Статистические методы проверки гипотез

Выносимые положения:

- Значение большого количества наблюдений в защите растений и необходимость перехода к малым выборкам.
- Принципы подбора малых выборок.
- t –распределение Стьюдента.
- F - распределение Фишера.

Положение 1

Значение большого количества наблюдений в защите растений и необходимость перехода к малым выборкам.

Закон нормального распределения

- Проявляется при большом количестве наблюдений
- При объеме выборки более 20
- Чем больше взято наблюдений, тем точнее опыт

При небольшом количестве наблюдений опыт может оказаться

малонадежным

Большое количество наблюдений имеет огромное значение в растениеводстве при изучении биометрии растений:

- длина стебля;
- длина колоса;
- размер листьев;
- биомасса сорных растений

В растениеводстве при изучении биометрии растений измеряют

100 растений

В защите растений:

при определении <u>пораженности растений болезнями</u>

в учет берется

- 100 растений.

при определении <u>численности вредителей</u>

в учет берется

- 30-100 растений;
- 30-50 площадок на метр квадратный.

при определении засоренности полей сорняками учитывается:

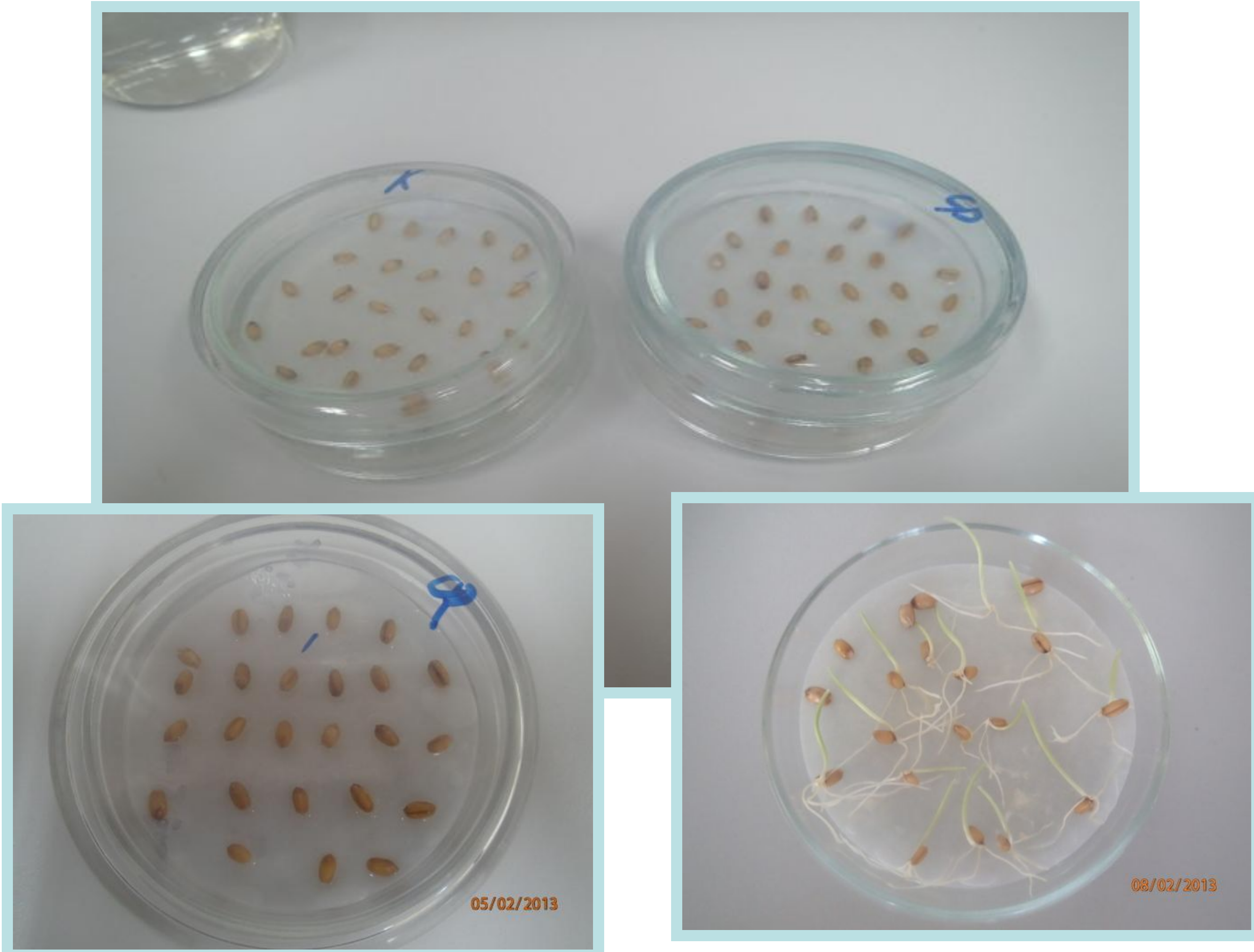
- **30-50 площадок на метр квадратный.**

