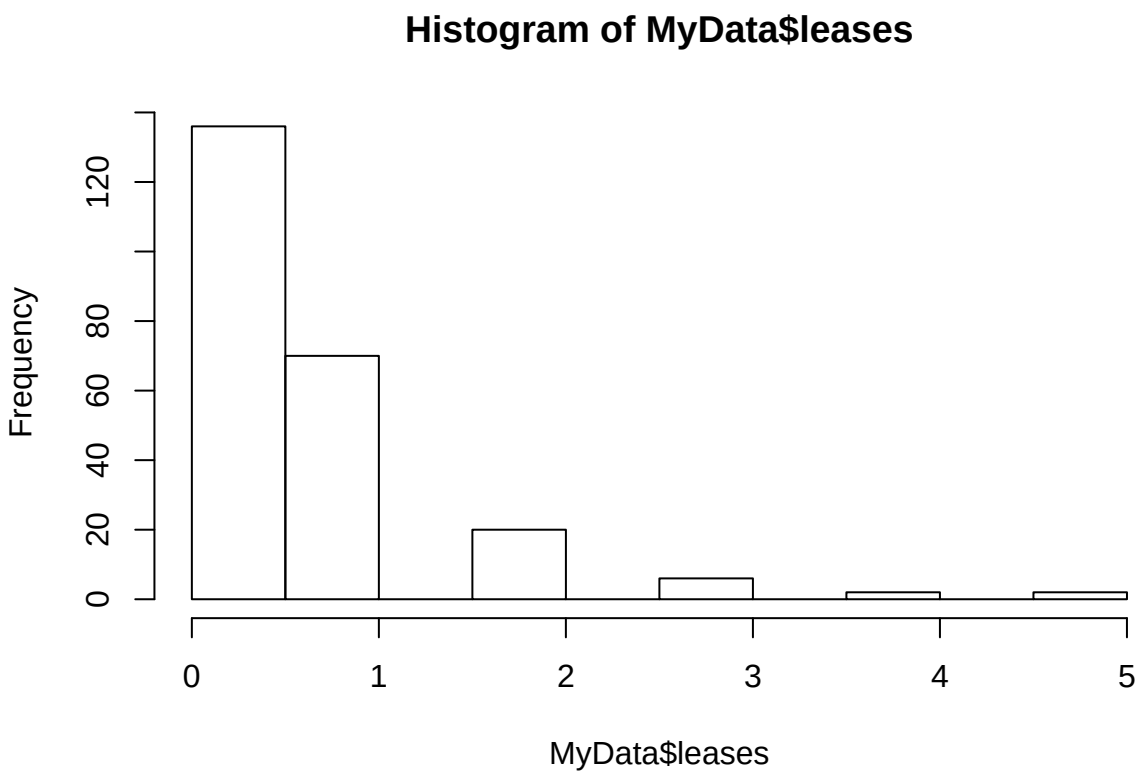
MoviesDemand-Exp-Feb_18-2.R

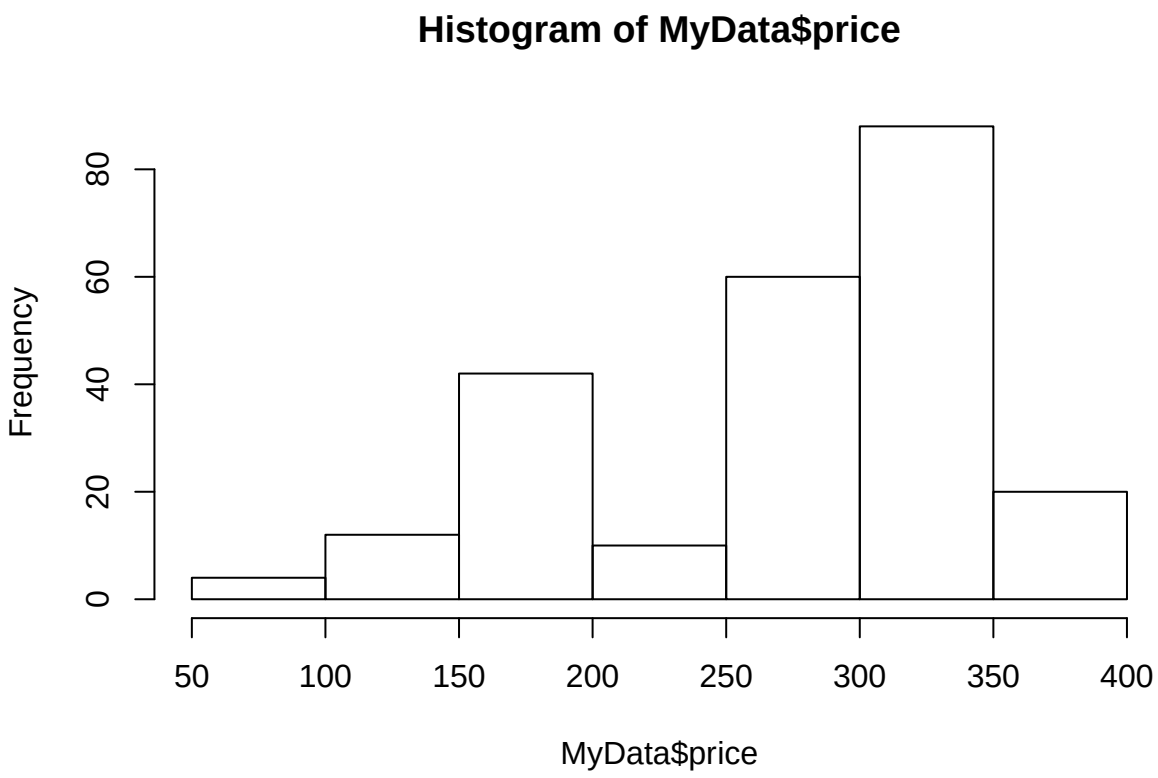
danny

2020-02-18

```
suppressWarnings(suppressPackageStartupMessages({
library(data.table)
library(stargazer)
library(ggplot2)
library(MESS)
} ))

#*** MSBA 6440 ***#
#*** Gordon Burtch and Gautam Ray***#
#*** Updated Feb 2020 ***#
#*** Code for Lecture 3 ***#

# Analyzing Movie Rental Pricing Experiment Data

#*** Load Dataset ***#
MyData<- read.csv("MovieData-Exp.csv")
View(MyData)

# Descriptive statistics / plots...
hist(MyData$leases)
```




```
hist(MyData$price)
```




```
# Let's make a treatment dummy to keep things simple for now.
# This helps us do some easy randomization checks.
# Let's also construct the discount variable.
MyData$disc <- MyData$base_price - MyData$price
summary(MyData[MyData$disc>0,]$disc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   30.00   50.00   52.88   70.00  120.00
```

```
MyData$treated <- (MyData$disc > 0)

# Let's check randomization...
t.test(likes~treated,data=MyData)
```

```
##
##  Welch Two Sample t-test
##
## data:  likes by treated
## 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 FALSE  mean in group TRUE
##             2343120             2328429
```

```r
t.test(base_price~treated,data=MyData)
```

```
##
##  Welch Two Sample t-test
##
## data:  base_price by treated
## 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 FALSE  mean in group TRUE
##            306.6271            309.3390
```

```r
summary(lm(disc~log(likes), data=MyData))
```

```
##
## Call:
## lm(formula = disc ~ log(likes), data = MyData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -27.73 -26.82 -17.32  25.80  94.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.4814    15.5443   0.996    0.320
## log(likes)    0.7839     1.1019   0.711    0.478
##
## Residual standard error: 31.93 on 234 degrees of freedom
## Multiple R-squared:  0.002158,   Adjusted R-squared:  -0.002106
## F-statistic: 0.5061 on 1 and 234 DF,  p-value: 0.4775
```

```r
summary(lm(disc~log(base_price), data=MyData))
```

```
##
## Call:
## lm(formula = disc ~ log(base_price), data = MyData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -28.65 -25.61 -20.67  24.84  93.55
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       57.299     47.616   1.203    0.230
## log(base_price)   -5.412      8.343  -0.649    0.517
##
## Residual standard error: 31.93 on 234 degrees of freedom
## Multiple R-squared:  0.001795,   Adjusted R-squared:  -0.002471
## F-statistic: 0.4208 on 1 and 234 DF,  p-value: 0.5172
```

```r
