# 第 1 篇
# EViews 数据分析基础

EViews 是最常用的数据分析和预测软件之一，在经济学和统计学领域及金融投资、政府部门宏观经济决策、工业制造等方面都有广泛应用。EViews 通过建立工作文件、对象、单个序列和序列组等，对数据进行图形化、描述性统计量和假设检验等基本分析。这部分内容是 EViews 软件的基本操作部分，也是进一步通过回归分析和时间序列分析等方法，建立模型、分析模型和预测的基础。

# 第 1 章　EViews 概述

EViews 是 Econometrics Views 的缩写，最早由 Quantitative Micro Software（QMS）开发，目前，EViews 归属于 IHS Markit 公司。IHS Markit 公司的总部位于伦敦，在中国设有办事机构，该公司的中文名称为埃信华迈，是世界领先的数据信息服务及解决方案提供商。2022 年初，标普全球（S&P Global）完成与 IHS Markit 的合并，合并后的公司市值超过 1000 亿美元，成为全球最大的金融信息和统计数据服务商之一。

EViews 最初由经济学家开发，主要应用于计量经济学的研究。经过多年的发展，目前，EViews 在计量经济学、统计学、经济和金融领域已成为一款重要的建模和预测软件包，在时间序列分析等领域也有广泛的应用。

EViews 的客户主要是政府部门和高等院校从事相关研究的企业用户和师生：目前超过 1600 所大学的经济系和商业系使用该软件，在每年的 U.S. News 世界大学排名中，有 78% 的高校在教学和研究中使用 EViews。国际货币基金组织、联合国和世界银行等机构也将 EViews 作为宏观经济预测的重要工具。EViews 同时广泛应用于能源、汽车制造、电信、航空、投资银行、零售和医药等行业。

EViews 的界面友好，操作简便，容易入门，通过菜单操作基本就能实现大部分的任务，可以满足计量分析的主要需求。EViews 软件同时支持菜单型操作和命令代码操作，适合计量和统计分析的初学者使用；对于中高级使用者，可以通过代码编程完成各种分析任务。EViews 拥有非常全面的回归分析和时间序列分析工具，因此，EViews 也是“计量经济学”和“时间序列分析”课程常用的配套软件。

## 1.1　EViews 基础

### 1.1.1　EViews 的版本和安装

EViews 软件有标准版、企业版、高校版和个人版等多种版本，它是一个英文软件，没有汉化版。2022 年，EViews 推出了最新的 EViews 13，该版本目前只有标准版（Standard Edition）和企业版（Enterprise Edition）两种，同时提供了针对学术机构用户使用的许可证。目前仍在使用的 EViews 12 包括高校版（University Edition）和学生版（Student Version Lite）。其中，高校版的价格多年未调整，半年的使用费用约为 50 美元。

安装软件时，依据提示在安装过程中输入序列号即可，图 1-1 为 EViews 13 的安装界面。EViews 13 做了以下调整和升级：

- 工作文件窗口有较大的调整，更适合越来越多的小型终端屏幕，同时增加了代码调

试功能。

- 在数据处理上增加了以天为单位的季节调整模型，并增加了多个机构的数据库接口。
- 在图形处理方面进行了部分功能升级。
- 在计量和统计模型方面增加了非线性 ARDL 模型估计，对 PMG 估计和 VEC 估计等进行了升级。
- 在检验和模型诊断方面增加了 ARDL 模型、面板 ARDL 模型的诊断，对协整检验、脉冲响应的计算和显示方式进行了升级。

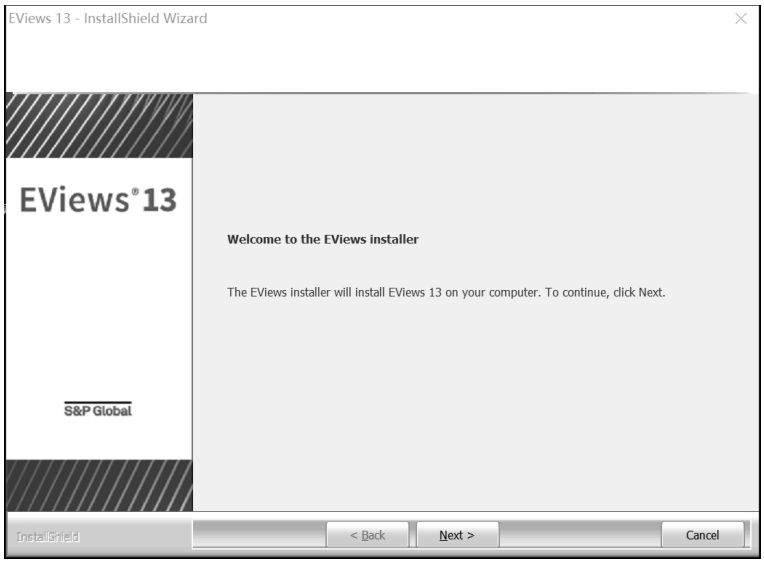



图 1-1　EViews 13 的安装界面

EViews 软件在高校中被广泛应用，它提供了免费的学生版供高校教师和学生使用，学生版软件可以运行和操作，但不能保存处理好的文件。目前，学生版的最新版本为 EViews 12 SV，下载地址为 http://www.eviews.com/download/student12/，有 Windows（只提供 64 位的操作系统下载文件）和 macOS 两个下载版本。在安装 EViews 前，需要在网上注册简单的个人信息，地址为 http://register1.eviews.com/Lite/。注册后，在几分钟内，邮箱会收到注册序列号。

EViews 13 和 EViews 12 的使用界面基本相同，考虑到目前国内主要使用 EViews 12 及以下的版本的软件，因此本书的案例和课后上机练习题大部分使用 EViews 12 在 Windows 系统上操作，另一部分使用 EViews 13 操作。

## 1.1.2　EViews 的启动与退出

EViews 软件全面支持 Windows 和 macOS 操作系统，在 Windows 操作系统下，启动和退出方式与 Windows 操作系统下的其他软件是一样的。

### 1．EViews 12的启动

启动 EViews 12 有两种方法，一种是双击桌面的 EViews 图标，直接运行 EViews 12 软

件；另一种是单击 Windows 的“开始”菜单，在 EViews 12 目录下运行“EViews 12(×64)”命令，如图 1-2 所示。EViews 的启动界面如图 1-3 所示，EViews 12 的用户界面仍然是传统的界面。

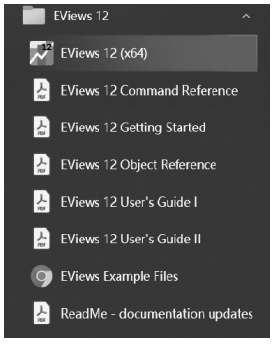



图 1-2　EViews 12 的启动菜单

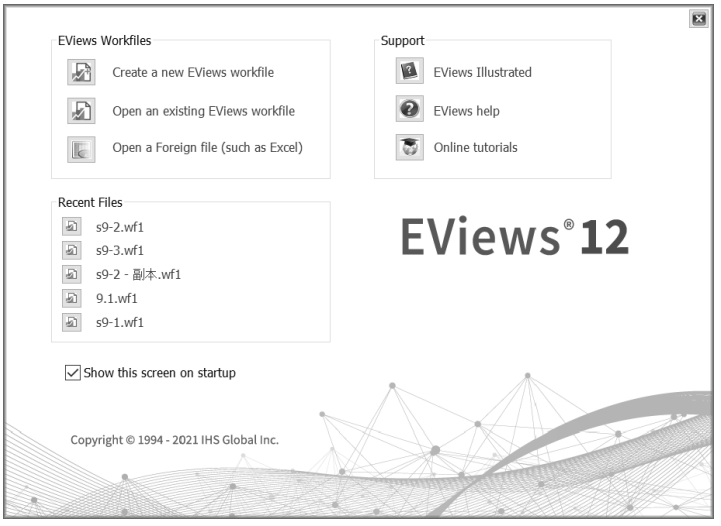



图 1-3　EViews 12 的启动界面

### 2．EViews 12的退出

在 EViews 12 的主菜单中依次选择 File | Exit 命令，或者单击菜单栏右上角的关闭按钮，可退出 EViews 12。

### 3．EViews 12的用户指南（使用手册）

进入 Windows 的“开始”菜单，在 EViews 12 目录下，可以直接运行 EViews 12 的用户指南（用户手册）：EViews User’s Guide Ⅰ和 EViews User’s Guide Ⅱ。EViews 的每个版本都有对应的操作指南，在学习和使用 EViews 的过程中如果有疑问，以操作指南为准。

## 1.1.3　EViews 的主窗口

软件启动后，进入 EViews 12 主窗口，如图 1-4 所示，下面具体介绍。

### 1．标题栏

标题栏（Title Bar）位于主窗口的最上方，当 EViews 12 在 Windows 中被激活时，标题栏会显现出与其他应用程序不同的深色色调，可以按住 Alt+Tab 组合键在不同的应用程序之间进行切换。

### 2．主菜单

EViews 12 的主菜单（Main Menu）包括 10 个选项，分别是 File（文件）、Edit（编辑）、Object（对象）、View（浏览）、Proc（处理或加工）、Quick（快速分析）、Options（参数设

定选项）、Add-ins（加载工具包）、Window（窗口）和 Help（帮助）。

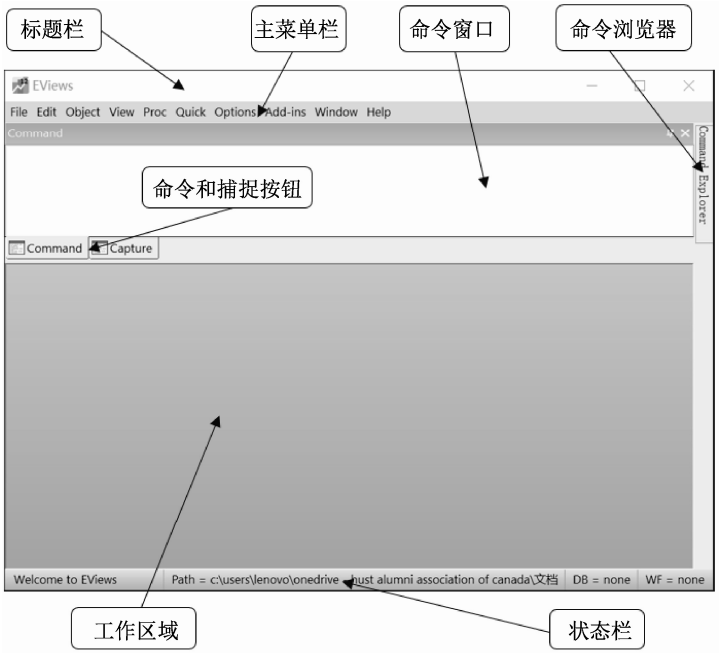



图 1-4　EViews 12 的主窗口

### 3．命令窗口

用户在 Command（命令）窗口中输入命令，完成操作过程。例如，在命令窗口中输入 e=nrnd，表示生成一个名称为 e 的白噪声序列。输入命令后，按 Enter 键执行命令。

### 4．命令和捕捉按钮

Command 窗口下方有 Command（命令）和 Capture（捕捉）两个按钮，二者可以相互切换。Capture 是捕捉按钮，非常实用。单击 Capture，捕捉功能将被激活，命令窗口切换为捕捉窗口（窗口上方出现 Capture 字样），也可以在主菜单中选择 Window | Display Command Capture Window 命令激活 Capture 功能。Capture 功能激活后，用户进行的每个操作都会以命令形式在捕捉窗口中显示，这可以帮助用户熟悉 EViews 各项命令的操作。Capture 捕捉到的命令，可以直接复制到 Command 状态下的命令窗口内。

### 5．状态栏

状态栏（Status Bar）位于主窗口最下端，主要包括三个部分：Path 是系统设定的 EViews 文件保存目录（Default Directory）；DB 是系统设定的数据库（Default Database）；WF 是当前活动的工作文件（Active Workfile）。

### 6．工作区域

工作区域（Work Area）位于主窗口的中部，类似一个文件夹。工作区域内显示的是在操作过程中生成的各类对象，每个对象类似于文件夹中的一份文件或一页纸。这些对象会

以标题或 Windows 子窗口的形式显示。如果想要激活一个对象，单击子窗口的标题栏或者其他可见的部分，子窗口可以通过按 F6 键或 Ctrl+Tab 组合键进行切换。

**7．命令浏览器**

命令浏览器是在 EViews 11 和 EViews 12 中新增加的功能，位于命令窗口的右方，相当于命令（Command）的帮助按钮。当光标移动到主窗口右上角的 Command Explorer 按钮上时，将会出现 EViews 所有的命令，单击想要打开的命令，将会打开 EViews 在互联网上的帮助页面，获得使用该命令的相关解释。

# 1.2　工 作 文 件

EViews 文件是带结构的数据工作文件（Workfile），数据在录入、导入、处理和分析之前，需要建立一个新的工作文件。建立新工作文件主要有两种方法：一种是直接在软件中建立一个新工作文件；另一种是通过读取外部数据（非 EViews 格式文件）来建立新工作文件。

## 1.2.1　新工作文件的建立

工作文件的基本结构类型（Workfile Structure Type）分为三种：时间序列数据（Dated-regular Frequency）、面板数据（Balanced Panel）和无结构数据（Unstructured/Undated，也称作截面数据）。在 EViews 12 的主菜单中依次选择 File | New | Workfile 命令，出现如图 1-5 所示的 Workfile Create（建立工作文件）对话框。默认结构是时间序列，在 EViews 操作的过程中，可以随时对数据类型进行调整。对话框中的各选项如下：

**1．时间序列工作文件的建立**

时间序列需要设置时间的起点（Start date）、终点（End date）和时间频率（Frequency）。在 Frequency 下拉列表框中，有 14 个时间频率选项（见图 1-6），分别是：Multi-year（多年）、Annual（每年）、Semi-annual（半年）、Quarterly（每季度）、Monthly（每月）、Bimonthly（每月两次）、Fortnightly（每两周）、Ten-day（Trimonthly）（每季度内以十天为周期）、Weekly（每周）、Daily-5 day week（每周 5 个工作日）、Daily-7 day week（每周 7 天）、Daily-custom week（每日自定义周期）、Intraday（一天的交易时间内）和 Integer date（整数日期）。时间序列分析是 EViews 重要的功能之一，软件默认的是年度数据的时间序列。

**2．无结构/无时间顺序数据（截面数据）工作文件的建立**

截面数据是没有顺序的数据，需要在对话框的文本框中输入观察值（Observations）的数量，如图 1-7 所示。

**3．面板数据工作文件的建立**

面板数据实际上是时间序列和截面数据的结合，它的选项设置包含时间序列的频率、时间起点、时间终点和截面数据的数量，如图 1-8 所示。

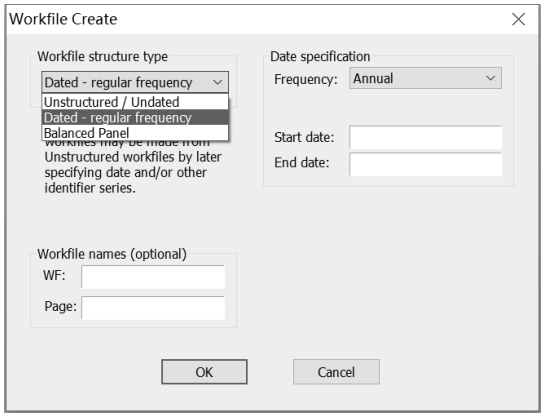



图 1-5　Workfile Creat 对话框

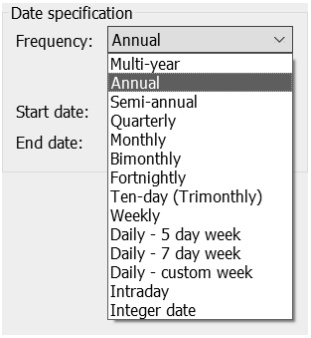



图 1-6　时间频率选项


图 1-7　建立无结构/无时间顺序工作文件

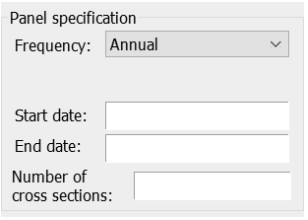



图 1-8　建立面板数据工作文件

## 1.2.2　读取外部数据

EViews 可以通过读取外部其他格式的数据文件，直接建立 EViews 工作文件，包括常见的 Excel、SAS、SPSS、Stata、TXT 等格式的文件。

### 1．直接复制、粘贴数据

EViews 支持直接从其他格式文件中复制和粘贴数据，从而生成工作文件中的一个序列。例如，可以直接复制 Excel 文件中的一列（或多列），然后粘贴到 EViews 工作窗口内，此时软件会出现一个对话框，由用户设置起始时间和文件名称等参数，然后在工作窗口中会生成一个（或多个）新序列。

### 2．直接读取外部数据建立新工作文件

有多种方式可以建立新工作文件：

- ❑ 依次在主菜单中选择 File | Open | Open Foreign Data as Workfile 命令。
- ❑ 依次在主菜单中选择 File | Import | Import from File 命令。
- ❑ 依次在主菜单中选择 Proc | Import | Import from File 命令。
- ❑ 直接用鼠标把外部数据文件拖入工作文件窗口。
- ❑ 右击选定的外部数据文件，用 EViews 程序打开该文件。

以上几种方式均会出现一个读取外部数据的对话框，用户通过该对话框可以对读取数

据的参数进行设定。此外，所有外部的各种数据库文件都可以通过在主菜单中选择 File | Open | Database 命令打开。

在工作文件中的每个序列内，可以手动录入或修改数据，但需要在可编辑状态下才可以操作。

**【例 1-1】**Excel 文件是最常用的外部数据文件，文件 S1-1.XLSX 是 2016 年 1 月至 2020 年 12 月中国的 CPI 数据（见表 1-1）。在 EViews 12 中直接读取该文件，注意观察变量名称的设置。

表 1-1　2016—2020 年中国的CPI（季度数据）

| Year | CPI |
|---|---|
| 2016-01 | 101.8 |
| 2016-02 | 102.3 |
| 2016-03 | 102.3 |
| 2016-04 | 102.3 |
| 2016-05 | 102 |
| 2016-06 | 101.9 |
| 2016-07 | 101.8 |
| 2016-08 | 101.3 |
| 2016-09 | 101.9 |
| … | … |

打开一个空白的 EViews 窗口，直接用鼠标将文件 S1-1.XLSX 拖入工作区域窗口内，弹出如图 1-9 所示的读取外部数据对话框，根据提示设置新工作文件的各项参数，最后单击“完成”按钮，生成一个新的工作文件，如图 1-10 所示。新的工作文件建立后，需要检查相关参数和数据的准确性。

