

模块二 - 线上女装销售

张笑竹 / 201618070114

2019年7月5日

背景介绍

本案例以某电商连衣裙的线上销售数据为研究对象，先对描述连衣裙的款式风格的文本变量进行词云分析，提取5个主要影响因素后，再加入商品价格，消费者适用年龄等控制变量进行多元回归。最后，还考虑了年龄分别与这5个关注的文本变量的交互作用。

数据来自某购物网站上月销售量超过100件的连衣裙，样本量为5880，即有5880款不同的连衣裙，我们使用连衣裙的月销售量作为响应变量，而解释变量为：商品价格、适用人群等商品本身属性，以及店铺评分、地点等于商品相关的线上店铺属性信息。

1 任务一：数据清洗及词云制作

进行简单的词频分析，去除评价中的无关字符（如英文、数字），去除诸如“女装”，“夏季”等无意义字符后，进行词云分析：

```
library(stringr)
setwd('C:\\Users\\Desktop\\统计仿真实验\\2')
dat <- read.csv('data.csv')
dat1 <- dat[!is.na(dat$price),]
s <- as.integer(str_extract_all(dat1$monthly_sales, '\\d+'))
dat1 <- cbind(dat1, s)

# 词云
library(jiebaR)
```

```
## Loading required package: jiebaRD
```

```
sg <- segment(paste(dat1$goodsName), worker())
fq <- freq(sg)
fq <- fq[order(fq$freq, decreasing = T),]
fq1 <- fq[1:30,]
fq1 <- fq1[-c(2, 4, 5, 8, 11, 17, 21, 25, 29),]
fq1
```

```
##      char freq
## 4245 连衣裙 6039
## 1872 新款 3796
## 1894 修身 2048
## 2034 夏装 1729
## 1906 裙子 1549
## 1251 短袖 1475
## 1974 中长款 1386
## 1756 印花 1316
## 1858 韩版 1293
## 3729 显瘦 1276
## 2147 无袖 985
## 1810 雪纺 932
## 2039 长裙 923
## 710 宽松 887
## 4517 气质 711
## 2348 时尚 661
## 1895 蕾丝 654
## 3737 大码 606
## 2385 条纹 521
## 2486 新品 519
## 1885 收腰 497
```

```
library(wordcloud2)
wordcloud2(fq1)
```



由词云图可知，“新款”、“中长款”、“修身”、“显瘦”、“韩版”为5个主要影响因素；因此我们根据词云结果，提取五个关键标签进行后续分析。

2 任务二：描述性统计和可视化

可视化之前对数据进行预处理，提取出商品的适用年龄变量（因为适用年龄是一个重要变量，因此将数据中缺失年龄的样本剔除，并重复上文中的关键词提取工作）。根据词云结果，将“新款”、“中长款”、“修身”、“显瘦”、“韩版”作为二值变量加入数据集，同时将价格和销量取对数，将0-1变量改为“是/否”后，重新组成分析用的新数据集，记为lgdata3。

操作过程如下：

```

# 提取年龄
loc <- str_locate_all(dat1$productDescription, '适用年龄:?.')
loc1 <- str_locate_all(dat1$productDescription, '周岁')
age <- vector('character', nrow(dat1))
for(i in 1:length(loc)){
  if(length(loc[[i]][,1])==0){
    age[i] <- NA
  }else{
    age[i] <- str_sub(dat1$productDescription[i], loc[[i]][,2]+2, loc1[[i]][,1]-1)
  }
}

dat1 <- cbind(dat1, age)
dat1 <- dat1[!is.na(age),]

# “新款”加入数据集
xinkuan <- vector('integer', nrow(dat1))
l_xinkuan <- str_locate_all(dat1$goodsName, '新款')
for(i in 1:length(l_xinkuan)){
  if(length(l_xinkuan[[i]][,1])==0){
    xinkuan[i] <- 0
  }else{
    xinkuan[i] <- 1
  }
}

# “中长款”加入数据集
zckuan <- vector('integer', nrow(dat1))
l_zckuan <- str_locate_all(dat1$goodsName, '中长款')
for(i in 1:length(l_zckuan)){
  if(length(l_zckuan[[i]][,1])==0){
    zckuan[i] <- 0
  }else{
    zckuan[i] <- 1
  }
}

# “修身”加入数据集
xiushen <- vector('integer', nrow(dat1))
l_xiushen <- str_locate_all(dat1$goodsName, '修身')
for(i in 1:length(l_xiushen)){
  if(length(l_xiushen[[i]][,1])==0){
    xiushen[i] <- 0
  }else{
    xiushen[i] <- 1
  }
}

# “显瘦”加入数据集
xianshou <- vector('integer', nrow(dat1))
l_xianshou <- str_locate_all(dat1$goodsName, '显瘦')
for(i in 1:length(l_xianshou)){
  if(length(l_xianshou[[i]][,1])==0){
    xianshou[i] <- 0
  }else{
    xianshou[i] <- 1
  }
}

```

```

# “韩版”加入数据集
hanban <- vector('integer', nrow(dat1))
l_hanban <- str_locate_all(dat1$goodsName, '韩版')
for(i in 1:length(l_hanban)){
  if(length(l_hanban[[i]][,1])==0){
    hanban[i] <- 0
  }else{
    hanban[i] <- 1
  }
}

# “印花”加入数据集
yinhua <- vector('integer', nrow(dat1))
l_yinhua <- str_locate_all(dat1$goodsName, '印花')
for(i in 1:length(l_yinhua)){
  if(length(l_yinhua[[i]][,1])==0){
    yinhua[i] <- 0
  }else{
    yinhua[i] <- 1
  }
}

# 将价格和销量取对数
log_price <- log(dat1$price)
log_sale <- log(dat1$s)

# 将0-1变量改为是否
xinkuan_ <- rep('是', length(xinkuan))
xinkuan_[xinkuan == 0] <- '否'

zckuan_ <- rep('是', length(zckuan))
zckuan_[zckuan == 0] <- '否'

xiushen_ <- rep('是', length(xiushen))
xiushen_[xiushen == 0] <- '否'

xianshou_ <- rep('是', length(xianshou))
xianshou_[xianshou == 0] <- '否'

hanban_ <- rep('是', length(hanban))
hanban_[hanban == 0] <- '否'

lgdata3 <- cbind(dat1, xinkuan, zckuan, xiushen, xianshou, hanban,
                  xinkuan_, zckuan_, xiushen_, xianshou_, hanban_,
                  log_price, log_sale)

