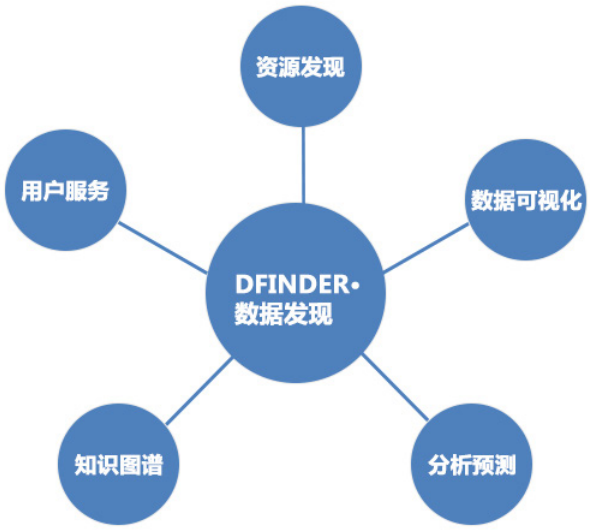
使用指南

2016 年 3 月 28 日

目录

DFINDER• 数据发现

DFINDER• 数据发现平台是由北京福卡斯特信息技术有限公司自主研发的创新型数据服务平台。它利用搜索引擎技术处理海量数据资源的精准定位，实现了时间序列、电子表格、文献资源资源的一站式发现。系统包含海量的数值型数据资源，涉及经济、贸易、金融、财政等数十个领域，覆盖全球 200 多个国家和国际组织、全国 31 个省自治区直辖市、400 多个地级市、2000 多个县级市与港澳台；收录了超过 700 万个统计指标的时间序列数据、超过 30 万张的统计表格数据、100 余万篇最新社科类核心期刊文献。
系统包含资源发现、数据可视化、分析预测、知识图谱、用户服务五大模块：



（1）资源发现是基于内部数据使用全文检索技术（ES）的搜索系统，可在高并发请求时毫秒内检索出千万级别的数据；
（2）数据可视化基于数据立方体和多维关联规则，结合最先进的前端展示技术，为用户提供表格、图表、地图、图谱四大类可视化选择；
（3）分析预测是基于 R 计算引擎为用户提供的在线数据分析平台，包含 8 大类、40 余种分析功能，提供图、表、序列等多样化结果输出形式；
（4）知识图谱是基于文献计量关联分析理论开发的文献挖掘功能，支持对大量相关文献的多角度多维度分析，支持用户生成数据；
（5）系统为用户提供了上传、下载、分享、历史查看等人性化功能，打通了用户、平台、资源之间的通道，让一切数据为用户所用。

系统基于 SaaS 软件即服务理念，无需在本地安装部署服务器，无需馆员维护；系统为所有用户提供高质量、高效率、低成本的数据资源发现、可视化、分析挖掘服务，给予一个简单易用、功能强大且容易定制化的整合平台，提供研究所需的从检索数据、处理分析数据到得出结论的一站式服务，为科学研究与论文撰写提供丰富的数据资源保障和强大的处理分析工具支持。
DFINDER• 数据发现，人人都是数据专家！

系统要求

为了有效地使用 DFINDER• 数据发现的所有功能，最基本的浏览器要求为 Internet Explorer 9.0（及以上版本），360 浏览器，谷歌浏览器，火狐浏览器等主流浏览器。