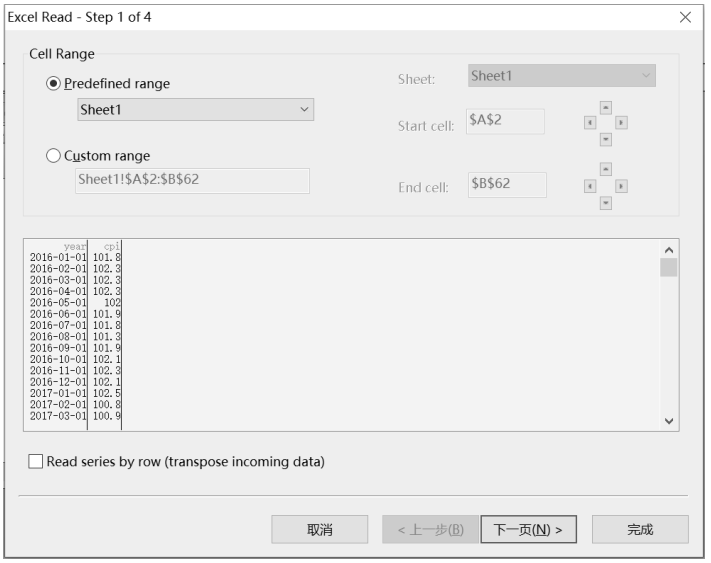



图 1-9　读取外部数据的对话框

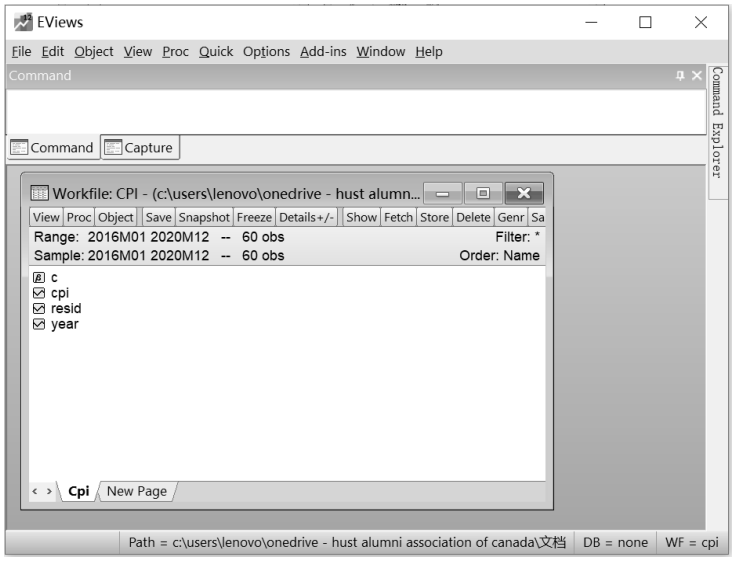



图 1-10　通过 Excel 建立的 CPI 工作文件

## 1.2.3　工作文件窗口

在 EViews 12 主窗口的工作区域建立 Workfile 后，将会出现如图 1-11 所示的工作文件窗口，这是进行数据分析和建模的窗口，所有的数据、序列、图形和方程等都保存在这个窗口中。EViews 13 之前的各个版本，工作文件窗口的界面均十分相似。

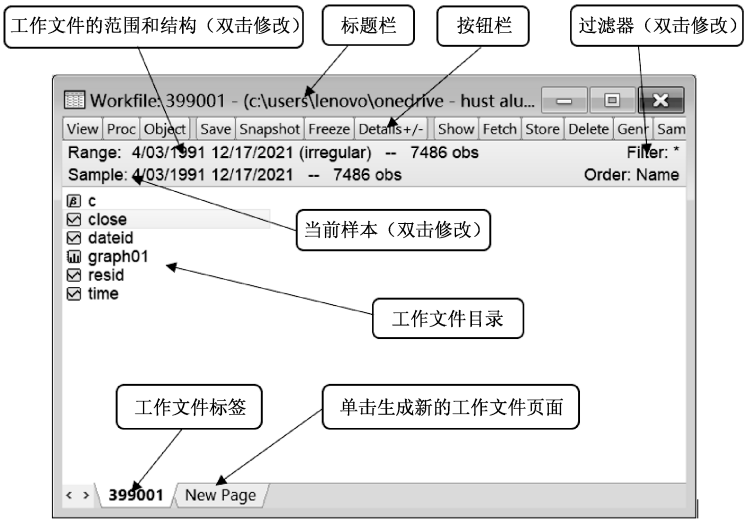



图 1-11　工作文件窗口

### 1．工作文件窗口

工作文件窗口包括标题栏、按钮栏、工作文件的范围和结构、当前样本等，下面具体

介绍。

- 标题栏：显示当前工作文件的名称和路径，如果工作文件尚未保存，则会显示“未命名”（UNTITLED）。
- 按钮栏：未对 EViews 单个对象进行操作时，工作窗口一共有 13 个操作按钮，分别是 View（浏览）、Proc（处理）、Object（对象）、Save（工作文件保存）、Snapshot（快照）、Freeze（冻结）、Details+/-（细节）、Show（展示）、Fetch（提取）、Store（操作对象保存）、Delete（删除）、Genr（新生成）和 Sample（取样）。
- 工作文件的范围和结构：双击 Range 及后面的数值，可以修改当前工作文件的结构和数据范围。
- 当前样本：双击 Sample 及后面的数值，可以修改当前操作的样本范围。
- 工作文件目录：包含打开的工作文件的所有对象的目录，单击 Details+/−，可以看到这些对象的详细信息。
- 工作文件标签：也是工作文件的名称，在打开多个工作文件时，方便用户识别和切换。
- 新工作文件页面：单击 New Page 按钮会出现一个建立新的工作文件的对话框，EViews 支持同时处理多个工作文件。
- 过滤器：当工作文件窗口有大量文件时，双击过滤器，可以对不同类型的文件进行过滤，以方便操作。

### 2．EViews英文字母的大小写说明

EViews 软件操作时英文字母不区分大小写。工作文件窗口默认为小写英文字母，依次选择 View | Name Display | Uppercase，可以将工作文件窗口内的所有文件名称显示为大写英文字母。在 EViews 的输出结果中，图形输出界面和函数（模型）的变量名称默认为大写英文字母。用户在操作过程中可以按自己的习惯输入大写或小写的英文字母。本书变量的输入一般采用小写的英文字母，输出结果为软件自动生成的界面，英文字母默认为大写。

### 3．工作文件的保存和退出

在 EViews 主菜单中依次选择 File | Save 或 File | Save as 命令，可以保存正在操作的文件，EViews 文件的后缀名是.WF1。

单击工作文件窗口右上角的关闭按钮，可以直接关闭当前的工作文件，但 EViews 软件并没有退出。

## 1.3　对　　象

对象（Object）是 EViews 的核心概念，也是实现数据分析和建模的基础。可以把它理解成工作文件中的一个单元，这些单元内分别储存了不同类型的信息。对象内的信息可以是序列，也可以是表格、图像和方程等，一个 EViews 工作文件包含若干个对象。

### 1.3.1　对象的建立

新建或者打开一个工作文件后，在 EViews 主窗口或工作文件窗口依次选择 Object | New Object 命令；或者在工作文件窗口右击，在弹出的快捷菜单中选择 New Object 命令，均可弹出 New Object 对话框，如图 1-12 所示。Type of object 下共有 24 种对象选项，这些对象在工作文件窗口的工作文件目录中对应不同的图标，如图 1-13 所示。

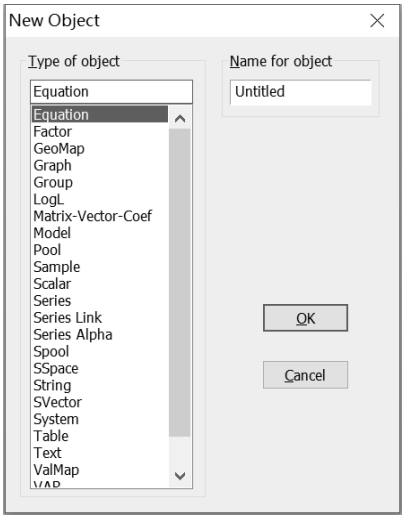



图 1-12　New Object 对话框

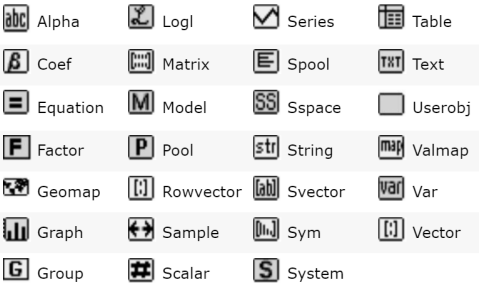



图 1-13　Object 在工作文件窗口中的图标

从选项菜单中选择希望建立的对象，将弹出相应的新对话框，可以根据提示，在工作文件窗口中建立一个新的对象。在 EViews 对象中，最常用的是 Series（序列）、Equation（方程）、Graph（图形）和 Table（表格）等。

在 New Object 对话框中，可以在 Name for object 文本框中输入对象的名称，也可以在建立新对象后，关闭该对象时再输入名称。对象的名称不区分大小写字母，同时不能使用软件的保留字符。

### 1.3.2　对象窗口

建立 EViews 对象后，对象名称会出现在工作文件目录中，单击该对象，弹出对象窗口（见图 1-14），不同的对象有不同的对象窗口。在图 1-14 中保存的图形对象 Graph01，包含一个脉冲响应函数的分析结果。对象窗口上方是对象工具栏，通过这个工具栏可以进行 EViews 软件的相关操作。

在工作文件的对象目录中，有两个对象由软件自动生成：一个是 c，代表系数向量；另一个是 resid，代表残差序列。进行模型估计后，模型的系数和残差会分别保存在 c 和 resid 序列内。

**【例 1-2】**建立一个名为 S1-1.WF1 的截面数据工作文件，数据范围设置为 20 个值。在工作文件窗口建立一个名为 a 的新序列（Serie）；打开序列 a，进入编辑模式，依次录入 47，

64，23，71，38，64，55，41，59，48，71，35，57，40，58，44，80，55，37，74；修改工作文件的结构，数据范围设置为 30；将样本区间修改为最后 10 个值，保存并退出。

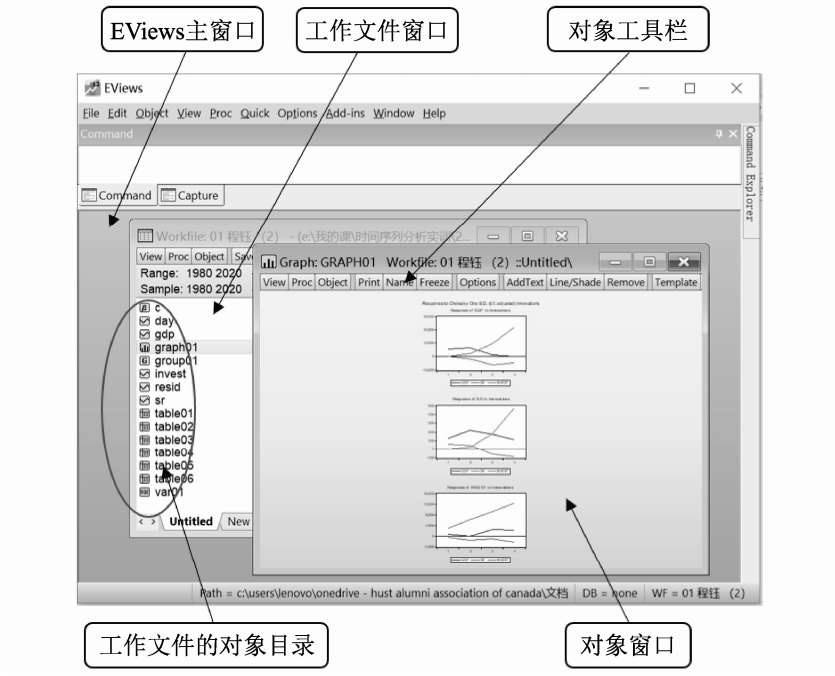



图 1-14　Object（对象）窗口

本例的操作步骤如下：

（1）打开 EViews 软件，在主菜单中依次选择 File | New | Workfile 命令，文件结构选择 Unstructured/Undated，在 Data range 的 Observations 文本框内输入 20，在工作文件窗口中输入 S1-1，单击 OK 按钮。

（2）在工作文件窗口中依次选择 Object | New Object 命令，在弹出的对话框中，在 Type of object 列表框中选择 Series，在 Name for object 文本框中输入 a，单击 OK 按钮（见图 1-15）。

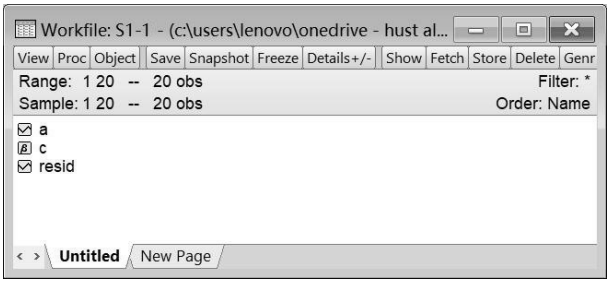



图 1-15　S1-1.WF1 的工作文件窗口

（3）双击打开序列 a，单击 Edit+/-按钮，切换为编辑状态（见图 1-16），依次将本例中的 20 个数值输入序列然后关闭序列。

（4）双击工作文件中的 Range，或在主菜单中选择 Proc-Structure | Resize Current Page 命令，在弹出的对话框的 Observations 文本框中输入 30，单击 OK 按钮；或者在 Command

命令窗口内输入 range 30 后按 Enter 键。

（5）双击工作文件中的 Sample，在 Sample range pairs 文本框内输入 11 20，中间用一个空格隔开，单击 OK 按钮；或者在 Command 命令窗口中输入 smpl 11 20 后按 Enter 键。可以发现，工作文件窗口的 Sample（样本）区间已经变为“11 20”，观察值调整为 10，如图 1-17 所示。

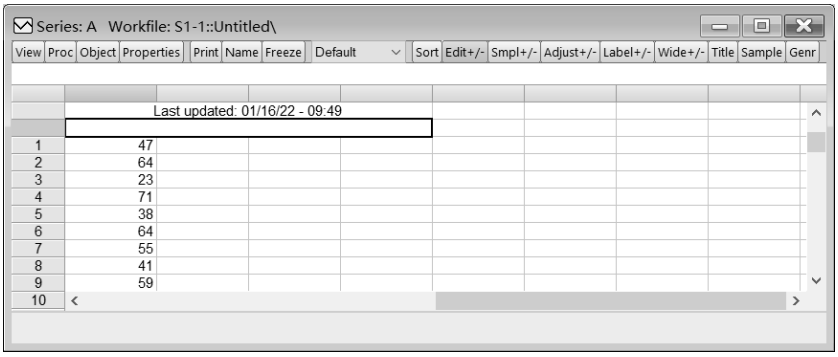



图 1-16　打开序列 a

（6）在主菜单中依次选择 File | Save，弹出 Workfile Save（保存工作文件）对话框（见图 1-18）。工作文件有两种数据存储类型：Single precision（7 digit accuracy）用于存储单精度（32 位）浮点数，每个数值用 32 位（4 字节）存储，适用于不需要高精度的数值存储，如图像和音频数据；Double precision（16 digit accuracy）用于存储双精度（64 位）浮点数，每个数值用 64 位（8 字节）存储，适用于高精度数值的存储，如金融领域的计算。EViews 软件默认的存储类型为双精度（64 位）浮点数。勾选 Use compression 复选框可以在保存工作文件时使用压缩技术，以减小数据的体积，提高数据传输的速度和安全。单击 OK 按钮，文件将会保存在 EViews 软件预设的路径中。

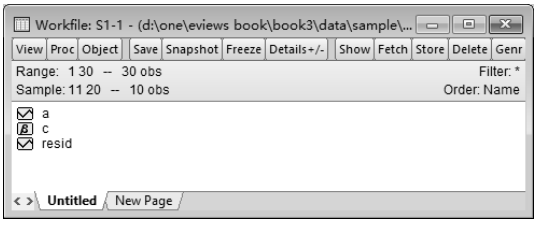



图 1-17　修改样本区间后的工作文件窗口

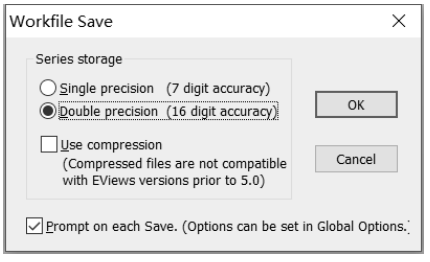



图 1-18　Workfile Save 对话框

## 1.3.3　生成新序列

对工作文件中已有的数据，通过数学运算，获得新序列。在主菜单中选择 Quick | Generate Series 命令，或者在工作文件窗口中单击工具栏上的 Genr 按钮，弹出 Generate Series by Equation 对话框。在其中输入方程，可以生成新的序列。

**【例 1-3】**在工作文件 S1-1.WF1 中，对序列 a 取自然对数，得到一个新的对数序列 lna。

打开如图 1-19 所示的 Generate 对话框，在 Enter equation 文本框中输入方程 lna=log(a)，单击 OK 按钮；或者在 Command 命令窗口输入 genr  lna=log(a)，按 Enter 键。可以发现，在工作文件窗口中多了一个新序列 lna，它是序列 a 的对数序列。

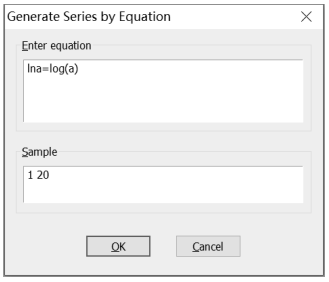



图 1-19　Generate 对话框

## 1.4　上 机 练 习

1．Excel 工作簿文件 E1-1.XLSX 存储的是美国 1950—1981 年的季度经济数据（见表 1-2），其中，GDP 为国内生产总值，INVEST 为总投资，CONS 为总消费，UNEMP 为失业率。

表 1-2　美国 1950—1981 年的季度经济数据

| TIME | GDP | INVEST | CONS | UNEMP |
|---|---|---|---|---|
| 1950-01 | 1350.9000 | 43.4000 | 183.6000 | 6.4000 |
| 1950-04 | 1393.5000 | 48.6000 | 187.5000 | 5.5667 |
| 1950-07 | 1445.2000 | 53.5000 | 201.2000 | 4.6333 |
| 1950-10 | 1484.5000 | 63.9000 | 198.6000 | 4.2333 |
| 1951-01 | 1504.1000 | 60.4000 | 209.7000 | 3.5000 |
| 1951-04 | 1548.3000 | 65.4000 | 205.3000 | 3.1000 |
| 1951-07 | 1585.4000 | 61.7000 | 207.9000 | 3.1667 |
| 1951-10 | 1596.0000 | 57.3000 | 211.9000 | 3.3667 |
| 1952-01 | 1607.7000 | 58.9000 | 213.3000 | 3.0667 |
| 1952-04 | 1612.1000 | 51.1000 | 217.4000 | 2.9667 |
| 1952-07 | 1621.9000 | 52.8000 | 219.9000 | 3.2333 |
| 1952-10 | 1657.8000 | 55.8000 | 228.0000 | 2.8333 |
| … | … | … | … | … |

（1）新建一个合适的工作文件。

（2）将 Excel 文件 E1-1.XLSX 导入工作文件中。

（3）将新建的工作文件命名为 E1-1.WF1。

2．Excel 工作簿文件 E1-2.XLSX 存储的是 2020 年 1 月至 2021 年 12 月深证成份指数（深证成指，代码 399001）历史行情的相关数据（见表 1-3）。

表 1-3　2020 年 1 月至 2021 年 12 月深证成指历史行情

| 日　　期 | 开　盘　价 | 收　盘　价 | 涨　跌　额 | 涨　跌　幅 | 成交金额（万） |
|---|---|---|---|---|---|
| 2020-01-02 | 10 509.12 | 10 638.83 | 208.06 | 1.99% | 23 772 802 |
| 2020-01-03 | 10 666.66 | 10 656.41 | 17.58 | 0.17% | 20 886 406 |
| 2020-01-06 | 10 599.41 | 10 698.27 | 41.86 | 0.39% | 25 401 960 |
| 2020-01-07 | 10 725.18 | 10 829.05 | 130.78 | 1.22% | 22 600 346 |
| 2020-01-08 | 10 776.71 | 10 706.87 | −122.18 | −1.13% | 24 351 342 |
| 2020-01-09 | 10 807.04 | 10 898.17 | 191.3 | 1.79% | 22 747 518 |
| 2020-01-10 | 10 927.98 | 10 879.84 | −18.33 | −0.17% | 20 125 750 |
| 2020-01-13 | 10 894 | 11 040.2 | 160.36 | 1.47% | 22 032 536 |
| 2020-01-14 | 11 074.89 | 10 988.77 | −51.43 | −0.47% | 22 227 506 |
| 2020-01-15 | 10 978.28 | 10 972.32 | −16.45 | −0.15% | 19 046 184 |
| 2020-01-16 | 10 986.65 | 10 967.44 | −4.88 | −0.04% | 19 366 554 |
| … | … | … | … | … | … |

