



数据分析岗位 招聘情况及薪资影响因素分析

目录



研究背景与意义



数据介绍



描述性统计分析



回归分析



产品商业化及应用

1

研究背景及意义

1

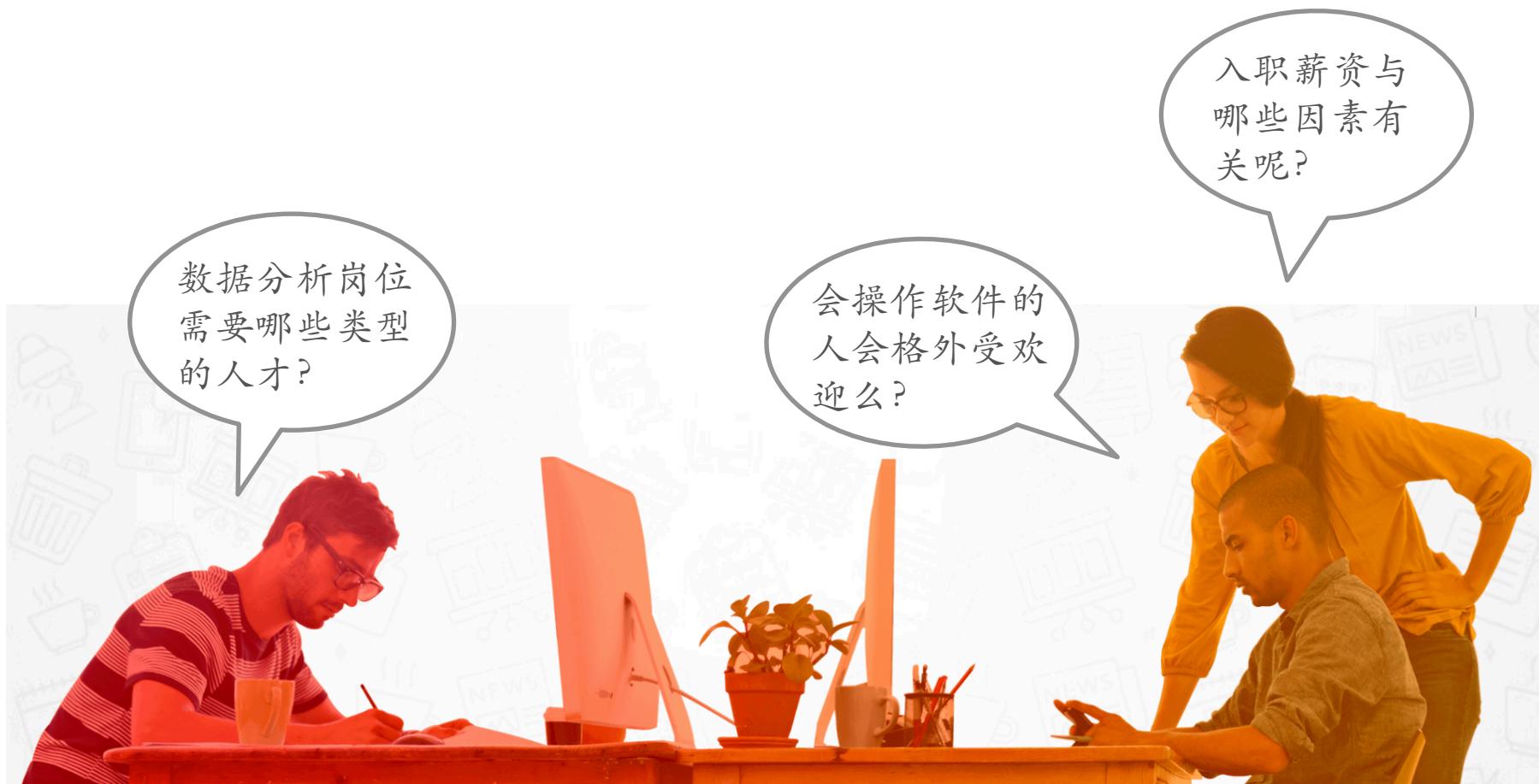
研究背景



- 随着“互联网+”和大数据时代的到来，越来越多的数据科学公司如雨后春笋般涌现。
- 传统行业也面临着“互联网+”时代下的创新转型，在数据分析相关领域有大量需求。
- 各行各业与数据分析相关的招聘岗位越来越多。

1

研究背景



1

研究意义

1 市场

了解数据分析人才市场现状，合理化市场资源配置

2 公司

为公司招聘提供借鉴，为数据分析人才定制薪资提供参考

3 应聘者

更科学地职业测评，以帮助其更准确地自我定位

4 学校

明确学生的培养方向，优化应用统计及数据分析方向学生的培养方案



2

数据介绍

2

数据来源与说明

数据来源

- 本案例的分析数据源自2016年9月各大招聘网站发布的数据分析岗位招聘的相关信息。

数据说明

- 本数据集包含了7493条岗位招聘信息。
- 该数据集涵盖了较有代表性的六个省市地区：数据分析岗位密集的北京市、上海市和深圳市，以及相对稀少的陕西省、山西省、河北省。



2

数据描述

变量名称	数据描述
招聘职位	字符型变量，例：数据分析员，战略分析经理。
公司名称	字符型变量，例：北京中航讯科技股份有限公司。
公司所在地	定性变量，共6个水平，分别为陕西、山西、河北、北京、上海、深圳。
公司类别	定性变量，共8个水平，分别为民营公司、上市公司、外资、国企、非营利机构、合资、事业单位、创业公司。
公司规模	定性变量，共7个水平，分别为少于50人、50-150人、150-500人、500-1000人、1000-5000人、5000-10000人，10000人以上。
公司行业类别	定性变量，共443个水平，例：计算机服务，互联网/电子商务。
经验要求	数值型变量。取值范围：0-10年。

2

数据描述

变量名称	数据描述
学历要求	定性变量，共7个水平，分别为无、高中、中专、大专、本科、硕士、博士。
招聘人数	数值型变量。取值范围：0-9人。
最低薪资	数值型变量。取值范围：1500-200000元/月。（存在异常值）
最高薪资	数值型变量。取值范围：1500-599999元/月。（存在异常值）
平均薪资	(最低薪资+最高薪资)/2。取值范围：1500-400000元/月。
岗位描述	字符型变量。包括职位描述，任职条件等。例：针对海量数据进行分析建模，挖掘数据潜在价值。掌握一定数据统计及分析工具，熟练运用SPSS或SAS或R等数据挖掘工具的优先等。
软件要求	定性变量，共12个水平，分别为R, SPSS, Excel, Python, MATLAB, Java, SQL, SAS, Stata, Eviews, Spark, Hadoop。

