Joerg Niessing
Affiliate Professor of Marketing

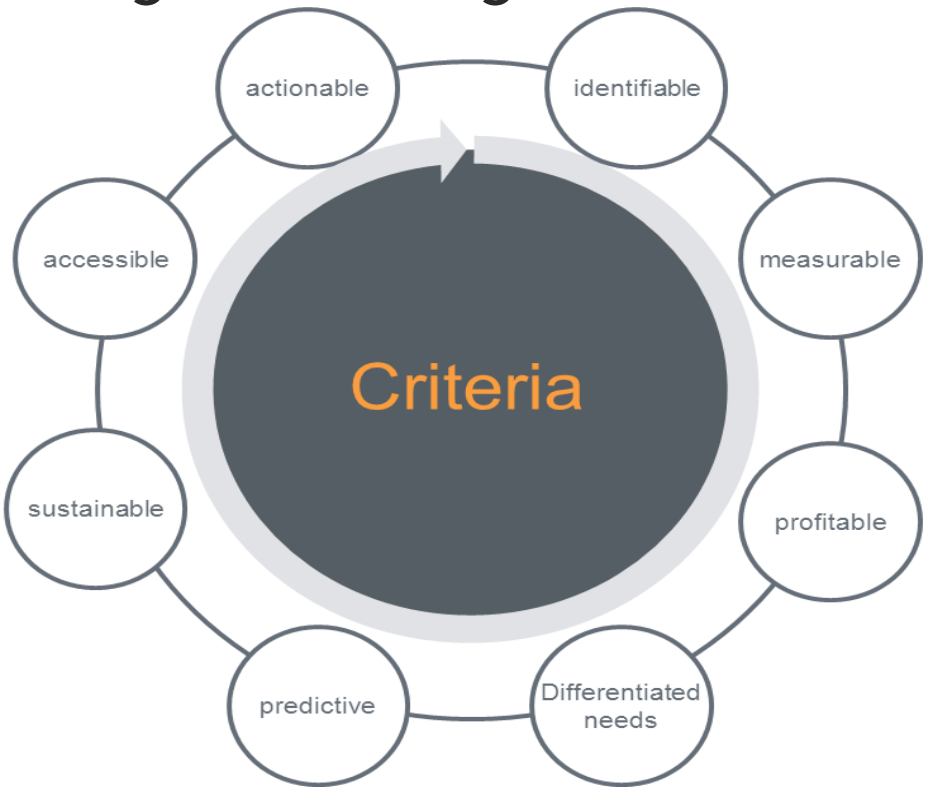
Theos Evgeniou
Professor of Decision Sciences

[Big]-Data Analytics for Businesses

What Makes a “Good” Segmentation?

What Makes a "Good" Segmentation?

Many different evaluation systems exist.
Most deem a segmentation good if it meets these criteria



Class Outline: three tools you will learn

1. Finding important factors that summarize your data, and visualizing your data:

   Factor Analysis (Sessions 2 and 3)

2. Finding a few clusters of similar data:

   Cluster Analysis (Sessions 4 and 5)

3. **Discriminating among and predicting successes vs failures:**

   Logistic Regression and Tree Analyses (Sessions 6 and 7)

Other Examples

- Who are most likely to click on an ad?
- Who are likely to respond to a direct mail campaign? What distinguishes those who responded to previous direct mail compared to those who do not?
- How are satisfied customers different from dissatisfied customers in terms of their demographics and attitudes towards your products' characteristics?
- Who are likely to default on a loan?
- To whom should we offer a particular promotion?
- Which transaction is most likely a fraud?
- Which applicants are most likely to fit in our organization and succeed?
- Which drug development project should we mainly invest in?

What is common to these problems?

- There is a dependent variable which is categorical
  - e.g. success vs failure (fit vs. non-fit; fraud vs. non-fraud, response vs. non-response, etc.)
- There are some independent variables which we can use to explain membership in the different categories.

Discrimination: 7 (Easy) Steps

1. Create an estimation and two validation samples in a balanced way
2. Set up the dependent variable ("what is a success? What is a failure?")
3. Select the independent variables
4. Estimate model (many methods, we do 2 here)
5. Assess significance of variables (tricky…)
6. Assess performance on first validation data
7. Repeat steps 2-6 with variations till performance on first validation data is satisfying
8. Assess performance on second validation data once

Various Methods

- Logistic regression
- Classification trees
- Discriminant Analysis (DA)
- Neural Networks
- Bayesian methods
- Support Vector Machines
- ….

Regression Modeling

- Regression is a technique that can be used to investigate the effect of one or more predictor variables (independent) on another variable (dependent).
  - Regression allows you to make statements about how well one or more independent variables will predict the value of a dependent variable.
  - The multiple regression equation takes the form.

$$y = b1*x1 + b2*x2 + c$$

y =dependent

b=regression coefficients, representing the amount the dependent variable y changes when the corresponding independent changes 1 unit

c=constant, representing the amount the dependent y will be when all the independent variables are 0.

