

1. 概述

本节以苏州和南京的二手房房价为例，利用 SPSS 软件演示简单的推断性统计检验的应用。SPSS 的数据文件为“house_price.sav”，同时还提供了同名的 csv 文件。

2. 正态性检验

在菜单栏中依次选择“Analyze → Descriptive Statistics → Explore”，打开数据探索对话框。该程序集合了许多在其他地方也能找到的基础描述性统计和绘图功能，便于用户一站式地快速建立对数据的直观认识。同时，该程序还包含一些特色功能，如本节所介绍的正态性检验。需要指出，虽然正态性检验被附属在了“Descriptive Statistics”的程序中，但是它属于推断性统计。

将“price023”和“price2024”变量选入“Dependent List”，它们是拟被探索分析的目标变量；将“city”变量选入“Factor List”，它是目标变量相应的组变量，如图 1 所示。

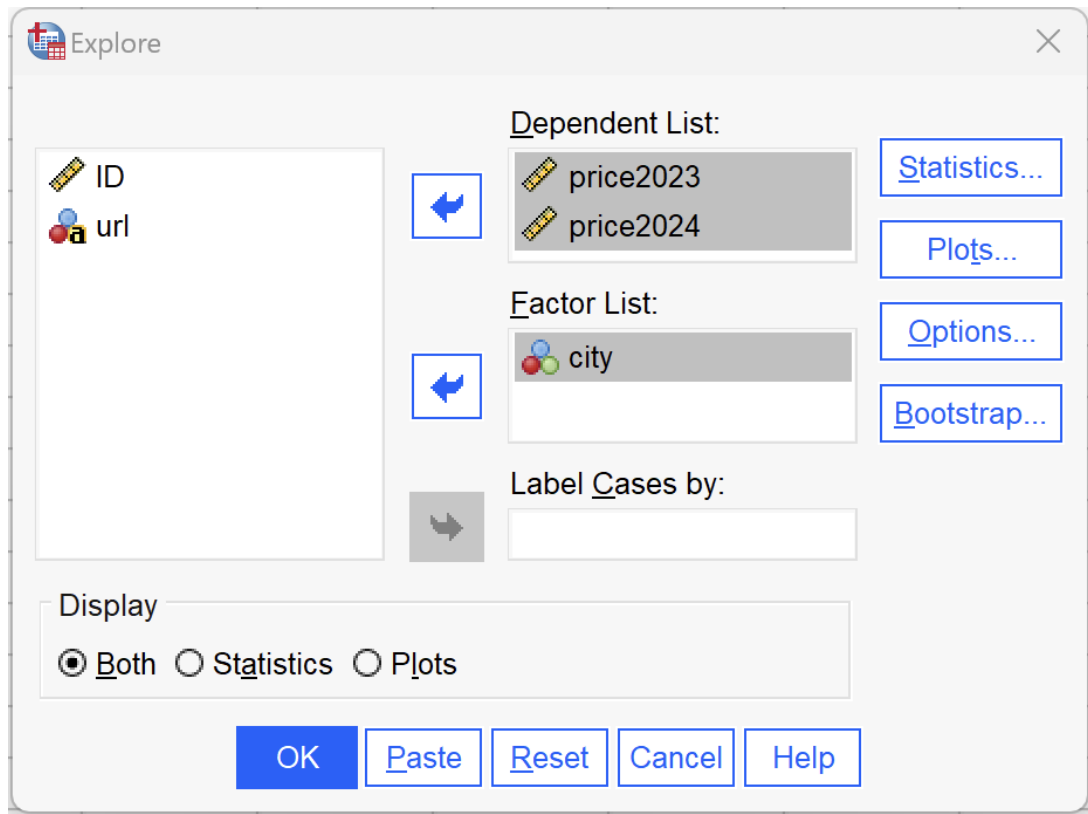


图 1 数据探索主对话框的设置

点击主对话框上的“Plot”按钮，打开绘图形对话框。SPSS 默认选中了箱形

图 (“Boxplots”) 和茎叶图 (“Stem-and-leaf”)。由于后者目前的使用场合较少，且反映的信息与直方图高度重合，因此我们取消对茎叶图的选中，转而勾选直方图 (“Histogram”)。另外，作为本次分析的主体，我们勾选 “Normality plots with tests”，要求绘制更直接地反映正态性的 Q-Q 图，以及相应的假设检验。设置完成的状态如图 2 所示。设置完成后点击 “Continue” 回到主对话框。

