Understanding Consumer Behaviour – what can statistics tell us?

Aspasia Angelakopoulou and Vivian Chiu
Swiss Re

14 May 2013

Agenda

- Consumer Behaviour

- Studying Consumer Behaviour
  - Non disclosure
  - Lapse

- What else can be done?

- How statistics can help insurance companies?

14 May 2013

Consumer Behaviour

Institute and Faculty of Actuaries

14 May 2013

What Is Consumer Behaviour?

"*Consumer behaviour* is the study of individuals, groups, or organizations and the processes they use to select, secure, and dispose of products, services, experiences, or ideas to satisfy needs and the impacts that these processes have on the consumer and society"
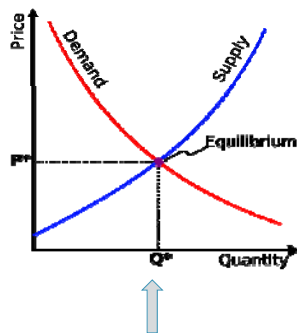
Kuester, Sabine (2012): *MKT 301: Strategic Marketing & Marketing in Specific Industry Contexts*, University of Mannheim, p. 110.

Institute and Faculty of Actuaries

14 May 2013

Studying Consumer Behaviour

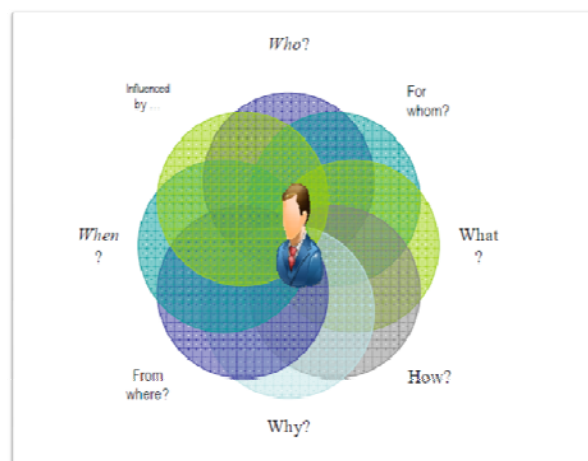| **Understanding market trends** | **Qualitative direct market research** | **Quantitative direct market research** |

Institute and Faculty of Actuaries

14 May 2013

5

Understanding Consumer Behaviour

Answers to these questions should not be taken separately but in correlation!
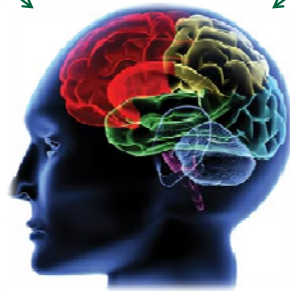
Institute and Faculty of Actuaries

14 May 2013

6

Factors Influencing Consumer Behaviour

Endogenous
- Perception
- Learning
- Attitudes
- Motivation
- Personality
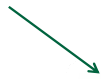
Exogenous
- Situational
- Marketing stimuli
- External macro-medium
- Personal characteristics

Institute
and Faculty
of Actuaries

14 May 2013

7

What can we actually measure?

- **Endogenous factors** usually tricky to measure

- **Exogenous factors** can be directly observable for example:

  – Information on the atmosphere in which transaction closed, unexpected events at the moment of purchase such as change in the premium

  – Information on distribution channel, premiums, process of communication

  – Information on factors of economic or technological nature (e.g. current economic crisis)

  – Information on sociological factors (e.g. social class, region) as well as lifestyle factors (e.g. bmi, alcohol, smoking, age)

Institute
and Faculty
of Actuaries

14 May 2013

8

Institute
and Faculty
of Actuaries

Studying Consumer Behaviour

14 May 2013

Qualitative vs. Quantitative Market Research

Which of the two is most important?

Your Opinion

☐ Strongly agree

☑ Agree

☐ Neither agree nor disagree

☐ Disagree

☐ Strongly disagree

OR

Institute
and Faculty
of Actuaries

14 May 2013

Statistics….

