

评估相关性

城市分析方法系列课程

苏州大学 王灿

学习目标

- 掌握方差分析
- 掌握列联表分析
- 掌握相关系数

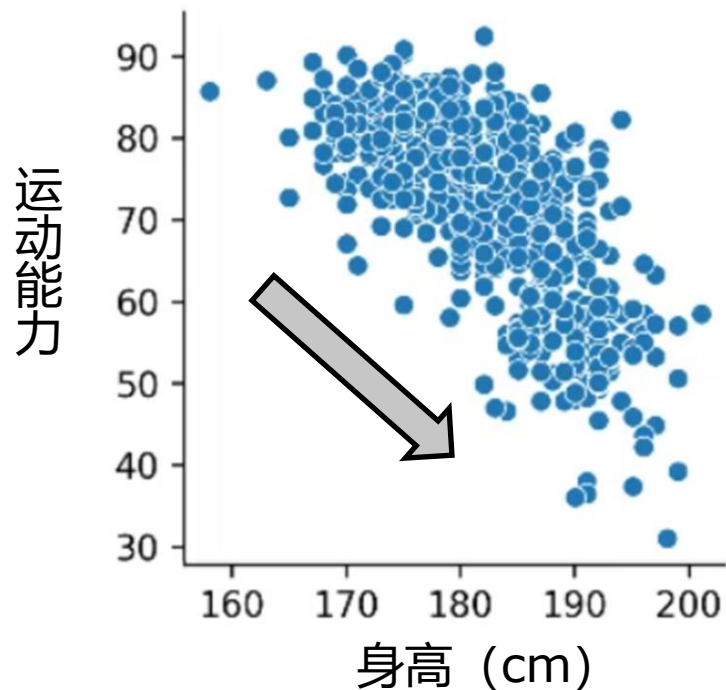
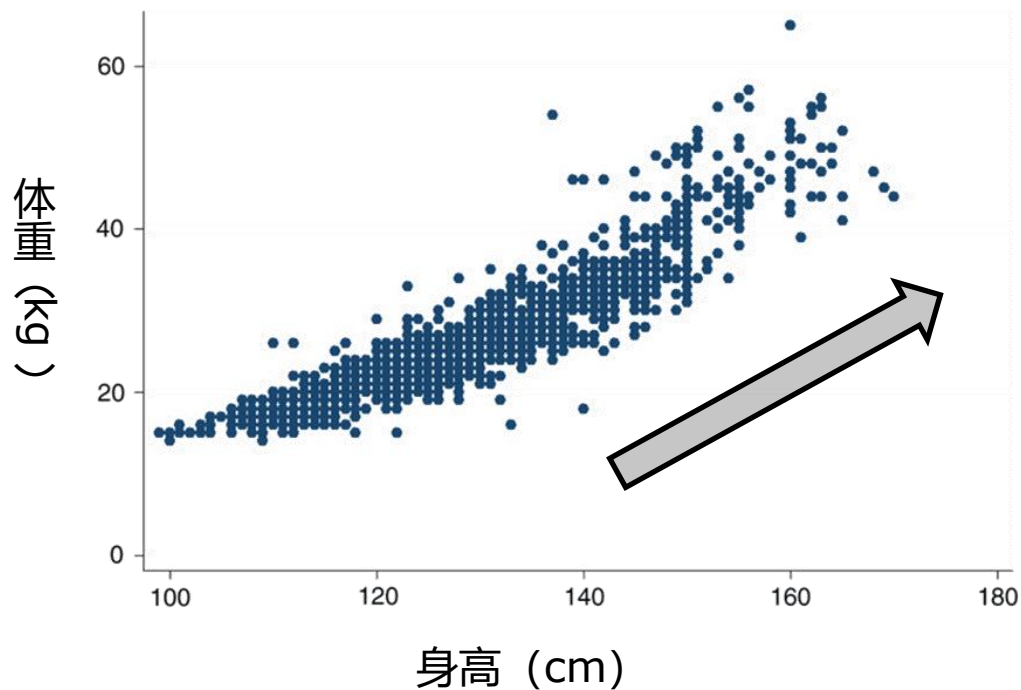
什么是相关性

相关性 (correlation)：两个变量之间的关联程度。它衡量的是当一个变量发生变化时，另一个变量是否也会随之发生变化，以及这种变化的方向和强度。

老师老师别生气，
咱娃就是这样滴，
叫他朝西他朝东，
叫他朝东他朝西。

什么是相关性

相关性 (correlation)：两个变量之间的关联程度。它衡量的是当一个变量发生变化时，另一个变量是否也会随之发生变化，以及这种变化的方向和强度。



相关性测度方法



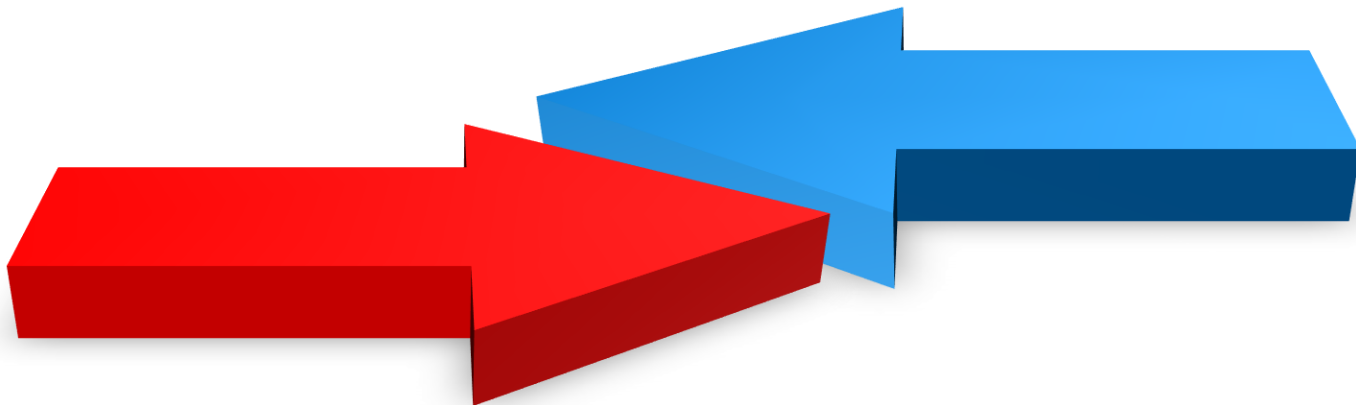
数值变量
(Numerical)

VS

数值变量
(Numerical)

相关系数 (Correlation Coefficient)

相关性测度方法



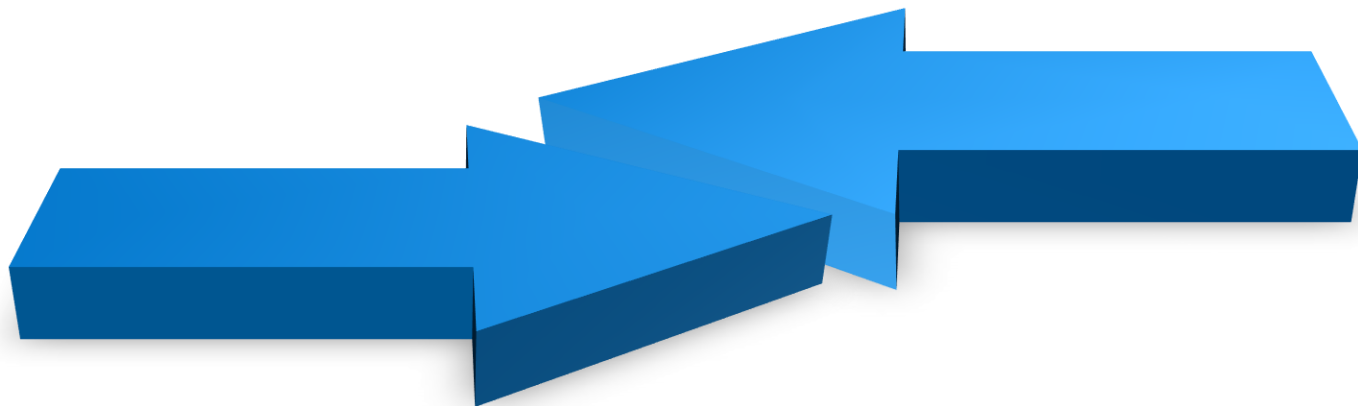
数值变量
(Numerical)

VS

分类变量
(Categorical)

方差分析 (Analysis of Variance, ANOVA)

相关性测度方法



分类变量

(Categorical)

VS

分类变量

(Categorical)

列联表分析 (Crosstabs Analysis)

案例

- 数据：工业用地功能更新



VS

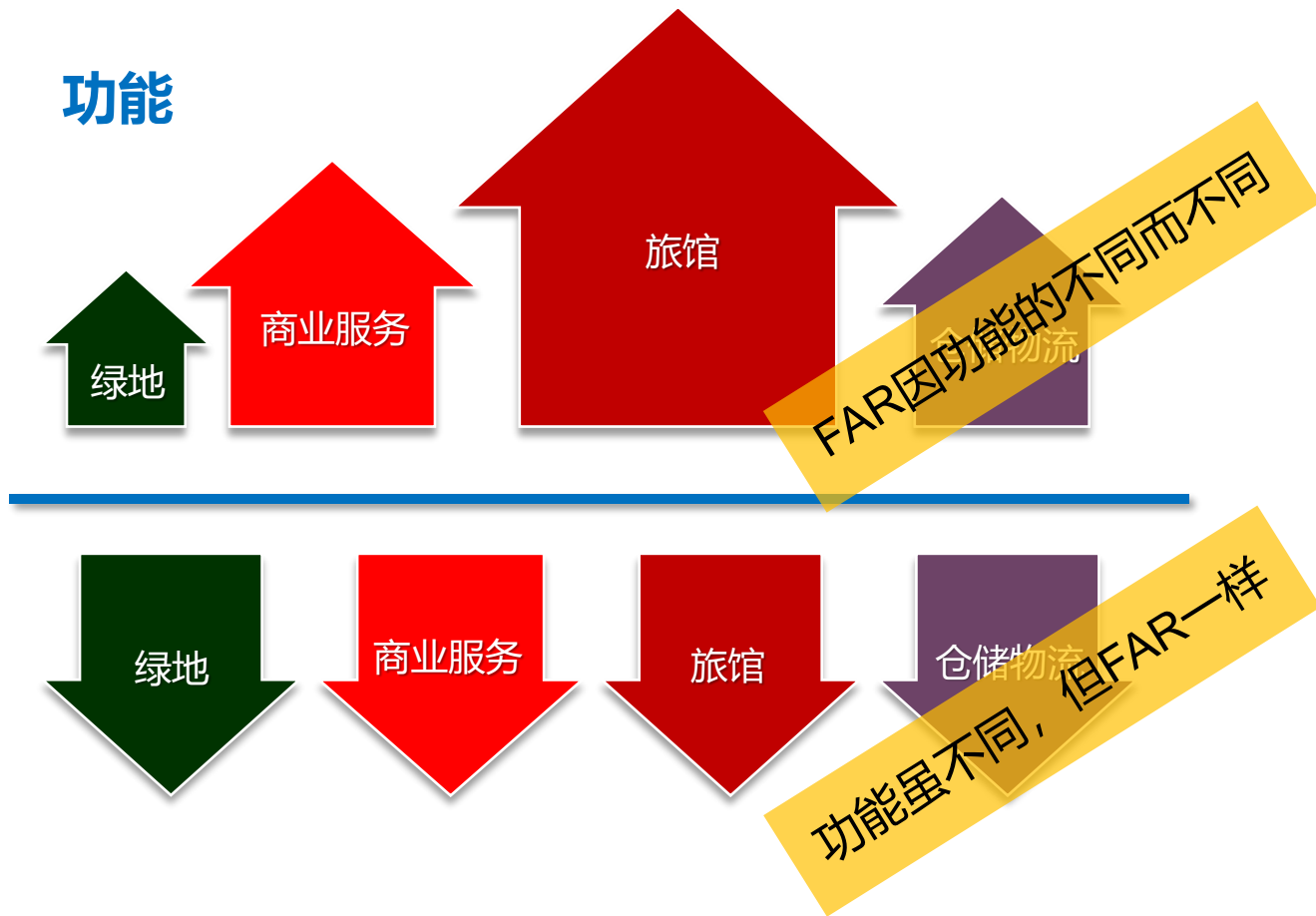


数值 vs. 分类：方差分析

现状容积率 Vs 功能



什么叫相关/不相关?