```

下面，通过统计图的绘制进行EDA：

2.1 销量整体的直方分布图

将月销量划分为多个区间，绘制不同区间下对应的的总销量的分布直方图：

```

yxliang <- vector('character', length(s))
yxliang[s<=150] <- '100-150'
yxliang[s>150 & s<=200] <- '150-200'
yxliang[s>200 & s<=250] <- '200-250'
yxliang[s>250 & s<=300] <- '250-300'
yxliang[s>300 & s<=400] <- '300-400'
yxliang[s>400 & s<=500] <- '400-500'
yxliang[s>500 & s<=1000] <- '500-1000'
yxliang[s>1000] <- '>1000'

df <- freq(yxliang)
df$char <- factor(df$char, levels = unique(df$char), ordered = T)
library(ggplot2)

```

```

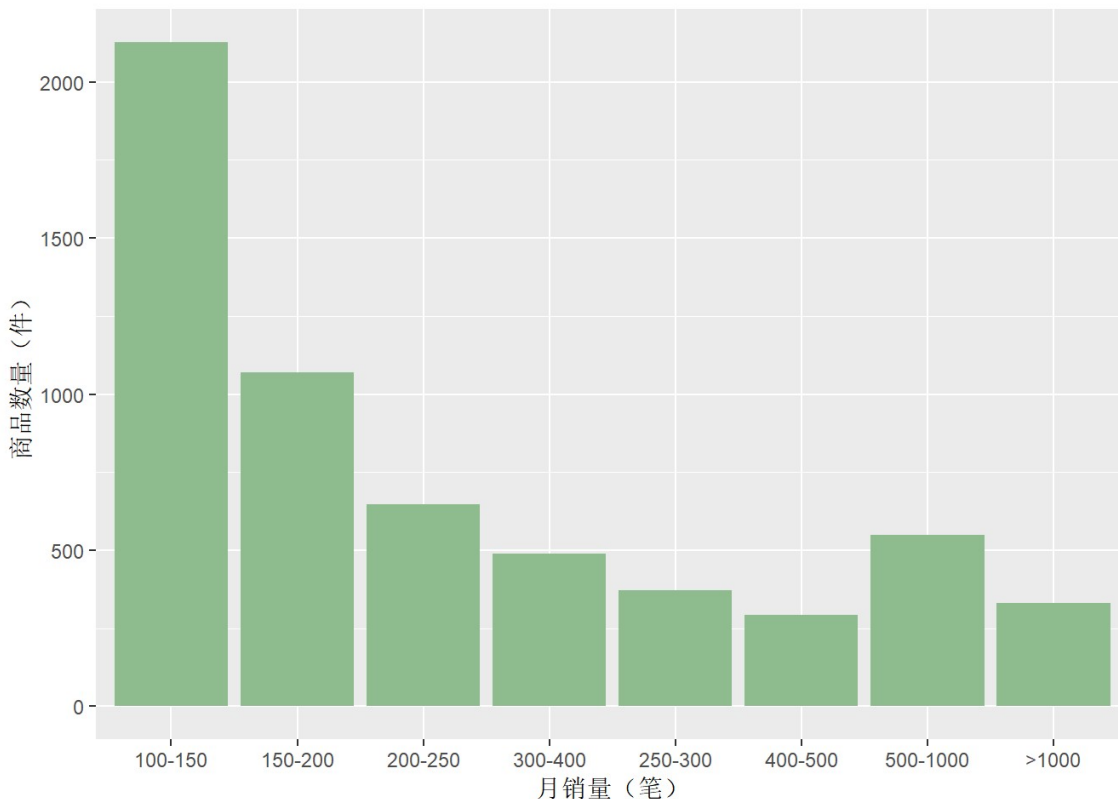
## Registered S3 methods overwritten by 'ggplot2':
##   method      from
## [.quosures    rlang
## c.quosures     rlang
## print.quosures rlang

```

```

ggplot(data = df, aes(x=char, y=freq)) +
  geom_bar(stat="identity", fill='darkseagreen') +
  xlab('月销量 (笔)') + ylab('商品数量 (件)')

```



根据该直方图，月销量的分布呈现尖峰厚尾的形态，销量处于100至150件之间的连衣裙占比35.4%，其余不同月销量区间连衣裙款式数量分布较平均，可见连衣裙市场的竞争是较为激烈的，不存在明显的垄断势力，因此对其销售量的影响因素分析是有意义的。

2.2 商品价格对销量的箱线图

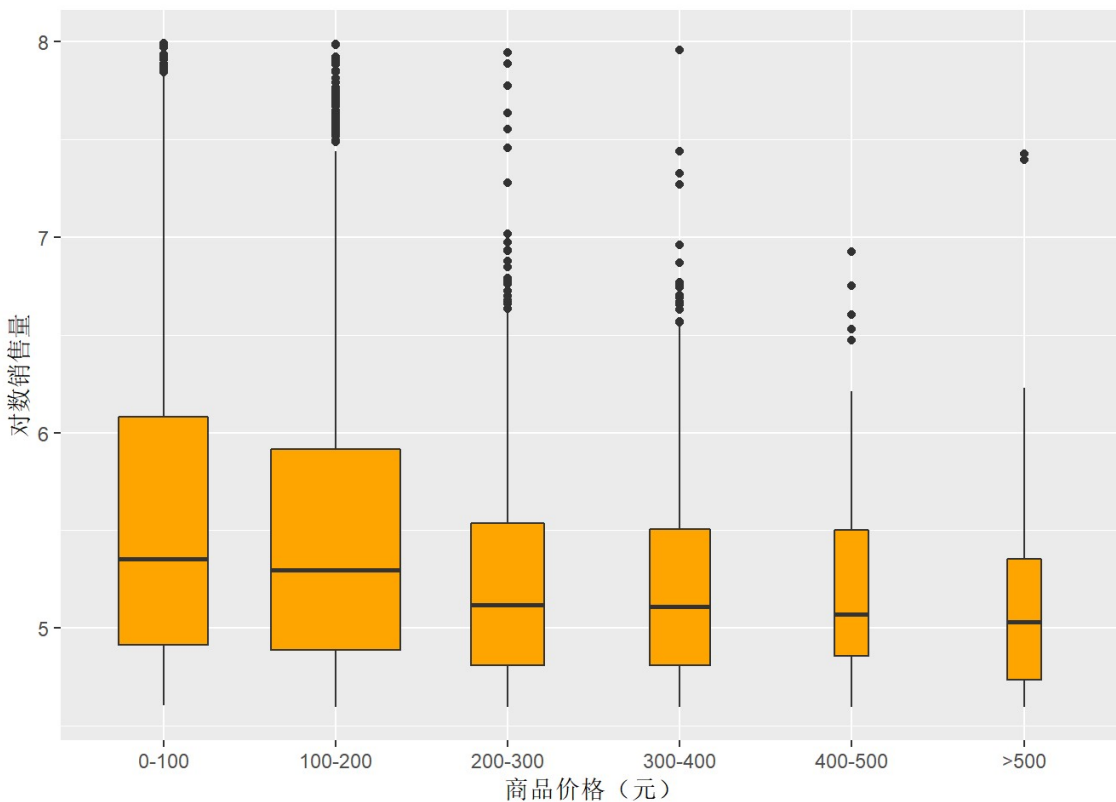
将商品价格分为6个区间，为了避免月销售量异常值的影响，我们将其取对数，再根据不同的商品价格区间的对数销