- Qualitative precedes quantitative research
- We could use the term "statistical analysis" instead of "quantitative analysis"
- So why is statistics so important?
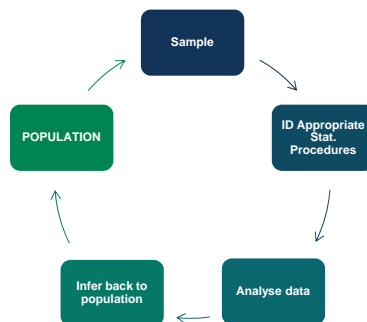
14 May 2013

Institute and Faculty of Actuaries

11

Why is statistics interesting for Insurance?

- Non-disclosure patterns
- Lapse rates
- Non Take Up rates
- STP rates
- Benchmarking
- Segmentation
- Mortality / Claims rates

….and others

Institute and Faculty of Actuaries

14 May 2013

12

The data

- Insurance data available over a 5 year time period for Accelerated Critical Illness

- Followed policy holders across all the way starting from the moment of application, through to in-force until they lapsed

- Data available on more than 700,000 in-force policies

- Information available on a number of factors at both *personal* and *company* level

- Artificially generated dataset!

Institute and Faculty of Actuaries

14 May 2013   13

Levels of data

Personal Characteristics
- Age
- Gender
- Smoking
- Alcohol
- BMI
- Impairment counts
- Occupation class
- Region
- Marital status
- Family history
- Medical concepts

Company Specific Characteristics
- Distribution channel
- Sum assured
- Rated
- Underwriting decision
- Company performance

Institute and Faculty of Actuaries

14 May 2013   14

Methodology

- Fitted three types of *Generalized Linear Models* (GLMs) depending on the question of interest. These include:
  - Negative Binomial models
  - Logistic models
  - Poisson models
- GLMs are useful when the relationship between our outcome of interest and the predictors is not linear
- In simple linear regression models we have the following:

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- This is NOT the case with GLMs as we need to take a transformation of the outcome and model that instead of the outcome itself
- This transformation is called the *link function* and it varies depending on the nature of the outcome variable e.g.
  - In Negative Binomial and Poisson models the link function is *log*
  - In Logistic models this is *logit*

Institute
and Faculty
of Actuaries

14 May 2013

15

Institute
and Faculty
of Actuaries

Non Disclosure

14 May 2013

16

Non disclosure

- Failure or refusal to declare or reveal some information that is required to be declared or revealed

- People often tend to non-disclose on a number of factors such as
  - number of cigarettes smoked
  - number of alcohol units consumed
  - weight

- Important to capture it as it can be used for pricing purposes or client information

Institute
and Faculty
of Actuaries

14 May 2013

Methodology

- Fitted *Negative Binomial* models, a member of the GLM family
- Useful to model counts, e.g. number of alcohol units consumed, number of cigarettes smoked etc
- General form of the model will be:

$$\log(counts) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

- As mentioned before, *log* link function is used to make the association between outcome and predictors linear (log-linear models)
- Models with more than one factors were fitted (i.e. multivariate models)
- *Poisson* models are usually used for counts modelling but due to their strong assumptions sometimes they fail to fit
- In these cases Negative Binomial models are the often the next best alternative

Institute
and Faculty
of Actuaries

14 May 2013

18

Weekly alcohol consumption

Compare actual and predicted weekly alcohol consumption
(adjusted for gender, region, occupation, cigarette consumption, age, sum assured, family history and medical concepts)

Evidence of non-disclosure from company E?