3

描述性统计分析

3

描述性统计分析概述

1

文本挖掘

- 公司行业类别
- 岗位描述

2

薪资及其相关因素分析

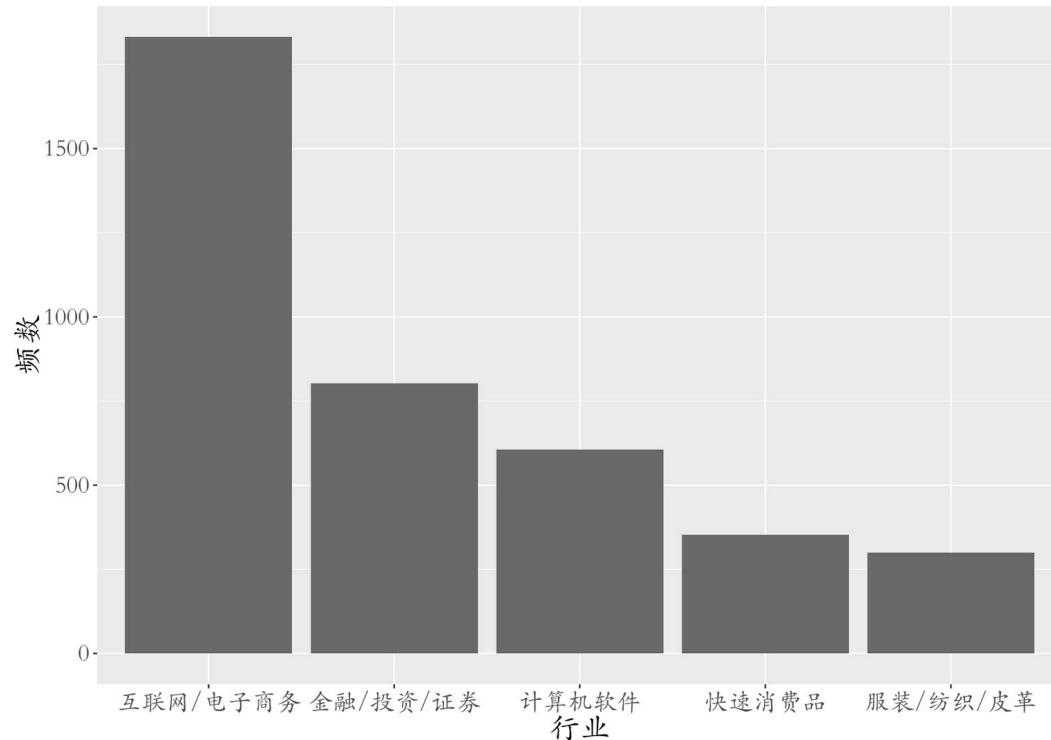
- 薪资 vs 地区
- 薪资 vs 公司类别
- 薪资 vs 经验要求
- 薪资 vs 学历要求
- 薪资 vs 软件要求

3

描述性分析小结

3-1

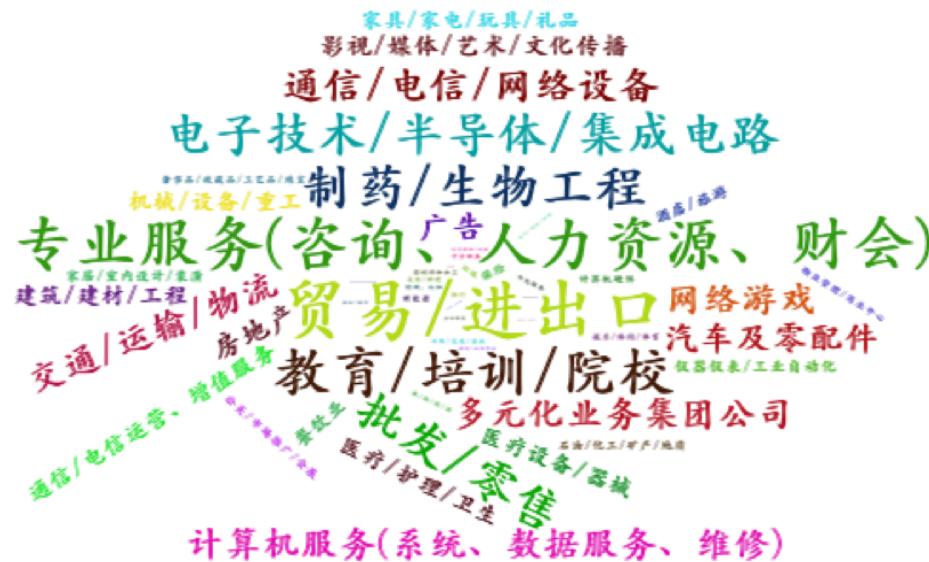
文本挖掘：公司行业类别_{1/2}



- 提取频数最高的五个行业绘制柱状图。可见数据分析岗位多分布于**互联网或电子商务，金融投资证券，以及计算机软件**等行业。
- 其中，互联网或电子商务行业的数据分析岗位多达1832，占整个样本集的24.79%。
- 在**快速消费品和服装**等传统行业中，尤其是随着电商产业的飞速发展，数据分析也逐渐凸显其重要性。

3-1

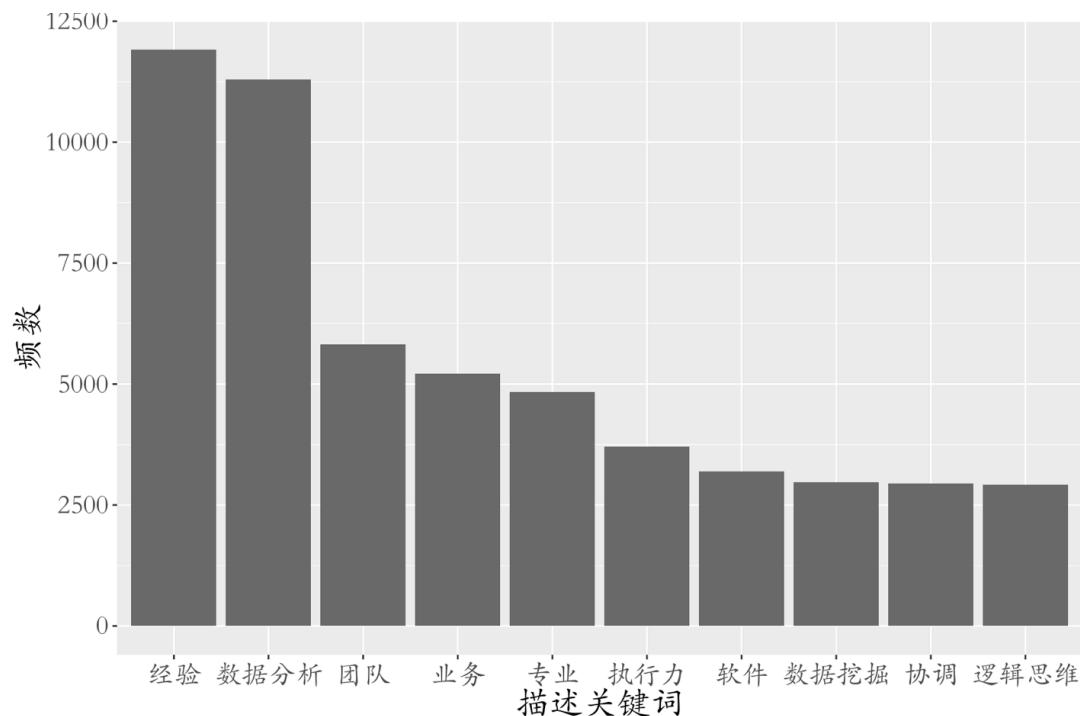
文本挖掘：公司行业类别_{2/2}



- 除以上五个占比最大的行业外，对其他行业绘制词云图。字体越大表示频数越大。
- 贸易及进出口行业，专业服务以及教育培训等行业也存在数据人才缺口。
- 越来越多的传统行业，如批发零售业等，也开始需要数据分析人才。

