

**基于回归分析方法研究高管薪酬对上市公司绩效影响**

**小组成员：钱泽昊、刘艳、宋新新、孙乙格阁、李哲勋、王曼、臧雨露**

**19级工业工程1班**

**郑州大学管理工程学院**

# **工业工程软件与应用小组作业报告**

**郑州大学管理工程学院**

# 工业工程软件与应用小组作业

#### ——基于回归分析方法研究高管薪酬对上市公司绩效影响

## 0 摘要

经过本学期工业工程软件与应用的学习，我们掌握了Minitab软件的使用（析因实验设计、响应曲面法）、优化类算法（模拟退火、遗传算法等）、回归分析、BP神经网络、贝叶斯预测等内容。

本篇小组报告主题是房地产企业高管薪酬对于上市公司绩效的影响，报告中部分经济审计学相关理论来源于郑州大学商学院18级审计学本科生于弋航，我们主要采用了课程中所学的回归分析对数据进行分析，使用Python编程求解以及SPSSPRO求解对比结果，通过以上操作复现了管薪酬对于上市公司绩效的影响这篇论文的研究。

数据来源：CSMAR国泰安数据库（沪深A股上市公司2016到2020年五年间的数据）。

小组成员分工：钱泽昊，程序编写，资料汇总；刘艳、宋新新，SPSSPRO软件验证；孙乙格阁、王曼、李哲勋，文档撰写、数据挖掘。

## 1 研究内容

### 1.1 研究背景

最近几年以来，社会各界对高管的薪资水平依旧十分关注。企业绩效是企业价值最直观的表现，提高企业绩效是公司发展的重要目标，在提高企业绩效过程中，管理层起着关键性的作用。高管人员是企业人力资源中的关键一环，因为高管身为一个公司的决策层，且作为公司经营管理的中心人物，肩负着指定公司的发展战略和执行董事会决策的关键任务，一定程上决定了公司未来的发展方向，优秀的高管人员很大程度上代表着公司的核心竞争力。因此，一个公司必需要调动高管人员的积极性，塑造其良好的价值观，进而能使高管人员的工作对本公司的绩效产生正向影响。尽管我国现在高管薪酬的激励制度较过去已有很大提升，但是和发达国家相比，我国的上市公司中高管薪酬制度仍然存在很大的缺陷，仍需不断的探索与完善。本文选用49家金融业和房地产上市公司2016-2020年的数据探究高管的薪酬水平和公司绩效的关系。研究结果表明，高管的薪酬激励会对公司绩效水平产生很强的促进作用。

### 1.2 模型的建立

|  |  |  |  |
| --- | --- | --- | --- |
| **变量选择表** | **变量名称** | **变量符号** | **表示** |
| 被解释变量 | 企业绩效 | ROE | 净资产收益率 |
| 解释变量 | 高管薪酬 | LNCOMP | 董事、监事、高级管理人员年度薪酬总额的自然对数 |
| 控制变量 | 公司规模 | SIZE | 年末总资产自然对数 |
| 控制变量 | 债务期限成本 | DC | 短期债务/长期债务+短期债务 |
| 控制变量 | 资本结构 | ALR | 资产负债率 |
| 控制变量 | 董事会规模 | BOD | 董事会总人数 |

### 1.3 模型的设计

模型设计

### 1.4 在经济和管理层次上的意义

研究上市公司高管薪酬与绩效之间的关系具有十分重要的理论和实践意义。

首先，从理论上看，由于我国的市场还处于初步发展阶段，仍然不成熟，导致研究人员对高管薪酬与公司绩效两者关系的研究成果不同，各学者看法不统一，存在许多弊端。本文总结了现存的相关研究成果，对公司的绩效以及高管薪酬现状进行分析，并结合相关理论，经过理论分析，假设提出和实证分析后，对高管薪酬水平设置如何为公司绩效发挥最大效用提出建议，是对前人丰富研究成果的补充，希望可以借此对提高公司的绩效发挥借鉴作用，为公司治理提供新的角度和思路。

