**Price elasticity of Demand:**

In Economics, it is a measure of how sensitive demand or supply is to price.

It is also defined as "Ratio of % change in units sold to the percent change in price

Mathematically Elasticity = % Δ UNIT\_SOLD / % Δ UNIT\_PRICE.

There are many factors that define elasticity that are both supply side and demand side

E.g.

* Availability of substitute goods
* Consumer Income
* Necessity
* Brand Loyalty
* Duration of price change

**Price elasticity[Ed]:** it gives the percentage change in quantity demanded in response to a one percent change in price.

*Note: Only goods which do not conform to the* ***law of demand*** *such as "Veblen" and "Giffen" will not fall true.*

There are two types of elasticity

1. **Own price elasticity**

*Measures the rate of response of quantity demanded due to a price change.*

**PEoD = (% Change in Quantity Demanded)/(% Change in Price)**

1. **Cross price elasticity**

*The rate of response of quantity demanded of one good, due to a price change of another good. If two goods are substitutes, we should expect to see consumers purchase more of one good when the price of its substitute increases..."*

**CPEoD = (% Change in Quantity Demand for Good X)/(% Change in Price for Good Y)**

The quantity demanded for margarine was originally 3500 kilos - it is now 16% less or 2940 kilos.

(3500 \* (1 - 0.16)) = 2940

**Elasticity will raise questions like**

1. If I lower the price of a product, how much more I will sell?
2. If I raise the price of one good, how will that affect sales of the other goods?
3. If the market price of a product goes down, how much will that affect the amount?

**PED(Ed) < 1**, the demand for the good is said to be inelastic when the change in price have a relatively small effect on the quantity of good demand

**PED(Ed) > 1**, the demand for the good is said to be elastic when changes in price have a relatively large effect on the quantity of a good demand.

|  |  |
| --- | --- |
| **Terminology** | **Meaning** |
| Un-responsive | Elasticity between 0 to -0.75 |
| Responsive | Elasticity between -0.75 to -1.5 |
| Very Responsive | Elasticity below -1.5 |
| Insufficient Observations | No. of weeks less than 105 |
| Insufficient Variation | No price change with 5% coefficient of variation |
| Inconclusive | Less than 70% confidence |

**Regression:**

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors').

**Robust regression:**

Robust regression is an alternative to least squares regression when data is contaminated with outliers or influential observations and it can also be used for the purpose of detecting influential observations.

Sales ~ Price of a SKU +

Calamities: Rain, Flood, Tornado, Snow, Natural disasters, promotions.

holiday: Christmas, Victoria day, Good friday, Easter, Thanks giving, Labour day.

Market factors: Disposable income, Average house hold income, Single furnished housing permit.

**Some terms in linear regression:**

**Residual:** The difference between the predicted value (based on the regression equation) and the actual, observed value.

**Outlier:** In linear regression, an outlier is an observation with large residual. In other words, it is an observation whose dependent-variable value is unusual given its value on the predictor variables. An outlier may indicate a sample peculiarity or may indicate a data entry error or other problem.

**Leverage:** An observation with an extreme value on a predictor variable is a point with high leverage. Leverage is a measure of how far an independent variable deviates from its mean. High leverage points can have a great amount of effect on the estimate of regression coefficients.

**Influence:** An observation is said to be influential if removing the observation substantially changes the estimate of the regression coefficients. Influence can be thought of as the product of leverage and outlierness.

**Cook's distance (or Cook's D)**: A measure that combines the information of leverage and residual of the observation.

We now have our model:  
**Sales of Eggs = 137.37 – (16.12)Price.Eggs + 4.15 (Ad.Type) – (8.71)Price.Cookies**

### **Own Price Elasticity**

To calculate Price Elasticity of Demand we use the formula:  
PE = (ΔQ/ΔP) \* (P/Q)

(ΔQ/ΔP) is determined by the coefficient -16.12 in our regression formula.  
To determine (P/Q) we will use the mean Price (4.43) and mean Sales (30).

Therefore we have PE = -16.12 \* 4.43/30 = -2.38

This means that an increase in the price of eggs by 1 unit will decrease the sales by 2.38 units.

**Sales Lift – Sales Impact Analysis**

Lift analysis is a way to measure how a campaign impacts a key metric. **Lift analysis means comparing users who receive a campaign to a group of users who do** *not* **receive the campaign (i.e. the control group) to see which group is better off**.

**Sales Impact Analysis:**

To measure the performance of Tests within Home depot

Tradition approaches include

Only test stores

* Dollar run rate
* Comp run rate

Test and Control stores

* Dollar run rate
* Comp run rate

*Dollar run rate with test and control stores is preferred, but the challenge is choosing the control stores.*

MuSigma proposed a statistical approach for the problem.

Analytical approach involves analyzing each test store individually using dummy variable reg model and club the findings.

Factors like economy, holidays, seasonality, calamity, commodity pricing are also considered.

The steps include:

1. Data preparation: factors considered in the model are at store-week level
2. Missing value treatment: for data continuity they are made 0
3. Outlier Treatment: Z-scores are take and values >5 are outliers
4. Seasonality Index: one way is by creating dummy variables and other way is calculating seasonality index for each month
5. Deweatherization: WDD data from Plananalytics is used,

**de-weatherized sales =  sales/(1+ wdd\_factor)**

**Control store Selection:**

1. Division Filter: Consider stores from same geographic division
2. Non Control Filter: remove few stores like the 45 stores which geographically apart and the stores in which the same even it planed few months back on in near future.
3. Competition Filter: remove the stores impacted by competitive stores. Effect of competition on test stores is measured by dummy variable and the store is removed if lift is found negative.
4. Calamity filter: Stores impacted by calamities are removed.
5. Store open date filter: Stores opened within 26 weeks of kickoff week are removed.
6. Correlation Filter: Co-relation on every test-control pair is taken and top 100 pairs are considered for next steps.
7. Dickey Fuller test: testing the stationary of time series. The Test store sales are regressed against Control sales.

