MEM 第二次作业

夏章印

2024-11-27

注：

- **回答使用中英文皆可**
- **推荐使用 Rmd 或者其他支持 markdown 的书写工具（如免费工具 MarkText，收费 Typora) 答题。**
- **请在 github 里提交你的作业**
- **提交期限是 12 月 2 日**

Question #1: BigBangTheory. (Attached Data: BigBangTheory)

The Big Bang Theory, a situation comedy featuring Johnny Galecki, Jim Parsons, and Kaley Cuoco-Sweeting, is one of the most-watched programs on network television. The first two episodes for the 2011–2012 season premiered on September 22, 2011; the first episode attracted 14.1 million viewers and the second episode attracted 14.7 million viewers. The attached data file BigBangTheory shows the number of viewers in millions for the first 21 episodes of the 2011–2012 season (*the Big Bang theory* website, April 17, 2012).

a. Compute the minimum and the maximum number of viewers.

```
##   The mininum of viewers is 13.3 millions.
```

```
##   The maxinum of viewers is 16.5 millions.
```

b. Compute the mean, median, and mode.

```
##   The mean of viewers is 13.3 millions.
```

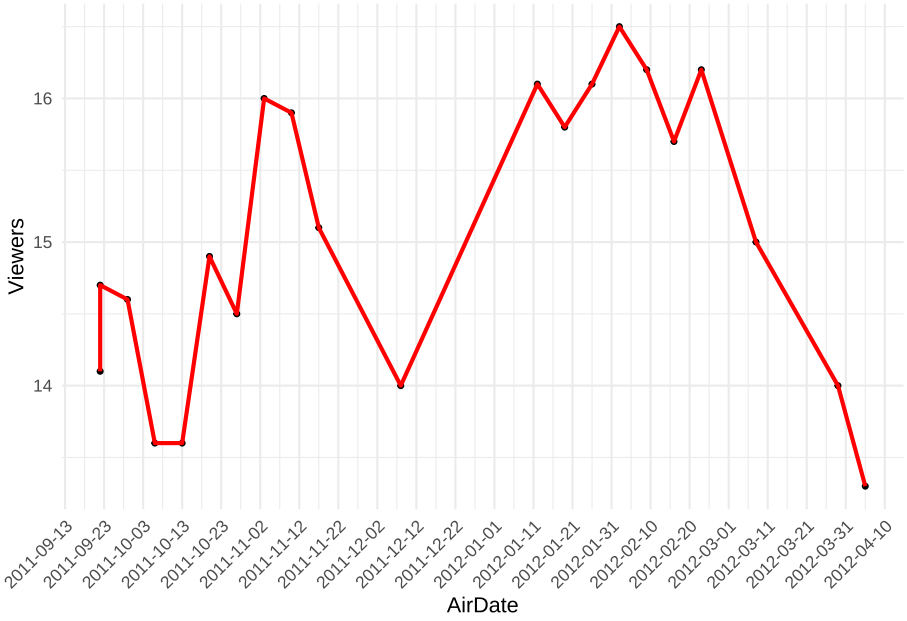
```
##   The median of viewers is 15 millions.
```

```
##   The mode of viewers is c(13.6, 14, 16.1, 16.2) millions.
```

c. Compute the first and third quartiles.

```
##   The first quantile is 14.1 millions, and the third quantile is 16 millions
```

d. has viewership grown or declined over the 2011–2012 season? Discuss.



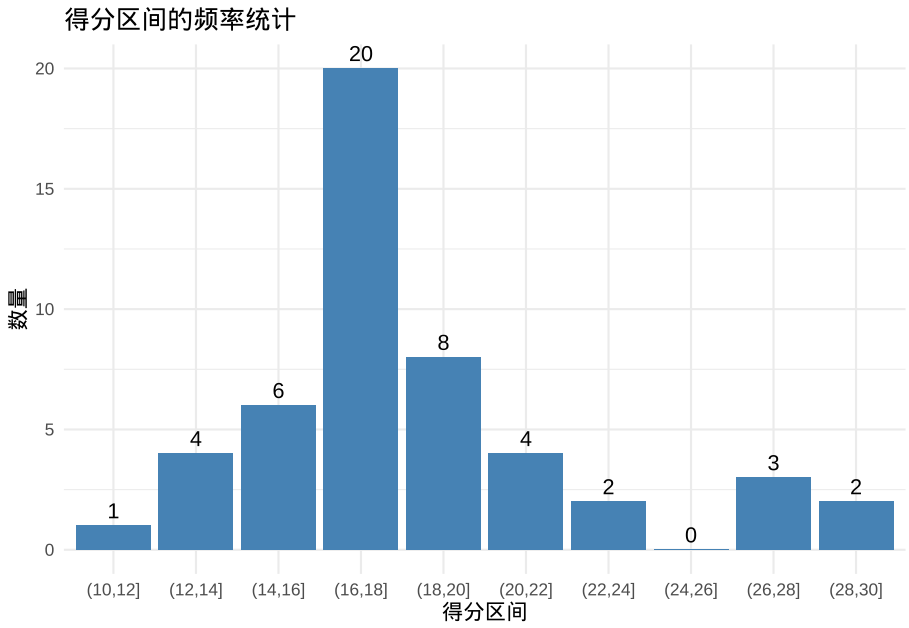

```
From September 2011 to the end of January 2012, the number of viewers of the Big Bang T
```

Question #2: NBAPlayerPts. (Attached Data: NBAPlayerPts)

CbSSports.com developed the Total Player Rating system to rate players in the National Basketball Association (NBA) based on various offensive and defensive statistics. The attached data file NBAPlayerPts shows the average number of points scored per game (PPG) for 50 players with the highest ratings for a portion of the 2012–2013 NBA season (CbSSports.com website, February 25, 2013). Use classes starting at 10 and ending at 30 in increments of 2 for PPG in the following.

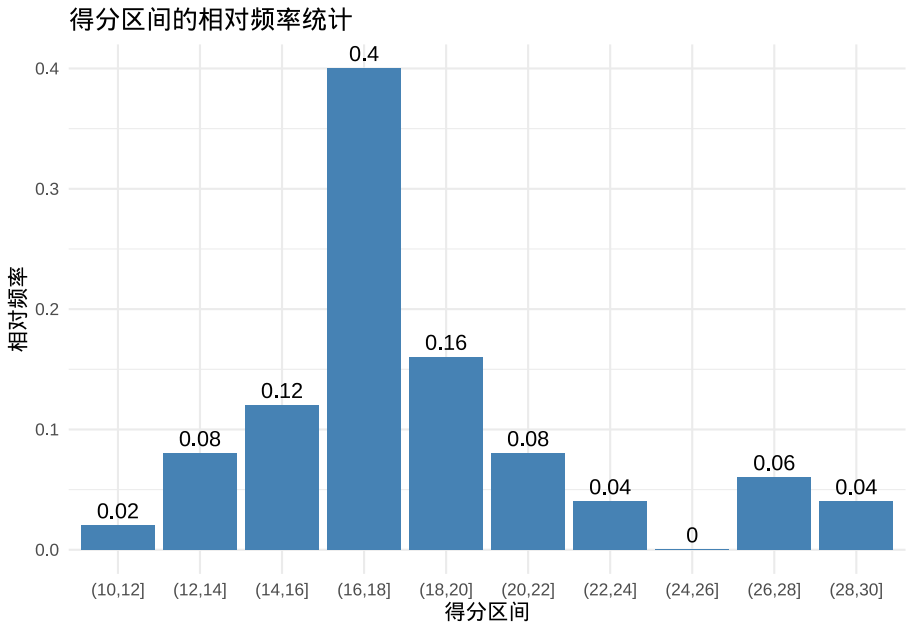
a. Show the frequency distribution.

```
##      group frequency
## 1  (10,12]         1
## 2  (12,14]         4
## 3  (14,16]         6
## 4  (16,18]        20
## 5  (18,20]         8
## 6  (20,22]         4
## 7  (22,24]         2
## 8  (24,26]         0
## 9  (26,28]         3
## 10 (28,30]         2
```