其次，从实践上来看，一家公司薪酬制度设计完美与否影响了企业的绩效水平，所以制定好与自身发展情况相符合的薪酬制度体系十分重要。尤其针对现如今劳动力市场供需不平衡，出现了“用人难”的现象，对此如果能设计出精准且科学化的薪酬激励体系，对激发员工的工作热情和提升企业绩效水平至关重要。从宏观层面讲，本文对高管人员高薪酬的现状进行了具体分析并发现其内在问题，并借此对高管收入分配地优化和改进提供了建议。从微观层面讲，由于房地产行业和金融行业的高管收入一直占据国内上市公司高管薪酬收入的前列，且这两个行业之间的相似点颇多，具有很大参考价值，因此该研究可以加深社会各界人士对高管薪酬的现状以及存在的问题的了解，同时也能使高管更准确地评估自身价值，因此能更好的发挥自身对企业绩效提升的贡献。

## 2 使用工具——SPSSPRO

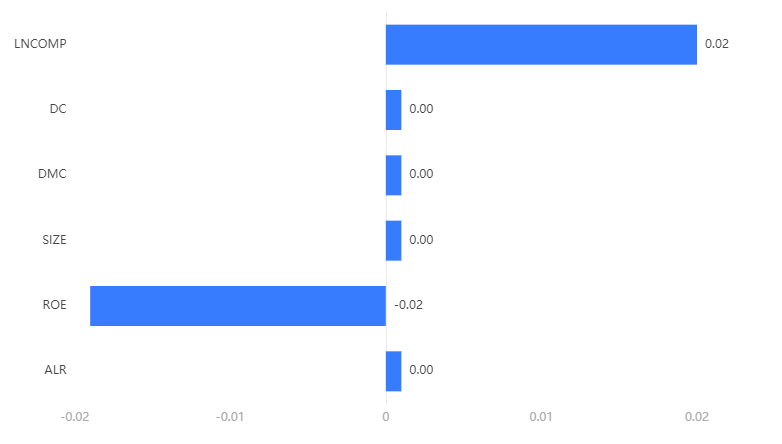
### 2.1 SPSSPRO软件介绍

SPSSPRO（ScientificPlatformServingforStatisticsProfessional）“专业统计服务的科学平台”，它是一款区别于SPSS、SAS传统客户端模式的全新在线数据分析平台。

### 2.2 使用SPSSPRO导出的结果

### 分析步骤 1. 通过训练集数据来建立线性回归模型。 2. 通过建立的线性回归模型计算特征重要性。 3. 将建立的线性回归模型应用到训练、测试数据，得到模型评估结果。 4. 若线性回归模型选择了数据洗牌功能，每次运算的结果不一样，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。

### 线性回归模型结果 **输出结果1：特征重要性**



**图表说明：**

上表展示了模型各个特征的权重，其中负号代表负向影响，正号代表正向影响。 **输出结果2：模型评估结果**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **MSE** | **RMSE** | **MAE** | **MAPE** | **R²** |
| **训练集** | 3.267 | 1.807 | 1.287 | 14.416 | 0.016 |
| **测试集** | 4.031 | 2.008 | 1.553 | 16.669 | -0.367 |