Logistic Regression: The Basic Idea

Model the probability p of being a "success":

$$p = \frac{1}{1 + \exp(-V)}$$

$$V = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

- This function for p is called logistic function
- Logistic Regression model is popularly also called the logit model

Logistic Regression: The Basic Idea

The estimated probability is:

$$p = 1/[1 + \exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$$

If $(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$ =0, then p = .50

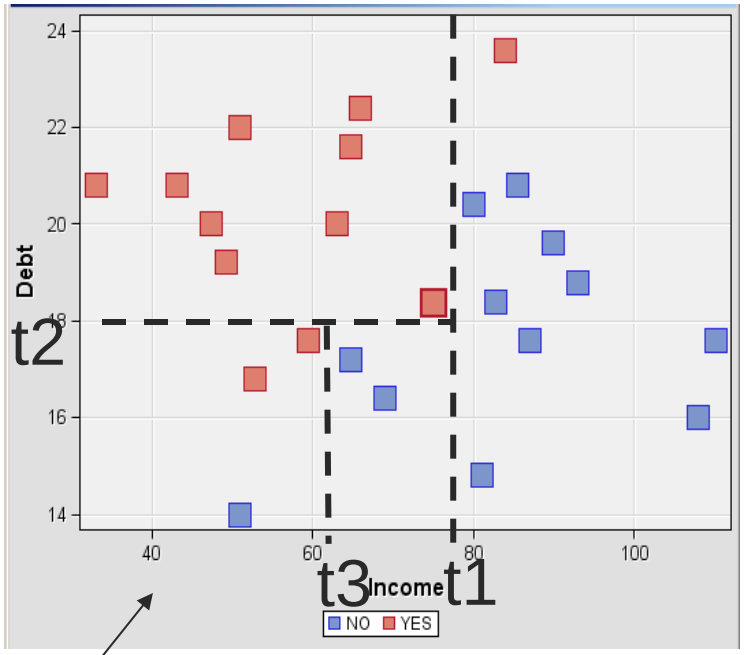
Logistic Regression Example

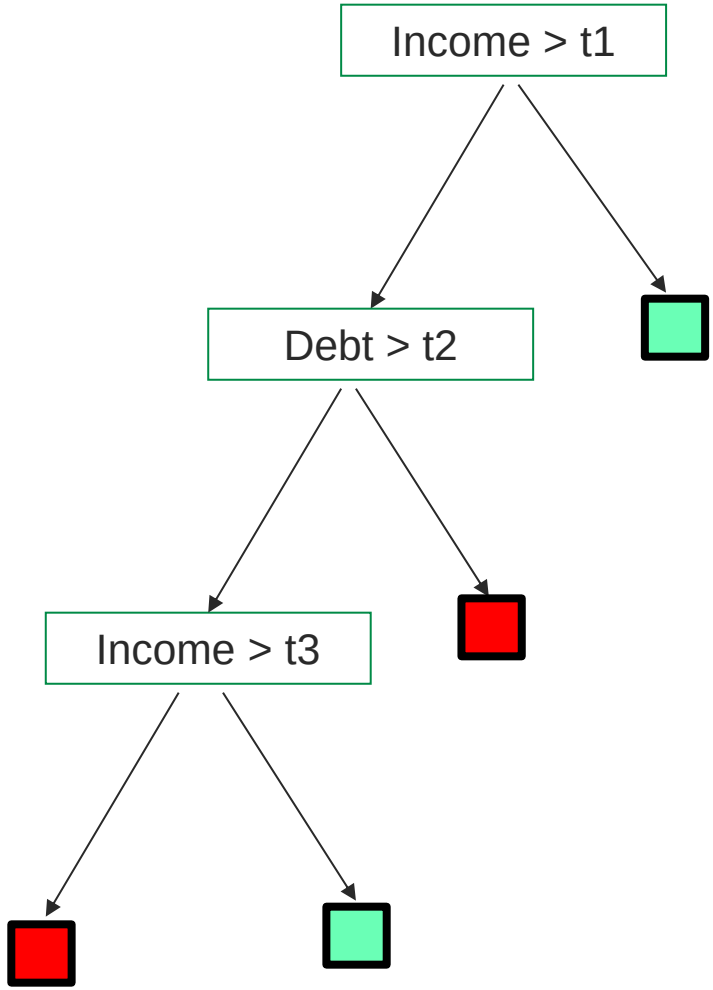
V=b0+b1X1+b2X2	p=1/(1+EXP(-V))	Classification
-0.11748	0.470663733	0
8.99804	0.999876363	1
2.1614	0.896729269	1
-1.6829	0.156711843	0
8.7704	0.999844763	1
2.93239	0.94942456	1
-2.9667	0.048953132	0
2.38904	0.915987722	1
-3.19434	0.039379275	0
-1.0117	0.26664729	0
0.88115	0.707060473	1
-3.51822	0.028798239	0

Various Methods

- Logistic regression
- Classification trees
- Discriminant Analysis (DA)
- Neural Networks
- Bayesian methods
- Support Vector Machines
- ….