3-1

文本挖掘：岗位描述_{1/2}



- 对招聘岗位描述的关键词词频统计，列出出现频数最高的10个关键词。
- 用人单位通常十分注重应聘者的**数据分析经验和数据挖掘能力**，希望应聘者能够掌握某些软件。
- 团队合作、沟通、协调以及执行力和逻辑思维等能力也很重要。**

3 - 1

文本挖掘：岗位描述_{2/2}

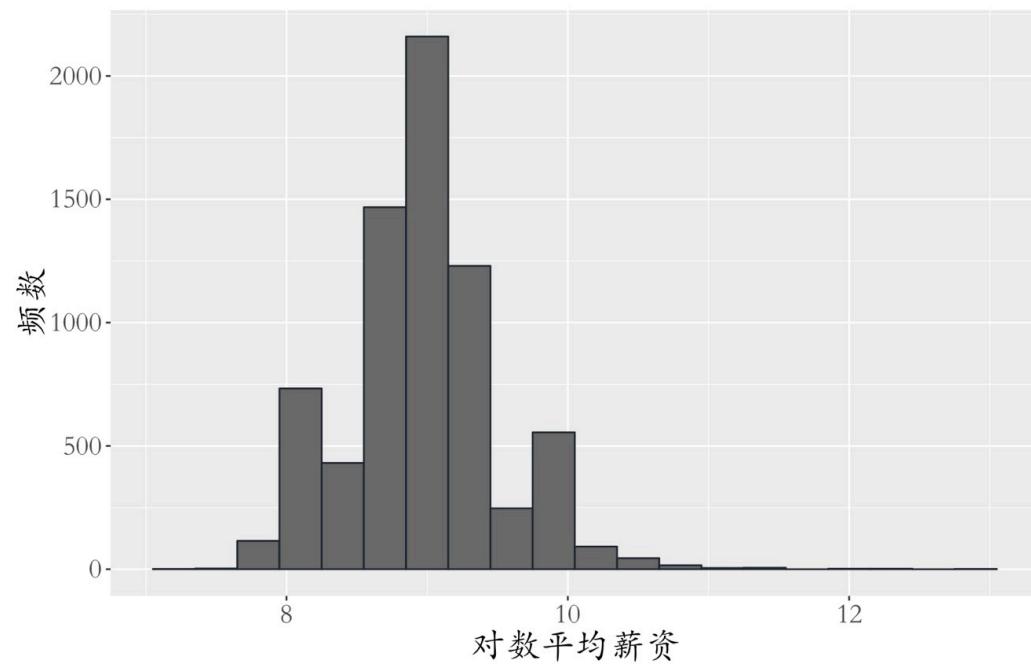


- 除此前10个占比最大的岗位描述关键词外，对其他频数大于500的关键词绘制词云图。字体越大表示出现频率越高。
 - 公司通常比较注重应聘者的**按时、独立、责任心**等品格，同时希望应聘者具有一定**统计或金融背景**和较强的**撰写及表达能力**。
 - 希望应聘者掌握特定**统计软件**，如Excel, SQL, SAS, R等。

3-2

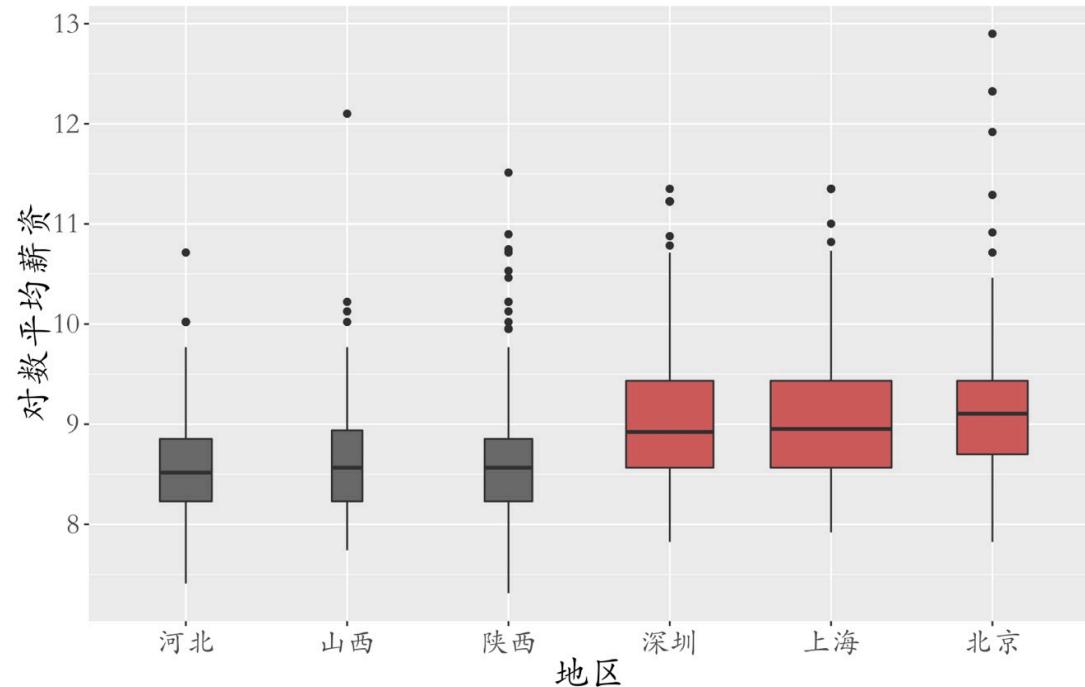
薪资分析

- 我们对岗位提供的平均薪资进行分析。平均薪资取该岗位提供的最高和最低薪资的平均值。
- 由于原始薪资中存在很多登记过程中人为因素造成的异常值，后续分析将基于**对数处理后的平均薪资**。
- 对数平均薪资大体呈正态分布。



