

Практическое занятие 4: Генеральная и выборочная средняя. Мода. Медиана

Цель занятия:

- вычисление генеральной и выборочной средней;
- вычисление медианы дискретного и интервального вариационного ряда;
- вычисление моды дискретного и интервального вариационного ряда.

Задание 1

Варианты задания 1

Задание 2

Варианты задания 2

Генеральная и выборочная средняя

Пусть исследуется некоторая *генеральная совокупность* объёма N , а именно её числовая характеристика X , не важно, дискретная или непрерывная.

Генеральной средней называется *среднее арифметическое* всех значений этой совокупности:

$$\bar{x}_G = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

Если среди чисел x_i есть одинаковые (*что характерно для дискретного ряда*), то формулу можно записать в более компактном виде:

$$\bar{x}_G = \frac{x_1 N_1 + x_2 N_2 + x_3 N_3 + \dots + x_k N_k}{N} = \frac{\sum_{i=1}^k x_i N_i}{N},$$

где *варианта* x_1 повторяется N_1 раз;

варианта x_2 – N_2 раз;

варианта x_3 – N_3 раз;

...

варианта x_k – N_k раз.

Обработка всей генеральной совокупности часто затруднена либо невозможна, и поэтому из неё организуют представительную выборку *объёма* n , и на основании исследования этой выборки делают вывод обо всей совокупности.

Выборочной средней называется *среднее арифметическое* всех значений выборки:

$$\bar{x}_e = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

и при наличии одинаковых вариантов формула запишется компактнее:

$$\bar{x}_e = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \dots + x_k n_k}{n} = \frac{\sum_{i=1}^k x_i n_i}{n} \quad \text{— как сумма произведений вариантов}$$

x_i на соответствующие частоты n_i .

Выборочная средняя \bar{x}_e позволяет достаточно точно оценить истинное значение \bar{x}_r , чего вполне достаточно для многих исследований. При этом, чем больше выборка, тем точнее будет эта оценка.

Пример

По результатам выборочного исследования у $n = 25$ рабочих цеха были установлены их квалификационные разряды: 4, 5, 6, 4, 4, 2, 3, 5, 4, 4, 5, 2, 3, 3, 4, 5, 5, 2, 3, 6, 5, 4, 6, 4, 3.

Требуется: вычислить выборочную среднюю, моду и медиану.

Если даны *первичные данные* (исходные необработанные значения), то их можно тупо просуммировать и разделить результат на объём выборки:

$$\bar{x}_e = \frac{4 + 5 + 6 + \dots + 4 + 3}{25} = 101 / 25 \approx 4,04 \approx 4 \quad \text{— среднестатистический}$$

квалификационный разряд рабочих цеха.

Но во многих задачах требуется составить вариационный ряд.

x_i	n_i
2	3
3	5
4	8
5	6
6	3
Σ	25

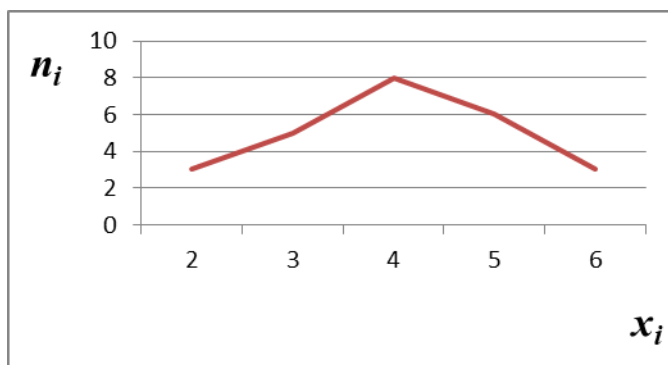
Тогда

$$\bar{x}_e = \frac{\sum_{i=1}^k x_i n_i}{n} = \frac{2 \cdot 3 + 3 \cdot 5 + 4 \cdot 8 + 5 \cdot 6 + 6 \cdot 3}{25} = 101 / 25 = 4,04$$

Мода

Мода M_0 дискретного вариационного ряда — это *варианта* с максимальной частотой. В данном случае $M_0 = x_3 = 4$.

