

## 1. 概述

本节以上海市工业用地更新数据为例，利用 SPSS 软件演示两变量之间的相关性分析方法，包括分类变量与数值变量的方差分析、分类变量与分类变量的列联表分析、数值变量与数值变量的相关分析等。SPSS 的数据文件为“correlation.sav”，同时还提供了同名的 csv 文件。

## 2. 分类变量与数值变量的相关性

更新后的功能是一个分类变量，而现状容积率是一个数值变量。本小节即以这 2 个变量为例，演示针对“分类变量 vs. 数值变量”的方差分析等方法的操作过程。

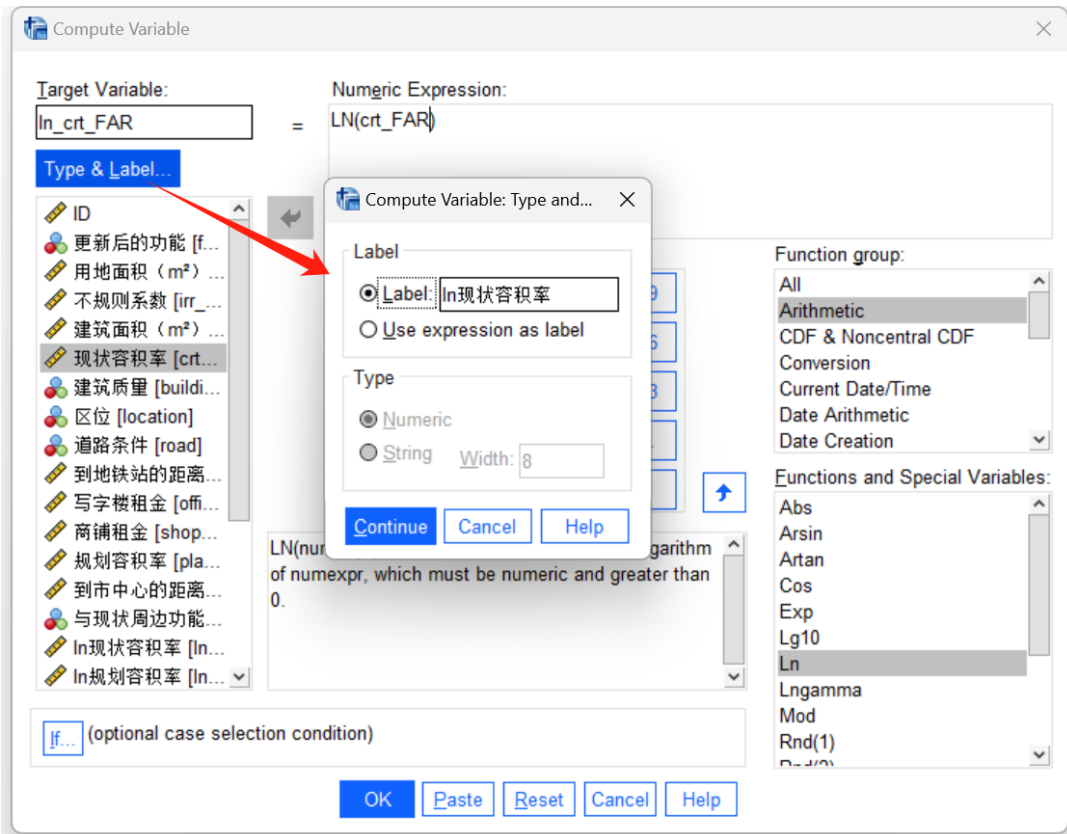


图 1 对现状容积率进行对数变换

为了更好的满足方差分析中正态性、方差齐性等条件，我们对现状容积率进行对数变化。在菜单栏中选择“Transform → Compute Variable”，打开计算变量对话框。在“Target Variable”中输入“ln crt\_FAR”，作为新变量的变量名，在“Numeric

Expression”中输入“LN crt\_FAR)”，作为计算新变量的表达式，如图 1 所示。为了保证输入正确，像“LN”这样的函数可以从右侧的“Function group → Arithmetic → Ln”中选择，像“crt\_FAR”这样的已有变量名可以从左侧的变量列表中选择。此外，我们还可以点击“Type & Label”按钮，为新变量设置类型和标签，如图 1 所示。这里设置的标签为“ln 现状容积率”。

## 2.1 方差分析

在菜单栏中依次选择“Analyze → Compare Means → One-Way ANOVA”，打开单因素方差分析对话框<sup>1</sup>。将“ln 现状容积率 [ln crt\_FAR]”选入“Dependent List”，将“更新后的功能 [function]”选入“Factor”，如图 2 所示。

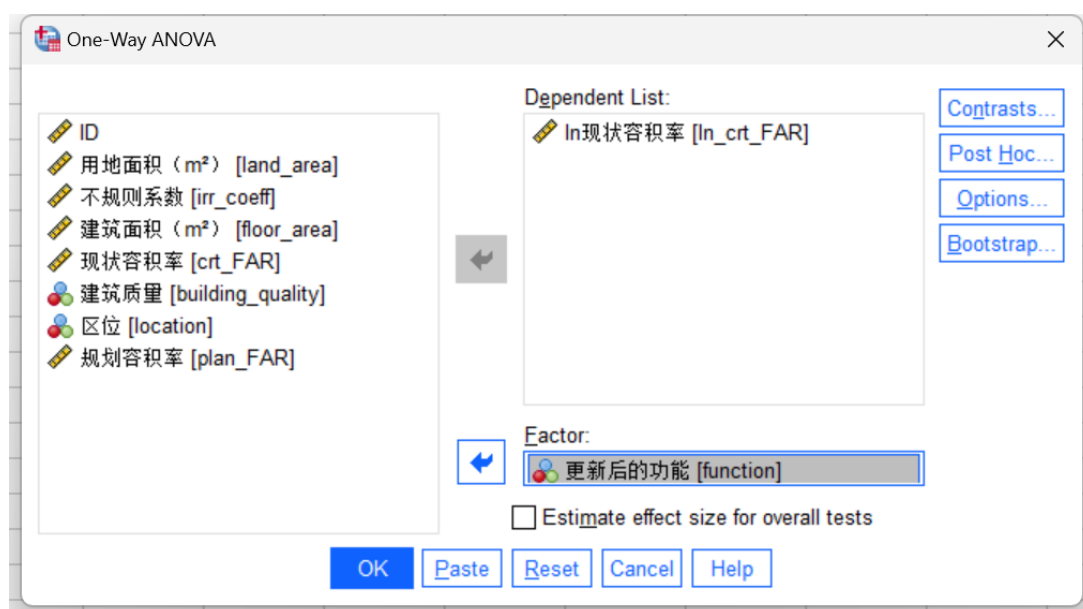


图 2 单因素方差分析的主对话框设置

点击“Options”按钮，进入方差分析的选项子对话框。在“Statistics”面板中勾选“Homogeneity of variance test”，要求执行方差齐性检验；同时勾选下方的“Means plot”，要求绘制各组的均值比较图，如图 3 所示。

<sup>1</sup> 这里的单因素是指只有一个分类变量。我们在课程上提到，方差分析有更一般的形式，可以同时考虑多个分类变量对一个数值变量的影响，即多因素方差分析，其归属于一般线性模型，在 SPSS 中可以通过“Analyze → General Linear Model → Univariate”实现。我们在这里只关注更基础的单因素方差分析。

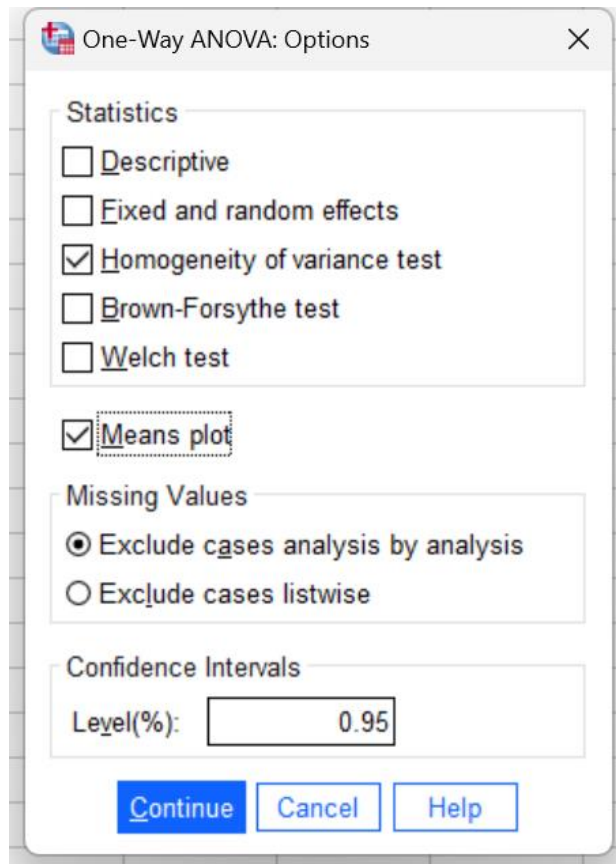


图 3 单因素方差分析的选项子对话框设置

点击“Continue”回到主对话框，再点击“OK”进行分析。SPSS 的输出窗口首先呈现了方差齐性检验的结果，如图 4 所示。表格中的各行是按不同的方法进行检验；各列中，“Levene Statistic”是检验统计量，“df1”与“df2”是检验所对应的 2 个自由度，“Sig.”是我们主要关注的提示显著性的 p 值。

#### Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
In现状容积率	Based on Mean	1.391	8	136	.206
	Based on Median	1.188	8	136	.311
	Based on Median and with adjusted df	1.188	8	128.409	.311
	Based on trimmed mean	1.359	8	136	.220

图 4 方差齐性检验结果

接下来，SPSS 输出了方差分析最主要的结果——方差分析表，如图 5 所示。表头的“ANOVA”即“Analysis of Variance”的意思；各行中，“Between Groups”、“Within Groups”和“Total”分别表示组间变异、组内变异、总变异；各列中，“Sum

of Squares”是以平方和形式表示的变异总量，“df”是对应的自由度，“Mean Square”是用平方和除以自由度得到的均方值，F 是方差分析的 F 统计量，由组间均方除以组内均方计算得到，“Sig.”是我们主要关注的 p 值。

