

# 金风数据类型简单分析

# 分析总括

- 分析对象：总计简单分析**172**个属性的数值，每个属性数值约为**11**万条，但中间有缺失值。
- 共分为5类
  - 傅里叶函数
  - 线性函数
  - 分段混合函数
  - 分段常数函数
  - 无具体的连续变化规律
- 工具：MySQL，Excel、Matlab

# 傅里叶函数

- 合计 6个

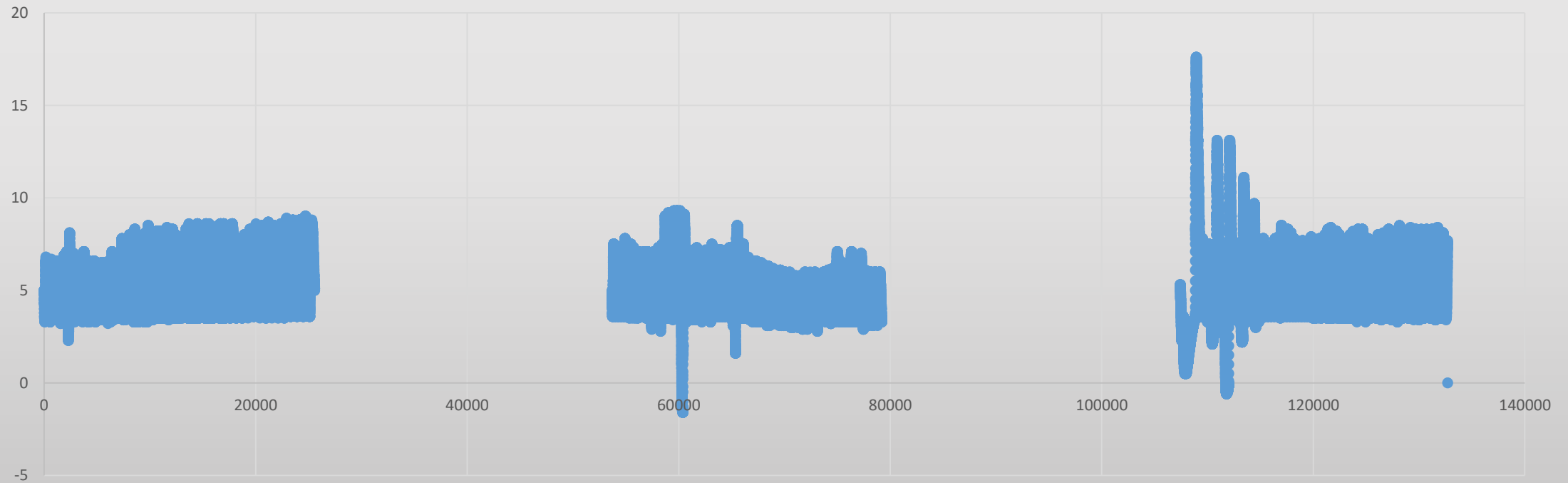
Column	函数类型	函数形式
column_497	傅里叶函数	$Y = 5.07 + 0.07563 \cdot \cos(x \cdot 0.02507) - 1.527 \cdot \sin(x \cdot 0.02507)$
column_498	傅里叶函数	$Y = 5.123 + 1.059 \cdot \cos(x \cdot 0.02503) - 1.103 \cdot \sin(x \cdot 0.02503)$
column_499	傅里叶函数	$Y = 4.901 - 1.211 \cdot \cos(x \cdot 0.0244) - 0.7806 \cdot \sin(x \cdot 0.0244)$
column_536	傅里叶函数	$Y = 5.07 - 0.07563 \cdot \cos(x \cdot 0.02507) - 1.527 \cdot \sin(x \cdot 0.02507)$
column_537	傅里叶函数	$Y = 5.123 - 1.059 \cdot \cos(x \cdot 0.02503) - 1.103 \cdot \sin(x \cdot 0.02503)$
column_538	傅里叶函数	$Y = 4.901 - 1.211 \cdot \cos(x \cdot 0.02444) - 0.7806 \cdot \sin(x \cdot 0.02444)$

