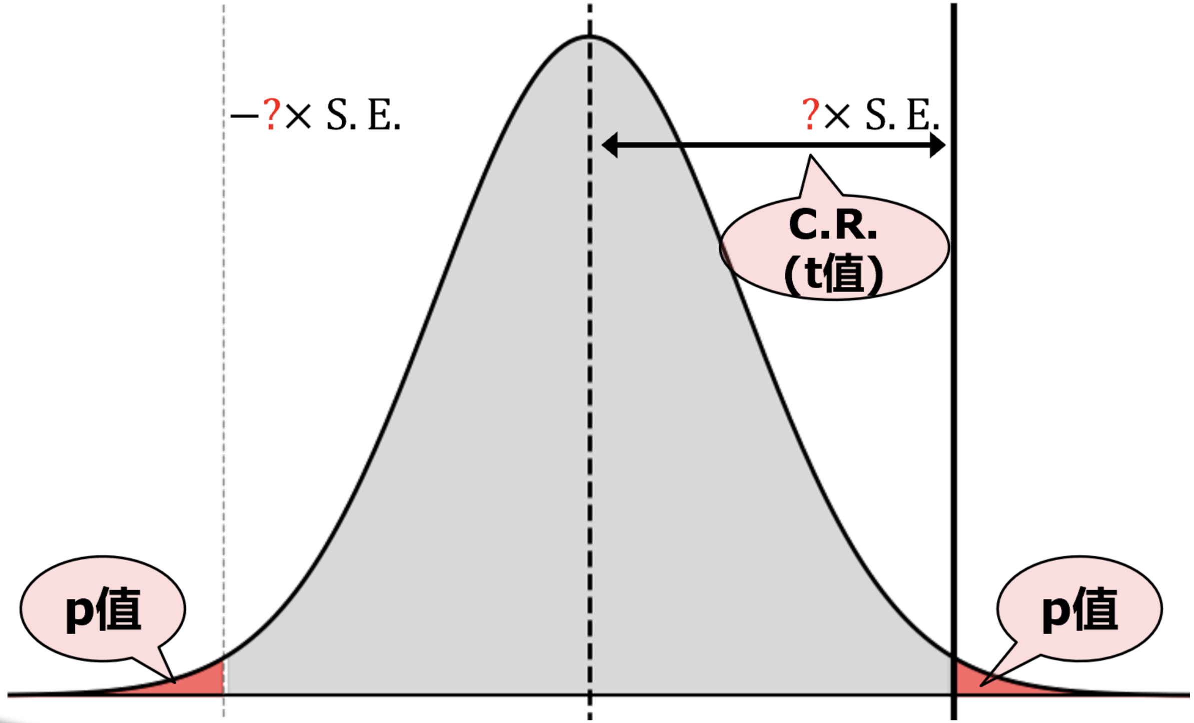


推断性统计

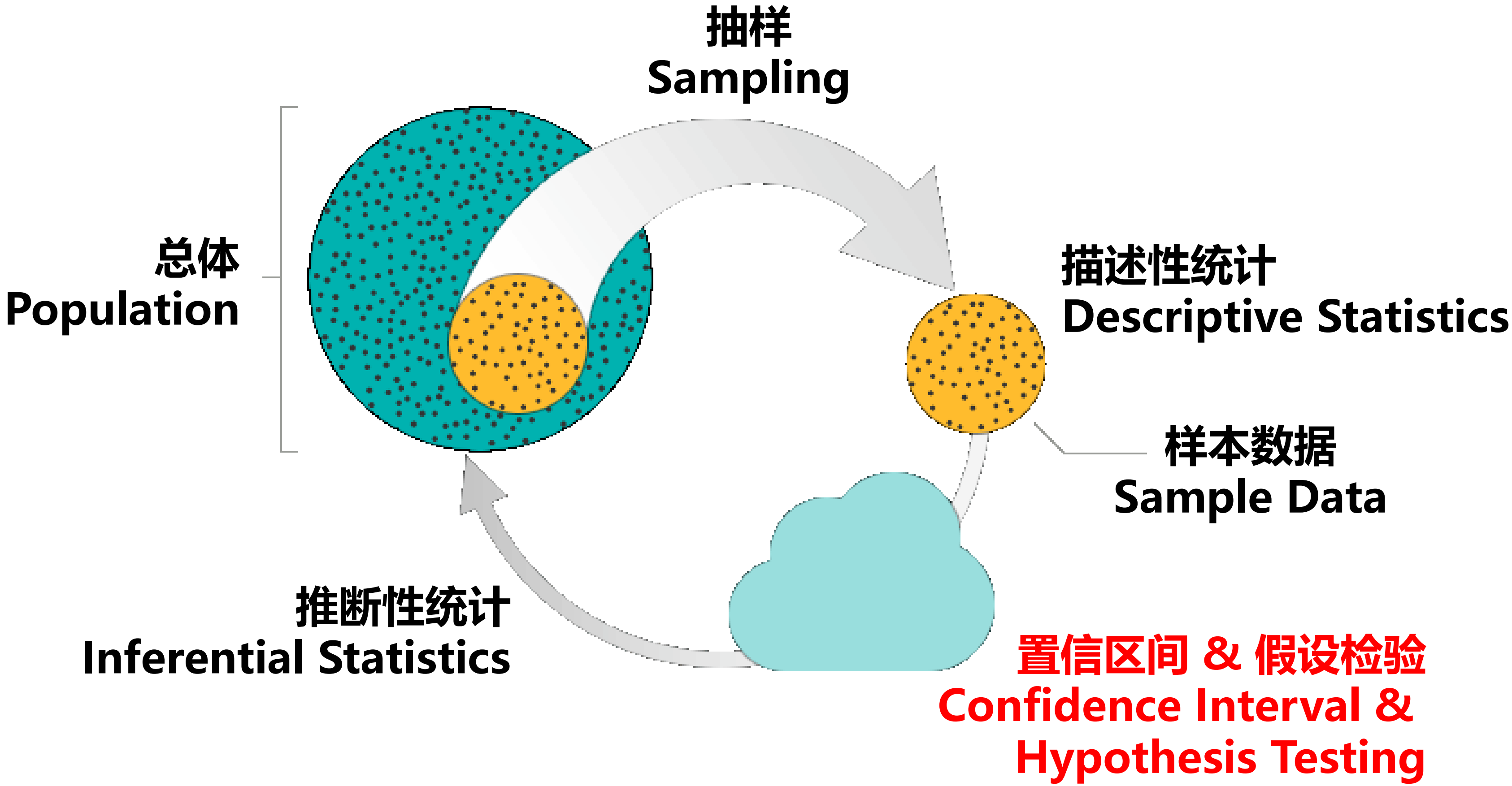
城市分析方法系列课程

苏州大学 王灿

学习目标

- 理解推断性统计的一般原理
- 掌握经典的假设检验方法

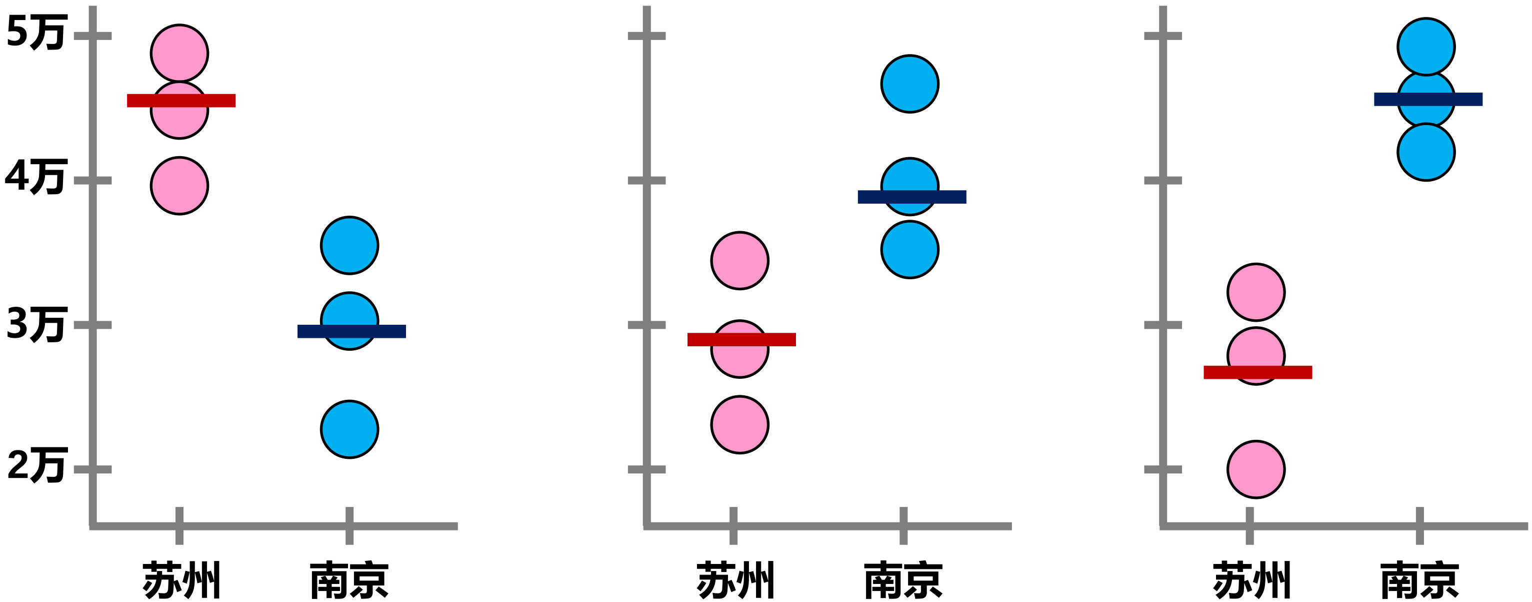
推断性统计



抽样
Sampling
总体
Population
描述性统计
Descriptive Statistics
样本数据
Sample Data
推断性统计
Inferential Statistics
置信区间 & 假设检验
Confidence Interval &
Hypothesis Testing

推断性统计

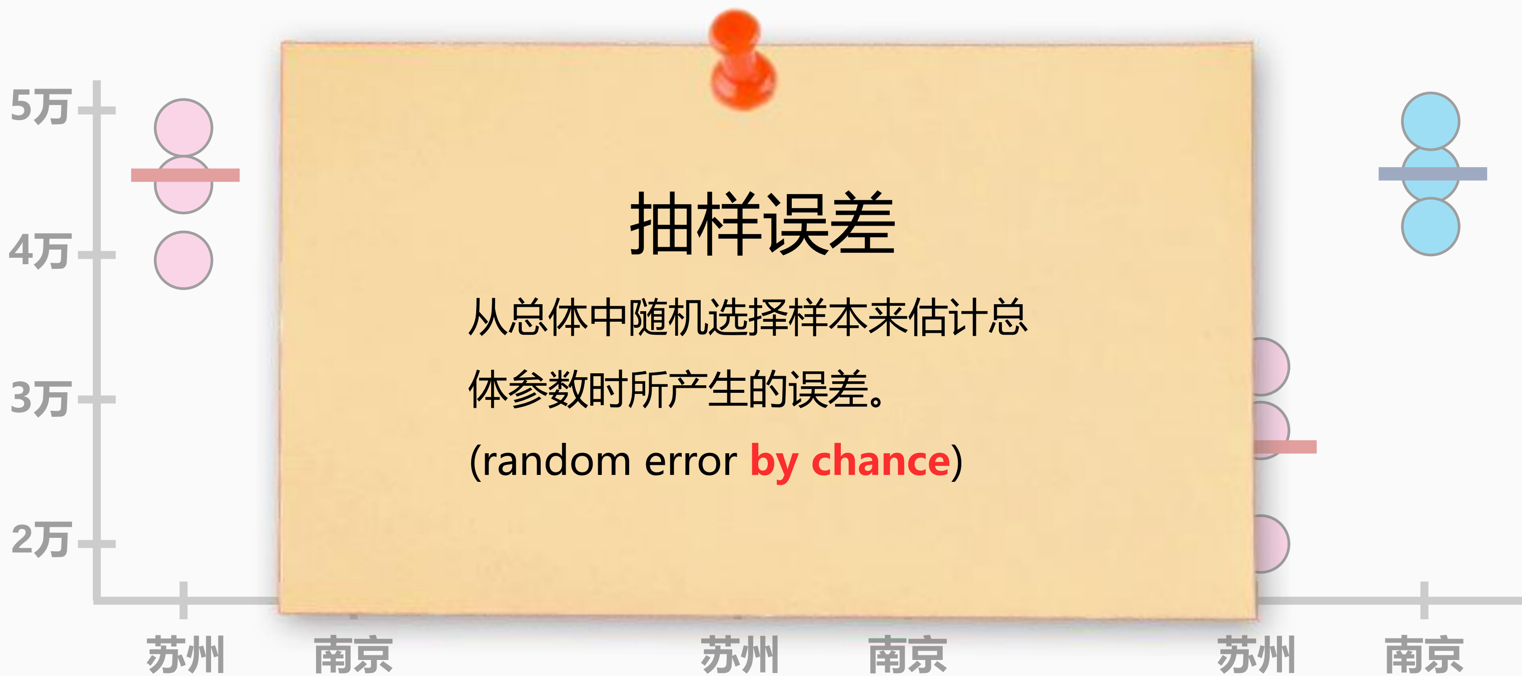
- 问题：南京与苏州，哪个城市的房价更高？



苏州房价高15000？　　南京房价高9000？　　南京房价高12000？

推断性统计

- 问题：南京与苏州，哪个城市的房价更高?