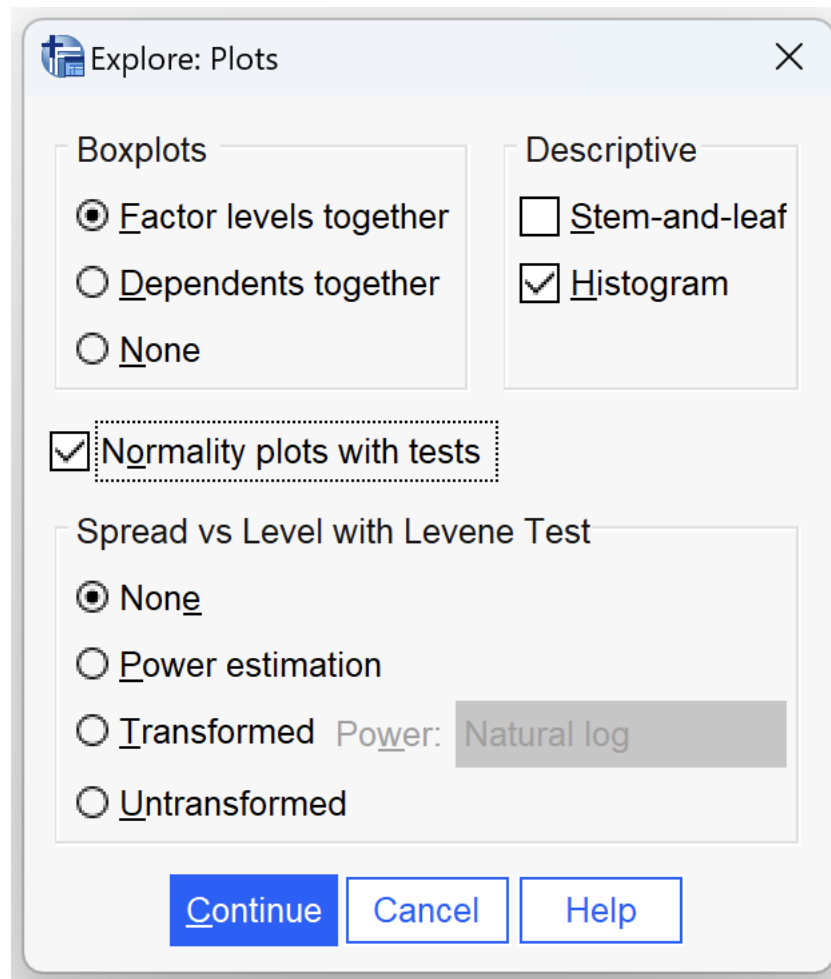


图 2 Plots 子对话框的设置

在执行分析之前，我们还需要对缺失值进行设置。缺失值是统计分析中的常见问题，在一些场合下需要慎重设置。我们的数据仅提供了南京市样本 2024 年房价，其相应的 2023 年房价 (“price2023”) 均为空，也就是缺失值。我们在主对话框中点击 “Options” 按钮，打开选项子对话框。可以看到，它主要控制了 “Missing values” 的处理方式。默认状态下，“Exclude cases listwise” 被选中，这的意思是一条数据中，只要有一个变量缺失，整条数据即被剔除。所以，如果我们直接保持默认状态继续分析，所有的南京样本都将被排除，这显示是不合适的。为此，我们转而选中 “Exclude cases pairwise”，如图 3 所示。它的意义是：

当我们使用某 2 个变量进行分析时，在这 2 个变量上存在缺失值的样本将被剔除；而在其他变量上存在缺失值，但在这 2 个变量上有记录的样本将会被保留。这样的设置是符合本例预期的。

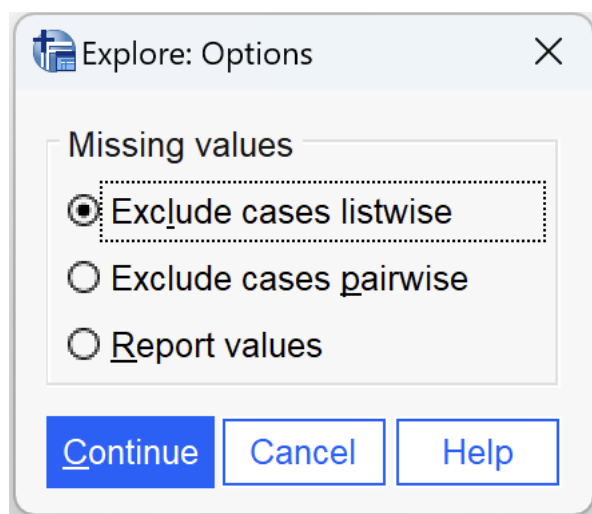


图 3 Options 子对话框的设置

点击“Continue”回到主对话框，然后点击“OK”运行分析。SPSS 的输出窗口中首先呈现了数据的基本情况（“Case Processing Summary”，如图 4），包括有效样本、缺失值样本、总样本的数量。由于我们把“city”变量设置为了“Factor”，所以，结果均按照“city”的两个组别——苏州和南京分别显示。接下来，SPSS 呈现了描述性统计量的计算结果（“Statistics”，如图 5），包括均值、中位数、方差等，绝大多数已在前一章介绍过，不再赘述。值得一提的是“5% Trimmed Mean”，如图 5 中的红框。该统计量通常被译成“5%截尾均值”，所谓的“截尾”是指去掉数据集中最极端的 5%（一般为最低的 2.5%和最高的 2.5%），以减少极端的离群值对均值的影响，使结果更加稳健。

		Case Processing Summary					
		Valid		Cases Missing		Total	
city		N	Percent	N	Percent	N	Percent
price2023	苏州	42	100.0%	0	0.0%	42	100.0%
price2024	苏州	42	100.0%	0	0.0%	42	100.0%
	南京	51	100.0%	0	0.0%	51	100.0%

图 4 数据的基本情况

Descriptives^a

city			Statistic	Std. Error	
price2023	苏州	Mean	26331.40	1564.435	
		95% Confidence Interval for Mean	Lower Bound	23171.96	
			Upper Bound	29490.85	
		5% Trimmed Mean		25718.48	
		Median		24675.50	
		Variance		102793168.78	
		Std. Deviation		10138.697	

图 5 描述性统计量

下面报告的是本节的主体内容——正态性检验结果，如图 6 所示。该表格报告了 K-S（Kolmogorov-Smirnov）检验和 S-W（Shapiro-Wilk）检验两种结果。对于每一种检验，“Statistic”列是检验统计量，“df”列是自由度，这两列无需特别关注。“Sig.”列即为 p 值，是确认显著性的关键。

Tests of Normality^b

city		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
price2023	苏州	.179	42	.002	.913	42	.004
price2024	苏州	.115	42	.183	.943	42	.035
	南京	.150	51	.006	.922	51	.002

a. Lilliefors Significance Correction

b. There are no valid cases for price2023 when city = 2.000. Statistics cannot be computed for this level.