售量的箱线图。

```
jiage <- vector('character', length(lgdata3$price))
jiage[lgdata3$price<=100] <- '0-100'
jiage[lgdata3$price<=200 & lgdata3$price>100] <- '100-200'
jiage[lgdata3$price<=300 & lgdata3$price>200] <- '200-300'
jiage[lgdata3$price<=400 & lgdata3$price>300] <- '300-400'
jiage[lgdata3$price<=500 & lgdata3$price>400] <- '400-500'
jiage[lgdata3$price>500] <- '>500'

jiage <- factor(jiage, levels = c('0-100','100-200','200-300','300-400',
                                '400-500','>500'),
               ordered = T)

df <- data.frame(jiage, log_sale)
ggplot(df, aes(x=jiage,y=log_sale)) +
  geom_boxplot(fill='orange',varwidth = T) + xlab('商品价格 (元)') + ylab('对数销售量')
```



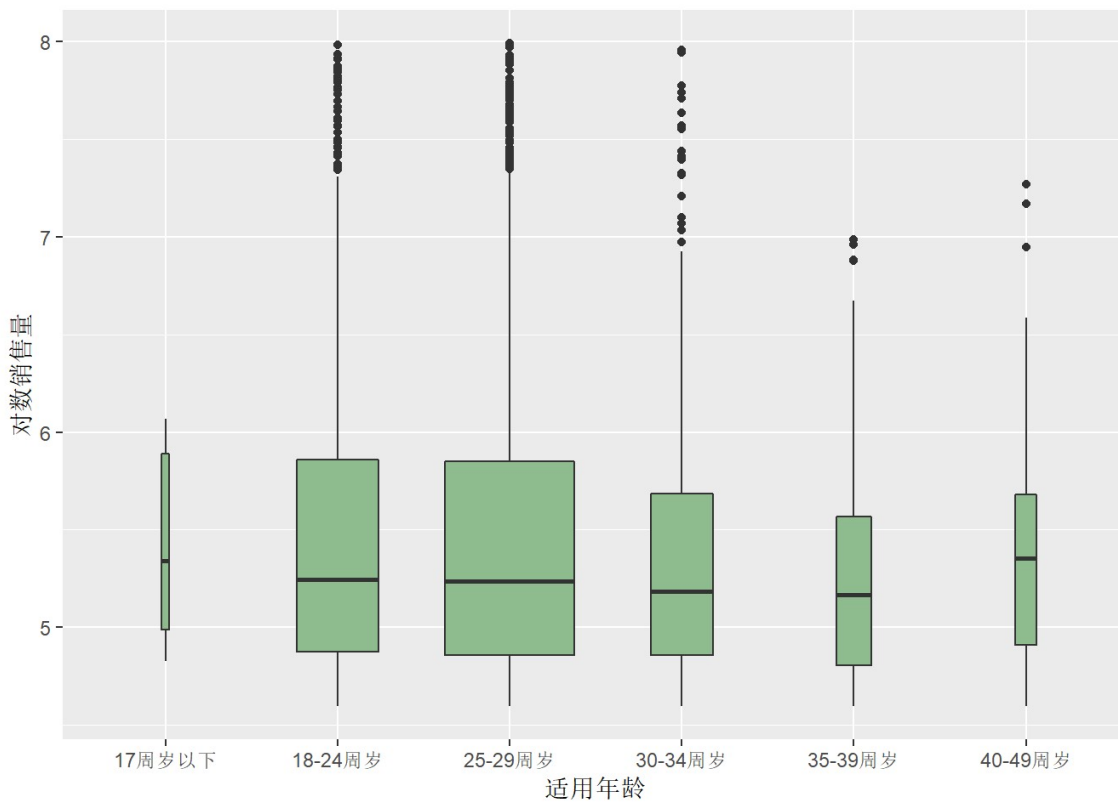
根据该箱线图，可以识别出经典的经济学规律——随着价格的升高，销售量逐渐降低。然而，销售量的变动幅度并不明显，其原因可能是销量对价格的不敏感性，同时也可能是取对数而导致的平缓趋势。

2.3 不同年龄层对月销量的影响

适用年龄分为6个年龄层，使用不同年龄层的箱线图描述不同年龄层对月销量的影响：

```
df <- data.frame(age = lgdata3$age, log_sale)
df$age <- factor(df$age, levels = c('17','18-24','25-29','30-34','35-39','40-49'),
               labels = c('17周岁以下','18-24周岁','25-29周岁','30-34周岁',
                          '35-39周岁','40-49周岁'),ordered = T)

ggplot(df, aes(x=age,y=log_sale)) +
  geom_boxplot(fill='darkseagreen',varwidth = T) + xlab('适用年龄') + ylab('对数销售量')
```

