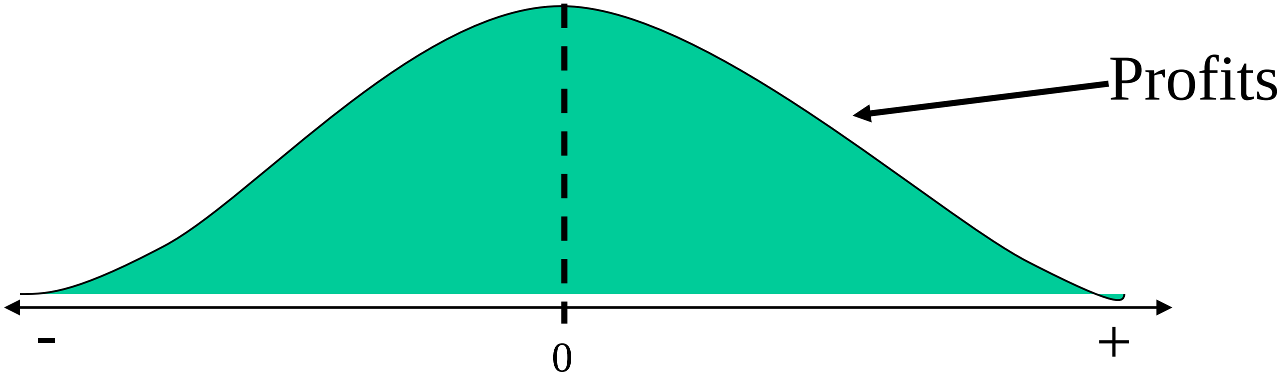
Logistic Regression

Linear Regression Assumption

- Linear regression assumes the dependent variable (DV) to be continuous (and normally distributed)

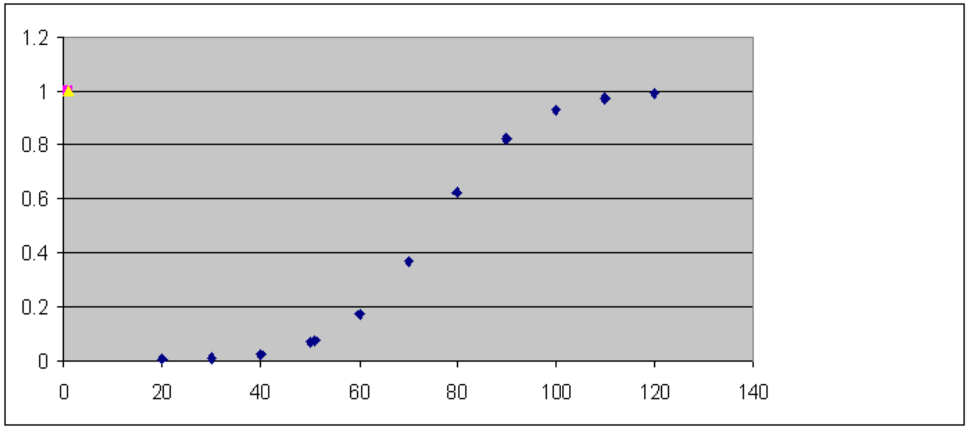


- Often we have variables where there are only 2 different values
  - Buy (1) vs no buy (0)
  - Retain (1) vs lose customer (0)

Logistic Regression

- Logistic Distribution

Choice Probability



Price Discount on
$877 Price of Miami Flight

- Do our choice preferences evolve in an “S” shaped manner?

Customer Retention: Logistic Regression

- With categorical (1/0) dependent variables, linear regression can result in nonsensical estimated probabilities (e.g. probability of retention > 100%)

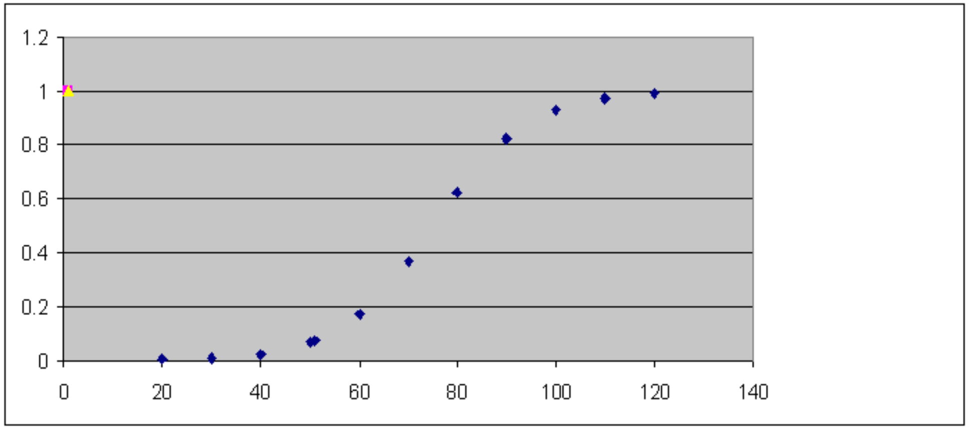
- A model that allows us to do this is the so-called “logistic regression”