Моду легко отыскать по таблице, и ещё легче на **полигоне частот** — это абсцисса самой высокой точки:



Поиск моды в среде MS Excel (без предварительного упорядочения):

4	5	6	4	4	2	3	5	4	4	5	2	3	3	4	5	5	2	3	6	5	4	6	4	3	4

=МОДА.ОДН(A1:Y1)

Иногда таковых значений несколько (с одинаковой максимальной частотой), и тогда модой считают каждое из них.

Медиана

Медиана m_e вариационного ряда – это значение, которая делит его на две равные части (по количеству вариантов). Не важно, *дискретного или интервального, генеральной совокупности или выборочной*.

Медиану можно отыскать несколькими способами.

Если даны первичные данные, то сортируем их по возрастанию либо убыванию и находим середину ранжированного ряда (для примера 8): $m_e = x_{13} = 4$. Почему именно 13-е число? Потому что перед ним находится 12 чисел и после него тоже 12 чисел, таким образом, значение $x_{13} = 4$ разделило ряд на две равные части, а значит, является медианой.

Этот номер можно найти аналитически:

– если совокупность содержит **нечётное** количество чисел, то делим её объём пополам: $n/2 = 25/2 = 12,5$ и округляем полученное значение в большую сторону: 13 – получая тем самым срединный номер.

– если совокупность содержит **чётное** количество чисел, например, 20, то делаем то же самое: $n/2 = 20/2 = 10$, и медианное значение здесь рассчитывается как среднее арифметическое 10-го и следующего числа:

$$m_e = (x_{10} + x_{11}) / 2.$$

Изложенная инструкция работает для упорядоченного (по возрастанию либо убыванию) ряда.

Поиск медианы в среде MS Excel (без предварительного упорядочения):

4	5	6	4	4	2	3	5	4	4	5	2	3	3	4	5	5	2	3	6	5	4	6	4	3	4

=МЕДИАНА(A7:Y7)

Полученные значения близки друг к другу, и это говорит о симметрии вариационного ряда относительно центра, что хорошо видно по полигону частот. И с высокой вероятностью можно утверждать, что примерно так же распределена и вся генеральная совокупность (все рабочие цеха).

И тут возникает следующий закономерный вопрос: а зачем вообще нужна мода с медианой? – ведь есть *средняя*.

А дело в том, что в ряде случаев среднее значение неудовлетворительно характеризует центральную тенденцию статистической совокупности:

4	5	6	4	4	2	3	5	4	4	5	2	3	3	4	5	5	2	3	6	5	4	6	4	3	4	4,04
4	5	6	4	4	2	3	5	4	4	5	2	3	3	4	5	5	2	3	6	5	4	6	4	3	4	4,04
100	5	6	100	4	2	3	5	100	100	5	100	3	3	100	5	5	2	3	6	5	100	6	4	3	100	31,00
100	5	6	100	4	2	3	5	100	100	5	100	3	3	100	5	5	2	3	6	5	100	6	4	3	5	31,00

Пример

Известны результаты продаж пиджаков в универмаге города:

X_i	1	2	3	4	5
f_i	225	32	82	145	16

X_i	1	2	3	4	5		
n_i	225	32	82	145	16	$\Sigma =$	500
$X_i * n_i$	225	64	246	580	80	$X_{cp} =$	2,4
						$Mo =$	1
						$me =$	82

Задание 1

По результатам выборочного исследования цен на ботинки в магазинах города получены следующие данные (ден. ед.). Найти среднюю, моду и медиану.

Методика выполнения

	A	B	C
1	7,5	7,6	8,7
2	6,1	10,6	9,8
3	7	6	8,3
4	6	8,2	8,5
5	7,4	7,1	9,5
6	6,8	9,6	6,3
7	6,3	8,5	5,8
8	7,5	9,2	7,2
9	7	8	7,5
10	7,5	8	6,5

$$\bar{x}_e = 7,666666667$$

=СРЗНАЧ(A1:C10)

$$m_e = 7,5$$

=МЕДИАНА(A1:C10)

Вариант ручного счёта.

