

1. 概述

本节以工业用地更新数据为例，利用 SPSS 软件演示常见的描述性统计方法及统计图的应用。SPSS 的数据文件为“manufacture.sav”，同时还提供了同名的 csv 文件。

2. 初识 SPSS

2.1 数据视图与变量视图

在 SPSS 中打开“manufacture.sav”，如图 1 所示。

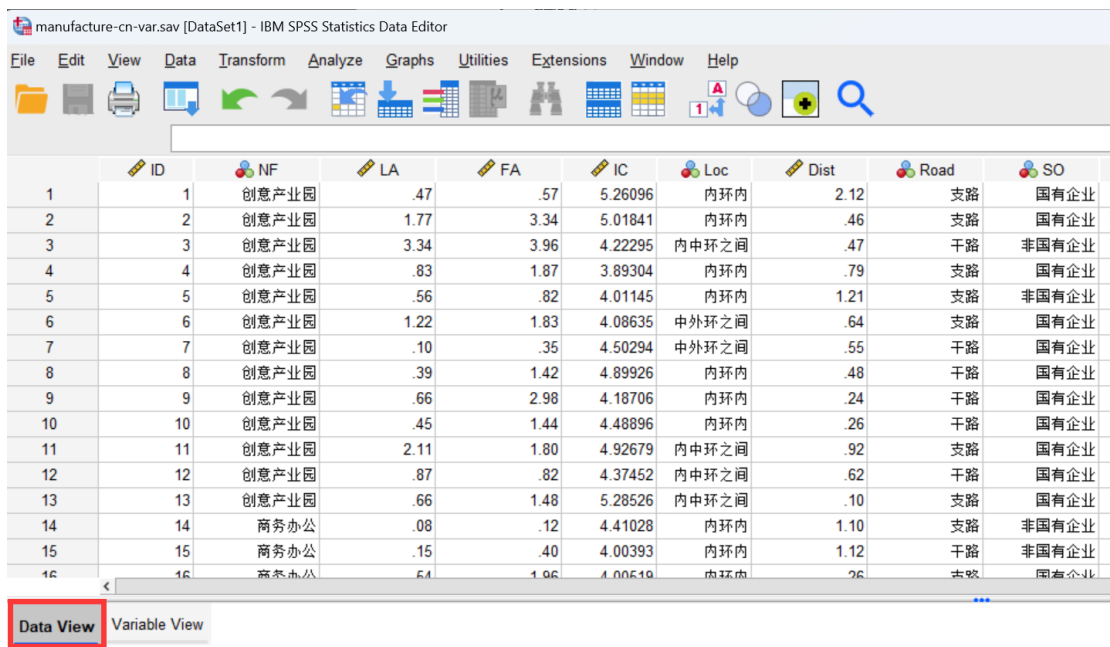


图 1 初始界面——数据视图

从左下角高亮显示的红框可以看到：SPSS 有两个基本视图——数据视图（Data View）和变量视图（Variable View）。当前我们正处于数据视图，它以表格的方式显示了数据内容。表格的每一行是一个样本，每一列是一个变量。

点击红框右侧的“Variable View”，切换至变量视图，如图 2 所示。该视图显示的是变量的基本属性，每一行是一个变量，每一列是一种属性，这里简单介绍几个最常用的属性。“Name”列是变量名称，通常是较为精炼的英文；“Label”列是变量的标签，用于对变量的含义做更详细的说明；“Values”列是分类变量的值标签，下面再重点介绍；“Measure”列是变量的测量方式，包括“Scale”（数

值变量）、“Ordinal”（有序分类变量）、“Nominal”（无序分类变量）三种。有一些分析对测量方式比较敏感，错误的设置（例如将无序分类变量设置成了“Scale”）会导致不正确的结果。本例中，用地面积（LA）、建筑面积（FA）、到市中心的距离（Dist）属于数值变量，而更新后的功能（NF）、区位（Loc）属于分类变量。

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0		None	None	8	Right	Scale	Input
2	NF	Numeric	8	0	更新后的使用功能	{1. 创意产业园...}	None	11	Right	Nominal	Input
3	LA	Numeric	8	2	用地面积	None	None	10	Right	Scale	Input
4	FA	Numeric	8	2	建筑面积	None	None	11	Right	Scale	Input
5	IC	Numeric	8	5	用地不规则系数	None	None	8	Right	Scale	Input
6	Loc	Numeric	8	0	区位	{1. 内环内...}	None	8	Right	Nominal	Input
7	Dist	Numeric	8	2	到市中心的距离	None	None	10	Right	Scale	Input
8	Road	Numeric	8	0	道路条件	{1. 干路...}	None	10	Right	Nominal	Input
9	SO	Numeric	8	0	单位性质	{1. 国有企业...}	None	10	Right	Nominal	Input
10											

图 2 变量视图


2.2 分类变量的值标签

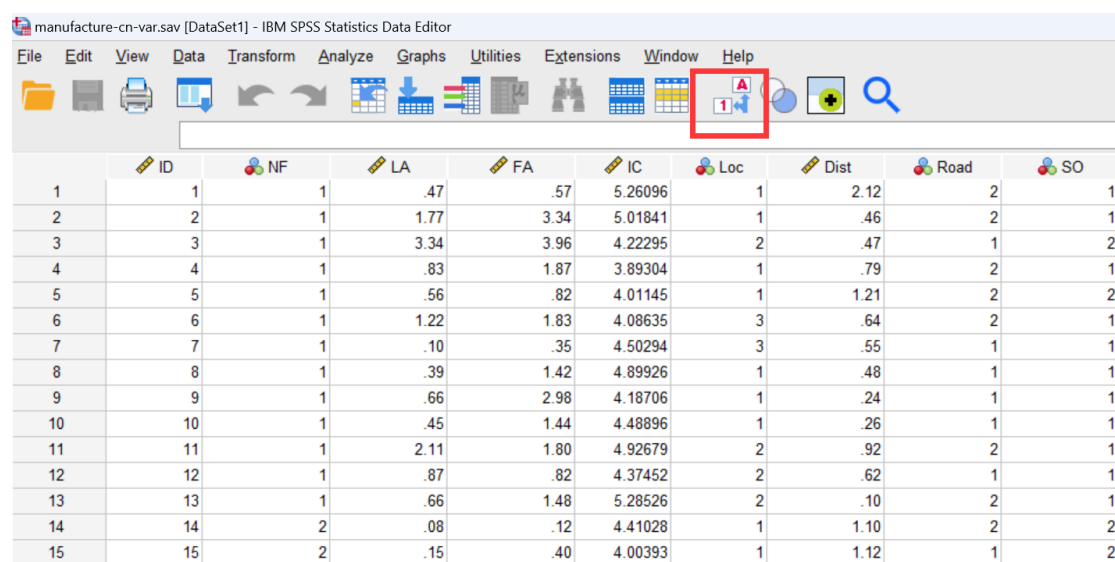
下面重点介绍值标签的应用。对于数值变量，我们直接使用其原始值即可；而对于分类变量，我们通常不直接使用文本格式的类别名称。这是因为与数字相比，文本非常不利于查询和分析，且大量重复的文本将占用过高的存储空间。作为替代，我们一般对类别进行编码，以一个数字代表一种类别，再通过值标签建立数字与类别名称的对应关系。以上图中高亮显示的区位（Loc）变量为例，点击该行所对应的“Values”列，如图 3 所示。



图 3 在变量视图中设置值标签

可以看到，该变量有 3 个取值：1 代表“内环内”；2 代表“内中环”之间，3 代表“中外环”之间。我们可以在“Value”框中输入新的数字，在“Label”框中输入该数字所代表的类别名称，再点击“Add”建立新的类别编码。也可以通过“Change”或“Remove”按钮编辑或删除已有的编码。

点击左下方的“Data View”回到数据视图，确认区位和其他分类变量显示的都是类别标签。我们可以通过点击工具栏上的  图标（见图 4 红框），或者在菜单栏中选择“View → Value Labels”，在数字编码和类别标签之间相互转换。图 4 显示的是数字编码的状态，再次点击则将回到类别标签状态。



	ID	NF	LA	FA	IC	Loc	Dist	Road	SO
1	1	1	.47	.57	5.26096	1	2.12	2	1
2	2	1	1.77	3.34	5.01841	1	.46	2	1
3	3	1	3.34	3.96	4.22295	2	.47	1	2
4	4	1	.83	1.87	3.89304	1	.79	2	1
5	5	1	.56	.82	4.01145	1	1.21	2	2
6	6	1	1.22	1.83	4.08635	3	.64	2	1
7	7	1	.10	.35	4.50294	3	.55	1	1
8	8	1	.39	1.42	4.89926	1	.48	1	1
9	9	1	.66	2.98	4.18706	1	.24	1	1
10	10	1	.45	1.44	4.48896	1	.26	1	1
11	11	1	2.11	1.80	4.92679	2	.92	2	1
12	12	1	.87	.82	4.37452	2	.62	1	1
13	13	1	.66	1.48	5.28526	2	.10	2	1
14	14	2	.08	.12	4.41028	1	1.10	2	2
15	15	2	.15	.40	4.00393	1	1.12	1	2

图 4 数据视图中显示分类变量的数字编码

3. 描述性统计方法

3.1 分类变量：频数分析

在菜单栏中选择“Analyze → Descriptive Statistics → Frequencies”，进入频数统计对话框。将“更新后的使用功能 [NF]”与“区位 [Loc]”从左侧的全体变量列表中选入右侧待分析的“Variable(s)”列表中，请注意各变量名左侧用以提示变量类型的小图标，这两个变量都是无序分类变量。确认左下角的“Display frequency tables”被勾选，如图 5 所示。

