

EXCEL 在一元线性回归分析中的应用

董凤鸣 周 萍
(菏泽学院经济系 山东 菏泽 274000)

摘要: 通过调查一个餐饮连锁店, 附近学校大学生的人数和季度销售额存在线性关系, 建立一个以大学生人数为自变量, 而季度销售额为因变量的回归模型, 用 EXCEL 进行回归分析并预测季度销售额, 说明 EXCEL 在回归分析中的应用。

关键词: 线性回归; 相关; 显著性水平; 一元线性回归

Abstract: This paper discusses the great effects of EXCEL on establishing a regressive equation by establishing a regressive equation with the sales and number of students nearby as contra - variables to estimate the present sales.

Key words: linear regression ; correlation ; variance ; level of significance

一、一元线性回归分析

1. 回归分析的特点

回归分析是在研究现象之间相关分析的基础上, 对自变量 x 和因变量 y 的变动趋势拟合数学模型进行数量推算的一种统计分析方法。进行回归分析, 要以现象之间存在相关关系为前提, 然后对自变量 x 和因变量 y 的变动拟合适宜的回归方程, 确定其定量关系式, 在对拟合的回归方程进行显著性检验, 最后利用所求得关系式进行推算和预测。

回归分析和相关分析的关系非常密切, 两者既有联系又有区别, 联系在于, 两者都是对客观事物数量依存关系的分析, 其中回归分析是在相关分析的基础上进行。如果没有定性的说明现象之间是否存在相关关系, 也没有对这种相关关系的密切程度作量的说明, 就不宜进行回归分析, 即使进行了回归分析, 也不会有什么实际意义。回归分析不仅可认识事物之间的关系, 更重要的是可运用这种关系推算、预测未来的发展趋势, 可见回归分析是相关分析的继续和拓展。通过回归分析对现象之间的相关关系拟合回归方程, 就有可能进行推算和预测, 相关分析才能更好地发挥作用。如果仅有相关分析而没有回归分析, 就如鱼有头无尾一样, 失去了统计分析的作用。

回归分析和相关分析的区别在于, 两者的概念和作用不同, 它们从不同的角度说明现象之间的依存关系。相关分析只能说明现象之间是否相关及相关方向和密切程度, 但不能说明一个现象发生一定量的变化, 另一个现象会对应发生多大变化。而回归分析通过建立适宜的回归方程则能够测出这种变化的量, 它是进行推算和预测的重要依据。

2. 一元线性回归模型

一元线性回归模型, 是分析两个变量之间相互关系的数学方程式, 其一般表达式为

$$\hat{y}=a+bx$$

式中, \hat{y} 表示因变量 y 的估计值, x 表示自变量, a, b 称为回归模型的待定参数, 其中 b 又称为回归系数。

上述的回归方程式在平面坐标系中表现为一条直线即回归直线。当 $b>0$ 时 y 随 x 的增加而增加, 两变量之间为正相关关系; 当 $b<0$ 时, y 随 x 的增加而减少, 两变量之间为负相关关系; 当 $b=0$ 时, y 为一个常量, 不随 x 的变动而变动。这样我们就判断现象之间的关系, 分析现象之间是否处于正常状态提供了一提一条标准。

3. 一元线性回归模型的建立

(1) 回归模型的建立程序

建立一元线性回归模型一般可分为四步、分析变量之间的相互关系, 通常是在理论定性分析的基础上采用相关表或相关图进行观察, 再计算相关系数; 通过检验相关系数的显著性, 判断相关系数的客观真实状况; 根据研究目的确定自变量和因变量; 根据调查的资料估计模型参数建立回归模型。

(2) 显著性检验

相关系数

相关系数, 是在线性相关的条件下, 说明两个现象之间相关关系紧密程度的统计分析指标。相关系数通常 r 或 r^2 , $r^2=SSR/SST=1-SSE/SST$, 其中

$$SSR=\sum(\bar{y}-\hat{y}_i)^2 \quad SSE=\sum(y_i-\hat{y}_i)^2 \quad SST=\sum(y_i-\bar{y})^2$$