（1）新建一个合适的 EViews 工作文件。

（2）将 Excel 文件 E1-2.XLSX 的所有数据导入工作文件中，并将导入的序列分别命名为 a1（开盘价）、a2（收盘价）、a3（涨跌额）、a4（涨跌幅）和 a5（成交金额）。

（3）将新建的工作文件命名为 E1-2.WF1。

（4）对收盘价序列 a2 进行对数处理，生成一个新的对数序列 lna2。

（5）对序列 lna2 进行一阶差分，生成新的序列 dlna2，差分方程为 dlna2=d(lna2)。

# 第 2 章　EViews 基本数据分析（单序列）

EViews 软件提供图形化、描述性统计量和假设检验等多种分析方法，是进一步建立模型、分析模型和预测的基础。EViews 基本数据分析在工作文件窗口内操作，其中最重要的是对序列（Series）的分析，包括单个序列和序列组。

在工作文件窗口中打开一个序列，单击对象窗口的工具栏，基本数据分析的选项位于 View 按钮的下拉菜单（见图 2-1 中），共 13 个选项。其中，有 3 个选项的右边有黑色三角形标志，表示它们还有子菜单。从 EViews 12 开始，增加了 Wavelet Analysis（小波分析）选项。这些选项分为 4 个部分：第一部分用于数据展示，包括电子表格和图形的切换；第二部分用于基本统计量的数据分析和检验；第三部分用于时间序列的数据分析和检验；第四部分用于对序列标签的操作。

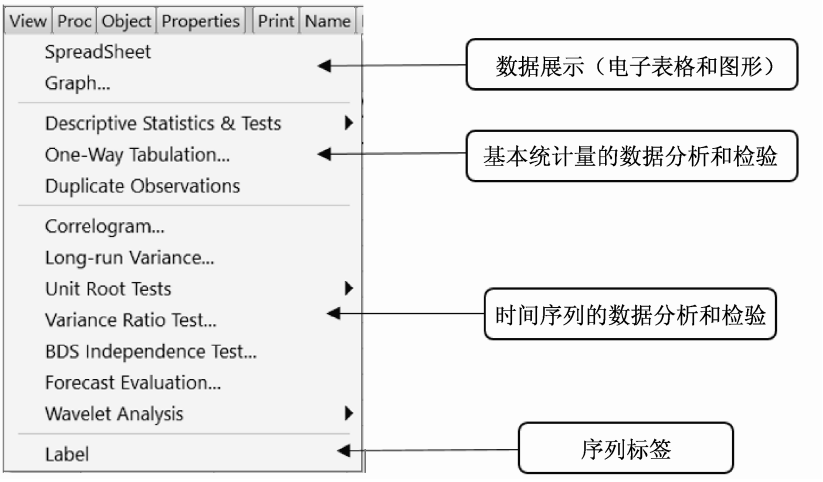



图 2-1　单序列的基本数据分析菜单

## 2.1　数据的展示

数据展示是指将序列内的数据直观地展示给用户，EViews 主要通过电子表格（SpreadSheet）和图形（Graph）来展示数据。

## 2.1.1　电子表格

电子表格类似 Excel 表格，是 EViews 中显示数据的基本形式。打开一个序列，在工具栏中依次选择 View | SpreadSheet，可以在原始数据、转换后的数据、图形与电子表格之间进行切换。

单击工具栏中的 Properties 按钮，可以对表格中栏的宽度、颜色和频率等基本属性进行设置。单击 Edit+/-按钮，可以对表格内的数据进行编辑。单击 Sort 按钮，可以对表格内的数据进行排序。

**【例 2-1】**工作文件 S2-1.WF1 存储的是 2011—2021 年中国 CPI 的月度数据，打开工作文件，单击序列 CPI，得到如图 2-2 所示的电子表格，可以继续在该表格中进行编辑。

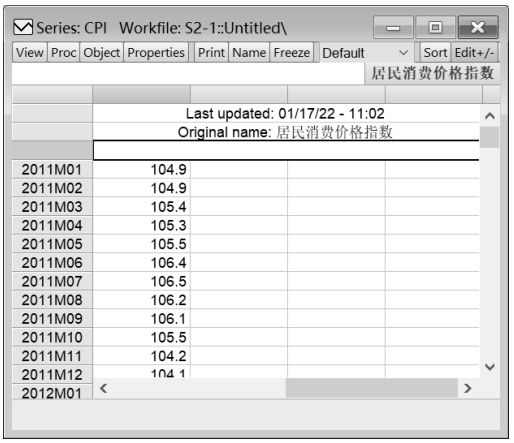



图 2-2　SpreadSheet 表格

## 2.1.2　绘图

可视化是数据分析的重要内容，EViews 的绘图功能也变得越来越强大。打开一个序列，选择 View | Graph，弹出绘图功能对话框，在其中进行相关操作即可。绘图功能后面会专门学习。

**【例 2-2】**打开工作文件 S2-1.WF1，单击序列 CPI，将会得到如图 2-2 所示的电子表格，选择 View | Graph 命令，进入 Graph Options 绘图对话框，如图 2-3 所示，默认为折线图（Line & Symbol）。单击 OK 按钮，得到输出的图形。单击 Freeze 按钮，可以将生成的图形冻结，然后将其命名为 GRAPH01 的图形对象并保存，如图 2-4 所示。

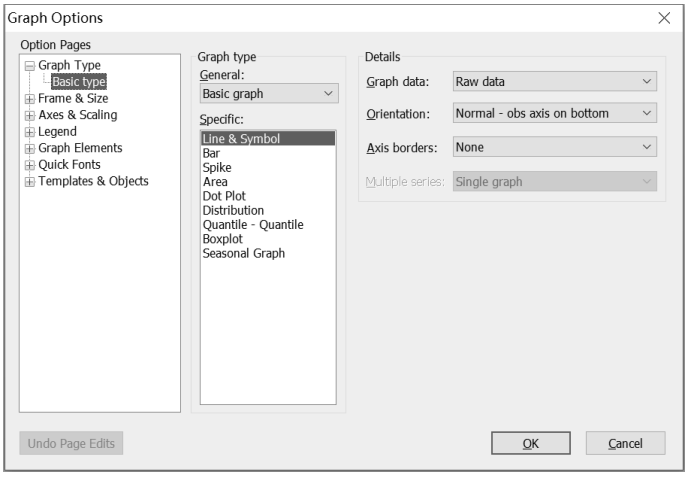



图 2-3　Graph Options 对话框

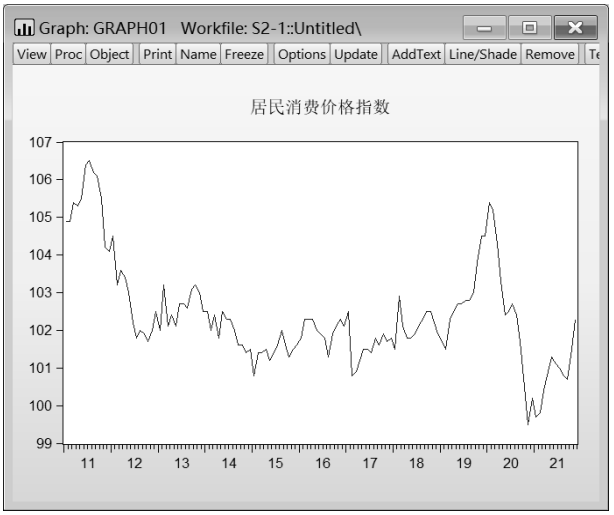



图 2-4　Graph 的输出图形

# 2.2　基本统计量分析和检验

EViews 软件拥有强大的统计分析功能。基本统计量分析和检验是数据分析的前提，包括描述性统计分析、频数分析、列联表分析、相关性分析、参数检验与非参数检验等。

## 2.2.1　描述性统计量和检验

数据的描述性统计量和检验（Descriptive Statistics & Tests）是进行统计分析的基础，数据的描述性统计量一般分为 3 类：

- 刻画数据集中趋势的描述性统计量（描述水平的统计量），主要包括均值（Mean）、中位数（Median）、最大值（Maximum）和最小值（Minimum）等。
- 刻画数据离散程度的描述性统计量（描述差异的统计量），主要包括标准差（Std. Dev.）和方差（Variance）。方差是标准差的平方，这两个统计量只需要一个就可以。
- 刻画数据分布形态的描述性统计量，主要包括偏度系数（Skewness）和峰度系数（Kurtosis），同时判断数据是否符合正态分布（Normal Distribution）。

掌握以上 3 类统计量后，就能够较为清晰地了解数据的分布特点。

在 EViews 工作文件窗口中打开任意一个序列，选择 View | Descriptive Statistics & Tests，出现描述性统计量和检验的子菜单，如图 2-5 所示，共有 6 个选项，下面具体介绍。

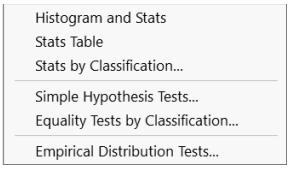



图 2-5　描述性统计量和检验的子菜单

### 1．Histogram and Stats（直方图和统计量）

Histogram and Stats 选项的输出结果，除了包括前面提到的 3 类基本描述统计量，也可以对数据是否呈现正态分布

进行假设检验。直方图是将序列的值按相等的组距进行划分，显示数据的频率分布情况。和条形图不同，直方图用面积表示数量的值。

**【例 2-3】**打开工作文件 S2-1.WF1 中的序列 CPI，进行直方图和统计量分析。

打开 S2-1.WF1 工作文件，接着打开序列 CPI，依次选择 View | Descriptive Statistics & Tests-Histogram and Stats，得到如图 2-6 所示的输出结果。图左边为序列 CPI 的直方图，右边为序列的基本统计量和 Jarque-Bera 检验。

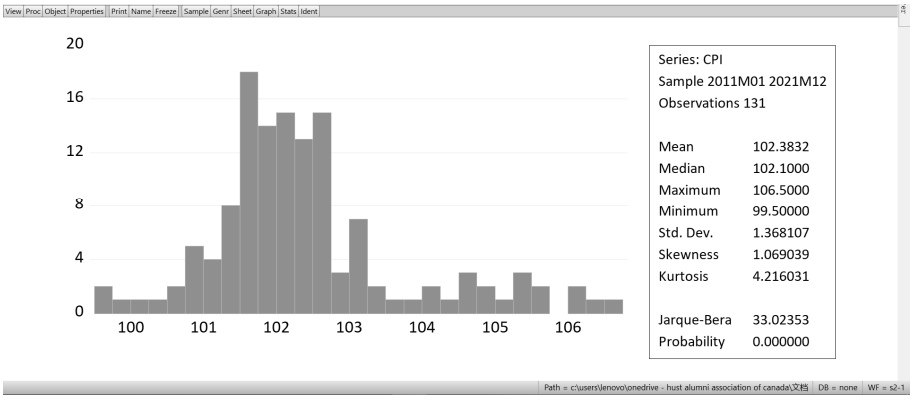



图 2-6　Histogram and Stats 输出结果

从图 2-6 右侧可以看出，序列 CPI 的样本从 2011 年 1 月至 2021 年 12 月，共有 131 个观察值，它的基本统计量解释如下：

- Mean：所有数据的算术平均数。在本例中，131 个观察值的平均值为 102.3832。
- Median：所有数据按升序排序后，处于中间位置的数据值（或最中间两个数据的平均值）。在本例中，131 个观察值的中位数是 102.1。
- Maximum：在 131 个观察值中，最大值为 106.5。
- Minimum：在 131 个观察值中，最小值为 99.5。
- Std. Dev.：标准差，表示序列观察值的平均离散程度，标准差越大，数据的离散程度越强。在本例中，序列 CPI 的标准差为 1.368107。
- Skewness：偏度系数描述数据分布形态对称性的统计量。当分布对称时，偏度系数为 0；如果偏度系数大于 0，则数据分布呈右偏（或正偏）分布，在直方图中有一条长尾拖在右边；如果偏度系数小于 0，则数据分布呈左偏（负偏）分布，在直方图中有一条长尾拖在左边。在本例中，序列 CPI 的偏度系数为 1.069039，大于 0，属于右偏分布。
- Kurtosis：峰度系数，是描述数据分布形态陡缓的统计量，标准正态分布的峰度系数等于 3。如果峰度系数大于 3，则数据的分布比标准正态分布更陡峭，称为尖峰分布；如果峰度系数小于 3，则数据的分布比标准正态分布更平缓，称为平峰分布。在本例中，序列 CPI 的峰度系数为 4.216031，大于 3，属于尖峰分布。
- Jarque-Bera（简称 JB）检验：对观察值是否符合正态分布的检验。原假设是序列服从正态分布，如果 JB 统计量的伴随概率 P-值大于设定的显著性水平，则不拒绝原假设，数据服从正态分布；如果伴随概率 P-值小于设定的显著性水平，则拒绝原假设，数据不服从正态分布。需要注意的是，JB 统计量一般用于大样本的检验。在

本例中，序列 CPI 的 JB 统计量的伴随概率接近于 0，如果设定的显著性水平为 0.05，则拒绝原假设，数据不服从正态分布。

如果希望在图 2-6 中加入正态分布的理论密度曲线，则需要通过 Graph 绘图功能进行单独设置，方法是在直方图和统计量的输出结果中，选择 View | Graph，弹出绘图对话框。在 Specific 下面的列表框中选择 Distribution（分布），右边出现 Distribution 的选择框，选择 Theoretical Distribution（理论分布），单击右边的 Option 选项，弹出 Distribution Plot Customize 对话框，单击 Added Elements 下方的 Add 按钮，在 Add 对话框中选择 Theoretical Density（理论密度），单击 OK 按钮。在右边 Specification Distribution 的选项框内，默认的理论分布就是 Normal（正态分布），单击 OK 按钮返回到绘图对话框，再单击 OK 按钮。最终得到如图 2-7 所示的包含正态分布曲线的直方图。

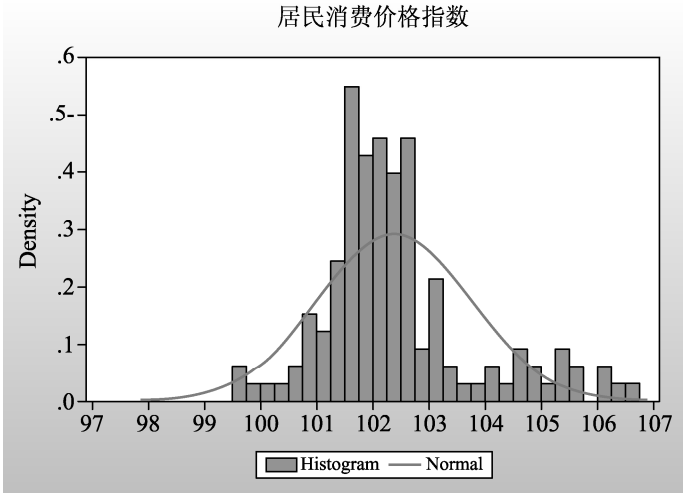



图 2-7　带正态分布曲线的直方图

### 2．Stats Table（统计量表）

Stats Table 是以电子表格的形式输出基本描述性统计量。除了图 2-6 所示的基本统计量外，还增加了所有观察值求和的总数（Sum）和离差平方和（Sum Sq. Dev.）。

**【例 2-4】**对工作文件 S2-1.WF1 中的序列 CPI 进行统计量表分析。

打开 S2-1.WF1 工作文件，继续打开序列 CPI，依次选择 View | Descriptive Statistics & Tests-Stats Table，得到如图 2-8 所示的输出结果，统计量的解释和例 2-3 相同。

### 3．Stats by Classification（分组统计量）

Stats by Classification 是对序列观察值的描述性统计量进行分组，至少需要设置一个分组变量，如果有多个分组变量，则中间间隔一个空格。

打开一个序列，依次选择 View | Descriptive Statistics & Tests-Stats by Classification，弹出如图 2-9 所示的对话框。在对话框的 Statistics 统计量选项中选择 Mean、Median、Std. Dev. 和 Observations。在分组序列 Series/Group for classify 下方填入分组变量，如果有多个变量，则中间用空格进行间隔。NA handling 下方的选项，表示可以选择是否将缺失数据作为一个

单独的分组类型。

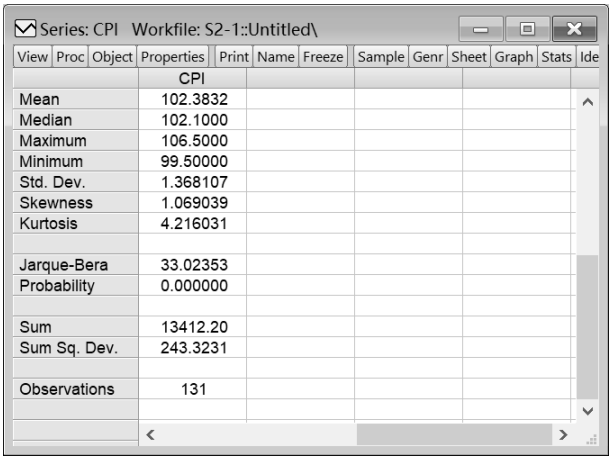



图 2-8　Stats Table 的输出结果

为了防止将分组单元设置得过小，Group into bins if 选项用于设定分组单元的个数和大小。其中，# of values 选项表示当分组序列内的观察值个数大于指定值时进行分组统计（默认值为 100），Avg. count 选项表示当每个分组序列的平均观察值小于指定值时进行分组统计（默认值为 2），Max # of bins 选项表示指定的最大分组数（这个选项只是进行大致控制）。

Output layout 选项用于选择输出的显示方式，可以选择 Table Display（表格方式显示）或 List Display（清单方式显示）。设置结束后，单击 OK 按钮。

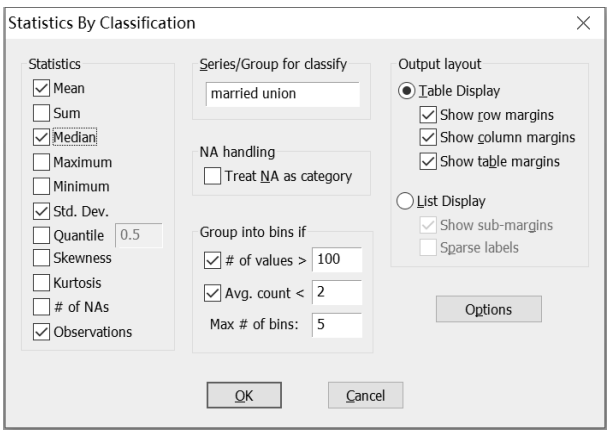



图 2-9　Statistics By Classification 对话框

**【例 2-5】**工作文件 S2-2.WF1 中存储的是搜集一些女性基本情况的数据（见图 2-10），一共有 1000 个样本。其中：lwage 为女性收入的对数序列；married 为婚姻状况，是分类变量，包括 married 和 single 两个类别；union 为参加工会情况，也是分类变量，包括 union 和 non-union 两个类别。要求对以上 3 个序列进行分组统计。

打开 S2-2.WF1 工作文件，接着打开序列 lwage，以序列 married 和 union 作为分组变量。在图 2-9 中的 Series/Group for classify 下方的文本框中输入分组变量 married union，中间用空格间隔。分组统计后得到图 2-11 所示的输出结果。