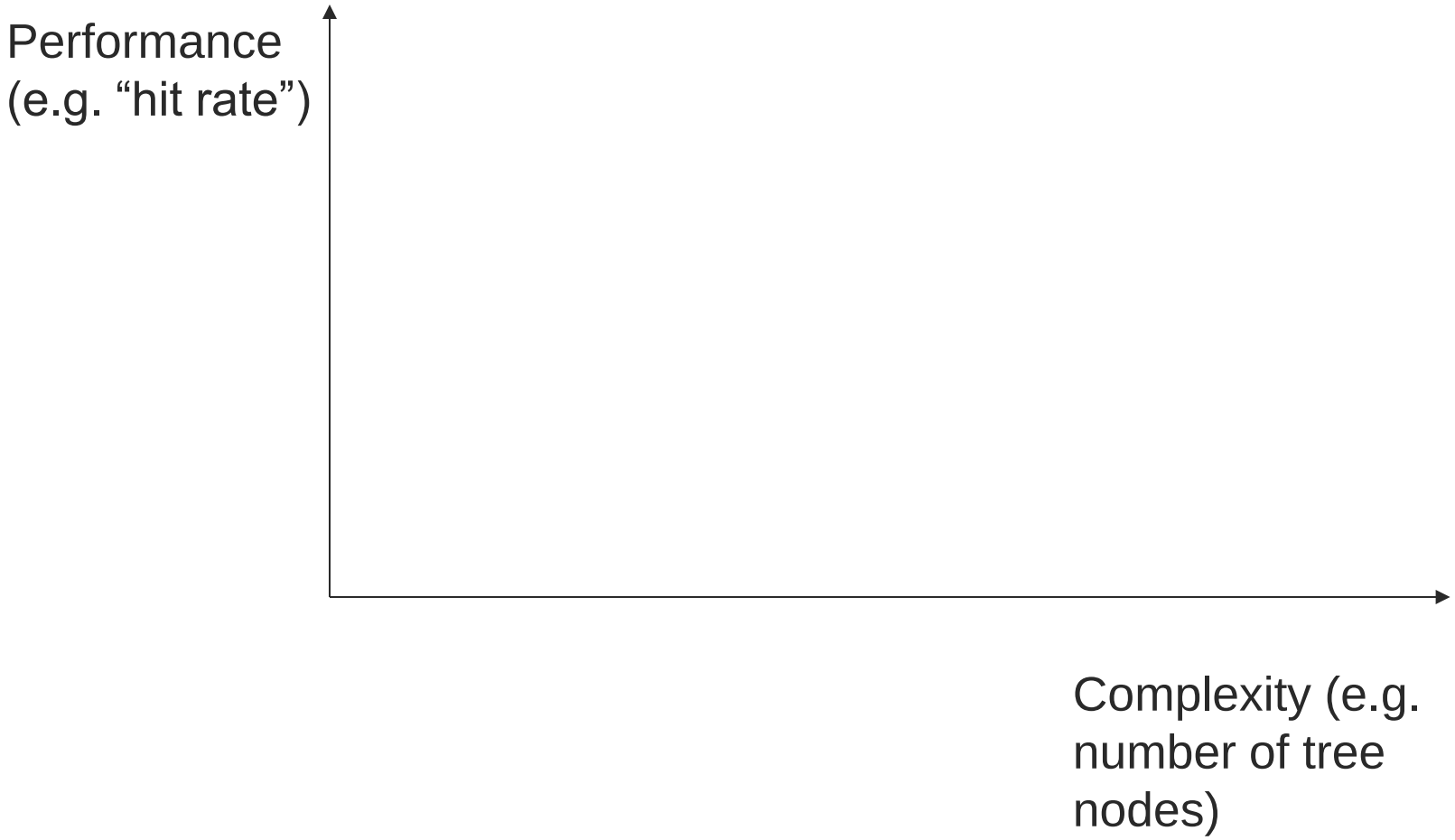
Classification Trees





Note: tree boundaries are piecewise linear and axis-parallel

Complexity and Predictability:
KEY SLIDE



Discrimination: 7 (Easy) Steps

1. Create an estimation and two validation samples in a balanced way
2. Set up the dependent variable ("what is a success? What is a failure?")
3. Select the independent variables
4. Estimate model (many methods, we do 2 here)
5. Assess significance of variables (tricky…)
6. Assess performance on first validation data
7. Repeat steps 2-6 with variations till performance on first validation data is satisfying
8. Assess performance on second validation data once

Example in SPSS: The Coffee Project

BRAND PERCEPTION

16. Below are several statements that could describe a brand (products, image, reputation, etc). Please indicate how much you agree or disagree with each of the statements (respondents have seen a brand that they own or a brand that they are familiar with):