Institute
and Faculty
of Actuaries
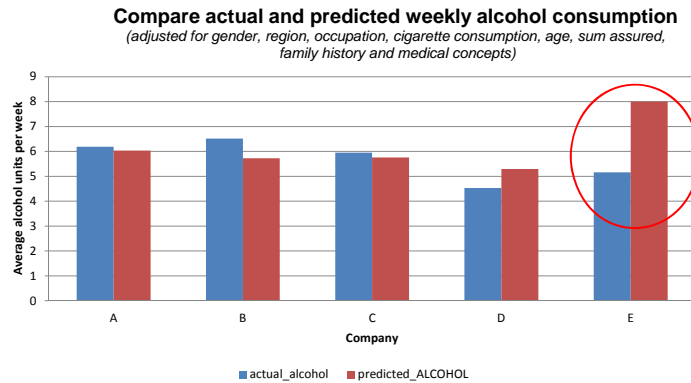
14 May 2013

19

Daily Cigarettes Consumption

Compare actual and predicted weekly cigarettes consumption
(adjusted for gender, region, occupation, age, sum assured, family history, medical concept and alcohol)

Evidence of non-disclosure from company E?

Institute
and Faculty
of Actuaries
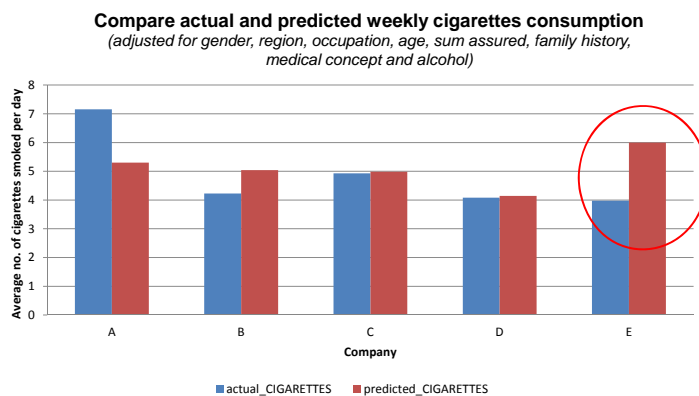
14 May 2013

20

BMI Disclosure

Compare actual and predicted BMI disclosure
(adjusted for gender, region, occupation, cigarette consumption, age, sum assured, family history, medical concepts and alcohol)
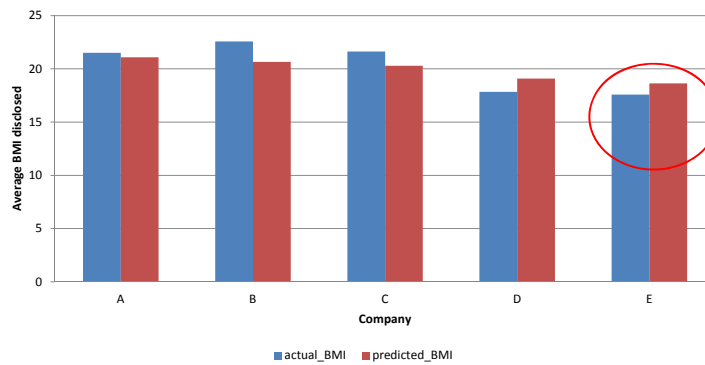
Slight evidence of non-disclosure from company E?

Institute
and Faculty
of Actuaries

14 May 2013   21

Comments

- Open dialogue with the company to identify whether true non-disclosure exists

- If non-disclosure does exist, the following actions may be considered:
  - which agent(s) is the culprit
  - what is the extent of non-disclosure
  - what are the causes of non-disclosure
    - e.g. agents / policyholders not understanding alcohol units
    - e.g. agents understanding the importance of accurate information

Institute
and Faculty
of Actuaries

14 May 2013   22

Lapses

Lapses

- Occurs when a policy stops being inforce for reason not related to death or claims

- A number of factors can affect lapse behaviour such as
  - smoking
  - age
  - gender
  - sum assured
  - rated
  - ….. and others

- Important to capture consumer lapse behaviour as it affects profitability of the company

Methodology

- Fitted *Poisson* regression models, a member of the GLM family

- Used to model time to a binary event i.e. lapse rates

- Link function=*log* & distribution= *poisson*

$$\log(\text{rate}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

- Modelling one Risk Factor (RF) at a time

- Produced Rate Ratios (RRs) with corresponding 95% Confidence Intervals (CIs)

