Week1

- Types of data: Categorical, Numerical. Categorical has two subtypes: Ordinal, Nominal. Numerical has two subtypes: Interval and Ratio.
- Typically all measured quantities are numerical, and all counted quantities are categorical.
- Umbrella principles of effective visualization – Know Purpose, Ensure Integrity, Maximize data/minimize non-data ink, Annotate data (preferred over legend)
- Information display steps – (a) Defining message (b) Choosing form (c) Creating design
- Choose a table representation when
    - displaying a complete (partial) dataset,
    - data is too wide to display on a graph,
    - explain how numbers are derived or calculated.
- Choose a chart representation when
    - Comparing a slice of information
    - Show cause vs effect
    - Show change over time
    - Show patterns of data distribution (eg. normal curve)
- Appropriate graph types for conveying messages

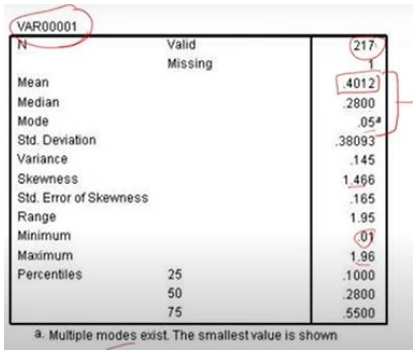
Message	*Chart type*
Components of one item	Pie chart
Components of multiple items	100% column/stacked column chart
Item comparison	Bar chart
Change over time	Column/line chart
Frequency/distribution	Histogram
Outliers	Box plot
Correlation	Paired bar, Scatter chart

- A concise definition of a dashboard: A visual display of the most important information needed to achieve one or more objectives that has been consolidated on a single screen so it can be monitored and understood at a glance.
- Common Issues plots : Axes not starting from 0, percentages in pie chart not adding up, Inappropriate chart type, 3D pie charts

Week2

- PMF (discrete) represent probability masses for each value of a random variable.
- PDF (continuous) represent probability density for a random variable. It's not possible to find the probability for a particular value of a continuous random variable, but for a small range of values.
- Trace-driven simulation is limited to provided data values. Of course, the disadvantage is, for values outside the historically observed data, prediction isn't possible.
- Alternatively, you might choose to "fit" a theoretical distribution, or even an empirical distribution (that you've built). Theoretical distributions are preferred over empirical distributions, due to the limitations of available data in the latter.

- To define a distribution, we need to define its density/distribution function and estimate its parameters (mean, std. deviation etc)
- While defining density functions, make sure that for $x < a_0$, $y = 0$ and for $x > a_k$, , $y = 1$, where a_0 and a_k are arbitrary values.
- Selecting the appropriate distribution to fit given data also depends on the domain.
- To "fit" a theoretical distribution, follow these steps.
    - Get the descriptive statistics of the data as follows.

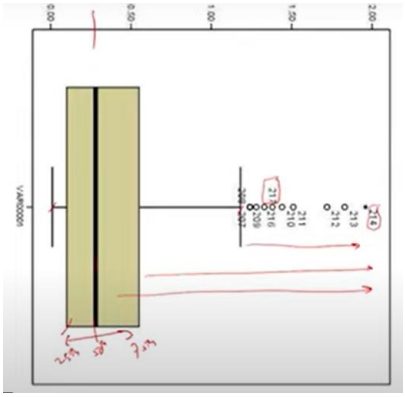


VAR00001

N	Valid	217
	Missing	1
Mean		.4012
Median		.2800
Mode		.05[a]
Std. Deviation		.38093
Variance		.145
Skewness		1.466
Std. Error of Skewness		.165
Range		1.95
Minimum		.01
Maximum		1.96
Percentiles	25	.1000
	50	.2800
	75	.5500

a. Multiple modes exist. The smallest value is shown

    - Is the distribution symmetric?
        - If the mean, median and mode are not equal to each other, then it is not symmetric. We can eliminate symmetric distributions like standard normal and uniform at this point. However, it's not necessary that the distribution is symmetric, if the mean and median are equal. To confirm symmetricity, plot a histogram.
    - What's the support of the distribution?
        - If the support of the distribution includes negative X-axis, we could expect at least some values in the sample-set to negative. If the minimum value is greater than 0, it could mean that the support is restricted to positive X-axis.
    - What could be the shape of the distribution?
        - Plot histogram and observe the shape.
        - If the skewness is positive, the right tail of the distribution is bigger than the left tail. Else, the left tail is bigger than right tail. Note that skewness of a standard normal distribution is 0, and that of a exponential distribution is 2.
        - For a std. normal distribution, kurtosis is 3. An increased kurtosis (>3) can be visualized as a thin "bell" with a high peak whereas a decreased kurtosis corresponds to a broadening of the peak and "thickening" of the tails.
    - Are there outliers?
        - Plot box-plot and see if there're datapoints beyond the whiskers. This will also agree with the skewness. If the distribution is right skewed, median will be placed on the left of the distribution.



