

# IBM SPSS Forecasting 20



注意：使用本信息及其支持的产品之前，请阅读注意事项第 100 页码下的一般信息。

此版本适用于 IBM® SPSS® Statistics 20 及所有后续发布和修订，除非在新版本中另有说明。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。

受许可保护材料 – IBM 所有

**Copyright IBM Corporation 1989, 2011.**

美国政府用户受限权利 – 使用、复制或披露受与 IBM Corp. 签订的 GSA ADP Schedule Contract 的限制。

---

# 前言

IBM® SPSS® Statistics 是一种用于分析数据的综合系统。Forecasting 可选附加模块提供本手册中描述的其他分析方法。此 Forecasting 附加模块必须与 SPSS Statistics Core 系统一起使用，并已完全集成到了该系统中。

## 关于 IBM Business Analytics

IBM Business Analytics 软件提供决策者赖以提高业务绩效的完整、一致和准确的信息。包括业务智能、预测分析、财务状况和战略管理以及分析应用程序在内的一整套产品组合让您即刻、清楚地了解当前绩效并依此采取行动，以及能够预测未来的成果。结合丰富的行业解决方案、被证明的实践经验 and 专业的服务，无论公司规模大小，都能促使其获得最高的产能、自信自觉地做出决定并得到更好的成绩。

作为产品组合的一部分，IBM SPSS Predictive Analytics 软件帮助公司预测未来实践并采取积极行动，促使其获得更好的业务成果。全世界的商业政府和学术客户依赖 IBM SPSS 技术，因其具有竞争力的优势，能够吸引、留住和发展客户，同时减少欺诈和减轻风险。通过将 IBM SPSS 软件融入日常运营中，公司成为具有预测性的企业，能够引导和自觉做出决策，以满足业务目标，实现可观的竞争优势。欲知更多信息或联系代表，请访问 <http://www.ibm.com/spss>。

## 技术支持

我们提供有“技术支持”以维护客户。客户可就 IBM Corp. 产品使用或某一受支持硬件环境的安装帮助寻求技术支持。如要联系技术支持，请访问 IBM Corp. 网站，网址为 <http://www.ibm.com/support>。在请求协助时，请准备好您和您组织的 ID 以及支持协议。

## 针对学生的技术支持

如果您是使用任何学生版 IBM SPSS 软件产品的学生，请访问我们专为学生提供的在线教育解决方案 (<http://www.ibm.com/spss/rd/students/>) 页面。如果您是使用大学提供的 IBM SPSS 软件副本的学生，请联系所在大学的 IBM SPSS 产品协调员。

## 客户服务

如果对发货或帐户存在任何问题，请联系您当地的办事处。请先准备好您的序列号以供识别。

## 培训讲座

IBM Corp. 提供公开的以及现场的培训讲座。所有讲座都是以实践小组为特色的。讲座将定期在各大城市开展。有关这些讲座的更多信息，请前往 <http://www.ibm.com/software/analytics/spss/training>。

## 附加出版物

SPSS Statistics: 数据分析指南、SPSS Statistics: Statistical Procedures Companion 和 SPSS Statistics: Advanced Statistical Procedures Companion (由 Marija Norušis 编写, 并已由 Prentice Hall 出版) 作为建议的补充材料提供。这些出版物涵盖 SPSS Statistics Base 模块、Advanced Statistics 模块和 回归模块中的统计过程。无论您是刚开始从事数据分析工作, 还是已准备好使用高级应用程序, 这些书籍都将帮助您最有效地利用在 IBM® SPSS® Statistics 产品中找到的功能。有关其他信息, 包括出版物的内容和示例章节, 请参阅作者的网站: <http://www.norusis.com>

---

# 内容

## 部分 I: 用户指南

<b>1</b>	<b>时间序列简介</b>	<b>1</b>
	时间序列数据 . . . . .	1
	数据转换 . . . . .	1
	估计期和验证期 . . . . .	2
	建立模型并生成预测值 . . . . .	2
<b>2</b>	<b>时间序列建模器</b>	<b>3</b>
	为专家建模器指定选项 . . . . .	5
	模型选择和事件说明 . . . . .	6
	使用专家建模器处理离群值 . . . . .	7
	定制的指数平滑法模型 . . . . .	8
	定制的 ARIMA 模型 . . . . .	9
	定制的 ARIMA 模型的模型规格 . . . . .	10
	定制的 ARIMA 模型中的转换函数 . . . . .	11
	定制的 ARIMA 模型中的离群值 . . . . .	13
	输出 . . . . .	14
	统计量和预测表 . . . . .	14
	图 . . . . .	16
	将输出限制为最佳或最差拟合模型 . . . . .	18
	保存模型预测值和模型规格 . . . . .	19
	选项 . . . . .	20
	TSMODEL 命令的附加功能 . . . . .	21
<b>3</b>	<b>应用时间序列模型</b>	<b>23</b>
	输出 . . . . .	25
	统计量和预测表 . . . . .	26
	图 . . . . .	28
	将输出限制为最佳或最差拟合模型 . . . . .	29
	保存模型预测值和模型规格 . . . . .	31
	选项 . . . . .	32
	TSAPPLY 命令的其他功能 . . . . .	33

<b>4</b>	<b>季节性分解</b>	<b>34</b>
	季节性分解保存 . . . . .	35
	SEASON 命令的其他功能 . . . . .	36
<b>5</b>	<b>谱图</b>	<b>37</b>
	SPECTRA 命令的其他功能 . . . . .	39
 <b>部分 II：示例</b>		
<b>6</b>	<b>使用专家建模器进行批量预测</b>	<b>41</b>
	检查数据 . . . . .	41
	运行分析 . . . . .	42
	模型摘要图 . . . . .	49
	模型预测值 . . . . .	50
	摘要 . . . . .	51
<b>7</b>	<b>通过应用保存的模型重新进行批量预测</b>	<b>52</b>
	运行分析 . . . . .	52
	模型拟合统计量 . . . . .	55
	模型预测值 . . . . .	56
	摘要 . . . . .	56
<b>8</b>	<b>使用专家建模器确定重要预测值</b>	<b>57</b>
	对数据进行绘图 . . . . .	57
	运行分析 . . . . .	58
	序列图 . . . . .	64
	模型描述表 . . . . .	64
	模型统计表 . . . . .	65

ARIMA 模型参数表 . . . . .	65
摘要 . . . . .	65
<b>9 通过应用保存的模型试验预测值</b>	<b>66</b>
延伸预测变量序列 . . . . .	66
在预测期中修改预测变量值 . . . . .	70
运行分析 . . . . .	72
<b>10 周期性分解</b>	<b>76</b>
移去销售数据中的季节性 . . . . .	76
确定和设置周期性 . . . . .	76
运行分析 . . . . .	80
了解输出 . . . . .	80
摘要 . . . . .	82
相关过程 . . . . .	82
<b>11 谱图</b>	<b>83</b>
使用频谱图验证周期性的期望值 . . . . .	83
运行分析 . . . . .	83
了解周期图和频谱密度 . . . . .	84
摘要 . . . . .	86
相关过程 . . . . .	86

## 附录

A	拟合优度测量	87
B	离群值类型	88
C	ACF/PACF 图指南	89
D	样本文件	93
E	注意事项	100
	参考书目	102
	索引	104

# 部分 I: 用户指南



# 时间序列简介

**时间序列**是可通过定期度量时间段中某个变量获得的一组观察值。例如，在一系列库存数据中，观察值可能表示若干个月的日常库存水平。显示产品市场份额的序列可能由过去几年来的每周市场份额组成。总销售额序列可能包含许多年来每个月的一个观察值。这些示例中的每个示例都有一个共同点，即都曾在一定时间长度内的已知间隔定期观察某个变量。因此，典型时间序列的数据形式是表示定期间隔内获得的测量值的单个观察值序列或观察值列表。

表 1-1  
每日库存数据序列

时间	星期	日期	库存水平
t <sub>1</sub>	1	星期一	160
t <sub>2</sub>	1	星期二	135
t <sub>3</sub>	1	星期三	129
t <sub>4</sub>	1	星期四	122
t <sub>5</sub>	1	星期五	108
t <sub>6</sub>	2	星期一	150
		...	
t <sub>60</sub>	12	星期五	120

进行时间序列分析的最重要原因之一是尝试预测序列的未来值。解释过去值的序列的模型还可以预测下几个值是增加还是减少，以及增减的幅度有多大。显然，成功进行此类预测的能力对于任何公司或科学领域来说都非常重要。

## 时间序列数据

在定义要与“预测”附加模块一起使用的时间序列数据时，每个序列均对应一个变量。例如，要在数据编辑器中定义时间序列，请单击**变量视图**选项卡，然后在任意空白行中输入变量名。时间序列中的每个观察值都对应一个个案（数据编辑器中的一行）。

如果打开包含时间序列数据的电子表格，则每个序列都应安排在该电子表格的一列中。如果已将时间序列安排在电子表格的行中，则可以采用任意方式打开该电子表格，并使用“数据”菜单中的“转置”，将行转置成列。

## 数据转换

Core 系统中提供的许多数据转换过程在进行时间序列分析时非常有用。

- 定义日期过程（在“数据”菜单上）生成用于建立周期性和用于区分历史、验证和预测期的日期变量。“预测”可以与由“创建日期”过程创建的变量一起使用。

- 创建时间序列过程（在“转换”菜单上）创建新的时间序列变量作为现有时间序列变量的函数。它包括使用相邻观察值来求平滑值、平均值或差分值的函数。
- 替换缺失值过程（在“转换”菜单上）将系统缺失值和用户缺失值替换为根据多种方法之一估计的值。序列开头和末尾的缺失数据不会引发特殊的问题，只会缩短序列的有效长度。序列中间的缺口（内嵌缺失数据）是更为严重的问题。

有关时间序列的数据转换的详细信息，请参见《Core 用户指南》。

## 估计期和验证期

通常，将时间序列划分成估计或历史期和验证期非常有用。在估计（历史）期的观察值的基础上建立一个模型，然后在验证期对其进行测试，以了解其性能。通过使用该模型对已知的点（验证期中的点）进行预测，即可了解该模型在预测方面的性能如何。

验证期的个案通常指保持个案，因为它们是从模型构建过程中保留下来的。估计期包含活动数据集中当前选定的个案。最后选定的个案后的所有剩余个案都可以用作保持个案。在对该模型满意之后，就可以进行充分的预测作业了，您可以重新定义估计期以包含保持个案，然后建立最终模型。

## 建立模型并生成预测值

“预测”附加模块提供了两个过程，用于完成创建模型和生成预测值的任务。

- **时间序列建模器** 过程为时间序列创建模型并生成预测值。它包括专家建模器，用于自动确定每个时间序列的最佳模型。对于希望在更高的级别上进行控制的有经验的人员，它还提供了定制模型建立的工具。
- **应用时间序列模型** 过程将时间序列建模器创建的现有时间序列模型应用于活动数据集。使用此过程，可以在不重新建立模型的情况下获得其新数据或修订数据可用的序列的预测。如果确认模型已发生更改，则可使用时间序列建模器重新建立该模型。

# 时间序列建模器

“时间序列建模器”过程会估计时间序列的指数平滑法模型、单变量自回归积分移动平均数 (ARIMA) 模型和多变量 ARIMA (或转换函数模型) 模型, 并生成预测值。该过程包含的专家建模器可自动为一个或多个因变量序列标识和估计最佳拟合 ARIMA 或指数平滑法模型, 因而就不必通过反复试验来标识适当的模型了。另外, 可以指定定制的 ARIMA 模型或指数平滑法模型。

**示例。**假定您是产品经理, 负责预测 100 种独立产品中每一种产品的下个月的单位销售额和收入, 但是您对建模时间序列又知之甚少。所有 100 种产品的单位销售额的历史数据都存储在一个 Excel 电子表格中。在 IBM® SPSS® Statistics 中打开电子表格后, 您使用专家建模器申请了未来一个月的预测。专家建模器会为每一种产品查找最合适的单位销售额模型, 并使用这些模型生成预测。由于专家建模器可以处理多个输入序列, 因此, 您只需要运行该过程一次即可获得所有产品的预测。通过选择将预测保存到活动数据集, 您可以轻松地将结果导回 Excel 中。

**统计量。**拟合优度测量: 平稳的 R 方、R 方 ( $R^2$ )、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对误差百分比 (MAPE)、最大绝对误差 (MaxAE)、最大绝对误差百分比 (MaxAPE)、标准化 BIC 准则。残差: 自相关函数、偏自相关函数、Ljung-Box Q。对于 ARIMA 模型: 因变量的 ARIMA 阶数、自变量的转换函数阶数以及离群值估计数。此外, 还有指数平滑法模型的平滑法参数估计值。

**图。**跨所有模型的摘要图: 平稳的 R 方、R 方 ( $R^2$ )、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对误差百分比 (MAPE)、最大绝对误差 (MaxAE)、最大绝对误差百分比 (MaxAPE)、标准化 BIC 准则的直方图; 残差自相关和偏自相关的箱图。单个模型的结果: 预测值、拟合值、观察值、置信区间的上限和下限、残差自相关和偏自相关。

## 时间序列建模器的数据注意事项

**数据。** 因变量和任何自变量必须是数值型。

**假设。** 因变量和任何自变量都视为时间序列, 即每个个案代表一个时间点, 连续的个案之间由常数时间区间分隔。

- **平稳性。** 对于定制的 ARIMA 模型, 要建模的时间序列应该是平稳序列。将不平稳序列转换成平稳序列的最有效方法是使用差分转换 — 可从“创建时间序列”对话框开始。
- **预测值。** 为了使用带有自 (预测) 变量的模型生成预测值, 活动数据集应包含预测期中全部个案的这些变量的值。此外, 自变量不应包括估计期中的任何缺失值。

## 定义日期

尽管不是必需的, 但还是建议使用定义日期对话框指定与第一个个案相关的日期以及连续个案之间的时间区间。此操作应在使用时间序列建模器之前执行, 并将导致一组变量上会标记每个个案的相关日期。此操作还会设置数据的假定周期, 例如, 如果连续个案之间的时间区间为一个月, 则周期为 12。如果您希望创建季节性模型, 则需要此周

期。如果不想创建季节性模型，输出中也不需要日期标签，则可以跳过“定义日期”对话框。这样，与每个个案相关的标签将仅是个案号。

### 使用时间序列建模器

- ▶ 从菜单中选择：  
分析 > 预测 > 创建模型...

图片 2-1  
时间序列建模器，“变量”选项卡



- ▶ 在“变量”选项卡上，选择一个或多个要建模的因变量。
- ▶ 从“方法”下拉框中，选择一种建模方法。如果要自动建模，请保留专家建模器的缺省方法。这样将调用专家建模器来为每个因变量确定最佳拟合模型。  
要生成预测，请执行以下操作：
  - ▶ 单击选项选项卡。
  - ▶ 指定预测期。这将生成一个包括预测和观察值的图表。  
根据需要，您可以：

- 选择一个或多个自变量。自变量的处理方法与回归分析中预测变量的处理方法十分类似，但自变量是可选的。它们可以包含在 ARIMA 模型中，但不能包含在指数平滑法模型中。如果您指定**专家建模器**作为建模方法并包括自变量，则将只考虑 ARIMA 模型。
- 单击**条件**，指定建模的详细信息。
- **保存预测值、置信区间和噪声残差**。
- **以 XML 格式保存估计的模型**。保存的模型可以应用于新数据或修订的数据，以便在不重新构建模型的情况下获取更新的预测。此操作可通过**应用时间序列模型**过程实现。
- **获取所有估计模型的摘要统计量**。
- **为定制的 ARIMA 模型中的自变量指定转换函数**。
- **启用自动检测离群值的功能**。
- **将特定时间点作为定制的 ARIMA 模型的离群值来建模**。

## 建模方法

可用的建模方法有：

**专家建模器**。专家建模器会自动查找每个相依序列的最佳拟合模型。如果指定了自变量（预测）变量，则专家建模器为 ARIMA 模型中的内容选择那些与该相依序列具有统计显著性关系的模型。适当时，使用差分和/或平方根或自然对数变换对模型变量进行转换。缺省情况下，专家建模器既考虑指数平滑法模型也考虑 ARIMA 模型。但是，您可以将专家建模器限制为仅搜索 ARIMA 模型或仅搜索指数平滑法模型。还可以指定自动检测离群值。

**指数平滑法**。使用此选项可指定定制的指数平滑法模型。您可以从各种指数平滑法模型中进行选择，它们在处理趋势和季节性上有所不同。

**ARIMA**。使用此选项可指定定制的 ARIMA 模型。其中包含显式指定自回归的阶和移动平均的阶，以及差分度。可以包含自变量（预测变量）并为它们当中的任何一个或全部定义转换函数。还可以指定自动检测离群值或指定显式离群值集合。

## 估计期和预测期

**估计期**。估计期定义用于确定模型的个案集。缺省情况下，估计期包含活动数据集中的所有个案。要设置估计期，请选择选择个案对话框中的**基于时间或个案范围**。取决于可用的数据，过程中使用的估计期可能会因因变量而异，因此会与显示的值不同。对于给定的因变量，真正的估计期是从该变量中消去指定估计期的开始或结束时发生的任何连续缺失值之后剩下的期间。

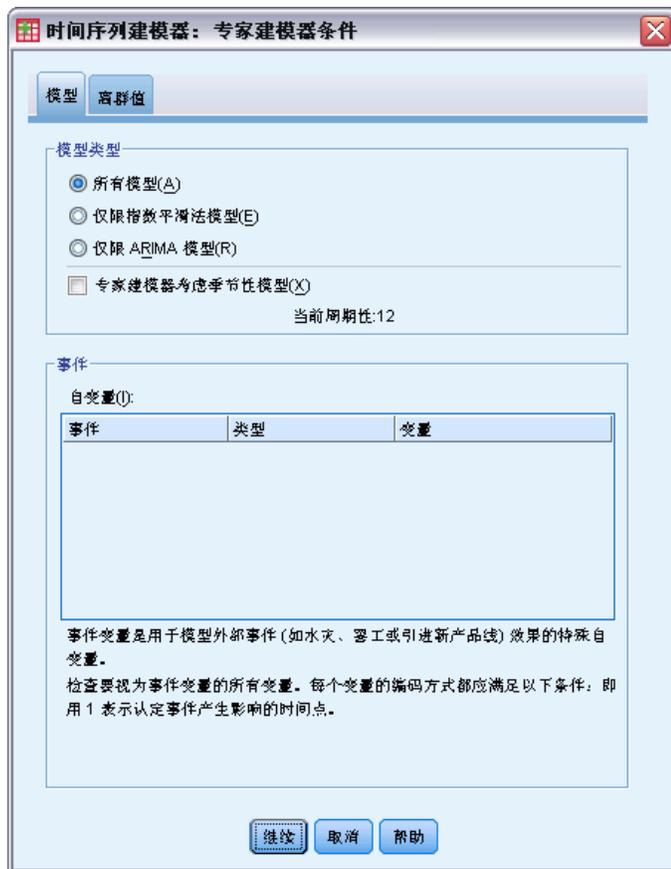
**预测期**。预测期从估计期后的第一个个案开始，缺省情况下，到活动数据集中的最后一个个案结束。您可以从“**选项**”选项卡设置预测期的结束时间。

## 为专家建模器指定选项

专家建模器提供用于约束备选模型集、指定离群值的处理方法以及包括事件变量在内的选项。

## 模型选择和事件说明

图片 2-2  
“专家建模器条件”对话框，“模型”选项卡



使用“模型”选项卡，可以指定专家建模器可以处理的模型类型并指定事件变量。

**模型类型。**可用选项包括：

- **所有模型。**专家建模器既考虑 ARIMA 模型也考虑指数平滑法模型。
- **仅指数平滑法模型。**专家建模器仅考虑指数平滑法模型。
- **仅限 ARIMA 模型。**专家建模器仅考虑 ARIMA 模型。

**专家建模器考虑季节性模型。**只有在为活动数据集定义了周期性时才启用此选项。选择（选中）此选项时，专家建模器既考虑季节性模型，又考虑非季节性模型。如果未选择此选项，则专家建模器仅考虑非季节性模型。

**当前周期性。**指示当前为活动数据集定义的周期性（如果有）。当前周期性以整数形式给出，例如，12 表示年度周期性，每个个案代表一个月份。如果尚未设置周期性，则显示值无。季节性模型要求具有周期性。您可以从“定义日期”对话框中设置周期性。

**事件。**选择要视为事件变量的任何自变量。对于事件变量，值为 1 的个案指示相依序列将受该事件影响的时间。1 以外的值指示无影响。

## 使用专家建模器处理离群值

图片 2-3  
“专家建模器条件”对话框，“离群值”选项卡



使用“离群值”选项卡，可以选择自动检测离群值以及要检测的离群值的类型。

**自动检测离群值。** 缺省情况下，不自动检测离群值。选择（选中）此选项可执行自动检测离群值的操作，并可选择以下一种或多种离群值类型：

- 加法
- 移位水平
- 创新的
- 瞬时的
- 季节性可加的
- 局部趋势
- 可加的修补

有关详细信息，请参阅第 88 页码附录 B 中的离群值类型。

## 定制的指数平滑法模型

图片 2-4  
“指数平滑法条件”对话框



**模型类型。** 指数平滑法模型 (Gardner, 1985) 分为季节性模型和非季节性模型。季节性模型只有在为活动数据集定义了周期时才可用 (请参见下文的“当前周期性”)。

- **简单的。** 该模型适用于没有趋势或季节性的序列。其唯一的平滑参数是水平。简单指数平滑法与 ARIMA 模型极为相似，包含零阶自回归、一阶差分、一阶移动平均数，并且没有常数。
- **Holt 线性趋势。** 该模型适用于具有线性趋势并没有季节性的序列。其平滑参数是水平和趋势，不受相互之间的值的约束。Holt 模型比 Brown 模型更通用，但在计算大序列时要花的时间更长。Holt 指数平滑法与 ARIMA 模型极为相似，包含零阶自回归、二阶差分以及二阶移动平均数。
- **Brown 线性趋势。** 该模型适用于具有线性趋势并没有季节性的序列。其平滑参数是水平和趋势，并假定二者等同。因此，Brown 模型是 Holt 模型的特例。Brown 指数平滑法与具有零阶自回归、二阶差分和二阶移动平均的 ARIMA 模型极为相似，且移动平均第二阶的系数等于第一阶的系数二分之一的平方。
- **阻尼趋势。** 此模型适用于具有线性趋势的序列，且该线性趋势正逐渐消失并且没有季节性。其平滑参数是水平、趋势和阻尼趋势。阻尼指数平滑法与具有一阶自回归、一阶差分和二阶移动平均的 ARIMA 模型极为相似。
- **简单季节性。** 该模型适用于没有趋势并且季节性影响随时间变动保持恒定的序列。其平滑参数是水平和季节。简单季节性指数平滑法与 ARIMA 模型极为相似，包含零阶自回归、一阶差分、一阶季节性差分和一阶、 $p$  阶和  $p + 1$  阶移动平均数，其中  $p$  是季节性区间中的周期数 (对于月数据， $p = 12$ )。

- **Winters 可加的。** 该模型适用于具有线性趋势和不依赖于序列水平的季节性效应的序列。其平滑参数是水平、趋势和季节。Winters 可加的指数平滑法与 ARIMA 模型极为相似，包含零阶自回归、一阶差分、一阶季节差分 and  $p + 1$  阶移动平均数，其中  $p$  是季节性区间中的周期数（对于月数据， $p = 12$ ）。
- **Winters 可乘的。** 该模型适用于具有线性趋势和依赖于序列水平的季节性效应的序列。其平滑参数是水平、趋势和季节。Winters 的可乘指数平滑法与任何 ARIMA 模型都不相似。

