



## Тема 10. Сравнение двух выборок

---

- 10.1. Гипотеза о равенстве средних. Независимые выборки
- 10.2. Доверительный интервал для разности средних
- 10.3. Гипотеза о равенстве средних. Парные выборки
- 10.4. Доверительный интервал для разности средних. Парные выборки
- 10.5. Гипотеза о равенстве дисперсий
- 10.6. Гипотеза о равенстве долей
- 10.7. Доверительный интервал для разности долей

# 1. Независимые выборки



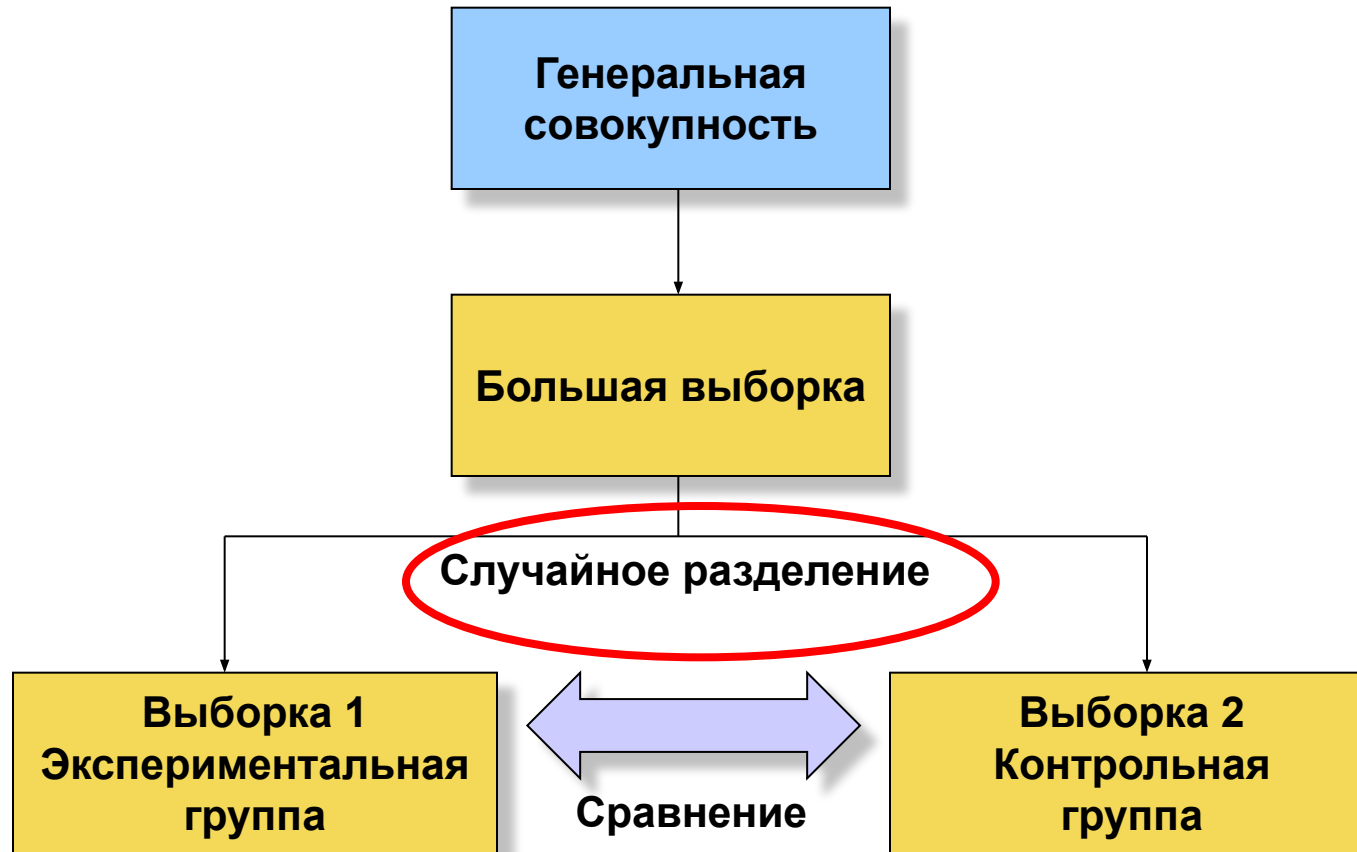
## 1.1. Две генеральные совокупности, две независимые выборки



# 1. Независимые выборки



## 1.2. Одна генеральная совокупность, две независимые выборки



## 2. Зависимые выборки



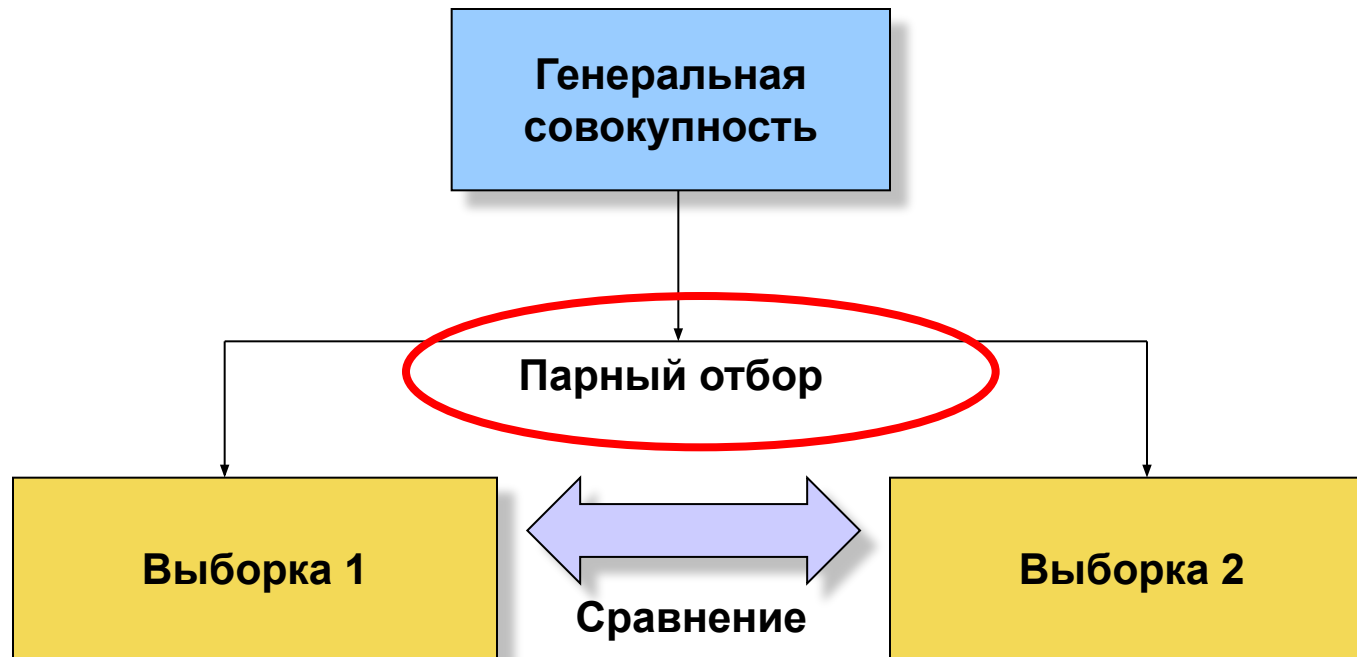
### 2.1. Две генеральные совокупности, две зависимые выборки