抽样误差

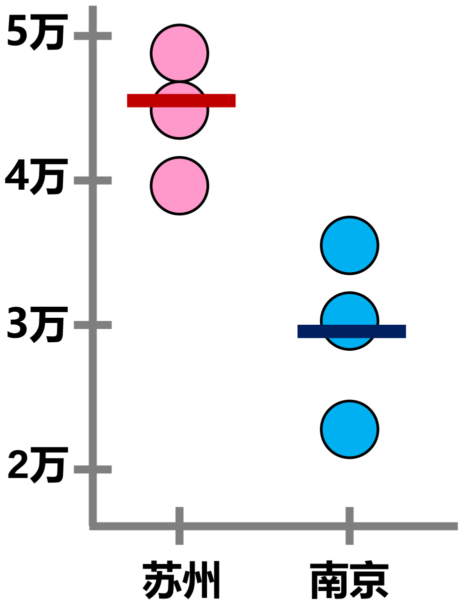
从总体中随机选择样本来估计总体参数时所产生的误差。

(random error **by chance**)

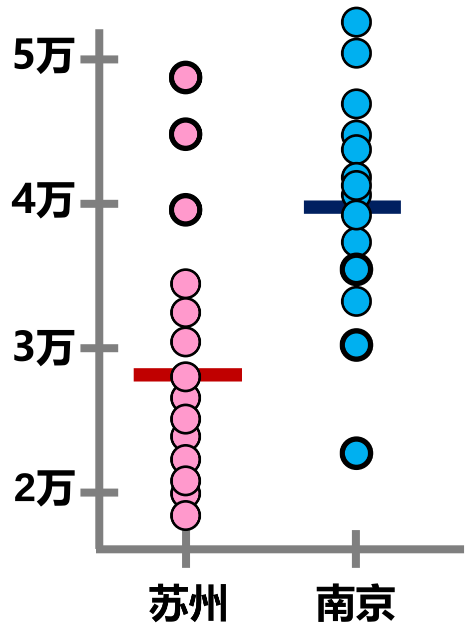
苏州房价高15000?　　南京房价高9000?　　南京房价高12000?

推断性统计

- 问题：南京与苏州，哪个城市的房价更高?



苏州房价高15000?



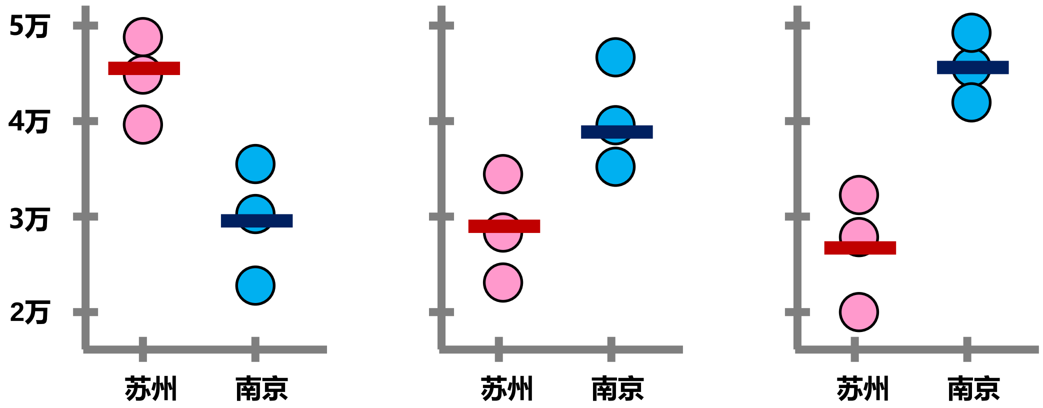
南京房价高11000?

更大的样本量能够提升统计推断的信心。

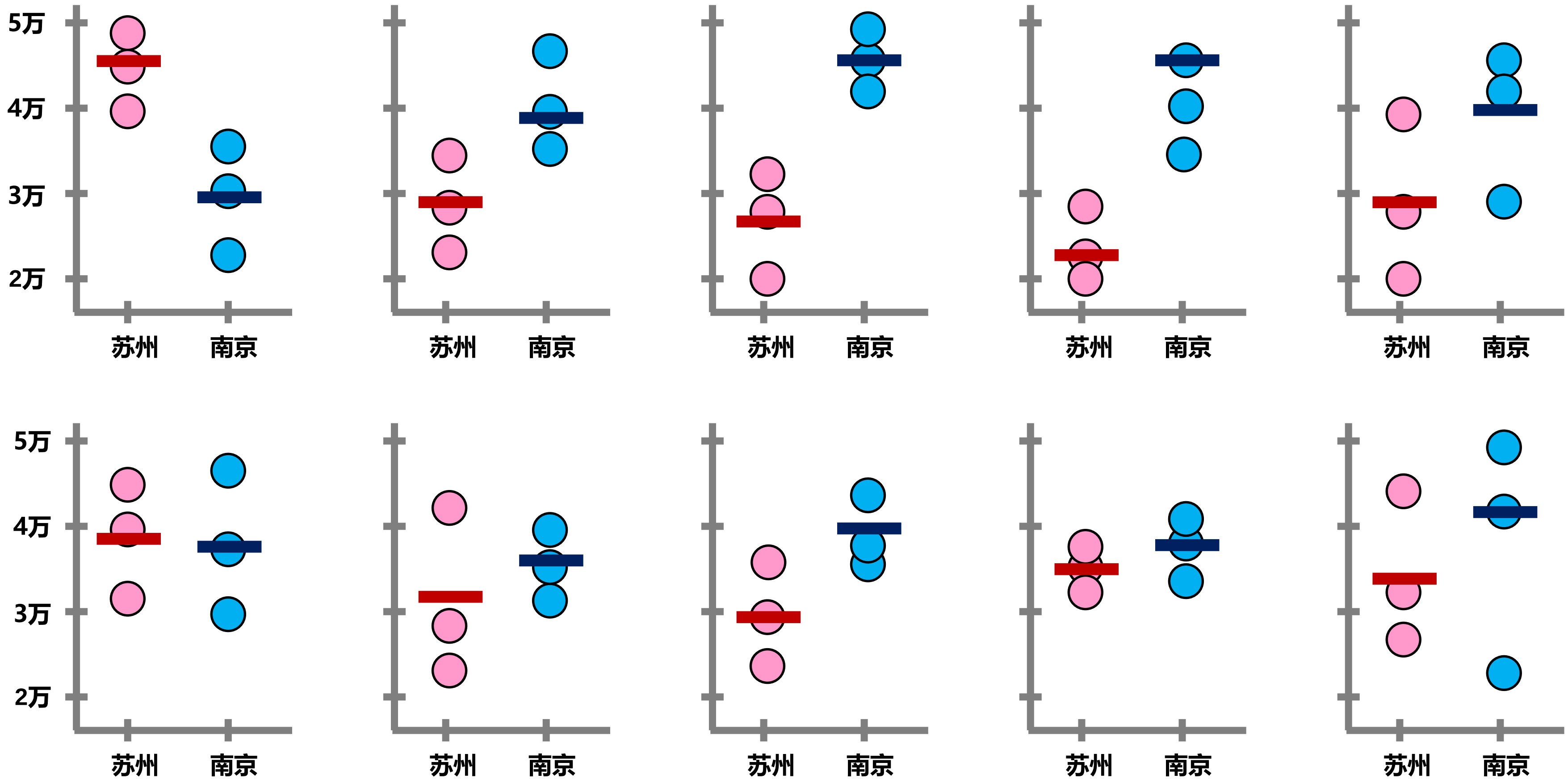
置信区间 (Confidence Interval, C.I.)

- **点估计**：用单个数值估计总体参数，是基于样本的最优猜测；
  - 南京与苏州的房价差：+11000, +9000, -15000, …
- **区间估计**：用一个范围估计总体参数，该范围称为置信区间。
  - 南京与苏州的房价差：95%的可能落在 (2000, 20000) 之间