3-2

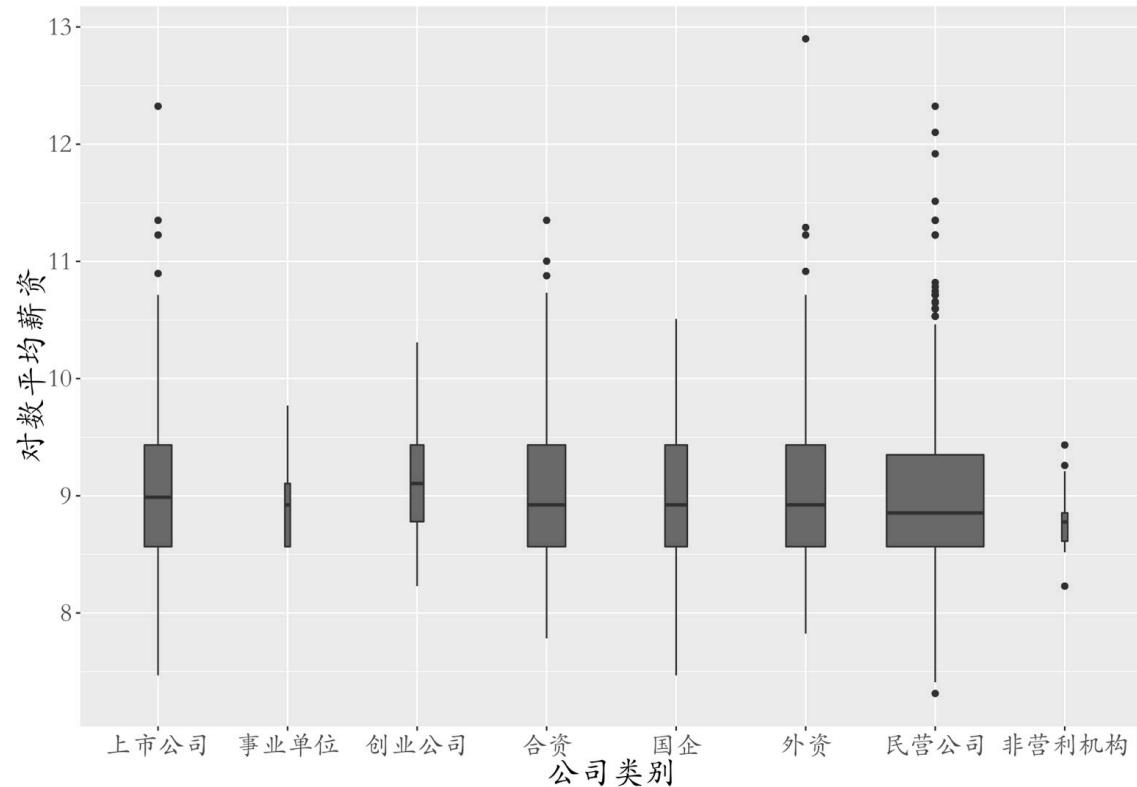
薪资分析：薪资与地区



- 箱体的水平宽度代表该地区岗位数量。可见[上海](#)的数据分析岗位是最多的。且北京、上海、深圳的工作机会明显多于其他三个地区。
- 北京、上海、深圳的薪资相对于河北、山西、陕西而言更高。
- 北京等三市和河北等三省可以划分为两类不同的区域，在各自类中薪资分布较为平均。

3-2

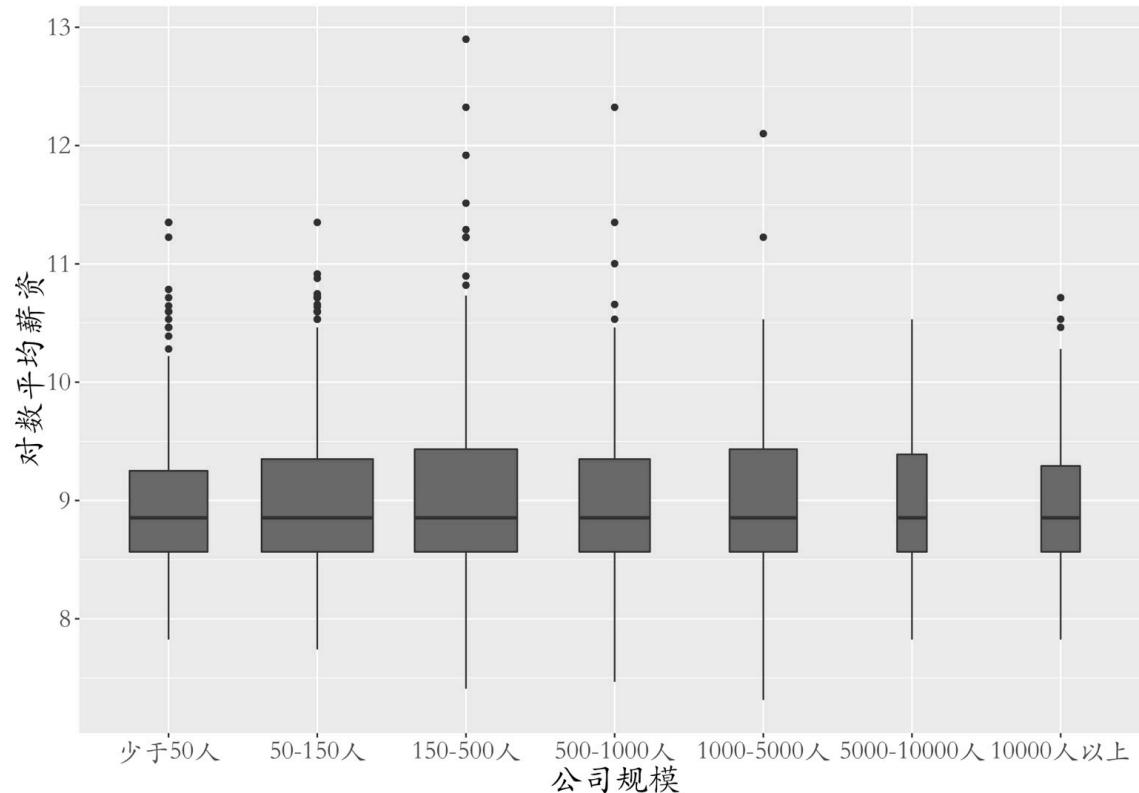
薪资分析：薪资与公司类别



- 箱体的水平宽度代表该类别公司岗位数量。**民营公司**提供的数据分析岗位最多，非营利化机构及事业单位较少。
- 各类别公司的薪资水平差距不大，创业公司的薪资相对较高。

3-2

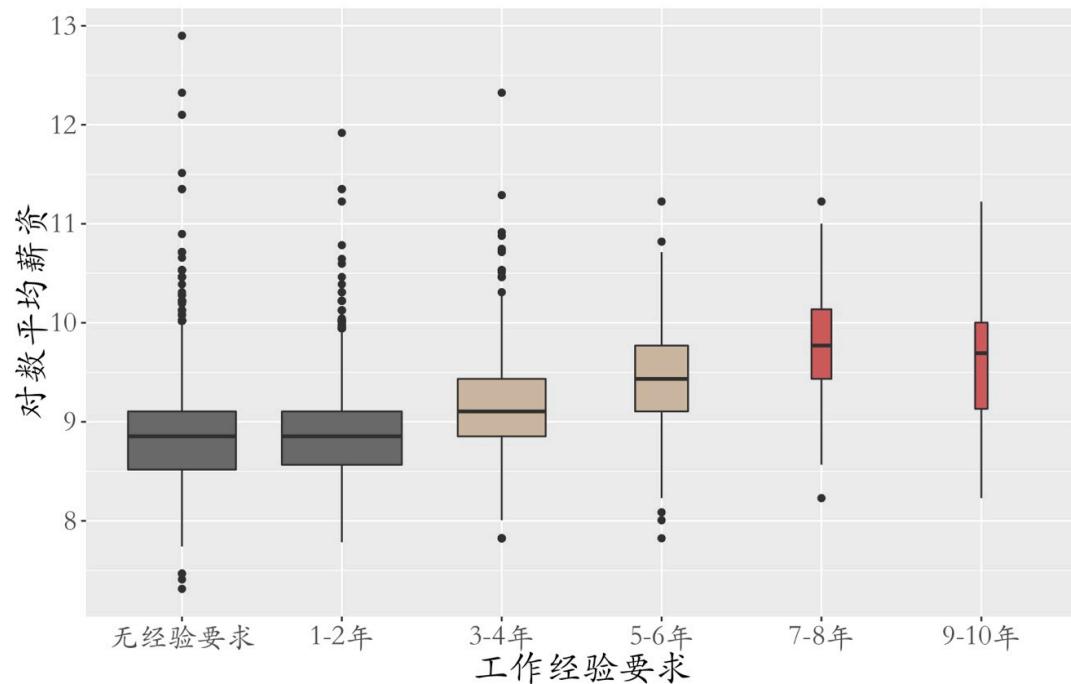
薪资分析：薪资与公司规模



- 箱体的水平宽度代表该规模公司岗位数量。其中**中小型公司**，例如50-500人的公司，数据分析的岗位较多，而5000人以上的大公司该类职位反而较少。
- 不同规模的公司薪资水平差距不大。