图 6 正态性检验结果

接下来，SPSS 为每个变量绘制了我们所要求的统计图，包括直方图、箱形图、Q-Q 图。我们仅对第一次出现的 Q-Q 图进行简要的介绍，供感兴趣的同学了解。Q-Q 图（Quantile-Quantile 图）是用来判断数据是否符合某种特定分布（通常是正态分布）的图形化工具。图 7 以苏州市样本 2023 年的房价数据为例，展示了用于辅助正态性检验的 Q-Q 图。图中的 x 轴代表样本数据的分位数，y 轴代表正态分布的对应分位数。如果样本数据服从正态分布，点将会落在图中的对角线上。图 7 中的点呈现出一个弯曲的形状，与对角线有一定的偏离，但程度并不大。因此，虽然 K-S 检验与 S-W 检验都拒绝了正态分布的零假设，但影响有限。

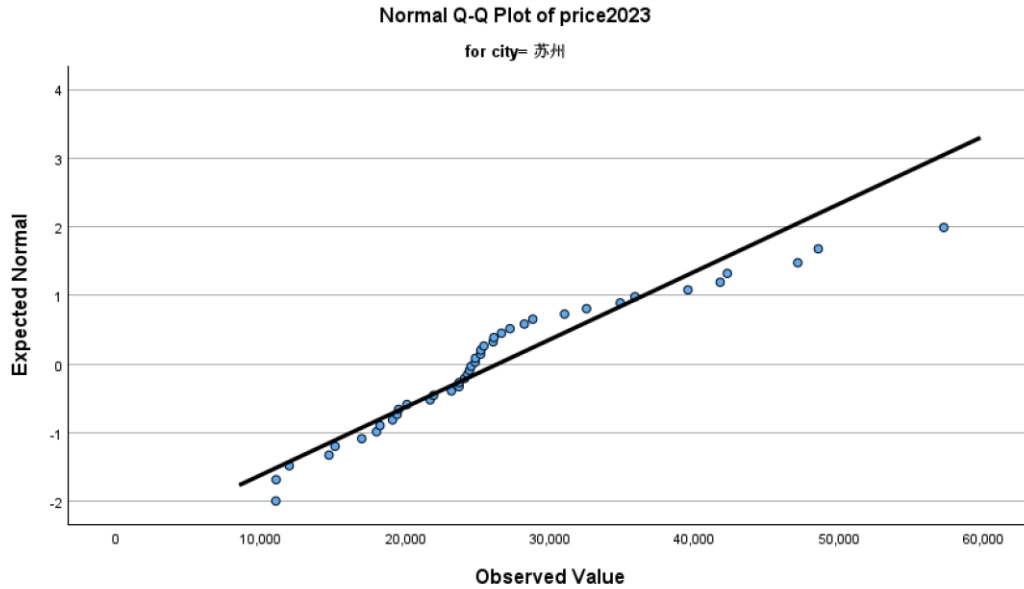


图 7 苏州市样本 2023 年房价的 Q-Q 图

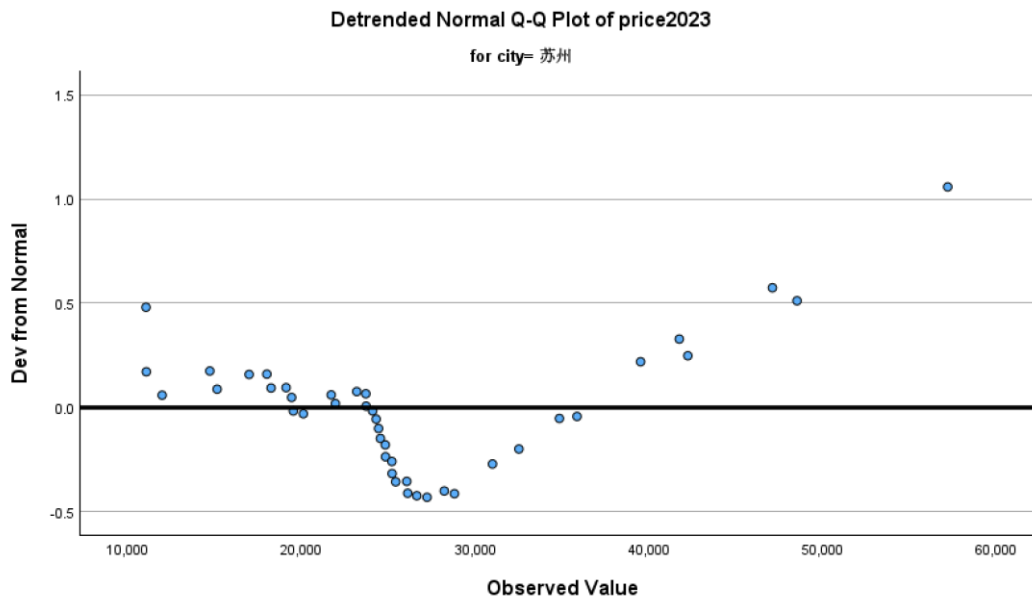


图 8 苏州市样本 2023 年房价的去趋势 Q-Q 图

除了标准的 Q-Q 图以外，SPSS 还会呈现去趋势的 Q-Q 图（Detrended Q-Q plot），如图 8 所示。该图计算了标准 Q-Q 图中同一点的 x 坐标与 y 坐标的差值，即样本数据的实际分位数与正态分布的理论分位数之间的差值，然后将这些差值绘制在 0 水平线上，从而更容易地观察样本与正态分布之间的偏差。如果点大致在 0 水平线附近随机波动，这表明数据与正态分布吻合得较好；反之，离 0 水平线越远的点代表这个区间段内的数据与正态分布的偏差越大。

在实践中，我们一方面要基于 K-S 和 S-W 这两个推断性假设检验的显著性结果，另一方面也要参考直方图、Q-Q 图等可视化结果，综合评估正态性。

最后，对于配对样本，最严谨的做法是检验差值（而非原始数据）是否服从正态分布。因此在本例中，我们需要计算苏州市同一样本 2024 年与 2023 年的房价差值。在菜单栏中依次选择“Transform → Compute Variable”，打开变量计算对话框。在“Target Variable”中输入“delta_price”，作为待生成变量的变量名；在“=”号右边的“Numeric Expression”中输入“price2024-price2023”，即变量计算的公式，如图 9 所示。点击“OK”后，即可在数据视图中看到新生成的变量“delta_price”。接下来，我们可以用相同的方法，检验该变量的正态性。

