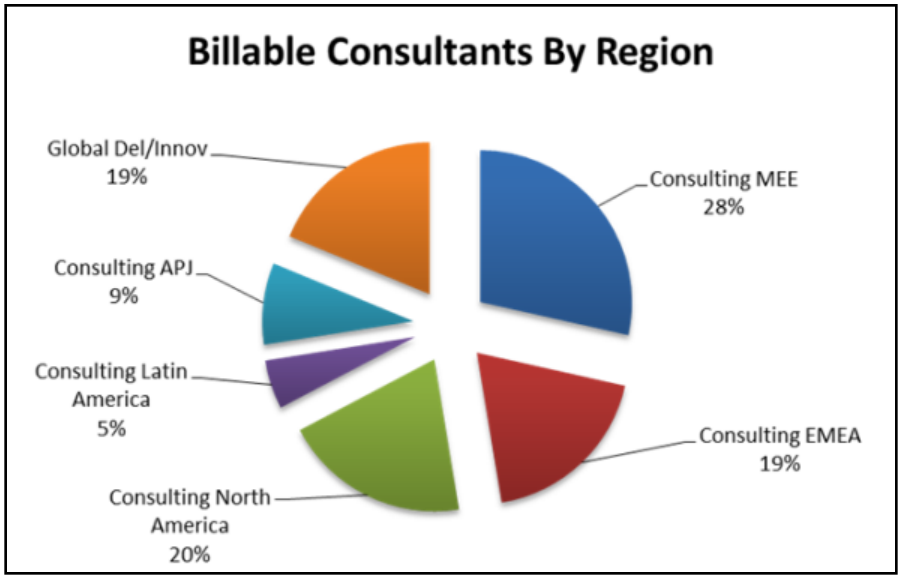
SAP Propensity

SAP Propensity Modeling: SAP Global Lead Platform/MEE Leads

- Project in 2013 for SAP Global Lead Platform—MEE Region—to better focus consultants on solutions w/in accounts
- Focus on “market categories” including applications, analytics, database and technology, cloud etc. plus core ERP and education and likelihood to buy
- Piloted on 64 top accounts and rating each account by **propensity and likelihood to buy** specific SAP solutions

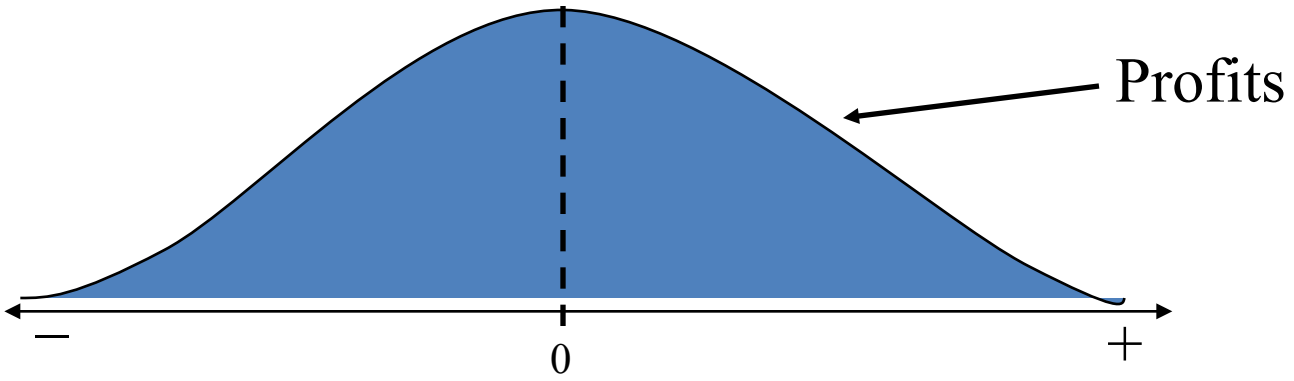
Applying propensity models to the sales funnel



Scoring	Explanation
🟢	high probability to buy
🟡	medium probability to buy
🔴	low probability to buy
⚫	has the solution already
*)	no information about solutions

