



湖南大学
HUNAN UNIVERSITY

统计仿真实验 2

实验报告

广义线性回归
双十一销售数据分析

张笑竹
统计 1601
201618070114

1. 引言

1.1 背景介绍

双十一购物狂欢节源起于 2009 年，阿里巴巴集团联合淘宝商城的 27 家品牌，进行打折促销，首次就取得了 5000 万的佳绩。自此之后双十一成为每年的固定购物盛大节日，交易额也一路攀升。2012 年，京东、当当、苏宁等加入鏖战；今年，经官方统计，交易额更是达到了 1207 亿的好成绩。那么，双十一销量与哪些因素有关？商家应选择怎样的促销策略？

1.2 数据集

为此，共收集了某线上电商平台 5 个品类相关的 611 个产品的销售数据。每一列分别对应变量的商品名称（name）、关键词（keyword）、品类（category）、原价（price_origin）、现价（price_sale）、折扣信息（discount）、双十一销量（dif）等。通过对该数据集的简要分析，我们可以得出一些较有意义的结论。代码见附录 Markdown 文件。

2. 任务分析

2.1 任务一

在任务一中，我们读入数据，并筛选出了最大销量对应的商品名称（name）、现价（price_sale）和销售量（dif）。根据程序输出结果，2016 秋季新款马丁女靴的销量最好，其销售量在 783 件，对应的单价为 127.43 元/件。可以认为，对于一双马丁靴而言，127 元属于较低的价格，而正式由于低廉的价格吸引了大批的顾客。

此外，为了双十一的销售基本情况，我们绘制出了对数销售量的分布直方图。

双十一期间销量分布直方图

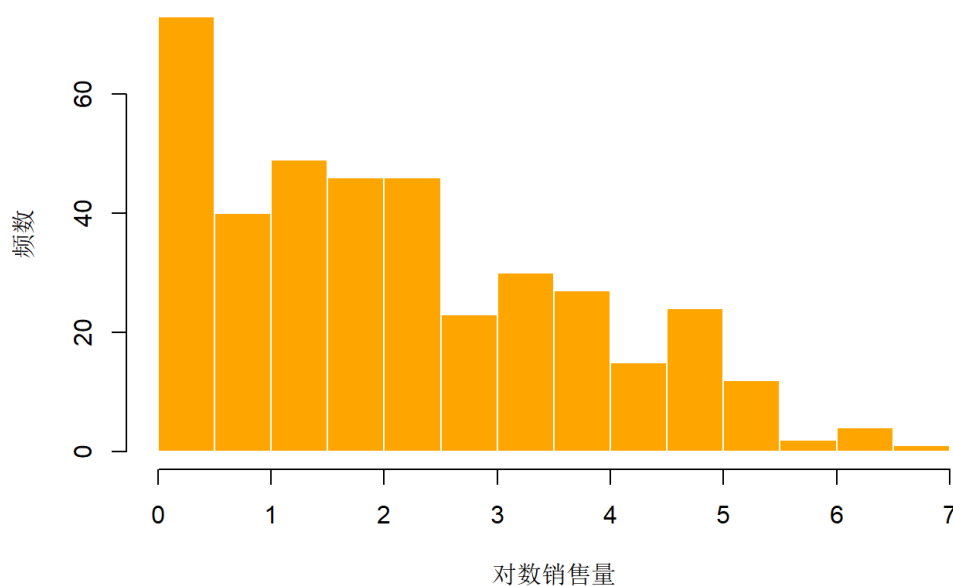


图 1 双十一期间销量分布直方图

通过图 1，不难发现销售量呈尖峰厚尾式的对数正态分布。大部分商品的销售量集中在 10 件以内，可见“双十一”的促销收效并不显著。

2.2 任务二

在任务二中，我们对双十一期间商品类型（category）进行分析。该数据集将商品类型分为“女鞋”、“女装”、“男士饰品”、“男装”以及“女士配饰”5 个类别。对每一个类别的平均销量进行统计，可以得到图 2 所示的条形图。

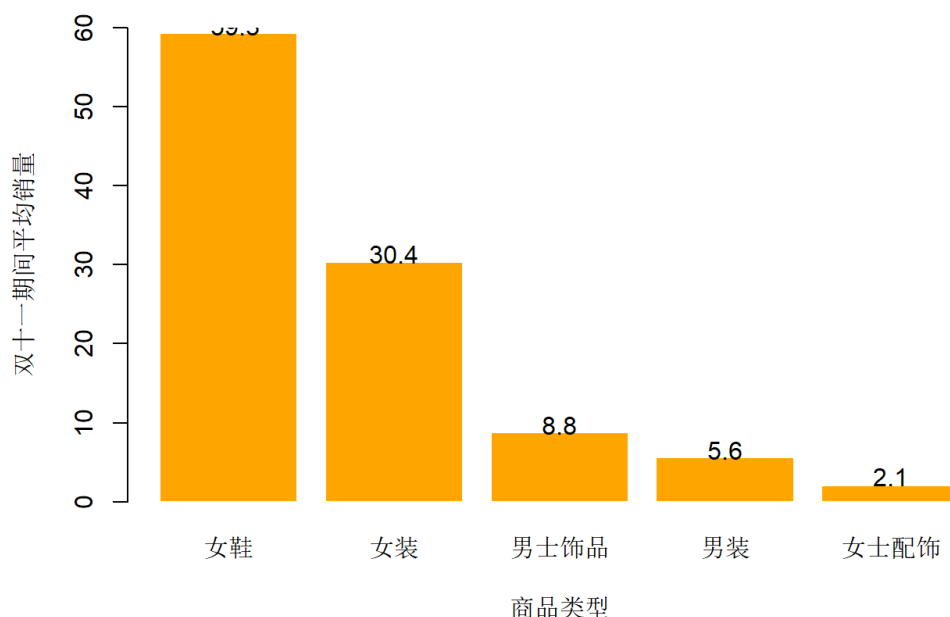


图 2 各个商品类型的双十一期间平均销量

根据图 2，女士用品如“女鞋”、“女装”等销量，要远远高于男士用品如“男士饰品”、“男装”等；而“女士配饰”销量较低的原因，是其需求量相对较少且价格往往较高。可见，女性在双十一中的消费要远远高于男性，而这也符合传统的女性印象。