检索 资源发现

基本检索

1、在主页 **[基本检索界面]** 上的 **[检索]** 栏中，输入您的检索词：

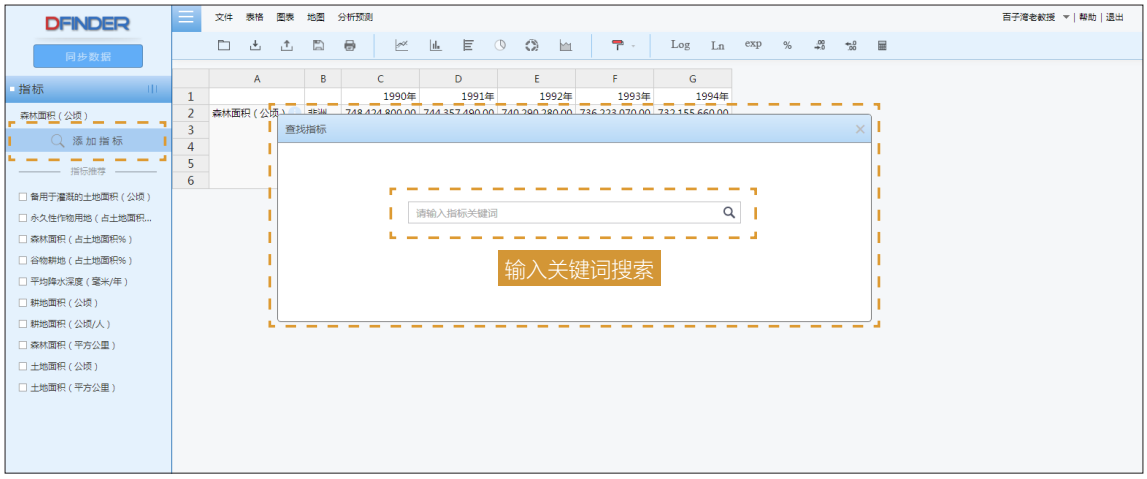
2、点击 **[检索]** 按钮，系统会跳转到检索结果页。

高级检索

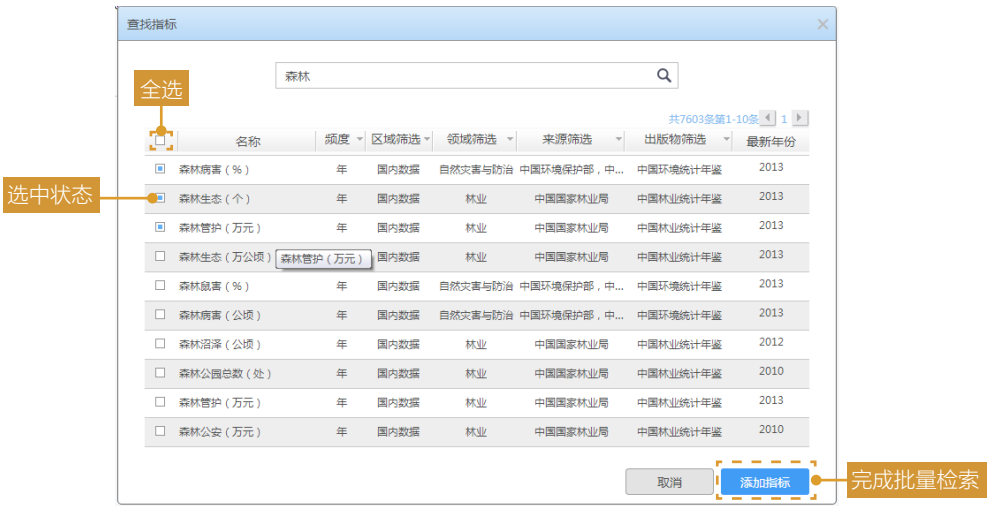
在文献资源栏目下，点击 [搜索] 左侧的向下箭头符号，选择需要的检索字段。

批量检索

1、在分析平台内，点击左侧边栏 **[添加指标]**，系统弹出 **[查找指标]** 弹出框。



