# Лабораторная работа.
# Использование электронных таблиц MS Excel для проверки статистических гипотез о сравнении средних статистических совокупностей с известными дисперсиями

**Цель** работы – научить с помощью MS Excel определять достоверности различий средних арифметических двух статистических совокупностей.

## 1. Проверка гипотезы о равенстве средних с известными дисперсиями

Пусть даны две нормально распределенные переменные $X_1$ и $X_2$, известны соответствующие значения их дисперсий ($D_1 = \sigma_1^2$ и $D_2 = \sigma_2^2$).

Гипотезы формулируются следующим образом. Нулевая гипотеза $H_0$: генеральные средние значения равны, альтернативная гипотеза $H_1$: генеральные средние значения различны.

Для проверки этой гипотезы используется процедура **Двухвыборочный *z*-тест для средних**. Для этого необходимо выполнить: **Данные / Анализ данных / Двухвыборочный *z*-тест для средних.**

### АЛГОРИТМ

1. Выбрать уровень значимости $\alpha$. Сформулировать гипотезы:

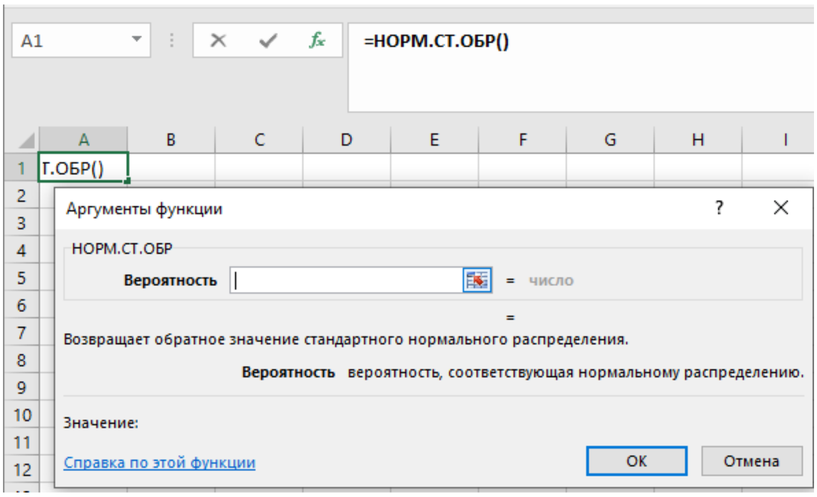
$H_0$: $\overline{X}_1 = \overline{X}_2$, то есть генеральные средние равны.

$H_1$: 1) $\overline{X}_1 \neq \overline{X}_2$, 2) $\overline{X}_1 > \overline{X}_2$, 3) $\overline{X}_1 < \overline{X}_2$.

2. Найти эмпирическое значение критерия по формуле: $z_{эмп} = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{D_1}{n_1} + \dfrac{D_2}{n_2}}}$,

где $\bar{x}_1$ и $\bar{x}_2$ – выборочные средние для нормально распределенных переменных $X_1$ и $X_2$ соответственно, $D_1$ и $D_2$ – известные генеральные дисперсии.

3. Найти критическое значение критерия с помощью статистической функции **НОРМ.СТ.ОБР**(*вероятность*), при аргументе *вероятность*, равном $\Phi(z) + 0{,}5$, где $\Phi(z)$ – интегральная функция Лапласа: 1) $\Phi(z) = \Phi(z_{крит}) = \dfrac{p}{2} = \dfrac{1-\alpha}{2}$; 2) и 3) случаи $\Phi(z) = \Phi(z_{крит}) = \dfrac{1-2\alpha}{2}$.



4. Сравнить эмпирическое и критическое значения критерия.

Если $|z_{эмп}| < z_{крит}$, то на уровне значимости $\alpha$ принимается нулевая гипотеза $H_0$, иначе – альтернативная.

**Пример.** В результате измерений роста $n_1 = 70$ случайным образом отобранных студентов I курса и $n_2 = 80$ студентов II курса были получены следующие результаты: для студентов I курса средний рост составил 176 см, оценки дисперсий роста генеральных совокупностей равны 1,1 см²; для студентов II курса – 178,5 см; 1,6 см². Выяснить, можно ли на основании полученных результатов утверждать о значимом различии средних значений роста студентов I и II курсов (представляющих генеральную совокупность) на уровне значимости $\alpha = 0{,}05$.

