678 Fianl Project

Abstract

The purpose of this study is to explore the influencing factors of sales profit of large stores, and screen out a series of factors through difference test and correlation analysis, such as sales market, type of goods sold, sales volume, sales quantity, discount and transportation cost. Therefore, a multiple linear regression analysis model is established with the sales market as a multi-level variable, the type of goods sold, sales volume, sales quantity, discount and transportation cost as independent variables, and the sales profit as dependent variables. According to the test results of the model, there are enough inter group differences in different sales markets, and different sales markets, types of goods sold, sales volumes, sales quantities, discounts, transportation costs and sales profits are shown. Sales markets, types of goods sold, sales volumes, sales quantities, discounts and transportation costs are important predictors of sales profits.

Introduction

Operating profit is an important cornerstone for the sustainable development of enterprises. How to obtain more profits at the minimum cost is the basis for shareholders to continue to operate the enterprise. Therefore, profit maximization is an important goal pursued by enterprises. The main factors that affect profit include product sales quantity, sales price, fixed cost, variable cost, marginal contribution, profit and loss critical point, etc. The change of any one of these factors will lead to the change of enterprise profit, and even make the enterprise change from profit to loss, and also make the enterprise change from loss to profit. In reality, the most closely related to profits are operating income and operating costs. Revenue is the most important basis for generating profits. If there is not enough sales revenue as a prerequisite, it is impossible to obtain high profits. In addition, the cost of sales is also an extremely important factor. Cost is the loss of sales revenue. If the cost is too high, no sales revenue will be lost. Therefore, in order to obtain a higher level of profits, lower costs and higher revenues are undoubtedly extremely important. In addition, the type of goods sold is also an important factor. If the added value of goods is very high, the marginal profit level that can be brought is very high, so the profit potential is greater. The sales volume is very close to the sales revenue. The sales volume and unit price of goods sold undoubtedly constitute the sales revenue. When the unit price of goods is fixed, the higher the sales quantity of goods, the higher the sales revenue. Discount is also related to profit, but this effect is undoubtedly double. The proposal of discount is undoubtedly the cost of sales, reducing the profit of the enterprise. But more importantly, the discount will promote the sales of goods, increase the sales volume, and further improve the sales revenue of enterprises, thus providing an important prerequisite for obtaining higher profit levels. In addition, the sales market is also an important factor affecting sales profit. If the consumption scale of the market is small and the consumption potential is low, the probability and scale of enterprises to obtain profits in this market will be reduced. If the consumption scale of the market is large and the consumption potential is also large, the market will bring huge profits to sellers. More importantly, in different sales markets, there may be different profit levels, and the degree of influence of various factors on sales profits is also different.

Method

The data set includes sales data of various products sold by stores, as well as follow-up information related to geography, product categories and subcategories, sales and profits, and segmentation between consumers. This sample dataset provides a common use case from which I can gather useful insights from sales data to

improve marketing and sales strategies. I can use this sample dataset to understand the various operations and elements and come up with better strategies to improve and grow the business.

The method used in this study is multi-level linear regression analysis. First of all, the sales market in this model is a second level variable, and in the first level variable, it distinguishes the independent variable and dependent variable in the study to explore the influencing factors of enterprise sales profit, and explores the relationship between the independent variable and dependent variable, as well as the performance of the relationship between the independent variable and dependent variable in different sales markets. In addition, this study also used ANOVA and other difference tests to select the independent variables in the multiple linear regression analysis model. In addition, there is also a correlation analysis method to explore the linear correlation between independent variables and dependent variables.

Descriptive statistical analysis

According to the results of descriptive statistical analysis of profits (Result 1, see appendix), the minimum value of profits is -6599.98 and the maximum value is 8399.98, which indicates that there is a large difference between the profits of all orders, with the median of 9.24 and the average of 28.64. It indicates that the overall profits of all orders are profitable.

```
hist(SuperStoreOrders$profit,main = "Histogram of profit")
```

According to the profit histogram, the profit distribution is mainly concentrated in the middle value, and the profit of most orders exceeds 0, that is, most orders are profitable.

