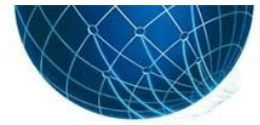


# 销售额预测的成功与失败案例 - 1

- ①多元回归模型
- ②构建多元回归模型的案例
- ③销售额预测成功与失败案例



选址赢家

China  
**MarketAnalyzer**<sup>™</sup>  
GIS商业选址与商圈分析软件

# 何为销售额预测?

所谓销售额预测就是使用各种数据域分析逻辑在开店前预测开店后的销售额，是在判断开店与否时的重要参考指标。

## 「各种各样的数据」

能够使用的数据到底是哪个 . . . ?

- 店铺属性数据
- 商圈数据
- 实地调查数据
- 问卷调查数据



## 「分析逻辑」

对我们来说到底哪个逻辑是对的?

- 哈夫模型
- 多元回归模型
- 类似店铺检索
- Neural模型

行行行色有很多、有不适合也有适合、花费的时间太长

# 门店数据的样本图

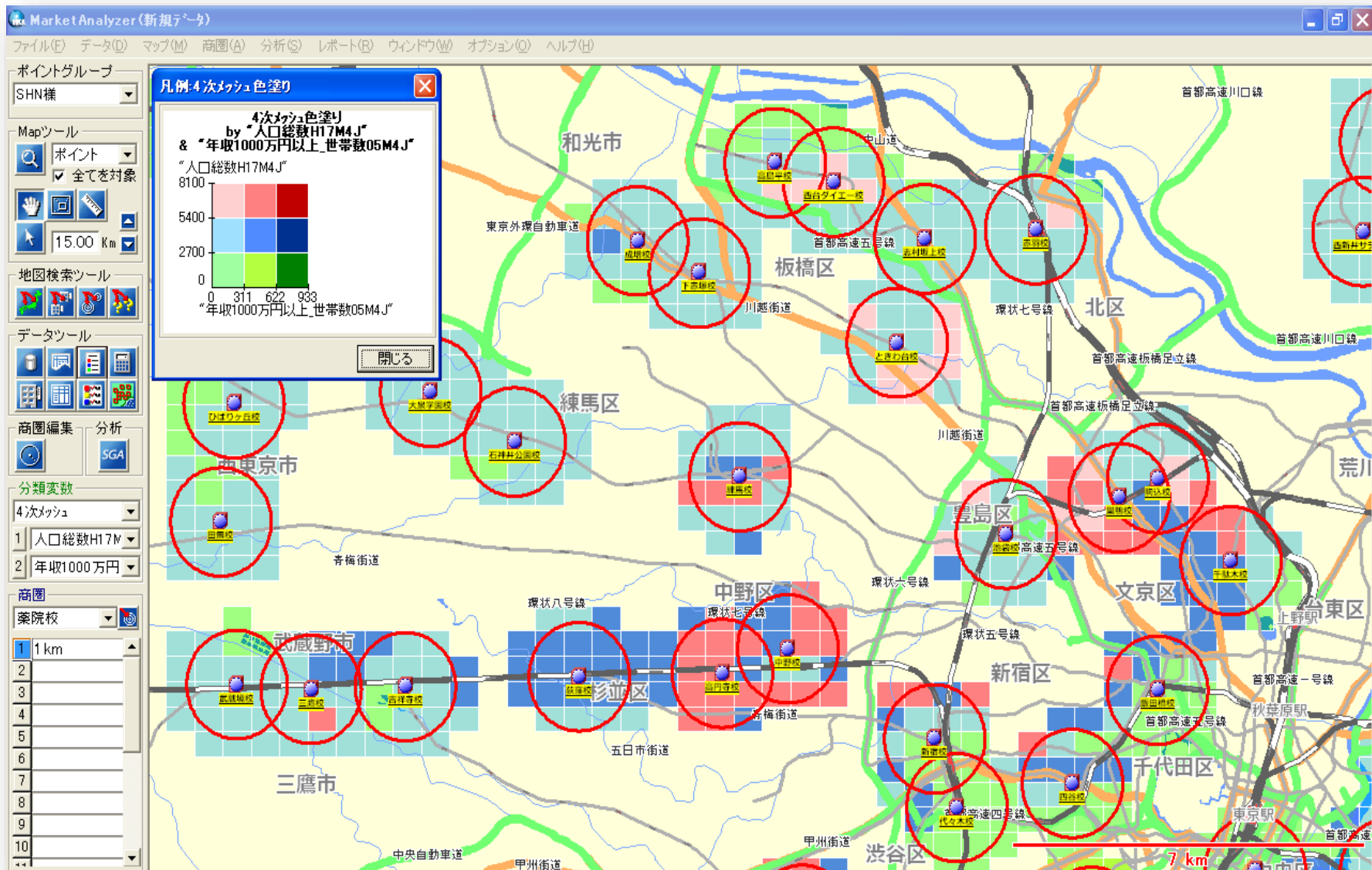
## 门店数据

本部 コード	本部 名称	店舗 コード	店舗 名称	商品カテゴリA				商品カテゴリB			住所
				商品A	商品B	商品C	小計	商品D	商品E	小計	
1	関東	1	店舗1	1,293	14,814	2,280	18,387	380	5,194	5,574	東京都目黒区原町1-2-3
1	関東	2	店舗2	15,095	30,992	20,920	67,007	8,724	11,824	20,548	東京都豊島区南池袋4-5-6
2	東北	3	店舗4	3,148	11,648	5,165	19,961	2,847	5,320	8,167	福島県いわき市土内町1-1