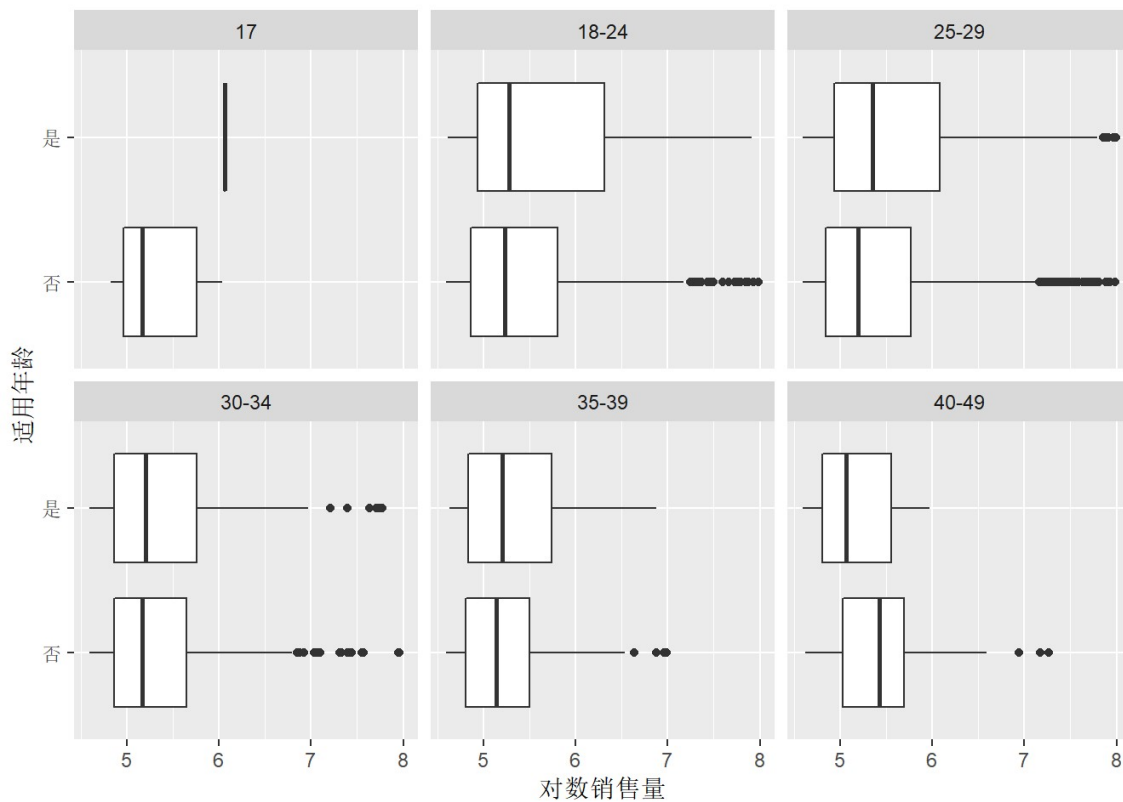


箱体宽度代表不同年龄层的商品数量，比较25-29周岁女性连衣裙线上商品数量最多；青少年及中老年消费者现有商品数量少，但平均销量高，市场潜力大；总体上来说，不同年龄的连衣裙整体销售量无明显差异。

2.4 不同款式对月销量的影响

① “显瘦”对销量的影响。

```
# 显瘦
df <- data.frame(xianshou=lgdata3$xianshou_, log_sale, age=lgdata3$age)
ggplot(df, aes(x=xianshou,y=log_sale)) +
  geom_boxplot() + xlab('适用年龄') + ylab('对数销售量') + coord_flip() +
  facet_wrap(~ age)
```

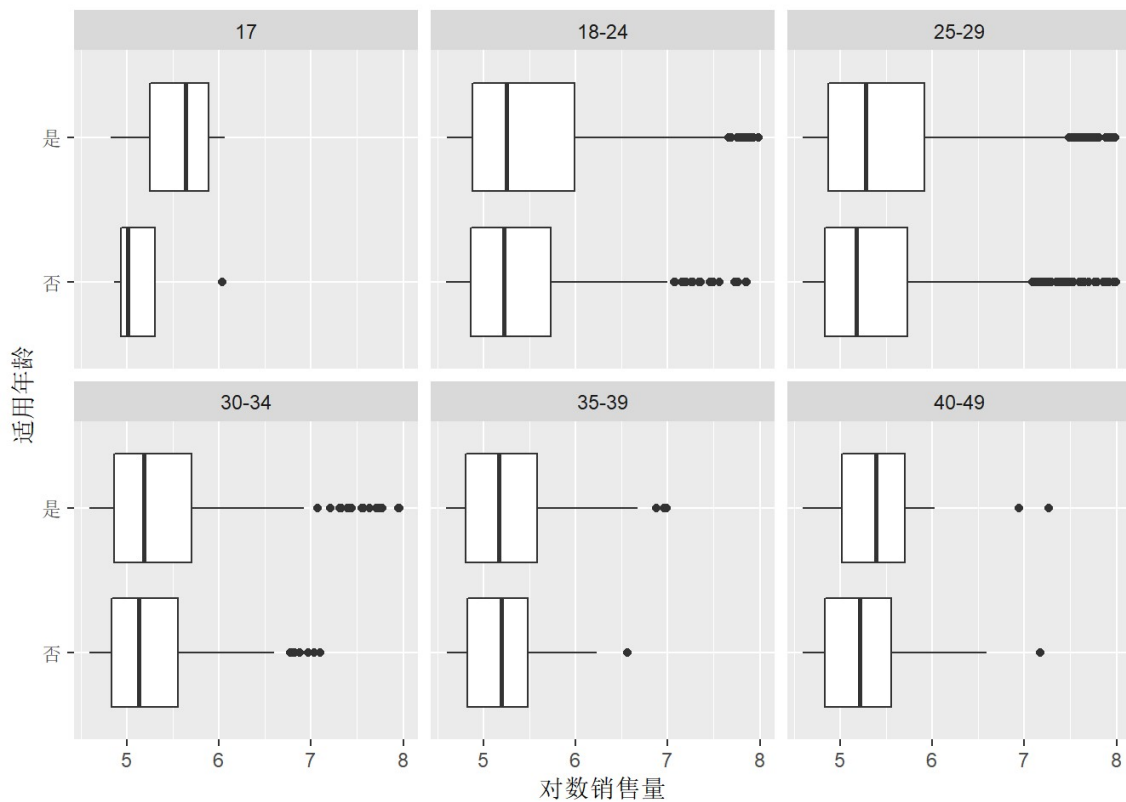



对17岁以下的年轻人，“显瘦”能够有效地提高销量；而对于40-49岁的中年人，“显瘦”却起到了相反的作用。对于其他年龄段的人群，“显瘦”的促销效果并不是十分明显。

② “新款”对销量的影响。

新款

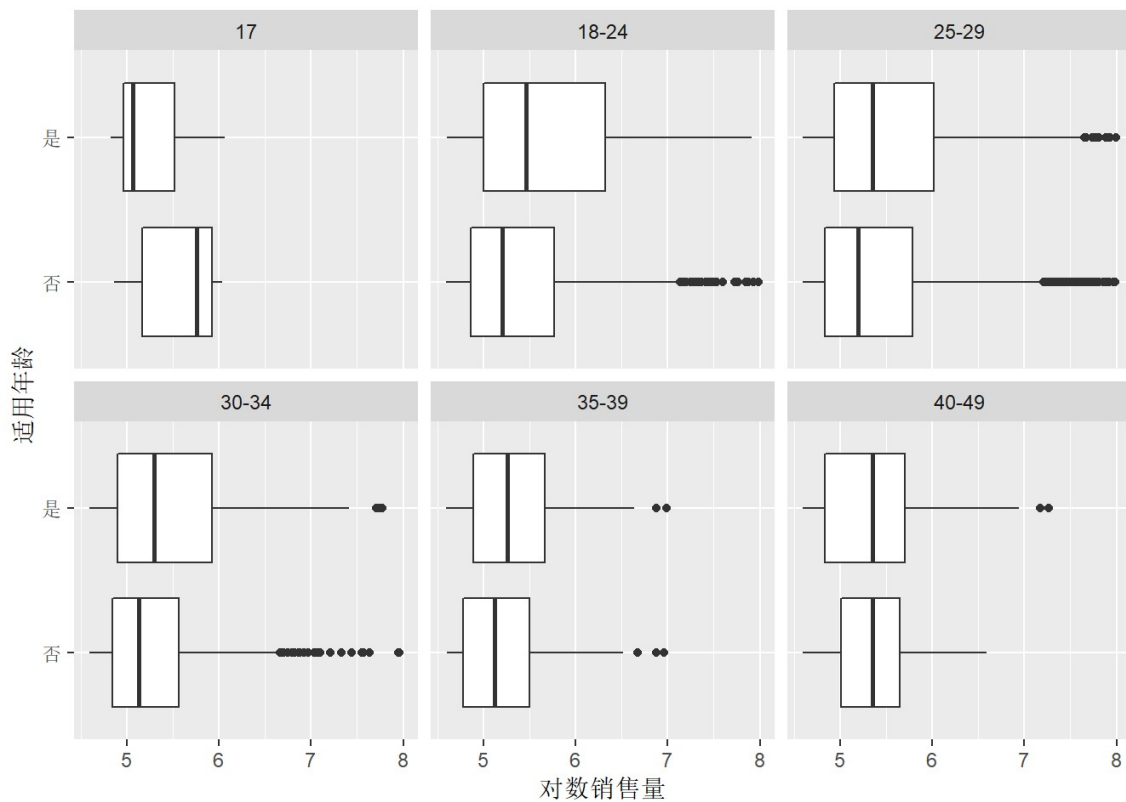
```
df <- data.frame(xinkuan=lgdata3$xinkuan_, log_sale, age=lgdata3$age)
ggplot(df, aes(x=xinkuan,y=log_sale)) +
  geom_boxplot() + xlab('适用年龄') + ylab('对数销售量') + coord_flip() +
  facet_wrap(~ age)
```