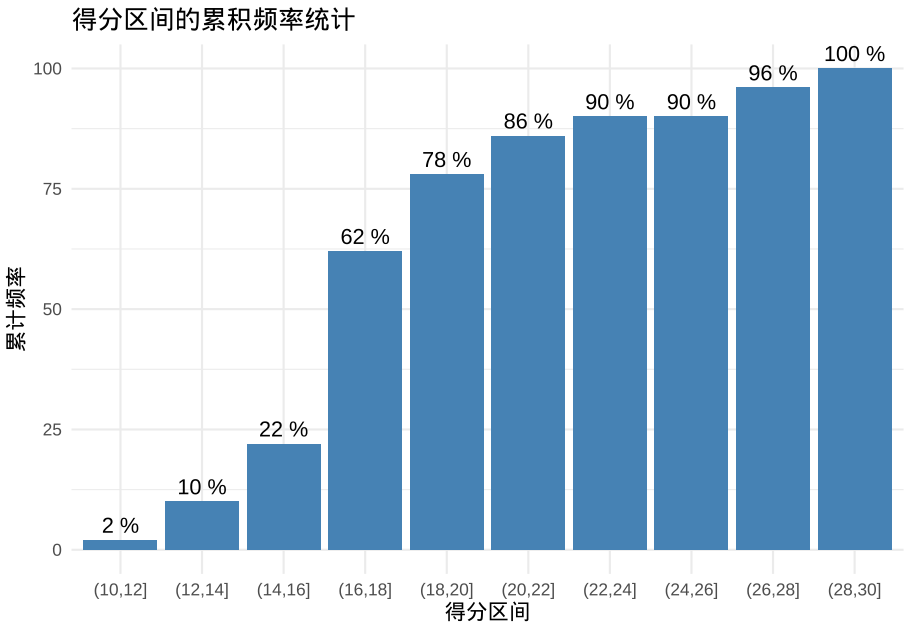
b. Show the relative frequency distribution.

```
##      group relative_frequency
## 1  (10,12]              0.02
## 2  (12,14]              0.08
## 3  (14,16]              0.12
## 4  (16,18]              0.40
## 5  (18,20]              0.16
## 6  (20,22]              0.08
## 7  (22,24]              0.04
## 8  (24,26]              0.00
## 9  (26,28]              0.06
## 10 (28,30]              0.04
```

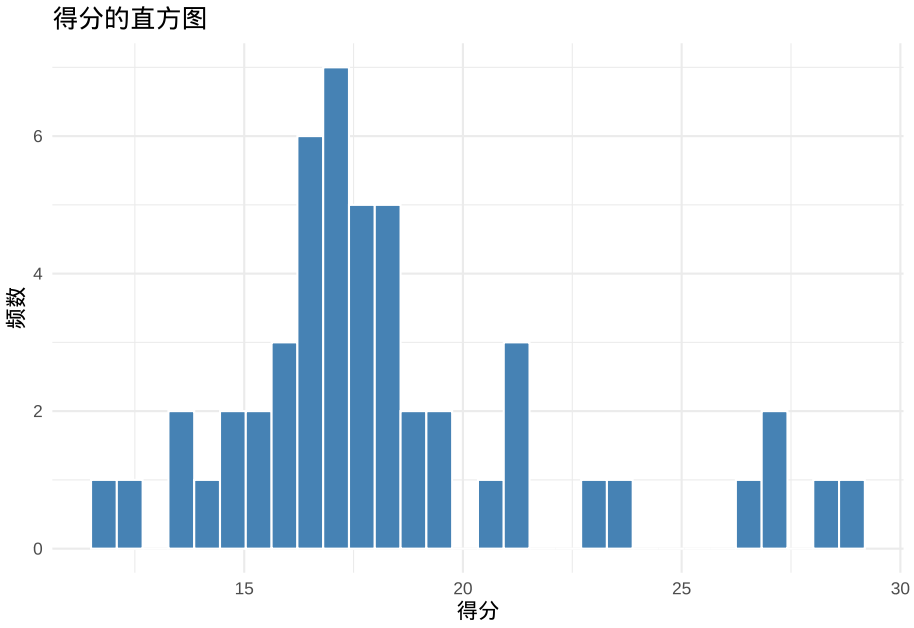


c. Show the cumulative percent frequency distribution.

```
##      group cumulative_percent_freq
## 1  (10,12]                      2
## 2  (12,14]                     10
## 3  (14,16]                     22
## 4  (16,18]                     62
## 5  (18,20]                     78
## 6  (20,22]                     86
## 7  (22,24]                     90
## 8  (24,26]                     90
## 9  (26,28]                     96
## 10 (28,30]                    100
```



d. Develop a histogram for the average number of points scored per game.



e. Do the data appear to be skewed? Explain.

```
## [1] 1.124025
```

```
##  skewness_value 为1.124025,  右偏
```

f. What percentage of the players averaged at least 20 points per game?

```
## The players averaged at least 20 points per game is 22%.
```

Question #3: A researcher reports survey results by stating that the standard error of the mean is 20. The population standard deviation is 500.

a. How large was the sample used in this survey?

```
## The number of the sample is  625
```

b. What is the probability that the point estimate was within ±25 of the population mean?

```
## [1] 0.7887005
```

Question #4: Young Professional Magazine (Attached Data: Professional)

Young Professional magazine was developed for a target audience of recent college graduates who are in their first 10 years in a business/professional career. In its two years of publication, the magazine has been fairly successful. Now the publisher is interested in expanding the magazine's advertising base. Potential advertisers continually ask about the demographics and interests of subscribers to *young Professionals*. To collect this information, the magazine commissioned a survey to develop a profile of its subscribers. The survey results will be used to help the magazine choose articles of interest and provide advertisers with a profile of subscribers. As a new employee of the magazine, you have been asked to help analyze the survey results.

Some of the survey questions follow:

1. What is your age?
2. Are you: Male_____ Female_____
3. Do you plan to make any real estate purchases in the next two years? Yes_____ No_____
4. What is the approximate total value of financial investments, exclusive of your home, owned by you or members of your household?
5. How many stock/bond/mutual fund transactions have you made in the past year?
6. Do you have broadband access to the Internet at home? Yes_____ No_____
7. Please indicate your total household income last year. _____
8. Do you have children? Yes_____ No_____

The file entitled Professional contains the responses to these questions.