Logistic Regression –
How do we get the S- Shaped Form?

$$\text{Prob(Retention)} = \frac{e^{(a+b_1 \Pr iceDiscount)}}{1+e^{(a+b_1 \Pr iceDiscount)}}$$

Predictions are bound between [0,1]

Choice Probability



Price Discount on
$877 Price of Miami Flight

Example:
What Predicts Above Median Sales of Xbox Games on Best Buy Mobile App?

							customer	customer
						regular	review	review
sku	game	numsales	abmedian	browsetime	new	price	count	average
1004622	Sniper: Ghost Warrior - Xbox 360	53	1	-0.00017	0	19.99	7	3.4
1010544	Monopoly Streets - Xbox 360	12	1	-0.00285	0	29.99	3	4
1011067	MySims: SkyHeroes - Xbox 360	3	1	0.00157	0	19.99	1	2
1011491	FIFA Soccer 11 - Xbox 360	85	1	-479.80822	0	12.99	18	4.6
1011831	Hasbro Family Game Night 3 - Xbox 360	6	1	0.00094	0	9.99	2	3.5
1012721	The Sims 3 - Xbox 360	140	1	-0.00031	0	19.99	13	3.8
1012876	Two Worlds II - Xbox 360	5	1	0.00047	0	39.99	8	3.4
1013666	Call of Duty: The War Collection - Xbox 360	41	1	0.00115	0	68.18	2	4.5
1014064	Castlevania: Lords of Shadow - Xbox 360	15	1	-0.00235	0	7.99	4	4.8
1032361	Need for Speed: Hot Pursuit - Xbox 360	168	1	-0.00039	0	19.99	45	4.2
1052221	Marvel vs. Capcom 3: Fate of Two Worlds - Xbox 360	28	1	-0.00092	0	19.99	11	4

Example:
What Predicts Above Median Sales of Xbox Games on Best Buy Mobile App?

Top Sellers	Bottom Sellers
Battlefield 3 Limited Edition - Xbox 360	Adrenalin Misfits - Xbox 360
Dead Island - Xbox 360	Dance Masters - Xbox 360
Call of Duty: Modern Warfare 3 - Xbox 360	Rango - Xbox 360
Batman: Arkham City - Xbox 360	MotionSports: Adrenaline - Xbox 360

Example: XLStat Output

Summary statistics:

Variable	Categories	Frequencies	%
nrx_ind	0	1128	44.183
	1	1425	55.817

Variable	Observations	Obs. with missing data	Obs. without missing data
sales calls	2553	0	2553

Minimum	Maximum	Mean	Std. deviation
0.000	12.000	2.396	2.128

Goodness of fit statistics (Variable nrx_ind):

Statistic	Independent	Full
Observations	2553	2553
Sum of weigh	2553.000	2553.000
DF	2552	2551
-2 Log(Likelih	3504.580	3216.666
R²(McFadden	0.000	0.082
R²(Cox and S	0.000	0.107
R²(Nagelkerk	0.000	0.000
AIC	3508.580	3220.666
SBC	3520.270	3232.356
Iterations	0	6

Example: XLStat Output

Model parameters (Variable abmedian):

Source	Value	SE	Wald Chi-Square	Pr > Chi²
Intercept	-1.707	0.814	4.397	0.036
new	-2.896	1.736	2.784	0.095
regular price	0.023	0.022	1.153	0.283
customer review count	0.175	0.073	5.695	0.017
customer review average	0.352	0.164	4.573	0.032

Example: Sales of Xbox Games

Coefficient of Customer Review Average (b_{review})	0.352	
	Customer Review Average = 3	Customer Review Average = 4
U = a+bx	-1.707 + 3*0.352 = -0.651	-1.707 + 4*0.352 = -0.299
P(sale) = exp(u)/(1+exp(u))	.34	0.43
difference	0.09	

Hit Rates – In Sample

		Observed	
		Above Median	Below Median
Predicted	Above Median	16	11
	Below Median	10	62

Hit Rate = (16+62)/(16+10+11+62)

= (78)/99 = 79%

Model Building

- Determine properties of dependent variable
  - Linear, + ve values, Dummy Variable, text data
- Select model that reflects dependent variable properties
  - Logistic regression for dummy variables

Model Building

- Include the decision variable of interest among the independent variable set
  - Price, advertising, etc

- Include common control variables
  - Quality, Distribution, Demographics, Tenure, Competition etc.

Model Building

- Does including lagged dependent variable lead to UNIT ROOT?
- If UNIT ROOT, use difference as the dependent variable

Marketing Mix Models - Summary

- Are independent variables correlated?
  - Is the sign of a variable not making sense?
  - Is the significance and sign of the coefficient changing with other variables in the model?

- Do we have an omitted variable bias?

- If no omitted variable bias-
  - Check for correlation among independent variables
  - If they are correlated; try combining them (add/subtract/divide/multiply etc.)

Model Building

- Does the model hint @ causality or is it a correlational model?
  - – Are dependent and independent variables measured at the same time?
  - – Are there sufficient controls or confounding variables included
  - – Can a reverse causation reasonably exist
  - – Do we need to recommend an experiment?