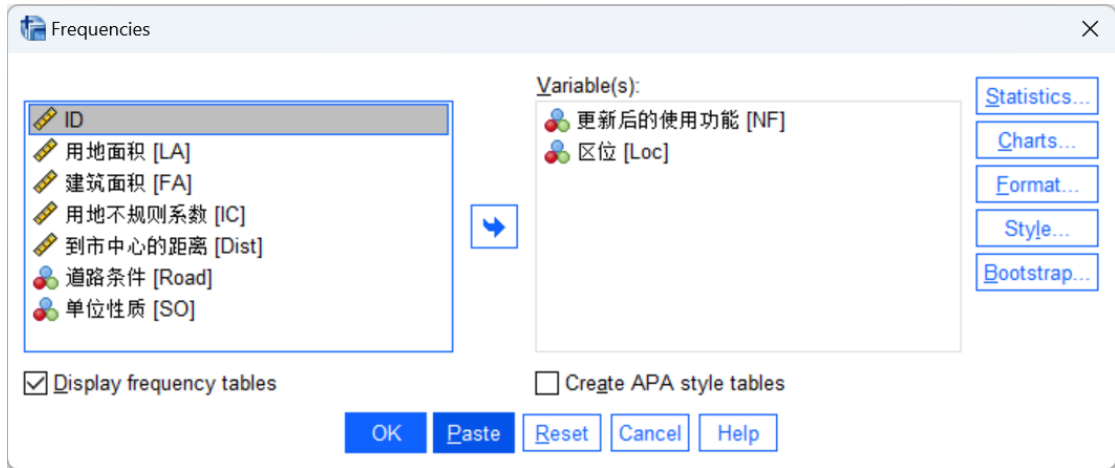


图 5 频数统计设定

点击“OK”运行分析，结果将在单独的输出窗口中呈现。首先显示的是样本量，如图 6 所示。可以看到，这两个变量均在 113 个样本（“Valid”），没有缺失样本（“Missing”）。

Statistics

		更新后的使用功能	区位
N	Valid	113	113
	Missing	0	0

图 6 频数统计结果：统计量

下面呈现的是主要分析结果——频数表。这里以更新的使用功能为例，如图 7 所示。

Frequency Table

更新后的使用功能

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	创意产业园	13	11.5	11.5	11.5
	商务办公	33	29.2	29.2	40.7
	商业服务	30	26.5	26.5	67.3
	生产物流	37	32.7	32.7	100.0
Total		113	100.0	100.0	

图 7 频数统计结果：频数表

表的左侧列出了各个类别的名称。“Frequency”列统计了每一类出现的频数，注意“Total”行对应的是所有类别的频数之和，亦即样本量。“Percent”和“Valid

Percent”均是某一类的频数占总频数的百分比，区别在于：前者的分母是全部样本量，包括缺失样本；后者的分母则是有效样本量，不包括缺失样本。由于本例没有缺失样本，所以二者的结果是一样的。“Cumulative Percent”是累积百分比，某一行上的取值是前面所有行“Valid Percent”的和，例如：商业服务的累积百分比为 67.3%，表明该类别及其上面的所有类别（创新产业园+商务办公+商业服务）的累计占比为 67.3%。

SPSS 中的每个分析功能通常都有一系列的选项，让我们对分析做更详细的设置。例如在频数分析中，结果默认按类别名称的自然顺序（数字编码的顺序，或文本的字母表顺序）升序显示，我们也可以要求其按照频数的大小降序显示，以便于发现占比较高的类别。再次选择“Analyze → Descriptive Statistics → Frequencies”，执行相同的操作，只是在运行分析之前，点击“Format”按钮，在弹出的对话框中，把左侧的“Order by”由默认的“Ascending values”改为“Descending counts”，如图 8 所示。

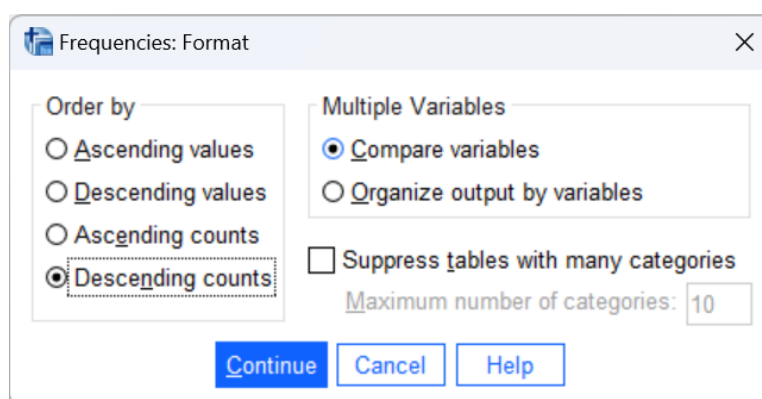


图 8 按频数降序排列的设定

点击“Continue”回到主界面，继续按原操作运行分析，此时呈现的频数表中，各类型的顺序已按其频数由高到低排列，如图 9 所示。

更新后的使用功能

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	生产物流	37	32.7	32.7	32.7
	商务办公	33	29.2	29.2	61.9
	商业服务	30	26.5	26.5	88.5
	创意产业园	13	11.5	11.5	100.0
	Total	113	100.0	100.0	

图 9 按频数降序排列的频数表结果

3.2 数值变量：统计量分析

在菜单栏中选择“Analyze → Descriptive Statistics → Descriptives”，进入统计量分析对话框。将“用地面积 [LA]”与“到市中心的距离 [Dist]”选入待分析的“Variable(s)”变量列表中，如图 10 所示。



图 10 统计量分析主窗口设定

SPSS 默认只会生成少数几个统计量。我们可以点击“Options”按钮，要求其报告更多的统计量，如图 11 所示。除了已经默认勾选的常用统计量外，再将“Variance”（方差）、“Range”（极差）、“Kurtosis”（峰度）、“Skewness”（偏度）也选中，点击“Continue”回到主对话框。

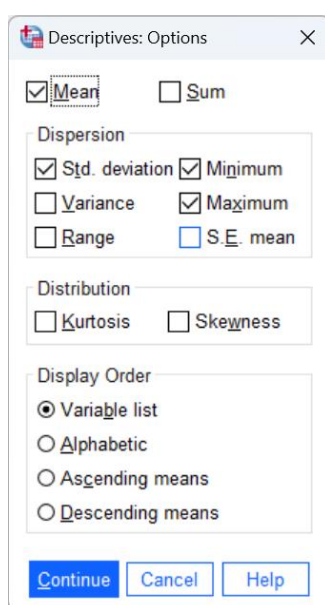


图 11 统计量选项设定

点击“OK”运行分析，结果如图 12 所示。表头是统计量的名称，其中，“Statistic”是统计量的值，也是我们主要关心的内容；“Std. Error.”是统计量的估计误差，属于推断性统计的内容，我们暂且跳过。下面的两行分别报告了用地面积与到市中心的距离这两个变量的统计结果。

Descriptive Statistics											
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
用地面积	113	3.31	.03	3.34	.5133	.56385	.318	2.759	.227	9.194	.451
到市中心的距离	113	3.21	.10	3.30	1.1509	.79891	.638	1.016	.227	.011	.451
Valid N (listwise)	113										

图 12 统计量分析结果

在 SPSS 中，某些简单的分析任务可以通过多种程序实现，它们的功能基本相同，但可能有一些细微的差异，统计量即是其中一个例子。我们不仅可以通过上面的方式计算统计量，也可以通过频数统计的附属功能实现。在菜单栏中选择“Analyze → Descriptive Statistics → Frequencies”，进入频数统计对话框。首先点击“Reset”按钮进行重置，然后在拟分析的“Variable(s)”变量列表中选入“用地面积 [LA]”与“到市中心的距离 [Dist]”这两个变量。作为数值变量，为它们生成频数表是没有意义的，因此，我们取消对于左下角“Display frequency tables”的勾选。同时，点击右侧的“Statistics”按钮，弹出统计量对话框，如图 13 所示。

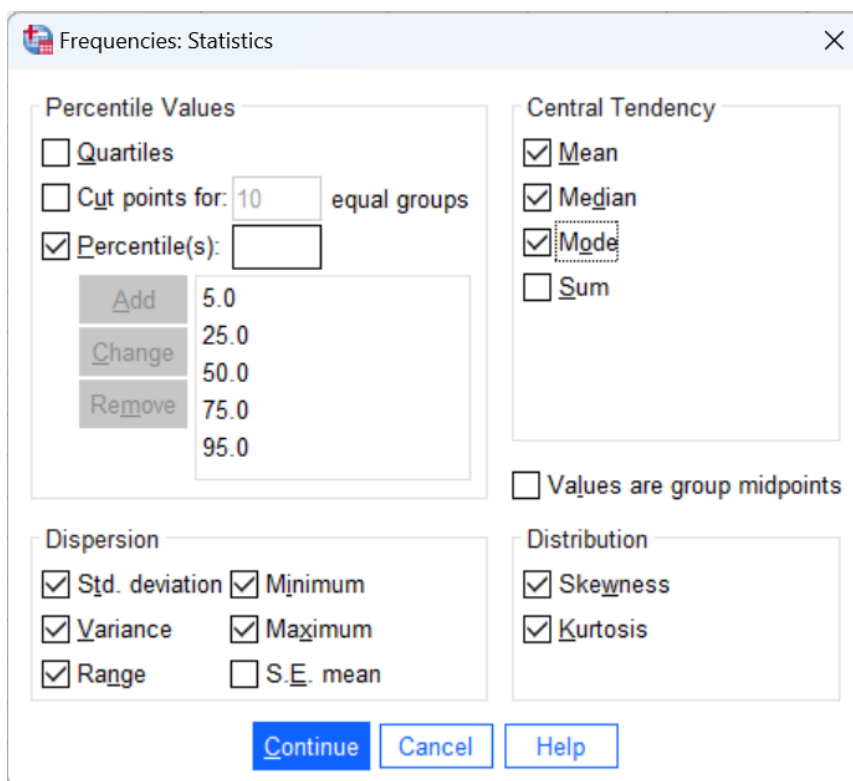


图 13 使用频数统计开展统计量分析的设置

可以看到，这里可供选择的统计量更加丰富。我们在“Central Tendency”（集中趋势）中选中“Mean”（均值）、“Median”（中位数）、“Mode”（众数）；在“Dispersion”（离散趋势）中选中“Std. deviation”（标准差）、“Variance”（方差）、“Range”（极差）、“Minimum”（最小值）、“Maximum”（最大值）；在“Distribution”（分布）中选中“Skewness”（偏度）、“Kurtosis”（峰度）。最后，我们在“Percentile Values”中设置分位数统计量：勾选“Percentile(s)”，在后面的框中输入“5”，再点击“Add”，即将5%分位数加入列表中，用类似的办法依次加入25%、50%、75%、95%分位数；也可以勾选“Quartiles”，直接将3个四分位数（25%、50%、75%）加入。

