



# МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОГРАФИИ

Карпиченко Александр Александрович

*доцент кафедры почвоведения и  
земельных информационных систем*

# Литература

- [elib.bsu.by](http://elib.bsu.by)
- Математические методы в географии: учебно-методическое пособие / Н. К. Чертко, А. А. Карпиченко. – Минск: БГУ, 2009.

## Вычисление основных показателей описательной статистики в MS Excel

Тему рассмотрим на примере двух выборок, в которых приведены площади фермерских хозяйств в Брестской и Гомельской областях. В MS Excel в ячейках A2–A3 набираем названия областей, в B2–G2 и B3–G3 – цифры площадей для каждой области.

	A	B	C	D	E	F	G
1	Площадь фермерских хозяйств						
2	Брестская обл.	300	305	315	320	330	335
3	Гомельская обл.	180	175	190	185	187	197

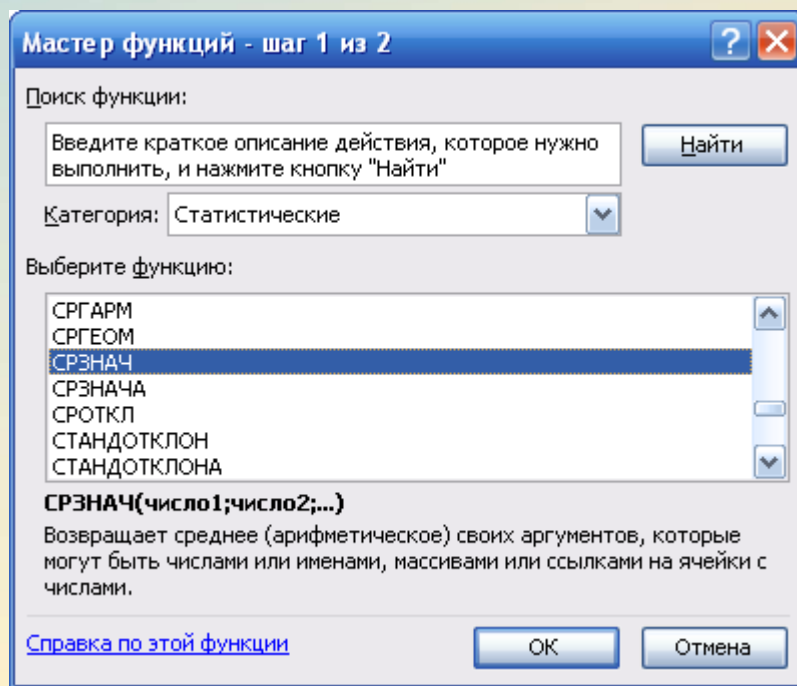
## Вычисление основных показателей описательной статистики в MS Excel

Основными статистическими показателями, характеризующими данные выборки являются: *среднее* ( $M$ ), *медиана* ( $Me$ ), *наименьшее*, *наибольшее*, *коэффициент вариации* ( $V$ ), *среднеквадратическое отклонение* ( $\sigma$ ), *дисперсия* ( $\sigma^2$ ).

В MS Excel их можно найти, используя функции **СРЗНАЧ**, **МЕДИАНА**, **МИН**, **МАКС**, **СТАНДОТКЛОН**, **ДИСП**. Коэффициент вариации находится путем деления  $\sigma$  на  $M$  с умножением на 100.

# Вычисление основных показателей описательной статистики в MS Excel

Среднее ( $M$ ) находится следующим образом: устанавливаем курсор в ячейку, где будет приведено знание средней (к примеру, H2), выполняем команду *Функция* из меню *Вставка* (или нажимаем на иконку  $f_x$  на панели инструментов), далее в категориях *Статистические* выбираем функцию СРЗНАЧ, нажимаем *ОК*, сворачиваем появившееся окно нажатием на кнопку напротив поля **Число 1**. Выделяем ячейки со значениями площадей для первой области (B2:G2), разворачиваем окно, нажав на эту же кнопку и ждем *ОК*.

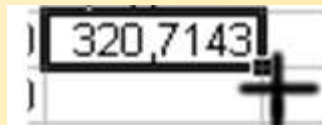


## Вычисление основных показателей описательной статистики в MS Excel

Показанный пример не является единственно возможным или оптимальным вариантом вычисления функций в MS Excel, большинство задач можно выполнить 2-3-мя способами, выбор которых скорее определяется навыком и привычками пользователя.

