Solution Assignment Lab 2

18^{th} March 2021

cavicchia@ese.eur.nl

ASSIGNMENT

Use the data set Grape Juice and answer to the following questions.

Data description A company is selling a new type of grape juice in some of its stores for pilot selling. Its marketing team wants to analyse:

- Which type of in-store advertisement is more effective?
- The Price Elasticity
- The Cross-price Elasticity
- How to find the best unit price to maximize the profit and the forecast of sales with that price.

There are 5 variables:

- **Sales**: Total unit sales of the grape juice in one week in a store;
- **Price**: average unit price of the grape juice in one week;
- **Ad type**: The in-store advertisement type to promote the grape juice, ad type=0 (natural production); ad type=1 (family health caring)
- **Price apples**: average unit price of the apple juice in the same store in one week
- **Price cookies**: average unit price of the cookies in the same store in one week

Work on yourself on the following tasks:

1. Data Exploration
2. Fit Multiple Linear Regression. Provide a brief interpretation of coefficients; evaluate the statistical significance of the model (t-tests and F-test and say in what they differ); assess the model assumptions (residual analysis).
3. With the fitted model, we can analysis the Price Elasticity(PE) and Cross-price Elasticity(CPE) to predict the reactions of sales quantity to price. Price elasticity is defined as $\%\Delta Q/\%\Delta P$, which indicates the percent change in quantity divided by the percent change in price; Cross-price Elasticity is the percent change in quantity divided by the change in the price of some other product - PE $= (\Delta Q/Q)/(\Delta P/P) = (\Delta Q/\Delta P) * (P/Q)$. Calculate also the CPE on apple juice and cookies to analyze the how the change of apple juice price and cookies price influence the sales of grape juice.

4. Optimal Pricing and Sales Prediction. Usually companies want to get higher profit rather than just higher sales quantity. So, how to set the optimal price for the new grape juice to get the maximum profit based on the dataset collected in the pilot period and the regression model above? To simplify the question, we can let the ad type = 1, the price apple = 7.659 (mean value), and the price cookies = 9.738 (mean value). Assume the marginal cost(C) per unit of grape juice is 5. We can calculate the profit (Y) by the following formula - Y = (price - C) * Sales Quantity = (price - 5) * (772.64 - 51.24*price). Find the optimal price.