# Let's evaluate statistical power now.
# Do we have enough data? Remember, we have 0.5 leases per movie-week on average prior
# to the experiment taking place, for this set of customers.
# Management wants to know about a 20% increase with 90% confidence.
# Thus, we need to detect an increase of 0.10 leases per movie in the week of the experiment.
# That is, 0.50 * 20% = 0.10. This is our delta parameter.
# 90% confidence implies an alpha of 0.10 (1 - 0.9 = 0.1).
# We assume a power of 80% absent other information.

# The first power test tells us what sort of difference we can reliably detect with our current
# sample size... 118 movies per group.
power_t_test(n=118,type=c("two.sample"),alternative="two.sided",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
```

```r
# The second tells 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"),alternative="two.sided",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
```

```r
# Note: we appear to be heavily underpowered to detect the effect management is looking for.
# I would thus caution management about reading too much into results from this experiment.
# I might even advise repeating it with the bigger, requisite sample.

# That said, moving on...
# Let's estimate the treatment effect.

#*** OLS of leases on price and log(price) ***#
ols <- lm(leases ~ price, data = MyData)
olslog <- lm(leases ~ log(price), data = MyData)
stargazer(ols,olslog,title="OLS leases on prices and log(price)",type="text",column.labels=c("price","l
```

```
##
## OLS leases on prices and log(price)
## ===============================================================
##                                    Dependent variable:
##                                ---------------------------
##                                           leases
##                                    price       log(price)
##                                     (1)            (2)
## ---------------------------------------------------------------
## price                               -0.001
##                                    (0.001)
##
## log(price)                                          -0.255
##                                                     (0.177)
##
## Constant                           0.973***        2.043**
##                                    (0.229)         (0.994)
##
## ---------------------------------------------------------------
## Observations                         236             236
## R2                                  0.011           0.009
## Adjusted R2                         0.007           0.005
## Residual Std. Error (df = 234)      0.909           0.910
## F Statistic (df = 1; 234)           2.571           2.062
## ===============================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

