

داده کاوی با نرم افزار

دکتر محمد ستاری
عضو هیات علمی گروه مدیریت و فناوری اطلاعات سلامت
دانشگاه علوم پزشکی اصفهان

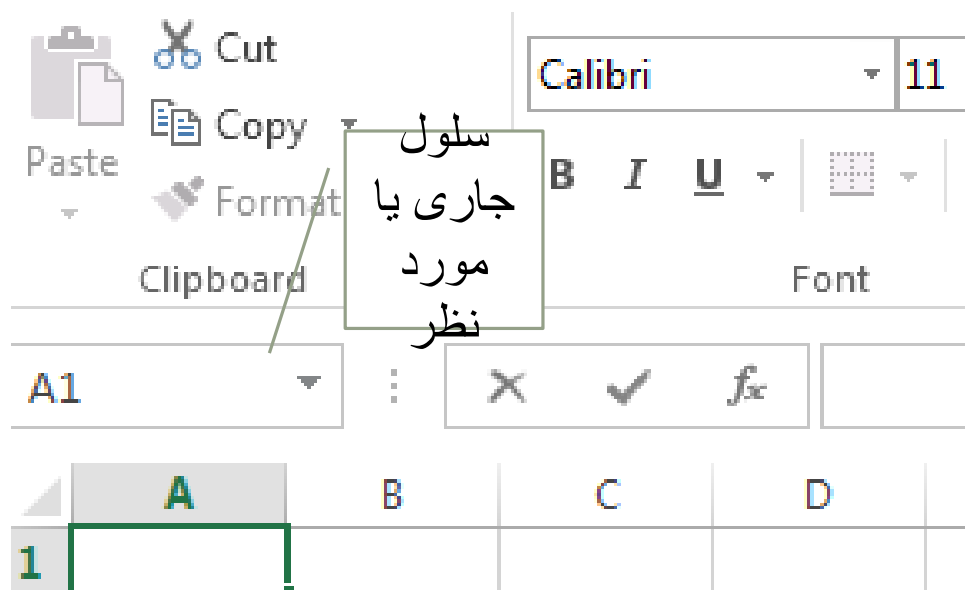
فرمول نویسی در اکسل

فرمولها، عبارتهای محاسباتی یا مقایسه ای هستند که روی انواع داده ها به کمک عملگرها، عملیاتی را انجام م ی دهند.

در اکسل فرمولها با علامت مساوی شروع می شوند.

فرمول نویسی در اکسل

هر فرمول بر روی سلول، سطر یا ستون جاری یا مورد نظر اعمال می شود



فرمول نویسی در اکسل

اجزای فرمول عبارت اند از :

ثابتها،

آدرس سلول ها (متغیرها،)

عملگرها

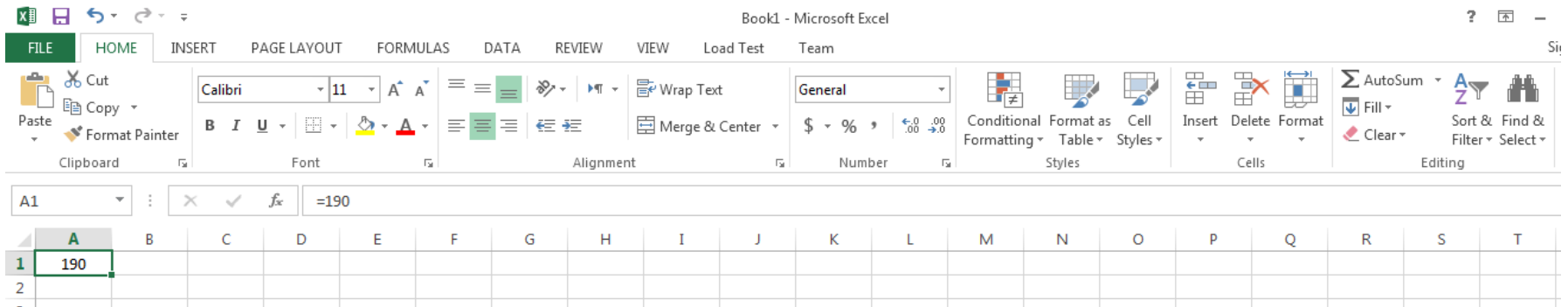
انواع تابع ها

فرمول نویسی در اکسل

ثابتها:

مقادیري هستند که تغییر نمی کنند. مثلا عدد 190

"ali"



فرمول نویسی در اکسل

آدرس سلول ها :

آدرس هر سلول براساس قاعده ی ذیل ایجاد می شود :

ColumnnumberRownumber

مثلا

A15

از طریق آدرس به محتویات سلول می توان دسترسی پیدا کرد

فرمول نویسی در اکسل

عملگرها:

علامت یا نشانه هایی هستند که نوع محاسبات را در يك عبارت مشخص مي کنند. عملگرها که میتوانید در فرمول ها به کار ببرید به سه دسته تقسیم میشوند:

عملگرهاي محاسبات رياضي
مقایسه اي

محدوده

فرمول نویسی در اکسل

توابع شرطی :

عملگر مقایسه ای در قالب تابع IF مطرح می شوند :

`IF(logicaltest,value_if_true,value_if_false)`

اگر `value_if_false` تعیین نشود FALSE نمایش داده می شود

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بزرگتر بودن سلول A2 از صفر را در سلول A3 بررسی کند
اگر مقدار سلول A2 بزرگتر از صفر بود مقدار سلول A3 "مثبت" شود

A3	:			<i>fx</i>	=IF(A2>0,"مثبت")
	A	B	C	D	E
1					
2	7				
3	مثبت				

A3	:			<i>fx</i>	=IF(A2>0,"مثبت")
	A	B	C	D	E
1					
2	-7				
3	FALSE				

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بزرگتر بودن سلول A2 از صفر را در سلول A3 بررسی کند
اگر مقدار سلول A2 بزرگتر از صفر بود مقدار سلول A3 "مثبت" شود در غیر اینصورت
مقدار سلول A3 "منفی یا صفر" شود

A3	:			<i>fx</i>	=IF(A2>0,"مثبت","منفی یا صفر")	
	A	B	C	D	E	F
1						
2	7					
3	مثبت					
4						
5						

A3	:			<i>fx</i>	=IF(A2>0,"مثبت","منفی یا صفر")	
	A	B	C	D	E	F
1						
2	-11					
3	منفی یا صفر					
4						
5						

فرمول نویسی در اکسل

توابع شرطی :

(عبارت شرطی 2, عبارت شرطی 1) AND

وقتی صحیح است که هر دو عبارت صحیح باشند

وقتی غلط است که شرط اول غلط باشد و شرط دوم صحیح باشد

وقتی غلط است که شرط اول صحیح باشد و شرط دوم غلط باشد

وقتی غلط است که هر دو شرط غلط باشد

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بزرگتر بودن A2 و A1 از صفر را در سلول A3 بررسی کند
اگر هر دو بزرگتر از صفر بودند مقدار سلول A3 "هر دو عدد مثبت هستند" شود در غیر
اینصورت مقدار سلول A3 "حداقل یکی از اعداد منفی است" شود