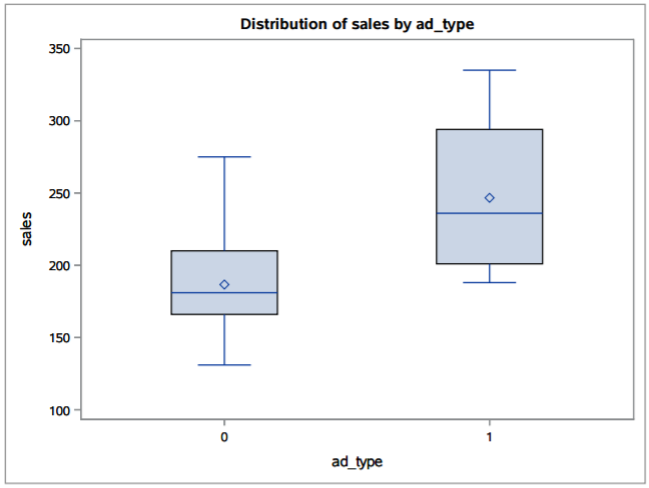
1 Data Exploration

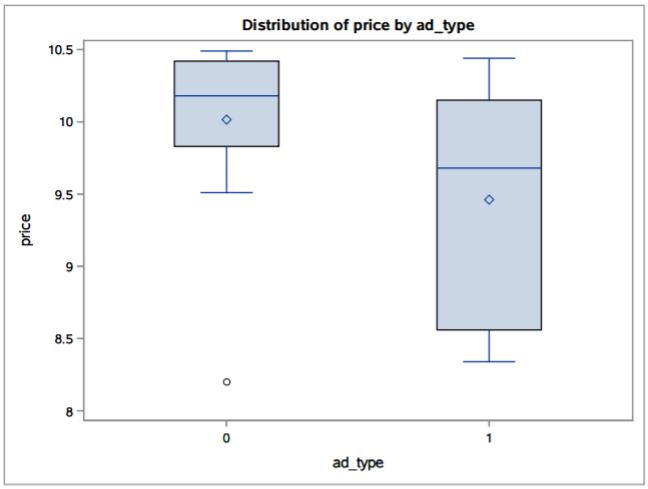
1.1 Univariate Analysis

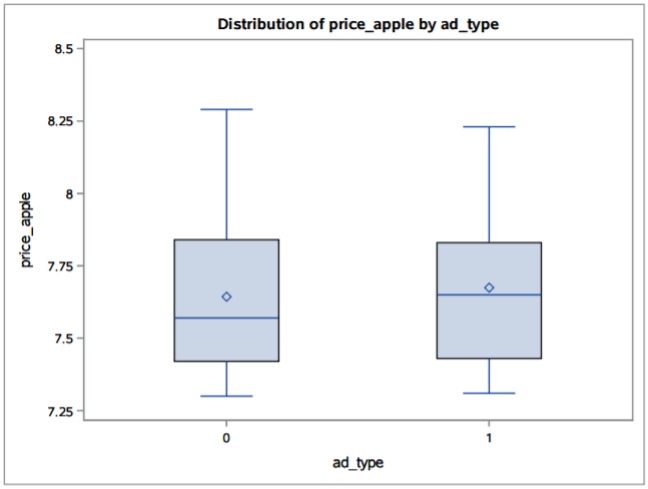
First it is possible to conduct univariate analysis by ad type. The mean of sales with nature product theme is about 187; the mean of sales with family health caring theme is about 247. The variable sales seems to be more skewed when ad type is 0, although is less variable (its standard deviation is 36 compared to 51 when ad type is 1).

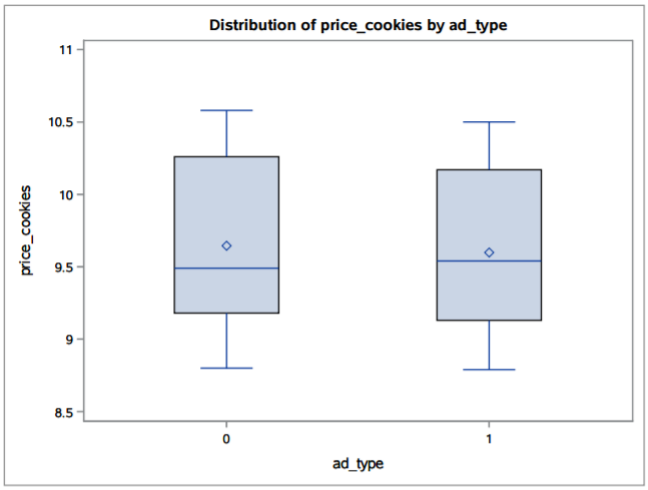
ad_type	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	Median	N	Lower Quartile	Upper Quartile	Quartile Range
0	15	sales	186.6666667	35.8641617	131.0000000	275.0000000	181.0000000	15	166.0000000	210.0000000	44.0000000
		price	10.0153333	0.5829588	8.2000000	10.4900000	10.1800000	15	9.8300000	10.4200000	0.5900000
		price_apple	7.6433333	0.2973854	7.3000000	8.2900000	7.5700000	15	7.4200000	7.8400000	0.4200000
		price_cookies	9.6453333	0.5673858	8.8000000	10.5800000	9.4900000	15	9.1800000	10.2600000	1.0800000
1	15	sales	246.6666667	50.5041252	188.0000000	335.0000000	236.0000000	15	201.0000000	294.0000000	93.0000000
		price	9.4613333	0.7328210	8.3400000	10.4400000	9.6800000	15	8.5600000	10.1500000	1.5900000
		price_apple	7.6746667	0.2910539	7.3100000	8.2300000	7.6500000	15	7.4300000	7.8300000	0.4000000
		price_cookies	9.5986667	0.5770846	8.7900000	10.5000000	9.5400000	15	9.1300000	10.1700000	1.0400000

To compare the variable over the two groups, it is convenient to analyse their boxblots by ad type. All the distributions seem to be a little skewed. Only the price variable is negatively skewed (mean is smaller than the median). The distributions of price apple and price cookies seem to be similar in the two groups. On the other hand, sales and price are differently distributed over the two groups. When ad type is 0, sales has a lower mean and smaller variance when ad type is 0, while price has a higher mean and smaller variance.



