Nicholas L. Brown; Peipei Cai; David Forteguerre; Jiaming Guo; Almantas Palubinskas
Professor A. Basu
MAR653
10/10/2018

50/50

Group assignment #1

1) (a) Florida Gold was on sale 37.40% of the time, Minute Maid was on sale 32.13% of the time, and Tropicana was on sale 34.38% of the time.

Count of Feat	Column Labels		
Row Labels	0	1	Grand Total
FG	62.60%	37.40%	100.00%
MINMAID	67.88%	32.13%	100.00%
TROPICANA	65.63%	34.38%	100.00%
Grand Total	65.37%	34.63%	100.00%

✓

1) (b) For each season, the percentage of time each brand was on sale was:

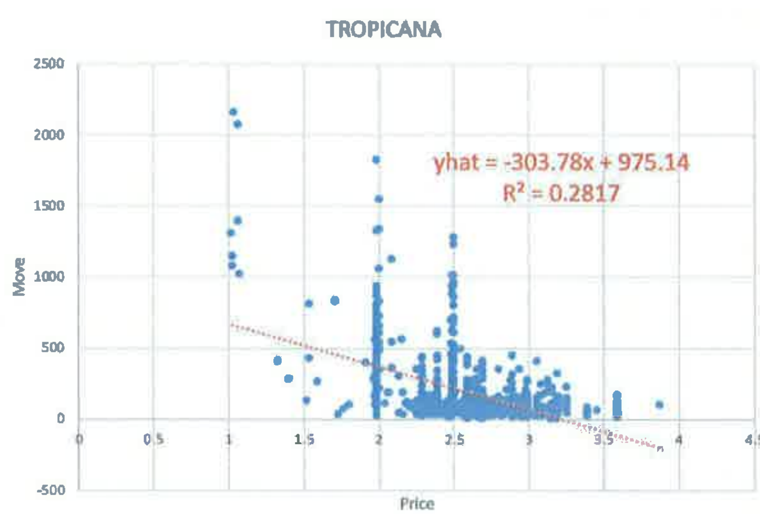
Count of CASEID	Column Labels		
Row Labels	0	1	Grand Total
FG	62.60%	37.40%	100.00%
Fall	66.40%	33.60%	100.00%
Spring	63.97%	36.03%	100.00%
Summer	62.84%	37.16%	100.00%
Winter	57.66%	42.34%	100.00%
MINMAID	67.88%	32.13%	100.00%
Fall	64.90%	35.10%	100.00%
Spring	66.60%	33.40%	100.00%
Summer	74.26%	25.74%	100.00%
Winter	66.41%	33.59%	100.00%
TROPICANA	65.63%	34.38%	100.00%
Fall	59.75%	40.25%	100.00%
Spring	72.74%	27.26%	100.00%
Summer	68.11%	31.89%	100.00%
Winter	61.83%	38.17%	100.00%
Grand Total	65.37%	34.63%	100.00%