数值 vs. 分类：方差分析

不同的**功能**，带来**现状容积率**的**差异**

分类变量不同的**分类**，带来**数值变量**在**数值**上的**差异**

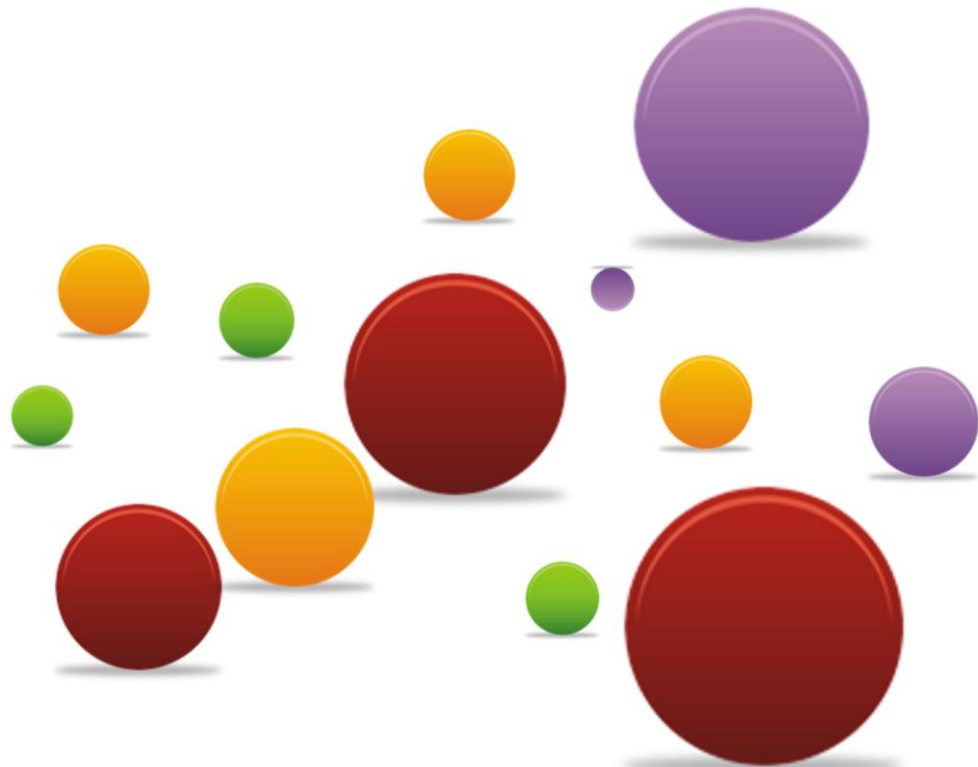
DIFFER

数值 vs. 分类：方差分析

现状容积率 Vs 功能



什么算有差异/没差异?

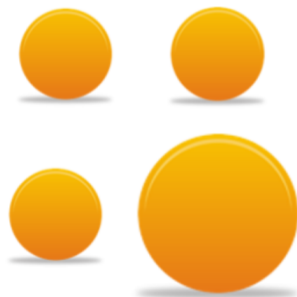


数值 vs. 分类：方差分析

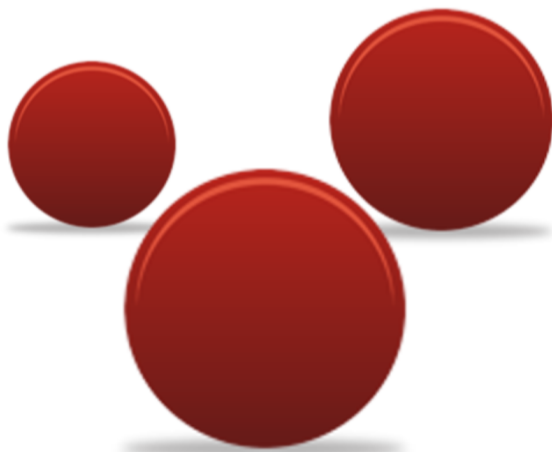
现状容积率

Vs

功能



什么算有差异/没差异?

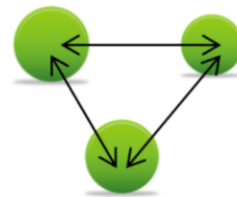
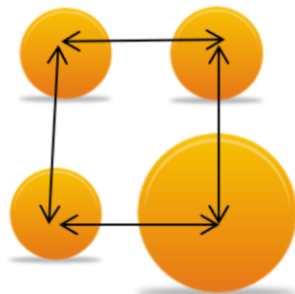


数值 vs. 分类：方差分析

现状容积率

Vs

功能



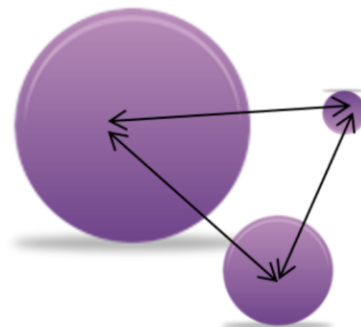
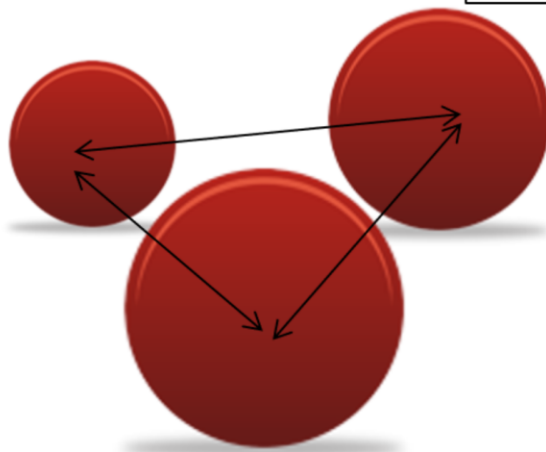
总差异

=

组内差异



什么算有差异/没差异?



数值 vs. 分类：方差分析

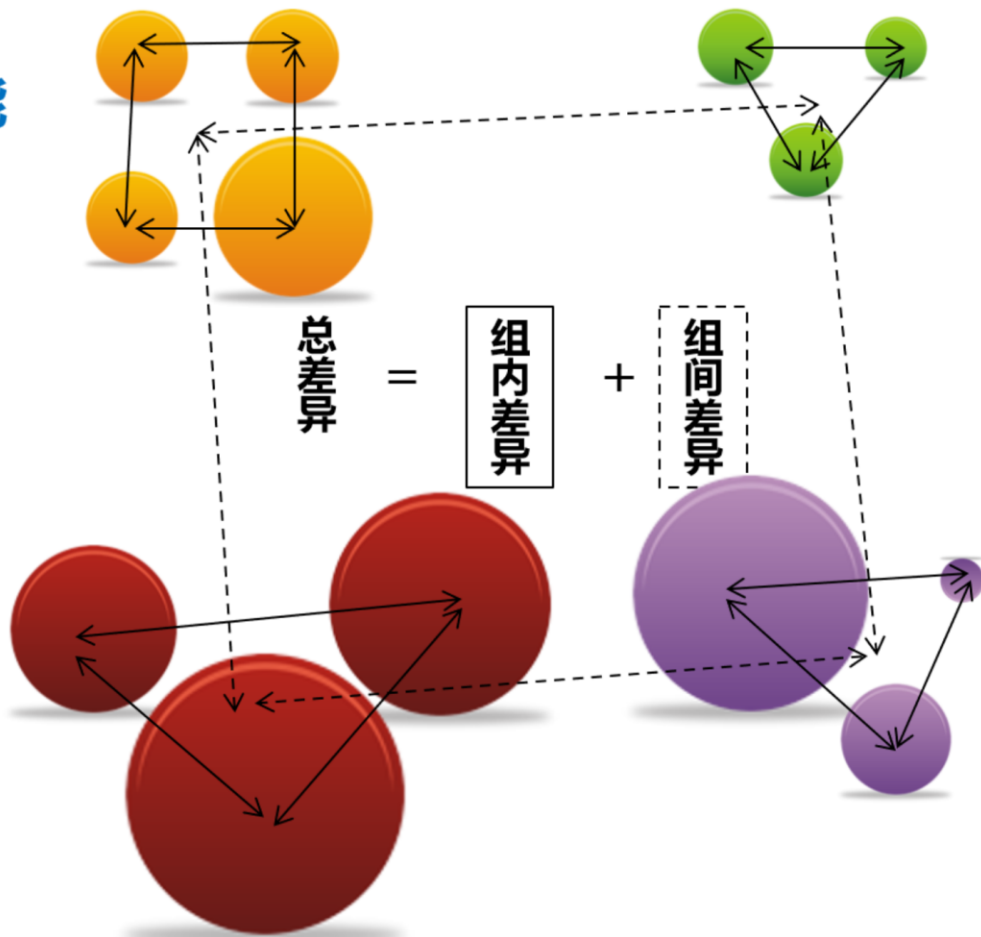
现状容积率

Vs

功能



什么算有差异/没差异?



数值 vs. 分类：方差分析

现状容积率 Vs 功能

$$\text{总差异} = \boxed{\text{组内差异}} + \text{组间差异}$$



方差分析

什么算有差异/没差异?