Institute
and Faculty
of Actuaries

14 May 2013   25

Lapse RRs for Personal Characteristics
(single factor models)

RFactors	Comparator	N		RR (95% CI)
Occupation				
2	vs 1	93508		1.16 (1.14, 1.18)
3	vs 1	130899		1.10 (1.08, 1.12)
4	vs 1	87134		1.24 (1.22, 1.26)
5	vs 1	13565		1.16 (1.12, 1.21)
H	vs 1	18927		1.40 (1.36, 1.45)
Age(yrs)				
>=16	vs 26-36	98550		1.22 (1.20, 1.23)
>=36	vs 26-36	294416		0.92 (0.91, 0.92)
46-56	vs 26-36	116554		0.86 (0.84, 0.87)
>=56	vs 26-36	19577		0.90 (0.88, 0.93)
Alcohol(units)				
2	vs 1	148445		1.00 (1.00, 1.01)
3	vs 1	32705		0.97 (0.95, 0.99)
4	vs 1	3294		0.90 (0.85, 0.96)
5	vs 1	1067		0.83 (0.74, 0.94)
BMI(kg/m2)				
<20	vs 20-25	46438		1.03 (1.01, 1.05)
25-30	vs 20-25	350497		1.00 (0.99, 1.01)
30-35	vs 20-25	100202		0.98 (0.97, 0.99)
>=35	vs 20-25	12161		0.83 (0.80, 0.86)
Impair.count				
1 count	vs no counts	224975		0.83 (0.82, 0.84)
2 counts	vs no counts	71336		0.76 (0.74, 0.77)
3 counts	vs no counts	22346		0.69 (0.68, 0.71)
>4 counts	vs no counts	10575		0.62 (0.60, 0.65)
Marital Status				
Single	vs Married	353669		0.89 (0.89, 0.90)
Gender				
Females	vs Males	435855		0.96 (0.95, 0.97)
Smoker				
Yes	vs No	153644		1.41 (1.40, 1.43)

0   1   1.5

Institute
and Faculty
of Actuaries

14 May   26

13

Lapse RRs for Company Characteristics
(single factor models)

RFactor	Comparator	N		RR (95% CI)
Company				
A	vs average			0.97 (0.96, 0.98)
B	vs average			1.20 (1.19, 1.21)
C	vs average			1.09 (1.07, 1.10)
D	vs average			0.83 (0.82, 0.85)
E	vs average			0.96 (0.96, 0.97)
Underwriting dec.				
Refer	vs Standard	272497		0.90 (0.89, 0.91)
Loaded	vs Standard	88147		1.06 (1.05, 1.08)
Follow up	vs Standard	28647		0.80 (0.78, 0.82)
Dist. channel				
BPOS	vs IFA	127088		0.93 (0.92, 0.94)
Direct	vs IFA	4284		0.58 (0.54, 0.62)
Point Of Sale	vs IFA	41898		0.79 (0.78, 0.81)
Partner	vs IFA	18706		1.10 (1.08, 1.13)
Planit	vs IFA	182530		0.96 (0.95, 0.98)
Bank	vs IFA	68129		0.71 (0.70, 0.72)
Sum assured(thousands)				
100-150	vs 0-100	182433		1.09 (1.08, 1.10)
150-250	vs 0-100	96826		1.09 (1.08, 1.10)
250-500	vs 0-100	20667		1.08 (1.05, 1.11)
>=500	vs 0-100	1739		0.82 (0.75, 0.90)
Rated				
50-100	vs <50	8167		0.95 (0.92, 0.99)
>=100	vs <50	871		0.64 (0.57, 0.71)

0   1   1.5

Institute
and Faculty
of Actuaries

Lapse RRs for Region
(single factor model)