- Other metrics?
  - Find co-efficient of variation (cv = std. dev/mean). If cv = 1, then distribution could be exponential. If cv > 1, the distribution could be lognormal. Note that this applies only to continuous distributions, and only when mean is non-zero. In the case of discrete distributions, cv is called Lexis ratio.
  - If mean and variance are equal (or approximately so), then the distribution could be Poisson.
  - If the standard deviation is low, and the intervals between the quartiles are more or less same, then the distribution could be uniform.

Refer to the following table for details about specific distributions:

Distribution	PMF ($f_X(k)$)	CDF ($F_X(x)$)	$E[X]$	$\mathrm{Var}(X)$
Poisson(λ)	$\frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, \ldots, \infty$	$\begin{cases} 0 & x < 0 \\ e^{-\lambda}\sum_{i=0}^{k}\frac{\lambda^i}{i!} & k \le x < k+1 \\ & k = 0, 1, \ldots, \infty \end{cases}$	λ	λ
Exp(λ)	$\lambda e^{-\lambda x}$, $x > 0$	$\begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal(μ, σ^2)	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$, $-\infty < x < \infty$	No closed form	μ	σ^2

- Once a distribution has been "fit", we should check how good is the fit – using a chi-square/Kolmogorov-Smirnov (KS) test or a probability plot.
- There are two kinds of probability plots = P-P plot and Q-Q plot.
  - Q-Q plot is plotted taking equivalent quantiles from model and sample distributions that have equal probabilities. If the equivalent quantiles are equal then the Q-Q plot will be a 45° line with an intercept 0. Q-Q plot can amplify the differences between the tails of the model and sample distributions.
  - P-P plot is plotted taking equivalent probabilities from model and sample distributions at the same quantiles. If the equivalent probabilities are equal then the P-P plot will be a 45° line with an intercept 0. This is valid for both continuous and discrete datasets. P-P plot can amplify the differences in the middle portion of the model and sample distributions.

- Chi-square tests are based on a null/alternative hypothesis. Null hypothesis is that the datapoints are from the model distribution. Alternative hypothesis is that the datapoints are not from the model distribution.
    - If the tabulated chi-squared statistic (scipy.stats.chi2.ppf) is more than the calculated value (stats.chisquare(obs_freq, expec_freq)), then we will fail to reject null hypothesis. Alternatively, as the calculated value approaches zero, there is more evidence that the null hypothesis is true.
    - In terms of p-value, if p-value is greater than (1 – confidence level), we fail to reject null hypothesis.

Week3

- Consider a B-school which shortlisted 1200 candidates (960 men and 240 women) for its post-graduate management program. Out of these, 324 candidates were given offer letters for admission. The data is included here:

	Male	Female	Total
Offers made	288	36	324
Not offered	672	204	876
Total	960	240	1200

Contingency table

- After reviewing the record, a women's forum raised the issue of gender discrimination on the basis that 288 male candidates were offered admission against only 36 female candidates.

- This can be proved or disproved using conditional probability.
- Calculate the joint and marginal probabilities as below.

- In terms of probabilities, the previous table can now be rewritten as:

	Male	Female	Total
Offers made	0.24	0.03	0.27
Not offered	0.56	0.17	0.73
Total	0.8	0.2	1.0

- Joint probabilities (of what?) appear in the main body of the table (e.g. 0.24, 0.03).
- Marginal probabilities (of what?) appear in the margin of the table (e.g. 0.8, 0.2).

- From the above table, we have P(A ∩ M) = 288/1200 = 0.24, P(M) = 960/1200 = 0.8 and P(A | M) = 288/960 = 0.3. Thus, we can write
    - P(A | M) * P(M) = P(A ∩ M).
- On similar grounds, P(A | F) = 0.15
- Since P(A | F) < P(A | M), offering admissions appears to be a biased towards the candidate being male.
- In other words, Conditional Probability = Joint Probability/Marginal Probability. This is the Baye's formula.
- Assuming two suppliers S1 and S2 supplying products to a factory,

- We are interested in the posterior probability that a particular supplier is guilty of supplying bad quality product *given that* we have bad quality raw material at our doorstep – $Pr(S1|B)$ or $Pr(S2|B)$.
- This is an application of Baye's theorem – finding posterior probability given some initial facts and numbers.