Scale:

1. Strongly Disagree
2. Disagree
3. Neither Agree or Disagree
4. Agree
5. Strongly Agree

Statements:

1. Has been around for a long time
2. Best in-class customer service
3. Strong dealer network
4. Offers cutting edge technology
5. Leader in safety
6. Offers innovative products
7. Is for people who are serious about boating
8. Is good for beginners
9. Is a brand I see in the water all the time
10. Provide a fast and powerful boating experience
11. Is great for socializing
12. Is great for water sports
13. Superior interior style
14. Superior exterior style
15. Stands out from the crowd
16. Offers boats that look cool
17. Can easily handle rough weather or choppy water
18. Can handle frequent and heavy usage
19. Offers a wide breadth of product offerings and accessories
20. Offers boats that I can move around safely
21. Boats are easy to maintain
22. Boats are easy to use
23. Boats are easy to clean up
24. Is low priced

ENGINES : BRAND FUNNEL

Now we would like to ask you some questions about boat engines

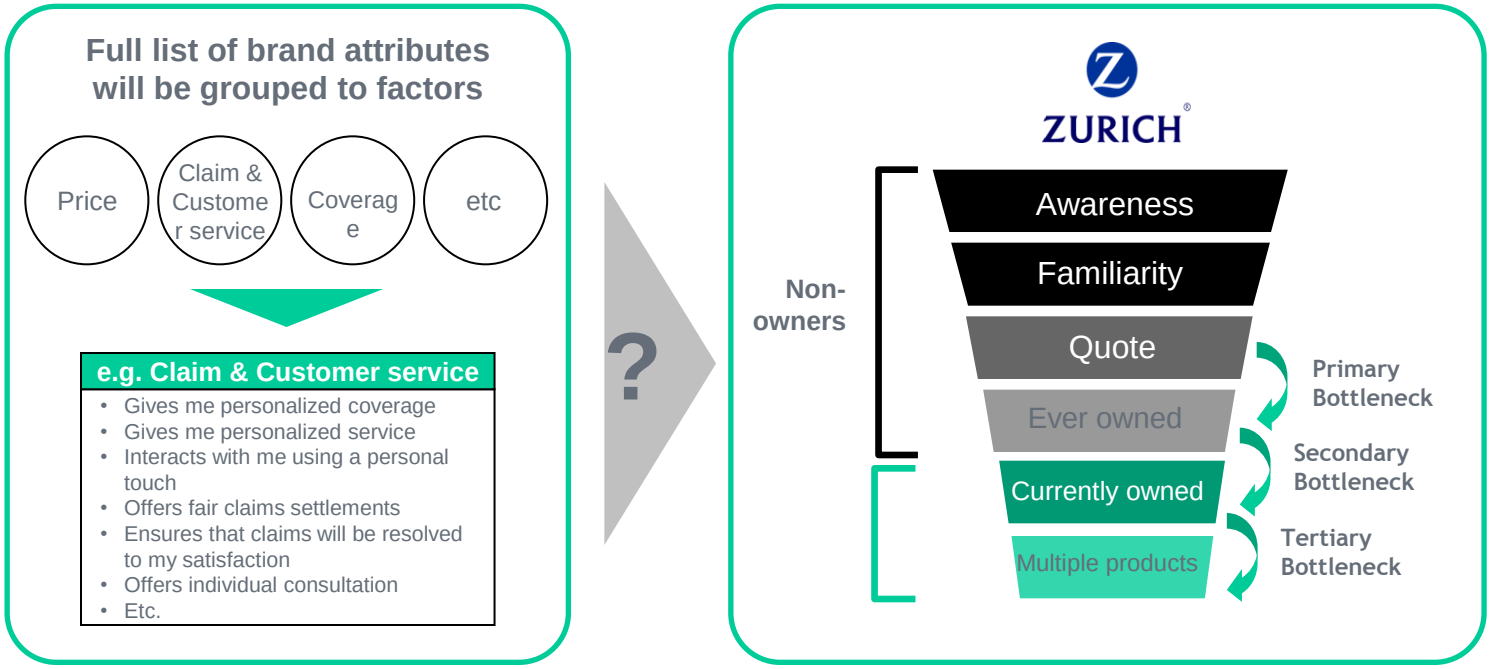
39. Which of the following marine engine brands have you **ever heard of**? (Select all that apply)

1. BRP Evinrude
2. Crusader
3. Cummins
4. CMD
5. Honda
6. Johnson
7. MerCruiser (Mercury)
8. Mercury
9. Parsons
10. Suzuki
11. Torqeedo
12. Volvo
13. Volkswagen TDI
14. Yamaha
15. Yanmar
16. Other. Please specify_____

FOR Q40 SHOW BRANDS SELECTED IN Q39

40. Which of the following brands are you **familiar with**? By familiar we mean knowing some information beyond brand name and logo? (Select all that apply)