Region	Comparator	N		RR (95% CI)
Channel Islands	vs average	3276		0.84 (0.79, 0.90)
Isle of Man	vs average	40847		1.01 (0.99, 1.03)
London	vs average	34348		1.03 (1.01, 1.05)
Midlands	vs average	142392		1.07 (1.05, 1.08)
North	vs average	145744		1.10 (1.09, 1.12)
Northeast & Cumbria	vs average	39446		1.03 (1.01, 1.05)
Scotland	vs average	83667		0.98 (0.97, 1.00)
South	vs average	43626		0.92 (0.91, 0.94)
SouthEast	vs average	94971		0.96 (0.94, 0.97)
SouthWest	vs average	38453		0.98 (0.96, 1.00)
Wales	vs average	42428		1.04 (1.02, 1.05)

0   1   1.5

Institute
and Faculty
of Actuaries

Key Messages

- **Smokers** have 41% higher lapse rates compared to non smokers
- Lapse behaviour appears similar between **males and females**
- **High risk jobs** seem to result to higher lapse rates than the low risk ones
- **Housepersons** have 40% higher lapse rates compared to low risk posts
- Number of **impairment counts** has a significant effect on lapse rates
- **Sum assured** band between 100-500K appears to lead to worst lapse behaviour than the one below 100K
- **Highly rated** policy holders have 36% lower lapse rates compared to low rated ones
- The majority of **regions** seem to have a good lapse behaviour compared to the average

Institute
and Faculty
of Actuaries

14 May 2013

29

Multivariate Analysis (MA)

- Important step of analysis as we cannot rely on crude analysis i.e. fitting one factor at a time
- MA involves fitting the model of interest with >1 predictors
- Model selection follows where the best model is selected
- Selection criteria vary, but some of the most popular ones include the *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC)
- Rule of thumb: "smaller is better"
- Select model with the smaller AIC or BIC value
- In this case **BIC** values are used

Institute
and Faculty
of Actuaries

14 May 2013

30

Example..

- Starting with a crude model including only the **company** variable
- Then adding up each of the *company specific factors* sequentially (stepwise selection)

Risk Factor(s) included in the model	Model	BIC Criterion	% change in BIC(vs. previous model)	% change in BIC(vs. model 0)
company	0	1339454		
company + rated	1	1339293	-0.012	
company + rated + dist.channel	2	1337869	-0.106	
company + rated + dist.channel + sum assured	3	1330932	-0.519	
company + rated + dist.channel + sum assured + underwriting dec.	4	1329426	-0.113	
company + rated + dist.channel + sum assured + underwriting dec.+ region	5	1158281	-12.874	-13.526

- Select model with the smallest BIC value
- In our case this is model 5!
- The following additional factors are included: rated, dist. channel, sum assured and region
- **So how does the new model affect the individual company effect on lapse rates?**

Institute and Faculty of Actuaries

14 May 2013

31

Example cont'd…

Risk Factor(s)	Level	RR	95% LCI	95% UCI	% change in RR
Crude Analysis (Model 0)					
company	A	0.969	0.961	0.976	
company	B	1.203	1.192	1.214	
company	C	1.084	1.074	1.095	
company	D	0.835	0.822	0.848	
company	F	0.965	0.959	0.970	
Adjusting for rated, dist.channel and sum assured (Model 5)					
company + rated + dist.channel + sum assured + underwriting dec.+ region	A	0.781	0.752	0.810	-19.401
company + rated + dist.channel + sum assured + underwriting dec.+ region	B	1.049	1.010	1.088	-12.801
company + rated + dist.channel + sum assured + underwriting dec.+ region	C	0.904	0.871	0.939	-16.605
company + rated + dist.channel + sum assured + underwriting dec.+ region	D	na	na	na	na
company + rated + dist.channel + sum assured + underwriting dec.+ region	E	1.350	1.298	1.404	39.896

- **Company E estimate dramatically changed by about 40%** after adjusting for the additional risk factors
- Other companies showed non-negligible negative changes in their estimates as well
- **Region** seems to be a really important predictor for a company's lapse behaviour
- This is followed by the impact of **distribution channel** which also seems to play a key role
- Similar selection approaches could be used for other factor combinations such as personal characteristics and others

Institute and Faculty of Actuaries

14 May 2013

32

Therefore…

- Crude analysis is a good starting point in order to investigate single factor effects

- Should always be followed by multiple factor analysis

- Different combination of predictors can be included in models

- Best model selection is the next step based on various model selection criteria

- **Factors included in the final model should not only be supported statistically but theoretically as well**

Institute and Faculty of Actuaries

14 May 2013

33

Institute and Faculty of Actuaries

What else can be done?