## 2. Зависимые выборки



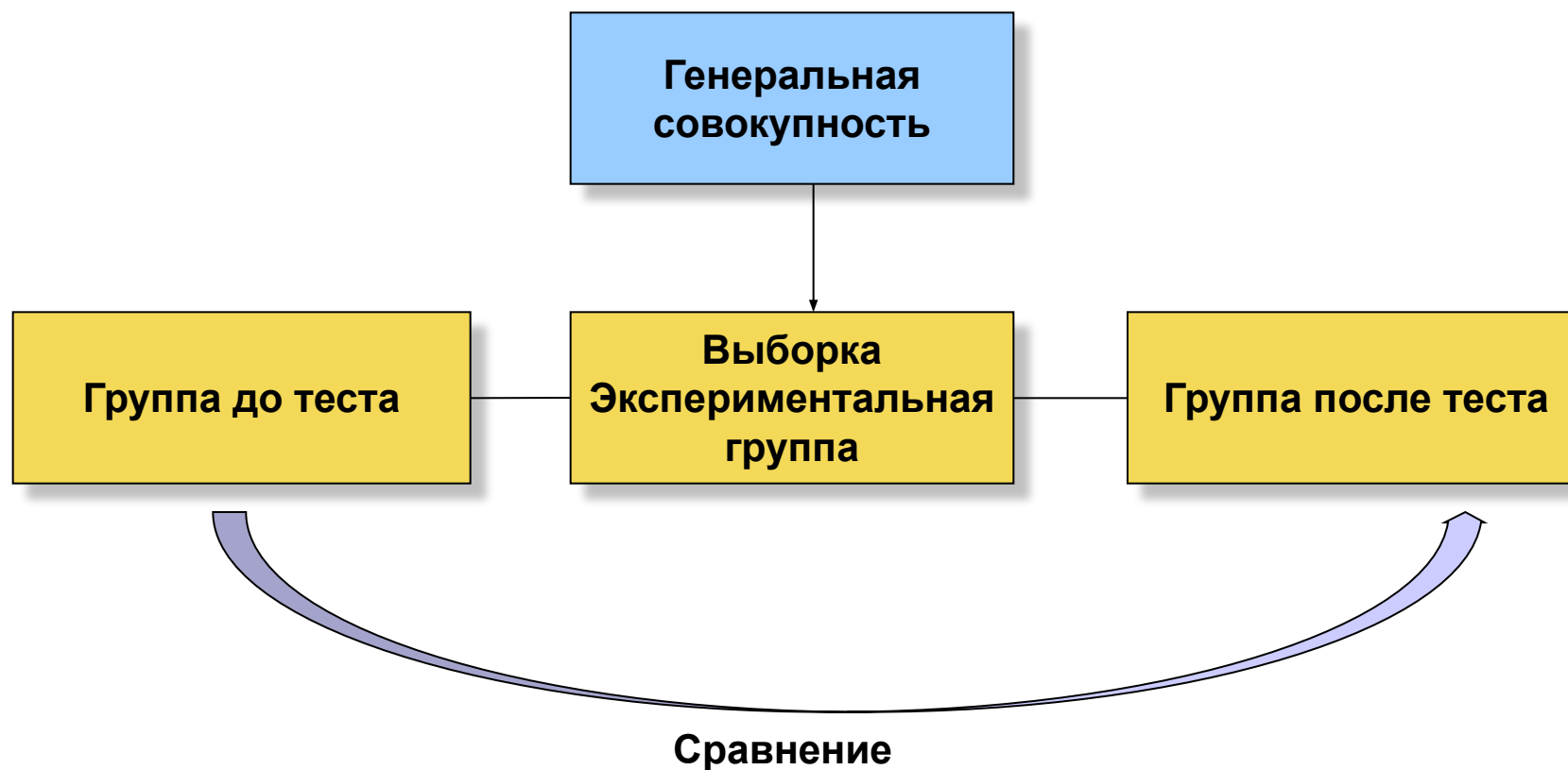
### 2.2. Одна генеральная совокупность, две зависимые выборки



## 2. Зависимые выборки



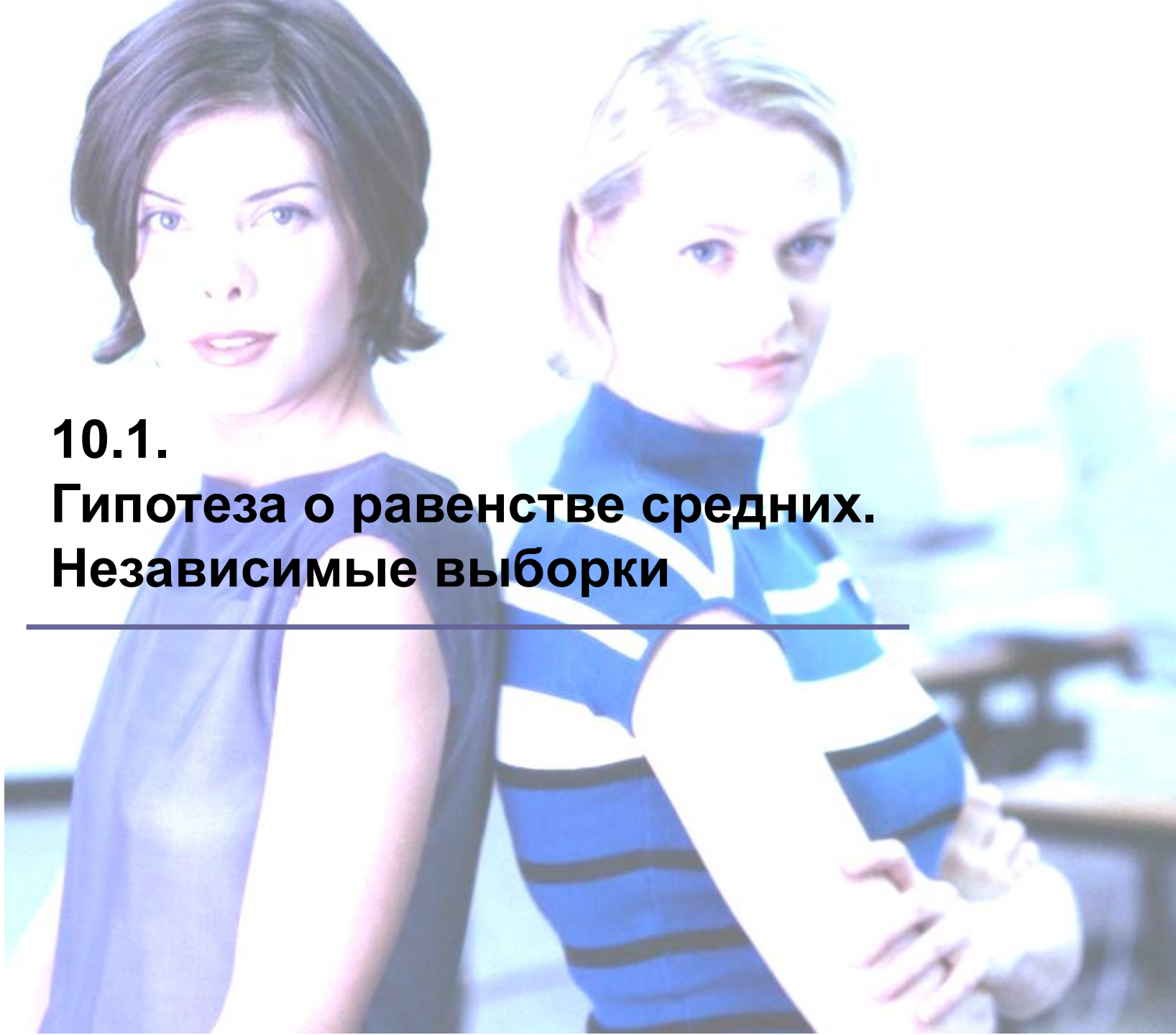
### 2.3. Одна генеральная совокупность, две зависимые выборки до и после теста