Managerial Report:

Prepare a managerial report summarizing the results of the survey. In addition to statistical summaries, discuss how the magazine might use these results to attract advertisers. You might also comment on how the survey results could be used by the magazine's editors to identify topics that would be of interest to readers. Your report should address the following issues, but do not limit your analysis to just these areas.

a. Develop appropriate descriptive statistics to summarize the data.

```
##       Age           Gender    RealEstatePurchases ValueOfInvestments
##  Min.   :19.00   Female:181   No :229             Min.   :     0
##  1st Qu.:28.00   Male  :229   Yes:181             1st Qu.: 18300
##  Median :30.00                                    Median : 24800
##  Mean   :30.11                                    Mean   : 28538
##  3rd Qu.:33.00                                    3rd Qu.: 34275
```

```
##  Max.   :42.00                                    Max.   :133400
##  NumberofTrans    BoardbandAccess HouseholdIncome  haveChildren
##  Min.   : 0.000   No :154         Min.   : 16200   No :154
##  1st Qu.: 4.000   Yes:256         1st Qu.: 51625   Yes:256
##  Median : 6.000                   Median : 66050
##  Mean   : 5.973                   Mean   : 74460
##  3rd Qu.: 7.000                   3rd Qu.: 88775
##  Max.   :21.000                   Max.   :322500
```

b. Develop 95% confidence intervals for the mean age and household income of subscribers.

```
## 95% confidence intervals of the mean age is :

## [1] 29.72153 30.50286

## 95% confidence intervals of the household income is :

## [1] 71079.26 77839.77
```

c. Develop 95% confidence intervals for the proportion of subscribers who have broadband access at home and the proportion of subscribers who have children.

```
## 95% confidence intervals for the proportion of subscribers who have broadband access

## 95% confidence intervals for  the proportion of subscribers who have children is: 0.
```

d. Would *Young Professional* be a good advertising outlet for online brokers? Justify your conclusion with statistical data.

```
##       Age           Gender    RealEstatePurchases ValueOfInvestments
##  Min.   :19.00   Female:181   No :229             Min.   :     0
##  1st Qu.:28.00   Male  :229   Yes:181             1st Qu.: 18300
```

```
##  Median :30.00                                    Median : 24800
##  Mean   :30.11                                    Mean   : 28538
##  3rd Qu.:33.00                                    3rd Qu.: 34275
##  Max.   :42.00                                    Max.   :133400
##  NumberofTrans    BoardbandAccess HouseholdIncome  haveChildren
##  Min.   : 0.000   No :154         Min.   : 16200   No :154
##  1st Qu.: 4.000   Yes:256         1st Qu.: 51625   Yes:256
##  Median : 6.000                   Median : 66050
##  Mean   : 5.973                   Mean   : 74460
##  3rd Qu.: 7.000                   3rd Qu.: 88775
##  Max.   :21.000                   Max.   :322500
```

```
##          n
## 1 0.995122
```

```
##           n
## 1 0.6243902
```

```
##           n
## 1 0.6243902
```

我认为 *Young Professional* 是在线经纪人的一个很好的广告渠道。从统计结果来看:
1. 过去一年中 99.5% 的人有投资经历，投资的金额均值在 $28538，有投资意愿;
2. 64% 的家中有网络，可以用于在线联络和投资，有投资的渠道；
3. 去年家庭的收入均值为 $74460，有投资的能力；
综上，我认为该杂志是一个很好的广告渠道。

e. Would this magazine be a good place to advertise for companies selling educational software and computer games for young children?

我认是该杂志可以用于投放教育软件和电脑游戏的广告，理由如下：
1. 从统计样本中可以得出，62.4% 的订阅用户，家中都有小孩并且家中有

网络，可以使用教育软件和玩游戏；
2. 订阅用户的收入均值为 $74460，收入还可以，并且都有投资，这种人群比较重视小孩教育，所以更愿意来投资在教育上; 综上，我认为该杂志可以用于投放教育软件和电脑游戏的广告。

f. Comment on the types of articles you believe would be of interest to readers of *Young Professional*.

结合所调研的读者的统计数据来看，我认为该杂志的订阅者感兴趣的文章类型包括：1. 投资建议类，例如：如何选择投资标的；如何进行资产配置实现收益最大化等；
2. 企业分析类，例如：企业财务状况分析；企业业务分析等；
3. 家庭关系类，例如：促进家庭关系；发现孩子的兴趣；亲子游戏等；
4. 实事热点类，例如：国内外的热点事件等。

Question #5: Quality Associate, Inc. (Attached Data: Quality)

Quality associates, inc., a consulting firm, advises its clients about sampling and statistical procedures that can be used to control their manufacturing processes. in one particular application, a client gave Quality associates a sample of 800 observations taken during a time in which that client's process was operating satisfactorily. the sample standard deviation for these data was .21; hence, with so much data, the population standard deviation was assumed to be .21. Quality associates then suggested that random samples of size 30 be taken periodically to monitor the process on an ongoing basis. by analyzing the new samples, the client could quickly learn whether the process was operating satisfactorily. when the process was not operating satisfactorily, corrective action could be taken to eliminate the problem. the design specification indicated the mean for the process should be 12. the hypothesis test suggested by Quality associates follows.

$$H_0 : \mu = 12 H_1 : \mu \neq 12$$

Corrective action will be taken any time H_0 is rejected.

Data are available in the data set Quality.

Managerial Report

a. Conduct a hypothesis test for each sample at the .01 level of significance and determine what action, if any, should be taken. Provide the p-value for each test.

```
## [1] "Sample.1" "Sample.2" "Sample.3" "Sample.4"
```

```
## [1] 0.2810 0.4547 0.0038 0.0339
```

b. compute the standard deviation for each of the four samples. does the assumption of .21 for the population standard deviation appear reasonable?

```
## [1] 0.22 0.22 0.21 0.21
```

The assumption of the standard deviation is reasonable.

c. compute limits for the sample mean $\overline{x}$ around $\mu = 12$ such that, as long as a new sample mean is within those limits, the process will be considered to be operating satisfactorily. if $\overline{x}$ exceeds the upper limit or if $\overline{x}$ is below the lower limit, corrective action will be taken. these limits are referred to as upper and lower control limits for quality control purposes.

```
## The upper limit is 12.12
```

```
## The lower limit is 11.88
```

d. discuss the implications of changing the level of significance to a larger value. what mistake or error could increase if the level of significance is increased?
增加显著性水平是假设检验中用于决定是否拒绝原假设的阈值。增加显著性水平，会增加 TypeIerror（假阳性）的概率，当原假设为真时，可能存在拒绝原假设的情况。

Question #6: Vacation occupancy rates were expected to be up during March 2008 in Myrtle Beach, South Carolina (*the sun news,* February 29, 2008). Data in the file Occupancy (Attached file **Occupancy**) will allow you to replicate the findings presented in the newspaper. The data show units rented and not rented for a random sample of vacation properties during the first week of March 2007 and March 2008.

a. Estimate the proportion of units rented during the first week of March 2007 and the first week of March 2008.

```
## The proportion of units rented during the first week of March 2007 is 0.35, and the
```

b. Provide a 95% confidence interval for the difference in proportions.

```
## The 95% confidence interval is -0.224, -0.016
```

c. On the basis of your findings, does it appear March rental rates for 2008 will be up from those a year earlier? 是的，由于 2007 年和 2008 年的比例差的置信区间不包含 0，说明在 95% 的置信水平下，2007 年 3 月和 2008 年 3 月是有差异的，而 2008 年 3 月会比 2007 年高，所以 2008 年的 3 月会比同期高。

Question #7: **Air Force Training Program** (data file: Training)

An air force introductory course in electronics uses a personalized system of instruction whereby each student views a videotaped lecture and then is given a programmed instruc-tion text. the students work independently with the text until they have completed the training and passed a test. Of concern is the varying pace at which the students complete this portion of their training program. Some students are able to cover the programmed instruction text relatively quickly, whereas other students work much longer with the text and require additional time to complete the course. The fast students wait until the slow students complete the introductory course before the entire group proceeds together with other aspects of their training.