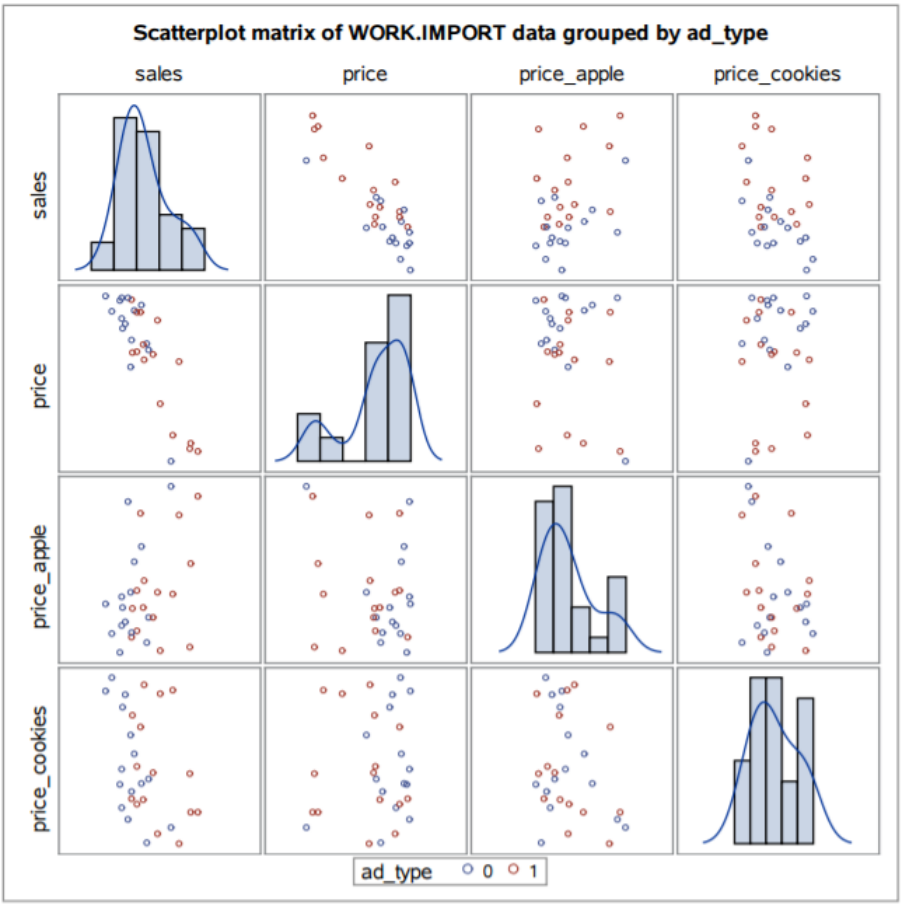






1.2 Bivariate Analysis

Scatterplot matrix and correlation matrix are useful to explore the dependencies between variables. On the main diagonal of the plot the histograms with the corresponding kernels are depicted. All variable distributions depart from the normality assumption: they mainly show skewness and biimodality (see for example price distribution). As regards the dependencies, price and sales results to be highly negative correlated; while the other variables are moderately correlated on each other as also confirmed by the correlation matrix (-0.85 vs. 0.37, -0.37, -0.30).



Pearson Correlation Coefficients, N = 30				
	sales	price	price_apple	price_cookies
sales	1.00000	-0.85105	0.36954	-0.37368
price	-0.85105	1.00000	-0.22046	0.08082
price_apple	0.36954	-0.22046	1.00000	-0.29933
price_cookies	-0.37368	0.08082	-0.29933	1.00000

2 Multiple Linear Regression Model

Looking ad the correlation matrix we could deduce that all variables could have an effect on sales. We fit a multiple linear model with all predictor (remember that ad type is a categorical predictor!).