```
#*** OLS of leases on price and log(price) with additional controls***#
olslogcontrols <- lm(leases ~ log(price) + log(likes), data = MyData)
stargazer(ols,olslog,olslogcontrols,title="OLS leases on prices, log(price) and controls",type="text",c
```

```
##
## OLS leases on prices, log(price) and controls
## ================================================================================
##                                          Dependent variable:
##                      ---------------------------------------------------------
##                                               leases
##                             price            log(price)         with controls
##                              (1)                 (2)                 (3)
## --------------------------------------------------------------------------------
## price                      -0.001
##                            (0.001)
##
## log(price)                                     -0.255              -0.266
##                                                (0.177)             (0.177)
##
## log(likes)                                                          0.043
##                                                                    (0.031)
##
## Constant                  0.973***             2.043**             1.506
##                            (0.229)             (0.994)             (1.066)
##
## --------------------------------------------------------------------------------
## yr dummies                   No                  No                  No
```

```
## ------------------------------------------------------------------------------
## Observations                236                 236                 236
## R2                         0.011               0.009               0.017
## Adjusted R2                0.007               0.005               0.008
## Residual Std. Error  0.909 (df = 234)    0.910 (df = 234)    0.909 (df = 233)
## F Statistic         2.571 (df = 1; 234) 2.062 (df = 1; 234) 1.978 (df = 2; 233)
## ==============================================================================
## Note:                                            *p<0.1; **p<0.05; ***p<0.01
```

```
# Hmmm... something is wrong!
# WAIT!!! We can't just look at price...
# Not all of the variation in price is from our experiment!
# The variation across movies in base-price is endogenous...
# We need to focus just on the price discount treatment itself...
t.test(leases~treated, data=MyData)
```

```
##
##  Welch Two Sample t-test
##
## data:  leases by treated
## t = -1.8645, df = 217.08, p-value = 0.0636
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.45326122  0.01258325
## sample estimates:
## mean in group FALSE  mean in group TRUE
##           0.5084746           0.7288136
```