2.3 任务三

在任务三中，我们需要着重分析“双十一红包”、“满减券”和“其他优惠”等优惠活动的效果。

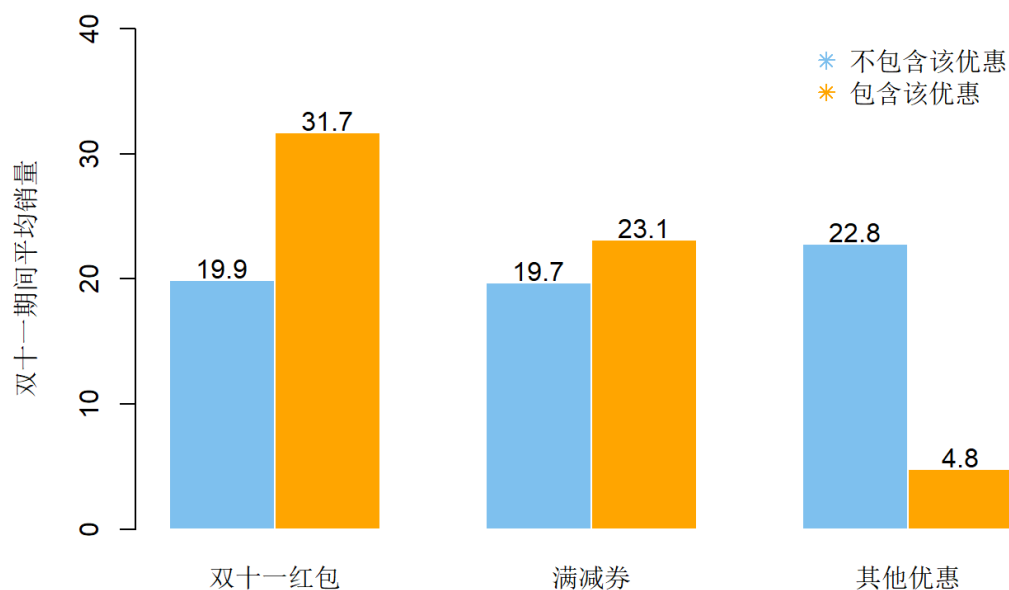


图 3 各种优惠的效果

在图 3 中，我们按照“不包含该优惠”和“包含该优惠”两种情况分别展示了“双十一红包”、“满减券”以及“其他优惠”对双十一期间平均销量的影响。对于“双十一红包”和“满减券”两种方式，包含优惠明显比不包含优惠更能够促进销量；而且“双十一红包”的效果更好。然而，对于“其他优惠”而言，不包含优惠的情况反而要好过包含优惠的情况——这或许由于相比于“双十一红包”和“满减券”，“其他优惠”几乎完全不具有吸引力。

2.4 任务四

在任务四中，我们针对“女装”这一类特定商品，通过提取热词的方式研究市场的潮流和促销的方式。

根据程序输出结果，“新款”、“外套”、“韩版”、“2016”、“秋冬”、“长袖”、“中长款”、“打底”以及“修身”占据了热词的 Top9，反映出了双十一期间女装的潮流。“外套”“秋冬”“长袖”“2016”和“中长款”与时令季节相关，自然称为搜索的火热关键词；而“韩版”“打底”“修身”则反映出了女装的潮流；最后，“新款”则是营销的重要手段，带有该关键词的商品更容易称为搜索的热点。

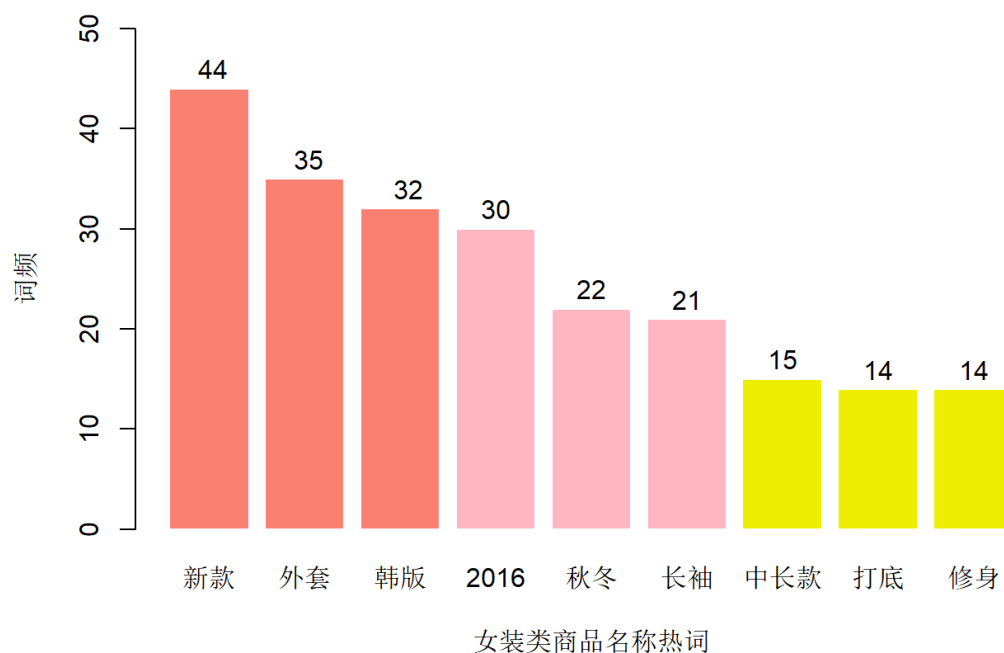


图 4 女装类商品热词词频分析

2.5 任务五

最后，在任务五中，我们着重探索更容易使得消费者产生购买的因素。通过分析，不难发现样本中约 35.8% 的商品在双十一没有销量，因而考虑建立 0-1 逻辑回归因变量：

$$Y = \begin{cases} 0, & \text{双十一销量大于 0} \\ 1, & \text{双十一销量为 0} \end{cases}$$

并将其余变量作为自变量，进行 Logistics 回归分析。经过 AIC（赤池信息准则）筛选，最后保留“category”，“price_cut”，“price_sale”，“service_score”，“包邮”，“韩版”，“冬装”，“学院”，“情侣”以及“dis_strength1”为自变量，建立最终的 Logistics 回归模型。在显著性为 0.05 的水平下，“category 女装”和“韩版”两个变量并不显著；在其余显著的自变量中，“category 男装”(+）、“category 女鞋”(+）、“price_cut” (-)， “price_sale” (-)， “service_score” (+) 都极为显著，p-value 接近于 0。可见，当顾客在双十一“剁手”时，男装、女鞋、价格和快递服务是最为关键的因素。