A3		fx =IF(AND(A1>0,A2>0),"هر دو عدد مثبت هستند","حداقل یکی از اعداد منفی یا صفر است")						
	A	B	C	D	E	F	G	H
1	12							
2	4							
3	هر دو عدد مثبت هستند							
4								
5								
6								

فرمول نویسی در اکسل

توابع شرطی :

(عبارت شرطی 2, عبارت شرطی 1) OR

وقتی صحیح است که حداقل یکی از عبارات صحیح باشند
وقتی غلط است که هر دو عبارت غلط باشند

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بزرگتر بودن A2 و A1 از صفر را در سلول A3 بررسی کند
اگر حداقل یکی از آنها بزرگتر از صفر بود مقدار سلول A3 "حداقل یکی از اعداد مثبت
هست" شود در غیر اینصورت مقدار سلول A3 "هر دو عدد منفی یا صفر هستند" شود

A3	:	  <i>fx</i>	=IF(OR(A1>0,A2>0),"حداقل یکی از اعداد مثبت هست","هر دو عدد منفی یا صفر هستند")						
	A	B	C	D	E	F	G	H	I
1	12								
2	-4								
3	حداقل یکی از اعداد مثبت هست								
4									
5									

فرمول نویسی در اکسل

توابع شرطی :

ISNUMBER(عبارت)

بررسی می کند عبارت عدد است

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بررسی کند در سلول A3، معدل یک دانشجو وارد شده باشد اگر معدل به صورت عددی وارد نشده است در B3 عبارت لطفاً "یک عدد وارد شود" نوشته شود

B3		fx =IF(NOT(ISNUMBER(A3)),"لطفاً یک عدد وارد کنید")					
	A	B	C	D	E	F	
1							
2	معدل						
3	ali	لطفاً یک عدد وارد کنید					
4							
5							
6							
7							

فرمول نویسی در اکسل

توابع شرطی :

ISTEXT(عبارت)

بررسی می کند عبارت متن است

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بررسی کند در سلول A3، نام یک دانشجو وارد شده باشد اگر نام به صورت عددی وارد شده باشد در B3 عبارت لطفا "نام به صورت صحیح وارد شود" نوشته شود

B3		fx =IF(NOT(ISTEXT(A3)),"نام به صورت صحیح وارد شود")				
	A	B	C	D	E	
1						
2	نام					
3	12	نام به صورت صحیح وارد شود				

فرمول نویسی در اکسل

توابع شرطی :

ISERROR(آدرس سلول)

بررسی می کند که آیا در سلول مورد نظر خطا رخ داده است یا خیر

فرمول نویسی در اکسل

مثال : فرمولی بنویسید که بررسی کند در سلول C5 خطا رخ داده است اگر خطایی رخ داده باشد عبارت "خطا رخ داده است" نوشته شود در غیر اینصورت عبارت عبارت "خطایی رخ نداده است" نوشته شود

D5 fx =IF(ISERROR(C5),"خطا رخ داده است","خطایی رخ نداده است")				
	A	B	C	D
1				
2				
3				
4				
5	12	0	#DIV/0!	خطا رخ داده است
6				
7				
8				

تکرار یک فرمول

با این تفاوت که در هنگام تکرار فرمول، آدرس سلول ها به صورت ذیل تغییر می یابد

یعنی اگر به سمت پایین حرکت کنیم یکی به شماره سطر اضافه میشود

اگر به سمت بالا حرکت کنیم یکی از شماره سطر کم می شود

تکرار یک فرمول

اگر به سمت چپ حرکت کنیم یکی از شماره ستون کم می شود (حرف الفبای قبلی جایگزین می شود)

اگر به سمت راست حرکت کنیم یکی به شماره ستون اضافه می شود (حرف الفبای بعدی جایگزین می شود)

تکرار یک فرمول

مثال : نمرات دانشجویان یک کلاس دریافت کرده و دانشجویان ممتاز را مشخص کنید ؟ (ممتاز = معدل بزرگتر یا مساوی ۱۷)

B1	:	X	✓	<i>f_x</i>	=IF(A1>=17,"ممتاز","ممتاز نیست")		
	A	B	C	D	E	F	
1	17	ممتاز					
2	16	ممتاز نیست					
3	17	ممتاز					
4	18	ممتاز					
5	20	ممتاز					
6	19	ممتاز					
7	11	ممتاز نیست					
8	13	ممتاز نیست					
9							

پاک کردن مقادیر تکراری

برای پاک کردن مقادیر تکراری در اکسل، ساده ترین و رایج ترین روش استفاده از ابزار داخلی **Remove Duplicates** است. مراحل کار به طور خلاصه عبارتند از:

محدوده یا جدول داده های خود را انتخاب کنید (برای کل جدول **Ctrl + A** را بزنید).
به تب **Data** در بالای نرم افزار بروید.

در بخش **Data Tools**، روی گزینه **Remove Duplicates** کلیک کنید.

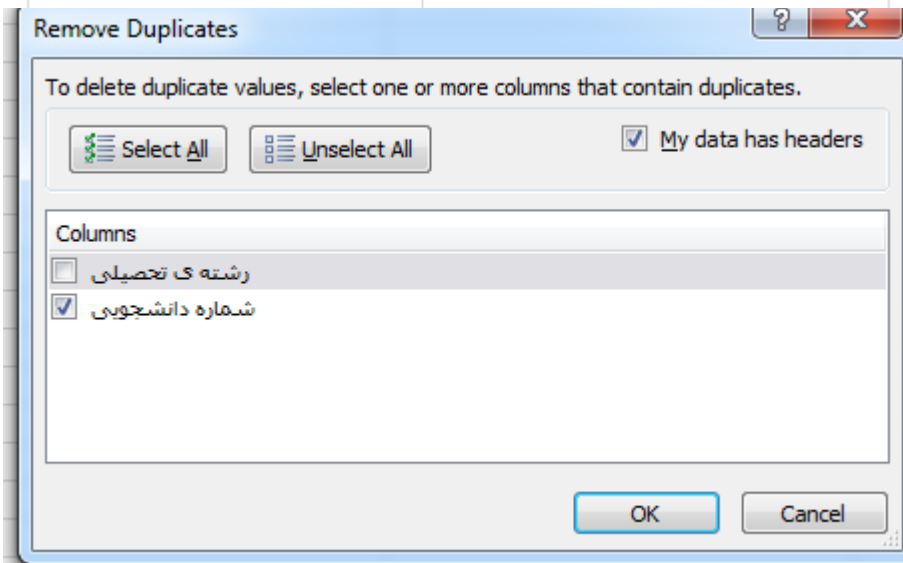
در پنجره باز شده، ستونهایی که می خواهید تکراریها بر اساس آنها شناسایی و حذف شوند را انتخاب کنید. اگر داده ها شامل هدر (عنوان ستون) هستند، گزینه **My data has headers** را فعال کنید.