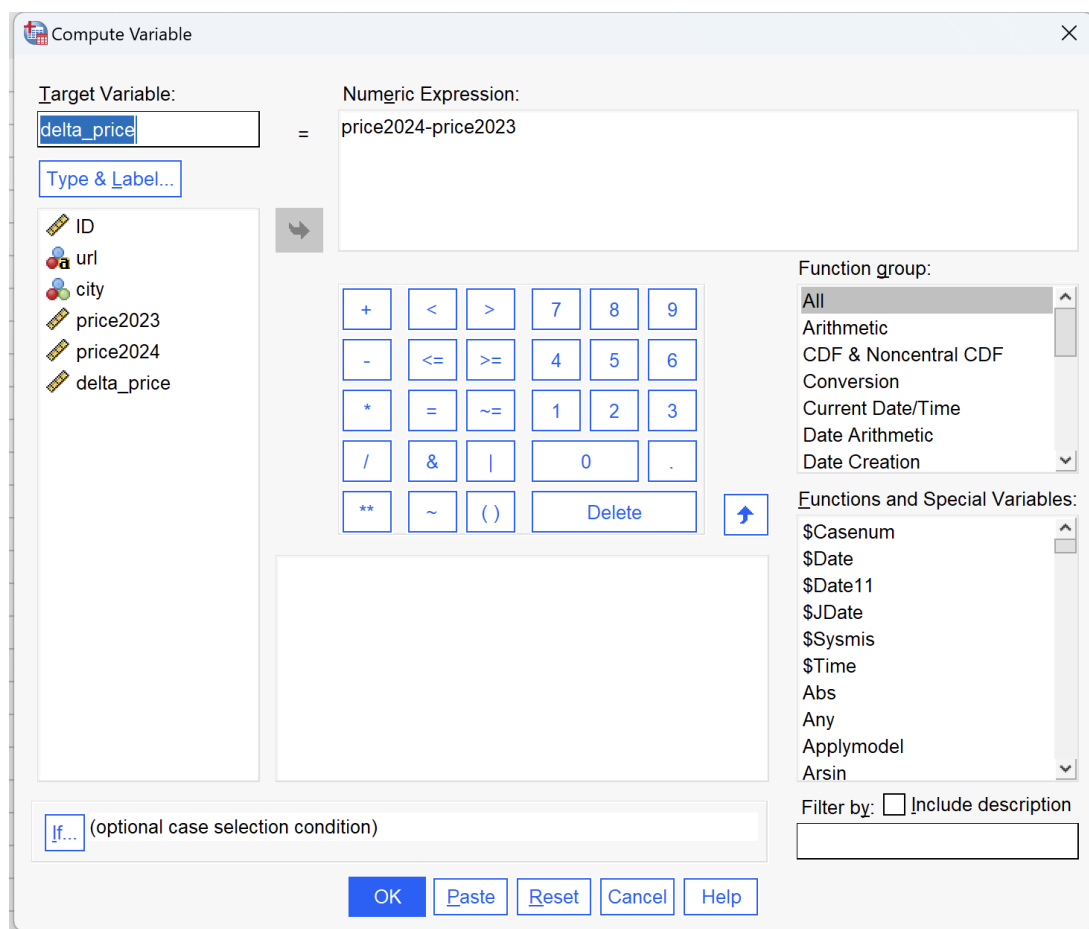


图 9 计算生成新的变量 delta_price

3. 独立样本的差异比较

本节将演示如何利用独立样本 T 检验和 Mann-Whitney 非参数检验两种方法，判断苏州和南京 2024 年的房价是否存在显著差异。

3.1 独立样本 T 检验

在菜单栏中选择“Analyze → Compare Means¹ → Independent-Samples T Test”，进入独立样本 T 检验对话框。将“price2024”选入“Test Variable(s)”，作为待检验的变量。将“city”选入“Group Variable”，作为分组变量，如图 10 所示。此时的组变量呈现“city(??)”的状态，表明具体的分组尚未定义。