**ANOVA**

In现状容积率

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.196	8	.650	3.644	<.001
Within Groups	24.242	136	.178		
Total	29.438	144			

图 5 方差分析表

最后，SPSS 呈现了各组均值比较折线图，如图 6。该图的横坐标是组变量——“更新后的功能”，纵坐标是每一组样本的平均“ln 现状容积率”。当方差分析的 F 检验显著时，我们应当能在该图中观察到不同组之间的明显差异。

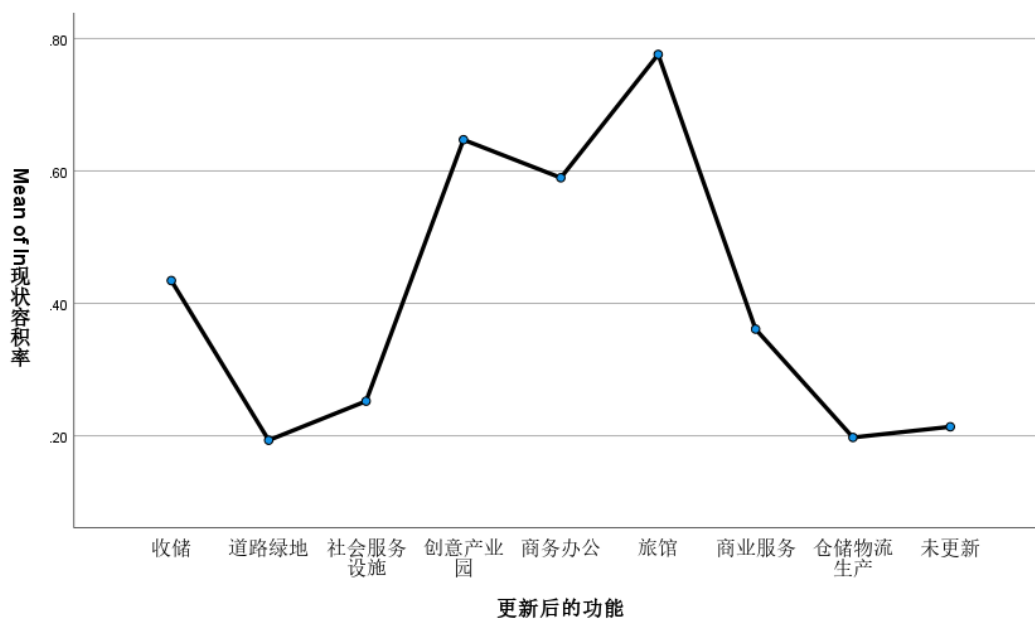


图 6 方差分析中的各组均值比较图

为了进一步确认哪些组之间的差异是显著的，我们需要执行事后多重比较。该步骤只有当方差分析的 F 检验显著时才有意义。再次在菜单栏中选择“Analyze → Compare Means → One-Way ANOVA”，重新打开方差分析对话框。保持原来的设置不变，点击“Post Hoc”按钮，进入多重比较子对话框，如图 7 所示。可以看到，该对话框提供了一系列事后多重比较的具体方法，归属于方差齐性条件得

到满足(“Equal Variances Assumed”)和未得到满足(“Equal Variances Not Assumed”)两类情况。由于本例通过了方差齐性检验(图 4), 因此我们可以选择“Equal Variances Assumed”中的“LSD”方法, 这是一种较为敏感的检验方法。

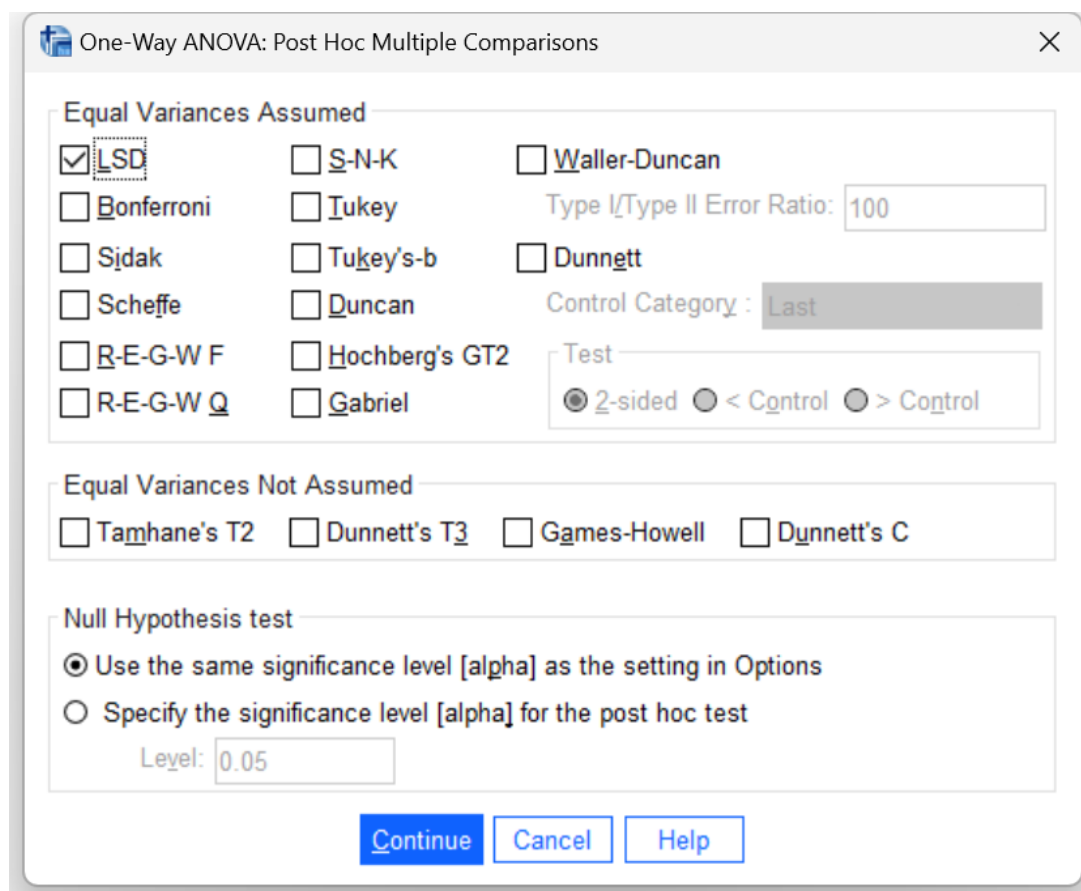


图 7 方差分析的事后多重比较对话框设置

点击“Continue”回到主对话框, 再点击“OK”执行分析。SPSS 除了输出之前所提到的内容以外, 还输出了 LSD 事后多重比较的结果, 如图 8 所示。表中的前两列分别为“(I)更新后的功能”和“(J)更新后的功能”, 表格中的每一行即是对(I)功能和(J)功能进行比较的结果, 例如, 红框高亮显示的即是更新为“收储”功能与更新为“旅馆”功能的样本的比较。“Mean Difference (I-J)”表示(I)功能样本与(J)功能样本相比, 在“ln 现状容积率”上的均值差异, 例如, 红框高亮显示的-0.34159表示, 收储用地的平均 ln 现状容积率比旅馆用地低 0.34159; “Std. Error”是这一差异的标准误; “Sig.”是判断这一差异是否具有统计显著性的 p 值; “95% Confidence Interval”是这一差异的 95%置信区间。

## Post Hoc Tests

### Multiple Comparisons

Dependent Variable: ln现状容积率  
LSD

(I) 更新后的功能	(J) 更新后的功能	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
收储	道路绿地	.24117	.27253	.378	-.2978	.7801
	社会服务设施	.18214	.32246	.573	-.4555	.8198
	创意产业园	-.21277	.16901	.210	-.5470	.1215
	商务办公	-.15533	.14927	.300	-.4505	.1399
	旅馆	-.34159	.18617	.069	-.7098	.0266
	商业服务	.07328	.14421	.612	-.2119	.3585
	仓储物流生产	.23693	.14026	.093	-.0404	.5143
	未更新	.22082	.16352	.179	-.1025	.5442
道路绿地	收储	-.24117	.27253	.378	-.7801	.2978
	社会服务设施	-.05903	.38541	.878	-.8212	.7031
	创意产业园	-.45393	.27042	.096	-.9887	.0808
	商务办公	-.39649	.25854	.127	-.9078	.1148
	旅馆	-.58275 <sup>a</sup>	.28146	.040	-1.1394	-.0261
	商业服务	-.16788	.25565	.512	-.6735	.3377
	仓储物流生产	-.00424	.25344	.987	-.5054	.4970
	未更新	-.02035	.26702	.939	-.5484	.5077

图 8 LSD 事后多重比较结果（部分）

## 2.2 Eta 的计算

Eta 是测度一个数值变量和一个分类变量之间相关性强度的指标，然而计算 Eta 的功能并不在方差分析中，而在 3.1 节将要介绍的列联表分析中。在菜单栏中依次选择“Analyze → Descriptive Statistics → Crosstabs”，打开列联表分析对话框。将“ln 现状容积率 [ln crt FAR]”选入“Row(s)”，把“更新后的功能 [function]”选入“Column(s)”，如图 9 所示<sup>2</sup>。然后，点击“Statistics”按钮，进入统计量子对话框。在“Nominal by Interval”中选择中“Eta”，如图 10 所示。点击“Continue”回到主对话框，将左下角的“Suppress tables”勾选，要求不要输出列联表。这是因为我们只是借用列联表分析中的子功能计算 Eta 指标，一旦输出列联表，数值变量中许多不重复的值会占据很大的篇幅。设置完成的主对话框如图 9 所示。

<sup>2</sup> 这里对于行与列的设定是随意的，我们也可以反过来把“ln 现状容积率 [ln crt FAR]”选入“Column(s)”，把“更新后的功能 [function]”选入“Row(s)”。