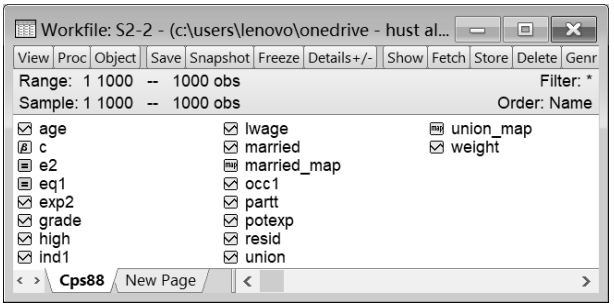



图 2-10　S2-2.WF1 工作文件窗口

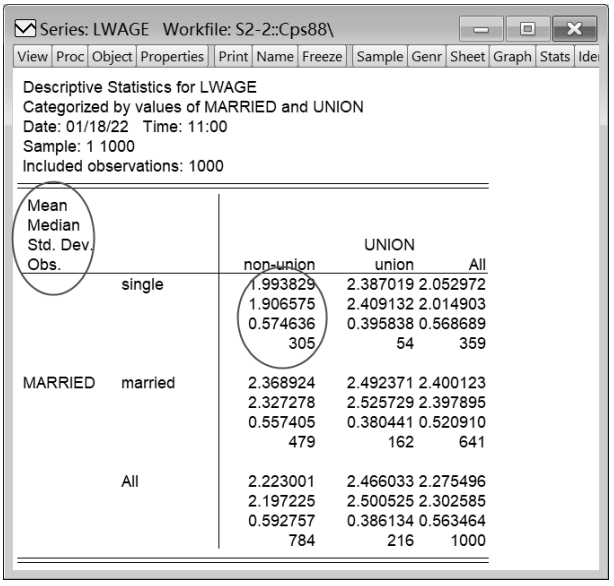


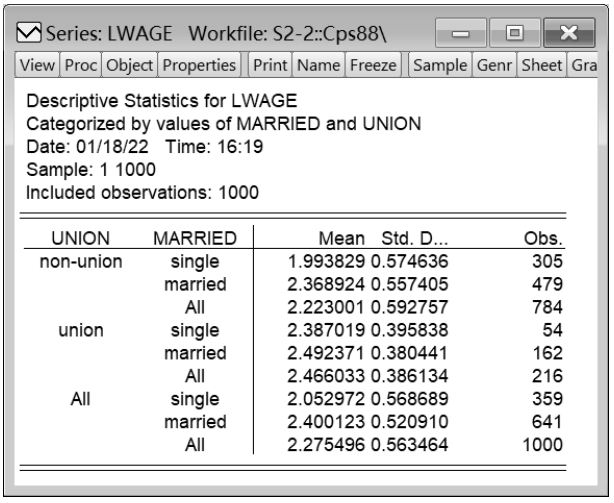

图 2-11　序列 lwage 分组统计量输出结果

在图 2-11 中，第一列为 MARRIED 变量，包括 single（单身）、married（已婚）和 All（合计）三类；第一行为 UNION 变量，包括 non-union（未参加工会）、union（参加工会）和 All（合计）三类。左边的椭圆内是统计量的名称，包括 Mean、Median 和 Std. Dev.，以及观察值 Obs.的个数。右边的椭圆内分别为对应统计量和观察的值，表示在未参加工会的单身女性中，工资均值（对数后）为 1.993829，工资中位数（对数后）为 1.906575，工资标准差为 0.574636，共有 305 位这样的女性（观察值）。

在图 2-9 所示的对话框中，右下方的显示方式选择 List Display（清单显示），同时选中下面的两个复选框，这样可以得到更为清晰的逐行清单式显示，如图 2-12 所示。

### 4．Simple Hypothesis Tests（简单假设检验）

简单假设检验、分组齐性检验和经验分布检验属于统计分析方法的推断统计部分，在统计分析中占有重要地位，除了经济和金融等领域，也广泛应用于医学、生物、农业和林业等领域。它的基本原理是通过样本数据的特征来推断总体数据的特征。在其他软件如 SPSS 中被称作参数检验和非参数检验。




| UNION | MARRIED | Mean | Std. D... | Obs. |
|---|---|---|---|---|
| non-union | single | 1.993829 | 0.574636 | 305 |
| | married | 2.368924 | 0.557405 | 479 |
| | All | 2.223001 | 0.592757 | 784 |
| union | single | 2.387019 | 0.395838 | 54 |
| | married | 2.492371 | 0.380441 | 162 |
| | All | 2.466033 | 0.386134 | 216 |
| All | single | 2.052972 | 0.568689 | 359 |
| | married | 2.400123 | 0.520910 | 641 |
| | All | 2.275496 | 0.563464 | 1000 |

图 2-12　LWAGE 序列分组统计量的列表式输出结果

简单假设检验包含 3 个检验：均值检验、方差检验和中位数检验。它们的含义是，用户任意提出一个检验值，检验其是否和样本来自总体的均值、中位数或方差存在显著差异。

1）均值检验

用户任意设定一个检验值，利用已有的序列值（样本），推断是否和样本来自总体的均值有显著差异（双侧检验），实际上是一个单样本 t-检验过程。原假设为检验值和总体均值没有显著差异，备择假设为检验值和序列的均值有显著差异。

**【例 2-6】**工作文件 S2-3.WF1 存储的是某地区信用卡月平均消费金额的抽样数据（见图 2-13），一共有 500 个客户样本。其中，credit 为客户的月平均刷卡金额。据估计，当地信用卡月平均刷卡金额约为 3000 元。根据抽样数据，判断是否支持这个估计（假设）。

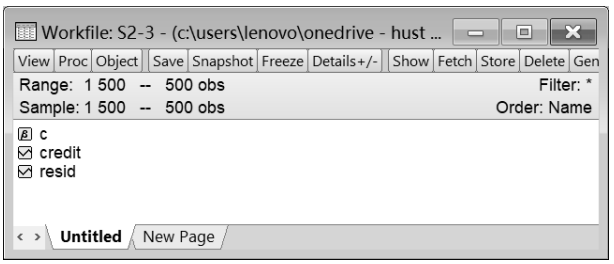



图 2-13　S2-3.WF1 工作文件窗口

这个案例并不是要求我们检验这 500 个样本的月刷卡金额均值和 3000 元有没有显著差异，因为这样的检验是没有意义的，我们可以直接得出这 500 个样本的月平均刷卡金额是 4781.879 元。

这个案例实际上是要求我们根据 500 个抽样数据的分布状况，检验当地所有信用卡持有人的月平均刷卡金额和 3000 元是否有显著差异，是一个典型的均值检验问题。

打开工作文件 S2-3.WF1，接着打开序列 credit，依次选择 View | Descriptive Statistics & Tests-Simple Hypothesis Tests，弹出 Series Distribution Tests（序列分布检验）对话框，如图 2-14 所示，这是一个数据分布的检验。在 Mean 文本框中输入需要检测的值 3000。单击

OK 按钮，得到图 2-15 所示的均值检验输出结果。

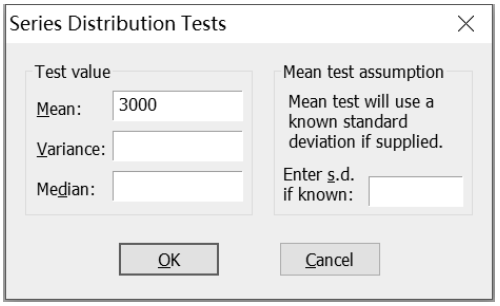



图 2-14　Series Distribution Tests 对话框

从图 2-15 中可以看到，序列 credit 有 500 个观察值，它们的均值为 4781.879 元，样本的标准差为 7418.718 元。原假设是检验值 3000 元和所有信用卡用户的月刷卡金额的均值没有显著差异，备择假设相反。检验的 t-统计量观察值为 5.370742，它的伴随概率 P-值接近 0，如果设定的显著性水平为 0.05，则显著拒绝原假设，接受备择假设，即该地区信用卡月消费金额的平均值和 3000 元有显著差异。

这是一个双侧检验，也就是该地区的信用卡月消费金额大于 3000 元，或者小于 3000 元，都符合题目要求。如果题目要求检测该地区信用卡月消费金额的平均值不低于 3000 元（或不高于 3000 元），则是单侧检验，这时候需要先把伴随概率 P-值除以 2，再和设定的显著性水平 0.05 进行比较。

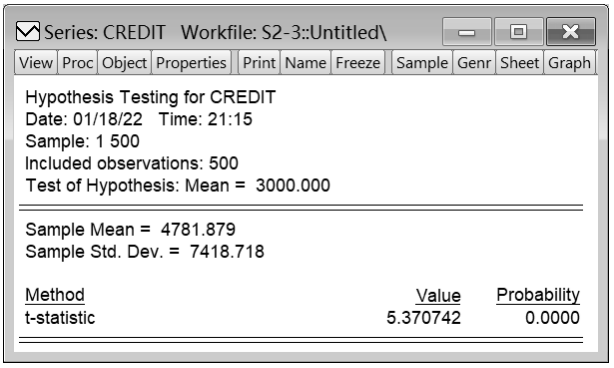



图 2-15　均值检验输出结果

2）方差检验

用户任意设定一个检验值，检验是否和序列的方差有显著差异（双侧检验），这是一个对卡方统计量的检验。原假设为检验值和序列的方差没有显著差异，备择假设为检验值和序列的方差有显著差异。

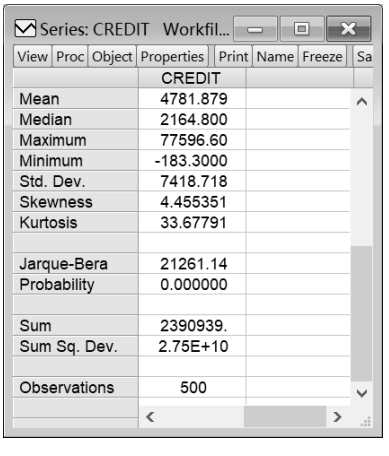



图 2-16　序列 credit 的基本统计量

**【例 2-7】**工作文件 S2-3.WF1 存储的是某地区信用卡月刷卡金额的抽样数据，判断该地区所有信用卡用户月刷卡金额的方差和 500 万是否有显著差异。

和上例的背景一样，要求我们根据序列 credit 的方差，判断样本来自总体的方差和 500 万是否有显著差异。

打开工作文件 S2-3.WF1 中的序列 credit，先做一个简单的描述性统计量表分析（见图 2-16），发现月信用卡消费最高金额为 77 596.6 元，最低金额为−183.3 元。最高金额和最低金额差距很大，这种情况下，序列总体方

差的值可能较大。

打开序列 credit，进入简单假设检验，在图 2-14 所示的 Variance 文本框中输入 500 万，得到图 2-17 所示的输出结果，序列 credit 的方差为 55037375，如果设定的显著性水平为 0.05，检验的伴随概率 P-值接近 0，远远小于设定的显著性水平，则显著拒绝原假设，即检验值 500 万和序列的方差有显著差异。

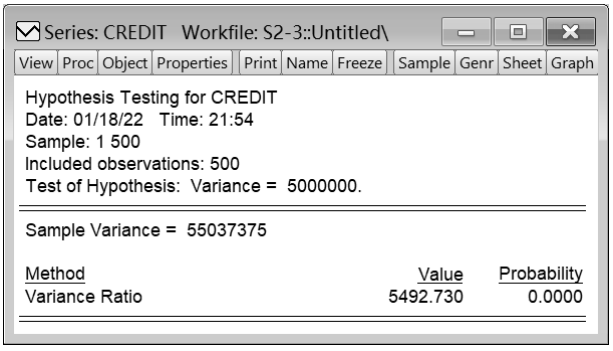



图 2-17　方差检验输出结果

3）中位数检验

用户任意设定一个检验值，检验是否和序列来自总体的中位数有显著差异（双侧检验）。原假设为检验值和总体的中位数没有显著差异，备择假设为检验值和总体的中位数有显著差异。

**【例 2-8】**工作文件 S2-3.WF1 存储的是某地区信用卡月平均消费金额的抽样数据，判断 2000 元是否为当地信用卡每月平均刷卡金额的中位数。

本例实际上是要求判断当地所有信用卡用户月消费金额的中位数，与 2000 元比较是否有显著差异。

均值是通过算术平均数得到的，容易受到极端值（最大值和最小值）的影响，因此，中位数也是描述数据集中趋势的重要统计量。

在本例中，抽样数据的最大值为 77 596.6 元，最小值为−183.3 元，方差的值非常大，均值可能并不是最好的描述数据集中趋势的统计量。这时候可以考虑中位数，中位数是通过排序得到的，它不受最大和最小两个极端数值的影响。部分数据的变动对中位数没有影响，当序列中的个别数据变动较大时，经常用中位数来描述这组数据的集中趋势。

打开工作文件 S2-3.WF1 中的序列 credit，依次选择 View | Descriptive Statistics & Tests-Simple Hypothesis Tests，在图 2-14 所示的 Median 文本框中输入 2000，得到图 2-18 所示的输出结果，序列 credit 的中位数为 2164.8。EViews 对中位数检验提供了几种常见的方法：

- Binomial sign test（二项符号检验）。该检验基于一个二项分布的假设，是一种用于比较两个相关的样本差异是否显著的非参数检验方法。它的原理是：如果样本是从一个二项分布总体中随机抽取的，那么样本数据也应当符合二项分布，即中位数上下的样本数据量应当是接近的。在图 2-18 中，两种二项符号检验结果的伴随概率均大于设定的显著性水平，不拒绝原假设，表明当地信用卡每月平均刷卡金额的中位数和 2000 元没有显著差异。

- Wilcoxon signed rank test（威尔科克森符号秩检验）。它的大致原理是：把所有观察值之间取差分后的绝对值，从高到低排序，接着对中位数上下的值进行求和，二者之间应当是相近的。原假设是输入的检验值和序列的中位数没有显著差异，备择假设是它们之间有显著差异。在本例中，假设显著性水平为 0.05，Wilcoxon signed rank 检验的伴随概率接近 0，小于设定的显著性水平，则拒绝原假设，即当地信用卡每月平均刷卡金额的中位数和 2000 元有显著差异。
- van der Waerden（normal scores）test（范德瓦尔登正态计分检验）。它的原理和 Wilcoxon signed rank 检验大致相同，只是它基于平滑排序：首先把数据转换为秩的排序，然后转换成标准正态分布的分位数（正态计分），可以用于非正态分布。原假设是输入的检验值和序列的中位数没有显著差异，备择假设是它们之间有显著差异。在本例中，假设显著性水平为 0.05，van der Waerden（normal scores）检验的伴随概率接近 0，小于设定的显著性水平，则拒绝原假设，即当地信用卡平均刷卡金额的中位数和 2000 元有显著差异。

可以看出，3 种检验方法的结果并不完全一致，后两种检验方法的精度更高。

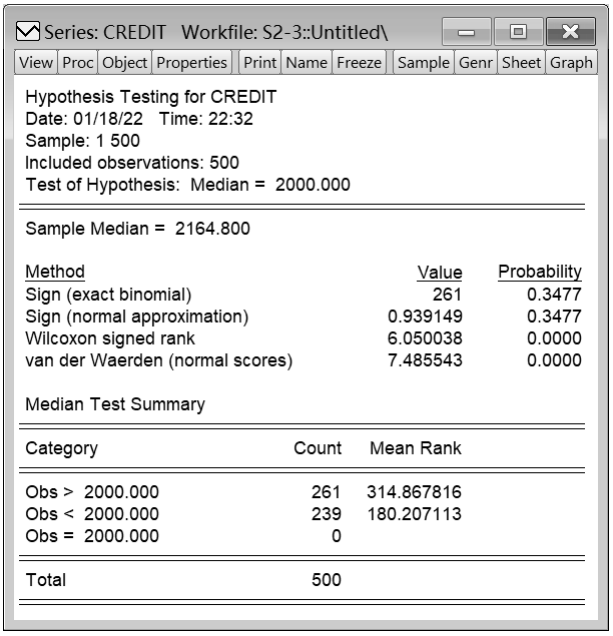



图 2-18　中位数检验输出结果

### 5．Equality Tests by Classification（分组齐性检验）

分组齐性检验是一个序列分组后，利用方差分析方法，对分组后的数据来自总体的均值、中位数和方差是否相同进行检验。

分组齐性检验需要设置至少一个分组变量，实际上是对分组后的数据进行差异性检验。例如，男性和女性的样本，检验各自总体收入的平均数或中位数是否有差异，教育则可以作为一个分类变量，检验它的方差是否和性别、种族有关。

打开一个序列，依次选择 View | Descriptive Statistics & Tests-Equality Tests by Classification，

弹出如图 2-19 所示的 Tests by Classification 对话框。在 Test equality of 下方选择均值、中位数或方差进行检验；在 Series/Group for classify 内输入分组变量，如果有两个以上分组变量，则中间用空格分隔；NA handling 选项表示可以选择将缺失值单独列为一个分组。

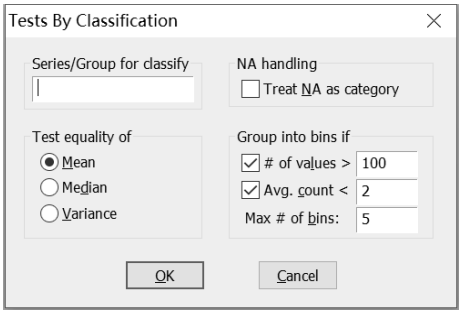



图 2-19　Tests By Classification 对话框

1）均值齐性检验

均值齐性检验（Mean Equality Test）是检验两个样本或多个样本的总体均值是否相等的检验方法。该检验的原假设是每个样本的总体均值都相等（没有显著差异），并且总体方差不等。如果样本的均值之间有显著差异，则拒绝原假设，表明总体均值不相等。

**【例 2-9】**经济学家和社会学家都十分关注男性和女性在收入上的差别，性别的收入鸿沟称作 Gender Wage Gap。根据案例 S2-4.WF1 中的数据（见图 2-20），以性别（sex）作为收入的分组变量，判断男性和女性的平均收入是否有显著差异。

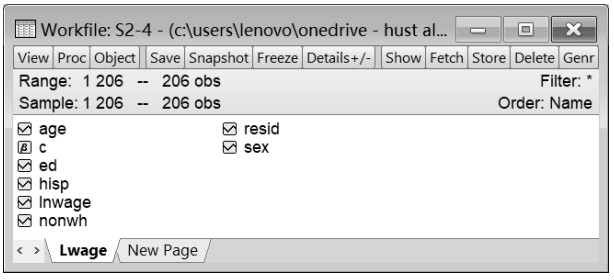



图 2-20　S2-4.WF1 工作文件窗口

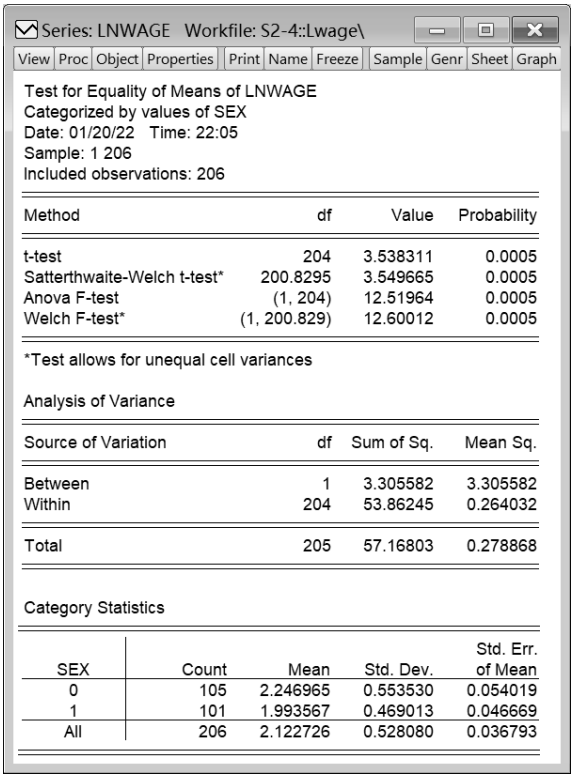
在本案例中，显然不是要求我们检验样本中男性和女性的平均收入是否有显著差异，而是希望通过样本推断总体，检验所有的男性和女性的平均收入是否有显著差异。