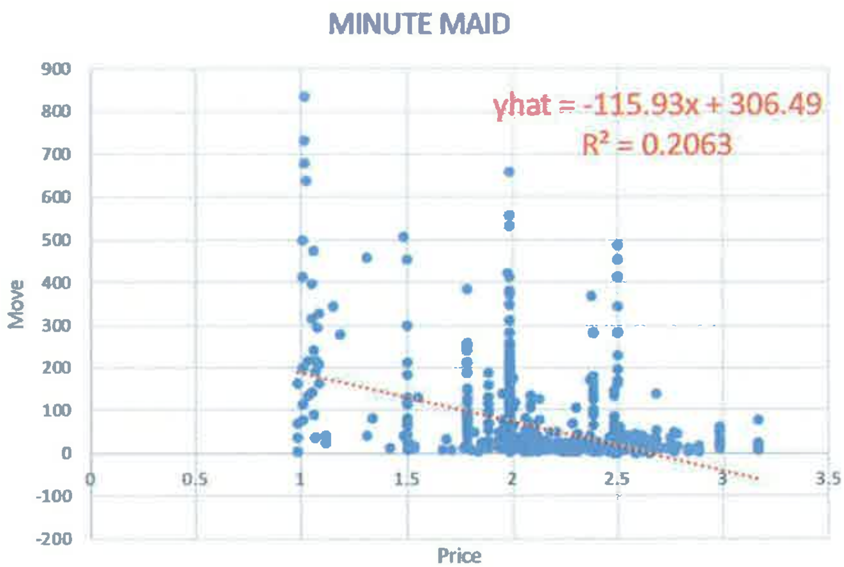
✓

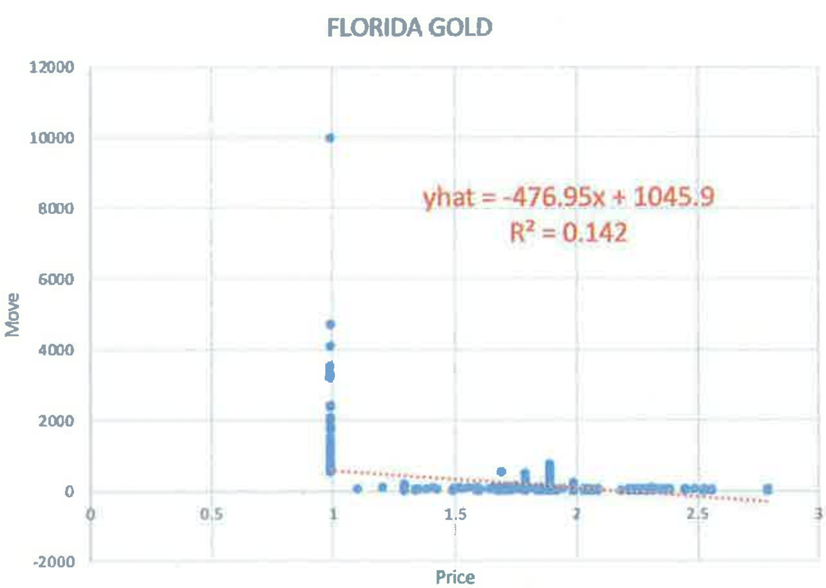
1) (c) For each season, the average price of each brand when it was on sale (Feat=1 column) and when it was not on sale (Feat=0 column) was:

Average of PRICE	Column Labels		
Row Labels	0	1	Grand Total
FG	**2.190626997**	**1.909926471**	**2.085645**
Fall	2.222030303	2.06754491	2.170120724
Spring	2.203855422	1.749545455	2.040163776
Summer	2.237007299	1.981080247	2.141915138
Winter	2.103718354	1.876056034	2.007335766
MINMAID	**2.336313076**	**1.961299611**	**2.21584**
Fall	2.341417565	1.841509972	2.16595
Spring	2.37832853	1.989252874	2.248387716
Summer	2.341622419	1.981829787	2.249014239
Winter	2.284337176	2.03962963	2.202143541
TROPICANA	**2.849382857**	**2.38224**	**2.6888025**
Fall	2.829561404	2.5228125	2.706090147
Spring	2.886899736	2.268767606	2.718426104
Summer	2.884522293	2.306836735	2.700314534
Winter	2.79077728	2.383244552	2.635221811
Grand Total	**2.461505609**	**2.08207411**	**2.330095833**

2) For this question, we used filter and created a new worksheet that only included move and price for each brand during winter. Then, we used scatter plots to plot move against unit price and used trendline to estimate the regression parameters.
Below are the scatter plots and the associated regression parameters.




MINUTE MAID
yhat = -115.93x + 306.49
R^2 = 0.2063
Move
Price
900
800
700
600
500
400
300
200
100
0
-100
-200
0
0.5
1
1.5
2
2.5
3
3.5


FLORIDA GOLD
yhat = -476.95x + 1045.9
R^2 = 0.142
Move
Price
12000
10000
8000
6000
4000
2000
0
-2000
0
0.5
1
1.5
2
2.5
3

3) (a) Based on our output, the price elasticity of demand of each brand is:

1. **Florida Gold:** -2.86563
2. **Minute Maid:** -2.86563 - 0.03421= -2.89984
3. **Tropicana:** -2.86563 + 0.59291 = -2.27272

Output:

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   -4.55233    1.24249  -3.664  0.00025 ***
logPRICE                      -2.86563    0.07170 -39.965  < 2e-16 ***
BRAND[T.MINMAID]               0.31546    0.07512   4.200 2.69e-05 ***
BRAND[T.TROPICANA]             1.71789    0.09256  18.560  < 2e-16 ***
Season[T.Spring]               0.09822    0.02296   4.278 1.90e-05 ***
Season[T.Summer]              -0.05587    0.02374  -2.354  0.01861 *  
Season[T.Winter]               0.10012    0.02279   4.394 1.12e-05 ***
Feat                           0.52766    0.01873  28.166  < 2e-16 ***
AGE9                           1.16234    0.99554   1.168  0.24301    
AGE60                          3.02475    0.38929   7.770 8.49e-15 ***
EDUC                           1.00126    0.14936   6.704 2.12e-11 ***
ETHNIC                         0.09843    0.10530   0.935  0.34993    
POVERTY                        1.71746    0.99036   1.734  0.08291 .  
SINGLE                         0.98018    0.50306   1.948  0.05138 .  
INCOME                         0.74235    0.10968   6.768 1.37e-11 ***
NOCAR                          1.33548    0.27994   4.771 1.86e-06 ***
logPRICE:BRAND[T.MINMAID]     -0.03421    0.09721  -0.352  0.72494    
logPRICE:BRAND[T.TROPICANA]    0.59291    0.10391   5.706 1.19e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8899 on 11982 degrees of freedom
Multiple R-squared:  0.5584,	Adjusted R-squared:  0.5577 
F-statistic: 891.1 on 17 and 11982 DF,  p-value: < 2.2e-16
```

3) (b) (i) From the regression output in 3(a), the demographic variables that are not significant at a 90% level of confidence are **AGE9** and **ETHNIC**, as their p-values are greater than 0.1. Let's now test the null hypothesis that the coefficients of these demographic variables are all zeros:

H_0: $\beta_{AGE9} = \beta_{ETHNIC} = 0$

H_A : β_{AGE9} *or* β_{ETHNIC} *does not equal* 0.