2、输入搜索关键词，点击【**搜索**】按钮，系统展示搜索结果简明列表。



3、点击左侧白色小方块 □ ，呈现选中颜色 ▣ ，或点击最上部的白色小方块实现本页全选。

4、点击弹出框右下角的 **[添加指标]** ，完成批量检索。

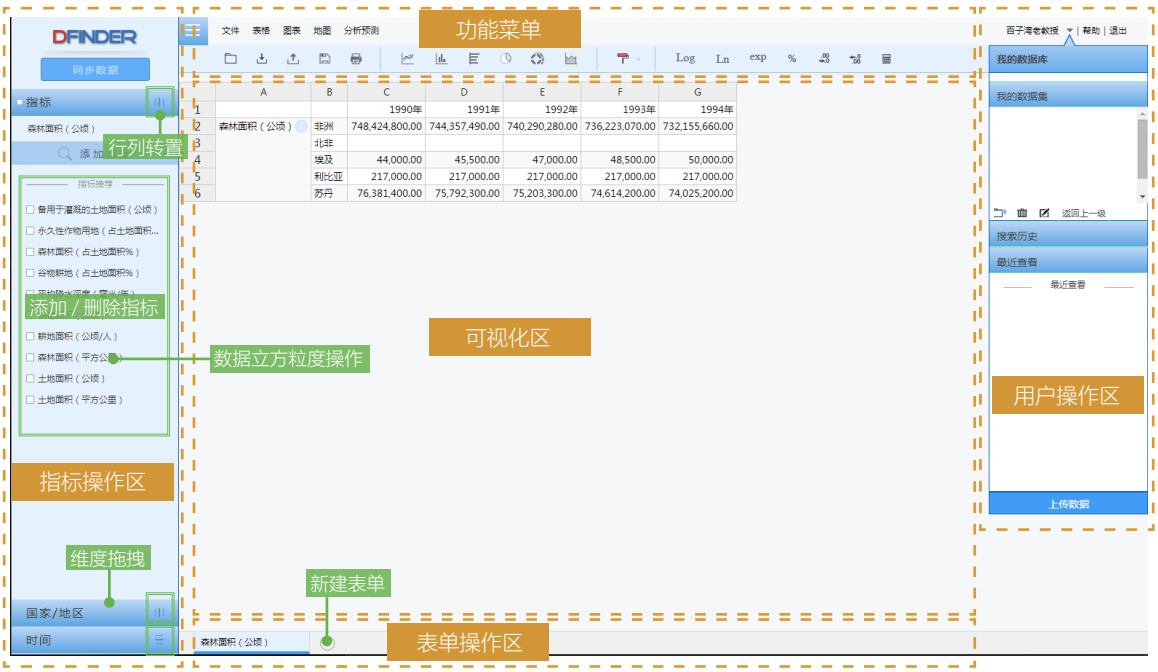
数据可视化

系统提供多样化的可视化展示工具，分别包括智能表格、统计图表、行政地图、知识图谱四大模块，其中智能表格是统计图表和行政地图的基础；此外，系统还提供电子表格的在线重现。