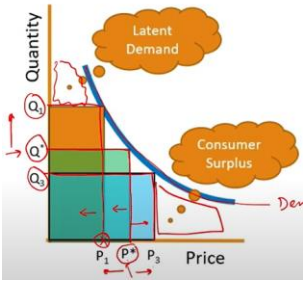
Thus, the probability that S1 is the supplier, given that the product is of bad quality is given by the following formula,

$$\Pr(S1 \mid B) = \frac{\Pr(S1) * \Pr(B \mid S1)}{\Pr(S1) * \Pr(B \mid S1) + \Pr(S2) * \Pr(B \mid S2)}.$$

- It's not possible to draw inferences purely based on statistical behavior on samples. We need do hypothesis tests.
- In hypothesis tests, two tests must be conducted, one each for the null hypothesis (H_0) and alternative hypothesis (H_A)
    - Let f_o be the observed frequencies
    - Let f_e be the expected frequencies, if the null hypothesis were true. (obtained by multiplying row total and column total, divided by the total sample size)
    - Chi-square statistic can now be calculated using the formula $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$.
- When H0 is true, f_o and f_e tend to be closer, and hence the calculated chi-square statistic is relatively small. Larger the calculated chi-square statistic is, greater is the evidence against H_0
- Degrees of freedom is given by the formula (k - p - 1), where k is the # observations(#bins), p is the number of parameters used in the distribution pdf. Alternatively, if data can be reproduced like in a contingency table, degrees of freedom is given by the product of (#rows-1) and (#columns-1)

Week4

- Demand response curve (shown below) deals with a single product in a single market, whereas a typical demand curve in Economics deals with the entire market.

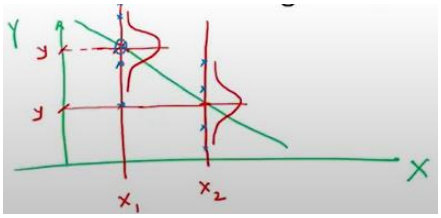


- Latent demand lies towards the top left of the demand-response curve, and gets realized as you reduce the price.
- Consumer surplus lies towards right bottom of the demand-response curve and gets realized (reduces) as you increase the price.
- Revenue-maximizing price is not necessarily same as Profit-maximizing price, and may even be conflicting objectives.
- Properties of demand-response curve are: Non-negative (first quadrant), Continuous, Differentiable, Downward sloping
- There are two types of demand-response curve.
    1. Linear curve
        - Slope of the curve is defined as $\partial(p_1, p_2) = \frac{D(p_2) - D(p_1)}{p_2 - p_1}$
        - Elasticity of the curve is defined as $\epsilon(p_1, p_2) = -\frac{[d(p_2) - d(p_1)]/d(p_1)}{(p_2 - p_1)/p_1}$
        - Elasticity of 2 means a 10% reduction in price leads to a 20% increase in demand (sales).

- Elasticity of certain items are very less (Eg. Salt), whereas that of certain others are very high (Eg. Movies)
- In certain cases, the short-term elasticity could differ from the long-term elasticity. Thus, short-term elasticity of airline travel is less, but the long-term elasticity is high. In the case of a two-wheeler, the short-term elasticity is high, but long-term elasticity is lower.
- $D(p) = D_0 - m * p$ is the simplest price response curve. This is a linear curve. D_0 is called the market price and represents the demand when the price is 0.
- The price at which the demand is 0 is called satiating price. $P_S = D_0/m$
- Elasticity can also be represented as $\epsilon = \frac{m*p}{D_0 - m*p}$. Note that elasticity depends on the price at the specific point.
- When p = 0, elasticity is 0. When p is P_S, elasticity is also infinite.
- Thus, keep in mind slope and elasticity are not same. While slope remains for a linear equations, elasticity can keep changing with price (X-axis)

2. Non-linear curve (constant elasticity curve)
   - We can keep elasticity unchanged, with a non-linear demand curve $D = Cp^{-\epsilon}$, where C is a constant given as demand when p = 1
   - In the case of a non-linear curve, we can't find the market by putting up 0 price; neither can we find the satiating price, since the demand never reduces to 0. It's asymptotic either way.
   - Revenue = p * D = $Cp^{(1-\epsilon)}$
   - When elasticity is less than 1 (inelastic), the revenue can be increased by simply increasing the prices.
   - When elasticity is greater than 1 (elastic), the revenue can only be increased by setting price closer to 0.

- We can use a simple linear regression (SLR) to find the slope and intercept of a linear response curve. We can perhaps use SLR even in the case of a non-linear curve with constant-elasticity.
- Use Excel's scatter chart and plot the trend of price-response curve; the trend-line should give the slope m and intercept D_0
- SLR plots $\mu_{y|x}$ as $\beta_0 + \beta_1 x$, where β_0 is equal to D_0, and β_1 is equal to m. Deviations of actual values from the predicted values are called errors. Expected value of sum of squared errors is 0.
- In order to use SLR, we assume that error terms are independent of each other, of equal variance and normally distributed.
- $y = \beta_0 + \beta_1 x + \epsilon \text{ where } \epsilon \sim N(0, \sigma_\epsilon^2)$ is the general representation of the linear model.
- In the above equation, β_1 is called marginal slope.
- Following picture represents this graphically.