推断性统计

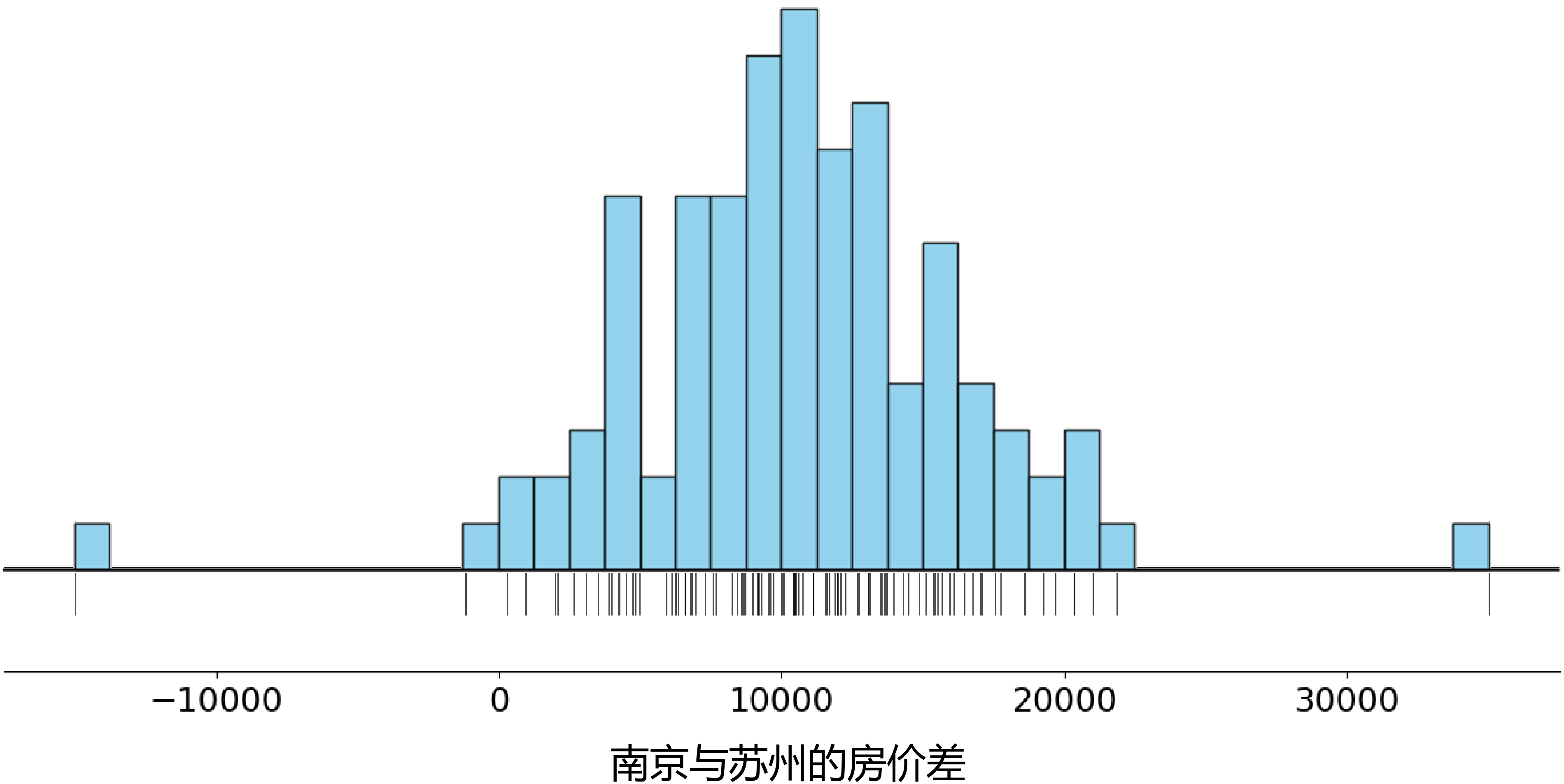


推断性统计

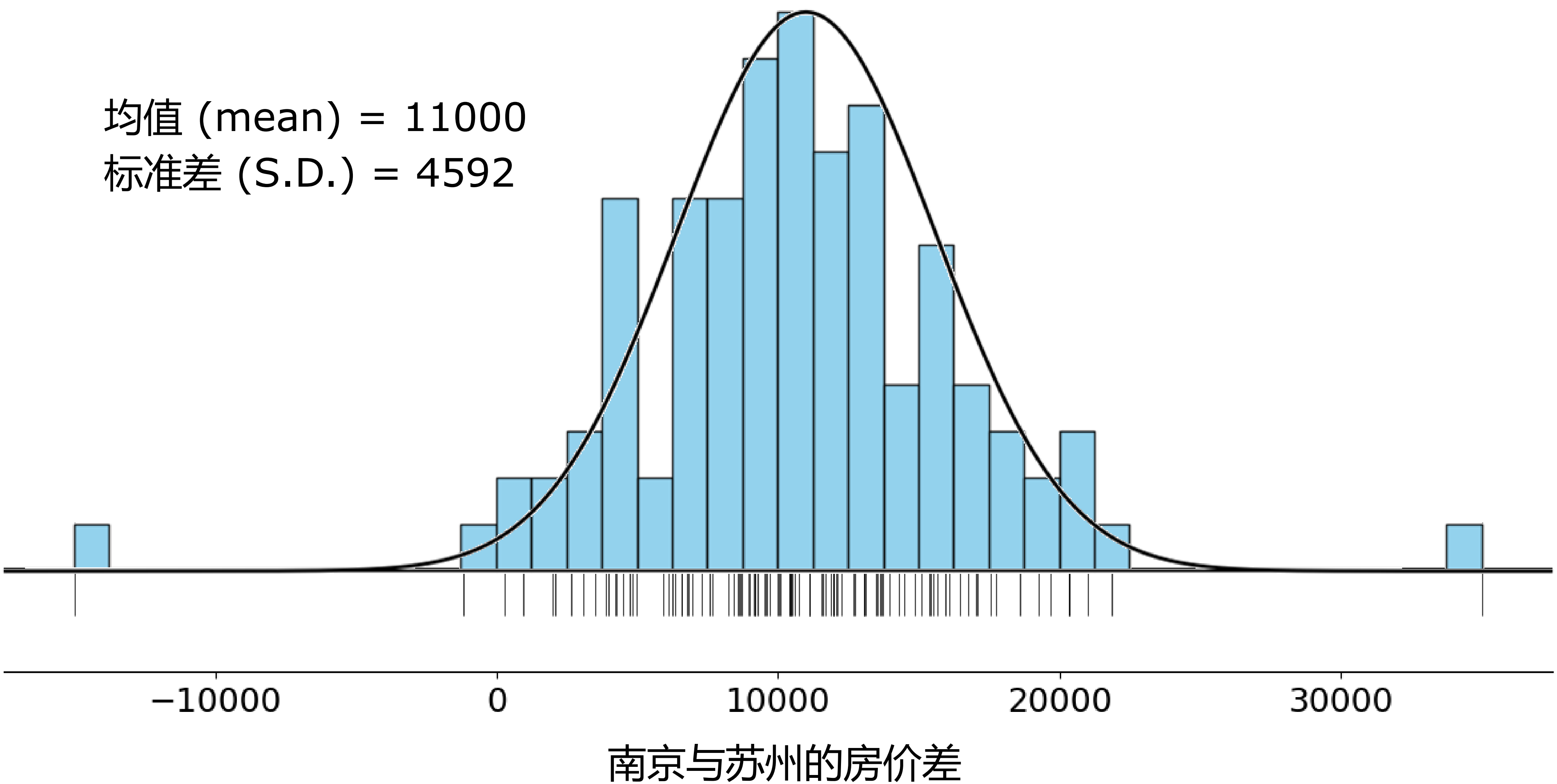


推断性统计

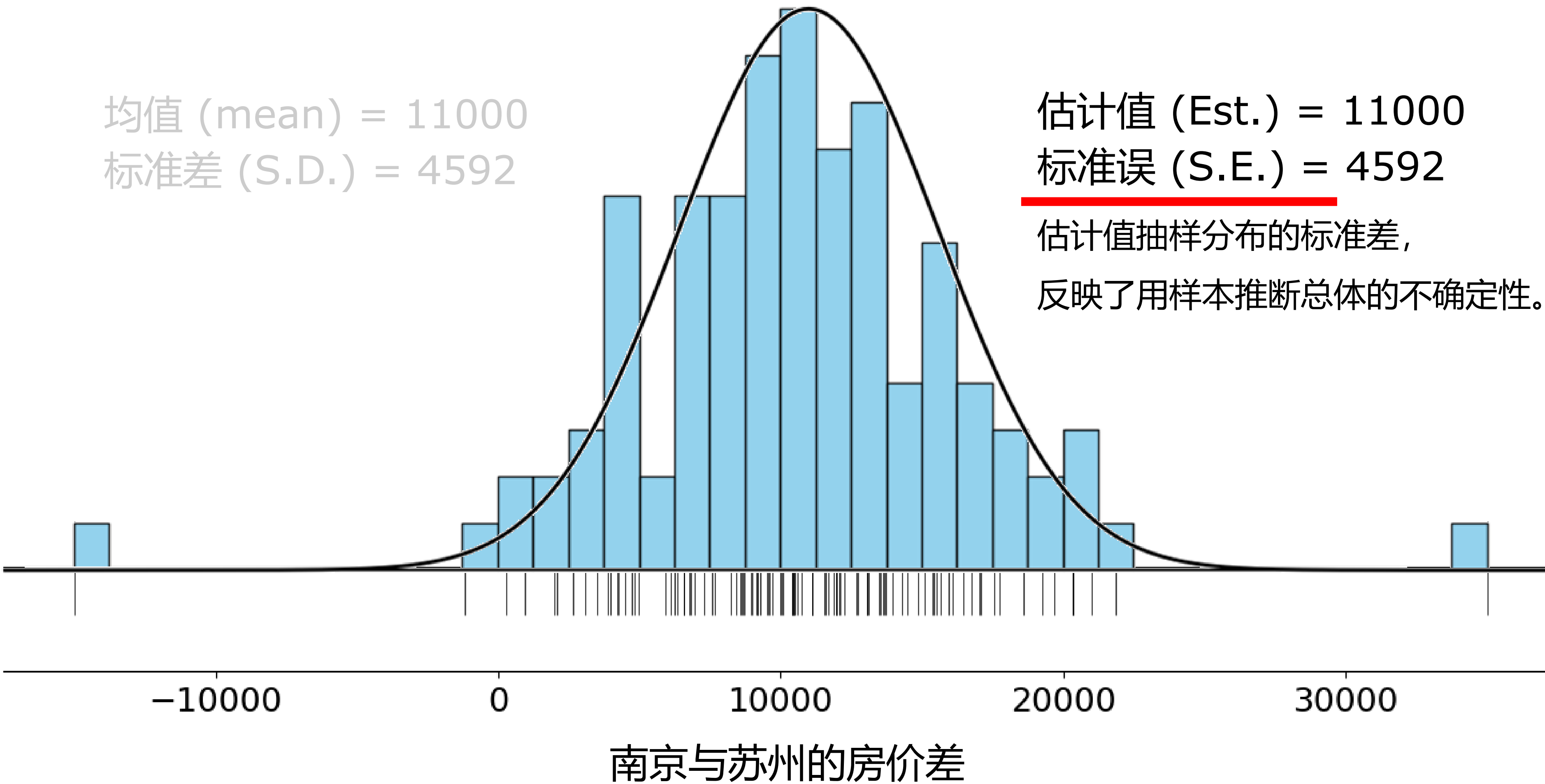
推断性统计



推断性统计

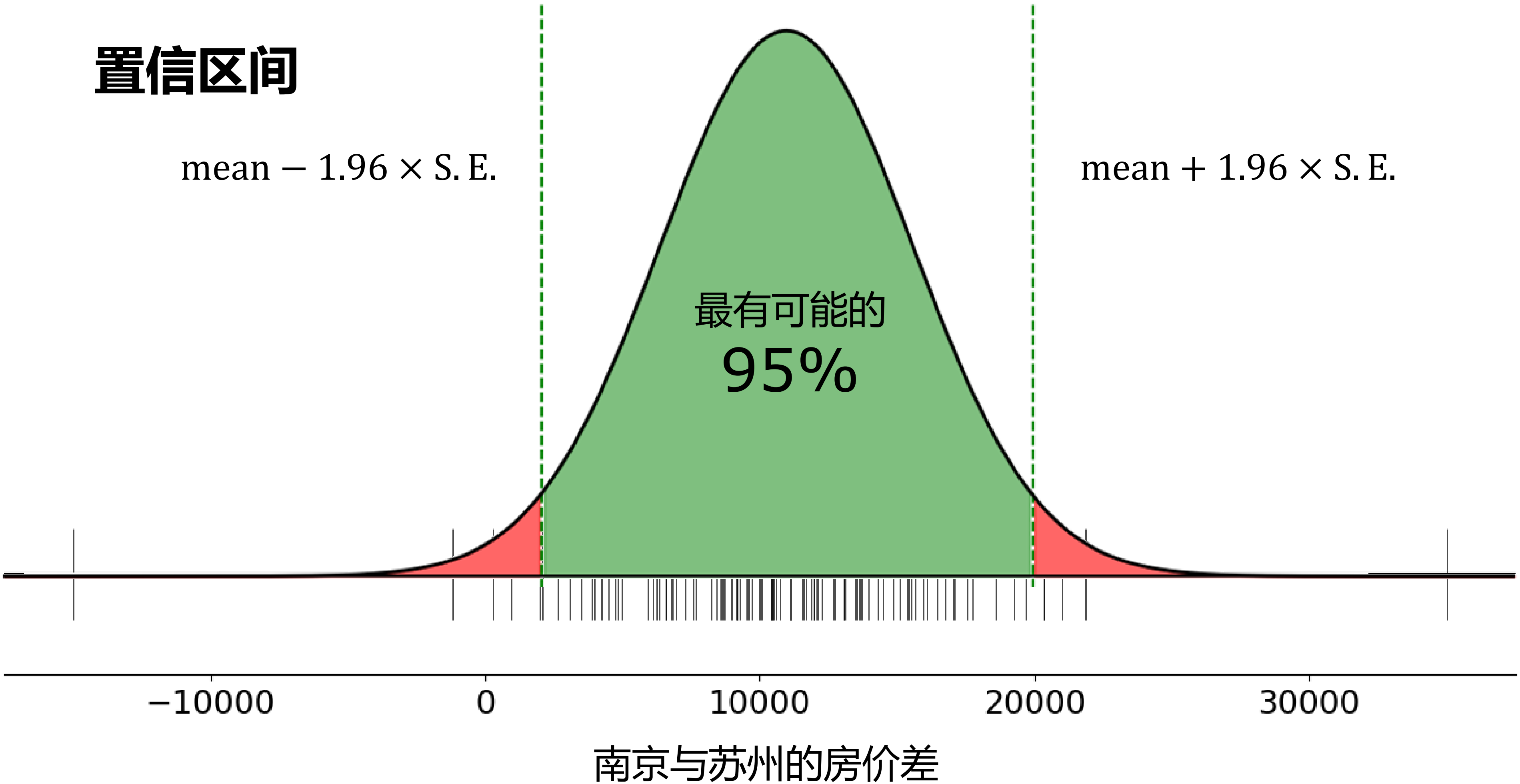

均值 (mean) = 11000
标准差 (S.D.) = 4592
−10000
0
10000
20000
30000
南京与苏州的房价差

推断性统计


均值 (mean) = 11000
标准差 (S.D.) = 4592
估计值 (Est.) = 11000
标准误 (S.E.) = 4592
估计值抽样分布的标准差，
反映了用样本推断总体的不确定性。
−10000
0
10000
20000
30000
南京与苏州的房价差

推断性统计

置信区间

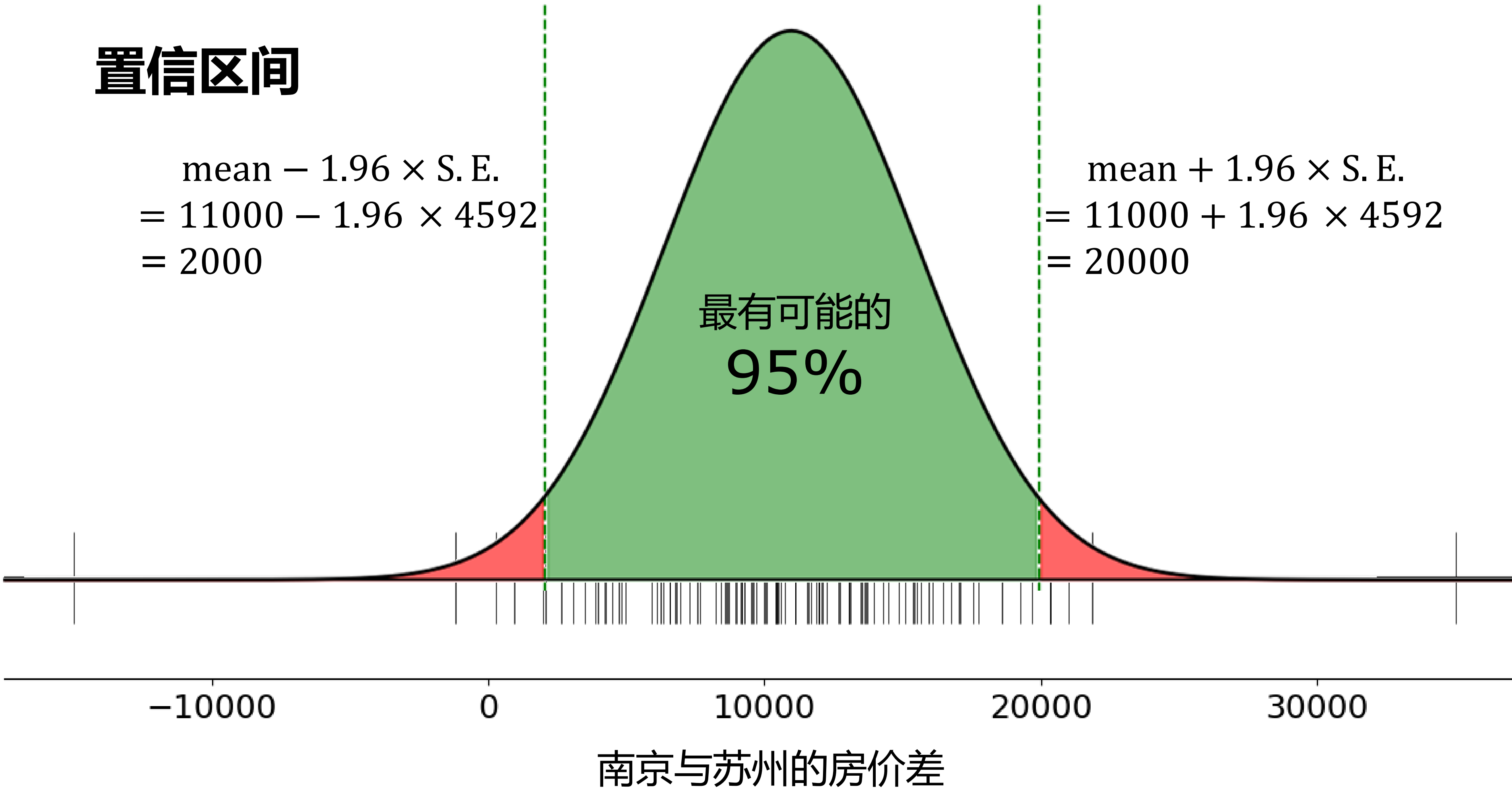


mean − 1.96 × S. E.
mean + 1.96 × S. E.
最有可能的
95%
−10000
0
10000
20000
30000
南京与苏州的房价差

置信区间

$$\begin{aligned}&\text{mean} - 1.96 \times \text{S.E.}\\ &= 11000 - 1.96 \times 4592\\ &= 2000\end{aligned}$$

$$\begin{aligned}&\text{mean} + 1.96 \times \text{S.E.}\\ &= 11000 + 1.96 \times 4592\\ &= 20000\end{aligned}$$