“新款”能够提高销量，且对于17岁以下人群，其效果增加明显。

③ “中长款”对销量的影响。

```
# 中长款
df <- data.frame(zckuan=lgdata3$zckuan_, log_sale, age=lgdata3$age)
ggplot(df, aes(x=zckuan,y=log_sale)) +
  geom_boxplot() + xlab('适用年龄') + ylab('对数销售量') + coord_flip() +
  facet_wrap(~ age)
```

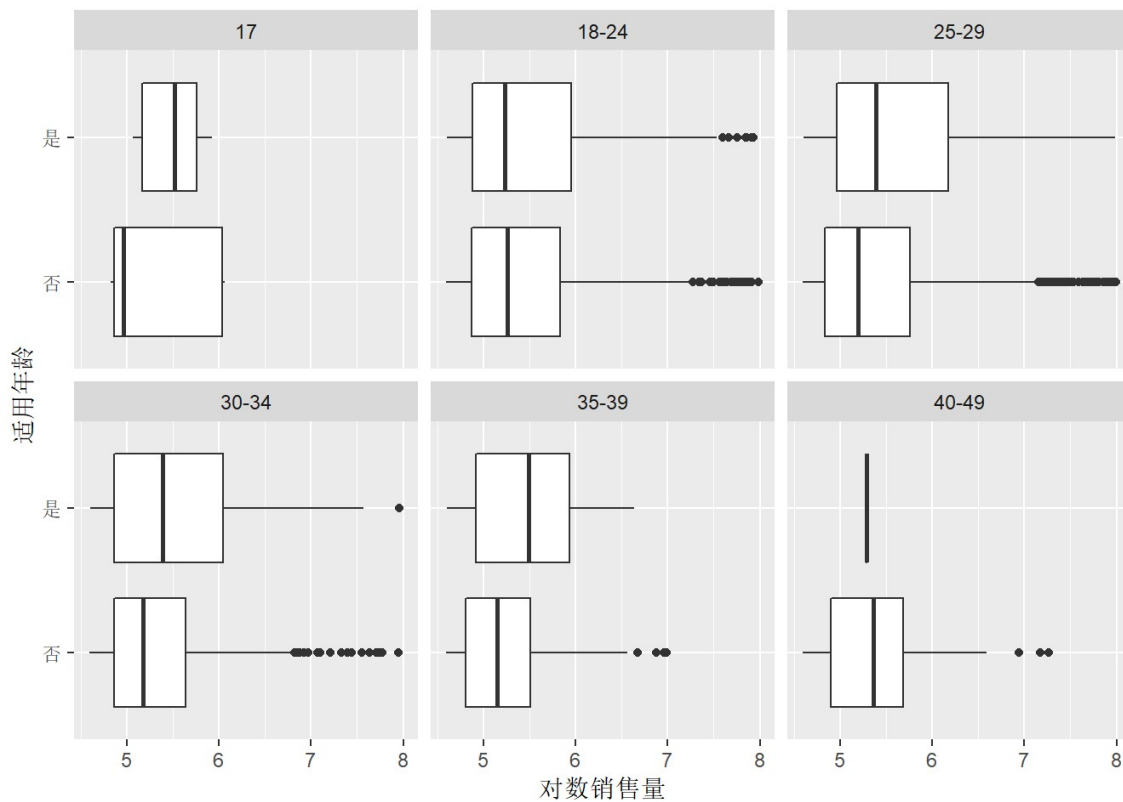


“中长款”对于18岁以上人群，能够提高销量；而对于17岁及以下人群，却起到了相反的作用。因此，不同年龄段的人群对此偏好明显不同。

④ “韩版”对销量的影响。

韩版

```
df <- data.frame(hanban=lgdata3$hanban_, log_sale, age=lgdata3$age)
ggplot(df, aes(x=hanban,y=log_sale)) +
  geom_boxplot() + xlab('适用年龄') + ylab('对数销售量') + coord_flip() +
  facet_wrap(~ age)
```



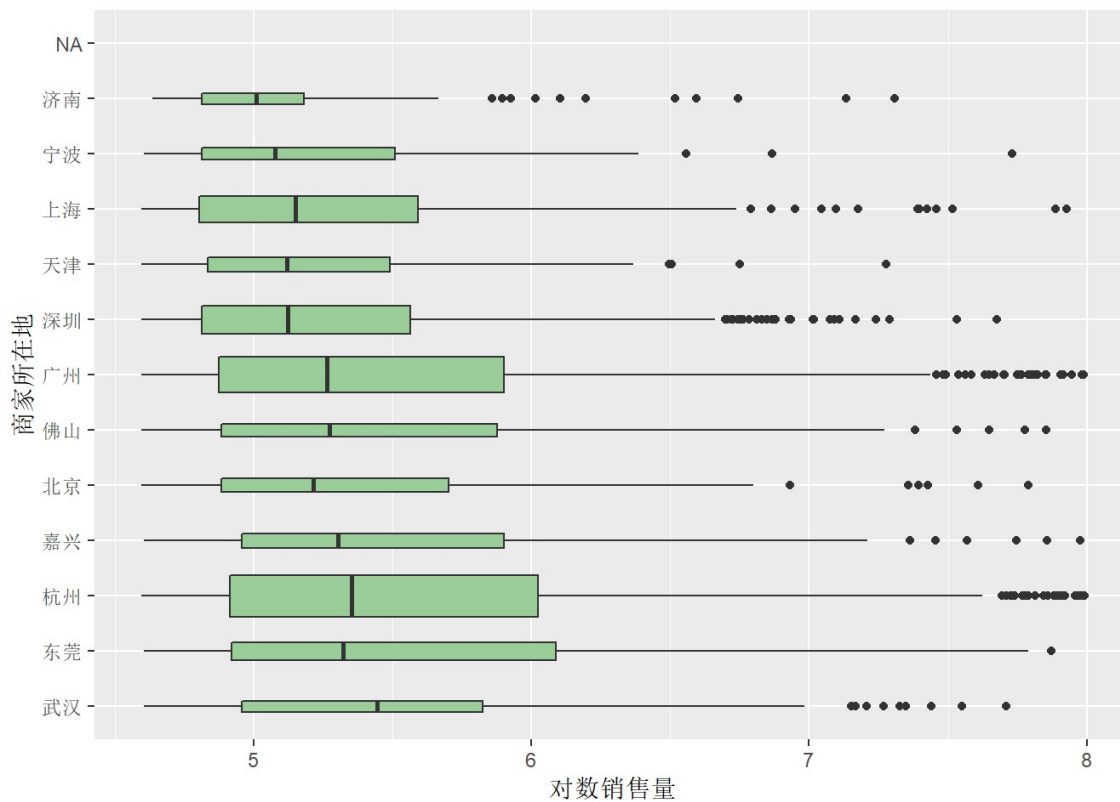
几乎对于任何年龄段，“韩版”都是块不错的招牌。不过，对于18-24岁以及40-49岁人群，“韩版”效果不如其他年龄段明显。