点击“Continue”回到主对话框，再点击“OK”运行分析，结果如图14所示。可以看到，两种方法得到的结果是相同的。下图的脚注提示：对于“到市中心的距离”这个变量，有多个众数存在，这里显示的是值最小的一个。事实上，由于数值型变量可以无限取值，众数的意义不大，一般不纳入分析。

		用地面积	到市中心的距离
N	Valid	113	113
	Missing	0	0
Mean		.5133	1.1509
Median		.3600	.8869
Mode		.66	.10 ^a
Std. Deviation		.56385	.79891
Variance		.318	.638
Skewness		2.759	1.016
Std. Error of Skewness		.227	.227
Kurtosis		9.194	.011
Std. Error of Kurtosis		.451	.451
Range		3.31	3.21
Minimum		.03	.10
Maximum		3.34	3.30
Percentiles	5	.0696	.2091
	25	.1583	.6119
	50	.3600	.8869
	75	.6475	1.5027
	95	1.8570	2.7610

a. Multiple modes exist. The smallest value is shown

图14 使用频数统计开展统计量分析的结果

4. 统计图

SPSS 的绘图功能均在 Graphs 菜单中。老版本的 SPSS 把每一种统计图作为一个单独的程序。从版本 16.0 开始，SPSS 引入了新的“Chart Builder”功能，将多种统计图的绘制过程整合到一起，提供更为直观和灵活的图表创建界面。旧的绘图功能被归类为“Legacy Dialogs”，以兼容旧版本的用户习惯。接下来，我们将以“Chart Builder”为主，兼顾“Legacy Dialogs”，演示两种方式的绘图操作。

4.1 分类变量：柱状图、折线图、饼状图

本小节以更新后的使用功能为例，绘制柱状图、折线图、饼状图。柱状图和折线图采用 Chart Builder 绘制，饼状图采用传统对话框绘制。

(1) 柱状图

在菜单栏中选择“Graphs → Chart Builder”，打开图表构建器对话框。首先将弹出如图 15 所示的提醒窗口，告诉我们需要妥善设置每个变量的测量方式(数值、有序分类、无序分类)，以及为分类变量设置值标签。

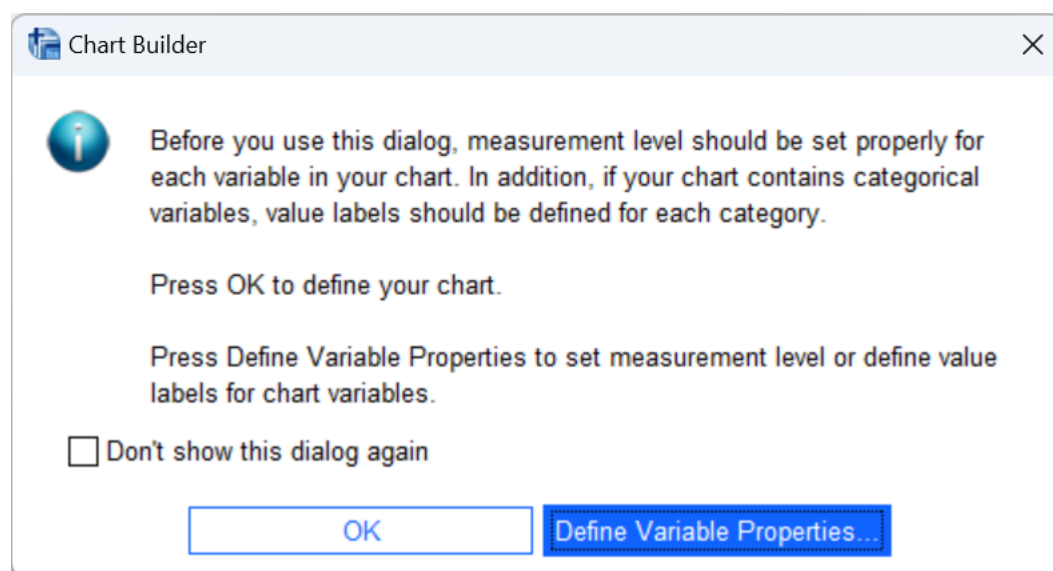


图 15 “Chart Builder”关于变量设置的提示信息

我们的数据集已经正确设置，因此，可点击“OK”继续，弹出如图 16 所示的对话框。可以看到：窗口的左上方是变量列表，右上方的白色空白区域是工作预览区，下方则是一系列选项卡，当然默认的“Gallery”选项卡中列出了各种统计图类型及缩略图。

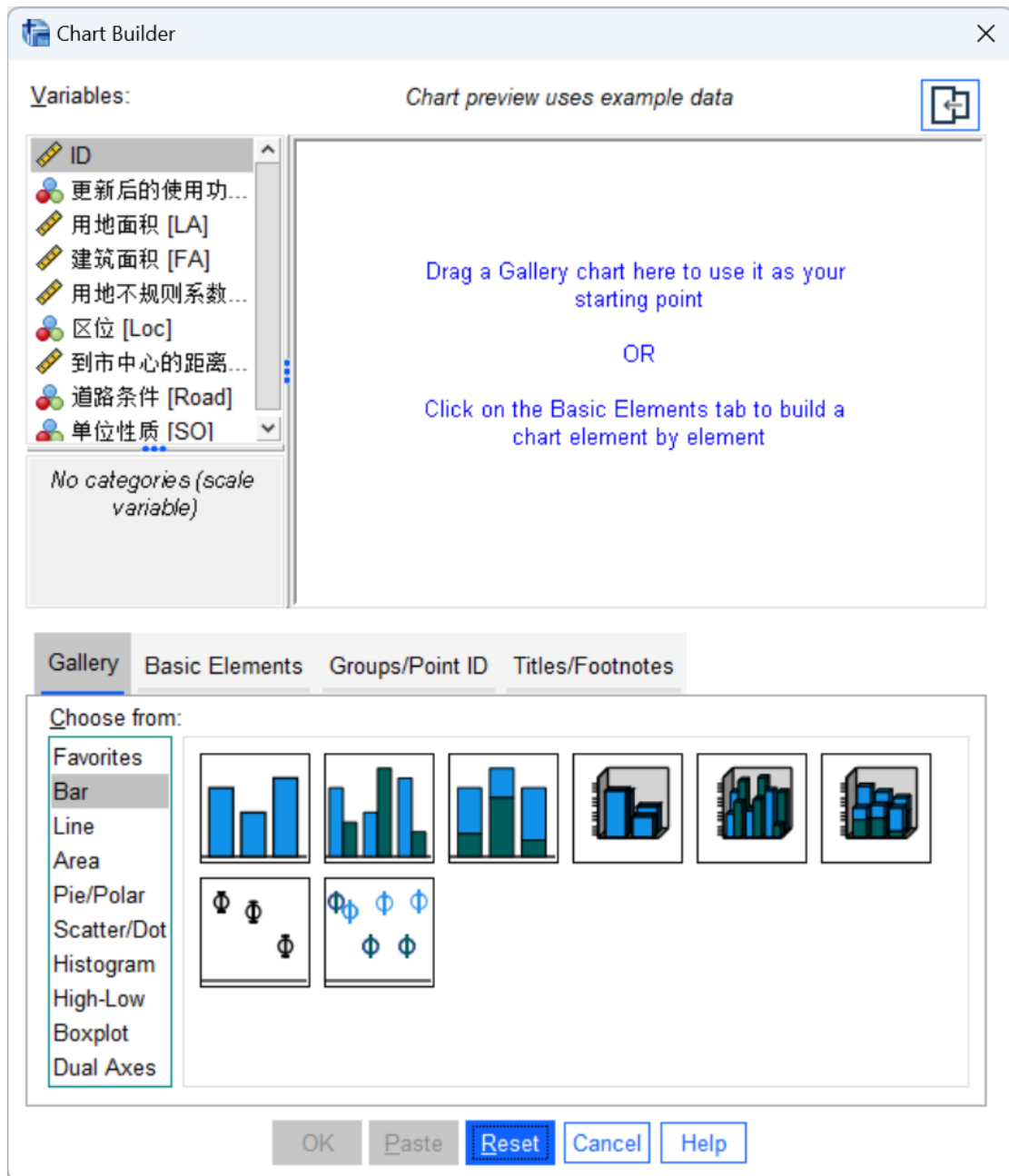


图 16 “Chart Builder” 的主窗口

按住鼠标左键，将“Bar”（柱状图）中第一个缩略图“Simple Bar”拖入工作预览区，如图 17 所示。预览区中的 X 轴、Y 轴带有问号（“X-Axis?”、“Y-Axis?”），等待我们进一步定义。