且 $SST=SSR+SSE$

r 表示变量 x 和 y 线性相关方向和相关程度, 取值范围在 $-1 < r < 1$ 。

当 $-1 < r < 0$ 时相关图分布呈现出 y 随 x 的增加而减少的趋势, 即为负相关

当 $r=0$ 时, 相关图分布呈现出非规则状态, 变量 y 不受 x 的影响, 表明 x 和 y 之间没有线性相关关系, 但不能随意排斥其它关系, 如可能存在曲线相关关系。

当 $0 < r < 1$ 时, 相关分布呈现出 y 随 x 的增加而增加的趋势, 即为正相关。

当 $|r|=1$ 时, 相关散布点呈现出一条直线, 即 y 与 x 呈完全线性相关。 $r=1$, 表现为完全正相关; $r=-1$, 表现为完全负相关。

为了判断现象之间相关程度的高低, 通过相关系数判断相关关系密切程度的标准是: $r=0$, 表明 x 和 y 完全不相关; $0 < |r| < 0.3$ 时, 认为 x 和 y 不相关; $0.3 < |r| < 0.5$ 时, 认为 x 和 y 低度相关; $0.5 < |r| < 0.8$ 时, 认为 x 和 y 显著相关; $0.8 < |r| < 1$ 时, 认为 x 和 y 高度相关。

显著性检验

显著性检验, 可以根据相关系数、自由度 $(n-m)$ 其中 n 为样本容量, m 为回归模型中待定参数的个数) 和给定的显著水平 α (在社会经济现象中, 给定的显著水平 α 一般为 0.05), 从相关系数检验表中查出临界值 $r_{(n,m)}$, 据此判断其线性关系是否成立。如果 $|r| < r_{(n,m)}$, 表明在显著水平 α 条件下变量之间的线性关系是显著的, 因此将要建立的线性回归模型是很有意义的; 如果 $|r| > r_{(n,m)}$, 表明不宜建立线性回归模型, 需要对其进一步分析, 然后再作处理。

(3) 回归模型的检验

回归方程建立以后还需要对模型进行检验, 检验回归模型的代表性, 用 t 检验法, 若两个变量之间相关程度为高度相关, 方程有很高的代表性, 还不能说明这种直线相关关系是否可靠, 为了说明这种相关关系的可靠性, 必须对相关系数进行 t 检验。

$$\text{公式为: } t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}; r \text{ 与 } b \text{ 同号}$$

设, 统计量 t 服从 $t(n-2)$ 分布。根据一组样本量计算出 t 值, 再根据所给定的显著性水平 α 和自由度 $n-2$, 查 t 分布表, 找到相应的临界值 t_{α} 。若

$|t| > t_{\alpha}$ 表明 t 在统计上是显著的, 即总体的两个变量间存在线性关系, 这种关系是可靠的, 否则就认为两个变量间不存在线性关系。

二、EXCEL 在一元线性回归分析中的应用

我们有了在一元线性回归分析的基本知识, 通过一个实例来介绍如何用 EXCEL 对数据进行回归分析。考虑如下问题:

“阿曼德匹萨”是一个制作和外卖意大利匹萨的餐饮连锁店, 其主要客户群是在校大学生。为了研究各店铺销售额与店铺附近地区大学生人数之间的关系, 随机抽取了十个分店的样本, 得到的数据如下:

店铺编号	区内大学生数(万人)	季度销售额(万元)
1	0.2	5.8
2	0.6	10.5
3	0.8	8.8

4	0.8	11.8
5	1.2	11.7
6	1.6	13.7
7	2	15.7
8	2	16.9
9	2.2	14.9
10	2.6	20.2

对数据进行回归分析并预测区内大学生数 1.8 万的店铺 11 季度销售预测为多少?

