

**Костанайский государственный  
университет им. А. Байтурсынова**

**Шилова Н.И.**

**Основы статистической обработки  
результатов исследований**

**Цель:** Определить задачи математической статистики;  
ознакомиться с основными статистическими  
обработками результатов исследований.

# План

- 1 Задачи математической статистики. Совокупность и выборка
- 2 Статистические характеристики количественной и качественной изменчивости
- 3 Теоретические распределения
- 4 Статистические методы проверки гипотез
- 5 Понятие о нулевой гипотезе и методах ее проверки. Точечная и интервальная оценка параметров распределения.
- 6 Оценка существенности разности выборочных средних по  $t$ -распределению Стьюдента.
- 7 Проверка гипотезы о принадлежности «сомнительной» даты к совокупности.
- 8 Оценка соответствия между наблюдаемыми и ожидаемыми (теоретическими) распределениями по критерию «хи-квадрат» ( $\chi^2$ ).
- 9 Оценка различий между дисперсиями по  $f$ -распределению Фишера.

# **ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ**

**Планирование экспериментов и  
обследований**

**Оценка параметров и проверка  
гипотез**

**Принятие решений при изучении  
сложных систем**

# СОВОКУПНОСТЬ И ВЫБОРКА

**Совокупность  
или  
генеральная  
совокупность**

- Это вся группа объектов, подлежащих изучению

**Выборочная  
совокупность  
или выборка**

- Это та часть объектов, которая попала на проверку, исследование

**Выборка**

```
graph LR; A[Выборка] --- B[Большая (n>30) единиц наблюдения]; A --- C[Малая (n < 30) единиц наблюдения];
```

**Большая ( $n > 30$ )  
единиц  
наблюдения**

**Малая  
( $n < 30$ ) единиц  
наблюдения**

**Вариант** – возможные значения варьирующего признака  $X$  и обозначают  $X_1, X_2$ .

- **Ранжирование** – это расположение вариант в порядке возрастания (или убывания).

**Вариационный ряд** – это такой ряд данных, в которых указаны возможные значения варьирующего признака в порядке возрастания или убывания и соответствующие им частоты.

- **Частота** – это числа, которые характеризуют, сколько раз повторяется каждое значение признака у членов данной совокупности.





# ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ИЗМЕНЧИВОСТИ

## Количественная

Средняя арифметическая –  $\bar{x}$

Дисперсия –  $S^2$

Стандартное отклонение –  $S$

Коэффициент вариации –  $V, \%$

Абсолютная ошибка выборочной средней –  $S_{\bar{x}}$

Относительная ошибка выборочной средней –  $S_{\bar{x}}, \%$

## Качественная

Доля признака –  $p$

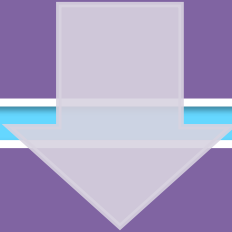
Показатель изменчивости –  $S$

Ошибка доли –  $Sp$

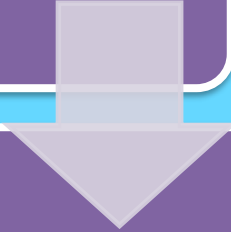
Коэффициент вариации –  $V, \%$

# СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ – $\bar{x}$

Представляет собой обобщенную, абстрактную характеристику всей совокупности в целом.



Если сумму всех вариантов ( $X_1+X_2+\dots+X_n$ ) обозначить через  $\sum X$ , а число всех вариантов через  $n$ .



$$\bar{x} = \sum X/n$$

Дисперсия представляет собой частное от деления суммы квадратов отклонений  $\sum (X - x)^2$  на число всех измерений без единицы ( $n - 1$ )

$$S^2 = \sum (X - x)^2 / (n - 1)$$

Стандартное отклонение представляет собой корень квадратный из дисперсии

$$S = \sqrt{S^2}$$

**Коэффициент вариации** – это стандартное отклонение, выраженное в процентах к средней арифметической данной совокупности.

$$V = S/\bar{x} * 100$$

>10%

незначительная

10 -20 % средняя

< 20%

значительная

## Абсолютная ошибка выборочной средней

- Мера отклонения выборочной средней от средней всей (генеральной) совокупности.
- $S_{\bar{x}} = S/\sqrt{n}$