```
ols_treat <- lm(leases ~ treated, data = MyData)
ols_log_discount <- lm(leases ~ log(disc+1), data = MyData)

# Does a positive coefficient make sense? Yes, discount is amount of money removed from price.
stargazer(ols_treat,ols_log_discount,type="text",column.labels=c("Binary","Log Discount"))
```

```
##
## ===============================================================
##                                   Dependent variable:
##                               ----------------------------
##                                          leases
##                                   Binary     Log Discount
##                                     (1)            (2)
## ---------------------------------------------------------------
## treated                            0.220*
##                                   (0.118)
##
## log(disc + 1)                                     0.055*
##                                                   (0.030)
##
## Constant                          0.508***       0.512***
##                                   (0.084)        (0.083)
##
## ---------------------------------------------------------------
## Observations                        236            236
```

```
## R2                                0.015          0.014
## Adjusted R2                       0.010          0.010
## Residual Std. Error (df = 234)    0.908          0.908
## F Statistic (df = 1; 234)         3.476*         3.401*
## ============================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

```
# What sort of heterogeneity might we look at here? And how?
# Let's check out base price.
ols_moderated_base <- lm(leases ~ treated*base_price, data=MyData)
ols_log_disc_moderated_base <- lm(leases ~ log(disc+1)*base_price, data=MyData)
stargazer(ols_moderated_base,ols_log_disc_moderated_base, type="text",column.labels=c("Treated Moderate
```

```
##
## ===============================================================================================
##                                                Dependent variable:
##                                    ---------------------------------------------------------
##                                                       leases
##                                    Treated Moderatedy by Base Price Disc Moderated by Base Price
##                                                  (1)                              (2)
## -----------------------------------------------------------------------------------------------
## treated                                         0.477
##                                                (0.552)
##
## log(disc + 1)                                                                    0.071
##                                                                                 (0.133)
##
## base_price                                      -0.001                           -0.001
##                                                (0.001)                          (0.001)
##
## treatedTRUE:base_price                          -0.001
##                                                (0.002)
##
## log(disc + 1):base_price                                                        -0.0001
##                                                                                 (0.0004)
##
## Constant                                        0.695*                          0.782**
##                                                (0.383)                          (0.382)
##
## -----------------------------------------------------------------------------------------------
## Observations                                     236                              236
## R2                                              0.021                            0.020
## Adjusted R2                                     0.009                            0.007
## Residual Std. Error (df = 232)                  0.909                            0.909
## F Statistic (df = 3; 232)                       1.673                            1.551
## ===============================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```

```
# Why does the treatment effect disappear? Because it's the effect of treatment when
# base price = 0... this never actually occurs in the data!
# Let's shift the base price variable so it is mean 0.
# Then, the coefficient on treatment's main effect reflects treatment on the average movie.
MyData$log_base_price_demean <- log(MyData$base_price)-mean(log(MyData$base_price))
```

```r
ols_moderated_dm <- lm(leases ~ treated*log_base_price_demean, data=MyData)
ols_log_disc_moderated_dm <- lm(leases ~ log(disc+1)*log_base_price_demean, data=MyData)
stargazer(ols_moderated_dm,ols_log_disc_moderated_dm, type="text",column.labels=c("Base Price Moderator
```

```
##
## ===============================================================================
##                                            Dependent variable:
##                                    -----------------------------------------
##                                                    leases
##                                    Base Price Moderator De-Meaned
##                                              (1)                  (2)
## -------------------------------------------------------------------------------
## treated                                    0.223*
##                                           (0.118)
##
## log(disc + 1)                                                    0.055*
##                                                                 (0.030)
##
## log_base_price_demean                      -0.141               -0.227
##                                           (0.336)              (0.335)
##
## treatedTRUE:log_base_price_demean          -0.289
##                                           (0.475)
##
## log(disc + 1):log_base_price_demean                              -0.024
##                                                                 (0.115)
##
## Constant                                  0.508***             0.513***
##                                           (0.084)              (0.083)
##
## -------------------------------------------------------------------------------
## Observations                                236                  236
## R2                                         0.022                0.020
## Adjusted R2                                0.010                0.008
## Residual Std. Error (df = 232)             0.908                0.909
## F Statistic (df = 3; 232)                  1.765                1.593
## ===============================================================================
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
##
## ======================================
## Log Disc Base Price Moderated De-Meaned
## --------------------------------------
```

```r
# Nope, the effect doesn't seem to be moderated by baseline price.
# You can try it with a log transformation and you'll come to the same conclusion.
# What can we conclude?
# Nothing! Don't draw conclusions from null results...