图 9 计算 Eta: 列联表分析主对话框的设定

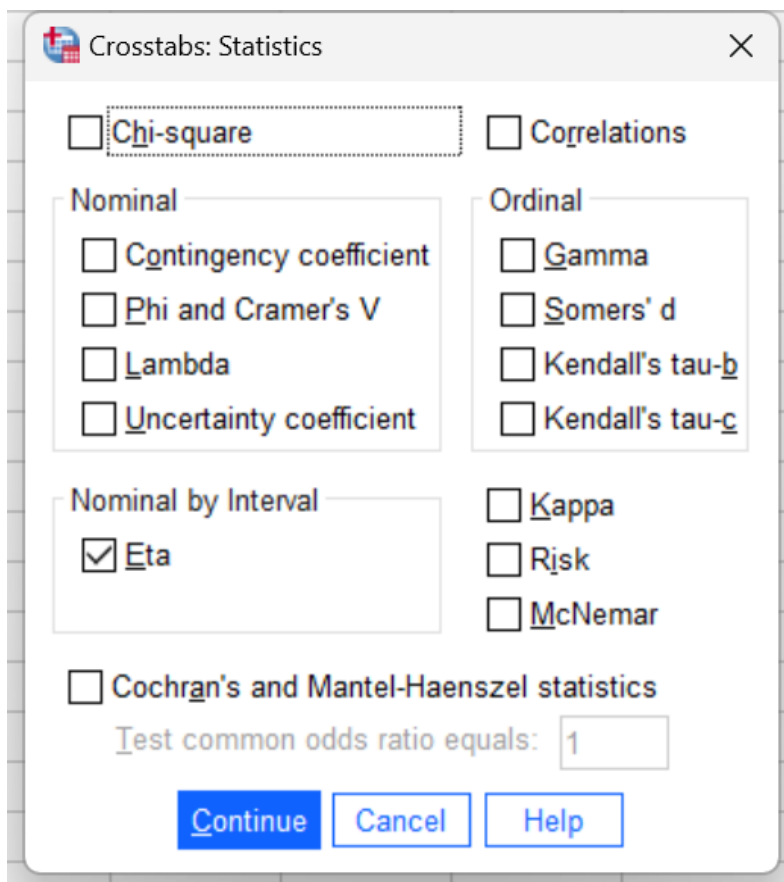


图 10 计算 Eta: 列联表分析统计量子对话框的设定

点击“OK”执行计算。SPSS 会输出如图 11 所示的指标计算结果。表头中的“Directional Measures”表示这是一个具有方向性的指标。表格中的两行——“更新后的功能 Dependent”和“ln 现状容积率 Dependent”分别把分类变量和数值变量作为被解释的角色。可以看到，当该角色变换时，Eta 值有很大的差异。在绝大多数情况下，我们把数值变量（本例中的“ln 现状容积率”）放在被解释的角色，用分类变量“更新后的功能”解释其在各类之间的差异。因此，本例的 Eta 值应当为 0.420。

Directional Measures			Value
Nominal by Interval	Eta	更新后的功能 Dependent	.984
		ln 现状容积率 Dependent	.420

图 11 Eta 计算结果

最后指出，Eta 指标的计算同样是以方差分析的 F 检验显著为前提的。如果检验不显著，则 Eta 指标将失去统计推断的意义。

## 2.3 非参数检验

如果方差分析的正态性和方差齐性条件未能满足且有较大的偏差，我们可以改用非参数检验进行替代。在菜单栏中依次选择“Analyze → Nonparametric Tests → Legacy Dialogs → K Independent Samples”，打开 K 组独立样本非参数检验对话框。将“现状容积率 [crt\_FAR]”选入“Test Variable List”（非参数检验没有必要再使用对数变换后的数值变量），将“更新后的功能 [function]”选入“Grouping Variables”，此时该变量呈现的状态是“function(?)”，即组别编码的范围未指定。我们点击“Define Range”，然后在“Minimum”中输入“1”，在“Maximum”中输入“9”，点击“Continue”回到主对话框，此时的组别变量呈现的状态变为“function(1 9)”，即更新后的功能的类别编码介于 1~9 之间的样本纳入分析<sup>3</sup>。确认“Test Type”中的“Kruskal-Wallis H”被选中，即要求执行 Kruskal-Wallis 秩检验，如图 12 所示。

<sup>3</sup> 这里设置 1~9，是因为“更新后的功能 [function]”这个变量的分类编码最小值为 1，最大值为 9。如果我们希望纳入所有分组，又不清楚实际的最小、最大值是多少，也可以简单地使用一些显而易见的低值和高值，例如 0~100。

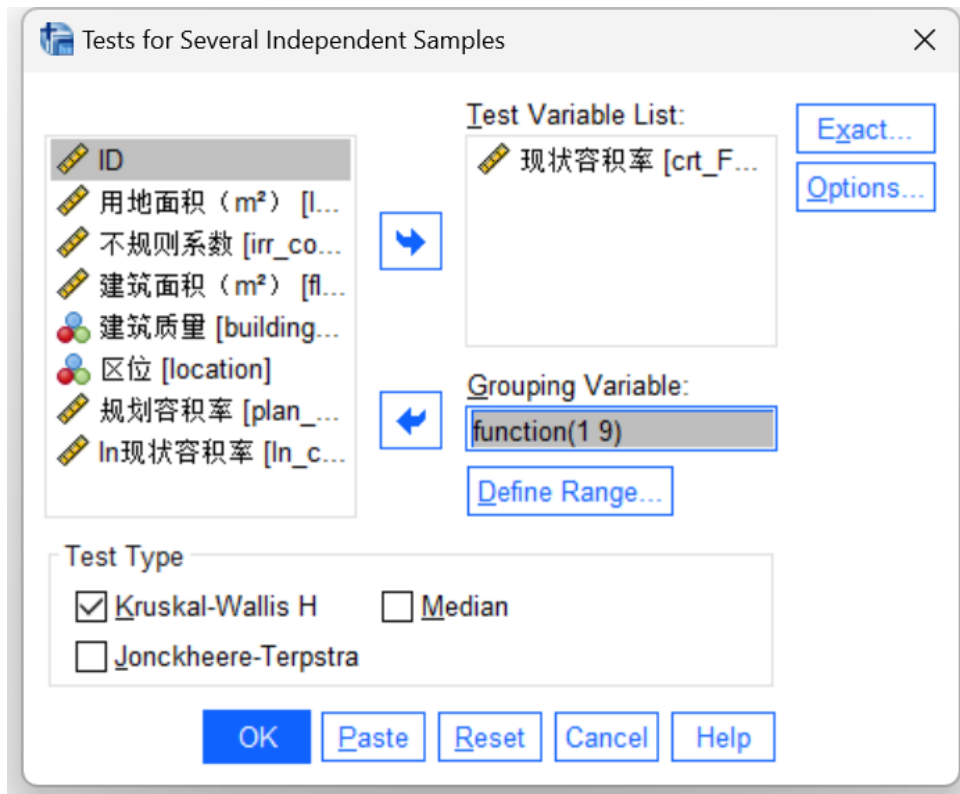


图 12 K 组独立样本的 Kruskal-Wallis 非参数检验设置

点击“OK”执行分析，SPSS 的输出窗口中首先报告了每一组的样本量（“N”）和平均秩（“Mean Rank”），如图 13 所示。

Ranks			
	更新后的功能	N	Mean Rank
现状容积率	收储	12	78.79
	道路绿地	3	52.67
	社会服务设施	2	55.00
	创意产业园	13	91.23
	商务办公	24	91.33
	旅馆	9	110.22
	商业服务	30	68.57
	仓储物流生产	37	54.47
	未更新	15	61.93
	Total		145

图 13 Kruskal-Wallis 非参数检验的基本统计量：样本量和平均秩

然后，SPSS 报告了该检验的检验结果，如图 14 所示。表中，“Kruskal-Wallis H”是检验统计量，“df”是自由度，“Asymp. Sig.”是我们主要关注的 p 值。

**Test Statistics<sup>a,b</sup>**

现状容积率

Kruskal-Wallis H	23.963
df	8
Asymp. Sig.	.002

a. Kruskal Wallis Test

b. Grouping Variable: 更新后的功能

图 14 Kruskal-Wallis 非参数检验结果

### 3. 分类变量与分类变量的相关性

更新后的功能和建筑质量均为分类变量。本小节即以这 2 个变量为例，演示针对“分类变量 vs. 分类变量”的列联表分析、对应分析等内容的操作过程。

#### 3.1 列联表分析与卡方检验

在菜单栏中依次选择“Analyze → Descriptive Statistics → Crosstabs”，打开列联表分析对话框。将“更新后的功能 [function]”选入“Row(s)”，将“建筑质量 [building\_quality]”选入“Column(s)”，确认左下角的“Suppress tables”未被选中，如图 15 所示。



图 15 列联表分析主对话框的设置

点击“OK”进行分析，SPSS 的结果窗口首先将呈现样本的汇总情况，即有效值、缺失值的数量和占比，我们略过这些基本信息。然后，SPSS 将报告列联表的表格，亦称为交叉表，如图 16 所示。

### 更新后的功能 \* 建筑质量 Crosstabulation

Count		建筑质量			Total
		差	好	中	
更新后的功能	社会服务设施	0	2	0	2
	创意产业园	0	13	0	13
	商务办公	3	8	13	24
	旅馆	1	5	3	9
	商业服务	20	1	9	30
	仓储物流生产	27	1	9	37
	未更新	9	1	5	15
Total		60	31	39	130

图 16 列联表表格：实际频数

在默认状态下，该表格的单元格报告的是实际频数。例如，图 16 红框中的“13”表示：实际共有 13 个样本更新后的功能为创意产业园，而且建筑质量为好。我们也可以在列联表中显示其他信息。再次进行列联表分析主对话框，点击“Cell”按钮，弹出单元格设置子对话框。可以看到，“Counts”下的“Observed”被默认选中。我们勾选“Expected”，要求报告每个单元格的期望频数；同时勾选“Percentages”下的“Row”，要求报告每个单元格的频数占所在行总频数的百分比，如图 17 所示。