2.5 不同地区对月销量的影响

```
# 地区变量
area <- vector('integer', nrow(lgdata3))
l_area <- str_locate_all(lgdata3$seller, '地区')
for(i in 1: length(l_area)){
  area[i] <- str_sub(lgdata3$seller[i], l_area[[i]][,2]+2, l_area[[i]][,2]+3)
}

df <- data.frame(area, log_sale)
df <- df[(df$area == '武汉' | df$area == '东莞' | df$area == '杭州' |
  df$area == '嘉兴' | df$area == '北京' |
  df$area == '佛山' | df$area == '广州' | df$area == '深圳' |
  df$area == '天津' | df$area == '上海' | df$area == '宁波' |
  df$area == '济南'),]
df <- df[order(df, decreasing = T),]
df$area <- factor(df$area, levels = c('武汉', '东莞', '杭州', '嘉兴', '北京',
  '佛山', '广州', '深圳', '天津', '上海',
  '宁波', '济南'), ordered = T)

ggplot(df, aes(x=area, y=log_sale)) +
  geom_boxplot(fill = 'darkseagreen3', varwidth = T) + xlab('商家所在地') + ylab('对数销售量') + coord_flip()
p()
```



在上图中，箱体宽度代表该城市拥有的商品数量，因而不难看出高销量店铺分布最多的城市是杭州、广州、深圳；事实上，高销量店铺集中在东部及南部，以沿海为主，反映出江浙沪一带女装产业尤为发达的现状。武汉虽然店铺不多，但商品平均销量最高；杭州店铺多且平均销量较高；广州、深圳等虽然店铺数量庞大，但平均销量不及其他地区。

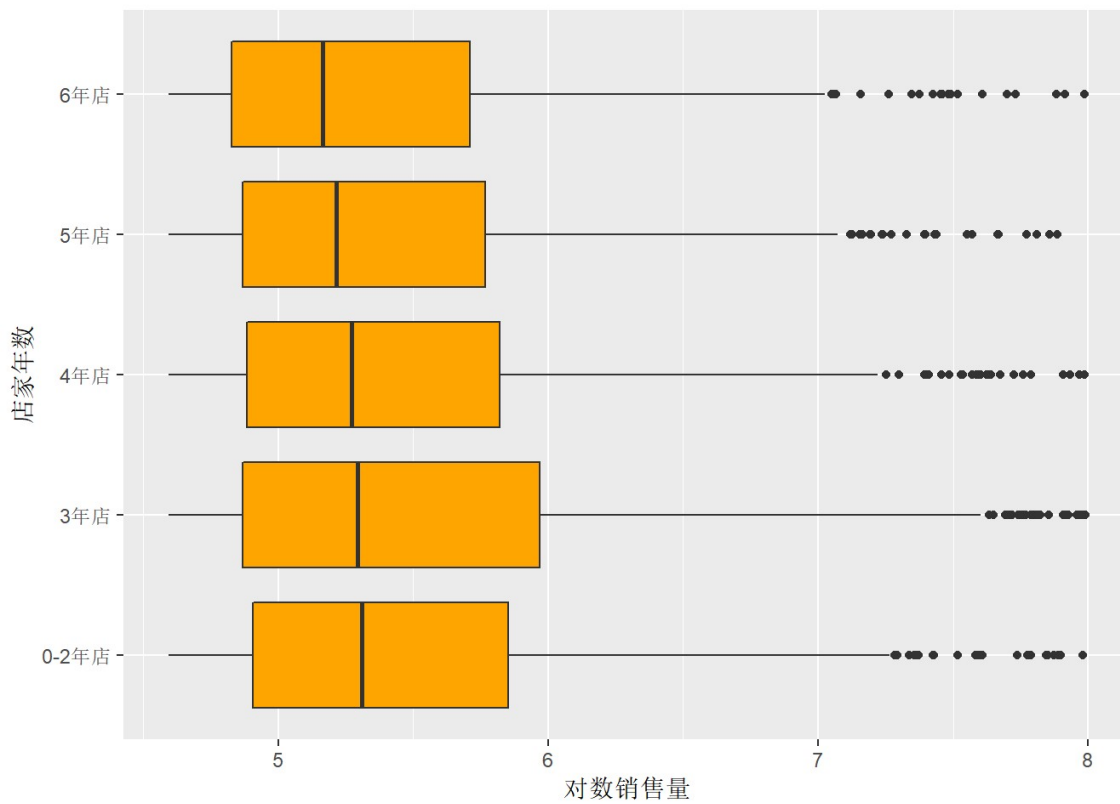
2.5 开店时常对月销量的影响

先提取开店时长的信息，再根据其分布，分为‘0-2年店’，‘2年店’，‘3年店’，‘4年店’，‘5年店’，‘6年店’六中情况绘制箱线图：

```
loc <- str_locate_all(lgdata3$seller, '年')
shijian <- vector('character', nrow(lgdata3))
for(i in 1:length(loc)){
  if(length(loc[[i]][,1])==0){
    shijian[i] <- '0-2年店'
  }else{
    shijian[i] <- str_sub(lgdata3$seller[i], loc[[i]][,1]-1, loc[[i]][,1]+1)
  }
}

df <- data.frame(shijian, log_sale)
df <- df[(df$shijian == '0-2年店' |
  df$shijian == '3年店' |
  df$shijian == '4年店' |
  df$shijian == '5年店' |
  df$shijian == '6年店'),]

ggplot(df, aes(x=shijian, y=log_sale)) +
  geom_boxplot(fill='orange') + xlab('店家年数') + ylab('对数销售量') + coord_flip()
```



