

## 数据有效性筛选

### 学习目标

本章主要学习如何筛选有效数据,通过工具栏和函数实现限定条件下的数据提取,并借助筛选条件的设定、高级筛选运用等,实现对疑点数据的定位和提取。

数据的有效性往往与数据本身代表的意义相关联。每类数据的存在都有其约束性条件,不合理范围的存在即为违规违法操作。本章借助经济业务中各项业务执行的红线标准,以此为数据有效性的成立条件对数据进行筛查,得到超越合理范围的疑点数据。

### 5.1 数据排序的有效性约束

表格数据的排序是对数据有效性进行筛选的一种常用方法。排序方式中有很多种,默认的即为升降序,或者按颜色,或者筛选数据内容和上下限等条件,从而获取想要的数据库。而某些数据需要选用特定的排列顺序,而且此类通用数据可以在相类似的数据中进行运用,这就可以使用自定义排序筛选功能。

图 5.1 为湖南省各地级市和自治州的社保人员数量统计表,数据分析人员想要得到按照编制排列的数据统计。自动的排序方式一般只能按照拼音顺序,因此并不能实现按照编制排序,还需要进行人工调整。因此,这里可以借用自定义的排序方式,将编制序列数据设置为固定模板。

行政地区	社保人数						
	2015年	2016年	2017年	2018年	2019年	2020年	2021年
安化县							
安仁县							
安乡县							
保靖县							
北湖区							
北塔区							
茶陵县							
常德市							
常宁市							
郴州市							
辰溪县							
城步苗族自治县							
慈利县							
大祥区							
道县							

图 5.1

从网络上下载湖南省的行政地级市和自治州名,已经按照固定的编制排位排序好的行政区域划分,如图 5.2 所示。

在要进行数据处理的表格中,单击“文件”菜单,选择“高级”命令,然后单击“编辑自定义列表”(图 5.3),打开该工具对话框。选择已下载好的行政区划名称,单击“导入”按钮,这样就将行政区划的固定排序以序列的方式采集进了模板中,如图 5.4 所示。

长沙市	
	岳麓区
	芙蓉区
	天心区
	开福区
	雨花区
	望城区
	浏阳市
	宁乡市
	长沙县
株洲市	
	天元区
	荷塘区
	芦淞区
	石峰区
	渌口区
	醴陵市
	攸县
	茶陵县
	炎陵县

图 5.2



图 5.3



图 5.4

打开要进行数据处理的表格,单击“筛选”,然后选择“按颜色排序”中的“自定义排序”,如图 5.5 所示。

打开“自定义排序”对话框,在“次序”框下选择“自定义”排序,这时会出现已经自定义排序的各类数据,选择前面定义好的行政区划次序,单击“确定”按钮(图 5.6)。此时就可以看到表格中的数据按照既定的排序规则已经重新排列了顺序(图 5.7)。

在数据审计工作中,数据的优先排序是获取关键数据的重要手段。在通过通用排序无法实现目的的情况下,可以将关键字段采用自定义形式固化成模板,以方便在执行类似工作时快捷操作。



图 5.5



图 5.6

行政地区	社保人数						
	2015年	2016年	2017年	2018年	2019年	2020年	2021年
长沙市							
岳麓区							
芙蓉区							
天心区							
开福区							
雨花区							
望城区							
浏阳市							
宁乡市							
长沙县							
株洲市							
天元区							
荷塘区							
芦淞区							
石峰区							
渌口区							
醴陵市							