According to the results of variance analysis (Result 2), the average order profit in Africa market is 19.375, the average order profit in APAC market is 39.773, the average order profit in Canada market is 46.399, the

average order profit in EMEA market is 8.730, the average order profit in EU market is 37.283, the average order profit in LATAM market is 21.531, and the average order profit in US market is 28.657. The F statistic of the difference between the average order profit in different markets is 28.24, And this statistic has reached the significance level of 1%, so it can be considered that there are significant differences in the average order profit of at least two of the seven markets.

According to the results of variance analysis (Result 3), the average order profit of Furniture is 29.038, the average order profit of Office Supplies is 16.580, and the average order profit of Techology is 65.455. The F statistic of the difference in the average profit of orders of different categories is 304.2, and this statistic reaches the significance level of 1%. Therefore, it can be considered that there are significant differences in the average order profit of at least two of the three categories.

According to the correlation analysis results (Result 4), the correlation coefficient between order profit and sales is 0.486, and the correlation coefficient reaches a significant level of 5%. Therefore, it can be considered that there is a significant positive correlation between order profit and sales.

According to the correlation analysis results (Result 5), the correlation coefficient between order profit and quantity is 0.105, and this correlation coefficient reaches the significance level of 5%. Therefore, it can be considered that there is a significant positive correlation between order profit and quantity.

According to the correlation analysis results (Result 6), the correlation coefficient between order profit and quantity is 0.105, and this correlation coefficient reaches the significance level of 5%. Therefore, it can be considered that there is a significant positive correlation between order profit and quantity.

According to the relevant analysis results (Result 7), it can be seen that the order profit and shipping_ The correlation coefficient of cost is 0.357, and the correlation coefficient reaches the significance level of 5%, so it can be considered that the order profit and shipping_ There is a significant positive correlation between costs.

Empty model

According to the results of the empty model (Result 8), ICC's calculation result is 0.0047, that is, the 4.7% variance of the sales profit margin can be explained by the difference between groups in different markets. Therefore, it can be considered that the difference between different sales markets can be used to explain the variation of the model sales profit, although this explanation proportion is relatively low. Therefore, the multilevel linear model of this study is constructed.

Compared with the empty model 0, the interpretation ability of the full model has been improved. Therefore, the full model can be used for interpretation in this study.

The test results (Result 9) show that there are indeed differences between different sales markets, that is, the above variables and sales profit have different performances in different sales markets.

The regression coefficient of the independent variable category office supplies is -0.009681, which reaches a significant level of 5%, indicating that the average order profit level of office supplies is significantly different from that of furniture types. Specifically, compared with the average order profit of furniture, the average order profit of the commodity category behavior of office supplies is lower.

The regression coefficient of independent variable category technology is -0.003966, which reaches a significant level of 5%. This indicates that the average order profit level of category technology is significantly different from that of furniture types. Specifically, compared with the average order profit of furniture, the average order profit of category technology is lower.

The regression coefficient of independent variable sales is 1.168118, which reaches the significance level of 5%, indicating that order sales can significantly and positively affect order profits. Specifically, when other variables remain unchanged, the average profit level of orders will increase by 1.168118 units as the average value of order sales increases by 1 unit.

The regression coefficient of the independent variable quantity is 0.058,702, which reaches the significance level of 5%, indicating that the quantity can significantly negatively affect the profit of the order. Specifically, with other variables unchanged, the average profit level of the order will increase by 4.232e+00 units as the average quantity increases by one unit.

The regression coefficient of the independent variable discount is -1.113843, and the regression coefficient reaches the significance level of 5%, which indicates that the discount can significantly and negatively affect the order profit. Specifically, with other variables unchanged, the average profit level of orders will decrease by 4.232e+00 units as the average discount value increases by one unit.

Argument shipping_ The regression coefficient of cost is 0.013865, which reaches the significance level of 5%, indicating that shipping_ Cost will have a significant negative impact on order profits. Specifically, while keeping other variables unchanged, if the average cost increases by one unit, the average profit level of the order will increase by 0.013865 units.