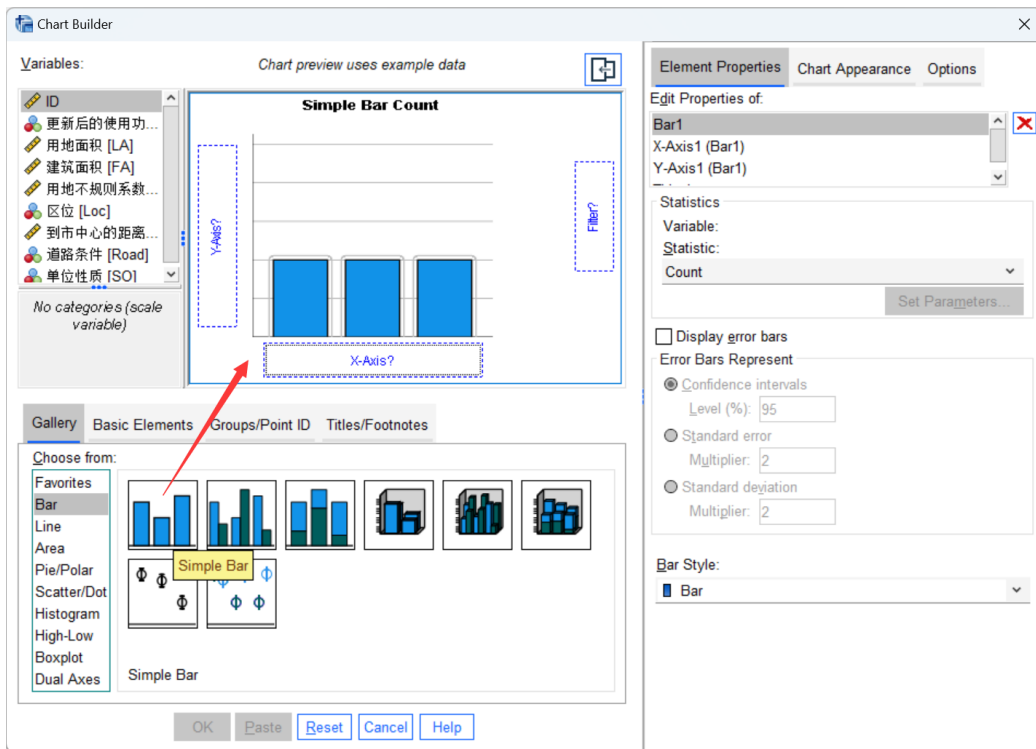


图 17 在“Chart Builder”中使用柱状图

按住鼠标左键，从左上部的“Variables”列表中把“更新后的使用功能 [NF]”拖到工作预览区的 X 轴（“X-Axis?”）上，如图 18 所示。由于该变量是一个分类变量，并且已经正确设置了测量方式和值标签，Y 轴自动设置为“Count”，即每一个类别的频数。

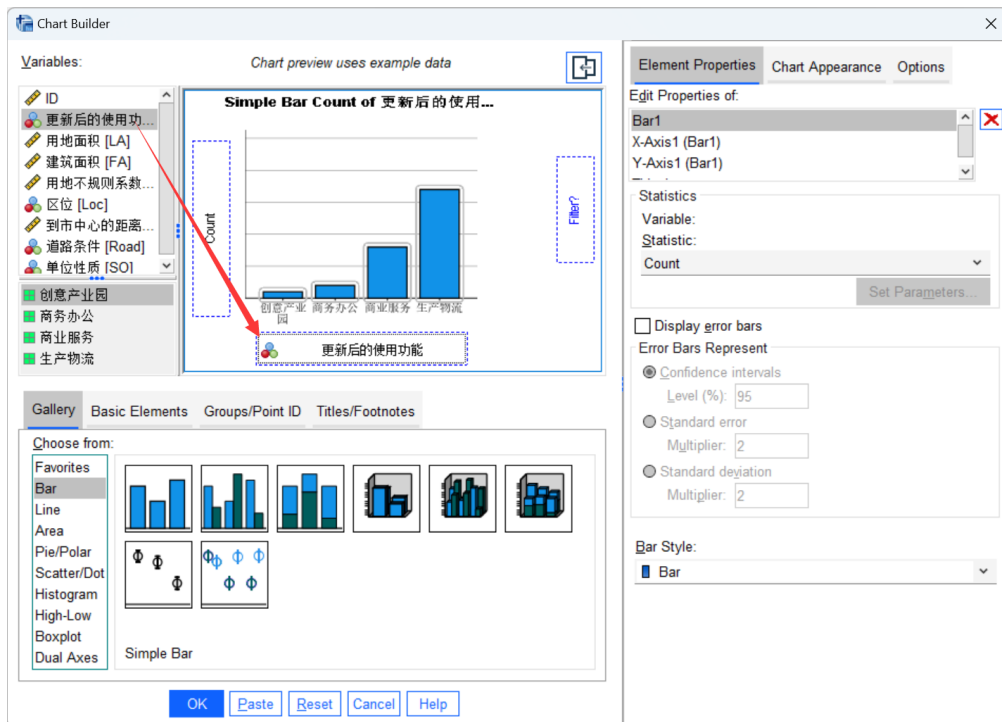


图 18 在“Chart Builder”中设置柱状图要分析的分类变量

至此，我们就完成了在 Chart Builder 中绘制柱状图的主要步骤。在此基础上，还可以进一步调整标题、坐标轴名称等内容，这里不再赘述。点击“OK”，SPSS 的输出窗口中将呈现刚才所绘制的柱状图，如图 19 所示。

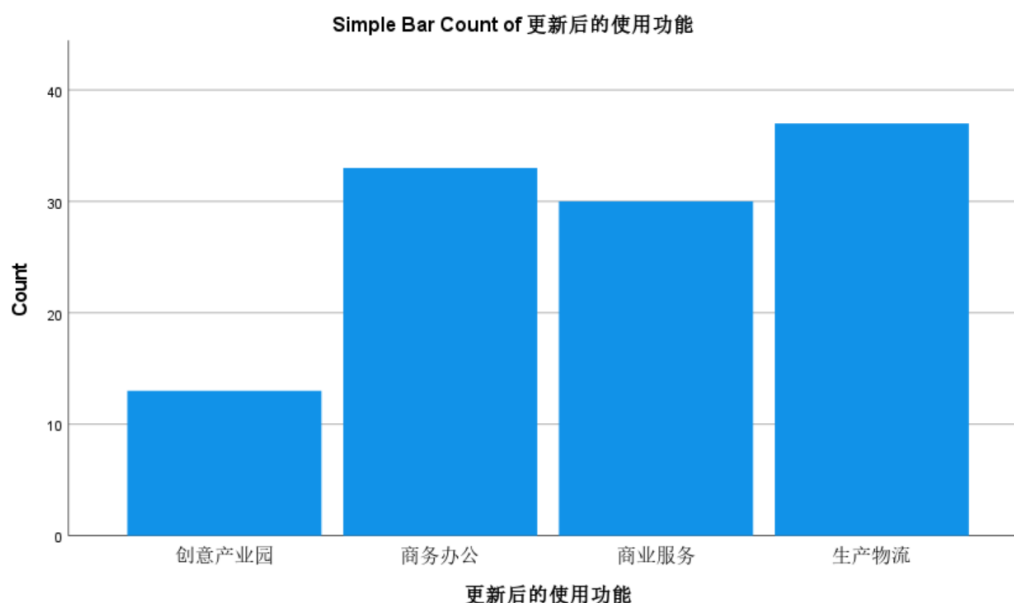


图 19 柱状图结果

许多情况下，我们可能希望对统计图的尺寸、字体、字号、颜色、文本内容、位置等做进一步的调整。对此，可以在输出窗口中双击统计图，弹出“Chart Editor”编辑窗口和“Properties”属性窗口，如图 20 所示。在这里，我们可以进行各种精细化的设置。

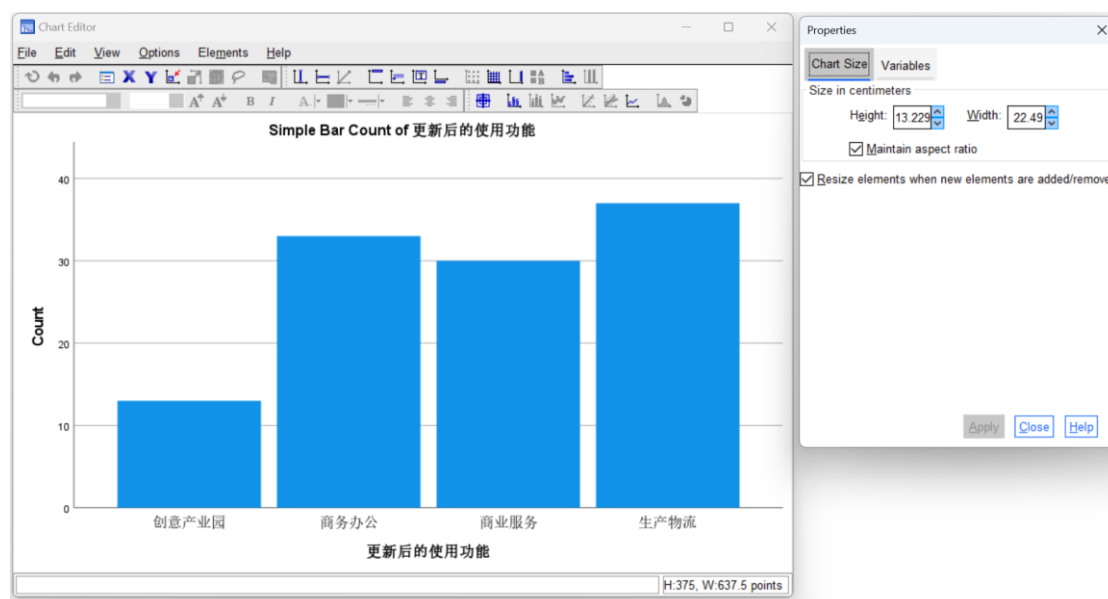


图 20 进入“Chart Editor”窗口

例如，如果希望调整 X 轴上类别标签的字体和字号，我们可以在任意一个类别标签上单击，此时的“Properties”属性窗口将显示可调整的内容。在“Font”中把“Family”改为我们想要的字体，把“Preferred Size”改为我们想要的字号，如图 21 所示。点击“Apply”应用修改，可以看到，“Chart Editor”窗口中的类别标签样式已经相应地发生变化。