**10.1.**  
**Гипотеза о равенстве средних.**  
**Независимые выборки**

---





## Что мы имеем

1. Две простые случайные выборки, полученные из двух генеральных совокупностей
2. Выборки являются независимыми. Это значит, что между субъектами в каждой из выборок нет связи.
3. Обе выборки имеют объем  $n \geq 30$ . Если нет, то обе выборки взяты из нормально распределенных генеральных совокупностей.

## Что мы хотим

Проверить гипотезу о равенстве средних двух генеральных совокупностей:

$$H_0 : \mu_1 = \mu_2$$





Нулевая гипотеза:

$$H_0 : \mu_1 = \mu_2$$

Это равносильно гипотезе:

$$H_0 : \mu_1 - \mu_2 = 0$$

Альтернативная гипотеза:

$$H_1 : \mu_1 \neq \mu_2$$

# Односторонние гипотезы



Нулевая гипотеза:

$$H_0 : \mu_1 \leq \mu_2$$

$$H_0 : \mu_1 \geq \mu_2$$

Или, что равносильно:

$$H_0 : \mu_1 - \mu_2 \leq 0$$

$$H_0 : \mu_1 - \mu_2 \geq 0$$

Альтернативная гипотеза:

$$H_1 : \mu_1 > \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

**Правосторонний критерий**

**Левосторонний критерий**

# 1. Статистика ( $\sigma_1$ и $\sigma_2$ известны)



Для проверки гипотезы используется статистика:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

где

$\bar{x}_1$   $\bar{x}_2$  - выборочные средние

$\sigma_1^2$   $\sigma_2^2$  - известные дисперсии генеральных совокупностей

$n_1$   $n_2$  - объемы выборок

# Почему выбирает этот критерий?



В качестве критерия мы выбираем:

$$\frac{\text{Наблюдаемое значение} - \text{Ожидаемое значение}}{\text{Стандартная ошибка}}$$

Наблюдаемое значение:

$$\bar{x}_1 - \bar{x}_2$$

Ожидаемое значение:

$$\mu_1 - \mu_2 = 0$$

Стандартная ошибка:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Это следует из формулы для дисперсии разности выборочных средних:

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

# Последовательность действий

---



**Шаг 1.** Сформулировать основную и альтернативную гипотезы.

**Шаг 2.** Задать уровень значимости  $\alpha$ .

**Шаг 3.** По таблице найти критические значения и построить критическую область.

**Шаг 4.** По выборке сосчитать значение статистики.

**Шаг 5.** Сравнить полученное значение с критической областью. Если значение попало в критическую область – отклонить основную гипотезу, не попало – принять.

**Шаг 6.** Написать ответ.

## Задача. Стоимость ремонта



Частная исследовательская фирма тестировала две различных марки автомобиля, с целью определить, есть ли различие в ущербе, получаемом машиной, если она попадает в аварию на скорости 10 миль в час. Ниже представлен средний ущерб в долларовом эквиваленте для каждой марки. Предполагается, что генеральная совокупность распределена нормально.

$$\bar{x}_1 = \$1560$$

$$\bar{x}_2 = \$1432$$

$$\sigma_1^2 = \$252$$

$$\sigma_2^2 = \$280$$

$$n_1 = 20$$

$$n_2 = 15$$

На уровне значимости  $\alpha = 0,05$  проверьте утверждение о том, что различия в долларовом эквиваленте понесенного каждым типом автомобиля ущерба не существует.



**Шаг 1.** Основная и альтернативная гипотезы:

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$$

**Шаг 2.** Задан уровень значимости  $\alpha=0,05$ .

**Шаг 3.** По таблице находим критическое значение  $z = 1,96$  и строим критическую область:

$$z < -1,96 \quad z > 1,96$$

**Шаг 4.** По выборке сосчитаем значение статистики:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(1560 - 1432) - 0}{\sqrt{\frac{252^2}{20} + \frac{280^2}{15}}} = 1,40$$

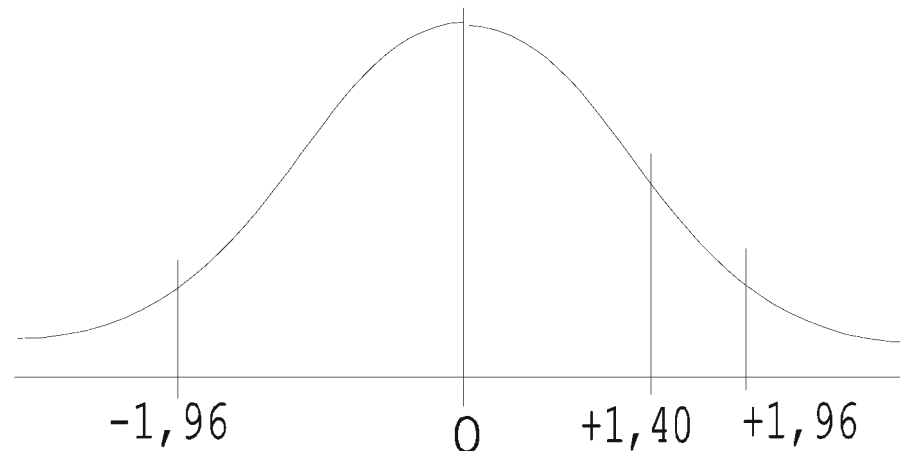


**Шаг 5.** Сравним полученное значение с критической областью.

$$1,40 < 1,96$$

Полученное значение статистики не попало в критическую область.  
Мы принимаем основную гипотезу.

**Шаг 6.** Формулируем вывод: мы не имеем достаточных оснований, чтобы отвергнуть утверждение о равенстве средних.





# Если дисперсии не известны...

---



Как проверить гипотезу о равенстве средних, если дисперсии генеральной совокупности не известны?

