

分析广告投入与企业总销售额、增长率的关系及预测

摘要：在当今激烈的市场竞争里，企业的广告投入对销售额有着重要影响。本文旨在通过数据分析探讨销售额与广告投入之间的关系，利用收集到的相关数据进行统计分析，包括相关性分析，回归分析等，以确定广告投入如何影响销售额，同时通过相关数据预估营业额，与实际数据进行对比，通过偏差分析得出原因，并为企业制定合理的广告投资策略提供参考建议。

关键词：广告投入；销售额；回归分析；投资建议

绪论：广告作为一种重要的营销手段，对企业的销售收入有着至关重要的影响。企业会投入资金用于广告宣传，期望通过广告提升知名度，以此吸引消费者购买，增加销售额。然而在实际操作中，企业往往难以准确把握广告投入与销售额之间的具体关系，不确定广告投入能在何种程度上带来销售额的增长。因此需要深入研究二者关系，以实现企业利益最大化。而在学术研究领域，对于广告投入与销售额关系的探讨由来已久，早期的研究多基于简单的线性模型，试图直接建议二者之间的线性关系。随着计算机技术的发展，越来越多的研究开始采用更加复杂的数据分析方法，如多元回归分析等，从不同角度挖掘二者关系，并考虑更多影响销售额的相关因素。

一、研究背景：广告投入与销售额关联之重要性

（一）广告对企业销售收入的关键作用

广告作为重要营销手段，对企业的销售收入有着至关重要的影响。企业会投入资金用于广告宣传，期望通过广告提升知名度，以此吸引消费者购买，增加销售额。然而在实际操作中，企业往往难以准确把握广告投入与销售额之间的具体关系，不确定广告投入能在何种程度上带来销售额的增长。因此需要深入研究二者关系，以实现企业利益最大化。

（二）学术研究发展脉络梳理

早期的研究多基于简单的线性模型，试图直接建议二者之间的线性关系。随着计算机技术的发展，越来越多的研究开始采用更加复杂的数据分析方法，如多元回归分析等，从不同角度挖掘二者关系，并考虑更多影响销售额的相关因素。