图 5.7

## 5.2 条件格式筛选异常

现在有一张医院采购药品明细表,表中罗列了各种药品的进货价格和零售价格,如图 5.8 所示,现在需要对这样一张表格进行数据分析。

入库单号	药品代码	规格	单位	进货价	零售价	进销加成比例	入库日期
RMZ7279	1705024AT0	10g:0.1mg	支	126.27	127.00	0.6%	2021/1/9
RMZ7280	1504066TA0	10mg	片	37.58	38.00	1.1%	2021/1/9
RMZ7281	0105038SL0	0.5g	支	58.32	62.00	6.3%	2021/1/9
RMZ7282	3012104EY0	1t	张	70.49	90.00	27.7%	2021/1/9
RMZ7282	3012104EY0	1t	张	70.49	71.00	0.7%	2021/1/9
RXYS2991	0502012IJ0	10ml	支	0.87	1.00	14.9%	2021/1/14
RXYF1806	5104030ZY0	10g	包	1.32	1.32	0.0%	2021/2/3
RXYS3036	0401015TA0	1t	片	0.66	0.80	21.2%	2021/4/7
RXYS3037	0204017VR0	10g	支	25.41	26.00	2.3%	2021/4/7
RXS4309	0703005IJ0	500ml:35g	瓶	27.99	33.00	17.9%	2021/4/27
RXS4310	1207008IJ0	20mg	支	101.00	117.00	15.8%	2021/4/27
RXS4311	0602014IJ0	1支	支	36.96	42.50	15.0%	2021/4/27

图 5.8

### 5.2.1 突出显示单元格规则

“突出显示单元格规则”在本案例中主要用于突出显示销售价格高于进货价格 20% 以上的药品。其操作步骤如下。

第一步:选中要突出的数据区域。选中“采购药品明细表”中的药品“进销加成比例”区域,即 H 列单元格区域。

第二步:在“条件格式”功能中选择“突出显示单元格规则”中的“大于”,在“为大于以下值的单元格设置格式”中输入 20%，“设置为”中选择“浅红填充色深红色文本”,如图 5.9 和图 5.10 所示。

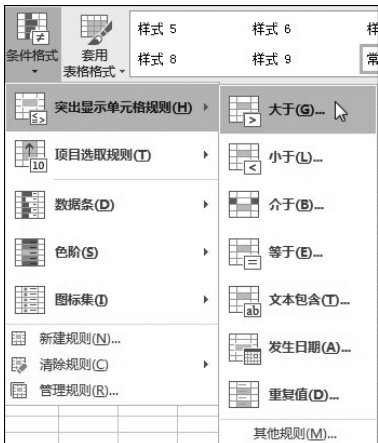


图 5.9



图 5.10

这时,进销价格超过 20% 的药品名就已经被标注出来了,如图 5.11 所示。

入库单号	药品代码	规格	单位	进货价	零售价	进销加成比例	入库日期
RMZ7279	1705024AT0	10g:0.1mg	支	126.27	127.00	0.6%	2021/1/9
RMZ7280	1504066TA0	10mg	片	37.58	38.00	1.1%	2021/1/9
RMZ7281	0105038SLO	0.5g	支	58.32	62.00	6.3%	2021/1/9
RMZ7282	3012104EY0	1t	张	70.49	90.00	27.7%	2021/1/9
RMZ7282	3012104EY0	1t	张	70.49	71.00	0.7%	2021/1/9
RXYS2991	0502012IJ0	10ml	支	0.87	1.00	14.9%	2021/1/14
RXYF1806	5104030ZY0	10g	包	1.32	1.32	0.0%	2021/2/3
RXYS3036	0401015TA0	1t	片	0.66	0.80	21.2%	2021/4/7
RXYS3037	0204017UR0	10g	支	25.41	26.00	2.3%	2021/4/7
RXS4309	0703005IJ0	500ml:35g	瓶	27.99	33.00	17.9%	2021/4/27
RXS4310	1207008IJ0	20mg	支	101.00	117.00	15.8%	2021/4/27
RXS4311	0602014IJ0	1支	支	36.96	42.50	15.0%	2021/4/27

图 5.11

## 5.2.2 项目选取规则

“项目选取规则”在本案例中主要用于突出显示药品“进销加成比例”排名前 5 的数据。其操作步骤如下。

第一步：清除之前设置的条件格式。选中上面已经设置条件格式的 H 单元格区域。在“条件格式”功能中选择“清除规则”中的“清除所选单元格的规则”，如图 5.12 所示，单击“确定”按钮后，刚设置的条件格式就被清除了。



图 5.12

第二步：选择数据区域。用鼠标选中“采购药品明细表”的“进销加成比例”区域，即 H 列单元格区域，同前面。

第三步：突出显示排名前 5 的单价。在“条件格式”功能中选择“项目选取规则”中的“前 10 项”，在“前 10 项”对话框中，将 10 改为 5，即仅突出显示前 5 个最大值项目，“设置为”中的填充颜色选择默认的即可，如图 5.13 和图 5.14 所示。