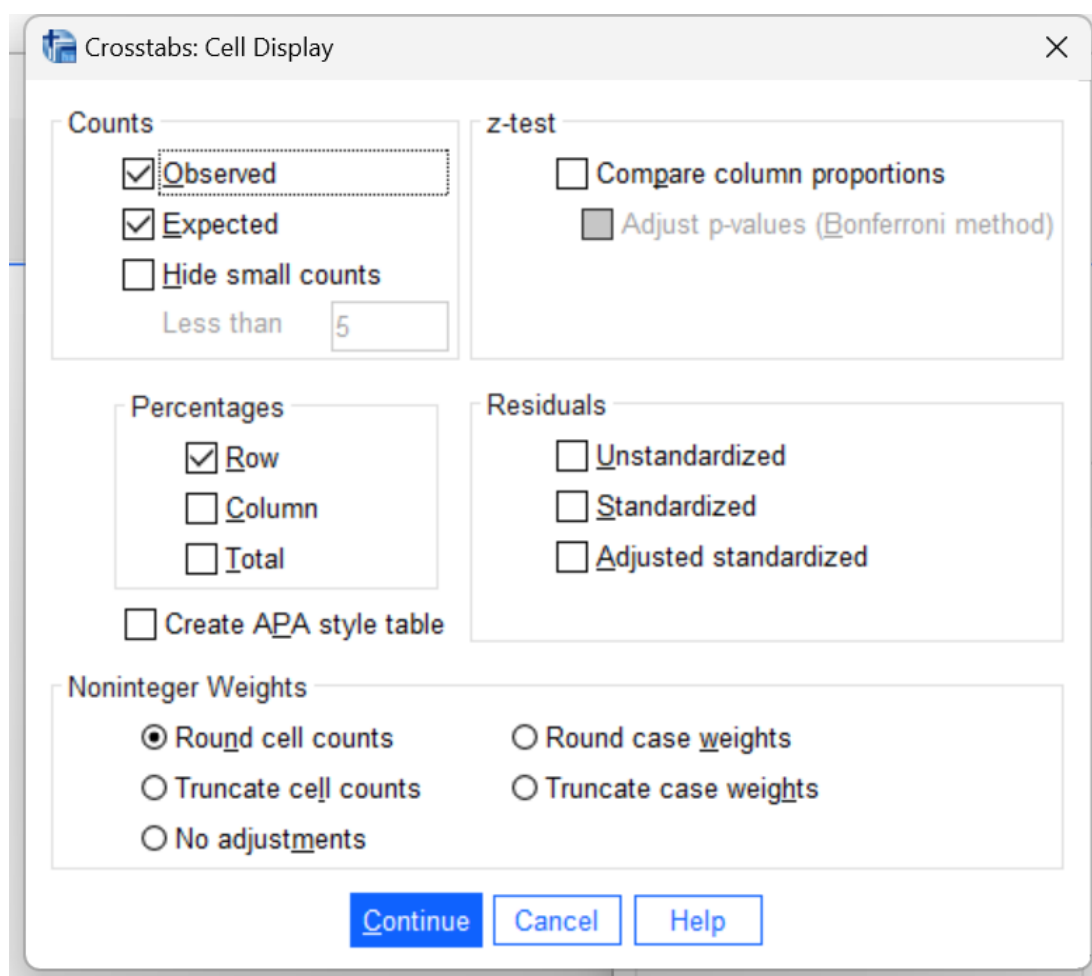


图 17 列联表分析：显示更多单元格信息的设置

点击“Continue”回到主对话框，再点击“OK”运行分析。SPSS 将在列联表的单元格中额外呈现期望频数和行百分比，即每一类更新功能内部不同建筑质量的占比，如图 18 所示。同学们可以根据需求，自己尝试其他单元格信息。

更新后的功能 \* 建筑质量 Crosstabulation

			建筑质量			Total
			差	好	中	
更新后的功能	社会服务设施	Count	0	2	0	2
		Expected Count	.9	.5	.6	2.0
		% within 更新后的功能	0.0%	100.0%	0.0%	100.0%
	创意产业园	Count	0	13	0	13
		Expected Count	6.0	3.1	3.9	13.0
		% within 更新后的功能	0.0%	100.0%	0.0%	100.0%
	商务办公	Count	3	8	13	24
		Expected Count	11.1	5.7	7.2	24.0
		% within 更新后的功能	12.5%	33.3%	54.2%	100.0%
	旅馆	Count	1	5	3	9
		Expected Count	4.2	2.1	2.7	9.0
		% within 更新后的功能	11.1%	55.6%	33.3%	100.0%
	商业服务	Count	20	1	9	30
		Expected Count	13.8	7.2	9.0	30.0
		% within 更新后的功能	66.7%	3.3%	30.0%	100.0%
	仓储物流生产	Count	27	1	9	37
		Expected Count	17.1	8.8	11.1	37.0
		% within 更新后的功能	73.0%	2.7%	24.3%	100.0%
未更新	Count	9	1	5	15	
	Expected Count	6.9	3.6	4.5	15.0	
	% within 更新后的功能	60.0%	6.7%	33.3%	100.0%	
Total	Count	60	31	39	130	
	Expected Count	60.0	31.0	39.0	130.0	
	% within 更新后的功能	46.2%	23.8%	30.0%	100.0%	

图 18 列联表表格：实际频数+期望频数+行百分比

以上内容均属于描述性统计的范畴，这也是列联表分析为什么位于“Descriptive Statistics”菜单下的原因。然而，我们最关心的其实是两个分类变量是否具有相关性，这就需要使用列联表分析的推断性统计功能——卡方检验。再次打开列联表分析的主对话框，保持之前的设置，然后点击“Statistics”按钮，打开统计量子对话框。勾选左上方的“Chi-square”，要求进行卡方检验。同时，勾选“Nominal”面板中的“Contingency coefficient”和“Phi and Cramer’s V”，要求计算反映相关程度的列联系数、Phi 指标、Cramer’s V 指标，如图 19 所示。

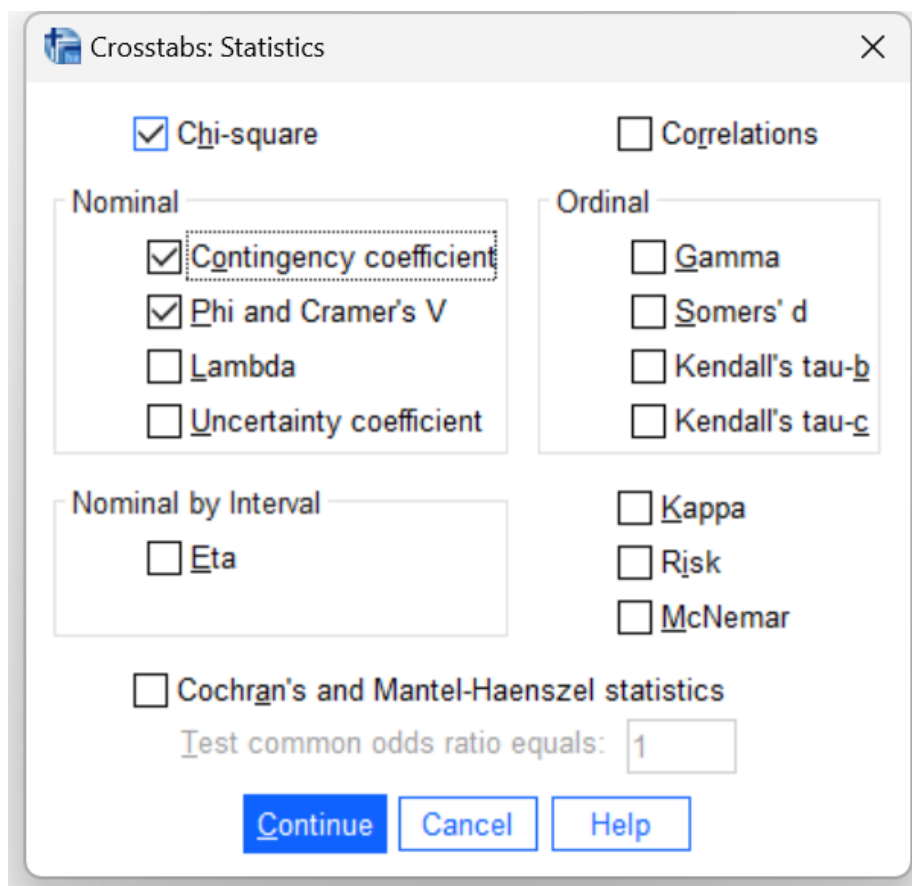


图 19 列联表分析：卡方检验和相关性程度指标的设定

点击“Continue”回到主对话框，再点击“OK”运行分析。SPSS 会在列联表的下方输出卡方检验的结果，包括卡方统计量 (“Value”)、自由度 (“df”)、双侧检验的 p 值 (“Asymptotic Significance (2-sided)”)，如图 20 所示。表中报告了 3 种卡方检验的结果，其各自的适用性请参考课件的内容。

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	89.257 <sup>a</sup>	12	<.001
Likelihood Ratio	92.741	12	<.001
Linear-by-Linear Association	11.546	1	<.001
N of Valid Cases	130		

a. 10 cells (47.6%) have expected count less than 5. The minimum expected count is .48.