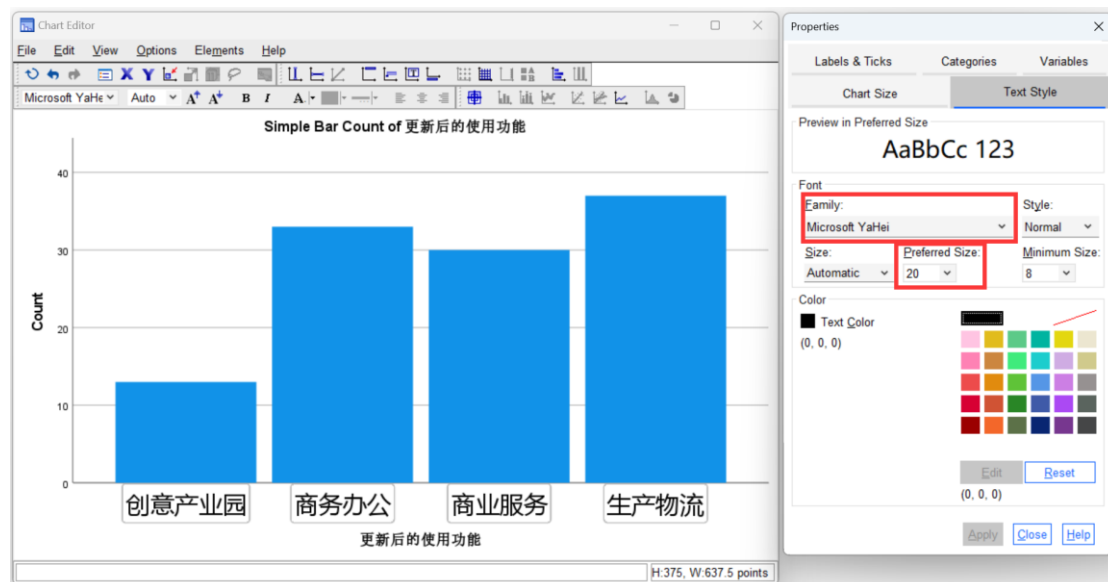


图 21 在“Chart Editor”窗口中调整字体和字号

如果觉得柱子的宽度太宽，可以在“Properties”窗口中切换至“Bar Options”选项卡，将“Widths”中的“Bars”宽度控制滑条向左拖动至满意位置，如图 22 所示。

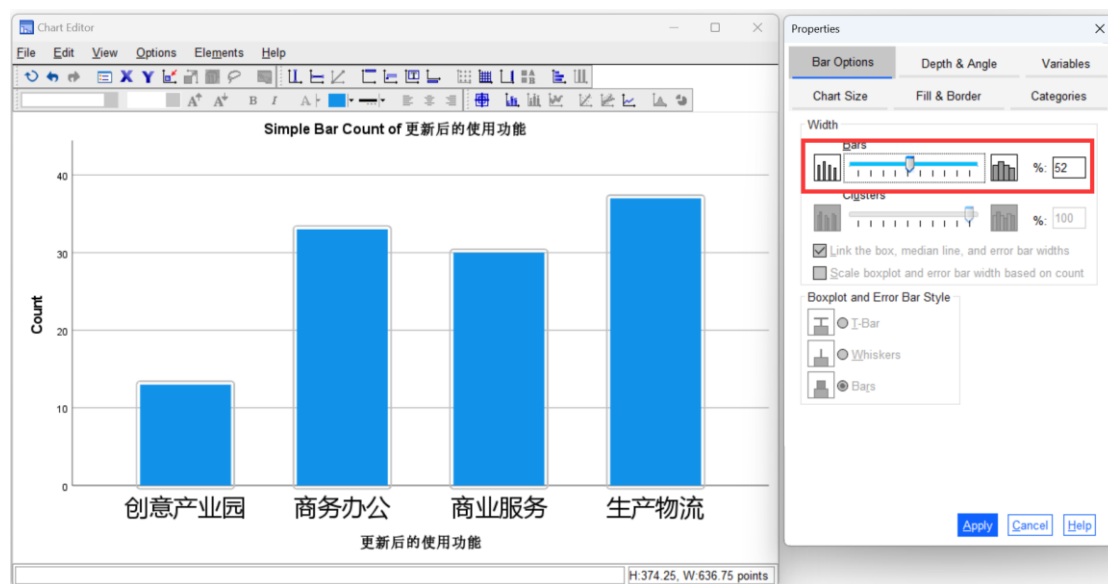


图 22 在“Chart Editor”窗口中调整柱状图内柱子的宽度

SPSS 的图形编辑功能是非常友好的。用户可以自己探索其他各种可调整的内容，不断提升软件操作技巧。在完成调整之后，关闭“Chart Editor”即可，输出窗口中的统计图将同步变化。

(2) 折线图

重新在菜单栏中选择“Graphs → Chart Builder”，打开图表构建器对话框。按照相同的操作逻辑绘制折线图：将“Gallery”中“Line”折线图系列的第一个缩略图“Simple Line”拖入工作预览区，再把 Variables 变量列表中的“更新后的使用功能 [NF]”拖入待设定的 X 轴（“X-Axis?”）上，如图 23 所示。