单击“确定”按钮，进销加成比例最大的前 5 种药品数据就被突出显示出来了，如图 5.15 所示。



图 5.13

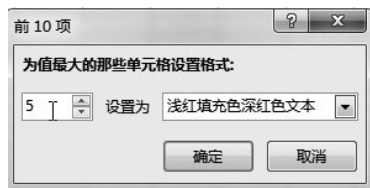


图 5.14

入库单号	药品代码	规格	单位	进货价	零售价	进销加成比例	入库日期
RMZ7279	1705024ATO	10g:0.1mg	支	126.27	127.00	0.6%	2021/1/9
RMZ7280	1504066TAO	10mg	片	37.58	38.00	1.1%	2021/1/9
RMZ7281	0105038SLD	0.5g	支	58.32	62.00	6.3%	2021/1/9
RMZ7282	3012104EYO	1t	张	70.49	90.00	27.7%	2021/1/9
RXYF1806	5119062ZY0	10g	包	1.43	1.50	4.9%	2021/2/3
RXYF1806	5120015ZY0	10g	包	2.16	2.50	15.7%	2021/2/3
RXYF1806	5110028ZY0	20g	包	1.02	1.10	7.8%	2021/2/3
RXYF1806	5108048ZY0	15g	包	0.60	0.80	33.3%	2021/2/3
RXYF1806	5108049ZY0	20g	包	0.74	1.00	35.1%	2021/2/3
RXYF1806	1802028LJD	250ml:2.25	袋	9.01	9.01	0.0%	2021/2/3
RXYF1806	0902003SLD	6wu	枚	43.70	44.00	0.7%	2021/2/3
RMZ7636	0503005LJD	0.2g	支	63.25	65.00	2.8%	2021/5/7
RMZ7637	1202011LJD	0.1g	支	3.68	4.00	8.7%	2021/5/7
RMZ7637	1205002LJD	10mg	支	0.54	1.00	85.2%	2021/5/7
RMZ7638	2108012LJD	18ug	支	212.49	215.00	1.2%	2021/5/7

图 5.15

**注意：**这里选择的突出颜色和查询数据个数，都可以根据需要进行相应选择。

## 5.2.3 数据条

“数据条”在本案例中主要用来查看各种药品的进销加成情况。其操作步骤如下。

第一步：清除之前设置的条件格式，同上面操作。

第二步：选择数据区域。用鼠标选中采购药品明细表的“进销加成比例”区域，即 H 列单元格区域。

第三步：将每种药品的加成比例用条形图展现。在“条件格式”功能中选择“数据条”，然后在“渐变填充”中选择“蓝色数据条”。

确认后，各种药品的加成比例已经用条形图标注出来了，加成多少一目了然，如图 5.16 和图 5.17 所示。



图 5.16

入库单号	药品代码	规格	单位	进货价	零售价	进销加成比例	入库日期
RMZ7279	1705024AT0	10g:0.1mg	支	126.27	127.00	0.6%	2021/1/9
RMZ7280	1504066TA0	10mg	片	37.58	38.00	1.1%	2021/1/9
RMZ7281	0105038SL0	0.5g	支	58.32	62.00	6.3%	2021/1/9
RMZ7282	3012104EY0	1t	张	70.49	90.00	27.7%	2021/1/9
RMZ7282	3012104EY0	1t	张	70.49	71.00	0.7%	2021/1/9
RKYS2991	0502012IJ0	10ml	支	0.87	1.00	14.9%	2021/1/14
RXYF1806	5104030ZY0	10g	包	1.32	1.32	0.0%	2021/2/3
RKYS3036	0401015TA0	1t	片	0.66	0.80	21.2%	2021/4/7
RKYS3037	0204017UR0	10g	支	25.41	26.00	2.3%	2021/4/7
RKS4309	0703005IJ0	500ml:35g	瓶	27.99	33.00	17.9%	2021/4/27
RKS4310	1207008IJ0	20mg	支	101.00	117.00	15.8%	2021/4/27
RKS4311	0602014IJ0	1支	支	36.96	42.50	15.0%	2021/4/27
RKS4312	0702003IJ0	10g	支	372.00	378.00	1.6%	2021/4/27
RKS4313	1407037CF0	10mg	粒	94.96	109.20	15.0%	2021/4/27
RKS4314	0606012IJ0	1000ml	袋	222.61	223.00	0.2%	2021/4/27
RKS4315	0102005IJ0	1g	瓶	90.10	105.00	16.5%	2021/4/27
RKS4316	1303020HQ0	120瓶	支	56.91	66.00	16.0%	2021/4/27