```
Linear hypothesis test

Hypothesis:
AGE9 = 0
ETHNIC = 0

Model 1: restricted model
Model 2: logMOVE ~ logPRICE + BRAND + Season + BRAND * logPRICE + Feat + 
    AGE9 + AGE60 + EDUC + ETHNIC + INCOME + SINGLE + POVERTY + 
    NOCAR

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1  11984 9491.6                           
2  11982 9489.1  2    2.5444 1.6064 0.2007
```

We fail to reject the null hypothesis at a 99% level of confidence (P-value for the F-statistic: 0.20 > P-value = 0.01).

3) (b) (ii) We will now test the null hypothesis that the price elasticity of demand is the same for all three brands at a 99% level of confidence.

$H_0 : \beta_{LogPrice*Florida\ Gold} = \beta_{LogPrice*Minute\ Maid} = \beta_{LogPrice*Tropicana}$

$H_A : \beta_{LogPrice*Florida\ Gold} \neq \beta_{LogPrice*Minute\ Maid} \neq \beta_{LogPrice*Tropicana}$ } At least one is different

```
Linear hypothesis test

Hypothesis:
logPRICE:BRAND[T.MINMAID] = 0
logPRICE:BRAND[T.TROPICANA] = 0

Model 1: restricted model
Model 2: logMOVE ~ logPRICE + BRAND + Season + BRAND * logPRICE + Feat +
    AGE9 + AGE60 + EDUC + ETHNIC + INCOME + SINGLE + POVERTY +
    NOCAR

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  11984 9524.7
2  11982 9489.1  2    35.673 22.523 1.725e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the null hypothesis at a 99% level of confidence, as P-value < 0.01.

3) (b) (iii) Let's now test the null hypothesis that the price elasticity of demand is same for Florida Gold and Minute Maid at a 99% level of confidence.

$H_0 : \beta_{LogPrice*Florida\ Gold} = \beta_{LogPrice*Minute\ Maid}$

$H_A : \beta_{LogPrice*Florida\ Gold} \neq \beta_{LogPrice*Minute\ Maid}$

```
Linear hypothesis test

Hypothesis:
logPRICE:BRAND[T.MINMAID] = 0

Model 1: restricted model
Model 2: logMOVE ~ logPRICE + BRAND + Season + BRAND * logPRICE + Feat +
    AGE9 + AGE60 + EDUC + ETHNIC + INCOME + SINGLE + POVERTY +
    NOCAR

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1  11983 9489.2
2  11982 9489.1  1  0.098051 0.1238 0.7249
```

We fail to reject the null hypothesis at a 99% level of confidence, as P-value = 0.7249 > 0.01.

3) (b) (iv) We now test the null hypothesis that the price elasticity of demand is same for Minute Maid and Tropicana at a 99% level of confidence.

$H_0 : \beta_{LogPrice*Minute\ Maid} = \beta_{LogPrice*Tropicana}$

$H_A : \beta_{LogPrice*Minute\ Maid} \neq \beta_{LogPrice*Tropicana}$

```
Linear hypothesis test

Hypothesis:
logPRICE:BRAND[T.MINMAID] - logPRICE:BRAND[T.TROPICANA] = 0

Model 1: restricted model
Model 2: logMOVE ~ logPRICE + BRAND + Season + BRAND * logPRICE + Feat +
    AGE9 + AGE60 + EDUC + ETHNIC + INCOME + SINGLE + POVERTY +
    NOCAR

  Res.Df    RSS Df Sum of Sq      F        Pr(>F)
1  11983 9518.7
2  11982 9489.1  1    29.591 37.365 0.00000000101 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the null hypothesis at a 99% level of confidence, as P-value < 0.01.

3) (c) In this question, we fit a logit model with Feat as the dependent variable, and Brand and Season as independent variables.

Output:

```
glm(formula = Feat ~ BRAND + Season, family = binomial(logit),
    data = dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0255  -0.9359  -0.8668   1.3945   1.5695