دکمه **OK** را بزنید تا اکسل موارد تکراری را حذف کند و پیام تعداد موارد حذف شده و باقی مانده را نمایش دهد.

در این روش، تنها کپی اول از هر داده تکراری باقی میماند و موارد بعدی حذف می شوند. این قابلیت در همه نسخه های اکسل وجود دارد و برای حذف تکراری ها در چند ستون یا کل جدول قابل استفاده است

مثال : یک نمونه از پاک کردن مقادیر تکراری در شماره دانشجویی

شماره دانشجویی	رشته ی تحصیلی
981111111	فناوری اطلاعات سلامت
98222222	فناوری اطلاعات سلامت
98333333	فناوری اطلاعات سلامت
98444444	فناوری اطلاعات سلامت
981111111	مدیریت خدمات بهداشتی و درمانی



شماره دانشجویی	رشته ی تحصیلی
981111111	فناوری اطلاعات سلامت
98222222	فناوری اطلاعات سلامت
98333333	فناوری اطلاعات سلامت
98444444	فناوری اطلاعات سلامت

فیلترهای متنی

در این فیلترها

? : یک کاراکتر

• * : مجموعه ای از کاراکترها

فلترهای متنی

Equals : با این الگو برابر است

نام	نام خانوادگی	شماره دانشجویی	رشته	نمره
علی	احمدی	981111111	فناوری اطلاعات سلامت	18
امیر	امیری	982222222	مدیریت خدمات بهداشتی و درمانی	16
سعید	سعیدی	983333333	فناوری اطلاعات سلامت	17
علی	اکبری	984444444	مدیریت خدمات بهداشتی و درمانی	15
احمد	احمدی	985555555	فناوری اطلاعات سلامت	17
محمد	محمدی	986666666	فناوری اطلاعات سلامت	19
سعید	اکبری	987777777	مدیریت خدمات بهداشتی و درمانی	15
امیر	علیزاده	988888888	فناوری اطلاعات سلامت	17
علی	امیری	989999999	مدیریت خدمات بهداشتی و درمانی	17

Custom AutoFilter

Show rows where:

نام خانوادگی

equals

☒ And ☐ Or

Use ? to represent any single character
Use * to represent any series of characters

OK Cancel

نام	نام خانوادگی	شماره دانشجویی	رشته	نمره
علی	احمدی	981111111	فناوری اطلاعات سلامت	18
امیر	امیری	982222222	مدیریت خدمات بهداشتی و درمانی	16
علی	اکبری	984444444	مدیریت خدمات بهداشتی و درمانی	15
احمد	احمدی	985555555	فناوری اطلاعات سلامت	17
سعید	اکبری	987777777	مدیریت خدمات بهداشتی و درمانی	15
علی	امیری	989999999	مدیریت خدمات بهداشتی و درمانی	17

مثال : الگوی دانشگاه در راهنمای تلفن را در جدول ذیل اعمال کنید ؟

نام	نام خانوادگی	شماره دانشجویی	رشته	نمره
علی	محمدی	981111111	فناوری اطلاعات سلامت	18
امیر	محمدیان	982222222	مدیریت خدمات بهداشتی و درمانی	16
سعید	دوست محمدی	983333333	فناوری اطلاعات سلامت	17
علی	محمدی فرد	984444444	مدیریت خدمات بهداشتی و درمانی	15
احمد	علیمحمدی	985555555	فناوری اطلاعات سلامت	17
محمد	محمدیان	986666666	فناوری اطلاعات سلامت	19
سعید	یارمحمدیان	987777777	مدیریت خدمات بهداشتی و درمانی	15
امیر	علیزاده	988888888	فناوری اطلاعات سلامت	17
علی	امیری	989999999	مدیریت خدمات بهداشتی و درمانی	17

Custom AutoFilter

Show rows where:

نام خانوادگی

equals *محمدی*

☒ And ☐ Or

Use ? to represent any single character
Use * to represent any series of characters

OK Cancel

نام	نام خانوادگی	شماره دانشجویی	رشته	نمره
علی	محمدی	981111111	فناوری اطلاعات سلامت	18
امیر	محمدیان	982222222	مدیریت خدمات بهداشتی و درمانی	16
سعید	دوست محمدی	983333333	فناوری اطلاعات سلامت	17
علی	محمدی فرد	984444444	مدیریت خدمات بهداشتی و درمانی	15
احمد	علیمحمدی	985555555	فناوری اطلاعات سلامت	17
محمد	محمدیان	986666666	فناوری اطلاعات سلامت	19
سعید	یارمحمدیان	987777777	مدیریت خدمات بهداشتی و درمانی	15

جدول محوری

جدول محوری ابزاری است که برای سازماندهی مجدد داده‌ها در یک صفحه گسترده برای دستیابی به گزارش مطلوب به طور مؤثرتر استفاده می‌شود.

این ممکن است شامل مجموعه‌ها، میانگین‌ها یا سایر آمارها باشد که جدول محوری آنها را به روشی معنادار در کنار هم قرار می‌دهد.

پس از ایجاد یک جدول محوری، می‌توانید به سرعت تعداد زیادی ردیف و ستون را به یک گزارش معنادار و با قالب‌بندی زیبا تبدیل کنید.

جدول محوری

1. ابتدا داده های خود را در یک محدوده از ردیف ها و ستون ها وارد کنید. این داده ها باید شامل سرفصل های مشترک در ستونها باشند (مثلاً نام، تاریخ، مقدار و غیره).
2. یکی از سلولهای داده ها را انتخاب کنید.
3. به تب Insert در نوار ابزار بالای اکسل بروید.
4. در بخش Tables روی گزینه PivotTable کلیک کنید.
5. در پنجره باز شده، محدوده داده ها به طور خودکار انتخاب می شود؛ در صورت نیاز آن را بررسی و اصلاح کنید.

جدول محوری

6. محل قرارگیری جدول محوری را مشخص کنید:

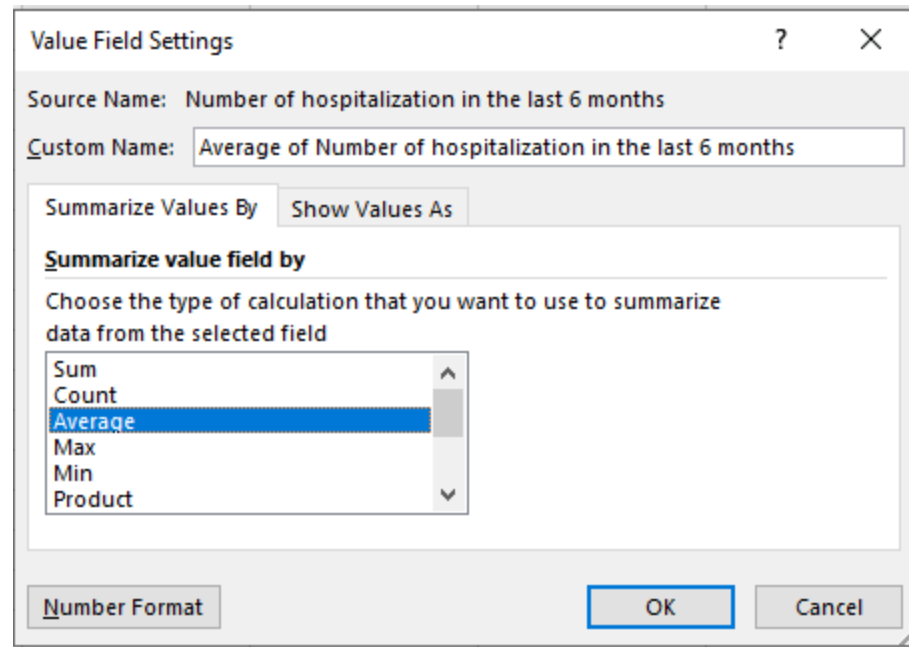
7. اگر گزینه New worksheet را انتخاب کنید، جدول محوری در یک صفحه کاری جدید ساخته میشود.