3-2

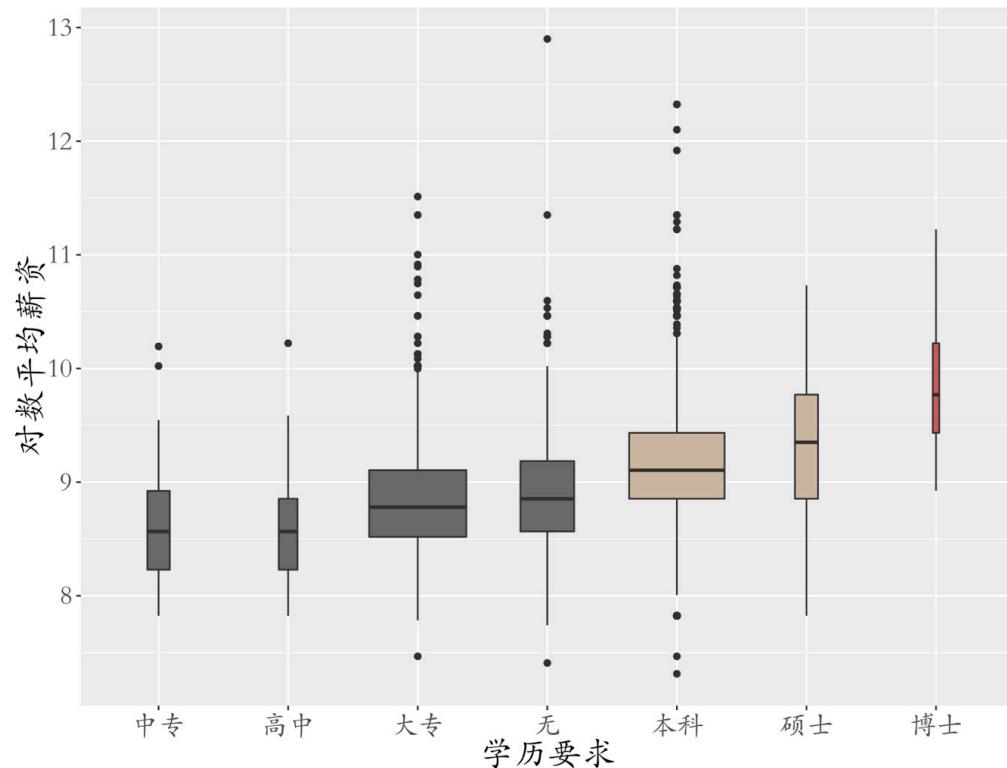
薪资分析：薪资与工作经验



- 箱体的水平宽度代表要求相应工作年限的公司岗位数量。由于数据科学属于较新型的产业，**对工作经验的要求不太高**，多数在3-4年或以下。
- 整体来看，公司招聘所要求的工作**年限越长，承诺薪资也越高**。
- 工作经验2年及以下的薪资水平差距不大，但多于5年之后薪资会有显著提高。工作年限要求达到7年以上之后薪资便不会再有明显上升。

3-2

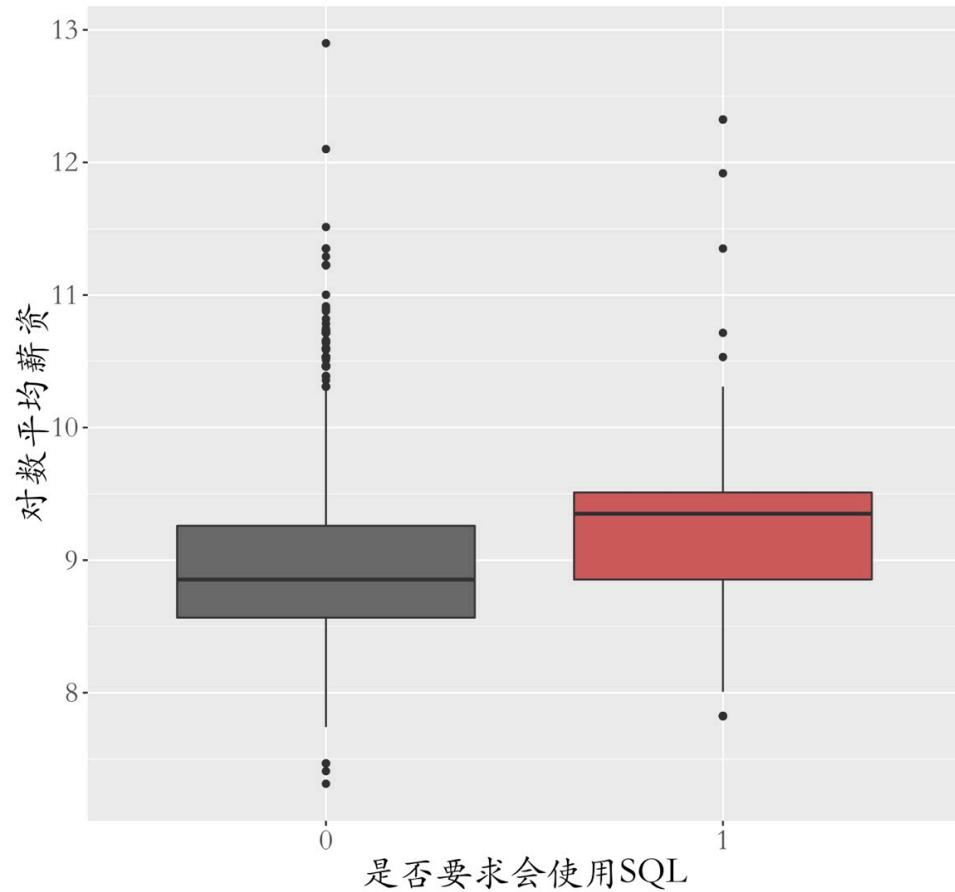
薪资分析：薪资与学历要求



- 箱体的水平宽度代表要求该学历的公司岗位数量。其中“无”表示未明确给出学历要求。
- 数据分析岗位**大多要求中等学历，如大专及本科**，要求硕博或高中及以下的岗位较少。
- 整体来看，**薪资会随学历水涨船高**。对学历无特别要求以及要求最低学历在大专以下的岗位薪资水平差距不大，本科和硕士逐步提升，博士学历薪资则明显高于其他学历。