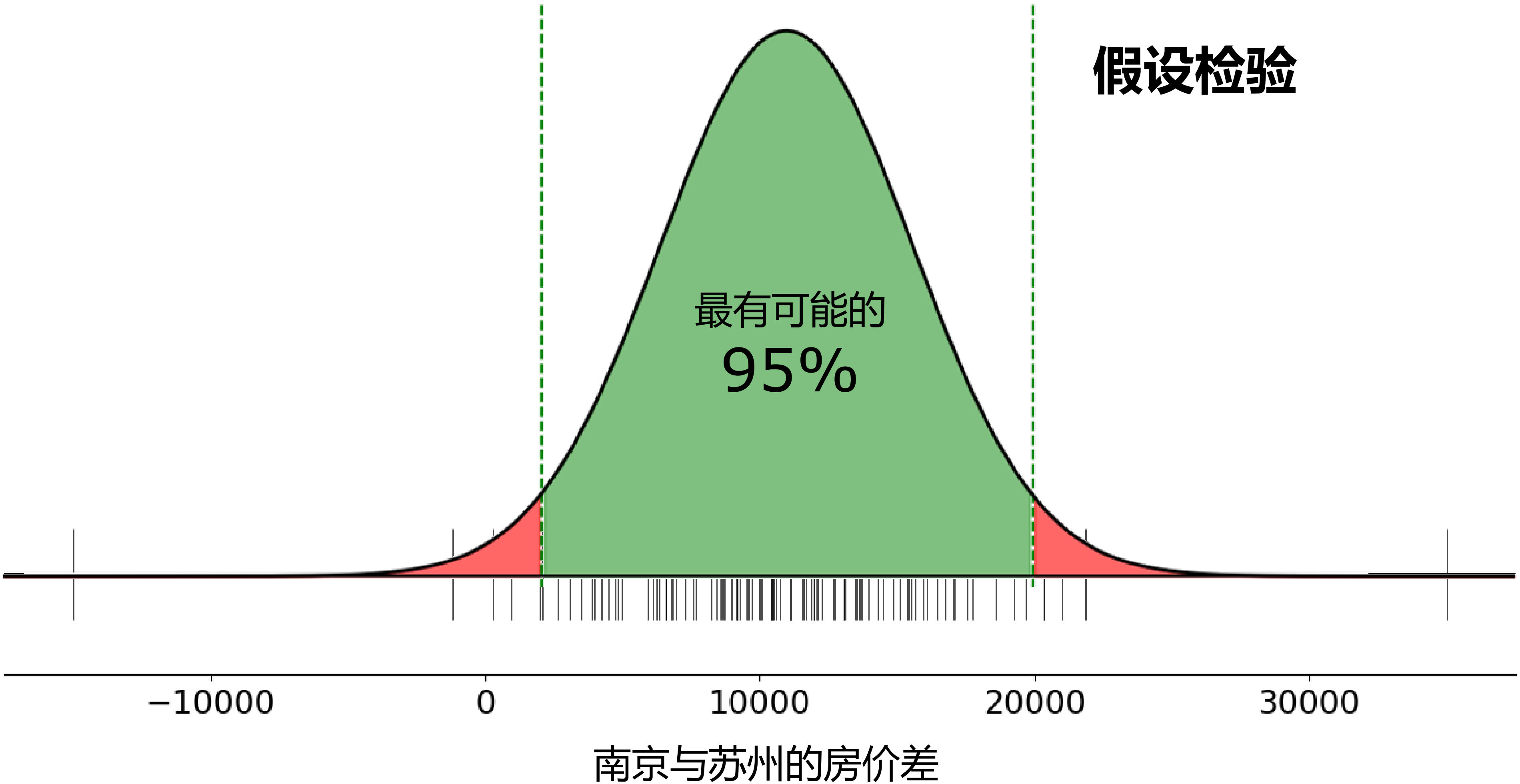


推断性统计

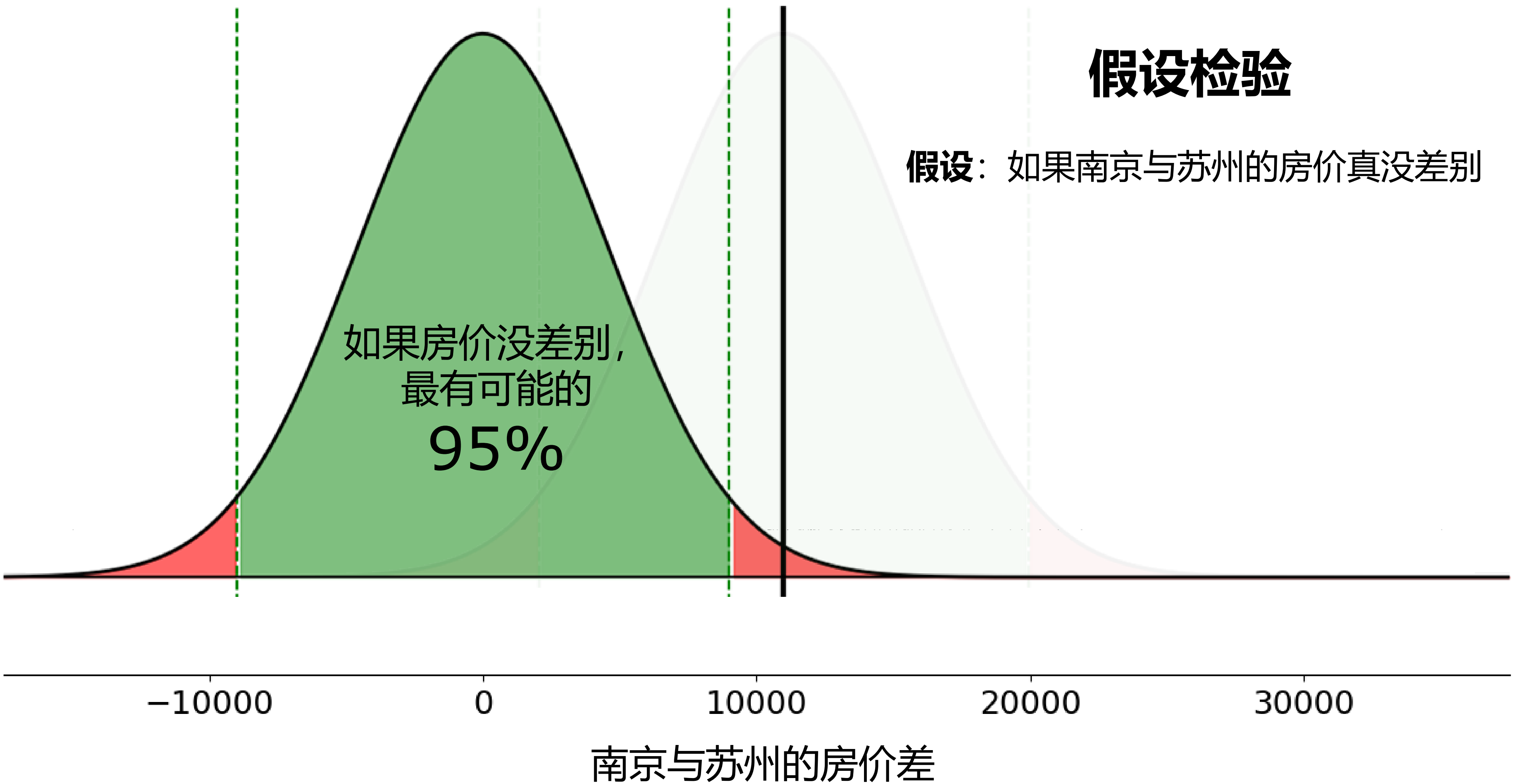
大概差不多是对的（Probably approximately correct, PAC）

- 样本中，南京房价平均比苏州高11000，那总体的真实情况呢?
  - 其实是高10000： 误差只有**10%**，这种情况还是**很有可能**的。
  - 其实是低15000： 误差高达**170%**，这种情况是**不太可能**的。
- **95%**置信区间：**2000 ~ 20000**
  - 无法绝对精确地估计，只能尽可能精确。
- 假设检验：结果真的有意义吗？是否只是抽样误差 (by chance)?