打开工作文件 S2-4.WF1，接着打开序列 lnwage（收入的对数序列），依次选择 View | Descriptive Statistics & Tests-Equality Tests by Classification，以 sex（性别）作为分组变量，在图 2-19 所示的 Test equality of 下面选择 Mean（均值），单击 OK 按钮，得到如图 2-21 所示的输出结果，软件使用了 4 种检验方法，分别是 t-test、Satterthwaite-Welch t-test*、Anova F-test 和 Welch F-test*。

从检验结果看，如果设定的显著性水平是 0.05，图 2-21 显示的通过 4 种检验方法得到的统计量伴随概率均小于设定的显著性水平，则拒绝原假设，即认为男性和女性的平均收入有显著差异。

2）中位数齐性检验

中位数齐性检验（Median/Distribution Equality Tests）是对不同分组序列来自总体的中位数之间是否有显著差异的检验。中位数齐性检验实际上是数据分布的非参数检验，因此也称作分布齐性检验。原假设是分组序列来自总体的中位数具有相同的分布，备择假设是至少有一组序列来自总体的数据具有不同的分布。



Series: LNWAGE　Workfile: S2-4::Lwage\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph

Test for Equality of Means of LNWAGE
Categorized by values of SEX
Date: 01/20/22　Time: 22:05
Sample: 1 206
Included observations: 206

| Method | df | Value | Probability |
|---|---|---|---|
| t-test | 204 | 3.538311 | 0.0005 |
| Satterthwaite-Welch t-test* | 200.8295 | 3.549665 | 0.0005 |
| Anova F-test | (1, 204) | 12.51964 | 0.0005 |
| Welch F-test* | (1, 200.829) | 12.60012 | 0.0005 |

*Test allows for unequal cell variances

Analysis of Variance

| Source of Variation | df | Sum of Sq. | Mean Sq. |
|---|---|---|---|
| Between | 1 | 3.305582 | 3.305582 |
| Within | 204 | 53.86245 | 0.264032 |
| Total | 205 | 57.16803 | 0.278868 |

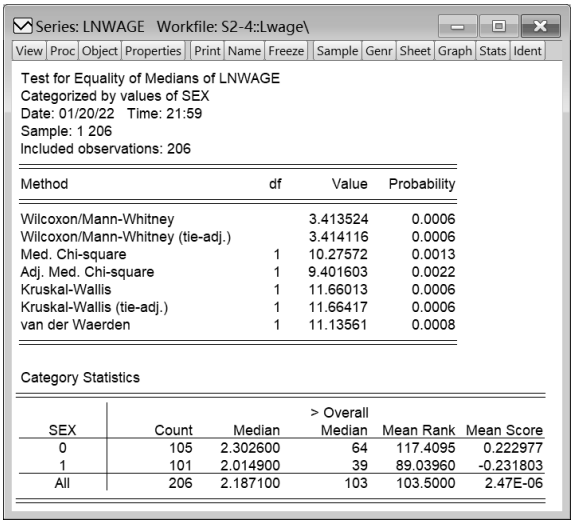
Category Statistics

| SEX | Count | Mean | Std. Dev. | Std. Err. of Mean |
|---|---|---|---|---|
| 0 | 105 | 2.246965 | 0.553530 | 0.054019 |
| 1 | 101 | 1.993567 | 0.469013 | 0.046669 |
| All | 206 | 2.122726 | 0.528080 | 0.036793 |

图 2-21　均值齐性检验的输出结果

**【例 2-10】**根据案例 S2-4.WF1 中的数据，以性别（sex）作为收入的分组变量，判断男性和女性收入的中位数是否有显著差异。

在图 2-19 所示的 Test equality of 下面选择 Median（中位数），单击 OK 按钮，得到图 2-22 所示的输出结果。EViews 软件使用了 4 种检验方法，分别是 Wilcoxon rank sum test、Chi-square test、Kruskal-Wallis one-way ANOVA by ranks 和 van der Waerden (normal scores) test。



Series: LNWAGE　Workfile: S2-4::Lwage\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Test for Equality of Medians of LNWAGE
Categorized by values of SEX
Date: 01/20/22　Time: 21:59
Sample: 1 206
Included observations: 206

| Method | df | Value | Probability |
|---|---|---|---|
| Wilcoxon/Mann-Whitney | | 3.413524 | 0.0006 |
| Wilcoxon/Mann-Whitney (tie-adj.) | | 3.414116 | 0.0006 |
| Med. Chi-square | 1 | 10.27572 | 0.0013 |
| Adj. Med. Chi-square | 1 | 9.401603 | 0.0022 |
| Kruskal-Wallis | 1 | 11.66013 | 0.0006 |
| Kruskal-Wallis (tie-adj.) | 1 | 11.66417 | 0.0006 |
| van der Waerden | 1 | 11.13561 | 0.0008 |

Category Statistics

| SEX | Count | Median | > Overall Median | Mean Rank | Mean Score |
|---|---|---|---|---|---|
| 0 | 105 | 2.302600 | 64 | 117.4095 | 0.222977 |
| 1 | 101 | 2.014900 | 39 | 89.03960 | -0.231803 |
| All | 206 | 2.187100 | 103 | 103.5000 | 2.47E-06 |

图 2-22　中位数齐性检验的输出结果

从检验结果看，如果设定的显著性水平是 0.05，4 种检验方法得到的统计量伴随概率均小于设定的显著性水平，则拒绝原假设，即认为男性和女性收入的中位数有显著差异。

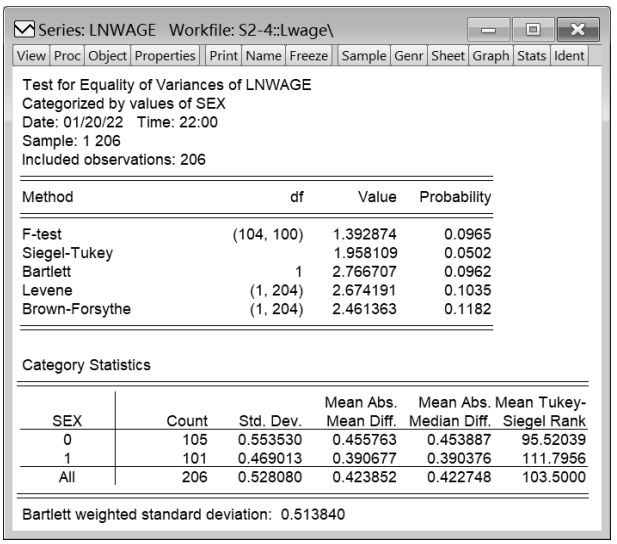
3）方差齐性检验

方差齐性检验（Variance Equality Tests）是对不同分组来自总体的方差之间是否有显著差异的检验。原假设是所有分组数据来自总体的方差之间没有显著差异，备择假设是至少有一个分组数据来自总体的方差具有显著差异。

**【例 2-11】**根据案例 S2-4.WF1 中的数据，以性别（sex）作为收入的分组变量，判断男性和女性的方差是否有显著差异。

在图 2-19 所示的 Test equality of 下面选择 Variance（方差），单击 OK 按钮，得到如图 2-23 所示的输出结果。EViews 软件使用了 5 种检验方法，分别是 F-test、Siegel-Tukey test、Bartlett test、Levene test 和 Brown-Forsythe (modified Levene) test。

从检验结果看，如果设定的显著性水平是 0.05，使用 5 种检验方法得到的统计量伴随概率均大于设定的显著性水平，则不拒绝原假设，即认为男性和女性收入的方差没有显著差异。


Series: LNWAGE　Workfile: S2-4::Lwage\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Test for Equality of Variances of LNWAGE
Categorized by values of SEX
Date: 01/20/22　Time: 22:00
Sample: 1 206
Included observations: 206

| Method | df | Value | Probability |
|---|---|---|---|
| F-test | (104, 100) | 1.392874 | 0.0965 |
| Siegel-Tukey | | 1.958109 | 0.0502 |
| Bartlett | 1 | 2.766707 | 0.0962 |
| Levene | (1, 204) | 2.674191 | 0.1035 |
| Brown-Forsythe | (1, 204) | 2.461363 | 0.1182 |

Category Statistics

| SEX | Count | Std. Dev. | Mean Abs. Mean Diff. | Mean Abs. Median Diff. | Mean Tukey-Siegel Rank |
|---|---|---|---|---|---|
| 0 | 105 | 0.553530 | 0.455763 | 0.453887 | 95.52039 |
| 1 | 101 | 0.469013 | 0.390677 | 0.390376 | 111.7956 |
| All | 206 | 0.528080 | 0.423852 | 0.422748 | 103.5000 |

Bartlett weighted standard deviation: 0.513840

图 2-23　方差齐性检验的输出结果

### 6．Empirical Distribution Tests（经验分布检验）

经验分布检验是检验样本来自的总体符合哪种数据的理论分布形态。原假设为样本来自的总体服从待检验的理论分布，备择假设是样本来自的总体不服从待检验的理论分布。软件提供的方法包括 Kolmogorov-Smirnov、Lilliefors、Cramer-von Mises、Anderson-Darling 和 Watson 经验检验。

打开一个序列，依次选择 View | Descriptive Statistics & Tests | Empirical Distribution Tests，弹出如图 2-24 所示的 EDF Test 对话框。

Test Specification 选项卡的 Distribution 下拉列表框包含 10 种可以进行检验的理论分

布，包括 Normal（正态分布）、Chi-Square（卡方分布）、指数分布（Exponential）、Extreme(Max)（最大极值分布）、Extreme(Min)（最小极值分布）、Gamma（伽马分布）、Logistic（逻辑分布）、Pareto（帕累托分布）、Uniform（均匀分布）和 Weibull（韦伯分布）。Parameters 用于设定检验理论分布的均值和标准差的参数或参数表达式，可以由系统自动设定。Estimation Options 选项卡用于重复渐进估计法的设定，一般使用系统的默认设定。

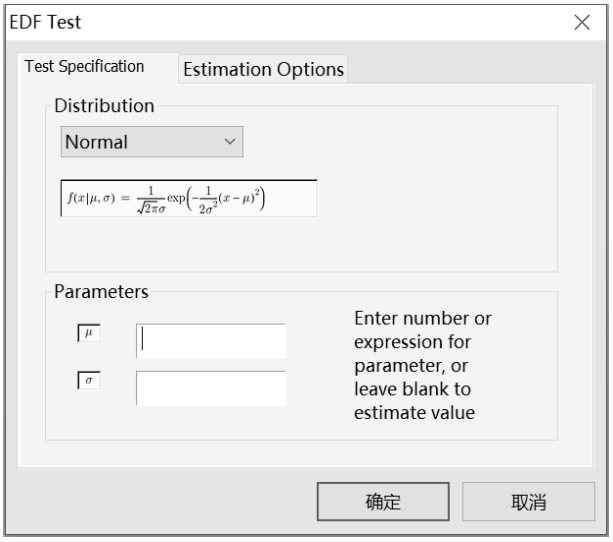



图 2-24　EDF Test 对话框

**【例 2-12】**正态分布是商业数据中最常见的数据分布，工作文件 S2-5.WF1 存储的是一家连锁快餐店的汉堡销售额和广告的数据（见图 2-25）。检验汉堡的销售额（sales）是否呈正态分布。

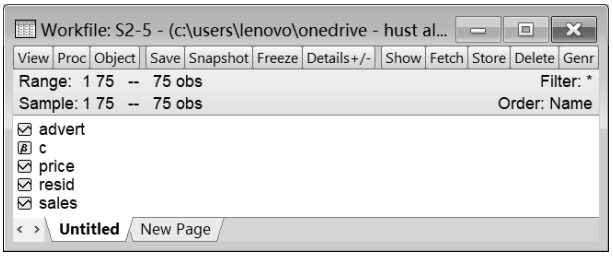



图 2-25　S2-5.WF1 工作文件窗口

本案例同样是从样本推断总体的分布情况，待检验的是该连锁店（或所有连锁店）的汉堡销售额是否符合正态分布。

打开工作文件 S2-5.WF1，接着对序列 sales 进行正态分布的经验分布检验，在图 2-24 所示的 Distribution 下拉列表框中选择默认的理论分布 Normal（正态分布），单击“确定”按钮，得到图 2-26 所示的输出结果。

在 Method 下方，软件显示了 4 种检验方法。从检验结果看，如果设定的显著性水平是 0.05，4 种检验方法得到的统计量伴随概率均大于设定的显著性水平，则不拒绝原假设，即认为该连锁店的汉堡销售额和正态分布没有显著差异。

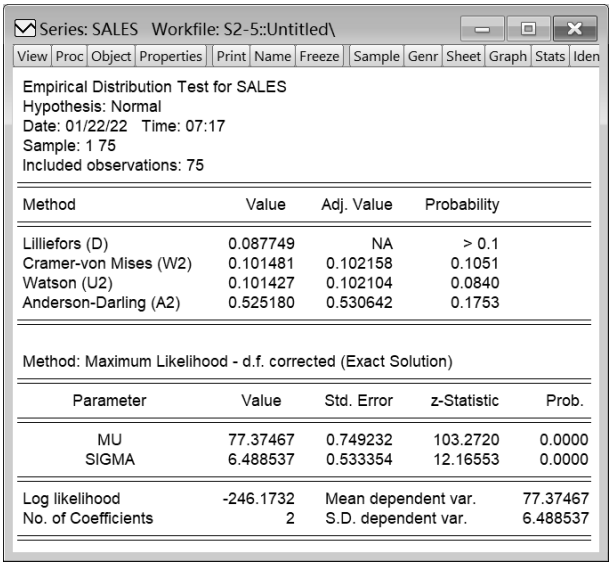



图 2-26　sales 经验分布检验输出结果

## 2.2.2　单因素统计表

单因素统计表（One-way Tabulation）是一个分类统计表，它将序列的值从低到高自动进行分类（点分类或区间分类），然后显示分类的值、百分比、累计值和累计百分比。单因素统计表也称作单因素列联表，单因素统计表分析也可以称作列联表分析。打开一个序列，选择 View | One-Way Tabulation，弹出如图 2-27 所示的 Tabulate Series 对话框。

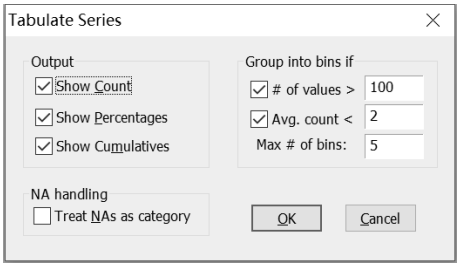



图 2-27　Tabulate Series 对话框

Output 选项组有三个输出结果显示选项，分别是 Show Count（显示观察值计数）、Show Percentages（显示百分比）和 Show Cumulatives（显示累计值和累计百分比）。NA handling 和 Group into bins if 选项参见分组统计量（Stats by Classification）部分的解释。

**【例 2-13】**工作文件 S2-6.WF1 存储的是 1972—2016 年度在 SAT 考试中数学部分的平均成绩（见图 2-28），包括男生平均成绩（male）、女生平均成绩（female）和总平均成绩（total）。要求将总平均成绩进行分类输出。

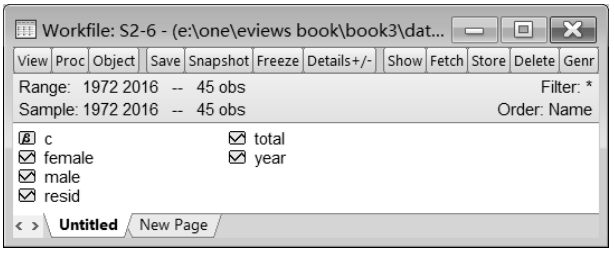



图 2-28　S2-6.WF1 工作文件窗口

打开工作文件 S2-6.WF1，在图 2-28 中对序列 total 进行单因素统计表分析，得到如图 2-29 所示的输出结果。在第 1 列中，软件将数学平均成绩自动分为 4 类，即 490～500 分、500～510 分、510～520 分和 520～530 分；第 2 列是每组的人数，分别为 10、16、18 和 1 人，总人数为 45；第 3 列是每组分类人数占总人数的百分比；第 4 列是每组人数的累计数；第 5 列是每组百分比的累计数。

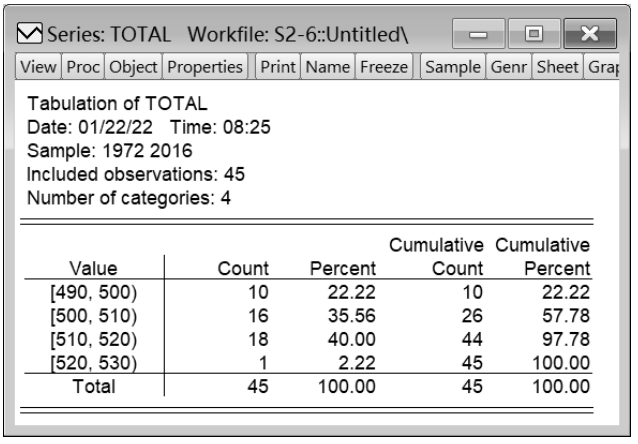



图 2-29　序列 total 的单因素统计表分析结果

## 2.2.3　重复值分析

从 EViews 11 开始，EViews 中就增加了重复值分析（Duplication Observations）功能。它属于数据清洗（Data Cleaning）的范畴，用户对获得的数据进行重新审核、校验，删除重复数据、纠正错误等。

打开一个序列，选择 View | Duplicate Observations，得到重复值分析摘要表，用户可以在摘要表中对重复数据进行分析和编辑等操作。

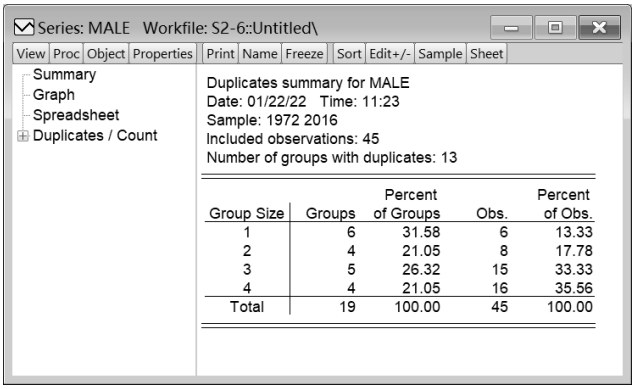
【例 2-14】工作文件 S2-6.WF1 是 1972—2016 年度在 SAT 考试中数学部分的平均成绩，对男生的平均成绩（male）进行重复值分析。

打开工作文件 S2-6.WF1，对序列 male 进行重复值分析，得到图 2-30 所示的重复值分析摘要表。

其中：第 1 列 Group Size 是组的容量；第 2 列 Groups 是同样容量的组的数量；第 3 列 Percent of Groups 是该组占总组数的百分比；第 4 列 Obs.是每个组的观察值数量；第 5 列 Percent of Obs.是该组观察值占总观察值数量的百分比。

本例对序列 male 进行重复值分析，共有 45 个观察值，分为 19 个组。没有出现重复值的有 6 个组，出现重复值的有 13 个组（4+5+4），其中：

- ❑ 包含 1 个值的有 6 个组，组内没有出现重复值，共有 6 个观察值。
- ❑ 包含 2 个值的有 4 个组，各有 2 个重复值，共有 8 个观察值。
- ❑ 包含 3 个值的有 5 个组，各有 3 个重复值，共有 15 个观察值。
- ❑ 包含 4 个值的有 4 个组，各有 4 个重复值，共有 16 个观察值。




| Group Size | Groups | Percent of Groups | Obs. | Percent of Obs. |
|---|---|---|---|---|
| 1 | 6 | 31.58 | 6 | 13.33 |
| 2 | 4 | 21.05 | 8 | 17.78 |
| 3 | 5 | 26.32 | 15 | 33.33 |
| 4 | 4 | 21.05 | 16 | 35.56 |
| Total | 19 | 100.00 | 45 | 100.00 |