## Относительная ошибка выборочной средней

- Ошибка выборки, выраженная в процентах от соответствующей средней.
- $S_{\bar{x}}\% = S_{\bar{x}}/\bar{x} * 100$

# КАЧЕСТВЕННАЯ ИЗМЕНЧИВОСТЬ

## Доли признака

- Это отношение численности каждого из членов ряда  $n_1, n_2, n_3$  и т.д. к численности совокупности  $N$ , т.е. вероятность появления данного признака в изучаемой совокупности.
- $p_1 = n_1/N; p_2 = n_2/N$

## Показатель изменчивости качественного признака

- Характеризует варьирование величин ряда относительно друг друга.
- $S = \sqrt{p \times q}$

## Коэффициент вариации качественных показателей

- Фактический показатель изменчивости, выраженный в процентах к максимально возможной изменчивости.
- $V, \% = S/S_{\max} * 100$

## Ошибка выборочной доли

Мера отклонения доли признака выборочной совокупности  $p$  от доли его по всей генеральной совокупности  $P$  вследствие неполной представительности (репрезентативности) выборки  $Sp = \sqrt{p \cdot q/N}$

**Распределения**

```
graph TD; A[Распределения] <--> B[Эмпирическое]; A <--> C[Теоретическое]; C --> D["Нормальное  
t- распределение  
F – распределение  
Распределение Пирсона(х  
– квадрат)  
Распределение Пуассона"]
```

**Эмпирическое**

**Теоретическое**

**Нормальное  
t- распределение  
F – распределение  
Распределение Пирсона(х  
– квадрат)  
Распределение Пуассона**

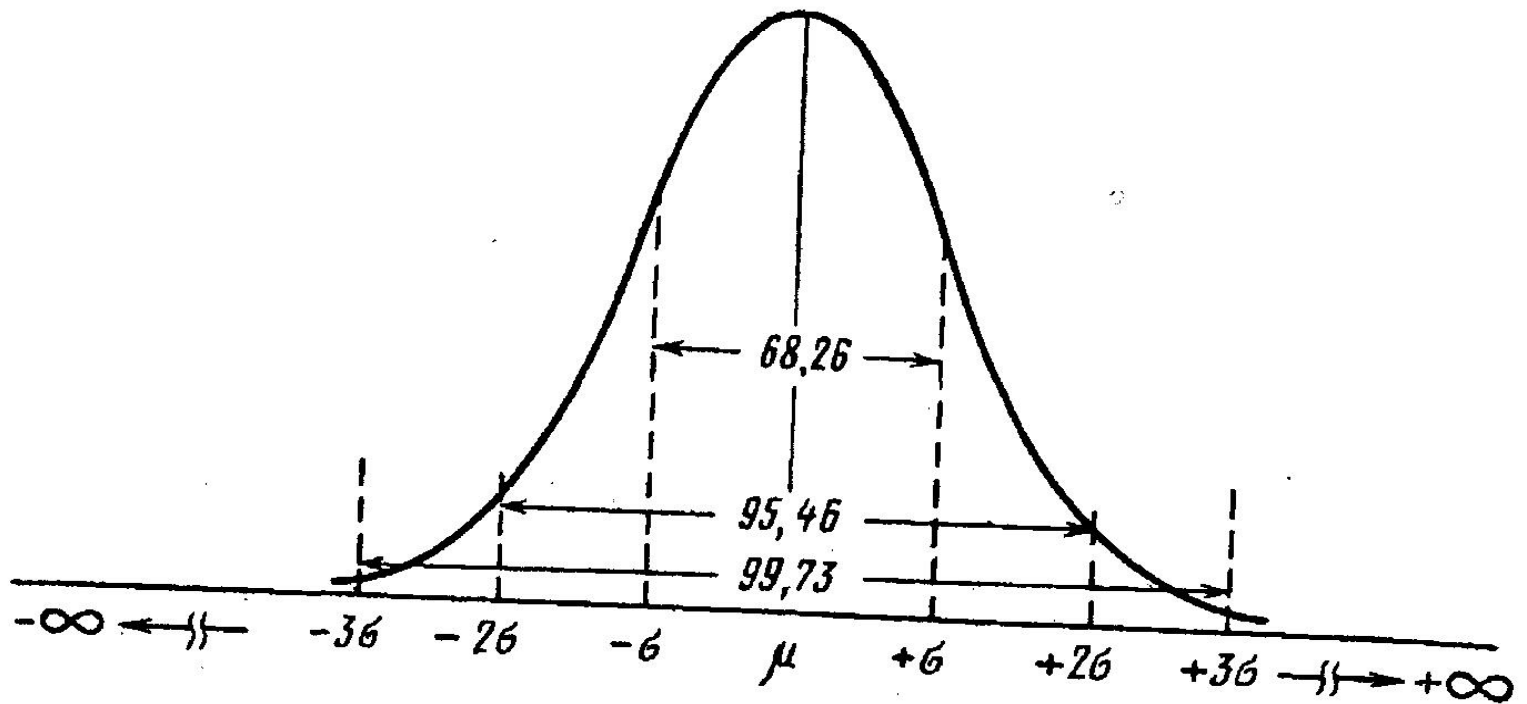
# НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