推断性统计



推断性统计

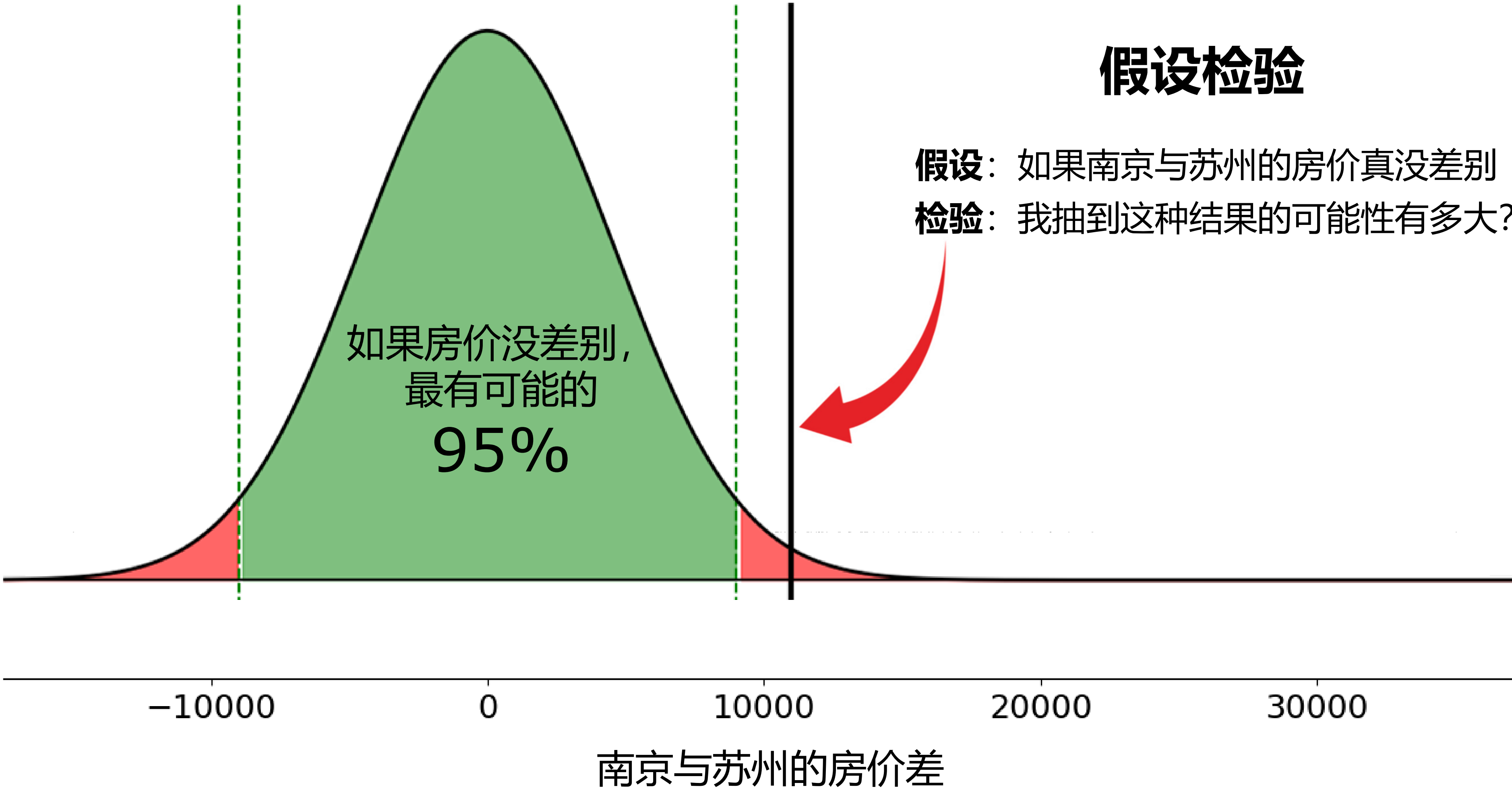

假设检验
假设：如果南京与苏州的房价真没差别
如果房价没差别，
最有可能的
95%
−10000
0
10000
20000
30000
南京与苏州的房价差

推断性统计

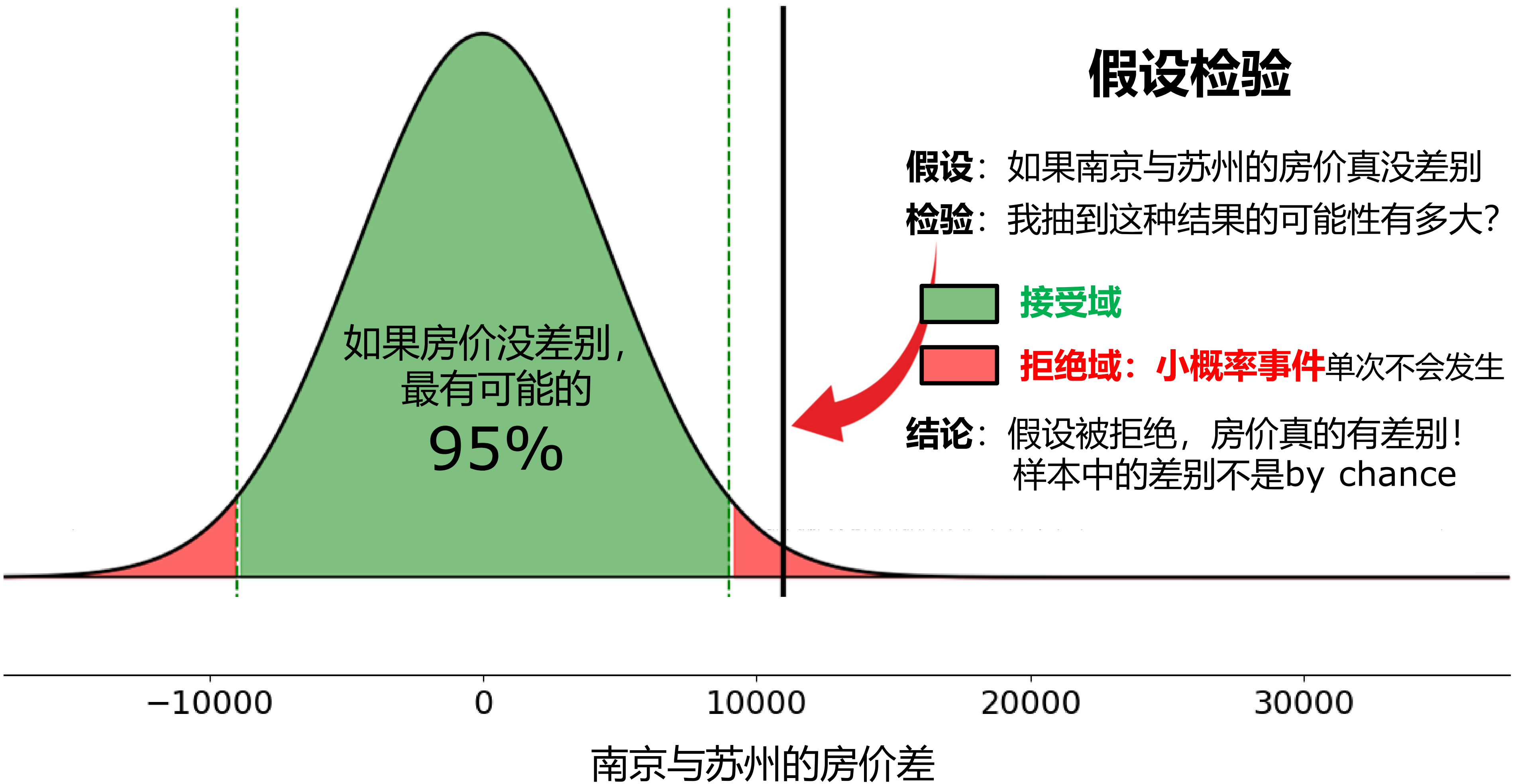
假设检验

假设：如果南京与苏州的房价真没差别

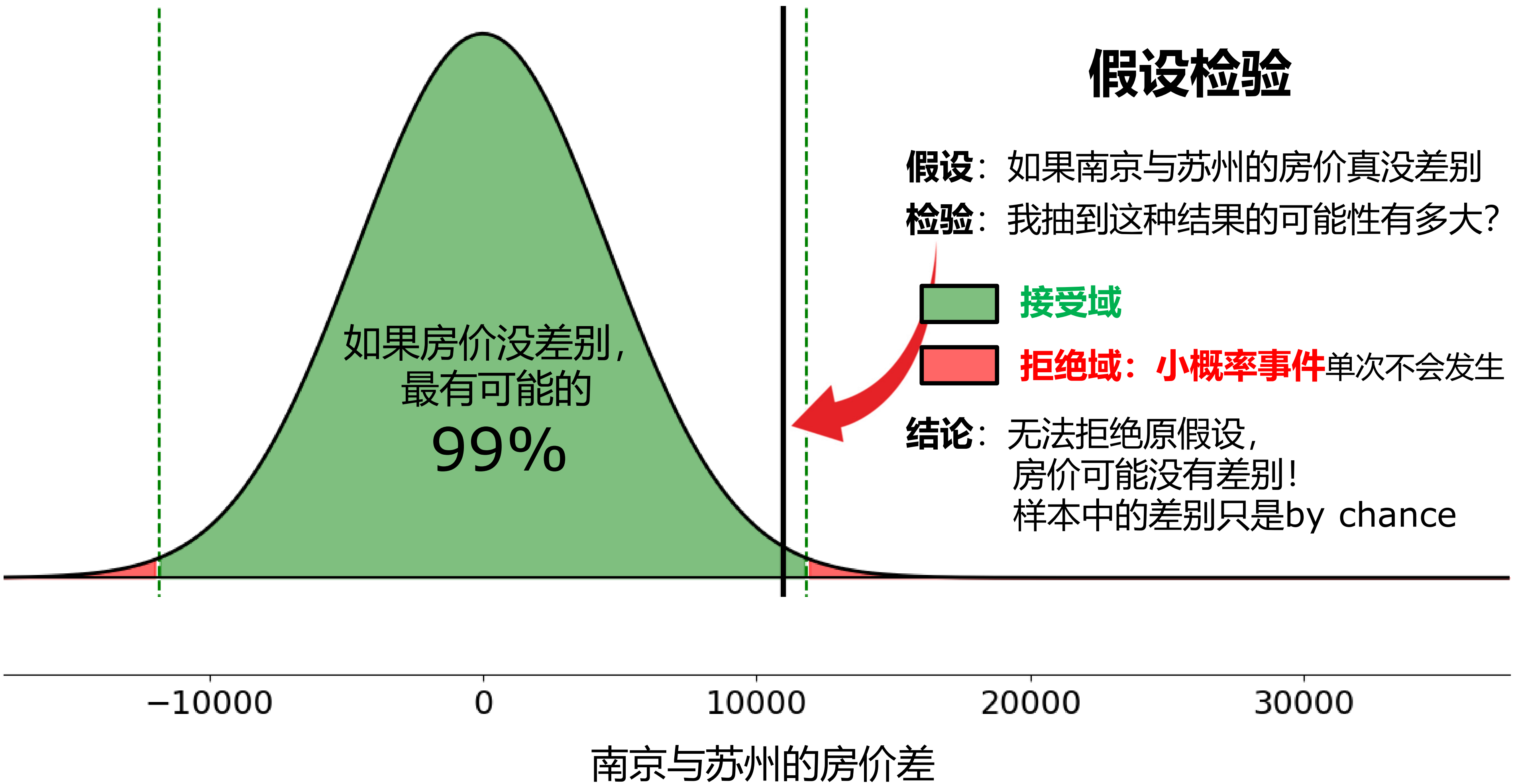
检验：我抽到这种结果的可能性有多大?



推断性统计

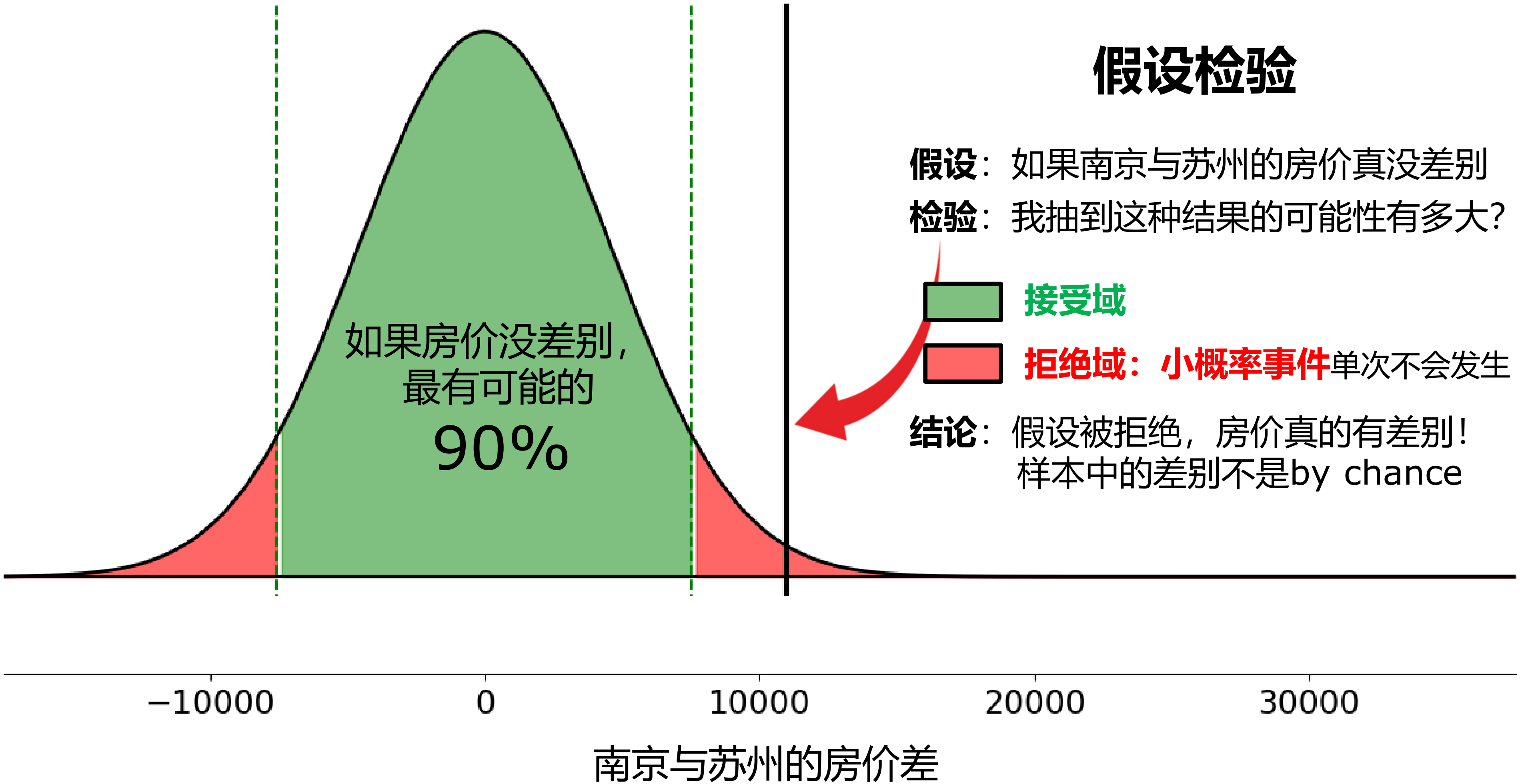

假设检验
假设：如果南京与苏州的房价真没差别
检验：我抽到这种结果的可能性有多大?
接受域
拒绝域：小概率事件单次不会发生
结论：假设被拒绝，房价真的有差别！
样本中的差别不是by chance
如果房价没差别，
最有可能的
95%
−10000
0
10000
20000
30000
南京与苏州的房价差

推断性统计



如果房价没差别，
最有可能的
99%
假设检验
假设：如果南京与苏州的房价真没差别
检验：我抽到这种结果的可能性有多大?
接受域
拒绝域：小概率事件单次不会发生
结论：无法拒绝原假设，
房价可能没有差别！
样本中的差别只是by chance
−10000
0
10000
20000
30000
南京与苏州的房价差

推断性统计



假设检验

假设：如果南京与苏州的房价真没差别

检验：我抽到这种结果的可能性有多大?

接受域

拒绝域：小概率事件单次不会发生

结论：假设被拒绝，房价真的有差别！
样本中的差别不是by chance

假设检验 (Hypothesis Testing)

- **零假设** (或原假设，null hypothesis)：假设检验的默认状态，通常表示没有效应、没有差异或没有关联。

“XX”与“XX”之间**没有显著差异**

假设检验（Hypothesis Testing）

- **零假设** (或原假设，null hypothesis)：假设检验的默认状态，通常表示没有效应、没有差异或没有关联。

“XX”与“XX”之间**没有显著差异**

“南京房价”与“苏州房价”之间没有显著差异
“南京、苏州的房价差”与“0”之间没有显著差异
“房价”对“幸福感”没有显著影响，or，
“房价对幸福感的影响系数”与“0”没有显著差异