- Excel's data analysis pack gives the following statistics when SLR is used. Here's the interpretation of these values.

Regression Statistics	
Multiple R	0.85668
R Square	0.7339
Adjusted R Square	0.72472
Standard Error	1290.45
Observations	31

- Multiple-R signifies the correlation between the price and demand (CORREL).
- R-Square is obtained by squaring Multiple-R.
- Standard error is σ_ε (the square-root of the variance mentioned earlier).
- R-square value is also called coefficient of determination and represents the power of the explanatory variable to explain dependent variable. In our case, R-square value of 0.7339 means that price values can explain ~73% of demand.
- R-square value can also be found as SS_Regression/Total_Regression, or 1 – (SS_Residual/Total_Regression)

ANOVA

	df	SS	MS	F	gnificance F
Regression	1	133188236.7	133188236.7	79.9806	7.8E-10
Residual	29	48292440.76	1665256.578		
Total	30	181480677.4			

- df_Regression=1, since there are two parameters for a linear regression (bias and weight).
- df_Total=30, since the #observations=31.
- df_Residual=29 (df_Total – df_Regression)
- SS_Regression is SSM (Sum of squares of model).
- SS_Residual is SSE (Sum of square of errors).
- SS_Total is sum of squares of variation of each point from the mean (same as variance of given data)
- MS_Regression is SSM/df_Regression.
- MS_Residual is SSE/df_Residual. This quantity is equal to variance of error. Standard_Error = sqrt(MS_Residual)
- F=test statistic = MS_Regression/MS_Residual

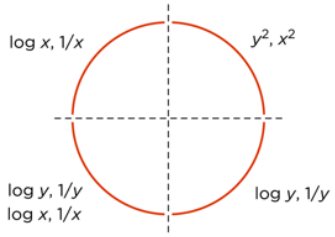
Now, we must take a hypothesis between x and y. H_0 = x and y are independent; H_A = x and y are dependent.
F indicates the test statistic. p-value (7.8E-10) is lesser than 0.05, so at 95% confidence level, we can reject the hypothesis and hence conclude the demand and price are dependent.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ow
Intercept	5842.8362	400.6837789	14.58216306	6.9E-15	5023.35	6662.33	
Price	-157.7009	17.63363083	-8.943187902	7.8E-10	-193.766	-121.636	-1

Above table indicates the slope and the intercept of the linear equation. t_Stat and P-Value are what a local hypothesis test of these values yields. $H_0 = \beta_0$ (or β_1) is zero; β_0 (or β_1) is non-zero. Since the p-values in both cases is lesser than 0.05, so at 95% confidence level, we can reject both hypothesis and hence conclude that β_0 (and β_1) are both non-zero.

Week5

- In the case of a constant elasticity (non-linear) model, we can apply transformations to the variables to convert it to a linear model. Use the following thumb-rule for deciding on the transformation.



- In the case of a constant elasticity model represented as $D = Cp^{-\epsilon}$, use a log transformation on D and p to obtain the representation $Log(D) = Log(C) - \epsilon Log(P)$.

Using optimization to maximize Revenue or Profit

- In the case of a linear relationship between demand and profit, revenue is represented as

  $R(p) = (D_0 + m * p) * p = D_0 * p + m * p^2$

  Price that yields maximum revenue is $D_0/2m$.

- In the case of a linear relationship between demand and profit, profit is represented as

  $\pi(p) = Total\ Revenue - Total\ Cost = D(p) * p - D(p) * c$

  Price that yields maximum profit is $(D_0 + m * c)/2m$.

  NOTE: Find the partial derivative of these functions and equate to zero, in order to obtain the optimum price in either cases.
  NOTE2: The optimum price that yields maximum revenue is often lesser than the optimum price that yield maximum profit.