Linear Regression Assumption

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



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

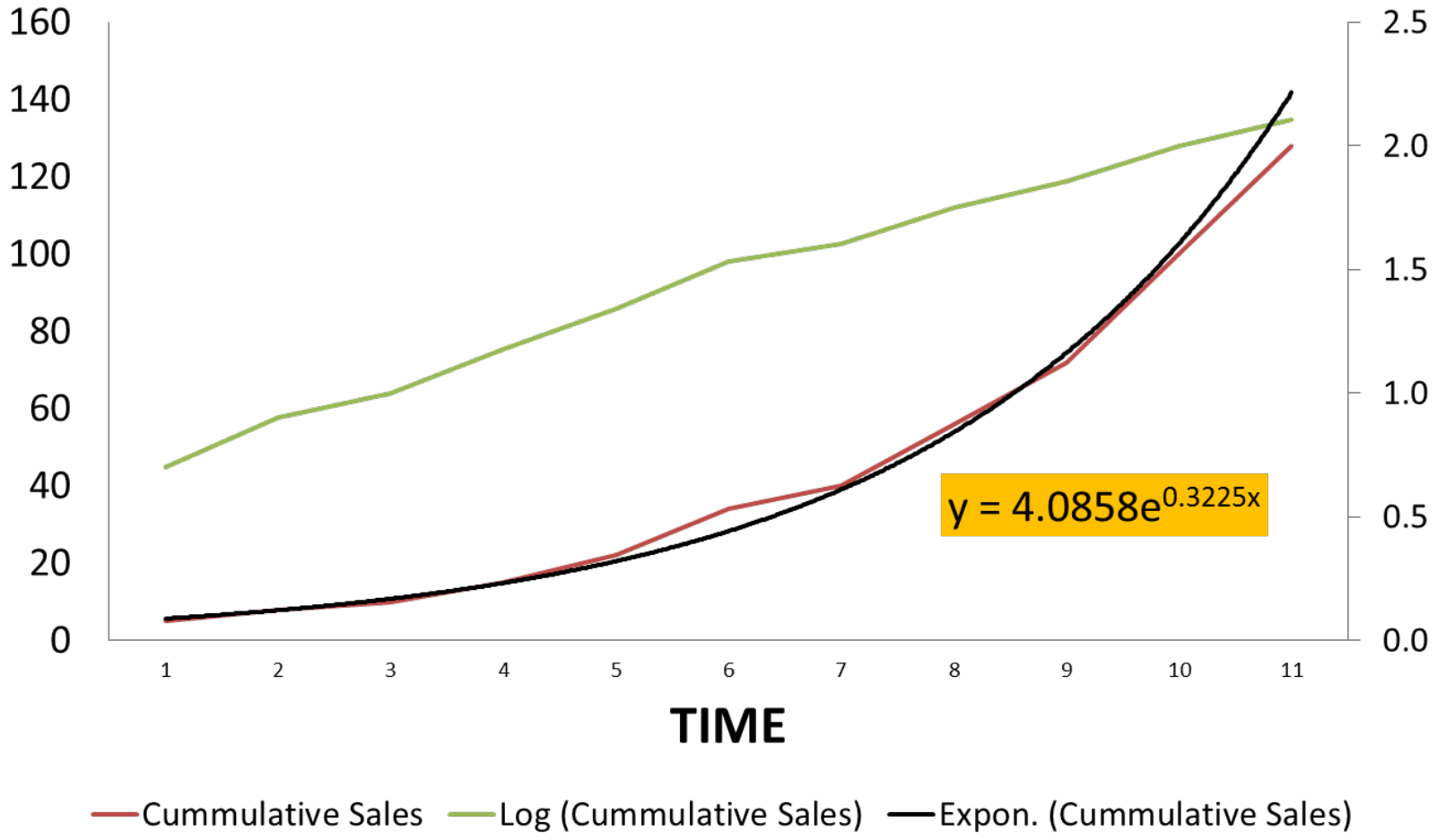
Customer Retention: Logistic Regression

- 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

Understanding S-Shape

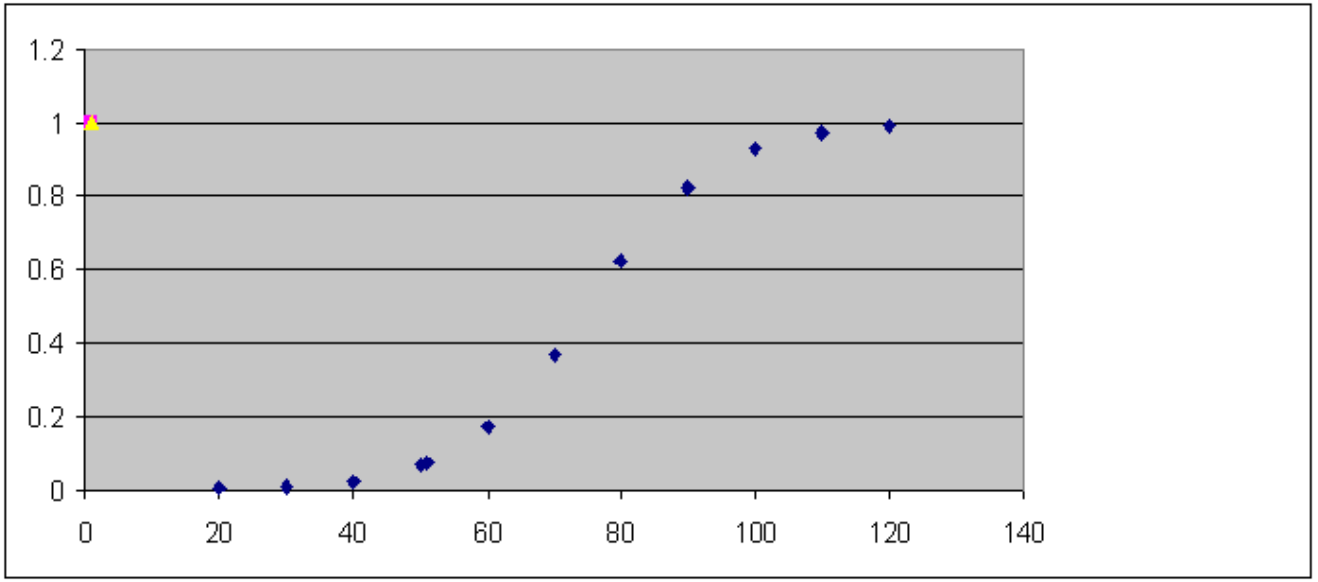
Cumulative Sales of Ultrasound Since Introduction (Millions)



Logistic Regression—How Do We Get the S-Shaped Form?

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

Predictions are bound between [0,1]



Choice Probability

Price Discount on $877 Price of Miami Flight

Best Buy

Example

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

sku	game	numsales	abmedian	browsetime	new	regular price	customer review count	customer review 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
R^2(McFadden	0.000	0.082
R^2(Cox and S	0.000	0.107
R^2(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