GEOCODING  
地理匹配

# 地图撒点 (门店落点)



# 结合门店数据 (合计)

门店数据



商圈数据

本部 コード	本部 名称	店舗 コード	店舗 名称	商品カテゴリA				商品カテゴリB			住所
				商品A	商品B	商品C	小計	商品D	商品E	小計	
1	関東	1	店舗1	1,293	14,814	2,280	18,387	380	5,194	5,574	東京都目黒区原町1-2-3
1	関東	2	店舗2	15,095	30,992	20,920	67,007	8,724	11,824	20,548	東京都豊島区南池袋4-5-6

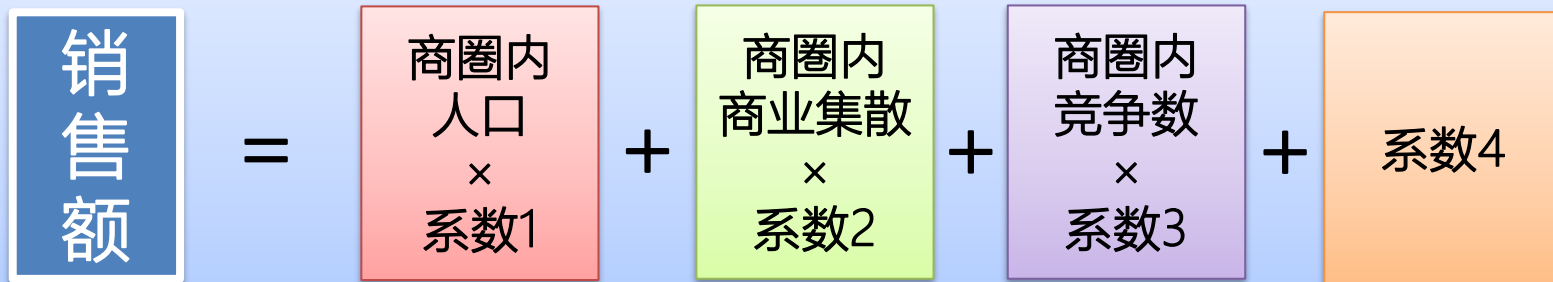
  

本部 コード	本部 名称	店舗 コード	店舗 名称	商品カテゴリA				商品カテゴリB			住所	人口総数	世帯数	15歳未満 人口	15-64歳 人口	65歳以上 人口
				商品A	商品B	商品C	小計	商品D	商品E	小計						
1	関東	1	店舗1	1,293	14,814	2,280	18,387	380	5,194	5,574	東京都目黒区原町1-2-3	6,442	2,342	1,219	4,421	765
1	関東	2	店舗2	15,095	30,992	20,920	67,007	8,724	11,824	20,548	東京都豊島区南池袋4-5-6	2,355	1,844	76	2,095	183
2	東北	3	店舗4	3,148	11,648	5,165	19,961	2,847	5,320	8,167	福島県いわき市土内町1-1	694	416	47	545	92