假设检验（Hypothesis Testing）

- **零假设** (或原假设，null hypothesis)：假设检验的默认状态，通常表示没有效应、没有差异或没有关联。

“XX”与“XX”之间**没有显著差异**

“南京房价”与“苏州房价”之间没有显著差异

“南京、苏州的房价差”与“[illegible]”之间没有显著差异

“房价”与“幸福感”没有显著影响

大部分情况希望拒绝

假设检验（Hypothesis Testing）

- **零假设** (或原假设，null hypothesis)：假设检验的默认状态，通常表示没有效应、没有差异或没有关联。

“XX”与“XX”之间**没有显著差异**

“南京房价”与“苏州房价”之间没有显著差异

“南京、苏州的房价差” 之间没有显著差异

“房价” “幸福感”没有显著影响

大部分情况希望拒绝

“南京房价的方差”与“苏州房价的方差”之间没有显著差异（方差齐性检验）

“房价的分布”与“正态分布”之间没有显著差异（正态性检验）

假设检验（Hypothesis Testing）

- **零假设** (或原假设，null hypothesis)：假设检验的默认状态，通常表示没有效应、没有差异或没有关联。

“XX”与“XX”之间**没有显著差异**

“南京房价”与“苏州房价”之间没有显著差异

“南京、苏州的房价差”与“0”之间没有显著差异

“房价”对“幸福感”没有显著影响

大部分情况希望拒绝

“南京房价的方差”与“苏州房价的方差”之间没有显著差异（方差齐性检验）

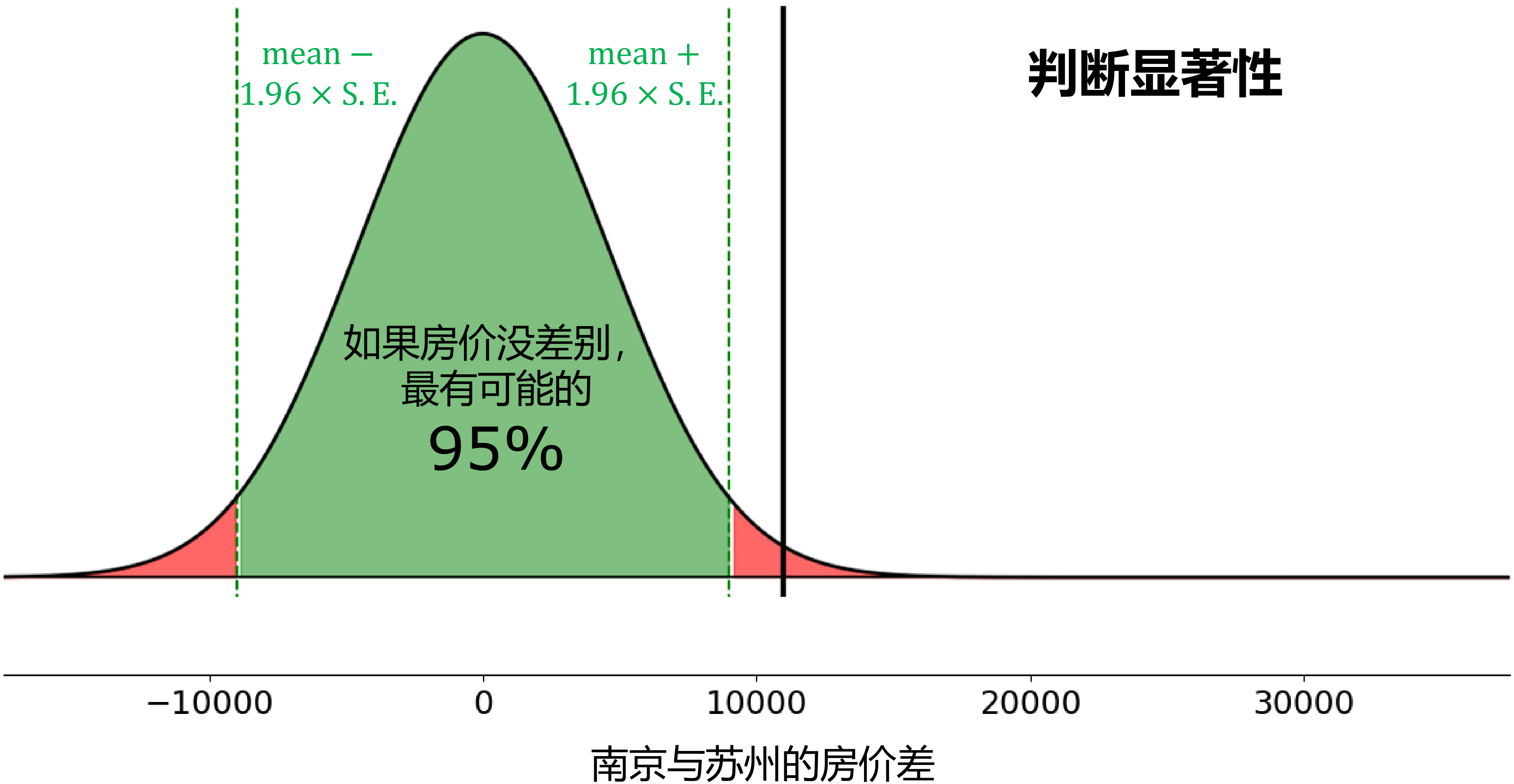
“房价的分布”与“正态分布”之间没有显著差异（正态性检验）

有些情况希望无法拒绝

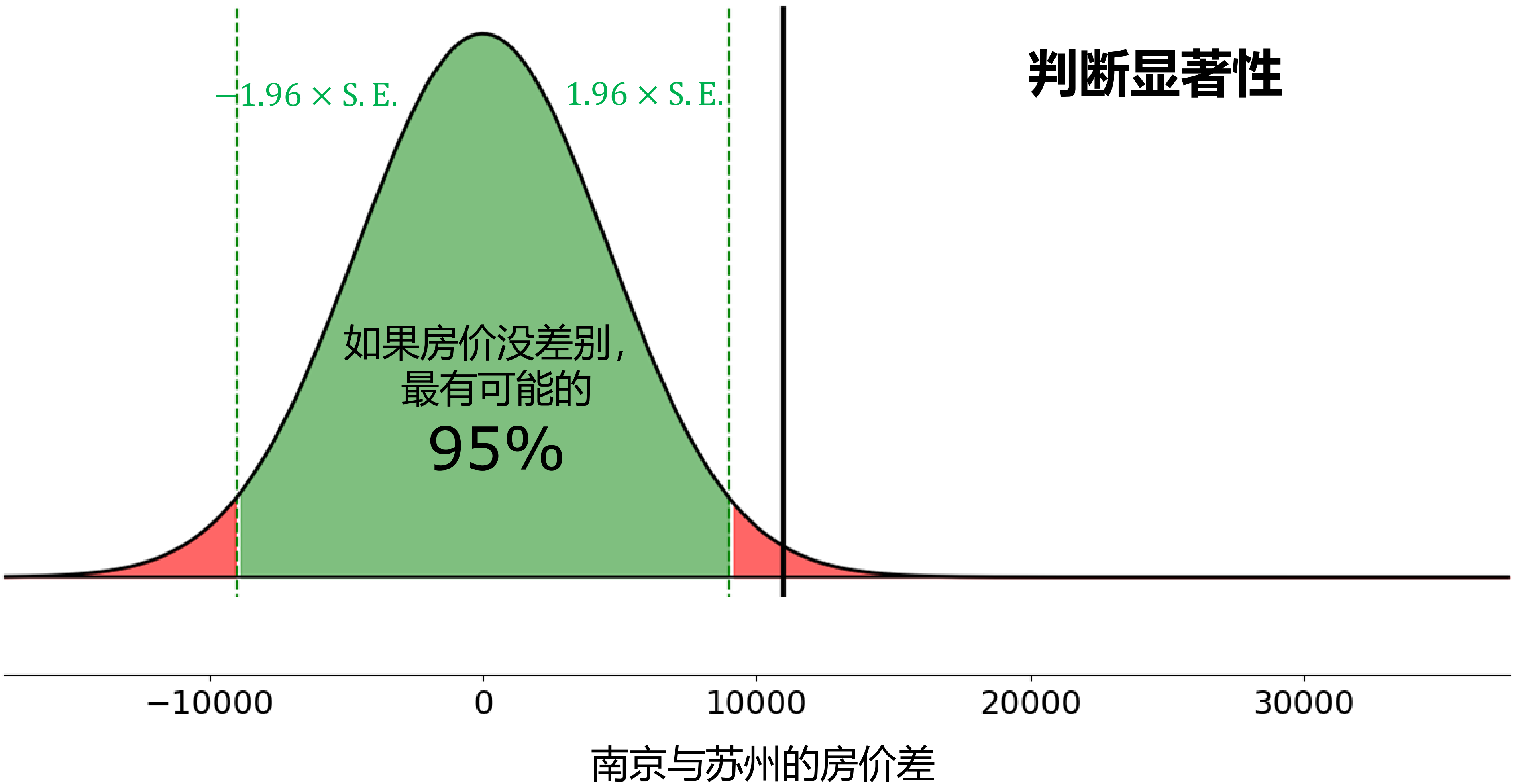
假设检验（Hypothesis Testing）

- **统计显著** (statistical significance)：零假设被拒绝，认为样本中的结果不太可能仅仅由偶然因素产生的，观察到的效应是真实存在的。
- **无法拒绝 ≠ 接受?** 证剧不足≠无罪！
- **差异统计显著 ≠ 差异很大?** 比如，房价显著高1块钱，说明效应**确实存在**，**但很小**。
- **显著性水平α** (significance level)：其实真没效应，但误以为有效应的概率，即“拒绝零假设”的判断其实是错误的概率。
  - **α=0.05（最常用）**：在5%的水平下（统计）显著，常标两颗星（**）
  - α=0.01（更严格）：在1%的水平下（统计）显著，常标一颗星（***）
  - α=0.1（最宽松）：在10%的水平下（统计）显著，常标三颗星（*）

推断性统计

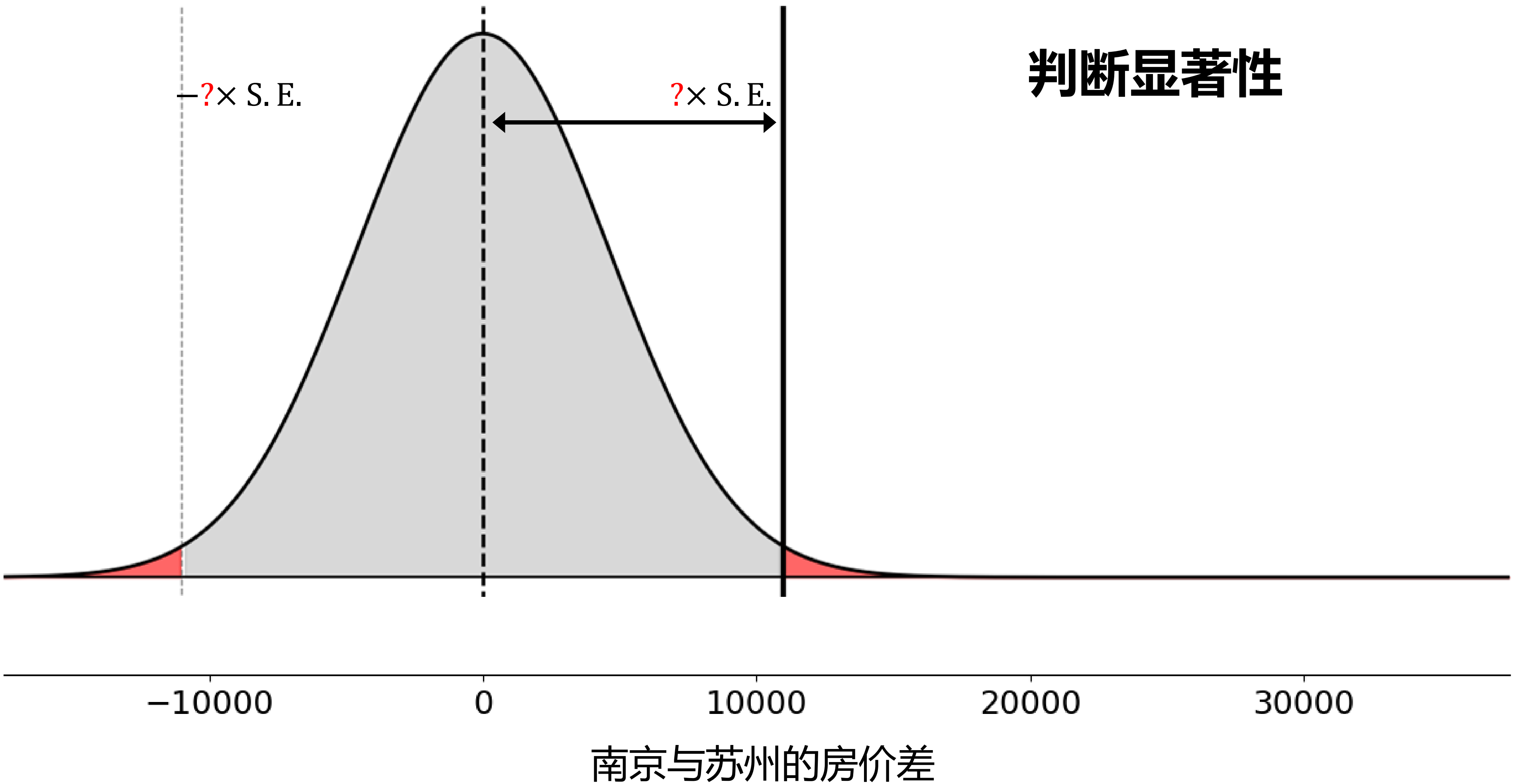

mean −
1.96 × S. E.
mean +
1.96 × S. E.
判断显著性
如果房价没差别，
最有可能的
95%
−10000
0
10000
20000
30000
南京与苏州的房价差

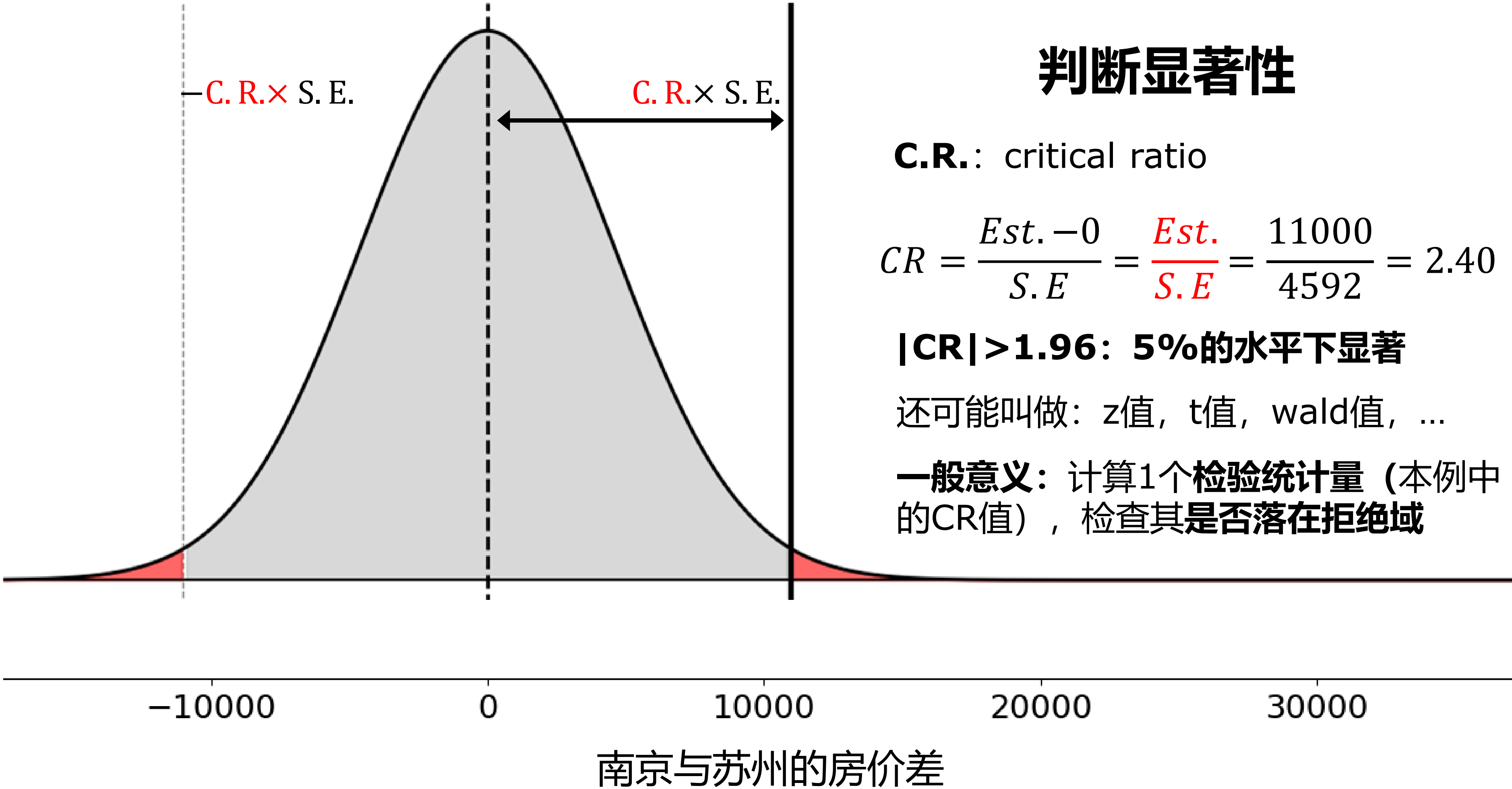
推断性统计



$-1.96 \times S.E.$
$1.96 \times S.E.$
判断显著性
如果房价没差别，
最有可能的
95%
−10000
0
10000
20000
30000
南京与苏州的房价差