## 如何得出-以column\_498为例

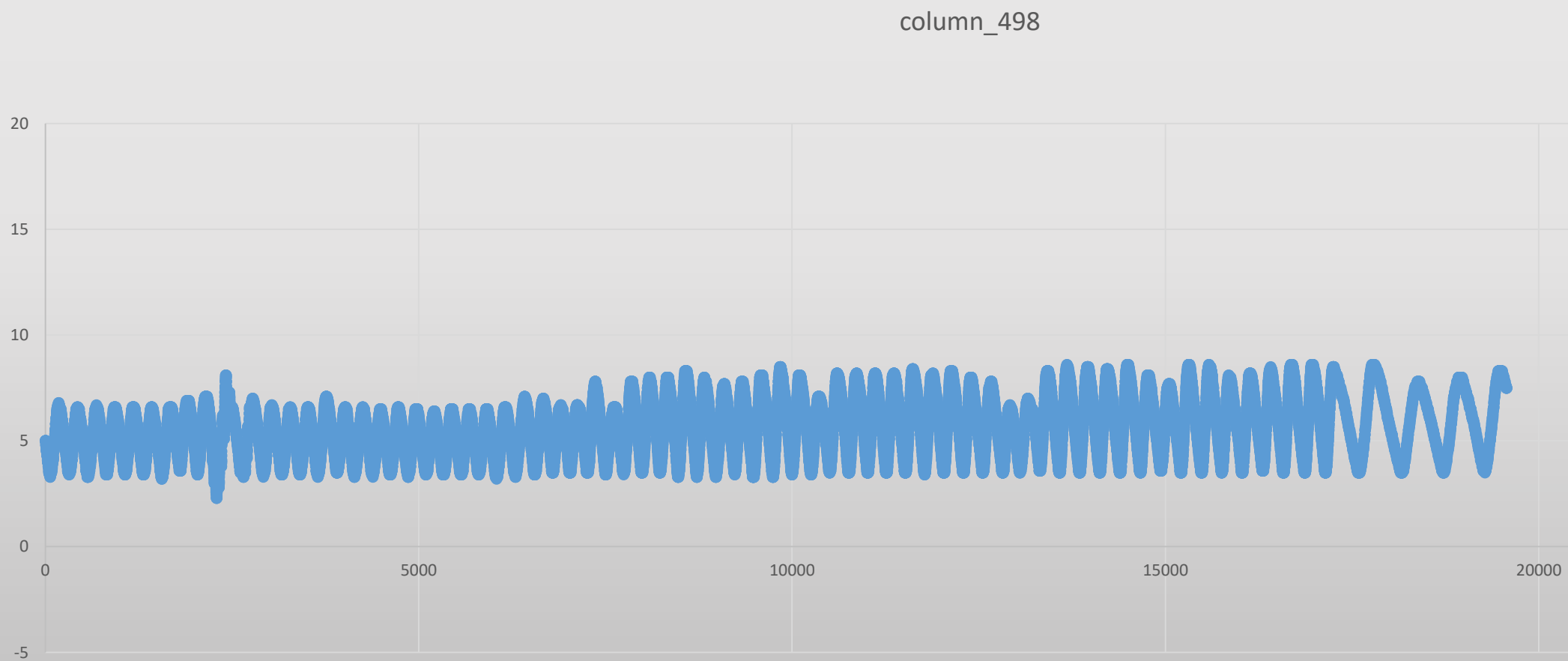
1. 从数据库中Navicat读取该属性所有数值
2. 存储到Excel中，先看所有数值的大致图像，判断是否可能有连续的变化规律
3. 在Excel中，继续取小样数据19000条，判断图像大致变化
4. Matlab连接MySQL读入1000条数据，对图像和数据进行详细的函数拟合和拟合评价分析