**当前周期性。** 指示当前为活动数据集定义的周期性（如果有）。当前周期性以整数形式给出，例如，12 表示年度周期性，每个个案代表一个月份。如果尚未设置周期性，则显示值无。季节性模型要求具有周期性。您可以从“定义日期”对话框中设置周期性。

**因变量转换。** 可以指定在建模之前对每个因变量执行的转换。

- **无。** 不执行任何转换。
- **平方根。** 平方根转换。
- **自然对数。** 自然对数转换。

## 定制的 ARIMA 模型

使用时间序列建模器，可以构建定制的非季节性或季节性 ARIMA（自回归积分移动平均数）模型，也称为 Box-Jenkins (Box, Jenkins, 和 Reinsel, 1994) 模型 - 带或不带固定预测变量集均可。您可以为任意或所有预测变量定义转换函数，指定自动检测离群值或指定一组确切的离群值。

- “变量”选项卡上指定的所有自变量（预测变量）都显式包括在该模型中。这点与使用专家建模器相反；使用专家建模器时，只有在自变量与因变量之间具有统计显著性关系时才会包括自变量。

## 定制的 ARIMA 模型的模型规格

图片 2-5  
ARIMA “条件” 对话框，“模型” 选项卡



“模型” 选项卡用于指定定制的 ARIMA 模型的结构。

**ARIMA 阶数。** 在“结构”网格的相应单元格中，输入模型的各个 ARIMA 成分的值。所有值都必须为非负整数。对于自回归和移动平均数成分，该值表示最大阶。模型中将包含所有正的较低阶。例如，如果指定 2，则模型包括阶数 2 和 1。只有在为活动数据集定义了周期性时，才会启用“季节性”列中的各个单元格（请参见下面的“当前周期性”）。

- **自回归 (p)**。模型中的自回归阶数。自回归阶指定要使用序列中以前的哪些值来预测当前值。例如，自回归阶为 2 时，指定序列中过去两个时段的价值用于预测当前值。
- **差分 (d)**。指定在估计模型之前应用于序列的差分的阶。在出现趋势（具有趋势的序列通常是不稳序列，而 ARIMA 建模假定其稳定）时需要差分，并将其用于去除其影响。差分的阶与序列趋势度相对应，一阶差分导致线性趋势，二阶差分导致二次趋势，等等。
- **移动平均数 (q)**。模型中的移动平均数的阶数。移动平均数的阶指定如何使用先前值的序列平均数的偏差来预测当前值。例如，如果移动平均数的阶为 1 和 2，则指定在预测序列的当前值时将考虑上两个时段的每个时段中的序列的平均值的偏差。

**季节性阶。** 季节性自回归成分、移动平均数成分和差分成分与其非季节性对应成分起着相同的作用。但对于季节性的阶，当前序列值受以前的序列值的影响，序列值之间间隔一个或多个季节性周期。例如，对于月数据（季节性周期为 12），季节性 1 阶表示当前序列值受自当前周期起 12 个周期之前的序列值的影响。因此，对于月数据，指定季节性 1 阶等同于指定非季节性 12 阶。

**当前周期性。** 指示当前为活动数据集定义的周期性（如果有）。当前周期性以整数形式给出，例如，12 表示年度周期性，每个个案代表一个月份。如果尚未设置周期性，则显示值无。季节性模型要求具有周期性。您可以从“定义日期”对话框中设置周期性。

**因变量转换。** 可以指定在建模之前对每个因变量执行的转换。

- **无。** 不执行任何转换。
- **平方根。** 平方根转换。
- **自然对数。** 自然对数转换。

**在模型中包含常数。** 除非您确定整个平均数序列值为 0，否则包含常数是标准方法。当应用差分时，建议不包含常数。

## 定制的 ARIMA 模型中的转换函数

图片 2-6  
ARIMA “条件” 对话框，“转换函数” 选项卡



使用“转换函数”选项卡（只有在指定了自变量时才显示），可以为“变量”选项卡上指定的任意或所有自变量定义转换函数。转换函数可用于指定如何使用自变量（预测变量）的过去值来预测相依序列的未来值。

**转换函数的阶数。** 在“结构”网格的相应单元格中，输入转换函数的各个成分的值。所有值都必须为非负整数。对于分子和分母成分，该值代表最大阶。模型中将包含所有正的较低阶。此外，0 阶始终包括在分子成分中。例如，如果您指定 2 作为分子，则模型会包括阶数 2、1 和 0。如果指定 3 作为分母，则模型会包括阶数 3、2 和 1。只有在为活动数据集定义了周期时，“季节性”列中的各个单元格才会启用（请参见下面的“当前周期性”）。

- **分子。** 转换函数的分子的阶。指定所选独立（预测）序列的哪些以前的值可用于预测相依序列的当前值。例如，分子的阶为 1 时，指定独立序列过去一个时间段的值（以及独立序列的当前值）用于预测每个相依序列的当前值。
- **分母。** 转换函数的分母的阶。指定如何使用所选独立（预测）序列的以前值的序列平均数偏差来预测相依序列的当前值。例如，分母的阶为 1 时，指定在预测每个相依序列的当前值时考虑独立序列过去一个时间段的平均值偏差。
- **差分。** 指定在估计模型之前应用于所选独立（预测）序列的差分的阶数。存在趋势时必须使用差分来去除其效果。

**季节性阶。** 季节性分子、分母和差分成分与其非季节性对应成分起着相同的作用。但对于季节性的阶，当前序列值受以前的序列值的影响，序列值之间间隔一个或多个季节性周期。例如，对于月数据（季节性周期为 12），季节性 1 阶表示当前序列值受自当前周期起 12 个周期之前的序列值的影响。因此，对于月数据，指定季节性 1 阶等同于指定非季节性 12 阶。

**当前周期性。** 指示当前为活动数据集定义的周期性（如果有）。当前周期性以整数形式给出，例如，12 表示年度周期性，每个个案代表一个月份。如果尚未设置周期性，则显示值无。季节性模型要求具有周期性。您可以从“定义日期”对话框中设置周期性。

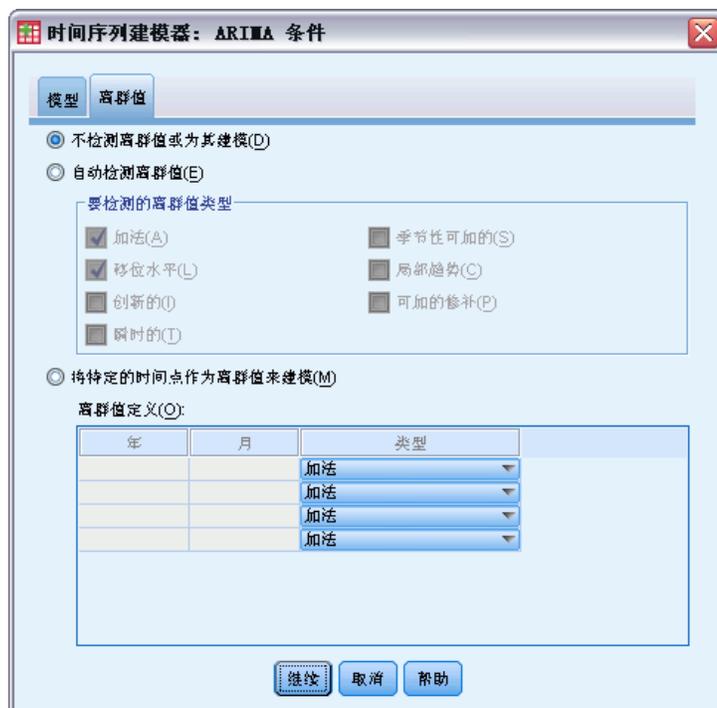
**延迟。** 设置延迟会导致指定的区间个数延迟自变量的影响。例如，如果延迟设置为 5，则时间  $t$  的自变量的值在五个周期过去 ( $t + 5$ ) 之前不会影响预测。

**转换。** 为一组自变量指定转换函数也包括在这些变量上执行可选的转换。

- **无。** 不执行任何转换。
- **平方根。** 平方根转换。
- **自然对数。** 自然对数转换。

## 定制的 ARIMA 模型中的离群值

图片 2-7  
ARIMA “条件” 对话框，“离群值” 选项卡



“离群值”选项卡提供以下选项来处理离群值 (Pena, Tiao, 和 Tsay, 2001): 自动检测离群值、指定特定点作为离群值, 或者不检测离群值或为其建模。

**不检测离群值或为其建模。** 缺省情况下, 不检测离群值, 也不为其建模。选择此选项可禁用任何检测离群值或为其建模的功能。

**自动检测离群值。** 选择此选项可执行自动检测离群值的操作, 并可选择以下一种或多种离群值类型:

- 加法
- 移位水平
- 创新的
- 瞬时的
- 季节性可加的
- 局部趋势
- 可加的修补

有关详细信息, 请参阅第 88 页码附录 B 中的离群值类型。

**将特定时间点作为离群值来建模。** 选择此选项可指定特定的时间点作为离群值。每个离群值在“离群值定义”网格中占单独的一行。在给定行中为所有单元格输入值。

- **类型。** 离群值类型。支持的类型有：可加的（缺省值）、移位水平、创新的、瞬时的、季节性可加以及局部趋势。

注 1：如果没有为活动数据集定义日期规范，则“离群值定义”网格只显示观察值一列。要指定离群值，请输入相关个案的行号（与数据编辑器中显示的相同）。

注 2：“离群值定义”网格中的循环列（如果存在）是活动数据集中的 CYCLE\_ 变量的值。

## 输出

可用输出包括单个模型的结果以及跨所有模型计算出的结果。单个模型的结果可以限制为基于用户特定标准的最佳或最差拟合模型的集合。

## 统计量和预测表

图片 2-8  
时间序列建模器，“统计量”选项卡



“统计量”选项卡提供显示建模结果的表的选项。

**按模型显示拟合测量、Ljung-Box 统计量和离群值个数。** 选择（选中）此选项可显示包含每个估计模型的所选拟合测量、Ljung-Box 值以及离群值数目的表。

**拟合测量。** 您可以将以下一项或多项纳入包含每个估计模型的拟合度量的表中：

- 固定的 R 方
- R 方
- 均方根误差
- 平均绝对误差百分比
- 平均绝对误差
- 最大绝对误差百分比
- 最大绝对误差
- 标准化的 BIC

有关详细信息，请参阅第 87 页码附录 A 中的拟合优度测量。

**比较模型的统计量。** 这组选项控制如何显示包含跨所有估计模型计算出的统计信息的表。每个选项分别生成单独的表。可以选择以下选项中的一个或多个：

- **拟合优度。** 固定的 R 方、R 方、均方根误差、平均绝对误差百分比、平均绝对误差、最大绝对误差百分比、最大绝对误差以及标准化的 BIC 准则的摘要统计量和百分位数表。
- **残差自相关函数 (ACF)。** 所有估计模型中残差的自相关摘要统计和百分位表。
- **残差部分自相关函数 (PACF)。** 所有估计模型中残差的部分自相关摘要统计和百分位表。

**个别模型的统计量。** 这组选项控制如何显示包含每个估计模型的详细信息的表。每个选项分别生成单独的表。可以选择以下选项中的一个或多个：

- **参数估计。** 显示每个估计模型的参数估计值的表。为指数平滑法和 ARIMA 模型显示不同的表。如果存在离群值，则它们的参数估计值也将在单独的表中显示。
- **残差自相关函数 (ACF)。** 按每个估计模型的延迟显示残差自相关表。该表包含自相关的置信区间。
- **残差部分自相关函数 (PACF)。** 按每个估计模型的延迟显示残差部分自相关表。该表包含部分自相关的置信区间。

**显示预测值。** 显示每个估计模型的模型预测和置信区间的表。预测期在“选项”选项卡中设置。



图片 2-9  
时间序列建模器，“图”选项卡



“图”选项卡提供显示建模结果的图的选项。

### 模型比较图

这组选项控制如何显示包含跨所有估计模型计算出的统计信息的图。每个选项分别生成单独的图。可以选择以下选项中的一个或多个：

- 固定的 R 方
- R 方
- 均方根误差
- 平均绝对误差百分比
- 平均绝对误差
- 最大绝对误差百分比
- 最大绝对误差
- 标准化的 BIC

- 残差自相关函数 (ACF)
- 残差部分自相关函数 (PACF)

有关详细信息，请参阅第 87 页码附录 A 中的拟合优度测量。

### 单个模型图

**序列。** 选择（选中）此选项可获取每个估计模型的预测值图。可以选择在图中包含以下一项或多项：

- **观察值。** 相依序列的观察值。
- **预测值。** 预测期的模型预测值。
- **拟合值。** 估计期的模型预测值。
- **预测值的置信区间。** 预测期的置信区间。
- **拟合值的置信区间。** 估计期的置信区间。

**残差自相关函数 (ACF)。** 显示每个估计模型的残差自相关图。

**残差部分自相关函数 (PACF)。** 显示每个估计模型的残差部分自相关图。

## 将输出限制为最佳或最差拟合模型

图片 2-10  
时间序列建模器，“输出过滤器”选项卡



“输出过滤器”选项卡提供的选项可用于将表格输出和图形输出限制为估计模型子集。可以选择根据提供的拟合标准将输出限制为最佳和/或最差拟合模型。缺省情况下，输出中包括所有估计模型。

**最佳拟合模型。**选择（选中）此选项将在输出中包含最佳拟合模型。选择拟合优度测量并指定要包含的模型数。选择此选项不妨碍同时选择最差拟合模型。如果同时选择两者，则输出将由最差拟合模型和最佳拟合模型组成。

- **模型的固定数量。**指定为  $n$  个最佳拟合模型显示结果。如果该数量超过估计模型的数量，则显示所有模型。
- **占模型总数的百分比。**指定为其拟合优度值在所有估计模型的前  $n$  个百分比范围内的模型显示结果。

**最差拟合模型。**选择（选中）此选项将在输出中包含最差拟合模型。选择拟合优度测量并指定要包含的模型数。选择此选项不妨碍同时选择最佳拟合模型。如果同时选择两者，输出将由最佳拟合模型和最差拟合模型组成。

- **模型的固定数量。**指定为  $n$  个最差拟合模型显示结果。如果该数量超过估计模型的数量，则显示所有模型。
- **占模型总数的百分比。**指定为其拟合优度值在所有估计模型的后  $n$  个百分比范围内的模型显示结果。

**拟合优度。**选择用于过滤模型的拟合优度测量。缺省值为固定的  $R$  方。

## 保存模型预测值和模型规格

图片 2-11  
时间序列建模器，“保存”选项卡



使用“保存”选项卡可以将模型预测值另存为活动数据集中的新变量，也可以将模型规格以 XML 格式保存到外部文件中。

**保存变量。**可以将模型预测值、置信区间和残差另存为活动数据集中的新变量。每个相依序列都会带来其自己的一组新变量，每个新变量都包含估计期和预测期的值。如果预测期超出了该相依变量序列的长度，则增加新个案。选中每一项所关联的“保存”复选框可以选择保存新变量。缺省情况下不保存新变量。

- **预测值。**模型预测值。
- **置信区间的下限。**预测值的置信区间下限。

- **置信区间的上限。** 预测值的置信区间上限。
- **噪声残值。** 模型残差。如果执行了因变量转换（例如自然对数），则为转换后的序列的残差。
- **变量名的前缀。** 指定要用于新变量名称的前缀或保留缺省的前缀。变量名由前缀、关联因变量的名称和模型标识组成。需要时可以扩展变量名以避免变量命名冲突。前缀必须符合有效变量名称规则。

**导出模型文件。** 所有估计模型的模型规格都将以 XML 格式导出到指定的文件中。保存的模型可用于通过[应用时间序列模型](#)过程在较新数据的基础上获得更新的预测。

## 选项

图片 2-12  
时间序列建模器，“选项”选项卡



“选项”选项卡可用于设置预测期、指定缺失值的处理方法、设置置信区间宽度、指定模型标识的定制前缀并设置为自相关显示的延迟个数。

**预测期。** 预测期始终以估计期（用于确定模型的一组个案）结束后的第一个个案开始，到活动数据集中的最后一个个案或用户指定的日期结束。缺省情况下，估计期的结束是活动数据集中的最后一个个案，但可以在以下位置更改它：选择个案对话框中的基于时间或个案范围。

- **估计期结束之后的第一个个案到活动数据集内的最后一个个案。** 如果估计期在活动数据集中的最后一个个案之前结束，而您需要直到最后一个个案的预测值，则请选择此选项。此选项通常用来生成保持期的预测，以便将模型预测与实际值子集进行比较。
- **估计期结束之后到指定日期之间的第一个个案。** 选择此选项可显式指定预测期的结束。此选项通常用于在实际序列结束后生成预测。在“日期”网格中为所有单元格输入值。

如果没有为活动数据集定义日期规范，则“日期”网格只显示观察值一列。要指定预测期的结束时间，请输入相关个案在数据编辑器中显示的行号。

“日期”网格中的循环列（如果存在）是活动数据集中的 CYCLE\_ 变量的值。

**用户缺失值。** 这些选项控制用户缺失值的处理。

- **视为无效。** 用户缺失值当作系统缺失值处理。
- **视为有效。** 用户缺失值当作有效数据处理。

**缺失值策略。** 下列规则适用于在建模过程期间处理缺失值（包括系统缺失值和视为无效的用户缺失值）：

- 模型中包含估计期发生的因变量存在缺失值的个案。缺失值的具体处理方式取决于估计方法。
- 如果自变量在估计期有缺失值，系统会发出警告。对于专家建模器，将在不使用变量的情况下对包含自变量的模型进行估计。对于定制 ARIMA，不会估计涉及自变量的模型。
- 如果在预测期内任何自变量有缺失值，建模过程将发出警告并进行最大程度的预测。

**置信区间宽度(%)。** 将为模型预测值和残差自相关计算置信区间。可以指定小于 100 的任何正数。缺省情况下使用 95% 的置信区间。

**输出中模型标识的前缀。** “变量”选项卡上指定的每个因变量都可带来一个单独的估计模型。模型都用唯一名称区别，名称由可定制的前缀和整数后缀组成。您可以输入前缀，也可以保留模型的缺省值。

**ACF 和 PACF 输出中显示的最大延迟数。** 可以设置在自相关和偏自相关表和图中显示的最大延迟数。

## TSMODEL 命令的附加功能

如果将选择内容粘贴到语法窗口并编辑随之产生的 TSMODEL 命令语法，则您可以定制自己的时间序列建模。使用命令语法语言可以：

- 指定数据的季节性周期（使用 AUXILIARY 子命令的 SEASONLENGTH 关键字）。该设置的优先级高于活动数据集的当前周期（如果有）。

- 指定定制 ARIMA 和转换函数成分的不连续延迟（使用 `ARIMA` 和 `TRANSFERFUNCTION` 子命令）。例如，您可以指定定制 ARIMA 模型的自回归延迟的阶数为 1、3 和 6；或转换函数的分子延迟阶数为 2、5 和 8。
- 为单次运行的数据序列建模器过程提供一组以上的建模规格，例如：建模方法、ARIMA 阶数、自变量等等（使用 `MODEL` 子命令）。

请参见命令语法参考以获取完整的语法信息。

# 应用时间序列模型

“应用时间序列模型”过程从外部文件加载现有的时间序列模型，并将它们应用于活动数据集。使用此过程，可以在不重新建立模型的情况下获得其新数据或修订数据可用的序列的预测值。模型是使用 [时间序列建模器](#) 过程生成的。

**示例。**假定您是一家大型零售店的库存经理，您负责管理 5,000 种产品。您曾使用专家建模器创建了一些模型，用来预测每种产品在未来三个月的销售情况。您的数据仓库每个月都会使用实际销售数据进行刷新，您希望使用这些数据来生成每月更新预测值。通过“应用时间序列模型”过程，您可以使用原有模型，然后只需重新估计模型参数以说明新数据即可实现此预测。

**统计量。**拟合优度测量：平稳的 R 方、R 方 ( $R^2$ )、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对误差百分比 (MAPE)、最大绝对误差 (MaxAE)、最大绝对误差百分比 (MaxAPE)、标准化 BIC 准则。残差：自相关函数、偏自相关函数、Ljung-Box Q。

**图。**跨所有模型的摘要图：平稳的 R 方、R 方 ( $R^2$ )、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对误差百分比 (MAPE)、最大绝对误差 (MaxAE)、最大绝对误差百分比 (MaxAPE)、标准化 BIC 准则的直方图；残差自相关和偏自相关的箱图。单个模型的结果：预测值、拟合值、观察值、置信区间的上限和下限、残差自相关和偏自相关。

## 应用时间序列模型数据注意事项

**数据。**对其应用模型的变量（自变量和因变量）应为数值变量。

**假设。**将模型应用到活动数据集中与模型中指定的变量同名的变量。所有此类变量均视为时间序列，即每个个案代表一个时间点，连续的个案之间由常数时间区间分隔。

- **预测值。**为了使用带有自（预测）变量的模型生成预测值，活动数据集应包含预测期中全部个案的这些变量的值。如果重新估计模型参数，则自变量不应包含估计期中的任何缺失值。

## 定义日期

“应用时间序列模型”过程要求活动数据集的周期性（如果有）与要应用的模型的周期性相匹配。如果只是使用建立模型时使用的同一数据集（或许包含新的数据或修订的数据）进行预测，则需满足此条件。如果活动数据集没有周期性则您将可能导航到定义日期对话框以创建一个周期性。但是，如果创建模型时没有指定周期性，则活动数据集也应该没有周期性。

## 应用模型

- ▶ 从菜单中选择：  
分析 > 预测 > 应用模型...