Необходимость перехода к малым выборкам связана:

- со значительными затратами средств и времени.

Например:

- **При проведении фитоэкспертизы семян озимой пшеницы при большом количестве вариантов – 18 и 4-х кратной повторности, количество чашек Петри достигает 72 шт.**


05/02/2013
08/02/2013

к малым выборкам связана:

- с возможностями приборов.

Прибор Варбурга не позволит провести большее количество наблюдений; для каждого варианта можно сделать 3 повторности.

Прибор Варбурга

Положение 2

Принципы подбора малых выборок.

Принципы подбора малых выборок

Принцип репрезентативности

Выборка должна отражать происходящее в природе

Принцип однородности

Выборки должны быть однородны

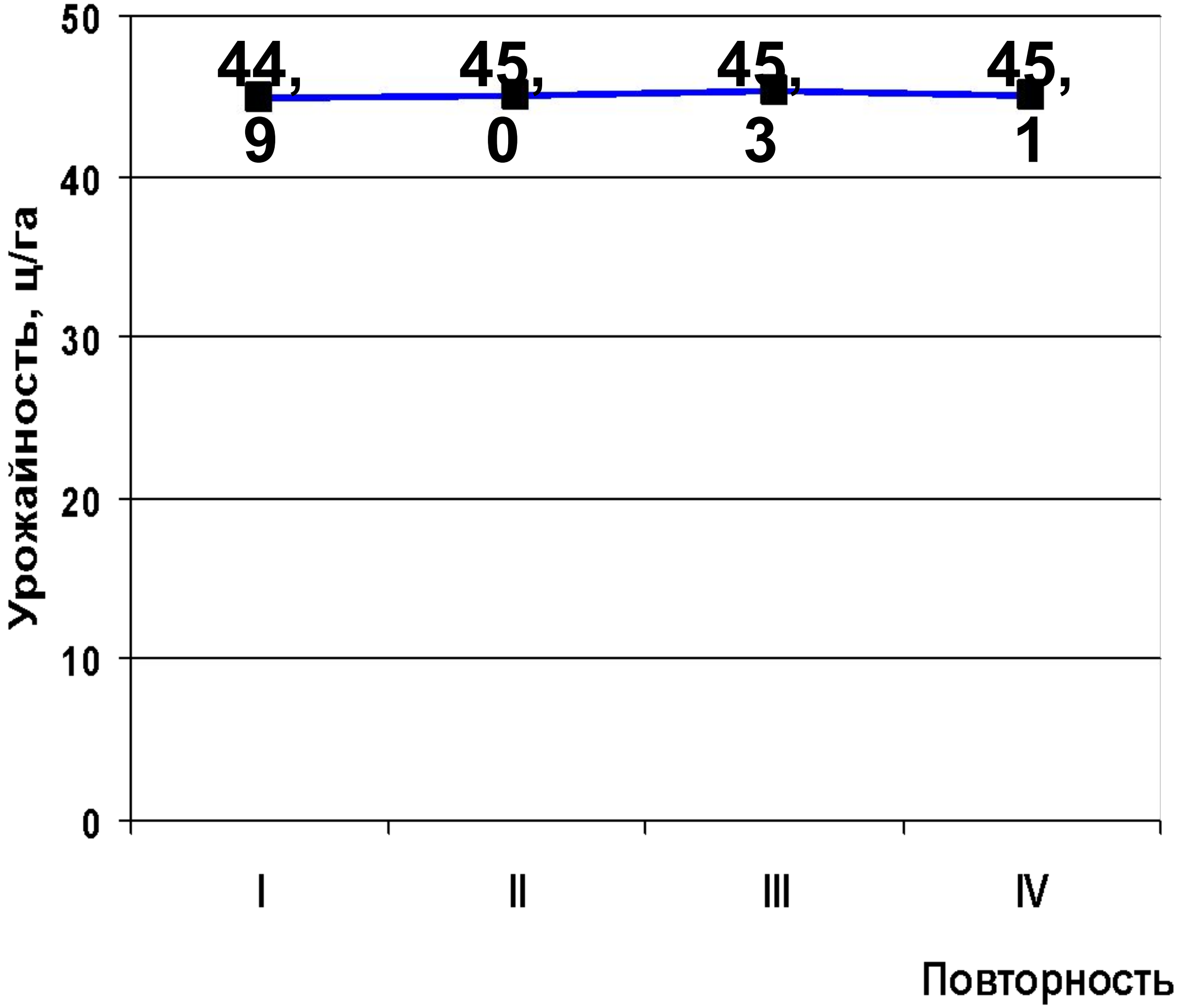
Положение 3

t –распределение Стьюдента.

Если, в полевом опыте выдержаны все основные принципы его закладки, то результаты опыта по повторностям будут колебаться незначительно.

Показатель	Повторность			
	I	II	III	IV
Урожайность в опыте, ц/га	**44,9**	**45,0**	**45,3**	**45,1**

Урожайность полевого опыта по повторностям



Плосковершинное t – распределение Стьюдента

Эти результаты недостаточно отражают объективную реальность в производственных условиях,

- т.к. стандартное отклонение, подсчитываемое по малой выборке S, меньше, чем по всей генеральной совокупности σ