# Excel中的第一个初步判断---可能能用函数关系表示

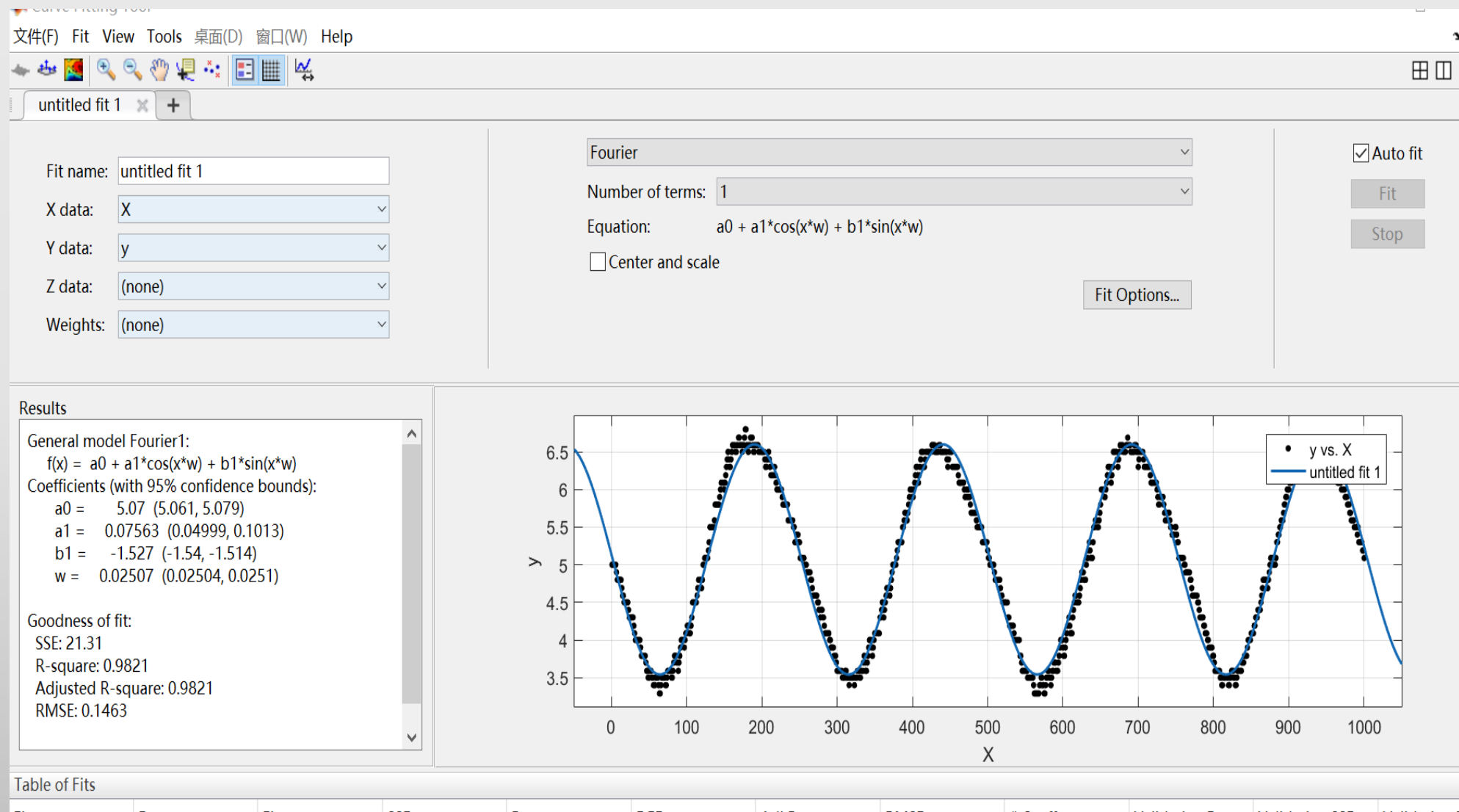
column\_498



# 很有可能



# Matlab中进一步图像拟合-傅里叶函数拟合最好



函数关系:  $Y = 5.123 + 1.059 \cdot \cos(x \cdot 0.02503) - 1.103 \cdot \sin(x \cdot 0.02503)$

# 正弦曲线逼近效果就差些了

Fit name:

X data:

Y data:

Z data:

Weights:

Sum of Sine

Number of terms:

Equation:  $a1*\sin(b1*x+c1)$

☐ Center and scale

Fit Options...