14 May 2013

34

What's possible?

- Looking at a **wider range of factors**, combining them and constructing models based on statistical evidence

- Looking at **interactions** between different factors e.g.
  - smoking x BMI x alcohol
  - gender x occupation
  - alcohol x gender x age
  - region x occupation x sum assured

- Repeat analyses using more advanced statistical techniques

- Compare results with the ones obtained from the traditional regression techniques described here

- An example of such is method is the **Multilevel Modelling** technique

Institute
and Faculty
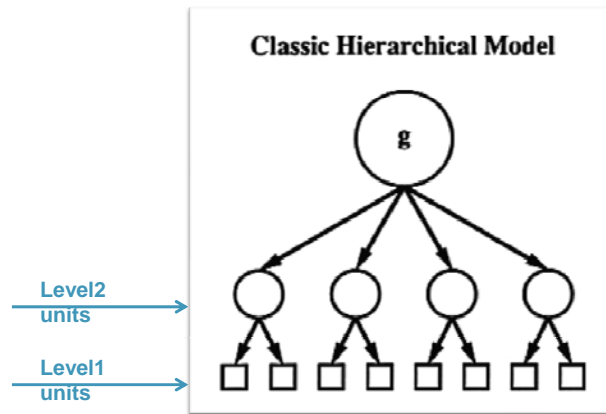of Actuaries

14 May 2013

What is Multilevel Modelling

- Multilevel Modelling is a statistical approach of parameters that vary at more than one level. In other words this is a useful tool when we have *nested* data

- Multilevel models are a generalisation of the linear models (in particular linear regression) but they can also extend to non-linear models as well

- They are also known as:
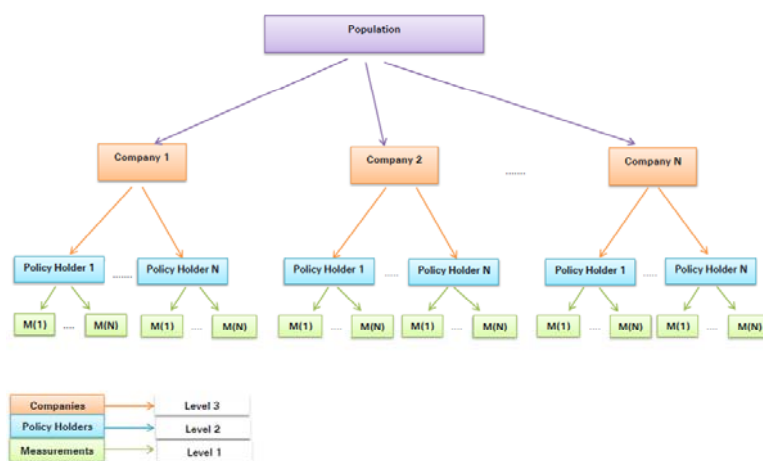 - hierarchical linear models
  - nested models
  - mixed models

Institute
and Faculty
of Actuaries

14 May 2013   36

What is nested data?

Classic Hierarchical Model

Level2 units

Level1 units

Institute and Faculty of Actuaries

14 May 2013

37

Example of *longitudinal* 3-level structure

Companies	Level 3
Policy Holders	Level 2
Measurements	Level 1

Institute and Faculty of Actuaries

14 May 2013

38

Multilevel Modelling in practice

- Complex statistical technique due to the dimensions of the data

- Can be powerful as it is expected to significantly improve precision of estimates obtained using the traditional statistical techniques

