Apteco

Whitepaper

Predictive weight of evidence (PWE) modelling for more relevant and personalised campaigns

How to create target models with speed, accuracy, and minimal user input.

Contents

Introduction 3

Why model and score? 4

How do you create and use a model? 5

What is a PWE model? 6

Technical details of PWE measure 7

Interpreting PWE measure from a profile report 8

Scoring and segmenting the prospect database 9

Summary of PWE Measure calculation 10

Interpretation of ZdExp values 11

Benefits of FastStats PWE models 12

Practical benefits 13

Integration of marketing processes 13

Fast analysis of large databases 13

PWE models can be easily interpreted 14

PWE models are easy to create and apply 14

Transactional data can be directly included in a model 14

Identifying hotspots within data 15

PWE models are well suited to marketing data 17

PWE models can directly use categorical and multi-response data 17

PWE models can cope with missing values 18

PWE models can cope with outliers 20

Records can still be scored even if only limited data is available 20

Disadvantages of PWE models 22

Is the univariate approach a problem? 22

Banding 23

Section one

Introduction

Why model and score? 4

How do you create and use a model? 5

What is a PWE model? 6

Why model and score?

Many organisations use direct marketing (direct mail, telemarketing, email marketing) to inform prospects about their products or services. Often a very large number of prospects are available in the prospect database, but marketing to all of them is expensive and even counterproductive when prospects realise they are subjected to irrelevant junk mail.

Modelling can be used to improve the targeting in a number of different situations:

- A common objective is to target direct marketing at those most likely to take up our offer and buy our product or service ("responders"). This could be to identify new prospects (who are not yet customers) or to generate additional revenue from existing customers.

- In other situations we may want to predict how likely a new customer is to have a high lifetime value. If we can predict these people then we can ensure that they obtain good customer service and perhaps we will send them special marketing communications.

- Alternatively, we may want to predict customers who are likely to lapse, so that we can contact them before this happens with an incentive to remain a customer.

In all these cases – and others – a model is used to score a database to identify the customers who should be included in the marketing activity (the target group).

How do you create and use a model?

In all modelling situations you work with a customer database, where a subset of customers has been identified as being the "target". In the examples above, this could be responders to a previous campaign, existing high value customers, or people who have already lapsed.

The modelling process compares the target group to the whole prospect database to identify the characteristics that differentiate the target group from other customers. This might show, for example, that the target group tends to be of a certain age or income band or are users of another product.

The end result of the modelling process is a way to generate a score for all people in the database. Usually, people are given a high score if they have similar characteristics to those identified in the modelling process as being typical of the target group.

We can test a model by applying it to data where we already know the customer status and measure how successfully the model classifies customers. A Gains Chart is typically used to show the performance of a model. Moving along the x-axis gradually includes all customers in the database, starting with those with the highest model score. A number of points are plotted for this increasing population of customers, showing for each population how many were actually in the target group.

What is a PWE model?

Apteco has developed powerful new predictive weight of evidence (PWE) modelling techniques that enable FastStats to create models at incredible speed and with a minimum of user input.

The models are used seamlessly to directly score the database. The new techniques make building a model a natural extension of profile analysis and therefore accessible to non-technical marketers. Modelling takes the information gained from the profile analysis and puts it into action.

PWE extracts predictive information from the profile report for each category of each variable. This is done by examining the proportion of people in the target group, comparing the proportion observed in each category with that observed overall.

This provides a PWE weight for each category, where a positive weight indicates that people from the category are more likely to be in the target group.

These PWE weights are used to create a score for each prospect record in the prospect database, which indicates how likely they are to be a responder. The prospects can then be ranked according to the PWE model score and grouped into segments, and those in the best scoring segments selected for the direct marketing.

Section two

Technical details of PWE measure

FastStats uses PWE to extract predictive information from the profile report for each category of each variable. This is used to create a score for each prospect record in the prospect database indicating the propensity for the record to be a responder.