图片 3-1  
应用时间序列模型，?模型?选项卡



- ▶ 输入模型文件的文件规范，或单击浏览，然后选择一个模型文件（模型文件是通过[时间序列建模器](#)过程创建的）。

根据需要，您可以：

- 使用活动数据集中的数据重新估计模型参数。使用重新估计的参数创建预测值。
- [保存预测值、置信区间和噪声残差。](#)
- [以 XML 格式保存重新估计的模型。](#)

### 模型参数和拟合优度测量

**从模型文件加载。**使用该模型文件中的模型参数（而不必重新估计这些参数）生成预测值。从该模型文件中获取显示在输出中并用于过滤模型（最佳拟合或最差拟合）的[拟合优度测量](#)，该测量反映了建立（或上次更新）每个模型时使用的数据。使用此选项，预测值将不考虑活动数据集中因变量或自变量的历史数据。如果您希望历史数据影响预测值，则必须选择[根据数据重新估计](#)。此外，预测值也不考虑预测期中相依序列的值，但会考虑预测期中自变量的值。如果相依序列的值已更新，并且您希望将它们包含在预测值中，则需要重新估计，将估计期调整为包含这些值。

**根据数据重新估计。**使用活动数据集中的数据重新估计模型参数。重新估计模型参数不会影响到模型结构。例如，ARIMA(1, 0, 1) 模型将保持原样，但会重新估计自回归参数和移动平均数参数。重新估计不会导致检测新的离群值。离群值（如果有）将始终从模型文件中获取。

- **估计期。**估计期定义用于重新估计模型参数的个案集。缺省情况下，估计期包含活动数据集中的所有个案。要设置估计期，请选择选择个案对话框中的基于时间或个案范围。根据可用数据的不同，过程中使用的估计期可能会因模型而异，因此会与显示的值不同。对于给定的模型，真正的估计期是从模型的因变量中排除指定估计期的开始或结束时发生的所有连续缺失值之后剩下的时段。

## 预测期

每个模型的预测期始终从估计期结束后的第一个个案开始，到活动数据集中的最后一个个案或用户指定的日期结束。如果没有重新估计参数（这是缺省情况），则每个模型的估计期是建立（或上次更新）模型时使用的个案集。

- **估计期结束之后的第一个个案到活动数据集内的最后一个个案。**如果估计期在活动数据集中的最后一个个案之前结束，而您需要直到最后一个个案的预测值，则请选择此选项。
- **估计期结束之后到指定日期之间的第一个个案。**选择此选项可显式指定预测期的结束。在?日期?网格中为所有单元格输入值。

如果没有为活动数据集定义日期规范，则?日期?网格只显示观察值一行。要指定预测期的结束时间，请输入相关个案在数据编辑器中显示的行号。

?日期?网格中的循环列（如果存在）是活动数据集中的 CYCLE\_ 变量的值。

## 输出

可用输出包括单个模型的结果以及所有模型的结果。单个模型的结果可以限制为基于用户特定标准的最佳或最差拟合模型的集合。

## 统计量和预测表

图片 3-2  
应用时间序列模型, ?统计量?选项卡



?统计量?选项卡提供的选项用于显示模型拟合统计量、模型参数、自相关函数以及预测值的表。除非重新估计模型参数（?模型?选项卡上的根据数据重新估计），否则，显示的拟合测量值、Ljung-Box 值和模型参数都是模型文件中的相应值，并反映每个模型建立（或上次更新）时使用的数据。始终从该模型文件获取离群值信息。

**按模型显示拟合测量、Ljung-Box 统计量和离群值个数。**选择（选中）此选项可显示包含每个模型的所选拟合测量、Ljung-Box 值以及离群值个数的表。

**拟合测量。**您可以将以下一项或多项纳入包含每个模型的拟合测量的表中：

- 固定的 R 方
- R 方
- 均方根误差
- 平均绝对误差百分比
- 平均绝对误差
- 最大绝对误差百分比
- 最大绝对误差
- 标准化的 BIC

有关详细信息，请参阅第 87 页码附录 A 中的拟合优度测量。

**比较模型的统计量。**这组选项控制如何显示包含所有模型的统计量的表。每个选项分别生成单独的表。可以选择以下选项中的一个或多个：

- **拟合优度。**固定的 R 方、R 方、均方根误差、平均绝对误差百分比、平均绝对误差、最大绝对误差百分比、最大绝对误差以及标准化的 BIC 准则的摘要统计量和百分位数表。
- **残差自相关函数 (ACF)。**所有估计模型中残差的自相关摘要统计和百分位表。此表只在重新估计模型参数时可用（?模型?选项卡上的**根据数据重新估计**）。
- **残差部分自相关函数 (PACF)。**所有估计模型中残差的部分自相关摘要统计和百分位表。此表只在重新估计模型参数时可用（?模型?选项卡上的**根据数据重新估计**）。

**个别模型的统计量。**这组选项控制如何显示包含每个模型的详细信息的表。每个选项分别生成单独的表。可以选择以下选项中的一个或多个：

- **参数估计。**显示每个模型的参数估计值的表。为指数平滑法和 ARIMA 模型显示不同的表。如果存在离群值，则它们的参数估计值也将在单独的表中显示。
- **残差自相关函数 (ACF)。**按每个估计模型的延迟显示残差自相关表。该表包含自相关的置信区间。此表只在重新估计模型参数时可用（?模型?选项卡上的**根据数据重新估计**）。
- **残差部分自相关函数 (PACF)。**按每个估计模型的延迟显示残差部分自相关表。该表包含部分自相关的置信区间。此表只在重新估计模型参数时可用（?模型?选项卡上的**根据数据重新估计**）。

**显示预测值。**显示每个模型的模型预测值和置信区间的表。



图片 3-3  
应用时间序列模型, ?图?选项卡



?图?选项卡提供的选项可用于显示模型拟合统计量、自相关函数以及序列值（包括预测值）的图。

### 模型比较图

这组选项控制如何显示包含所有模型的统计值的图。除非重新估计模型参数（?模型?选项卡上的根据数据重新估计），否则，显示的值均为模型文件中的值，并反映每个模型建立（或上次更新）时使用的数据。此外，只有在重新估计模型参数时自相关图才可用。每个选项分别生成单独的图。可以选择以下选项中的一个或多个：

- 固定的 R 方
- R 方
- 均方根误差
- 平均绝对误差百分比
- 平均绝对误差
- 最大绝对误差百分比
- 最大绝对误差
- 标准化的 BIC

- 残差自相关函数 (ACF)
- 残差部分自相关函数 (PACF)

有关详细信息，请参阅第 87 页码附录 A 中的拟合优度测量。

### 单个模型图

**序列。**选择（选中）此选项可获取每个模型的预测值的图。只有在重新估计模型参数时（?模型?选项卡上的根据数据重新估计），观察值、拟合值、拟合值的置信区间以及自相关才可用。可以选择在图中包含以下一项或多项：

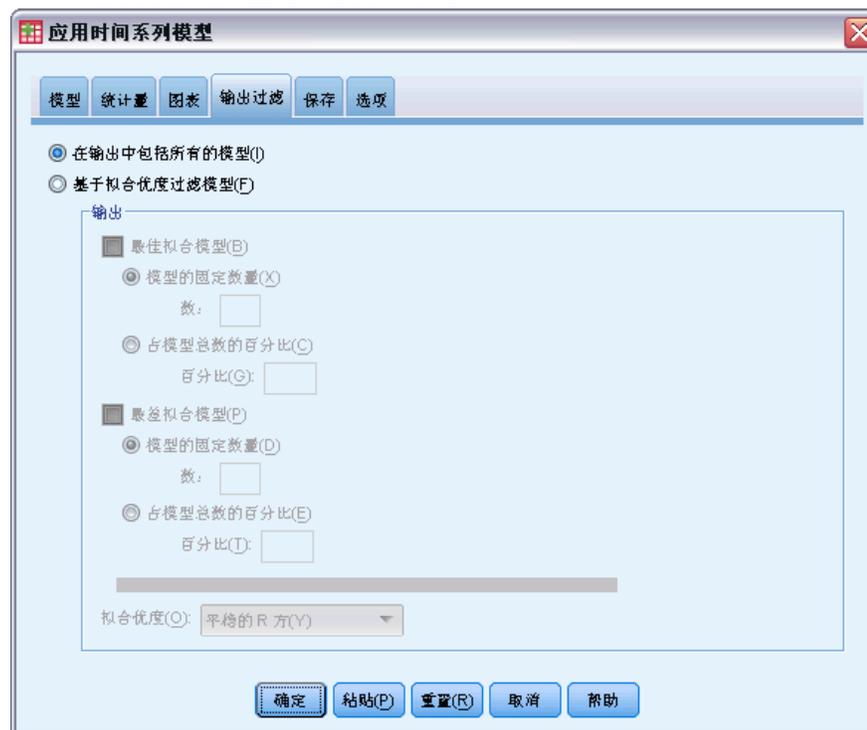
- **观察值。**相依序列的观察值。
- **预测值。**预测期的模型预测值。
- **拟合值。**估计期的模型预测值。
- **预测值的置信区间。**预测期的置信区间。
- **拟合值的置信区间。**估计期的置信区间。

**残差自相关函数 (ACF)。**显示每个估计模型的残差自相关图。

**残差部分自相关函数 (PACF)。**显示每个估计模型的残差部分自相关图。

## 将输出限制为最佳或最差拟合模型

图片 3-4  
应用时间序列模型，?输出过滤器?选项卡



?输出过滤器?选项卡提供的选项可用于将表格输出和图表输出限制为模型的子集。可以选择根据提供的拟合标准将输出限制为最佳和/或最差拟合模型。缺省情况下，输出中包含所有模型。除非重新估计模型参数（?模型?选项卡上的根据数据重新估计），否则，用于过滤模型的拟合测量的值均为模型文件中的那些值，并反映每个模型建立（或上次更新）时使用的数据。

**最佳拟合模型。**选择（选中）此选项将在输出中包含最佳拟合模型。选择拟合优度测量并指定要包含的模型数。选择此选项不妨碍同时选择最差拟合模型。如果同时选择两者，则输出将由最差拟合模型和最佳拟合模型组成。

- **模型的固定数量。**指定为  $n$  个最佳拟合模型显示结果。如果指定的数量超过模型的总数，则显示所有模型。
- **占模型总数的百分比。**指定为其拟合优度值在所有模型的前  $n$  个百分比范围内的模型显示结果。

**最差拟合模型。**选择（选中）此选项将在输出中包含最差拟合模型。选择拟合优度测量并指定要包含的模型数。选择此选项不妨碍同时选择最佳拟合模型。如果同时选择两者，输出将由最佳拟合模型和最差拟合模型组成。

- **模型的固定数量。**指定为  $n$  个最差拟合模型显示结果。如果指定的数量超过模型的总数，则显示所有模型。
- **占模型总数的百分比。**指定为其拟合优度值在所有模型的后  $n$  个百分比范围内的模型显示结果。

**拟合优度。**选择用于过滤模型的拟合优度测量。缺省值为固定的  $R$  方。

## 保存模型预测值和模型规格

图片 3-5  
应用时间序列模型，?保存?选项卡



使用?保存?选项卡可以将模型预测值另存为活动数据集中的新变量，也可以将模型规格以 XML 格式保存到外部文件中。

**保存变量。**可以将模型预测值、置信区间和残差另存为活动数据集中的新变量。每个模型都会带来其自己的一组新变量。如果预测期超出与该模型关联的因变量序列的长度，则添加新个案。除非重新估计模型参数（?模型?选项卡上的**根据数据重新估计**），否则仅为该预测期创建预测值和置信区间上下限。选中每一项所关联的?保存?复选框可以选择保存新变量。缺省情况下不保存新变量。

- **预测值。**模型预测值。
- **置信区间的下限。**预测值的置信区间下限。
- **置信区间的上限。**预测值的置信区间上限。
- **噪声残值。**模型残差。如果执行了因变量转换（例如自然对数），则为转换后的序列的残差。此选项只在重新估计模型参数时可用（?模型?选项卡上的**根据数据重新估计**）。
- **变量名的前缀。**指定要用于新变量名称的前缀或保留缺省的前缀。变量名由前缀、关联因变量的名称和模型标识组成。需要时可以扩展变量名以避免变量命名冲突。前缀必须符合有效变量名称规则。

**导出包含重新估计的参数的模型文件。** 包含重新估计的参数和拟合统计量的模型规格将以 XML 格式导出到指定的文件中。此选项只在重新估计模型参数时可用（?模型?选项卡上的根据数据重新估计）。

## 选项

图片 3-6  
应用时间序列模型, ?选项?选项卡



?选项?选项卡可用于指定缺失值的处理方法，设置置信区间宽度并设置为自相关显示的延迟个数。

**用户缺失值。** 这些选项控制用户缺失值的处理。

- **视为无效。** 用户缺失值当作系统缺失值处理。
- **视为有效。** 用户缺失值当作有效数据处理。

**缺失值策略。** 以下规则用于处理缺失值（包括视为无效的系统缺失值和用户缺失值）：

- 模型中包含估计期发生的因变量存在缺失值的个案。缺失值的具体处理方式取决于估计方法。
- 对于 ARIMA 模型，如果预测值在估计期内有任何缺失值，则将发出警告。不会重新估计任何包含该预测值的模型。
- 如果在预测期内任何自变量有缺失值，建模过程将发出警告并进行最大程度的预测。

**置信区间宽度 (%)。** 将为模型预测值和残差自相关计算置信区间。可以指定小于 100 的任何正数。缺省情况下使用 95% 的置信区间。

**ACF 和 PACF 输出中显示的最大延迟数。**可以设置在自相关和偏自相关表和图中显示的最大延迟数。此选项只在重新估计模型参数时可用（?模型?选项卡上的根据数据重新估计）。

## TSAPPLY 命令的其他功能

如果将选择内容粘贴到语法窗口并编辑生成的 TSAPPLY 命令语法，则可以使用其他功能。使用命令语法语言可以：

- 指定仅将模型文件中的模型子集应用于活动数据集（使用 MODEL 子命令的 DROP 和 KEEP 关键字）。
- 将两个或多个模型文件中的模型应用于数据（使用 MODEL 子命令）。例如，一个模型文件可能包含表示单位销售额的序列的模型，而另一个模型文件则可能包含表示收入的序列的模型。

请参见命令语法参考以获取完整的语法信息。

# 季节性分解

“季节性分解”过程可将一个序列分解成一个季节性成分、一个组合趋势和循环的成分和一个“误差”成分。此过程是对统计方法 I（也称为比率与移动平均数方法）的实现。

**示例。**科学家想要对特定气象站的臭氧层每月测量结果进行分析。目标是确定数据中是否存在任何趋势。为了揭示真实趋势，由于季节性影响，科学家首先需要考虑所读取资料中的变异。可使用“季节性分解”过程来删除任何系统性的季节性变化。然后对季节性调整序列执行趋势分析。

**统计量。** 一组季节性因子。

**数据。** 变量应为数值型。

**假设。** 变量不应包含任何内嵌的缺失数据。至少必须定义一个周期性日期成分。

## 估计季节性因子

- ▶ 从菜单中选择：  
分析 > 预测 > 季节性分解...

图片 4-1  
“季节性分解”对话框



- ▶ 从可用列表选择一个或多个变量，并将其移至“变量”列表中。注意，该列表只包含数值变量。

**模型类型。**“季节性分解”过程提供了用于对季节性因子建模的两种不同方法：乘法或加法。

- **乘法。** 季节性成分是一个因子，用来与经过季节性调整的序列相乘以得到原始序列。实际上，“趋势”会评估与序列的总体水平成正比的季节性成分。无季节性变动的观察值的季节性成分为 1。
- **加法。** 将季节性调整项加到到季节性调整的序列以获取观察值。此调整尝试从序列中移去季节性影响，以查看可能被季节性成分“掩盖”的其他兴趣特征。实际上，“趋势”会评估不依赖于序列的总体水平的季节性成分。无季节性变动的观察值的季节性成分为 0。

**移动平均权重。** “移动平均权重”选项允许您指定在计算移动平均数时如何处理序列。这些选项仅在序列的周期为偶数时才可用。如果周期为奇数，则所有点的权重都相等。

- **所有点都相等。** 使用等于周期的跨度以及所有权重相等的点来计算移动平均数。如果周期是奇数，则始终使用此方法。
- **按 0.5 对端点加权。** 使用等于周期加 1 的跨度以及以 0.5 加权的跨度的端点计算具有偶数周期的序列的移动平均。

根据需要，您可以：

- 单击保存指定如何保存新变量。

## 季节性分解保存

图片 4-2  
“季节保存”对话框



**创建变量。** 允许您选择如何处理新变量。

- **添加到文件。** 将季节性分解所创建的新序列作为常规变量保存在活动数据集中。变量名由三字母前缀、下划线和数字组成。
- **替换已有的。** 由季节性分解创建的新序列在活动数据集中保存为临时变量。同时，将删除由“预测”过程创建的任何现有的临时变量。变量名由一个三个字母的前缀、一个井字符（#）和一个数字组成。
- **不创建。** 不向活动数据集添加新序列。

### 新变量名称

“季节性分解”过程创建了四个新变量（序列），并且每个指定的序列都带有以下三个字母的前缀：

- SAF。** 季节性调整因子。这些值指示每个周期对序列水平的影响。
- SAS。** 季节性调整序列。这些值是在删除序列的季节性变化之后获得的。
- STC。** 平滑的趋势循环成分。这些值显示序列中出现的趋势和循环行为。

**ERR**。残差或“误差”值。这些值是在从序列中删除季节性、趋势和循环成分之后保留的。

## SEASON 命令的其他功能

使用命令语法语言还可以：

- 在 **SEASON** 命令中指定任意周期，而不是在“定义日期”过程所提供的选项中选择一个周期。

请参见命令语法参考以获取完整的语法信息。

# 谱图

“频谱图”过程用于标识时间序列中的周期行为。它不需要分析一个时间点与下一个时间点之间的变异，只要按不同频率的周期性成分分析整体序列的变异。平滑序列在低频率具有更强的周期性成分；而随机变异（“白噪声”）将成分强度分布到所有频率。

不能使用该过程分析包含缺失数据的序列。

**示例。** 建造新住房的比率是一个国家/地区经济的重要晴雨表。有关住房的数据开始时通常会表现出一个较强的季节性成分。但在估计当前数字时，分析人员需要注意数据中是否呈现了较长的周期。

**统计量。** 正弦和余弦变换、周期图值和每个频率或周期成分的谱密度估计。在选择双变量分析时：交叉周期图的实部和虚部、余谱密度、正交谱、增益、平方一致和每个频率或周期成分的相位谱。

**图。** 对于单变量和双变量分析：周期图和频谱密度。对于双变量分析：平方一致性、正交谱、交叉振幅、余谱密度、相位谱和增益。

**数据。** 变量应为数值型。

**假设。** 变量不应包含任何内嵌的缺失数据。要分析的时间序列应该是平稳的，任何非零均值应该从序列中删除。

- **平稳。** 要用 ARIMA 模型进行拟合的时间序列所必须满足的条件。纯的 MA 序列是平稳的，但 AR 和 ARMA 序列可能不是。平稳序列的均值和方差不随时间改变。

## 获取谱分析

- ▶ 从菜单中选择：  
分析 > 时间序列 > 频谱分析...

图片 5-1  
“频谱图”对话框



- ▶ 从可用列表选择一个或多个变量，并将它们移至“变量”列表中。注意，该列表只包含数值变量。
- ▶ 选择其中一个“频谱窗口”选项来选择如何平滑周期图，以便获得谱密度估计值。可用的平滑选项有“Tukey-Hamming”、“Tukey”、“Parzen”、“Bartlett”、“Daniell（单元）”和“无”。
  - **Tukey-Hamming.** 权重为  $W_k = .54D_p(2 \pi f_k) + .23D_p(2 \pi f_k + \pi/p) + .23D_p(2 \pi f_k - \pi/p)$ ,  $k = 0, \dots, p$ , 其中  $p$  是一半跨度的整数部分,  $D_p$  是阶数  $p$  的 Dirichlet 内核。
  - **Tukey.** 权重为  $W_k = 0.5D_p(2 \pi f_k) + 0.25D_p(2 \pi f_k + \pi/p) + 0.25D_p(2 \pi f_k - \pi/p)$ ,  $k = 0, \dots, p$ , 其中  $p$  是一半跨度的整数部分,  $D_p$  是阶数  $p$  的 Dirichlet 内核。
  - **Parzen.** 权重为  $W_k = 1/p(2 + \cos(2 \pi f_k)) (F[p/2](2 \pi f_k))^{**2}$ ,  $k = 0, \dots, p$ , 其中  $p$  是一半跨度的整数部分, 而  $F[p/2]$  是阶数  $p/2$  的 Fejer 内核。
  - **Bartlett.** 谱窗口的形状, 窗口上半部分的权重按如下公式计算:  $W_k = F_p(2 * \pi * f_k)$ ,  $k = 0, \dots, p$ , 其中  $p$  是半跨度的整数部分,  $F_p$  是阶数  $p$  的 Fejer 内核。下半部分与上半部分对称。
  - **Daniell (单元).** 所有权重均等于 1 的频谱窗口形状。
  - **无.** 无平滑。如果选择了此选项, 则频谱密度估计与周期图相同。

**跨度.** 一个连续值范围, 在该范围上将执行平滑。通常使用奇数。较大的跨度对谱密度图进行的平滑比较小的跨度程度大。

**变量中心化.** 调整序列以使在计算谱之前其均值为 0, 并且移去可能与序列均值关联的较大项。

**双变量分析—将第一个变量与每个变量。** 如果选择两个或多个变量，则可以选择此选项来请求双变量谱分析。

- “变量”列表中的第一个变量将被视为自变量，而其他所有变量将被视为因变量。
- 系统将使用独立于其他序列命名的第一个序列对第一个序列之后的每个序列进行分析。还会对每个序列执行单变量分析。

**图。** 周期图和谱密度对单变量分析和双变量分析均可用。其他所有选项仅对双变量分析可用。

- **周期图。** 针对频率或周期绘制的未平滑谱振幅图（绘制在对数刻度中）。低频率变动是平滑序列的特征。均匀地分布在所有频率上的变动则表示“白噪音”。
- **平方一致性。** 两个序列的增益的乘积。
- **正交谱。** 交叉周期图的虚部，是两个时间序列的异相频率成分的相关性的测量。成分的异相为  $\pi/2$  弧度。
- **交叉振幅。** 余谱密度平方和正交谱平方之和的平方根。
- **谱密度。** 已进行平滑而移去了不规则变动的周期图。
- **余谱密度。** 交叉周期图的实部，是两个时间序列的同相频率分量的相关性的测量。
- **相位谱。** 一个序列的每个频率成分提前或延迟另一个序列的程度的测量。
- **增益。** 用一个序列的谱密度除以跨振幅的商。这两个序列都有自己的获得值。

**按频率。** 所有图都是根据频率生成的，范围从频率 0（常数或中项）到频率 0.5（两个观察值的循环项）。

**按周期。** 所有图都是按周期生成的，范围从 2（两个观察值的循环项）到等于观察值数量的期间（常数或中项）。周期显示在对数刻度上。

## SPECTRA 命令的其他功能

使用命令语法语言还可以：

- 将计算的谱分析变量保存到活动数据集中以供将来使用。
- 为谱窗指定定制权重。
- 按频率和周期生成图。
- 打印图中显示的每个值的完整列表。

请参见命令语法参考以获取完整的语法信息。

# 部分 II:

## 示例

# 使用专家建模器进行批量预测

全国宽带提供商的分析员需要对用户订阅进行预测，以便预测带宽的利用率。需要对组成全国订户库的 85 个本地市场中的每个市场进行预测。每月历史数据收集在 broadband\_1.sav 中。有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。

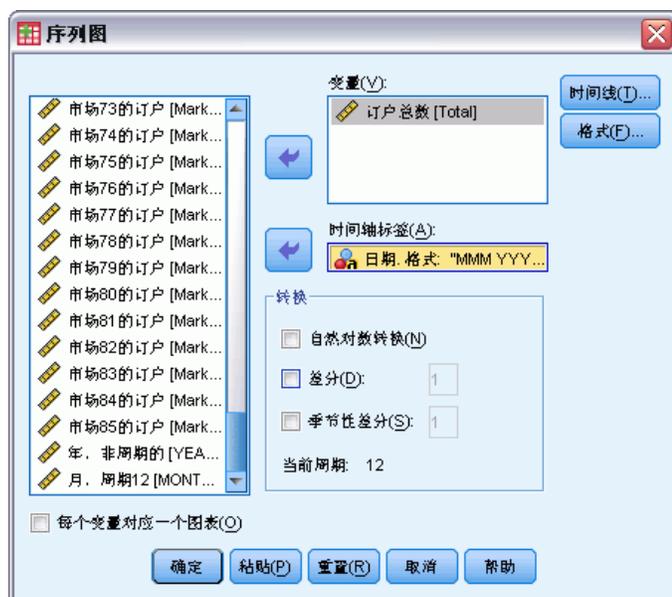
在本示例中，您将使用专家建模器对每个本地市场（共 85 个）在接下来的三个月的趋势进行预测，并将生成的模型保存到外部 XML 文件。一旦完成操作，即可以进行下一个示例通过应用保存的模型重新进行批量预测，第 7 章第 52 页码，该示例将保存的模型应用于一个更新的数据集，以便延伸预测三个月而无需重新构建模型。

## 检查数据

在构建模型前了解数据的性质始终是一种好的做法。数据是否表现出季节性变化？尽管“专家建模器”将自动为每个序列查找最佳的季节性或非季节性模型，但是，如果数据中不存在季节性，通过限制对非季节性模型的搜索通常可更快地获得结果。不需要对这 85 个本地市场中的每个市场的数据进行检查，只需通过绘制所有市场的总订户数量图，即可了解总体概况。

- ▶ 从菜单中选择：  
分析 > 预测 > 序列图...

