

上市公司分析报告

孙智

一、 指标的选择

本次报告选取北京地区从事互联网、酒店餐饮、服饰三个行业的 21 家上市公司总资产、营业收入、净利润、净收益率等指标, 针对北京市上市企业的盈利水平及跨行业差异进行分析, 并提供数据分析结果。

二、 分析过程

1. 数据的读取

```
1. dir.create("E:/Rdata/Exp9")
2. setwd("E:/Rdata/Exp9")
3. data<-
  read.csv('E:/data/homework/2021finalexam.csv', header = TRUE, stringsAsFactors
    = FALSE)
4. head(data, 10)
5. # 选取北京地区从事互联网、服饰、酒店餐饮这三个行业的企业总资产及净收益率
6. data1<-subset(data, data$地区 == '北京', select = c(1, 2, 4, 16 ))
7. data2<-subset(data1, data1$细分行业 == '互联网' | data1$细分行业 == '服饰
  ' | data1$细分行业 == '酒店餐饮', select = c(1, 2, 3, 4))
8. colnames(data2)<-c('名称', '行业', '总资产', '净收益率')
```

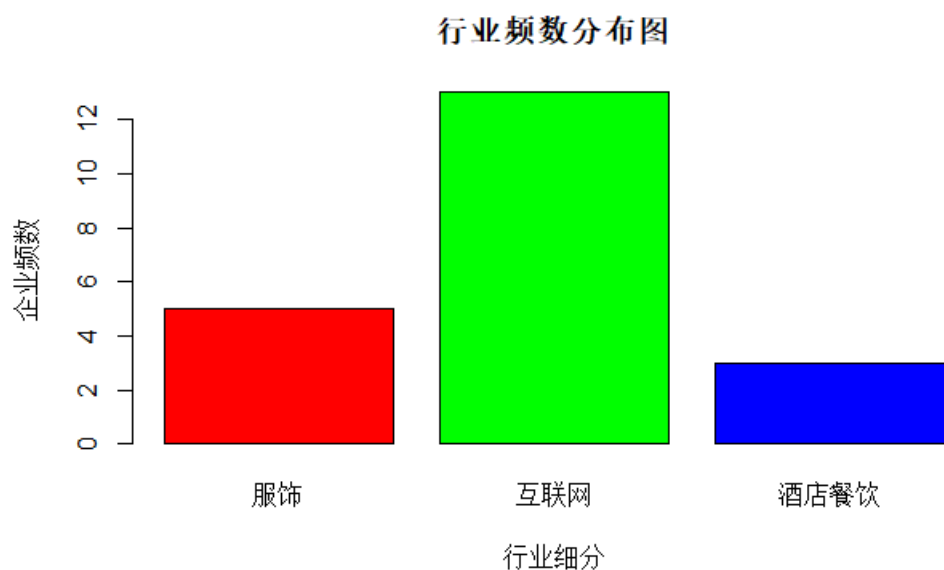
2. 频数分布图

对数据中北京地区从事互联网、服饰、酒店餐饮行业的上市企业的频数进行统计, 并绘制频数分布图。

```
1. # 统计北京市三个行业频数
2. a<-table(data2$行业)
3. # 绘制三个行业在北京市的频数分布图
4. barplot(a, col = rainbow(3), xlab = '行业细分', ylab = '企业频数', main = '行
  业频数分布图')
```

表格 1 各行业企业频数分布表

行业	频数
服饰	5
互联网	13
酒店餐饮	3



根据表格 1 各行业企业频数分布表，北京市三种行业中，互联网企业上市数量最多，酒店餐饮上市数量最少，反映了近年来北京市互联网行业的蓬勃发展，以及北京地区互联网行业丰富的资源，受疫情等各方面因素影响，餐饮行业呈现出“不景气”的现状，因而上市公司中酒店餐饮业所占比重较少。