二、研究方法：数据收集与分析工具运用

（一）研究问题的确立

我们从官方的上市企业报表出发整合数据，通过对企业类型、销售规模进行研究，观察它们的联系。同时，应用在爬虫网站上对微博、微信、其他平台的投入与收入再进行一次分析。

（二）数据来源及处理方式

1. 官方报表数据收集与整理

收集一百家公司两年来的营业数据，生成 excel 表格，用于数据分析。

2. 爬虫数据采集与整合

利用爬虫技术从微博、微信等平台获取广告投入与收入数据。

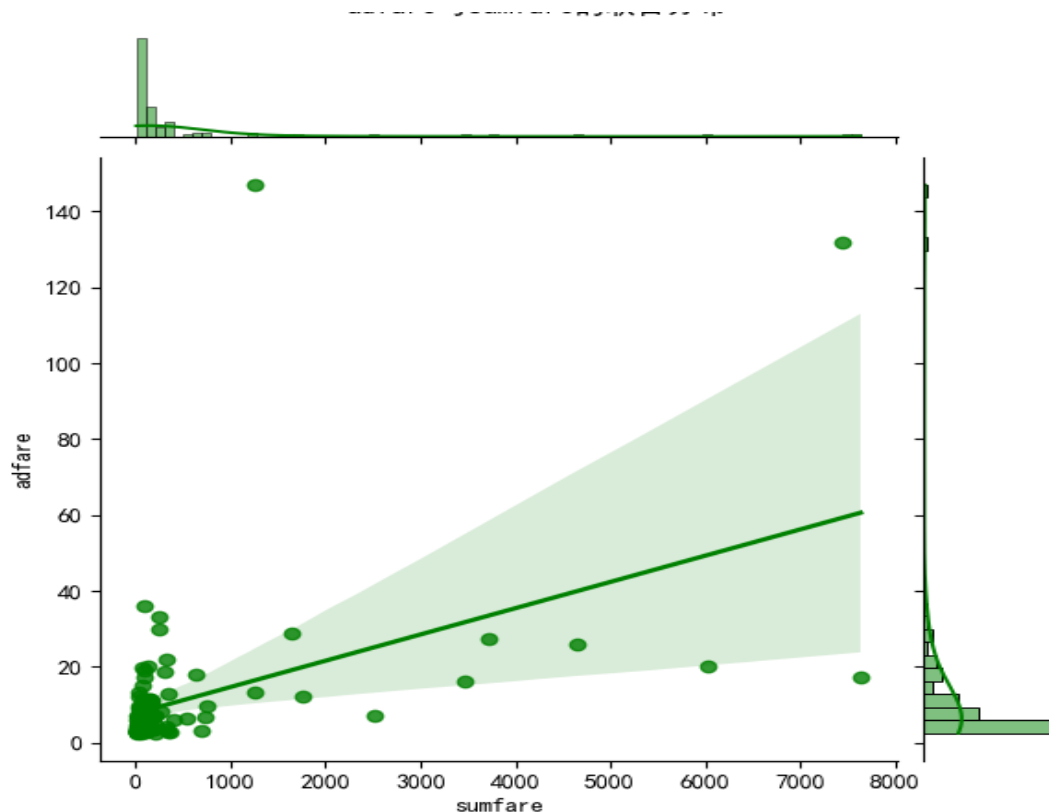
（三）分析工具与技术选型

运用 excel 与 python 编程进行数据分析，得到最终结果。python 用于数据可视化，生成多种图表；excel 用于回归分析等深入数据处理。

三、数据分析：多维度深度探究与结果解读

（一）数据可视化呈现

通过 python 将多个变量融合分析后，我们得到了十二张大概的曲线图、气泡图、热力图，得以顺利进行可视化分析，为后续继续用 excel 深入分析打下基础。



```

1111.py x
27 print(df['sumfare'].describe())
28
29 # 相关性分析
30 correlation = df[['adfare', 'sumfare', 'uprate']].corr()
31 print("\nadfare与sumfare的相关性:")
32 print(correlation)
33
34 # 散点图: adfare与sumfare的关系
35 plt.figure(figsize=(14, 8))
36 sns.scatterplot(x='sumfare', y='adfare', data=df, alpha=0.7)
37 plt.title('adfare与sumfare的关系')
38 plt.xlabel('sumfare(亿元)')
39 plt.ylabel('adfare(亿元)')
40 plt.show()
41
42 # 气泡图: adfare与sumfare的关系, 气泡大小表示adfare
43 plt.figure(figsize=(14, 8))
44 sns.scatterplot(x='sumfare', y='adfare', size='adfare', data=df, sizes=(100, 1000), alpha=0.7)
45 plt.title('企业规模与adfare的关系 (气泡图)')
46 plt.xlabel('sumfare(亿元)')
47 plt.ylabel('adfare(亿元)')
48 plt.show()
49
50 # adfare与uprate的关系
51 plt.figure(figsize=(14, 8))

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	7.7510	1.912	4.054	0.000	3.956	11.546
sumfare	0.0069	0.001	5.381	0.000	0.004	0.009

Omnibus: 137.387 Durbin-Watson: 0.579

（二）回归分析详述

1. 广告花费与营业额的一元线性回归剖析

基于 2023 年度不同企业的广告花费及营业额数据，建立模型 u ，运用 Excel 软件，进行一元线性回归，得到了一系列数据，并对这些数据进行详细

解释，包括回归统计中的相关系数、决定系数、调整后的决定系数，方差分析中的统计量和显著性，以及系数中的截距和自变量系数等，全面评估广告花费对营业额的影响程度及模型的显著性。

(1) 回归统计