图 2-30　male 序列重复数据分析摘要表

选择图 2-30 摘要表左侧的 Graph，可以看到图 2-31 所示的重复值分析结果，从图中可以看出重复值的各组容量（Size）和观察值之间的关系，纵轴代表包含 1～4 个相同观察值的组，横轴代表每个不同容量组的观察值数（竖线的条数），如最长的线条有 16 根，代表有 4 个相同观察值的总数为 16。

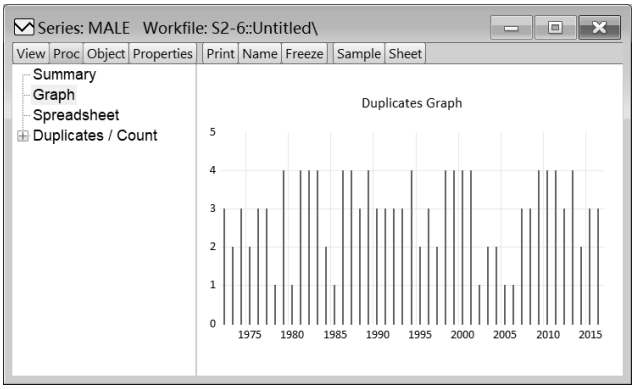



图 2-31　重复数据输出分析结果

选择图 2-30 摘要表左侧的 Spreadsheet，在图 2-32 所示的重复值分析输出表中，所有重复值均以红色和阴影背景显示。

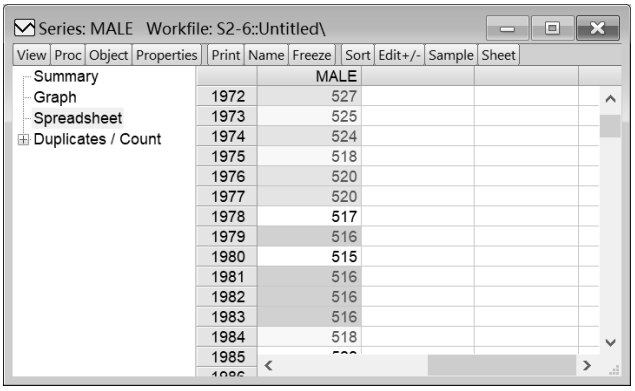



图 2-32　重复数据输出表

选择图 2-30 摘要表左侧的 Duplicates/Count，则会出现如图 2-33 所示的所有重复数据分组的树状图。单击每个组，可以看到分组的观察值详细信息。例如，单击（520）3，说明平均分为 520 的有 3 个年度，分别为 1976 年、1977 年和 1991 年。单击 Edit+/-按钮，可以对图 2-33 所示的数据进行编辑。

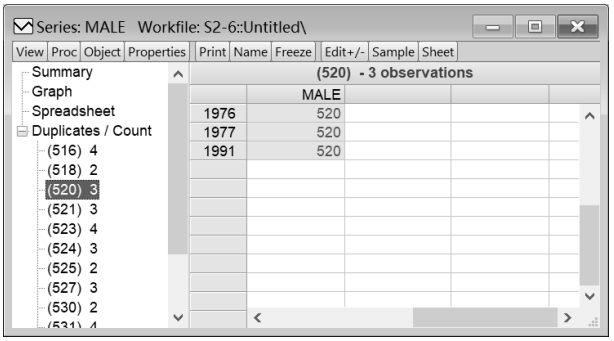



图 2-33　分组重复值的观察值

# 2.3　时间序列分析

时间序列是经济和金融实践中最常见的数据形式，时间序列分析也是 EViews 软件的重要功能。时间序列的基本数据分析主要包括相关图分析、单位根检验和 BDS 独立性检验等。

## 2.3.1　相关图

相关图（Correlogram）用于展示序列的自相关图和偏自相关图，包括原序列和残差序列，分析序列观察值和滞后项之间的关系。相关图分析是时间序列分析和建模的基本工具之一，一般用于以下几个方面：

- ❑ 分析序列是否为白噪声序列。
- ❑ 分析序列是否存在自相关。
- ❑ 分析序列适用于哪种时间序列模型。
- ❑ 分析序列是否存在异方差。

打开一个序列，选择 View | Correlogram，弹出如图 2-34 所示的 Correlogram Specification 对话框。Correlogram of 的选项组用于选择相关图分析的序列类型，Level 表示对原序列进行分析，1st difference 表示分析原序列的 1 阶差分序列，2nd difference 表示分析原序列的 2 阶差分序列。Lags to include 表示进行分析的滞后阶数，一般使用默认值。设置完成后单击 OK 按钮输出结果。

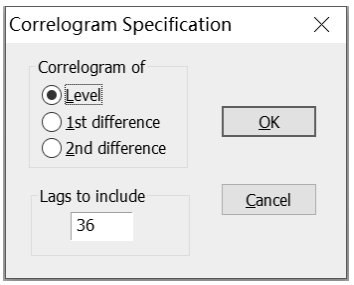



图 2-34　Correlogram Specification 对话框

**【例 2-15】**工作文件 S2-7.WF1 存储的是 1985—2020 年浙江省主要沿海港口货运吞吐量的数据（见图 2-35），分析序列 ttl（货运吞吐量）的相关图。

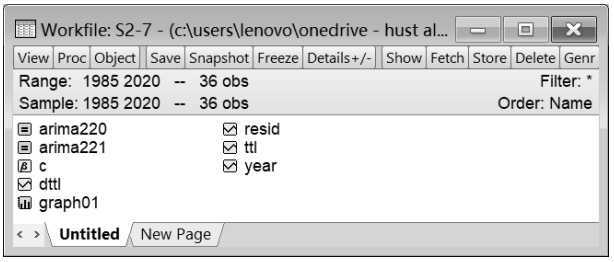



图 2-35　S2-7.WF1 工作文件窗口

打开工作文件 S2-7.WF1，接着打开如图 2-34 所示的相关图对话框，使用默认设置，得到序列 ttl 的相关图（见图 2-36）。其中：第 1 列 Autocorrelation 表示相关图；第 2 列 Partial Correlation 表示偏相关图；第 3 列是自然序列 1、2、3…表示滞后阶数；第 4 列 AC 为自相关系数，是第 1 列的数值；第 5 列 PAC 为偏相关系数，是第 2 列的数值；第 6 列 Q-Stat 是 Q-统计量的值；第 7 列 Prob 是第 6 列 Q-统计量对应的伴随概率 P-值。

对时间序列相关图的分析，在第 3 篇中还会专门学习。从图 2-36 中可以看出，Q-统计量的伴随概率 P-值均接近 0，说明 ttl 是一个非白噪声序列，可以进一步建立模型。

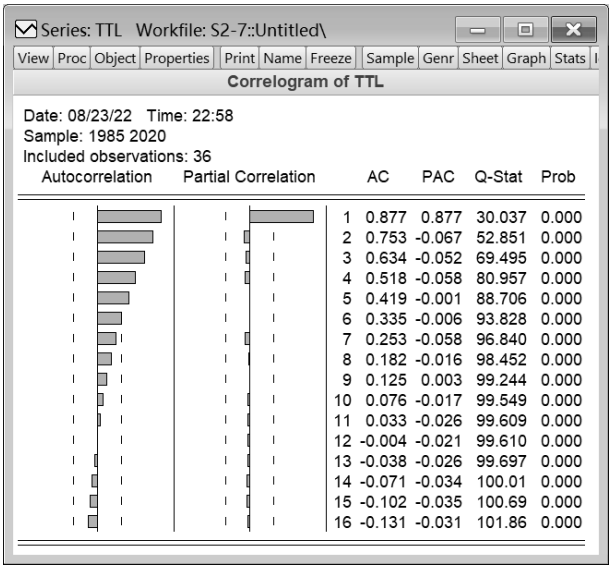



图 2-36　序列 ttl 的相关图

## 2.3.2　长期方差

长期方差（Long-run Variance）指时间序列数据在长期时间内的方差，它是对时间序列数据波动的一个度量，是金融时间序列分析中的一个重要概念。长期方差可以用来检验金融市场的波动性，如果金融市场的长期方差较大，则说明市场波动性较高，投资风险也

较高。长期方差也是风险模型中的一个重要参数，用于衡量投资组合的风险。在风险模型中，通常使用历史数据计算长期方差，并以此预测未来的波动性和风险。基于长期方差可以建立一些金融交易策略。例如，可以采用波动率策略，即在市场波动率较低时增加仓位，在市场波动率较高时减少仓位，以此获得更好的收益。长期方差还可以用于评估金融市场政策的效果，如政策调整是否能够降低市场波动性，从而提高市场效率和稳定性。长期方差分析方法也广泛应用于气象、环境、能源等领域的数据分析。

在 EViews 中打开需要分析的序列，选择 View | Long-run Variance，弹出 Long-run Variance 对话框（见图 2-37），在对话框中进行设置或者使用默认值即可。

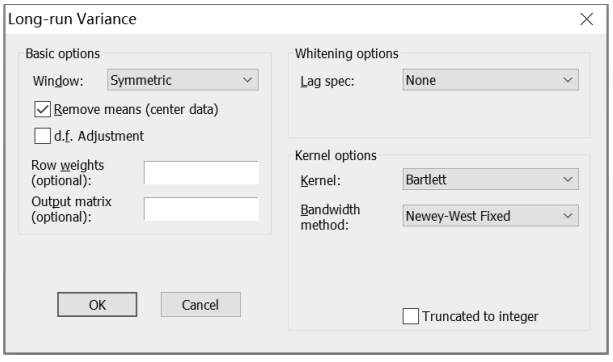



图 2-37　Long-run Variance 对话框

## 2.3.3　单位根检验

从 EViews 11 开始，EViews 在单位根检验（Unit Root Test）中设置了 3 个子菜单，分别是 Standard Unit Root Test（标准单位根检验）、Breakpoint Unit Root Test（断点单位根检验）和 Seasonal Unit Root Test（季节单位根检验）。如果没有特别说明，单位根检验一般指标准单位根检验。

单位根检验用于检验时间序列的平稳性（Stationary），是时间序列分析基本工具之一。例如，ARMA 模型估计就是基于序列的平稳性，当我们说一个序列平稳的时候，就意味着序列的均值和自协方差与时间参数是无关的。

单位根检验的原假设是序列（至少）存在一个单位根，序列非平稳；备择假设是序列不存在单位根，序列平稳。

在 EViews 中打开一个序列，依次选择 View | Unit Root Tests | Standard Unit Root Test，弹出如图 2-38 所示的 Unit Root Test 对话框。其中，Test type 用于选择单位根检验的方法，包括 ADF 检验、DF 检验、PP 检验等。ADF 检验是默认选项，也是应用最多的单位根检验方法；DF 检验一般只用于滞后 1 阶的情况；PP 检验一般应用于残差序列存在异方差的情况。

Test for unit root in 选项组用于选择检验的序列，Level 表示原序列，1st difference 表示 1 阶差分后的序列，2nd difference 表示 2 阶差分后的序列。

Include in test equation 选项组用于设置时间序列的形式，即对应序列的 3 种形式：包含截距项（Intercept）、包含趋势项和截距项（Trend and intercept），以及无截距项和无趋势

项（None）。一般情况下，对时间序列的单位根检验，这 3 种情形都需要检验。

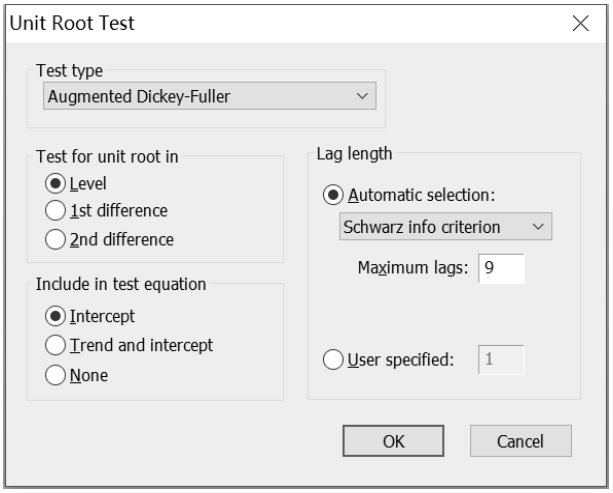



图 2-38　Unit Root Test 对话框

Lag length 选项组用于设置最优滞后期的方法，一般使用默认选项 Automatic selection（自动选择）。单击 OK 按钮输出结果。

**【例 2-16】**工作文件 S2-7.WF1 存储的是 1985—2020 年浙江省主要沿海港口货运吞吐量的数据（见图 2-35），对序列 ttl（货运吞吐量）进行标准单位根检验。

打开工作文件 S2-7.WF1，对序列 ttl 进行单位根检验。观察 ttl 的序列图，发现带有截距项和明显的时间趋势。对序列 ttl 的原序列和 1 阶差分后的序列进行标准单位根检验，发现均为非平稳序列。因此，Test for unit root in 选择检验 2 阶差分后的序列，Include in test equation 选择 Trend and intercept。检验后，得到图 2-39 和图 2-40 所示的 ADF 单位根检验的输出结果。

如图 2-39 所示，运用 ADF 单位根检验方法对序列 D(TTL，2)，即序列 ttl 的 2 阶差分序列进行单位根检验，原假设（Null Hypothesis）有一个单位根；外生变量（Exogenous）包含常量（Constant）和线性趋势（Linear Trend）；软件给出的最优滞后期为 1；t-统计量的值为−7.425728，它的伴随概率接近 0。

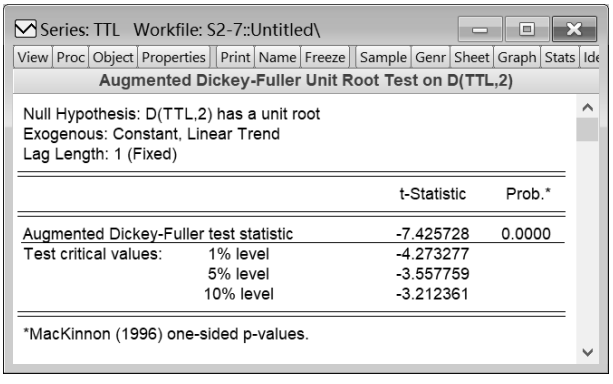



图 2-39　Unit Root Test 输出结果 1

从检验结果中看，如果设定的显著性水平是 0.05，ADF 单位根检验得到的统计量的伴随概率，远远小于设定的显著性水平，则拒绝原假设，可以认为序列 ttl 的 2 阶差分序列是一个平稳的时间序列。

图 2-40 显示的是在标准单位根检验中，ADF 检验计算 t-统计量的过程。在本例中，序列 D(TTL,2)的值作为因变量，检验的方法是最小二乘法，调整后有 29 个观察值。

在标准单位根检验中，如果选择其他的检验方法，输出结果是不同的，这里不再演示。

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TTL,2)
Method: Least Squares
Date: 01/25/22 Time: 15:27
Sample (adjusted): 1992 2020
Included observations: 29 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| D(TTL(-1)) | -0.510092 | 0.201791 | -2.527830 | 0.0196 |
| D(TTL(-1),2) | 0.266031 | 0.294612 | 0.902989 | 0.3768 |
| D(TTL(-2),2) | -0.440347 | 0.345129 | -1.275889 | 0.2159 |
| D(TTL(-3),2) | 0.908470 | 0.374058 | 2.428688 | 0.0242 |
| D(TTL(-4),2) | 0.148110 | 0.373771 | 0.396260 | 0.6959 |
| D(TTL(-5),2) | 1.044355 | 0.447829 | 2.332041 | 0.0297 |
| C | -8063.725 | 5362.938 | -1.503602 | 0.1476 |
| @TREND("1985") | 783.6246 | 361.2866 | 2.168983 | 0.0417 |

| | | | |
|---|---|---|---|
| R-squared | 0.592925 | Mean dependent var | 1596.769 |
| Adjusted R-squared | 0.457234 | S.D. dependent var | 10173.27 |
| S.E. of regression | 7494.918 | Akaike info criterion | 20.91079 |
| Sum squared resid | 1.18E+09 | Schwarz criterion | 21.28797 |
| Log likelihood | -295.2064 | Hannan-Quinn criter. | 21.02892 |
| F-statistic | 4.369656 | Durbin-Watson stat | 2.051392 |
| Prob(F-statistic) | 0.003948 | | |

图 2-40　Unit Root Test 输出结果 2

## 2.3.4　断点单位根检验

断点单位根检验（Breakpoint Unit Root Test）是在 EViews 11 之后增加的功能。时间序列数据随着时间的推进，经常会出现结构性变化，这种结构性变化同时会引发序列不平稳问题。断点单位根检验的原理，是利用统计学方法检测时间序列数据是否存在一个“断点”，即数据结构从一种非线性结构变为另一种非线性结构的时刻。如果数据存在这样的 “断点”，则说明数据存在非线性结构，而不存在单位根。具体使用两个步骤来实现：第一步是对时间序列进行 1 阶差分（或者非 1 阶差分），以检测自相关；第二步是在具有自相关的序列上进行断点检验，以检测单位根是否存在。在 EViews 中打开需要检验的序列，依次选择 View | Unit Root Tests | Breakpoint Unit Root Test...，弹出如图 2-41 所示的 Breakpoint Unit Root Test 对话框。

其中，Trend specification 选项组的 Basic 有 Intercept 和 Trend and intercept 两个选项，用来设置检验中是否包含趋势项。如果选择 Trend and intercept，则还需要设置 Breaking 选项，在 Intercept、Trend and intercept 和 Trend3 个选项中指出断点的形式。

Break type 选项用于设置断点的形式，包含 Innovation Outlier（新息异常值）和 Additive Outlier（附加异常值）两个异常值选项。Breakpoint selection 选项用于设置检验方法的断点。Lag length 选项用于设置滞后期的判断方法。

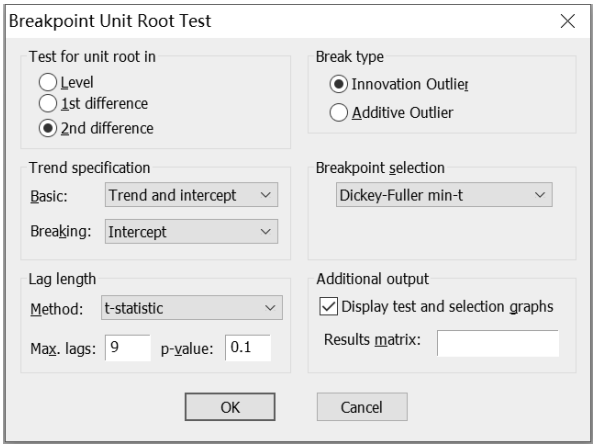



图 2-41　Breakpoint Unit Root Test 对话框

**【例 2-17】**工作文件 S2-7.WF1 存储的是 1985—2020 年浙江省主要沿海港口货运吞吐量的数据（见图 2-35），对序列 ttl（货运吞吐量）进行断点单位根检验。

打开工作文件 S2-7.WF1，接着打开序列 ttl 进行断点单位根检验，在对话框的 Test for unit root in 下方选择 2nd difference（2 阶差分），Trend specification 下方的 Basic 下拉列表框中选择 Trend and intercept，在 Breaking 下拉列表框中选择 Intercept，在 Lag length 的 Method 下拉列表框中选择 t-statistic（见图 2-41）。Trend specification 的选项不同，输出结果也不同。其他选项采用默认值，单击 OK 按钮，得到如图 2-42、图 2-43 和图 2-44 所示的输出结果。