**图表说明：**

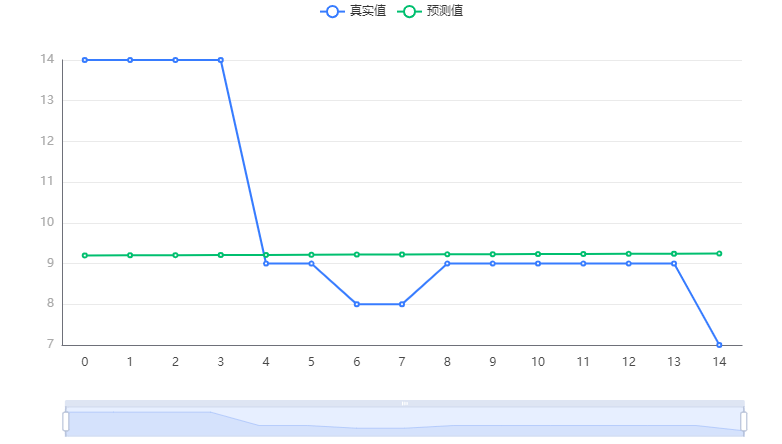
上表中展示了交叉验证集、训练集和测试集的预测评价指标，通过量化指标来衡量线性回归模型的预测效果。其中，通过交叉验证集的评价指标可以不断调整超参数，以得到可靠稳定的模型。  
●MSE（均方误差）：预测值与实际值之差平方的期望值。取值越小，模型准确度越高。  
●RMSE（均方根误差）：为MSE的平方根，取值越小，模型准确度越高。  
●MAE（平均绝对误差）：绝对误差的平均值，能反映预测值误差的实际情况。取值越小，模型准确度越高。  
●MAPE（平均绝对百分比误差）：是MAE的变形，它是一个百分比值。取值越小，模型准确度越高。  
●R²：将预测值跟只使用均值的情况下相比，结果越靠近1模型准确度越高。  
**输出结果3：测试数据预测结果**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **预测结果Y** | **BOD** | **LNCOMP** | **DC** | **DMC** | **SIZE** | **ROE** | **ALR** |
| 9.200973634597608 | 14 | 171 | 172 | 172 | 172 | 171 | 172 |
| 9.204106368652912 | 14 | 172 | 173 | 173 | 173 | 172 | 173 |
| 9.207239102708215 | 14 | 173 | 174 | 174 | 174 | 173 | 174 |
| 9.21037183676352 | 14 | 174 | 175 | 175 | 175 | 174 | 175 |
| 9.213504570818824 | 9 | 175 | 176 | 176 | 176 | 175 | 176 |
| 9.216637304874126 | 9 | 176 | 177 | 177 | 177 | 176 | 177 |
| 9.21977003892943 | 8 | 177 | 178 | 178 | 178 | 177 | 178 |
| 9.222902772984735 | 8 | 178 | 179 | 179 | 179 | 178 | 179 |
| 9.226035507040038 | 9 | 179 | 180 | 180 | 180 | 179 | 180 |
| 9.229168241095342 | 9 | 180 | 181 | 181 | 181 | 180 | 181 |
| 9.232300975150647 | 9 | 181 | 182 | 182 | 182 | 181 | 182 |
| 9.23543370920595 | 9 | 182 | 183 | 183 | 183 | 182 | 183 |
| 9.238566443261254 | 9 | 183 | 184 | 184 | 184 | 183 | 184 |
| 9.241699177316557 | 9 | 184 | 185 | 185 | 185 | 184 | 185 |
| 9.244831911371861 | 7 | 185 | 186 | 186 | 186 | 185 | 186 |

**图表说明：**

上表格为预览结果，只显示部分数据，全部数据请点击下载按钮导出。  
上表展示了线性回归模型对测试数据的预测情况。

**输出结果4：测试数据预测图**



## 3线性回归相关数学原理与Python代码实现

### 3.1 线性回归的简单介绍

线性回归是利用数理统计中回归分析，来确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法，运用十分广泛。

回归分析中，只包括一个自变量和一个因变量，且二者的关系可用一条直线近似表示，这种回归分析称为一元线性回归分析。如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为多元线性回归分析定义：线性回归在假设特证满足线性关系，根据给定的训练数据训练一个模型，并用此模型进行预测。为了了解这个定义，我们先举个简单的例子；我们假设一个线性方程Y=2x+1,x变量为商品的大小，y代表为销售量；当月份x=5时，我们就能根据线性模型预测出y=11销量；对于上面的简单的例子来说，我们可以粗略把y=2x+1看到回归的模型；对于给予的每个商品大小都能预测出销量；当然这个模型怎么获取到就是我们下面要考虑的线性回归内容。