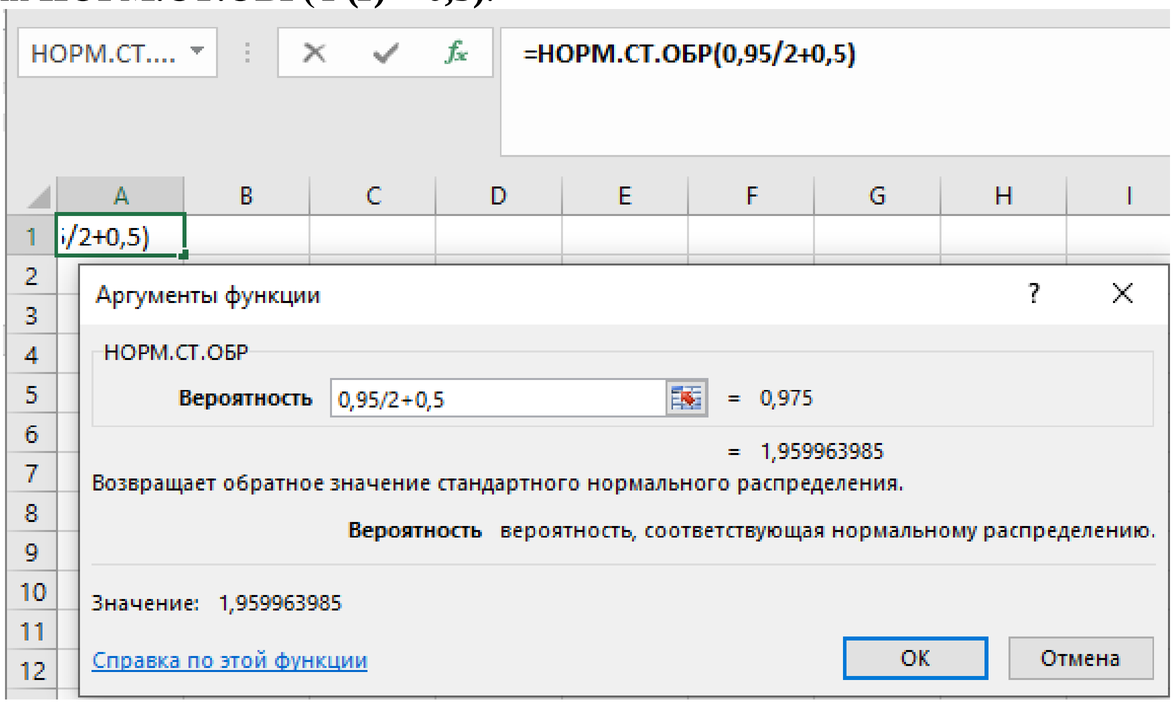
Решение. Иными словами, необходимо ответить на вопрос о значимости или незначимости (существенности или несущественности) полученного при измерениях различия в средних значениях роста.

Сформулируем гипотезы. Нулевая гипотеза $H_0$: различия средних значений роста студентов I и II курсов (представляющих генеральную совокупность) статистически незначимы. Альтернативная гипотеза: различия средних значений роста студентов I и II курсов (представляющих генеральную совокупность) статистически значимы (первый случай).

Найдём эмпирическое значение критерия:

$$z_{эмп} = \frac{|176 - 178{,}5|}{\sqrt{\frac{1{,}1}{70} + \frac{1{,}6}{80}}} = 13{,}23.$$

Критическое значение критерия можем вычислить с помощью статистической функции **НОРМ.СТ.ОБР(*Ф*(z) + 0,5)**:




Таким образом, для первого случая получаем $\Phi\left(z_{\text{крит}}\right) = \frac{1 - 0{,}05}{2} = \frac{0{,}95}{2} = 0{,}475$, то есть $z_{крит} = 1{,}96$.

Следовательно, $z_{эмп} > z_{крит}$, так как 13,23 > 1,96, и на уровне значимости 0,05 отвергаем нулевую гипотезу, то есть можем сделать вывод о возможной значимости экспериментально наблюдаемого различия в среднем росте студентов.