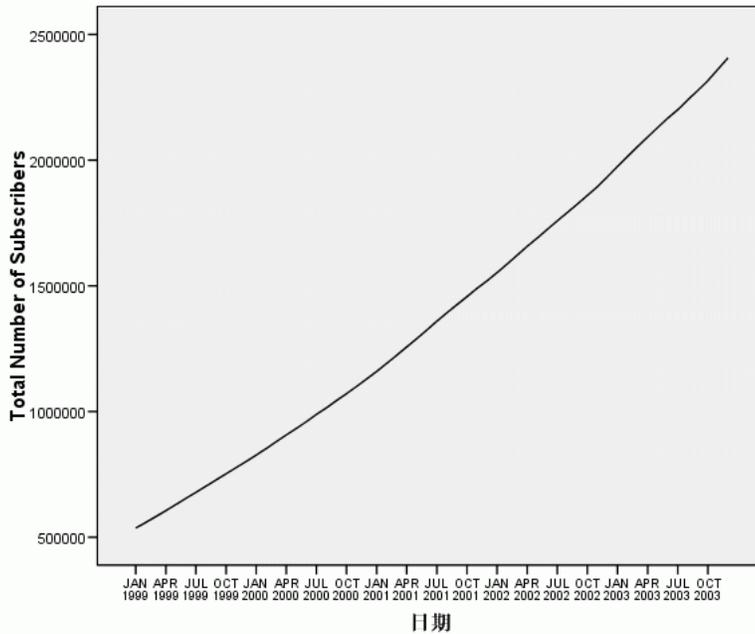
图片 6-1  
“序列图”对话框



- ▶ 选择 Total Number of Subscribers 并将其移到“变量”列表中。

- ▶ 选择 Date 并将其移到“时间轴标签”框中。
- ▶ 单击确定。

图片 6-2  
所有市场中的宽带订户总数



此序列表现出了一个很平滑的上升趋势，没有出现季节性变化的迹象。可能在个别序列中存在季节性，但总的看来，季节性不是数据的显著特征。当然，应该在排除季节性模型之前检查每个序列。然后分离出表现季节性的序列，并分别对这些序列建模。在提供的个案中，对 85 个序列进行检查得到的结果是，没有任何序列表现出季节性。

## 运行分析

使用专家建模器：

- ▶ 从菜单中选择：  
分析 > 预测 > 创建模型...

图片 6-3  
“时间序列建模器”对话框



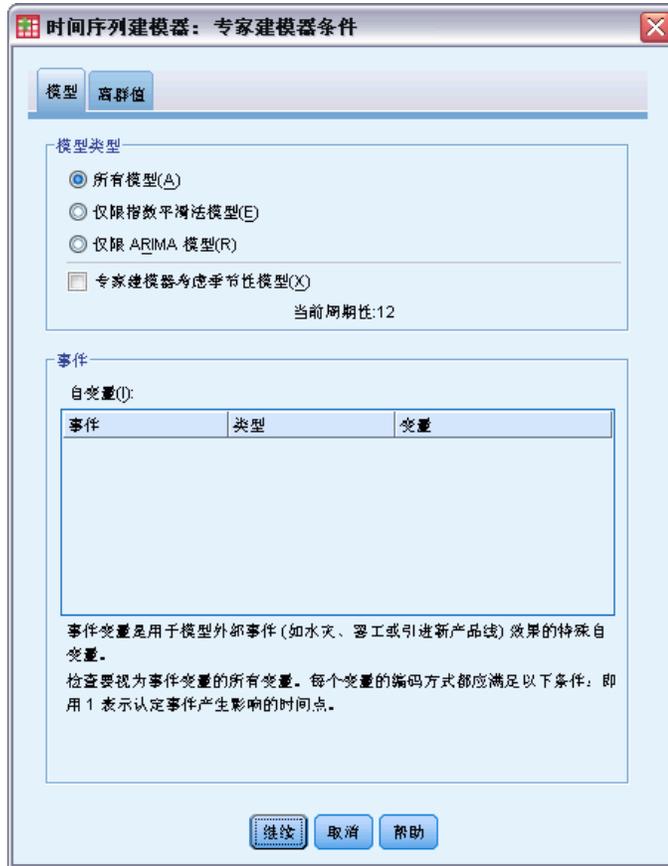
- ▶ 选择 Subscribers for Market 1 到 Subscribers for Market 85 作为因变量。
- ▶ 验证在“方法”下拉列表表中选择了专家建模器。“专家建模器”将自动为每个因变量序列查找最佳拟合模型。

用于估计模型的个案组称为**估计期**。默认情况下，它包含活动数据集中的所有个案。您可以通过在“选择个案”对话框中选择基于时间或个案范围来设置估计期。对于本示例，我们将继续使用默认值。

还要注意，默认的预测期是在估计期结束之后开始的，并持续到活动数据集中的最后一个个案。如果在最后一个个案之后进行预测，则需要延伸预测期。此操作需要在“选项”选项卡中完成，本示例稍后将进行介绍。

- ▶ 单击条件。

图片 6-4  
“专家建模器条件”对话框，“模型”选项卡



- ▶ 在“模型类型”组中，取消选择专家建模器考虑季节性模型。

尽管数据是按月份收集的，并且当前周期性为 12，但我们还是看到了这些数据没有表现出任何季节性，因此，不需要考虑季节性模型。这将减少“专家建模器”搜索的模型空间，并大幅度减少计算时间。

- ▶ 单击继续。
- ▶ 单击“时间序列建模器”对话框中的选项选项卡。

图片 6-5  
时间序列建模器，“选项”选项卡



- ▶ 在“预测期”组中选择估计期结束之后到指定日期之间的第一个个案。
- ▶ 在“日期”网格中，输入 2004 作为年份，3 作为月份。  
数据集包含从 1999 年 1 月到 2003 年 12 月之间的数据。使用当前设置，预测期将是 2004 年 1 月到 2004 年 3 月。
- ▶ 单击保存选项卡。

图片 6-6  
时间序列建模器，“保存”选项卡



- ▶ 选择（选中）保存列中的“预测值”条目，并保留默认值预测值作为“变量名的前缀”。  
通过对变量名使用前缀预测值，模型预测将在活动数据集中另存为新的变量。还可以将每个模型的规格保存到外部 XML 文件中。这将允许您在新数据可用时重新使用这些模型来延伸预测。
- ▶ 单击“保存”选项卡上的浏览按钮。  
此操作将向您显示用于保存文件的标准对话框。
- ▶ 浏览至要保存 XML 模型文件的文件夹，输入文件名，然后单击保存。
- ▶ 单击统计量选项卡。

图片 6-7  
时间序列建模器，“统计量”选项卡



- 选择显示预测值。

此选项将为每个因变量序列生成一个预测值表，并提供另一选项（而不是将预测值保存为新变量）来获得这些值。

拟合优度（位于“比较模型的统计量”组中）的默认选择可生成一个包含以下拟合统计量的表：如 R 方、平均绝对误差百分比和标准化 BIC，这些统计量是根据所有模型计算得出的。它提供了模型与数据之间的拟合程度的简明摘要。

- 单击图选项卡。

图片 6-8  
时间序列建模器，“图”选项卡



- ▶ 在“单个模型图”组中取消选择序列。

这将禁止为每个模型生成序列图。在本例中，我们更关心的是将预测值另存为新变量，而不是生成预测值图。

“模型比较图”组提供了多个显示根据所有模型计算得出的拟合统计量的图（以直方图形式）。

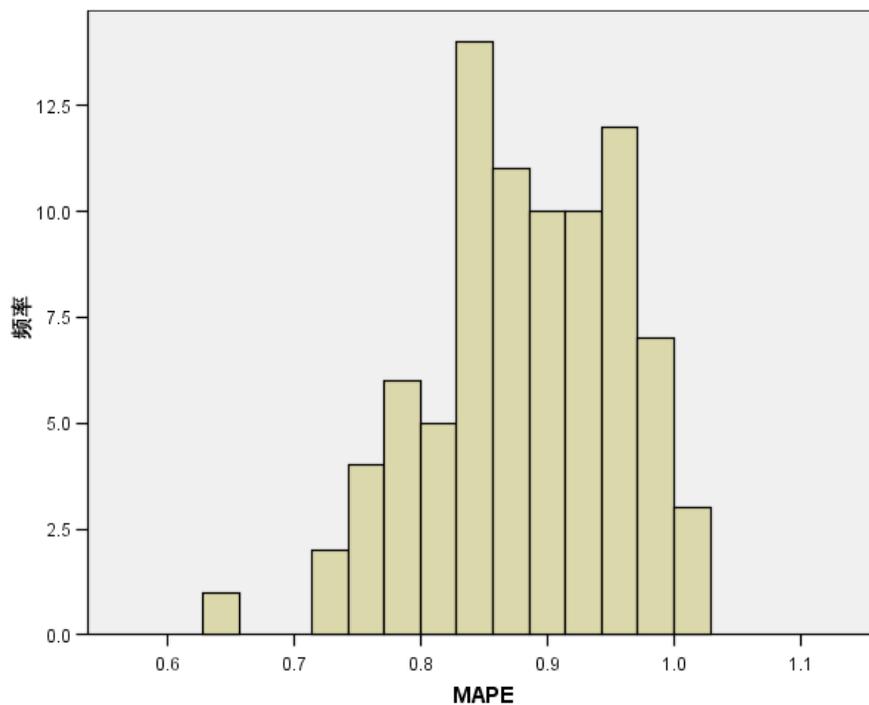
- ▶ 在“模型比较图”组中选择平均绝对误差百分比和最大绝对误差百分比。

绝对误差百分比是对因变量序列与其模型预测水平的相差程度的测量。通过检查所有模型的均值和最大值，可获得预测中不确定性的指示。由于因变量序列表示不同规模市场的订户数量，因此建议查看误差百分比的摘要图，而不是查看绝对误差。

- ▶ 单击“时间序列建模器”对话框中的确定。

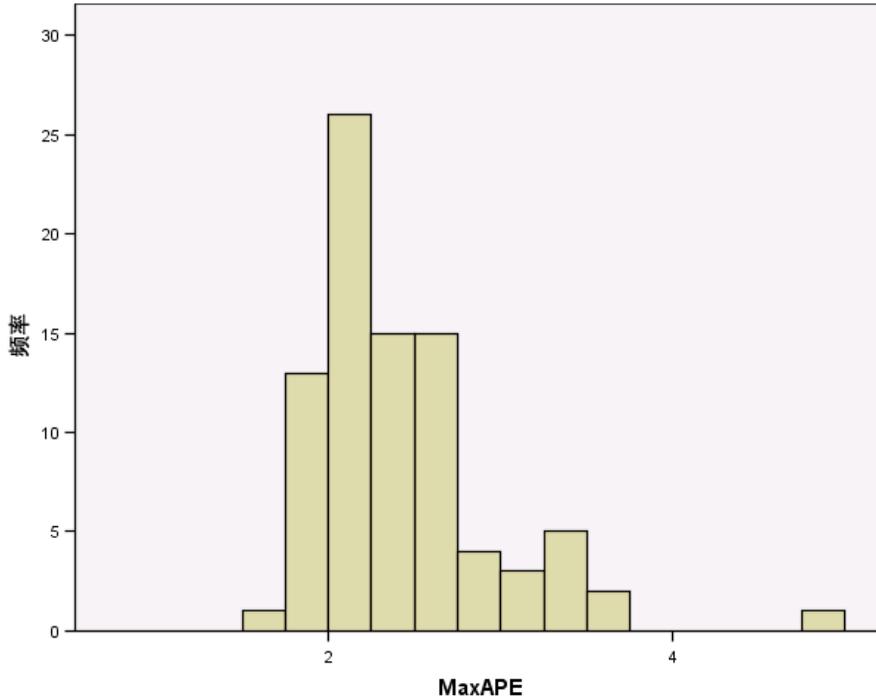
## 模型摘要图

图片 6-9  
平均绝对误差百分比的直方图



此直方图显示所有模型的平均绝对误差百分比 (MAPE)。它表明，所有模型都显示约为 1% 的平均不确定性。

图片 6-10  
最大绝对误差百分比的直方图



此直方图显示所有模型的最大绝对误差百分比 (MaxAPE)，它对于设想预测的最坏情况方案非常有用。它表明，每个模型的最大误差百分比介于 1% 到 5% 之间。这些值是否代表可接受的不确定性程度？由于可接受的风险随问题而变化，因此应根据业务情况确定可接受的不确定性值。

## 模型预测值

图片 6-11  
包含模型预测的新变量

YEAR_	MONTH_	DATE_	Predicted_Market_1_Model_1	Predicted_Market_2_Model_2
2003	10	OCT 2003	11820	51804
2003	11	NOV 2003	11857	51273
2003	12	DEC 2003	11687	53082
2004	1	JAN 2004	11503	54893
2004	2	FEB 2004	11447	55856
2004	3	MAR 2004	11390	56704

数据编辑器显示了包含模型预测的新变量。虽然此处只显示了两个新变量，但实际上有 85 个新变量，分别与 85 个相依序列对应。变量名称由默认前缀预测值、关联因变量的名称（例如 Market\_1）和模型标识（例如 Model\_1）顺序构成。

包含 2004 年 1 月到 2004 年 3 月的预测的三个新个案已添加到了数据集中，并且还带有自动生成的日期标签。每个新变量都包含估计期（1999 年 1 月到 2003 年 12 月）的模型预测，从而允许您查看模型与已知值之间的拟合程度。

图片 6-12  
预测表

模型		2004 年 1 月	2004 年 2 月	2004 年 3 月
Subscribers for Market 1-模型_1	预测	11503	11447	11390
	UCL	11686	11767	11870
	LCL	11321	11126	10910
Subscribers for Market 2-模型_2	预测	54893	55856	56704
	UCL	55632	57195	58575
	LCL	54154	54518	54832
Subscribers for Market 3-模型_3	预测	59656	59305	58954
	UCL	60457	60753	61158
	LCL	58856	57857	56750
Subscribers for Market 4-模型_4	预测	18235	18424	18628
	UCL	18413	18731	19121
	LCL	18058	18116	18136

还可以选择使用预测值创建表。表中包含预测期中的预测值，但与包含模型预测的新变量不同，它不包含估计期中的预测值。按模型组织所得结果，并用模型名称标识这些结果，模型名称由关联因变量的名称（或标签）后跟模型标识组成，就像包含模型预测的新变量的名称一样。此表还包括预测值（默认情况下为 95%）的置信区间上限（UCL）和置信区间下限（LCL）。

您现在已了解了获取预测值的两种方法：在活动数据集中将预测值另存为新变量和创建预测表。使用任一方法，都会为您提供一些可用于导出预测的选项（例如，导出到 Excel 电子表格）。

## 摘要

您已学习了如何使用专家建模器为多个序列生成预测，并且已将所得的结果模型保存到了外部 XML 文件中。在下一个示例中，您将学习如何在新数据可用时使用“应用时间序列模型”过程延伸预测，而无需重新构建模型。

# 通过应用保存的模型重新进行批量预测

您已经使用“时间序列建模器”为时间序列数据创建模型，并根据可用数据生成初始预测。您计划在具有更多可用的当前数据时重新使用这些模型来延伸您的预测，因此您将这些模型保存到一个外部文件中。现在您就可以应用这些保存的模型了。

本示例实质上是前一个示例“[使用专家建模器进行批量预测 in 第 6 章第 41 页码](#)”的扩展，但也可以独立使用。在此方案中，您是全国宽带提供商的分析员，负责为 85 个本地市场中的每个市场生成每月用户订阅预测。您已经使用专家建模器创建了模型，要对未来的三个月进行预测。您的数据仓库已经使用了原始预测期的实际数据进行了刷新，因此可以使用该数据将预测水平延伸三个月。

更新的每月历史数据收集在 broadband\_2.sav 中，保存的模型位于 broadband\_models.xml 中。[有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。](#)当然，如果完成了上一示例，并保存了自己的模型文件，则可以使用您自己的模型文件，而不使用 broadband\_models.xml。

## 运行分析

应用模型：

- ▶ 从菜单中选择：  
分析 > 预测 > 应用模型...

图片 7-1  
“应用时间序列模型”对话框



- ▶ 单击浏览，导航到并选择 broadband\_models.xml（或选择在上一示例中保存的您自己的模型文件）。有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。
- ▶ 选择根据数据重新估计。  
 要将新的时间序列值合并到预测中，“应用时间序列模型”过程将必须重新估计模型参数。但模型的结构保持不变，因此，要重新估计的计算时间比构建模型所需的原始计算时间短得多。  
 用于重新估计的这组个案需要包含这些新数据。如果您使用“第一个个案”到“最后一个个案”的缺省估计期，则可以确保这一点。如果还需要将估计期设置为非缺省的估计期，则可以通过在“选择个案”对话框中选择基于时间或个案范围来进行此操作。
- ▶ 在“预测期”组中选择估计期结束之后到指定日期之间的第一个个案。
- ▶ 在“日期”网格中，输入 2004 作为年份，6 作为月份。  
 数据集包含从 1999 年 1 月到 2004 年 3 月之间的数据。使用当前设置，预测期将是 2004 年 4 月到 2004 年 6 月。
- ▶ 单击保存选项卡。

图片 7-2  
应用时间序列模型，“保存”选项卡



- ▶ 选择（选中）保存列中的“预测值”条目，并保留缺省值预测值作为“变量名的前缀”。通过对变量名使用前缀预测值，模型预测将在活动数据集中另存为新的变量。
- ▶ 单击图选项卡。

图片 7-3  
应用时间序列模型，“图”选项卡



- ▶ 在“单个模型图”组中取消选择序列。

这将禁止为每个模型生成序列图。在本例中，我们更关心的是将预测值另存为新变量，而不是生成预测值图。

- ▶ 单击“应用时间序列模型”对话框中的确定。

## 模型拟合统计量

图片 7-4  
模型拟合表

拟合统计量	均值	SE	最小值	最大值	百分点						
					5	10	25	50	75	90	95
平滑 R 方	.181	.141	-4.2E-015	.609	-8.2E-016	-4.4E-016	.070	.193	.252	.359	.453
R 方	.999	.000	.998	1.000	.999	.999	.999	.999	.999	.999	1.000
RMSE	187.766	149.959	45.048	764.698	52.558	59.848	93.155	138.913	205.742	433.883	493.904
MAPE	.886	.074	.669	1.026	.754	.792	.832	.894	.948	.976	.996
MaxAPE	2.446	.512	1.742	4.373	1.840	1.952	2.103	2.321	2.663	3.313	3.676
MAE	146.373	113.749	35.573	612.989	41.209	48.742	73.553	111.569	162.306	331.207	364.582
MaxAE	499.714	434.112	116.966	2143.993	131.261	144.918	226.960	345.221	574.071	1165.207	1587.793
正态化的 BIC	10.086	1.343	7.749	13.412	8.049	8.285	9.193	9.968	10.717	12.210	12.494

“模型拟合”表提供了对所有模型计算的拟合统计量。它提供了使用重新估计参数的模型与数据之间的拟合程度的简明摘要。对于每个统计量，该表提供了所有模型的均值、标准误 (SE)、最小值和最大值。它还包含提供有关模型的统计量分布的信息的百分位

值。对于每个百分位，模型百分比具有一个小于规定值的拟合统计量值。例如，95% 的模型具有一个小于 3.676 的 MaxAPE 值（最大绝对误差百分比）。

在报告统计量数量时，我们将关注以下两项：MAPE（平均绝对误差百分比）和 MaxAPE（最大绝对误差百分比）。绝对误差百分比是相依序列与其模型预测水平的差异程度的测量，并提供有关预测不确定性的指示。所有模型的平均绝对误差百分比的范围是：最小值为 0.669%，最大值为 1.026%。所有模型的最大绝对误差百分比的范围是：最小值为 1.742%，最大值为 4.373%。因此，每个模型的预测中的平均不确定性约为 1%，最大不确定性约为 2.5%（MaxAPE 的平均值），最坏情况方案的不确定性约为 4%。这些值是否表示可接受的不确定性程度取决于您愿意接受的风险度。

## 模型预测值

图片 7-5  
包含模型预测的新变量

YEAR_	MONTH_	DATE_	Predicted_Market_1_Model_1	Predicted_Market_2_Model_2
2004	1	JAN 2004	11513	54947
2004	2	FEB 2004	11806	56810
2004	3	MAR 2004	11950	57344
2004	4	APR 2004	12312	59631
2004	5	MAY 2004	12501	60717
2004	6	JUN 2004	12689	61659

数据编辑器显示了包含模型预测的新变量。虽然此处只显示了两个新变量，但实际上有 85 个新变量，分别与 85 个相依序列对应。变量名称由缺省前缀预测值、关联因变量的名称（例如 Market\_1）和模型标识（例如 Model\_1）顺序构成。

包含 2004 年 4 月到 2004 年 6 月的预测的三个新个案已添加到数据集中，并且还带有自动生成的日期标签。

## 摘要

您已经了解了如何在具有更多可用的当前数据时应用保存的模型来延伸您以前的预测。并且，您已在没有重新构建模型的情况下做到了这一点。当然，如果考虑到模型已更改的原因，则应该使用“时间序列建模器”过程重新构建模型。

# 使用专家建模器确定重要预测值

一个需要开发预测模型的编目公司已收集了有关男式服装月销售量的数据，以及用于解释一些销售变化的几个序列。可能的预测变量包括邮寄的目录数、目录中的页数、订购热线的条数、印刷广告花费的金额以及客户服务代表的数量。这些预测变量是否对预测都有用？

在本示例中，您将结合使用“专家建模器”和所有的候选预测变量来找出最佳模型。由于“专家建模器”仅选择与因变量序列有显著统计关系的那些预测变量，因此您将知道哪些预测变量有用，并将获得一个使用这些预测变量的预测模型。一旦完成操作，即可以进行下一个示例[通过应用保存的模型试验预测值](#)，第 9 章第 66 页码，该示例将使用此示例中构建的模型调查不同预测变量方案下销售所受到的影响。

当前样本的数据收集在 catalog\_seasfac.sav 中。[有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。](#)

## 对数据进行绘图

对数据进行绘图始终是一种好的做法，特别是在仅处理一个序列时更是如此：

- ▶ 从菜单中选择：  
分析 > 预测 > 序列图...

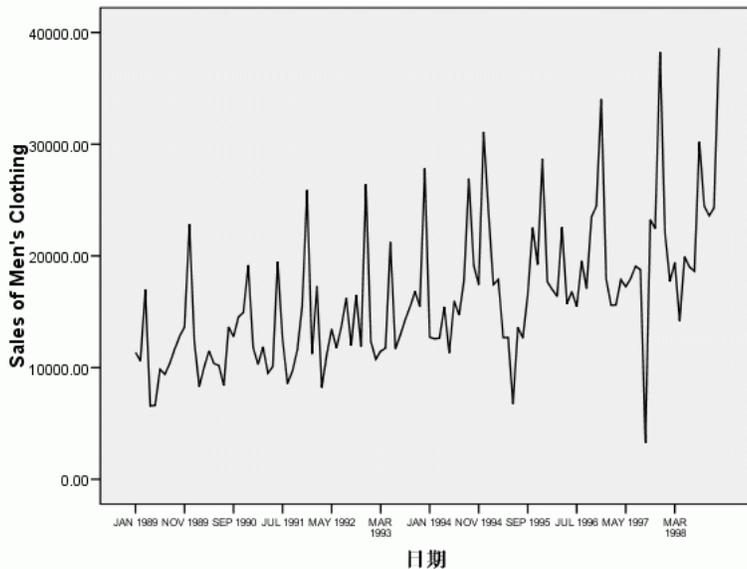
图片 8-1  
“序列图”对话框



- ▶ 选择 Sales of Men' s Clothing 并将其移到“变量”列表中。
- ▶ 选择 Date 并将其移到“时间轴标签”框中。

- ▶ 单击确定。

图片 8-2  
Sales of men' s clothing (美元)



此序列表现出了大量的峰值，其中的许多峰值以等间距显示，并且有明显的上升趋势。等间距峰值表明时间序列中存在周期性成分。已知销售具有季节性，通常在节假日期间出现频率很高，您应该不会对在数据中发现年度季节性成分感到奇怪。

还存在一些并不作为季节性模式一部分显示的峰值，并与相邻数据点之间呈现出明显偏差。这些点可能是离群值，可以并应该通过“专家建模器”来进行确定。

## 运行分析

使用专家建模器：

- ▶ 从菜单中选择：  
分析 > 预测 > 创建模型...