用显著性说话：显著差异/差异不显著

F检验：

$$F = \frac{\boxed{\text{组间均方差异}}}{\boxed{\text{组内均方差异}}}$$

数值 vs. 分类：方差分析

现状容积率 Vs 功能

In现状容积率

ANOVA

	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			

判定：相关

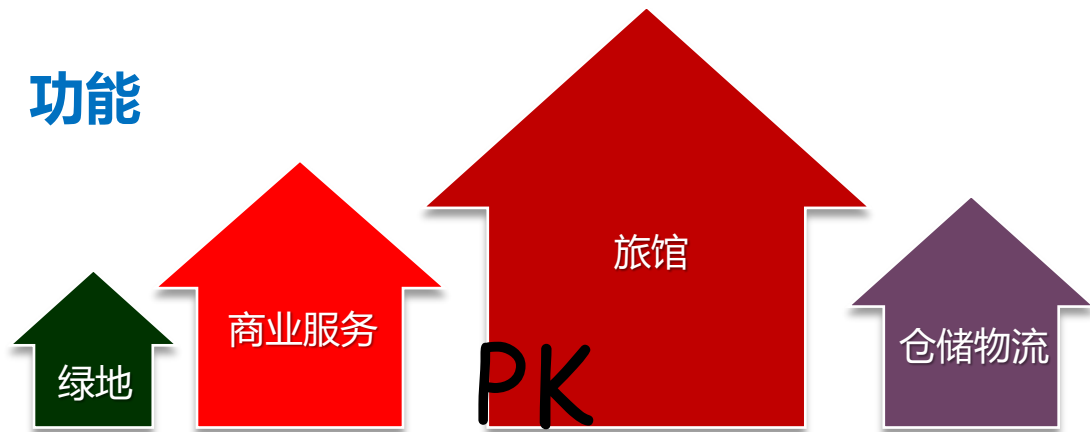
不同的**功能**之间，**现状容积率**存在**显著差异**。

数值 vs. 分类：方差分析

现状容积率 Vs 功能

HOW

已经相关了，然后.....



不是每两类之间
都有很大差异!



究竟是哪些类别之
间存在显著差异?

数值 vs. 分类：方差分析

现状容积率 Vs 功能



事后多重比较：
LSD、SNK.....

Post Hoc Tests Multiple Comparisons

Dependent Variable: In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
旅馆	收储	.34159	.18617	.069	-.0266	.7098
	道路绿地	.58275*	.28146	.040	.0261	1.1394
	社会服务设施	.52372	.33005	.115	-.1290	1.1764
	创意产业园	.12882	.18308	.483	-.2332	.4909
	商务办公	.18626	.16502	.261	-.1401	.5126
	商业服务	.41487*	.16046	.011	.0976	.7322
	仓储物流生产	.57852*	.15692	<.001	.2682	.8888
	未更新	.56240*	.17801	.002	.2104	.9144

*. The mean difference is significant at the 0.05 level.

数值 vs. 分类：方差分析

现状容积率 Vs 功能

不同功能现状容积率的两两比较分析表

	仓储物流 生产	未更新	商业服务	道路绿地	社会服务 设施	收储	创意产业 和科技园	商务办公	旅馆
仓储物流 生产	-						●	●	●
未更新		-					●	●	●
商业服务			-				●	●	●
道路绿地				-					●
社会服务 设施					-				
收储						-			
创意产业 和科技园	●	●	●				-		
商务办公	●	●	●					-	
旅馆	●	●	●	●					-



事后两两比较：
LSD、SNK.....

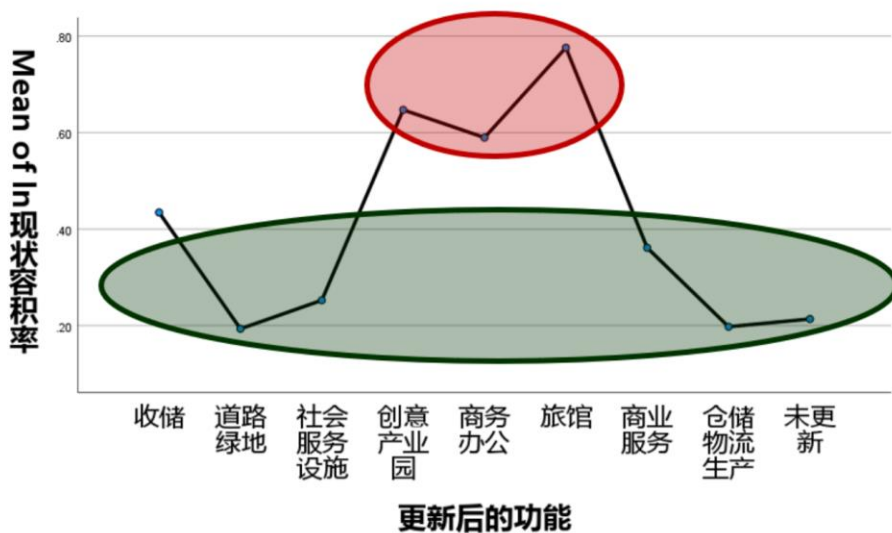
数值 vs. 分类：方差分析

现状容积率 Vs 功能



不同功能现状容积率的两两比较分析表

	仓储物流 生产	未更新	商业服务	道路绿地	社会服务 设施	收储	创意产业 和科技园	商务办公	旅馆
仓储物流 生产	-						•	•	•
未更新		-					•	•	•
商业服务			-				•	•	•
道路绿地				-					•
社会服务 设施					-				
收储						-			
创意产业 和科技园	•	•	•				-		
商务办公	•	•	•					-	
旅馆	•	•	•	•					-



事后两两比较：
LSD、SNK.....

数值 vs. 分类：方差分析

现状容积率 Vs 功能

How much

数值与分类变量

Eta

$$Eta = \sqrt{SS_{Between} / SS_{Total}}$$

$$Eta_{(现状容积率, 功能)} = 0.42$$

Nominal by Interval. When one variable is categorical and the other is quantitative, select **Eta**. The categorical variable must be coded numerically.

[Hide details](#)

- **Eta.** A measure of association that ranges from 0 to 1, with 0 indicating no association between the row and column variables and values close to 1 indicating a high degree of association. Eta is appropriate for a dependent variable measured on an interval scale (for example, income) and an independent variable with a limited number of categories (for example, gender). Two eta values are computed: one treats the row variable as the interval variable, and the other treats the column variable as the interval variable.

数值 vs. 分类：方差分析

规划容积率 Vs 功能

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			

不同的**功能**之间，

规划容积率

没有显著差异。

ANOVA

In规划容积率

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.385	8	.173	1.142	.343
Within Groups	14.093	93	.152		
Total	15.477	101			

判定：不相关

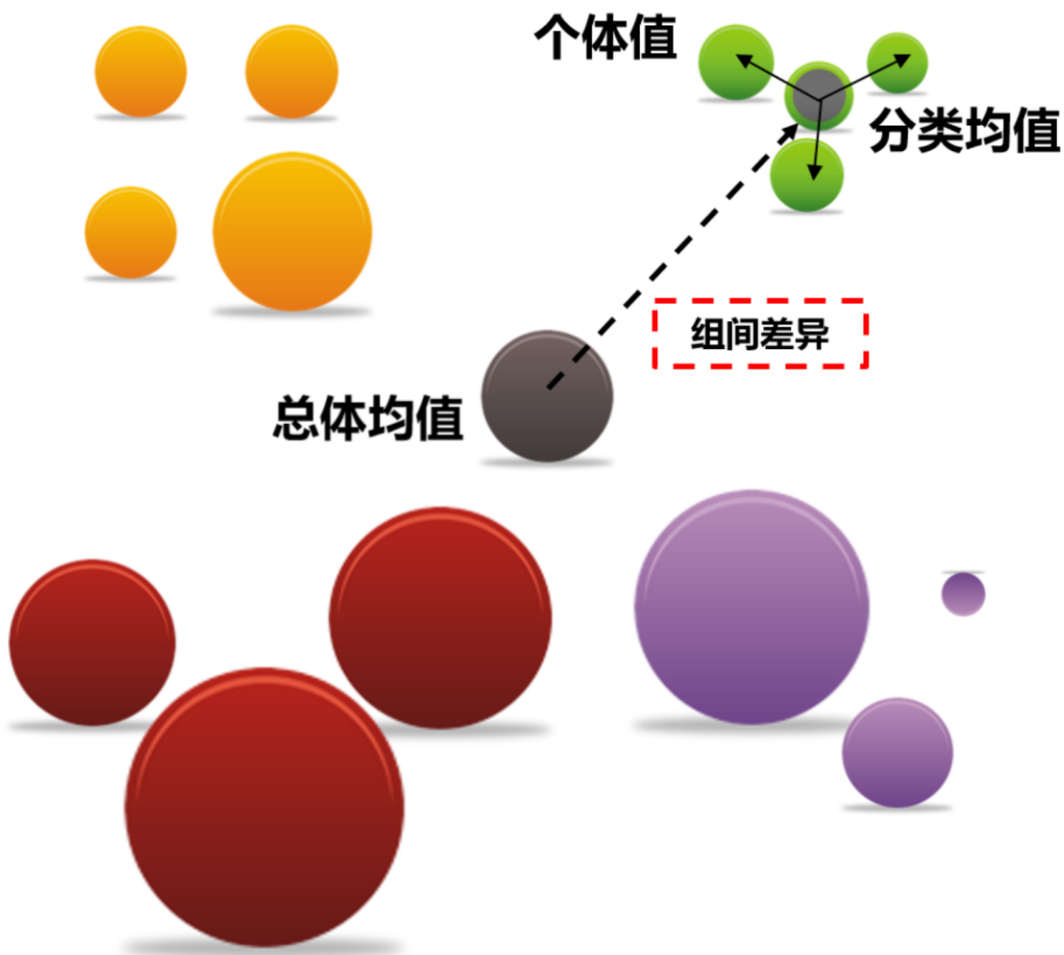
然后？

就没有然后了.....