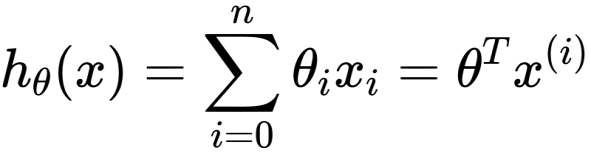
### 3.2 线性回归的数学推导

假设一个银行的贷款额度y，是由年龄x1，工资x2决定的。θ1为年龄的参数，θ2为工资的参数。θ可以理解为权重参数（θ0为偏置项）

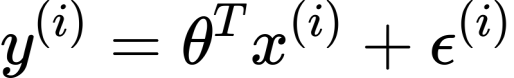
**拟合平面公式：**

****拟合平面公式****

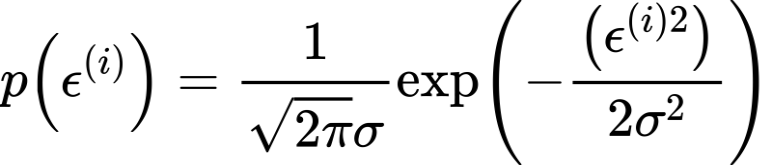
**整合（矩阵计算更高效所以转化成矩阵公式）：**

****

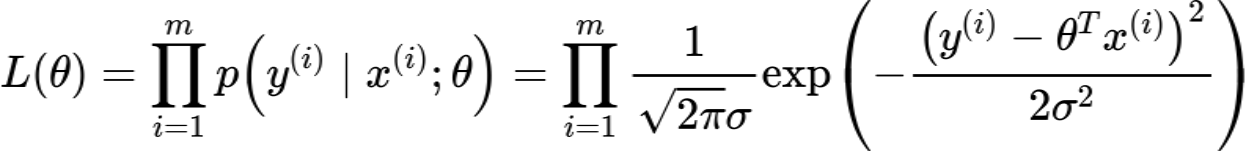
**误差：真实值和预测值之间肯定是存在误差的（用ε来表示误差），对于每个样本yi为真实值**