3-2

薪资分析：薪资与软件要求



- 以是否要求SQL软件为例绘制箱线图。
- 要求会使用SQL的岗位薪资相对于不要求该软件的岗位薪资要高。
- 其他11种软件要求结果类似。

描述性分析小结

- ① 文本挖掘中的**词云分析**总结了数据分析岗位的行业分布、业务要求及招聘公司注重的能力。
- ② 通过对岗位提供的平均薪资的描述性分析，**地区、经验、学历、是否掌握数据分析软件对薪资有明显影响**。在北上深这样的经济发达地区，高学历、工作经验丰富、掌握某种数据分析软件的人才会拿到更高的薪资。后续的回归分析会将其关系进一步量化。
- ③ **数据分析岗位的薪资与公司规模以及公司类别的关系不大**。并不是大公司就会提供更多的就业岗位和更高的薪酬。但它们对薪资的影响，还需通过定量分析的方法进一步探究。

4

回归分析

4

回归分析

- 1 | 变量描述
- 2 | 回归结果及解读
- 3 | 回归分析小结

4-1

变量描述

变量类型	变量名	详细说明
因变量	对数薪资	对平均薪资取对数，以减少异常值的影响。
自变量	学历要求	公司要求的最低学历。共7个水平，分别为无、高中、中专、大专、本科、硕士、博士。以“无”为基准创建6个虚拟变量。
	经验要求	公司要求的最低工作年限。取值0-10年。
	软件要求	公司对软件能力的要求，包括R, SPSS, Excel, Python, MATLAB, Java, SQL, SAS, Stata, Eviews, Spark, Hadoop。分别用12个虚拟变量表示。
	公司地区	公司所在地。共6个水平，分别为河北，陕西，山西，北京，上海，深圳。以河北为基准创建5个虚拟变量。
	公司规模	共7个水平，分别为少于50人、50-150人、150-500人、500-1000人、1000-5000人、5000-10000人，10000人以上。以少于50人为基准创建6个虚拟变量。
	公司类别	共8个水平，分别为民营公司、上市公司、外资、国企、非营利机构、合资、事业单位、创业公司。以国企为基准创建7个虚拟变量。

4-2

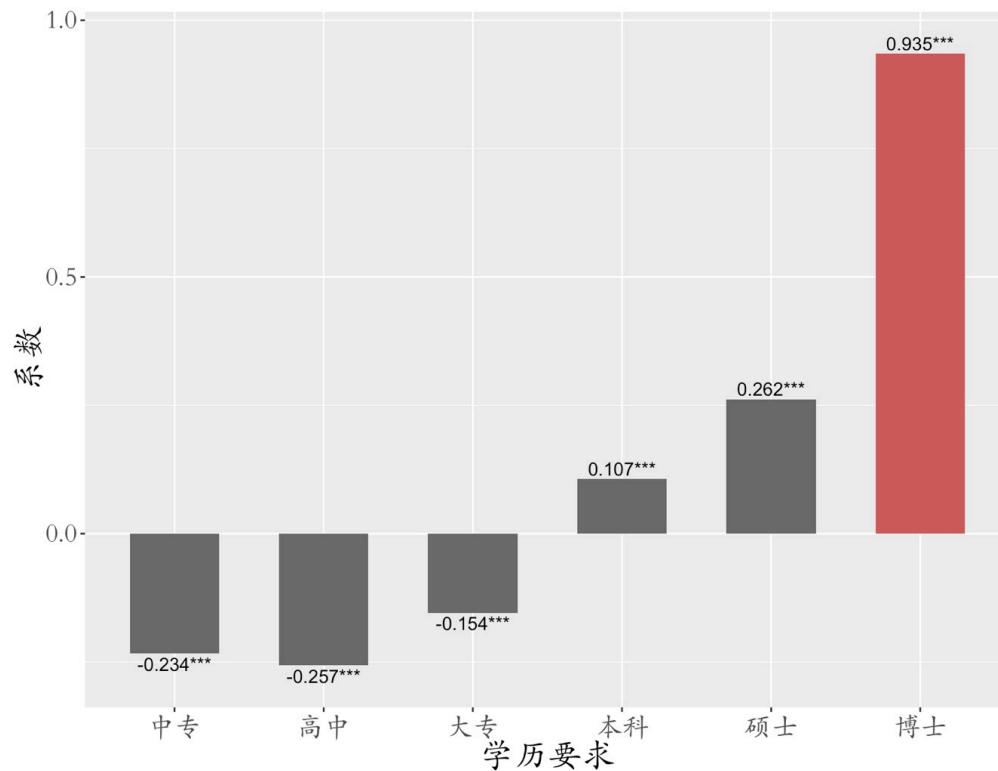
回归结果及解读

- 除“经验要求”外，其余自变量都用虚拟变量表示，其系数估计用柱状图表示。
- “经验要求”的系数估计为0.099，在0.01的显著水平下显著，表明工作经验要求每增加1年，平均薪资相应近似增长9.9%。
- 由于因变量薪资进行了对数变换，结果解读时所有系数的估计近似解释为“**薪资增长率**”。

友情提示：因变量增长率的严格计算方式为 $\exp\{\text{系数}\}-1$ ，当系数不太大时该数值泰勒展开近似等于系数本身。

4-2

回归结果及解读：学历要求

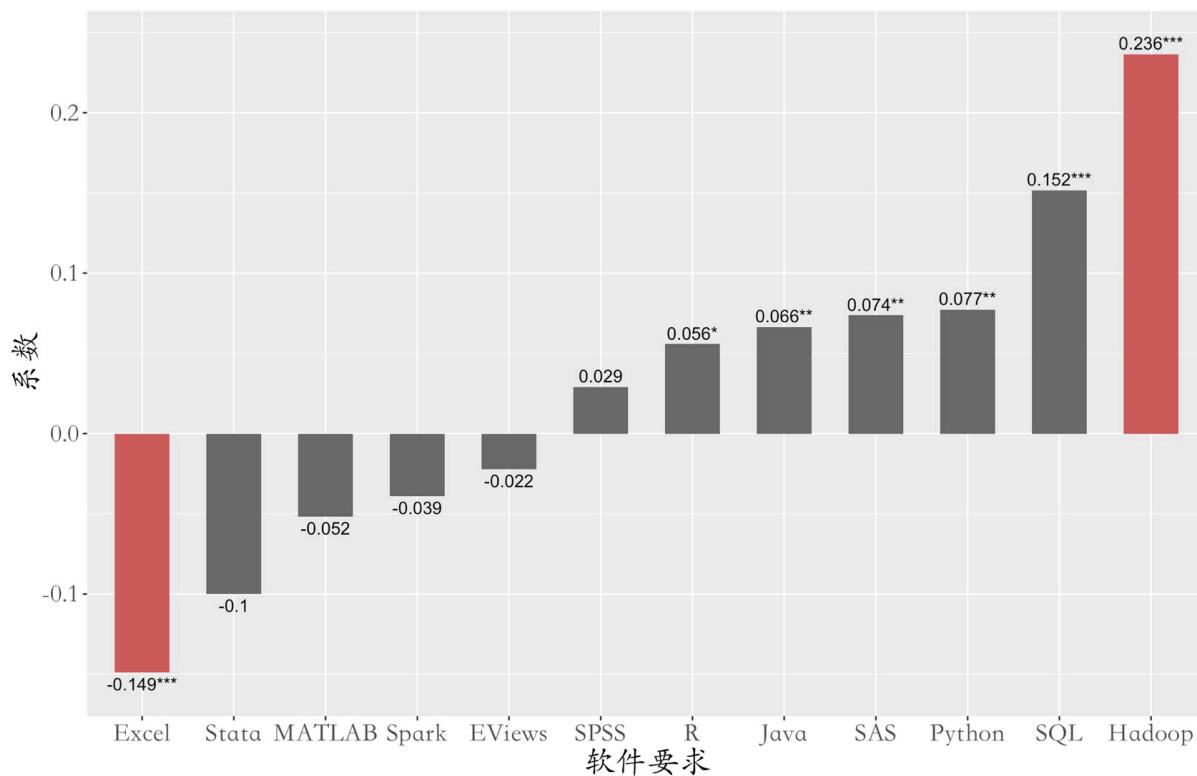


- 回归系数估计值在柱状图中标出，在5%的显著性水平下所有系数显著。
- 若将无学历明确要求的水平当做行业平均水平，则大专及以下学历的薪资低于平均水平，要求本科及硕士学历的薪资较高，要求博士学历的薪资明显走高。
- 例如，要求博士学历的岗位比要求本科学历的薪资要高 $0.935 - 0.107 = 82.8\%$ 。