门店数据 + 商圈数据

# 何为多元回归分析?

本部 コード	本部 名称	店舗 コード	店舗 名称	商品カテゴリA				商品カテゴリB			住所	人口総数	世帯数	15歳未満 人口	15-64歳 人口	65歳以上 人口
				商品A	商品B	商品C	小計	商品D	商品E	小計						
1	関東	1	店舗1	1,293	14,814	2,280	18,387	380	5,194	5,574	東京都目黒区原町1-2-3	6,442	2,342	1,219	4,421	765
1	関東	2	店舗2	15,095	30,992	20,920	67,007	8,724	11,824	20,548	東京都豊島区南池袋4-5-6	2,355	1,844	76	2,095	183
2	東北	3	店舗4	3,148	11,648	5,165	19,961	2,847	5,320	8,167	福島県いわき市土内町1-1	694	416	47	545	92



结合多个要素说明销售额的时候，多元回归是非常有效的方法。

# 成功构建多元回归分析的要点

## 数据整理

即使是实际数据,原封不动使用的话在统计处理上有时没有效果。

## 规避多重共线性

投入项目间的相关性较高的话,实际的精度会较低。

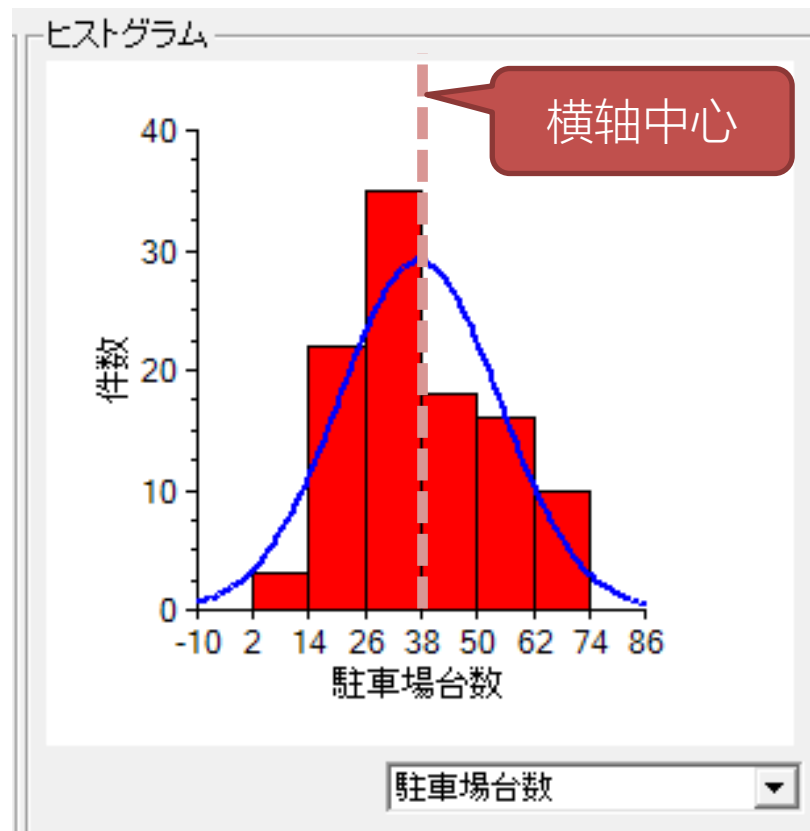
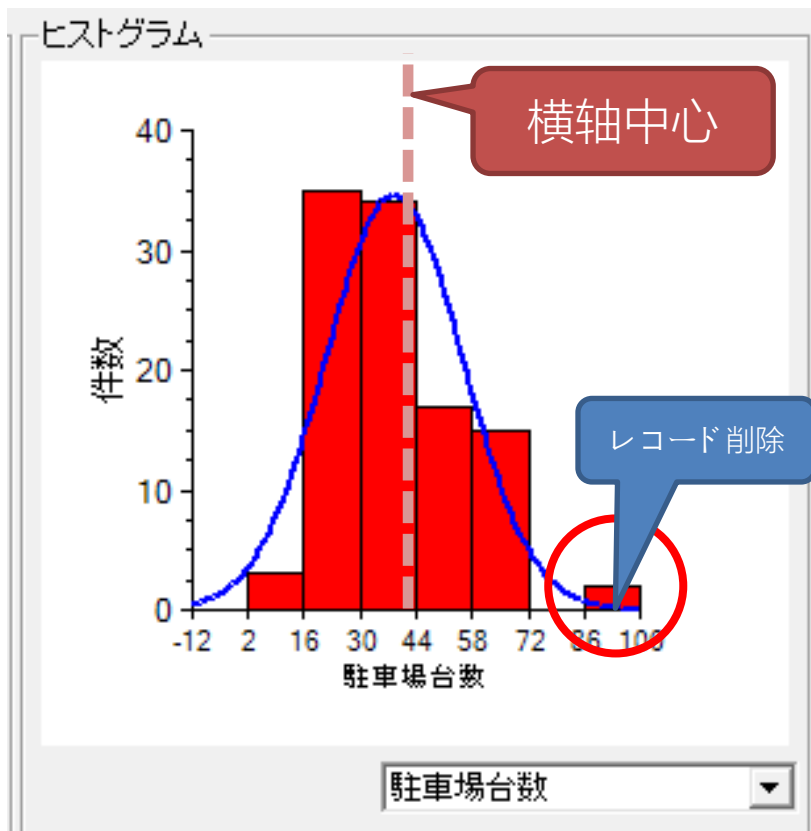
## 店铺分类