☒ Auto fit

Fit

Stop

## Results

General model Sin1:

$$f(x) = a1*\sin(b1*x+c1)$$

Coefficients (with 95% confidence bounds):

$a1 = 12.2$  (-474.5, 498.9)

$b1 = 6.045e-05$  (-0.002867, 0.002988)

$c1 = 0.4022$  (-16.56, 17.37)

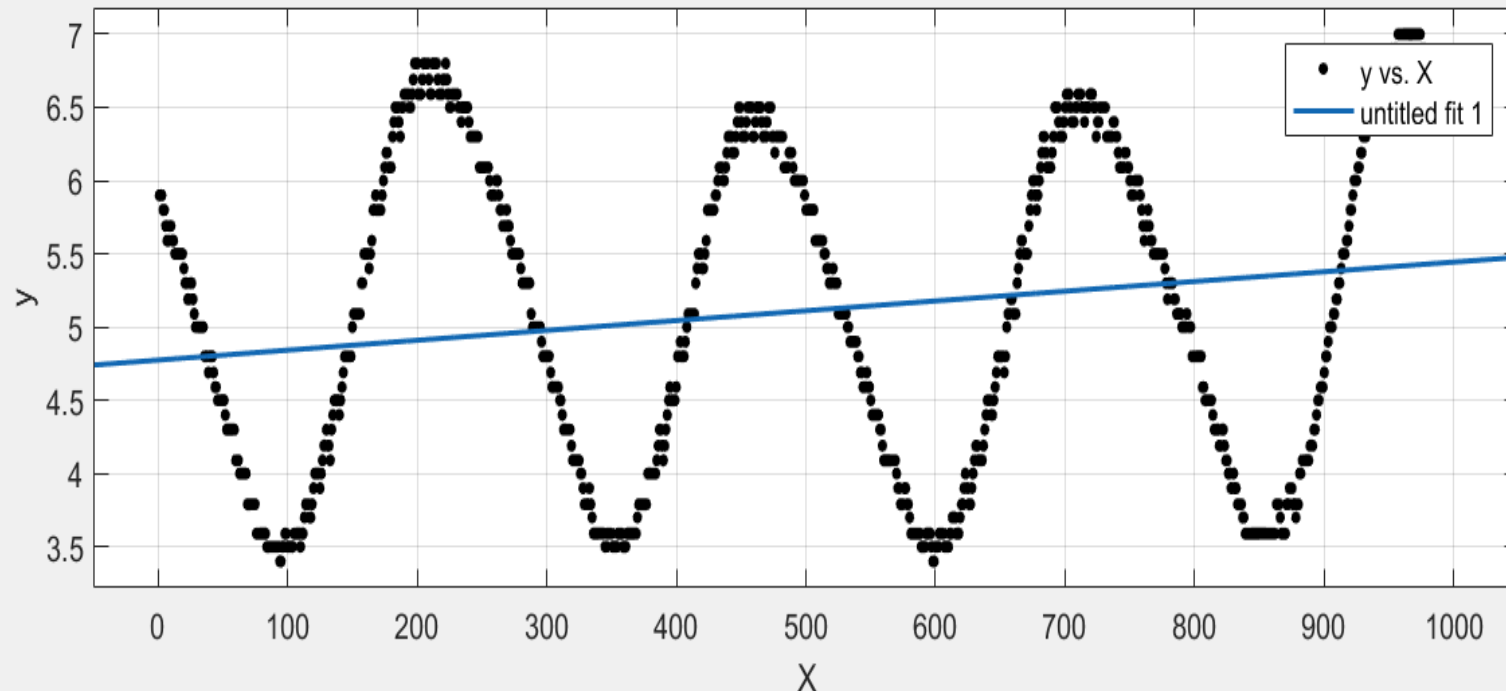
Goodness of fit:

SSE: 1177

R-square: 0.02705

Adjusted R-square: 0.0251

RMSE: 1.086





# Matlab中编程代码

- conn= database('ruc\_test', 'zw\_benchmark', '123456', 'com.mysql.jdbc.Driver', 'jdbc:mysql://114.115.137.143:3306/ruc\_test')
- cursy= exec(conn,'select column\_498 from ruc\_test\_3 order by column\_5 limit 1000')
- cursy = fetch(cursy)
- cur =cursy. Data
- y = cell2mat(cur)

•	X=[ 1	2	3	4	5	6	7	8	9	10	11	12
		13	14	15	16	17	18	19	20	21	22	23
		24	25	26	27	28	29	30	31	32	33	34
		35	36	37	38	39	40	41	42	43	44	45
		46	47	48	49	50	51	52	53	54	55	56
		57	58	59	60	61	62	63	64	65	66	67
		68	69	70	71	72	73	74	75	76	77	78
		79	80	81	82	83	84	85	86	87	88	89
		90	91	434	435	436	437	438	439	440	441	442
		443	444	445	446	447	448	449	450	451	452	453
		454	455	456	457	458	459	460	461	462	463	464
		465	466	467	468	469						
•	.....											
•		470	471	472	473	474	475	476	477	478	479	480
		481	482	483	484	485	486	487	488	489	490	491
		492	493	494	495	496	497	498	499	500	501	502
		503	504	505	506	507	508	509	510	511	512	513
		514	515	516	517	518	519	520	521	522	523	524
		525	526	527	528	529	530	531	532	533	534	535
		536	537	538	539	540	541	538	939	940	941	942
		943	944	945	946	947	948	949	950	951	952	953
		954	955	956	957	958	959	960	961	962	963	964
		965	966	967	968	969	970	971	972	973	974	975
		976	977	978	979	980	981	982	983	984	985	986
		987	988	989	990	991	992	993	994	995	996	997
		998	999	1000]								

# 线性函数

- 合计9个

Column	函数类型	函数形式
column_5	线性函数	$Y = 2016-01-13T00:00:02 + 0.7s * x$
column_337	线性函数	$Y = 17366.5 + (x/50) * 0.1$
column_421	线性函数	除前10个点外，每隔50约较为规律的增长 $0.1, Y = 33909.6 + (x/50) * 0.1$
column_423	线性函数	除前6个点外，每隔50约较为规律的增长 $0.1, Y = 35051 + (x/50) * 0.1$
column_425	线性函数	除前33个点外，每隔50约较为规律的增长 $0.1, Y = 35051 + (x/50) * 0.1$
column_507	线性函数	每隔6000个数据点增加1
column_543	线性函数	$Y = 0.001941 * x + 33909.7$
column_545	线性函数	$Y = 0.001941 * x + 33501$
column_547	线性函数	$Y = 0.001941 * x + 33847.4$

# column\_337为例

Fit name:

X data:

Y data:

Z data:

Weights:

## Linear Fitting

= f(  )

=

E

## Results

Linear model:

$f(x) = c \cdot x/50 + a$

Coefficients (with 95% confidence bounds):

c = 0.09687 (0.09655, 0.09718)

a = 1.737e+04 (1.737e+04, 1.737e+04)