图 20 卡方检验结果

在卡方检验的下方，SPSS 输出了测度两个分类变量相关性强度的三个指标：Phi、Cramer's V、列联系数（Contingency Coefficient）的取值（“Value”）和 p 值显著性（“Approximate Significance”），如图 21 所示。这里的 p 值其实就是图 20 中“Pearson Chi-Square”的 p 值。

		Value	Approximate Significance
Nominal by Nominal	Phi	.829	<.001
	Cramer's V	.586	<.001
	Contingency Coefficient	.638	<.001
N of Valid Cases		130	

图 21 两个分类变量相关性程度指标的计算结果

### 3.2 绘制簇形或堆积柱状图

在分析两个分类变量的相关模式时，柱状图能够给予直观的提示。我们可以在执行列联表分析时，在主对话框中勾选左下角的“Display clustered bar charts”，要求绘制簇形柱状图（图 22）。

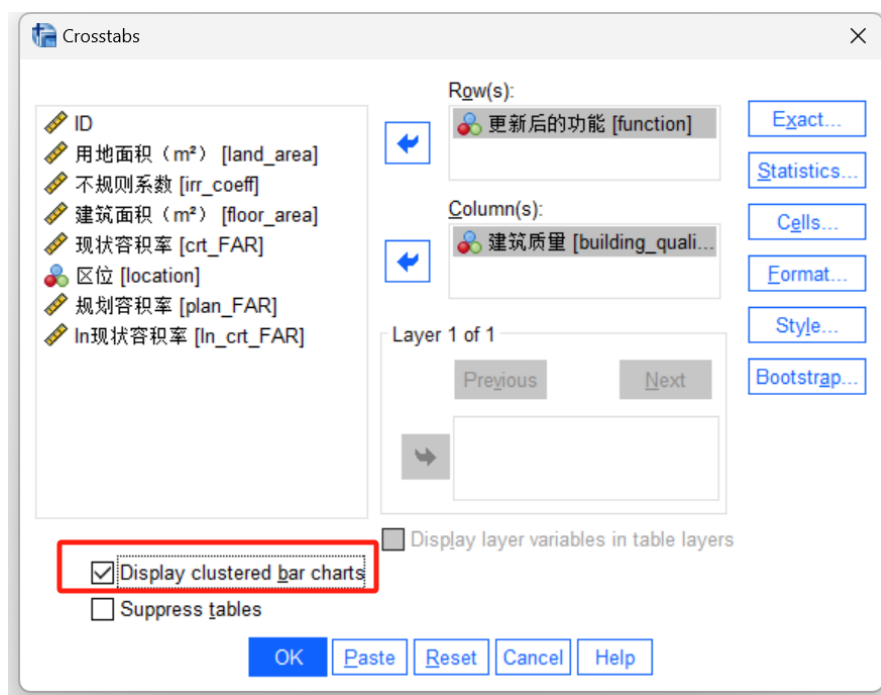


图 22 列联表分析：绘制簇形柱状图的设置

根据该设定，SPSS 会在输出窗口中呈现如图 23 所示的簇形柱状图。该图也可以通过菜单栏中的“Graph → Chart Builder”或“Graph → Legacy Diagos → Bar → Clustered”进行绘制。

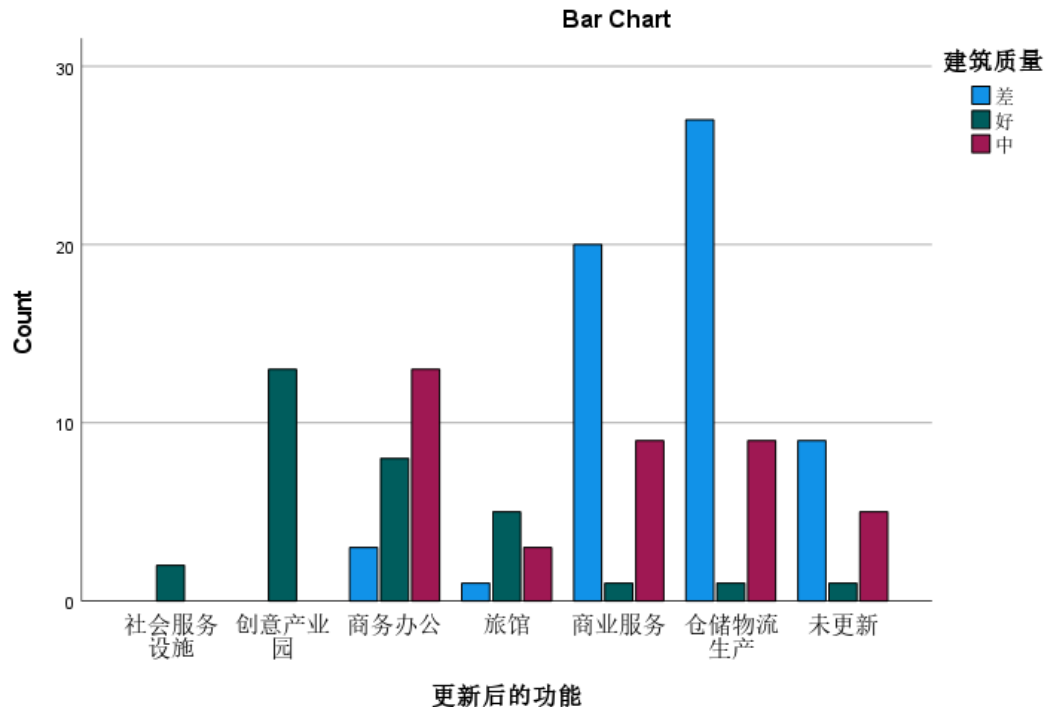


图 23 簇形柱状图

另一方面，堆积柱状图可能比簇形柱状图更容易识别具体的关联模式。我们可以通过菜单栏中的“Graph → Chart Builder”或“Graph → Legacy Dialogs → Bar → Stacked”绘制，这里演示前者的操作过程。打开如图 24 所示的“Chart Builder”对话框：在下方的“Gallery”中，把柱状图组（“Bar”）中的第 3 个图标——堆积柱状图（“Stacked Bar”）拖入工作区；然后，将左侧“Variables”列表中的“更新后的功能 [function]”拖入待定义的 X 轴（“X-Axis ?”）；最后，将“建筑质量 [building\_quality]”拖入工作区右上角待定义的堆积颜色（“Stack: set color”）中。