****

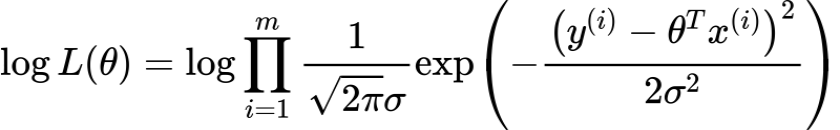
**将误差带入高斯分布：**

****

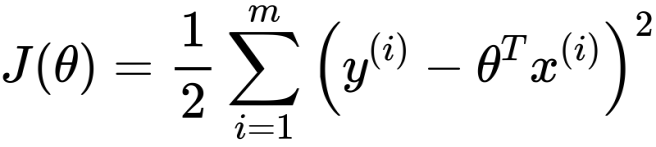
**似然函数：**

****

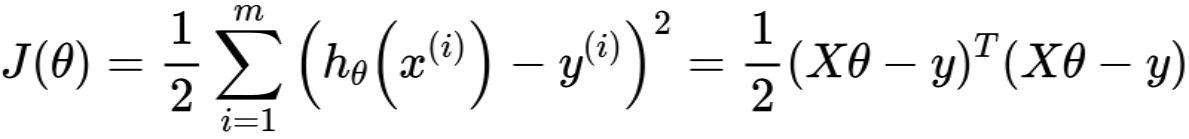
**对数似然：**

****

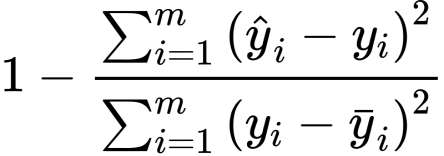
**最小二乘法：**

****

**目标函数：**

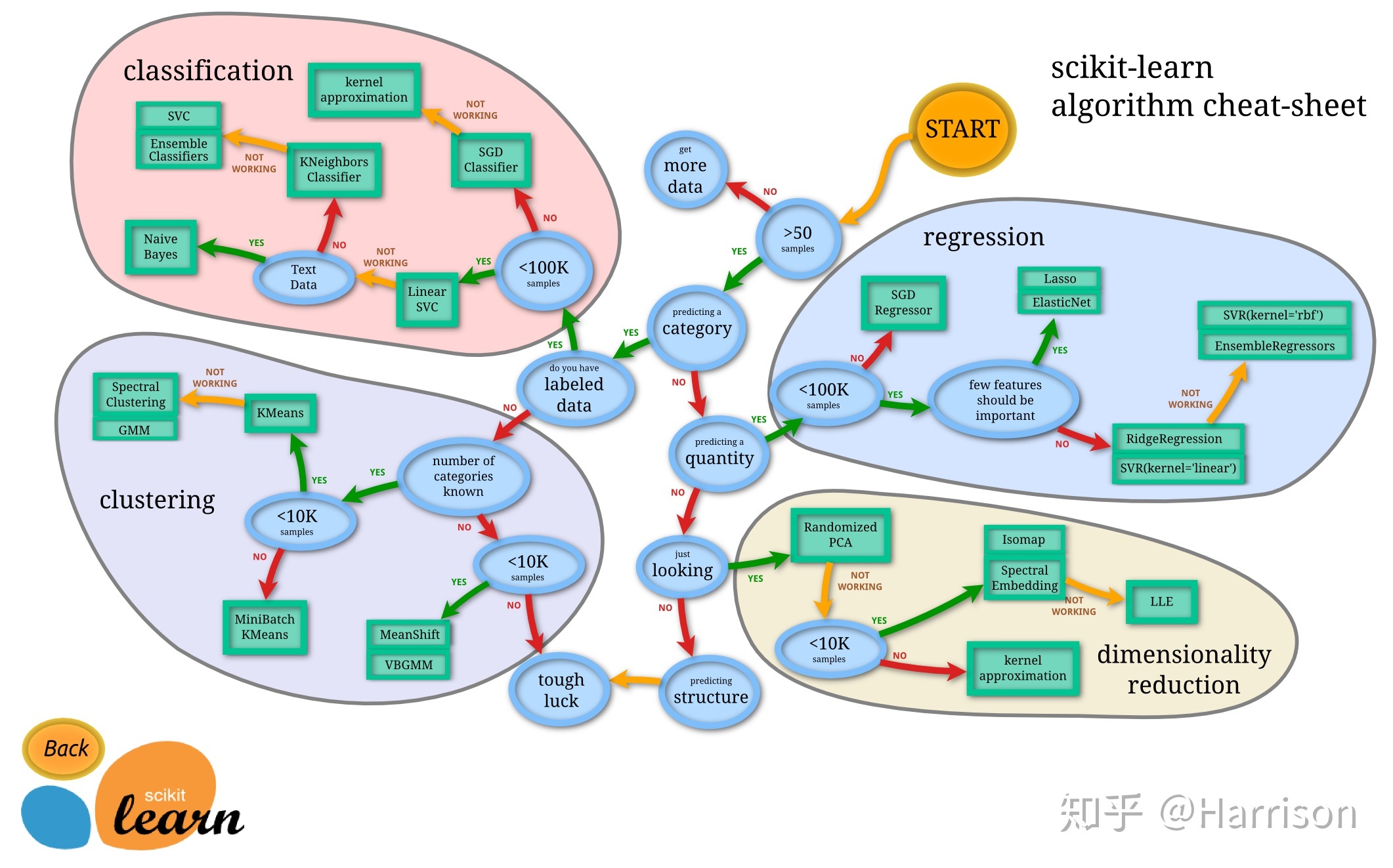
****

**评估方法R2:**

****

### 3.3 Python机器学习库Sklearn简介

scikit-learn是Python重要的机器学习库，我们使用sklearn中的linear\_model引入线性回归模块，scikit-learn简称sklearn，支持包括分类，回归，降维和聚类四大机器学习算法。还包括了特征提取，数据处理和模型评估者三大模块。  
　　sklearn是Scipy的扩展，建立在Numpy和matplolib库的基础上。利用这几大模块的优势，可以大大的提高机器学习的效率。  
　　sklearn拥有着完善的文档，上手容易，具有着丰富的API，在学术界颇受欢迎。sklearn已经封装了大量的机器学习算法，包括LIBSVM和LIBINEAR。同时sklearn内置了大量数据集，节省了获取和整理数据集的时间。



### 3.4 一元线性回归算法实例

1. **import** pandas as pd #导入pandas库
2. **import** numpy as np #导入numpy库
3. **from** sklearn.linear\_model **import** LinearRegression #导入机器学习库中的线性回归模块
4. data=pd.DataFrame({'square\_feet':[150,200,250,300,350,400,600],
5. 'price':[5450,6850,8750,9650,10450,13450,16450]})
6. #创建一组7行2列的数据，square\_feet为房屋面积，price为对应价格
7. data\_train=np.array(data['square\_feet']).reshape(data['square\_feet'].shape[0],1)#这里是将数据转化为一个1维矩阵
8. data\_test=data['price']
10. #创建线性回归模型，拟合面积与价格并通过面积预测价格
12. regr=LinearRegression() #创建线性回归模型，参数默认
13. regr.fit(data\_train,data\_test)#拟合数据，square\_feet将房屋面积作为x,price价格作为y；也可以理解用面积去预测价格
