космоса. Их результаты сравниваются с теми же сигналами, фиксируемыми искусственным спутником Земли. Проверить гипотезу о равенстве средних.

|  |  |
| --- | --- |
| Земные наблюдения | Космическиенаблюдения |
| А | Б | В | Г |
| 323231 | 344343 | 445443 | 764647 |

1. Три группы студентов, обучающихся машинописи с помощью трех различных систем, были подвергнуты проверке. Результаты в таблице. Проверить, влияет ли система обучения на количество ошибок при печатании.

|  |  |  |
| --- | --- | --- |
| I система | II система | III система |
| 67778 | 45558 | 66666 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 75% | 80% | 85% | 90% | 95% | 97,5% | 99% | 99,5% | 99,75% | 99,9% | 99,95% |
| 1 | 1,000 | 1,376 | 1,963 | 3,078 | 6,314 | 12,71 | 31,82 | 63,66 | 127,3 | 318,3 | 636,6 |
| 2 | 0,816 | 1,061 | 1,386 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 14,09 | 22,33 | 31,60 |
| 3 | 0,765 | 0,978 | 1,250 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 7,453 | 10,21 | 12,92 |
| 4 | 0,741 | 0,941 | 1,190 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 5,598 | 7,173 | 8,610 |
| 5 | 0,727 | 0,920 | 1,156 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 4,773 | 5,893 | 6,869 |
| 6 | 0,718 | 0,906 | 1,134 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 4,317 | 5,208 | 5,959 |
| 7 | 0,711 | 0,896 | 1,119 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 4,029 | 4,785 | 5,408 |
| 8 | 0,706 | 0,889 | 1,108 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 3,833 | 4,501 | 5,041 |
| 9 | 0,703 | 0,883 | 1,100 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 3,690 | 4,297 | 4,781 |
| 10 | 0,700 | 0,879 | 1,093 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 3,581 | 4,144 | 4,587 |
| 11 | 0,697 | 0,876 | 1,088 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 | 3,497 | 4,025 | 4,437 |
| 12 | 0,695 | 0,873 | 1,083 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 | 3,428 | 3,930 | 4,318 |
| 13 | 0,694 | 0,870 | 1,079 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 | 3,372 | 3,852 | 4,221 |
| 14 | 0,692 | 0,868 | 1,076 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 | 3,326 | 3,787 | 4,140 |
| 15 | 0,691 | 0,866 | 1,074 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 | 3,286 | 3,733 | 4,073 |
| 16 | 0,690 | 0,865 | 1,071 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 | 3,252 | 3,686 | 4,015 |
| 17 | 0,689 | 0,863 | 1,069 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 | 3,222 | 3,646 | 3,965 |
| 18 | 0,688 | 0,862 | 1,067 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 | 3,197 | 3,610 | 3,922 |
| 19 | 0,688 | 0,861 | 1,066 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 | 3,174 | 3,579 | 3,883 |
| 20 | 0,687 | 0,860 | 1,064 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 | 3,153 | 3,552 | 3,850 |
| 21 | 0,686 | 0,859 | 1,063 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 | 3,135 | 3,527 | 3,819 |
| 22 | 0,686 | 0,858 | 1,061 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 | 3,119 | 3,505 | 3,792 |
| 23 | 0,685 | 0,858 | 1,060 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 | 3,104 | 3,485 | 3,767 |
| 24 | 0,685 | 0,857 | 1,059 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 | 3,091 | 3,467 | 3,745 |
| 25 | 0,684 | 0,856 | 1,058 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 | 3,078 | 3,450 | 3,725 |
| 26 | 0,684 | 0,856 | 1,058 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 | 3,067 | 3,435 | 3,707 |
| 27 | 0,684 | 0,855 | 1,057 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 | 3,057 | 3,421 | 3,690 |
| 28 | 0,683 | 0,855 | 1,056 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 | 3,047 | 3,408 | 3,674 |
| 29 | 0,683 | 0,854 | 1,055 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 | 3,038 | 3,396 | 3,659 |
| 30 | 0,683 | 0,854 | 1,055 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,030 | 3,385 | 3,646 |
| 40 | 0,681 | 0,851 | 1,050 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 | 2,971 | 3,307 | 3,551 |
| 50 | 0,679 | 0,849 | 1,047 | 1,299 | 1,676 | 2,009 | 2,403 | 2,678 | 2,937 | 3,261 | 3,496 |
| 60 | 0,679 | 0,848 | 1,045 | 1,296 | 1,671 | 2,000 | 2,390 | 2,660 | 2,915 | 3,232 | 3,460 |
| 80 | 0,678 | 0,846 | 1,043 | 1,292 | 1,664 | 1,990 | 2,374 | 2,639 | 2,887 | 3,195 | 3,416 |
| 100 | 0,677 | 0,845 | 1,042 | 1,290 | 1,660 | 1,984 | 2,364 | 2,626 | 2,871 | 3,174 | 3,390 |
| 120 | 0,677 | 0,845 | 1,041 | 1,289 | 1,658 | 1,980 | 2,358 | 2,617 | 2,860 | 3,160 | 3,373 |
| ∞ | 0,674 | 0,842 | 1,036 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 2,807 | 3,090 | 3,291 |