Провести сортировку по возрастанию и разделить объём выборки пополам: $n/2=15$, и поскольку она состоит из чётного количества вариантов, то медиана равна среднему арифметическому 15-й и 16-й варианты упорядоченного (!) вариационного ряда:

$$m_e = (x_{15} + x_{16})/2 = (7,5 + 7,5)/2 = 7,5 \text{ ден. ед.}$$

Ситуация вторая. Когда дан готовый интервальный ряд (типичная учебная задача).

Продолжаем анализировать тот же пример с ботинками, где по исходным данным был составлен ИВР. Для вычисления *средней* потребуются середины x_i интервалов:

Интервалы		x_i	n_i
5,7	6,7	6,2	7
6,7	7,7	7,2	11
7,7	8,7	8,2	6
8,7	9,7	9,2	4
9,7	10,7	10,2	2
		$\Sigma=$	30

Что касается моды, то её оценка по исходным данным, становится непригодна. Хотя мы и видим среди чисел одинаковые, но среди них запросто может найтись несколько вариантов с одинаковой максимальной частотой, например, частотой 2. Кроме того, цены могут быть округлёнными. Поэтому модальное значение рассчитывается по сформированному интервальному ряду.

Чтобы найти моду, нужно найти **модальный интервал** (с максимальной частотой) – в данной задаче это интервал (6,7 – 7,7) с частотой 11, и воспользоваться следующей формулой:

$$Mo = x_0 + \frac{n_M - n_{M-1}}{(n_M - n_{M-1}) + (n_M - n_{M+1})} h,$$

x_0 – нижняя граница модального интервала;

$h = 7,7 - 6,7 = 1$ – длина модального интервала;

$n_M = 11$ – частота модального интервала;

$n_{M-1} = 7$ – частота предыдущего интервала;

$n_{M+1} = 6$ – частота следующего интервала.

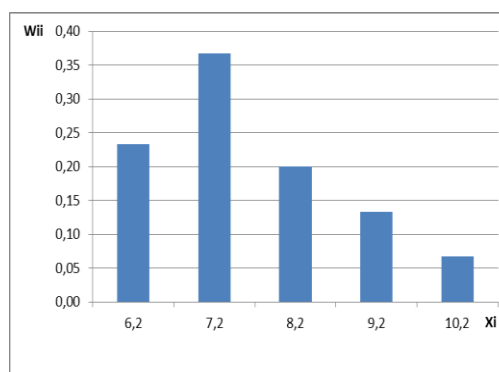
Таким образом:

$$Mo = 6,7 + \frac{11 - 7}{(11 - 7) + (11 - 6)} 1 = 6,7 + 4 / (4 + 5) \approx 7,14 \text{ ден. ед.}$$

– т.е., «модная» цена на ботинки заметно отличается от средней арифметической.

Другое решение – откуда хорошо видно, что мода смещена относительно центра модального интервала в сторону левого интервала с большей частотой.

Интервалы		x_i	n_i	w_i
5,7	6,7	6,2	7	0,23
6,7	7,7	7,2	11	0,37
7,7	8,7	8,2	6	0,20
8,7	9,7	9,2	4	0,13
9,7	10,7	10,2	2	0,07
		$\Sigma=$	30	1



Справка по редким случаям:

- если модальный интервал крайний, то $n_{M-1} = 0$ либо $n_{M+1} = 0$;
- если обнаружатся 2 модальных интервала, которые находятся рядом, например, (6,7; 7,7) и (7,7; 8,7), то рассматриваем модальный интервал,

при этом близлежащие интервалы (слева и справа) по возможности тоже укрупняем в 2 раза;

- если между модальными интервалами есть расстояние, то применяем формулу к каждому интервалу, получая тем самым 2 или большее количество мод.