智能表格

1、用户在登录的情况下，点击搜索结果的某条具体指标如图所示。

2、进入智能表格界面。智能表格界面分为五大区块：指标操作、功能菜单、用户操作、可视化区、表单操作。



2.1 指标操作

指标操作区域可以完成多类数据立方体操作功能：添加 / 删除指标、行列转置、维度拖拽、数据立方粒度操作；

2.2 功能菜单

功能菜单区包含多类功能：文件、表格、图表、地图、分析预测、以及快捷工具栏。

2.3 用户操作

用户操作包括：我的数据库、搜索历史、最近查看、上传数据、用户退出。

2.4. 可视化区

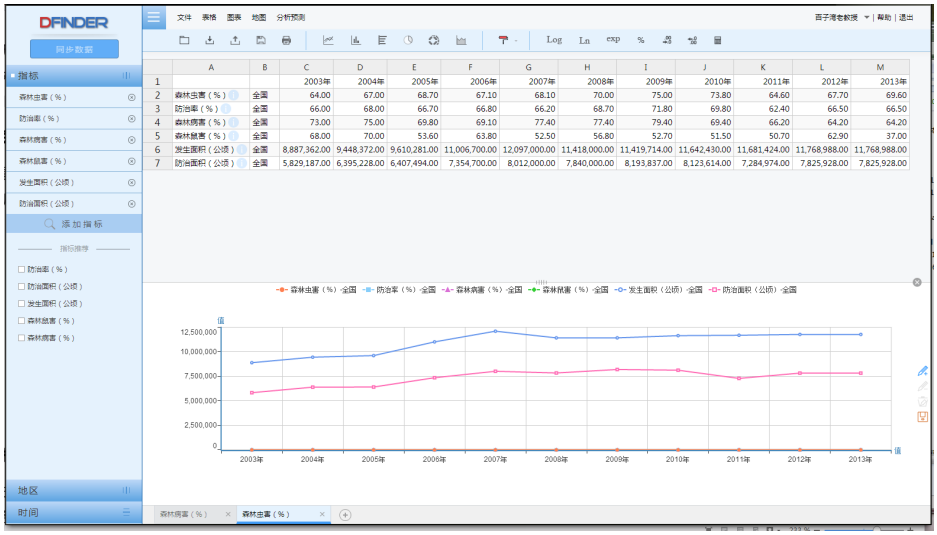
可视化区展示数据指标及相关信息：表格、图表、地图、information。

2.5. 表单操作

表单操作主要有新建表单以及关闭表单。

统计图表

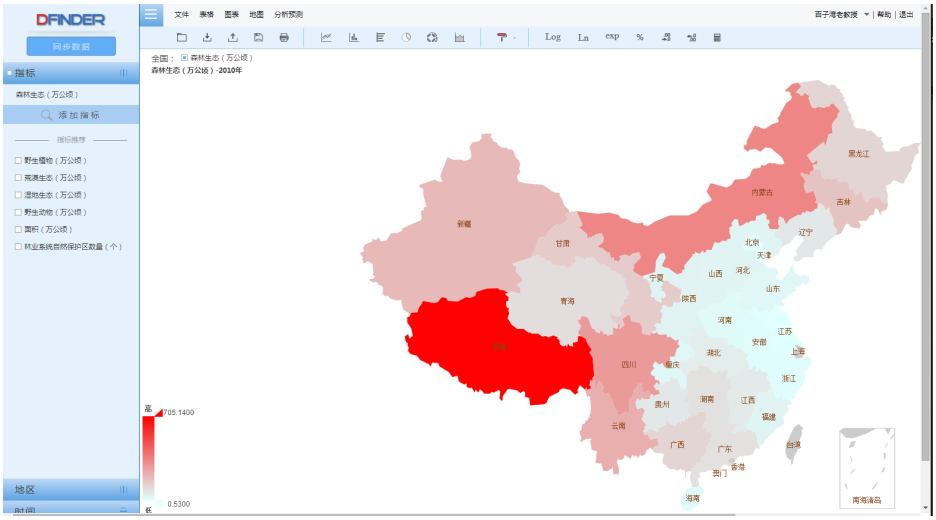
1、点击 **[图表]**，在下拉菜单中点击 **[折线图]**，出现如下所示图表：



统计图表是 DFINDER• 数据发现对数据的另一类可视化展现方式，共有折线图、面积图、柱形图、条形图、饼图、环形图、雷达图、散点图等 13 种图表。系统的图与表格具备联动展示特效，能够展示用户选中的任意数据区域。