最后，通过观察图 5，我们可以对该二分类 Logistics 回归的精确度有一定的了解。由于图中的 ROC 接近右上角，几乎可以认为，该分类器的精度较高。

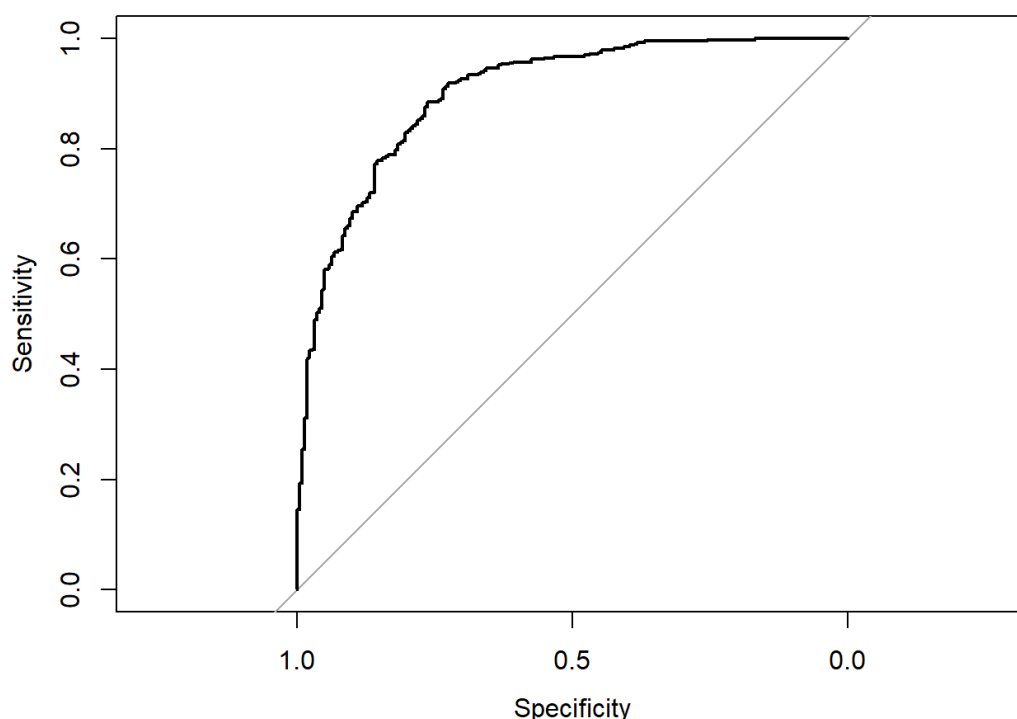


图 5 ROC 图

3 结论

通过上述 5 个任务，我们不难发现，在双十一期间，尽管总成交量很大，对于绝大多数商家，个体成交量却并不高，因此掌握一定的市场规律和营销手段是很有必要的。事实上，女性顾客比男性顾客具有更强的购买力，因此可以对女性顾客精准投放广告，增加商品对她们的吸引力。此外，“双十一红包”和“满减券”是效果显著的促销方式。对于店家而言，商品名称和标题的名称同样关键，“新款”、“外套”、“韩版”、“2016”无疑会带来更多的点击量。最后，增加男装、女鞋的商品种类，尽量控制价格，基于更高的折扣和优惠，提供更优质的快递服务（包邮），是双十一市场制胜的关键。

广义线性回归 - 双十一销售数据分析

张笑竹 / 201618070114

2019年7月2日

双十一购物狂欢节源起于2009年，阿里巴巴集团联合淘宝商城的27家品牌，进行打折促销，首次就取得了5000万的佳绩。自此之后双十一成为每年的固定购物盛大节日，交易额也一路攀升。在本次案例中，使用的数据集收集了2016年某电商平台上611个产品的双十一销售数据，每一列分别对应变量商品名称（name）、关键词（keyword）、品类（category）、原价（price_origin）、现价（price_sale）、折扣信息（discount）、双十一销量（dif）等。

下面，给出5个任务对应的实现代码。

1 任务一

读入数据 double11_addvar.csv，并命名为 datDesc。

```
setwd('C:\\Users\\小竹子\\Desktop\\统计仿真实验\\案例难度：__\\1. 双十一销售数据分析')
datDesc <- read.csv('double11_addvar.csv')
```

为了解双十一的销售基本情况，查看最大销量对应的商品名称（name）、现价（price_sale）和销售量（dif）。

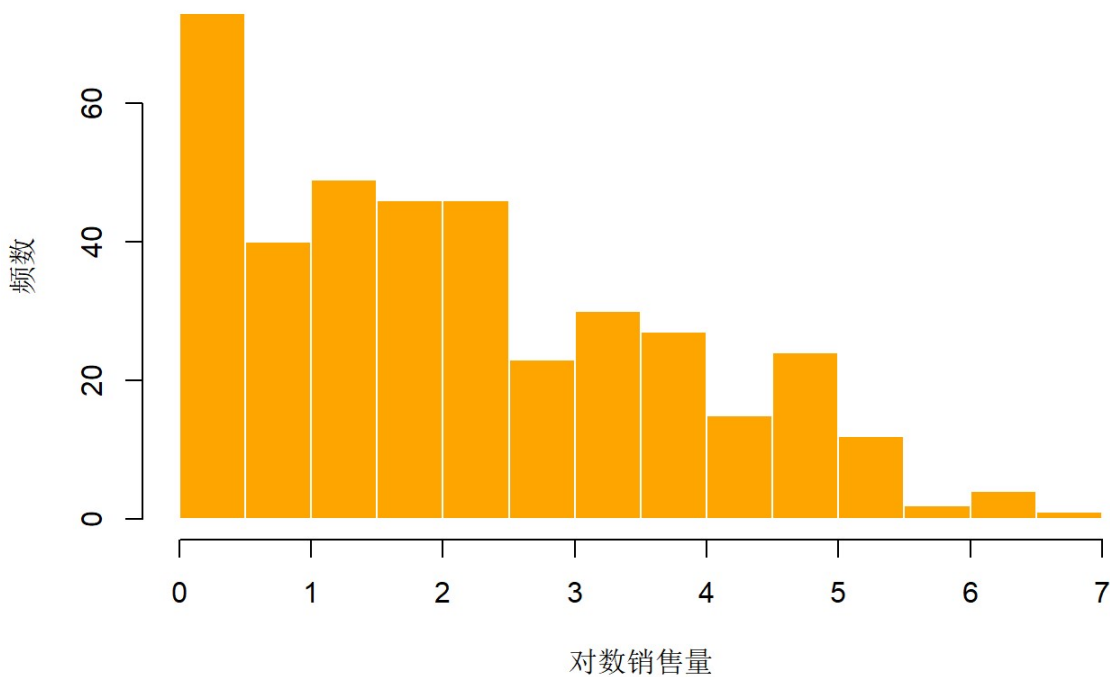
```
datDesc[datDesc$dif == max(datDesc$dif), c('name', 'price_sale', 'dif')]
```

```
##
## 94 2016秋季新款女靴子马丁靴短靴子英伦风粗跟及裸靴女鞋子单靴    name price_sale dif
## 127.4257 783
```

绘制对数销售量的分布直方图。

```
hist(log(datDesc$dif), col = 'orange', border = 'white', xlab = '对数销售量', ylab = '频数',
     main = '双十一期间销量分布直方图')
```