Model: MODEL1
Dependent Variable: sales

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation	95% Confidence Limits	
Intercept	Intercept	B	804.55429	144.05610	5.59	<.0001	0	0	507.86520	1101.24338
price	price	1	-51.23935	5.32094	-9.63	<.0001	-0.68861	1.24608	-62.19802	-40.28068
price_apple	price_apple	1	22.08917	12.51227	1.77	0.0897	0.12124	1.14925	-3.68034	47.85867
price_cookies	price_cookies	1	-25.27664	6.29589	-4.01	0.0005	-0.26965	1.09925	-38.24327	-12.31001
ad_type_0	ad_type 0	B	-29.74170	7.24851	-4.10	0.0004	-0.28669	1.18969	-44.67029	-14.81310
ad_type_1	ad_type 1	0	0	.	.	.	.	.	.	.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	72436	18109	54.67	<.0001
Error	25	8280.70278	331.22811		
Corrected Total	29	80717			

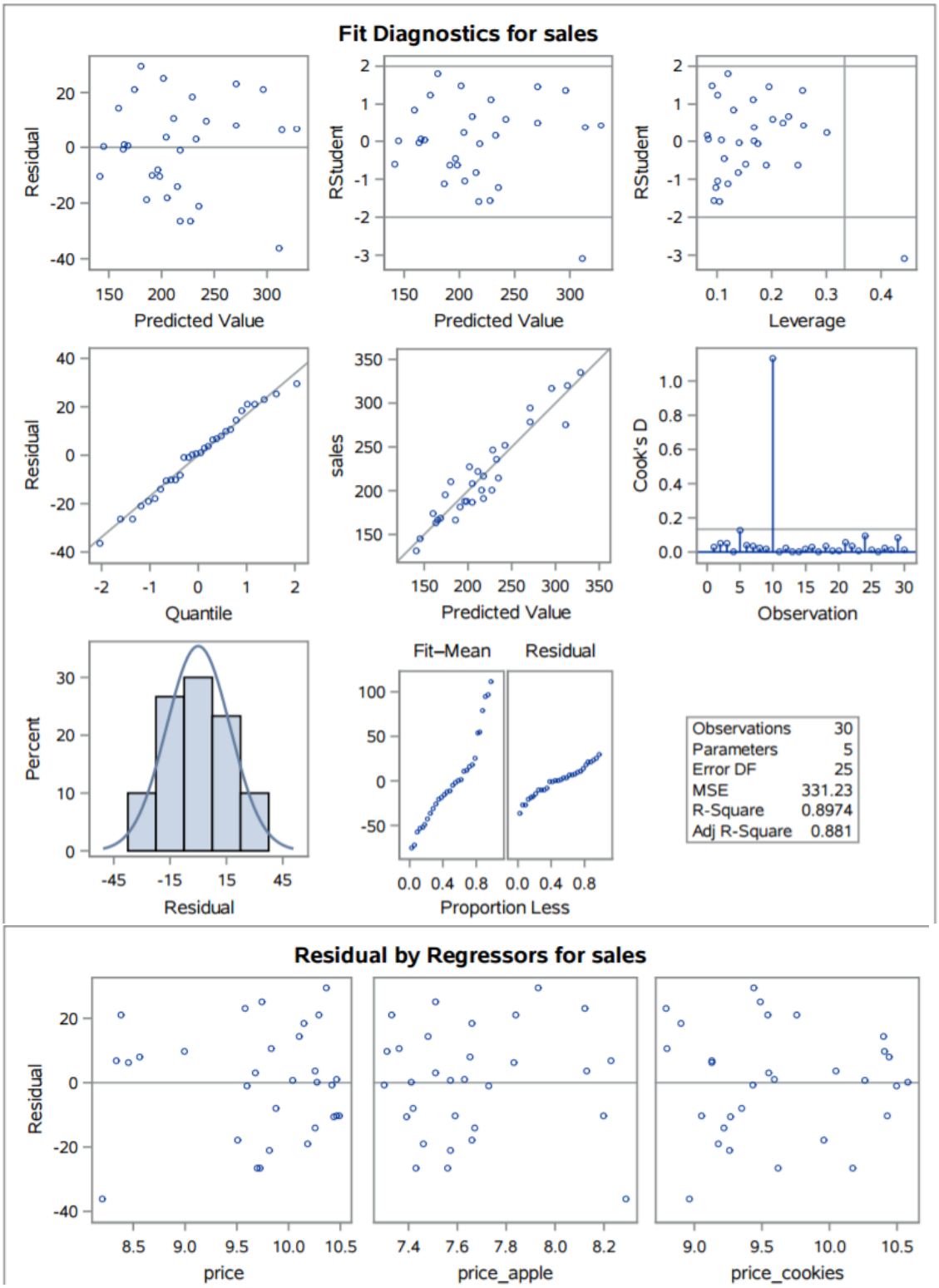
Root MSE	18.19967
Dependent Mean	216.66667
R-Square	0.8974
Adj R-Sq	0.8810