Interpreting PWE measure from a profile report 8

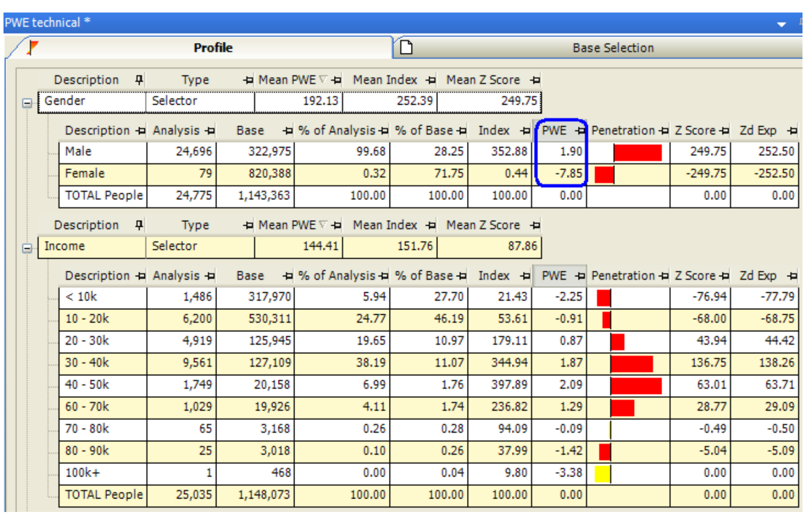
Summary of PWE Measure calculation 10

Interpreting PWE measure from a profile report

A PWE score is calculated for each category of each predictor variable. For example:

- If a prospect record is male, the PWE is 1.90. This is interpreted as evidence of 1.90 for the record being a responder.
- If they are female the PWE is -7.85. This is interpreted as evidence of 7.85 against the record being a responder.

This suggests that responders are more likely to be male. In addition, the PWE scores for the income variable suggest that responders are more likely to have incomes between £20k and £70k.


PWE technical *

Profile | Base Selection

Description	Type	Mean PWE	Mean Index	Mean Z Score
Gender	Selector	192.13	252.39	249.75

Description	Analysis	Base	% of Analysis	% of Base	Index	PWE	Penetration	Z Score	Zd Exp
Male	24,696	322,975	99.68	28.25	352.88	1.90		249.75	252.50
Female	79	820,388	0.32	71.75	0.44	-7.85		-249.75	-252.50
TOTAL People	24,775	1,143,363	100.00	100.00	100.00	0.00		0.00	0.00

Description	Type	Mean PWE	Mean Index	Mean Z Score
Income	Selector	144.41	151.76	87.86

Description	Analysis	Base	% of Analysis	% of Base	Index	PWE	Penetration	Z Score	Zd Exp
< 10k	1,486	317,970	5.94	27.70	21.43	-2.25		-76.94	-77.79
10 - 20k	6,200	530,311	24.77	46.19	53.61	-0.91		-68.00	-68.75
20 - 30k	4,919	125,945	19.65	10.97	179.11	0.87		43.94	44.42
30 - 40k	9,561	127,109	38.19	11.07	344.94	1.87		136.75	138.26
40 - 50k	1,749	20,158	6.99	1.76	397.89	2.09		63.01	63.71
60 - 70k	1,029	19,926	4.11	1.74	236.82	1.29		28.77	29.09
70 - 80k	65	3,168	0.26	0.28	94.09	-0.09		-0.49	-0.50
80 - 90k	25	3,018	0.10	0.26	37.99	-1.42		-5.04	-5.09
100k+	1	468	0.00	0.04	9.80	-3.38		0.00	0.00
TOTAL People	25,035	1,148,073	100.00	100.00	100.00	0.00		0.00	0.00

Figure: The PWE score is based on information in a profile report

The PWE scores are shown in the profile report. A positive score for a category is interpreted as evidence for prospects in that category being a responder, whereas a negative score is evidence against them being a responder. In this example, responders are more likely to be male with incomes between £20k and £70k.

The PWE method explores the statistical significance of each category of each variable. The PWE measures are calculated to indicate the strength of the relationship relative to the other predictor variables. The PWE score combines the measures calculated for the statistically significant categories.

Scoring and segmenting the prospect database

The result of running the PWE model is a score for every prospect record in the prospect database calculated according to each individual's characteristics.

The score is the sum of the PWE measures for each individual characteristic. For example, a male with an income of £40-£50k would have a score of 1.90 + 2.09 = 3.99.

The PWE score is initially calculated as a real number (e.g. 3.99) but is then often automatically banded (e.g. 3.0>4.0) to provide a segmentation of the dataset by propensity to be classified as a responder. The PWE score band is then appended to the prospect database as a new data field.

The whole process of scoring, banding, appending occurs automatically and rapidly in FastStats without user intervention.

Summary of PWE measure calculation

A PWE measure is calculated for each category of each predictor variable.

We are interested in the predictive weight of evidence PWE that membership of category Ai, of a variable A, gives that a record should be classified as a responder. In probabilistic terms we are interested in the probability of the "classification to Y (responder)" or "classification to N (not responder)" event occurring.

