1) 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

INCOME 0.74235 0.10968 6.768 1.37e-11 \*\*\*

NOCAR 1.33548 0.27994 4.771 1.86e-06 \*\*\*

SINGLE 0.98018 0.50306 1.948 0.05138 .

POVERTY 1.71746 0.99036 1.734 0.08291 .

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 \*\*\*

1(a) Price elasticity of demand of MINMAID=logprice-logprice:BRAND[T.MINMAID]=-2.86563-0.03421=-2.89984

Price elasticity of demand of TROPICANA=logprice-logprice:BRAND[T.TROPICANA]=-2.86563+0.59291=-2.27272

Price elasticity of demand of FG=logprice=-2.86563

1(b) Not significant at a 90% level of confidence -> p-value >0.1

These variables that are not significant at a 90% level of confidence are: AGE9 and ETHNIC

Test H0:

Method 1 Output:

Linear hypothesis test

Hypothesis:

AGE9 = 0

ETHNIC = 0

Model 1: restricted model

Model 2: logmove ~ logprice + Feat + BRAND + Season + BRAND \* logprice +

AGE9 + AGE60 + EDUC + ETHNIC + INCOME + NOCAR + POVERTY +

SINGLE

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

**Method 2 Output:**

Analysis of Variance Table

Model 1: logmove ~ logprice + Feat + BRAND + Season + BRAND \* logprice +

AGE9 + AGE60 + EDUC + ETHNIC + INCOME + NOCAR + POVERTY +

SINGLE

Model 2: logmove ~ logprice + Feat + BRAND + Season + BRAND \* logprice +

AGE60 + EDUC + INCOME + NOCAR + POVERTY + SINGLE

Res.Df RSS Df Sum of Sq F Pr(>F)

1 11982 9489.1

2 11984 9491.6 -2 -2.5444 1.6064 0.2007

**Pr(>F) is greater than 0.01, failed to reject the null hypothesis that the coefficients of AGE9 and ETHNIC are all zeros at a 99% level of confidence.**

**1 (b)(ii).**

local({

+ .Hypothesis <- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,

+ 0,0,0,0,0,0,0,0,0,0,1), 2, 18, byrow=TRUE)

+ .RHS <- c(0,0)

+ linearHypothesis(LinearModel.10, .Hypothesis, rhs=.RHS)

+ })

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 +

AGE60 + EDUC + INCOME + NOCAR + SINGLE + POVERTY + AGE9 +

ETHNIC

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

P-value<0.01, reject null hypothesis that the price elasticity of demand is same for all three brands at 99% level of confidence.

**1 (b)(iii)**

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 + NOCAR + SINGLE +

POVERTY

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

P-value>0.01, failed to reject null hypothesis that the price elasticity of demand is same for Florida Gold and Minute Maid

**1 (c)**

**Fit a logic model:**

Call:

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

data = orange)

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

> exp(coef(GLM.2)) # Exponentiated coefficients ("odds ratios")

(Intercept) BRAND[T.MINMAID] BRAND[T.TROPICANA] Season[T.Spring]

0.6405752 0.7942411 0.8790605 0.8354217

Season[T.Summer] Season[T.Winter]

0.8098016 1.0800956

I=-0.44539-0.23037MINMAID-0.1289TROPICANA-0.17982Spring-0.21097Summer+0.07705Winter

The larger the I, the larger P(Y=1)

**Brand in Fall:**

**FG Fall: I = -0.44539 🡨 Highest likely to be on sale in Fall**

**Minmaid Fall: I = -0.44539 – 0.23037 🡨 Lowest likely to be on sale in Fall**

**TRO Fall: I = -0.44539 – 0.12890 🡨 Medium likely to be on sale in Fall**

Minmaid:

Spring: I=0.44539-0.23037-0.17982

Summer: I=0.44539-0.23037-0.21097

Fall: I=0.44539-0.23037

Winter: I=0.44539-0.23037+0.07705

FG:

Spring: I=0.44539-0.17982

Summer: I=0.44539-0.21097

Fall: I=0.44539

Winter: I=0.44539+0.07705

Tropicana:

Spring: I=0.44539-0.1289-0.17982

Summer: I=0.44539-0.1289-0.21097

Fall: I= I=0.44539-0.1289

Winter: I= I=0.44539-0.1289+0.07705

All three brands are most likely to be on sale on winter

**1(d)(i):**

R Output:

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

P-value<0.01, reject the null hypothesis that a brand is equally likely to be on sale in all four seasons

**1(d)(ii):**

Linear hypothesis test

Hypothesis:

BRAND[T.MINMAID] = 0

BRAND[T.TROPICANA] = 0

Model 1: restricted model

Model 2: Feat ~ BRAND + Season

Res.Df Df Chisq Pr(>Chisq)

1 11996

2 11994 2 24.075 0.000005917 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

P-value<0.01, reject the null hypothesis that season being same, Minmaid and tropicana are equally likely to be on sale

**2.**

Call:

lm(formula = logmove ~ BRAND + Feat + logprice, data = orange)

Residuals:

Min 1Q Median 3Q Max

-4.6036 -0.5760 -0.0087 0.5556 4.0641

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.55428 0.03942 115.53 <2e-16 \*\*\*

BRAND[T.MINMAID] 0.27105 0.02090 12.97 <2e-16 \*\*\*

BRAND[T.TROPICANA] 2.21267 0.02383 92.85 <2e-16 \*\*\*

Feat 0.56347 0.01937 29.10 <2e-16 \*\*\*

logprice -2.52854 0.04635 -54.55 <2e-16 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9276 on 11995 degrees of freedom

Multiple R-squared: 0.5196, Adjusted R-squared: 0.5195

F-statistic: 3244 on 4 and 11995 DF, p-value: < 2.2e-16

>predict(LinearModel.4,interval='prediction',level=.95,newdata=data1)

fit lwr upr

1 3.985557 2.1670171 5.804097

2 5.307591 3.4889478 7.126234

3 2.549643 0.7311073 4.368178

4 3.871676 2.0530234 5.690329

5 2.657873 0.8393360 4.476411

6 3.853480 2.0348311 5.672129

**3.**

**1)** The flight origin does not have effect on the likelihood of flight delay at a 95% level of confidence?

Linear hypothesis test

Hypothesis:

d1 = 0

d2 = 0

Model 1: restricted model

Model 2: delaynew ~ d1 + d2 + d3 + d4 + d5

Res.Df Df Chisq Pr(>Chisq)

1 2197

2 2195 2 12.604 0.001833 \*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Chi-square value<0.05, reject the null hypothesis

**2)** The flight destination does not have effect on the likelihood of flight delay at a 95% level of confidence?

**Linear hypothesis test**

**Hypothesis:**

**d3 = 0**

**d4 = 0**

**Model 1: restricted model**

**Model 2: delaynew ~ d1 + d2 + d3 + d4 + d5**

**Res.Df Df Chisq Pr(>Chisq)**

**1 2197**

**2 2195 2 6.142 0.04637 \***

**---**

**Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1**

Chi-square value<0.05, reject the null hypothesis

**3)** B1=B4=0 .

B1=0. Given time of departure and destination, flights from IAD and BWI origins are

equally likely to be delayed.

B4=0. Given time of departure and origins, flights to LGA and JFK destinations are equally

likely to be delayed.

Combined: Given time of departure, for flights from either IAD or BWI (origin) and to

either LGA or JFK (destination), they are equally likely to be delayed

**Linear hypothesis test**

**Hypothesis:**

**d1 = 0**

**d4 = 0**

**Model 1: restricted model**

**Model 2: delaynew ~ d1 + d2 + d3 + d4 + d5**

**Res.Df Df Chisq Pr(>Chisq)**

**1 2197**

**2 2195 2 1.5887 0.4519**

Chi-square value>0.05, failed to reject the null hypothesis

**4)** For 4: I would be same for two combinations of origin and destination (B0 = B0 + B1 + B3). Identify those. P(y=1) is same for the two. B2+B3=0 Given time of departure, flights are equally likely to be delayed for the

following two combinations of origin and destination: 1) origin=IAD and

destination=LGA; 2) origin=DCA and destination=EWR.

Linear hypothesis test

Hypothesis:

d2 + d3 = 0

Model 1: restricted model

Model 2: delaynew ~ d1 + d2 + d3 + d4 + d5

Res.Df Df Chisq Pr(>Chisq)

1 2196

2 2195 1 0.1899 0.663

Chi-square value>0.05, failed to reject the null hypothesis

**5)** B0+B3+B5=0 A flight has a 50% likelihood to be delayed if the flight departures

from IAD to EWR in the evening (6:00pm or later).

For 5: This is I for a specific combination of origin, destination and time. I = 0 means P(y=1) = .5.

Linear hypothesis test

Hypothesis:

(Intercept) + d3 + d5 = 0

Model 1: restricted model

Model 2: delaynew ~ d1 + d2 + d3 + d4 + d5

Res.Df Df Chisq Pr(>Chisq)

1 2196

2 2195 1 12.689 0.0003678 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Chi-square value>0.05, reject the null hypothesis