Multiple R (相关系数): 0.4813478。这表示自变量 (广告花费) 和因变量 (营业额) 之间的线性关系强度。相关系数的范围在 -1 到 1 之间, 0.4813478 表明两者之间存在一定的正相关关系。R Square (决定系数): 0.2316957, 这是相关系数的平方, 表示模型解释因变量变异的比例。即营业额的 23.17% 的变异可以由广告花费来解释。Adjusted R Square (调整后的决定系数): 0.2236926, 考虑了模型中自变量的数量, 对 R Square 进行了调整。这个值略小于 R Square, 因为它考虑了模型的复杂性。

(2) 方差分析

F (F 统计量): 28.950 用于检验回归模型的显著性, 即判断广告投入是否对销售额有显著影响。Significance F (F 统计量的显著性): 5.23805E-07。这是 F 统计量对应的 p 值, 非常小, 说明模型在统计上是显著的。Coefficients (系数): 156.90651, 即当广告花费为 0 时, 预测的营业额为 156.90651。X Variable 1 (自变量, 即广告花费) Coefficients (系数): 33.499008 即广告花费每增加 1 单位, 营业额预计增加的量。t -Stat (t 统计量): 5.380566 用于检验自变量系数的显著性, 若其绝对值较大, 就表明对应的回归系数显著不为零, 即自变量对因变量有显著影响。P - value (p 值): 5.24E-07 自变量系数的 p 值, 非常小, 说明自变量在统计上是显著的。Lower 95% (95%置信区间下限): 21.1406428, 自变量系数的 95%置信区间下限。Upper 95% (95%置信区间上限): 45.8573729, 自变量系数的 95%置信区间上限。当置信区间不包含零时, 通常意味着自变量对因变量有显著影响。这里进一步支撑了广告花费对销售额有显著影响的结论。再进行一次线性回归之后, 我们抓取了 python 程序进行营业额预估。

(3) 总体结论

广告花费与营业额之间存在一定的正相关关系, 相关系数为 0.4813478。模型在统计上是显著的 (F 统计量的 p 值非常小), 说明广告花费对营业额有显著影响。营业额的 23.17% 的变异可以由广告花费来解释 (R Square 值)。广告花费每增加 1 单位, 营业额预计增加 33.499008 单位, 且这个系数在统计上是显著的 (自变量系数的 p 值非常小)。使用 Python 程序根据企业的广告投入估计总营业额与实际营业额的偏差与企业类别的关系。不同行业的广告效果可能有显著差异。例如, 快速消费品 (FMCG) 可能更依赖于高频次、视觉冲击强的

广告来刺激即时销售，而高科技产品或专业服务可能更注重内容营销和口碑传播。分析企业所属行业的广告敏感度，可以揭示哪些行业的预测误差较大。

2. 微信、微博及其他平台广告投入与销售额的线性回归解读

使用 statsmodels 库中的 OLS（普通最小二乘法）进行线性回归模型的构建和拟合，分析产品在微博、微信、其他平台的广告投入以及最终销量数据，明确各平台广告投入变量与销售额的相关性及显著性，包括常数项、各平台系数的值、t 值以及置信区间等关键指标的分析，为企业在不同平台的广告投放决策提供依据。我们利用一个 python 程序和一个 excel 文件对此进行分析。

（1）总体模型评估

从回归结果来看，模型整体具有一定的统计显著性。常数项（const）的系数为 4.62507，标准误差为 0.307507，t 值为 15.04055，对应的 p 值极低

(1.69×10^{-34}) ，这表明常数项在模型中具有显著意义。

（2）自变量分析