Z-критерий не подойдет. Вместо него используем t-критерий.

Существует два варианта.

**Первый**, когда мы ничего не знаем о дисперсиях. **Второй**, когда мы не знаем значения дисперсий генеральных совокупностей, но у нас есть основания полагать их равными. В этих случаях статистика строится по разному.

Разберем подробнее.

## 2. Статистика ( $\sigma_1$ и $\sigma_2$ не известны и не равны)



Для проверки гипотезы используется статистика:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

где

$\bar{x}_1$   $\bar{x}_2$  выборочные средние

$s_1^2$   $s_2^2$  выборочные дисперсии

$n_1$   $n_2$  объемы выборок

$$df = \min(n_1 - 1, n_2 - 1)$$

### 3. Статистика (предполагаем $\sigma_1 = \sigma_2$ )



Для проверки гипотезы используется статистика:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

где

$\bar{x}_1$   $\bar{x}_2$  - выборочные средние

$s_p^2$  - объединенная дисперсия двух выборок

$n_1$   $n_2$  - объемы выборок

$$df = n_1 + n_2 - 1$$

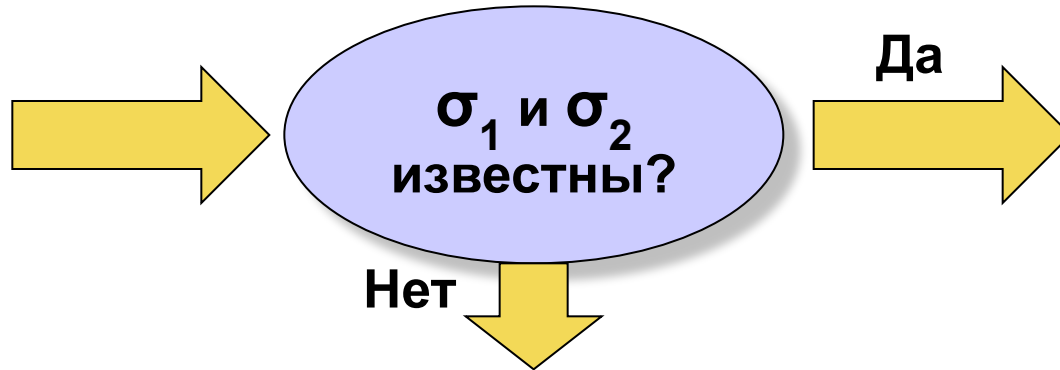
# Объединенная дисперсия (Pooled variance)



Если нам известно, что дисперсии генеральных совокупностей равны, или мы проверяем гипотезу, что случайные выборки получены из одной совокупности, нам следует вычислить для начала объединенную дисперсию для двух выборок:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

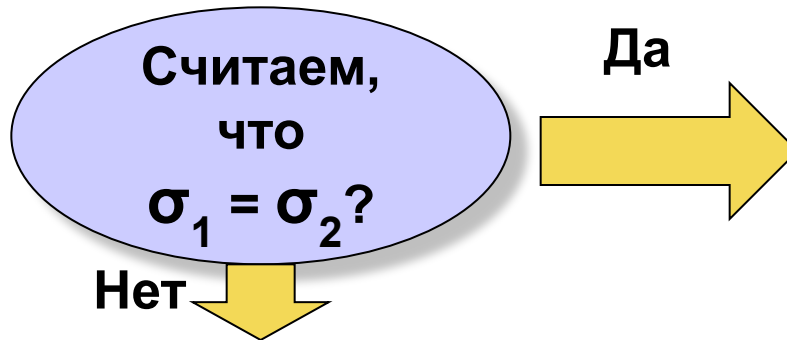
# Выводы о средних (независимые выборки)



Используем нормальное распределение со стандартной ошибкой:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

**Редкий случай!**



Используем t-распределение с объединенной дисперсией и ошибкой:

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Используем t-распределение со стандартной ошибкой:

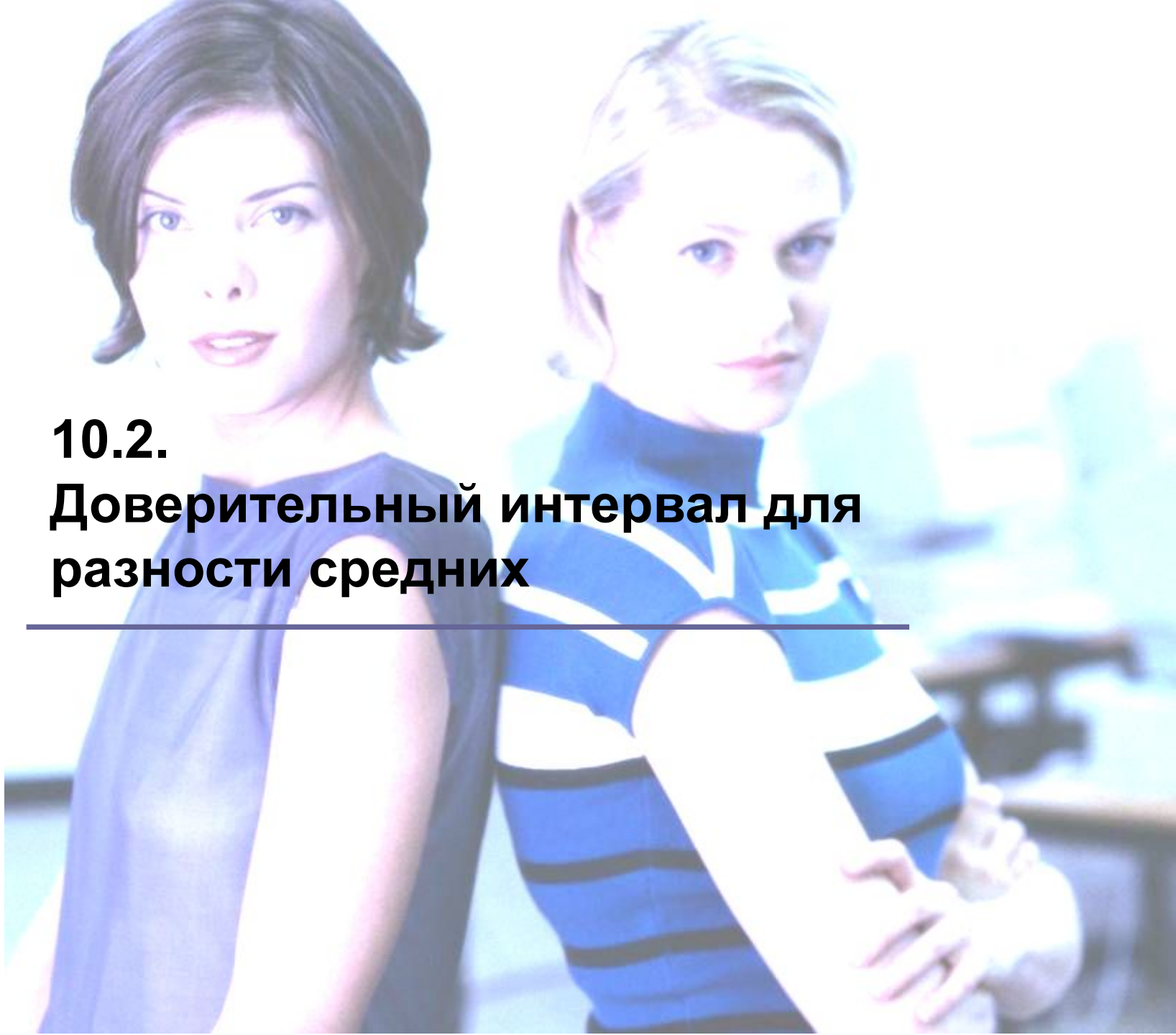
$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

**Наиболее часто!**



**10.2.**  
**Доверительный интервал для  
разности средних**

---





## Что мы имеем

Имеем две **простые случайные, независимые** выборки объема  $n_1$  и  $n_2$  из двух генеральных совокупностей.