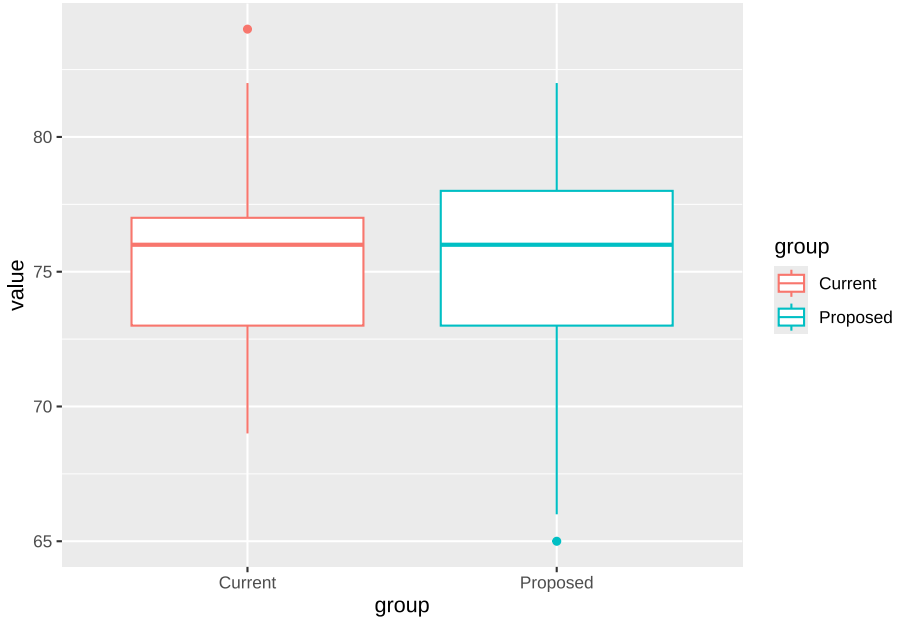
A proposed alternative system involves use of computer-assisted instruction. In this method, all students view the same videotaped lecture and then each is assigned to a computer terminal for further instruction. The computer guides the student, working independently, through the self-training portion of the course.

To compare the proposed and current methods of instruction, an entering class of 122 students was assigned randomly to one of the two methods. one group of 61 students used the current programmed-text method and the other group of 61 students used the proposed computer-assisted method. The time in hours was recorded for each student in the study. Data are provided in the data set training (see Attached file).

Managerial Report

a. use appropriate descriptive statistics to summarize the training time data for each method. what similarities or differences do you observe from the sample data?

```
##             group Min Max  Mean       Var qu1st qu3st
## 25%...1  Current  65  84 75.07 15.562295    72    78
## 25%...2 Proposed  69  82 75.43  6.281967    74    77
```



b. Comment on any difference between the population means for the two methods. Discuss your findings.

```
##
##  Welch Two Sample t-test
##
## data:  data$Current and data$Proposed
## t = -0.60268, df = 101.65, p-value = 0.5481
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.5476613  0.8263498
## sample estimates:
## mean of x mean of y
##  75.06557  75.42623
```

由于 P 值 0.5481 大于 0.05，并且 95% 置信区间包含 0，没有证据说明两种教学方式的存在差异。

c. compute the standard deviation and variance for each training method. conduct a hypothesis test about the equality of population variances for the two training methods. Discuss your findings.

```
## 标准差：

## [1] 3.944907

## [1] 2.506385

## 方差：

## [1] 15.5623

## [1] 6.281967

## F统计量：

## [1] 2.477296

## F检验拒绝域(最小、最大）：

## [1] 0.5999553 1.6667908
```

F 统计量为 2.477296 >1.6667908, 落在拒绝域内，所以两种实验结果是有差异的。

d. what conclusion can you reach about any differences between the two methods? what is your recommendation? explain.

从数据来看：
1. Proposed 的平均得分 (95.43) 略高于 Current（75.07），差距非常小；
2. Current 的方差 (15.56) 大于 Proposed 的方差 (6.28)，说明 Proposed 的数据更集中，稳定性更好；
3. F 统计量位于拒绝域内，说明两组数据有显著差异；
综上，建议采用 Proposed 的方法。

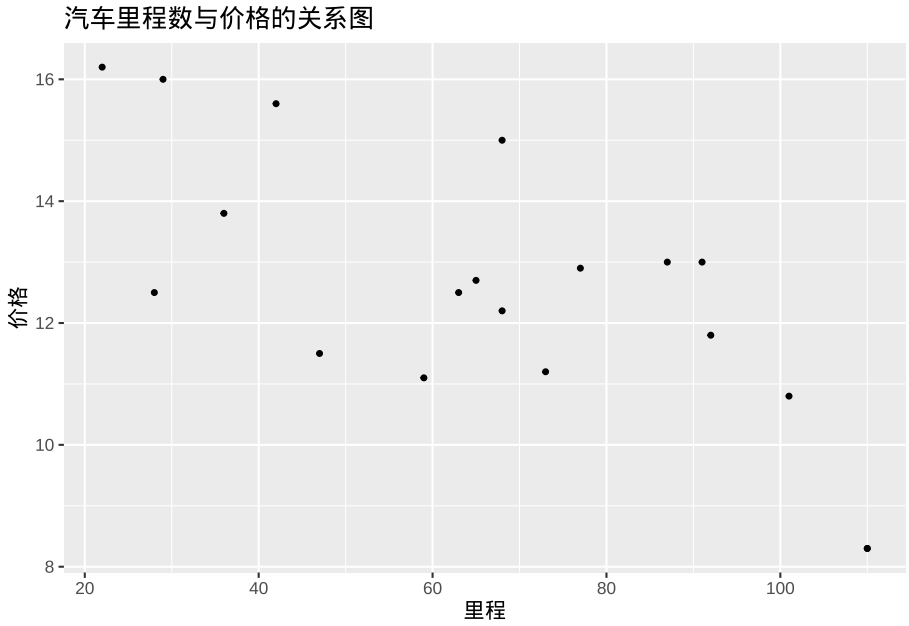
e. can you suggest other data or testing that might be desirable before making a final decision on the training program to be used in the future?
可能还需要：

1. 学习的时间；

2. 结束 1 段时间后（例如 1 个月、3 个月）的得分情况。

Question #8: The Toyota Camry is one of the best-selling cars in North America. The cost of a previously owned Camry depends upon many factors, including the model year, mileage, and condition. To investigate the relationship between the car's mileage and the sales price for a 2007 model year Camry, Attached data file Camry show the mileage and sale price for 19 sales (Pricehub website, February 24, 2012).

a. Develop a scatter diagram with the car mileage on the horizontal axis and the price on the vertical axis.



b. what does the scatter diagram developed in part (a) indicate about the relationship between the two variables?

