BUSINESS ANALYTICS

(SEPTEMBER 2022 TERM)

BY GAGNEET KAUR

Date
September 19, 2022

MONDAY

PAGE NO:
DATE:

WEEK 3

Determining & Inferring Association

JOINT PROBABILITY

Joint probability is the probability of two events happening together. The two events are usually designated event A and event B. In probability terms, it is :

$$p(A \text{ and } B) \quad \text{or} \quad p(A \cap B)$$

Joint probability can also be described as the probability of the INTERSECTION of two (or more) events

$$P(A \cap B) = P(A|B) * P(B)$$

MARGINAL PROBABILITY

Marginal Probability is the probability of an event irrespective of the outcome of another variable.

PAGE NO.
DATE:

CONDITIONAL PROBABILITY

Conditional probability is known as the possibility of an event or outcome happening, based on the existence of a previous event or outcome.

The probability of occurrence of any event A when another event B in relation to A has already occurred is known as conditional probability.
It is depicted by $P(A|B)$

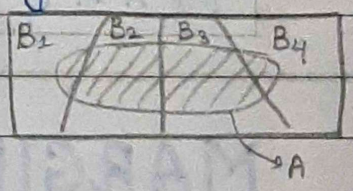
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{or} \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$P(A|B)$ ↓ probability of A given B

BAYES' RULE

* Bayes rule is used to calculate the posterior probability if we have the initial probability & additional sample information!!

When a sample space is a disjoint union of events & Event A overlaps this disjoint union, then the probability that one of the disjoint partitions events is true, given A is true, is



$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{P(A)}$$

$P(A \cap B_j)$

$$\sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)$$

PAGE NO.
DATE:

CHI-SQUARE DISTRIBUTION

Chi Square Test Statistic :

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

→ When the H_0 is true, expected and observed frequencies tend to be closed for each cell & the test statistic value is relatively small.

→ If H_0 is false, at least some cells have a big gap between expected & observed frequencies, leading to a large test statistic value.

→ The larger the χ^2 value, greater is the evidence against the null hypothesis of independence.

→ Degrees of freedom for the chi squared distribution is given by the expression : $df = (r-1)*(c-1)$
$r \rightarrow$ no. of rows ; $c \rightarrow$ no. of columns

→ If χ^2 > critical value, then reject H_0.

(χ^2 : calculated value) (critical value : tabulated value)

→ If p-value < level of significance, then reject H_0.

Date: September 27, 2022

TUESDAY

PAGE NO.
DATE:

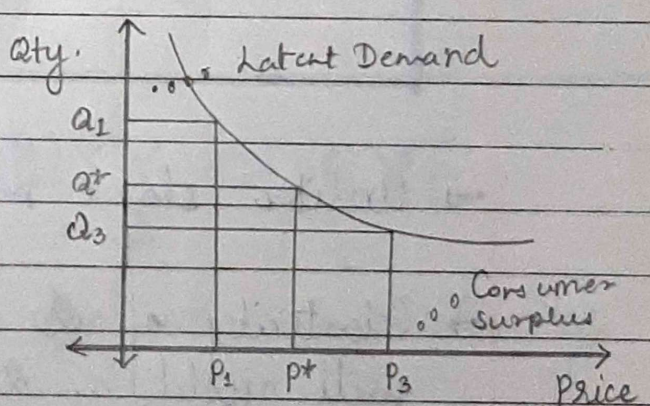
Week 4

Demand Response Curve

RELATIONSHIP BETWEEN PRICE & DEMAND

DEMAND RESPONSE CURVE

- A function that describes how demand for a product $D(p)$ varies as a function of its price.
- Similar to the Demand curve in Economics, but for a single seller, in a single market
- PROPERTIES :



1. Non-negative
2. Continuous
3. Differentiable
4. Downward Sloping

PRICE SENSITIVITY

- Slope : Measures how demand changes in response to a price change

$$\partial(p_1, p_2) = \frac{D(p_2) - D(p_1)}{p_2 - p_1}$$

→ $p_1 > p_2$ implies that $D(p_1) < D(p_2)$, hence slope is always negative.