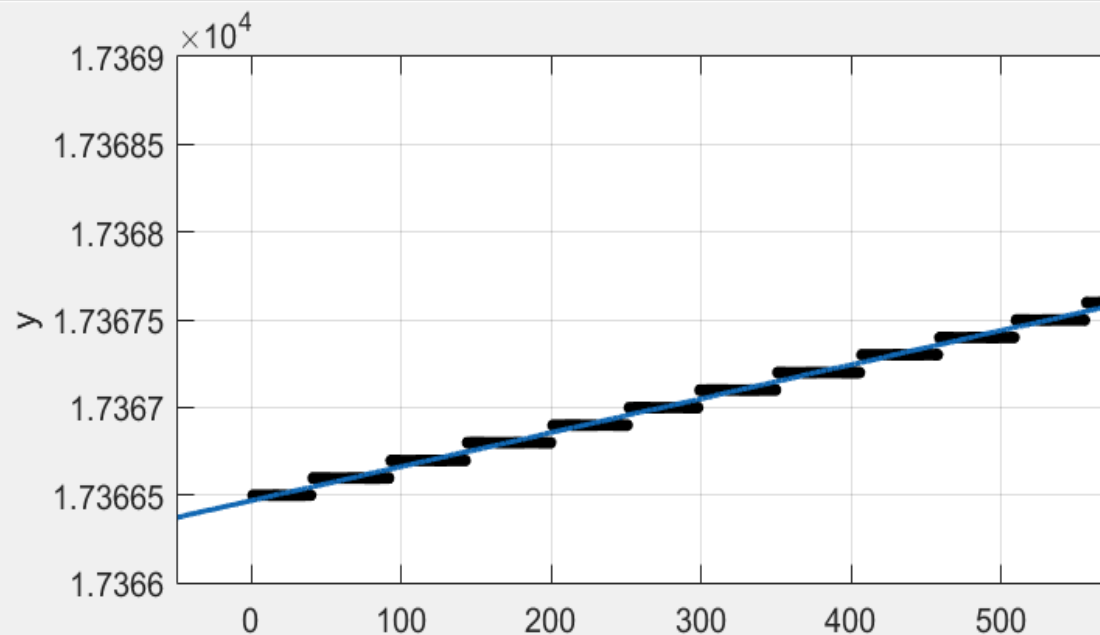
Goodness of fit:

SSE: 0.8456

R-square: 0.9973

Adjusted R-square: 0.9973

RMSE: 0.02911



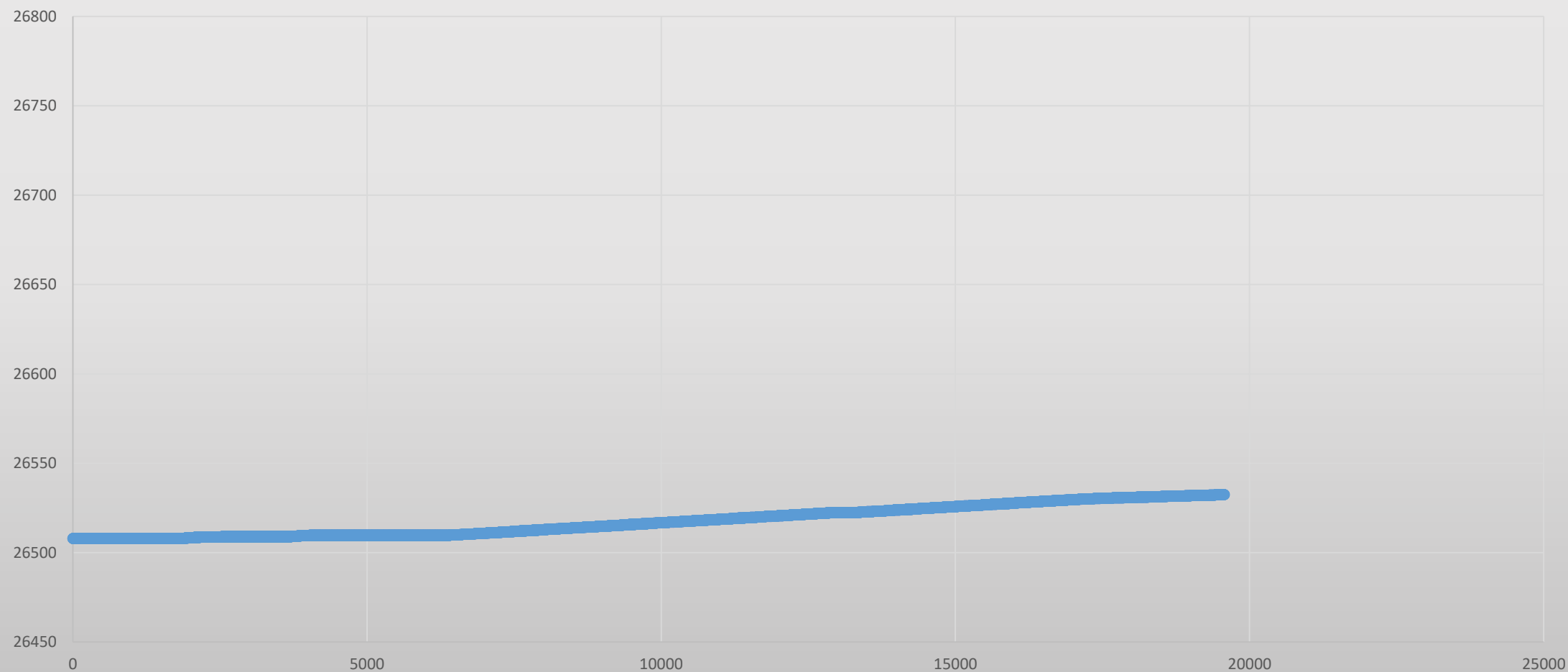
# 分段混合函数

- 合计3个

Column	函数类型	函数形式
column_356	分段混合函数	常数与线性函数结合。前1740个点为常数26507.8，后面成线性函数,6458位每隔不定个点稍微增加0.1。当 $x > 6458$ , $Y = 26510.4 + (x - 6458) / 120 * 0.1$
column_420	分段混合函数	常数与线性函数结合。前25578个点为常数1141.4，后期无连续规律的增长，中间仍有间隔的常数值，当 $x > 25578$ , $Y = 1141.4 + (x - 25578) / 5 * 0.1$
column_474	分段混合函数	常数与线性函数结合。前8116个点值为26513左右,后期每隔50个点，约增长0.1,当 $x > 8116$ , $Y = 26513 + (x - 8116) / 50 * 0.1$

# 以column\_474为例

column\_474



column\_474



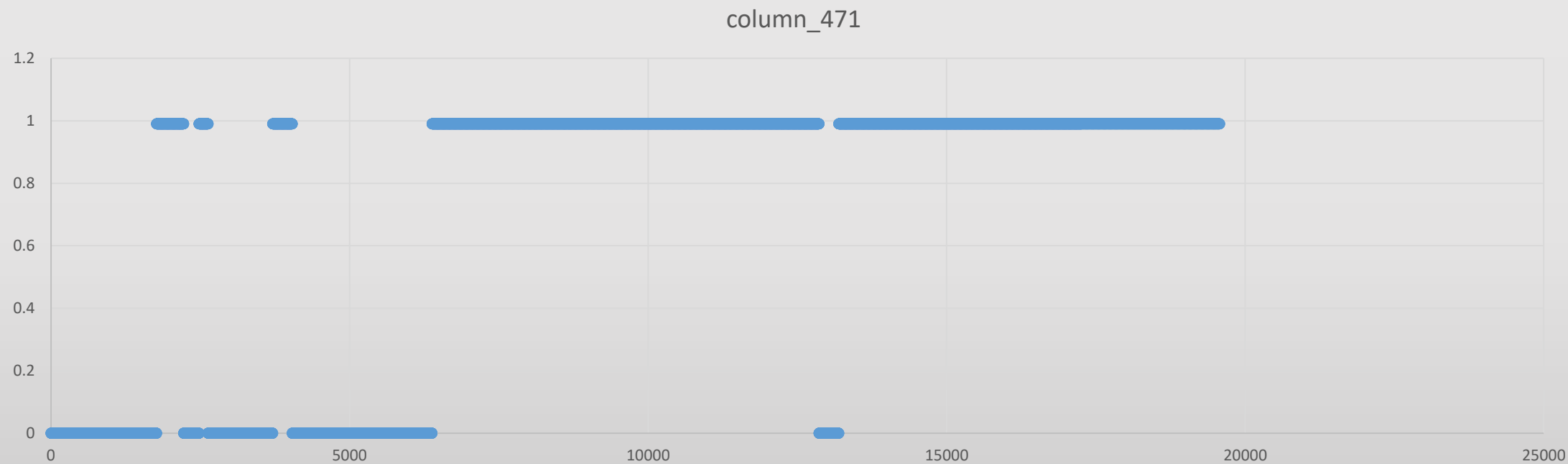
$$x > 8116, Y = 26513 + (x - 8116) / 50 * 0.1$$