推断性统计





判断显著性

C.R.： critical ratio

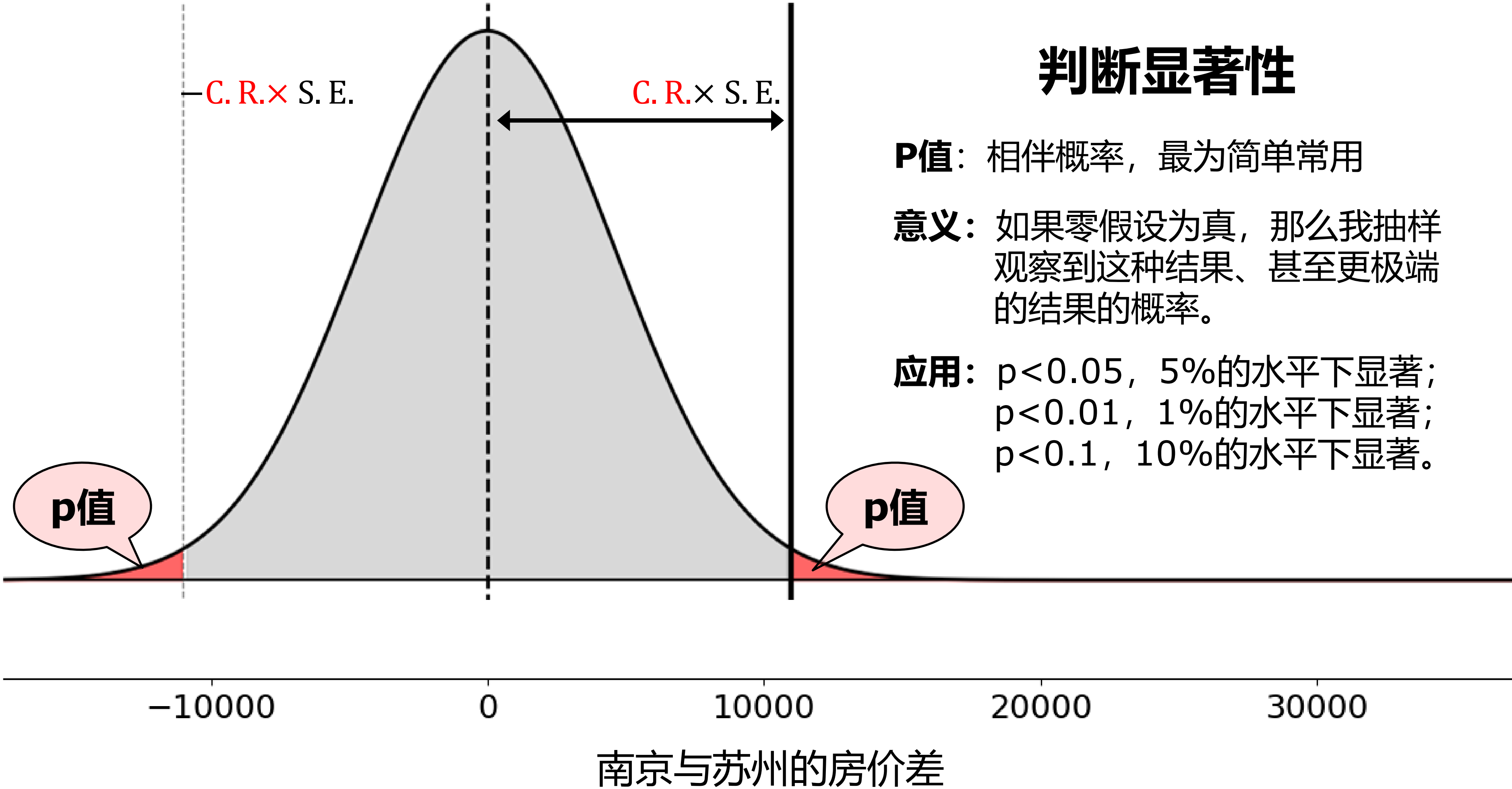
$$CR=\frac{Est.-0}{S.E}=\frac{Est.}{S.E}=\frac{11000}{4592}=2.40$$

|CR|>1.96：5%的水平下显著

还可能叫做：z值，t值，wald值，...

一般意义：计算1个**检验统计量**（本例中的CR值），检查其**是否落在拒绝域**

推断性统计



判断显著性

P值：相伴概率，最为简单常用

意义： 如果零假设为真，那么我抽样观察到这种结果、甚至更极端的结果的概率。

应用： $p<0.05$，5%的水平下显著；
$p<0.01$，1%的水平下显著；
$p<0.1$，10%的水平下显著。

小结

- **推断性统计**：考虑了抽样误差，利用样本数据来估计、分析总体的特征。
- **点估计**：对总体参数的单一值估计；**置信区间**：估计了一个范围。
- **假设检验**：样本数据是否能够拒绝“xx”与“xx”无显著差异的零假设。
- **统计显著**：拒绝零假设，认为样本结果不是偶然（by chance）。
- 如何判断是否在5%（1%更严格, 10%更宽松）的水平下显著?
  - P值<0.05；
  - C.R. (z值，t值, wald值)的绝对值 > 1.96;
  - 95%置信区间包含0（从负跨到正）。

推断性统计

TL;DR（太长不看版）：

只有$p<0.05$，结果才有意义

推断性统计

如果不显著……

- P值太大，Critical Ratio太小
- 可能是分子太小：样本表现出的效应不够强烈。
- 可能是分母太大：样本表现出的不确定性太强。
    - 可能是数据过于离散。
    - 可能是样本量太小。

$$CR = \frac{\boldsymbol{Est.}}{\boldsymbol{S.E.}}$$

$$= \frac{Est.}{\boldsymbol{s}/\sqrt{\boldsymbol{n}}}$$

经典的假设检验方法

正态性检验

独立样本的T检验/非参数检验

配对样本的T检验/非参数检验

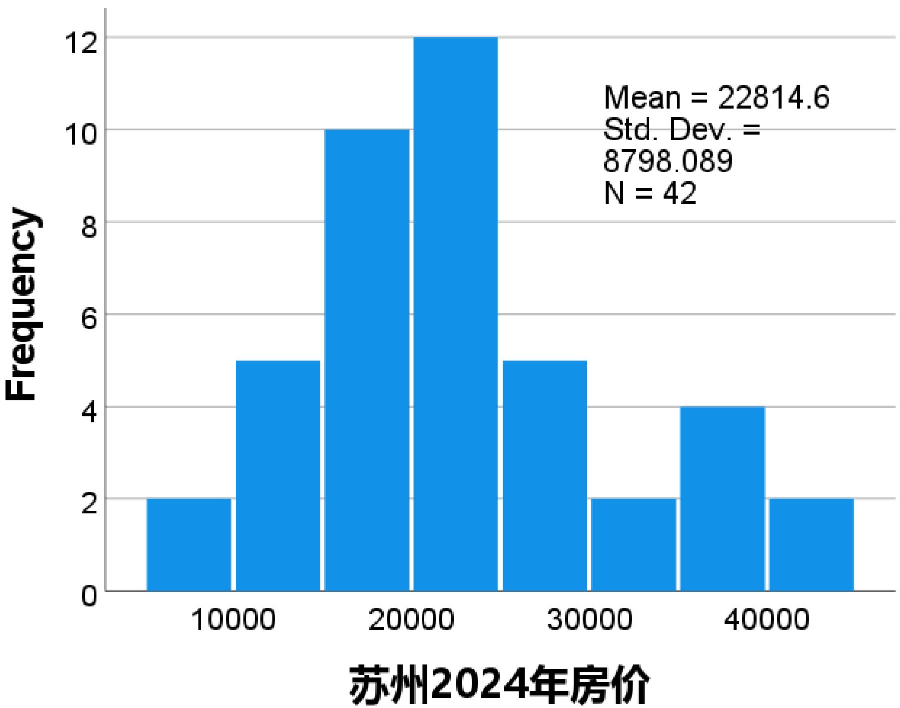
案例数据：链家二手房房价

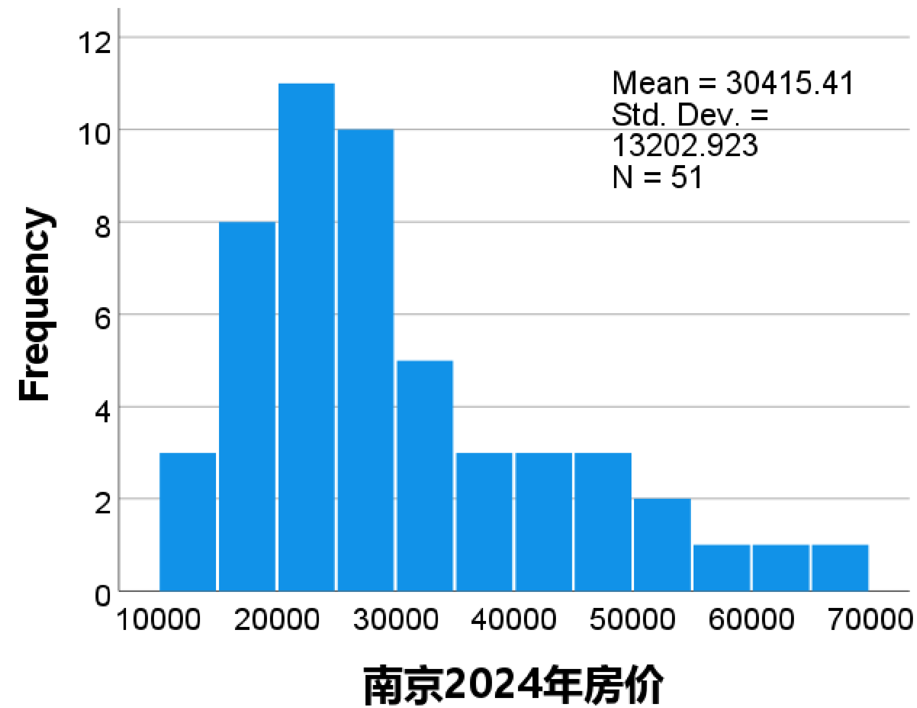
URL	城市	2023年房价	2024年房价
https://su.lianjia.com/ershoufang/107104989690.html	苏州	19100	16526
https://su.lianjia.com/ershoufang/107105838927.html	苏州	11953	10202
https://su.lianjia.com/ershoufang/107102921086.html	苏州	11042	10787
https://su.lianjia.com/ershoufang/107105809069.html	苏州	41764	37331
https://su.lianjia.com/ershoufang/107107190896.html	苏州	25196	24383
https://su.lianjia.com/ershoufang/107106563080.html	苏州	24535	22170
……	……	……	……
https://nj.lianjia.com/ershoufang/103133571262.html	南京	-	26861
https://nj.lianjia.com/ershoufang/103133503147.html	南京	-	44545
https://nj.lianjia.com/ershoufang/103135525061.html	南京	-	28371
https://nj.lianjia.com/ershoufang/103130041974.html	南京	-	18746
https://nj.lianjia.com/ershoufang/103135068524.html	南京	-	29859
……	……	-	……