图 2-42 上方是检验的基本信息，原假设有一个单位根，检验包含趋势项，断点的确定性成分是截距项，断点形式是 Innovational outlier（新息异常值）。图 2-42 的中间部分是 ADF 单位根 t-检验，检验的断点时间是 1996 年，检验结果显示，t-统计量的值为−7.171094，伴随概率小于 0.01，在显著性水平为 0.05（或为 0.01）的水平下拒绝原假设，说明不存在断点单位根。

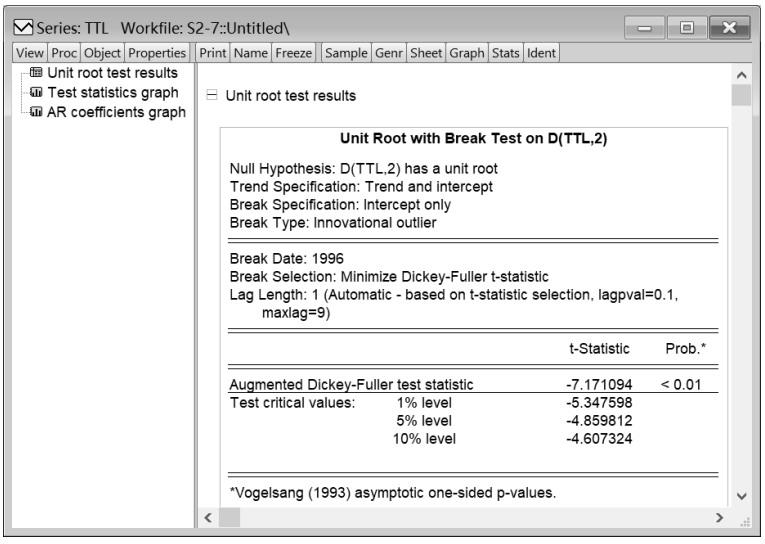



图 2-42　断点单位根检验结果 1

选择图 2-42 左栏的 Test statistics graph 和 AR coefficients graph，得到对每个时间点检验的 ADF 统计量图（见图 2-43）及自回归（AR）系数图（见图 2-44）。

在本例中继续对序列 ttl 进行不包含时间趋势的检验，并采用了多种检验方法，虽然检验的断点不同，但是最终结果均显示序列不存在断点单位根。

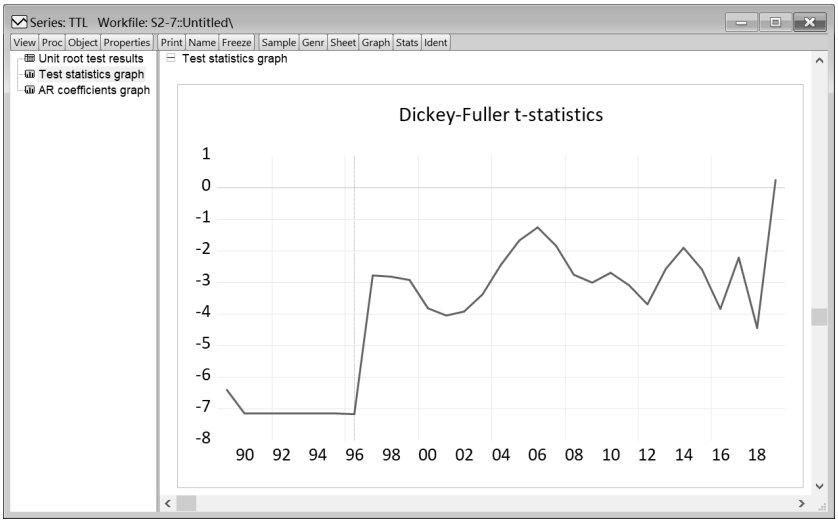



图 2-43　断点单位根检验结果 2

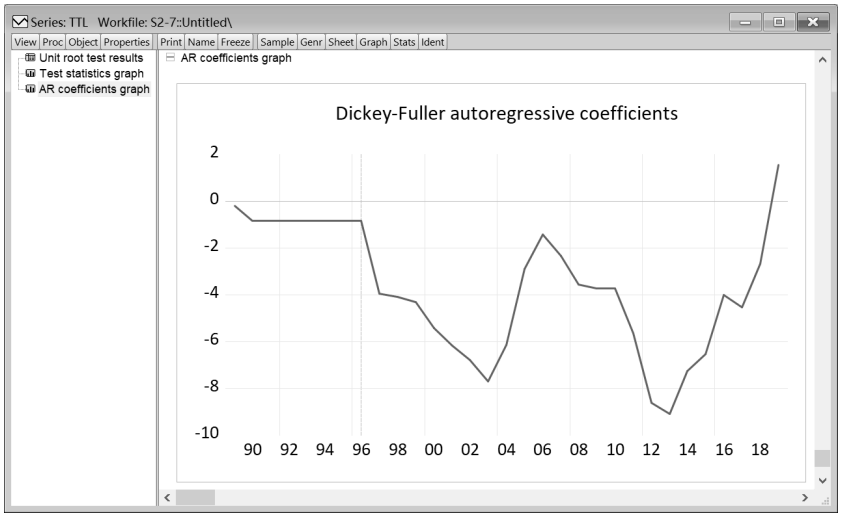



图 2-44　断点单位根检验结果 3

## 2.3.5　季节单位根检验

从 EViews 11 开始，就可以在 EViews 中进行季节单位根检验（Seasonal Unit Root Tests）了。季节性或周期性是很多时间序列的基本特征。在传统时间序列模型中，大部分带季节性的时间序列都被看作平稳时间序列。但是季节性单位根的存在，会给时间序列模型带来

较大的误差。

在 EViews 中进行季节单位根检验的步骤是，打开待检验的序列，依次选择 View | Unit Root ests | Seasonal Unit Root Tests，弹出季节单位根检验对话框，如图 2-45 所示。在 Test type 下拉列表框中有 4 种检验方法，分别是 Traditional HEGY（Hylleberg，Engle，Granger and Yoo）、HEGY Likelihood Ratio test、Canova-Hansen test 和 Variance ratio test。

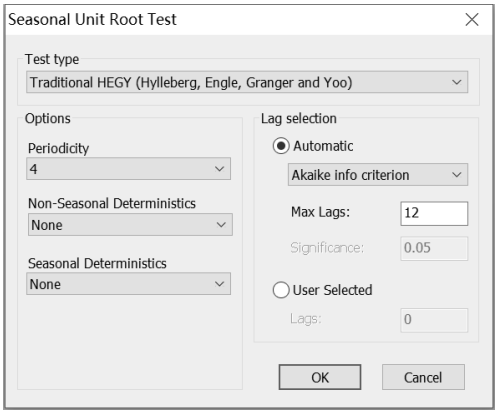



图 2-45　Seasonal Unit Root Test 对话框

在 Options 选项组中，Periodicity 用于选择季节频率，包括 2、4、5、6、7、12 这几个选项；Non-Seasonal Deterministics（非季节性确定因素）选项用于选择是否有确定性因素或时间趋势因素，可以选择 None、Constant 或 Constand and Trend；Seasonal Deterministics（季节性确定因素）选项可以选择 None、Seasonal Dummies（季节虚拟变量）、Spectral Intercepts（虚拟截距）或 Spectral Intercepts and Trends（虚拟截距和趋势）。Lag selection 用于选择滞后期的方法。

**【例 2-18】**工作文件 S2-7.WF1 存储的是 1985—2020 年浙江省主要沿海港口货运吞吐量的数据（见图 2-35），对序列 ttl（货运吞吐量）进行季节单位根检验。

在 S2-7.WF1 工作文件窗口，打开序列 ttl 进行季节单位根检验，在图 2-45 中采用软件的默认值，检验方法 Test type 采用传统的 HEGY 方法，单击 OK 按钮，得到如图 2-46、图 2-47 和图 2-48 所示的季节单位根检验结果。

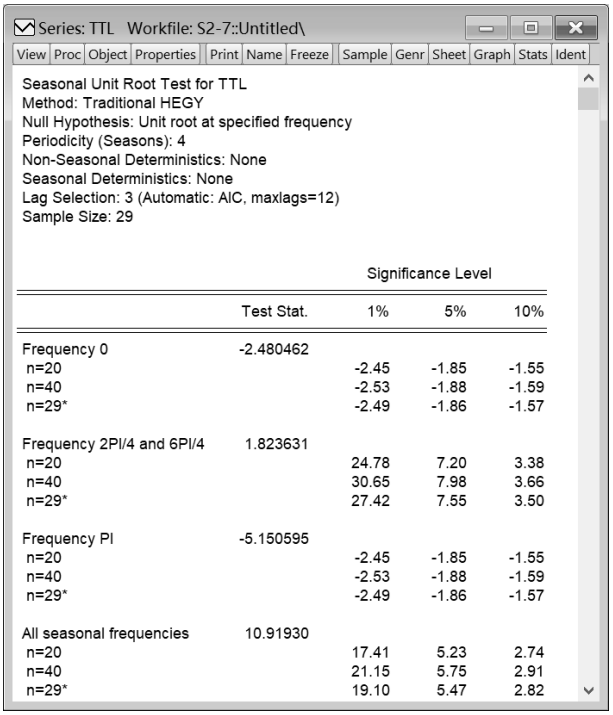



图 2-46　季节单位根检验结果 1

图 2-46 显示的是不同频率情况下的检验结果。Frequency 是频率，对于所有的检验统计量，都给出了临界值在 1%、5%和 10%的水平。因为所有的临界值都是在样本量基础上增加约 20 个值进行估计的，本例有 29 个样本，所以对 29 上下浮动 10 个值左右，即对 20、40 和 29 个样本量都进行估计。

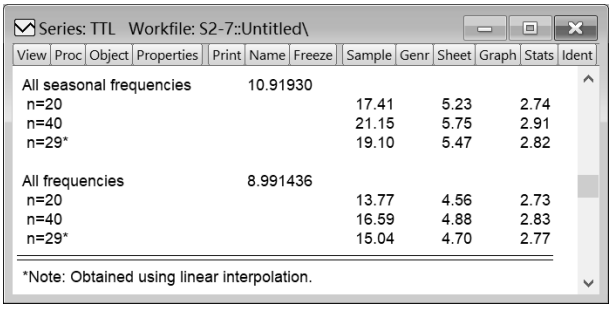



图 2-47　季节单位根检验结果 2

图 2-47 显示的是所有季节性频率和所有频率情况下的检验结果。原假设是在指定的频率下存在单位根，备择假设相反。从图 2-44 和图 2-45 中可以看到，以传统的 HEGY 方法进行检验，序列 ttld 在 0.01、0.05 和 0.1 的显著性水平下，存在季节单位根。

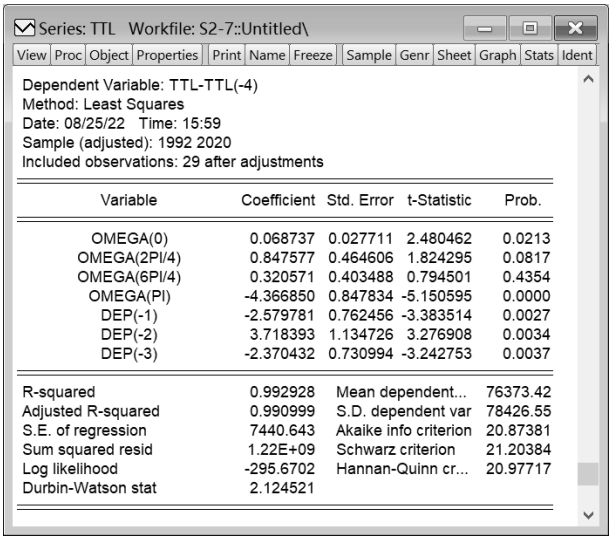



图 2-48　季节单位根检验结果 3

图 2-48 显示的是传统的 HEGY 检验的回归输出结果，以及整个季节单位根检验的统计量。

## 2.3.6　方差比率检验

方差比率检验（Variance Ratio Test）是一种用于检验时间序列数据是否存在随机游走（Random Walk）行为的统计方法。它的主要作用是判断一个时间序列是不是随机游走过程，即序列中的未来值与当前值之间是否存在一定的可预测关系。如果时间序列是随机游走过

程，那么其未来值与当前值之间的关系将是随机的，即无法通过历史数据对未来数据进行预测。在金融领域中，随机游走假说是一种重要的假设，它认为股票价格或汇率等金融资产的价格走势是随机游走过程，随机游走模型下的金融资产收益率将是不可预测的。方差比率检验可以用来检验这个假设是否成立，从而判断股票价格或汇率等金融资产的价格是否具有可预测性。除了金融领域，方差比率检验还可以应用于其他领域，如经济学、气象学和医学等，用于检验时间序列数据是否具有可预测性。

经济学家安德鲁·W.罗（Andrew W.Lo）和艾·克雷格·麦金雷（A.Craig MacKinlay）在 1988 年提出了时间叠加方差比率检验。它的原理是通过比较在不同时间区间内数据的方差差异，来研究时间序列数据的可预测性。如果股价的自然对数遵循随机游走过程，则其方差与时间成正比。具体地说，如果股价的自然对数满足随机游走假设，即股价的变动是一个随机游走过程，且未来的价格变动无法被预测，那么此时的方差会随着时间的增加而增加。

在 EViews 中打开时间序列，依次选择 View | Variance Ratio Test，弹出如图 2-49 所示的 Variance Ratio Test 对话框。

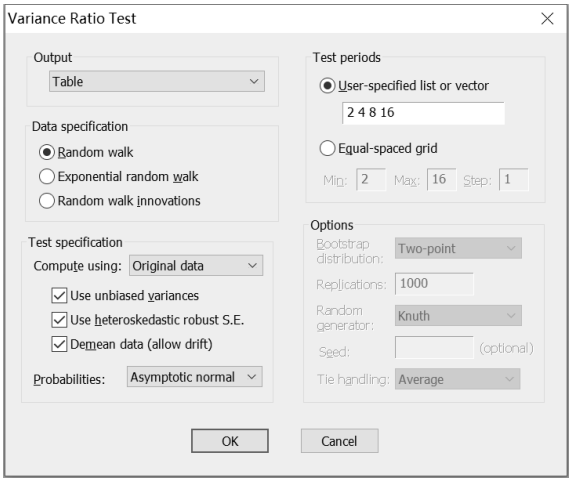



图 2-49　Variance Ratio Test 对话框

Output 选项用于检验的输出形式，有 Table（表格）和 Graph（图形）两个选项。Data specification 选项用于描述时间序列数据的属性，如果选择默认的 Random walk，则 EViews 假定时间序列的数据服从 Random walk，方差会以差分的形式进行估计；如果选择 Exponential random walk（指数随机游走），则数据以自然对数的形式进行估计；最后一个选项表示假定数据服从 Random walk innovations（随机游走创新）。

在 Test specification 选项组内，Compute using 用于设置估计的方法，默认选项是 Original data，即 Lo 和 MacKinlay 的方差比率方法。其他的估计方法包括 Ranks（排序检验）、Rank scores（van der Waerden 分数排序检验）和 Signs（符号检验）。下方复选框是对使用方差的定义，包括 Use unbiased variances（使用无偏方差）、Use heteroskedastic robust S.E.（使用异方差稳健标准误）和 Demean data（allow drift）（零均值化）。

Test periods 选项组用于设置方差比较的时间区间，可以设置一个时间区间，也可以设置多个时间区间。

**【例 2-19】**工作文件 S2-8.WF1 存储的是 1974 年 8 月至 1996 年 5 月期间，加拿大（can）、德国（deu）、法国（fra）、日本（jp）、英国（uk）的货币和美元的汇率，为每周收盘数据（见图 2-50）。检验它们的汇率是否服从指数随机游走（Exponential random walk）。

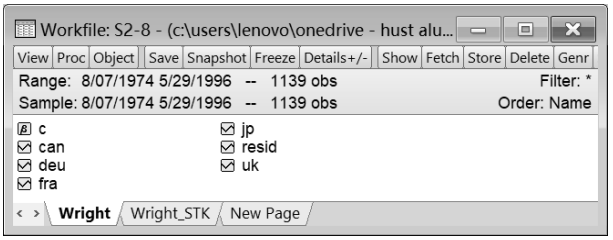



图 2-50　S2-8.WF1 的工作文件窗口

数据来源于乔纳森 • 赖特（Wright Jonathan H.）于 2000 年在《商业与经济统计杂志》（*Journal of Business and Economic Statistics*）上发表的一篇关于方差比率检验的论文中的案例。

这里以日元汇率为例进行检验，在 S2-8.WF1 工作文件中打开序列 jp，弹出 Variance Ratio Test 对话框。Data specification 选择 Exponential random walk，对序列 jp 的自然对数序列进行检验；Test specification 的估计方法选择 Original data，复选框选择 Demean data(allow drift)；Test periods 按照 WRIGHT 在论文中的设置“2 5 10 30”。设置结束后如图 2-51 所示，单击 OK 按钮，得到如图 2-52 和图 2-53 所示的检验结果。

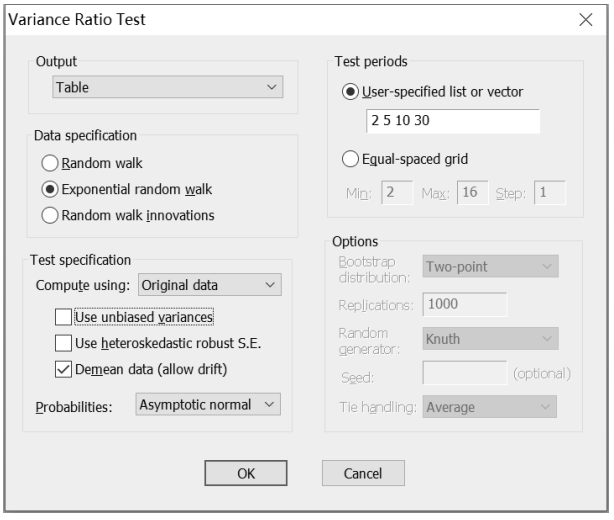
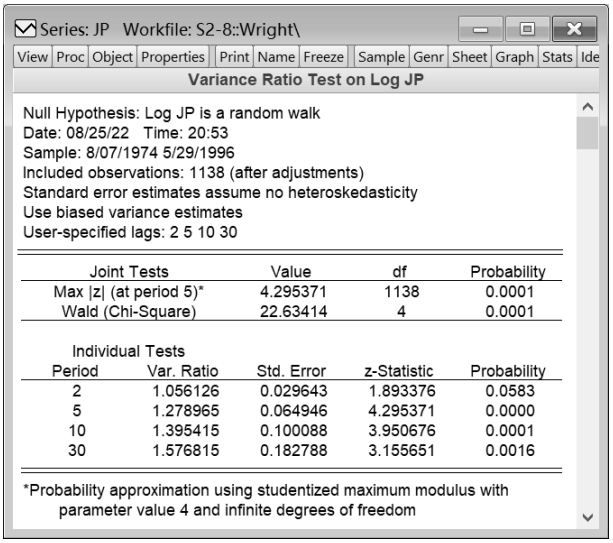



图 2-51　Variance Ratio Test 对话框

在图 2-52 中，原假设是序列 jp 的自然对数服从随机游走。检验结果中的 Joint Tests 是对所有时间区间的综合检验结果，Chow-Denning Max |z|统计量的值是 4.295371，伴随概率为 0.0001，强烈拒绝原假设，说明序列 jp 的对数序列不服从随机游走。Individual Tests 是每个时间区间的检验结果，除了区间“2”的方差比率的统计量的伴随概率略大于 0.05 以外，其余区间的检验均拒绝原假设。

图 2-53 是每个区间的方差比率检验的中间结果，从第 2 列开始，依次为方差的均值、

每个区间的方差、每个区间的观察值。


Series: JP　Workfile: S2-8::Wright\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ide