行政地图

点击 **[地图]**，在下拉栏中点击 **[绘制地图]**，系统依据当前表单数据绘制地图。

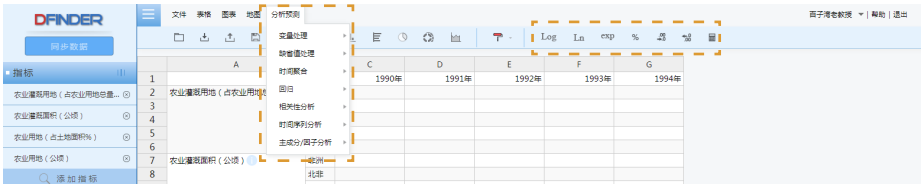


知识图谱

正在开发，敬请期待！

分析预测

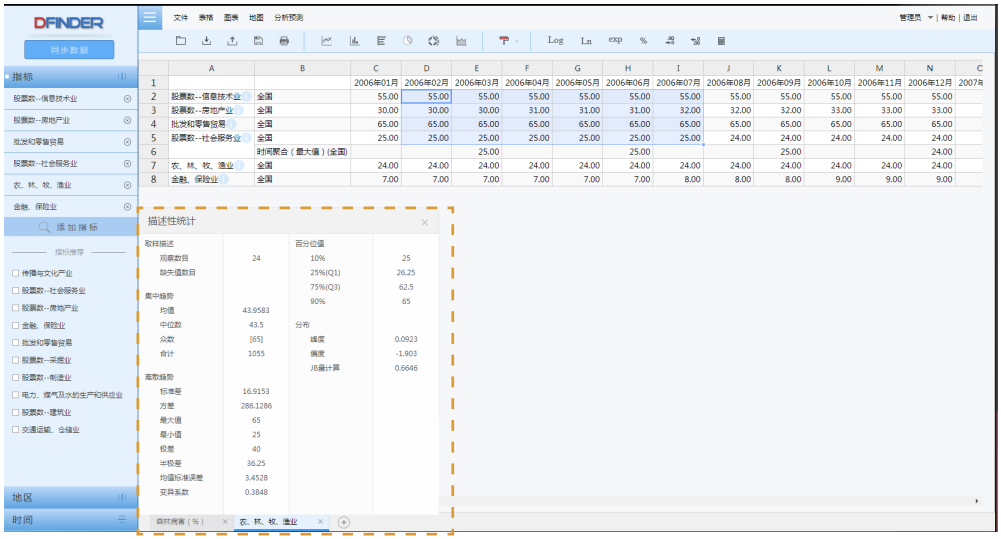
数据的价值在其背后隐藏的规律，DFINDER 为用户提供了功能强大的分析预测模型库。包括描述性统计、变量处理、缺省值处理、时间聚合、回归分析、相关分析、时间序列分析、主成分 / 因子分析、高级计算器等 9 大模块，共计 40 余类常见分析模型，涉及时间序列数据、截面数据及面板数据等多种数据模型的计算处理。



描述性统计

描述性统计分析，就是对一组数据的各种特征进行分析，以便于描述测量样本的各种特征及其所代表的总体的特征。

在表格中按住左键，拖动鼠标，选中一部分数据，单击右键，在弹出菜单中选中 **[描述性统计]**，系统得出被选中数据的描述性统计：

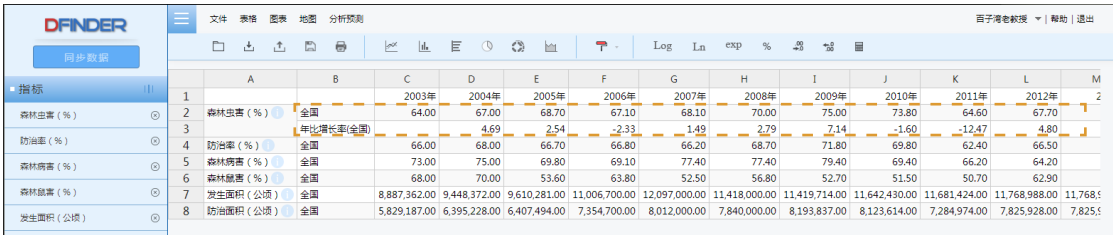


变量处理

变量处理模块主要用于对时间序列变量进行初步简单处理，主要包含了增长率、差分、对数转换、滞后函数、先行函数、发展速度、累积方法、标准化等。

选中需要处理的时间序列，鼠标左击该序列上的任一位置即代表选中该序列。

点击 **[增长率] → [年比增长]**，系统自动将结果放置在被选中的序列下方。



缺省值处理

缺省值处理模块用于对时间序列变量存在的缺省值进行相关处理，主要包含了删除、序列均值、三次样条内插、几何插补、线性插值、线性趋势、前一个值、后一个值、前 N 个增长率、后 N 个增长率、相邻 N 点均值、相邻 N 点中位数、随机值。