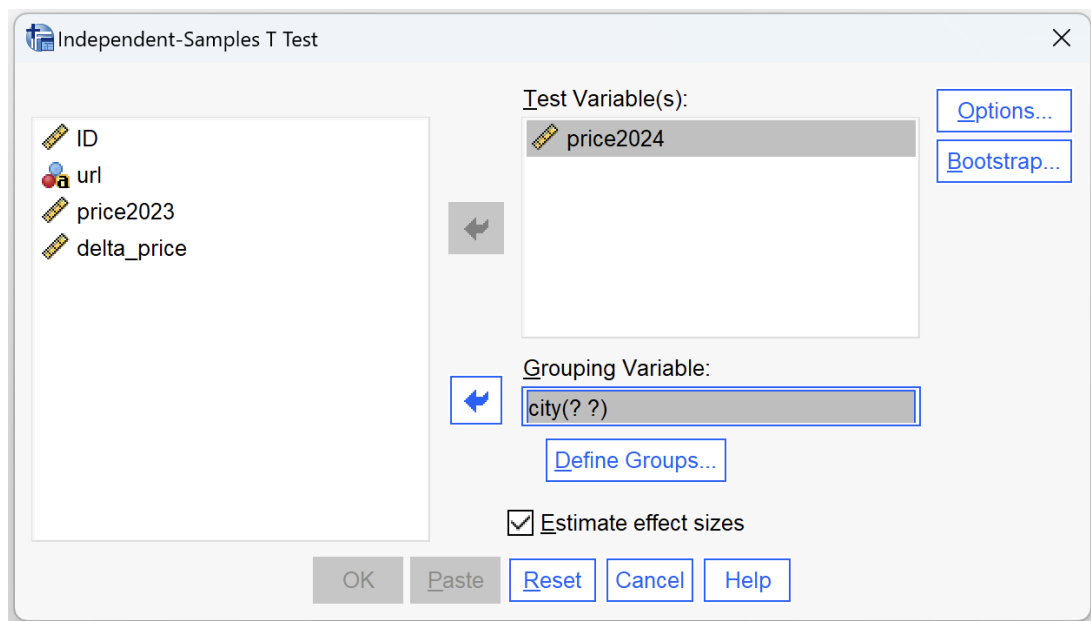


图 10 独立样本 T 检验的主体设置（分组未定义）

我们点击“Define Groups”，对分组进行定义。SPSS 提供了 2 种定义方式：第一种是“Use specified values”，即分别指定两组的数字编码，这里为 1 和 2，分别代表苏州和南京；第二种是“Cut point”，即指定一个分断点数值，高于此值和低于此值的样本分别构成两个组。

¹ 在最新版本的 SPSS29 中，该菜单的名称为“Compare Means and Proportions”。

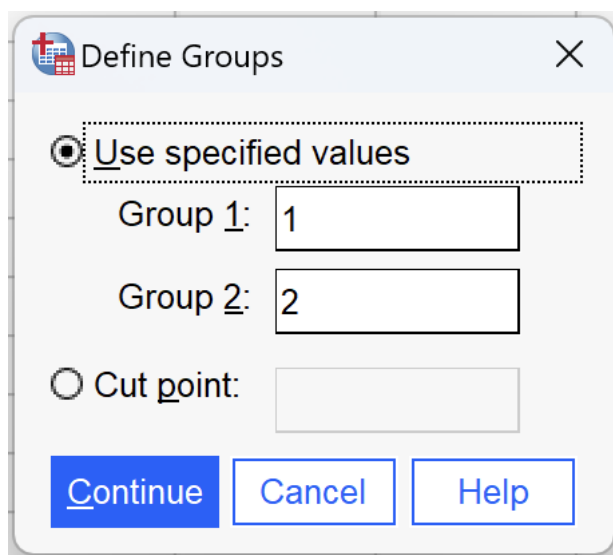


图 11 在独立样本 T 检验中定义分组

点击“Continue”后回到主对话框，此时的分组变量已经变成“city(1 2)”，即对“city=1”和“city=2”这两组进行比较，如图 12 所示。此外，我们可以取消对“Estimate effect size”的勾选，因为这一部分通常不是分析的重点。

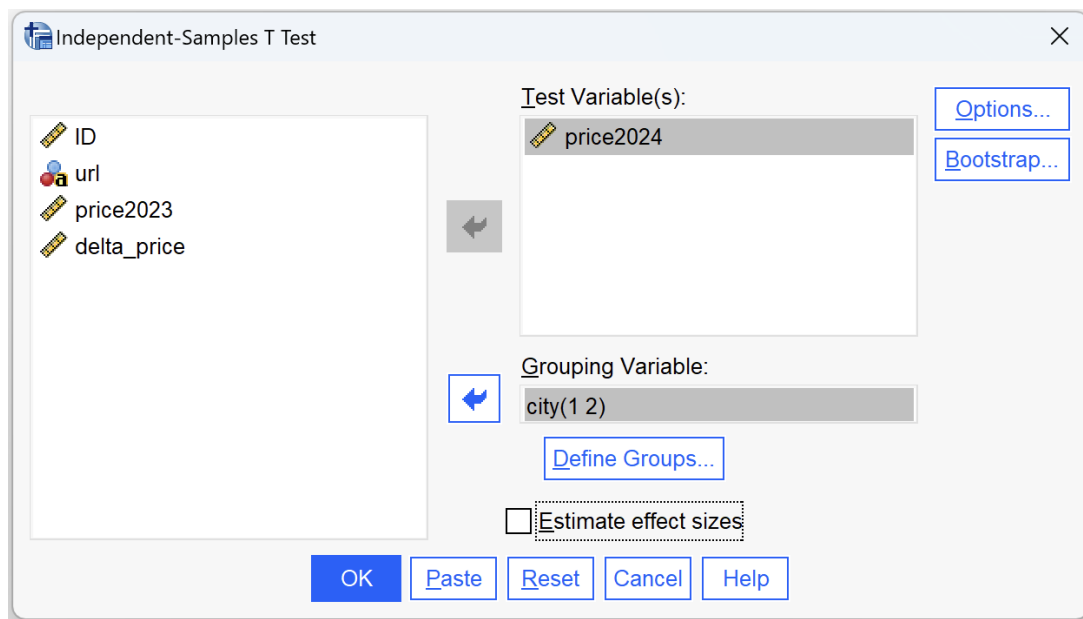


图 12 独立样本 T 检验的主体设置（分组已定义）

点击“OK”运行分析，SPSS 的输出窗口中首先报告了每一组的基本统计信息（“Group Statistics”），包括组内的样本量、均值、标准差，以及均值的标准误，如图 13 所示。

Group Statistics

	city	N	Mean	Std. Deviation	Std. Error Mean
price2024	苏州	42	22814.60	8798.089	1357.575
	南京	51	30415.41	13202.923	1848.779

图 13 独立样本 T 检验的分组统计量结果

接下来，SPSS 报告了 T 检验的结果，如图 14 所示。表左侧的“Levene’s Test for Equality of Variance”是方差齐性检验，“F”和“Sig.”分别是该检验的统计量和 P 值。如果 P 值大于 0.05，则无法拒绝方差齐的零假设，后面的结果参照第 1 行；反之，如果 P 值小于 0.05（即本例的情况），则拒绝零假设，后面的结果参照第 2 行。