图片 8-3  
“时间序列建模器”对话框



- ▶ 选择 Sales of Men' s Clothing 作为因变量。
- ▶ 选择 Number of Catalogs Mailed 到 Number of Customer Service Representatives 作为自变量。
- ▶ 验证在“方法”下拉列表中选择了专家建模器。专家建模器将自动为因变量序列查找最佳拟合季节性或非季节性模型。
- ▶ 单击条件，然后单击离群值选项卡。

图片 8-4  
“专家建模器条件”对话框，“离群值”选项卡



- ▶ 选择自动检测离群值，然后保留默认选择作为要检测的离群值类型。

我们对数据进行的直观检查表明可能存在离群值。使用当前选择，“专家建模器”将搜索最常用的离群值类型，并将所有离群值合并到最终的模型中。离群值检测会明显增加专家建模器所需的计算时间，因此在使用此功能时需要慎重考虑，特别是在一次性对许多序列进行建模时更是如此。默认情况下，不对离群值进行检测。

- ▶ 单击继续。
- ▶ 单击“时间序列建模器”对话框中的保存选项卡。

图片 8-5  
时间序列建模器，“保存”选项卡



您要将估计模型保存到外部 XML 文件中，以便使用“应用时间序列模型”过程试验不同的预测变量值，而无需重新构建模型。

- ▶ 单击“保存”选项卡上的浏览按钮。  
此操作将向您显示用于保存文件的标准对话框。
- ▶ 浏览至要保存 XML 模型文件的文件夹，输入文件名，然后单击保存。
- ▶ 单击统计量选项卡。

图片 8-6  
时间序列建模器，“统计量”选项卡



- ▶ 选择参数估计。

此选项将生成一张表，显示“专家建模器”所选模型的所有参数，其中包括显著性预测变量。

- ▶ 单击图选项卡。

图片 8-7  
时间序列建模器，“图”选项卡



- ▶ 取消选择预测值。

在当前示例中，我们只需要确定显著性预测变量和构建模型。我们将不做任何预测。

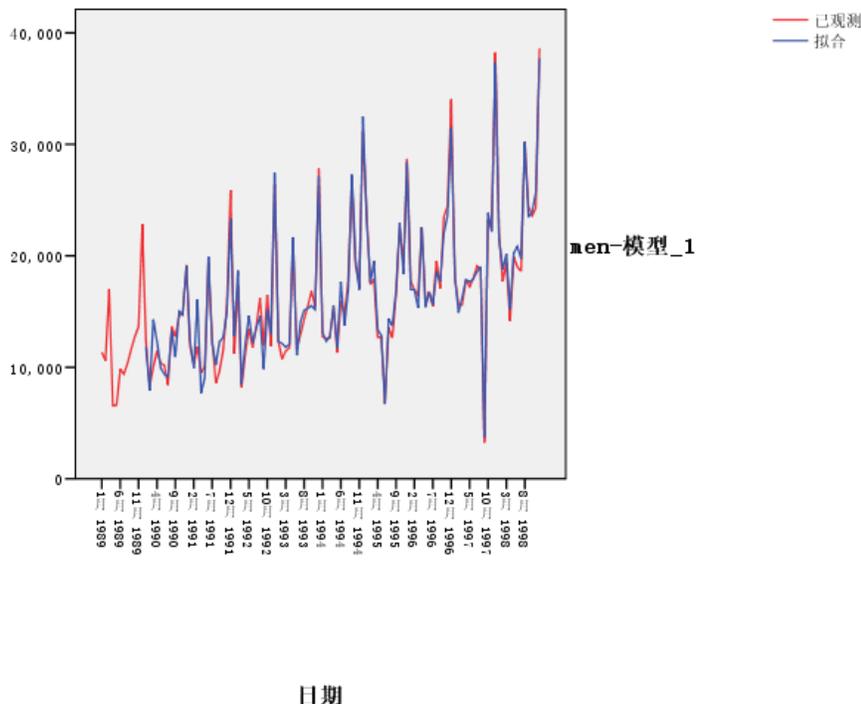
- ▶ 选择拟合值。

此选项将显示估计模型所用期间的预测值。此期间称为**估计期**，它包含此示例的活跃数据集中的所有个案。这些值可以指示模型与观察值之间的拟合程度，因此，这些值被称为**拟合值**。生成的图将由观察值和拟合值组成。

- ▶ 单击“时间序列建模器”对话框中的**确定**。

## 序列图

图片 8-8  
预测值和观察值



预测值与观察值表现出了良好的符合性，表示该模型具有令人满意的预测能力。请注意模型对季节性峰值的预测性能。因为它是获取数据上升趋势的好办法。

## 模型描述表

图片 8-9  
模型描述表

模型 ID	Sales of Men's Clothing	模型_1	模型类型
			ARIMA(0,0,0)(0,1,0)

模型描述表中包含每个估计模型的条目，还包含模型标识和模型类型。模型标识由相关因变量的名称（标签）和系统分配的名称组成。在当前示例中，因变量为 Sales of Men's Clothing，系统分配的名称为 Model\_1。

时间序列建模器支持指数平滑法和 ARIMA 模型。指数平滑法模型类型按其常用名称列出，如 Holt 模型和 Winters 可加的模型。ARIMA 模型类型使用 ARIMA 的标准符号 (p, d, q) (P, D, Q) 列出，其中，p 为自回归的阶，d 为差分（或积分）的阶，q 为滑动平均的阶，(P, D, Q) 为这些阶的季节性对应符号。

专家建模器已确定具有一阶差分的季节性 ARIMA 模型可对男式服装销售量进行最佳描述。模型的季节性性质考虑了我们在序列图中看到的季节性峰值，并且单阶差分反映了数据中呈明显的上升趋势。

## 模型统计表

图片 8-10  
模型统计表

模型	预测变量数	模型拟合 统计量	Ljung-Box Q(18)			离群值数
		平稳 R 方	统计量	DF	Sig.	
Sales of Men's Clothing 模型_1	2	.948	7.589	18	.984	9

模型统计表为每个估计模型提供了摘要信息和拟合优度统计量。每个模型的结果使用模型描述表中提供的模型标识进行标记。首先，请注意模型中包含最初指定的五个候选预测变量中的两个预测变量。它表明“专家建模器”已标识出了证明可能对预测有用的两个自变量。

尽管时间序列建模器提供了许多不同的拟合优度统计量，但我们只选择固定的 R 方值。如此处的个案所示，在存在趋势或季节性模式时，此统计量提供了序列中总变异比例估计值，该估计值由模型解释，并优先使用普通 R 方。固定的 R 方值越大（最大值为 1），表明拟合度越好。值 0.948 意味着模型对解释序列中观察到的变异起到了很好的作用。

Ljung-Box 统计量也称作改进的 Box-Pierce 统计量，它可指示是否正确指定了模型。显著性值小于 0.05 表示在观察到的序列中存在该模型没有考虑到的结构。值 0.984 表明此处没有显著性，因此，我们可以确信已正确指定了模型。

“专家建模器”检测到九个被认为是离群值的点。已相应地对这些点进行了建模，因此，不需要从序列中移去这些点。

## ARIMA 模型参数表

图片 8-11  
ARIMA 模型参数表

				估计	SE	t	Sig.
Sales of Men's Clothing 模型_1	Sales of Men's Clothing	无转换	季节性差分	1			
	Number of Catalogs Mailed	无转换	分子 季节性差分	1.549	.071	21.943	.000
			滞后 0	1			
	Number of Phone Lines Open for Ordering	无转换	分子 季节性差分	315.282	15.298	20.607	.000
			滞后 0	1			

ARIMA 模型参数表显示模型中所有参数的值，并为每个由模型标识标记的估计模型显示一个条目。在本示例中，它将列出模型中的所有变量，包括“专家建模器”已确定具有显著性的因变量和任何自变量。我们已经从模型统计表中了解到存在两个显著性预测变量。模型参数表显示，这两个显著性预测变量为 Number of Catalogs Mailed 和 Number of Phone Lines Open for Ordering。

## 摘要

您已经学习了如何使用专家建模器来构建模型并标识显著性预测变量，并且已将所得的结果模型保存到了外部文件中。现在，您可以使用“应用时间序列模型”过程来试验预测变量序列的其他方案，并查看其他方案对销售预测有何影响。

# 通过应用保存的模型试验预测值

您已经使用“时间序列建模器”为时间序列数据创建模型，并标识哪些预测变量已被证明对预测有用。预测变量表示在您控制范围内的因子，因此，您可以在预测期中试验其值，以查看因变量的预测受到怎样的影响。使用“应用时间序列模型”过程可以很容易地完成此任务，方法是通过使用“时间序列建模器”过程创建的模型文件。

本示例实质上是前一个示例“[使用专家建模器确定重要预测值 in 第 8 章第 57 页码](#),”的扩展，但也可以独立使用。此方案涉及一个编目公司，该公司收集了有关 1989 年 1 月到 1998 年 12 月期间男式服装的月销售量的数据，还收集了一些认为是有用的序列作为将来销售的预测变量。专家建模器已确定，五个候选预测变量中只有两个变量是有意义的：邮寄的目录个数和订购热线的条数。

在计划下一年度的销售策略时，您用来印制目录和保持订购热线的资源是有限的。1999 年前三个月的预算允许在初始突出线上增加 2000 个目录或 5 个订购热线。在这三个月的销售期间，哪种选择可以产生更多的销售收入？

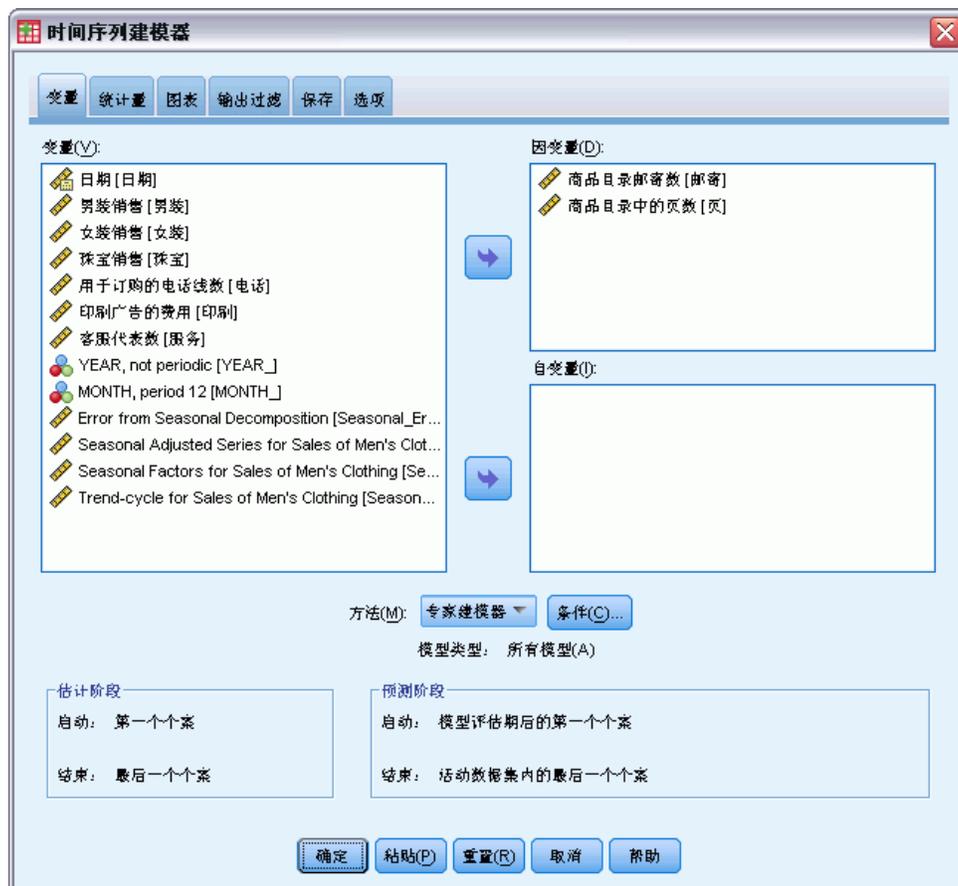
本示例的数据收集在 catalog\_seasfac.sav 中，并且 catalog\_model.xml 中包含使用专家建模器构建的月销售模型。[有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20](#)。当然，如果完成了上一示例，并保存了自己的模型文件，则可以使用您自己的模型文件，而不使用 catalog\_model.xml。

## 延伸预测变量序列

在您使用预测变量为因变量序列创建预测时，每一个预测变量序列都需要在预测期中延伸。除非您很准确地知道将来的预测变量的值是什么，否则，您需要估计这些值。然后，您可修改这些估计值，以检验不同的预测变量方案。使用专家建模器可以很容易地创建初始突出。

- ▶ 从菜单中选择：  
分析 > 预测 > 创建模型...

图片 9-1  
“时间序列建模器”对话框



- ▶ 选择 Number of Catalogs Mailed 和 Number of Phone Lines Open for Ordering 作为因变量。
- ▶ 单击保存选项卡。

图片 9-2  
时间序列建模器，“保存”选项卡



- ▶ 在保存列中，选择（选中）“预测值”条目，并保留缺省值预测值作为“变量名的前缀”。
- ▶ 单击选项选项卡。

图片 9-3  
时间序列建模器，“选项”选项卡



- ▶ 在“预测期”组中选择估计期结束之后到指定日期之间的第一个个案。
- ▶ 在“日期”网格中，输入 1999 作为年份，3 作为月份。  
数据集包含从 1989 年 1 月到 1998 年 12 月之间的数据。因此，使用当前设置，预测期将是 1999 年 1 月到 1999 年 3 月。
- ▶ 单击确定。

图片 9-4  
包含预测变量序列的预测的新变量

	Predicted_mail_Model_1	Predicted_mail_Model_2
121	11742	51
122	11853	45
123	11965	45

数据视图 变量视图

“数据编辑器”显示新变量预测值\_mail\_Model\_1 和预测值\_phone\_Model\_2，其中包含邮寄的目录个数和订购热线的条数的模型预测值。要延伸我们的预测变量序列，仅需要 1999 年 1 月到 1999 年 3 月期间的值，其中的量对应于个案 121 到个案 123 中的量。

- ▶ 从预测值\_mail\_Model\_1 中复制这三个个案的值，并将这些值附加到变量 mail 中。
- ▶ 对预测值\_phone\_Model\_2 重复此过程，复制最后三个个案，并将其附加到变量 phone 中。

图片 9-5  
已在预测期中延伸的预测变量序列

	mail	page	phone	print	service	YEAR_	MONTH_	DATE_
121	11742	.	51	.	.	1999	1	JAN 1999
122	11853	.	45	.	.	1999	2	FEB 1999
123	11965	.	45	.	.	1999	3	

数据视图 变量视图

现在，预测变量已在预测期中延伸。

## 在预测期中修改预测变量值

检验邮寄多个目录的两个方案或提供多条热线需要分别修改预测变量 mail 或 phone 的估计值。由于我们仅修改三个个案（月）的预测变量值，所以可以很容易地将新的值直接输入到“数据编辑器”的相应单元格中。我们使用“计算变量”对话框作为示例进行说明。在要修改许多值的情况下，您可能会发现使用“计算变量”对话框更加方便。

- ▶ 从菜单中选择：  
转换 > 计算变量...

图片 9-6  
“计算变量”对话框



- ▶ 输入 mail 作为目标变量。
- ▶ 在“数字表达式”文本框中，输入 mail + 2000。
- ▶ 单击如果。

图片 9-7  
“计算变量：If 个案”对话框



- ▶ 选中如果个案满足条件则包含。
- ▶ 在该文本框中，输入 `$CASENUM > 120`。  
这会将变量 `mail` 的更改限制在此预测期中的个案中。
- ▶ 单击继续。
- ▶ 单击“计算变量”对话框中的确定，并在系统询问是否要更改现有变量时单击确定。  
这会导致预测期的三个月中每个月的 `mail`（邮寄的目录个数）值增大 2000。您已经准备好了用于检验第一个方案的数据，并准备运行分析。

## 运行分析

- ▶ 从菜单中选择：  
分析 > 预测 > 应用模型...

图片 9-8  
“应用时间序列模型”对话框



- ▶ 单击浏览，导航到并选择 catalog\_model.xml（或选择在上一示例中保存的您自己的模型文件）。有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。
- ▶ 在“预测期”组中选择估计期结束之后到指定日期之间的第一个个案。
- ▶ 在“日期”网格中，输入 1999 作为年份，3 作为月份。
- ▶ 单击统计量选项卡。

图片 9-9  
应用时间序列模型，“统计量”选项卡



- ▶ 选择显示预测值。

这将生成一份因变量的预测值表。

- ▶ 单击“应用时间序列模型”对话框中的确定。

图片 9-10  
预测表

模型		1999年1月	1999年2月	1999年3月
Sales of Men's Clothing-Model_1	预测	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

预测表中包含相依序列的预测值，并考虑了预测期中的两个预测变量 mail 和 phone 的值。此表还包括预测的置信区间上限 (UCL) 和置信区间下限 (LCL)。

您已经为每月多邮寄 2000 个目录的方案生成了销售预测。现在，您还要为增加热线条数的方案准备数据，这意味着要将变量 mail 重置为原始值，并将变量 phone 的值增大 5。您可以通过复制预测期中的预测值\_mail\_Model\_1 值来重置 mail。并且将这些值粘贴以代替预测期中 mail 的当前值。然后，将可以直接在数据编辑器中或使用“计算变量”对话框中增加热线条数值（预测期中每个月增加 5 条），就像我们增加目录数的操作一样。

要运行分析，请按照以下步骤重新打开“应用时间序列模型”对话框：

- ▶ 单击“对话框调用”工具栏按钮。

- ▶ 选择应用时间序列模型。

图片 9-11  
“应用时间序列模型”对话框



- ▶ 单击“应用时间序列模型”对话框中的确定。

图片 9-12  
两个方案的预测表

		预测		
模型		1999年1月	1999年2月	1999年3月
Sales of Men's Clothing-Model_1	预测	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

		预测		
模型		1999年1月	1999年2月	1999年3月
Sales of Men's Clothing-Model_1	预测	23757.25	20542.06	20058.29
	UCL	26068.95	22853.76	22370.00
	LCL	21445.55	18230.35	17746.59

显示这两个方案的预测表将说明，在三个预测月份的每个月中，增加邮寄的目录数带来的销售收入比增加订购热线带来的销售收入多大约 1500 多美元。根据此分析，将资源分配到增加 2000 个邮寄目录上是明智的做法。

# 周期性分解

## 移去销售数据中的季节性

编目公司希望根据一组预测变量（例如邮寄的目录数和订购热线的条数）对其男式服装系列销售的上升趋势建模。到目前为止，公司收集了 10 年期间的男式服装月销售量的数据。该信息收集在 catalog.sav 中。[有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。](#)

要执行趋势分析，必须移去数据中出现的所有季节性变化。通过使用“季节性分解”过程可以很容易地完成此任务。

## 确定和设置周期性

“季节性分解”过程要求在活动数据集中存在周期性日期成分，例如，年周期性 12（个月）、周周期性 7（天）等。由于查看时间序列图通常可以对基本周期性进行合理猜测，因此首先对时间序列进行绘图是一种好的做法。

获取随时间变化的男式服装销售量图：

- ▶ 从菜单中选择：  
分析 > 预测 > 序列图...

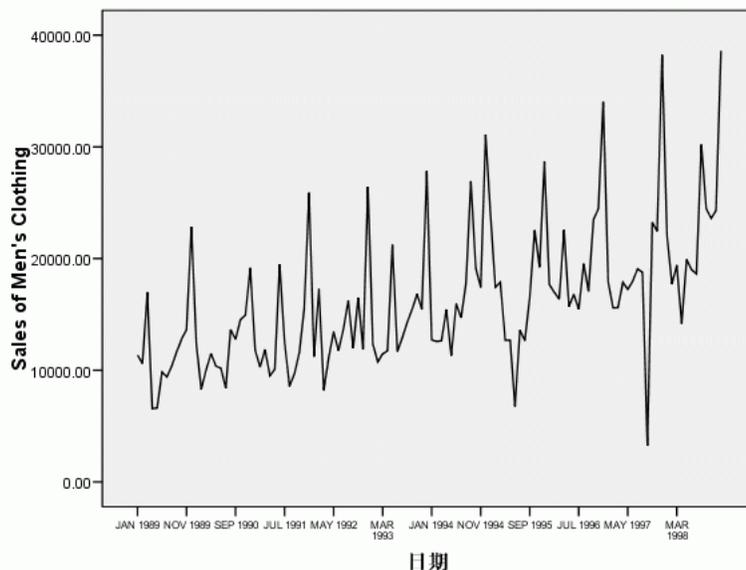
图片 10-1  
“序列图”对话框



- ▶ 选择 Sales of Men's Clothing 并将其移到“变量”列表中。

- ▶ 选择数据并将其移到“时间轴标签”列表中。
- ▶ 单击确定。

图片 10-2  
Sales of men' s clothing (美元)

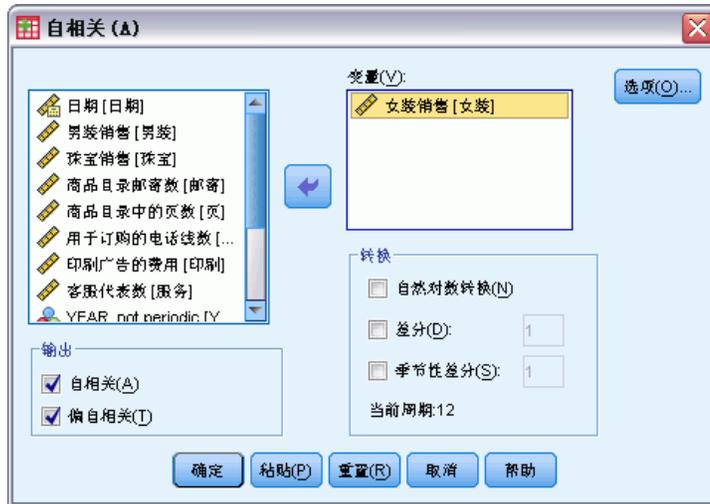


此序列表现出了许多峰值，但这些峰值并非等间距显示。此输出表明，如果序列中存在周期性成分，则它也具有非周期性的波动，这是典型的实时序列个案。除了小幅度的波动之外，显著性峰值按好几个月分隔显示。已知销售具有季节性，通常在十二月份假期内出现频率很高，因此时间序列可能具有年度周期性。还要注意，季节性变化会随序列增长表现出上升趋势，表明季节性变化与序列水平成比例，这表明是乘法模型而不是加法模型。

检查时间序列的自相关和部分自相关可得出有关基本周期性的更具定量性的结论。

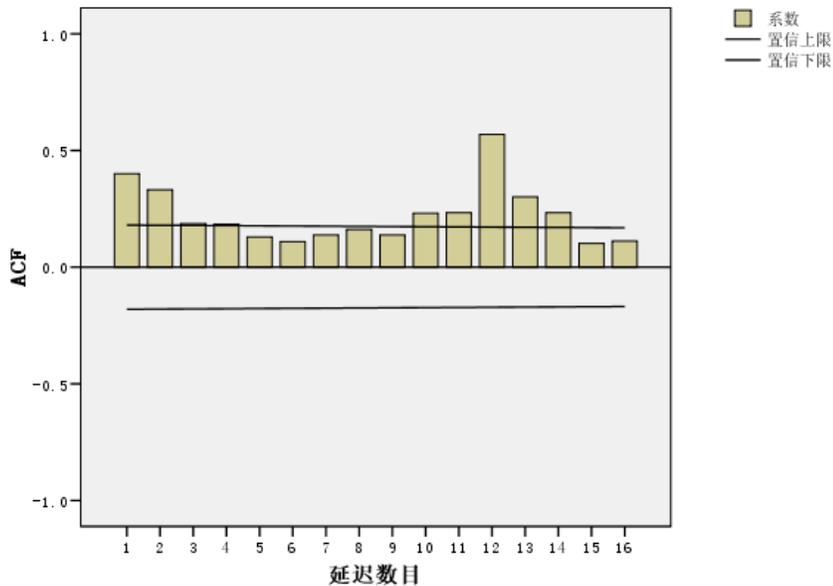
- ▶ 从菜单中选择：  
分析 > 预测 > 自相关 (0)...