点击 **[缺省值处理] → [序列均值]**，系统自动将结果填充到缺省位置。

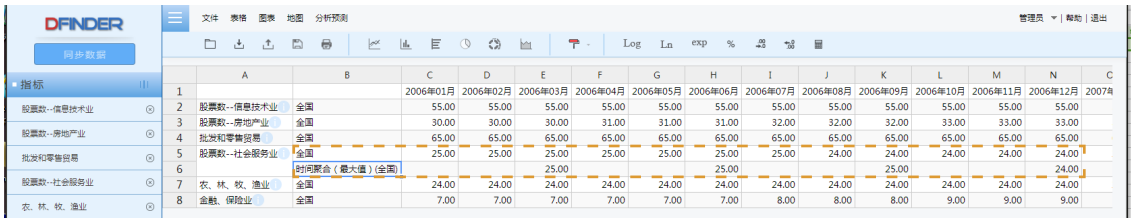
时间聚合

时间聚合的功能主要用于考察时间序列的周期特性，用于寻找某个周期（如年度、半年度、季度）内该序列的最大值、最小值、首值、末值、平均值、标注差、和值。

第一步，选择想要处理的时间序列；

第二步，选中想要应用的处理方法，如最大值聚合；

第三步，选择想要聚合的目标频度，如季度。



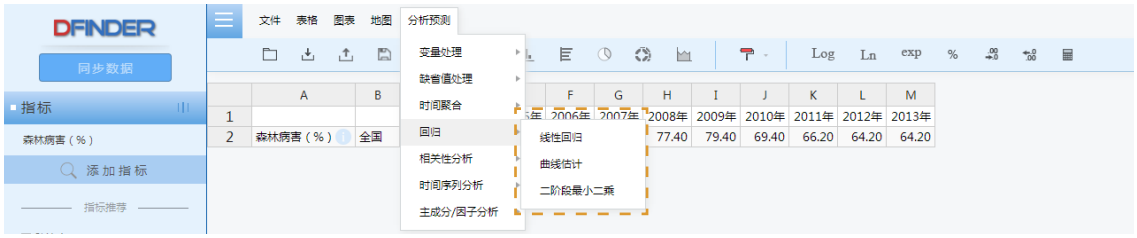
结果将会以断续时间序列的形式展示在数据页上。

回归分析

回归分析（regression analysis）是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。

结合平台特性，DFINDER 当前提供三大类回归分析方法，分别是线性回归、曲线估计、二阶段最小二乘回归，其中线性回归及曲线估计又包含多种方法级多种模型。

DFINDER 数据发现平台允许线性回归应用截面数据模型与时间序列数据模型。



基本操作步骤为：

第一步，搜索数据，准备要分析的因变量与自变量。

第二步，选择方法，确定使用线性回归、曲线估计还是二阶段最小二乘法。

第三步，选择参数，确定因变量与自变量（必选），其他可选参数（可选）。

第四步，点击确认运行，在当前 sheet 表单和弹出页得到结果。

1、线性回归

统计学中，线性回归 (Linear Regression) 是利用称为线性回归方程的最小平方函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析。这种函数是一个或多个称为回归系数的模型参数的线性组合。只有一个自变量的情况称为简单回归，大于一个自变量情况的叫做多元回归。

2、曲线估计

曲线估计是一类特殊一元回归模型。平台提供了指数分布、增长曲线、对数曲线、线性、立方、复合曲线、幂函数、逆模型、S 函数、二次项曲线共十种曲线估计模型。

3、二阶段最小二乘回归

二阶段最小二乘法（two stage least square，TSLS）本质上属于工具变量法，它包括两个阶段：

第一阶段找一组变量（称为工具变量），模型中每个解释变量分别关于这组变量作最小二乘回归；

第二阶段，所有变量用第一阶段回归得到的拟合值来代替，对原方程进行回归，这样求得的回归系数就是 TSLS 估计值。

相关分析

相关分析（correlation analysis），相关分析是研究现象之间是否存在某种依存关系，并对具体有依存关系的现象探讨其相关方向以及相关程度，是研究随机变量之间的相关关系的一种统计方法。

当前 DFINDER 平台提供两类相关分析，一类为线性相关分析（后文 4.7.2 的相关分析），另一类为偏相关分析。

1、双变量相关分析

平台提供多变量的双变量相关分析，可求 pearson、Kendall、spearman 三种类型的相关系数，可对相关分析进行双侧或单侧检验，分析结果通过表格的形式展现在弹出页面上，可对时间序列模型及截面数据模型进行相关分析。