**Нормальным, или гауссовым, называют распределение вероятностей непрерывной случайной величины  $X$ .**

$$X = \mu, \text{ при } n \rightarrow \infty$$

$$x \rightarrow \mu$$





Для нормального распределения характерны следующие закономерности:

$$\bar{x} \pm 1S = 68,26\%; \quad \bar{x} \pm 2S = 95,46 \%; \quad \bar{x} \pm 3S = 99,74\%$$

$$\bar{x} \pm t S$$

# t -РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА

Максимумы частоты нормального и t -распределения совпадают, но форма кривой t -распределения зависит от числа степеней свободы.

При очень малых значениях степеней свободы она принимает вид плосковершинной кривой, причем площадь, отграниченная кривой, больше, чем при нормальном распределении

при увеличении числа наблюдений ( $n > 30$ ) распределение t приближается к нормальному и переходит в него при  $n \rightarrow \infty$ .

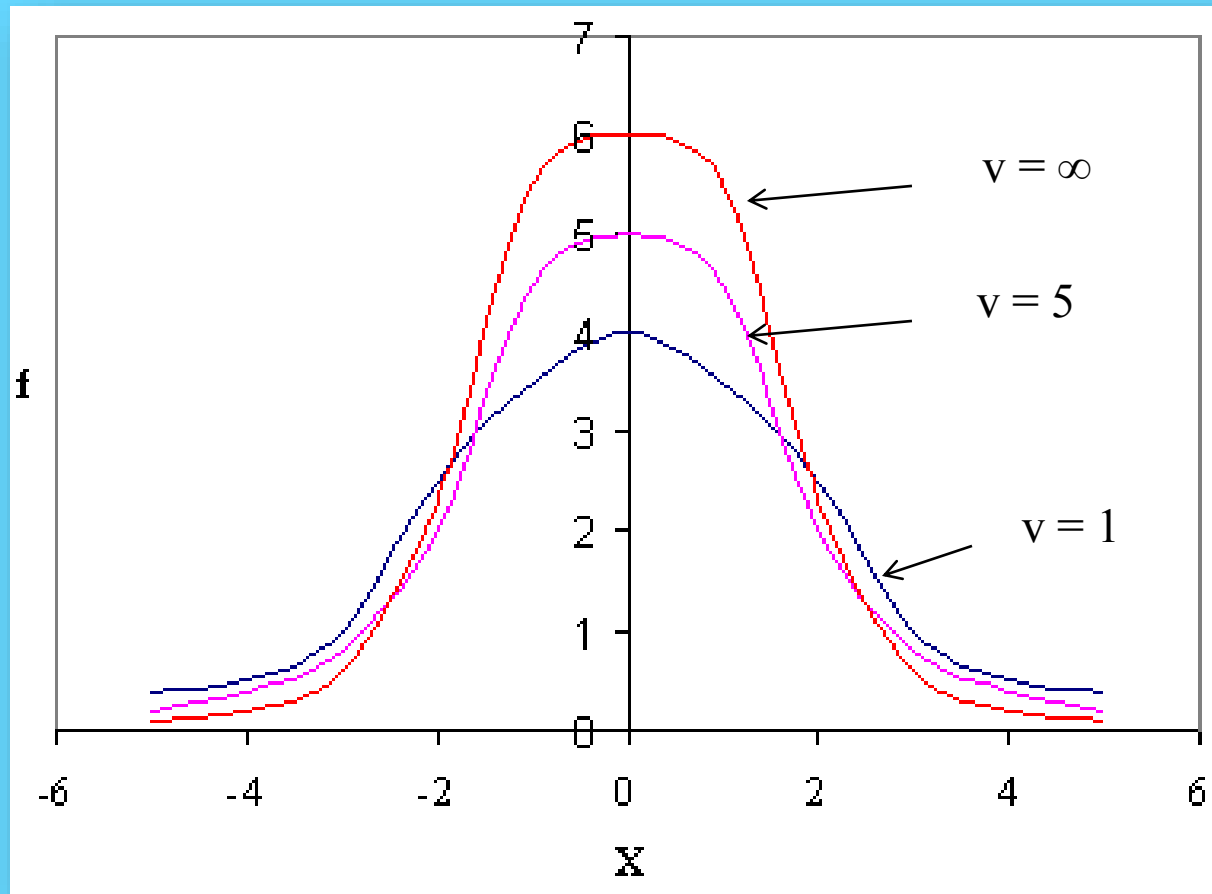


Рисунок 2 - Соотношение между нормальным  $v = \infty$  и t-распределением Стьюдента ( $v = 1$  и  $v = 5$ ).

## F – распределение Фишера

- $F = s_1^2/s_2^2$

## $\chi^2$ - распределение К.Пирсона

- $\chi^2 = \sum (f - F)^2 / F$
- Если
- $\chi^2 \text{ факт} \geq \chi^2 \text{ теор}$  – гипотеза опровергается
- $\chi^2 \text{ факт} < \chi^2 \text{ теор}$  – гипотеза не опровергается.
- $\chi^2 = 0$ , когда совпадают

## Распределение Пуассона

- $P_x = a^x e^{-a} / x!$
- Где P – вероятность значениях;
- x- число редких событий, происшедших в каждой большой группе (x=0, 1, 2, 3 и т.д.);
- a – среднее число редких событий на каждую большую группу;
- x! — произведение чисел от 1 до x (факториал); считается, что факториал нуля равен единице: 0!=1; e – основание натуральных логарифмов
- (e ≈ 2,718).

# СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОВЕРКИ ГИПОТЕЗ

Отсутствию реального различия между фактическими и теоретически ожидаемыми наблюдениями называют *нулевой гипотезой  $H_0$* .