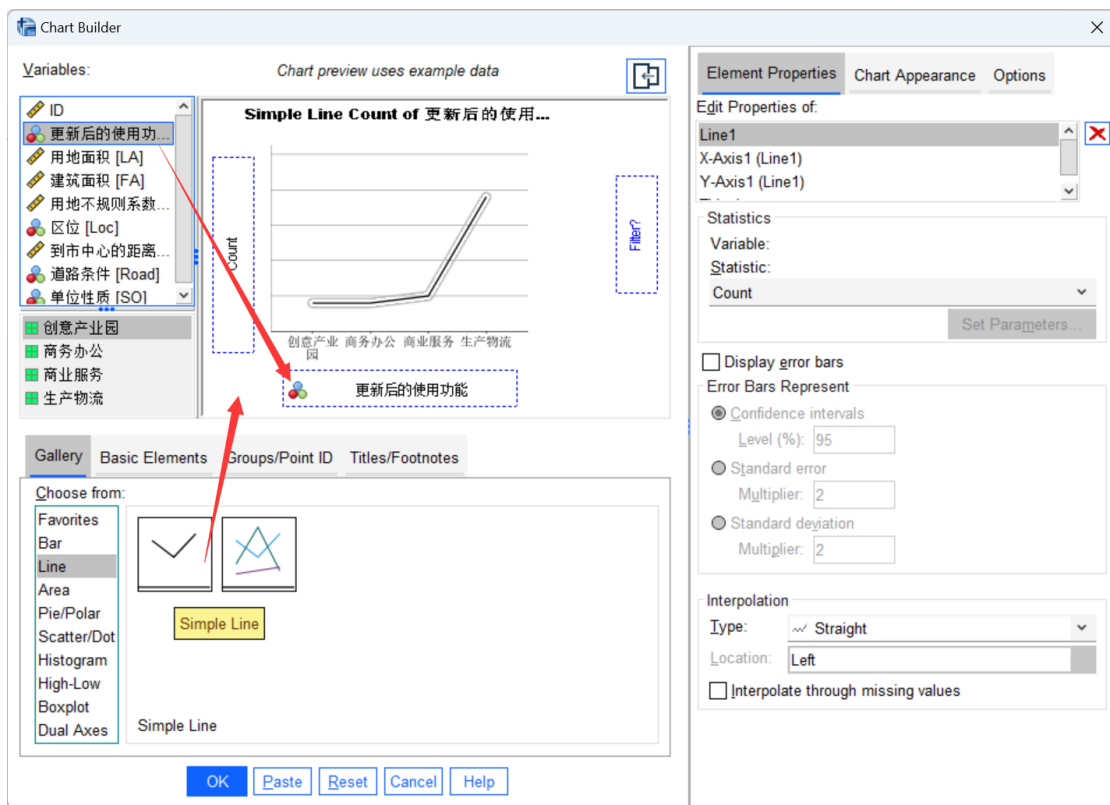


图 23 在“Chart Editor”窗口中设置折线图

点击“OK”，SPSS 的输出窗口将呈现所绘制的线状图，如图 24 所示。

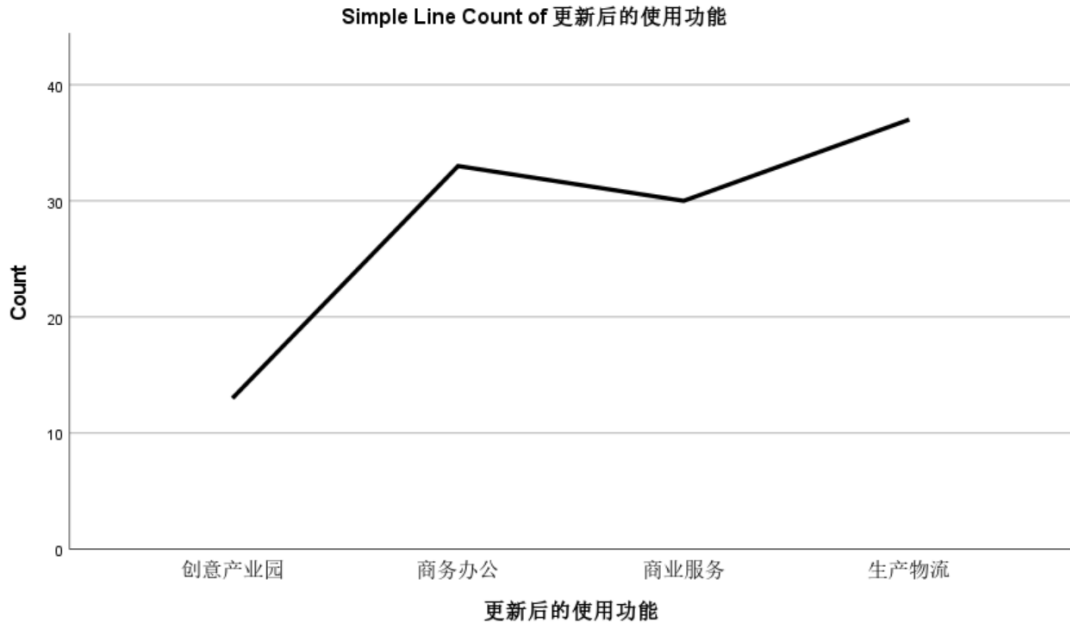


图 24 折线图结果

(3) 饼状图

在菜单栏中依次选择“Graphs → Legacy Dialogs → Pie”，以传统方式绘制饼状图。首先，SPSS 会弹出如图 25 所示的对话框，要求我们设定数据类型。本例中，我们要对“更新后的使用功能 [NF]”这一个变量的不同组别进行汇总统计，因此选择“Summaries for groups of cases”，这也是分类变量最为典型的情况。

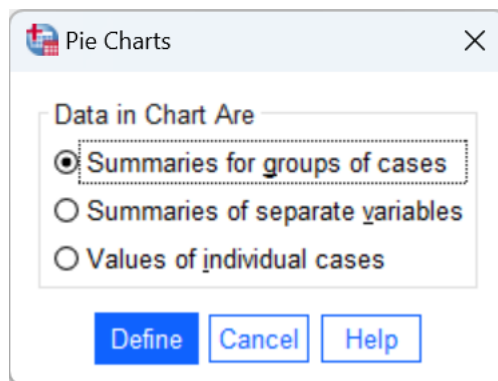


图 25 定义数据类型

点击“Define”后，进入如图 26 所示的主对话框。这里只需要将“更新后的使用功能 [NF]”选入“Define Slices by”即可。

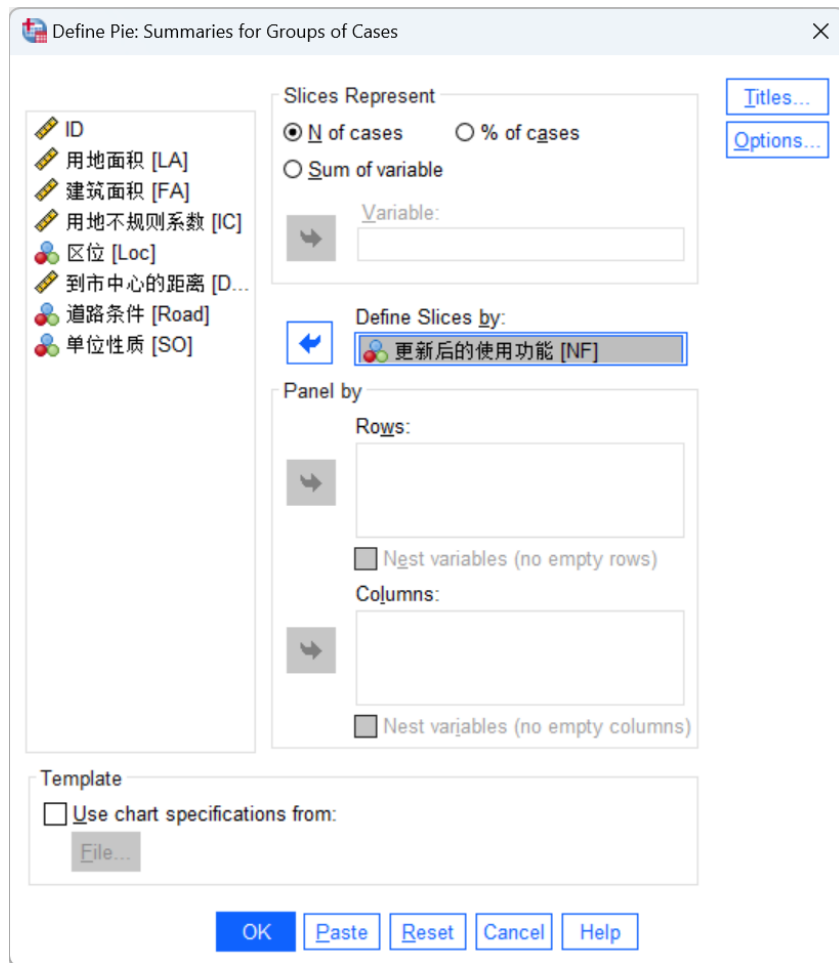


图 26 饼状图设定

点击“OK”，SPSS 的输出窗口即呈现饼状图，如图 27 所示。

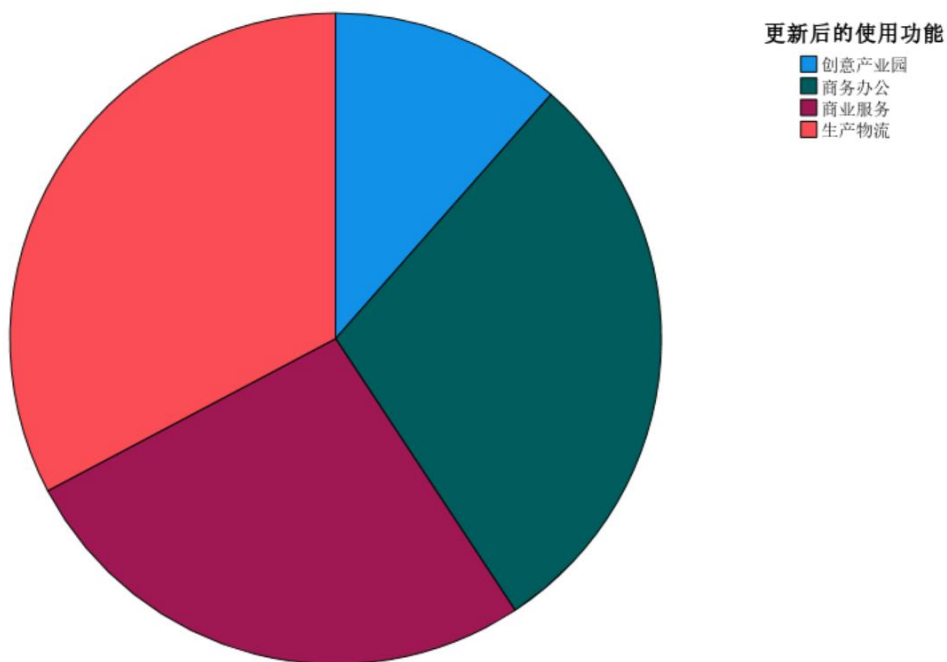


图 27 饼状图结果

如果希望在切片上显示每个类别的比例，可以双击该图进入“Chart Editor”窗口。选中任何一个切片，在“Chart Editor”的菜单栏中选择“Elements → Show Data Labels”。然后，在“Properties”对话框中选择“Data Value Labels”选项卡，在“Labels”面板中将默认显示的“Count”从“Displayed”中移除，并将“Percent”从下方的“Not Displayed”中上移至“Displayed”，如图 28 所示。最后点击“Apply”按钮即可。

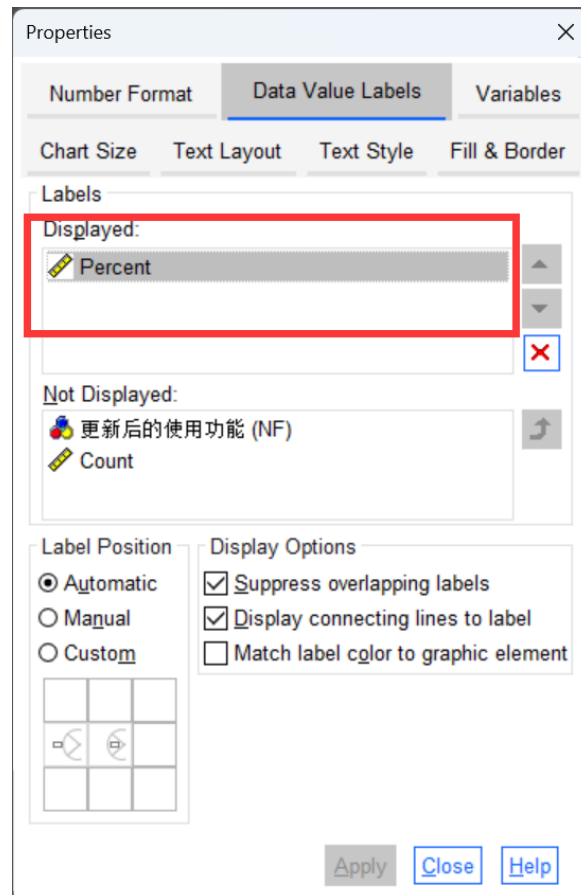


图 28 在饼状图中显示百分比标签

4.2 数值变量：直方图、箱形图

本小节以到市中心的距离为例，采用“Chart Builder”演示直方图和箱形图的绘制。此外，SPSS 中不能直接绘制核密度图，需要使用 Stata 等其他软件。