Miles 和 Price 呈负相关关系

c. Develop the estimated regression equation that could be used to predict the price ($1000s) given the miles (1000s).

```
##
## Call:
## lm(formula = Price ~ Miles, data = data8)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32408 -1.34194  0.05055  1.12898  2.52687
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.46976    0.94876  17.359 2.99e-12 ***
## Miles       -0.05877    0.01319  -4.455 0.000348 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.541 on 17 degrees of freedom
## Multiple R-squared:  0.5387, Adjusted R-squared:  0.5115
## F-statistic: 19.85 on 1 and 17 DF,  p-value: 0.0003475
```

从回归模型的摘要信息可知，线性规划的方程为：price = -0.05877 * miles + 16.46976

d. Test for a significant relationship at the .05 level of significance.

p-value(0.000348) 小于 0.05

e. Did the estimated regression equation provide a good fit? Explain. 根据 R-squared: 0.5387，说明是合适的。

f. Provide an interpretation for the slope of the estimated regression equation. Miles 每增加一个单位，价格减少 0.05877*1000=58.77 美元。

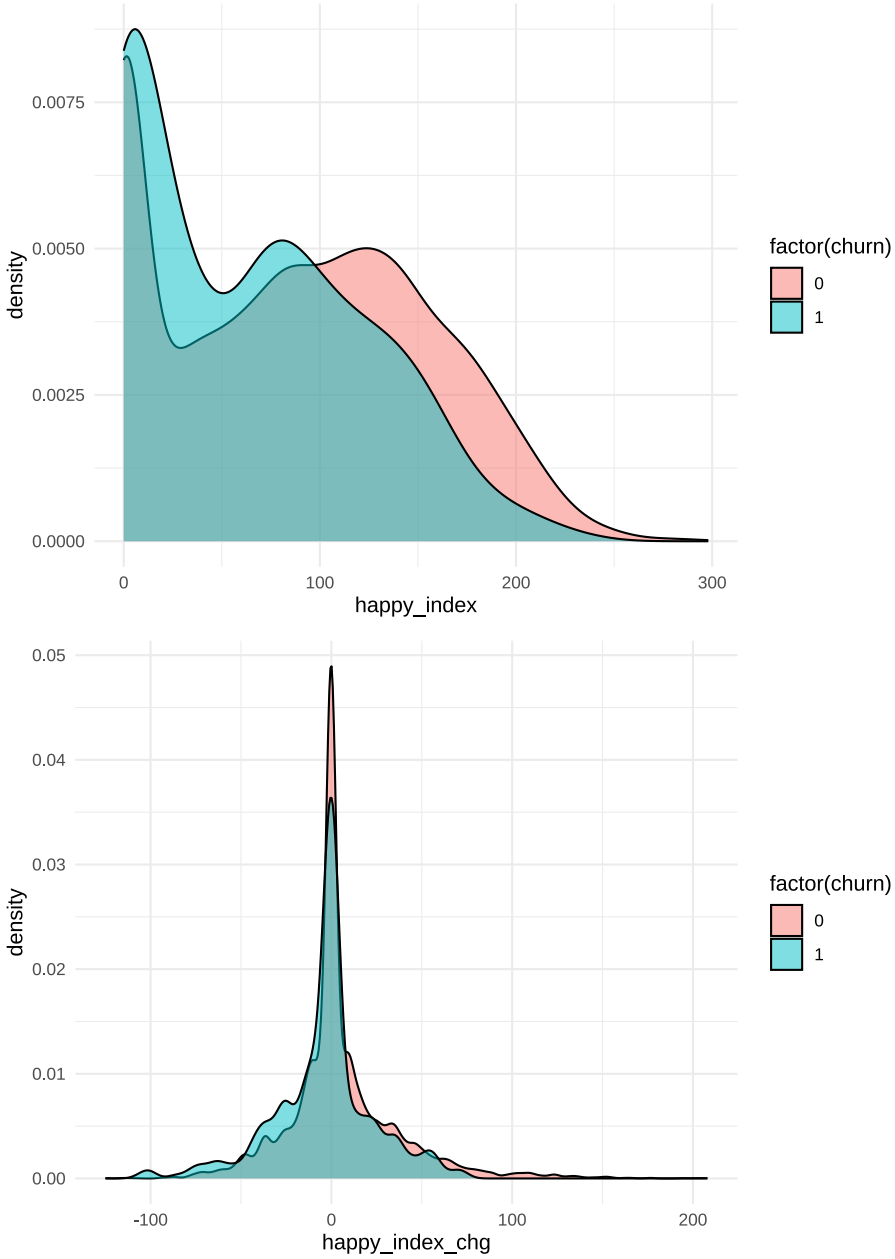
g. Suppose that you are considering purchasing a previously owned 2007 Camry that has been driven 60,000 miles. Using the estimated regression equation developed in part (c), predict the price for this car. Is this the price you would offer the seller.