8. اگر گزینه Existing worksheet را انتخاب کنید، با تعیین آدرس سلول، جدول محوری در همان صفحه فعلی نمایش داده میشود.

9. روی OK کلیک کنید تا پنجره فیلدهای PivotTable باز شود.

جدول محوری

در قسمت Value Field Setting می توانید نوع تابع تجمعی را انتخاب نمایید



vlookup

تابع VLOOKUP در اکسل یک تابع بسیار کاربردی برای جستجوی یک مقدار مشخص در ستون اول یک جدول یا محدوده داده است و سپس مقدار متناظر را از همان ردیف، اما از ستونی که شما مشخص می کنید، برمی گرداند. حرف V در VLOOKUP به معنای Vertical یا عمودی است، به این معنی که جستجو به صورت عمودی در ستون ها انجام میشود.

vlookup

فرمول کلی تابع VLOOKUP به صورت زیر است:

`VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])`

`lookup_value` مقداری که می خواهید در ستون اول جستجو کنید، مثل نام یا شناسه محصول.

`table_array` بازه یا جدول داده ای که می خواهید در آن جستجو انجام شود.

`col_index_num` شماره ستونی که می خواهید مقدار متناظر از آن استخراج شود. ستون اول جدول عدد ۱ در نظر گرفته می شود.

`range_lookup` مشخص می کند جستجو دقیق باشد یا تقریبی. برای جستجوی دقیق معمولاً `FALSE` استفاده میشود.

vlookup

	A	B	C	D	E	F	G	H
1	کد محصول	نام محصول	قیمت		جستجوی موردی			
2	106	آبمیوه آناناس	15999		نام محصول	قیمت		
3	108	آبمیوه پرتقال	47003		بستنی وانیلی	FALSE)		
4	120	آبمیوه آلبالو	7635					
5	143	آبمیوه هلو	9614					
6	116	بستنی آناناس	22604					
7	119	بستنی موز	24135					
8	125	بستنی انبه	43599					
9	127	بستنی توت فرنگی	12393					
10	173	بستنی شائوت	31228					
11	187	بستنی نوتلا	14280					
12	207	بستنی وانیلی	17563					
13	208	بستنی کاکائویی	6290					
14	215	بستنی قیفی ساده	41485					
15	131	بیسکویت شهد توت	8935					
16	153	بیسکویت ستاک	17797					
17	176	بیسکویت بدون قند	5679					
18	191	بیسکویت ساده	28961					

Function Arguments

VLOOKUP

Lookup_value: E3 = 'بستنی وانیلی'

Table_array: B:C = [-]

Col_index_num: 2 = 2

Range_lookup: FALSE = FALSE

= 17563

Looks for a value in the leftmost column of a table, and then returns a value in the same row from a column you specify. By default, the table must be sorted in an ascending order.

Range_lookup is a logical value: to find the closest match in the first column (sorted in ascending order) = TRUE or omitted; find an exact match = FALSE.

Formula result = 17563

[Help on this function](#)

OK Cancel

تابع Match

تابع MATCH در اکسل برای جستجو کردن یک مقدار مشخص در یک محدوده (یک سطر یا یک ستون) استفاده می شود و جایگاه یا موقعیت نسبی آن مقدار را به صورت عدد برمی گرداند. به بیان دیگر، این تابع مشخص می کند که مقدار مورد نظر در محدوده ی جستجو چندمین عضو است

تابع MATCH معمولاً اولین تطابق را برمیگرداند و بیشتر مواقع با تابع INDEX ترکیب میشود تا مقدار متناظر را بر اساس موقعیتی که MATCH پیدا میکند بازیابی کنند.

تابع Match

فرمول کلی تابع MATCH به شکل زیر است:

`=MATCH(lookup_value,lookup_array,[match_type])`

lookup_value مقدار یا عبارتی که می خواهید موقعیت آن را جستجو کنید (عدد، متن، مقدار منطقی یا آدرس یک سلول).

lookup_array محدوده ای از سلول ها که در آن جستجو انجام می شود (باید یک بعدی باشد، یعنی یک سطر یا یک ستون).

تابع Match

match_type نوع تطبیق را مشخص میکند و میتواند یکی از سه مقدار زیر باشد:

- ۰: جستجوی تطابق دقیق مقدار (نیاز به مرتب کردن محدوده نیست).
- ۱: جستجوی نزدیکترین مقدار کوچکتر یا مساوی مقدار مورد جستجو (محدوده باید به صورت صعودی مرتب شده باشد).
- -۱: جستجوی نزدیکترین مقدار بزرگتر یا مساوی مقدار مورد جستجو (محدوده باید به صورت نزولی مرتب شده باشد).

تابع Match

MATCH(D6,B6:B14,0)

تابع Match برای تطابق دقیق

E6					=MATCH(D6,B6:B14,0)
	A	B	C	D	E
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					

MATCH function
MATCH (lookup_value, lookup_array, match_type)

Fruit	
Apple	1
Pear	2
Grape	3
Lemon	4
Peach	5
Lime	6
Kiwi	7
Mango	8
Pineapple	9

Lookup	Result
Peach	5



تابع Match

MATCH(E2,B3:B11,1)

تابع Match برای تطابق تقریبی

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1						
2		Values		Lookup	575	
3	1	100		Result	5	
4	2	200				
5	3	300				
6	4	400				
7	5	500				
8	6	600				
9	7	700				
10	8	800				
11	9	900				
12						

استفاده از جستجوی تقریبی با توجه به مرتب سازی صعودی لیست

تابع index

تابع INDEX در اکسل یک تابع جستجو و مرجع است که برای بازگرداندن مقدار یک سلول در یک محدوده مشخص بر اساس شماره ردیف و ستون استفاده می‌شود)

`INDEX(array, row_num, [column_num])`

Array محدوده سلول‌هایی که می‌خواهید از آن‌ها مقدار را بازیابی کنید.
row_num شماره ردیف سلول مورد نظر.
column_num شماره ستون سلول مورد نظر (اختیاری).