数值 vs. 分类：方差分析

方差分析的一般形式

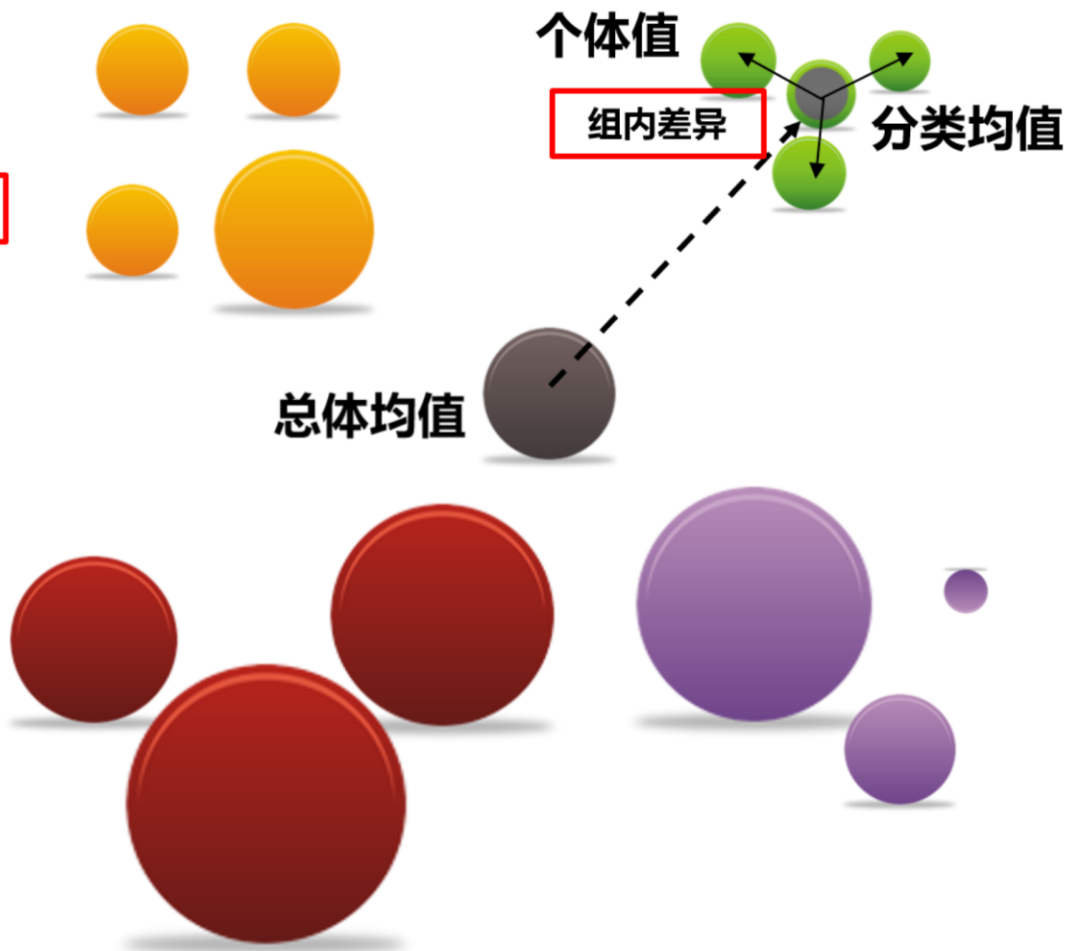
个体值 = 总体均值 + 分类效应 + 误差



数值 vs. 分类：方差分析

方差分析的一般形式

个体值 = 总体均值 + 分类效应 + 误差

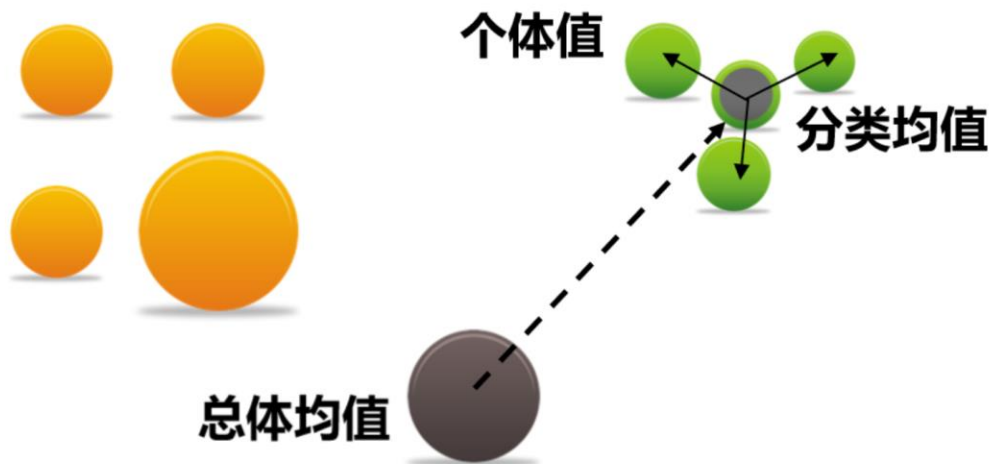


数值 vs. 分类：方差分析

方差分析的一般形式

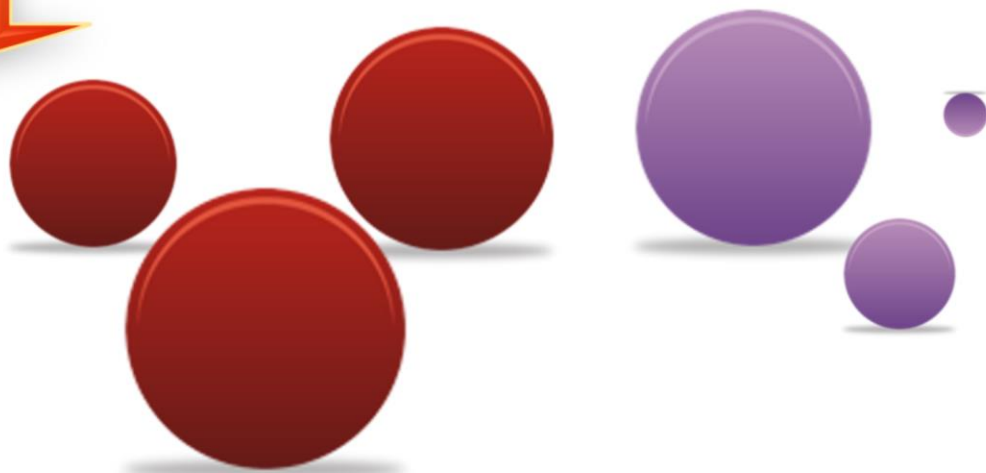
个体值 = 总体均值 + 分类效应 + 误差

个体值 = 总体均值
+ 分类A效应 + 分类B效应 + ...
+ 误差



多因素方差分析

一般线性模型 (GLM)



数值 vs. 分类：方差分析

方差分析的前提条件：

- 样本之间是独立的。
- 每组样本来自正态分布总体。

方差分析对正态性假设较为稳健（即使不完全正态，结果依然有一定的可信度），尤其是在样本量较大的情况下。

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
现状容积率	.137	145	<.001	.898	145	<.001
ln现状容积率	.054	145	.200 [*]	.992	145	.625

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

数值 vs. 分类：方差分析

方差分析的前提条件：

- 样本之间是独立的。
- 每组样本来自正态分布总体。

方差分析对正态性假设较为稳健（即使不完全正态，结果依然有一定的可信度），尤其是在样本量较大的情况下。

- 各组之间的方差相等（方差齐性）。

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
现状容积率	Based on Mean	3.032	8	136	.004
	Based on Median	2.441	8	136	.017
	Based on Median and with adjusted df	2.441	8	107.470	.018
	Based on trimmed mean	2.946	8	136	.005

Tests of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
In现状容积率	1.391	8	136	.206
	1.188	8	136	.311
	1.188	8	128.409	.311
	1.359	8	136	.220

数值 vs. 分类：非参数检验

K个独立样本的非参数检验

- **Kruskal-Wallis检验**是一个基于秩的**非参数检验**，用于比较多组样本的**中位数**，适用于数据不满足正态分布或方差齐性的情况下，是方差分析的非参数替代方法。

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

Test Statistics^{a,b}

	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

判定：相关

数值 vs. 分类：非参数检验

K个独立样本的非参数检验

- **Kruskal-Wallis检验**是一个基于秩的**非参数检验**，用于比较多组样本的**中位数**，适用于数据不满足正态分布或方差齐性的情况下，是方差分析的非参数替代方法。

Ranks

	更新后的功能	N	Mean Rank
规划容积率	收储	9	57.11
	道路绿地	1	94.50
	社会服务设施	2	39.00
	创意产业园	11	57.91
	商务办公	15	51.53
	旅馆	5	55.30
	商业服务	23	55.37
	仓储物流生产	26	39.63
	未更新	10	57.60
	Total		102

Test Statistics^{a,b}

	规划容积率
Kruskal-Wallis H	8.606
df	8
Asymp. Sig.	.377

a. Kruskal Wallis Test

b. Grouping Variable: 更新后的功能

判定：不相关

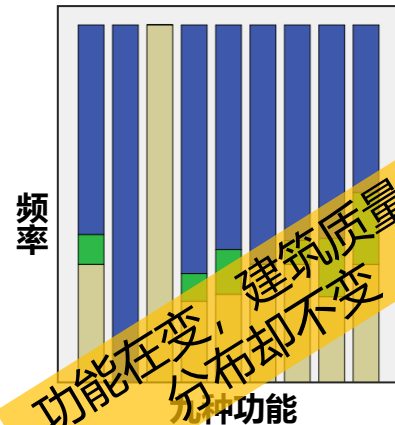
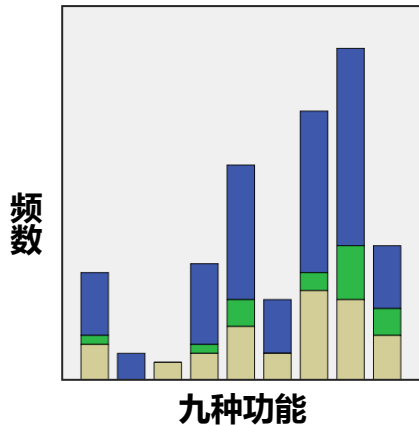
分类 vs. 分类：列联表分析

功能 Vs 建筑质量

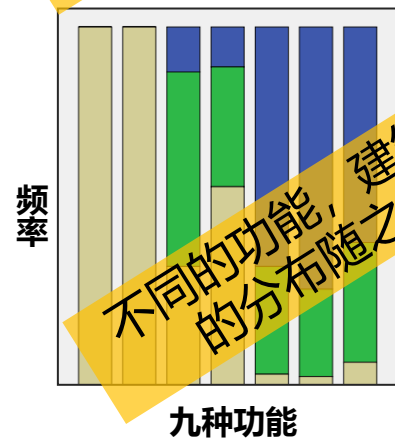
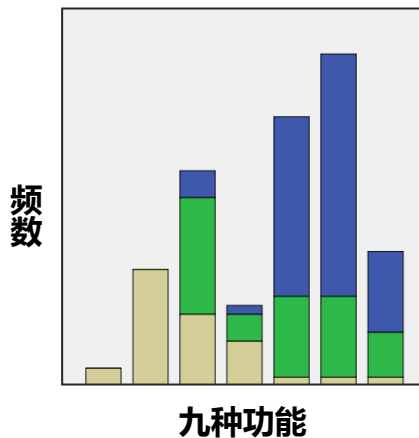


什么叫相关/不相关?