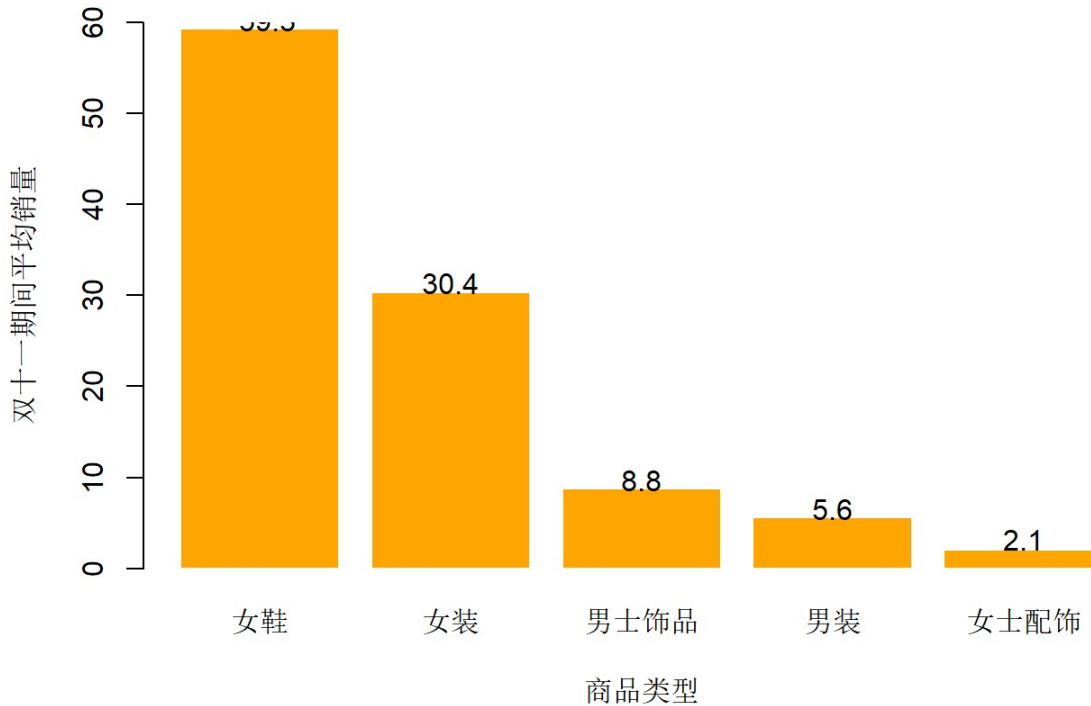
双十一期间销量分布直方图



2 任务二

对双十一期间商品类型 (category) 进行分析。画出各类型商品的平均销量条形图，尝试分析商品类型和销量之间的关系。注意，条形图横坐标按照销量大小排序，并在柱子上方添加销量数字标签。

```
s <- aggregate(datDesc$dif, by = list(datDesc$category), FUN = mean)
ss <- s$x
names(ss) <- s$Group.1
ss <- sort(ss, decreasing = T)
barplot(height = ss, col = 'orange', border = 'white', ylim = c(0,60),
        xlab = '商品类型', ylab = '双十一期间平均销量')
text(0.75, 60.3, '59.3')
text(1.9, 31.4, '30.4')
text(3.1, 9.8, '8.8')
text(4.3, 6.6, '5.6')
text(5.5, 3.2, '2.1')
```

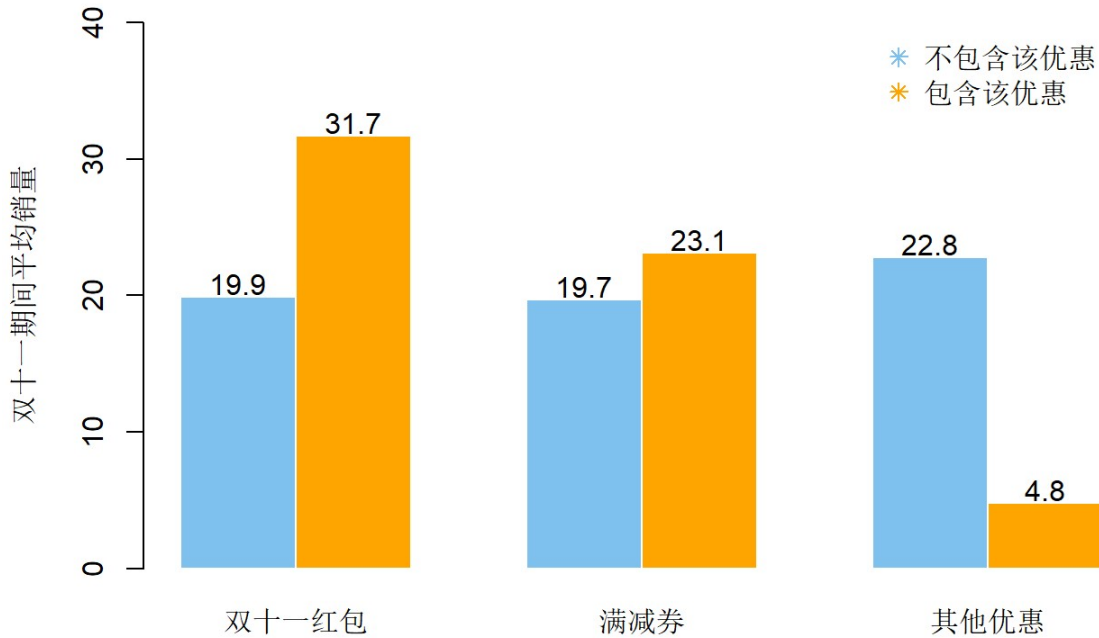



3 任务三

双十一期间优惠活动，包括“双十一红包”、“满减券”、“其他优惠”，每个优惠活动都被处理成了数据的一列。你需要对于包含或不包含某种优惠活动，分别计算对应的销量平均值。请绘制复式条形图，对优惠活动的效果进行简要分析。同样地，你需要添加销量数字标签。

```
s1 <- c(mean(datDesc$dif[datDesc$双十一红包 == 0]),
        mean(datDesc$dif[datDesc$双十一红包 == 1]),
        mean(datDesc$dif[datDesc$满减券 == 0]),
        mean(datDesc$dif[datDesc$满减券 == 1]),
        mean(datDesc$dif[datDesc$其他优惠 == 0]),
        mean(datDesc$dif[datDesc$其他优惠 == 1]))
leibie <- factor(c(rep('双十一红包', 2), rep('满减券', 2), rep('其他优惠', 2)),
                 levels = c('双十一红包', '满减券', '其他优惠'), ordered = T)
youwu <- factor(rep(c('不包含该优惠', '包含该优惠'), 3),
                 levels = c('不包含该优惠', '包含该优惠'), ordered = T)

barplot(s1 ~ youwu + leibie, beside = T, ylim = c(0, 40), col = c('skyblue2', 'orange'), border = F,
        legend = F, xlab = '', ylab = '双十一期间平均销量')
legend(7, 40, legend = c('不包含该优惠', '包含该优惠'), bty = 'n', col = c('skyblue2', 'orange'), pch = 8)
text(1.5, 20.9, '19.9')
text(2.5, 32.7, '31.7')
text(4.5, 20.7, '19.7')
text(5.5, 24.1, '23.1')
text(7.5, 23.8, '22.8')
text(8.5, 5.8, '4.8')
```



4 任务四

对商品名称提取双十一销售热词。你需要借助“jiebaR”程序包，对女装类销量>0的商品的名称进行分词，统计出词频Top10并输出。

```
# 提取女装类
nz <- datDesc[datDesc$category=='女装' & datDesc$dif >0,]

# 进行分词
library(jiebaR)
```

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

```
sg <- segment(paste(nz$name),worker())
fq <- freq(sg)
fq <- fq[order(fq$freq,decreasing = T),]
head(fq,10)
```