Independent Samples Test											
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
price2024	Equal variances assumed	5.949	.017	-3.191	91	<.001	.002	-7600.817	2381.734	-12331.838	-2869.795
	Equal variances not assumed			-3.314	87.451	<.001	.001	-7600.817	2293.685	-12159.433	-3042.200

图 14 独立样本 T 检验的检验结果

对于后面的 T 检验结果（“t-test for Equality of Means”）：“t”列是本检验的统计量；“df”是自由度；“Significance（或 Sig.）”是直接反映显著性的 P 值；“Mean Difference”是两组之间的均值差；“Std. Error Difference”是该差值的标准误，即推断总体时的误差；“95% Confidence Interval of the Difference”是该差值的 95% 置信区间，即均值差有 95% 的可能性落在“Lower”和“Upper”之间的范围内。需要指出，在显著性结果上，较新版本的 SPSS 会同时报告单侧检验的 P 值（“One-Sided p”）和双侧检验的 P 值（“Two-Sided p”）；而之前的 SPSS 一般只报告双侧检验的 P 值（“2-tailed”）。这里简单解释二者的区别：双侧检验只考虑两组之间是否存在差异，而不预设差异的方向性，即组 1 与组 2 相比，不管是高是低，双侧检验都是敏感的，不会因为偏向某个方向的期望而忽略了另一方向的可能性，这种中立性和全面性使之在实践中更为常用；另一方面，单侧检验对差异有明确的方向性预期，当我们认为组 1 的均值必然大于或小于组 2 时，可以使用单侧检验，但也需要因此承担一定的偏差风险。综合而言，除非我们有强烈的方向倾向，否则直接使用双侧检验结果即可。

3.2 非参数检验

SPSS 中的非参数检验功能位于菜单栏的“Analyze → Nonparametric Tests”中。老版本的 SPSS 为每一种非参数检验单独设置了一个程序，而从版本 16.0 开始，SPSS 对它们做了一定程度的整合，同时把旧的程序归类为“Legacy Dialogs”。为了操作的清晰性和便利性，我们仍然按旧的操作习惯进行讲解。本小节演示的是两组独立样本的非参数检验；在 4.2 节将演示两组配对样本的非参数检验。

在菜单栏中选择“Analyze → Nonparametric Tests → Legacy Dialogs → 2 Independent Samples”，打开两独立样本非参数检验对话框。把“price2024”变量选入“Test Variable List”中，作为待检验的目标变量；把“city”变量选入“Grouping Variable”中，作为分组变量。与 T 检验类似，此时的分组仍处于未定义的状态，显示为“city(?)?”。我们点击“Define Groups”，在弹出的对话框中为“Group 1”输入“1”，为“Group 2”输入“2”，此时的分组变量显示为“city(1 2)”。确认“Test Type”中，“Mann-Whiney U”检验处于选中状态，如图 15 所示。

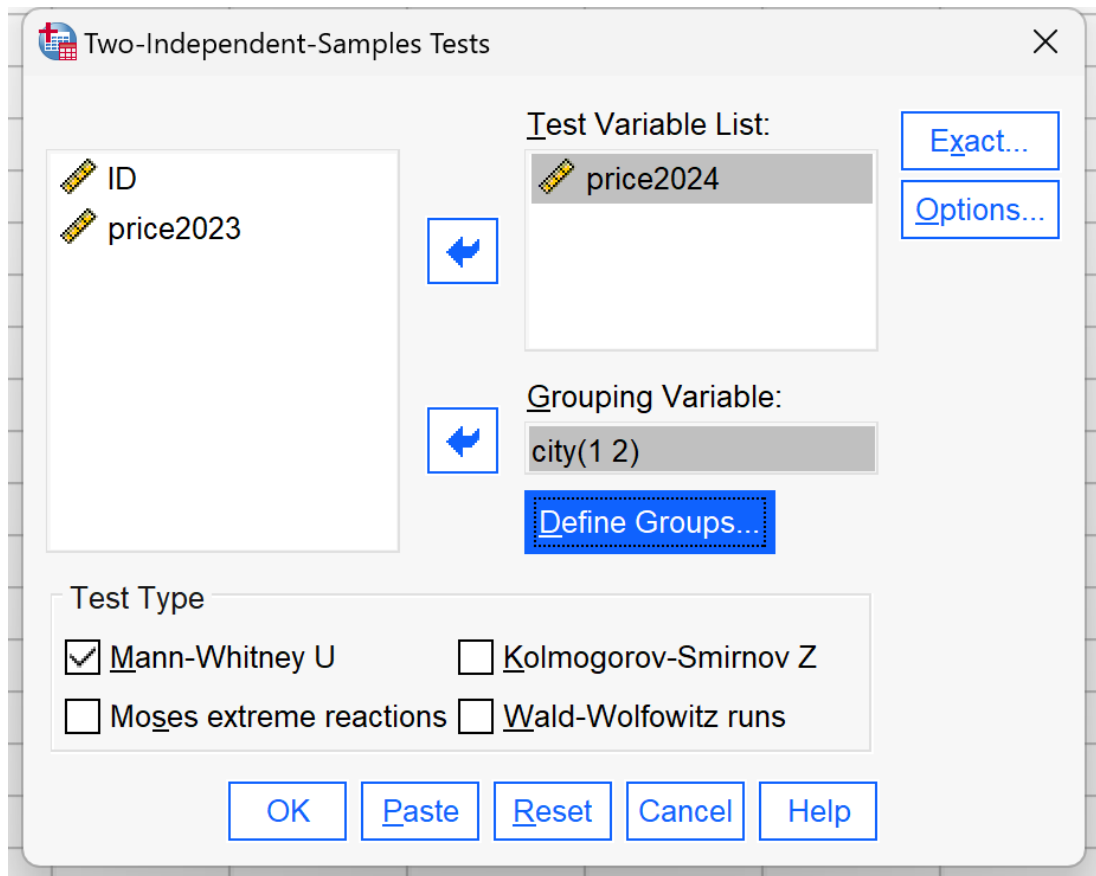


图 15 两组独立样本非参数检验的设置

点击“OK”按钮运行分析。SPSS 的输出窗口首先报告了一些秩（“Rank”）统计量，如图 16 所示。该表格，“Mean Rank”是平均秩，即将苏州和南京样本混合排序后，每组样本的平均排序；“Sum of Ranks”则是秩和。

	city	N	Mean Rank	Sum of Ranks
price2024	苏州	42	37.93	1593.00
	南京	51	54.47	2778.00
	Total	93		