不要同时投入概念与商圈特性完全不同的店铺。

## 试错

需要有多次试错的环境。

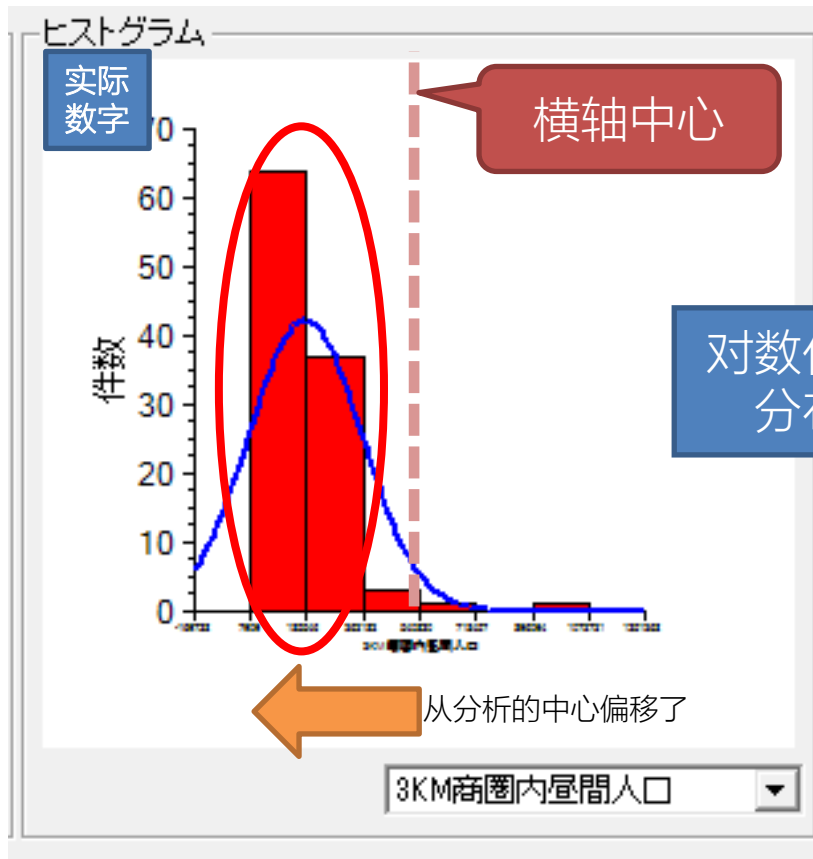
# 数据整理的要点 (排除异常值)