有显著差异吗？

正态性检验

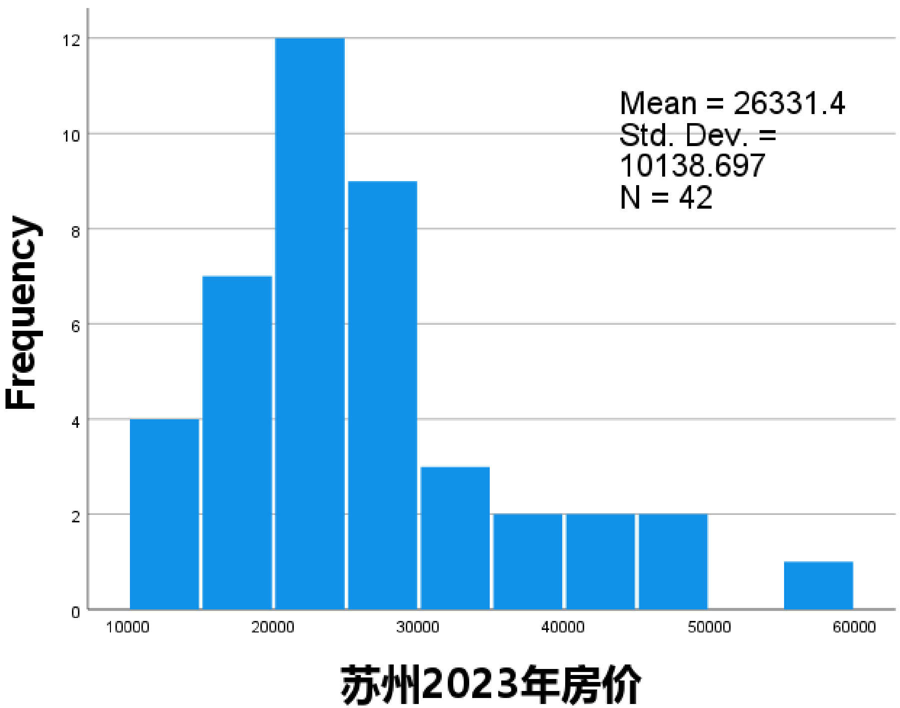
- 正态分布是许多统计模型的前提，因此，我们有时需要做正态性检验。
- 本例：苏州、南京的房价是否服从正态分布？

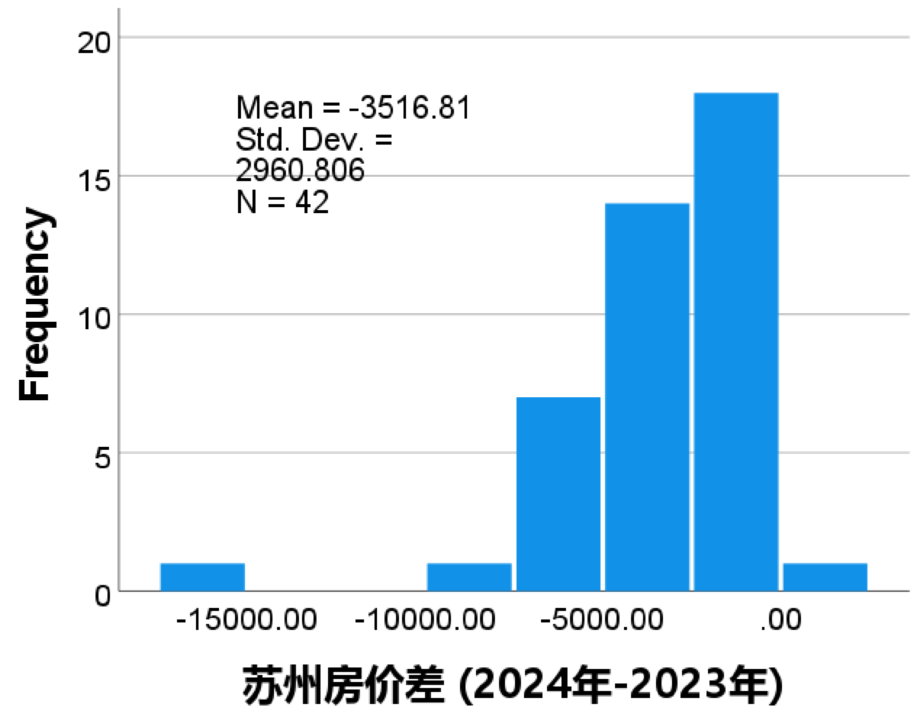




正态性检验

- 正态分布是许多统计模型的前提，因此，我们有时需要做正态性检验。
- 本例：苏州、南京的房价是否服从正态分布？





正态性检验

- 正态分布是许多统计模型的前提，因此，我们有时需要做正态性检验。
- 本例：苏州、南京的房价是否服从正态分布？
- **假设检验**，零假设：

“房价的分布”与“正态分布”之间

没有显著差异

正态性检验

Tests of Normality

	city	Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
price2024	苏州	.115	42	.183	.943	42	.035
	南京	.150	51	.006	.922	51	.002
price2023	苏州	.179	42	.002	.913	42	.004
delta_price	苏州	.142	42	.034	.793	42	<.001

- 检验方法：K-S检验（Kolmogorov-Smirnov）、S-W检验（Shapiro-Wilk）。
- 计算统计量，看它们是否落在拒绝域。然而我们也不知道拒绝域在哪😂
- **看P值！看P值！看P值！**即SPSS中的**Sig.**值。

正态性检验

总体要是正态分布的话，能抽到这种奇葩样本就见鬼了！

Tests of Normality

	city	Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
price2024	苏州		[illegible]	.183	.943	42	.035
	南京			.006	.922	51	.002
price2023	苏州			.002	.913	42	.004
delta_price	苏州			.034	.793	42	<.001

不见得……依我看，能抽到这样甚至更奇葩的样本的概率也有18%，不低了

- **S-W检验**：在0.05的显著性水平下拒绝零假设。“苏州2024年房价”、“南京2024年房价”、“苏州2023年房价”、“苏州的房价差”不服从正态分布。
- **K-S检验**：结论基本相同。但是对于“苏州2024年房价”，在0.1的显著性水平下能拒绝零假设，认为其服从正态分布。

正态性检验

总体要是正态分布的话，能抽到这种奇葩样本就见鬼了！

Tests of Normality

	city	Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
price2024	苏州	[illegible]	[illegible]	.183	.943	42	.035
	南京	[illegible]	[illegible]	.006	.922	51	.002
price2023	苏州	[illegible]	[illegible]	.002	.913	42	.004
delta_price	苏州	[illegible]	[illegible]	.034	.793	42	<.001

不见得……依我看，能抽到这样甚至更奇葩的样本的概率也有18%，不低了

If test of normality is not statistically significant (p > .05), you have *insufficient evidence* to reject the null hypothesis—so you proceed as if the population is normal

- **K-S检验**：结论基本相同。但是对于“苏州2024年房价”，在10%的显著性水平下能拒绝零假设，**认为其服从正态分布**。

案例数据：链家二手房房价

城市	2023年房价	2024年房价
苏州	19100	16526
苏州	11953	10202
苏州	11042	10787
苏州	41764	37331
苏州	25196	24383
苏州	24535	22170
……	……	……
南京	-	26861
南京	-	44545
南京	-	28371
南京	-	18746
南京	-	29859
……	-	……

配对样本
paired samples

同样的苏州住宅，2023年和2024年的房价有显著差异吗？

- 每1组对应1个值变量。
- 每组的样本量必然相同。

独立样本
independent samples

苏州与南京的房价有显著差异吗？

- 1个值变量和1个组别变量。
- 每组的样本量可以不同。

独立样本的差异性比较

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

- 描述性统计：单从本样本来看，南京房价高于苏州。
- 这个结论能扩展到总体吗？

独立样本的差异性检验

独立样本T检验 (Independent Sample T Tests)

- 零假设：

"苏州房价的**均值**" 与 "南京房价的**均值**" 之间

没有显著差异

独立样本的差异性比较

独立样本T检验 (Independent Sample T Tests)

- 前提：
  - 样本之间是独立的。
  - 每组样本来自正态分布总体。
  - 各组之间的方差相等（方差齐性）。

不正态，怎么办？

样本量越大，重要性越小。可以参考使用**n=30**的临界值。n>30，只要分布不是过于不对称，没有不合理的极端值，即可不做要求。

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
price2024	Equal variances assumed	5.949	.017	-3.191	91	.002	-7600.817	2381.734	-12331.838	-2869.795
	Equal variances not assumed			-3.314	87.451	.001	-7600.817	2293.685	-12159.433	-3042.200

独立样本的差异性比较

独立样本T检验 (Independent Sample T Tests)

- 前提：各组之间的方差相等（方差齐性）。
  - Levene方差齐性检验：零假设是“苏州房价的方差”与“南京房价的方差”之间没有显著差异。
  - 结论：在0.05的显著性水平下拒绝H0，方差不齐。

		Levene's Test for Equality of Variances	
		F	Sig.
price2024	Equal variances assumed	5.949	.017
	Equal variances not assumed		

独立样本的差异性比较

独立样本T检验 (Independent Sample T Tests)

- 根据方差齐性检验结果选择合适的行，查看T检验的结果。
  - P值：在0.01的水平下拒绝零假设，可以认为2024年苏州与南京的房价均值存在显著差异。
  - 差多少？苏州比南京平均便宜7601元。
  - 置信区间：95%的可能，苏州比南京便宜3042~12159元。

		t-test for Equality of Means					
						95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
price2024	Equal variances assumed	-3.191	91	.002	-7600.817	-12331.838	-2869.795
	Equal variances not assumed	-3.314	87.451	.001	-7600.817	-12159.433	-3042.200

独立样本的差异性比较

如果担心前提条件，**还有别的选择吗?**

有！属于**非参数检验**的Mann-Whitney检验申请出战!

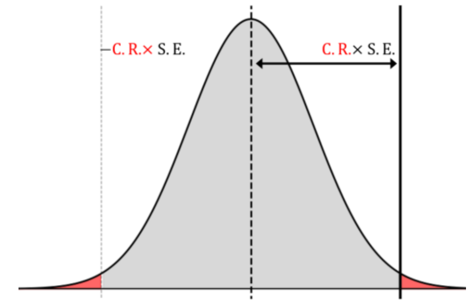
“苏州房价的**中位数**”与“南京房价的**中位数**”之间

没有显著差异

独立样本的差异性比较

参数检验 (Parametric tests)

- 要求样本来自特定的分布，通常是正态分布。
  - 检验过程依赖于分布的参数（如均值、标准差）。



$$CR = \frac{Est. - \mu}{s/\sqrt{n}} \sim t(n-1)$$

T检验

非参数检验 (Nonparametric tests)

- 不需要假设样本来自特定的分布。
  - 基于样本数据的排序或其他统计量进行分析，不依赖于分布的参数。

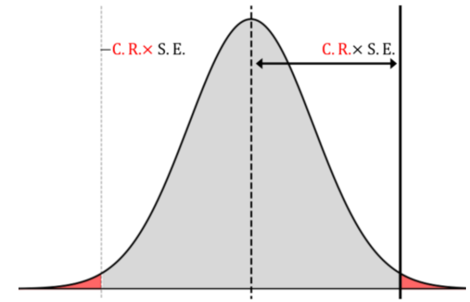
		排序(rank)
苏州	2.1万, 2.2万, 2.3万, 2.4万, 2.5万	1, 2, 3, 4, 5
南京	3.1万, 3.2万, 3.3万, 3.4万, 3.5万	6, 7, 8, 9, 10

M-W检验

独立样本的差异性比较

参数检验（Parametric tests)

- 要求样本来自特定的分布，通常是正态分布。
  - 检验过程依赖于分布的参数（如均值、标准差）。
  - **通常较为敏感，但是相对不够稳健。**



$$CR = \frac{Est. - \mu}{s/\sqrt{n}} \sim t(n-1)$$

T检验

非参数检验（Nonparametric tests)

- 不需要假设样本来自特定的分布。
  - 基于样本数据的排序或其他统计量进行分析，不依赖于分布的参数。
  - **通常较为稳健，但是相对不够敏感。**

		排序(rank)
苏州	2.1万, 2.2万, 2.3万, 2.4万, **17.5万**	1, 2, 3, 4, 10
南京	3.1万, 3.2万, 3.3万, 3.4万, 3.5万	5, 6, 7, 8, 9

M-W检验

独立样本的差异性比较

非参数检验

Mann-Whitney Test

Ranks

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

Test Statistics

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

- Mann-Whitney检验（M-W检验）是独立样本T检验的非参数版本。
- 该检验使用排序而非原始数值，存在信息损失，但对异常值更稳健。
- P值：在0.01的水平下拒绝零假设，可以认为2024年苏州和南京的房价存在显著差异。

配对样本的差异性检验

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

- 描述性统计：单从本样本来看，2024年苏州相同房产的价格低于2023年。
- 这个结论能扩展到总体吗?

配对样本的差异性检验

配对样本T检验 (Paired Sample T Tests)

- 零假设：

对于同一套苏州房产，

“2024年房价的均值”与“2023年房价的均值”之间

没有显著差异

配对样本的差异性检验

配对样本T检验 (Paired Sample T Tests)

- 零假设：

对于同一套苏州房产，

“2024和2023年房价之差的均值”与“0”之间

没有显著差异

- 前提：
  - 样本具有配对性。
  - 样本差值服务正态分布。

配对样本的差异性检验

配对样本T检验 (Paired Sample T Tests)

- P值：在0.01的水平下拒绝零假设，可以认为苏州市同一套房产的价格在2024年相比于2023年发生了显著下跌。
- 差多少？下均下跌3517元。
- 置信区间：95%的可能下跌2594~4439元。

Paired Samples Test

			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Lower	Upper	t	df	Sig. (2-tailed)
price2023 - price2024	3516.810	2960.806	2594.158	4439.461	7.698	41	<.001

独立样本的差异性比较

非参数检验

Wilcoxon Signed Ranks Test

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

Test Statistics[a]

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

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

- Wilcoxon符号秩检验是配对样本T检验的非参数版本。
- 该检验使用排序而非原始数值，存在信息损失，但对异常值更稳健。
- P值：在0.01的水平下拒绝零假设，可以认为苏州市同一套房产的价格在2024年相比于2023年发生了显著下跌。

总结

- 推断性统计：考虑了抽样误差，利用样本数据来估计、分析总体的特征。
- 假设检验：样本数据是否能够拒绝“xx”与“xx”无显著差异的零假设。
- 统计显著： P值<0.05 。
- 实务：经典的假设检验——如何比较两组数据是否有显著差异?
  - 独立样本、配对样本。
  - T检验更灵敏，非参数检验更稳健。
  - 小样本（n<30）要检查正态性，大样本要留心对称性和极端值。