根据上图结果，开店时间越短，连衣裙的月销量就相对越高，这或许是由于新店折扣力度较大，顾客新鲜度较高；当然，这样的关系并不显著。

2.6 商品评价对月销量的影响

将商品描述、物流、服务评分的箱线图放在同一张图上：

```

wuliupf <- vector('numeric',nrow(lgdata3))
wuliupf <- lgdata3$wuliu
wuliupf[wuliupf<=4.6] <- '4.6及以下'
df <- data.frame(wuliupf,log_sale)
df <- df[complete.cases(df),]
p1 <- ggplot(df, aes(x=wuliupf,y=log_sale)) +
  geom_boxplot(fill='darkseagreen', varwidth = T) +
  xlab('物流评分') + ylab('') + coord_flip() + labs(title = '商品评价对销量的影响')

fuwupf <- vector('numeric',nrow(lgdata3))
fuwupf <- lgdata3$fuwu
fuwupf[fuwupf<=4.6] <- '4.6及以下'
df <- data.frame(fuwupf,log_sale)
df <- df[complete.cases(df),]
p2 <- ggplot(df, aes(x=fuwupf,y=log_sale)) +
  geom_boxplot(fill='darkseagreen', varwidth = T) +
  xlab('服务评分') + ylab('') + coord_flip()

miaoshupf <- vector('numeric',nrow(lgdata3))
miaoshupf <- lgdata3$miaoshu
miaoshupf[miaoshupf<=4.6] <- '4.6及以下'
df <- data.frame(miaoshupf,log_sale)
df <- df[complete.cases(df),]
p3 <- ggplot(df, aes(x=miaoshupf,y=log_sale)) +
  geom_boxplot(fill='darkseagreen', varwidth = T) +
  xlab('描述评分') + ylab('') + coord_flip()

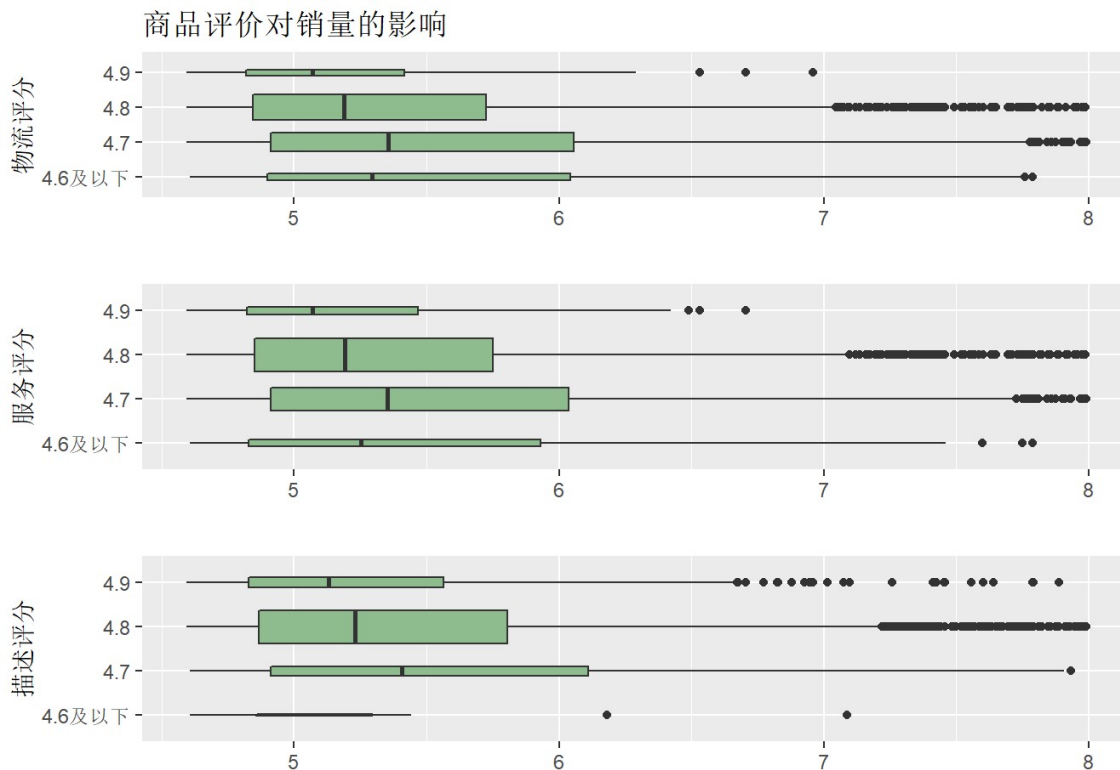
```

```
library(Rmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: plyr
```

```
multiplot(p1, p2, p3)
```



在上图中，箱体宽度代表不同评分的商品数量，因而不难看出，各项评分均较大程度集中在4.7或4.8；评分4.7的商品平均销量最高，而评价低于4.7的商品很难达到月销量100以上。这诠释了顾客在价格和口碑之间的平衡——口碑越好意味着价格越高，而口碑越差意味着服务质量越差——因此，销量最多的应该是口碑处于中值的店家。

3 任务三：回归分析

3.1 简单的线性回归

经过上述的EDA，我们对数据集以及其内部各个变量之间的关系有了较为全面的了解。下面，整合新的数据集，以销量（对数）为相应变量进行线性回归。

```
a <- vector('integer',length(lgdata3$age))
a[lgdata3$age == '17'] <- 0
a[lgdata3$age == '18-24'] <- 1
a[lgdata3$age == '25-29'] <- 2
a[lgdata3$age == '30-34'] <- 3
a[lgdata3$age == '35-39'] <- 4
a[lgdata3$age == '40-49'] <- 5
lgdata6 <- data.frame(sales = log_sale, price = log_price,
                      是否韩版 = lgdata3$hanban, 是否新款 = lgdata3$xinkuan,
                      是否显瘦 = lgdata3$xianshou, 是否中长款 = lgdata3$zckuan,
                      是否修身 = lgdata3$xiushen, 是否印花 = yinhua,
                      old = a)
lm <- lm(sales ~ . - sales, data = lgdata6)
summary(lm)
```



```
##
## Call:
## lm(formula = sales ~ . - sales, data = lgdata6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2841 -0.5403 -0.1837  0.3714  2.6419
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.46614    0.09946   65.013 < 2e-16 ***
## price        -0.23052    0.02029  -11.363 < 2e-16 ***
## 是否韩版      0.06725    0.02479    2.713  0.00669 **
## 是否新款      0.03298    0.02135    1.545  0.12251
## 是否显瘦      0.10822    0.02339    4.627 3.80e-06 ***
## 是否中长款    0.11789    0.02221    5.307 1.16e-07 ***
## 是否修身      0.05165    0.02165    2.385  0.01711 *
## 是否印花      0.04429    0.02414    1.835  0.06657 .
## old           0.01638    0.01434    1.142  0.25359
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7158 on 5184 degrees of freedom
## Multiple R-squared:  0.0573, Adjusted R-squared:  0.05585
## F-statistic: 39.39 on 8 and 5184 DF,  p-value: < 2.2e-16
```