14. a=regr.predict(268.5)
15. **print**(a)#查看预测结果
16. **print**(regr.score(data\_train,data\_test))#查看拟合准确率情况,这里的检验是 R^2 ，趋近于1模型拟合越好
18. #预测的结果：268.5平的房子价格为8833.54， R^2 =0.967
19. #我们来画个图看一下数据最后是什么样的
21. plt.scatter(data['square\_feet'],data['price']) #画散点图看实际面积和价格的分布情况
22. plt.plot(data['square\_feet'],regr.predict(np.array(data['square\_feet']).reshape(data['square\_feet'].shape[0],1)),color='red') #画拟合面积与价格的线型图

### 3.5 二元线性回归及多元实例

1. **import** pandas as pd #导入pandas库
2. **import** numpy as np #导入numpy库
3. **from** sklearn.linear\_model **import** LinearRegression #导入机器学习库中的线性回归模块
5. data=pd.DataFrame({'square\_feet':[150,200,230,255,260,340,700,720,790,850,900,950,1000],
6. 'price':[5450,6850,8750,9650,10450,13450,16450,16930,17200,17943,18320,18412,18900],
7. 'bedrooms':[2,2,3,4,4,5,6,6,6,7,7,8,9]})
9. data\_train=np.array(data[['square\_feet','bedrooms']]).reshape(len(data),2)#不管什么方法将list或DataFrame或Series转化成矩阵就行
10. data\_test=np.array(data['price']).reshape(len(data),1)
11. regr=LinearRegression() #创建线性回归模型，参数默认
12. regr.fit(data\_train,data\_test)#拟合数据，square\_feet将房屋面积作为x,price价格作为y；也可以理解用面积去预测价格
13. a=regr.predict([[268.5,3]])#预测多增加一个卧室数量3
14. **print**(a)
16. #预测的结果：268.5平的房子价格为8594， R^2 =0.9456418694547066