И медиана. Если дан готовый интервальный ряд, то сначала нужно найти **медианный интервал** – это интервал, содержащий варианту (либо 2 варианты), которая делит вариационный ряд на две равные части.

Выше было показано, как определить медиану, ориентируясь на *относительные накопленные частоты* w_n , здесь же сподручнее рассчитать «обычные» накопленные частоты n_n .

	B		C	D	E
39	Интервалы		x_i	n_i	n_n
40	5,7	6,7	6,2	7	7
41	6,7	7,7	7,2	11	18
42	7,7	8,7	8,2	6	24
43	8,7	9,7	9,2	4	28
44	9,7	10,7	10,2	2	30
			Σ=	30	

=D40

=E40+D41

Смысл чисел в правом столбце – это количество вариантов, которые успели «накопиться» на всех «пройденных» интервалах, включая текущий.

Поскольку имеется чётное количество вариантов (30 штук), то медианным будет тот интервал, который содержит $30/2 = 15$ -ю и 16-ю варианту. И ориентируясь по накопленным частотам, легко прийти к выводу, что эти варианты содержатся в интервале (6,7; 7,7).

Формула медианы: $m_e = x_0 + \frac{0,5n - n_{m-1}^*}{n_m} h$,

где:

$n = 30$ – объём статистической совокупности;

x_0 – нижняя граница медианного интервала;

$h = 7,7 - 6,7 = 1$ – длина медианного интервала;

$n_m = 11$ – частота медианного интервала;

$n_{m-1}^* = 7$ – накопленная частота **предыдущего** интервала.

Таким образом:

$m_e = 6,7 + (0,5 \cdot 30 - 7) \cdot 1 / 11 \approx 7,43$ ден. ед. – заметим, что медианное значение, наоборот, оказалось смещено правее, т.к. по правую руку находится значительное количество вариантов:

И справочно особые случаи:

- если медианным является крайний левый интервал, то $n \cdot m - 1 = 0$;
- если вариационный ряд содержит чётное количество вариантов и две средние варианты попали в разные интервалы, то объединяем эти интервалы, и по возможности удваиваем предыдущий интервал.

Ответ: $\bar{x}_e \approx 7,67$, $M_o \approx 7,14$, $m_e \approx 7,43$ ден. ед.

Здесь центральные показатели оказались заметно отличны друг от друга, и это говорит об **асимметрии** распределения, которая хорошо видна по гистограмме.

Варианты задания 1

Вариант 1			Вариант 2			Вариант 3			Вариант 4		
6,9	7,6	8,7	10,7	7,6	8,7	9,0	7,6	8,7	9,5	7,6	8,7
6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8
7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3
6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5
7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5
6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3
6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8
7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2
7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5
7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5
Вариант 5			Вариант 6			Вариант 7			Вариант 8		
10,0	7,6	8,7	8,8	7,6	8,7	6,9	7,6	8,7	6,6	7,6	8,7
6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8
7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3
6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5
7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5
6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3
6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8
7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2
7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5
7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5
Вариант 9			Вариант 10			Вариант 11			Вариант 12		
5,7	7,6	8,7	8,4	7,6	8,7	10,8	7,6	8,7	6,6	7,6	8,7
6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8
7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3
6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5
7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5
6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3
6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8
7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2
7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5
7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5

Вариант 13			Вариант 14			Вариант 15			Вариант 16		
8,8	7,6	8,7	7,8	7,6	8,7	8,2	7,6	8,7	6,7	7,6	8,7
6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8
7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3
6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5
7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5
6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3
6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8
7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2
7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5
7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5
Вариант 17			Вариант 18			Вариант 19			Вариант 20		
7,6	7,6	8,7	8,3	7,6	8,7	5,0	7,6	8,7	10,8	7,6	8,7
6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8
7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3
6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5
7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5
6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3
6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8
7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2
7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5
7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5
Вариант 21			Вариант 22			Вариант 23			Вариант 24		
6,2	7,6	8,7	8,5	7,6	8,7	5,8	7,6	8,7	10,9	7,6	8,7
6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8	6,1	10,6	9,8
7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3	7,0	6,0	8,3
6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5	6,0	8,2	8,5
7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5	7,4	7,1	9,5
6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3	6,8	9,6	6,3
6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8	6,3	8,5	5,8
7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2	7,5	9,2	7,2
7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5	7,0	8,0	7,5
7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5	7,5	8,0	6,5