Using linear programming to find optimum values

- Primal or dual representations of an optimization problem can be used to solve for the optimum values of the variables.
- Primal and dual forms are interchangable, and both yields the same value of the objective function.
- To convert one form to another, use

Primal

Maximize $500000X_1 + 1000000X_2 + 2500000X_3$

Subject to

$$15X_1 + 20X_2 + 60X_3 \leq 120000$$
$$20X_1 + 50X_2 + 100X_3 \leq 140000$$
$$X_1, X_2, X_3 \geq 0$$

Dual

Minimize $120000Y_1 + 140000Y_2$

Subject to

$$15Y_1 + 20Y_2 \geq 500000$$
$$20Y_1 + 50Y_2 \geq 1000000$$
$$60Y_1 + 100Y_2 \geq 2500000$$
$$Y_1, Y_2 \geq 0$$

Primal	Dual
Maximization	Minimization
Number of constraints	Number of variables
Number of variables	Number of constraints
Objective function coefficient	Right hand side in constraints
Right hand side in constraints	Objective function coefficient

- Dual of a dual is primal
- If one of the constraints in the above set of inequalities is changed to an equality constraint, the corresponding shadow price can never be 0, in the dual representation of the problem. If there are only two constraints, this implies that the shadow price corresponding to the other variable will be 0.
- In excel, you may use solver add-in to optimize using primal-dual representations of the problem. Use *sensitive report* to get the dual variables (shadow prices), after solving for the primal variables.
- In Python, use *pulp* package to optimize using primal-dual representations of the problem.

Week6

- Use multiple linear regression (MLR) to describe the relationship between multiple explanatory variables and response variable. Recollect simple linear regression (SLR) is used when there's a single explanatory variable.
- $Y = \beta_0 + \beta_1 X_1 + \cdots \beta_k X_k + \epsilon$ where $\epsilon \sim N(0, \sigma_\epsilon^2)$ is the general representation of the MLR model.
- MLR can be thought of like an SLR, where all but one explanatory variables are bundled into standard error. Thus, in the above MLR, $\beta_1 X_1 + \cdots \beta_k X_k + \epsilon$ can be considered as part of the standard error when represented as an SLR problem.
- In the above equation, $\beta_1 \ldots \beta_k$ are called partial slopes. Note that, in an SLR, β_1 is called marginal slope.
- The β-values are the change in the response variable, with a unit change in the corresponding explanatory variable, keeping all the other explanatory variables constant.
- Partial and marginal slopes are same, only when the explanatory variables are uncorrelated. When there's high correlation among the explanatory variables, the MLR output becomes difficult to interpret. In this case, the explanatory variables are said to be collinear.
- Two methods exist to quantify this correlation between explanatory variables
    - Through path diagrams, where correlation between X_j and response variable has two components to it, a direct effect and an indirect effect (brought over by correlation of X_j and all other explanatory variables)
    - Variance inflation factor (VIF). VIF for X_j is given by the formula $VIF(X_j) = \frac{1}{1-R_j^2}$, where R_j^2 is the coefficient of determination in the regression of X_j on all other explanatory variables.

i.e., consider X_j as the response variable, while keeping all other explanatory variables as its explanatory variables.

- When correlation between explanatory variables is significant, standard error in estimation of partial slope gets inflated due to VIF according to the formula, $se(b_1) = \frac{s_e}{\sqrt{n}} \times \frac{1}{s_x} \times \sqrt{VIF(X_1)}$ Thus, if the explanatory variables are totally *uncorrelated*, then $R_j^2 = 0$, and hence VIF = 1. In this case, standard error remains unchanged. However, if the explanatory variables are *correlated*, $R_j^2 > 0$, which implies VIF > 1 and hence standard error increases, thereby making those predictions unreliable.
- Similar to the case of SLR, the assumptions are that the error terms are independent of one another, have equal variance and are normally distributed around the regression equation.
- In the case of MLR, residuals departing from normality suggests that an important explanatory variable has been omitted.
- Similar to the case of SLR, R^2 (called R-squared) indicates how much does the fitted MLR equation explain the variation in response. $\bar{R}^2$ (called adjusted R-squared) adjusts R^2 for both sample size n and #explanatory variables k. s_e indicates the standard error.
- In order to calculate $\bar{R}^2$ use this formula: $\text{Adjusted R Squared Formula} = 1 - \left[\frac{(1 - R^2) \times (n - 1)}{(n - k - 1)}\right]$), where n is the number of observations and k is the number of explanatory variables.
- Another way to interpret R-squared value is that it helps represent the correlation between y and y_hat (predicted y)
- When $\bar{R}^2$ increases s_e reduces, and vice-versa.
- In the case of MLR, F-statistic is used to test the null hypothesis that all partial slopes are 0. It's calculated using the formula $F = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k}$ and t-statistic is used to test the null hypothesis that individual partial slopes are 0. It's calculated using the fomula $t_j = \frac{b_j - 0}{se(b_j)}$
- To solve problems due to collinearity, either remove redundant explanatory variables, or re-express some in terms of others.