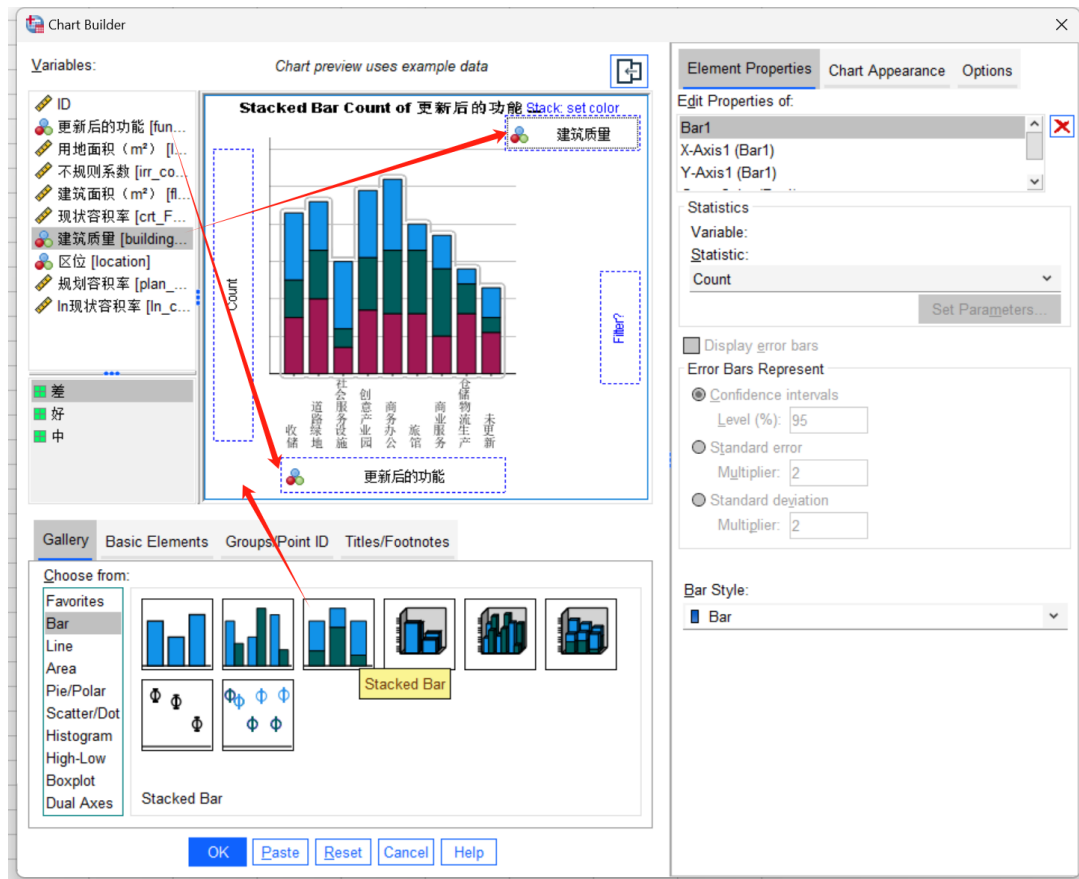


图 24 在 Chart Builder 中绘制堆积柱状图

点击“OK”执行分析，SPSS 的结果窗口中将呈现图 25 所示的堆积柱状图。

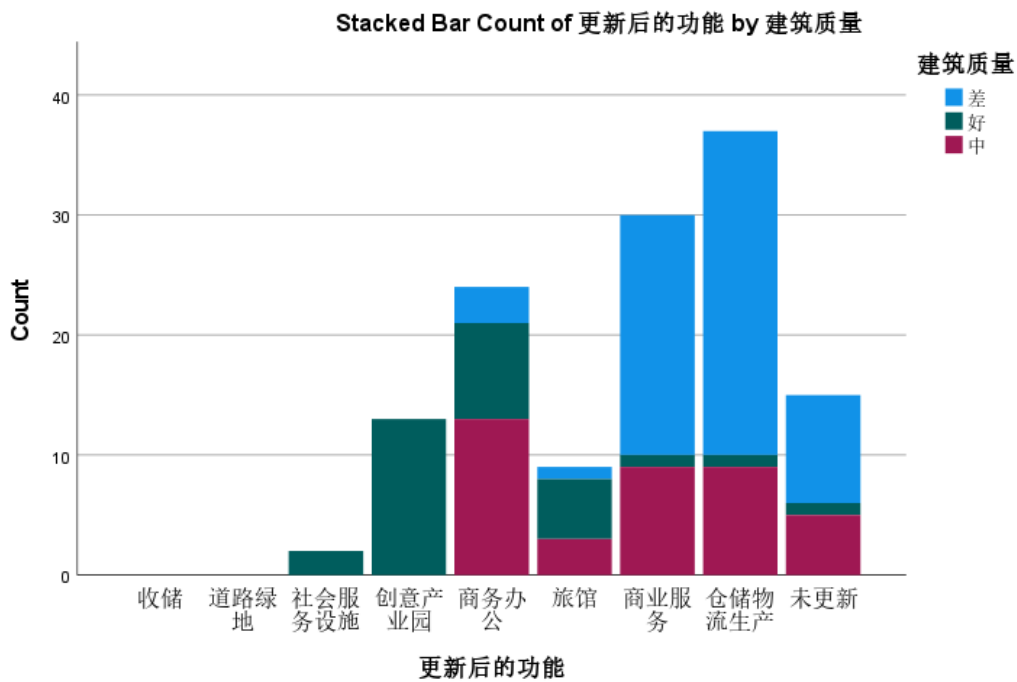



图 25 尺度未缩放的原始堆积柱状图

图 25 中的柱高表示样本的绝对数量，我们也可以将其转化为相对意义上的占比。在输出窗口中双击该图，打开“Chart Editor”，在其工具栏中点击图标，或者在其菜单栏中选择“Options → Scale to 100%”，结果如图 26 所示。值得注意的是，图 25 和图 26 中“收储”和“道路绿地”两类功能的柱子为空，这是因为此两类用地上没有建筑，建筑质量的数据自然是缺失的，我们可以在“Chart Editor”中将它们移除。

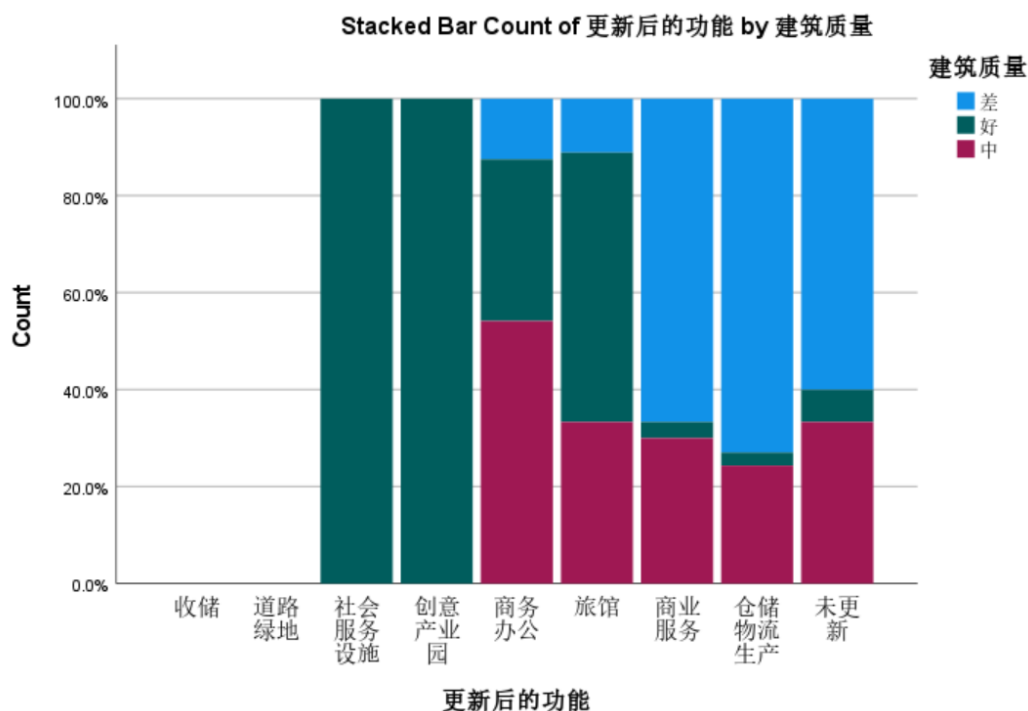


图 26 尺度缩放至 100%后的堆积柱状图

### 3.3 对应分析

对应分析是另一种直观反映两个分类变量之间关联模式的描述性统计方法，主要输出结果是二维对应分析图。在 SPSS 的菜单栏中依次选择“Analyze → Dimension Reduction → Correspondence Analysis”，打开对应分析对话框。分别将“更新后的功能 [function]”和“建筑质量 [building\_quality]”这两个变量选入“Row”和“Column”中，此时将呈现“function(? ?)”和“building\_quality(? ?)”，提示类别编码的取值范围未定义。点击行变量“function”下方的“Define Range”按钮，在“Minimum value”和“Maximum value”中分别输入“1”和“9”，并点击“Update”进行更新，再点击“Continue”回到主对话框，此时的行变量状态变为“function(1 9)”，

提示我们已经将“function”的类别编码范围设置为 1~9。用相同的方法，将列变量“building\_quality”的类别编码范围设置为 1~3，如图 27 所示。

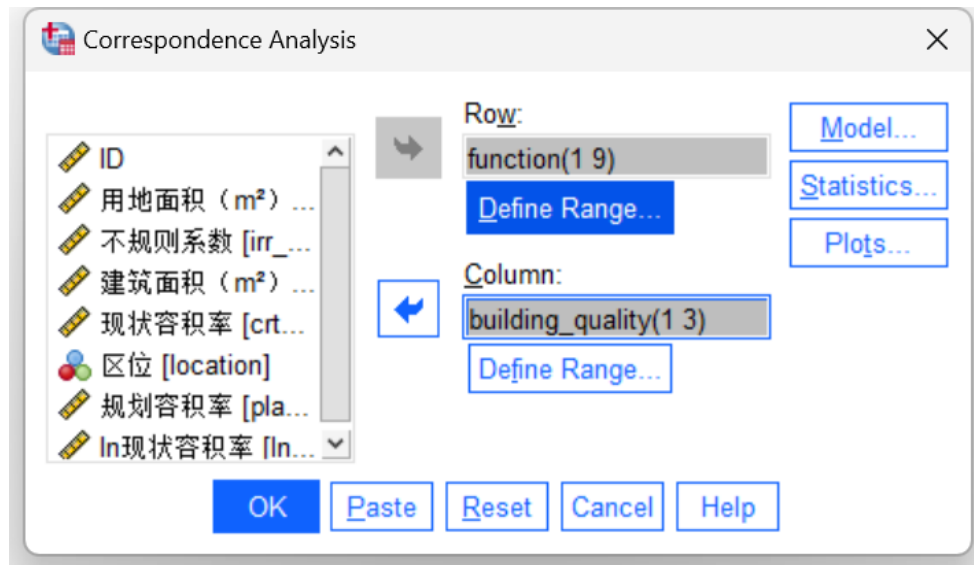


图 27 对应分析设置

点击“OK”进行分析，SPSS 的输出窗口中会报告一系列结果，但我们仅仅关注二维对应分析图。其基本原理是：对应分析算法将高维数据（原始维度等于行数乘以列数）近似地压缩至二维平面，平面中距离越近的点之间的关联性越强。

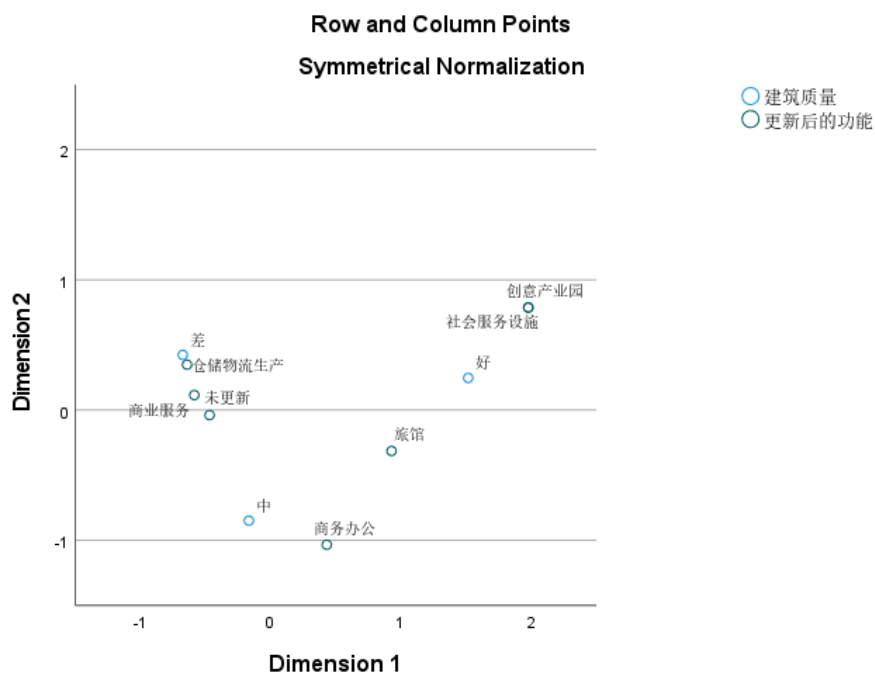


图 28 二维对应分析图

本例中，SPSS 输出的原始对应分析图如图 28 所示。我们可以通过“Chart

Editor”对其做进一步的修饰和美化，使结果更加清晰，这里不再赘述。

## 4. 数值变量与数值变量的相关性

用地面积和建筑面积均为数值变量，本小节以这 2 个变量为例，演示针对“数值变量 vs. 数值变量”的散点图、相关分析等方法的操作过程。

### 4.1 散点图

散点图是直观认识两个数值变量之间关系的最主要的方法，在 SPSS 中可以通过在菜单栏中选择“Graph → Chart Builder”或“Graph → Legacy Dialogs → Scatter/Dot → Simple Scatter”进行绘制，这里演示前者的绘制过程。