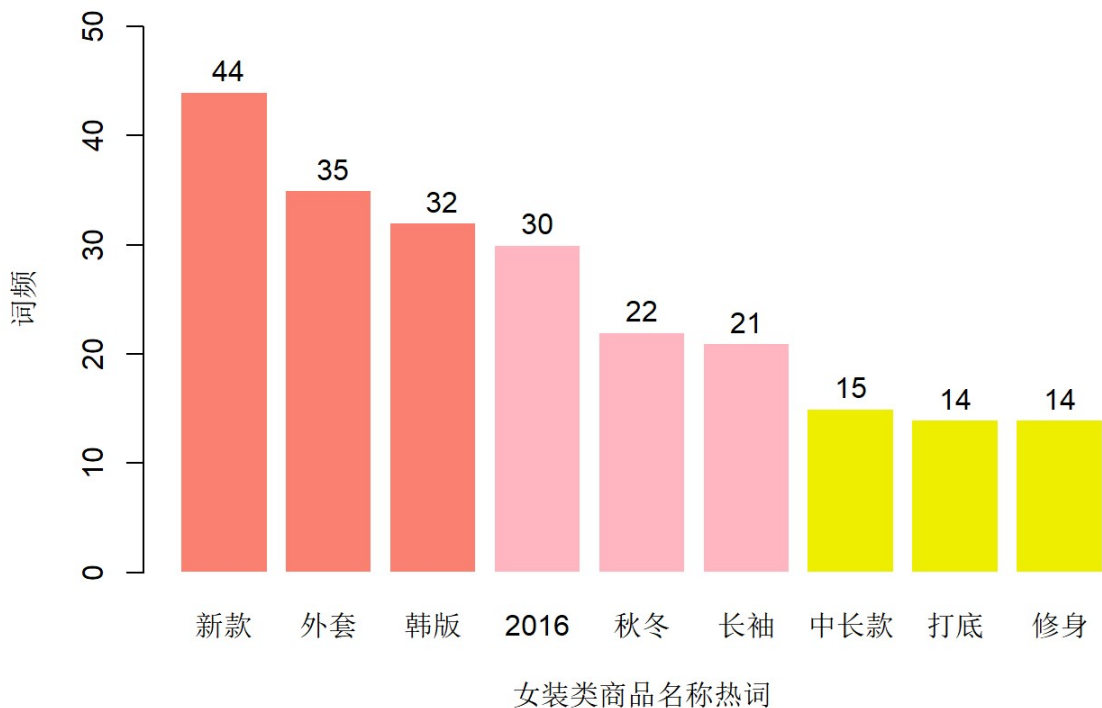
```
##      char freq
## 171  女装    73
## 121  新款    44
## 135  外套    35
## 161  韩版    32
## 215  2016    30
## 170  秋冬    22
## 64   长袖    21
## 213  中长款  15
## 71   打底    14
## 172  修身    14
```

```
nvzhuang_key <- fq$freq[1:10]
names(nvzhuang_key) <- fq$char[1:10]
nvzhuang_key
```

```
##   女装   新款   外套   韩版   2016   秋冬   长袖  中长款   打底   修身
##    73    44    35    32    30    22    21    15    14    14
```

请绘制女装类商品名称词频条形图。注意，由于词汇“女装”在此没有意义，在绘图时将该词忽略。

```
barplot(nvzhuang_key[-1], ylim = c(0,50), xlab='女装类商品名称热词',
        ylab='词频', col=c(rep('salmon',3), rep('lightpink',3),rep('yellow2',3)),
        border=F)
text(0.75,46,'44')
text(1.95,37,'35')
text(3.2,34,'32')
text(4.3,32,'30')
text(5.5,24,'22')
text(6.7,23,'21')
text(7.9,17,'15')
text(9.1,16,'14')
text(10.3,16,'14')
```



5 任务五

哪些因素更容易使得消费者产生购买呢？我们发现样本中约35.8%的商品在双十一没有销量，因而考虑建立0-1逻辑回归：因变量 = 1，代表双十一销量 > 0。首先，你需要生成逻辑回归数据集 datReg，数据集共包含以下变量。其中 Y 为因变量，是根据双十一是否有销量生成的0-1变量；其余变量为自变量，均选自原始数据集 datDesc。

```
Y <- vector('integer',nrow(datDesc))
Y[datDesc$dif > 0] <- 1
attach(datDesc)
datReg <- data.frame(Y, category, inventory, price_cut, price_sale,
                     des_score, quality_score, service_score, 包邮, 双十一,
                     新款, 韩版, 英伦, 欧美, 冬装, 礼盒, 明星, 均码, 通勤,
                     学院, 情侣, 双十一红包, 其他优惠, dis_num1, dis_strength1,
                     easy_satisfy1)
colnames(datReg)
```

```
## [1] "Y"           "category"    "inventory"   "price_cut"
## [5] "price_sale"  "des_score"   "quality_score" "service_score"
## [9] "包邮"       "双十一"     "新款"       "韩版"
## [13] "英伦"      "欧美"       "冬装"       "礼盒"
## [17] "明星"      "均码"       "通勤"       "学院"
## [21] "情侣"      "双十一红包" "其他优惠"   "dis_num1"
## [25] "dis_strength1" "easy_satisfy1"
```

在逻辑回归数据集上，建立逻辑回归模型。

```
ml <- glm(Y ~., family = binomial(), data = datReg)
summary(ml)
```

```
##
## Call:
## glm(formula = Y ~ ., family = binomial(), data = datReg)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6710  -0.3457   0.2724   0.5773   2.8081
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.350e+01  9.463e+00  -3.540 0.000401 ***
## category男装      1.339e+00  5.178e-01   2.586 0.009712 **
## category女士配饰  -1.067e-01  6.486e-01  -0.164 0.869371
## category女鞋      2.717e+00  8.803e-01   3.087 0.002023 **
## category女装      4.615e-01  9.117e-01   0.506 0.612723
## inventory        5.127e-05  3.964e-05   1.294 0.195812
## price_cut       -3.615e-02  1.089e-02  -3.319 0.000904 ***
## price_sale      -7.291e-03  1.002e-03  -7.278 3.39e-13 ***
## des_score       -4.600e+00  8.216e+00  -0.560 0.575586
## quality_score   -5.434e-01  7.887e+00  -0.069 0.945074
## service_score    1.244e+01  5.722e+00   2.174 0.029715 *
## 包邮            -2.095e+00  1.171e+00  -1.789 0.073577 .
## 双十一           9.355e+00  5.354e+02   0.017 0.986060
## 新款             1.926e-01  2.982e-01   0.646 0.518445
## 韩版             4.016e-01  2.707e-01   1.483 0.137958
## 英伦            -1.621e-01  5.222e-01  -0.310 0.756229
## 欧美            -7.223e-02  3.428e-01  -0.211 0.833136
## 冬装             8.400e-01  3.085e-01   2.723 0.006468 **
## 礼盒             7.443e-01  4.942e-01   1.506 0.132046
## 明星            -3.692e-02  1.011e+00  -0.037 0.970874
## 均码             2.673e-01  4.252e-01   0.628 0.529687
## 通勤            -4.218e-01  5.063e-01  -0.833 0.404777
## 学院             9.133e-01  4.301e-01   2.124 0.033694 *
## 情侣            -1.129e+00  6.364e-01  -1.774 0.076124 .
## 双十一红包      -6.178e-01  1.196e+00  -0.517 0.605484
## 其他优惠         4.193e-01  5.412e-01   0.775 0.438500
## dis_num12种优惠   1.026e+00  7.744e-01   1.325 0.185046
## dis_num1不超过1种优惠 1.262e-01  7.417e-01   0.170 0.864887
## dis_strength1    -5.549e-01  4.202e-01  -1.320 0.186671
## easy_satisfy1(1,2] -5.504e-01  4.877e-01  -1.129 0.259072
## easy_satisfy1(2,8] -6.784e-01  1.008e+00  -0.673 0.501031
## easy_satisfy10    -3.417e-01  8.220e-01  -0.416 0.677597
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 797.37  on 610  degrees of freedom
## Residual deviance: 441.44  on 579  degrees of freedom
## AIC: 505.44
##
## Number of Fisher Scoring iterations: 12
```