我们将通过以下步骤对该问题进行解决:

1.回归

第一步,录入数据,录入结果见下图 1。

	A	B	C
1	店铺编号	区内大学生数 (X)	季度销售额 (Y)
2	1	0.2	5.8
3	2	0.6	10.5
4	3	0.8	8.8
5	4	0.8	11.8
6	5	1.2	11.7
7	6	1.6	13.7
8	7	2	15.7
9	8	2	16.9
10	9	2.2	14.9
11	10	2.6	20.2

图 1

第二步,作散点图,选中数据(包括自变量和因变量),选中数据后,数据变为蓝色。点击“图表向导”图标;或者在“插入”菜单中打开

“图表(H)”。图表向导的图标为。在弹出的图框左边一栏中选中“XY 散点图”,点击“完成”按钮,立即出现散点图的原始形式(图 2):

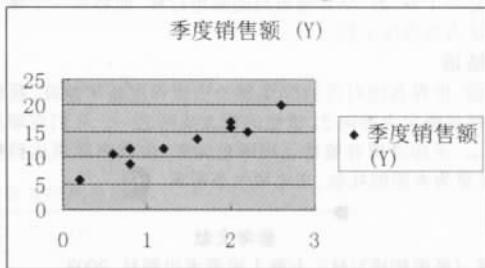


图 2

第三步,回归

观察散点图,判断点列分布是否具有线性趋势。只有当数据具有线性分布特征时,才能采用线性回归分析方法。从图中可以看出,本例数据具有线性分布趋势,可以进行线性回归。回归的方法如下

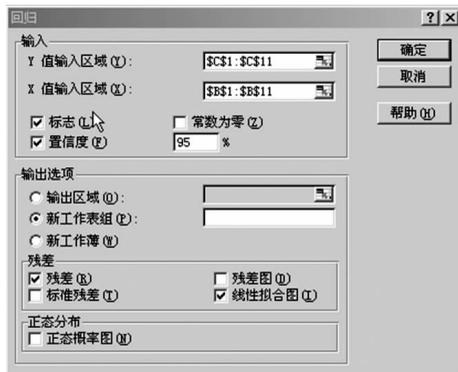


图 3-1 包括数据“标志”



图 3-2 包括数据“标志”

首先,打开“工具”下拉菜单,用鼠标双击“数据分析”选项(如果没有该选项,需要加载宏——>分析工具),弹出“数据分析”对话框。然后,选择“回归”,确定,弹出选项表:

进行如下选择: X、Y 值的输入区域(B1:B11, C1:C11), 标志, 置信度(95%), 新工作表组, 残差, 线性拟合图(图 3-1)

或者: X、Y 值的输入区域(B2:B11, C2:C11), 置信度(95%), 新工作表组, 残差, 线性拟合图(图 3-2)

注意:选中数据“标志”和不选“标志”, X、Y 值的输入区域是不一样的:前者包括数据标志:区内大学生数(X),季度销售额(Y)

然后,确定,取得回归结果(图 4)

回归统计	
Multiple R	0.950122955
R Square	0.90273363
Adjusted R Square	0.890575334
标准误差	1.382931669
观测值	10

方差分析					
	df	SS	MS	F	Significance F
回归分析	1	142	142	74.24837	2.55E-05
残差	8	15.3	1.9125		
总计	9	157.3			

Coefficients: 标准误差		t Stat	P-value	Lower 95%	Upper 95%
Intercept	6.0922603	6.503336	0.000187	3.872473	8.127527
区内大学生数 (X)	5.0580265	8.616749	2.55E-05	3.661906	6.338094

RESIDUAL OUTPUT		
观测值	Y 季度销售额	残差
1	7	-1.2
2	9	1.5
3	10	-1.2
4	10	1.8
5	12	-0.3
6	14	-0.3
7	16	-0.3
8	16	0.9
9	17	-2.1
10	19	1.2

图 4 线性回归结果图

最后,读取回归结果如下:

截距: a=6; 斜率: b=5; 相关系数: R=0.950; 测定系数: R²=0.90273363; F 值: F=74.24837; t 值: t=8.616749; 标准离差(标准误差): s=1.3829; 回归平方和: SSR=142; 剩余平方和: SSE=15.3; y 的误差平方和即总平方和: SST=157.3。