Задание 2

Для изучения затрат времени на изготовление одной детали рабочими завода проведена выборка, в результате которой получено следующее статистическое распределение:

Затраты на одну деталь, мин.	Число деталей, шт.
до 20	10
от 20 до 24	20
от 24 до 28	50
от 28 до 32	15
свыше 32	5
Итого	100

Найти среднюю, моду и медиану.

Методика выполнения

Решение. Поскольку длина внутренних интервалов равна $h = 4$, то длины крайних интервалов полагаем такими. Заполним расчётную таблицу:

	D E		F	G	H	I
1	Интервалы		n_i	x_i	$n_i x_i$	n_n
2	16	20	10	18	180	10
3	20	24	20	22	440	30
4	24	28	50	26	1300	80
5	28	32	15	30	450	95
6	32	36	5	34	170	100
7	$\Sigma =$		100		2540	

=F2

=I2+F3

Выборочная средняя

$$\bar{x}_e = \frac{\sum_{i=1}^k x_i n_i}{n} = 2540/100 = 25,4 \text{ мин.}$$

Моду вычислим по формуле

$$M_o = x_0 + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} h,$$

где

$x_0=24$ – нижняя граница модального интервала;

$h=4$ – длина модального интервала;

$n_m=50$ – частота модального интервала;

$n_{m-1}=20$ – частота предшествующего интервала;

$n_{m+1}=15$ – частота следующего интервала.

Таким образом:

$$M_o = 24 + \frac{50 - 20}{(50 - 20) + (50 - 15)} \cdot 4 = 24 + \frac{30}{65} \cdot 4 \approx 24 + 1,85 = 25,85 \text{ мин.}$$

Анализируя накопленные частоты, приходим к выводу, что медианным является интервал (24; 28) (именно он содержит 50-ю и 51-ю варианты, которые делят ряд пополам).

Медиану вычислим по формуле

$$m_e = x_0 + \frac{0,5n - n_{m-1}^*}{n_m} h,$$

где:

$x_0=24$ – нижняя граница медианного интервала;

$h=4$ – длина этого интервала;

$n=100$ – объём статистической совокупности;

$n_m=50$ – частота медианного интервала;

$n_{m-1}=30$ – накопленная частота предыдущего интервала.

Таким образом:

$$m_e = 24 + (0,5 \cdot 100 - 30) \cdot 4 / 50 = 25,6 \text{ мин.}$$

Ответ: среднее время изготовления детали характеризуется следующими центральными характеристиками: $x_e=25,4$, $M_o \approx 25,85$, $m_e \approx 25,6$.

Варианты задания 2

№ вар.	Затраты на 1 деталь мин.				
	до 19	от 19 до 23	от 23 до 27	от 27 до 31	свыше 31
	Число деталей, шт.				
1	13	18	6	50	22
2	43	16	15	24	8
3	6	5	11	49	45
4	31	33	9	34	36
5	34	31	21	43	6
6	39	29	24	48	44
7	19	13	33	10	33
8	8	13	48	45	33
9	46	13	26	30	30
10	15	47	38	48	29
11	46	5	45	44	48
12	23	20	20	8	25
13	46	38	23	5	19
14	25	42	31	21	23
15	21	25	8	18	38
16	9	48	11	49	37
17	30	36	20	31	36
18	15	25	44	50	20
19	16	21	11	45	38
20	47	42	29	45	39
21	44	47	42	21	39
22	49	48	23	30	13
23	49	20	33	42	22
24	20	13	29	33	42
25	8	21	41	15	9