图片 10-3  
“自相关”对话框



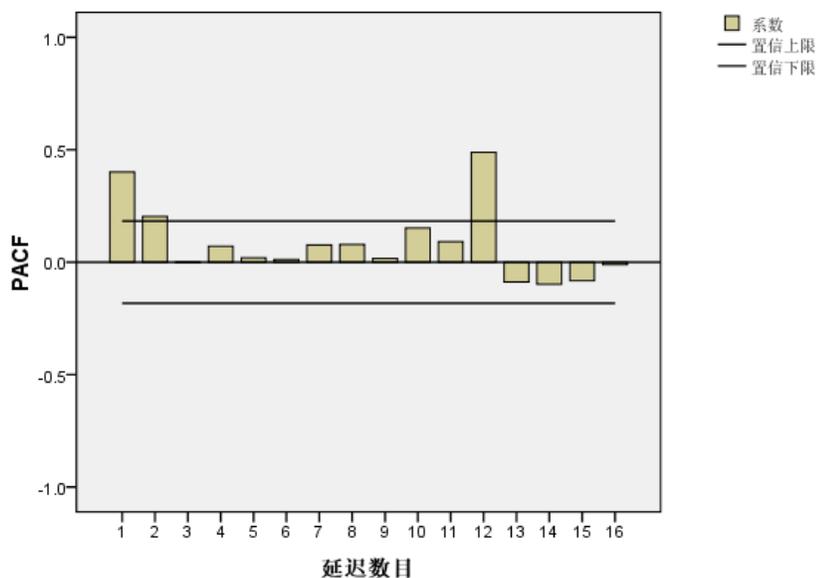
- ▶ 选择 Sales of Men' s Clothing 并将其移到“变量”列表中。
- ▶ 单击确定。

图片 10-4  
男式服装的自相关图



自相关函数在延迟 1 处显示一个具有长指数尾部的显著性峰值，这是时间序列的典型模式。延迟 12 处的显著性峰值表明数据中存在年度季节性成分。检查部分自相关函数将可得出更具确定性的结论。

图片 10-5  
男式服装的部分自相关图

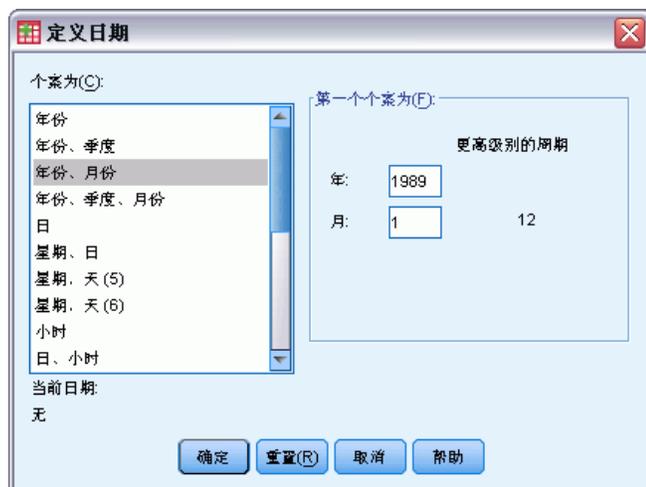


部分自相关函数中延迟 12 处的显著性峰值证实了数据中存在年度季节性成分。

设置年度周期性：

- ▶ 从菜单中选择：  
数据 > 定义日期...

图片 10-6  
“定义日期”对话框



- ▶ 在“个案为”列表中选择年份、月份。
- ▶ 输入 1989 作为年份，输入 1 作为月份。

- ▶ 单击确定。

这将把周期设置为 12，并创建专用于“预测”过程的一组日期变量。

## 运行分析

运行“季节性分解”过程：

- ▶ 从菜单中选择：  
分析 > 预测 > 季节性分解...

图片 10-7  
“季节性分解”对话框



- ▶ 右键单击源变量列表中的任何位置并从上下文菜单中选择显示变量名称。
- ▶ 选择 men 并将其移到“变量”列表中。
- ▶ 在“模型类型”组中选择乘法。
- ▶ 单击确定。

## 了解输出

“季节性分解”过程为此过程分析的每个初始变量创建了四个新的变量。默认情况下，新变量将添加到活动数据集中。新序列的名称以下列前缀开头：

**SAF。** 季节性调整因子，表示季节性变化。对于乘法模型，值 1 表示不存在季节性变化；对于加法模型，值 0 表示不存在季节性变化。

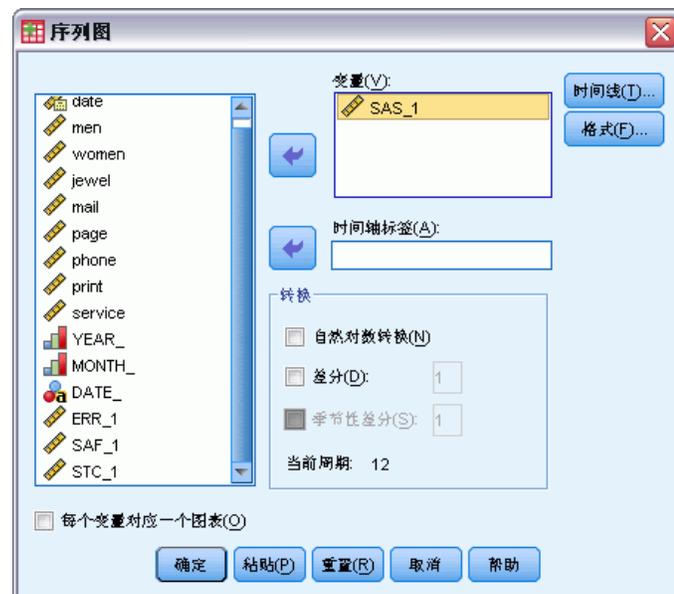
**SAS。** 季节性调整序列，表示已移去季节性变化的初始序列。例如，使用季节性调整序列可以隔离趋势成分，并独立于所有的季节性成分对趋势成分进行分析。

**STC。**平滑的趋势循环成分，它是季节性调整序列的平滑版本，可显示趋势和循环成分。

**ERR。**特定观察值的序列的残差成分。

对于所提供的个案，季节性调整序列是最合适的，因为它表示初始序列已移去了季节性变化。

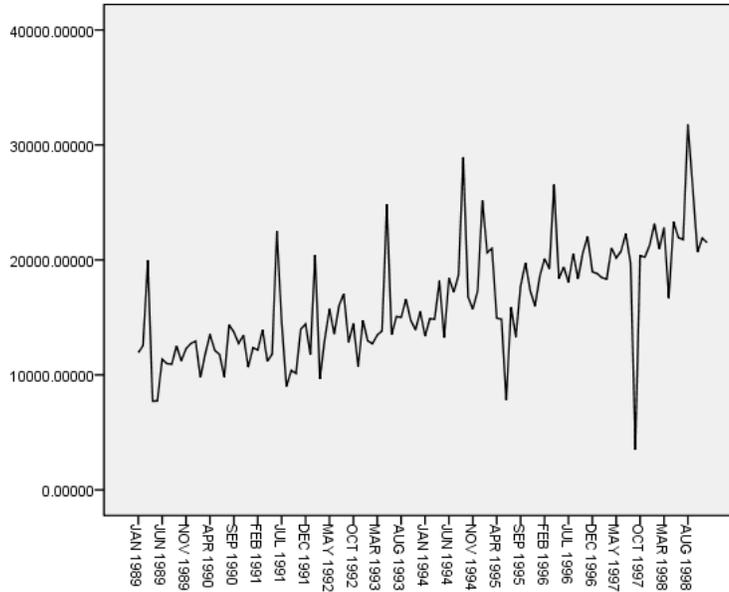
图片 10-8  
“序列图”对话框



对季节性调整序列绘图：

- ▶ 打开“序列图”对话框。
- ▶ 单击重置以清除以前的所有选择。
- ▶ 右键单击源变量列表中的任何位置并从上下文菜单中选择显示变量名称。
- ▶ 选择 SAS\_1 并将其移到“变量”列表中。
- ▶ 单击确定。

图片 10-9  
季节性调整序列



季节性调整序列呈现明显的上升趋势。许多峰值都很明显，但它们以随机的间隔显示，表明没有年度模式的迹象。

## 摘要

通过使用“季节性分解”过程，您已移去周期性时间序列的周期性成分，以生成一个更适合趋势分析的序列。在确定基本周期性时，检验时间序列的自相关和部分自相关是很有用的。在此个案中，周期性为年度。

## 相关过程

在从周期性时间序列中移去单个季节性成分时，“季节性分解”过程非常有用。

- 要对时间序列周期性执行更深入的分析（相对于部分自相关函数所执行的分析而言），请使用“频谱图”过程。有关详细信息，请参见第 11 章。

# 谱图

## 使用频谱图验证周期性的期望值

由于节假日的销售中通常存在峰值，因此表示零售额的时间序列具有基本的年度周期性。生成销售投影意味着构建时间序列模型，这意味着标识所有的周期性成分。由于时间序列包含通常会掩盖基础结构的随机波动，因此时间序列图可能不会始终揭示年度周期性。

编目公司的每月销售量数据存储在 catalog.sav 中。[有关详细信息，请参阅附录 D 中的样本文件中的 IBM SPSS Forecasting 20。](#)在处理销售投影之前，您需要确认销售数据是否表现出了年度周期性。时间序列图显示了具有不规则间隔的许多峰值，因此没有任何明显的基本周期性。使用“频谱图”过程可标识销售数据中的所有周期性。

## 运行分析

运行“频谱图”过程：

- ▶ 从菜单中选择：  
分析 > 预测 > 频谱分析(T)...

图片 11-1  
“频谱图”对话框



- ▶ 选择 Sales of Men's Clothing 并将其移到“变量”列表中。

- ▶ 在“图”组中选择频谱密度。
- ▶ 单击确定。

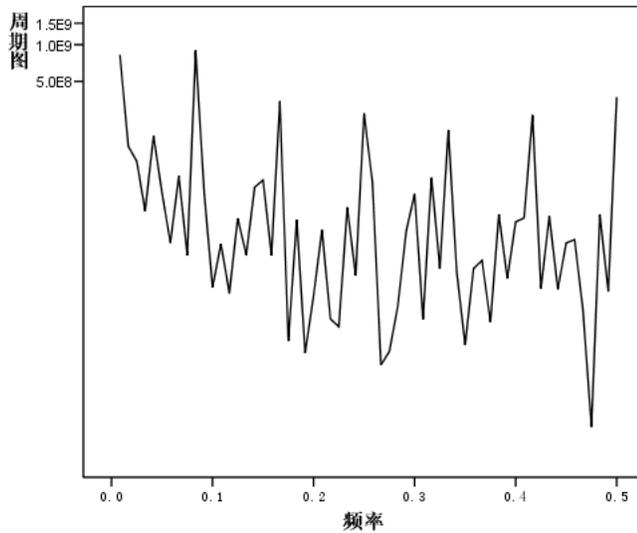
这些选择将生成以下命令语法：

```
* Spectral Analysis.  
TSET PRINT=DEFAULT.  
SPECTRA  
  /VARIABLES=men  
  /WINDOW=HAMMING(5)  
  /CENTER  
  /PLOT=P S BY FREQUENCY.
```

注意，为了在输出中获得单变量统计表，TSET 命令需要更改为读取 TSET PRINT=DETAILED。

## 了解周期图和频谱密度

图片 11-2  
周期图



周期图显示突出于背景噪声的一系列峰值，并在频率小于 0.1 时出现最低频率峰值。您会怀疑这些数据是否包含年度周期性成分，因此需要考虑年度成分对周期图所做的贡献。时间序列中的每个数据点表示一个月，因此，年度周期性在当前数据集中对应周期 12。由于周期和频率互为倒数，因此周期 12 对应于频率  $1/12$ （或 0.083）。因此，年度成分意味着周期图中的频率为 0.083 的峰值，这与存在小于频率 0.1 的峰值相符。

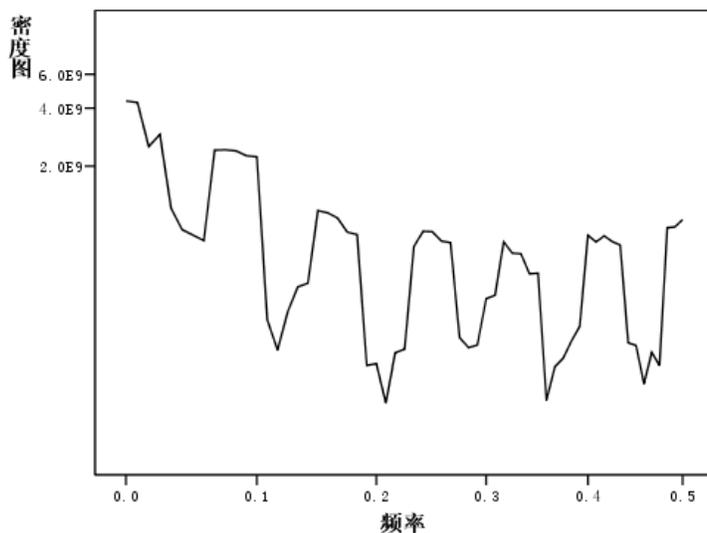
图片 11-3  
单变量统计表

序列名: Sales of Men's Clothing

	频率	正弦转换	余弦转换	周期图	谱密度估计
1	.00000	.000	18242.813	.000	4.361E+009
2	.00833	-3696.643	370.153	8.3E+008	4.278E+009
3	.01667	-1496.277	454.831	1.5E+008	2.526E+009
4	.02500	-1336.400	252.087	1.1E+008	2.921E+009
5	.03333	-662.146	529.734	43143316	1.210E+009
6	.04167	-1654.614	501.571	1.8E+008	9.359E+008
7	.05000	-784.814	-636.729	61281367	8.755E+008
8	.05833	-335.646	532.062	23744655	8.204E+008
9	.06667	-1094.178	-451.489	84064109	2.420E+009
10	.07500	264.554	492.876	18774934	2.429E+009
11	.08333	-3053.934	2370.483	9.0E+008	2.401E+009
12	.09167	-978.882	-267.035	62435898	2.263E+009
13	.10000	-403.128	93.036	10270064	2.235E+009

单变量统计表包含用于绘制周期图的数据点。请注意，对于小于 0.1 的频率，周期图列中的最大值在频率为 0.08333 时出现。如果存在年度周期性成分，这就是您所查找的内容。此信息证实了可以使用年度周期性成分标识最低频率峰值。但是更高频率的其他峰值又表示什么呢？

图片 11-4  
谱密度



对于剩余的峰值，可以使用频谱密度函数进行最有效的分析，频谱密度函数只是周期图的平滑版本。平滑法提供了去除周期图背景噪声的方法，从而可以更清晰地隔离基本结构。

频谱密度由五个等距离的相异峰值构成。最低频率峰值仅显示频率为 0.08333 的平滑版本的峰值。为了了解四个更高频率峰值的显著性，请记住，周期图是通过将时间序列建模为正弦函数和余弦函数之和来计算的。具有正弦函数或余弦函数形状（正弦曲线）的周期成分将在周期图中显示为单个峰值。不是正弦曲线的周期成分显示为一组不同高度的等距离峰值，并且此序列中的最低频率峰值以周期成分频率显示。因此，频谱密度中的四个更高频率的峰值仅指示年度周期性成分不是正弦曲线。

现在，您已经考虑了频谱密度图中所有可辨别的结构，并得出周期为 12 个月时数据包含单一周期性成分的结论。

## 摘要

通过使用“频谱图”过程，您已证实时间序列中存在年度周期性成分，并已验证了没有其他显著性周期存在。对于揭示基本结构而言，谱密度比周期图更实用，因为频谱密度消除了数据的非周期性成分所导致的波动。

## 相关过程

“频谱图”过程对于标识时间序列的周期性成分非常有用。

- 要移去时间序列中的周期性成分（例如，执行趋势分析），请使用“季节性分解”过程。有关详细信息，请参见第 10 章。

# 拟合优度测量

本节提供了时间序列建模中使用的拟合优度测量的定义。

- **固定的 R 方.** 将模型的平稳部分与简单均值模型相比较的测量。当具有趋势或季节性模式时，该度量适用于普通 R 方。固定的 R 方可以是负无穷大到 1 范围中的负值。负值表示考虑中的模型比基线模型差。正值表示考虑中的模型比基线模型好。
- **R 方.** 总变动在由模型解释的序列中的比例估计。当序列很平稳时，此度量最有用。R 方可以是负无穷大到 1 范围中的负值。负值表示考虑中的模型比基线模型差。正值表示考虑中的模型比基线模型好。
- **RMSE.** 均方根误差。均方误差的平方根。度量因变量序列与其模型预测水平的相差程度，用和因变量序列相同的单位表示。
- **MAPE.** 平均绝对误差百分比。度量因变量序列与其模型预测水平的相差程度。它与使用的单位无关，因此可用于比较具有不同单位的序列。
- **MAE.** 平均绝对误差。度量序列与其模型预测水平的差别程度。MAE 以原始序列单位报告。
- **MaxAPE.** 最大绝对误差百分比。最大的预测误差，以百分比表示。该度量对于想象预测的最坏情况方案很有用。
- **MaxAE.** 最大绝对误差。最大的预测误差，以和因变量序列相同的单位表示。与 MaxAPE 相同，它对于想象预测的最坏情况方案很有用。最大绝对误差和最大绝对误差百分比可能发生在不同的序列点上，例如，当较大序列的绝对误差比较小值的绝对误差稍微大一些时。在此情况下，最大绝对误差将发生在较大序列值处，而最大绝对误差百分比将发生在较小序列值处。
- **标准化的 BIC.** 标准化的 BIC (BIC 准则)。尝试代表模型复杂性的模型整体拟合的一般度量。它是基于均方误差的分数，包括模型中参数数量的罚分和序列长度。罚分去除了具有更多参数的模型优势，从而可以容易地比较相同序列的不同模型的统计量。

# 离群值类型

本节提供了时间序列建模中使用的离群值类型的定义。

- **可加的.** 影响单个观察值的离群值。例如，可能将数据编码错误标识为可加离群值。
- **移位水平.** 从某个特定的序列点开始将所有观察值移动一个常数的离群值。移位水平可能由于策略的更改而造成的。
- **创新的.** 在某个特定的序列点附加到噪声项的离群值。对于平稳的序列，创新离群值将影响多个观察值。对于不平稳的序列，它可能影响在某个特定的序列点开始的每个观察值。
- **瞬时的.** 其影响按指数衰减到 0 的离群值。
- **季节性可加的.** 一个离群值，它影响特定观察值以及通过一个或多个季节性期间与之分隔的所有后续观察值。该离群值对所有这些观察值具有同等的影响。如果从某年开始，每个一月的销售额都增加，则可能发生季节性可加离群值。
- **局部趋势.** 从某个特定的序列点开始局部趋势的离群值。
- **可加的修补.** 由两个或更多连续可加离群值构成的组。选择此离群值类型将导致除了检测可加离群值的变量值组以外，还检测单独的可加离群值。

## ACF/PACF 图指南

此处所示的图是纯 ARIMA 过程或理论 ARIMA 过程的图。以下为过程的一些常用标识方法：

- 不稳序列的 ACF 的显著性会保留 6 个或一打以上的延迟，而不是迅速降低为 0。您必须区分此类序列，直至其稳定后，才可标识该过程。
- 自回归过程有一个指数不断衰减的 ACF，而且 PACF 的第一个或更多延迟中存在花序。花序数表示自回归的阶数。
- 移动平均数过程的 ACF 的第一个或更多延迟中存在花序，且有一指数不断衰减的 PACF。花序数表示移动平均数的阶数。
- 混合 (ARMA) 过程通常显示 ACF 和 PACF 中的指数衰减。

在标识阶段，您无需担心 ACF 或 PACF 的符号，也无需担心 ACF 或 PACF 的指数衰减至 0 的速度。这些都取决于 AR 和 MA 系数的符号和实际值。在某些实例中，指数不断衰减的 ACF 会在正值和负值之间来回变换。

根据实际数据绘制的 ACF 和 PACF 图决不会像此处显示的图那样规则。您必须学会在任何给定的图中分辨出基本规律。始终检查残差的 ACF 和 PACF，以防标识有误。请记住：

- 季节性过程将显示这些季节性延迟（季节性周期的倍数）的模式。
- 您有权将不显著值视为 0。也就是说，您可以忽略图中置信区间内的值。但是，您不必忽略它们，尤其在它们延续统计显著值的模式时。
- 偶然自相关有时会在统计上具有显著性。如果统计显著性是独立的，经常出现在高延迟上，而且未出现在季节性延迟中，则可以忽略此统计显著性。

有关 ACF 和 PACF 图的更完整的讨论，请参见与 ARIMA 分析有关的所有文本。

表 C-1  
ARIMA(0, 0, 1),  $q > 0$

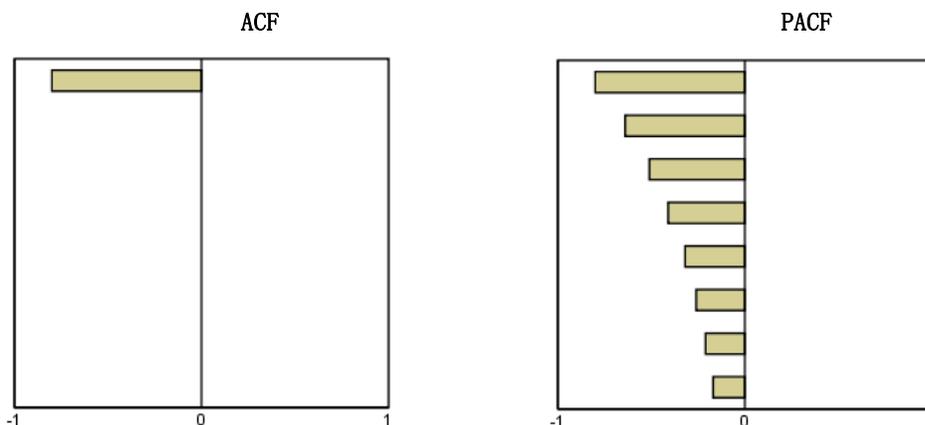
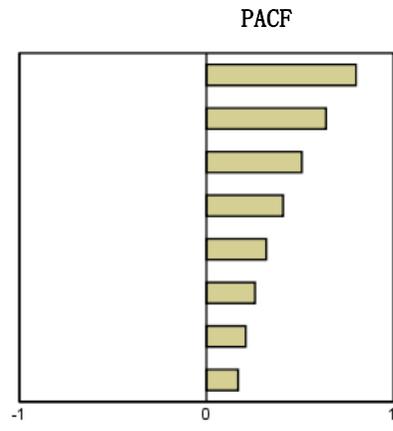
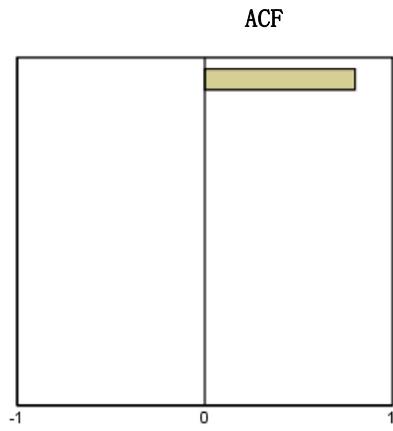