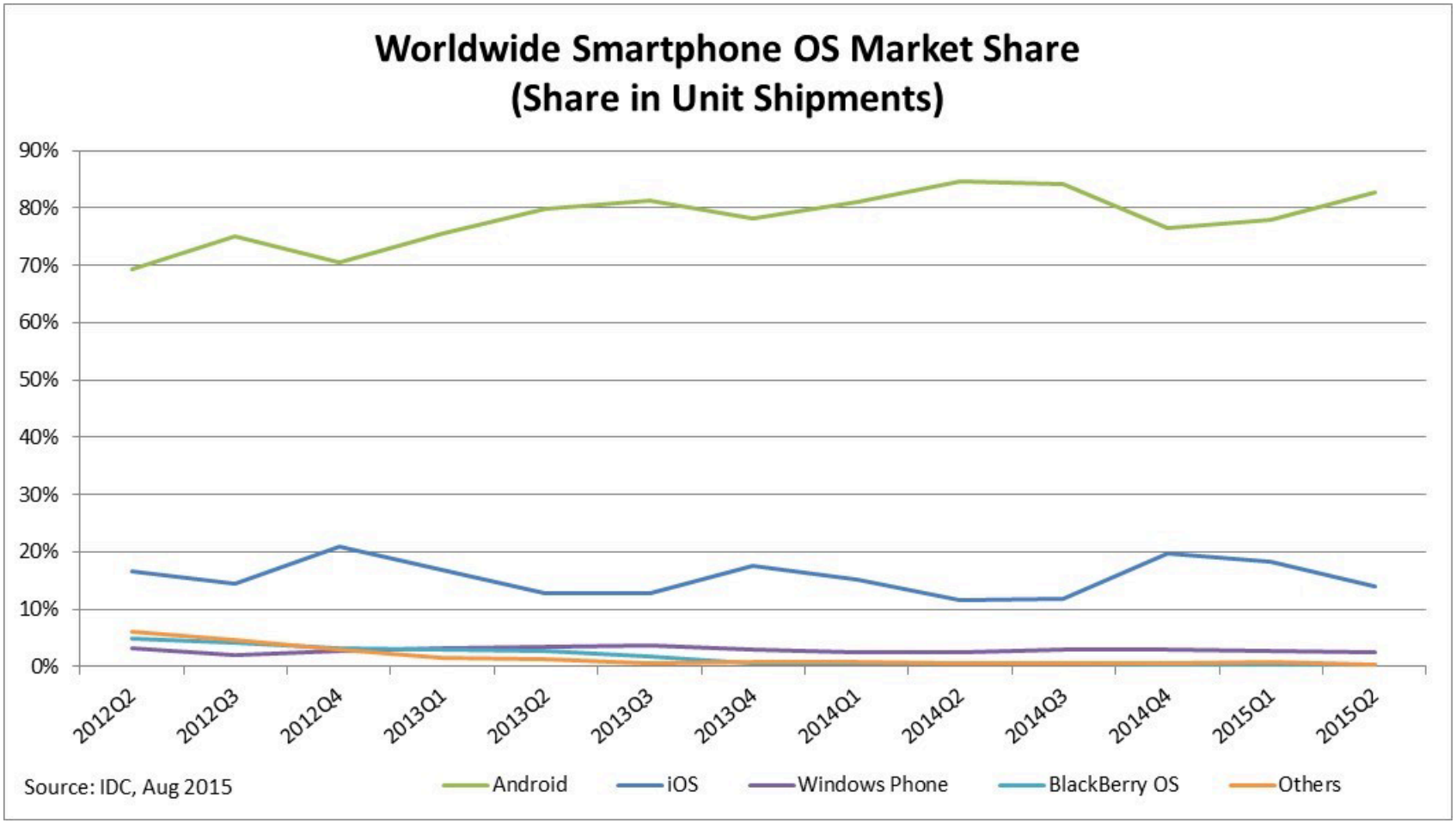
Interpreting Coefficients of Logistic Regression

Output: Sales of Xbox Games

Coefficient of Customer Review Average (b_{review})	0.399	
	Customer Review Average = 3	Customer Review Average = 4
U = a + bx	0.76	1.159
P(sale) = exp(u)/(1 + exp(u))	0.68	0.76
difference	0.079	

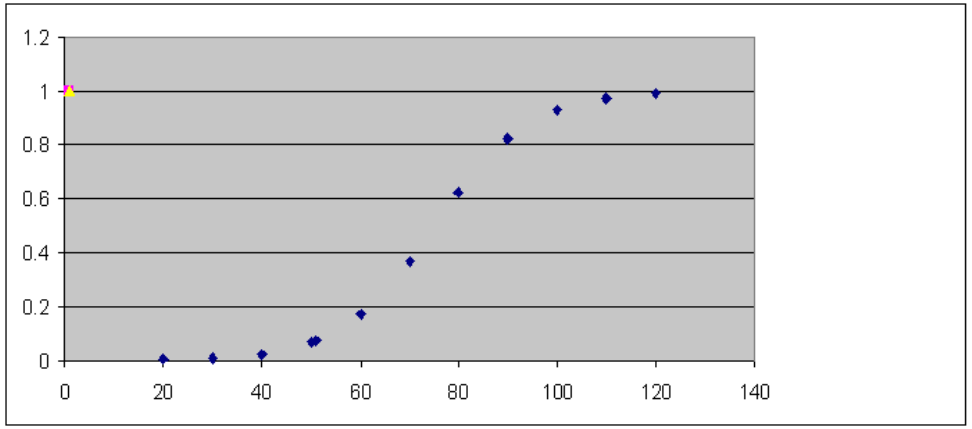
Market Share Predictions

What If One Has Only Share Data?