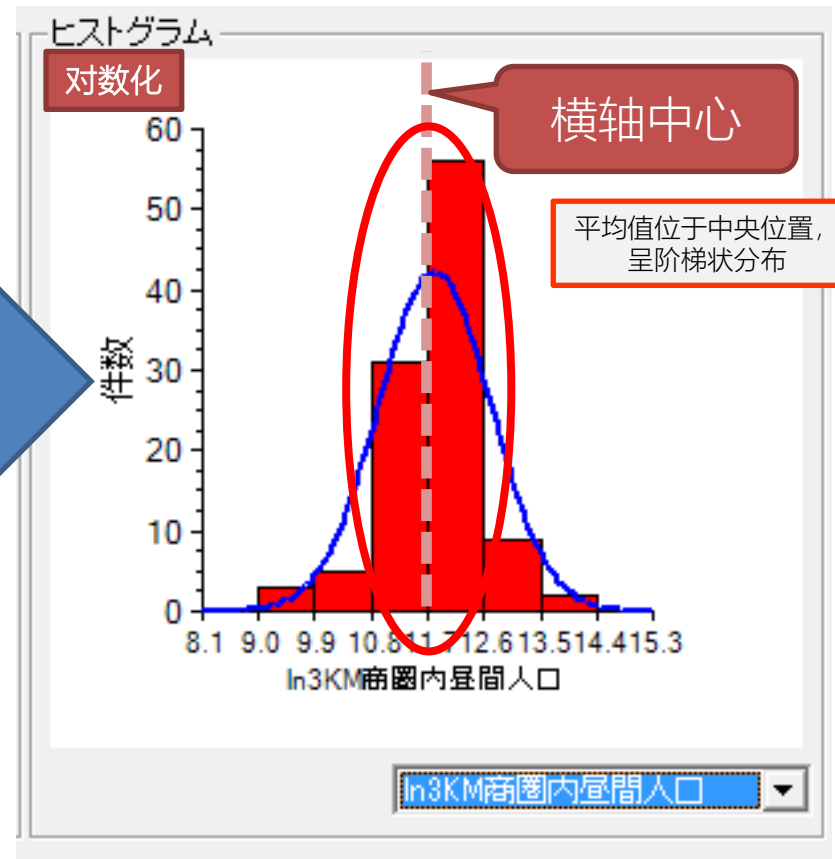
通过去除异常值有时可实现正规化分布。



# 数据整理的要点 (对数化)



对数化后  
分布

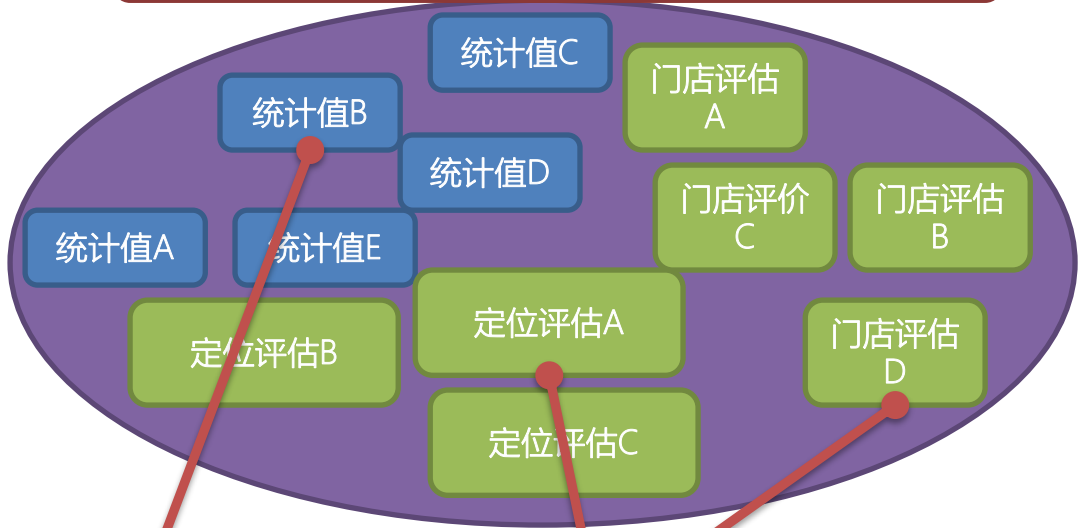


在实际数字的柱状图中值较大的数量较少、  
往值较少方偏移的时候，使用了对数的正规化分布比较有效。

# 计算出预测模型的概要图

## 多元回归分析 (阶梯法) 计算出极少异常值且最可信赖的销售额预测模型

### 销售额预测公式中使用的候选指标



用阶梯法来最优化指标间的组合后计算出每个指标的加乘系数

$$\text{统计值B} + \text{门店评估D} + \text{定位评估A} + \text{切片} = \text{销售额预测值}$$

抽出销售额预测公式中可以采用的指标。



在指标群中  
自动计算最优的组合。



计算出销售额预测公式

# 代表性的预测模型评估指标的看法和要点

R	R2乘	调整后R2乘	推算值的标准误差	AIC	Cp统计量
0.873	0.762	0.752	2206.443	1639.095	5

項目	偏回归系数	标准误差	标准化系数	t	有意确率	VIF
定数	9849.936	836.462		11.776	0	
居住人口_1km	0.105	0.039	0.318	2.719	0.008	5.805
白天人口_3km	0.033	0.009	0.99	3.652	0	31.193
到访人口_3km	-0.011	0.004	0.478	-2.366	0.02	17.327
卖场面积	5.777	0.88	0.323	6.561	0	1.03