Sales (Test sales) = intercept + Beta \* sales (control stores) +Residual ( eq:1)

The residuals of eq 1 is regressed against the lag of residuals

1st diff of residuals = intercept + Beta \* 1st lag of residual + Error (eq: 2)

t- Value = beta/Standard error (cutoff of 3.14 is used and top 50 control stores are selected)

1. Euclidean distance filter: the distance on the 13 standardized variables (mean 0, sd 1) is calculated and nearest 10 control stores are selected.

**Model calibration:**

Sometimes the even that is running in parallel may impact the event of interest. To calculate the correct lift we use this approach, the regression model accounts for parallel events that are running.

1. Create a calibration dummy assuming an unknown event from 8 week before kickoff week is impacting. Regress this with sales. Check for calibration dummy significance by p-value < 0.3. if dummy is not significant impact of parallel event is not considered and model is called "organically calibrated".
2. Decay factor is used for capturing the correct impact and is made to 10%.

***Regression model***: an approach to find the linear relationship between dependent and independent variables.

* R-square: quantifies explicability of the model. Should be closer to 1.0
* P - value: quantifies the significance of a variable to prove NULL hypothesis.

***Moving average model:*** if the variance in the sales is >0.5 we go with moving average model else the normal regression.

1. 5 weeks Moving Average is calculated and dead period is increased to 4 week from 1.Kickoff week is increased to original kickoff week+4.
2. Control gap% = (test sales - Control sales) / Test sales.
3. Run the weekly model and get the R square and control gap metrics. If CGP metric is >10% then freeze and go for weekly model or else

* Weekly R^2 <= 40% and CGP >10% - 5 week MA
* Weekly R^2 between 41 -50%, CGP > 10% - 4 week MA
* Weekly R^2 between 51 -60%, CGP > 10% - 3 week MA
* Weekly R^2 >60%, CGP > 10% - 2 week MA

Based on CGP:

* If CGP <= 10% freeze on first level of MA
* If CGP >10 and <15% then goto one more order of MA
* If CGP > 15% and <= 20% then goto two more order of MA
* If CGP > 20% then goto 5 week MA and % week is max.

In case of Lift analysis, depending the availability of control stores we have two broad ways of modelling:

1. Test/Control regression modeling: when you have good number of control stores.
2. Pre-post regression modeling: too many test stores and less control stores to choose.

**Test/Control Regression**:

1. Sales data from 52 weeks of the kickoff weeks is pulled for each test store and regression dataset is prepared. Ie: one test and 10 control stores data at store-week combination level. The data points of kickoff week are ignored, for some events specified period is ignored(dead period).
2. You might still have outliers, which are removed using cooks distance.
3. The scale gap between test and control stores might cause erroneous calculations, so percentage calculations are preferred.

Actual: lift = (Test post sales - Test pre sales) - (Control post sales - Control pre sales)

Preferred: lift = % change in test store sales - % change in control store sales

To bring test and control sales to same scale,  we use normalized control sales:

Normalized control sales = Actual control sales \* ( Avg. Test pre 52 sales / Avg. Control Pre 52 Sales)

1. Dependent variable is actual sales in dollars and independent variables are weather, macro, seasonality, holidays, calamity, commodity price, competition, test-post dummy, test control dummy, pre-post dummy.

Variables remaining after removing multicollinearity and forward step wise regression, pre-post dummy, 10 test control dummies, test-post dummy, Bias dummy(if effect exists) are used in the regression.

1. Measuring Lift and Impact duration: The parameter estimate of Test-post dummy is taken as the AVG weekly lift, if its significance is >50%. If not the average weekly lift of the store is rounded off to 0, as the lift won't be statistically correct.

**Pre Post Regression model:**

To measure the impact of an event that is rolled out in large number of stores. Historical data can be used to measure trend. By removing the seasonality and trend in the data we can quantify the impact of our event.

The framework accounts for the effect of weather, calamities, seasonal variations, sales trend, and year over year trend.

Pre-post methodology has three steps:

1. Clustering: stores are clustered on the bases of region and event kick-off week. A minimum of 5 stores are required in a cluster
2. Model selection: 7 models are built over pre 104 weeks, excluding 13 weeks validation period. De-weatherization and De-seasonalization of sales is done. MAPE is calculated for each model, model with least MAPE and maximum R-square are selected for final regression.
3. Final Regression: the independent variables going into regression are from calamity, holiday, promotions, trend etc. Best model is selected based on R-square and the lift is calculated.

**Competitor Pricing -**

Problem - Dynamic pricing that adjusts with competitor price changes

Data required - Competitor SKU leved prices, SKU description, SKU attributes

Approach - Identify the markets that are important to the competitor by looking at competitor's major promotional programs, number of stores, share of wallet etc. and then map the prices across attributes of products against our prices

Post which identify opportunities for a quick win- markdown in markets where we are priced lower than competition

Besides, we can also observe the trend in competitor prices and predict their prices across categories to adjust our long term pricing strategy

Alongside, we can identify the KVI for the competition, understand any gap in assortment based on competition data as well