Use of purchase funnels and regression modeling allows us to derive key drivers based on brand attributes



1. From the funnel section we are able to learn which are the critical bottlenecks across the purchase process by country

2. Therefore, to determine what the key attributes are, we analyze which attributes drive customers through these **specific bottlenecks by country**

Purchase drivers will be compared by segment

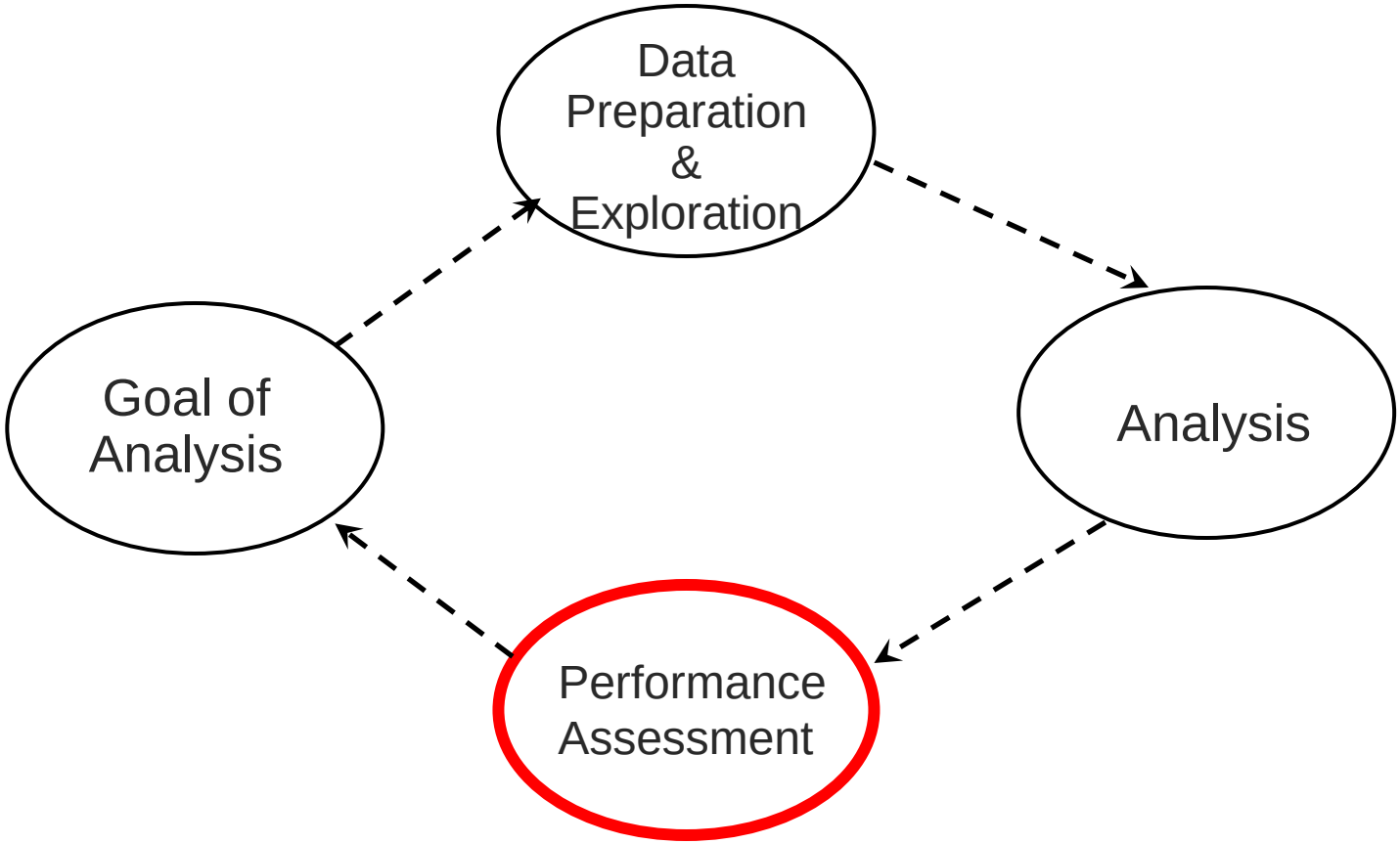
Index to Mean	Overall Drivers	Segment 1	Segment 2	etc
Bottleneck	Quote → Own	Quote → Own	Quote → Own	etc
Above average	•Customer focused (144%) •Self-expression (137%) •Claim Process (130%) •Prestige & Leadership (118%)	•Self-expression (158%) •Claim Process (136%) •Customer focused (136%) •Prestige & Leadership (122%)	•Customer focused (170%) •Prestige & Leadership (146%) •Global Presence (141%) •Claim Process (127%) •Self-expression (124%)	• etc
Average	•Global Presence (107%) •Approachable (106%) •Good relationship managers (97%) •Technology support (96%) •Innovation (93%)	•Innovation (101%) •Good relationship managers (98%) •Technology support (96%) •Stable (93%)	•Innovation (94%) •Approachable (93%) •Good relationship managers (91%)	• etc
Below average	•For everybody (87%) •Wide offering range (82%) •Stable (81%) •Value for money (75%) •Independence (73%) •Old-fashioned (73%)	•Approachable (90%) •Global Presence (84%) •Value for money (84%) •Independence (82%) •Wide offering range (80%) •For everybody (74%) •Old-fashioned (65%)	•Technology support (88%) •Wide offering range (84%) •Old-fashioned (76%) •For everybody (72%) •Stable (71%) •Value for money (66%) •Independence (65%)	• etc