## 2. Использование инструмента Пакет анализа

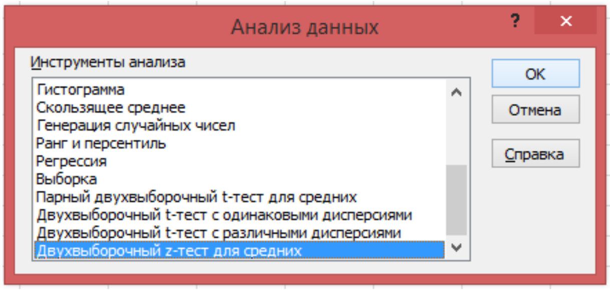
Для использования статистического пакета анализа данных необходимо:

- указать курсором мыши на пункт главного меню **Данные** и щелкнуть левой кнопкой мыши;

- в крайнем правом углу главного меню выбрать команду **Анализ данных** (если команда Анализ данных отсутствует, то необходимо установить в Excel пакет анализа данных);

- выбрать строку:

для *произвольных* статистических совокупностей с известными дисперсиями применяется **Двухвыборочный *z*-тест для средних**.




В диалоговом окне задаются следующие параметры:

*Интервал переменной <u>1</u>* – адреса ячеек, содержащих выборочные значения случайной величины $X_1$;

*Интервал переменной <u>2</u>* – адреса ячеек, содержащих выборочные значения случайной величины $X_2$;

*Гипотетическая средняя разность* задаёт число, равное предполагаемой разности математических ожиданий (их разность равна 0);

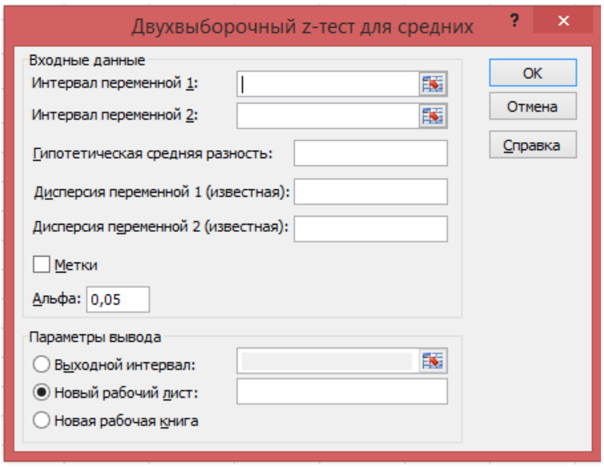
*Дисперсия переменной 1 (известная)* – известное значение дисперсии первой переменной;

*Дисперсия переменной 2 (известная)* – известное значение дисперсии второй переменной;

*Метки* – включается, если учитываются заголовки данных;

*Альфа* – уровень значимости (по умолчанию 0,05, то есть надёжность 95%);

*Выходной интервал* – указывается, куда выводятся результаты вычислений.



## Задачи для самостоятельного решения

**Задача 1.** Были собраны данные о диаметре валиков (мм), изготовленных двумя автоматическими станками. Предварительным анализом было установлено, что размер валиков имеет нормальное распределение с дисперсиями $\sigma_1^2$=5 мм$^2$ (для первого станка), $\sigma_2^2$= 7 мм$^2$ (для второго станка). Выяснить, можно ли утверждать о значимом различии средних.

| № | 1 станок | 2 станок |
|---|---|---|
| 1 | 178,9 | 184,0 |
| 2 | 179,7 | 184,1 |
| 3 | 179,8 | 184,2 |
| 4 | 179,8 | 184,5 |
| 5 | 180,9 | 185,2 |
| 6 | 181,4 | 183,3 |
| 7 | 181,5 | 185,6 |
| 8 | 181,6 | 185,8 |
| 9 | 181,7 | 186,2 |
| 10 | 181,8 | 186,4 |
| 11 | 182,3 | |
| 12 | 182,4 | |
| 13 | 182,5 | |
| 14 | 183,0 | |
| 15 | 183,2 | |

**Задача 2.** Были составлены две выборки: равномерно и случайным образом со всей площади посева каждого изучаемого сорта берут на исследование по 50 или более растений. Пусть для анализа сорта пшеницы выбран хозяйственно важный признак – число колосков в колосе, и собрано по 100 колосьев с разных растений каждого сорта, далее подсчитывается число колосков ($X_1$ для первого и $X_2$ для второго сортов, их объёмы $n_1$ и $n_2$) в каждом колосе и составляются два вариационных ряда для первого и второго сортов. Необходимо исследовать на возможную высокую урожайность два сорта пшеницы.

| $X_1$ | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| $n_1$ | 3 | 8 | 16 | 19 | 21 | 17 | 9 | 5 | 2 |

| $X_2$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_2$ | 1 | 4 | 4 | 7 | 12 | 18 | 22 | 15 | 8 | 6 | 3 |