图 5.17

## 5.2.4 图标集

“图标集”在本案例中主要用于区分药品进销加成是否合理。

自 2017 年以来国家政策规定,医院禁止对药品进行加价销售。在此用“图标集”来进行数据分析。

第一步:清除之前设置的条件格式,同前面操作。

第二步:依旧选择数据区域 H 列。

第三步:将各种药品的加价比例用图标标注。在“条件格式”功能中选择“图标集”,然后选择“其他规则”,打开“新建格式规则”对话框,在“图标样式”中指定一个样式,这里选择“对、错、感叹号”图标集,如图 5.18 和图 5.19 所示。

第四步:设置图标显示数值区间。在“选择规则类型”中选定第一个“基于各自值设置所有单元格的格式”。在“图标”选项中,第一个选择符号“×”,值“>0”,类型选择“百分比”,



图 5.18



图 5.19

表示不符合政策的数据；第二个选择符号“√”，这里需要注意的是，选项中没有单独的“=0”，所以在这里选择一个交集，“ $\leq 0$  且  $\geq 0$ ”，其交集默认的值即为“=0”，表示符合政策不加价销售，如图 5.20 所示。



图 5.20

这里只需要保留两个图标,第三个选项这里不需要,所以直接选择“无单元格图表”。

单击“确定”按钮,到这里,药品是否进行加价销售已经被条件格式清晰地标注出来了。效果如图 5.21 所示。

入库单	药品代码	规格	单位	进货	零售	进销加成比例	入库日期
RXS4316	1303020HQ0	120喷	支	56.91	66.00	✖ 15.97%	2021/4/27
RZC4779	3011024LFO	8g	袋	40.00	42.00	✖ 5.00%	2021/4/28
RXS4319	2108019IJ0	15000u	瓶	1095.00	1170.00	✖ 6.85%	2021/4/28
RMZ7636	0503005IJ0	0.2g	支	63.25	65.00	✖ 2.77%	2021/5/7
RMZ7637	1202011IJ0	0.1g	支	3.68	4.00	✖ 8.70%	2021/5/7
RMZ7637	1205002IJ0	10mg	支	0.54	1.00	✖ 85.19%	2021/5/7
RMZ7638	2108012IJ0	18ug	支	212.49	215.00	✖ 1.18%	2021/5/7
RMZ7654	1407027TA0	10mg	片	254.00	254.00	✔ 0.00%	2021/5/15
RMZ7655	1903067NJ0	2.5g	支	18.97	20.00	✖ 5.43%	2021/5/15
RMZ7656	0702003IJ0	5g	支	222.00	222.00	✔ 0.00%	2021/5/15
RMZ7657	1299015IJ0	10ml:0.4g	支	100.05	102.00	✖ 1.95%	2021/5/15
RXYS3073	0103006IJ0	8wu	支	0.42	0.50	✖ 19.05%	2021/5/17
RXYS3073	0207005IJ0	10ml	支	54.42	55.00	✖ 1.07%	2021/5/17
RMZ7660	1802028IJ0	250ml:2.25	袋	9.01	9.01	✔ 0.00%	2021/5/18
RMZ7660	1702002IJ0	1.25wu	支	10.86	10.86	✔ 0.00%	2021/5/18
RMZ7660	1404006IJ0	10mg	支	0.35	0.35	✔ 0.00%	2021/5/18
RMZ7660	1303020HQ0	120喷	支	65.44	65.44	✔ 0.00%	2021/5/18
RMZ7660	0808044IJ0	300iu	支	177.99	177.99	✔ 0.00%	2021/5/18
RMZ7660	0808017IJ0	300iu	支	55.20	55.20	✔ 0.00%	2021/5/18