- В этих случаях полагаться на критерии нормального распределения в своих выводах нельзя.

С начала 20 века в математической статистике стало разрабатываться новое направление – статистика малых выборок.

- Наиболее практическое значение для экспериментальной работы имело в 1908 году открытие

t- распределение Стьюдента

- Данный критерий был разработан Уильямом Госсетом для оценки качества пива в компании Гиннес. В связи с обязательствами перед компанией по неразглашению коммерческой тайны, статья Госсета вышла в 1908 году в журнале «Биометрика» под псевдонимом «Student» (Студент).

•вывел статистику t (t-Стьюдента), широко используемую в критериях различия средних для малых выборок.

Уильям Госсет, 1876-1937 гг.

- Плосковершинное распределение - это t- распределение Стьюдента

- Для большой выборки отклонения будут записываться - $\bar{x} \pm S\bar{x}$.
- Для малой выбоки - $\bar{x} \pm tS\bar{x}$

t- распределение Стьюдента – поправка для того, чтобы результаты отклонения, полученные по малой выборке, распространить на отклонения полученные по большой выборке.

Формула

t- распределение Стьюдента

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}} = \frac{\bar{x} - \mu}{S_{\bar{x}}}$$

- Значении t- критерия рассчитаны теоретически для 5 % и 1 % уровня значимости.