The PWE category score is developed by considering the difference in information gain when a record characterised by category Ai is assigned to Y ("responder") as opposed to the alternative class N ("not responder").

Using Shannon's Information Theory, this information gain can be expressed in terms of the conditional probability of a record being a responder given the prior knowledge that it is in category Ai. Bayesian probability estimates can be made using the observed frequencies for category Ai and the marginal totals. Thus the PWE measure to be estimated directly from the counts in the profile report.

The "adjusted residual" ZdExp can be used to test the significance of the deviation of the observed frequency from the expected frequency and provides a confidence measure for the PWE scores.

Interpretation of ZdExp values

Large positive values of [ZdExp]i indicate that the observed frequency is significantly greater than the expected value. Large negative values of [ZdExp]i indicate that the observed frequency is significantly less than the expected value.

The values of [ZdExp]i indicate which PWE scores can be used confidently in a PWE model. Any category where the absolute [ZdExp]I is greater than a threshold value (typically 3.0) are included in the model.

Destination | Drop your variable here

	Statistics	United States	Kuwait	Mongolia	Australia	New Zealand
United States	People	-	17,737	126	220,620	630
	Zd Exp of Pe	-	-9.19	1.15	81.42	-3.90
Kuwait	People	17,737	-	61	7,679	291
	Zd Exp of Pe	-9.19	-	15.86	-203.74	30.67
Mongolia	People	126	61	-	64	13
	Zd Exp of Pe	1.15	15.86	-	-13.06	19.98
Australia	People	220,620	7,679	64	-	453
	Zd Exp of Pe	81.42	-203.74	-13.06	-	-22.78
New Zealand	People	630	291	13	453	-
	Zd Exp of Pe	-3.90	30.67	19.98	-22.78	-
Denmark	People	299	64	1	208	7
	Zd Exp of Pe	-49.40	-3.11	0.50	-63.09	1.93
France	People	17	2	0	2	1
	Zd Exp of Pe	-123.36	-17.74	-1.40	-139.26	-3.01
Germany	People	13	1	0	4	1
	Zd Exp of Pe	-117.70	-16.96	-1.33	-132.67	-2.84
Greece	People	56,242	14,646	61	91,455	408
	Zd Exp of Pe	-261.26	101.25	1.85	21.97	9.07

Figure: Example ZdExp values

Section three

Benefits of FastStats PWE models

Practical benefits	13
Identifying hotspots within data	15
PWE models are well suited to marketing data	17
Disadvantages of PWE models	22

Practical benefits

There are a number of both statistical and practical benefits of using FastStats PWE models.

Integration of marketing processes

FastStats provides a single environment to perform all of the key modelling steps in a marketing process. The same FastStats system that holds the customer data provides the functionality to create and apply scoring models and to select customers for marketing campaigns. FastStats streamlines the process avoiding the need to extract data from one system in order to build a model in another system, only to later apply the model back on to the original database. This greatly simplifies the process of creating and applying a model.

Fast analysis of large databases

The proprietary data storage method and optimised analysis used by FastStats means millions of records can be processed in seconds.

The ability to build models using large volumes of data, instead of having to work with just a sample, means that small hotspots within the data can be identified and complex interactions can be discovered. At the same time, there is sufficient data for models to be tested against a holdout sample to ensure that they have not been over-fitted.

Not only is the underlying data processing fast, but the process of calculating PWE measures is also very efficient. This means it is possible to experiment with a number of models, resulting in a better final choice of model.

PWE models can be easily interpreted

When a PWE model is used to score prospects it is easy to understand why some customers have been given a high score and others a low score. This is important in a business environment where it is often necessary to justify decisions, for example, why someone has been refused credit.