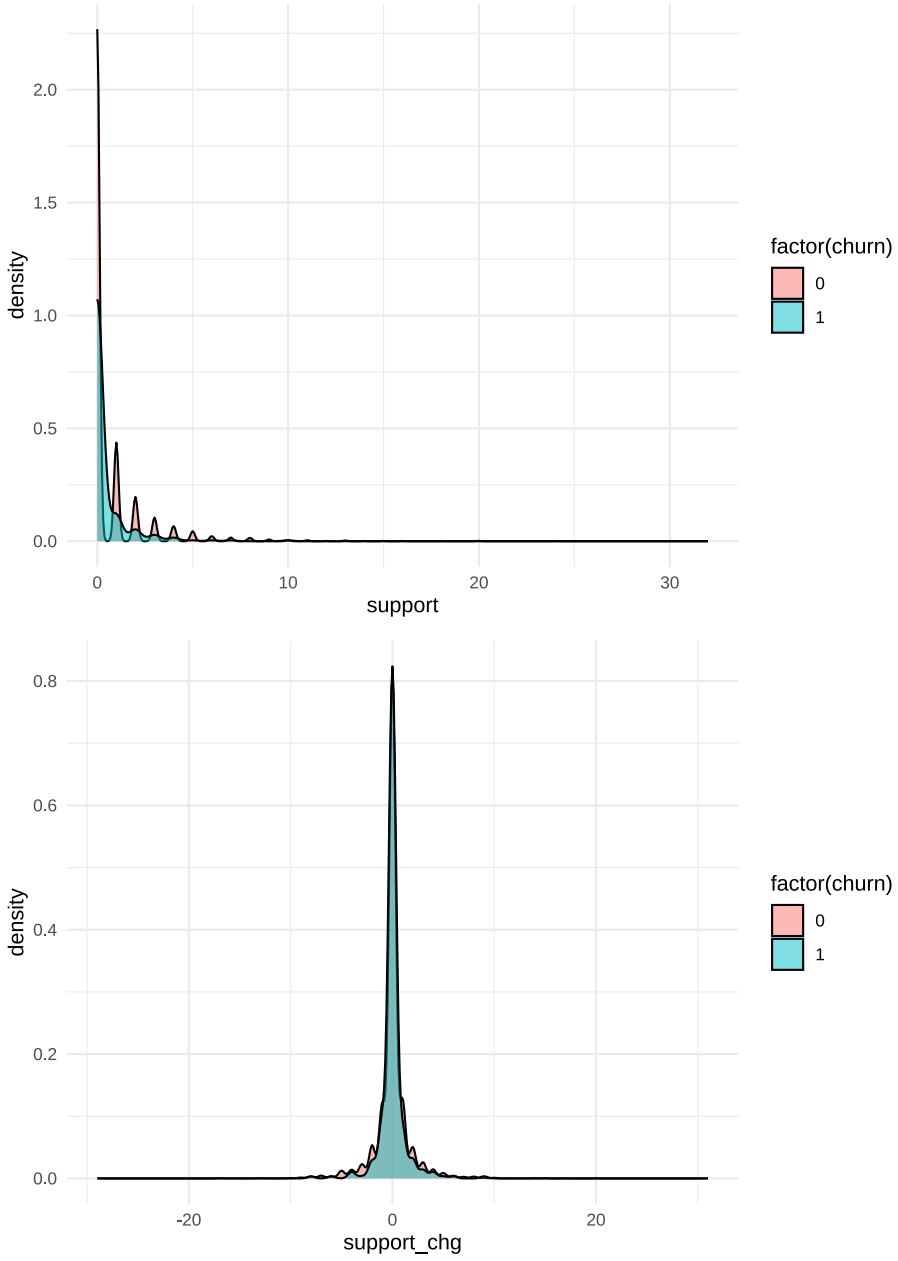
```
##        1
## 12.94332
```

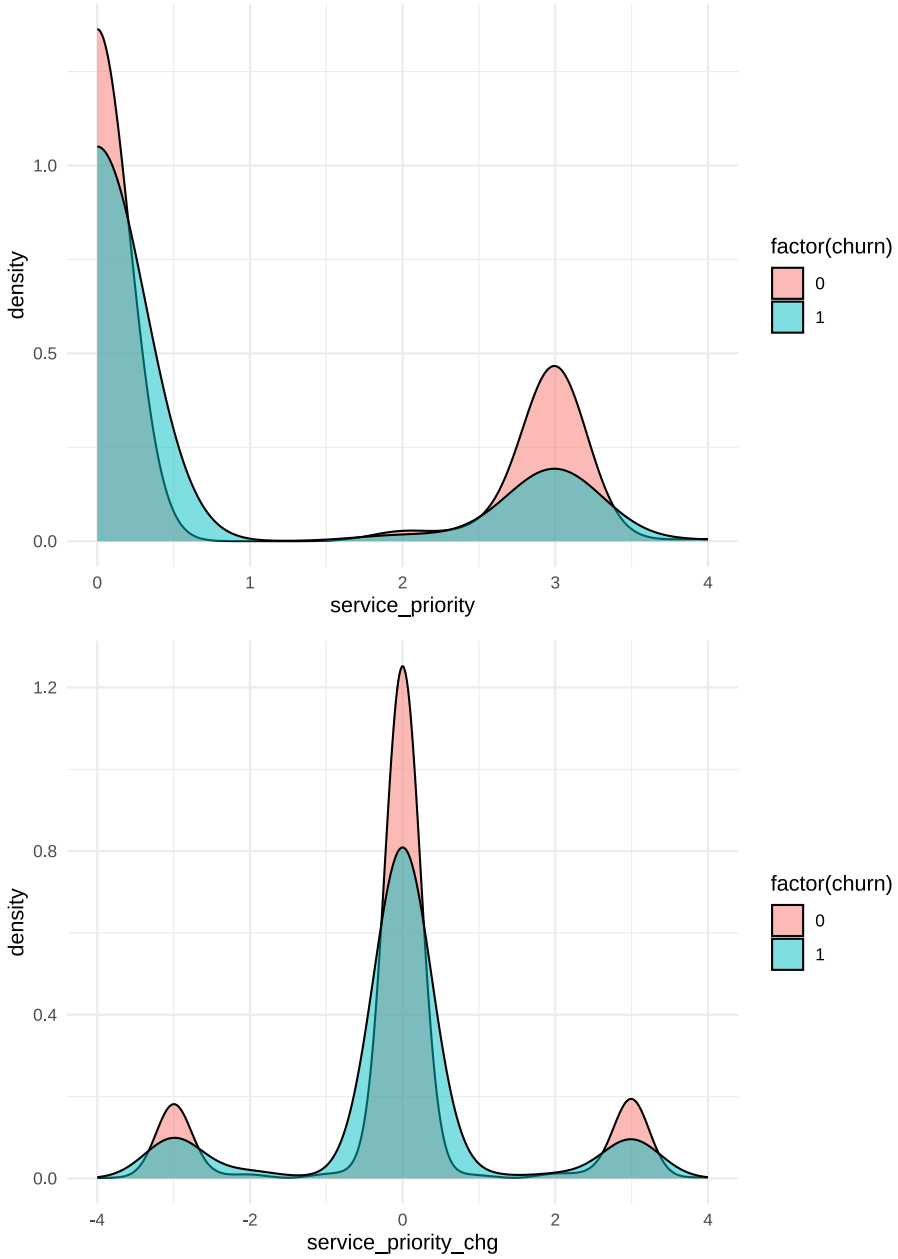
6 万公里，需要 1.294 万美元。这个价格我觉得有点贵，不喜欢二手车。

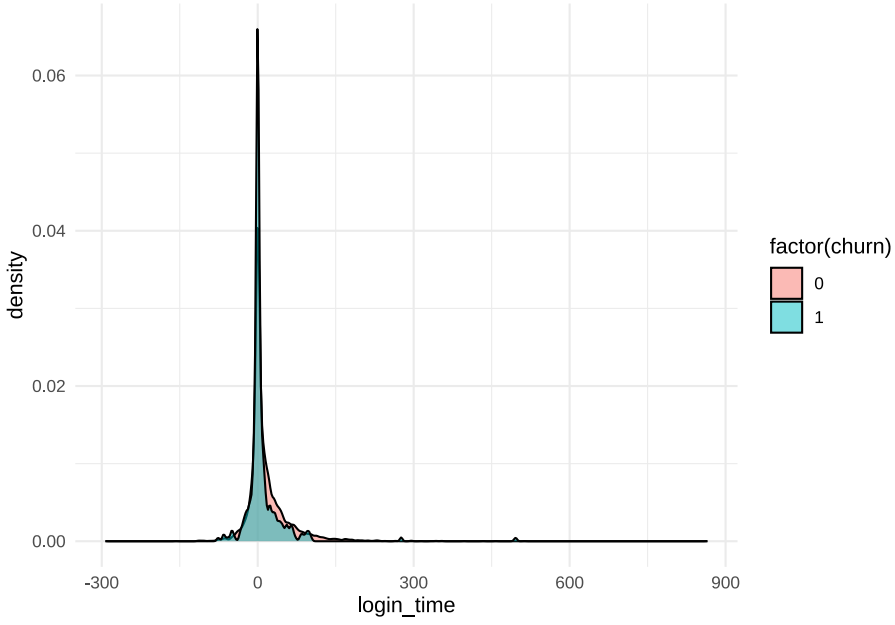
Question #9: 附件 WE.xlsx 是某提供网站服务的 Internet 服务商的客户数据。数据包含了 6347 名客户在 11 个指标上的表现。其中" 流失 "指标中 0 表示流失，"1"表示不流失，其他指标含义看变量命名。