Results

This study found that the order profits of the large supermarkets investigated in this study are basically concentrated around 0, while those data points that exceed 0 are large, that is, most order profits of the large supermarkets are at the profit level. When further discussing the influencing factors of the order profit of large supermarkets, the sales market is an important stratification factor, that is, different sales markets can be used to explain the order profit of supermarkets. In all markets, the profit level brought by the EU market is very low. In addition, the type of goods is also an important factor affecting sales profit. Compared with furniture, the profit level of office supplies and technological products is higher. Sales revenue is a positive factor affecting sales profit, that is, the higher the sales revenue, the higher the sales profit. Sales volume is a negative factor affecting sales profit, that is, the higher the sales volume, the lower the sales profit. Sales discount is also a negative factor affecting sales profit. The transportation cost of sales is a negative predictor of sales revenue.

Discussion

First of all, this study found that the sales market is an important factor affecting sales profit, and this factor will have a certain impact on the relationship between sales market, type of goods sold, sales volume, sales quantity, discount, transportation cost and sales profit.

Secondly, this study also found that the type of goods sold is also an important factor affecting the sales profit of enterprises. The results of this study also confirmed that the higher the added value of goods, the higher the sales profit. Compared with furniture sales, the higher the added value of office supplies, especially scientific and technological products, the greater the profit space of enterprises.

Third, this study also found that sales revenue is a positive factor affecting sales profit. There is no doubt about the result. Sales revenue is the premise of sales profit. Good sales revenue provides sufficient guarantee for sales profit, and also provides an important prerequisite for the improvement of profit level.

Lastly, this study unexpectedly found that sales volume has become an obstacle to improve sales profit. If an enterprise's sales volume is higher, its profit level will be lower, and this result is undoubtedly contrary to the actual situation. The possible reason is that the increase of sales volume does not bring greater marginal profit, but increases the subsequent transportation cost and discount cost.

Discount has become a factor hindering the improvement of profits. This result is also very strange. The results show that the discount level has a negative impact on the sales profit of enterprises, which is also predictable. The discount level is undoubtedly put forward for promotion, and the important reason for promotion is the difficulty of sales itself. Therefore, the discount in the investigation order may be to promote the goods, and the purpose is to sell the goods without considering the profit level.

Transportation cost is undoubtedly a negative predictor of sales profit, which is also extremely reasonable. The increase of cost will inevitably reduce the profit level, thus reducing the profit brought by the enterprise's sales revenue.

Appendix

Result 1

```
summary(SuperStoreOrders$profit)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -6599.98     0.00     9.24    28.64    36.81  8399.98
```

Result 2

```
aggregate(SuperStoreOrders$profit,by=list(SuperStoreOrders$market),mean)
```

```
##   Group.1         x
## 1  Africa 19.374674
## 2    APAC 39.772549
## 3  Canada 46.399453
## 4    EMEA  8.728966
## 5      EU 37.282974
## 6   LATAM 21.531328
## 7      US 28.656896
```

```
summary(aov(SuperStoreOrders$profit~SuperStoreOrders$market))
```

```
##                             Df    Sum Sq Mean Sq F value Pr(>F)
## SuperStoreOrders$market      6 5.139e+06  856559   28.24 <2e-16 ***
## Residuals                51283 1.555e+09   30327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result 3

```
aggregate(SuperStoreOrders$profit,by=list(SuperStoreOrders$category),mean)
```

```
##           Group.1        x
## 1       Furniture 29.03830
## 2 Office Supplies 16.57896
## 3      Technology 65.45496
```

```
summary(aov(SuperStoreOrders$profit~SuperStoreOrders$category))
```

```
##                               Df    Sum Sq Mean Sq F value Pr(>F)
## SuperStoreOrders$category      2 1.830e+07 9147662   304.2 <2e-16 ***
## Residuals                  51287 1.542e+09   30068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result 4

```
cor.test(as.numeric(SuperStoreOrders$profit),as.numeric(SuperStoreOrders$sales))
```

```
##
##  Pearson's product-moment correlation
##
## data:  as.numeric(SuperStoreOrders$profit) and as.numeric(SuperStoreOrders$sales)
## t = 125.92, df = 51288, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4793056 0.4925272
## sample estimates:
##       cor
## 0.4859442
```