(1) 直方图

在菜单栏中选择“Graphs → Chart Builder”，打开图表构建器对话框。在下方的“Gallery”选项卡中选择“Histogram”（直方图），将第一个缩略图“Simple

Histogram”拖拽至工作预览区，然后再将左侧“Variables”列表中的“到市中心的距离 [Dist]”变量拖拽至等待定义的 X 轴上（“X-Axis?”），如图 29 所示。

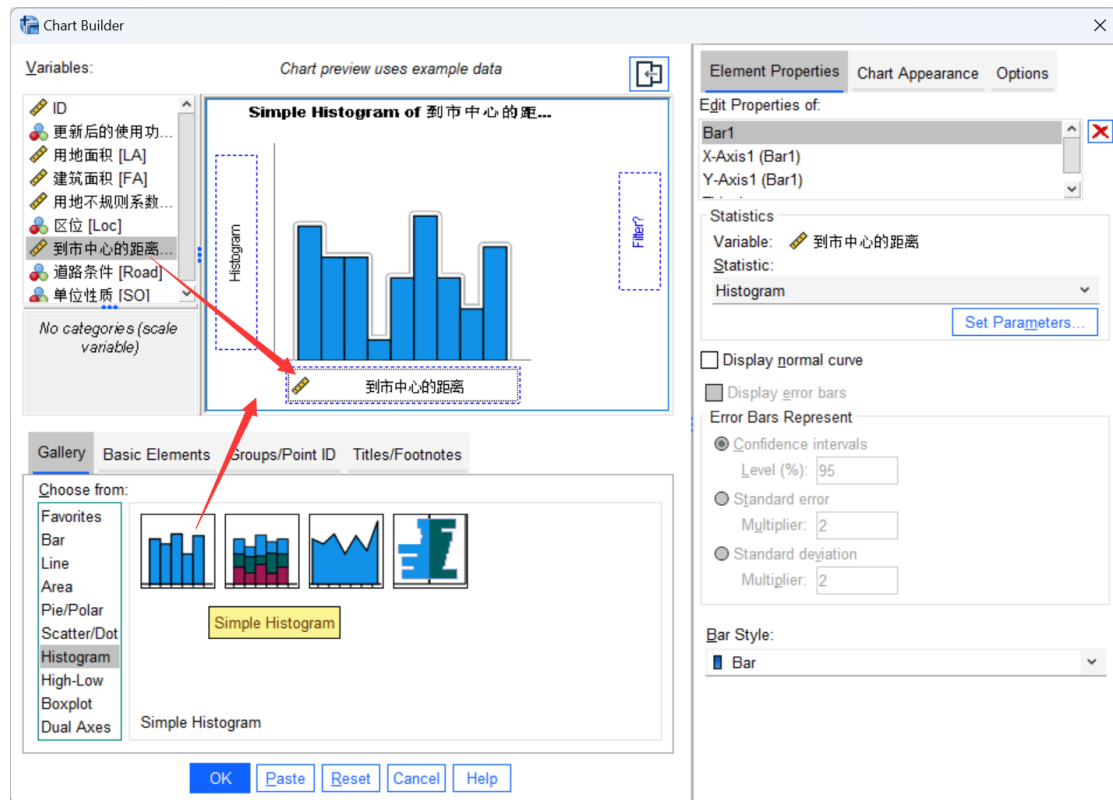


图 29 在“Chart Builder”窗口设置直方图

点击“OK”，SPSS 的输出窗口将呈现如图 30 所示的直方图结果。

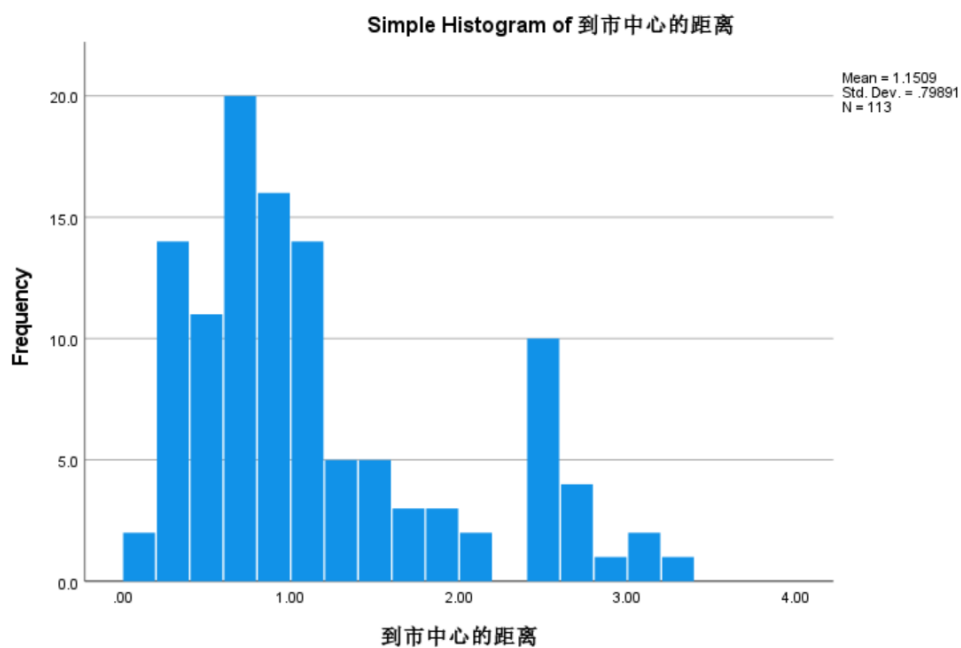


图 30 直方图结果

如果希望调整直方图中区间的划分方式,可以双击直方图,进入“Chart Editor”窗口。点击任意一个柱子,然后即可在“Properties”窗口的“Binning”选项卡中进行调整,如下图所示。这里将“X-Axis”的区间划分由自动(“Automatic”)转变为自定义(“Custom”),然后在“Number of intervals”中填入 50,即要求划分 50 个区间,如图 31 所示。点击“Apply”后即可看到,此时的直方图区间更加细致。

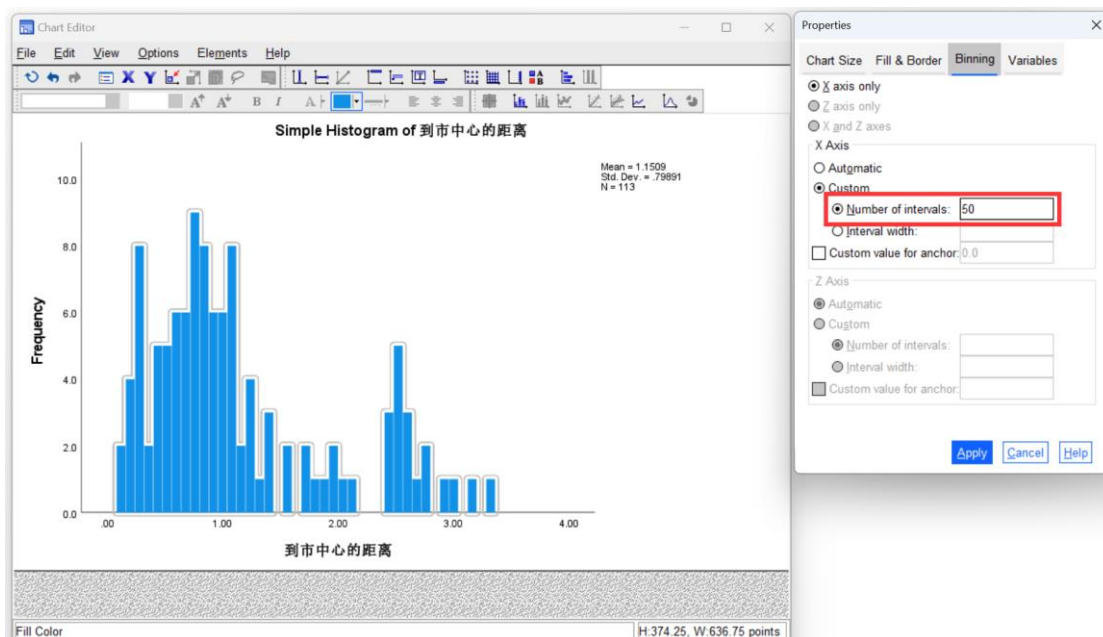


图 31 在“Chart Editor”中设置直方图的区间划分方式

(2) 箱形图

在菜单栏中选择“Graphs → Chart Builder”,打开图表构建器对话框。在下方的“Gallery”选项卡中选择“Boxplot”(箱形图),将第一个缩略图“Simple Boxplot”拖拽至工作预览区。与其他统计图不同的是,箱形图中的分析对象本身显示在 Y 轴(而非 X 轴)上,因此,我们需要将“到市中心的距离 [Dist]”拖拽至等待定义的 Y 轴(“Y-Axis?”)上,如图 32 所示。