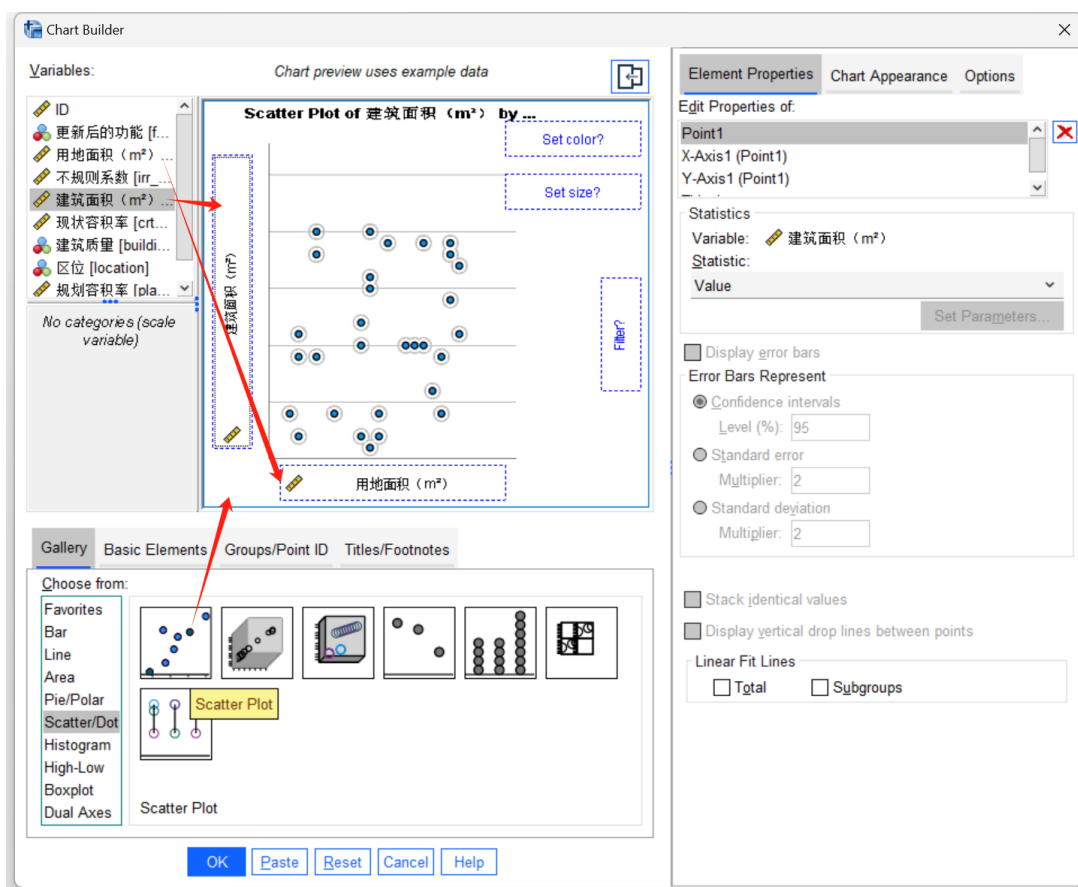


图 29 在 Chart Builder 中绘制散点图的设置

打开如图 29 所示的“Chart Builder”对话框。在下方的“Gallery”中，将散点图 (“Scatter/Dot”) 组的第 1 个图标“Scatter Plot”拖拽入工作区；然后，将左侧“Variables”列表中的“用地面积(m<sup>2</sup>) [land\_area]”拖入未定义的 X 轴 (“X-Axis ?”) 上，将“建筑面积 (m<sup>2</sup>) [floor\_area]”拖入未定义的 Y 轴 (“Y-Axis ?”) 上，点击

“OK”进行分析。SPSS 的输出窗口将呈现如图 30 所示的散点图。

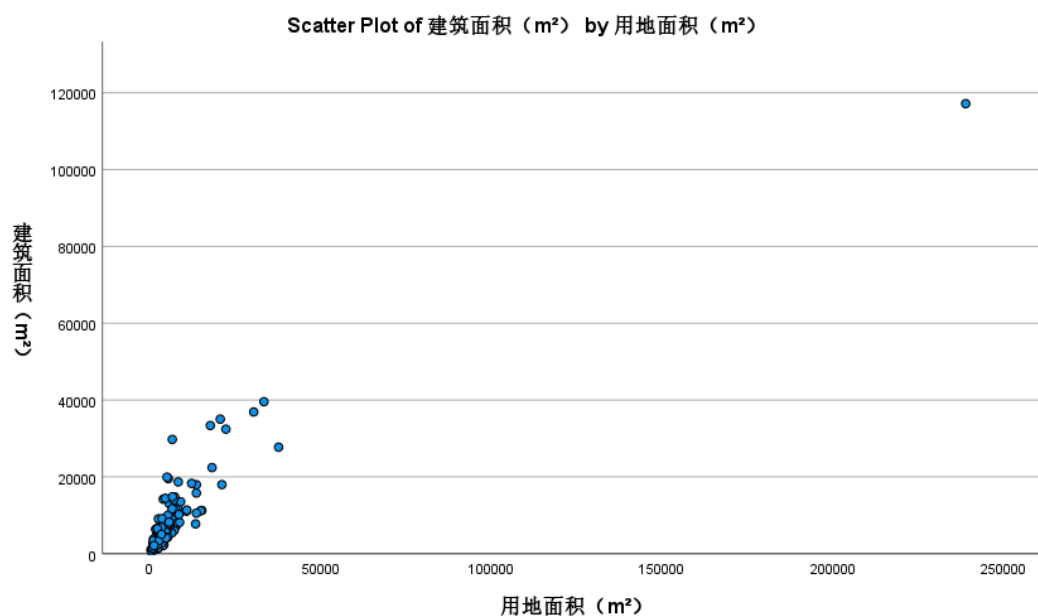


图 30 用地面积和建筑面积的散点图

从图 30 可以明确地看到，数据中存在一个离群样本，其用地面积和建筑面积都远远高于其他样本，可能会影响后续分析。对此，我们可以通过数据筛选将其滤除。在菜单栏中选择“Data → Select Cases”，打开数据筛选对话框。默认状态下，“All cases”被选中，即使用所有数据。我们选择“If condition is satisfied”，并点击其下方的“IF”按钮，弹出筛选条件对话框，在其输入框中设置过滤条件“land\_area < 200000”，如图 31 所示。为避免输入错误，可以使用如下的安全方式：将左侧变量列表中的“用地面积 (m2) [land\_area]”选入右侧的输入框中，再在下方点击“<”号，最后再键入 200000。点击“Continue”回到主对话框，再点击“OK”执行筛选。此时，我们可以在数据视图中发现：第 89 行样本的行号上有一道斜线（图 32），表明该样本被滤除了，其用地面积为 239000，不满足图 31 设定的过滤条件。

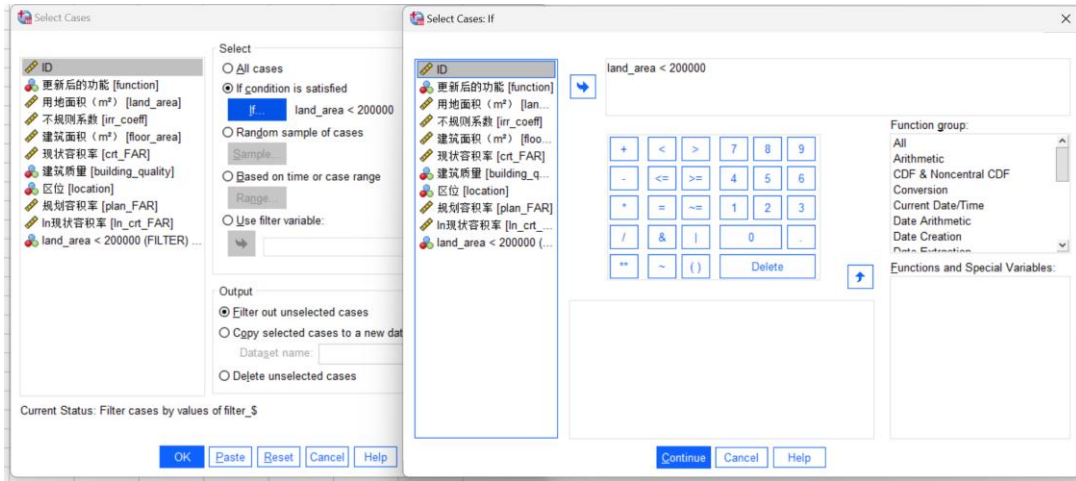



图 31 基于用地面积对离群数据进行过滤

	ID	function	land_area	irr_coef	floor_area	crt_FAR	building_quality
88	97	仓储物流生...	20600	4.9235800	35053	1.70	差
89	98	未更新	239000	4.2216700	117184	.49	中
90	99	未更新	13600	4.8425400	17969	1.32	中

图 32 被滤除的离群样本

为了更好地满足正态性和同方差的条件，除了过滤离群值以外，我们还分别对用地面积和建筑面积进行对数变换，这两个变量已经存在于数据中，变量名为“ln\_land\_area”和“ln\_floor\_area”。为二者绘制散点图，如图 33 所示。对于图中的趋势线，我们可以双击散点图进入“Chart Editor”，然后点击工具栏上的图标，或者在菜单栏中选择“Options → Reference Line from Equation”进行添加。