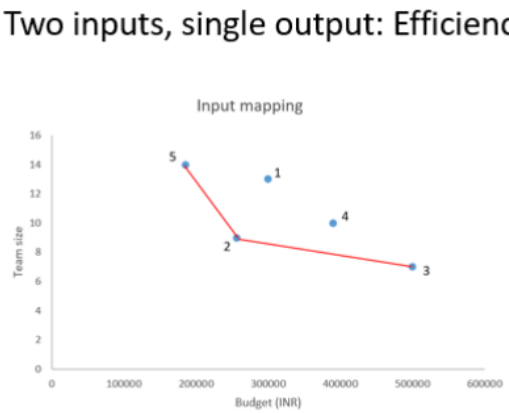
Week7

- To make categorical predictions (like Yes/No), make use of logistic regression models.
- A typical formulation of logistic regression problem starts with a set of explanatory variables, each of which have a corresponding weight (β1…βm), jointly predicting the odds of a successful outcome. Odds is defined as the ratio P(Y=1)/P(Y=0). Note that it's normal to deal with log(odds) instead of odds.
- Thus we can write the model as $Log(Odds) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$
- From the above model equation above, it follows that $\Pr(Y = 1) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}$
- The β-values are the change in the log(odds), with a unit change in the corresponding explanatory variable, keeping all the other explanatory variables constant. Thus, if the explanatory variable increases by 1 unit, the odds of Y=1 increases by a factor of e^{β}
- To start with building a model, run the regression model and find the initial β-values

- Objective function in the case of a logistic regression is maximization of log-likelihood. To calculate the likelihood, convert all cases of P(0) to 1-P(1). Take a log of this quantity. Sum of log-likelihood across all samples must be maximized, by manipulating the β-values.
- Now, to find the classification output(Yes/No), compare P(Y=1) with a threshold probability. If P(Y=1) is above the threshold, Y=1. If P(Y=1) is below the threshold, Y=0.
- To evaluate the model, use classification metrics - accuracy, precision and recall.
- Accuracy is a ratio of the number of times, predicted and actual Y values matched (for both *Y = 0* and *Y = 1*) to the total observations in the sample.
- Recall is a ratio of the number of times the prediction is positive, to the number of times when it's actually positive.
- Precision is a ratio of the number of times the actual is positive, to the number of times when it's predicted to be positive.

Week8

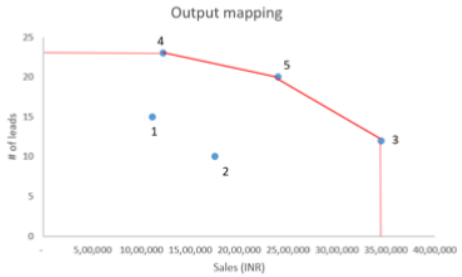
- The productive efficiency "frontier" are all the combinations of outputs such that the production of one product cannot be increased without sacrificing the output of the other. It can also be thought of in terms of the input combinations, wherein one input cannot be reduced without increasing another input.
- Efficiency is defined as ratio of (weighted) output to (weighted) input.
- In a 2 inputs/1 output case, imagine that output (sales target) is fixed at 10L INR. When the LP problems for all the DMUs are solved, we get the below graph. Note that the objective function in this case is minimization of the inputs, and DMUs 5, 2, 3 are in the frontier. DMUs 1 and 4 are considered inefficient.



Sales office	Budget (INR)	Team size
1	3,00,000	13
2	2,56,000	9
3	5,00,000	7
4	3,90,000	10
5	1,85,000	14

- In a 1 inputs/2 output case, imagine that input(budget) is fixed at 2L INR. When the LP problems for all the DMUs are solved, we get the below graph. Note that the objective function in this case is maximization of the outputs, and DMUs 4, 5, 3 are in the frontier. DMUs 1 and 2 are considered inefficient.

One input, two outputs: Efficiency Frontier



Sales office	Sales (INR)	No of leads
1	11,10,000	15
2	17,50,000	10
3	34,50,000	12
4	12,24,000	23
5	24,00,000	20

- DEA is a non-parametric mathematical method to find the production frontier. This measures the relative efficiency of DMUs through LP.
- Since various inputs/outputs can't be directly added, weights (also called *decision variables*) must be defined on each.
- There are separate sets of weights for each DMU, but the weights of one DMU shouldn't yield an efficiency more than 1, for any DMU, including itself.
- Using a DMU's own weights, if it can't achieve an efficiency of 1, it's truly inefficient.
- Using a DMU's weights, if another DMU gets an efficiency of 1, then the latter is much more efficient than the former.
- $E_k = \frac{y_{1k}O_{1k}+y_{2k}O_{2k}+y_{3k}O_{3k}\ldots+y_{Mk}O_{Mk}}{x_{1k}I_{1k}+x_{2k}I_{2k}+x_{3k}I_{3k}\ldots+x_{Nk}I_{Nk}}$ gives the efficiency of k[th] DMU, where y1k, y2k…represent the weights assigned to each output for the k[th] DMU and x1k, x2k… represent the weights assigned to each input for the k[th] DMU.
- Since maximizing E_k is a non-linear problem (due to this being a ratio), we linearize it by equating the denominator to 1.
- For a case, where we're consider 2 outputs and 2 inputs depicted by the following table