`INDEX(A1:C5, 3, 2)`

این فرمول مقدار موجود در سلول B3 را برمی‌گرداند

ترکیب تابع Index و Match

رشته‌ی ما در سلول C10 قرار دارد. سپس یک مجموعه از سلول‌ها را برای جست‌وجوی خود انتخاب می‌نماییم، عدد ۰ نیز مشخص می‌کند که دقیقاً به دنبال همان رشته می‌گردیم، نه کمتر و نه بیشتر. این تابع محل سلول را بر می‌گرداند که مقدار آن را در سلول C12 مشاهده می‌کنید، یعنی جنسی که آی‌دی آن BE99 است، در سطر چهارم از محدوده‌ی مشخص شده‌ی ما قرار داد؛ به زبان ساده‌تر منظور سلول A5 است.

C12

✕

✓

f_x

=MATCH(C10,A2:A8,0)

	A	B	C	D	E	F
1	Stock ID	Item Name	Price			
2	TS22	T-shirt	\$4.80			
3	JE99	Jeans	\$28.90			
4	SO77	Socks	\$2.76			
5	BE99	Belt	\$12.99			
6	BA77	Bandanna	\$7.76			
7	SU66	Sunglasses	\$4.60			
8	JA88	Jacket	\$88.88			
9						
10		Stock ID	BE99			
11		Price				
12		Item Pos.	4			
13						

تابع Forecast

تابع FORECAST در اکسل برای پیشبینی یک مقدار جدید از متغیر وابسته بر اساس رابطه خطی با یک متغیر مستقل استفاده می شود. این تابع بر اساس داده های تاریخی مقادیر متغیر مستقل و وابسته یک مدل رگرسیون خطی ساده می سازد و مقدار y پیش بینی شده را برای مقدار جدید x برمی گرداند. ساختار تابع به صورت زیر است:

`FORECAST.LINEAR(x,known_ys,known_xs)`

تابع Forecast

x مقداری از متغیر مستقل که برای آن می خواهید مقدار y را پیش بینی کنید.
known_ys: آرایه یا بازه ای از مقادیر متغیر وابسته در یک بازه ی تاریخی
known_xs: آرایه یا بازه ای از مقادیر متغیر مستقل در یک بازه ی تاریخی

تابع Forecast - مثال

B	C	D
ماه	تعداد فروش محصول	
1	1000	
2	1320	
3	1500	
4	1600	
5	876	
6	500	
7	323	
8	600	
9	800	
10	400	
11	540	

تابع Forecast- مثال

محدوده مقادیر متغیر در دوره‌های گذشته : از اونجایی که ما مقادیر یازده ماه گذشته رو داریم پس این پارامتر شامل محدوده A2:A12 یعنی عددی ۱ تا ۱۱ می شود.

محدوده مقدار تابع برای متغیرهای گذشته : این محدوده باید مقدار تابع یعنی تعداد فروش رو به ازای عددی یک تا یازده به ما نشان بدهد. پس این محدوده رنج B2:B12 رو شامل می شود.

مقدار متغیر که برای آن قصد پیش بینی تابع را داریم : ما می‌خواهیم مقدار تعداد فروش رو برای ماه ۱۲ حساب کنیم پس مقدار متغیری که ما می‌خواهیم برایش پیش بینی کنیم سلول A13 هست.

مقدار دهی فرمول : در نهایت فرمول به شکل زیر خواهد بود

FORECAST.LINEAR(A13;B2:B12;A2:A12)=

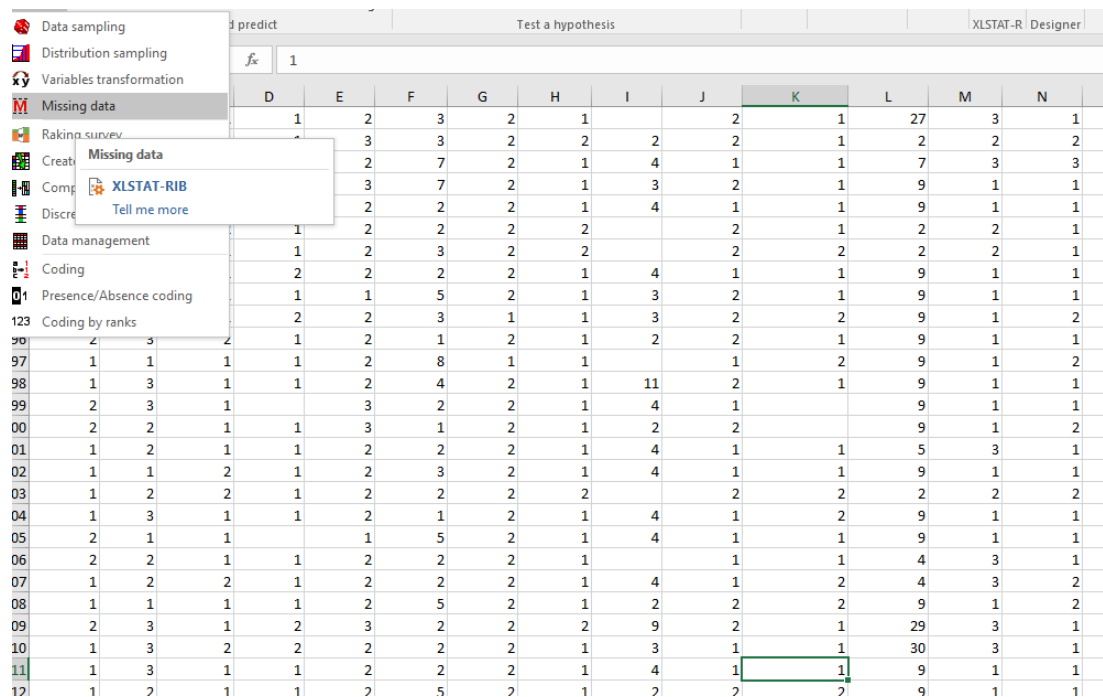
12	1439.891
----	----------

تابع Forecast - مثال

B	C	D
ماه	تعداد فروش محصول	
1	1000	
2	1320	
3	1500	
4	1600	
5	876	
6	500	
7	323	
8	600	
9	800	
10	400	
11	540	
12	1439.890909	

Missing data

وارد قسمت Missing data شوید



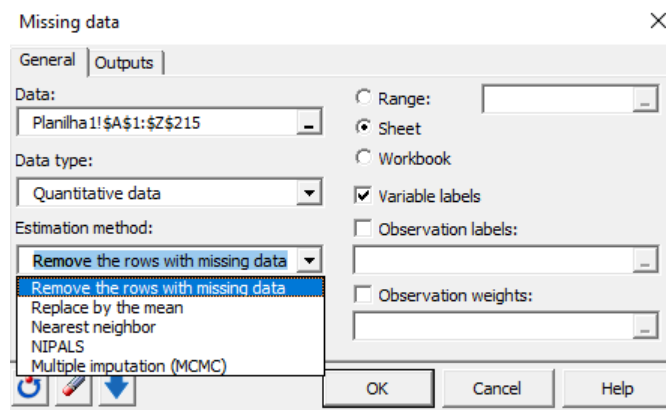
The screenshot shows the XLSTAT software interface. On the left, a menu is open with 'Missing data' selected. The main window displays a data table with columns D through N and rows 1 through 12. The table contains numerical data, with some cells highlighted in green.