→ Slope can be used as a local estimator of demand change for a small change in price.

⇒ Elasticity : Ratio of the percentage change in demand to the % age change in price.

$$\epsilon(p_1, p_2) = \frac{[d(p_2) - d(p_1)] / d(p_1)}{(p_2 - p_1) / p_1}$$

→ Unlike slope, elasticity is independent of units.

→ Elasticity of 2 means that a 10% red^n in price will yield a 20% increase in sales.

→ Elasticity also depends on TIME. There can be different elasticities of same product for short-term & for long-term.

Shrisneh
PAGE NO.
DATE:

LINEAR RESPONSE CURVE

- Simplest Price Response Curve :

$$D(p) = D_0 - m * p$$

where D_0 is the demand at price = 0 (this is called the market size) & m is the slope.

→ the price at which demand = 0 is called the satiating price, $P_s = \frac{D_0}{m}$

→ the elasticity of this curve is $\epsilon = \frac{m * p}{D_0 - m * p}$

We see that $\epsilon = 0$ when $p = 0$.
As $p \longrightarrow P_s$, $\epsilon \longrightarrow \infty$.

CONSTANT ELASTICITY CURVE

- After algebraic transition, the constant elasticity curve is given by :

$$D = C p^{-\epsilon}$$

where C is a constant (it is the Demand when price = 1)

→ it is not guaranteed that the demand is either finite or satiated ($D \to \infty$, as $p \longrightarrow 0$). Also, $D \neq 0$, for any p.

→ Revenue is $R = p * D = Cp^{(1-\epsilon)}$

- We notice that :

→ When $\epsilon < 1$, (inelastic product demand), the revenue can be increased by simply increasing prices.

→ When $\epsilon > 1$, (elastic demand), the revenue can only be increased by setting price close to zero.

ESTIMATION PROBLEM

- Assume that we conduct ~~an~~ a market experiment where we offer different prices and check the realized demand at that time.
- So we have the price offered & the corresponding demand values.
- Price can be considered as an explanatory variable & demand can be considered to be the dependent variable.
- Can we estimate the slope & the elasticity.

Date
September 28, 2022

Wednesday.

Simple Linnear Regression

LINEAR DEMAND RESPONSE CURVE

- A simple Linear Regression (SLR) model can help identify D_0 (y-intercept) & m (the slope).
- The SLR can tell us if the linear relationship is a good fit for the data available from the market experiment.

SIMPLE REGRESSION MODEL

LINEAR ON AVERAGE

→ The eqn of the SRM describes how the conditional mean of Y depends on X.

→ The SRM shows that these means lie on a line with the intercept β_0 & the slope β_1

$$\mu_{y|x} = E(Y|X=x) = \beta_0 + \beta_1 x$$

DEVIATIONS FROM THE MEAN

→ The deviations of responses around $\mu_{y|x}$ are called errors.

→ **ERRORS** : The vertical gap between the actual value of demand realized at that particular price point & the predicted value of y coming from the SLR is called Error.

→ Error, is denoted by ϵ, & $E(\epsilon) = 0$

→ In general we minimize the sum of squared errors. That is the objective with which we run the regression model.

→ The SRM makes three assumptions about the error term :

1. **Independent**
   - Errors are independent of each other.

2. **Equal Variance**
   - All errors have the same variance, $Var(\epsilon) = \sigma_\epsilon^2$.

3. **Normal**
   - The errors are normally distributed.

Coefficient of Determination (R^2)

"price explains $R^2 \times 100\%$ of variation/variability in y"

→ Observed values of the response Y are linearly related to the values of the explanatory variable X by the equation

$$y = \beta_0 + \beta_1 x + \epsilon$$

(β_0: y-intercept; β_1: slope; x: Market size)

where $\epsilon \sim N(0, \sigma_\epsilon^2)$ (mean = 0, variance = σ_ϵ^2; N: normal distribution)

→ The observations are independent of one another, have equal variance around the regression line, and are normally distributed around the regression line.