Common drivers across segment

Boating Case: Part II

Boating results in excel

The Eternal Iterative Process Cycle



Performance Measures

- Hit ratio
- False positive
- False negative
- Confusion Matrix
- Maximum chance criterion
- Lift curve
- “Profit matrix”
- “Profit curve”

Confusion Matrix

Confusion matrix for the estimation sample:

from \ to	0	1	Total	% correct
0	12	1	13	92.31%
1	2	15	17	88.24%
Total	14	16	30	90.00%

False positive: 7.69%

False negative: 11.76%

Hit Ratio

Maximum Chance Criterion:

Is the hit ratio better than the proportion of the largest group?

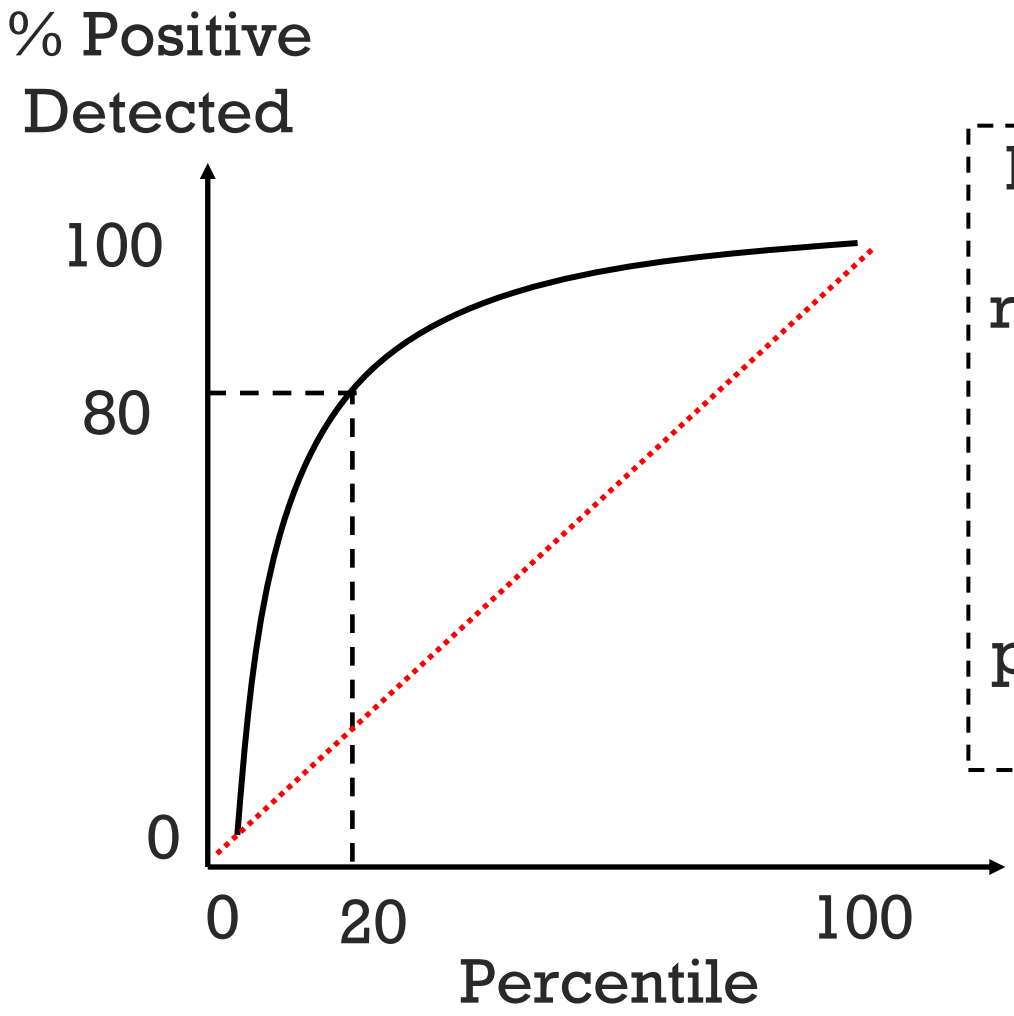
(Otherwise we could assign all observations to the largest group!!)