Генеральные совокупности имеют нормальный закон распределения с параметрами  $\mu_1, \sigma_1$  и  $\mu_2, \sigma_2$  либо объемы обеих выборок  $\geq 30$ .

## Что мы хотим

Оценить разницу  $(\mu_1 - \mu_2)$  между средними двух генеральных совокупностей. Для этого построить доверительный интервал для разности средних в виде:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

# Доверительный интервал для разности средних (1)



Среднее генеральной совокупности с надежностью  $1-\alpha/2$  находится в доверительном интервале:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

Стандартные отклонения  $\sigma_1$  и  $\sigma_2$  известны. Тогда:

$$E = z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$



## Доверительный интервал для разности средних (2)



Среднее генеральной совокупности с надежностью  $1-\alpha/2$  находится в доверительном интервале:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

Стандартные отклонения  $\sigma_1$  и  $\sigma_2$  неизвестны и не подразумеваются равными. Тогда:

$$E = t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \min(n_1 - 1, n_2 - 1)$$

## Доверительный интервал для разности средних (3)



Среднее генеральной совокупности с надежностью  $1-\alpha/2$  находится в доверительном интервале:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

Стандартные отклонения  $\sigma_1$  и  $\sigma_2$  неизвестны, но подразумеваются равными. Тогда:

$$E = t_{\alpha/2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$df = n_1 + n_2 - 2$$

# Пример



Преподаватель хочет оценить различия в оценках студентов-вечерников и студентов-дневников. Ниже приведены результаты экзамена. Построить 95% доверительный интервал для разности средних баллов.

**Дневники**

**Вечерники**

$$n_1 = 26$$

$$n_2 = 18$$

$$\bar{x}_1 = 4,2$$

$$\bar{x}_2 = 4,4$$

$$s_1 = 0,3$$

$$s_2 = 0,25$$

**Решение. На семинаре**



**10.3.**  
**Гипотеза о равенстве средних.**  
**Парные выборки**

---





## Что мы имеем

1. Две простые случайные выборки, полученные из двух генеральных совокупностей
2. Выборки являются парными (зависимыми)
3. Обе выборки имеют объем  $n \geq 30$ . Если нет, то обе выборки взяты из нормально распределенных генеральных совокупностей.

## Что мы хотим

Проверить гипотезу о разности средних двух генеральных совокупностей:

$$H_0 : \mu_d = 0$$



Для проверки гипотезы используется статистика:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} \quad df = n - 1$$

где

- $d$  - разность между двумя значениями  $x - y$  в одной паре
- $\mu_d$  - среднее для парных разностей генеральной совокупности
- $\bar{d}$  - среднее для парных разностей для выборки
- $s_d$  - стандартное отклонение разностей для выборки
- $n$  - количество пар

# Пример. Тренинг студентов



Студент	До	После	$d$	$d^2$
1	90	93	-3	9
2	91	90	1	1
3	93	89	4	16
4	89	88	1	1
5	85	88	-3	9
6	89	86	3	9
7	83	84	-1	1
8	88	83	5	25
9	84	83	1	1
10	82	80	2	4
11	83	77	6	36
12	81	76	5	25
13	72	74	-2	4
14	70	70	0	0
15	71	69	2	4
			$\Sigma=21$	$\Sigma=145$

Группа из 15 студентов прошла тест до тренинга и после. Результаты теста в таблице. Проверим гипотезу для парных выборок на отсутствие влияния тренинга на подготовку студентов на уровне значимости 0,05.

**Решение.** Подсчитаем разности и их квадраты.



**Шаг 1.** Основная и альтернативная гипотезы:

$$H_0 : \mu_d = 0 \quad H_1 : \mu_d > 0$$

**Шаг 2.** Задан уровень значимости  $\alpha=0,05$ .

**Шаг 3.** По таблице для  $df = 15 - 1=14$  находим критическое значение  $t = 2,145$  и строим критическую область:

$$t > 2,145$$

**Шаг 4.** По выборке сосчитаем значение статистики.

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} \Rightarrow s_d = \sqrt{\frac{145 - \frac{21^2}{15}}{15-1}} = 2,87$$





Статистика принимает значение:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{1,4 - 0}{\frac{2,87}{\sqrt{15}}} = 1,889$$

Среднее значение разностей получено делением 21 на 15 и равно 1,4.



**Шаг 5.** Сравним полученное значение с критической областью.

$$1,889 < 2,145$$

Полученное значение статистики не попало в критическую область.