3. 互联网企业分布直方图

资产总额是指企业资产负债表的资产总计项。公司过去的交易、事项形成并由企业拥有或控制的资源，资产总额即是指企业拥有或控制的全部资产，因此资产总额在一定程度上反映了企业的经营规模。本报告通过对互联网企业资产总额进行划分，对北京地区互联网企业经营现状进行分析。

```

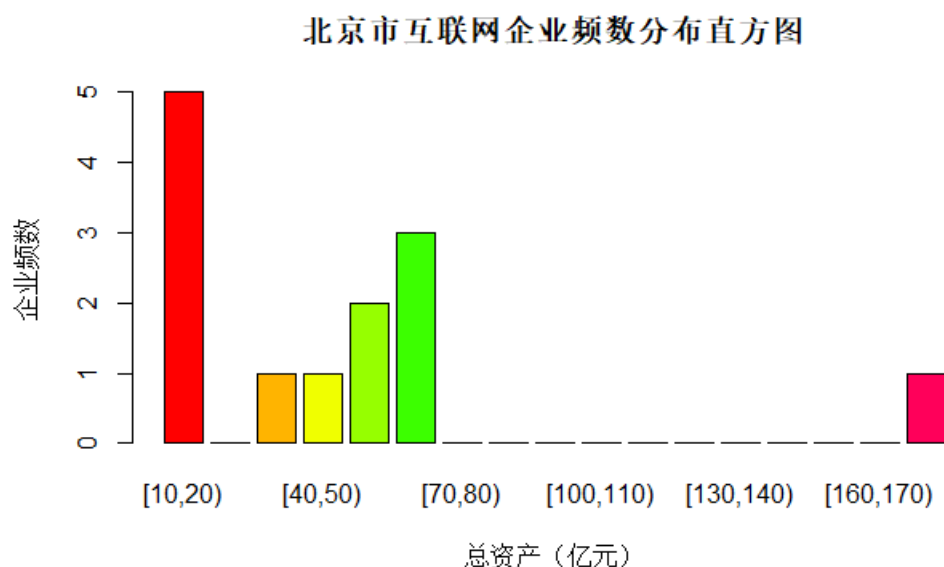
1. # 绘制互联网行业总资产的直方图
2. data3<-data2[data2$行业 == '互联网', ]
3. range(data3$总资产)
4. length(data3$总资产)

```

```

5. data3$总资产<-cut(data3$总资产, seq(10, 180, by = 10), right = FALSE)
6. table(data3$总资产)
7. plot(data3$总资产, col = rainbow(17), xlab = '总资产（亿元）', ylab = '企业频数', main = '北京市互联网企业频数分布直方图')

```



```

1. library(fBasics)
2. skewness(as.numeric(data3$总资产))

```

如图所示，北京市互联网企业分布呈现出明显的右偏趋势，根据偏度计算得到数据的偏态系数等于 1.76，即为高度右偏，说明在北京市从事互联网行业的企业中，大多数企业处于中小规模，尤其以小规模企业为主，有极少数上市企业拥有较大的资产总额，反映了北京地区互联网行业发展的不均衡问题。

4. 方差分析

近年来随着 5G 的普及，“互联网”成为热词，本报告利用数据中从事互联网、酒店餐饮、服饰的上市企业净益率进行方差分析，对从事各行业的净益率是否有显著差异进行方差分析。

```

1. # 比较北京市互联网、酒店餐饮、服饰企业净益率是否有显著差异
2. y<-data2$净益率
3. x<-data2$行业
4. data2.aov<-aov(y ~ x, data = data2[, c(2, 4)])
5. summary(data2.aov)

```

表格 2 ANOVA 表

	Df	SumSq	MeanSq	F value	Pr(>F)
x	2	0.51	0.256	0.018	0.982
Residuals	18	250.35	13.908		

根据

表格 2 ANOVA 表，在显著性水平为 0.05 的情况下，不能认为各个行业净益率存在显著差异，净益率作为衡量上市公司盈利能力的重要指标，反映了北京市各个上市公司的盈利能力在各行业之间并不存在显著差异，笼统地认为互联网上市公司盈利能力超过传统酒店餐饮和服饰公司是不严谨的。

5. 相关性分析

为了进一步研究上市企业能力的影响因素，本报告选取 3 个行业总资产、营业收入、净利润、净益率作相关性分析，判断指标之间是否存在显著的线性关系，采用 Pearson 相关系数进行分析。

```
1. # 选取北京市从事互联网、服饰、酒店餐饮这三个行业的企业总资产、营业收入、净利润、净益率做相关性分析
2. data4<-subset(data, data$地区 == '北京', select = c(1, 2, 4, 9, 11, 16 ))
3. data5<-subset(data4, data4$细分行业 == '互联网' | data4$细分行业 == '服饰' | data4$细分行业 == '酒店餐饮')
4. data5$净益率.<-as.numeric(data5$净益率.)
5. cor(x = data5[, 3:6], y = NULL, method = 'pearson')
6. cor.test(data5$净利润.亿., data5$净益率., method = 'pearson')
```

表格 3 相关系数表

	总资产	营业收入	净利润	净益率
总资产	1	0.408042	-0.02349	-0.0724
营业收入	0.408042	1	0.090054	0.092544
净利润	-0.02349	0.090054	1	0.659531
净益率	-0.0724	0.092544	0.659531	1

表格 4 相关系数检验

t	df	p-value	lower	upper
3.8245	19	0.001144	0.318534	0.849388

由表格 3 相关系数表 可知净利润与净益率之间 Pearson 相关系数等于 0.6595，说明两个指标之间存在正线性相关关系，进而利用 T 检验对相关系数的显著性进行检验，根据表格 4 相关系数检验结果，在显著性水平为 0.05 的情况下，可以认为净利润与净益率之间的线性关系是显著的。