Result 5

```
cor.test(SuperStoreOrders$profit,SuperStoreOrders$quantity)
```

```
##
##  Pearson's product-moment correlation
##
## data:  SuperStoreOrders$profit and SuperStoreOrders$quantity
## t = 23.852, df = 51288, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.09617555 0.11329434
## sample estimates:
##       cor
## 0.1047427
```

Result 6

```
cor.test(SuperStoreOrders$profit,SuperStoreOrders$discount)
```

```
##
##  Pearson's product-moment correlation
##
## data:  SuperStoreOrders$profit and SuperStoreOrders$discount
## t = -75.529, df = 51288, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3241418 -0.3085655
## sample estimates:
##       cor
## -0.316375
```

Result 7

```
cor.test(SuperStoreOrders$profit,SuperStoreOrders$shipping_cost)
```

```
##
##  Pearson's product-moment correlation
##
## data:  SuperStoreOrders$profit and SuperStoreOrders$shipping_cost
## t = 86.562, df = 51288, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3494588 0.3645613
## sample estimates:
##       cor
## 0.3570334
```

Result 8

```
model0 <- lmer(profit~(1|market),data=SuperStoreOrders)
summary(model0)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: profit ~ (1 | market)
##    Data: SuperStoreOrders
##
## REML criterion at convergence: 674874.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -38.063  -0.189  -0.098   0.053  48.070
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  market   (Intercept)   146.1   12.09
##  Residual             30327.2  174.15
## Number of obs: 51290, groups:  market, 7
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   28.021      4.745   5.905
```

```
ICC = 146.1/(146.1+30327.2)
ICC
```

```
## [1] 0.004794361
```

Result 9

```
SuperStoreOrders$sales =log(as.numeric(SuperStoreOrders$sales)+1)
SuperStoreOrders$quantity=log(SuperStoreOrders$quantity+1)
SuperStoreOrders$discount =log(SuperStoreOrders$discount+1)
SuperStoreOrders$shipping_cost =log(SuperStoreOrders$shipping_cost+1)
SuperStoreOrders=na.omit(SuperStoreOrders)
modelfull <- lmer(profit~ category + sales + quantity + discount + shipping_cost+(1|market),data=SuperS
summary(modelfull)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: profit ~ category + sales + quantity + discount + shipping_cost +
##     (1 | market)
##    Data: SuperStoreOrders
##
##       AIC       BIC    logLik  deviance  df.resid
##  666662.8  666742.4 -333322.4  666644.8     51281
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -41.380  -0.235  -0.090   0.164  50.881
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  market   (Intercept)    62.09   7.88
##  Residual             25830.64 160.72
## Number of obs: 51290, groups:  market, 7
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)               -68.3676     5.3197 -12.852
## categoryOffice Supplies    21.7480     2.1518  10.107
## categoryTechnology         23.4267     2.2805  10.273
## sales                      18.5672     1.3264  13.998
## quantity                   -0.7513     1.8499  -0.406
## discount                 -282.4554     4.4203 -63.900
## shipping_cost              12.4841     1.3045   9.570
##
## Correlation of Fixed Effects:
##             (Intr) ctgrOS ctgryT sales  quntty discnt
## ctgryOffcSp -0.498
## ctgryTchnlg -0.205  0.519
## sales       -0.589  0.273 -0.019
## quantity    -0.064 -0.255  0.023 -0.295
## discount    -0.258  0.185  0.038  0.177 -0.133
## shippng_cst  0.365 -0.002 -0.005 -0.837  0.019 -0.037
```

```
AIC(modelfull)
```

```
## [1] 666662.8
```

```
BIC(modelfull)
```

```
## [1] 666742.4
```

```
anova(model0,modelfull,REML=FALSE)
```

```
## refitting model(s) with ML (instead of REML)

## Data: SuperStoreOrders
## Models:
## model0: profit ~ (1 | market)
```

```
## modelfull: profit ~ category + sales + quantity + discount + shipping_cost + (1 | market)
##           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model0       3 674885 674912 -337440   674879
## modelfull    9 666663 666742 -333322   666645 8234.6  6  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```