图 5.21

### 5.3 高级筛选识别异常

平时进行的筛选都是简单运用,其实使用的都是自动筛选功能,而且每次都只能对一列数据设置筛选条件,如果要设置多列数据的筛选条件,就要用到“高级筛选”。在“数据”菜单下有个“高级”选项,单击进去后,会出现如图 5.22 所示的界面。

从对话框中可以看到,“高级筛选”有多个必须要输入的内容,分别是:在原有区域显示筛选结果、将筛选结果复制到其他位置、列表区域、条件区域。

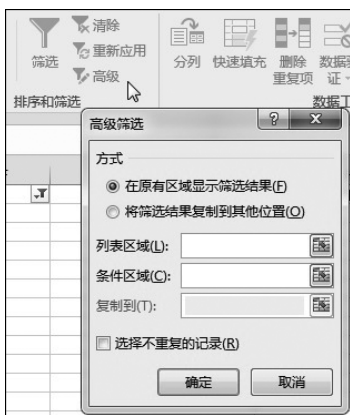


图 5.22

(1) 列表区域就是原始数据区域。

(2) 条件区域是指用什么条件来筛选,通常会在原始数据区域的下面或者右面空白处添加一个区域,专门放置条件区域,以免执行筛选后覆盖了原始数据。

(3) 条件区域的基本格式是“表头标题+条件”,“表头标题”必须与原始数据区域的表头标题内容一致,通常直接复制原始数据区域的标题。

(4) 筛选条件为两个以上时,条件区域就对应为两个以上。如果写在同一行,表示是同时满足的“与”关系。条件区域标题不一定要跟原始数据区域表头标题顺序一致,但是其内容必须完全一致。如果将几个筛选条件排列在同一列,则表示筛选条件为“或”关系,即满足其中一个条件即可。因此,一定要注意行与列的区别。

### 5.3.1 检测试剂过期/临期

近年来,新冠病毒的出现,让防疫物资在某段时间内非常稀缺,价格也是成倍上涨。经过国家和企业的努力,相关物资的生产保证了市场供应充足。但是在多次的防疫物资审计中,惊现众多机构囤积物资、高价转售、虚报数据获取国家补贴的情况,严重扰乱了社会秩序和民众利益。本节从被审计单位获取“在库防疫物资明细表”,如图 5.23 所示,通过分析其过期/临期物资的数量和金额,来审计其造成的经济利益损失情况,为其他事项审计提供事实依据。下面以“检测试剂”专题为例。

药品代码	品名	厂家	入库日期	有效期	包装规格	进货价格	库存数量
0101038TA0	酒精	广东万方	2021/12/17	2023/12/17	100g/瓶	5.4	600
0102036IF0	检测试剂	武汉明德	2019/12/26	2020/6/23	20人份/盒	163.6	210
0102055TA0	检测试剂	广州万孚	2021/9/25	2022/3/24	20人份/盒	196	120
0102055TA0	酒精	华北金坦	2021/10/20	2023/10/20	100g/瓶	4.9	2
0103014IT0	口罩	海南万隆	2021/5/18	2024/5/17	100个/盒	36	300
0105038SL0	棉签	河南立宁	2021/1/9	2024/1/9	100根/盒	3.8	11
0106002TA0	酒精	西安兆兴	2021/5/18	2023/5/18	100g/瓶	6	45
0109006LF0	手套	厦门星鲨	2020/5/18	2023/5/18	100双/盒	3.8	300
0109007IT0	口罩	广东万方	2019/5/18	2022/5/17	100个/盒	36	400
0202038IT0	口罩	四川万通太平洋	2021/5/20	2024/5/19	100个/盒	27	600
0204017US0	防护服	北大西南合	2021/4/7	2024/4/6	10套/包	250	1

图 5.23

利用“高级筛选”工具,设定两个筛查条件统计品名和有效期,品名选择“检测试剂”,有效期设置为“<2022/7/1”,两个条件需要同时满足,所以两个条件区域并列在一行中。

第一步:创建条件区域,即在药品入库明细表数据区域的上边空白区域设置两个条件区域“A1:B3”。复制表格中的表头标题“品名”和“有效期”至“A2:B2”单元格。接着在条件区域表头标题下分别输入条件值“检测试剂”和“<2022/7/1”,完成条件区域的设置,如