- Для проверки гипотезы  $H_0$



**Параметрические**

- критерии, которые основаны на предположении, что распределение признака в совокупности подчиняется некоторому известному закону, например закону нормального распределения.

**Непараметрические**

- критерии, использование которых не требует предварительного вычисления оценок неизвестных параметров распределения и даже приближенного значения закона распределения признака.

# ТОЧЕЧНАЯ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

**Интервальный**

оценка, которая характеризуется двумя числами – концами интервала, покрывающего оцениваемый параметр.

**Доверительный**

интервал, который с заданной вероятностью покрывает оцениваемый параметр.

$$\bar{x} - ts_{\bar{x}} \leq \mu \leq \bar{x} + ts_{\bar{x}}$$

Крайние точки интервала – начало  $\bar{x} - ts_{\bar{x}}$  и конец  $\bar{x} \pm ts_{\bar{x}}$  называются *доверительными границами*

## Наименьшая существенная разность НСР

Величина,  
указывающая  
границу  
предельным  
случайным  
отклонениям

$$\text{НСР} = t_{sd}$$

$d \geq \text{НСР}$ , то  $H_0$   
отвергается,  
а если  $d < \text{НСР}$  - не  
отвергается

Доверительный интервал для  
разности генеральных средних  
определяется по соотношению:

$$d - \text{НСР} \leq D \leq d + \text{НСР} \text{ или } d \pm \text{НСР}$$

По величине стандартного отклонения  $s$   
оценивается *интервал* для *отдельного*  
*значения X* и *всей совокупности*:


$$\bar{x} - ts \leq \mu \leq \bar{x} + ts$$

# ОЦЕНКА СУЩЕСТВЕННОСТИ РАЗНОСТИ ВЫБОРОЧНЫХ СРЕДНИХ ПО T - РАСПРЕДЕЛЕНИЮ СТЬЮДЕНТА

В первом случае по критерию Стьюдента оценивается существенность разности средних ( $\bar{d} = \bar{x}_1 - \bar{x}_2$ ), а во втором существенность средней разности ( $\bar{d} = \Sigma d : n$ )

