Questions (to be turned in)

1. Consider the following regression and output in R. The data here is the reshaped data where we’re only looking at the Tropicana sales. Features include price of Dominicks, MM and Trop, though, and feat is only for Tropicana being featured.
   1. What is the expected change in quantity (in percents) for a 10% increase in the price of Trop?

***-23.2%***

* 1. What is the expected change in quantity (in percents) for a 10% increase in the price of MM?

***2.7%***

* 1. Why should we be concerned about inferring causality from the coefficient on log(P\_Trop):feat?

***Usually when a product is being featured, they might also be on sale.***

* 1. If Trop is a normal good, what would you expect the sign of the interaction of INCOME and log(P\_Trop) to be?

***The sign would be positive, since the greater the income, the less price sensitive they would be.***

1. Call:
2. glm(formula = logmove ~ log(P\_Dom) + log(P\_MM) + log(P\_Trop) \*
3. feat + AGE60 + EDUC + ETHNIC + INCOME + HHLARGE + WORKWOM +
4. HVAL150 + SSTRDIST + SSTRVOL + CPDIST5 + CPWVOL5, data = trop\_cross)
5. Deviance Residuals:
6. Min 1Q Median 3Q Max
7. -2.64300 -0.29531 -0.01999 0.26977 2.75790
8. Coefficients:
9. Estimate Std. Error t value Pr(>|t|)
10. (Intercept) 16.036404 0.405539 39.543 < 2e-16 \*\*\*
11. log(P\_Dom) 0.099452 0.021932 4.535 5.84e-06 \*\*\*
12. log(P\_MM) 0.265621 0.027778 9.562 < 2e-16 \*\*\*
13. log(P\_Trop) -2.318683 0.029309 -79.113 < 2e-16 \*\*\*
14. feat 1.736249 0.061270 28.337 < 2e-16 \*\*\*
15. AGE60 2.667011 0.156525 17.039 < 2e-16 \*\*\*
16. EDUC 1.597824 0.124527 12.831 < 2e-16 \*\*\*
17. ETHNIC 0.422341 0.045676 9.246 < 2e-16 \*\*\*
18. INCOME -0.509394 0.040865 -12.465 < 2e-16 \*\*\*
19. HHLARGE -1.310843 0.282861 -4.634 3.63e-06 \*\*\*
20. WORKWOM -0.778644 0.180130 -4.323 1.56e-05 \*\*\*
21. HVAL150 1.090685 0.051409 21.216 < 2e-16 \*\*\*
22. SSTRDIST -0.023534 0.001803 -13.053 < 2e-16 \*\*\*
23. SSTRVOL -0.112973 0.011990 -9.422 < 2e-16 \*\*\*
24. CPDIST5 0.087307 0.007720 11.309 < 2e-16 \*\*\*
25. CPWVOL5 -0.534881 0.031534 -16.962 < 2e-16 \*\*\*
26. log(P\_Trop):feat -1.367762 0.066919 -20.439 < 2e-16 \*\*\*
27. ---
28. Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1
29. (Dispersion parameter for gaussian family taken to be 0.2164328)
30. Null deviance: 6927.8 on 9648 degrees of freedom
31. Residual deviance: 2084.7 on 9632 degrees of freedom
32. AIC: 12634
33. Number of Fisher Scoring iterations: 2
34. Subset the oj data to only look at Dominicks sales (to be turned in as Rmd output).
    1. Find the quartiles of INCOME.
    2. Average sales within each quartile.
    3. Construct the MSE within each quartile for the model described above?
    4. Which quartile has the lowest MSE?
       1. What does this mean about the distribution of sales within that quartile?
    5. Which quartile has the highest MSE?
       1. What does this mean about the distribution of sales with that quartile?

***Question 2 see Rmarkdown***