	D	E	F	G	H	I	J	K	L	M	N	O
1	1	2	3	2	1		2	1	27	3	1	
2		3	3	2	2	2	2	1	2	2	2	
3		2	7	2	1	4	1	1	7	3	3	
4		3	7	2	1	3	2	1	9	1	1	
5		2	2	2	1	4	1	1	9	1	1	
6		2	2	2	2		2	1	2	2	1	
7		1	2	3	2	2		2	2	2	1	
8		2	2	2	2	1	4	1	9	1	1	
9		1	1	5	2	1	3	2	1	9	1	1
10		2	2	3	1	1	3	2	9	1	2	
11		1	1	1	2	1		1	9	1	1	
12		1	1	1	2	1	2	2	9	1	1	
13		1	1	1	2	1	2	2	9	1	2	
14		1	3	1	1	2	11	2	9	1	1	
15		2	3	1		3	4	1	9	1	1	
16		2	2	1	1	2	2		9	1	2	
17		1	2	1	1	2	4	1	5	3	1	
18		1	1	2	1	2	1	1	9	1	1	
19		1	2	2	1	2	2	2	2	2	2	
20		1	3	1	1	2	1	2	9	1	1	
21		2	1	1		1	5	2	1	9	1	1
22		2	2	1	1	2	2	2	1	4	3	1
23		1	2	2	1	2	2	1	2	4	3	2
24		1	1	1	1	2	1	2	2	9	1	2
25		2	3	1	2	3	2	2	29	3	1	
26		1	3	2	2	2	1	3	1	30	3	1
27		1	3	1	1	2	2	1	9	1	1	
28		1	2	1	1	2	2	2	9	1	1	

Missing data

از قسمت Data، قسمتی که قرار است missing value مشخص شود را انتخاب نمایید

از قسمت Estimation method، روشی که قرار است برای مدیریت داده های از دست رفته انتخاب شود را مشخص نمایید



Missing data

یکی از خروجی ها این است که هر صفت چند missing data دارد همچنین شاخص های دیگر آماری نیز محاسبه می شود

Variable	bservation	ith miss	in	hout miss	Minimum	Maximum	Mean	d. deviation
Gender	214	0	214	0/000	2/000	1/416	0/503	
Age Range	214	0	214	0/000	3/000	1/776	0/837	
Place of re	214	0	214	0/000	2/000	1/313	0/475	
Race	214	10	204	1/000	10/000	1/181	0/710	
Marital sta	214	1	213	1/000	4/000	2/070	0/523	
Schooling	214	0	214	0/000	8/000	3/252	1/826	
Paid Work	214	0	214	0/000	2/000	1/855	0/366	
Type of Tr	214	1	213	1/000	2/000	1/268	0/444	
Cancer Sta	214	34	180	1/000	34/000	3/767	2/910	
Metastasi	214	5	209	1/000	5/000	1/603	0/597	
Comorbid	214	3	211	1/000	3/000	1/479	0/510	
Main Com	214	2	212	1/000	30/000	7/901	5/730	
Main Com	214	2	212	1/000	3/000	1/882	0/876	
Continuin	214	0	214	0/000	3/000	1/322	0/507	
Diagnosis	214	0	214	0/000	21/000	5/187	5/135	
Diagnosis	214	0	214	0/000	7/000	3/430	1/542	
Cancer Ty	214	2	212	1/000	3/000	1/401	0/501	
Year of Di	214	0	214	0/000	6/000	2/014	1/200	
Oncology	214	10	204	1/000	10/000	2/936	1/978	
First Hosp	214	0	214	0/000	2/000	1/575	0/523	
Readmissi	214	2	212	1/000	2/000	1/840	0/368	
Number o	214	4	210	0/000	4/000	0/295	0/750	
Readmissi	214	3	211	1/000	3/000	1/763	0/437	
Number o	214	6	208	0/000	12/000	0/942	2/350	
Readmissi	214	3	211	1/000	3/000	1/829	0/389	
Number o	214	6	208	0/000	29/000	1/341	4/659	

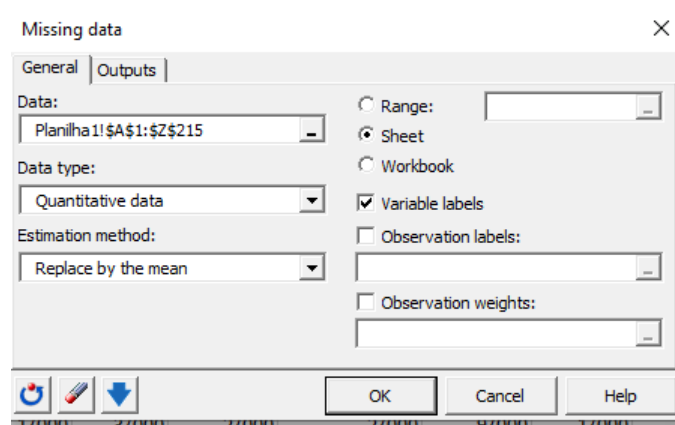
Missing data

داده ها پس از missing در قسمت completed data قرار می گیرد.