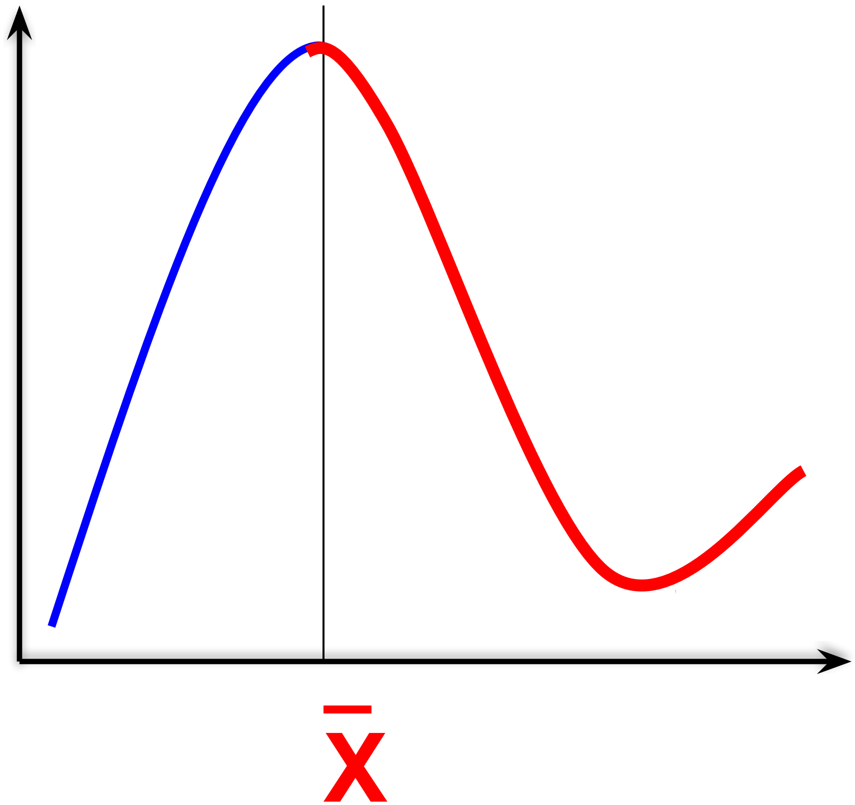
- **t- распределение Стьюдента имеет важное значение при работе с малыми выборками, позволяет определить генеральную среднюю .** σ

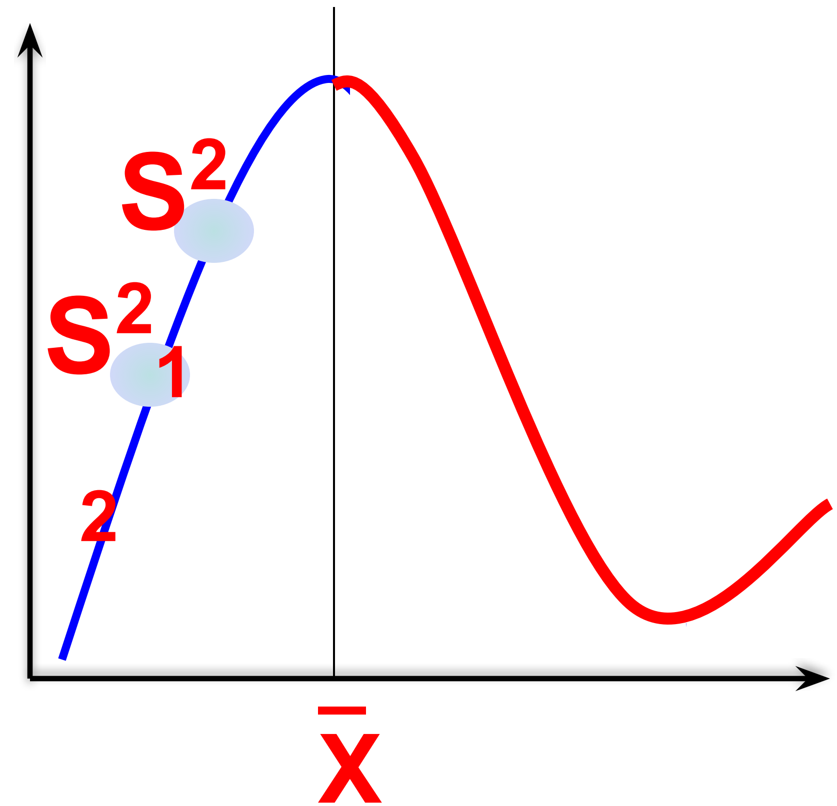
Положение 4

F - распределение Фишера

F – критерий

позволяет по части выборки судить о принадлежности ее к одной генеральной совокупности (σ) и подчинении ее закону нормального распределения (S^2)