微信（wechat）系数为 0.011966，标准误差为 0.003002，t 值为 39.50989，p 值为 1.9×10^{-95} 。这表明微信变量与销售额之间存在显著的正相关关系。t 值较高且 p 值极小，说明该变量对销售额有显著的正向影响。即微信相关指标每增加一个单位，销售额预计会增加 0.011966 个单位。微博（weibo）系数为 0.013375，标准误差为 0.001061，t 值为 12.60397，p 值为 4.6×10^{-27} 。微博变量同样与销售额呈显著正相关。微博相关指标每增加一个单位，销售额预计增加 0.013375 个单位。其 t 值和 p 值也表明该变量对销售额的影响非常显著。其他（others）系数为 4.2×10^{-5} ，标准误差为 0.000724，t 值为 0.05802，p 值为 0.953792。与前两个变量不同，其他变量的 t 值较小，p 值较大，说明该变量对销售额的影响不显著。其系数接近于 0，意味着在当前模型下，其他变量对销售额的影响可以忽略不计。

（3）置信区间

微信（wechat）95%置信区间下限为 0.01137，上限为 0.012562。这意味着在 95%的置信水平下，微信变量对销售额影响系数的真实值有 95%的可能性落在这个区间内。微博（weibo）95%置信区间下限为 0.011283，上限为 0.015468。同理，微博变量对销售额影响系数的真实值大概率在此区间内。其他（others）95%置信区间下限为 -0.00138，上限为 0.001469。该区间包含 0，进一步验证了其他变量对销售额影响不显著的结论。

（4）系数

Coefficients (系数): 156.90651, 即当广告花费为 0 时, 预测的营业额为 156.90651。X Variable 1 (自变量, 即广告花费), Coefficients (系数): 33.499008, 即广告花费每增加 1 单位, 营业额预计增加的量。t Stat (t 统计量): 5.380566, 用于检验自变量系数的显著性, 若其绝对值较大, 就表明对应的回归系数显著不为零, 即自变量对因变量有显著影响。P - value (p 值): 5.24E-07, 自变量系数的 p 值, 非常小, 说明自变量在统计上是显著的。

(5) 方差

F (F 统计量): 28.9505, 用于检验回归模型的显著性, 即判断广告投入是否对销售额有显著影响。Significance F (F 统计量的显著性): 5.23805E-07, 这是 F 统计量对应的 p 值, 非常小, 说明模型在统计上是显著的。

(6) 预测程序

通过 python 线性回归分析, 我们发现微信和微博变量对销售额有显著的正向影响, 而其他变量对销售额的影响不显著。在实际业务操作中, 可以考虑加大在微信和微博平台上的营销投入, 以期获得更高的销售额。然而, 需要注意的是, 线性回归模型是基于现有数据建立的, 实际情况可能受到其他未考虑因素的影响, 因此在决策时应综合考虑多方面因素。

通过 excel 多元回归分析, 我们得到了销售额与在微信、微博和其他平台上的广告投入的关系, 并通过 visual studio 写出了相应的 C++ 程序, 其具有实用的功能: 当用户输入预期销售的收入数值时, 程序能够依据之前建立的多元回归模型, 准确地计算并输出在微信、微博和其他平台上分别需要投入的广告费用。

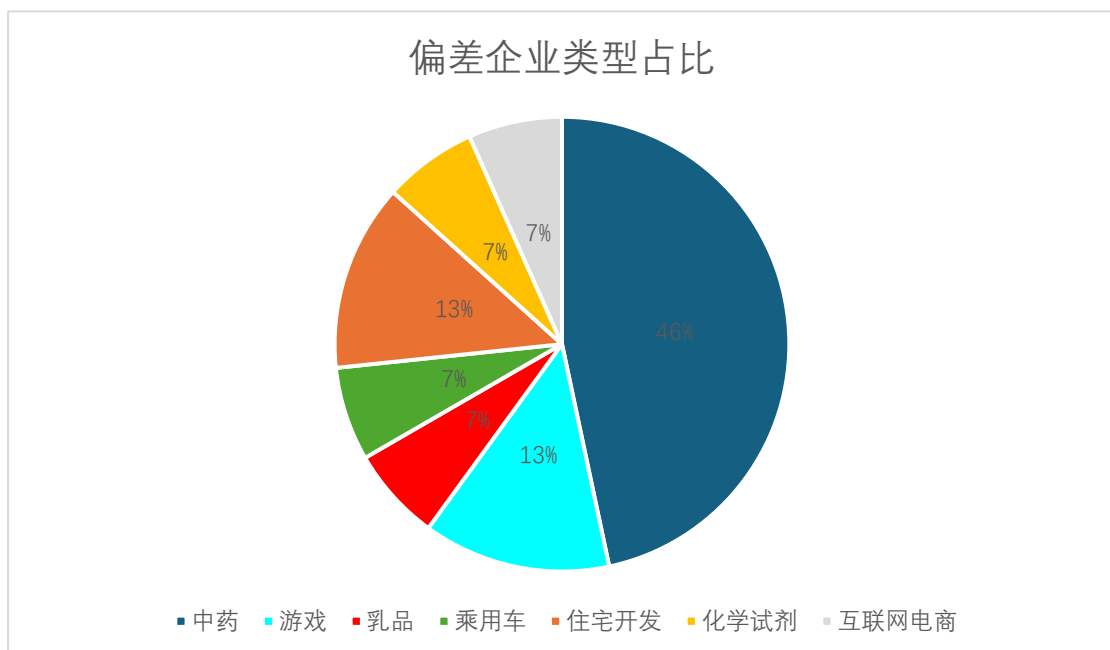
例如, 当用户在程序运行时输入预期销售的收入为 9 时, 程序会迅速进行计算, 并输出在微信平台上的投入应为 0.59 单位, 在微信平台上的投入应为 0.52 单位, 在其他平台上的投入应为 0.18 单位。