## Оценка разности средних независимых выборок

Ошибка разности или суммы  
средних арифметических  
независимых выборок при  
 $n_1 = n_2$

$$\rightarrow s_d = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}$$

Критерия существенности разности

$$\rightarrow t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}} = \frac{d}{s_d}$$

Нулевая гипотеза  
опровергается

$$\rightarrow t_{\text{факт}} \geq t_{\text{теор}}$$

Нулевая гипотеза не  
опровергается

$$\rightarrow t_{\text{факт}} < t_{\text{теор}}$$



# ОЦЕНКА СУЩЕСТВЕННОСТИ СРЕДНЕЙ РАЗНОСТИ (СОПРЯЖЕННЫЕ ВЫБОРКИ – РАЗНОСТНЫЙ МЕТОД)

Ошибку средней разности для сопряженных выборок  
*вычисляют разностным методом.*

Значение  
средней  
разности

$$\bar{d} = \Sigma d : n$$

Ошибка  
средней  
разности

$$s_{\bar{d}} = \sqrt{\frac{\Sigma (d - \bar{d})^2}{n(n-1)}}$$

Критерий  
существенности  
Стьюдента

$$t = \frac{\bar{d}}{s_{\bar{d}}}$$

Относительную ошибку опыта:

$$S_{x, \%} = \frac{100 \times l \times \Sigma Sd}{1,41 \times (l - 1) \times \Sigma \bar{x}}$$

# ПРОВЕРКА ГИПОТЕЗЫ О ПРИНАДЛЕЖНОСТИ «СОМНИТЕЛЬНОЙ» ДАТЫ К СОВОКУПНОСТИ

Гипотезу о принадлежности «сомнительных», наиболее уклоняющихся (крайних) вариант  $X_1$  и  $X_n$  к данной совокупности в малых выборках проверяют по критерию  $\tau$ .

Сравнивают с теоретическим на 5%-ном или 1% - ном уровне значимости.

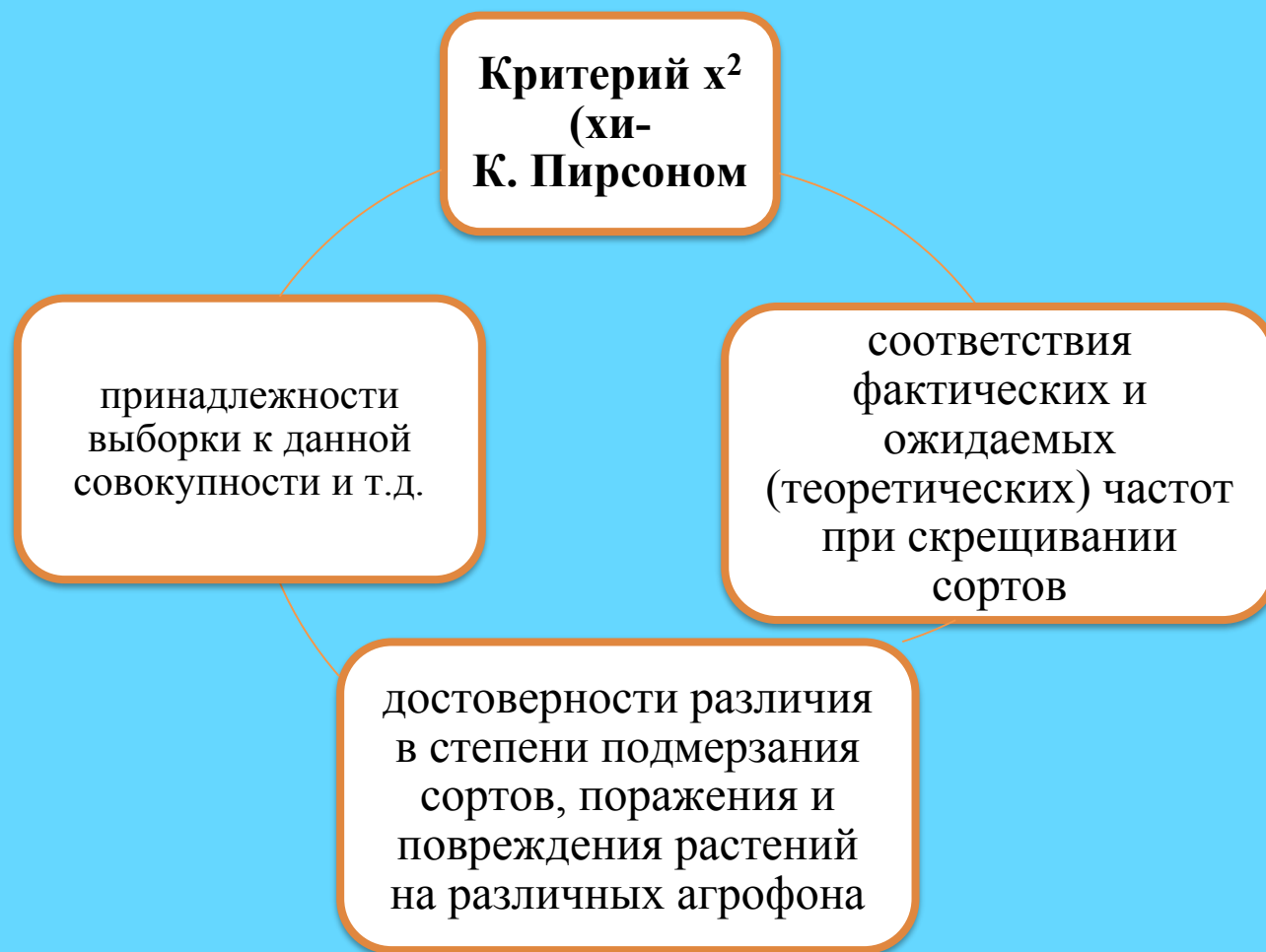
Если  $\tau_{\text{факт}} \geq \tau_{\text{теор}}$ , то дата отбрасывается, если  $\tau_{\text{факт}} < \tau_{\text{теор}}$ , то дата оставляется.