表 C-2  
ARIMA (0, 0, 1),  $q < 0$



ARIMA (0, 0, 2),  $\theta_1\theta_2 > 0$

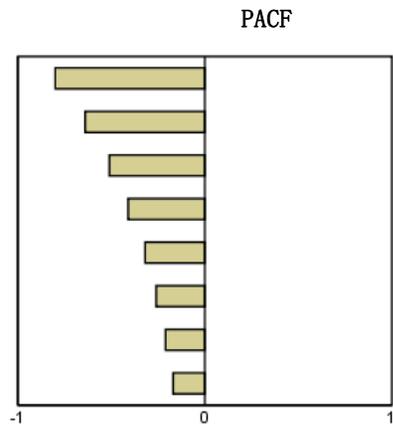
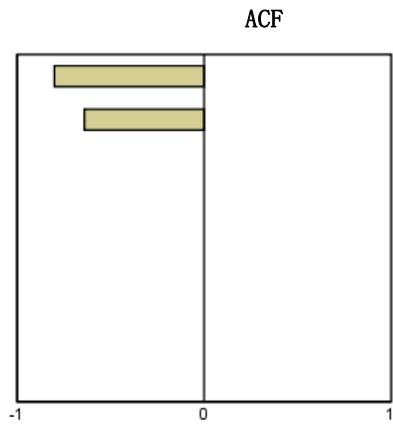


表 C-3  
ARIMA (1, 0, 0),  $f > 0$

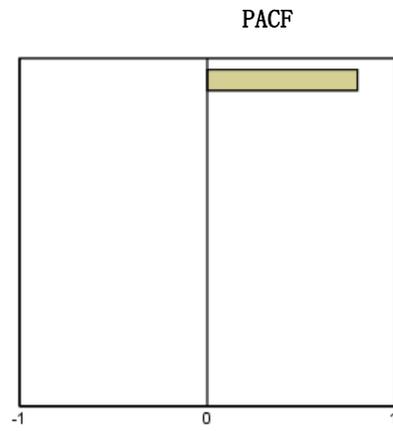
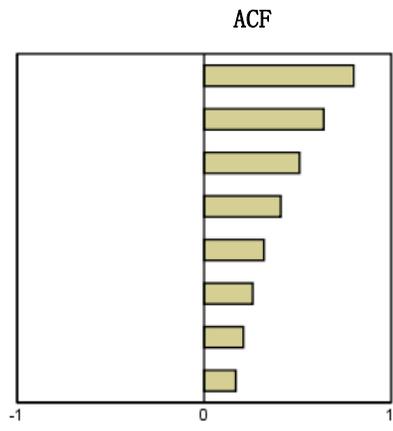
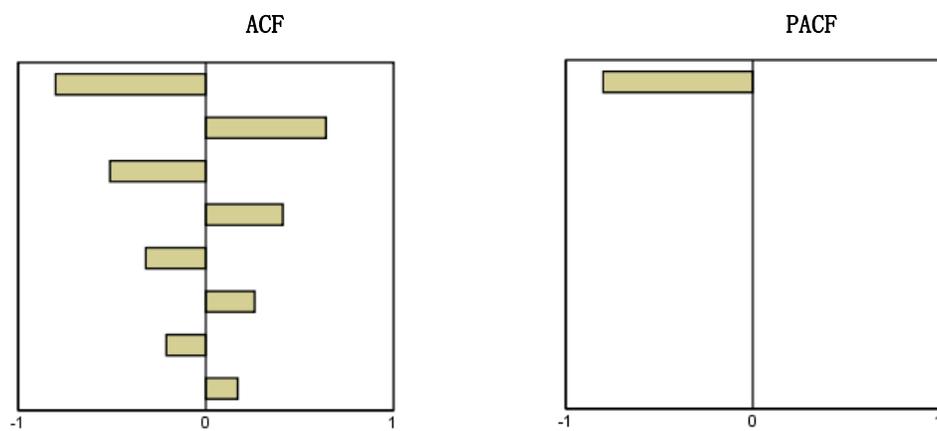
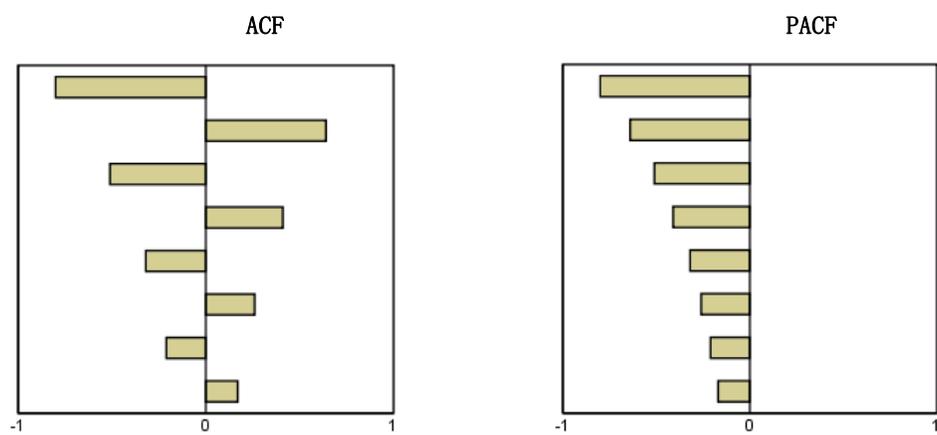


表 C-4  
ARIMA(1, 0, 0),  $\phi < 0$



ARIMA(1, 0, 1),  $\phi < 0, \theta > 0$



ARIMA(2, 0, 0),  $\phi_1\phi_2 > 0$

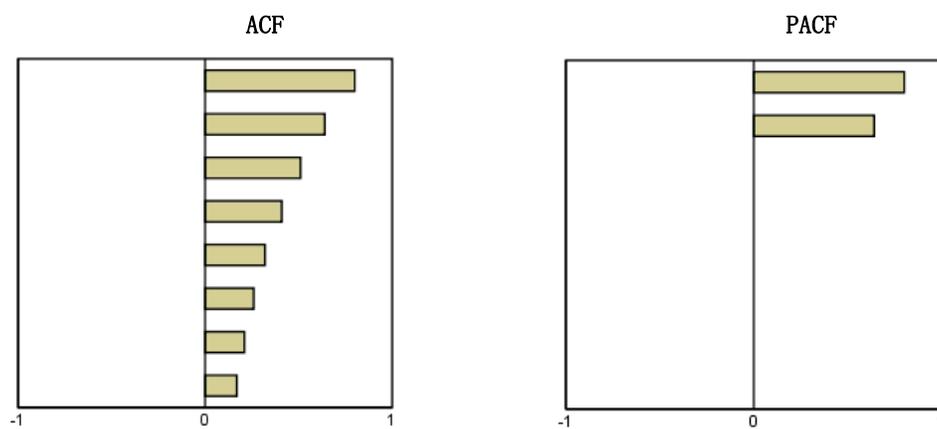
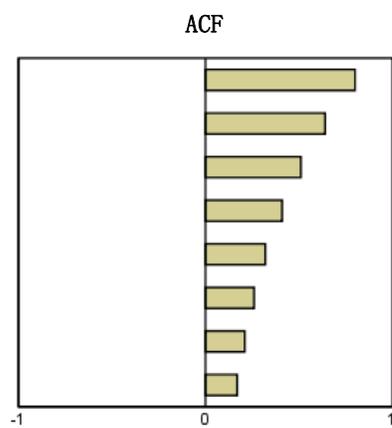


表 C-5  
ARIMA (0, 1, 0) (积分序列)



# 样本文件

随产品一起安装的样本文件可以在安装目录的 Samples 子目录中找到。对于以下每种语言在“样本”子目录中有单独的文件夹：英语、法语、德语、意大利语、日语、韩语、波兰语、俄语、简体中文、西班牙语和繁体中文。

并非所有样本文件均提供此处的全部语言版本。如果样本文件未提供某种语言的版本，则相应语言文件夹中包含该样本文件的英语版本。

## 描述

以下是对在整个文档的各种示例中使用的样本文件的简要描述。

- **accidents.sav**。该假设数据文件涉及某保险公司，该公司正在研究给定区域内汽车事故的年龄和性别风险因子。每个个案对应一个年龄类别和性别类别的交叉分类。
- **adl.sav**。该假设数据文件涉及在确定针对脑卒中患者的建议治疗类型的优点方面的举措。医师将女性脑卒中患者随机分配到两组中的一组。第一组患者接受标准的物理治疗，而第二组患者则接受附加的情绪治疗。在进行治疗的三个月时间里，将为每个患者进行一般日常生活行为的能力评分并作为原始变量。
- **advert.sav**。该假设数据文件涉及某零售商在检查广告支出与销售业绩之间的关系方面的举措。为此，他们收集了过去的销售数据以及相关的广告成本。
- **aflatoxin.sav**。该假设数据文件涉及对谷物的黄曲霉毒素的检测，该毒素的浓度会因谷物产量的不同（不同谷物之间及同种谷物之间）而有较大变化。谷物加工机从 8 个谷物产量的每一个中收到 16 个样本并以十亿分之几 (PPB) 为单位来测量黄曲霉毒素的水平。
- **anorectic.sav**。在研究厌食/暴食行为的标准症状参照时，研究人员 (Van der Ham, Meulman, Van Strien, 和 Van Engeland, 1997) 对 55 名已知存在进食障碍的青少年进行了调查。其中每名患者每年都将进行四次检查，因此总观测数为 220。在每次观测期间，将对这些患者按 16 种症状逐项评分。但 71 号和 76 号患者的症状得分均在时间点 2 缺失，47 号患者的症状得分在时间点 3 缺失，因此有效观测数为 217。
- **bankloan.sav**。该假设数据文件涉及某银行在降低贷款拖欠率方面的举措。该文件包含 850 位过去和潜在客户的财务和人口统计信息。前 700 个个案是以前曾获得贷款的客户。剩下的 150 个个案是潜在客户，银行需要按高或低信用风险对他们进行分类。
- **bankloan\_binning.sav**。该假设数据文件包含 5,000 位过去客户的财务和人口统计信息。
- **behavior.sav**。在一个经典示例中 (Price 和 Bouffard, 1974)，52 名学生被要求以 10 分的标度对 15 种情况和 15 种行为的组合进行评价，该 10 分的标度介于 0 = 平均值在个人值之上，值被视为相异性。
- **behavior\_ini.sav**。该数据文件包含 behavior.sav 的二维解的初始配置。

- **brakes.sav**。该假设数据文件涉及某生产高性能汽车盘式制动器的工厂的质量控制。该数据文件包含对 8 台专用机床中每一台的 16 个盘式制动器的直径测量。盘式制动器的目标直径为 322 毫米。
- **breakfast.sav**。在一项经典研究中 (Green 和 Rao, 1972), 21 名 Wharton School MBA 学生及其配偶被要求按照喜好程度顺序对 15 种早餐食品进行评价, 从 1 = 他们的喜好根据六种不同的情况加以记录, 从“全部喜欢”到“只带饮料的快餐”。
- **breakfast-overall.sav**。该数据文件只包含早餐食品喜好的第一种情况, 即“全部喜欢”。
- **broadband\_1.sav**。该假设数据文件包含各地区订制了全国宽带服务的客户的数量。该数据文件包含 4 年期间 85 个地区每月的订户数量。
- **broadband\_2.sav**。该数据文件和 broadband\_1.sav 一样, 但包含另外三个月的数据。
- **car\_insurance\_claims.sav**。在别处被提出和分析的 (McCullagh 和 Nelder, 1989) 关于汽车损坏赔偿的数据集。平均理赔金额可以当作其具有 gamma 分布来建模, 通过使用逆联接函数将因变量的均值与投保者年龄、车辆类型和车龄的线性组合关联。提出理赔的数量可以作为尺度权重。
- **car\_sales.sav**。该数据文件包含假设销售估计值、订价以及各种品牌和型号的车辆的物理规格。订价和物理规格可以从 edmunds.com 和制造商处获得。
- **car\_sales\_uprepared.sav**。这是 car\_sales.sav 的修改版本, 不包含字段的任何已转换版本。
- **carpet.sav**。在一个常用示例中 (Green 和 Wind, 1973), 一家公司非常重视一种新型地毯清洁用品的市场营销, 希望检验以下五种因素对消费者偏好的影响—包装设计、品牌名称、价格、优秀家用品标志和退货保证。包装设计有三个因子水平, 每个因子水平因刷体位置而不同; 有三个品牌名称 (K2R、Glory 和 Bissell); 有三个价格水平; 最后两个因素各有两个级别 (有或无)。十名消费者对这些因素所定义的 22 个特征进行了排序。变量优选包含对每个特征的平均等级的排序。低排序与高偏好相对应。此变量反映了对每个特征的偏好的总体度量。
- **carpet\_prefs.sav**。该数据文件所基于的示例和在 carpet.sav 中所描述的一样, 但它还包含从 10 位消费者的每一位中收集到的实际排列顺序。消费者被要求按照从最喜欢到最不喜欢的顺序对 22 个产品特征进行排序。carpet\_plan.sav 中定义了变量 PREF1 到 PREF22 包含相关特征的标识符。
- **catalog.sav**。该数据文件包含某编目公司出售的三种产品的假设每月销售数据。同时还包括 5 个可能的预测变量的数据。
- **catalog\_seasfac.sav**。除添加了一组从“季节性分解”过程中计算出来的季节性因子和附带的日期变量外, 该数据文件和 catalog.sav 是相同的。
- **cellular.sav**。该假设数据文件涉及某便携式电话公司在减少客户流失方面的举措。客户流失倾向分被应用到帐户, 分数范围从 0 到 100。得到 50 分或更高分数的帐户可能会更换提供商。
- **ceramics.sav**。该假设数据文件涉及某制造商在确定新型优质合金是否比标准合金具有更高的耐热性方面的举措。每个个案代表对一种合金的单独检验; 个案中会记录合金的耐热极限。
- **cereal.sav**。该假设数据文件涉及一份 880 人参与的关于早餐喜好的民意调查, 该调查记录了参与者的年龄、性别、婚姻状况以及生活方式是否积极 (根据他们是否每周至少做两次运动)。每个个案代表一个单独的调查对象。

- **clothing\_defects.sav**。这是关于某服装厂的质量控制过程的假设数据文件。检验员要对工厂中每次大批量生产的服装进行抽样检测并清点不合格的服装的数量。
- **coffee.sav**。这是关于六种冰咖啡的认知品牌形象 (Kennedy, Riquier, 和 Sharp, 1996) 的数据文件。对于 23 种冰咖啡特征属性中的每种属性, 人们选择了由该属性所描述的所有品牌。为保密起见, 六种品牌用 AA、BB、CC、DD、EE 和 FF 来表示。
- **contacts.sav**。该假设数据文件涉及一组公司计算机销售代表的联系方式列表。根据这些销售代表所在的公司部门及其公司的秩来对每个联系方式进行分类。同时还记录了最近一次的销售量、最近一次销售距今的时间和所联系公司的规模。
- **creditpromo.sav**。该假设数据文件涉及某百货公司在评价最新信用卡促销的效果方面的举措。为此, 随机选择了 500 位持卡人。其中一半收到了宣传关于在接下来的三个月内降低消费利率的广告。另一半收到了标准的季节性广告。
- **customer\_dbase.sav**。该假设数据文件涉及某公司在使用数据仓库中的信息来为最有可能回应的客户提供特惠商品方面的举措。随机选择客户群的子集并为其提供特惠商品, 同时记录下他们的回应。
- **customer\_information.sav**。该假设数据文件包含客户邮寄信息, 如姓名和地址。
- **customer\_subset.sav**。来自 customer\_dbase.sav 的拥有 80 个个案的子集。
- **debate.sav**。该假设数据文件涉及在某政治辩论前后对该辩论的参与者所做的调查的成对回答。每个个案对应一个单独的调查对象。
- **debate\_aggregate.sav**。该假设数据文件分类汇总了 debate.sav 中的回答。每个个案对应一个辩论前后的偏好的交叉分类。
- **demo.sav**。这是关于购物客户数据库的假设数据文件, 用于寄出每月的商品。将记录客户对商品是否有回应以及各种人口统计信息。
- **demo\_cs\_1.sav**。该假设数据文件涉及某公司在汇编调查信息数据库方面的举措的第一步。每个个案对应不同的城市, 并记录地区、省、区和城市标识。
- **demo\_cs\_2.sav**。该假设数据文件涉及某公司在汇编调查信息数据库方面的举措的第二步。每个个案对应来自第一步中所选城市的不同的家庭单元, 并记录地区、省、区、市、子区和单元标识。还包括设计前两个阶段的抽样信息。
- **demo\_cs.sav**。该假设数据文件包含用复杂抽样设计收集的调查信息。每个个案对应不同的家庭单元, 并记录各种人口统计和抽样信息。
- **dmdata.sav**。该假设数据文件包含直销公司的人口统计学和购买信息。dmdata2.sav 包含收到测试邮件的联系人子集的信息, dmdata3.sav 包含未收到测试邮件的其余联系人的信息。
- **dietstudy.sav**。该假设数据文件包含对 “Stillman diet” (Rickman, Mitchell, Dingman, 和 Dalen, 1974) 的研究结果。每个个案对应一个单独的主体, 并记录其在实行饮食方案前后的体重 (磅) 以及甘油三酸酯的水平 (毫克/100 毫升)。
- **dvdplayer.sav**。这是关于开发新的 DVD 播放器的假设数据文件。营销团队用原型收集了焦点小组数据。每个个案对应一个单独的被调查用户, 并记录他们的人口统计信息及其对原型问题的回答。
- **german\_credit.sav**。该数据文件取自加州大学欧文分校的 Repository of Machine Learning Databases (Blake 和 Merz, 1998) 中的 “German credit” 数据集。
- **grocery\_1month.sav**。该假设数据文件是在数据文件 grocery\_coupons.sav 的基础上加上了每周购物 “累计”, 所以每个个案对应一个单独的客户。所以, 一些每周更改的变量消失了, 而且现在记录的消费金额是为期四周的研究过程中的消费金额之和。

- **grocery\_coupons.sav**。该假设数据文件包含由重视顾客购物习惯的杂货连锁店收集的调查数据。对每位顾客调查四周，每个个案对应一个单独的顾客周，并记录有关顾客购物地点和方式的信息（包括那一周里顾客在杂货上的消费金额）。
- **guttman.sav**。Bell (Bell, 1961) 创建了一个表，用来阐释可能的社会群体。Guttman (Guttman, 1968) 引用了该表的一部分，其中包括五个变量，用于描述以下七个理论社会群体的社会交往、对群体的归属感、成员的物理亲近度以及关系正式性：观众（比如在足球比赛现场的人们）、听众（比如在剧院或听课堂讲座的人们）、公众（比如报纸或电视观众）、组织群体（与观众类似但具有紧密的关系）、初级群体（关系密切）、次级群体（自发组织）及现代社区（因在物理上亲近而导致关系松散并需要专业化服务）。
- **health\_funding.sav**。该假设数据文件包含关于保健基金（每 100 人的金额）、发病率（每 10,000 人的比率）以及保健提供商拜访率（每 10,000 的比率）的数据。每个个案代表不同的城市。
- **hivassay.sav**。该假设数据文件涉及某药物实验室在开发用于检测 HIV 感染的快速化验方面的举措。化验结果为八个加深的红色阴影，如果有更深的阴影则表示感染的可能性很大。用 2,000 份血液样本来进行实验室试验，其中一半受到 HIV 感染而另一半没有受到感染。
- **hourlywagedata.sav**。该假设数据文件涉及在政府机关和医院工作的具有不同经验水平的护士的时薪。
- **insurance\_claims.sav**。该假设数据文件涉及某保险公司，该公司希望构建一个模型用于标记可疑的、具有潜在欺骗性的理赔。每个个案代表一次单独的理赔。
- **insure.sav**。该假设数据文件涉及某保险公司，该公司正在研究指示客户是否会根据 10 年的人寿保险合同提出理赔的风险因子。数据文件中的每个个案代表一副根据年龄和性别进行匹配的合同，其中一份记录了一次理赔而另一份则没有。
- **judges.sav**。该假设数据文件涉及经过训练的裁判（加上一个体操爱好者）对 300 次体操表演给出的分数。每行代表一次单独的表演；裁判们观看相同的表演。
- **kinship\_dat.sav**。Rosenberg 和 Kim (Rosenberg 和 Kim, 1975) 开始分析 15 个亲属关系项（伯母、兄弟、表兄妹、女儿、父亲、孙女、祖父、祖母、孙子、母亲、侄子或外甥、侄女或外甥女、姐妹、儿子和叔叔）。他们让四组大学生（两组女同学，两组男同学）根据相似程度将各项排序。他们让其中的两组同学（一组女同学，一组男同学）进行了两次排序，第二次排序和第一次排序采取的标准不同。这样，一共得到六组“源”。每个源对应一个  $15 \times 15$  的相似性矩阵，其单元格中的值等于源中的人数减去此源中对象被划分的次数。
- **kinship\_ini.sav**。该数据文件包含 kinship\_dat.sav 的三维解的初始配置。
- **kinship\_var.sav**。该数据文件包含自变量 gender、gener(ation) 和 degree (of separation)，这些变量可用于解释 kinship\_dat.sav 的解的维数。具体而言，它们可用来将解的空间限制为这些变量的线性组合。
- **marketvalues.sav**。该数据文件涉及 1999 - 2000 年间 Algonquin, Ill. 地区新的房屋开发中的住房销售。这些销售仅仅来自公众记录。
- **nhis2000\_subset.sav**。美国健康访问调查 (NHIS) 是针对美国全体公民的大型人口调查。该调查对美国的具有全国代表性的家庭样本进行了面对面的访问，并获取了每个家庭的成员的健康行为和健康状态的人口统计信息和观察数据。该数据文件包含取自 2000 年调查信息的子集。国家健康统计中心。2000 年美国健康访问调查。公用数据文件和文档。

ftp://ftp.cdc.gov/pub/Health\_Statistics/NCHS/Datasets/NHIS/2000/。2003 年发布。

- **ozone.sav**。这些数据包含了用来根据其余变量预测臭氧浓度的六个气象变量的 330 个观察值。在以前的研究人员中, (Breiman 和 Friedman(F), 1985) 和 (Hastie 和 Tibshirani, 1990) 发现了这些变量之间的非线性, 这妨碍了标准回归方法。
- **pain\_medication.sav**。该假设数据文件包含用于治疗慢性关节炎疼痛的抗炎药的临床试验结果。我们感兴趣的是该药见效的时间以及它和现有药物的比较。
- **patient\_los.sav**。该假设数据文件包含被医院确诊为疑似心肌梗塞(即 MI 或“心脏病发作”)的患者的治疗记录。每个个案对应一位单独的患者, 并记录与其住院期有关的一些变量。
- **patlos\_sample.sav**。该假设数据文件包含在治疗心肌梗塞(即 MI 或“心脏病发作”)期间收到溶解血栓剂的患者样本的治疗记录。每个个案对应一位单独的患者, 并记录与其住院期有关的一些变量。
- **poll\_cs.sav**。该假设数据文件涉及民意测验专家在确定正式立法前公众对法案的支持水平方面的举措。个案对应注册的选民。每个个案记录选民居住的县、镇、区。
- **poll\_cs\_sample.sav**。该假设数据文件包含在 poll\_cs.sav 中列出的选民的样本。该样本是根据 poll\_csplan 中指定的设计来选取的, 而且该数据文件记录包含概率和样本权重。请注意, 由于该抽样计划使用与大小成正比(PPS)方法, 因此, 还有一个文件(poll\_jointprob.sav)包含联合选择概率。在选取了样本之后, 对应于选民人群统计信息及其对提交法案的意见的附加变量将被收集并添加到数据文件。
- **property\_assess.sav**。该假设数据文件涉及某县资产评估员在利用有限的资源不断更新资产价值评估方面的举措。个案对应过去一年中县里所出售的资产。数据文件中的每个个案记录资产所在的镇、最后评估资产的评估员、该次评估距今的时间、当时的估价以及资产的出售价格。
- **property\_assess\_cs.sav**。该假设数据文件涉及某州资产评估员在利用有限的资源不断更新资产价值评估方面的举措。个案对应该州的资产。数据文件中的每个个案记录资产所在的县、镇和区, 最后一次评估距今的时间以及当时的估价。
- **property\_assess\_cs\_sample.sav**。该假设数据文件包含在 property\_assess\_cs.sav 中列出的资产的样本。该样本是根据 property\_assess\_csplan 中指定的设计来选取的, 而且该数据文件记录包含概率和样本权重。在选取了样本之后, 附加变量 Current value 将被收集并添加到数据文件。
- **recidivism.sav**。该假设数据文件涉及某政府执法机构在了解其管辖区域内的屡犯率方面的举措。每个个案对应先前的一名罪犯, 并记录其人口统计信息和第一次犯罪的详细资料; 如果在第一次被捕后两年内又第二次被捕, 则还将记录两次被捕间隔的时间。
- **recidivism\_cs\_sample.sav**。该假设数据文件涉及某政府执法机构在了解其管辖区域内的屡犯率方面的举措。每个个案对应应在 2003 年 6 月期间第一次被捕释放的先前的一名罪犯, 并记录其人口统计信息和第一次犯罪的详细资料, 及其第二次被捕的数据(如果发生在 2006 年 6 月底之前)。根据 recidivism\_cs\_csplan 中指定的抽样计划从抽样部门选择罪犯; 该计划使用与大小成正比(PPS)方法, 因此, 还有一个文件(recidivism\_cs\_jointprob.sav)包含联合选择概率。
- **rfm\_transactions.sav**。此假设数据文件包含购买交易数据, 即每笔交易的购买日期、购买商品和消费金额。