PWE models are easy to create and apply

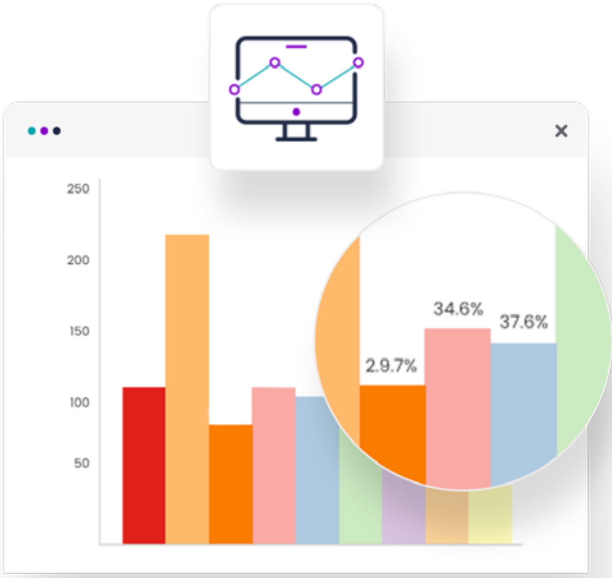
PWE models are presented as a natural progression from profiling and so are accessible to non-technical marketers. Variables may be studied in the profile report and those to be included in the model can be chosen.

The model is all generated automatically once the user has chosen the variables. The patented PWE approach takes care of the statistics behind the scenes; the user just has to click the button.

Transactional data can be directly included in a model

PWE models can be built using a mixture of customer level and transactional level data. The pre-processing of transactional data needed by other techniques can be avoided, so speeding up the model building process.

Identifying hotspots within data



PWE models are particularly effective at recognising small but important "hotspots" that could be overlooked by other techniques since they consider each category (answer) of each variable (question) separately to understand which have a bearing on whether a customer is a prospect.

Figure: Hotspots within data

For example, the question "What is your occupation?" may only have a bearing on whether you are a good prospect, if your response is "I'm a student". In a PWE model, the student category of the occupation variable could receive a positive score even if all other responses had little bearing on whether you are a prospect. Even if there were very few students in the database, a PWE score would still be applied, if this is a significant factor.

Other modelling techniques might overlook these effects in their fitting of a global function. For example, logistic regression fits a global function through the predictor space, with the same function spanning all values of all variables. A "hotspot" is defined as being a group of records with a tight range of values on one or more variables.

Typically this would not be significant enough to affect the global response surface. The global function is doing its best to model all values of all variables, and so would often ignore a small "hotspot" of responders since they would only distort the response surface in one small section of one dimension.

The potential downside to the PWE approach is that by analysing each category of each variable in isolation, you are ignoring the multivariate effects between variables. On the one hand this can lead to you failing to discover interactions between variables, and on the other hand it can result in you double counting when there is multicollinearity between variables.

See the section "Is the univariate approach a problem?" for fuller details.

PWE models are well suited to marketing data

PWE models can directly use categorical and multi-response data

Categorical and multi-response data are very common in marketing, and these types of data are ideal for PWE models. Other modelling methods often require pre-processing in order to handle multiple responses or categorical data, and even then the number of categories or responses can limit the effectiveness of using this type of data.

A multi-response variable is, for example, which newspaper(s) a customer reads. This type of data is similar to transactional data in that a person could read more than one newspaper.

A single “flag array” variable can be used in FastStats to represent multi and the PWE model can process this directly to pick out the newspapers whose readers are most likely to be prospects.

Other modelling techniques, such as logistic regression or neural networks, require real valued data and so need dummy variables to be created. This effectively creates a set of new variables, one for each category, that have either a value of 0 or 1. For example, a person who read “The Times” newspaper would have a value of 1 for the Times variable.

This approach means that such techniques can in theory use categorical data, but practical and performance issues can quickly arise when variables with many categories are used.

For example, the inclusion of a SIC variable would generate hundreds or thousands of dummy variables, each representing a single SIC code (each type of company business). This will slow down processing and create practical problems for the analyst in reviewing the model. A PWE model can examine each category within a categorical variable directly, creating an individual score for the category and providing summary statistics to say how significant the variable was as a whole.

PWE models can cope with missing values

PWE models are well suited to being built with data that contain variables that are sparsely populated, which is very common in marketing data.

Since model scores are built up from information gleaned from, for example, what a customer's occupation is, with scores being higher for some occupations and lower for others, if someone's occupation is unknown then their score is not affected either way.

The approach is taken that, if for example your occupation is unknown, then no inference can be made as to whether you are a prospect or not.

Other techniques, such as logistic regression, can require a missing value to be imputed before a record can be used in creating a model.

To do this requires some expertise and certainly time and introduces potential bias to the model. It also makes it difficult to justify scores, since a component of the score will be based on, for example, what income you were assumed to have given that your income was otherwise unknown.

An alternative is to treat missing values as a valid category. This can distort how significant a variable appears to be. For example, number of children might appear to be a significant predictor of your customers in a prospect database, but this could simply be because you only have this information for your customers and the inclusion of the variable with a missing category has made it appear significant. PWE models give you the option of whether to exclude or include missing values when scores are being calculated.