Критерий  $\tau$ :

$$\tau_{x_{\min}} = \frac{x_2 - x_{\min}}{x_{n-1} - x_{\min}} \quad \tau_{x_{\max}} = \frac{x_{\max} - x_{n-1}}{x_{\max} - x_2}$$

где  $x_2$  – дата ближайшая по величине к  $x_{\min}$ , а  $x_{n-1}$  к  $x_{\max}$

# ОЦЕНКА СООТВЕТСТВИЯ МЕЖДУ НАБЛЮДАЕМЫМИ И ОЖИДАЕМЫМИ (ТЕОРЕТИЧЕСКИМИ) РАСПРЕДЕЛЕНИЯМИ ПО КРИТЕРИЮ «ХИ-КВАДРАТ».



Если  $\chi^2_{\text{факт}} > \chi^2_{\text{теор}}$  – гипотеза отвергается,  
если  $\chi^2_{\text{факт}} < \chi^2_{\text{теор}}$  - гипотеза не отвергается.

# ОЦЕНКА РАЗЛИЧИЙ МЕЖДУ ДИСПЕРСИЯМИ ПО $F$ -РАСПРЕДЕЛЕНИЮ ФИШЕРА

Если  $F_{\text{факт}} \geq F_{\text{теор}}$

- имеются существенные различия



$F_{\text{факт}} < F_{\text{теор}}$

- различия не существенны и нулевая гипотеза не отвергается

## Контрольные вопросы:

1. Дайте определение понятиям – совокупность и выборка?
2. Какие ряды данных называют ранжированным и вариационным?
3. Какие типы изменчивости вы знаете, дайте им определение?
4. Назовите основные статистические характеристики количественной изменчивости?
5. Дайте определения статистической и нулевой гипотезе?
6. Что означает принятие и отбрасывание нулевой гипотезы?
7. Какие критерии существуют для проверки статистической гипотезы  $H_0$ ?
8. Что представляют собой точечная и интервальная оценка параметров совокупности?
9. Понятие доверительного интервала и доверительных границ?
10. Понятие наименьшей существенной разности (НСР)?
11. Оценка существенности разности средних и средней разности?
12. Сущность проверка гипотезы о принадлежности «сомнительной» даты к совокупности?
13. Область и сущность применение критерия  $\chi^2$ ?
14. Оценка различий между дисперсиями по  $F$ -распределению Фишера?

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

- 1 Доспехов Б.А. Методика полевого опыта. – М.: Агропромиздат, 1985.- 351 с.
- 2 Моисейченко В.Ф., Трифонова М.Ф., Заверюха А.Х., Ещенко В.Е. Основы научных исследований в агрономии. – М.: Колос, 1996. - 336 с.