- Widely used in education and medicine at present

- Not well known in insurance world

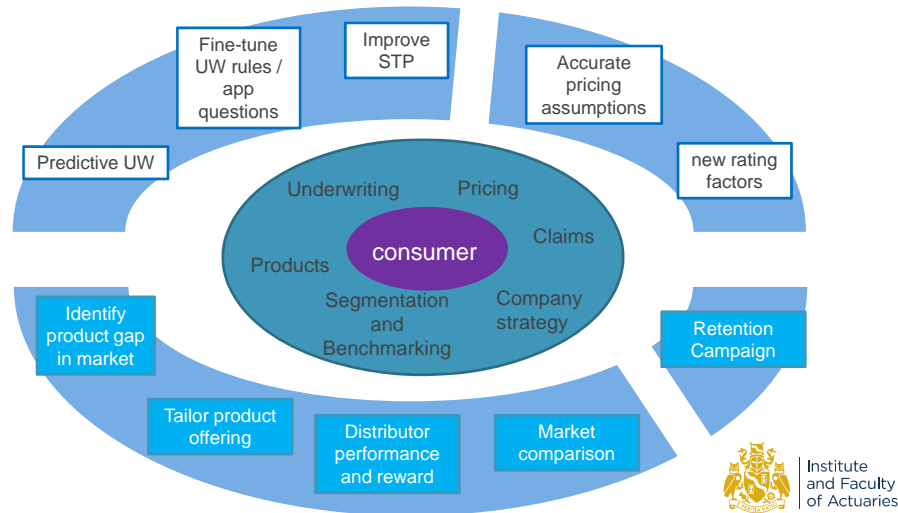- We are exploring its potential at the moment

Institute
and Faculty
of Actuaries

14 May 2013   39

Institute
and Faculty
of Actuaries

How statistics can help insurance companies?

14 May 2013   40

A lot to gain!

Fine-tune UW rules / app questions

Improve STP

Accurate pricing assumptions

Predictive UW

new rating factors

Underwriting

Pricing

consumer

Claims

Products

Company strategy

Segmentation and Benchmarking

Identify product gap in market

Retention Campaign

Tailor product offering

Distributor performance and reward

Market comparison

Institute and Faculty of Actuaries

14 May 2013

41

Questions

Comments

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.

Institute and Faculty of Actuaries

14 May 2013

42

Contact Details and Acknowledgements

- **Aspasia Angelakopoulou**
  Aspasia_Angelakopoulou@swissre.com

- **Vivian Chiu**
  Vivian_Chiu@swissre.com

- **Anita Munford,** Head of Insights Delivery**,** Life & Health, Swiss Re
- **Stephen Baldry,** Senior Underwriter, Life & Health Underwriting, Swiss Re
- **Rupert Holden,** IT consultant**,** Life & Health IT, Swiss Re

Institute
and Faculty
of Actuaries

14 May 2013

43

Institute
and Faculty
of Actuaries

Appendix

14 May 2013

44

Exogenous factors

- **Situational**
  - Related to the moment before the sale of the insurance policy, the moment of purchase of the insurance policy and the moment of use of insurance services
- **Marketing Stimuli**
  - Includes factors related to the marketing mix of insurance companies such as their offers, insurance premiums, the means of distribution and the process of communication
- **External macro-medium**
  - These include demographical, economic, political, technological factors and others
- **Personal characteristics of the consumer**
  - These include age, gender, culture, subculture, social class, reference group family and others

Institute
and Faculty
of Actuaries

14 May 2013

45

Endogenous factors

- **Learning**
  - Refers to a relatively permanent change in behaviour which comes with experience and the experience does not have to affect the learner directly: we can learn by observing events that affect others
- **Perception**
  - Includes the mental activity of observing, understanding and discerning between stimuli, achieved through the system of sensory receptors
- **Attitudes**
  - There are learned predispositions which are relatively consistent with the behaviour they reflect but they are not permanent they can change
- **Motivation**
  - An inner state which drives an organism towards a goal
- **Personality**
  - The distinctive and enduring patterns of thoughts, emotions and behaviours that characterize individual's adaptation to the situations of his or her life

Institute
and Faculty
of Actuaries

14 May 2013

46