Незавершенный вид кривой выборки с большими значениями	Незавершенный вид кривой выборки с малыми значениями.

Роналд Эйлмер Фишер, 1890-1962 гг.

Роналд Эйлмер Фишер – один из создателей современной математической статистики.

Р.Э. Фишер предложил по отношению дисперсий 2-х малых выборок из этой выборки определить принадлежит ли эта выборка к одной генеральной

совокупности.

Если рассчитать дисперсии 2-х выборок S^2_1 и S^2_2, то можно определить:

$$F_ф = \frac{S^2_1}{S^2_2}$$

$F_{ф} \geq 1$

т.к. отношение дисперсий берут таким образом, чтобы в числителе была большая дисперсия.

Если две сравниваемые выборки взяты из общей генеральной совокупности (σ), с генеральной средней (μ), то

$$F_{фак.} < F_{теор.}$$

Если генеральные средние (*μ*) сравниваемых выборок различны, то

$$F_{\text{фак.}} \geq F_{\text{теор.}}$$

F **– критерий Фишера применяется для проверки выборок на однородность .**

- Критерий Стьюдента

Двухвыборочный t-критерий для независимых выборок

С незначительно отличающимся размером выборки применяется упрощённая формула приближенных расчётов:

$$t = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

в случае неравночисленных выборок $n_1 \neq n_2$, применяется более сложная и точная формула:

$$t = \frac{|M_1 - M_2|}{\sqrt{\frac{(N_1-1)\sigma_1^2+(N_2-1)\sigma_2^2}{N_1+N_2-2}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

$$t_{\text{эмп}} = \left|\frac{\overline{X}-\overline{Y}}{Sd}\right|$$

$$Sd = \sqrt{S_x^2 + S_y^2}$$

$$Sd = \sqrt{S_x^2 + S_y^2} = \sqrt{\frac{\sum(x_i - \overline{x})^2 + \sum(y_i - \overline{y})^2}{(n_1 + n_2 - 2)} \cdot \frac{(n_1 + n_2)}{(n_1 \cdot n_2)}}$$

Двухвыборочный t-критерий для зависимых выборок

$$t = \frac{|M_d|}{\sigma_d/\sqrt{N}}$$

$$t_{\text{эмп}} = \frac{\overline{d}}{Sd}$$

$$\overline{d} = \frac{\sum d_i}{n} = \frac{\sum (x_i - y_i)}{n}$$

где Md - средняя разность значений,
а σd - стандартное отклонение разностей.
Количество степеней свободы рассчитывается как $df = N - 1$

$$d_i = x_i - y_i$$

$$Sd = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n \times (n-1)}}$$

Одновыборочный t-критерий

$$t = \frac{|M_x - A|}{\sigma/\sqrt{N}}$$

- **Пример.**

  Каждый их двух сортов ячменя высевался на пяти делянках. Получены следующие данные по урожайности.

Номер делянок	1	2	3	4	5	n
Сорт 1	**9,1**	**7,3**	**8,0**	**7,9**	**9,4**	5
Сорт 2	**7,6**	**8,2**	**5,7**	**6,1**	-	4

Задание.

Оценить достоверность различий.

Выполнение.

В случае с незначительно отличающимся размером выборки применяется упрощённая формула приближенных расчётов:

$$t = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

Наименьшая существенная разность (НСР)

НСР=t*Sd.

Sd=√S2x + S2x

df=n1+n2-2

t – значение критерия Стьюдента, соответствующее числу степеней свободы.