根据回归初步分析，我们可以得到以下结论：

- （1）价格对销量有负向影响，这符合先前的观测以及经济学规律；由于销量和价格均基于对数变换，对应回归系数体现了销量对价格的弹性，即价格增长1%，连衣裙月销量相应减少约0.23%；
- （2）“韩版”、“新款”、“显瘦”、“中长款”、“修身”均对连衣裙销量有正向影响，其中“中长款”、“显瘦”对连衣裙销量影响最为明显，“中长款”特质能有效提高月销量约11.79%；“显瘦”特质提高月销量10.82%，对应回归系数体现销量的增长率；
- （3）是否带有“印花”标签对销量影响并不显著；
- （4）年龄对销量影响不显著。

3.2 考虑交互项

基于任务二中EDA的结果，不同年龄层可能通过商品的标签间接影响到连衣裙的月销量，因此我们考虑加入年龄和各项标签的交互项后，再进行一次回归分析：

```
lm1 <- lm(sales ~ . + 是否韩版 * old + 是否中长款 *
          old + 是否印花 * old + 是否修身 * old + 是否显瘦 *
          old + 是否新款 * old, data = lgdata6)
summary(lm1)
```

```
##
## Call:
## lm(formula = sales ~ . + 是否韩版 * old + 是否中长款 * old +
##     是否印花 * old + 是否修身 * old + 是否显瘦 * old + 是否新款 *
##     old, data = lgdata6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2821 -0.5399 -0.1837  0.3745  2.6196
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.505218   0.111682  58.248 < 2e-16 ***
## price         -0.231930   0.020364 -11.389 < 2e-16 ***
## 是否韩版       -0.135922   0.067249  -2.021  0.04331 *
## 是否新款        0.003685   0.057297   0.064  0.94873
## 是否显瘦        0.274014   0.063029   4.347  1.4e-05 ***
## 是否中长款     0.165521   0.058209   2.844  0.00448 **
## 是否修身        0.004218   0.058908   0.072  0.94292
## 是否印花       -0.020747   0.062312  -0.333  0.73919
## old            0.001175   0.026961   0.044  0.96524
## 是否韩版:old    0.113081   0.035079   3.224  0.00127 **
## 是否中长款:old -0.024243   0.025851  -0.938  0.34840
## 是否印花:old    0.030485   0.026754   1.139  0.25456
## 是否修身:old    0.021996   0.026380   0.834  0.40441
## 是否显瘦:old   -0.080262   0.028479  -2.818  0.00485 **
## 是否新款:old    0.015836   0.026362   0.601  0.54806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7148 on 5178 degrees of freedom
## Multiple R-squared:  0.06105,    Adjusted R-squared:  0.05852
## F-statistic: 24.05 on 14 and 5178 DF,  p-value: < 2.2e-16
```

不难发现，回归结果与EDA分析时一致：

- （1）价格对月销量的影响与未加交互项时基本一致；
- （2）消费年龄层对销量无明显影响；
- （3）“显瘦”“中长款”对销量的正向影响最为显著。“显瘦”能明显提高销量，并且该效应会根据年龄层不同而有所改变；
- （4）随着消费者年龄的增加，“中长款”“显瘦”对销售量的促进作用减弱；
- （5）“韩版”对销量的积极影响随着消费者年龄的增加更为明显。

3.3 回归交互项系数的进一步解读

最后，我们希望能够进一步解读交互作用。为此，计算“是否韩版”以及“是否显瘦”（其他款式与年龄的交互项统计不显著）在每个年龄层的销量比值：

```
# 交互作用
ratio_han <- aggregate(hanban, by=list(lgdata3$age),
                        function(x) mean(lgdata3$s[x==1])/mean(lgdata3$s[x==0]))
ratio_shou <- aggregate(xianshou, by=list(lgdata3$age),
                        function(x) mean(lgdata3$s[x==1])/mean(lgdata3$s[x==0]))
colnames(ratio_han) <- c('年龄', '比值')
colnames(ratio_shou) <- c('年龄', '比值')

# -----
# 是否韩版
# 总比值
mean(lgdata3$s[hanban==1])/mean(lgdata3$s[hanban==0])
```

```
## [1] 1.310265
```

```
# 按年龄分组
ratio_han
```

```
##      年龄      比值
## 1      17 1.0016416
## 2 18-24 1.0084401
## 3 25-29 1.2453944
## 4 30-34 1.0885568
## 5 35-39 1.0464534
## 6 40-49 0.9890088
```

```
# -----
# 是否显瘦
# 总比值
mean(lgdata3$s[xianshou==1])/mean(lgdata3$s[xianshou==0])
```

```
## [1] 1.313418
```

```
# 按年龄分组
ratio_shou
```

```
##      年龄      比值
## 1      17 1.0075483
## 2 18-24 1.1296257
## 3 25-29 1.1967537
## 4 30-34 1.0326210
## 5 35-39 1.0240413
## 6 40-49 0.9843862
```

几乎与2.4节中的分析一致，首先，总体而言，“显瘦”（“韩版”）特质对其他年龄层连衣裙销量均有明显促进作用（>1）；按分组来看，只有40-49周岁年龄层比值小于1，即“显瘦”（“韩版”）特质对连衣裙销量有抑制作用。该标签对25-29周岁人群正向影响最为显著，这或许是由于，相比于更高年龄段人群，该年龄段人群更加青睐“显瘦”“韩版”，而相比于较低年龄段人群，他们又有更强的财务支配自由和消费能力。

4 总结

通过上述分析，我们对线上女装销售的概况，及其人群划分的特征有了较为全面的了解。基于此，我们可以对线上女装的销售提出建议及策略。首先，商家应巧妙设计产品名称，使其包含更多的“热词”（例如“新款”），从而增加

产品被搜索、点击的概率。此外，商家应当针对不同年龄段人群设计不同的营销策略，精准投放合适产品。例如，对于中年顾客，可以着重推销“韩版”服装；对于20岁至30岁左右的顾客，可以着重推销“显瘦”和“中长款”服装；而对于20岁以下的年轻顾客，推销“中长款”服装显然并不明智，但是“韩版”、“显瘦”等关键词必能快速吸引他们。最后，对于店家本身而言，经营时间越长并不能占据优势，处于东南沿海江浙沪才是王道；此外，对于评价不高的店家，应主攻薄利多销，而对于评价满分的店家，则可能遭受销量较低的窘迫局面。