调整后R2乘

取0~1的值。数值离1越近说明预测公式预测精度越高

标准化系数

各说明变量对预测公式的影响程度

有意确率

发生异常值的概率（危险率）、一般允许在0.05以内

VIF

多重共线性的指标。在5以下的话无需怀疑多重共线性

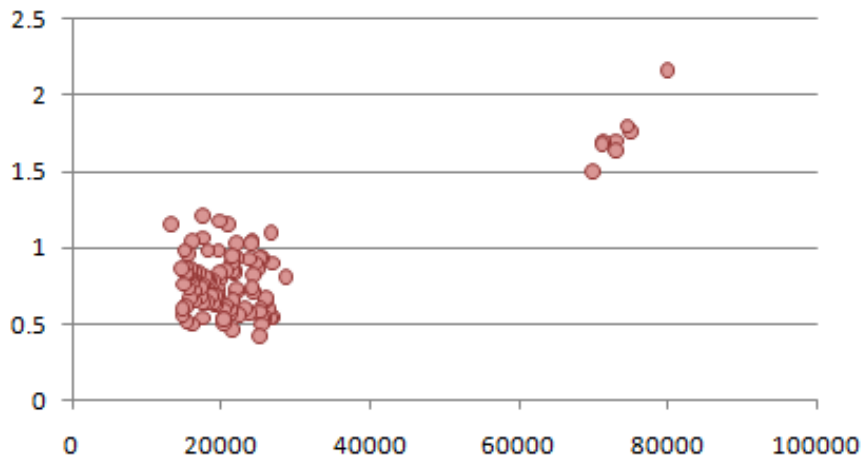
**要点**

调整后R2乘指就算再高，其他指标不高的话就欠缺可信度了

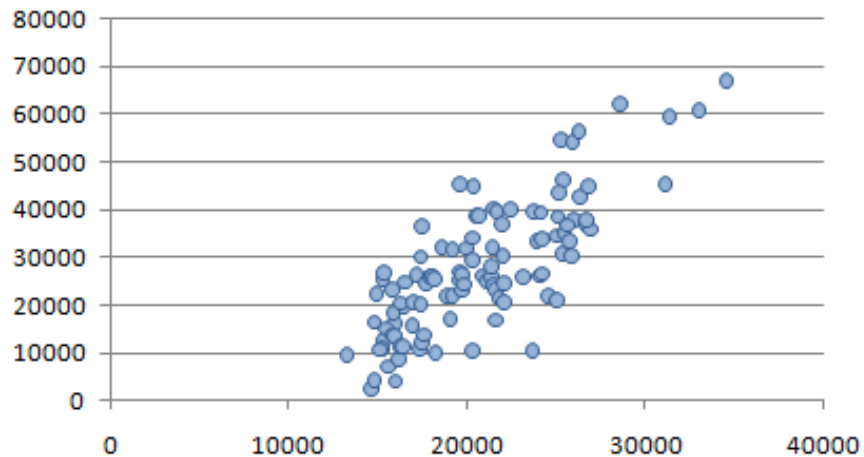
# 相关性分析的陷阱①

即使是相同的相关系数用散布图查看后有时也有很大的差异。

相関係数\_0.78



相関係数\_0.78



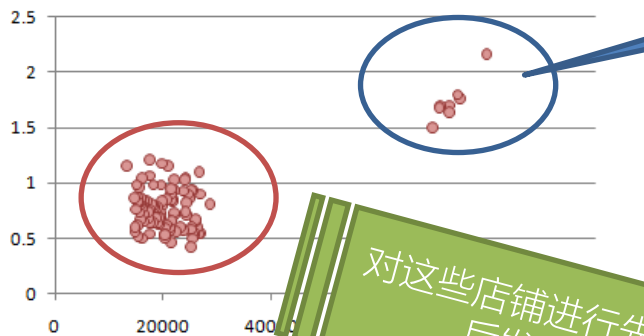