■ 差
■ 中
■ 好



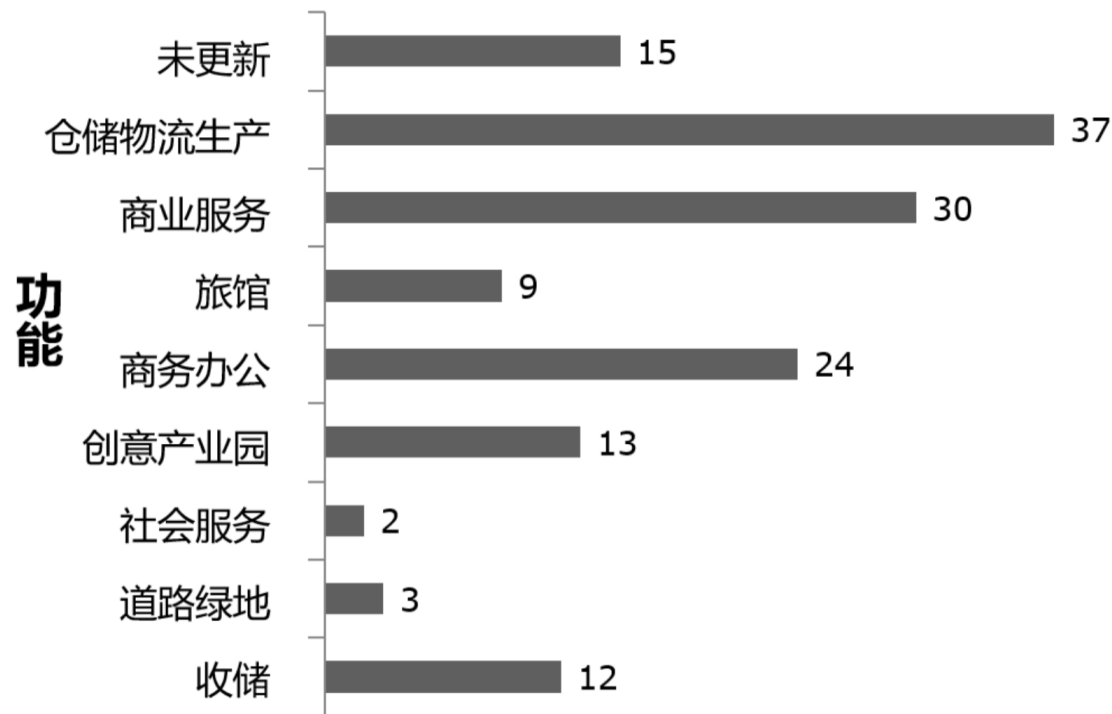
功能在变，建筑质量的分布却不变



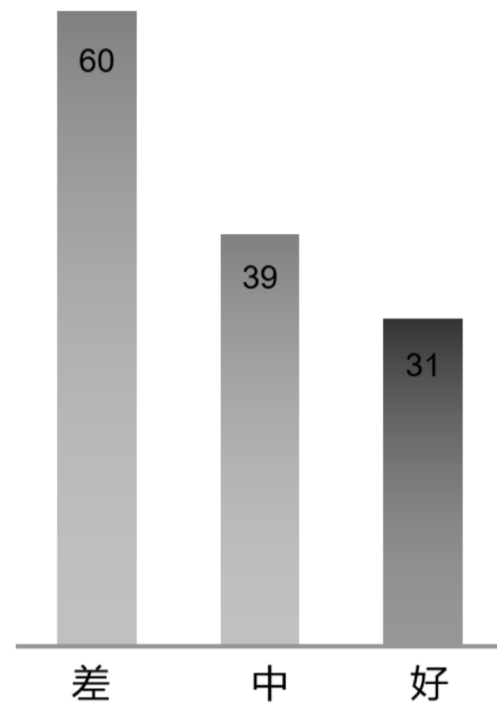
不同的功能，建筑质量的分布随之不同

分类 vs. 分类：列联表分析

功能 Vs 建筑质量



建筑质量



分类 vs. 分类：列联表分析

功能 Vs 建筑质量

建筑质量

功能

	差	中	好
未更新			
仓储物流生产			
商业服务			
旅馆			
商务办公		13	
创意产业园			
社会服务			
道路绿地			
收储			

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

			建筑质量			Total
			差	中	好	
功能	社会服务设施	频数	0	0	2	2
		%	0.0%	0.0%	1.5%	1.5%
	创意产业园	频数	0	0	13	13
		%	0.0%	0.0%	10.0%	10.0%
	商务办公	频数	3	13	8	24
		%	2.3%	10.0%	6.2%	18.5%
	旅馆	频数	1	3	5	9
		%	0.8%	2.3%	3.8%	6.9%
	商业服务	频数	20	9	1	30
		%	15.4%	6.9%	0.8%	23.1%
	仓储物流生产	频数	27	9	1	37
		%	20.8%	6.9%	0.8%	28.5%
	未更新	频数	9	5	1	15
		%	6.9%	3.8%	0.8%	11.5%
Total	频数	60	39	31	130	
	%	46.2%	30.0%	23.8%	100.0%	

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

			建筑质量			Total
			差	中	好	
功能	社会服务设施	频数				
		%				1.5%
	创意产业园	频数				
		%				10.0%
	商务办公	频数				
		%				18.5%
	旅馆	频数				
	%				6.9%	
	商业服务	频数				
		%				23.1%
	仓储物流生产	频数				
		%				28.5%
	未更新	频数				
		%				11.5%
Total		频数				130
		%	46.2%	30.0%	23.8%	100.0%

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

			建筑质量			Total
			差	中	好	
功能	社会服务设施	频数				
		%	0.7%	0.5%	0.4%	1.5%
	创意产业园	频数				
		%	4.6%	3.0%	2.4%	10.0%
	商务办公	频数				
		%	8.5%	5.6%	4.4%	18.5%
	旅馆	频数				
		%	3.2%	2.1%	1.6%	6.9%
	商业服务	频数				
		%	10.7%	6.9%	5.5%	23.1%
	仓储物流生产	频数				
		%	13.2%	8.6%	6.8%	28.5%
	未更新	频数				
		%	5.3%	3.5%	2.7%	11.5%
Total		频数			31	130
		%	46.2%	30.0%	23.8%	100.0%

独立假设

独立事件乘法公式

$$P(AB)=P(A) \times P(B)$$

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

			建筑质量			Total
			差	中	好	
功能	社会服务设施	频数	0.9	0.6	0.5	2
		%	0.7%	0.5%	0.4%	1.5%
	创意产业园	频数	6.0	3.9	3.1	13
		%	4.6%	3.0%	2.4%	10.0%
	商务办公	频数	11.1	7.2	5.7	24
		%	8.5%	5.6%	4.4%	18.5%
	旅馆	频数	4.2	2.7	2.1	9
		%	3.2%	2.1%	1.6%	6.9%
	商业服务	频数	13.9	9.0	7.1	30
		%	10.7%	6.9%	5.5%	23.1%
	仓储物流生产	频数	17.1	11.1	8.8	37
		%	13.2%	8.6%	6.8%	28.5%
	未更新	频数	6.9	4.5	3.6	15
		%	5.3%	3.5%	2.7%	11.5%
Total		频数	60	39	31	130
		%	46.2%	30.0%	23.8%	100.0%

独立假设

独立事件乘法公式

$$P(AB)=P(A) \times P(B)$$

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

实际频数分布

			建筑质量			Total
			差	中	好	
功能	社会服务设施	频数	0	0	2	2
		%	0.0%	0.0%	1.5%	1.5%
	创意产业园	频数	0	0	13	13
		%	0.0%	0.0%	10.0%	10.0%
	商务办公	频数	3	13	8	24
		%	2.3%	10.0%	6.2%	18.5%
	旅馆	频数	1	3	5	9
		%	0.8%	2.3%	3.8%	6.9%
	商业服务	频数	20	9	1	30
		%	15.4%	6.9%	0.8%	23.1%
	仓储物流生产	频数	27	9	1	37
		%	20.8%	6.9%	0.8%	28.5%
	未更新	频数	9	5	1	15
		%	6.9%	3.8%	0.8%	11.5%
Total	频数	60	39	31	130	
	%	46.2%	30.0%	23.8%	100.0%	

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

实际频数分布

独立期望频数分布

			建筑质量			Total
			差	中	好	
功能	社会服务设施	频数	0.9	0.6	0.5	2
		%	0.7%	0.5%	0.4%	1.5%
	创意产业园	频数	6.0	3.9	3.1	13
		%	4.6%	3.0%	2.4%	10.0%
	商务办公	频数	11.1	7.2	5.7	24
		%	8.5%	5.6%	4.4%	18.5%
	旅馆	频数	4.2	2.7	2.1	9
		%	3.2%	2.1%	1.6%	6.9%
	商业服务	频数	13.9	9.0	7.1	30
		%	10.7%	6.9%	5.5%	23.1%
	仓储物流生产	频数	17.1	11.1	8.8	37
		%	13.2%	8.6%	6.8%	28.5%
	未更新	频数	6.9	4.5	3.6	15
		%	5.3%	3.5%	2.7%	11.5%
Total		频数	60	39	31	130
		%	46.2%	30.0%	23.8%	100.0%

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

实际频数分布

独立期望频数分布



如果没差异，则不能拒绝独立假设，二者不相关



列联表分析

卡方检验 (Chi-square) :

$$\chi^2 = \sum \frac{(\text{实际频数} - \text{期望频数})^2}{\text{期望频数}}$$

什么算有差异/没差异?

用显著性说话：显著差异/差异不显著

分类 vs. 分类：列联表分析

功能 Vs 建筑质量

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	89.257 ^a	12	<.001
Likelihood Ratio	89.257	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.

实际频数分布与独立假设下的期望频数分布间存在**显著差异**。

分类 vs. 分类：列联表分析

卡方检验 (Chi-Square Test)

- 卡方检验本身就是非参数检验。
- Pearson卡方
 - 条件：对于2x2的列联表，每个单元格的期望频数大于5；对于其他列联表，80%以上的单元格的期望频数大于5，且不能有单元格的期望频数为0。
 - 最经典的卡方统计量，只要满足条件，优先使用。
- 似然比卡方 (likelihood ratio)
 - 在样本量较小或频数较低时更加稳健。
 - 在Pearson卡方的条件不满足时，可作为替代。
- Linear-by-Linear卡方
 - 适用于有序分类变量之间。

分类 vs. 分类：列联表分析

功能 Vs 建筑质量
表

HOW

已经相关了，然后……

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

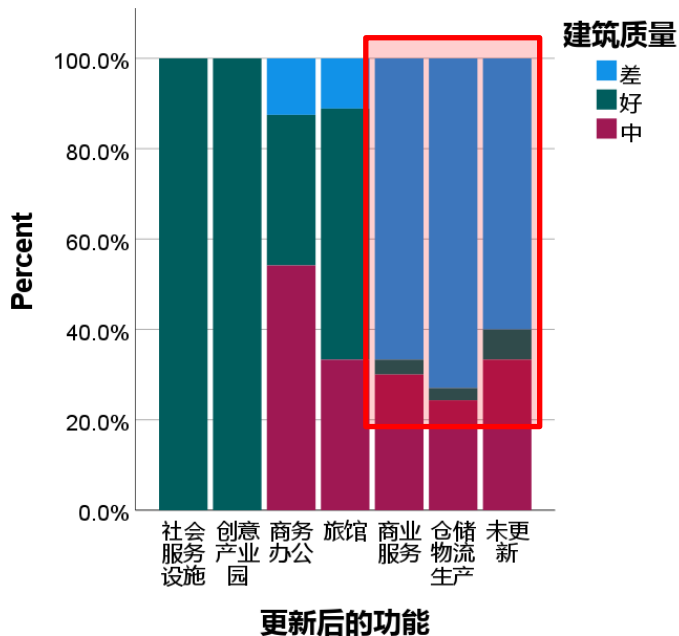
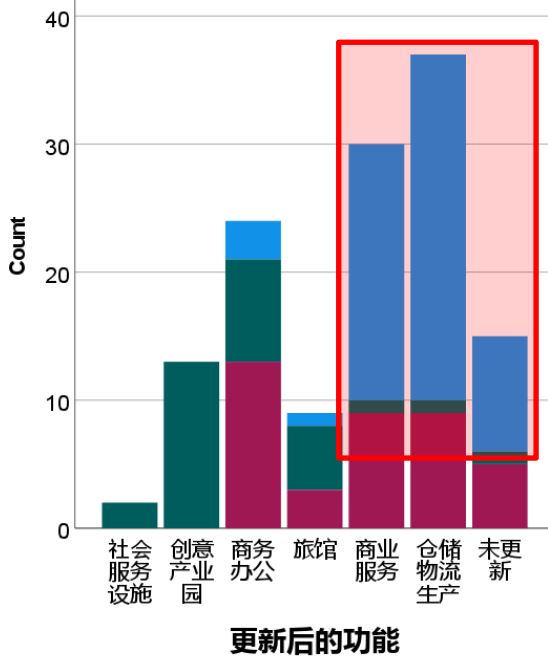
分类 vs. 分类：列联表分析

