

## Практическое задание 1.5.2

В файле `water.txt` представлено 61 наблюдение.

Каждое наблюдение – город в Англии и Уэльсе.

Города поделены на северные и южные.

Для каждого города известны средняя годовая смертность на 100000 населения (по данным 1958–1964) и концентрация кальция в питьевой воде (в частях на миллион).

Чем выше концентрация кальция, тем жёстче вода.

1) Используя метод `.describe()` вычислите описательные статистики для северных и южных городов. Сравните средние значения смертности в северных и южных городах и значения концентрации кальция в питьевой воде.

mortality    hardness South			mortality    hardness North		
count	26.000000	26.000000	count	35.000000	35.000000
mean	1376.807692	69.769231	mean	1633.600000	30.400000
std	140.269175	40.360682	std	136.936911	26.134494
min	1096.000000	5.000000	min	1378.000000	6.000000
25%	1259.250000	40.250000	25%	1557.500000	12.500000
50%	1364.000000	75.500000	50%	1637.000000	17.000000
75%	1485.750000	99.750000	75%	1718.000000	44.000000
max	1627.000000	138.000000	max	1987.000000	94.000000

– Среднее значение смертности в северных городах выше, чем в южных.

– Среднее значение концентрации кальция в питьевой воде в северных городах выше, чем в южных.

Размах изменения значений смертности – достаточно большой, т.е. данные неоднородны:

в южных городах =  $1627 - 1096 = 531$

в северных городах =  $1987 - 1378 = 609$

Кроме того, среднеквадратическое отклонение концентрации кальция в питьевой воде тоже достаточно большое!

2) Постройте 95% доверительные интервалы для средней годовой смертности по всем южным и северным городам. Отличаются ли границы интервалов?

*Доверительный интервал – это интервал, который строится вокруг оценочного значения параметра и показывает, где находится истинное значение оцениваемого параметра с априори заданной вероятностью.*

- 95% доверительные интервалы для средней годовой смертности по южным городам (1376.80): (1320.15, 1433.46); длина интервала – 113.31
- по северным городам (1633.60): (1586.56, 1680.63); длина интервала – 94.07
- Границы доверительного интервала для средней годовой смертности в северных городах сдвинуты в сторону больших значений смертности.

3) Постройте 95% доверительные интервалы для средней концентрации кальция в питьевой воде для южных и северных городов. Отличаются ли границы интервалов?

- 95% доверительные интервалы для средней концентрации кальция в питьевой воде – достаточно большой, т.е. данные неоднородны:

для южных городов (69.76): (53.46, 86.07); длина интервала – 32.61

для северных городов (30.40): (21.42, 39.37); длина интервала – 17.95

Размах изменения значений концентрации кальция

в южных городах =  $138 - 5 = 133$

в северных городах =  $94 - 6 = 88$

4) Какие можно сделать выводы?

### Интерпретация доверительных интервалов

При интерпретации доверительного интервала нас интересует следующий вопрос:

#### Насколько широк доверительный интервал?

Широкий доверительный интервал указывает на то, что оценка неточна; узкий указывает на точную оценку.

В прикладных статистических задачах **длина доверительного интервала** играет важную роль: чем меньше его длина, тем точнее оценка. Если длина этого интервала велика, то ценность такой оценки незначительна.

Ширина доверительного интервала зависит от размера стандартной ошибки, которая, в свою очередь, зависит от объёма выборки и при рассмотрении числовой переменной от изменчивости данных дают более широкие доверительные интервалы, чем исследования многочисленного набора данных немногих переменных.

В общем случае длина доверительного интервала определяется, главным образом, размером выборки  $n$  и неопределенностью генеральной совокупности. При прочих равных условиях, если размер выборки  $n$  большой, то доверительный интервал будет меньше, поскольку формула для вычисления стандартной ошибки включает деление на  $\sqrt{n}$ . Уменьшение интервала свидетельствует о меньшей неопределенности из-за наличия большего количества информации.