Prior and posterior classification, and membership probabilities

Observation	Prior	Posterior	Pr(0)	Pr(1)
Obs4	1	**0**	0.998	***0.002***
Obs5	1	1	0.010	***0.990***
Obs12	1	1	0.000	***1.000***
Obs19	0	0	0.979	***0.021***
Obs20	0	0	0.952	***0.048***
Obs21	0	0	0.999	***0.001***
Obs22	0	**1**	0.068	***0.932***

What should be the cut off probability to decide whether or not "the customer is good"?

Lift Curve: “Moving” the cut off



How many trials (i.e. emails, recommendations, investments, etc) should we do in order to capture a large percentage of all the successes?

"Profit Matrix"

		Actual	
		0	1
Predicted	0	0	-$20
	1	-$20	$20

What is now the $ performance?

To Whom to Give Credit?

- **What if the "cost" of Act0_Pred1 is 0?**

  Give credit to everyone!

- **What if the "cost" of Act0_Pred1 is huge?**

  Give credit to noone!

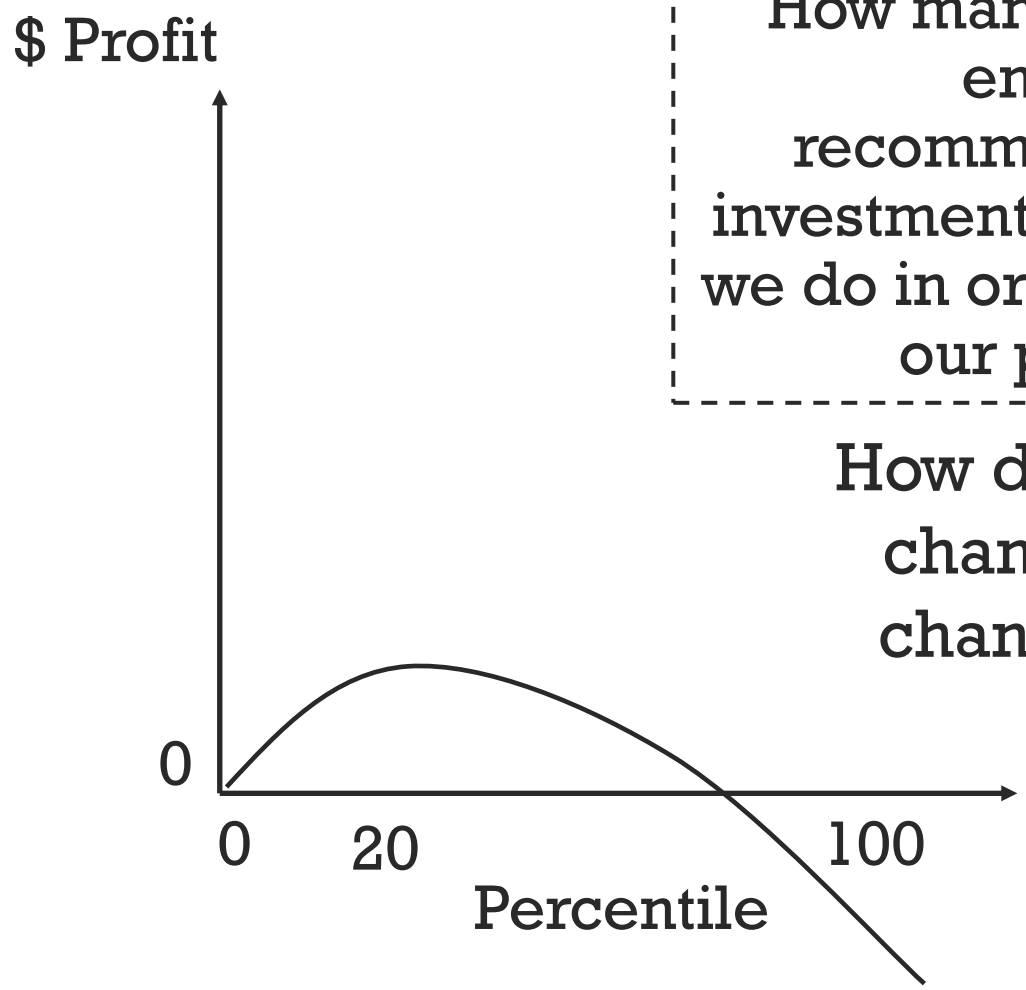
- **What happens in the mid-range?**

  Give credit to some

- **What if the "cost" of Act1_Pre0 is huge?**

  Give credit to everyone!

