


表计算高级案例：MTD-ANR和YTD-ANR计算




By 喜乐君

Oct 3, 2022


说明：本案例涉及到两次表计算嵌套，还可以结合数据混合计算达成率，再使用if判断计算线性插值得分和排名

 完整案例，作为2023年图书案例之典型（高级计算和绩效分析章节）

基本的计算逻辑

1.  各机构、各月份的 ENR 余额
2.  计算各机构，各月份的 MTD-ANR 余额（BNR 和 ENR 的均值）
 - a. 使用一次表计算，lookup，上月的 ENR 就是当月的 BNR
3.  基于年初到现在各月的 MTD-ANR，计算 YTD-ANR 余额
 - a. 使用第二次表计算，WINDOW_AVG 计算，从年初到现在
4. 基于 YTD-ANR 目标值，计算 YTD-ANR 达成率
5. 基于 YTD-ANR 达成率，计算 YTD-ANR 得分
 - a. 计算线性插值

9.9 高级嵌套：不同计算依据的表计算嵌套

 补充进入《数据可视化分析》第二版第9章9.9，作为嵌套表计算案例。

快速表计算“YTD增长率”和帕累托的“累计贡献占比”都是表计算嵌套组合的典型示例，它们都可以通过编辑表计算、添加附加表计算快速完成。表计算嵌套的高级形式，是多个依据不同的表计算的自定义组合。这里以笔者项目中的一个案例为代表介绍：金融行业的ENR和ANR余额分析。

本案例也适用于包含期初、期末的财务期间分析。

1、业务背景和示例数据表

在金融机构发放贷款之后，要实时跟踪借款人的还款情况和质量。随时间变化的应收贷款余额，就是 ENR（Ending Net Receivable），金融机构会和某个期初比较贷款余额的变化，对应的期初贷款余额（通常是月初或者年初），就是 BNR（Beginning Net Receivable）。

为了平滑波动，常用平均贷款余额 ANR（Average Net Receivable）表示各月的贷款余额，并以此为基础计算各种比率，比如提前结清比率（EPO%ANR）、净损失比率（IIP%ANR）等。

ANR 也会随着时间范围不同而有所差异，比如 MTD-ANR 就是上月期末和本月期末的算术平均，而 YTD-ANR 则是年初至今各月 MTD-ANR 的算术平均。常见指标关系如表 9-3 所示。

表 9-3 不同统计日期的贷款余额指标计算及其关系

指标名称	2021/12/31	2022/1/31	2022/2/28	2022/3/31	……	
BNR	O	X	Y	Z	K	
ENR	X	Y	Z	K	……	
MTD-ANR	(O+Y)/2	(X+Y)/2	(Y+Z)/2	(Z+K)/2		
YTD-ENR	AVG (Σ [各月的MTD-ANR])					
EPO	E1	E2	E3	E4	……	
EPO%ANR	E1/ANR	E2/ANR	E3/ANR	E4/ANR	……	

数据表可以是基于借据号的明细表，也可以是聚合到机构等特定详细级别的聚合表，核心的数据表结果是【统计日期】【贷款余额】字段，它们可以反映贷款余额的历史变化，也是BNR、ANR的计算来源。基于借据号的明细表，如表9-4所示。

表9-4 不同统计日期的借据号余额明细

分支机构	统计日期	贷款人ID	借据编号	贷款日期	贷款余额
XX支行	2021/12/31	……	……	……	80
XX支行	2022/1/31	……	……	……	100
XX支行	2022/2/28	……	……	……	150
XX支行	2022/3/31	……	……	……	160
XX支行	2022/4/30	……	……	……	100
XX支行	2022/5/31	……	……	……	150
XX支行	2022/6/30	……	……	……	160
YY支行	2022/1/31	……	……	……	300
YY支行	2022/2/28	……	……	……	350
YY支行	2022/3/31	……	……	……	320