备注：*, **, ***分别表示在0.10, 0.05, 0.01显著性水平下显著；以下同。

4-2

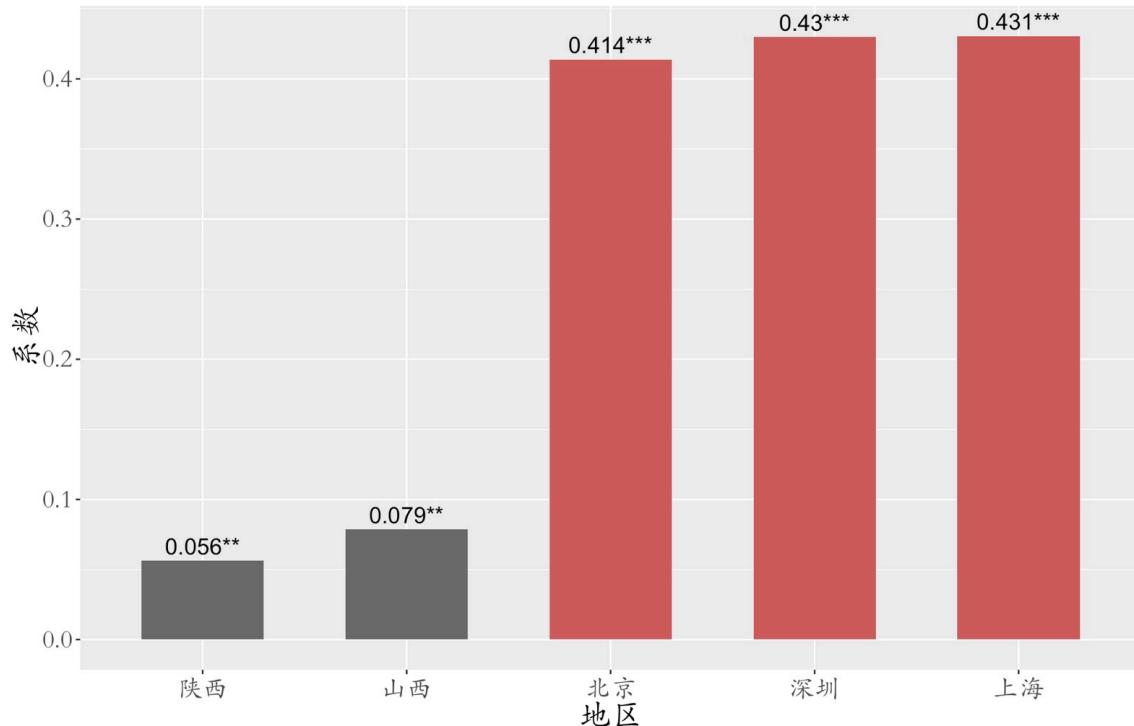
回归结果及解读：软件要求



- 每个系数表示有无该项软件要求的岗位平均薪资差距。
- 要求Excel技能的岗位薪资相对于无此要求显著为负，猜测因为指明要求Excel的岗位多为初级数据处理工作，因此薪资相对较低。
- 要求R, Java, SAS, Python, SQL以及Hadoop的数据分析岗位的需求通常为专业技术人才，故倾向于提供较高薪资。
- 例如，要求Hadoop的数据分析岗位比未要求的平均薪资高23.6%。

4-2

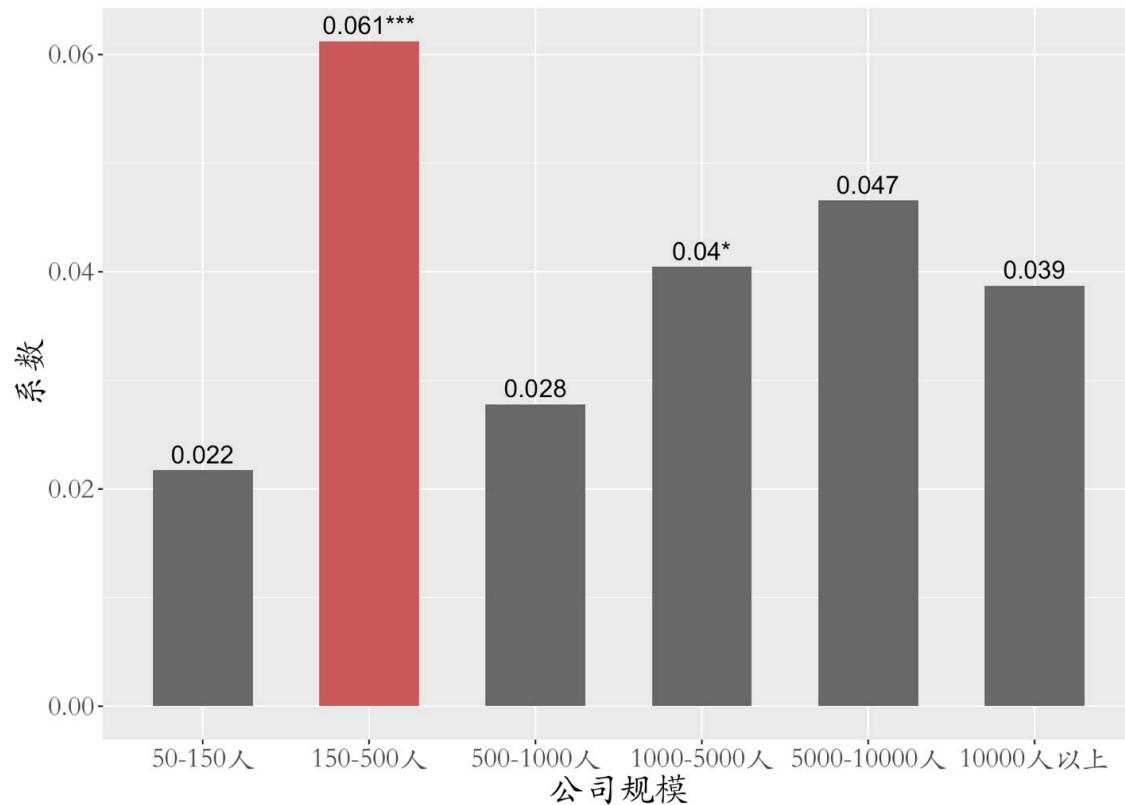
回归结果及解读：公司所在地区



- 以河北为基准，所有地区的系数都为正，说明平均薪资水平均高于河北。但陕西和山西与河北相差不大。
- 北京、上海和深圳的系数在1%的显著性水平下仍显著，说明其薪资水平远高于另外三个地区，但组内差别很小。
- 例如，北京地区的岗位薪资平均比河北高 41.4% ，比山西高 $0.414 - 0.079 = 33.5\%$ 。