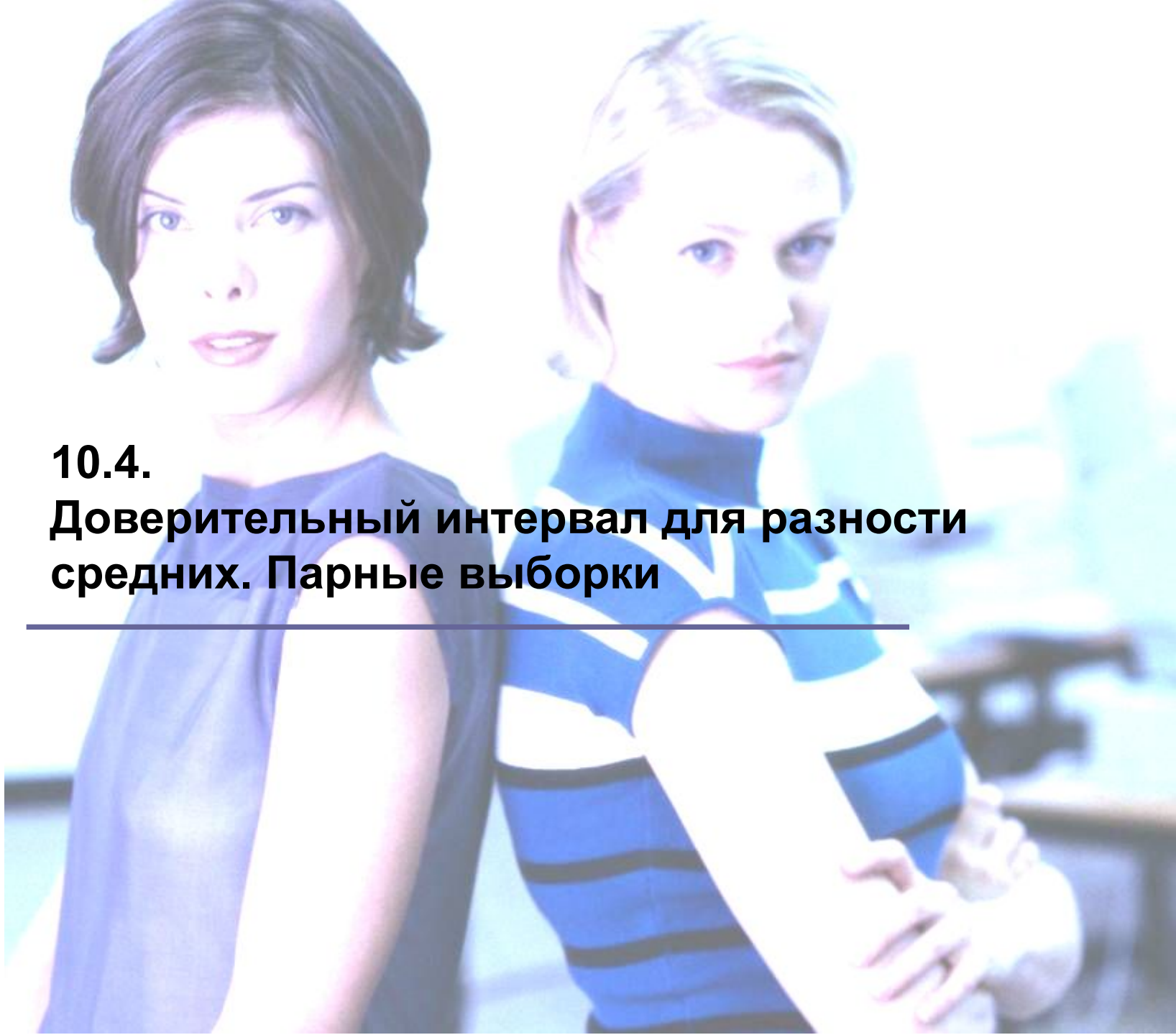
**Шаг 6.** Формулируем вывод.

**Мы не имеем достаточных оснований, чтобы отвергнуть нулевую гипотезу. Это означает, что влияние тренинга не значимо на уровне значимости 0,05.**



**10.4.  
Доверительный интервал для разности  
средних. Парные выборки**

---





## Что мы имеем

Имеем две **случайные парные (зависимые)** выборки объема  $n$  из двух генеральных совокупностей.

Генеральные совокупности имеют нормальный закон распределения с параметрами  $\mu_1, \sigma_1$  и  $\mu_2, \sigma_2$  либо объемы обеих выборок  $\geq 30$ .

## Что мы хотим

Оценить среднее значение парных разностей для двух генеральных совокупностей. Для этого построить доверительный интервал для среднего в виде:

$$\bar{d} - E < \mu_d < \bar{d} + E$$

# Доверительный интервал



Среднее разности парных значений между двумя генеральными совокупностями с надежностью  $1-\alpha/2$  находится в доверительном интервале:

$$\bar{d} - E < \mu_d < \bar{d} + E$$

Точность оценки находится по формуле:

$$E = t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$df = n - 1$$

# Пример построения доверительного интервала



Выборка включает 15 студентов, следовательно  $df = 15 - 1 = 14$ .

Находим t-значение по таблице для доверительной вероятности 95% ( $\alpha/2 = 2,5$ ):

$$t_{\alpha/2} = 2,145$$

Точность оценки:

$$E = t_{\alpha/2} \frac{s_d}{\sqrt{n}} = 2,145 \cdot \frac{2,87}{\sqrt{15}} = 1,590$$

Доверительный интервал:

$$\begin{aligned} \bar{d} - E &< \mu_d < \bar{d} + E \\ 1,4 - 1,59 &< \mu_d < 1,4 + 1,59 \\ -0,19 &< \mu_d < 1,99 \end{aligned}$$



**10.5.**  
**Гипотеза о равенстве дисперсий**

---



## В ходе исследования...

---



Исследователю может понадобиться проверить предположение, о равенстве дисперсий двух изучаемых генеральных совокупностей.

В случае, когда эти генеральные совокупности имеют нормальное распределение, для этого существует F-критерий, называемый также критерием Фишера.

В отличие от Стьюдента, Фишер не работал на пивном заводе.





## Что мы имеем

1. Две простые случайные выборки, полученные из двух **нормально распределенных** генеральных совокупностей.
2. Выборки являются независимыми. Это значит, что между субъектами выборок нет связи.

## Что мы хотим

Проверить гипотезу о равенстве дисперсий генеральных совокупностей:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

# Обозначения



Дисперсии генеральных совокупностей:

$$\sigma_1^2 \quad \sigma_2^2$$

Дисперсии двух выборок:

$$s_1^2 \quad s_2^2$$

Объемы двух выборок:

$$n_1 \quad n_2$$

Подбираем, чтобы обязательно:

$$s_1^2 > s_2^2$$



Нулевая гипотеза:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Альтернативная гипотеза:

$$H_1 : \sigma_1^2 > \sigma_2^2$$

Других альтернативных гипотез в этом критерии не рассматривается.



Для проверки гипотезы используется статистика:

$$F = \frac{s_1^2}{s_2^2}$$

Если гипотеза верна, эта статистика имеет F-распределение (распределение Фишера) с количеством степеней свободы:

числителя  $n_1 - 1$

знаменателя  $n_2 - 1$

# Критическая область



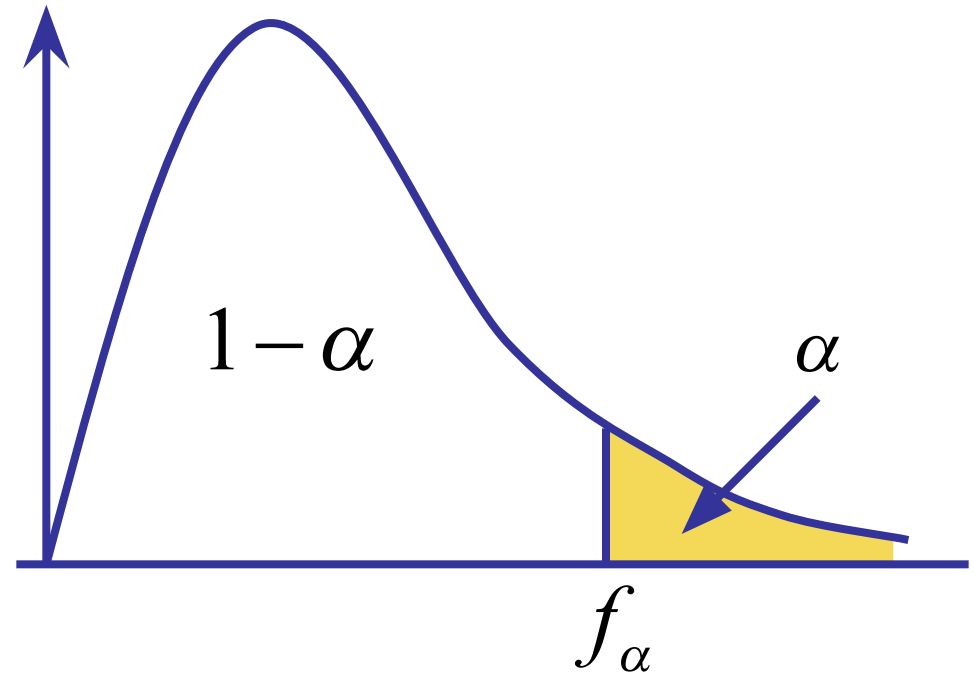
Альтернативная гипотеза:

$$H_1 : \sigma_1^2 > \sigma_2^2$$

Уравнение критической области:

$$P(F > f_\alpha) = \alpha$$

Критическое значение находим по таблице F-распределения



# Таблицы F-распределения



Приложение 0

Критические значения F-критерия ( $\alpha = 0,05$ )