Another alternative is to build the model with records that are fully populated. This will introduce its own bias, as for example you may be discarding a whole group of records, perhaps internet enquiries, simply because you don't have their town of residence.

The ability to create models without having to reassign missing values is important in marketing applications where data is often sparsely populated.

Figure:
Working with missing values

PWE models can cope with outliers

Outliers are common in marketing data, but a PWE model can handle these without the scores for the main categories being affected. With a PWE model, all categories are scored separately so outliers are given their own score depending on how likely they are to be prospects, without affecting the other scores.

For example, the number of purchases made in a year by the majority of customers may be in single figures, but a few customers may purchase daily and so have a comparatively high frequency.

This is again important in marketing and transactional data where outliers are common. These outlying customers may well receive a significant PWE score if this information does have a bearing on whether or not they are a prospect.

In any case, the presence of these few outliers will not affect the scores assigned to the bulk customers. Other techniques, which produce a global function for each variable, are susceptible to the model being distorted by outliers.

Records can still be scored even if only limited data is available

Each customer will be scored appropriately using as much or as little data is available about them. For example, a model may have identified that age, income and gender are all important factors in determining whether a person is a prospect; nevertheless customers can still be scored with a PWE model, even if all that is known is their gender.

Other techniques, for example logistic regression, would require full information on all variables in the model in order to score customers. In cases where perhaps age or income is not known, it either has to be imputed (through use of a separate model) or a score is assigned on the basis of the unknown value.

PWE models also perform well in situations where only one of a number of related variables is populated. For example, information on where someone lives could be any of the following:

A full address attained from a mail order

A telephone std code attained from a phone enquiry

An indication of region based on attendance at an event

A model could be built using three different geographic variables, perhaps only one of which would be populated. If geography was important in identifying prospects, then people would be given a positive score on the basis of whichever geographic variable you have their data for.

Disadvantages of PWE models

Is the univariate approach a problem?

PWE models calculate a score for each category of each variable and so do not use the multivariate effects between variables.

This has considerable advantages enabling PWE models to identify hotspots, to handle missing values and outliers easily, and making the scores easy to interpret. But what are the potential downsides to the PWE approach?

On the one hand this can lead to you failing to discover interactions between variables. For example, young men and older women may both be good customers, meaning that neither age nor gender can be treated in isolation when determining how likely a prospect is. These effects would be found automatically by a decision tree or a correctly configured neural network. In the same way that logistic regression requires some user intervention to model interactions, PWE models can handle this situation by the user creating a composite variable. These interactions are relatively rare and on a small scale in marketing data, and fortunately tend to be well known by the business experts who would be involved in the model building.

Another potential downside to the univariate approach is the risk of double counting.

For example, if students are good prospects then it is likely that both “Occupation = student” and “Income = low” will have positive scores. A PWE model will include both these factors, so is effectively double counting since the characteristics are related. Other techniques such as logistic regression may have only included one of the characteristics in the model, or if both factors were included the scores associated with each may not both be positive, making the model harder to interpret.

There is a trade-off here between the benefits of PWE models being easy to interpret and their risk of double counting where effects are related. However in practice, especially where there are large numbers of variables (with large numbers of categories), it produces good models that tend to generalise well.

Banding

PWE models require numeric variables to be banded before they can be processed. In marketing the numeric variables are often in the minority, and so this limitation needs to be considered alongside the considerable advantage of being able to readily analyse the majority of categorical variables. FastStats provides facilities for banding numeric variables so they can easily be used in models.

There is a potential advantage here, that the user can choose the bandings so that they either coincide with business logic (e.g. length of relationship as a customer = 0 to 52 weeks,...) rather than the modelling technique deciding the significant boundary values (e.g. creating a model where people who have been customers for more than 41.7 weeks are treated differently).

Start building PWE models with Apteco FastStats

We’ve spoken in great detail about the potential benefits of PWE models. The best way to discover their true value though is to start creating your own models to help you refine your target audience and deliver more relevant marketing offers and messages. Apteco Marketing Suite™ gives you the tools to segment your customer database and then turn those insights into action, with the power to create campaigns quickly and manage multi-channel engagement and responses.

Request a one-to-one demo to learn more about how Apteco FastStats can enable you to carefully target prospects and customers to drive improved performance.

BOOK A DEMO

Or call us on +44 (0)1926 407 565
info@apteco.com

Apteco

apteco.com