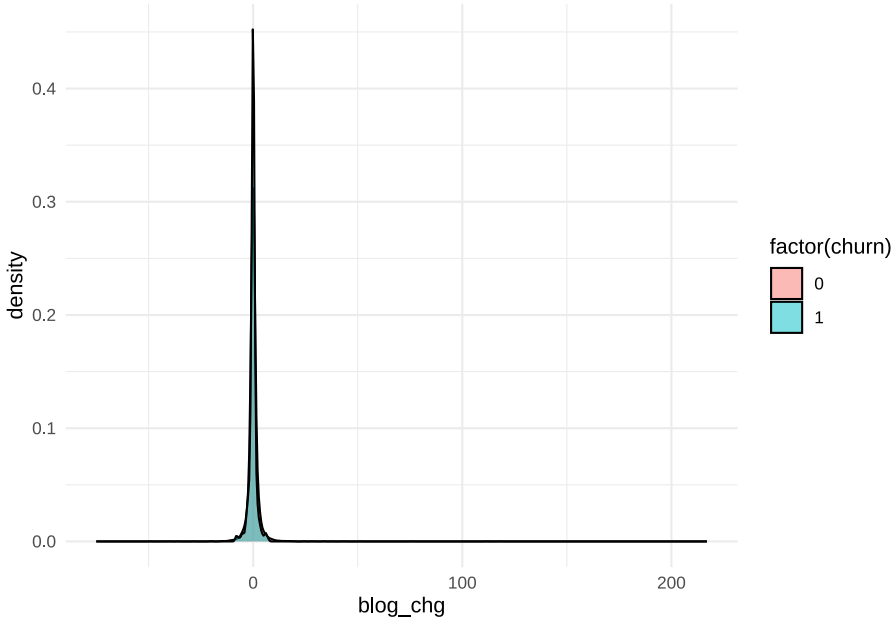
a. 通过可视化探索流失客户与非流失客户的行为特点（或特点对比），你能发现流失与非流失客户行为在哪些指标有可能存在显著不同？

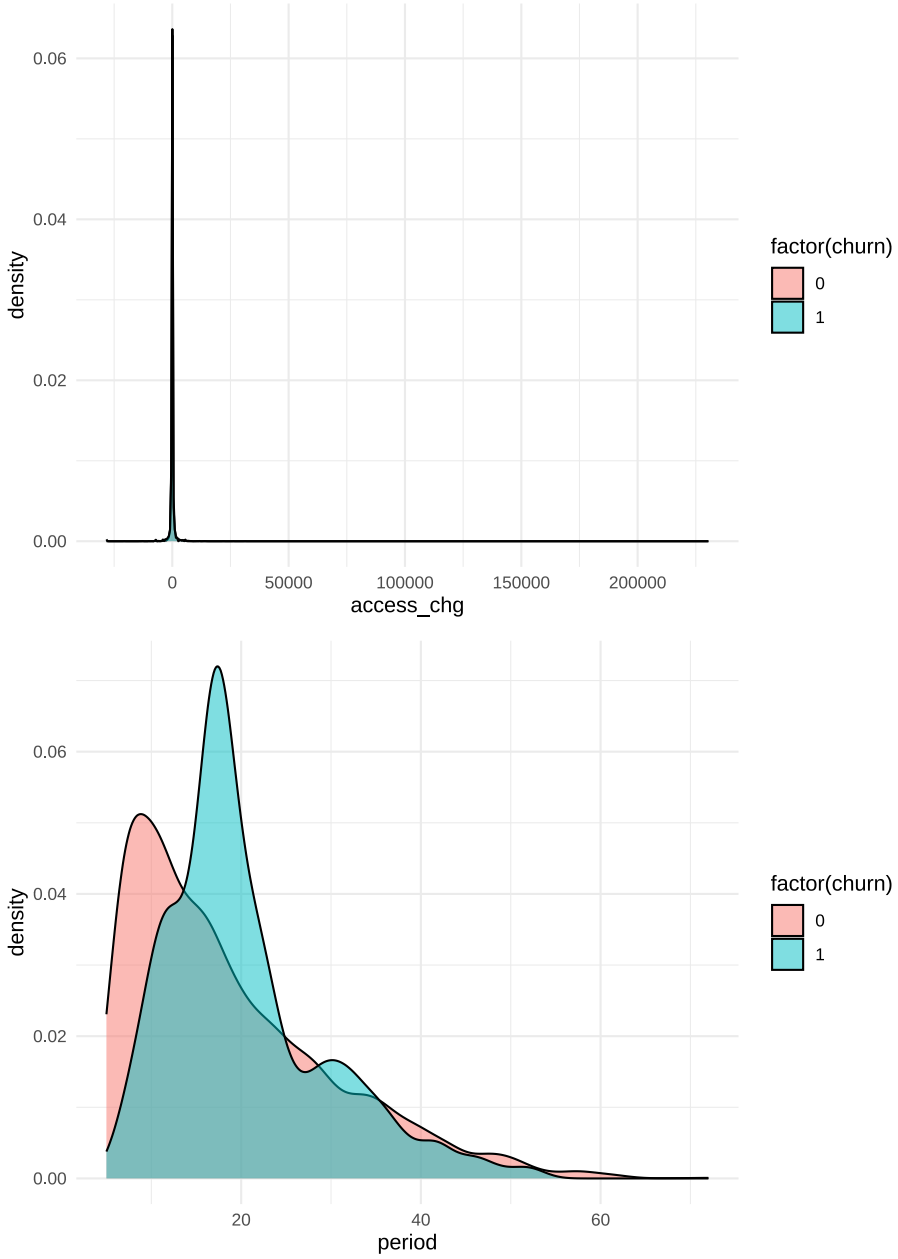

0.0075
0.0050
0.0025
0.0000
density
0
100
200
300
happy_index
factor(churn)
0
1
0.05
0.04
0.03
0.02
0.01
0.00
density
-100
0
100
200
happy_index_chg
factor(churn)
0
1

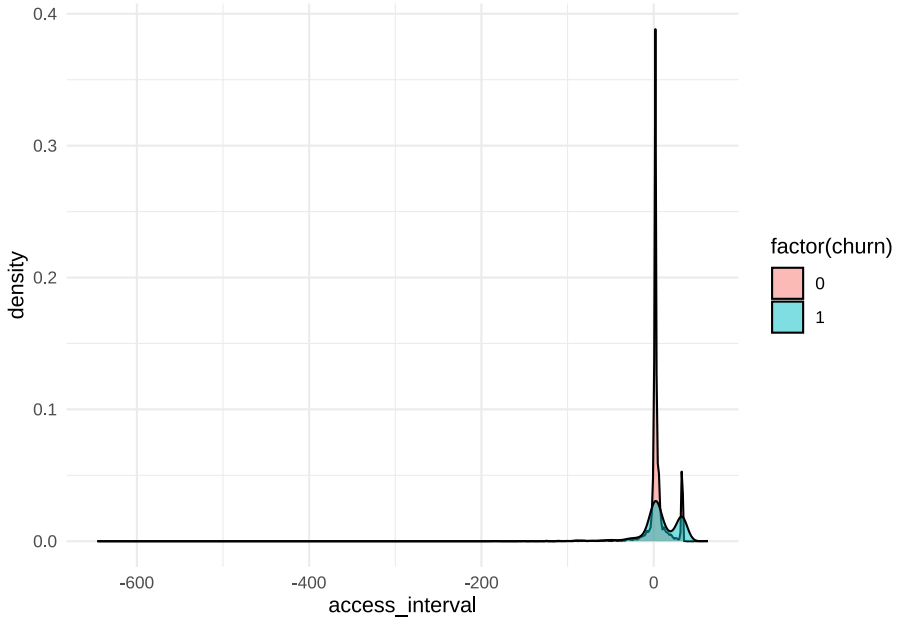

density
2.0
1.5
1.0
0.5
0.0
0
10
20
30
support
factor(churn)
0
1
density
0.8
0.6
0.4
0.2
0.0
-20
0
20
support_chg
factor(churn)
0
1