6. 一元线性回归

基于相关分析中净收益率与净利润之间存在显著线性关系, 此部分利用一元线性回归模型给出二者线性关系的数学表达式, 并对模型的显著性进行检验。

```
1. y1<-data5$净收益率.
2. x1<-data5$净利润.亿.
3. g_lm<-lm(y1 ~ x1, data = data5[, c(5, 6)])
4. summary(g_lm)
5. aov(g_lm)# 对回归系数进行方差分析
```

表格 5 线性回归检验结果

	Estimate	Std.Error	t value	Pr(> t)	F-statistic	p-value
(Intercept)	-0.2861	0.5986	-0.478	0.63808		
x1	2.2204	0.5806	3.825	0.00114	14.63	0.00114

根据表格 5 线性回归检验结果, 净收益率与净利润之间的数学表达式为:

$$\hat{y}_1 = 2.2204 \times x_1 - 0.2861$$

并且根据 T 检验和 F 检验结果, 在显著性水平为 0.05 的情况下, 可以认为净收益率与净利润之间的线性关系是显著的。

三、 结论

根据以上针对北京地区从事互联网、酒店餐饮、服饰等上市企业的相关数据进行分析, 主要得出以下结论。

1. 北京地区在上市企业中, 以上三个行业存在企业数量严重不平衡现象, 反映出近年来互联网行业的火热以及传统行业的式微。
2. 北京地区互联网行业经营状况也存在显著差异, 以中小型企业为主, 大规模企业较少, 存在空白值的现象, 反映了互联网行业拥有巨大潜力。
3. 针对三个行业中在北京地区的上市企业, 不能认为其盈利能力存在显著差异。
4. 对于上市企业而言, 其盈利能力主要受净利润的影响, 因此扩大经营提高净收益对于企业提高盈利能力有着重要意义。

四、 R 代码

```
1. dir.create("E:/Rdata/Exp9")
2. setwd("E:/Rdata/Exp9")
3. data<-
  read.csv('E:/data/homework/2021finalexam.csv', header = TRUE, stringsAsFactors = FALSE)
4. head(data, 10)
```

```

5. # 选取北京地区从事互联网、服饰、酒店餐饮这三个行业的企业总资产及净益率
6. data1<-subset(data, data$地区 == '北京', select = c(1, 2, 4, 16 ))
7. data2<-subset(data1, data1$细分行业 == '互联网' | data1$细分行业 == '服饰'
  '| data1$细分行业 == '酒店餐饮', select = c(1, 2, 3, 4))
8. colnames(data2)<-c('名称', '行业', '总资产', '净益率')
9. # 统计北京市三个行业频数
10. a<-table(data2$行业)
11. # 绘制三个行业在北京市的频数分布图
12. barplot(a, col = rainbow(3), xlab = '行业细分', ylab = '企业频数', main = '行业频数分布图')
13. # 绘制互联网行业总资产的直方图
14. data3<-data2[data2$行业 == '互联网', ]
15. range(data3$总资产)
16. length(data3$总资产)
17. data3$总资产<-cut(data3$总资产, seq(10, 180, by = 10), right = FALSE)
18. table(data3$总资产)
19. plot(data3$总资产, col = rainbow(17), xlab = '总资产（亿元）', ylab = '企业频数',
  main = '北京市互联网企业频数分布直方图')
20. # 计算偏态系数
21. library(fBasics)
22. skewness(as.numeric(data3$总资产))
23. # 比较北京市互联网、酒店餐饮、服饰企业净益率是否有显著差异
24. y<-data2$净益率
25. x<-data2$行业
26. data2.aov<-aov(y ~ x, data = data2[, c(2, 4)])
27. summary(data2.aov)
28. # 选取北京市从事互联网、服饰、酒店餐饮这三个行业的企业总资产、营业收入、净利润、净益率做相关性分析
29. data4<-subset(data, data$地区 == '北京', select = c(1, 2, 4, 9, 11, 16 ))
30. data5<-subset(data4, data4$细分行业 == '互联网' | data4$细分行业 == '服饰'
  '| data4$细分行业 == '酒店餐饮')
31. data5$净益率.<-as.numeric(data5$净益率.)
32. cor(x = data5[, 3:6], y = NULL, method = 'pearson')
33. cor.test(data5$净利润.亿., data5$净益率., method = 'pearson')
34. # 利用简单线性回归对净利润和净益率做回归分析
35. y1<-data5$净益率.
36. x1<-data5$净利润.亿.
37. g_lm<-lm(y1 ~ x1, data = data5[, c(5, 6)])
38. summary(g_lm)
39. aov(g_lm)# 对回归系数进行方差分析

```