功能 Vs 建筑质量



已经相关了，然后……

表 + 图



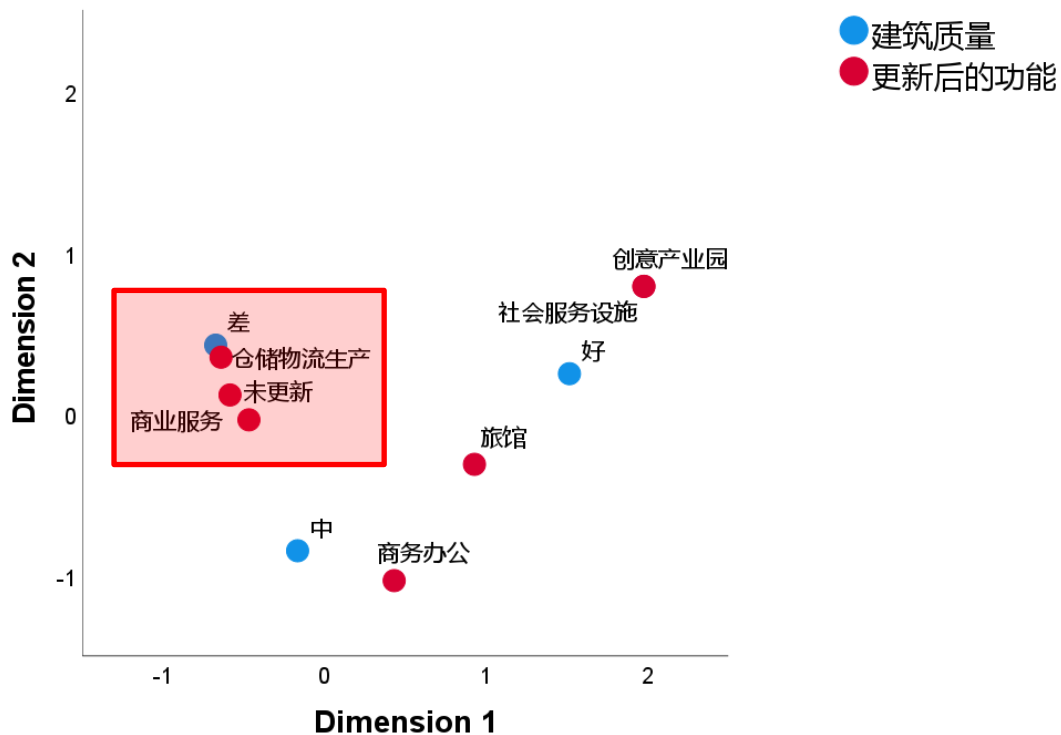
分类 vs. 分类：列联表分析

功能 Vs 建筑质量

表 + 图 + 对应分析



已经相关了，然后……



分类 vs. 分类：列联表分析

功能 Vs 建筑质量

分类与分类变量

Cramer's V

How much

Nominal. For nominal data (no intrinsic order, such as Catholic, Protestant, and Jewish), you can select Contingency coefficient, Phi (coefficient) and Cramér's V, Lambda (symmetric and asymmetric lambdas and Goodman and Kruskal's tau), and Uncertainty coefficient.

分类 vs. 分类: 列联表分析

功能 Vs 建筑质量

How much

Phi is the ratio of the chi-square statistic to the weighted total number of observations. It is the most "optimistic" of the symmetric measures, and unlike most association measures, does not have a theoretical upper bound when either of the variables has more than two categories.

Phi=0.829

Cramer's V is a rescaling of phi so that its maximum possible value is always 1. As the number of rows and columns increases, Cramer's V becomes more conservative with respect to phi.

分类与分类变量

Cramer's V

$$V = \sqrt{\chi^2 / (k - 1)n}$$

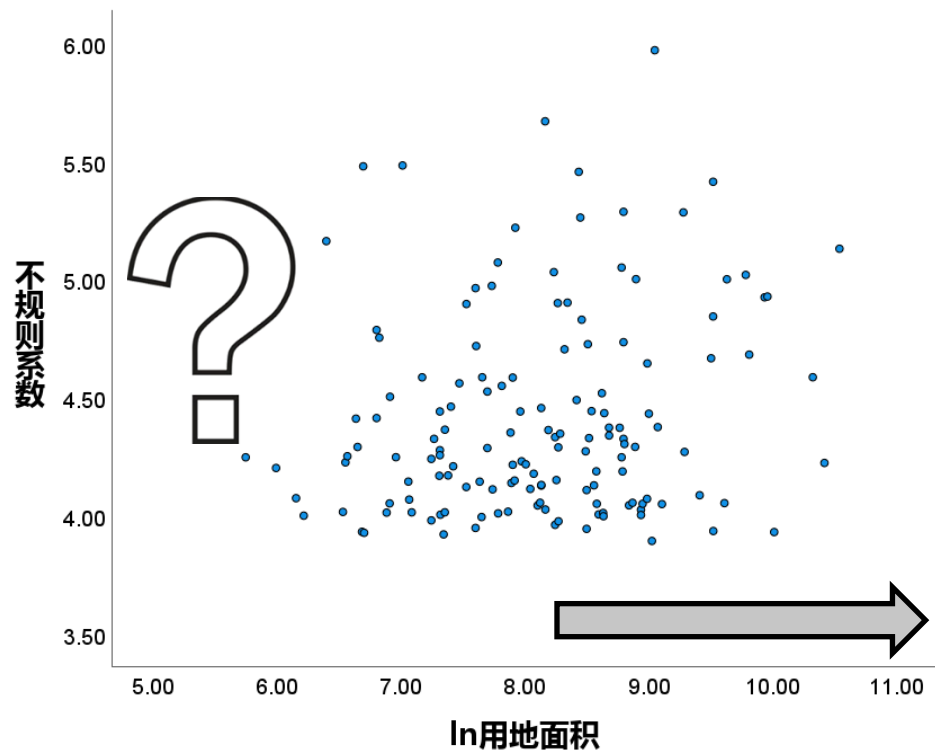
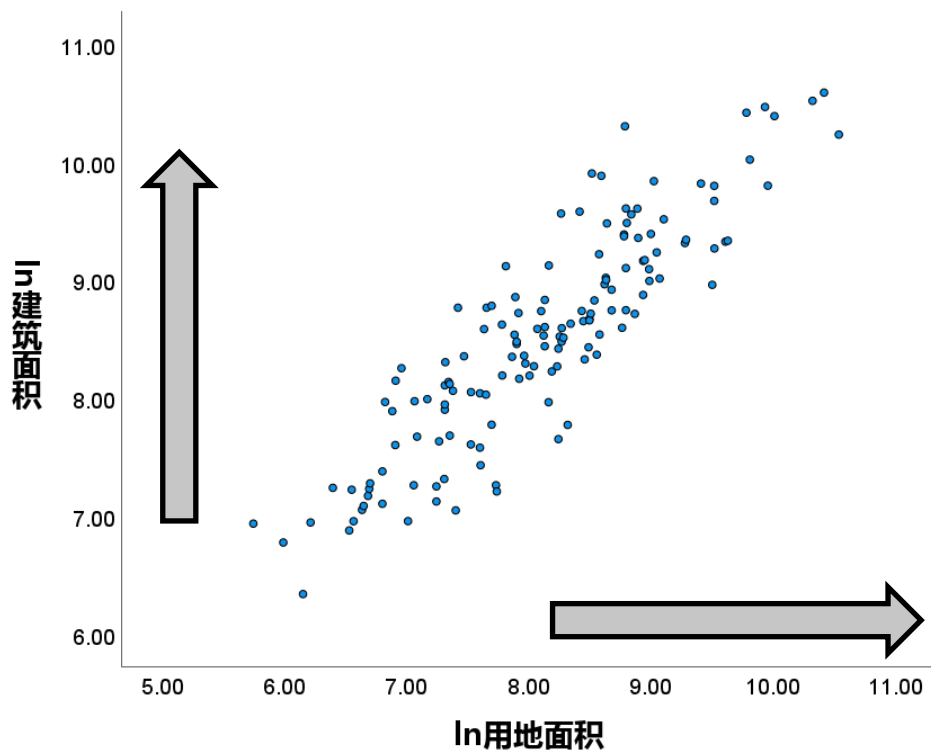
$$V_{(\text{功能}, \text{建筑质量})} = \mathbf{0.586}$$

The **contingency coefficient** takes values between 0 and SQRT[(k-1)/k], where k = the number of rows or columns, whichever is smaller. It becomes more conservative with respect to phi as the associations between the variables become stronger.

Contingency=0.638

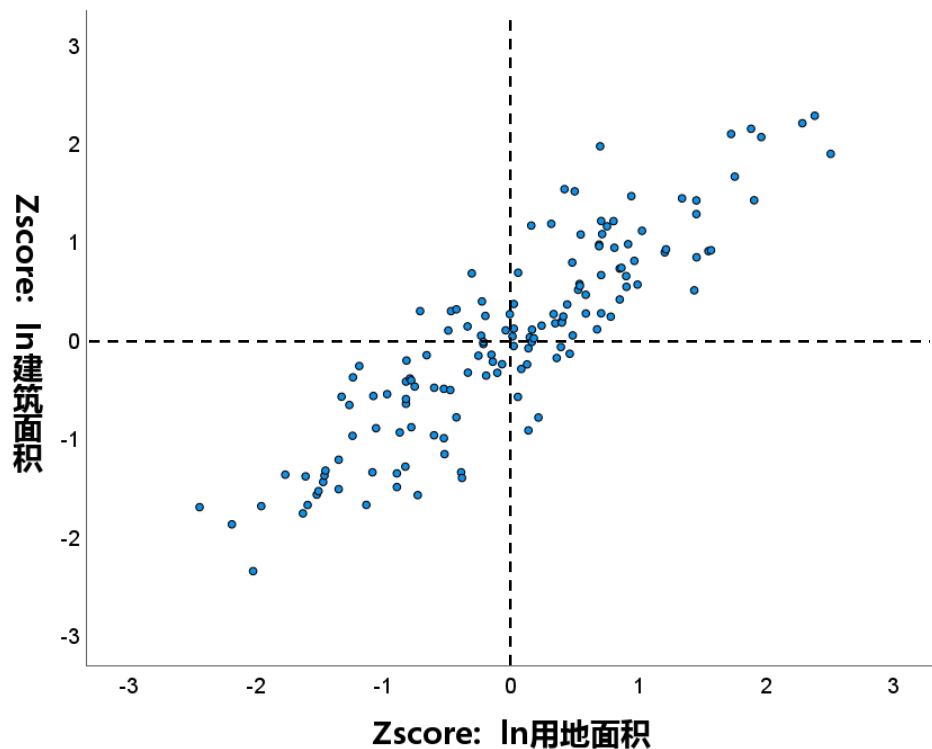
数值 vs. 数值：相关分析

用地面积 **Vs** 建筑面积 / 用地不规则系数



数值 vs. 数值：相关分析

用地面积 **Vs** 建筑面积



Z-score标准化：减去均值，除以标准差。

$$ZX_i = \frac{X_i - \bar{X}}{S_x} \quad ZY_i = \frac{Y_i - \bar{Y}}{S_y}$$

$$r = \frac{1}{n-1} \sum_{i=1}^n ZX_i ZY_i$$

皮尔逊
相关系数

数值 vs. 数值：相关分析

皮尔逊相关系数 (Pearson Correlation Coefficient)