Completed data:																				
	Gender	Age Range	ee of reside	Race	larital stat	Schooling	Work Act	e of Treat	ncer Stagi	Metastasis	morbidity	in Compl	plaint grow	ing Medi	Diagnosis	osis group	Cancer Typ	of Diagn	ology Treat	Hospitaliz
Obs1	1/000	2/000	1/000	1/000	2/000	2/000	2/000	1/000	3/000	2/000	1/000	4/000	3/000	1/000	2/000	3/000	1/000	2/000	7/000	2/000
Obs2	1/000	1/000	1/000	1/000	2/000	6/000	1/000	1/000	9/000	2/000	2/000	4/000	3/000	2/000	4/000	5/000	1/000	1/000	2/000	2/000
Obs5	1/000	3/000	1/000	1/000	2/000	3/000	2/000	1/000	4/000	1/000	2/000	5/000	3/000	1/000	1/000	3/000	2/000	2/000	2/000	2/000
Obs6	1/000	1/000	1/000	1/000	2/000	5/000	2/000	1/000	4/000	1/000	2/000	3/000	3/000	1/000	1/000	3/000	2/000	2/000	2/000	1/000
Obs7	2/000	1/000	2/000	2/000	2/000	2/000	2/000	1/000	11/000	2/000	2/000	3/000	3/000	2/000	3/000	5/000	1/000	5/000	2/000	2/000
Obs8	1/000	3/000	1/000	1/000	2/000	2/000	1/000	1/000	3/000	2/000	1/000	4/000	3/000	1/000	2/000	3/000	2/000	2/000	2/000	2/000
Obs9	1/000	1/000	1/000	1/000	1/000	2/000	2/000	2/000	4/000	1/000	2/000	6/000	3/000	1/000	5/000	7/000	1/000	1/000	5/000	1/000
Obs10	1/000	1/000	1/000	1/000	2/000	2/000	2/000	1/000	11/000	2/000	2/000	7/000	3/000	2/000	3/000	5/000	1/000	3/000	6/000	1/000
Obs11	1/000	3/000	1/000	1/000	3/000	2/000	2/000	1/000	2/000	2/000	2/000	9/000	1/000	1/000	1/000	3/000	2/000	1/000	2/000	1/000
Obs13	2/000	1/000	1/000	1/000	2/000	7/000	2/000	1/000	3/000	2/000	2/000	9/000	1/000	2/000	1/000	3/000	2/000	2/000	2/000	2/000
Obs14	1/000	1/000	1/000	2/000	2/000	2/000	2/000	1/000	4/000	2/000	1/000	9/000	1/000	1/000	2/000	3/000	2/000	3/000	2/000	2/000
Obs15	1/000	3/000	1/000	1/000	2/000	3/000	2/000	1/000	4/000	1/000	1/000	8/000	3/000	1/000	7/000	4/000	2/000	4/000	4/000	2/000
Obs17	2/000	1/000	1/000	1/000	2/000	8/000	2/000	1/000	3/000	2/000	2/000	9/000	1/000	2/000	1/000	3/000	1/000	2/000	2/000	2/000
Obs18	1/000	1/000	1/000	1/000	2/000	2/000	2/000	1/000	3/000	2/000	2/000	9/000	1/000	2/000	2/000	3/000	2/000	2/000	1/000	2/000
Obs19	1/000	1/000	2/000	1/000	2/000	2/000	1/000	1/000	4/000	1/000	2/000	4/000	3/000	2/000	6/000	1/000	1/000	1/000	3/000	1/000
Obs20	2/000	1/000	1/000	1/000	2/000	4/000	1/000	1/000	4/000	1/000	1/000	9/000	1/000	1/000	1/000	3/000	2/000	2/000	1/000	2/000
Obs22	2/000	2/000	1/000	1/000	2/000	2/000	2/000	1/000	4/000	1/000	1/000	5/000	3/000	1/000	1/000	3/000	2/000	3/000	2/000	2/000
Obs23	1/000	1/000	2/000	1/000	2/000	2/000	2/000	1/000	3/000	2/000	2/000	9/000	1/000	2/000	9/000	3/000	1/000	1/000	2/000	2/000
Obs24	1/000	1/000	1/000	1/000	2/000	3/000	2/000	1/000	4/000	1/000	2/000	9/000	1/000	2/000	1/000	3/000	1/000	1/000	1/000	2/000

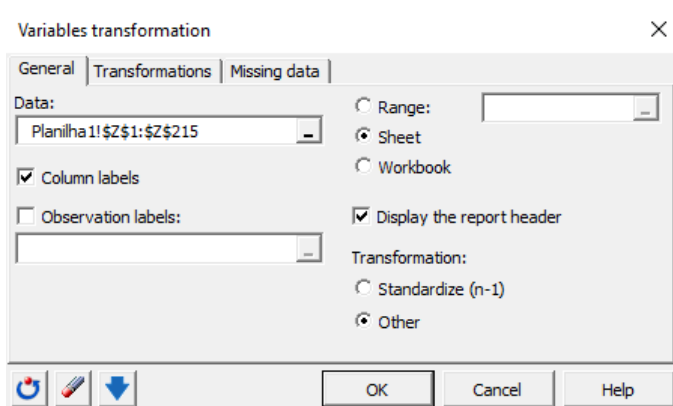
Missing data

از متد میانگین برای داده های از دست رفته استفاده می شود.



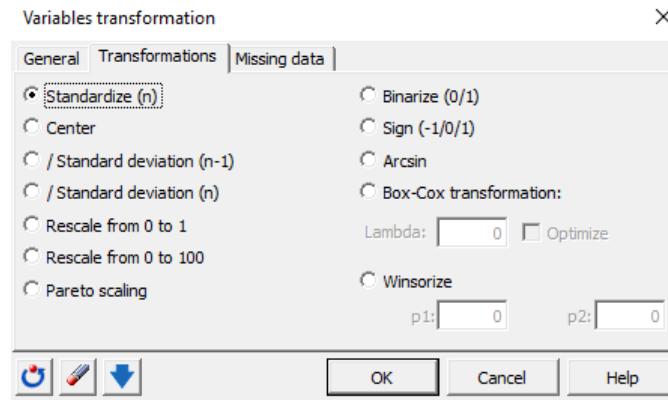
Transformation

Transformation: قسمت other انتخاب شود



Transformation

روش های Transformation



Transformation

روش های Transformation:

تبدیل باکس-کاکس Box-Cox Transformation

تبدیل با هدف نزدیک کردن توزیع داده به حالت نرمال انجام می شود. پارامتر λ لامبدا به گونه ای انتخاب میشود که توزیع داده بهبود یابد

Standard Deviation

استفاده از انحراف معیار

نرمال سازی به یک بازه ی مشخص مثلا صفر تا یک

Transformation

روش های Transformation:

باینری سازی: Binarization
تبدیل مقادیر به ۰ یا ۱ که در مسائل طبقه بندی مفید است

تبدیل‌های غیرخطی دیگر مانند Arcsin و Winsorize
برای بهبود نرمال بودن و کاهش اثر مقادیر پرت استفاده می شوند.

گسسته سازی

گسسته سازی

Discretization

General | Options | Missing data | Outputs | Charts

Data: Planilha1!\$Z:\$Z

Method: Constant range

Range: 1

☐ Minimum: 0

☐ Range:

☒ Sheet

☐ Workbook

☒ Column labels

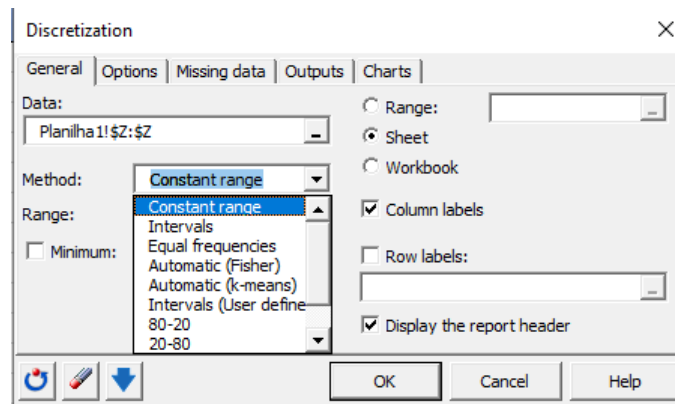
☐ Row labels:

☒ Display the report header

OK Cancel Help

روش های گسسته سازی

روش های گسسته سازی



روش های گسسته سازی

گسسته سازی بر اساس بازه های تعریف شده توسط کاربر User-defined Intervals
کاربر بازه های دلخواهی را مشخص میکند و داده ها بر اساس آنها دسته بندی می شوند.