	Inputs		Outputs	
Sales office	Budget (INR)	Team size	Sales (INR)	No of leads
1	3,00,000	13	11,10,000	15
2	2,56,000	9	17,50,000	10
3	5,00,000	7	34,50,000	12
4	3,90,000	10	12,24,000	23
5	1,85,000	14	24,00,000	20

the mathematical formulation for LP for DMU-1 looks like below. All other DMUs have similar LP formulation.

$$Max\ y_{11} * 1110000 + y_{21} * 15$$

$$subject\ to \quad x_{11} * 300000 + x_{21} * 13 = 1$$

$$y_{11} * 1110000 + y_{21} * 15 \le x_{11} * 300000 + x_{21} * 13$$

$$y_{11} * 1750000 + y_{21} * 10 \le x_{11} * 256000 + x_{21} * 9$$

$$y_{11} * 3450000 + y_{21} * 12 \le x_{11} * 500000 + x_{21} * 7$$

$$y_{11} * 1224000 + y_{21} * 23 \le x_{11} * 390000 + x_{21} * 10$$

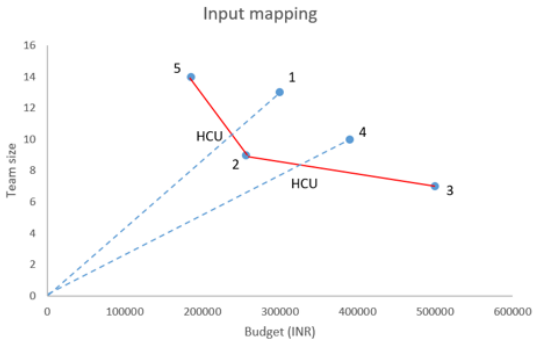
$$y_{11} * 2400000 + y_{21} * 20 \le x_{11} * 185000 + x_{21} * 14$$

$$Decision\ variables: x_{11}, x_{21}, y_{11}, y_{21} \ge 0$$

- We can solve this problem and obtain the weights by using *Solver add-in* in Excel.

Week9

- Inefficient DMUs can be rendered more efficient by moving the envelopes to a position called "Hypothetical Composite Unit (HCU)". But, neither horizontal nor vertical movement of the envelope works.
- As shown in the below graph, move inefficient DMU-1 towards the line joining DMU-2 and DMU-5 (called as reference units). From the graph, it can be found that, DMU-1 will be efficient, when the current budget is reduced to Rs.2,37540/- and team size to 10.3.



- The following sensitivity report generated in Excel for DMU-1 contains the details to arrive at the above revised figures for budget and team-size.

Variable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
G5	inp1 wt	2.06356E-06	0	0	138076.9231	48307.69231
H5	inp2 wt	0.029302518	0	0	2.093333333	5.983333333
I5	out wt	7.91993E-07	0	1000000	1E+30	1000000

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
H11	inp2 wt	0.791993397	0	0	1E+30	0.208006603
H12	inp2 wt	0.791993397	0.740817169	0	0.174011299	1E+30
H13	inp2 wt	0.791993397	0	0	1E+30	0.444903013
H14	inp2 wt	0.791993397	0	0	1E+30	0.305819232
H15	inp2 wt	0.791993397	0.259182831	0	0.802547771	0.285790032
H9	inp2 wt	1	0.791993397	1	1E+30	1

The highlighted portion in the first table contains the input weights ($2.06 * 10^{-6}$, 0.029) and output weight ($7.91 * 10^{-7}$) for DMU-1. However, using these weights DMU-1 is able to achieve only ~0.792 efficiency. Hence, it can be concluded that DMU-1 is not an efficient unit.

Now, see the portion highlighted in the second table. In order to convert DMU-1 into an efficient unit, use DMU-2 and DMU-5 as reference units, use the shadow prices of 0.74 and 0.26 respectively and compute its revised budget and team-size as follows.

