MoviesDemand-Exp_full.R

danny

2020-02-29

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

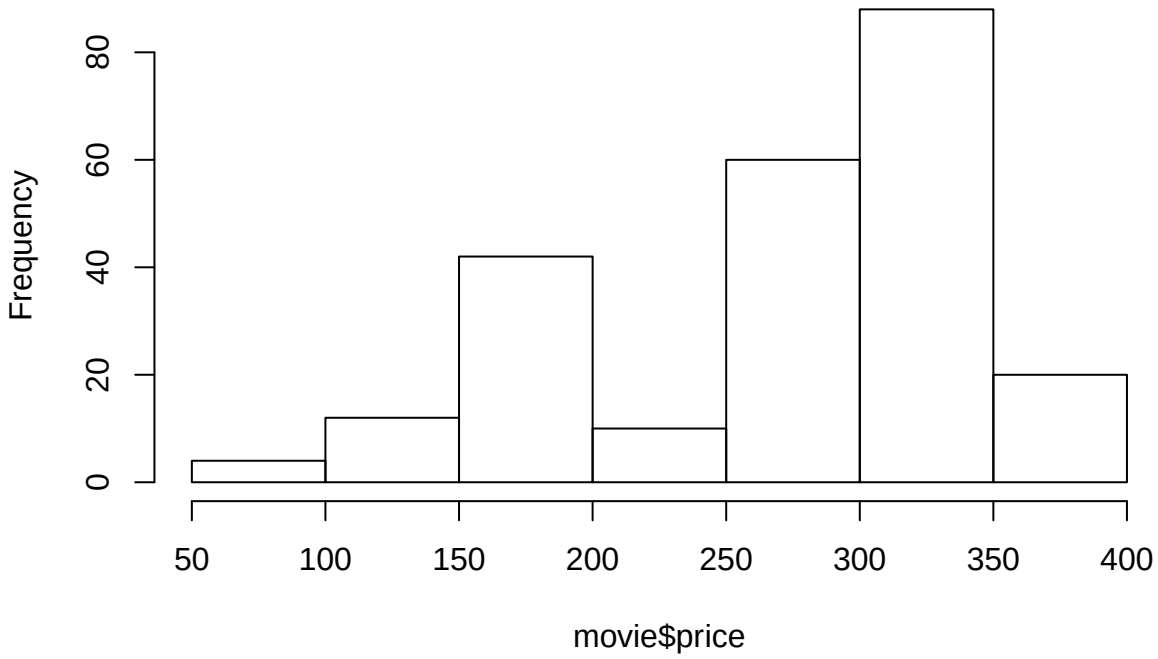
```
#*** MSBA 6440 ***#
#*** Mochen Yang ***#
#*** Original code by Gordon Burtch ***#

# Analyzing Movie Rental Pricing Experiment Data

# import data into R
movie = read.csv("MovieData-Exp.csv")

# Descriptive plots: distributions of leases and prices
hist(movie$leases)
```


```
hist(movie$price)
```




```
# randomization check
# "treatment" is discount on price, we can make a dummy variable of "receiving treatment or not" to fac
movie = movie %>% mutate(discount = base_price - price,
                         has_discount = ifelse(discount > 0, 1, 0))
# check randomization effort on base_price and likes
t.test(likes ~ has_discount, data = movie)
```

```
##
##  Welch Two Sample t-test
##
## data:  likes by has_discount
## t = 0.060292, df = 233.42, p-value = 0.952
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -465366.0  494747.5
## sample estimates:
## mean in group 0 mean in group 1
##         2343120         2328429
```

```
t.test(base_price ~ has_discount, data = movie)
```

```
##
##  Welch Two Sample t-test
##
## data:  base_price by has_discount
```

```
## t = -0.30694, df = 233.79, p-value = 0.7592
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -20.11877  14.69504
## sample estimates:
## mean in group 0 mean in group 1
##        306.6271        309.3390
```

```
# randomization check looks OK


# Let's evaluate statistical power now.
# How big a sample we would need to detect the 20% change they hope to find?
power.t.test(n=NULL,type=c("two.sample"),power=0.8,sig.level=0.1,delta=0.1)
```

```
##
##      Two-sample t test power calculation
##
##               n = 1237.188
##           delta = 0.1
##              sd = 1
##       sig.level = 0.1
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# sample size... 118 movies per group.
# What sort of difference we can reliably detect with our current?
power.t.test(n=118,type=c("two.sample"),power=0.8,sig.level=0.1,delta=NULL)
```

```
##
##      Two-sample t test power calculation
##
##               n = 118
##           delta = 0.324651
##              sd = 1
##       sig.level = 0.1
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# Do we have sufficient sample? What's the implication / advice for management?
# No, we have a highly insufficient sample to detect the desired effect. The general advice is to colle

# Let's estimate the treatment effect
m1 = lm(leases ~ has_discount, data = movie)
summary(m1)
```

```
##
## Call:
```

```
## lm(formula = leases ~ has_discount, data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7288 -0.7288 -0.5085  0.4915  4.2712
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.50847    0.08356   6.085 4.72e-09 ***
## has_discount  0.22034    0.11818   1.864   0.0635 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9077 on 234 degrees of freedom
## Multiple R-squared:  0.01464,    Adjusted R-squared:  0.01043
## F-statistic: 3.476 on 1 and 234 DF,  p-value: 0.06351
```

```
m2 = lm(leases ~ discount, data = movie)
summary(m2)
```

```
##
## Call:
## lm(formula = leases ~ discount, data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8690 -0.5746 -0.5479  0.4521  4.2648
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.547881   0.077052   7.111 1.39e-11 ***
## discount    0.002676   0.001862   1.437    0.152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9104 on 234 degrees of freedom
## Multiple R-squared:  0.008749,   Adjusted R-squared:  0.004513
## F-statistic: 2.065 on 1 and 234 DF,  p-value: 0.152
```

```
m3 = lm(leases ~ log(discount+1), data = movie)
summary(m3)
```

```
##
## Call:
## lm(formula = leases ~ log(discount + 1), data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7778 -0.6449 -0.5120  0.4880  4.2518
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.51196    0.08270   6.191 2.66e-09 ***
```

```