一部分店铺的数值有偏向性。

平均分散。  
没有发现有异常值的店铺。

# 相关性分析的陷阱②

排除有偏向性的店铺并进行相关性分析后会怎么样？

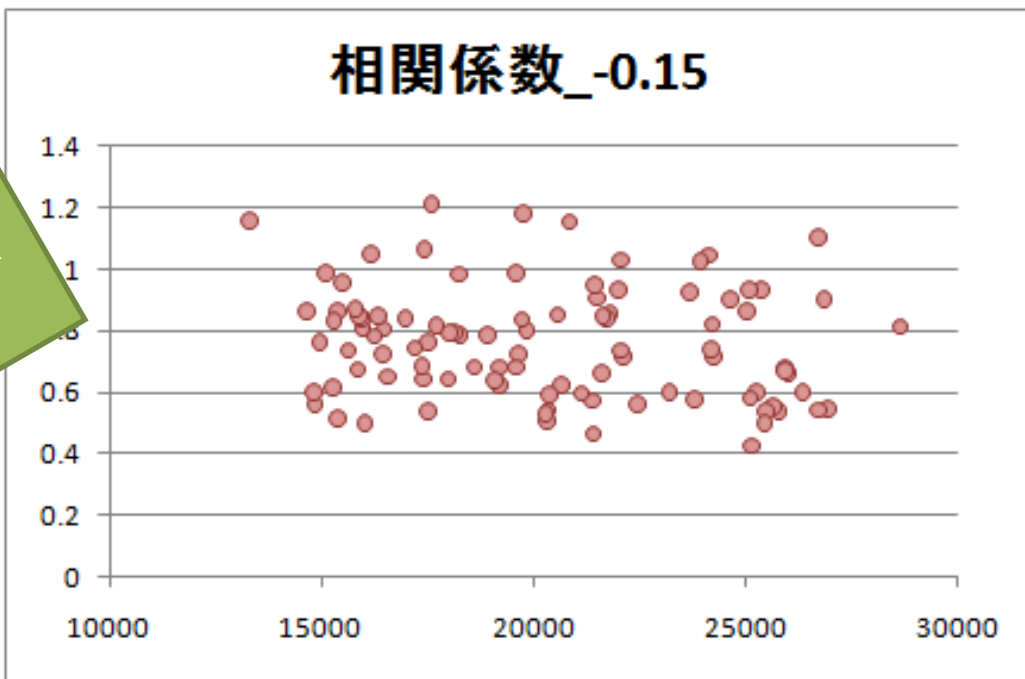
相関係数\_0.78



作为异常值排除

对这些店铺进行先关心分析  
后发现 . . .

相関係数\_-0.15



相关系数一下子降低了。

**0.78 → -0.15**

在判断是否作为预测模型的变量时，  
需要结合相关系数和散布图一起来查看。

# 销售额模型构建工具SiteAnalyzer™

完善销售额预测公式是非常重要的。  
公司内部的决心、不断变化的门店·商圈数据的再次投入· · ·  
必须有可反复试错的环境。