Число степеней свободы знаменателя	Число степеней свободы числителя								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	10,13	9,55	9,26	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,46	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,96	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

Критические значения находим по таблице F-распределения.

Например, критическое значение для двух выборок объема 14 и 10 равно 2,71.

Таблицы «трехмерные». Учитесь пользоваться!

## Пример.



Исследователь-медик хочет проверить, есть ли различие между частотой биения сердца курящих и некурящих пациентов (кол-во ударов в минуту). Результаты двух случайно отобранных групп приведены ниже. Используя  $\alpha = 0,05$ , выясните, прав ли медик.

**Курящие**

**Не курящие**

$$n_1 = 26$$

$$n_2 = 18$$

$$s_1^2 = 36$$

$$s_2^2 = 10$$



**Шаг 1.** Основная и альтернативная гипотезы:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 > \sigma_2^2$$

**Шаг 2.** Задан уровень значимости  $\alpha=0,05$ .

**Шаг 3.** По таблице для количества степеней свободы числителя 25 и знаменателя 17 находим критическое значение  $f = 3,08$  и строим критическую область:

$$f > 3,08$$

**Шаг 4.** По выборке сосчитаем значение статистики.

$$F = \frac{s_1^2}{s_2^2} = \frac{36}{10} = 3,6$$





**Шаг 5.** Сравним полученное значение с критической областью.

$$3,6 > 3,08$$

Полученное значение статистики попало в критическую область.

**Шаг 6.** Формулируем вывод. **Различие дисперсий двух генеральных совокупностей значимо.**



**10.6.**  
**Гипотеза о равенстве долей**

---



# Вопрос

---



Из 100 случайно отобранных студентов социологического факультета 43 посещают спецкурсы.

Из 200 случайно отобранных студентов-экономистов 90 посещают спецкурсы.

Отличается ли доля студентов, посещающих спецкурсы, на социологическом и экономическом факультетах? Похоже, что существенно не отличается.

## Как это проверить?

Доля посещающих спецкурсы – доля признака. 43 – количество «успехов». 43/100 – доля успехов. Терминология такая же, как в схеме Бернулли.



## Что мы имеем

1. Две простые случайные выборки, полученные из двух **нормально распределенных** генеральных совокупностей.
2. Выборки являются независимыми. Это значит, что между субъектами выборок нет связи.
3. Для выборок выполнено  $np \geq 5$  и  $nq \geq 5$ . Это означает, что, по крайней мере, 5 элементов выборки имеют изучаемое значение признака, и, по крайней мере, 5 не имеют.

## Что мы хотим

Проверить гипотезу о равенстве долей признака в двух генеральных совокупностях:

$$H_0 : p_1 = p_2 \quad H_1 : p_1 \neq p_2$$



$n_1$   $n_2$  - объемы выборок

$m_1$   $m_2$  - количество «успехов» в каждой выборке

$\hat{p}_1 = \frac{m_1}{n_1}$  - доля «успехов» в первой выборке

$\hat{p}_2 = \frac{m_2}{n_2}$  - доля «успехов» во второй выборке

$\bar{p} = \frac{m_1 + m_2}{n_1 + n_2}$  - общая доля «успехов» в обеих выборках



В качестве статистики выбираем следующую случайную функцию:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}(1 - \bar{p})}{n_1} + \frac{\bar{p}(1 - \bar{p})}{n_2}}}$$

Для проверки гипотезы пользуемся таблицей нормального распределения.

## Пример.

---



Из 100 случайно отобранных студентов социологического факультета 43 посещают спецкурсы. Из 200 студентов-экономистов 90 человек посещают спецкурсы. На уровне значимости  $\alpha = 0,05$ , проверьте гипотезу о том, что нет различия между долей посещающих спецкурсы на двух этих факультетах.



Вычислим необходимые значения:

$$\hat{p}_1 = \frac{m_1}{n_1} = \frac{43}{100} = 0,43$$

$$\hat{p}_2 = \frac{m_2}{n_2} = \frac{90}{200} = 0,45$$

$$\bar{p} = \frac{m_1 + m_2}{n_1 + n_2} = \frac{43 + 90}{100 + 200} = \frac{133}{300} = 0,443$$





**Шаг 1.** Основная и альтернативная гипотезы:

$$H_0 : p_1 = p_2 \quad H_1 : p_1 \neq p_2$$

**Шаг 2.** Задан уровень значимости  $\alpha=0,05$ .

**Шаг 3.** По таблице нормального распределения находим критические значения  $z = -1,96$  и  $z = 1,96$  строим критическую область:

$$z < -1,96 \quad z > 1,96$$

**Шаг 4.** По выборке сосчитаем значение статистики.

$$z = \frac{(0,43 - 0,45) - 0}{\sqrt{\frac{0,443 \cdot 0,557}{100} + \frac{0,443 \cdot 0,557}{200}}} = -0,33$$



**Шаг 5.** Сравним полученное значение с критической областью.

$$-1,96 < -1,33$$

Полученное значение статистики не попало в критическую область.

**Шаг 6.** Формулируем вывод. **Нет оснований отвергнуть основную гипотезу. Доля посещающих спецкурсы не отличается.**



**10.7.**  
**Доверительный интервал для  
разности долей**

---



# Доверительный интервал для разности долей



Доля значений признака в генеральной совокупности с надежностью  $1-\alpha/2$  находится в доверительном интервале:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

где

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

# Пример

---



Предположим, по результатам исследования 40% из 200 мужчин и 56% из 100 женщин высказались против смертной казни. Найдите 95%-ый доверительный интервал для действительной разности долей.

**Решение.** На семинаре.



# Задание на 5 минут

---



От чего и каким образом зависит критическая область?





10.1. Преподаватель хочет понять, действительно ли студенты-вечерники получают более высокие баллы по сравнению с учащимися дневного отделения. Ниже приведены результаты экзамена. Может ли преподаватель на их основе заключить, что балл вечерников выше? Используйте  $\alpha = 0,02$ .