- 用来衡量两个变量之间线性关系强度的一种统计量。
- 取值范围：-1~1
 - **正相关**： $r > 0$ ，一个变量增加时，另一个变量更有可能也增加。
 - **负相关**： $r < 0$ ，一个变量增加时，另一个变量更有可能降低。

r value	Interpretation
0.8 ~ 1.0	Very strong relationship (share most of the things in common)
0.6 ~ 0.8	Strong relationship (share many things in common)
0.4 ~ 0.6	Moderate relationship (share something in common)
0.2 ~ 0.4	Weak relationship (share a little in common)
0.0 ~ 0.2	Weak or no relationship (share very little or nothing in common)

数值 vs. 数值：相关分析

皮尔逊相关系数 (Pearson Correlation Coefficient)

- 用来衡量两个变量之间线性关系强度的一种统计量。
- 取值范围：-1~1
 - 正相关： $r > 0$ ，一个变量增加时，另一个变量更有可能也增加。
 - 负相关： $r < 0$ ，一个变量增加时，另一个变量更有可能降低。
- 假设检验：
 - 判断两个数值变量之间是否相关的主要依据。
 - 零假设：“皮尔逊相关系数”与“0”没有显著差异。
 - 前提假设：线性，正态性（大样本可放宽），同方差，无异常值。

数值 vs. 数值：相关分析

皮尔逊相关系数 (Pearson Correlation Coefficient)

Correlations

		In用地面积	In建筑面积	不规则系数
In用地面积	Pearson Correlation	--		
	N	144		
In建筑面积	Pearson Correlation	.890**	--	
	Sig. (2-tailed)	<.001		
	N	144	144	
不规则系数	Pearson Correlation	.178*	.111	--
	Sig. (2-tailed)	.033	.184	
	N	144	144	144

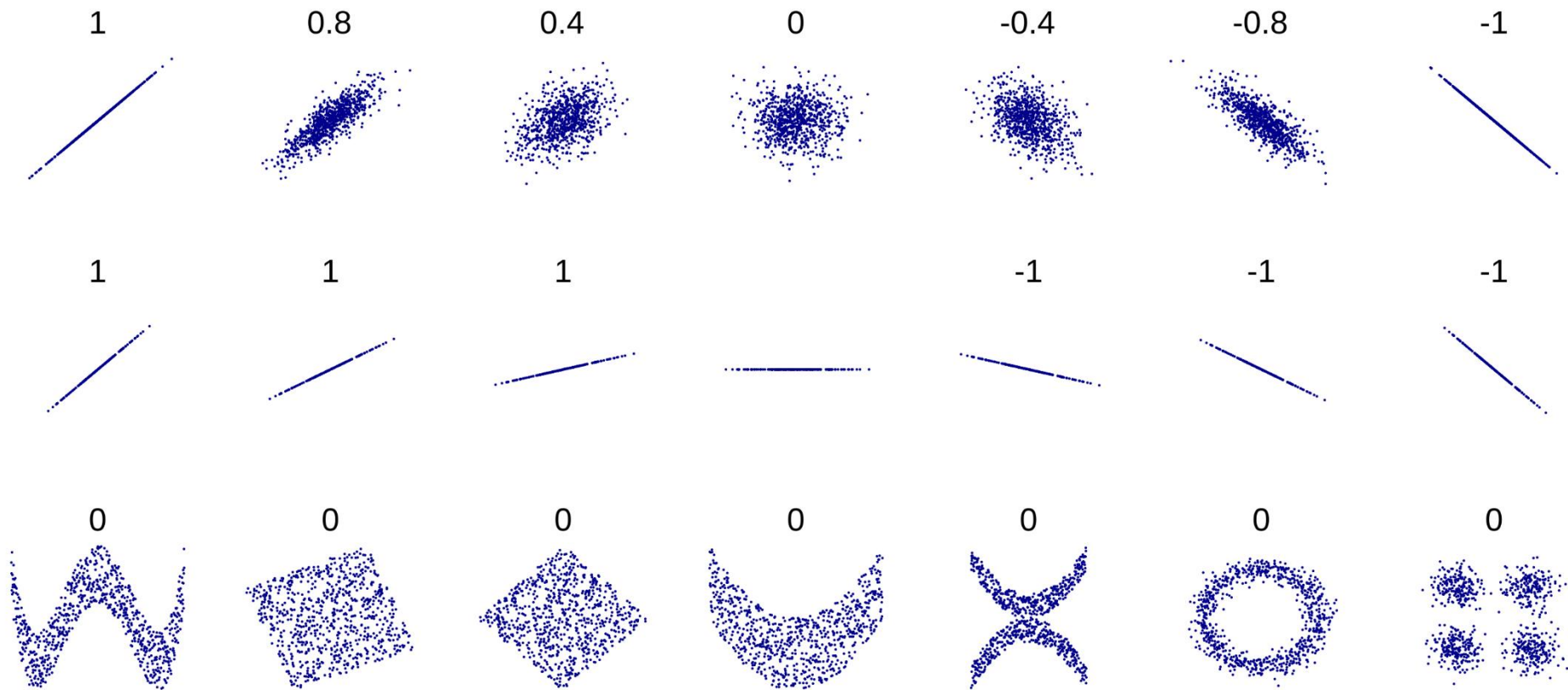
** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

- 用地面积与建筑面积在0.01的水平下强正相关。
- 用地面积与不规则系数在0.05的水平下弱正相关。
- 建筑面积与不规则系数在0.1的水平下不相关。

数值 vs. 数值：相关分析

皮尔逊相关系数 (Pearson Correlation Coefficient)



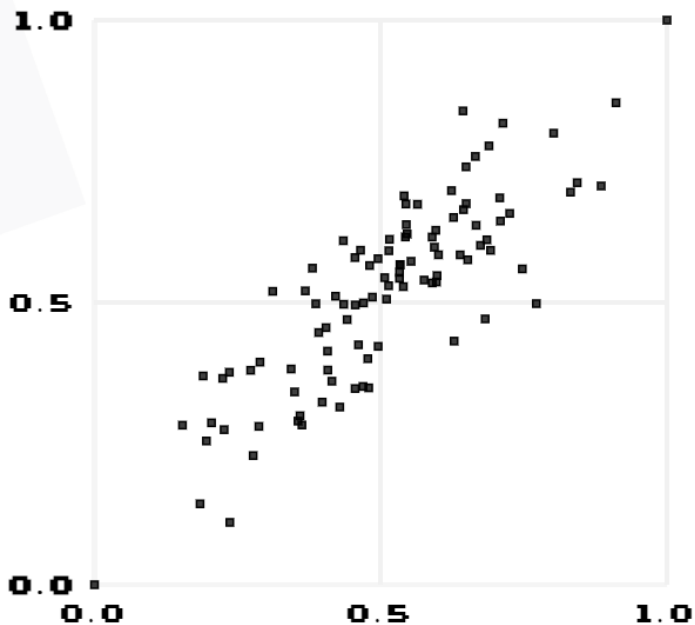
数值 vs. 数值：相关分析

皮尔逊相关系数 (Pearson Correlation Coefficient)



HIGH SCORE MAIN MENU

0



NEXT

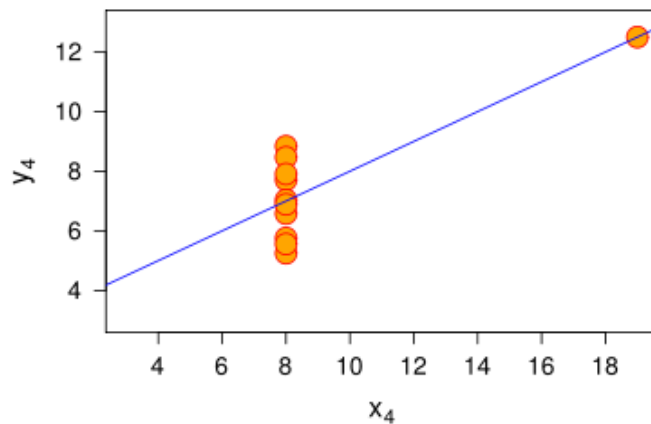
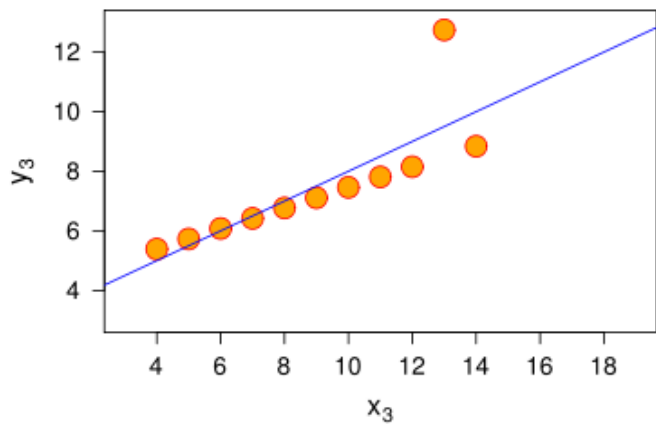
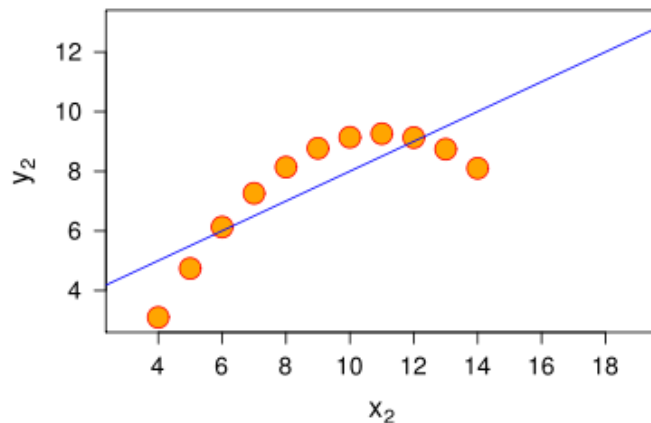
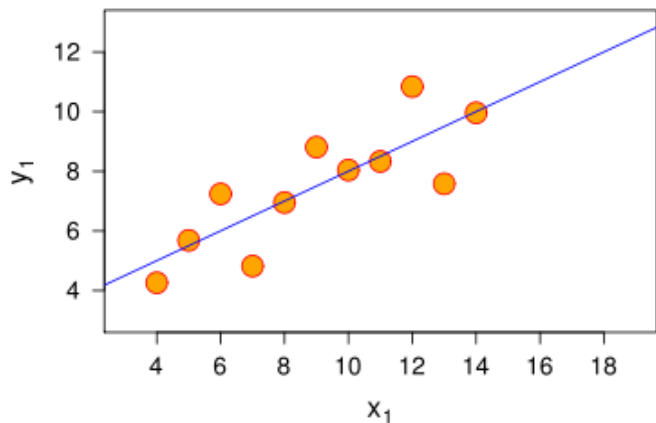
TRUE R	0.85
GUESSED R	0.90
DIFFERENCE	0.05
STREAKS	0
MEAN ERROR	0.09