## log(discount + 1)  0.05542    0.03005   1.844   0.0664 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9079 on 234 degrees of freedom
## Multiple R-squared:  0.01433,    Adjusted R-squared:  0.01011
## F-statistic: 3.401 on 1 and 234 DF,  p-value: 0.06641
```

```
# discount does have a positive effect on leases, but marginally significant (again, due to small sampl

# Does treatment effect vary with base price?
m4 = lm(leases ~ has_discount + base_price + has_discount*base_price, data = movie)
summary(m4)
```

```
##
## Call:
## lm(formula = leases ~ has_discount + base_price + has_discount *
##     base_price, data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9586 -0.6003 -0.4827  0.4260  4.2564
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.6950289  0.3832371   1.814    0.071 .
## has_discount              0.4770672  0.5522508   0.864    0.389
## base_price               -0.0006084  0.0012197  -0.499    0.618
## has_discount:base_price -0.0008246  0.0017513  -0.471    0.638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9086 on 232 degrees of freedom
## Multiple R-squared:  0.02117,    Adjusted R-squared:  0.008517
## F-statistic: 1.673 on 3 and 232 DF,  p-value: 0.1736
```

```
m5 = lm(leases ~ log(discount+1) + base_price + log(discount+1)*base_price, data = movie)
summary(m5)
```

```
##
## Call:
## lm(formula = leases ~ log(discount + 1) + base_price + log(discount +
##     1) * base_price, data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9221 -0.6238 -0.4760  0.3923  4.2420
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  7.823e-01  3.819e-01   2.048   0.0416 *
## log(discount + 1)            7.102e-02  1.332e-01   0.533   0.5945
## base_price                  -8.776e-04  1.215e-03  -0.723   0.4707
```

```
## log(discount + 1):base_price -5.077e-05  4.242e-04  -0.120   0.9048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9093 on 232 degrees of freedom
## Multiple R-squared:  0.01966,    Adjusted R-squared:  0.006984
## F-statistic: 1.551 on 3 and 232 DF,  p-value: 0.2021
```

```
# What can we conclude?
# Nothing! Don't draw conclusions from null results...

# Does treatment effect vary with movie popularity?
m6 = lm(leases ~ has_discount + likes + has_discount*likes, data = movie)
summary(m6)
```

```
##
## Call:
## lm(formula = leases ~ has_discount + likes + has_discount * likes,
##     data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9260 -0.7199 -0.3336  0.2810  4.2661
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.493e-01  1.310e-01   1.903   0.0583 .
## has_discount        4.688e-01  1.878e-01   2.496   0.0133 *
## likes               1.106e-07  4.334e-08   2.552   0.0113 *
## has_discount:likes -1.060e-07  6.289e-08  -1.685   0.0933 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8991 on 232 degrees of freedom
## Multiple R-squared:  0.04159,    Adjusted R-squared:  0.0292
## F-statistic: 3.356 on 3 and 232 DF,  p-value: 0.01964
```

```
m7 = lm(leases ~ log(discount+1) + likes + log(discount+1)*likes, data = movie)
summary(m7)
```

```
##
## Call:
## lm(formula = leases ~ log(discount + 1) + likes + log(discount +
##     1) * likes, data = movie)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9366 -0.7054 -0.3348  0.3559  4.2619
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.491e-01  1.295e-01   1.923  0.05574 .
## log(discount + 1)        1.219e-01  4.766e-02   2.558  0.01117 *
```

```
## likes                    1.124e-07  4.293e-08   2.618  0.00943 **
## log(discount + 1):likes -2.847e-08  1.599e-08  -1.780  0.07642 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8986 on 232 degrees of freedom
## Multiple R-squared:  0.04261,    Adjusted R-squared:  0.03023
## F-statistic: 3.442 on 3 and 232 DF,  p-value: 0.01754
```

```
# What can we conclude?
# If a movie is really good, "I don't care what it costs!"
# The strength of the moderation, however, is pretty small from a practical point of view.
```