2.建立回归模型,并对结果进行检验

模型为 $\hat{y}=6+5x$

至于检验, R、R²、F 值、t 值等均可直接从回归结果中读出。实际上, R=0.950, R²>0.90 检验通过。有了 R 值, F 值和 t 值均可计算出来,其中: t 值的计算公式和结果为:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-k-1}}} = \frac{0.950}{\sqrt{\frac{1-0.950^2}{10-1-1}}} = 8.61 > 2.306$$

回归结果中给出了残差(图 5), 据此可以计算标准离差。首先求残差的平方 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$, 然后求残差平方和 $S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 15.30$, 于是标准

$$离差为 s = \sqrt{\frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{10-1-1} \cdot 15.30} = \sqrt{\frac{15.30}{8}} = 1.383$$

$$\text{于是 } \frac{s}{y} = \frac{1.383}{11.3} = 0.1222 < 15\% = 0.15$$

图 5 预测值及残差

观测值	预测 季度 销售额 (Y)	残差	残差平方		
1	7	- 1.2	1.44	标准差 S	1.383
2	9	1.5	2.25	s/y 的均值	0.1222
3	10	- 1.2	1.44		
4	10	1.8	3.24		
5	12	- 0.3	0.09		
6	14	- 0.3	0.09		
7	16	- 0.3	0.09		
8	16	0.9	0.81		
9	17	- 2.1	4.41		

10	19	1.2	1.44		
		残差平方和	15.3		

3. 预测分析

显著性检验后,我们要对数据进行预测分析。考虑预测区内学生 1.8 万人的时候季度销售额预计为多少?

由于回归方程建立之后, $\hat{y}=a+bx$ 就是 y 的无偏估计, 故当 $X_0=1.8$ 时, 季度销售额的期望值 EY_0 的点估计值为 $\hat{y}_0=a+bx_0=6+1.8 \times 5=15$ 。

而 y_0 的 0.95 预测区间近似为 $(\hat{y}_0 - 1.96(Q/(n-2))^{1/2}, \hat{y}_0 + 1.96(Q/(n-2))^{1/2}) = (15 - 1.96 \times \sqrt{15.3/8}, 15 + 1.96 \times \sqrt{15.3/8}) = (12.289, 17.711)$, $Q=SSE$ 。

三、总结

本文只是通过一个具体的实例介绍 EXCEL 在一元线性回归中的应用, 如何建立数据表, 如何分析并回归, 最后如何对得到的结果进行检验分析和预测。当然我们可以用其他软件来做, 但基本原理一样。至于多元线性回归要比这复杂, 但采用矩阵论的话, 原理都是一样的。

(上接第 131 页)以组成任意的、多边的、几何的等等各种各样的形, 所以在景观整合中可以起到渗透功能。本文所称绿色元素的面主要是指草坪、绿墙等。

(1) 草坪 草坪也称草皮, 栽植人工选育的草种作为矮生密集型的植被, 经养护修剪形成整齐均匀的覆盖。草坪除了具有滞尘、阻滞地表径流、防止土壤冲蚀、补充地下水、净化地面水、降低地面水的温度、减少地面的辐射热和减弱地面对噪声的反射等生态功能外, 在观赏上如绿色的地毯, 柔美轻快, 对其它景物能起到很好的衬托作用。另外它还给人们提供了一个洁净舒适的休闲场所。景观设计中, 景区之间并没有十分明显的界限, 而是你中有我, 我中有你, 渐而变之。而草坪的延伸、渗透, 恰恰起到连接空间的作用, 给人一种在不知不觉中景物已发生变化的感觉, 有着良好的空间体验。

(2) 垂直绿化 垂直绿化是指利用攀缘植物绿化墙壁、栏杆、棚架、杆柱及陡直的山石等。在房屋外壁进行垂直绿化能起到降温、保温、减少噪声的反射、减缓墙面本身的风化等作用。垂直绿化不但自身可创造生动活泼的造型, 具有独特的山林景观效果, 还有效地把平面景观渗透到立面空间, 使空间景物融为一体。