随后用AIC准则进行逐步回归选择变量。

```
mstep <- step(m1)
```

```
## Start:  AIC=505.44
## Y ~ category + inventory + price_cut + price_sale + des_score +
##     quality_score + service_score + 包邮 + 双十一 + 新款 + 韩版 +
##     英伦 + 欧美 + 冬装 + 礼盒 + 明星 + 均码 + 通勤 + 学院 + 情侣 +
##     双十一红包 + 其他优惠 + dis_num1 + dis_strength1 + easy_satisfy1
##
##           Df Deviance    AIC
## - easy_satisfy1  3    442.88 500.88
## - dis_num1      2    443.38 503.38
## - 明星          1    441.44 503.44
## - quality_score  1    441.45 503.45
## - 双十一        1    441.47 503.47
## - 欧美          1    441.49 503.49
## - 英伦          1    441.54 503.54
## - 双十一红包    1    441.71 503.71
## - des_score     1    441.76 503.76
## - 均码          1    441.84 503.84
## - 新款          1    441.86 503.86
## - 其他优惠      1    442.05 504.05
## - 通勤          1    442.13 504.13
## - dis_strength1 1    443.23 505.23
## <none>          441.44 505.44
## - inventory     1    443.45 505.45
## - 韩版          1    443.64 505.64
## - 礼盒          1    443.73 505.73
## - 包邮          1    444.45 506.45
## - 情侣          1    444.79 506.79
## - service_score 1    446.16 508.16
## - 学院          1    446.28 508.28
## - 冬装          1    448.98 510.98
## - category      4    457.68 513.68
## - price_cut     1    453.70 515.70
## - price_sale    1    520.88 582.88
##
## Step:  AIC=500.88
## Y ~ category + inventory + price_cut + price_sale + des_score +
##     quality_score + service_score + 包邮 + 双十一 + 新款 + 韩版 +
##     英伦 + 欧美 + 冬装 + 礼盒 + 明星 + 均码 + 通勤 + 学院 + 情侣 +
##     双十一红包 + 其他优惠 + dis_num1 + dis_strength1
##
##           Df Deviance    AIC
## - dis_num1      2    444.16 498.16
## - 明星          1    442.88 498.88
## - 欧美          1    442.88 498.88
## - quality_score  1    442.88 498.88
## - 双十一        1    442.90 498.90
## - 英伦          1    442.95 498.95
## - 双十一红包    1    443.12 499.12
## - des_score     1    443.19 499.19
## - 均码          1    443.24 499.24
## - 其他优惠      1    443.30 499.30
## - 新款          1    443.44 499.44
## - 通勤          1    443.87 499.87
## - inventory     1    444.60 500.60
## <none>          442.88 500.88
## - dis_strength1 1    444.92 500.92
```

```
## - 韩版          1    444.94 500.94
## - 礼盒          1    445.24 501.24
## - 包邮          1    445.83 501.83
## - 情侣          1    447.09 503.09
## - 学院          1    447.16 503.16
## - service_score 1    447.59 503.59
## - 冬装          1    449.83 505.83
## - price_cut     1    454.80 510.80
## - category      4    468.66 518.66
## - price_sale    1    526.38 582.38
##
## Step:  AIC=498.16
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       quality_score + service_score + 包邮 + 双十一 + 新款 + 韩版 +
##       英伦 + 欧美 + 冬装 + 礼盒 + 明星 + 均码 + 通勤 + 学院 + 情侣 +
##       双十一红包 + 其他优惠 + dis_strength1
##
##              Df Deviance    AIC
## - 欧美          1    444.17 496.17
## - quality_score 1    444.17 496.17
## - 双十一        1    444.19 496.19
## - 明星          1    444.21 496.21
## - 英伦          1    444.31 496.31
## - 双十一红包    1    444.43 496.43
## - 其他优惠      1    444.48 496.48
## - des_score     1    444.53 496.53
## - 均码          1    444.69 496.69
## - 新款          1    445.04 497.04
## - 通勤          1    445.36 497.36
## - dis_strength1 1    446.15 498.15
## <none>          444.16 498.16
## - inventory     1    446.17 498.17
## - 礼盒          1    446.48 498.48
## - 韩版          1    446.52 498.52
## - 包邮          1    447.24 499.24
## - 情侣          1    447.38 499.38
## - 学院          1    448.24 500.24
## - service_score 1    449.02 501.02
## - 冬装          1    452.30 504.30
## - price_cut     1    456.23 508.23
## - category      4    472.89 518.89
## - price_sale    1    532.87 584.87
##
## Step:  AIC=496.17
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       quality_score + service_score + 包邮 + 双十一 + 新款 + 韩版 +
##       英伦 + 冬装 + 礼盒 + 明星 + 均码 + 通勤 + 学院 + 情侣 + 双十一红包 +
##       其他优惠 + dis_strength1
##
##              Df Deviance    AIC
## - quality_score 1    444.17 494.17
## - 双十一        1    444.19 494.19
## - 明星          1    444.21 494.21
## - 英伦          1    444.33 494.33
## - 双十一红包    1    444.43 494.43
## - 其他优惠      1    444.48 494.48
```



```
## - des_score      1    444.53 494.53
## - 均码           1    444.69 494.69
## - 新款           1    445.04 495.04
## - 通勤           1    445.36 495.36
## - dis_strength1  1    446.16 496.16
## <none>           444.17 496.17
## - inventory      1    446.17 496.17
## - 礼盒           1    446.49 496.49
## - 韩版           1    446.53 496.53
## - 包邮           1    447.24 497.24
## - 情侣           1    447.38 497.38
## - 学院           1    448.26 498.26
## - service_score  1    449.03 499.03
## - 冬装           1    452.37 502.37
## - price_cut      1    456.29 506.29
## - category       4    473.29 517.29
## - price_sale     1    533.17 583.17
##
## Step:  AIC=494.17
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 双十一 + 新款 + 韩版 + 英伦 + 冬装 +
##       礼盒 + 明星 + 均码 + 通勤 + 学院 + 情侣 + 双十一红包 + 其他优惠 +
##       dis_strength1
##
##              Df Deviance    AIC
## - 双十一      1    444.20 492.20
## - 明星         1    444.22 492.22
## - 英伦         1    444.34 492.34
## - 双十一红包  1    444.43 492.43
## - 其他优惠     1    444.49 492.49
## - 均码         1    444.71 492.71
## - 新款         1    445.04 493.04
## - des_score    1    445.18 493.18
## - 通勤         1    445.36 493.36
## <none>         444.17 494.17
## - inventory    1    446.17 494.17
## - dis_strength1 1    446.19 494.19
## - 礼盒         1    446.49 494.49
## - 韩版         1    446.53 494.53
## - 包邮         1    447.25 495.25
## - 情侣         1    447.42 495.42
## - 学院         1    448.26 496.26
## - service_score 1    449.67 497.67
## - 冬装         1    452.38 500.38
## - price_cut    1    456.35 504.35
## - category     4    473.52 515.52
## - price_sale   1    533.35 581.35
##
## Step:  AIC=492.2
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 新款 + 韩版 + 英伦 + 冬装 + 礼盒 +
##       明星 + 均码 + 通勤 + 学院 + 情侣 + 双十一红包 + 其他优惠 +
##       dis_strength1
##
##              Df Deviance    AIC
## - 明星         1    444.25 490.25
```

```
## - 英伦          1    444.36 490.36
## - 双十一红包    1    444.46 490.46
## - 其他优惠      1    444.52 490.52
## - 均码          1    444.74 490.74
## - 新款          1    445.07 491.07
## - des_score     1    445.21 491.21
## - 通勤          1    445.39 491.39
## <none>          444.20 492.20
## - inventory     1    446.20 492.20
## - dis_strength1 1    446.23 492.23
## - 礼盒          1    446.51 492.51
## - 韩版          1    446.57 492.57
## - 包邮          1    447.29 493.29
## - 情侣          1    447.45 493.45
## - 学院          1    448.29 494.29
## - service_score 1    449.69 495.69
## - 冬装          1    452.40 498.40
## - price_cut     1    456.38 502.38
## - category      4    473.58 513.58
## - price_sale    1    533.44 579.44
##
## Step:  AIC=490.25
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 新款 + 韩版 + 英伦 + 冬装 + 礼盒 +
##       均码 + 通勤 + 学院 + 情侣 + 双十一红包 + 其他优惠 + dis_strength1
##
##              Df Deviance    AIC
## - 英伦          1    444.41 488.41
## - 双十一红包    1    444.51 488.51
## - 其他优惠      1    444.57 488.57
## - 均码          1    444.76 488.76
## - 新款          1    445.15 489.15
## - des_score     1    445.23 489.23
## - 通勤          1    445.42 489.42
## <none>          444.25 490.25
## - inventory     1    446.25 490.25
## - dis_strength1 1    446.29 490.29
## - 韩版          1    446.57 490.57
## - 礼盒          1    446.61 490.61
## - 包邮          1    447.32 491.32
## - 情侣          1    447.71 491.71
## - 学院          1    448.32 492.32
## - service_score 1    449.69 493.69
## - 冬装          1    452.47 496.47
## - price_cut     1    456.41 500.41
## - category      4    473.59 511.59
## - price_sale    1    533.58 577.58
##
## Step:  AIC=488.41
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 新款 + 韩版 + 冬装 + 礼盒 + 均码 +
##       通勤 + 学院 + 情侣 + 双十一红包 + 其他优惠 + dis_strength1
##
##              Df Deviance    AIC
## - 双十一红包    1    444.68 486.68
## - 其他优惠      1    444.71 486.71
```

```
## - 均码          1    444.92 486.92
## - 新款          1    445.32 487.32
## - des_score     1    445.43 487.43
## - 通勤          1    445.64 487.64
## - inventory     1    446.33 488.33
## <none>          1    444.41 488.41
## - dis_strength1 1    446.54 488.54
## - 韩版          1    446.66 488.66
## - 礼盒          1    446.72 488.72
## - 包邮          1    447.49 489.49
## - 情侣          1    447.81 489.81
## - 学院          1    448.46 490.46
## - service_score 1    449.99 491.99
## - 冬装          1    452.48 494.48
## - price_cut     1    456.51 498.51
## - category      4    474.24 510.24
## - price_sale    1    533.58 575.58
##
## Step:  AIC=486.68
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 新款 + 韩版 + 冬装 + 礼盒 + 均码 +
##       通勤 + 学院 + 情侣 + 其他优惠 + dis_strength1
##
##              Df Deviance    AIC
## - 其他优惠     1    444.98 484.98
## - 均码         1    445.24 485.24
## - 新款         1    445.59 485.59
## - des_score    1    445.68 485.68
## - 通勤         1    445.87 485.87
## - inventory    1    446.65 486.65
## <none>         1    444.68 486.68
## - dis_strength1 1    446.81 486.81
## - 礼盒         1    446.94 486.94
## - 韩版         1    447.20 487.20
## - 包邮         1    447.85 487.85
## - 情侣         1    448.16 488.16
## - 学院         1    448.69 488.69
## - service_score 1    450.31 490.31
## - 冬装         1    452.55 492.55
## - price_cut    1    457.75 497.75
## - category     4    474.93 508.93
## - price_sale   1    543.16 583.16
##
## Step:  AIC=484.98
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 新款 + 韩版 + 冬装 + 礼盒 + 均码 +
##       通勤 + 学院 + 情侣 + dis_strength1
##
##              Df Deviance    AIC
## - 均码         1    445.49 483.49
## - 新款         1    445.83 483.83
## - des_score    1    446.10 484.10
## - 通勤         1    446.17 484.17
## - inventory    1    446.77 484.77
## <none>         1    444.98 484.98
## - dis_strength1 1    447.15 485.15
```

```
## - 礼盒          1    447.15 485.15
## - 韩版          1    447.40 485.40
## - 包邮          1    448.00 486.00
## - 情侣          1    448.38 486.38
## - 学院          1    448.85 486.85
## - service_score 1    450.88 488.88
## - 冬装          1    452.71 490.71
## - price_cut     1    458.13 496.13
## - category      4    477.04 509.04
## - price_sale    1    546.59 584.59
##
## Step:  AIC=483.49
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 新款 + 韩版 + 冬装 + 礼盒 + 通勤 +
##       学院 + 情侣 + dis_strength1
##
##              Df Deviance    AIC
## - 新款          1    446.36 482.36
## - 通勤          1    446.71 482.71
## - des_score     1    447.01 483.01
## - inventory     1    447.08 483.08
## <none>          445.49 483.49
## - 礼盒          1    447.51 483.51
## - 韩版          1    447.80 483.80
## - 包邮          1    448.25 484.25
## - dis_strength1 1    448.26 484.26
## - 情侣          1    448.54 484.54
## - 学院          1    449.63 485.63
## - service_score 1    452.43 488.43
## - 冬装          1    453.53 489.53
## - price_cut     1    458.38 494.38
## - category      4    478.29 508.29
## - price_sale    1    547.89 583.89
##
## Step:  AIC=482.36
## Y ~ category + inventory + price_cut + price_sale + des_score +
##       service_score + 包邮 + 韩版 + 冬装 + 礼盒 + 通勤 + 学院 +
##       情侣 + dis_strength1
##
##              Df Deviance    AIC
## - 通勤          1    447.48 481.48
## - inventory     1    447.79 481.79
## - des_score     1    447.96 481.96
## - 礼盒          1    448.14 482.14
## <none>          446.36 482.36
## - 包邮          1    448.94 482.94
## - 韩版          1    449.02 483.02
## - 情侣          1    449.06 483.06
## - dis_strength1 1    449.51 483.51
## - 学院          1    450.58 484.58
## - service_score 1    453.68 487.68
## - 冬装          1    455.15 489.15
## - price_cut     1    459.82 493.82
## - category      4    479.98 507.98
## - price_sale    1    548.82 582.82
##
```

```
## Step: AIC=481.48
## Y ~ category + inventory + price_cut + price_sale + des_score +
##     service_score + 包邮 + 韩版 + 冬装 + 礼盒 + 学院 + 情侣 +
##     dis_strength1
##
##           Df Deviance    AIC
## - des_score      1    448.83 480.83
## - inventory      1    449.02 481.02
## - 礼盒            1    449.30 481.30
## <none>           447.48 481.48
## - 韩版            1    450.09 482.09
## - 包邮            1    450.10 482.10
## - 情侣            1    450.12 482.12
## - dis_strength1  1    450.88 482.88
## - 学院            1    452.09 484.09
## - service_score  1    454.41 486.41
## - 冬装            1    456.38 488.38
## - price_cut      1    460.14 492.14
## - category       4    483.14 509.14
## - price_sale     1    549.00 581.00
##
## Step: AIC=480.83
## Y ~ category + inventory + price_cut + price_sale + service_score +
##     包邮 + 韩版 + 冬装 + 礼盒 + 学院 + 情侣 + dis_strength1
##
##           Df Deviance    AIC
## - inventory      1    450.10 480.10
## <none>           448.83 480.83
## - 礼盒            1    450.86 480.86
## - 韩版            1    451.46 481.46
## - 情侣            1    451.47 481.47
## - 包邮            1    451.97 481.97
## - dis_strength1  1    452.95 482.95
## - 学院            1    453.35 483.35
## - 冬装            1    457.26 487.26
## - price_cut      1    461.01 491.01
## - service_score  1    467.84 497.84
## - category       4    484.89 508.89
## - price_sale     1    550.16 580.16
##
## Step: AIC=480.1
## Y ~ category + price_cut + price_sale + service_score + 包邮 +
##     韩版 + 冬装 + 礼盒 + 学院 + 情侣 + dis_strength1
##
##           Df Deviance    AIC
## - 礼盒            1    452.01 480.01
## <none>           450.10 480.10
## - 韩版            1    452.54 480.54
## - 情侣            1    452.88 480.88
## - 包邮            1    452.95 480.95
## - dis_strength1  1    454.57 482.57
## - 学院            1    454.67 482.67
## - 冬装            1    458.53 486.53
## - price_cut      1    462.54 490.54
## - service_score  1    469.18 497.18
## - category       4    492.22 514.22
```

```
## - price_sale      1    561.17 589.17
##
## Step:  AIC=480.01
## Y ~ category + price_cut + price_sale + service_score + 包邮 +
##      韩版 + 冬装 + 学院 + 情侣 + dis_strength1
##
##              Df Deviance    AIC
## <none>              452.01 480.01
## - 韩版              1    454.43 480.43
## - 包邮              1    454.67 480.67
## - 学院              1    456.19 482.19
## - 情侣              1    456.39 482.39
## - dis_strength1    1    458.05 484.05
## - 冬装              1    460.50 486.50
## - price_cut        1    464.45 490.45
## - service_score    1    474.57 500.57
## - category         4    492.22 512.22
## - price_sale      1    573.13 599.13
```