```
Microsoft Visual Studio 调试 × + - □ ×
请输入预期销售收入：9
在微博平台上的投入：0.59 单位
在微信平台上的投入：0.52 单位
在其他平台上的投入：0.18 单位

C:\Users\...source\repos\Project3\x64\Debug\Project3.exe
(进程 31176)已退出，代码为 0 (0x0)。
按任意键关闭此窗口。 . . |
```

这样的程序为企业在进行广告投放预算时提供了有力的决策支持，能够帮助企业更加科学、合理地分配广告资源，以达到预期的销售目标。

（三）企业类别与偏差分析探究



1. 总体偏差状况概览

对 101 家企业研究发现，较大偏差（30%-50%）的有 9 个，严重偏差（50% 以上）的有 5 个，分布于多个行业，提示行业差异可能导致广告投放与营业额偏差显著。中药：3 家企业实际值高于预估值，4 家低于预估值，其中 2

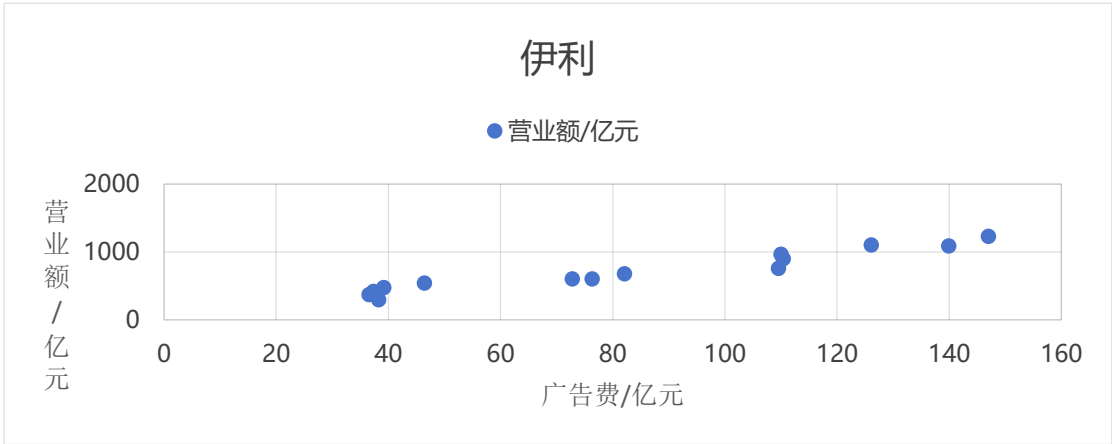
家远低于预估值, 1 家远高于预估值。住宅开发: 2 家均高于预估值, 其中一家远高于预估值。互联网电商: 1 家较高于预估值。乘用车: 远低于预估值。游戏: 均高于预估值, 其中 1 家远高于预估值。乳品: 较低于预估值接下来分析同类型企业的实际值和预估值之间的大小关系, 进而判断偏差类别是否与企业类型有关。

2. 不同行业偏差根源剖析

分别深入分析中药、游戏、乳品、乘用车、住宅开发、化学试剂等行业偏差原因, 从用户特性、市场竞争、品牌影响力、市场趋势等多维度揭示不同行业广告效果差异根源, 如游戏行业的用户黏性、口碑传播、多渠道投放等因素导致低投入高回报; 乳品行业的激烈竞争、品牌认知度低等因素致使高投入低回报等, 为企业针对性优化广告策略提供参考。游戏行业, 呈低投入高回报态势, 归因于用户强黏性、口碑速传播、多渠道协同、创意互动优、需求持续涨及年轻受众社交媒体高活跃等因素共振, 削弱广告投入依赖同时强力提升销售额, 塑造独特行业优势。乳品行业, 多为高投入低回报困境, 根源在于激烈市场竞争、新兴品牌认知局限、部分企业品牌力弱致营销成本高但市场拓展艰, 投入产出失衡, 掣肘行业盈利提升。乘用车行业 (比亚迪领衔), 展现低投入高回报佳绩, 得益于强劲品牌影响力、品牌溢价评估不足、新能源市场蓬勃与政策利好环境, 驱动销售额超预期增长, 彰显企业特质与市场趋势融合对广告效果的深度塑造力。

四、案例聚焦：伊利企业实证分析

(一) 散点图与相关性深度洞察，初步建立模型



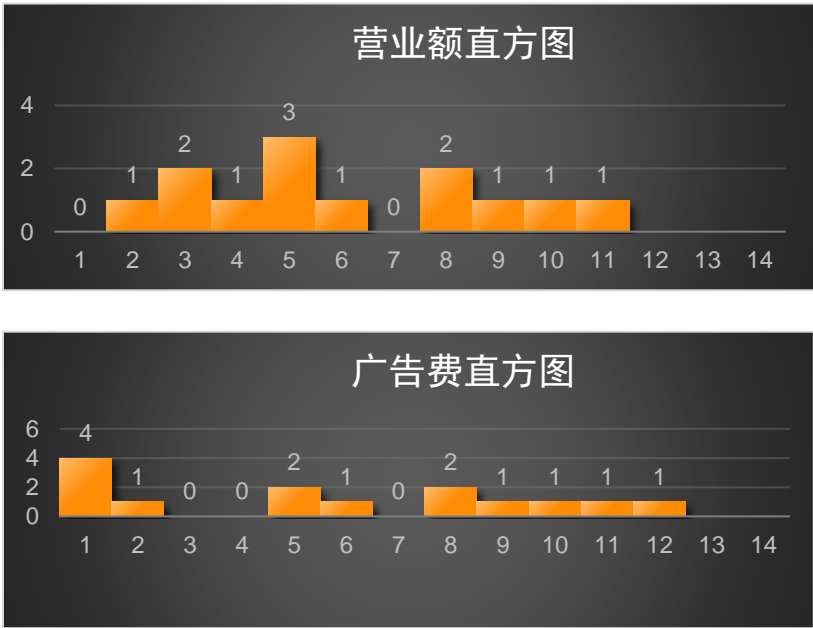