图 16 M-W 检验的秩统计量结果

下面报告的是检验结果，如图 17 所示。其中，“Mann-Whitney U”和“Wilcoxon W”都是基于秩的更复杂的统计量，反映了一组数据相对于另一组数据，其排名是否通常更高或更低。“Z”值的角色相当于 T 检验中的 t 值，它反映了“Mann-Whitney U”统计量是否显著偏离了它在零假设下的期望值，而落在拒绝域中。实务中，上述统计量都不是重点，我们应当重点关注“Asymp. Sig. (2-tailed)”的值，即 P 值²。

	price2024
Mann-Whitney U	690.000
Wilcoxon W	1593.000
Z	-2.941
Asymp. Sig. (2-tailed)	.003

a. Grouping Variable: city

图 17 M-W 检验的检验结果

² 这里的“Asymp”是 Asymptotic 的缩写，表明该 P 值是用渐进方法（而非精确方法）估计的。渐进方法是统计学中一种常用的方法，其核心是：数据量越来越大时，可以用一些简化的数学方法估计结果，样本量越大越准确。一般而言，我们对此不需要特别关注，直接使用即可。

4. 配对样本的差异比较

本节将演示如何利用配对样本 T 检验和 Wilcoxon 非参数检验两种方法，判断苏州市同一房产 2023 年和 2024 年的房价是否存在显著差异。

4.1 配对样本 T 检验

在菜单栏中依次选择 “Analyze → Compare Means → Paired-Samples T Test”，进入配对样本 T 检验对话框。可以看到，“Paired Variables” 是一条待设定的表格，每一行是 2 个拟被比较的配对变量。将 “price2023” 和 “price2024” 选入该表格中，组成第 1 个 “Pair”，如图 18 所示。类似的，我们可以取消对 “Estimate effect sizes” 的勾选，因为这一结果并不是分析的重点。此外，我们不需要对 “city” 进行特别设置，因为所有的南京样本在 “price2023” 变量上均为缺失，因此它们都将被排除，我们所分析的自然就是苏州数据。

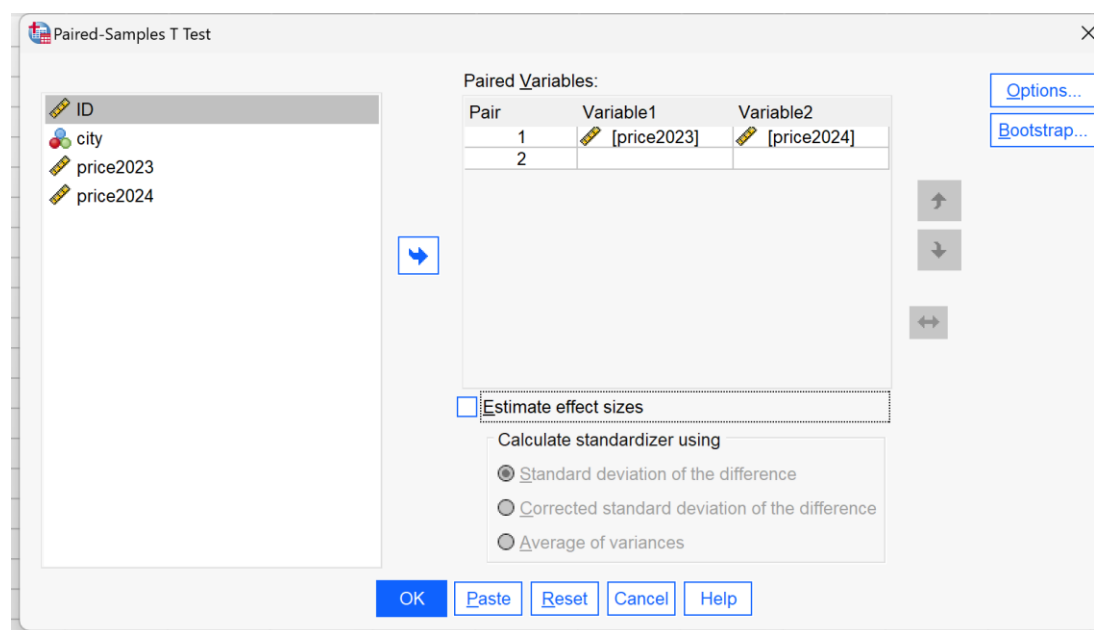


图 18 配对样本 T 检验的设置

点击 “OK” 键执行分析。SPSS 的输出窗口中首先报告了配对样本的基本统计量 (“Paired Samples Statistics”)，包括均值 (“Mean”)、样本量 (“N”)、标准差 (“Std. Deviation”)，以及均值的标准误 (“Std. Error Mean”)，如图 19 所示。

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	price2023	26331.40	42	10138.697	1564.435
	price2024	22814.60	42	8798.089	1357.575

图 19 配对样本 T 检验的基本统计量结果

下面报告的两个配对变量之间的相关性 (“Paired Samples Correlations”), 如图 20。我们尚未学习到此部分内容, 这里仅做简单的解释。表中的 “Correlation” 是相关系数, 它反映了两个变量之间的关系有多密切, 取值范围在 -1 到 1 之间; “Significance” 是相关系数的 P 值, 反映其是否具有统计显著性。本例中的相关系数为 0.961, 且在 0.01 的水平下高度显著, 提示苏州市同一房产 2023 年和 2024 年的房价具有高度正相关性, 即一个变量取值越高 (低), 另一个变量的取值也越高 (低)。这一结果当然是正常的, 但不是我们关注的重点。

