Files are separated by folders that are related to Groups in which they are separated:

Group 1 - 324 instances

Group 2 - 16 instances

Group 3 - 18 instances

In each folder you can find the results of the exact method, GATeS heuristics, HR-1 and HR-2, as well as the codes that were used.

Description of Group 1 instances:

Talla Nobibon, F., Leus, R. and Spieksma, F.C.R. (2011). Optimization models for targeted offers in direct marketing: exact and heuristic algorithms. European Journal of Operational Research, 210, 670-683.

The instances are saved as .txt files. Each instance is identified by:

- its group; S1 for 100 clients, S2 for 200 clients, S3 for 300 clients, M1 for 1000 clients, M2 for 2000 clients and L for 10000 clients.
- its hurdle rate *R*, which is either 5 for 5%, 10 for 10% or 15 for 15%.
- its number of products *n*, which is either 5, 10 or 15.
- an integer indicating whether we have a tight budget (1), an average budget (2) or a large budget (3).
- either the letter *s* indicating that each client can receive only few offers or the letter *l* if each client can receive a higher number of offers.

Therefore, an instance named S1-5-10-2-*l* refers to an instance with 100 clients, a hurdle rate of 5%, with 10 products, an average budget and where each client can receive a higher number of offers. We next present the instances per subgroup.

- S1, n = 5 (this subgroup contains 18 instances)
- S1, n = 10 (18 instances)
- S1, n = 15 (18 instances)
- S2, n = 5 (18 instances)
- S2, n = 10 (18 instances)
- S2, n = 15 (18 instances)
- S3, n = 5 (18 instances)
- S3, n = 10 (18 instances)
- S3, n = 15 (18 instances)
- M1, n = 5 (18 instances)

- M1, n = 10 (18 instances)
- M1, n = 15 (18 instances)
- M2, n = 5 (18 instances)
- M2, n = 10 (18 instances)
- M2, n = 15 (18 instances)
- L, n = 5 (18 instances)
- L, n = 10 (18 instances)
- L, n = 15 (18 instances)

Each instance has m+4 lines, where m is the number of clients. The first line with three columns contains successively the number of clients, the number of products and the hurdle rate. The next m lines each have 2n+1 columns. The i^{th} line contains the data corresponding to the i^{th} client. At that line, the first n columns represent the cost of offering each product to that client while the next n columns report expected revenue per product if the considered product is accepted by that client. Finally, the last column contains an integer representing the maximum number of offers that can be made to that client.
The last three lines contain n columns each. The last line reports the fixed cost for each product. The last but one line reports for each product the budget available while line m+2 gives for each product the minimum number of clients who must receive an offer of that product is used during the campaign. In case of a cannibalism variation of the instance, there will be an additional line at the end that will denote the cannibal pairs (each sequence of two numbers will represent a pair).

Description of Group 2 instances:

The instances of Group 2, named XL, follows the generation method of (Nobibon et al., 2011), all using the intermediate random value of B_j and the tighter M_i with: m = 15000, 20000; n = 10, 20, 30, 40, 50; H = 10% (10 instances); m = 40000; n = 5, 10, 15, 40; H = 10% (4 instances); m = 50000; n = 15; H = 15% (1 instance); and m = 100000; n = 15; H = 15% (1 instance) totaling 16 instances called Group 2. They are organized as follows:

XL-15k-10-n-s (5 instances, where m = 15000, H = 10%,n = 10, 20, 30, 40, 50, for s see Group 1)

XL-20k-10-n-s (5 instances, where m = 20000, H = 10%, n = 10, 20, 30, 40, 50, for s see Group 1)

XL-40k-10-n-s (4 instances, where m = 40000, H = 10%, n = 5, 10, 15, 40, for s see Group 1)

XL-50k-15-15-l (1 instance, where m = 50000, H = 15%, n = 15, for l see Group 1)

XL-100k-15-15-l (1 instance, where m = 100000, H = 15%, n = 15, for l see Group 1)

The format of each instance file is the same as in Group 1.

Description of Group 3 instances:

The instances of Group 3, named EST, maintaining the intermediate random value of B_j and the tighter M_i as in (Nobibon et al., 2011) with: m = 5000; n = 40; R = 15% (1 instance); and m = 15000; n = 10; R = 10% (16 instances); and m = 100000; n = 15; R = 15% (1 instance); totaling 18 as follows:

EST-5K-15-40-8-3-10-15-40-35 (1 instance, where m = 5000, H = 15%, n = 40, ts = 8, cs = 3, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-15K-10-10-*ts*-*ps*-10-15-40-35 (2 instances, where m = 15000, H = 10%, n = 10, ts = 4, cs = 3 and 4, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-15K-10-10-*ts*-*ps*-10-15-40-35 (2 instances, where m = 15000, H = 10%, n = 10, ts = 5 and 6, cs = 5, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-15K-10-10-*ts*-*ps*-10-15-40-35 (3 instances, where m = 15000, H = 10%, n = 10, ts = 7, cs = 5, 6 and 7, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-15K-10-10-*ts*-*ps*-10-15-40-35 (3 instances, where m = 15000, H = 10%, n = 10, ts = 8, cs = 3, 7 and 9, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-15K-10-10-*ts*-*ps*-10-15-40-35 (1 instance, where m = 15000, H = 10%, n = 10, ts = 9, cs = 10, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-15K-10-10-*ts*-*ps*-10-15-40-35 (5 instances, where m = 15000, H = 10%, n = 10, ts = 8, cs = 8, 9, 10, 11 and 12, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)
EST-100K-15-15-8-3-10-15-40-35 (1 instance, where m = 100000, H = 15%, n = 15, ts = 8, cs = 3, percentage of products type 1, 2, 3, and 4 are 10%, 15%, 40%, and 35% respectively.)