以最后一次迭代结果为准，构建新的逻辑回归模型。

```
m2<- glm(formula = Y ~ category + price_cut + price_sale + service_score +
          包邮 + 韩版 + 冬装 + 学院 + 情侣 + dis_strength1, family = binomial(),
          data = datReg)
summary(m2)
```

```
##
## Call:
## glm(formula = Y ~ category + price_cut + price_sale + service_score +
##       包邮 + 韩版 + 冬装 + 学院 + 情侣 + dis_strength1, family = binomial(),
##       data = datReg)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7730  -0.3922   0.2703   0.5845   2.8653
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.721e+01  8.461e+00  -4.398 1.09e-05 ***
## category男装      1.203e+00  3.422e-01   3.517 0.000437 ***
## category女士配饰 -8.087e-01  4.478e-01  -1.806 0.070931 .
## category女鞋      2.273e+00  5.091e-01   4.464 8.03e-06 ***
## category女装      1.413e-02  4.597e-01   0.031 0.975479
## price_cut      -3.356e-02  1.019e-02  -3.294 0.000988 ***
## price_sale     -7.268e-03  8.437e-04  -8.615 < 2e-16 ***
## service_score   8.204e+00  1.806e+00   4.542 5.58e-06 ***
## 包邮           -1.866e+00  1.107e+00  -1.685 0.091916 .
## 韩版            4.015e-01  2.584e-01   1.554 0.120212
## 冬装            8.296e-01  2.874e-01   2.886 0.003898 **
## 学院            8.229e-01  4.168e-01   1.974 0.048359 *
## 情侣           -1.080e+00  5.229e-01  -2.065 0.038905 *
## dis_strength1  -7.085e-01  2.902e-01  -2.442 0.014626 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 797.37  on 610  degrees of freedom
## Residual deviance: 452.01  on 597  degrees of freedom
## AIC: 480.01
##
## Number of Fisher Scoring iterations: 6
```