# Try doing the same thing with likes...
MyData$likes_demean <- MyData$likes-mean(MyData$likes)
ols_moderated_likes_dm <- lm(leases ~ treated*likes_demean, data=MyData)
ols_log_disc_moderated__likes_dm <- lm(leases ~ log(disc+1)*likes_demean, data=MyData)
stargazer(ols_moderated_likes_dm,ols_log_disc_moderated__likes_dm, type="text",column.labels=c("Like Mo
```

```
##
## ================================================================================================
##                                                  Dependent variable:
##                                     ---------------------------------------------------------
##                                                          leases
##                                     Like Moderator De-Meaned Log Price, Like Moderator De-Mean
##                                               (1)                             (2)
## ------------------------------------------------------------------------------------------------
## treated                                      0.221*
##                                             (0.117)
##
## log(disc + 1)                                                                0.055*
##                                                                             (0.030)
##
## likes_demean                               0.00000**                       0.00000***
##                                            (0.00000)                       (0.00000)
##
## treatedTRUE:likes_demean                   -0.00000*
##                                            (0.00000)
##
## log(disc + 1):likes_demean                                                  -0.00000*
##                                                                             (0.00000)
##
## Constant                                     0.508***                        0.512***
##                                              (0.083)                         (0.082)
##
## ------------------------------------------------------------------------------------------------
## Observations                                   236                             236
## R2                                             0.042                           0.043
## Adjusted R2                                    0.029                           0.030
## Residual Std. Error (df = 232)                 0.899                           0.899
## F Statistic (df = 3; 232)                     3.356**                         3.442**
## ================================================================================================
## Note:                                                        *p<0.1; **p<0.05; ***p<0.01
```

```r
# We do find that popular movies respond less strongly to the discount treatment.
# This makes some sense... if a movie is really good, "I don't care what it costs!"
# The strength of the moderation is pretty weak, however, in practical terms...
# We are going out to many significant digits... you can try the log transform here,
# But a better option might also be to just rescale the variable (e.g., 1,000's of likes)

MyData$likes_demean <- MyData$likes/1000000-mean(MyData$likes/1000000)
ols_moderated_likes_dm <- lm(leases ~ treated*likes_demean, data=MyData)
ols_log_disc_moderated__likes_dm <- lm(leases ~ log(disc+1)*likes_demean, data=MyData)
stargazer(ols_moderated_likes_dm,ols_log_disc_moderated__likes_dm, type="text",column.labels=c("Like Mo
```

```
##
## ================================================================================================
##                                                  Dependent variable:
##                                     ---------------------------------------------------------
##                                                          leases
##                                     Like Moderator De-Meaned Log Disc, Like Moderator De-Mean
##                                               (1)                             (2)
## ------------------------------------------------------------------------------------------------
```

```
## treated                                0.221*
##                                       (0.117)
##
## log(disc + 1)                                                  0.055*
##                                                               (0.030)
##
## likes_demean                           0.111**                 0.112***
##                                       (0.043)                 (0.043)
##
## treatedTRUE:likes_demean               -0.106*
##                                       (0.063)
##
## log(disc + 1):likes_demean                                     -0.028*
##                                                               (0.016)
##
## Constant                               0.508***                0.512***
##                                       (0.083)                 (0.082)
##
## ------------------------------------------------------------------------------------
## Observations                             236                     236
## R2                                      0.042                   0.043
## Adjusted R2                             0.029                   0.030
## Residual Std. Error (df = 232)          0.899                   0.899
## F Statistic (df = 3; 232)              3.356**                 3.442**
## ====================================================================================
## Note:                                                *p<0.1; **p<0.05; ***p<0.01
```