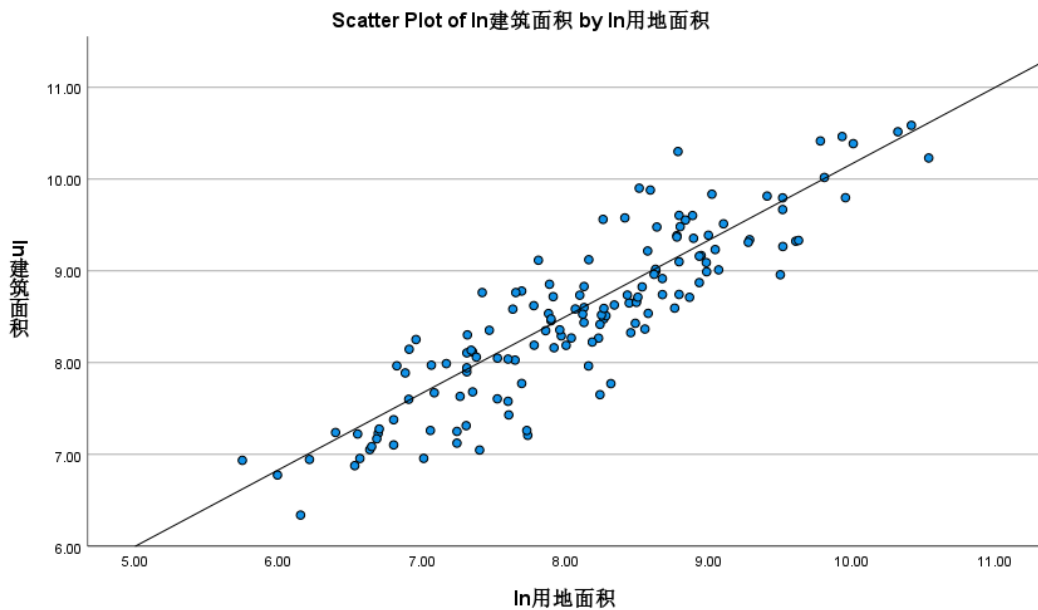


图 33 ln 用地面积和 ln 建筑面积的散点图（已过滤离群值）

## 4.2 Pearson 相关系数

在菜单栏中选择“Analyze → Correlation”，打开相关分析对话框。将“ln 用地面积 [ln\_land\_area]”和“ln 建筑面积 [ln\_floor\_area]”选入“Variables”中，确认下方的“Correlation Coefficients”中，“Pearson”被选中，如图 34 所示。

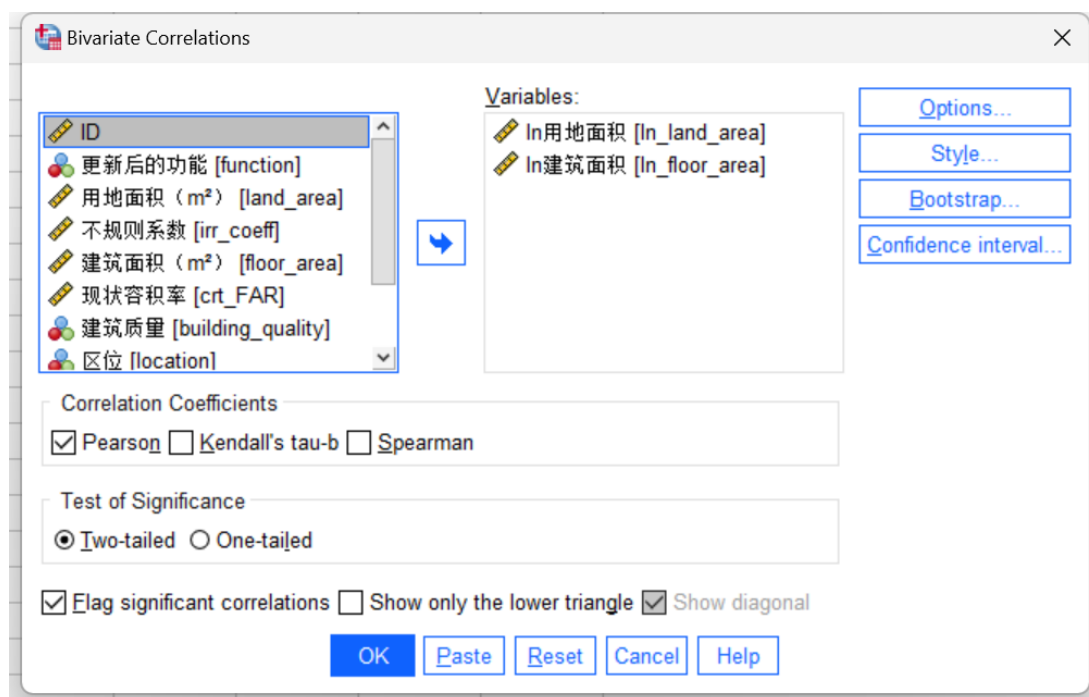


图 34 相关分析设置：Pearson 相关系数

点击“OK”运行分析，SPSS 的输出窗口将报告如图 35 所示的结果，包括相关系数的取值（“Pearson Correlation”）和判断显著性的 p 值（“Sig. (2-tailed)”）。表格下方的脚注提示：在 0.01 水平下显著的结果被标记为两颗星（“\*\*”）。

		ln用地面积	ln建筑面积
ln用地面积	Pearson Correlation	1	.890**
	Sig. (2-tailed)		<.001
	N	144	144
ln建筑面积	Pearson Correlation	.890**	1
	Sig. (2-tailed)	<.001	
	N	144	144

\*\* . Correlation is significant at the 0.01 level (2-tailed).

图 35 相关分析结果：Pearson 相关系数

我们可以同时对 N 个数值变量进行两两相关分析，结果将是一个 N 行 N 列的相关系数矩阵。该矩阵沿着对角线上下对称，即交换变量的顺序不会影响相关系数的取值。基于这一特点，有时候我们希望仅显示对角线以下的部分，以简化输出，这种输出样式可以通过在图 34 中勾选下方的“Show only the lower triangle”实现。

### 4.3 非参数方法

当数据存在严重的非正态、异方差、极端值问题时，我们也可以改用非参数方法，即 Spearman 相关系数和 Kendall's tau-b 相关系数。再次在菜单栏中选择“Analyze → Correlation”，打开相关分析对话框。如果对数变化后的变量还在“Variable”中，请先将它们移出，然后再将“用地面积 (m<sup>2</sup>) [land\_area]”和“建筑面积 (m<sup>2</sup>) [floor\_area]”这两个原始变量选入“Variables”中。这是因为非参数方法使用的是秩信息，而原始变量和对数变换后的变量具有相同的秩信息，因此不需要对数变换。在“Correlation Coefficients”中取消对“Pearson”的勾选，转而选中“Spearman”和“Kendall's tau-b”，如图 36 所示。

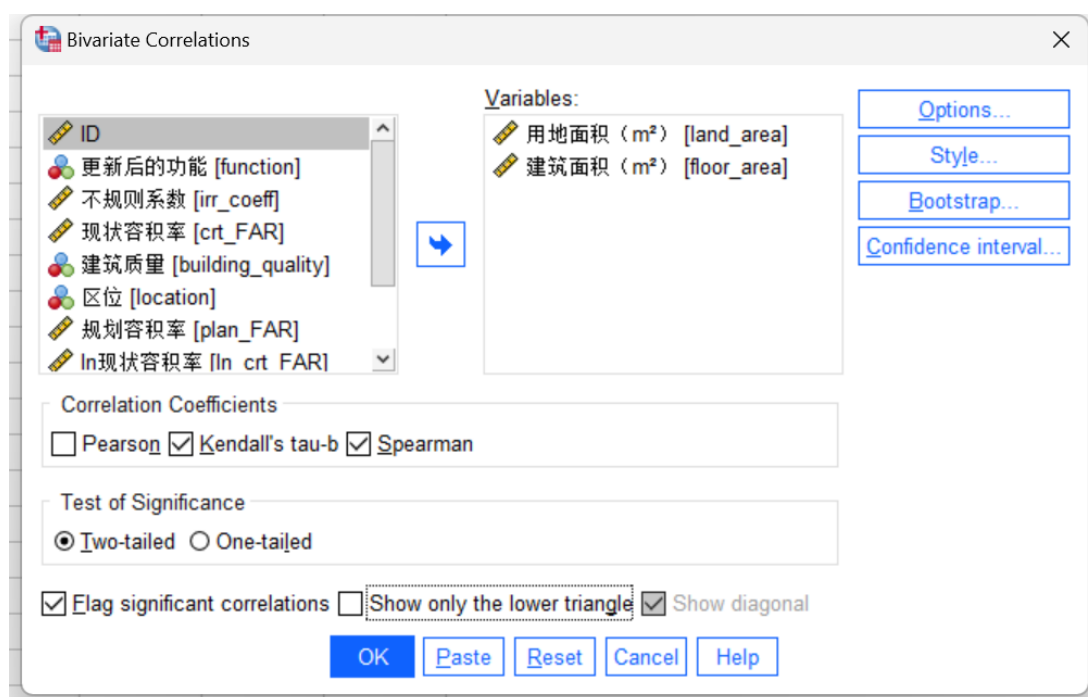


图 36 相关分析设置：Spearman 相关系数和 Kendall's tau-b 相关系数

点击“OK”运行分析，SPSS 的输出窗口将报告如图 37 所示的结果。上方为 Kendall's tau-b 相关系数，下方为更常用的 Spearman 相关系数。两组结果均包括

相关系数的取值 (“Correlation Coefficient”) 和 p 值 (Sig. (2-tailed))。表格下方的脚注提示：在 0.01 水平下显著的结果被标记为两颗星 (“\*\*”)。

**Correlations**

			用地面积 (m <sup>2</sup> )	建筑面积 (m <sup>2</sup> )
Kendall's tau_b	用地面积 (m <sup>2</sup> )	Correlation Coefficient	1.000	.712**
		Sig. (2-tailed)	.	<.001
		N	144	144
	建筑面积 (m <sup>2</sup> )	Correlation Coefficient	.712**	1.000
		Sig. (2-tailed)	<.001	.
		N	144	144
Spearman's rho	用地面积 (m <sup>2</sup> )	Correlation Coefficient	1.000	.890**
		Sig. (2-tailed)	.	<.001
		N	144	144
	建筑面积 (m <sup>2</sup> )	Correlation Coefficient	.890**	1.000
		Sig. (2-tailed)	<.001	.
		N	144	144

\*\* . Correlation is significant at the 0.01 level (2-tailed).

图 37 相关分析结果：Spearman 相关系数和 Kendall' s tau-b 相关系数