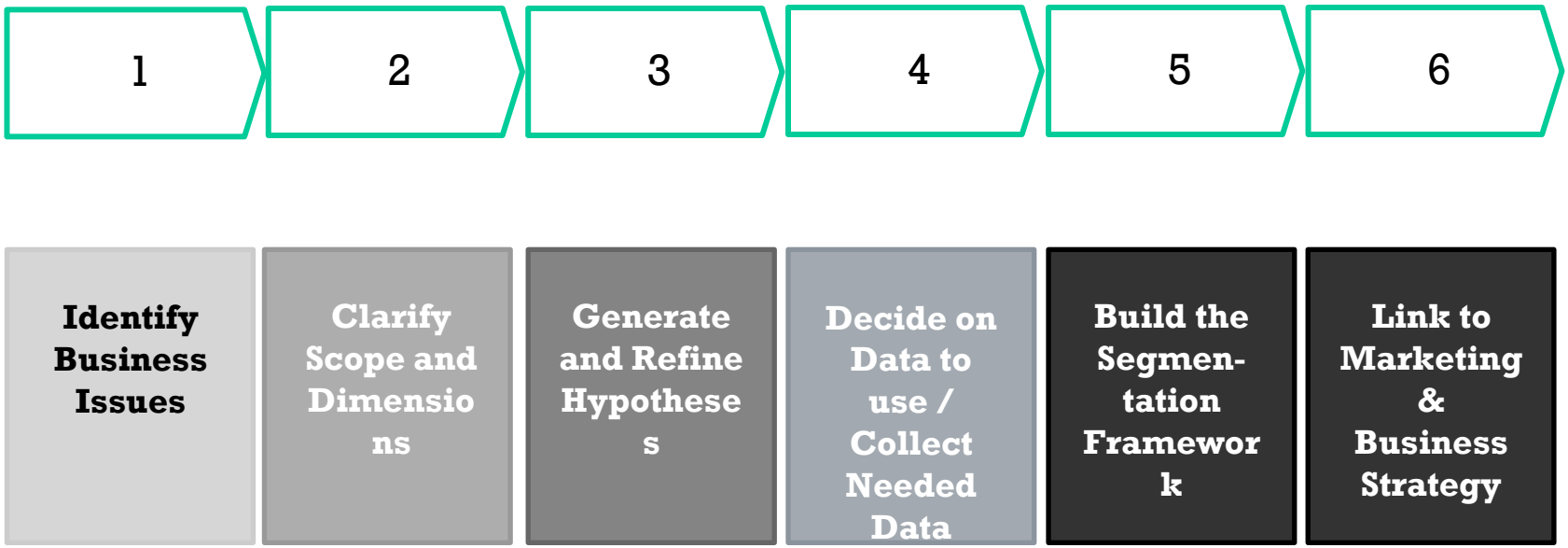
Profit Curve: "Moving" the cut off



How many trials (i.e. emails, recommendations, investments, etc) should we do in order maximize our profits?

How does the curve change when we change the profit matrix?

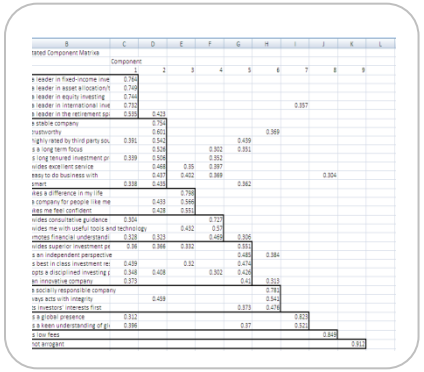
Segmentation Methodology – A(nother) Process

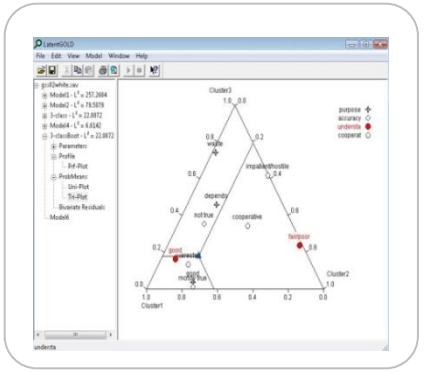


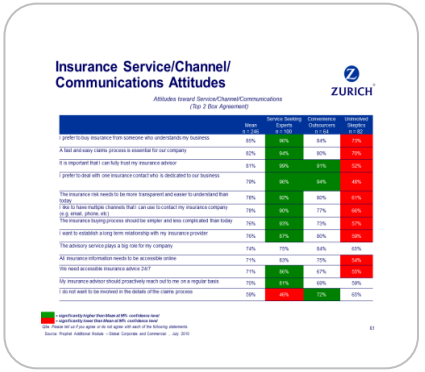
Step 5 – Build the Framework

- Segmentation solution is created through a rigorous and iterative process

Data Processing/ Factor Analysis → Cluster analyses ⇄ Review and refine







Step 6 – Link to Business Strategy

- Just building the segmentation is only half the battle

1 Product/Offer	2 Pricing/ASS	3 Marketing	4 Dealer/Channel
Breadth & Depth of Offer	Pricing & Financing	Marketing Communication	Channel Store/Environment
Product Requirements	Customer Service & Support	Events	Inventory Management
Product Naming & Architecture	Maintenance & Repair	Licencing	Dealer Network
			Dealer Relationship Building

Next Class (last class!)

Review and Project Presentations

INSEAD

The Business School
for the World®