	Reference units		HCU for 1	Actual values	Excess inputs used
	2 (2,56,000,9)	5 (1,85,000, 14)			
Dual variable	0.74	0.26			
Sales	10,00,000*0.74 +	10,00,000*0.26	10,00,000	10,00,000	-
Budget	2,56,000*0.74 +	1,85,000*0.26	2,37,540	3,00,000	62,460
Team size	9*0.74 +	14*0.26	10.3	13	2.7

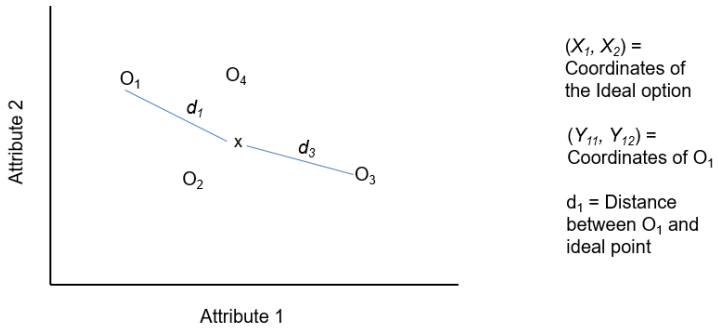
Similarly, look at the sensitivity report generated for DMU-4, and it can be computed that revised budget and team-size should be Rs.327980/- and 8.41 respectively.

	Reference units		HCU for 1	Actual values	Excess inputs used
	2 (2,56,000,9)	3 (5,00,000, 7)			
Dual variable	0.705	0.295			
Sales	10,00,000*0.705	+ 10,00,000*0.295	10,00,000	10,00,000	-
Budget	2,56,000*0.705	+ 5,00,000*0.295	3,27,980	3,90,000	62,020
Team size	9*0.705	+ 7*0.295	8.41	10	1.6

NOTE: If the factors given by the sensitivity report don't add to 1, scale them such that they add to 1.

Week10

- Family of techniques that model choice by decomposing overall preference in terms of the relative values of the attributes to respondents.
- Conjoint analysis, in that sense, constructs a value system by asking about preferences on a small subset of products and then using the system to make predictions about the relative choices.
- Conjoint analysis, in the sense of optimization, can also be used to arrive at the "best product" – a product that has all the attributes at a level desirable (preferable) to the customer.
- Once the preferred attributes are known, either
    - focus on refining these attributes and developing something that the consumers would like.
    - choice of attributes can be used to focus on can be narrowed down using conjoint analysis on the products available in the market.
    - reveal's customer's willing to pay for specific attributes.
    - decide to price the product based on the level of attribute present in that variant
- Here is a geometric explanation to the process of conjoint analysis. O1, O2, O3, O4 are 4 different *subjects*, each representing a different product variant with different values for the two attributes we consider - *Attribute1* and *Attribute2*. When asked to rate these 4 product variants, the customer places them at suitable locations on the space, guided by the importance of the attributes to her. Note that Ideal product is the point (X1, X2) that is placed at the least distance from the preferred variant for the customer.



The problem can be solved either as an optimization problem or LinearRegression.

- When approached as an optimization problem, the final formulation looks like this:

Let,

$$a_{jkp} = y_{kp}^2 - y_{jp}^2, \forall (j,k) \in \Omega \text{ and } p \in P$$

$$b_{jkp} = -2(y_{kp} - y_{jp}), \forall (j,k) \in \Omega \text{ and } p \in P$$

$$V = \{v_p\} = \{w_p x_p\}, p \in P$$

$$z_{jk} = max\left[0, -\left[\sum_{p \in P} w_p a_{jkp} + \sum_{p \in P} v_p b_{jkp}\right]\right]$$

and let,

$$A_p = \sum_{(j,k) \in \Omega} a_{jkp} \ for \ p \in P$$

$$D_p = \sum_{(j,k) \in \Omega} b_{jkp} \ for \ p \in P$$

then, objective function is given as

$$Min \sum_{(j,k) \in \Omega} z_{jk}$$

Subject to:

$$\sum_{p \in P} w_p a_{jkp} + \sum_{p \in P} v_p b_{jkp} + z_{jk} \geq 0 \ for \ (j,k) \in \Omega$$

$$\sum_{p \in P} w_p A_p + \sum_{p \in P} v_p D_p = 1$$

$$w_p \geq 0 \ and \ v_p \ unrestricted \ for \ p \in P$$

$$z_{jk} \geq 0 \ for \ (j,k) \in \Omega$$

Week11

- Conjoint analysis can also be through of as a multiple linear regression problem, where product attributes are the independent variables and respondent's ratings are the dependent variables.
- The regression coefficients are called 'part-worth', or alternatively referred to as a *level utilities* for the attributes.
- Following table show how to calculate the importance of each attribute, given the part-worth of each.

Attribute	Partworth range
Brand	1.834 – 0 = 1.834
Battery	2.334 – 0 = 2.334
Camera	4.834 – 0 = 4.834

Total of ranges = 1.834 + 2.334 + 4.834 = 9.002

Attribute	Importance
Brand	1.834/9 = 20.37%
Battery	2.334/9 = 25.92%
Camera	4.834/9 = 53.7%