Coefficients:
                    Estimate Std. Error z value   Pr(>|z|)
(Intercept)         -0.44539    0.04655  -9.568    < 2e-16 ***
BRAND[T.MINMAID]    -0.23037    0.04714  -4.887 0.00000102 ***
BRAND[T.TROPICANA]  -0.12890    0.04673  -2.758   0.005808 **
Season[T.Spring]    -0.17982    0.05423  -3.316   0.000913 ***
Season[T.Summer]    -0.21097    0.05645  -3.737   0.000186 ***
Season[T.Winter]     0.07705    0.05282   1.459   0.144605
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 15484  on 11999  degrees of freedom
Residual deviance: 15419  on 11994  degrees of freedom
AIC: 15431

Number of Fisher Scoring iterations: 4


Rcmdr>  exp(coef(GLM.6))  # Exponentiated coefficients ("odds ratios")
      (Intercept)   BRAND[T.MINMAID] BRAND[T.TROPICANA]   Season[T.Spring]   Season[T.Summer]   Season[T.Winter]
        0.6405752          0.7942411          0.8790605          0.8354217          0.8098016          1.0800956
```

Estimated indicator function:

I = -0.44539 -0.23037* MINMAID -0.12890*TROPICANA -0.17982*Spring -0.21097*Summer +0.07705*Winter

Interpretation:

Florida Gold has the highest probability of being on sale. The second most likely brand to be on sale when the is Tropicana. Minute Maid has the lowest probability to be on sale.

Orange juice has the highest probability of being on sale during the winter, the second highest probability of being on sale during the fall, the third highest probability of orange juice being on sale is during the spring, and the lowest probability that orange juice will be on sale is during the summer.

As the weather gets colder, there is an increased probability that orange juice will be on sale.

3) Let's test the following null hypotheses at a 99% level of confidence:

3) (d) (i) "A brand is equally likely to be on sale (Feat = 1) in all four seasons."

$H_0 : \beta_{Season[winter]} = \beta_{Season[Spring]} = \beta_{Season[Summer]} = 0$

H_A : *At least one of* $\beta_{Season[winter]}, \beta_{Season[Spring]}, \beta_{Season[Summer]}$ *does not equal* 0.

```
Linear hypothesis test

Hypothesis:
Season[T.Spring] = 0
Season[T.Summer] = 0
Season[T.Winter] = 0

Model 1: restricted model
Model 2: Feat ~ BRAND + Season

  Res.Df Df  Chisq    Pr(>Chisq)
1  11997
2  11994  3 39.524 0.00000001344 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the null hypothesis at a 99% level of confidence, as P-value<0.01.

3) (d) (ii) "Season being the same, Minute Maid and Tropicana are equally likely to be on sale."

$H_0 : \beta_{Brand[MinMaid]} = \beta_{Brand[Tropicana]}$

$H_A : \beta_{Brand[MinMaid]} \neq \beta_{Brand[Tropicana]}$

```
Linear hypothesis test

Hypothesis:
BRAND[T.MINMAID] - BRAND[T.TROPICANA] = 0

Model 1: restricted model
Model 2: Feat ~ BRAND + Season

  Res.Df Df  Chisq Pr(>Chisq)
1  11995
2  11994  1 4.5506    0.03291 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We fail to reject the null hypothesis at a 99% level of confidence, as P-value>0.01.

4) In this question, we fit a regression model with dependent variable log of move, and independent variables BRAND, Feat, logPRICE and BRAND*logPRICE. For the six cases provided, we used R to construct 99% prediction intervals for logMOVE.

Output:

```
> problem4<-lm(logMOVE~BRAND+Feat+logPRICE+BRAND*logPRICE, X653f18hw1_1_)
> problem4

Call:
lm(formula = logMOVE ~ BRAND + Feat + logPRICE + BRAND * logPRICE,
    data = X653f18hw1_1_)

Coefficients:
            (Intercept)              BRANDMINMAID          BRANDTROPICANA
                4.65134                   0.33057                 1.65267
                   Feat                  logPRICE   BRANDMINMAID:logPRICE
                0.57350                  -2.66954                -0.06498
BRANDTROPICANA:logPRICE
                0.61335

> predict(problem4, interval = "prediction", level = .99,   newdata = Problem4data)
       fit       lwr      upr
1 3.623111 1.2373048 6.008916
2 2.649181 0.2635486 5.034813
3 3.750622 1.3647951 6.136449
4 2.684909 0.2992835 5.070534
5 5.232549 2.8466188 7.618479
6 4.145003 1.7593565 6.530649
```

Case	BRAND	logPRICE	Feat
1	FG	.6	1
2	FG	.75	0
3	MINMAID	.66	1
4	MINMAID	.84	0
5	TROPICANA	.80	1
6	TROPICANA	1.05	0