Например, можно не использовать “Мастер функций”, а просто поставить в требуемой ячейке знак “=” и вручную набрать требуемую формулу, помня, что команды обычно пишутся кириллицей, а обозначения ячеек таблицы – латиницей.

Для второй области можно не делать описанную выше процедуру, а воспользоваться функцией автозаполнения: выделяем ячейку с найденным средним значением для первой области (H2), и наведя курсор на правый край клетки H2 до превращения курсора в «крестик», удерживая левую кнопку мыши растягиваем выделение на нижележащую клетку (I3), в которой появится значение для второй области.



## Вычисление основных показателей описательной статистики в MS Excel

Расчет базовых статистических показателей также может производиться с использованием надстройки (опции) «**Пакет анализа**», которая позволяет оперативно получить значения показателей описательной статистики. По умолчанию эта опция не установлена, поэтому для ее активации необходимо с помощью команды **Надстройки** из меню *Сервис* открыть окно диалога «*Надстройки*» и в нем установить флажок для компоненты «*Пакет анализа*». После нажатия кнопки *ОК* меню *Сервис* будет дополнено командой *Анализ данных*.

В Excel 2010: Закладка «*Данные*».

## Вычисление основных показателей описательной статистики в MS Excel

Для расчета показателей выполняем последовательность команду *Анализ данных* из меню *Сервис* в диалоговом окне *Анализ данных* выбираем *Описательная статистика – ОК*, в поле «*Входной интервал*» указываем наш (клетки **A2: G3**), в поле «*группирование*» выбираем «*по строкам*», ставим галочку у «*Метки в первом столбце*» в «*Параметрах вывода*» выбираем «*Выходной интервал*» и указываем там ячейку **B5** или другую свободную, отмечаем параметры «*Итоговая статистика*» и «*Уровень надежности*» (значение можно изменять, в нашем случае указываем **95%**), нажимаем **ОК**. **Выставление показателя метки помечает столбцы или строки, при этом первые значения в столбцах/строках в расчете не участвуют.**



# Вычисление основных показателей описательной статистики в MS Excel

Нахождение сходства или отличия между двумя выборками проводится с помощью t-теста (критерия Стьюдента). **Выбор конкретной команды зависит от типа выборки** (независимая/сопряженная) и от величин дисперсий. Так, если в вашем задании **независимая** выборка с **различными дисперсиями** выполняются следующие действия: *Сервис – Анализ данных – Двухвыборочный t-тест с различными дисперсиями – ОК*. Для **независимой** выборки с **одинаковыми дисперсиями** алгоритм следующий: *Сервис – Анализ данных – Двухвыборочный t-тест с одинаковыми дисперсиями – ОК*, для **сопряженной** выборки: *Сервис – Анализ данных – Парный двухвыборочный t-тест для средних – ОК*.

# Вычисление основных показателей описательной статистики в MS Excel

Двухвыборочный t-тест с различными дисперсиями

	Брестская обл	Гомельская обл
Среднее	320,71	187,71
Дисперсия	228,57	78,57
Наблюдения	7	7
Гипотетическая разность средних	0	
df	10	
t-статистика	20,08	
P(T<=t) одностороннее	1,03273E-09	
t критическое одностороннее	1,81	
P(T<=t) двухстороннее	2,06546E-09	
t критическое двухстороннее	2,23	

В полученных данных  $df$  – число степеней свободы; *t-статистика* – критерий Стьюдента (фактический); *t критическое двухстороннее* – критерий Стьюдента (табличный). На основании соотношения *t-статистики* (берется по модулю) и *t критического двухстороннего* делается вывод о достоверности различия выборок.

В данном случае *t-статистика 20,08 > t критического двухстороннего 2,23*, следовательно, отличие двух областей по площади фермерских хозяйств доказано, при районировании их нельзя объединять в одну группу.

# Вычисление основных показателей описательной статистики в MS Excel

По первой лабораторной работе надо будет сделать выводы по:

Коэффициентам вариации ( $V$ ) первой и второй выборок – установить характер варьирования признака.

По наличию сходства или отличия выборок, если *t-статистика*  $>$  или  $=$  *t критическое двухстороннее*, то считается, что **отличие между выборками доказано** и при классификации их нужно выделять в разные группы.

Если же *t-статистика*  $<$  *t критическое двухстороннее*, то **отличие не доказано** (т.е. доказано сходство) и выборки нужно объединять в одну группу.