گسسته سازی Equal frequency:
داده ها به تعدادی کلاس با تعداد نمونه تقریباً برابر تقسیم می شوند.

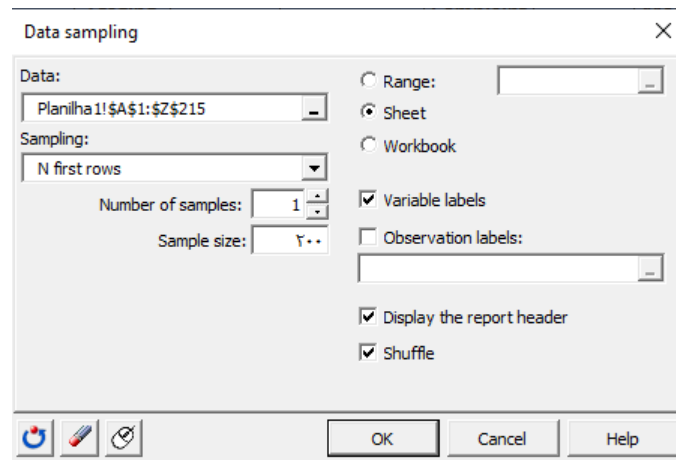
روش های گسسته سازی

روش گسسته سازی بر اساس بازه ثابت Constant Range:

شما داده های پیوسته را به کلاس هایی تقسیم می کنید که هر کلاس یا بازه دارای طول (محدوده) یکسانی است. به عبارت دیگر، در این روش دامنه متغیر به بخش های مساوی یا با اندازه ثابتی تقسیم می شود، مثلاً اگر داده از ۰ تا ۱۰۰ باشد و بخواهیم آن را در ۱۰ کلاس دسته بندی کنیم، هر کلاس بازه ۱۰ واحدی خواهد داشت (۰-۱۰، ۱۰-۲۰، ۲۰-۳۰ و ...).

نمونه برداری

نمونه برداری با n ردیف اول



نمونه برداری

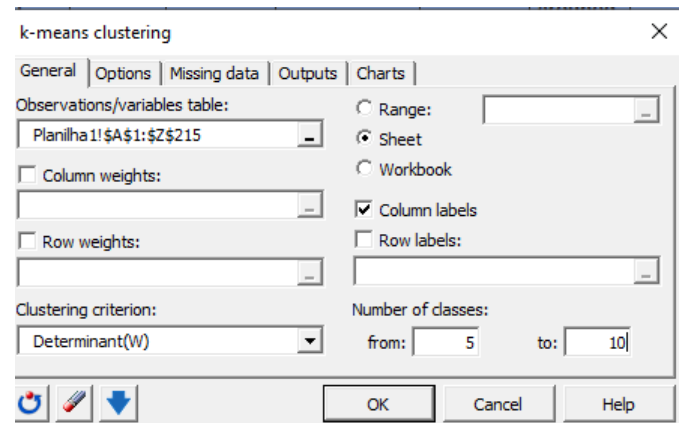
نمونه برداری با n ردیف اول

Sample 1:																				
Gender	Age	Range of residence	Race	Marital status	Schooling	Work	Act of Treatment	Cancer Stage	Metastasis	Morbidity	in Compliance	Complaint	giving Medicine	Diagnosis	osis group	Cancer Type	of Diagnosis	ology Treatment	Hospitalization	
1	2	1	2	2	2	2	2	1	3	2	1	9	1	1	2	3	1/400943	1	3	2
1	2	2	2	2	2	2	2	1	3	2	1	9	1	1	1	3	2	2	1	2
1	2	1	1	2	2	2	2	2	3/766667	2	2	2	2	1	7	4	1	2	5	1
1	2	2	1	2	2	2	2	1	3	2	2	9	1	2	1	3	1	2	2	1
1	2	2	1	4	5	2	1	2	2	2	1	9	1	1	1	3	2	2	1	1
2	2	1	1	3	3	2	2	2	2	2	1	2	2	2	1	3	1	1	5	1
1	1	2	1	2	4	2	1	3	2	2	9	1	2	13	7	1	2	2	2	2
1	1	2	2	2	4	2	1	3	2	1	9	1	1	12	3	2	1	1	2	2
2	1	1	1	2	4	1	1	4	1	1	9	1	1	1	3	2	2	1	2	2
1	3	2	1	2	2	2	1	2	2	1	7	3	1	6	1	1	5	2	1	1
1	1	2	1	2	2	2	1	3	2	2	9	1	2	9	3	1	1	2	2	2
2	1	1	1	3	3	2	1	3	2	2	9	1	1	12	3	1	2	2	2	2
2	3	1	1	2	2	2	2	4	1	1	2	2	1	8	4	1	2	5	2	2
2	1	1	2	2	5	2	2	3/766667	1/602871	1	2	2	1	8	4	1	2	5	1	1
1	2	1	2	2	2	2	1	4	1	1	9	1	1	1	3	2	2	2	2	2
1	3	2	1	3	1	2	2	3/766667	1	1	13	3	2	1	3	1	2	6	2	2
1	1	1	1	2	3	2	1	4	1	2	9	1	2	1	3	1	1	1	2	2
2	1	1	1	2	5	2	2	4	1	2	2	2	1	17	7	2	2	5	2	2
1	1	1	1	1	5	2	1	4	1	2	4	3	2	1	3	2	3	2/936275	2	2
2	2	1	1	2	2	2	1	4	1	1	6	3	1	12	3	2	1	6	1	1
1	1	2	1/181373	2	2	1	1	4	1	1	3	3	1	1	3	2	3	2	1	1
1	1	1	1	2	1	2	1	2	2	1	9	1	1	1	3	1	2	2	2	2
2	2	1	1	3	1	2	1	2	2	1/478673	9	1	2	9	3	1	1	1	1	1

خوشه بندی k-means

خوشه بندی k-means: تعداد کلاس ها از 5 تا 10 باشد

معیار خوشه بندی Determinant



رگرسیون خطی

Linear regression

General | Options | Validation | Prediction | Missing data | Outputs | Charts

Y / Dependent variables:

Quantitative:

Planilha1!\$I:\$I

X / Explanatory variables:

☒ Quantitative:

Planilha1!\$A\$1:\$H\$214

☐ Qualitative:

☐ Range:

☒ Sheet

☐ Workbook

☒ Variable labels

☐ Observation labels:

☐ Observation weights:

☐ Regression weights:

OK Cancel Help

k-نزدیکترین همسایه

تعداد همسایه ها 5 است

K Nearest Neighbors

General Options Validation Missing data Outputs

Training set: ☐ Range:

Y / Qualitative variables: ☐ Sheet ☐ Workbook

Data type:

X / Explanatory variables: ☒ Variable labels

☒ Quantitative: ☐ Observation labels:

☐ Qualitative: ☐ Weights:

Prediction set:

Quantitative variables:

Qualitative variables:

Number of neighbors:

   OK Cancel Help