## 4 结果分析

### 4.1 描述性统计

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **4** | **5** |
| **VARIABLES** | **N** | **mean** | **sd** | **min** | **max** |
| **LNCOMP** | 245 | 16.41 | 1.002 | 14.65 | 18.74 |
| **DC** | 245 | 0.00915 | 0.00933 | -0.0118 | 0.036 |
| **SIZE** | 245 | 24.63 | 1.607 | 21.5 | 28.06 |
| **ROE** | 245 | 0.112 | 0.0869 | -0.116 | 0.367 |
| **ALR** | 245 | 0.709 | 0.139 | 0.314 | 0.908 |
| **BOD** | 245 | 8.755 | 1.817 | 5 | 14 |
| **Number of code** | 49 | 49 | 49 | 49 | 49 |

### 4.2 相关性分析

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ROE** | **LNCOMP** | **DC** | **SIZE** | **ALR** | **BOD** |
| **ROE** | 1 |  |  |  |  |
| **LNCOMP** | 0.538\*\*\* | 1 |  |  |  |
| **DC** | 0.378\*\*\* | 0.358\*\*\* | 1 |  |  |
| **SIZE** | 0.546\*\*\* | 0.707\*\*\* | 0.386\*\*\* | 1 |  |
| **ALR** | 0.309\*\*\* | 0.474\*\*\* | 0.301\*\*\* | 0.642\*\*\* | 1 |
| **BOD** | 0.156\*\*\* | 0.427\*\*\* | 0.287\*\*\* | 0.306\*\*\* | 0.296\*\*\* |

### 4.3 多重共线性分析

|  |  |  |
| --- | --- | --- |
| **Variable** | **VIF** | **1/VIF** |
| **SIZE** | 2.71 | 0.368753 |
| **LNCOMP** | 2.23 | 0.449089 |
| **ALR** | 1.74 | 0.575085 |
| **BOD** | 1.27 | 0.784801 |
| **DC** | 1.23 | 0.813042 |

### 4.4 回归分析

|  |  |
| --- | --- |
| **VARIABLES** | **y** |
| **LNCOMP** | 0.028\*\*\* |
|  | -3.71 |
| **DC** | 1.790\*\*\* |
|  | (-3.93) |
| **SIZE** | 0.018\*\*\* |
|  | -4.4 |
| **ALR** | -0.05 |
|  | (-1.19) |
| **BOD** | 0.005\*\* |
|  | (-2.01) |
| **Constant** | 0.687\*\*\* |
|  | (-6.78) |
| **Observations** | 245 |
| **R-squared** | 0.384 |

### 4.5 研究结论

通过对金融和房地产行业2016至2020年的相关数据的分析表明，企业的绩效会受高管工资水平的影响，一家公司的高管薪酬结构安排很大程度上上影响到公司的发展。自我国学者对其展开研究以来，从一开始的研究发现毫无相关性可言，到逐步发现二者之间存在有一定的关系，再到如今大部分研究都认为二者之间有明显的相关性，从此可以看出，我国的薪酬制度改革取得了一定的成就，尤其是金融和地产这类高管享有高薪酬的企业中，也逐步建立起了高管薪酬和公司绩效挂钩的制度，企业的内部治理结构逐渐公开公正，走向成熟和完善。其中原因有以下几点，第一，高管薪酬的增加有利于提高高管对工作的热情，并增强公司内在的凝聚了，员工的归属感，为此公司高管会提高工作的热情；第二，高级管理人员在公司重大决策中起着重要的作用，公司支付的薪酬是对管理人员劳动成果的实际回报，他们的薪酬水平越高，为公司追求利润的动力就越大，高管若对自身的收入感到满意，也能对公司发展起到最大程度的促进作用。此外，由于公司的规模与企业绩效的相关性依旧比起高管薪酬与企业的相关性水平更加显著，这说明高管薪酬制度仍然需要进一步的优化，强化高管薪酬和公司绩效的联系。