最后，绘制出拟合ROC曲线图。

```
Y_P <- predict(m2,data = datReg)
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

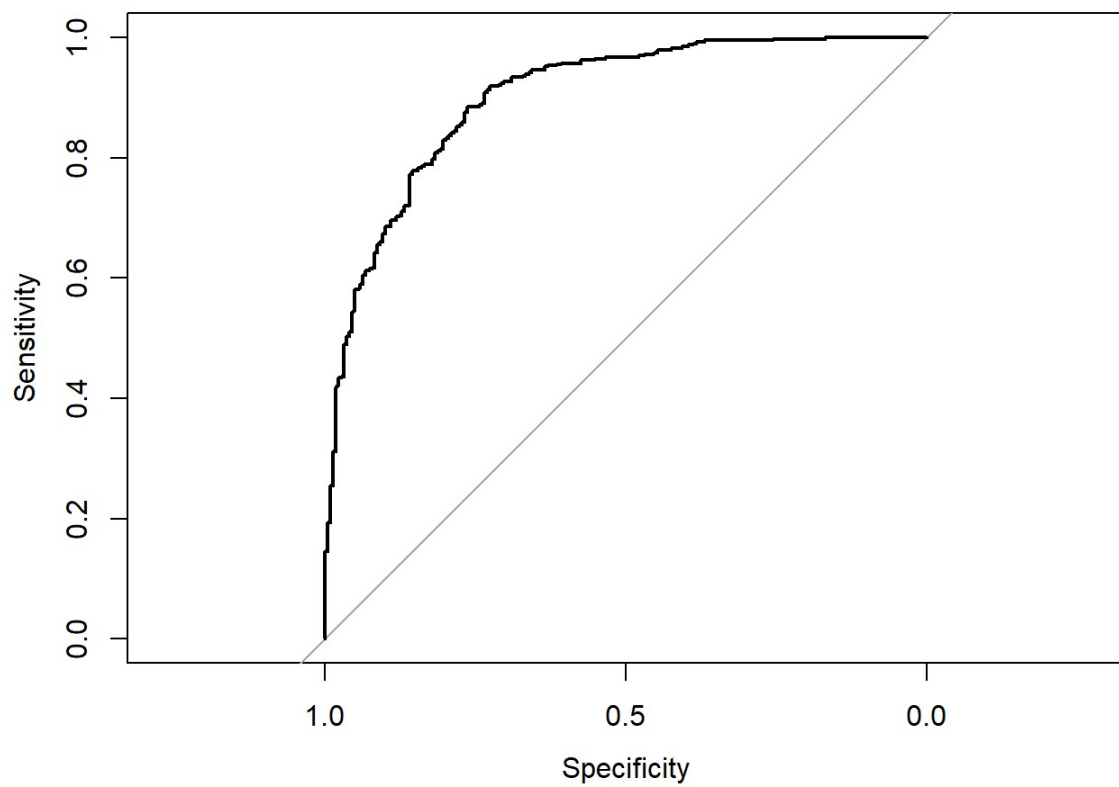
```
## The following objects are masked from 'package:stats':
##
##      cov, smooth, var
```

```
roc_m <- roc(Y,Y_P)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
plot(roc_m)
```



具体分析详见报告。