- **salesperformance.sav**。这是关于评估两个新的销售培训课程的假设数据文件。60 名员工被分成 3 组且都接受标准的培训。另外，组 2 接受技术培训；组 3 接受实践教程。在培训课程结束时，对每名员工进行测验并记录他们的分数。数据文件中的每个个案代表一名单独的受训者，并记录其被分配到的组以及测验的分数。
- **satisf.sav**。该假设数据文件涉及某零售公司在 4 个商店位置所进行的满意度调查。总共对 582 位客户进行了调查，每个个案代表一位单独客户的回答。
- **screws.sav**。该数据文件包含关于螺钉、螺栓、螺母和图钉的特征的信息 (Hartigan, 1975)。
- **shampoo\_ph.sav**。这是关于某发制品厂的质量控制的假设数据文件。在规定的间隔对六批独立输出的产品进行检测并记录它们的 pH 值。目标范围是 4.5 - 5.5。
- **ships.sav**。在别处被提出和分析的 (McCullagh 等., 1989) 关于波浪对货船造成的损坏的数据集。在给定了船的类型、建造工期和服务期后，可以根据泊松比率发生来为事件计数建模。在因子交叉分类构成的表格中，每个单元格的分类汇总服务月数提供遇到风险的值。
- **site.sav**。该假设数据文件涉及某公司在为扩展业务而选择新址方面的举措。该公司聘请了两名顾问分别对选址进行评估，除了提供长期报告外，他们还要以“前景颇佳”、“前景良好”或“前景不佳”来对每个选址进行总结。
- **smokers.sav**。该数据文件摘自 1998 年全国家庭药物滥用调查并且是美国家庭的概率样本。(<http://dx.doi.org/10.3886/ICPSR02934>) 因此，分析该数据文件的第一步应该是对数据进行加权以反映总体趋势。
- **stocks.sav** 该假设数据文件包含某一年的股票价格和成交量。
- **stroke\_clean.sav**。该假设数据文件包含某医学数据库在经过“数据准备”选项中的过程清理后的状态。
- **stroke\_invalid.sav**。该假设数据文件包含某医学数据库的初始状态及一些数据输入错误。
- **stroke\_survival**。此假设数据文件涉及正在研究结束缺血性中风后复元计划的患者存活时间的研究人员面临着很多挑战。中风后，记录心肌梗塞、缺血性中风或出血性中风的发生及其时间。样本为左侧截短，因为只包含在中风后管理的复元计划结束后存活的患者。
- **stroke\_valid.sav**。该假设数据文件包含在使用“验证数据”过程检查值后，某医学数据库的状态。它仍包含潜在异常个案。
- **survey\_sample.sav**。此数据文件包含调查数据，包括人口统计学数据和各种态度测量。它基于 1998 NORC 综合社会调查的变量子集，但某些数据值已经过修改，并添加了其他虚拟变量以供演示用途。
- **telco.sav**。该假设数据文件涉及某电信公司在减少客户群中的客户流失方面的举措。每个个案对应一个单独的客户，并记录各类人口统计和服务用途信息。
- **telco\_extra.sav**。该数据文件与 telco.sav 数据文件类似，但删除了“tenure”和经对数转换的客户消费变量，代替它们的是标准化的对数转换客户消费变量。
- **telco\_missing.sav**。该数据文件是 telco.sav 数据文件的子集，但某些人口统计数据值已被缺失值替换。

- **testmarket.sav**。该假设数据文件涉及某快餐连锁店为其菜单添加新项目的计划。有三种可能的促销新产品的活动，所以会在多个随机选择的地点引入新的项目。在每个地点采用不同的促销方式，并记录新项目四周的每周销售情况。每个个案对应单独地点的一周。
- **testmarket\_1month.sav**。该假设数据文件是在数据文件 `testmarket.sav` 的基础上加上了每周销售“累计”，所以每个个案对应一个单独的地点。所以，一些每周更改的变量消失了，而且现在记录的销售是为期四周的研究过程中的销售之和。
- **tree\_car.sav**。该假设数据文件包含人口统计和车辆购买价格数据。
- **tree\_credit.sav**。该假设数据文件包含人口统计和银行贷款历史数据。
- **tree\_missing\_data.sav**。该假设数据文件包含具有大量缺失值的人口统计和银行贷款历史数据。
- **tree\_score\_car.sav**。该假设数据文件包含人口统计和车辆购买价格数据。
- **tree\_textdata.sav**。这是一个只有两个变量的样本数据文件，主要打算在指定测量级别和值标签之前显示变量的默认状态。
- **tv-survey.sav**。该假设数据文件涉及由某电视演播室进行的一项关于是否要继续制作一档成功的节目的调查。906 位调查对象被问及他们在各种情况下是否会收看该节目。每行代表一位单独的调查对象；每列代表一种单独的情况。
- **ulcer\_recurrence.sav**。此文件包含某项研究的部分信息，该研究旨在比较两种用来防止溃疡复发的治疗的疗效。它提供了区间数据的优秀示例并且已在别处被提出和分析 (Collett, 2003)。
- **ulcer\_recurrence\_recoded.sav**。该文件重新组织 `ulcer_recurrence.sav` 中的信息以允许为研究的每个区间的事件概率建模而不是简单地研究结束事件概率建模。它已在别处被提出和分析 (Collett 等., 2003)。
- **verd1985.sav**。该数据文件涉及某项调查 (Verdegaal, 1985)。该调查记录了 15 个主体对 8 个变量的响应。需要处理的变量被分成 3 个集。数据集 1 包含年龄和婚姻；数据集 2 包含宠物和新闻；数据集 3 包含音乐和居住。宠物被尺度化为多名义而年龄被尺度化为有序；所有其他变量都被尺度化为单名义。
- **virus.sav**。该假设数据文件涉及某因特网服务提供商 (ISP) 在确定病毒对其网络的影响方面的举措。他们从发现病毒到威胁得以遏制这段时间内跟踪其网络上受感染的电子邮件的流量的 (近似) 百分比。
- **wheeze\_steubenville.sav**。这是关于空气污染对儿童健康影响的纵向研究的一个子集 (Ware, Dockery, Spiro III, Speizer, 和 Ferris Jr., 1984)。这些数据包含儿童的气喘状况的重复二分类测量 (这些儿童来自 Steubenville, Ohio, 年龄为 7 到 10 岁)，以及母亲在研究的第一年中是否为吸烟者的固定记录。
- **workprog.sav**。该假设数据文件涉及一份尝试为弱势群体提供较好的工作的政府工作计划。文件后还有一个潜在计划参与者的样本，其中一些参与者是被随机选择来参加该计划的，而其他参与者则不是。每个个案代表一位单独的计划参与者。
- **worldsales.sav**。该假设数据文件包含按不同大洲和产品列出的销售收入。

# 注意事项

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家/地区中不提供在本文档中讨论的产品、服务或功能。请咨询您当地的 IBM 代表以了解有关您所在地区当前可用产品和服务的信息。任何对 IBM 产品、程序或服务的引用，并不意味着仅可使用这些 IBM 产品、程序或服务。作为替代，可以使用任何功能相当的产品、程序或服务，前提是不侵犯任何 IBM 知识产权。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

在本文档中介绍的主题可能涉及 IBM 的专利或申请中的专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY  
10504-1785, U. S. A.

要查询双字节字符集 (DBCS) 相关许可证信息，请联系所在国家/地区中的 IBM 知识产权部门，或者以书面形式将查询函件发送至：

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan  
Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

**以下段落不适用于英国或此类条款与当地法律不符的其他国家/地区：** INTERNATIONAL BUSINESS MACHINES 公司“按原样”提供本出版物，不保证任何明示或暗示，包括但不限于对非侵权性、适销性或对特定用途适用性的暗示担保。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可能随时对本出版物中所述的产品和/或程序进行改进和/或更改，恕不另行通知。

在本信息材料中对任何非 IBM 网站的引用仅为了方便用户，并不以任何方式表明对这些网站的认可。这些网站上的材料并非本 IBM 产品材料的一部分，您对这些网站的使用需自担风险。

IBM 可以自认为适当并且不会对您构成任何约束的任何方式使用或分发您提供的任何信息。

如果本程序的受许可方试图了解有关程序的信息以启用：(i) 在独立创建的程序和其他程序（包括本程序）之间交换信息；(ii) 相互使用交换的信息，则应联系：

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606,  
USA.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

在本文档中介绍的受许可保护程序，及其所有受许可保护材料由 IBM 在双方签署的“IBM 客户协议”、“IBM 国际程序许可证协议”或任何其他等同协议下提供。

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。IBM 并未测试过这些产品，无法确认有关非 IBM 产品的性能准确性、兼容性或任何其他声明。有关非 IBM 产品功能的问题应由这些产品的供应商负责。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

如果您正在查阅此信息的软拷贝，照片和彩色插图可能不会显示。

## 商标

IBM、IBM 徽标、ibm.com 和 SPSS 是 IBM Corporation 的商标，在全球许多司法辖区注册。有关最新的 IBM 商标列表，请访问网页 <http://www.ibm.com/legal/copytrade.shtml>。

Adobe、Adobe 徽标、PostScript 和 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Java 和所有基于 Java 的商标和徽标是 Sun Microsystems, Inc. 在美国、其他国家或地区或两者的商标。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

此产品使用 Polar 工程咨询公司的 WinWrap Basic，版权 1993 年-2007 年，<http://www.winwrap.com>。

其他产品和服务名称可能是 IBM 或其他公司的商标。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。



---

# 参考书目

- Bell, E. H. 1961. Social foundations of human behavior: Introduction to the study of sociology. New York: Harper & Row.
- Blake, C. L., 和 C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Box, G. E. P., G. M. Jenkins, 和 G. C. Reinsel. 1994. Time series analysis: Forecasting and control, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.
- Breiman, L., 和 J. H. Friedman(F). 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 80, .
- Collett, D. 2003. Modelling survival data in medical research, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Gardner, E. S. 1985. Exponential smoothing: The state of the art. Journal of Forecasting, 4, .
- Green, P. E., 和 V. Rao. 1972. Applied multidimensional scaling. Hinsdale, Ill.: Dryden Press.
- Green, P. E., 和 Y. Wind. 1973. Multiattribute decisions in marketing: A measurement approach. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. Psychometrika, 33, .
- Hartigan, J. A. 1975. Clustering algorithms. New York: John Wiley and Sons.
- Hastie, T., 和 R. Tibshirani. 1990. Generalized additive models. London: Chapman and Hall.
- Kennedy, R., C. Riquier, 和 B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. Journal of Targeting, Measurement, and Analysis for Marketing, 5, .
- McCullagh, P., 和 J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. London: Chapman & Hall.
- Pena, D., G. C. Tiao, 和 R. S. Tsay, eds. 2001. A course in time series analysis. New York: John Wiley and Sons.
- Price, R. H., 和 D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. Journal of Personality and Social Psychology, 30, .
- Rickman, R., N. Mitchell, J. Dingman, 和 J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. Journal of the American Medical Association, 228, .
- Rosenberg, S., 和 M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. Multivariate Behavioral Research, 10, .
- Van der Ham, T., J. J. Meulman, D. C. Van Strien, 和 H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. British Journal of Psychiatry, 170, .

---

Verdegaal, R. 1985. Meer sets analyse voor kwalitatieve gegevens (in Dutch). Leiden: Department of Data Theory, University of Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, 和 B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .

- ACF  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16  
 纯 ARIMA 过程图, 89
- ARIMA 模型, 5  
 季节性阶, 10  
 差分阶, 10  
 常数, 10  
 离群值, 13  
 移动平均数阶, 10  
 自回归阶, 10  
 转换函数, 11
- ARIMA 模型参数表  
 在“时间序列建模器”中, 65
- Box-Ljung 统计量  
 在?应用时间序列模型?中, 26  
 在时间序列建模器中, 14  
 在“时间序列建模器”中, 65
- Brown 的指数平滑法模型, 8  
 Holt 的指数平滑法模型, 8
- MAE, 87  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16
- MAPE, 87  
 在?应用时间序列模型?中, 26, 28  
 在“应用时间序列模型”中, 56  
 在时间序列建模器中, 14, 16  
 在“时间序列建模器”中, 48
- MaxAE, 87  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16
- MaxAPE, 87  
 在?应用时间序列模型?中, 26, 28  
 在“应用时间序列模型”中, 56  
 在时间序列建模器中, 14, 16  
 在“时间序列建模器”中, 48
- outliers  
 专家建模器, 60
- PACF  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16  
 纯 ARIMA 过程图, 89
- $R^2$ , 87  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16
- RMSE, 87  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16
- Winters 的指数平滑法模型  
 可乘, 8  
 可加, 8
- XML  
 以 XML 格式保存时间序列模型, 19, 46, 61  
 以 XML 格式保存重新估计的模型, 31
- 专家建模器, 5, 41  
 outliers, 60  
 离群值, 7  
 限制模型空间, 6, 44
- 事件, 6  
 在时间序列建模器中, 6
- 估计期, 2  
 在?应用时间序列模型?中, 24  
 在时间序列建模器中, 5  
 在“时间序列建模器”中, 43
- 保存  
 XML 格式的模型规格, 19  
 XML 格式的重新估计的模型, 31  
 新变量名称, 19, 31  
 模型预测值, 19, 31
- 保持个案, 2
- 偏自相关函数  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16  
 纯 ARIMA 过程图, 89
- 创新离群值, 88  
 在时间序列建模器中, 7, 13
- 历史数据  
 在?应用时间序列模型?中, 28  
 在时间序列建模器中, 16
- 历史期, 2
- 变量名  
 在?应用时间序列模型?中, 31  
 在时间序列建模器中, 19
- 可加的修补离群值, 88  
 在时间序列建模器中, 7, 13
- 可加离群值, 88  
 在时间序列建模器中, 7, 13
- 周期性  
 在时间序列建模器中, 6, 8, 10 - 11
- 周期性分解  
 周期性日期成分, 76  
 新变量, 80  
 相关过程, 82
- 商标, 101
- 均方根误差, 87  
 在?应用时间序列模型?中, 26, 28  
 在时间序列建模器中, 14, 16
- 季节性分解, 34 - 36  
 保存新变量, 35

- 假设, 34
- 创建变量, 35
- 模型, 34
- 计算移动平均数, 34
- 季节性可加离群值, 88
  - 在时间序列建模器中, 7, 13
- 季节性差分变换
  - ARIMA 模型, 10
- 季节性阶
  - ARIMA 模型, 10
  
- 对数转换
  - 在时间序列建模器中, 8, 10 - 11
  
- 局部趋势离群值, 88
  - 在时间序列建模器中, 7, 13
  
- 差分变换
  - ARIMA 模型, 10
  
- 平均绝对误差, 87
  - 在?应用时间序列模型?中, 26, 28
  - 在时间序列建模器中, 14, 16
- 平均绝对误差百分比, 87
  - 在?应用时间序列模型?中, 26, 28
  - 在“应用时间序列模型”中, 56
  - 在时间序列建模器中, 14, 16
  - 在“时间序列建模器”中, 48
- 平方根转换
  - 在时间序列建模器中, 8, 10 - 11
- 平稳的  $R^2$ , 87
  - 在?应用时间序列模型?中, 26, 28
  - 在时间序列建模器中, 14, 16
  - 在“时间序列建模器”中, 65
  
- 应用时间序列模型, 23, 52, 66
  - Box-Ljung 统计量, 26
  - 以 XML 格式保存重新估计的模型, 31
  - 估计期, 24
  - 保存预测值, 31, 54
  - 拟合优度统计量, 26, 28, 55
  - 拟合值, 28
  - 新变量名称, 31, 56
  - 最佳和最差拟合模型, 29
  - 模型参数, 26
  - 模型拟合表, 55
  - 残差偏自相关函数, 26, 28
  - 残差自相关函数, 26, 28
  - 缺失值, 32
  - 置信区间, 28, 32
  - 跨所有模型的统计量, 26, 28, 55
  - 重新估计模型参数, 24, 53
  - 预测值, 26, 28, 74
  - 预测期, 24, 53, 73
  - 预测表, 74
  
- 拟合优度
  - 在?应用时间序列模型?中, 26, 28
  - 在“应用时间序列模型”中, 55
  - 在时间序列建模器中, 14, 16
  - 在“时间序列建模器”中, 47
  - 定义, 87
- 拟合值
  - 在?应用时间序列模型?中, 28
  - 在时间序列建模器中, 16
  - 在“时间序列建模器”中, 63
- 指数平滑法模型, 5, 8
  
- 时间序列建模器, 3
  - ARIMA, 5, 10
  - ARIMA 模型参数表, 65
  - Box-Ljung 统计量, 14
  - outliers, 60
  - 专家建模器, 5, 41, 57
  - 事件, 6
  - 以 XML 格式保存模型规格, 19, 46, 61
  - 估计期, 5, 43
  - 保存预测值, 19, 46
  - 周期性, 6, 8, 10 - 11
  - 序列转换, 8, 10 - 11
  - 拟合优度统计量, 14, 16, 47, 65
  - 拟合值, 16, 63
  - 指数平滑法, 5, 8
  - 新变量名称, 19, 50
  - 最佳和最差拟合模型, 18
  - 模型参数, 14, 62
  - 模型名称, 20
  - 模型描述表, 64
  - 模型统计表, 65
  - 残差偏自相关函数, 14, 16
  - 残差自相关函数, 14, 16
  - 离群值, 7, 13
  - 缺失值, 20
  - 置信区间, 16, 20
  - 跨所有模型的统计量, 14, 16, 47, 49
  - 转换函数, 11
  - 预测值, 14, 16, 47
  - 预测期, 5, 20, 43, 45
  - 预测表, 51
  
- 最大绝对误差, 87
  - 在?应用时间序列模型?中, 26, 28
  - 在时间序列建模器中, 14, 16
- 最大绝对误差百分比, 87
  - 在?应用时间序列模型?中, 26, 28
  - 在“应用时间序列模型”中, 56
  - 在时间序列建模器中, 14, 16
  - 在“时间序列建模器”中, 48
  
- 标准化 BIC (BIC 准则), 87
  - 在?应用时间序列模型?中, 26, 28
  - 在时间序列建模器中, 14, 16

## 索引

- 样本文件
  - 位置, 93
- 模型
  - ARIMA, 5, 10
  - 专家建模器, 5
  - 指数平滑法, 5, 8
- 模型参数
  - 在“应用时间序列模型”中, 26
  - 在时间序列建模器中, 14
  - 在“时间序列建模器”中, 62
- 模型名称
  - 在时间序列建模器中, 20
- 模型拟合表
  - 在“应用时间序列模型”中, 55
- 模型描述表
  - 在“时间序列建模器”中, 64
- 模型统计表
  - 在“时间序列建模器”中, 65
  
- 残差
  - 在“应用时间序列模型”中, 26, 28
  - 在时间序列建模器中, 14, 16
  
- 水平移位离群值, 88
  - 在时间序列建模器中, 7, 13
- 法律注意事项, 100
  
- 瞬时离群值, 88
  - 在时间序列建模器中, 7, 13
  
- 离群值
  - ARIMA 模型, 13
  - 专家建模器, 7
  - 定义, 88
  
- 积分
  - ARIMA 模型, 10
- 移动平均数
  - ARIMA 模型, 10
  
- 简单的季节性指数平滑法模型, 8
- 简单的指数平滑法模型, 8
  
- 缺失值
  - 在“应用时间序列模型”中, 32
  - 在时间序列建模器中, 20
  
- 置信区间
  - 在“应用时间序列模型”中, 28, 32
  - 在时间序列建模器中, 16, 20
  
- 自回归
  - ARIMA 模型, 10
- 自然对数转换
  - 在时间序列建模器中, 8, 10 - 11
- 自相关函数
  - 在“应用时间序列模型”中, 26, 28
  - 在时间序列建模器中, 14, 16
  - 纯 ARIMA 过程图, 89
  
- 调和分析, 37
- 谱图, 37, 39
  - 假设, 37
  - 双变量谱分析, 38
  - 周期图, 84
  - 居中变换, 38
  - 相关过程, 86
  - 谱密度, 84
  - 频谱窗口, 37
  
- 转换函数, 11
  - 分子的阶, 11
  - 分母的阶, 11
  - 季节性阶, 11
  - 差分阶数, 11
  - 延迟, 11
  
- 重新估计模型参数
  - 在“应用时间序列模型”中, 24
  - 在“应用时间序列模型”中, 53
  
- 阻尼指数平滑法模型, 8
  
- 预测值
  - 在“应用时间序列模型”中, 26, 28
  - 在“应用时间序列模型”中, 74
  - 在时间序列建模器中, 14, 16
  - 在“时间序列建模器”中, 47
- 预测期
  - 在“应用时间序列模型”中, 24
  - 在“应用时间序列模型”中, 53, 73
  - 在时间序列建模器中, 5, 20
  - 在“时间序列建模器”中, 43, 45
- 预测表
  - 在“应用时间序列模型”中, 74
  - 在“时间序列建模器”中, 51
  
- 验证期, 2