<https://www.guessthecorrelation.com>

数值 vs. 数值：相关分析

皮尔逊相关系数 (Pearson Correlation Coefficient)

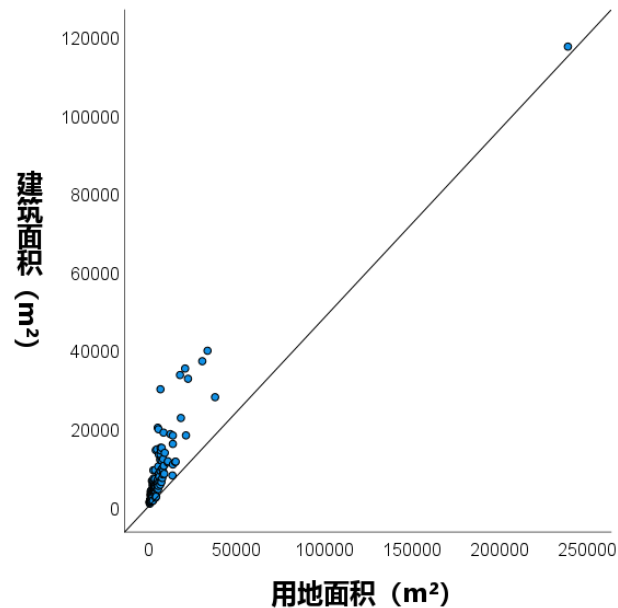


$r=0.816$

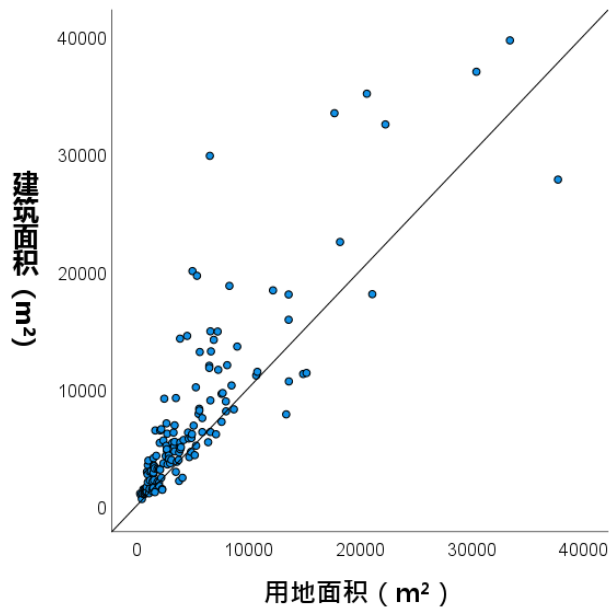
数值 vs. 数值：相关分析

用地面积 Vs 建筑面积

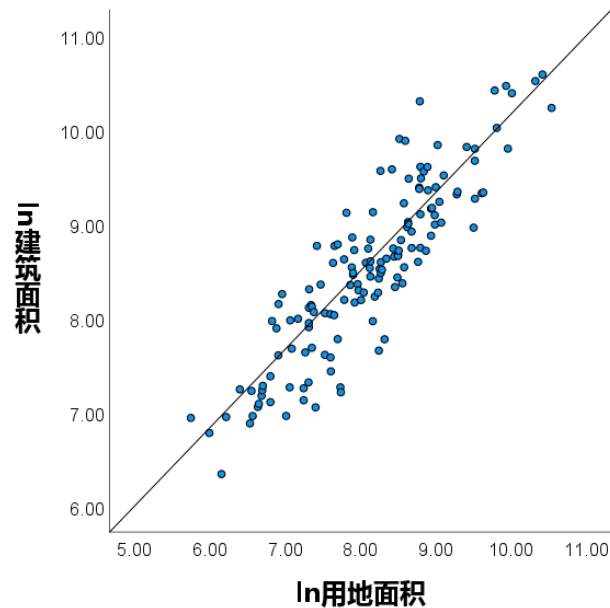
Pearson相关系数



$r=0.897$



$r=0.854$

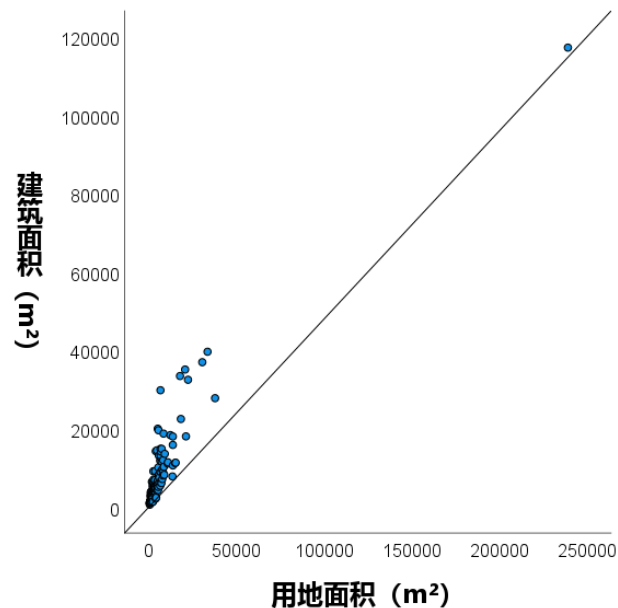


$r=0.890$

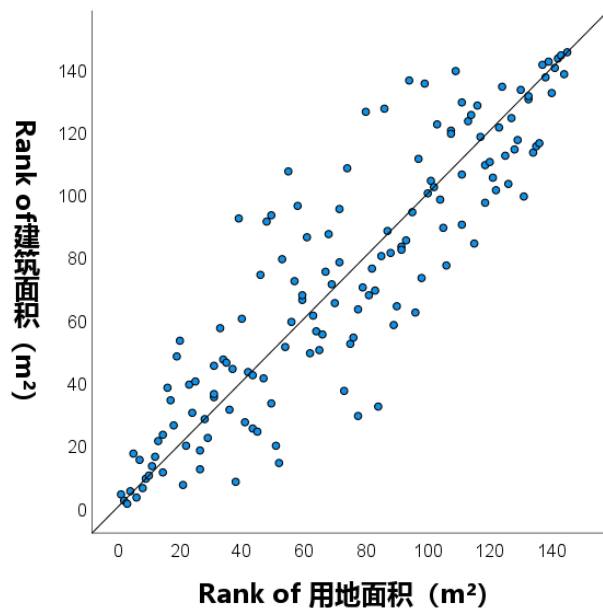
数值 vs. 数值：相关分析

用地面积 Vs 建筑面积

Spearman相关系数



$r=0.897$



$r=0.892$

- 对两个原始变量各自排序，计算秩 (rank) ；
- 利用秩变量计算 Pearson 相关系数；
- 结果即为斯皮尔曼相关系数，常记为 ρ 。

非参数方法的相关系数

- Spearman相关系数
 - 转变秩变量后，用秩变量计算的Pearson相关系数。
- Kendall's tau-b 相关系数
 - 计算一致对的数量 C 和不一致对的数量 D 。
$$\begin{array}{ll} x_1 < x_2 \text{ 且 } y_1 < y_2 & x_1 < x_2 \text{ 且 } y_1 > y_2 \\ x_1 > x_2 \text{ 且 } y_1 > y_2 & x_1 > x_2 \text{ 且 } y_1 < y_2 \end{array}$$
 - 利用公式计算： $\tau = (C - D)/n(n - 1)$
- 基于秩信息计算，虽然损失了部分信息，但更加稳健。
- 特别适合于原始数据严重非正态、异方差、离群值问题。

数值 vs. 数值：相关分析

用地面积 **Vs** 建筑面积

Correlations

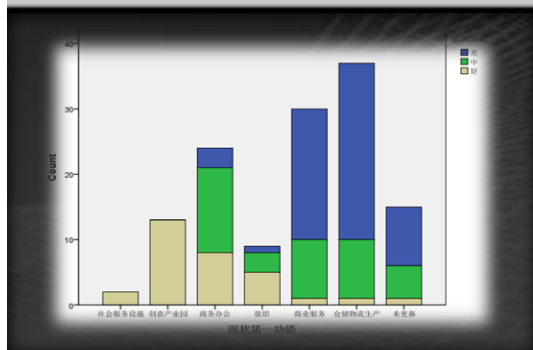
		用地面积 (m ²)	建筑面积 (m ²)	Rank of 用地面积 (m ²)	Rank of 建筑面积 (m ²)	In用地面积	In建筑面积	
Spearman's rho	用地面积 (m ²)	Correlation Coefficient	--					
		Sig. (2-tailed)	.					
		N	145					
建筑面积 (m ²)		Correlation Coefficient	.892**	--				
		Sig. (2-tailed)	<.001	.				
		N	145	145				
Rank of 用地面积 (m ²)		Correlation Coefficient	1.000**	.892**	--			
		Sig. (2-tailed)	.000	<.001	.			
		N	145	145	145			
Rank of 建筑面积 (m ²)		Correlation Coefficient	.892**	1.000**	.892**	--		
		Sig. (2-tailed)	<.001	.000	<.001	.		
		N	145	145	145	145		
In用地面积		Correlation Coefficient	1.000**	.892**	1.000**	.892**	--	
		Sig. (2-tailed)	.000	<.001	.000	<.001	.	
		N	145	145	145	145	145	
In建筑面积		Correlation Coefficient	.892**	1.000**	.892**	1.000**	.892**	--
		Sig. (2-tailed)	<.001	.000	<.001	.000	<.001	.
		N	145	145	145	145	145	

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

总结

两变量相关性评估方法

分类 Vs. 分类



是否
相关

列联表卡方检验显著性

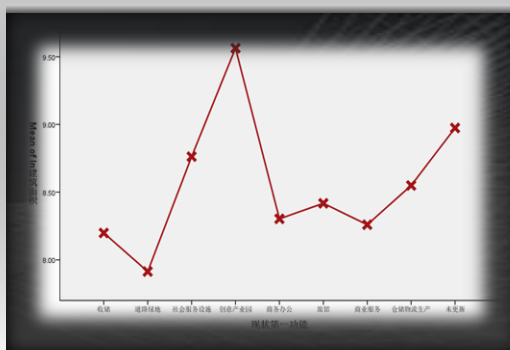
相关
程度

Cramer's V

如何
相关

直接观察、对应分析等

分类 Vs. 数值

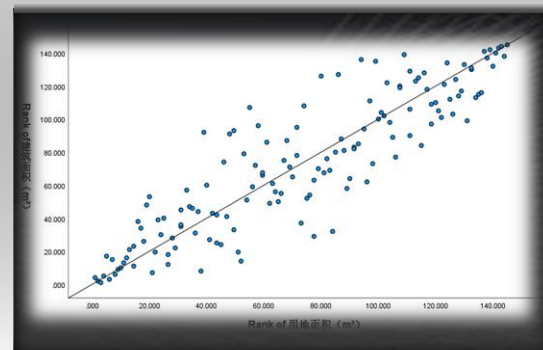


方差分析F检验显著性

Eta

事后两两比较, 具有推断性

数值 Vs. 数值



相关系数的显著性

Pearson/Spearman相关系数

直接观察