density
1.0
0.5
0.0
0
1
2
3
4
service_priority
factor(churn)
0
1
density
1.2
0.8
0.4
0.0
-4
-2
0
2
4
service_priority_chg
factor(churn)
0
1


0.06
0.04
0.02
0.00
density
-300
0
300
600
900
login_time
factor(churn)
0
1


0.4
0.3
0.2
0.1
0.0
density
0
100
200
blog_chg
factor(churn)
0
1


0.06
0.04
0.02
0.00
density
0
50000
100000
150000
200000
access_chg
factor(churn)
0
1
0.06
0.04
0.02
0.00
density
20
40
60
period
factor(churn)
0
1



如上图所示，流失客户与非流失客户在多个特征上都存在差，通过可视化探索发现：
1. 幸福指数在 100 以下，非流失客户大于流失客户；幸福指数在 100 以上，容易流失；2. 当月客户支持数越少，越容易流失；3. 当月服务优先级越低，流失客户越多；4. 当月服务优先级相比上月没有变化的，客户越容易流失；5. 当月登录次数越多，越不容易流失；6. 客户使用期限小于 15，越容易流失，使用在 20 天左右越不容易流失

b. 通过均值比较的方式验证上述不同是否显著。

```
##  [1] "id"                  "churn"               "happy_index"
##  [4] "happy_index_chg"     "support"             "support_chg"
##  [7] "service_priority"    "service_priority_chg" "login_time"
## [10] "blog_chg"            "access_chg"          "period"
## [13] "access_interval"

## happy_index  t-test p-value: 0
## happy_index_chg  t-test p-value: 0
```

```
## support  t-test p-value: 0
## support_chg  t-test p-value: 0.528
## service_priority  t-test p-value: 0
## service_priority_chg  t-test p-value: 0.522
## login_time  t-test p-value: 0
## blog_chg  t-test p-value: 0.012
## access_chg  t-test p-value: 0.056
## period  t-test p-value: 0.003
## access_interval  t-test p-value: 0
```

从11个指标的平均值来看，均有差异;
采用t.test检验，根据p值小于0.05存在显著差异，得出除了"客户支持相比上月的变化", "服务优先

c. 以" 流失 "为因变量，其他你认为重要的变量为自变量（提示：a、b 两步的发现），建立回归方程对是否流失进行预测。

```
##
## Call:
## glm(formula = churn ~ happy_index + happy_index_chg + support +
##     service_priority + login_time + blog_chg + period + access_interval,
##     family = binomial(link = "logit"), data = data9)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.8763327  0.1212590 -23.721  < 2e-16 ***
## happy_index      -0.0051988  0.0011558  -4.498 6.86e-06 ***
## happy_index_chg  -0.0093063  0.0024124  -3.858 0.000114 ***
## support          -0.0221691  0.0714550  -0.310 0.756369
## service_priority -0.0447524  0.0741355  -0.604 0.546072
## login_time        0.0008545  0.0019376   0.441 0.659211
## blog_chg         -0.0009717  0.0205099  -0.047 0.962213
## period            0.0142559  0.0052396   2.721 0.006513 **
## access_interval   0.0169505  0.0042787   3.962 7.44e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2553.1  on 6346  degrees of freedom
## Residual deviance: 2452.2  on 6338  degrees of freedom
## AIC: 2470.2
##
## Number of Fisher Scoring iterations: 6
```

d. 根据上一步预测的结果，对尚未流失（流失 =0）的客户进行流失可能性排序，并给出流失可能性最大的前 100 名用户 ID 列表。

rownum	id	prob
1	1363	0.1919757
2	1672	0.1802687
3	299	0.1744655
4	2922	0.1623822
5	2951	0.1618436
6	1021	0.1589334
7	335	0.1563708
8	156	0.1541017
9	1488	0.1469821
10	3340	0.1453980
11	2296	0.1448915
12	1069	0.1379684
13	3811	0.1362973
14	2653	0.1346123
15	3604	0.1338222
16	1405	0.1328413
17	2636	0.1313254
18	2077	0.1308082
19	1987	0.1297076

rownum	id	prob
20	4292	0.1296641
21	2082	0.1268756
22	1782	0.1265470
23	2766	0.1257334
24	2166	0.1249558
25	2120	0.1234054
26	1303	0.1231142
27	904	0.1230854
28	3092	0.1226184
29	2624	0.1222536
30	2371	0.1218715
31	2928	0.1218715
32	1563	0.1203540
33	1711	0.1188529
34	1532	0.1170892
35	891	0.1158993
36	945	0.1158993
37	947	0.1158993
38	948	0.1158993
39	2084	0.1156708
40	896	0.1144465
41	402	0.1139450
42	227	0.1130096
43	979	0.1130096
44	938	0.1127398
45	2902	0.1121649
46	257	0.1115885
47	317	0.1115885
48	363	0.1115885
49	371	0.1115885
50	523	0.1115885
51	543	0.1115885

rownum	id	prob
52	548	0.1115885
53	787	0.1115885
54	1214	0.1115885
55	1760	0.1115885
56	3312	0.1115885
57	3313	0.1115885
58	4500	0.1115885
59	3569	0.1106295
60	1659	0.1102284
61	3163	0.1101830
62	3235	0.1101830
63	3349	0.1101830
64	3228	0.1099191
65	3267	0.1099191
66	640	0.1097650
67	5312	0.1091330
68	930	0.1088256
69	3772	0.1087931
70	3363	0.1074185
71	3487	0.1072170
72	2179	0.1066551
73	4280	0.1062829
74	2707	0.1061770
75	4273	0.1060593
76	3861	0.1055119
77	2189	0.1053949
78	4263	0.1048756
79	4289	0.1048756
80	4291	0.1047153
81	453	0.1038621
82	1831	0.1037784
83	4680	0.1033863

rownum	id	prob
84	3584	0.1019346
85	2316	0.1015223
86	3317	0.1008967
87	5052	0.0995455
88	1523	0.0951064
89	1709	0.0950096
90	1909	0.0949265
91	387	0.0941401
92	5313	0.0935990
93	412	0.0935846
94	105	0.0935244
95	4893	0.0927391
96	1823	0.0926003
97	2003	0.0910678
98	1456	0.0910644
99	430	0.0909350
100	5460	0.0885247