4-2

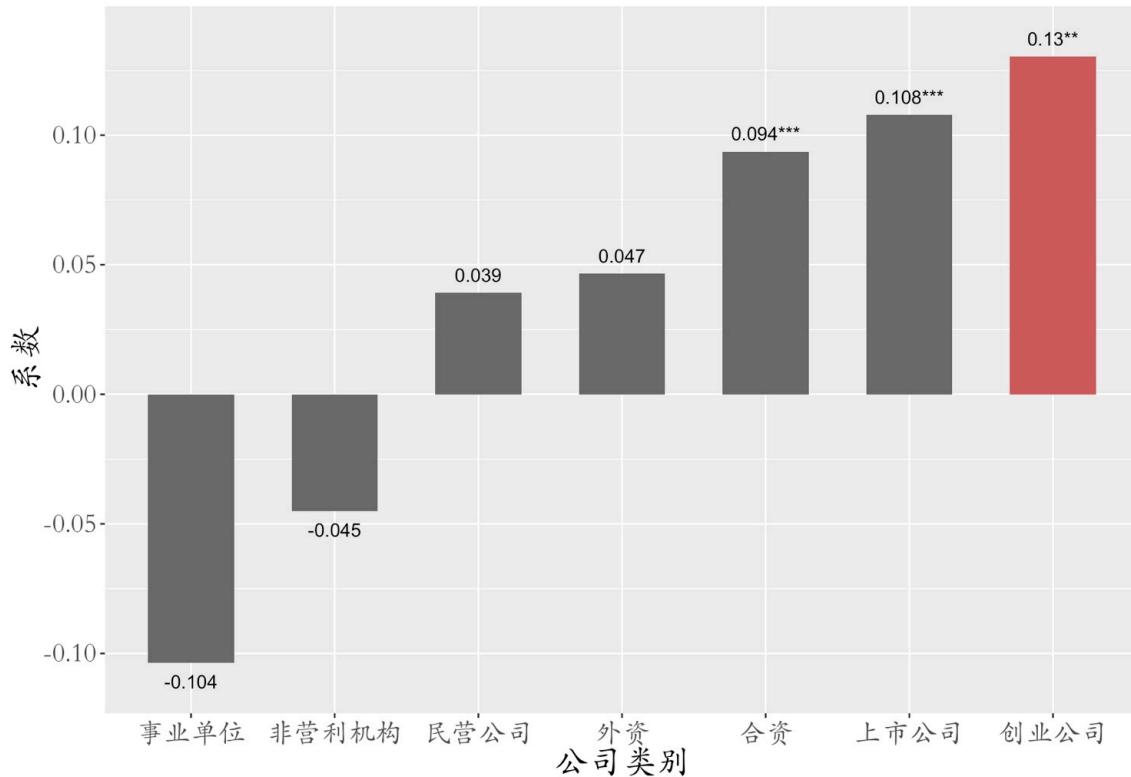
回归结果及解读：公司规模



- 以50人以下公司为基准，仅50-150人规模的公司系数在5%的显著性水平下显著。
- 并不是公司规模越大，提供薪资越高。**中等规模的公司可能更适合数据分析人才。**
- 150-500人规模的公司比50人以下的平均薪资高 $0.061=6.1\%$ 。

4-2

回归结果及解读：公司类别



- 以国企为基准，**创业公司、上市公司和合资企业的系数显著为正**，其他公司与国企没有明显差别。
- 例如，创业公司比国企的数据分析岗位平均薪资高13%。

4-3

回归分析小结

- ① 影响数据分析岗位薪资的主要因素为**公司所在地区，工作经验，学历和软件技能**。
- ② 北京，上海，深圳的薪资相对于河北，山西，陕西而言更高。工作经验越丰富，薪资越高，但达到7年之后经验对薪资的影响减弱。学历越高，薪资越高，博士学历尤其受高薪工作的青睐。通过回归分析，这些比较的指标得以量化。
- ③ R，Hadoop等数据分析软件的掌握也有利于寻找高薪的工作。有无某种软件的平均薪资差异也由回归分析的系数体现。

5

产品商业化及应用

5

数据产品的应用场景

- 1 • | 提供应聘者的专业职业测评及画像
- 2 • | 为相关岗位薪资定制提供参考
- 3 • | 优化学校或培训机构培养方案
- 4 • | 结论及建议

5-1

职业测评及应聘者画像

- 根据招聘岗位文本挖掘和回归模型系数估计确定招聘者的**五维打分机制**，对应聘者打分并画像，帮助其自我定位。
- 评分各维度分别为：**专业技能、职业素养、学历、经验、软件掌握情况**。其中前两项由文本挖掘结果赋予权重，后三项由回归系数估计赋予权重。
- 专业技能包括专业背景、数据分析和数据挖掘能力、撰写能力等。职业素养包括责任心、执行力、团队合作、表达能力等。
- 与传统的职业测评相比，本产品更具专业性，有垂直深入的优势。



5-1

职业测评之职场新人



5-1

职业测评之职场“高富帅”



5-2

薪资定制

- 将应聘者信息与招聘岗位情况带入回归模型得出薪资预测，为个人薪资预期和公司薪资定制提供参考。
- 不仅可以给出平均薪资的**单点预测值**，还可以通过**预测区间**给出薪资的范围。



5-2

薪资定制之职场新人



回归模型得出的平均薪资单点预测为9533元/月，
区间预测为8782元/月~10350元/月。

5-2

薪资定制之职场“高富帅”



回归模型得出的平均薪资的单点预测为26793元/月，
区间预测为23075元/月~31110元/月。

5-3

相关专业培养方案优化

- 根据文本挖掘和回归分析的结果，为学校相关专业（例如应用统计硕士、大数据统计硕士等）及职场培训项目的培养方案提供参考。
- 培养方向应更多关注文本挖掘得出的岗位需求关键词和回归分析发掘的显著因素。



5-3

相关专业培养方案优化

- 建议增加数据分析软件操作课程的比重，特别是高级统计软件的训练。
- 注重实际数据分析及数据挖掘能力的培养，以案例教学为途径。
- 除专业技能外，应同时训练表达能力、PPT设计能力、团队协调等软实力。
- 尽可能为学生提供实习机会，丰富其工作经历。



5-4

结论及建议

- ① 对于**应聘者**，可通过该职业测评和画像更清晰地自我定位。在职业发展方面，除掌握必备的专业知识和修炼个人品格外，同时应熟悉多种数据分析软件，积累工作经验。在读学生可以考虑继续深造来获得高薪工作。
- ② 对于**学校**，可以更加注重学生的软件能力培养与训练，并为学生提供更多实习机会。
- ③ 对于**招聘公司**，招聘时可根据不同应聘者的各方面条件由此回归模型计算定制其薪资水平。



谢谢观赏

案例出品：狗熊会
如需转载，请联系小编

扫描二维码，关注狗熊会
获取更多精品案例