**Дневники**

**Вечерники**

$$n_1 = 26$$

$$n_2 = 18$$

$$\bar{x}_1 = 4,2$$

$$\bar{x}_2 = 4,4$$

$$s_1 = 0,3$$

$$s_2 = 0,25$$





10.2. Исследователь хочет сравнить скорость реакции таксистов и полицейских. Полученные им результаты представлены ниже. Может ли он при  $\alpha = 0,02$  сделать вывод о том, что таксисты обладают меньшей скоростью реакции, чем полицейские. Предполагается, что генеральные совокупности распределены нормально.

Таксисты	Полицейские
----------	-------------

$$n_1 = 10$$

$$n_2 = 9$$

$$\bar{x}_1 = 0,38$$

$$\bar{x}_2 = 0,37$$

$$\sigma_1 = 0,02$$

$$\sigma_2 = 0,01$$



10.3. Исследователь предполагает, что среди учеников средней школы девочки чаще, чем мальчики, прогуливают занятия. Выборочное исследование 16-ти девочек показало, что их не бывает в школе примерно 3,9 дня в году, а мальчиков (22 человека) 3,6 дня. Стандартные отклонения 0,6 и 0,8 дня соответственно. Проверьте предположение исследователя на уровне значимости  $\alpha = 0,01$ . Предполагается, что дисперсии равны.



10.4. Налоговый инспектор желает проверить, есть ли различие в тарифных ставках на частную собственность в двух больших городах. Величина налогов в обоих городах представлена ниже (млн \$). При  $\alpha = 0,05$  проверьте, действительно ли налоги в двух городах различны?

	Город А				Город Б			
113	22	14	8	82	11	5	15	
25	23	23	30	295	50	12	9	
44	11	19	7	12	68	81	2	
31	19	5	2	20	16	4	5	



10.5. Преподаватель считает, что студенты, специализация которых – математика, могут написать компьютерную программу быстрее, чем те, чья специализация – экономика. Двенадцать студентов-математиков, попавшие в выборку, потратили в среднем по 36 минут на то, чтобы написать и отладить определенную программу. Восемнадцать студентов-экономистов справились с тем же заданием в среднем за 39 минут каждый. Стандартное отклонение каждой группы равно 4 и 9 минут соответственно. При  $\alpha = 0,10$  проверьте предположение преподавателя, считая, что дисперсии не равны.



10.6. Местное отделение налоговой инспекции потратило примерно по 21 минуте на то, чтобы помочь каждому из 10-ти человек заполнить их налоговую декларацию. Стандартное отклонение равно 5,6 минуты. Независимая служба подготовки налоговых деклараций потратила на каждого из 14 человек по 27 минут. Стандартное отклонение равно 4,3 минуты. При  $\alpha = 0,02$  найдите, есть ли разница во времени, которое тратят две этих службы. Предполагается, что дисперсии равны.



10.7. Преподаватель утверждает, что когда преподавание курса идет с использованием лекций, то дисперсия успеваемости больше, чем когда курс идет без лекций. Случайным образом были выбраны две группы студентов. Дисперсия успеваемости первой группы (с лекциями) равна 103, а дисперсия второй группы (без лекций) равна 73. В каждой группе учатся 20 студентов. При  $\alpha = 0,05$  проверить предположение преподавателя.



10.8. Преподаватель физкультуры утверждает, что тяжелоатлеты, принимающие витамин Е, могут увеличить свою силу, то есть поднимать более тяжелый вес. Было отобрано восемь атлетов, и была измерена их сила с помощью выжимания штанги. После двух недель регулярных тренировок и приема витамина Е, их силу измерили еще раз. Проверьте эффективность такого режима, считая, что  $\alpha = 0,05$ . Значения, которые даны ниже, – это максимальный вес, который может поднять спортсмен (в фунтах). Предполагается, что переменные распределены нормально.

Тяжелоатлеты	1	2	3	4	5	6	7	8
До	210	230	182	205	262	253	219	216
После	219	236	179	204	270	250	222	216



10.9. Социологу интересно узнать, повлияет ли показ фильма об управлении стрессом, на установки двенадцати человек, участвующих в исследовании. Результаты в таблице. Здесь большим числовым значениям соответствует более позитивное отношение к управлению стрессом. При  $\alpha = 0,05$  проверьте утверждение, что просмотр данного фильма изменит установки испытуемых.

До	12	15	6	20	2	5	9	16	14	17	8	5
После	11	13	3	21	5	7	6	10	9	72	4	1





10.10. Офис-менеджер хочет узнать, можно ли увеличить скорость печатания десяти секретарей, заменив печатные машинки компьютерами. В таблице число слов в минуту. На уровне значимости  $\alpha = 0,10$  проверьте утверждение, что, используя компьютер, секретари могут печатать большее количество слов в минуту.

Секретарь	Печатная машинка	Компьютер
1	63	68
2	72	80
3	85	95
4	97	93
5	82	80
6	101	106
7	73	82
8	62	78
9	58	65
10	75	83



**10-11.** В результате исследования выяснилось, из 100 опрошенных мужчин 83% предпочитают лекциям обучение с помощью компьютера, а из 100 женщин 75%. При  $\alpha = 0,05$  проверьте утверждение о равенстве долей мужчин и женщин, предпочитающих лекциям компьютерное обучение.

**10-12.** Из 200 хирургов 15% считают, что государство должно контролировать здравоохранение. А из 200 терапевтов так думает 21%. Существует ли различие в долях на уровне значимости  $\alpha = 0,05$ ?

**10-13.** Из 80 американцев 55% хотели бы разбогатеть. Из 90 европейцев, хотели бы разбогатеть 45%. При  $\alpha = 0,01$  есть ли различие в долях?

**10-14.** Из 200 мужчин 130 сказали, что пользуются ремнями безопасности. Из 300 женщин отметили, что пользуются ремнями безопасности, 63 человека. При  $\alpha = 0,01$  проверьте утверждение, что мужчины более осторожны, нежели женщины.



**10-15.** В первой выборке из 100 человек 30% были в Диснейленде, а во второй (тоже 100 человек опрошенных) 24% посещали Диснейуолд. Отличаются ли доли людей, посетивших каждый из двух парков? Возьмите  $\alpha = 0,02$ .

**10-16.** Из 200 подростков 59 считают, что война неизбежна. А из 300 человек старше 60-ти, так думают 93. Отличается ли доля подростков, считающих войну неизбежной, от доли людей старше 60-ти лет? Возьмите  $\alpha = 0,01$ ?

**10-17.** В выборке из 59 старшеклассников у восьми оказался свой собственный автомобиль, а среди 75 первокурсников колледжа свои машины есть у 20 человек. Можно ли на уровне значимости  $\alpha = 0,05$  сделать вывод, что доля первокурсников с собственными автомобилями выше?



**10-18.** Найдите 95%-ый доверительный интервал для действительной разности долей по результатам исследования, в котором 40% из 200 мужчин и 56% из 100 женщин высказались против смертной казни.

**10-19.** Найдите 99%-ый доверительный интервал для разности генеральных долей по результатам исследования, в котором 80% из 150-ти республиканцев и 60% из 200 демократов одобрили закон о повышении зарплаты.