接下来，笔者介绍两种计算ANR的方法，并阐述其优劣。

2、基于已有字段聚合MTD-ANR数据值

很多人计算ANR，特别是计算单月的ANR，习惯使用SUMIF的方法，把日期范围和聚合值合并在一起，于是就有了如下的判断样式：

```
- [2022年5月ENR]: SUM( IF [统计日期] = #2022-05-31# THEN [贷款余额] END) )
```

```
- [2022年6月ENR]: SUM( IF [统计日期] = #2022-06-30# THEN [贷款余额] END) )
```

之后，使用算术计算，获得2022年6月ANR值：

```
- 2022年6月ANR: ([2022年5月ENR] + [2022年6月ENR]) / 2
```

并用类似的方法，以此为创建多个计算列，分别获得多个包含日期范围的字段【2022年1月ANR】、【2022年2月ANR】、【2022年2月ANR】、【2022年3月ANR】……

在这样的“习惯”之下，分析师甚至于预先写好全年的指标，然后在分别创建各月的YTD- ANR值。这种方法适合于计算单月的指标，但在构建ANR趋势时，局限性就会暴露无疑；在大数据分析中，这种方法更大的问题在于，严重的拖累数据库查询的性能——因为上述方法是借助于IF行级别判断间接完成筛选，大量的行级别计算，正是数据库查询的陷阱。

推荐的方法是，充分利用聚合的结果，再把日期参与其中，通过聚合的二次计算的方式，既无需创建每个月的计算结果字段，又避免了大量的行级别计算拖累数据库计算性能。

如图9-81所示，以“各统计日期（年月）的贷款余额总和”为问题构建交叉表（由于数据表的统计日期为各月月末日期，因此这里统计日期精确值代表各月末），通过LOOKUP函数即可获得上一期的贷款余额总和，上一期的期末正是当前期间的期初。



图9 81 使用LOOKUP函数计算期初值和平均贷款余额

基于当前的贷款余额总和、上一期的贷款余额总和，就可以创建平均贷款余额。

相比之前SUMIF的方法，这个方法特别适合于构建趋势分析。

3、嵌套表计算，计算YTD-ANR

这里的关键是，在上述表计算的基础上，进一步计算YTD-ANR。

YTD- ANR计算年初当前月份的累积ANR平均，对于当年中的每个月份，计算的起点相同、终点不同，这正是9.5.3小节WINDOW_AVG移动聚合的应用场景。

计算的难点在于，YTD-ANR需要嵌套之前的MTD-ANR表计算，但是前者以年度为计算范围、月份为计算依据，后者则要以年度、月份为计算依据（1月份的期初正是上年12月的期末值）。两个计算依据不同的表计算组合，就要像快速表计算一样分别指定。

如图9-82所示，首先把此前的MTD-ANR计算拖入左侧保存为已有字段，重命名为“MTD-ANR”，之后使用WINDOW_AVG计算嵌套MTD-ANR计算，由于YTD-ANR是从年初开始，也就是区域内第一个值，可以使用FIRST参数指定完成：

WINDOW_AVG([MTD-ANR], FIRST(), 0)



图9 82 创建MTD-ANR计算，并嵌套构建YTD-ANR计算

难点在于，MTD-ANR和YTD-ANR的表计算依据不同。这种情况下的嵌套，必须在字段设置中预先设置，而且只能使用明确指定字段，类似于SQL中的OVER语法。

如图9-83所示，点击嵌套的表计算，选择“编辑表计算...”，可以分别选择被嵌套的表计算以及完整的表计算，分别设置计算的依据。这样，才能获得完整具有业务意义的计算值。

图9 83 分别编辑“嵌套表计算”的计算依据

编辑嵌套表计算的前提，是把被嵌套的表计算保存为独立的字段引用。

当然，相当于之前的SUMIF方法，表计算性能更好，但在展现单一月份的计算结果时，往往难以控制，这也是业务分析师在使用表计算时常见的困扰。

后续补充：ANR的目标值、考核得分计算（使用线性插值）

1. ANR分月目标值

机构名称	日期	余额
山东	2022-1-31	300
山东	2022-2-28	400
山东	2022-3-31	450
江苏	2022-1-31	
江苏	2022-2-28	
江苏	2022-3-31	