4.4 体的围合功能 体是最丰富的元素, 其中包含着点、线、面的各种组合。它能把景观中的碎片有机地围合为一个整体。本文所称绿色元素的体主要是指公园、空中花园等。

(1) 公园 公园是为城市居民提供室外休息、观赏、游戏、运动、娱乐的绿化用地。公园里可包括名胜古迹、纪念碑、植物园、动物园、展览馆、剧场等等各式各样的人文景观和自然景观, 就象是一个被围合成的“景观集合体”, 成为城市中的不可缺少的绿肺。

(2) 空中花园 空中花园是指利用主体建筑物的屋顶、平台、阳台、露台等多层次空间, 结合园林植物、微地形、水体和园林小品等造园因素, 创造出不同使用功能和性质的园林景观。空中花园不但把建

筑技术和绿化美化融合为一体, 更把地面与空间连成一片绿色世界。

5. 发展前景

西蒙兹说“我们可以说, 景观设计师的终生目标和工作就是帮助人类, 使人、建筑物、社区、城市以及他们的生活同生活的地球和谐共处。”所以现代意义上的景观规划设计是以协调人与自然的相互关系为己任。绿色元素作为自然环境的代表, 其生态效应、视觉感受和人文景观渗透和影响着人类生活和建筑环境的方方面面, 它是把高度机械化生活中的人同大自然联系起来的枢纽。近年来, 植物栽培技术的不断发展使得绿色植物的生长范围更加宽广, 屋顶种植和室内种植的发展, 使得建筑和绿化的融合更趋全方位和立体化, 建筑内部空间的营造也更趋人性和丰富。利用市花市树与其他植物或小品、构筑物相得益彰地配置, 使城市景观在整合的过程中, 又同时发展了自己的特色, 可以赋予城市景观浓郁的文化气息。绿是生命之道, 绿色元素的运用不再只是一个乔、灌、草合理布局的简单过程, 而将是一个赋予城市景观生命活力的系统工程。

6. 结语

当前, 世界各国对可持续发展的重视程度越来越高, 我国也把可持续发展战略定为面向 21 世纪的两大战略之一。我们应顺应这种时代的要求, 更加合理有效地运用绿色元素, 使城市景观达到整合的同时, 满足更为本质的社会、文化和生态需求。

参考文献

- [1] 毛溪, 《平面构成》[M], 上海人民美术出版社, 2003.
- [2] 刘捷, 《城市形态的整合》[M], 东南大学出版社, 2004.
- [3] 曹杰勇, 城市滨水住宅区设计探析[J], 建筑学报, 2004(4).

作者简介: 邱海玲(1974-), 女, 新疆农业大学硕士研究生。

(上接第 127 页)险问题我们可以提出一些针对性的措施:

1. 推进资本市场改革, 拓宽中小企业的融资渠道 较快资本市场改革, 降低中小企业入市融资的门槛, 不但有利于发挥资本市场在全社会高效率调节资源配置的功能, 而且可以从根本上决定广大中小企业在快速成长中长期性权益资本短缺的发展瓶颈, 有利于财务风险的降低。

2. 完善市场机制, 适度减轻宏观调控的频率与力度 进一步完善市场机制, 理顺市场机制调整社会经济运行的功能与作用, 可以很大程度上减少经济局部问题现象的发生, 从而减少政府对宏观经济的干预。这可以减少中小企业因宏观调控而带来的不确定因素, 从而降低财务风险。

3. 增强中小企业管理人员财务风险意识, 建立适合于中小企业的财务预警体系 增强中小企业管理人员尤其是财务管理

风险意识、重视与防范财务风险是降低中小企业财务风险的一个重要途径; 同时建立一个合适于中小企业的财务预警体系, 使中小企业的管理者能够随时了解企业的实时财务风险指标水平, 将有助于将财务风险防患于未然。

参考文献

- [1] 王贤平, 中小企业融资难问题研究, [J], 企业改革与发展, 2006.3.
- [2] 李善民, 中小企在泛珠三角经济区投融资行为分析, [J], 广东社会科学, 2004.04.

作者简介: 刘桂信, 江苏, 无锡, 会计师, 无锡市建瓴房地产有限公司财务主管。