经由简单的散点图易见随着广告费用的增加, 营业额总体上呈现上升趋势, 这表明广告投入对伊利公司的营业额有正向的影响。散点图中的数据点大致呈现出一种线性分布的趋势, 这说明广告费用和营业额之间可能存在较强的正相关性。接下来我们需要通过具体的数值计算来确定其关联的紧密程度。

借由 C++ 程序计算，得到伊利企业的广告费和营业额的皮尔逊相关系数约为 0.97，表示广告费和营业额之间存在非常强的正线性相关关系。决定系数高达 0.941863 说明伊利企业的营业额变化中，大约有 94.19% 可以通过广告费用的变化来解释，企业可以比较有依据地利用广告投入的调整来预估营业额的大致走向，但同时也要关注剩下约 5.81% 的其他影响因素，比如市场环境变化、竞争对手策略、消费者偏好变动等，来更全面地把握营业额的变化趋势。但是 MSE 值为 4784.19，这表明目前基于该线性回归模型去预估营业额，与真实情况存在很大的差距，意味着还需要进一步优化。

总的来说，这个线性回归模型能在一定程度上帮助伊利企业认识广告投入与营业额之间的紧密联系以及量化关系，但考虑到均方误差情况，仍有必要进一步完善分析、优化模型，使其在企业决策等方面发挥更精准有效的作用。

(二) 模型优化

(1) 数据预处理：提升数据质量的关键步骤



从广告费直方图与营业额直方图可见数据呈偏态分布，多数数据集中于低值区间，少数高值拉高整体分布，违背线性回归正态假设，影响模型准确性。采用平方根变换等手段，降低高值影响，使数据分布更对称，契合模型要求，提升分析可靠性与有效性。运用箱线图法检测广告费和营业额数据中的异常值，将超出上下四分位数 1.5 倍四分位距范围的数据视为异常值。对于可修正的录入错误类异常值进行修正；对于真实但极端的数据（如特殊营销活动引发），采取 Winsorize 缩尾处理，以上下四分位数 ± 1.5 倍四分位距的值替换异常值，在保留数据真实性同时减少对模型的过度干扰，确保数据质量可靠、分析结论坚实。

（2）模型再建

利用处理后的数据再次建立回归模型得均方误差为 2.01702，这个值相对较小。这表明该模型在预测数据上有一定的准确性。决定系数为 0.930936，比较接近 1。这意味着该线性回归模型能够解释约 93.09%的营业额变异，说明模型对处理后的数据拟合程度较好。但是由于对数据进行了基于箱线图的异常值处理，虽然有助于减少极端值对模型的影响，但也可能改变了数据的原始分布。处理后的数据所呈现的线性关系，可能在原始数据中并不存在，或者原始数据中的关系可能更为复杂。