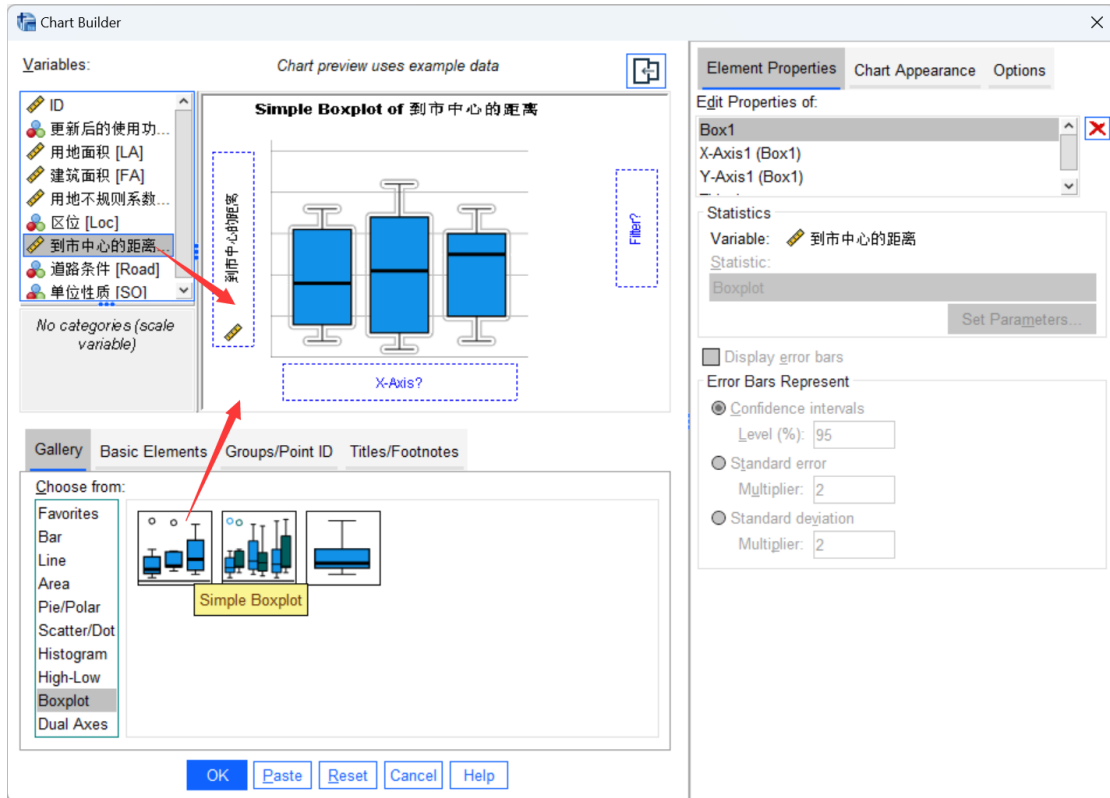


图 32 在“Chart Builder”中设置箱形图

点击“OK”运行分析，结果呈现在 SPSS 的输出窗口，如图 33 所示。图中离群点上所标注的数字是样本编号，如不需要，可以进入“Chart Editor”删除。

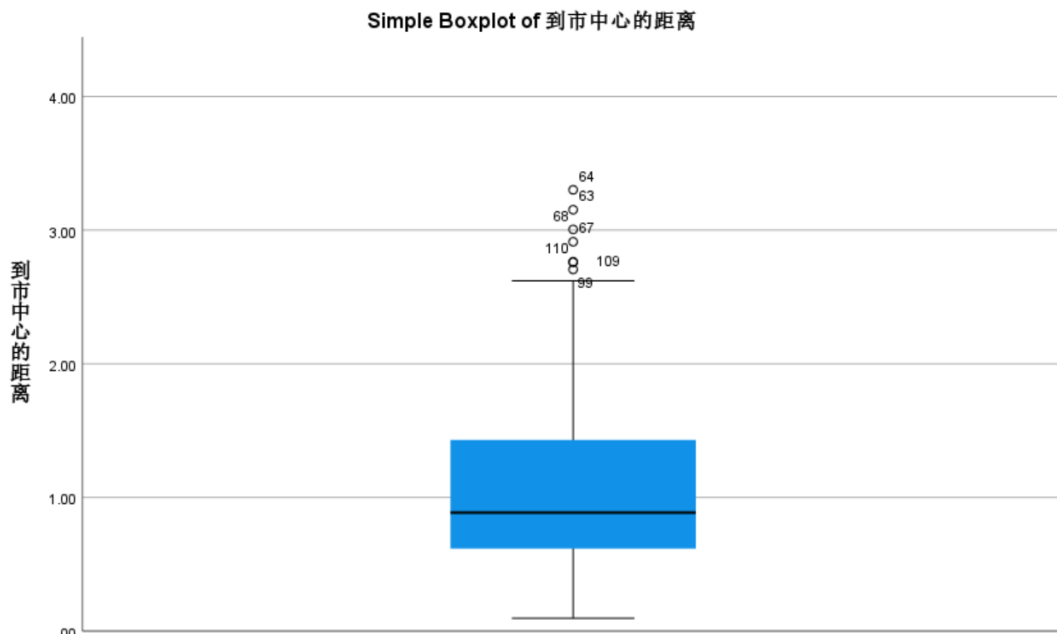


图 33 箱形图结果

箱形图的 X 轴具有什么意义呢？如课上所讲，箱形图的一大功能是用于多组之间的比较，因此，X 轴常常定义为一个类别变量。我们再次打开“Chart

Builder”，此时的界面应该保持记忆，还保留上一个箱形图的相关设定。在此基础上，我们把“更新后的使用功能 [NF]”拖拽至等待定义的 X 轴上(“X-Axis?”)，如图 34 所示。

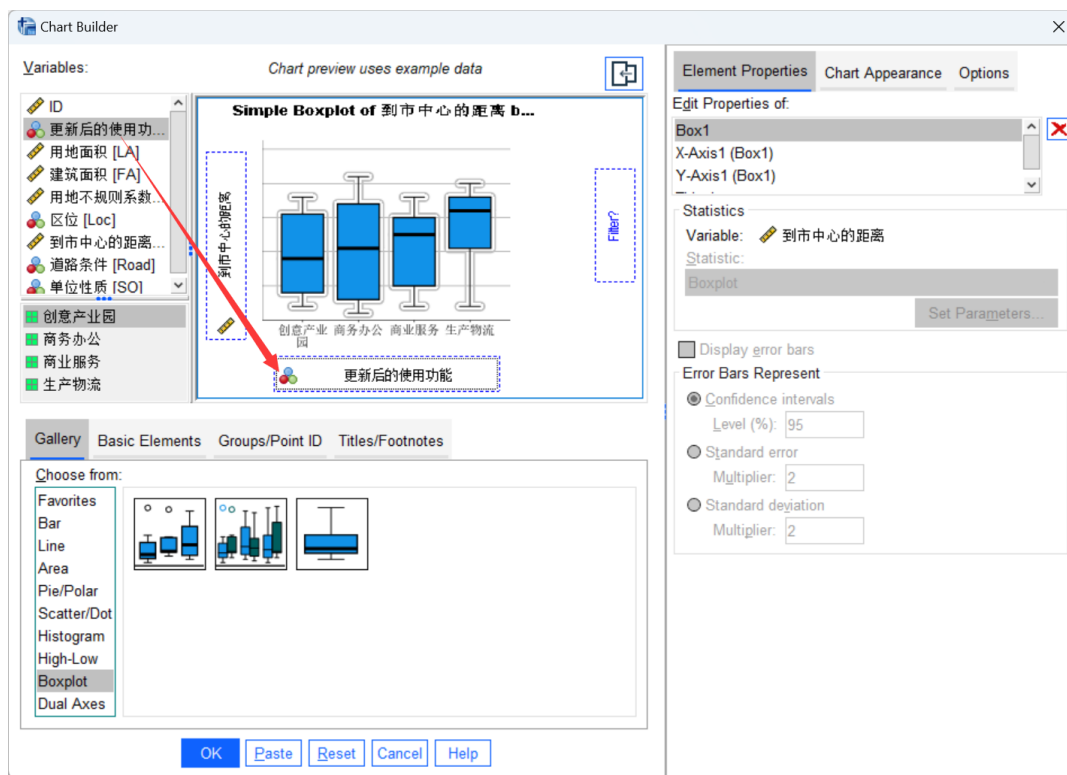


图 34 在“Chart Builder”窗口中为箱形图设置不同的组别

点击“OK”按钮运行分析，结果如图 35 所示。由此，我们可以比较更新为不同功能的用地在距离变量上的分布差异。

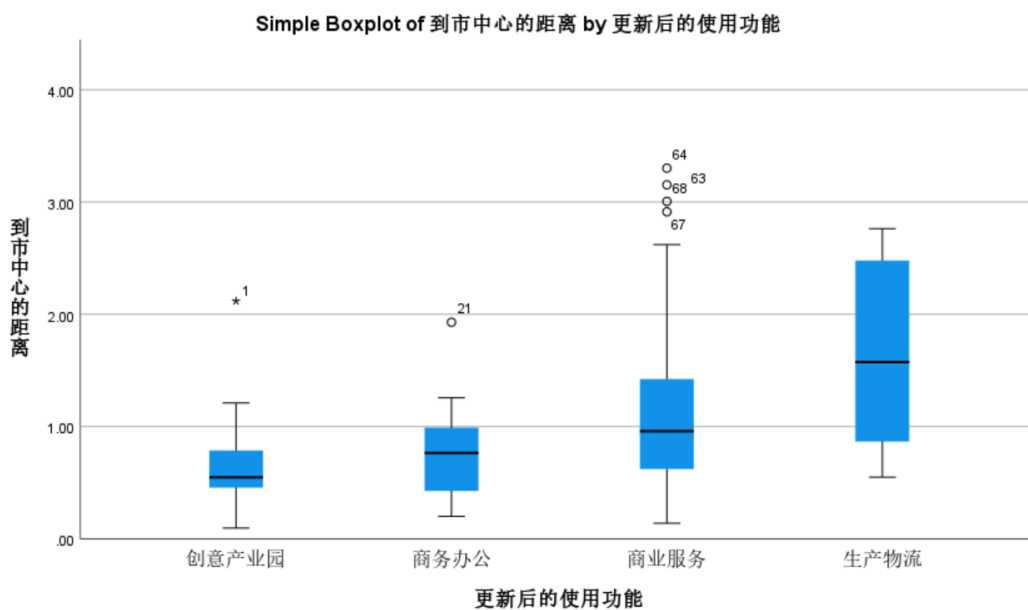


图 35 不同的组别的箱形图结果