**Variance Ratio Test on Log JP**

Null Hypothesis: Log JP is a random walk
Date: 08/25/22　Time: 20:53
Sample: 8/07/1974 5/29/1996
Included observations: 1138 (after adjustments)
Standard error estimates assume no heteroskedasticity
Use biased variance estimates
User-specified lags: 2 5 10 30

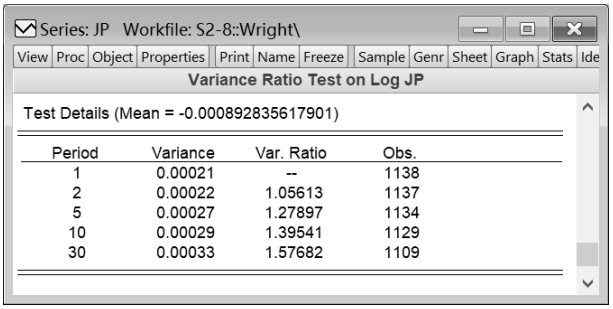
| Joint Tests | Value | df | Probability |
|---|---|---|---|
| Max \|z\| (at period 5)* | 4.295371 | 1138 | 0.0001 |
| Wald (Chi-Square) | 22.63414 | 4 | 0.0001 |

Individual Tests

| Period | Var. Ratio | Std. Error | z-Statistic | Probability |
|---|---|---|---|---|
| 2 | 1.056126 | 0.029643 | 1.893376 | 0.0583 |
| 5 | 1.278965 | 0.064946 | 4.295371 | 0.0000 |
| 10 | 1.395415 | 0.100088 | 3.950676 | 0.0001 |
| 30 | 1.576815 | 0.182788 | 3.155651 | 0.0016 |

*Probability approximation using studentized maximum modulus with parameter value 4 and infinite degrees of freedom

图 2-52　Variance Ratio Test 结果 1


Series: JP　Workfile: S2-8::Wright\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ide

**Variance Ratio Test on Log JP**

Test Details (Mean = -0.000892835617901)

| Period | Variance | Var. Ratio | Obs. |
|---|---|---|---|
| 1 | 0.00021 | -- | 1138 |
| 2 | 0.00022 | 1.05613 | 1137 |
| 5 | 0.00027 | 1.27897 | 1134 |
| 10 | 0.00029 | 1.39541 | 1129 |
| 30 | 0.00033 | 1.57682 | 1109 |

图 2-53　Variance Ratio Test 结果 2

## 2.3.7　BDS 独立性检验

BDS 独立性检验（BDS Independence Test）用来检验基于时间变化的序列是否独立分布，主要用于检验序列值独立分布的偏差，包括它们的线性相关、非线性相关或者混乱状态。BDS 独立性检验可以运用于对序列残差的检验，检验残差是独立分布还是均匀分布的（Independent and Identically Distributed，IID）。例如，可以测试一个拟合后的 ARMA 模型的残差是否存在非线性独立分布。在 EViews 中打开需要进行检验的序列，选择 View | BDS Independence Test，在弹出的对话框中进行相应设置即可，不再赘述。

## 2.3.8　预测效果评估

建立经济计量模型是为了对变量进行预测，不同模型提供的预测结果不尽相同，EViews 提供了对单变量预测效果评估（Forecast Evaluation）的多种方法，由使用者决定选

择哪种预测方法。

EViews 提供了 4 种预测精度的评估方法，分别是根均方误差（Root Mean Squared Error，RMSE）、平均绝对误差（Mean Absolute Error，MAE）、平均绝对百分比误差（Mean Absolute Percentage Error，MAPE）和泰尔不等式系数（Theil Inequality Coefficient）。

打开一个序列，选择 View | Forecast Evaluation，弹出如图 2-54 所示的预测效果评估对话框。

在 Forecast data objects（预测数据目标）下方的文本框中直接输入多个序列、序列组、模型的名称或者方程式对象的名称，软件会自动对数据进行预测并给出一个预测结果。Evaluation sample（评估样本）下方的文本框用于设置评估的数据范围。Averaging methods（optional）用于设置求平均值的方法，可以多选，选项包括 Simple mean、Trimmed mean、Simple median、Least-squares、Mean square error 和 MSE ranks，如果输入的是方程式，那么还会出现 Smooth AIC weights 和 SIC weights 两个选项。

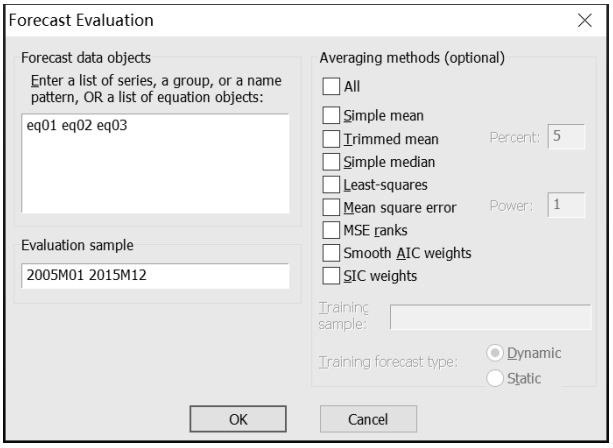



图 2-54　Forecast Evaluation 对话框

**【例 2-20】**工作文件 S2-9.WF1 存储的是有关英格兰和威尔士电力需求的数据（见图 2-55），其中包含 2005 年 1 月至 2015 年 12 月的月度数据样本（序列 elecdmd），还包含 5 个样本内的预测序列（elecf_fe1 至 elecf_fe5），5 个样本外预测序列（elecf_ff1- elecf_ff5）。其中，每个样本内预测序列均包含 2011 年 12 月之前的电力需求的真实数据，并对 2012 年 1 月至 2013 年 12 月的电力需求进行了预测。要求对 5 个预测模型的精度进行检验。

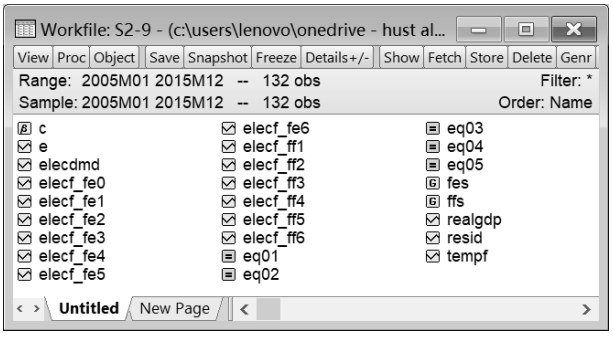



图 2-55　S2-9.WF1 工作文件窗口

打开工作文件 S2-9.WF1 中的序列 elecdmd，接着打开预测效果评估对话框，如图 2-56 所示。在 Forecast data objects 下方输入 elecf_fe1 elecf_fe2 elecf_fe3 elecf_fe4 elecf_fe5；Evaluation sample 设置为 2013 年 1 月至 12 月；Averaging methods(optional)的选项全部打勾；Training sample（训练样本）设置为 2012 年 1 月至 12 月。

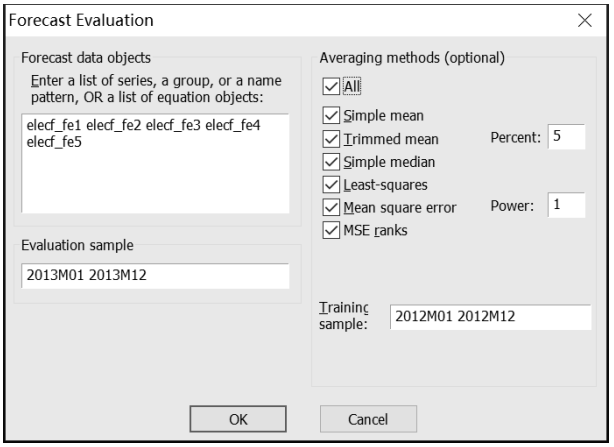



图 2-56　Forecast Evaluation 对话框

单击 OK 按钮，得到如图 2-57 和图 2-58 所示的检验输出结果。图 2-57 的上半部分是汇总信息，下半部分的 Combination tests 展示了 5 个预测序列的检验结果。原假设是在 0.05 的显著性水平下，一个预测包含相关信息的全部内容，即预测值与实际值之间没有显著差异，或者说预测模型的预测能力较差，无法对实际值进行准确的预测。在图 2-57 所示的 F-统计量的伴随概率中，只有第 1 个序列 elecf_fe1 的伴随概率是 0.0495，小于 0.05，拒绝原假设，其余 4 个序列的伴随概率均大于 0.05，不拒绝原假设。

Forecast Evaluation
Date: 03/09/22　Time: 21:02
Sample: 2013M01 2013M12
Included observations: 12
Evaluation sample: 2013M01 2013M12
Training sample: 2012M01 2012M12
Number of forecasts: 10

Combination tests
Null hypothesis: Forecast i includes all information contained in others

| Forecast | F-stat | F-prob |
|---|---|---|
| ELECF_FE1 | 4.138355 | 0.0495 |
| ELECF_FE2 | 1.443315 | 0.3146 |
| ELECF_FE3 | 1.533069 | 0.2913 |
| ELECF_FE4 | 1.558450 | 0.2851 |
| ELECF_FE5 | 0.420428 | 0.7898 |

图 2-57　Forecast Evaluation 结果 1

在输出结果中还显示了评估统计量的值，从图 2-58 中可以看出，因为数据量不足，Trimmed mean 统计量被排除在外，只保留了 RMSE、MAE、MAPE 和 Theil 统计量的值。可以看出，在 5 种求平均值的方法中，Mean square error 是最优的评估方法（阴影部分）。

Evaluation statistics

| Forecast | RMSE | MAE | MAPE | SMAPE | Theil U1 | Theil U2 |
|---|---|---|---|---|---|---|
| ELECF_FE1 | 329.3182 | 299.3388 | 6.582262 | 6.332612 | 0.034294 | 0.969059 |
| ELECF_FE2 | 108.1292 | 94.06769 | 2.046802 | 2.034273 | 0.011568 | 0.307378 |
| ELECF_FE3 | 219.1630 | 193.9669 | 4.293012 | 4.418975 | 0.024040 | 0.674861 |
| ELECF_FE4 | 154.1023 | 126.3301 | 2.793056 | 2.853796 | 0.016757 | 0.479016 |
| ELECF_FE5 | 145.5928 | 123.5517 | 2.692389 | 2.643105 | 0.015454 | 0.414718 |
| Simple mean | 103.3829 | 89.38450 | 1.940429 | 1.930543 | 0.011073 | 0.297969 |
| Simple median | 108.1292 | 94.06769 | 2.046802 | 2.034273 | 0.011568 | 0.307378 |
| Least-squares | 862.9722 | 858.1411 | 18.75875 | 17.11785 | 0.084943 | 2.486276 |
| Mean square error | 98.15188 | 80.04863 | 1.759876 | 1.764384 | 0.010547 | 0.294911 |
| MSE ranks | 99.04570 | 84.72580 | 1.855211 | 1.852648 | 0.010624 | 0.291829 |

*Trimmed mean could not be calculated due to insufficient data

图 2-58　Forecast Evaluation 结果 2

### 2.3.9　小波分析

小波分析（Wavelet Analysis）是近年来应用数学发展很快的一个领域，以前需要通过编程才能实现小波分析。从 EViews 12 开始，可以通过菜单直接进行小波分析，这是一款非常实用的分析工具。

经济和金融时间序列的统计特征，一般会随着时间的变化而变化，如非平稳性、波动性、季节性和结构不连续性等。小波分析是根据时间序列的自然特征进行分析，而不是通过传统的对序列进行简化方法进行分析。小波滤波器还可以分解和重构时间序列，以及进行跨时间尺度的相关结构分析。

打开一个序列，选择 View | Wavelet Analysis，出现 4 个选项，如图 2-59 所示，这就是 EViews 软件进行小波分析的四个方面。

- Transforms：小波变换。
- Variance Decomposition：方差分解。
- Outlier Detection：异常值检验。
- Thresholding（Denoising）：阈值（去噪）。

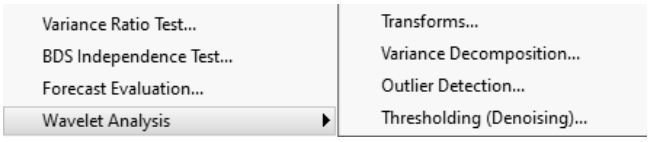



图 2-59　Wavelet Analysis 选项

## 2.4　标　　签

标签（Label）用于展示序列对象的一些基本描述性特征，如 Name（名称）、Display Name（展示的名称）、Last Update（上一次升级的时间）、Description（描述）和 Remarks（评论）。以例 2-20 为例，打开序列 elecdmd，选择 View/Label，弹出如图 2-60 所示的标签对话框，其中，Name 是序列名称，Description 可以对序列进行描述性备注。

在标签内填写必要的信息是一个好习惯，平时在学习和工作中会处理大量的数据，标

签内的信息可以有效地解决这个问题。

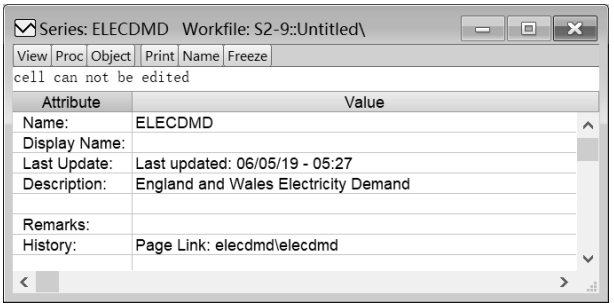
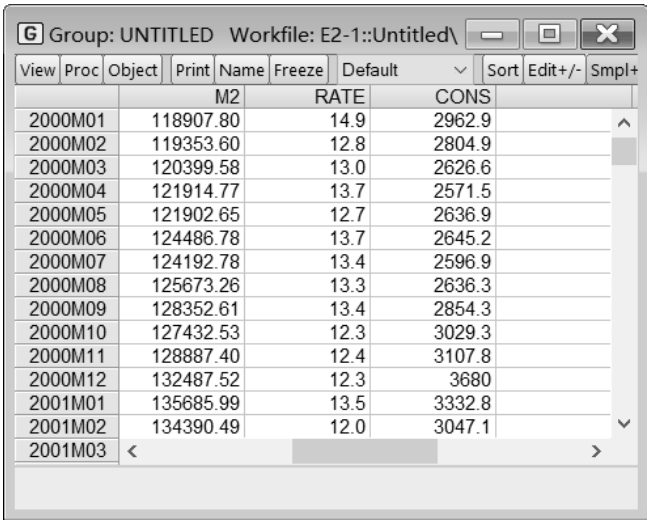



图 2-60　标签窗口

## 2.5　上 机 练 习

1．工作文件 E2-1.WF1（见图 2-61）存储的是 2000 年 1 月至 2022 年 7 月中国货币供应量和社会消费品零售总额的月度数据，其中，序列 M2 是中国货币和准货币的供应量(亿元)，序列 RATE 是 M2 供应量的同比增长值（%），序列 CONS 是中国社会消费品零售总额（亿元）。




| | M2 | RATE | CONS |
|---|---|---|---|
| 2000M01 | 118907.80 | 14.9 | 2962.9 |
| 2000M02 | 119353.60 | 12.8 | 2804.9 |
| 2000M03 | 120399.58 | 13.0 | 2626.6 |
| 2000M04 | 121914.77 | 13.7 | 2571.5 |
| 2000M05 | 121902.65 | 12.7 | 2636.9 |
| 2000M06 | 124486.78 | 13.7 | 2645.2 |
| 2000M07 | 124192.78 | 13.4 | 2596.9 |
| 2000M08 | 125673.26 | 13.3 | 2636.3 |
| 2000M09 | 128352.61 | 13.4 | 2854.3 |
| 2000M10 | 127432.53 | 12.3 | 3029.3 |
| 2000M11 | 128887.40 | 12.4 | 3107.8 |
| 2000M12 | 132487.52 | 12.3 | 3680 |
| 2001M01 | 135685.99 | 13.5 | 3332.8 |
| 2001M02 | 134390.49 | 12.0 | 3047.1 |
| 2001M03 | | | |

图 2-61　2000 年 1 月至 2022 年 7 月中国货币和标准货币的 M2 情况

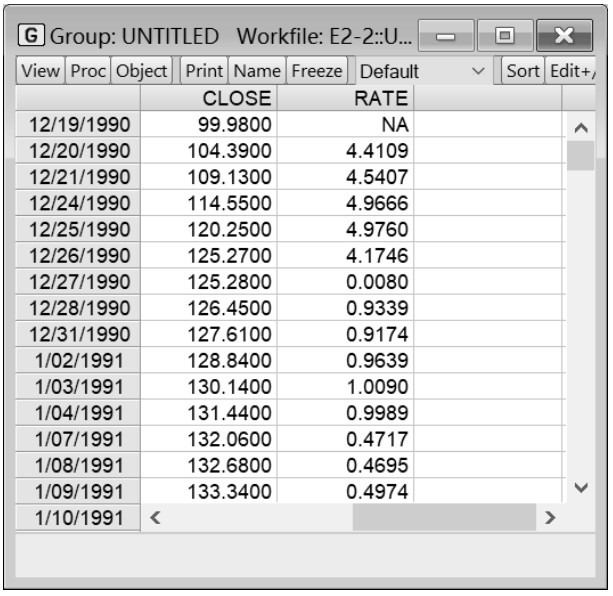
（1）绘制序列 M2 的折线图。

（2）绘制序列 M2 的直方图和统计表并判断是否为正态分布序列。

（3）对 M2 的增长值 RATE 进行分类输出，其中有几个月份的增长值分别超过 20%和 25%？

（4）对 M2 的增长值 RATE 进行重复值分析，其中，增长值为 26%和 28.5%各有多少次？分别为哪几个月？

2．工作文件 E2-2.WF1（见图 2-62）存储的是 1990 年 12 月至 2022 年 9 月上证指数（000001）数据，其中，序列 CLOSE 是当日收盘价，序列 RATE 是当日收盘价较前一日的涨跌幅。




| | CLOSE | RATE |
|---|---|---|
| 12/19/1990 | 99.9800 | NA |
| 12/20/1990 | 104.3900 | 4.4109 |
| 12/21/1990 | 109.1300 | 4.5407 |
| 12/24/1990 | 114.5500 | 4.9666 |
| 12/25/1990 | 120.2500 | 4.9760 |
| 12/26/1990 | 125.2700 | 4.1746 |
| 12/27/1990 | 125.2800 | 0.0080 |
| 12/28/1990 | 126.4500 | 0.9339 |
| 12/31/1990 | 127.6100 | 0.9174 |
| 1/02/1991 | 128.8400 | 0.9639 |
| 1/03/1991 | 130.1400 | 1.0090 |
| 1/04/1991 | 131.4400 | 0.9989 |
| 1/07/1991 | 132.0600 | 0.4717 |
| 1/08/1991 | 132.6800 | 0.4695 |
| 1/09/1991 | 133.3400 | 0.4974 |
| 1/10/1991 | | |

图 2-62　1990 年 12 月至 2022 年 9 月上证指数收盘情况

（1）将样本的区间设置为 2020 年 1 月 2 日至 2022 年 9 月 19 日（单击工作文件窗口的 Sample，在弹出的对话框中的 Sample range pairs 下方输入“1/2/2020 9/19/2022”，中间以空格间隔）。

（2）分别画出序列 CLOSE 和 RATE 的折线图。

（3）分别画出序列 CLOSE 和 RATE 的相关图并判断是否为纯随机序列。

（4）分别对序列 CLOSE 和 RATE 进行标准单位根检验，判断原序列是否存在单位根。如果存在单位根，则对其 1 阶差分进行标准单位根检验。

（5）对序列 CLOSE 进行断点单位根检验，判断是否存在断点单位根。