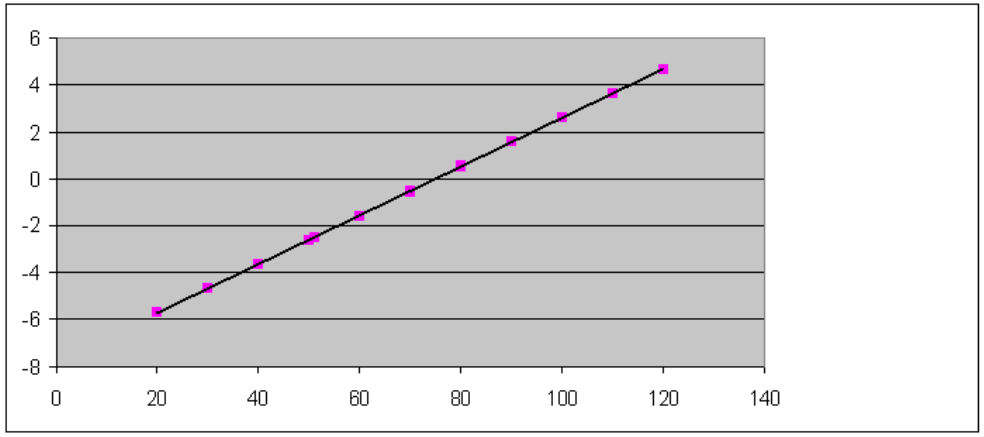
Logistic Regression

Logistic Distribution P(Y = 1)



Transformed, however, the "log odds" are linear.

$$\ln[p/(1 - p)]$$



Super Bowl 2016 Odds

Team	Odds Against
Green Bay Packers	6-1
Seattle Seahawks	13-2
Indianapolis Colts	8-1
New England Patriots	9-1
Dallas Cowboys	10-1
Denver Broncos	12-1

Source: Bovado (July 2015)

Logistic Regression

What if one does not have individual choice data but share data?

$$\text{Prob(Choosing Android)} = \frac{e^{(a+b_1X)}}{1+e^{(a+b_1X)}}$$

Predictions are bound between [0,1]

This is called → the "odds"

$$\frac{P}{1-P} = e^{a+b_1X} \longleftarrow$$

Chance of choosing to chance of not choosing

$$\text{where, P = Share of Android OS}$$

This is called → the "log odds"

$$\ln [p/(1-p)] = a + b_1\text{Price} + b_2 \text{ \# Vendors}$$

Hit Rates

Hit Rates in Sample

		Observed	
		Above median	Below median
Predicted	Above median	Match	Mismatch
	Below median	Mismatch	Match

Hit Rates in Sample

		Observed	
		Above median	Below median
Predicted	Above median	Match	Mismatch
	Below median	Mismatch	Match

Hit rate = (# Matches)/(Total # of Predictions)

Hit Rates in Sample

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

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%

Beers and Diapers

Correlation vs. Causation

- Does skipping breakfast cause obesity?

  http://www.webmd.com/diet/news/20080303/eating-breakfast-may-beat-teen-obesity

- Alternative explanations:
  - Physical activity
  - Lack of sleep

What Establishes Causality?

- Change in marketing mix produces change in sales
  - Increasing Advertising $ ⟶ Increased Sales
- No sales increase when there is no change in the marketing mix
  - No Increase in Advertising $ ⟶ Same Sales
- Time Sequence
  - Increased advertising $ today leads to higher sales tomorrow.
- No other external factor
  - When advertising was increased, one of the competitors left the market. So sales increased because of lesser competition, not because of increased advertising.