# 分段常数函数

- 合计9个

Column	函数类型	函数形式
column_299	分段常数函数	分段区间可看成常数值内前1762条值均为2，1764为3,1765为4,1766-2034为5,2231-3657为2，3713-4027为5,4050-6313为2,6382-12854为5，12872-13188为2,13268到后面19566为5
column_315	分段常数函数	只有两个可能值0和2,前17124个点是常数值0,17125-31001有每隔连个0后在两个2，呈现方波状,31002-53455是常数值2,53456-79079是常数值0
column_338	分段常数函数	前1761个点是常数值6,中间1762-4033有间隔出现0和1，4052-6368为6,6369-12847为0
column_422	分段常数函数	前25578个点为常数 1473.54，后面数据不全，到约55241有新的常数值1483.42
column_429	分段常数函数	前24510个点为常数 0，后面数据不全，到约55241有新的常数值25.93，再到后面75772-77261新的常数值36.74
column_471	分段常数函数	前1到1767值为0，1768到2591值为0.99,2592到6445值为0,6446到13193值为0.99
column_508	分段常数函数	1-2182，值为781.54,2183-11453，值为781.64,11455-19534，值为787.74
column_509	分段常数函数	1-2182，值为781.54,2183-11453，值为781.64,11455-19534，值为787.74
column_510	分段常数函数	1-2182，值为781.54,2183-11453，值为781.64,11455-19534，值为787.74

# 以column\_471为例





# 无具体的连续变化规律

- 合计：145个

Column	函数类型	函数形式
column_290	无具体的连续变化规律	-5到1534之间
column_377	无具体的连续变化规律	变化范围为3.23到11.86
column_418	无具体的连续变化规律	前6434个点，值从49834到49871，每约隔150个点的间隔增长1。后面为常数49872,49873，增长速度愈加缓慢
column_419	无具体的连续变化规律	变化范围为11340000到11357720
column_432	无具体的连续变化规律	前30000个点，值均为969.61，后面有增加，规模很小，7万个点才增加不到40，
column_437	无具体的连续变化规律	范围时间内从27.86-34.64
column_438	无具体的连续变化规律	范围时间内从27.86-34.64
column_439	无具体的连续变化规律	范围时间内从27.37-33.62
column_440	无具体的连续变化规律	范围时间内从27.76-34.45
column_441	无具体的连续变化规律	前6000个点为常数3.29，后面变化范围在3.29-942.65
column_442	无具体的连续变化规律	-18到18.39

# 总括

- |               |      |
|---------------|------|
| 1. 傅里叶函数      | 6个   |
| 2. 线性函数       | 9个   |
| 3. 分段混合函数     | 3个   |
| 4. 分段常数函数     | 9个   |
| 5. 无具体的连续变化规律 | 145个 |