The p-value for price, ad type, and price cookies in the is much less than 0.05. They are significant in explaining the sales, as also confirmed by the confidence intervals.

The p-value of price apple is a bit larger than 0.05, seems there are no strong evidence for apple juice price to explain the sales. However, according to some marketing strategies, we know that when apple juice price is lower, consumers likely to buy more apple juice, and then the sales of other fruit juice will decrease. So we can also add it into the model to explain the grape juice sales.

The Adjusted R-squared is 0.881, which indicates a reasonable goodness of fit and 88% of the variation in sales can be explained by the four variables. The remaining 12% can be attributed to other factors or inherent variability.

Furthermore, looking at the VIF values, it is possible to say that there is not multicollinearity. Looking at the standardized estimates, price results to be the most important variable (its coefficient is -0.69).



The assumptions for the regression seem to be met. The residuals look to be independent, with constant variance and normally distributed.

3 Price Elasticity

The estimated regression equation is the following one

$$\hat{Sales} = 804.55 - 51.24 * price - 29.74 * adtype + 22.1 * priceapple - 25.28 * pricecookies$$

. It follows that

$$PE = (\Delta Q/Q)/(\Delta P/P) = (\Delta Q/\Delta P) * (P/Q) = -51.24 * 0.045 = -2.3,$$

where: P is price; Q is sales quantity; $\Delta Q/\Delta P = -51.24$, the estimated coefficient of price (remember its interpretation - increment in y when the predictor increases by one unit); $P/Q = 9.738/216.7 = 0.045$, P is the mean of prices in the dataset, so does Q.
The PE indicates that 10% decrease in price will increase the sales by 23%, and vice versa.
Let us further calculate the CPE on apple juice and cookies to analyze the how the change of apple juice price and cookies price influence the sales of grape juice.

$$CPE_{apple} = (\Delta Q/\Delta P_{apple}) * (P_{apple}/Q) = 22.1 * (7.659/216.7) = 0.78$$

$$CPE_{cookies} = (\Delta Q/\Delta P_{cookies}) * (P_{cookies}/Q) = -25.28 * (9.622/216.7) = -1.12$$

The CPEapple indicates that 10% decrease in apple juice price will decrease the sales by 7.8%, and vice versa. So the grape juice and apple juice are substitutes.
The CPEcookies indicates that 10% decrease in cookies price will increase the sales by 11.2%, and vice verse. So the grape juice and cookies are compliments. Place the two products together will likely increase the sales for both.
We can also know that the sales increase 29.74 units when using the ad with the family health caring theme (ad type = 1).

4 Optimal Pricing and Sales Prediction

We can let the ad type = 1, the price apple = 7.659 (mean value), and the price cookies = 9.738 (mean value).
The model is simplified as follows:

$$\hat{Sales} = 804.55 - 51.24 * price + 22.1 * 7.659 - 25.28 * 9.738$$
$$\hat{Sales} = 727.64 - 51.24 * price$$

By assuming the marginal cost(C) per unit of grape juice is 5. We can calculate the profit (Y) by the following formula:

$$Y = (price - C) * Sales$$

$$Quantity = (price - 5) * (727.64 - 51.24 * price).$$

Rearranging the terms, we have

$$Y = -51.24 * price^2 + 983.84 * price - 3638.2.$$

By maximizing the previous function wrt price, the optimal price is 9.60; the maximum profit will be 1084.39 according to the above output. In reality we can set the price to be 9.50. We can further use the model to predict the sales while the price is 9.50.
The sales forecast will be 241 units with a CI between 231 and 251.