2、偏相关分析

偏相关分析也称净相关分析，它在控制其他变量的线性影响的条件下分析两变量间的线性相关性，所采用的工具是偏相关系数（净相关系数）。

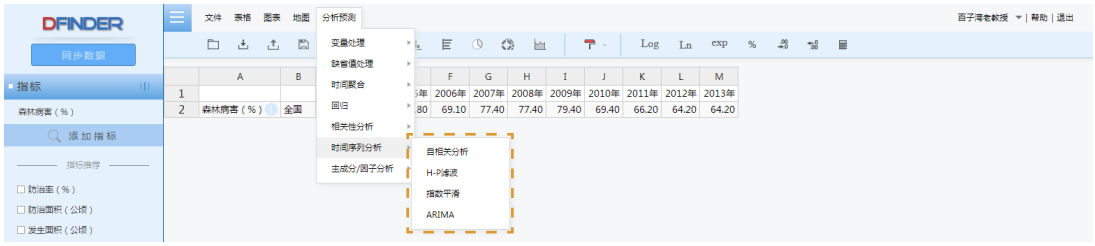
平台分析偏相关的时候，能够选择多个控制变量。分析结果以弹出页的形式展现，分为变量说明、描述性统计、偏相关系数及零阶相关系数等部分。

时间序列

时间序列是按时间顺序的一组数字序列。

时间序列分析是根据系统观测得到的时间序列数据，通过曲线拟合和参数估计来建立数学模型的理论和方法。

DFINDER• 数据发现平台提供四类时间序列分析方法：自相关及偏自相关分析、H-P 滤波、指数平滑以及 ARIMA 模型。基本操作步骤可以参考回归分析。



1、自相关及偏自相关分析

自相关及偏自相关分析是指利用自相关函数和偏自相关函数是分析随机过程和识别模型。

2、H-P 滤波

在宏观经济学中，人们非常关心序列组成成分中的长期趋势，Hodrick-Prescott 滤波方法（简称 H-P 滤波）是被广泛使用的一种。该方法在分析战后美国经济周期的论文中首次使用，DFINDER• 数据发现为用户提供了 HP 滤波分析方法。

3、指数平滑

有些经济时间序列数据（如股票数据）不具有明显的季节波动和趋势变动。对于这样的单指标时间序列数据，一般采用指数平滑方法进行拟合及预测。DFINDER• 数据发现为用户提供了五类指数平滑方法：单指数平滑、双指数平滑、H-W 加法模型、H-W 乘法模型、H-W 无季节模型。

4、ARIMA

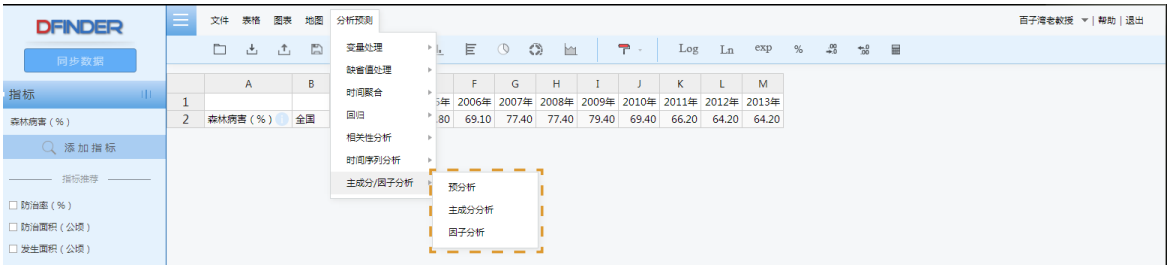
ARIMA 全称为自回归积分滑动平均模型 (Autoregressive Integrated Moving Average Model, 简记 ARIMA)，是由博克思 (Box) 和詹金斯 (Jenkins) 于 70 年代初提出一著名时间序列预测方法，所以又称为 box-jenkins 模型、博克思 - 詹金斯法。

ARIMA 模型的基本思想是：将预测对象随时间推移而形成的数据序列视为一个随机序列，用一定的数学模型来近似描述这个序列。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。现代统计方法、计量经济模型在某种程度上已经能够帮助企业对未来进行预测。

DFINDER 平台 ARIMA 模块还提供了季节 ARIMA 模型的接口。

主成分 / 因子分析

在信息爆炸的时代，在分析相关问题时，有效信息的提取非常重要。主成分与因子分析一致，都是用来从大量信息中提取有效信息的方法。在数据信息提取模型上，DFINDER 平台提供了主成分分析、因子分析两种方法。