Paired Samples Correlations

		N	Correlation	Significance	
				One-Sided p	Two-Sided p
Pair 1	price2023 & price2024	42	.961	<.001	<.001

图 20 配对样本 T 检验的相关性结果

接下来报告的则是重点内容——T 检验结果, 如图 21 所示。可以看到, “Pair1” 的行首名是 “price2023 - price2024”, 因此, 该检验相当于用 2023 年的房价减去 2024 年的房价, 再对这一差值进行分析检验。表格中报告了房价差值的均值 (“Mean”)、标准差 (“Std. Deviation”)、均值的标准误 (“Std. Error Mean”)、95 置信区间 (“95% Confidence Interval of the Difference”)。最后, 对房价差的均值是否等于 0 进行检验, “t” 为 T 值, 即检验统计量, “df” 是自由度, “Significance” 是检验的 P 值, 包括单侧检验的 P 值 (“One-Sided p”) 和双侧检验的 P 值 (“Two-Sided p”), 其区别与应用已在 3.1 节介绍。

		Paired Differences					Significance			
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	One-Sided p	Two-Sided p
					Lower	Upper				
Pair 1	price2023 - price2024	3516.810	2960.806	456.862	2594.158	4439.461	7.698	41	<.001	<.001

图 21 配对样本 T 检验的检验结果

4.2 非参数检验

最后，我们演示配对样本 T 检验的非参数版本——Wilcoxon 符号秩检验。在菜单栏中选择“Analyze → Nonparametric Tests → Legacy Dialogs → 2 Related Samples”，打开两组配对样本非参数检验对话框，如图 22 所示。可以看到，该对话框与配对样本 T 检验的对话框（图 18）具有很高的相似性。我们用相同的方式把“price2023”和“price2024”两个变量选入“Test Pairs”中，分别作为第 1 个“Pair”的“Variable1”和“Variable2”。

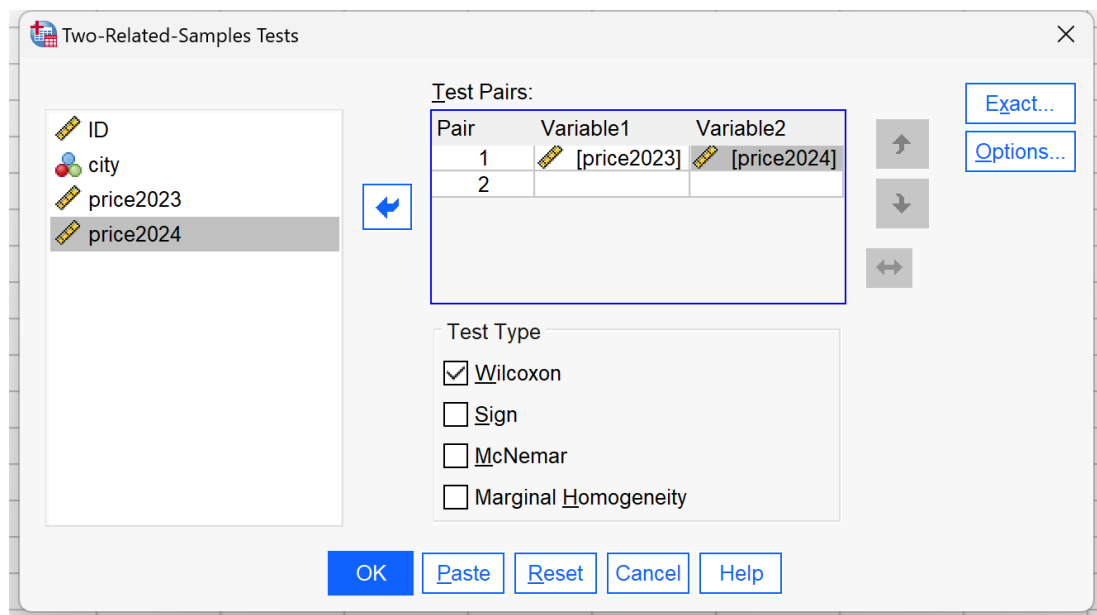


图 22 两组配对样本非参数检验的设置

确认“Test Type”中的“Wilcoxon”被选中，然后点击“OK”按钮，运行 Wilcoxon 符号秩检验。SPSS 的输出窗口中首先报告了一些简单的秩统计量（“Ranks”），如图 23 所示。可以看到，该表格报告了负秩（“Negative Ranks”，即“price2024 < price2023”）、正秩（“Positive Ranks”，即“price2024 > price2023”）、平秩（“Tie”，即“price2024 = price2023”）的样本量，以及它们的平均秩（“Mean Rank”）和秩和（“Sum of Ranks”）。

Ranks

		N	Mean Rank	Sum of Ranks
price2024 - price2023	Negative Ranks	41 ^a	21.00	861.00
	Positive Ranks	0 ^b	.00	.00
	Ties	1 ^c		
	Total	42		

a. price2024 < price2023

b. price2024 > price2023

c. price2024 = price2023

图 23 Wilcoxon 符号秩检验的秩统计量结果

接下来报告的是检验结果，如图 24 所示。表中的“Z”值是基于秩计算的较复杂的检验统计量，其角色相当于 T 检验中的 t 值；下方的“Asymp. Sig. (2-tailed)”是 P 值，是我们判断显著性的最直接的依据。

Test Statistics^a

		price2024 - price2023
Z		-5.579 ^b
Asymp. Sig. (2-tailed)		<.001

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

图 24 Wilcoxon 符号秩检验的检验结果