This set relays in the idea that inside the clients' database can exist different consumption behaviors but also different purchase power that would generate highly different consumption patterns. There are also differences among the products, e.g., similar products can have differences in quality and price reaching different market segments. The m clients are randomly split into three Stratus: Stratus 1 – high purchase power, corresponding to a uniformly distributed random number in the interval [0.06,0.15] times m; Stratus 2 – intermediate purchase power, corresponding to [0.15,0.30] times m; Stratus 3 – low purchase power, corresponding to the remaining m. The maximum number of offers M_i that a client can receive is directly proportional to its purchase power, then M_i is a random number generated in the integer intervals [4,6], [2,4], and [1,2] for Stratus 1, 2, and 3, respectively.

Products are divided into 4 types that define the range of penetration among each client stratus, the minimum turnover of positive offer responses, the maximum turnover for positive offer responses, and the cost of the product offer for each client stratus. Each product will have the following attributes defined: product penetrations for each client stratus (P_j^S); the range of turnover for positive responses to the product offer ($[T_j, T_j^M]$); the cost of a single product offer for each client stratus (c_j^S); fixed cost for including the product on the set of offered products; product budget for offers; minimum offer quantity requirement (MC_j). These attributes can be summarized in Table 1.

Table 1: Parameters for the instance generator.

	Product – Type 1			**Product – Type 2**			**Product – Type 3**			**Product – Type 4**		
Client Stratus	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
P_j^S	[80,100]%	[15,25]%	[0]%	[60,95]%	[65,95]%	[5,10]%	[30,40]%	[75,95]%	[40,60]%	[0]%	[75,90]%	[85,100]%
c_j^S	$6cs$	$[2cs , 4cs]$	$(c_j^1+c_j^2)/2$	$4cs$	$[cs , 2cs]$	$(c_j^1+c_j^2)/2$	$3cs$	$[1.5cs,2cs]$	$(c_j^1+c_j^2)/2$	$3cs$	$[0.5cs , cs]$	$(c_j^1+c_j^2)/2$
T_j	$P_j^S.(6ts + 10ts.H)$			$P_j^S(3ts + 7ts.H)$			$P_j^S.(2ts + 3ts.H)$			$P_j^S(ts + ts.H)$		
T_j^M	$[16ts , 24ts]$			$[8ts , 12ts]$			$[3ts , 6ts]$			$[2ts , 3ts]$		
MC_j	$[0.01m , 0.04m]$			$[0.05m , 0.15m]$			$[0.15m , 0.35m]$			$[0.15m , 0.55m]$		

where:

$[a , b]$ - uniformly distributed random number generated between a and b

P_j^S - penetration value of product j for client stratus s

T_j - minimum turnover of the positive response to offers for the product j

T_j^M - maximum turnover value

c_j^S - the cost of offer product j to stratus s

MC_j - minimum number of clients at product j

ts - turnover seed for the products (defined by the user to generate diversity between instances)
cs - cost seed for the products (defined by the user to generate diversity between instances)

To define which clients will be selected as positive responses to the product j we first determine the number of clients, on given client stratus, that will respond positively to the product offer given by the formula: $C_j^S = P_j^S.C^S$, where C^S is the client number in stratus s. For each client selected and include in the set PC_j^S, an individual offer will be randomly chosen in the interval $[T_j, T_j^M]$.

After defining the number of offers to the clients in PC_j^S we can set the budget and the fixed cost. The budget for each product must attend at least 40% of the clients that would respond positively to the offer, and is computed as: $B_j = \sum_{i \in PC_j^s} \frac{T_{ij}}{c_j^{s_i}}$, where: T_{ij} is the turnover of offering product j to the client i and $c_j^{s_i}$ represents the cost of offering the product j to the client i belonging to the stratus s_i. The fixed cost, F_j, is randomly generated in the interval $[1.6B_j, 3.0B_j]$.

We test the feasibility of each instance and if there are some products which the set of offers can not reach the hurdle rate, they will pass through a repair step where each offer will have a value correction applied to its turnover, respecting the upper bounds. The turnover value of the most lucrative offers is adjusted to make the client set corresponding to the minimum offer requirement to achieve the hurdle rate.

Cannibalism:

For each one of these instances, we generated a set with (n/5) cannibal pair of products through a statistical analysis identifying the most similar one by Euclidean Distance (sufix -CAN after instance name) and also by a method called Similarity (sufix -SIMILARITY after instance name) and Dissimilarity (sufix -DISSIMILARITY after instance name). Similarity and Dissimilarity make use of the best solution found for each instance, then the cannibal pairs are defined as the pair of products offered for the biggest number of identical clients (Similarity) or different clients (Dissimilarity). The format of the instances is the same as Group 1, but we add a last line containing the cannibal pair in sequence.

In total we have 1432 instances divided as following:

- 1296 instances of Group 1, 342 instances of what we call Original Problem (OP) without the cannibalism constraint, and 324 for each type of cannibalism presented, i.e., Euclidean Distances (ED), Similarity (Sim) and Dissimilarity (Diss);
- 64 instances of Group 2, following the same logic of Group 1, namely, 16 OP, 16 ED, 16 Sim, and 16 Diss;
- 72 instances of Group 3, i.e, 18 of each OP, ED, Sim, and Diss.