1、预分析

DFINDER 提供的预分析模块能够对被用来进行主成分 / 因子分析的数据做一个预分析。预分析中提供了平行检验、相关系数矩阵等计算信息，能够对后续分析方法具体方法的选择及相应参数设置提供一定的参考。

2、主成分分析

主成分分析（Principal Component Analysis，PCA），将多个变量通过线性变换以选出较少个数重要变量的一种多元统计分析方法，又称主分量分析。

3、因子分析

因子分析是指研究从变量群中提取共性因子的统计技术。因子分析可在许多变量中找出隐藏的具有代表性的因子。将相同本质的变量归入一个因子，可减少变量的数目，还可检验变量间关系的假设。

高级计算器

DFINDER• 数据发现考虑到很多用户需要对平台内变量进行多样化的操作，普通的变量转换无法完成这类需求，高级计算器就是为了满足变量的多元化计算需求而创造的。DFINDER 高级计算器以平台内的变量为计算单元，提供加减乘除、三角函数、对数转换、求余、求绝对值、开方等多样化的计算方法。

（1）输入计算表达式；

（2）输入目标变量，点击等号"="即可。

文献计量分析

敬请期待！

我的账户

DFINDER• 数据发现为用户提供更多更加人性化的功能操作，如保存数据、上传本地数据、下载数据、分享数据等。

上传 / 下载

目前仅支持单地区时间序列上传。

将会陆续支持多地区截面数据上传，面板综列数据上传。

保存

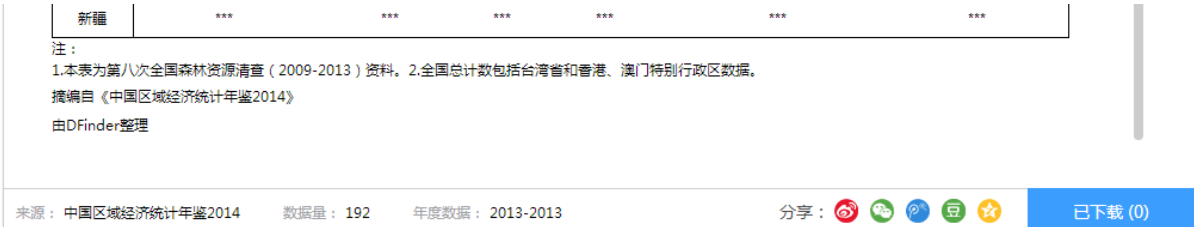
用户保存在账户内的数据集会被存储"我的收藏夹"内。

我的收藏夹

用户可以分享数据集、分析预测结果到多个互联网平台上。

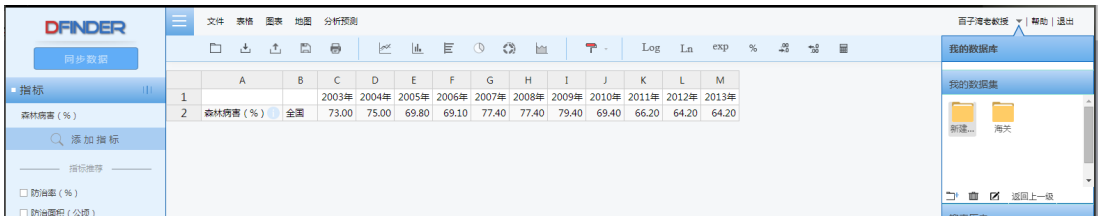
分享

用户可以分享数据集、分析预测结果到多个互联网平台上。



我的数据集

用户可以分享数据集、分析预测结果到多个互联网平台上。



意见反馈

感谢您使用 DFINDER• 数据发现！

如果您发现文档中有错误之处、产品运行不正确，或者您对该文档有任何疑问和建议，请与我们联系。您的意见是我们不断改进的动力，感谢您的支持，联系方法为：

邮件：service@dfinder.cn

电话：010-85786021

传真：010-85786020

地址：北京市海淀区知春路 9 号坤讯大厦 15 层，1502 室

邮编：100191