总体而言，该模型在处理后的数据范围内，可以用来大致预测在一定广告费投入下的营业额情况，但仍需进一步考察。查阅资料知，在数据分析和建模中，比较不同类型的模型是很常见的做法，是科学验证过程的一部分，于是选择比较线性和非线性模型的预测效果，如果线性模型在这个比较过程中表现出更好的预测效果，那么在一定程度上说明，在当前的数据和评估指标下，线性关系是一个相对更优的描述方式。基于简图只考虑与二次多项式模型进行比较。利用最小二乘法得出其决定系数为 0.930868、均方误差为 2.68284

从均方误差和决定系数来看，线性模型在这个数据集上的表现优于二次多项式模型。这意味着在当前的数据和模型比较下，线性关系更能准确地描述广告费和营业额之间的关系。模型建立成功。

如图所示：



（三）投资预测及合理性解释

通过创建代码（代入反算）可以实现输入实际商业场景中的广告投资额得到相应的营业额。

在实证研究中，由于现实情况的复杂性，将销售额假设为只受广告费用影响是一种常见的简化策略。这样可以使模型构建和分析更具可操作性。如果试图考

虑所有影响因素，模型会变得极其复杂，可能会出现数据获取困难、模型过拟合等问题。

由于本文研究的重点是广告费用与销售额之间的直接关系，并且希望突出广告费用的作用，这种假设是合理的。例如，在评估一个新的广告策略或者广告渠道的有效性时，暂时将其他因素固定或者忽略，集中观察广告费用变化对销售额的影响，可以为决策提供较为直观的依据。

且在数据收集阶段发现，广告费用的变化与销售额的变化呈现出比较明显的线性关联，而且其他因素相对稳定或者难以量化，所以这种假设在一定范围内是可以被接受的。

实际来说，这种假设显然与实际情况存在偏差。在现实商业环境中，如市场自然变动（如经济周期、季节变化等）、产品或服务的性能和质量、销售网络等因素对销售额的影响是不可忽视的。基于此假设构建的模型在应用范围上会受到限制。由于没有考虑其他重要因素，模型可能只适用于特定的市场条件、产品类型或者广告策略。一旦外部条件发生变化，模型的预测准确性可能会大大降低。同样，如果完全依赖这种假设下的模型进行决策，可能会带来风险。故本文只是一个简单建议，具有一定合理性。

六、研究总结与展望

（一）研究总结

通过计算和预判，我们对数据进行了处理，并进行了回归预测分析，然后将其应用在了伊利这一大牌企业上，得到了强线性相关的结论。与此同时，我们也对微博微信与其他平台的投入和销售得到了预测分析，可以输入一个投入数据，进行预估收入，具有实际意义。

（二）研究展望

我们需要更多的样本，更多的细化分析，以及后续的深入了解，才能得到更显著的结论。此次大作业，我们对于企业类型的数据分析还比较浅显，需要继续深入挖掘各大企业的相关与不相关。