

## A NEW MEDIA OPTIMIZER BASED ON THE MEAN-VARIANCE MODEL

**Pedro Jesus Fernandez**

Fundação Getulio Vargas (FGV-RJ)

Ipsos Novaction Brasil

São Paulo – SP

**Marcelo de Souza Lauretto**

**Carlos Alberto de Bragança Pereira**

**Julio Michael Stern\***

Instituto de Matemática e Estatística (IME)

Universidade de São Paulo (USP)

São Paulo – SP

[jstern@ime.usp.br](mailto:jstern@ime.usp.br)

\* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

*Recebido em 05/2006; aceito em 09/2007*

*Received May 2006; accepted September 2007*

### Abstract

In the financial markets, there is a well established portfolio optimization model called generalized mean-variance model (or generalized Markowitz model). This model considers that a typical investor, while expecting returns to be high, also expects returns to be as certain as possible. In this paper we introduce a new media optimization system based on the mean-variance model, a novel approach in media planning. After presenting the model in its full generality, we discuss possible advantages of the mean-variance paradigm, such as its flexibility in modeling the optimization problem, its ability of dealing with many media performance indices – satisfying most of the media plan needs – and, most important, the property of diversifying the media portfolios in a natural way, without the need to set up *ad hoc* constraints to enforce diversification.

**Keywords:** media planning; mean-variance optimization; parametric quadratic programming.

### Resumo

No mercado financeiro, existem modelos de otimização de portfólios já bem estabelecidos, denominados modelos de média-variância generalizados, ou modelos de Markowitz generalizados. Este modelo considera que um investidor típico, enquanto espera altos retornos, espera também que estes retornos sejam tão certos quanto possível. Neste artigo introduzimos um novo sistema otimizador de mídia baseado no modelo de média-variância, uma abordagem inovadora na área de planejamento de mídia. Após apresentar o modelo em sua máxima generalidade, discutimos possíveis vantagens do paradigma de média-variância, como sua flexibilidade na modelagem do problema de otimização, sua habilidade de lidar com vários índices de performance – satisfazendo a maioria dos requisitos de planejamento – e, o mais importante, a propriedade de diversificar os portfólios de mídia de uma forma natural, sem a necessidade de estipular restrições *ad hoc* para forçar a diversificação.

**Palavras-chave:** planejamento de mídia; otimização de média-variância; programação quadrática paramétrica.

## 1. Introduction

In every media advertising campaign, after the several strategic aspects like market, target, timing, budget, expected coverage, frequency of exposition and other parameters are defined, media planners must build the media portfolio, a schedule that determines where, when and how many advertisement insertions should be part of the media plan (Surmanek, 1993). The planner's challenge is to obtain a media portfolio within budget that reaches the target audience at the right time. Usually previous information (like product, marketing objectives, target, budget and eventually special tactics) helps the media planners to decide what media types would be effective in reaching the planned target.

Currently, several models for media optimization are available, constituting valuable tools for media planners (Leckenby & Ju, 1990). Their goal is to suggest budgets allocations, based on performance indices of vehicles like rating, reach, cost per thousand, coverage, composition and others. However, most of these models are based on two major paradigms:

- a) Greedy algorithms (Nie & Kambhampati, 2001): these algorithms start from a beginning (small) media portfolio and fill it incrementally by including vehicles (i.e. programs, newspapers, magazines) insertions with greater marginal gain until the desired reach-frequency level is achieved;
- b) Linear programming (Brown & Warshaw, 1965): optimizers based on this model return media portfolios with optimum expected return, usually expressed as a linear combination of cost, exposures and other indexes. Linear constraints like maximum investment in a media vehicle, minimum TRP in a particular TV day-part, etc, may be imposed.

The major limitation of these paradigms is that they only take into account the expected returns (first statistical moment), see Leckenby & Ju (1990). The variance of the returns (second statistical moment) is not considered. Since in general the vehicles with the greatest expected returns are also those with the greatest variances, these algorithms tend to return media portfolios with high risk, i.e. with high probability of not achieving the expected reach-frequency rates at the campaign exhibition time. This forces the planners to diversify their media plans by means of *ad hoc* constraints and bounds.

In the financial markets, there is a well established portfolio optimization model called generalized mean-variance model, or generalized Markowitz model (Markowitz, 1987; Sharpe & Alexander, 1990; Stern & Silva, 1995). This model considers that a typical investor, while expecting returns to be high, also expects returns to be as certain as possible. This means that the investor, in seeking both to maximize expected return and minimize uncertainty (that is, risk), has two conflicting objectives that must be balanced against each other when making the purchase decision. One interesting consequence of having these two conflicting objectives is that the investor should diversify by purchasing not just one asset but several ones. In the Markowitz model, assets expected returns are expressed as their mean returns and assets risks are expressed as their covariance matrix. Optimizers of this family usually return a set of efficient portfolios: an efficient portfolio is one which gives a certain expected (mean) return with the lowest risk (variance).

It is possible to establish an analogy between financial and media portfolios. In a way similar to investors, media planners must purchase assets (advertisement insertions in vehicles) with high cost effectiveness and high probability of achieving the desired goals (reach-frequency results).

In this paper we introduce a new media optimization system based on Markowitz model, a novel approach in media planning. After presenting the model in its full generality, we discuss possible advantages of the mean-variance paradigm, such as its flexibility in modeling the optimization problem, its ability of dealing with many media performance indices – satisfying most of the media plan needs – and, more important, the property of diversifying the media portfolios in a natural way, without the need of set up *ad hoc* constraints to enforce diversification.

In Section 2 we present the Markowitz financial model, and in Section 3 we discuss its formulation in the media optimization problem. In Section 4, an application to media planning in Brazilian market is analyzed in details, on real data (which will be available at the corresponding author's web page). In the Appendix, we present a brief summary of the media planning terms used in this article (Surmanek, 1993).

## 2. Mean-variance optimization concepts

In the Markowitz financial model (Markowitz, 1987), the objective function describes a trade-off between expected return and the corresponding risk (or volatility) of each portfolio. One expects to maximize the expected return and simultaneously to minimize the volatility. The decision variables are interpreted as amounts to be invested in several assets from a universe of potential candidate assets (i.e. stocks, bonds, commodities, currencies, etc.).

Formally, the investor should choose fractions  $x_1, x_2, \dots, x_M$  to be invested in the assets  $1 \dots M$ , subject to the constraints  $x_j \geq 0$  ( $j = 1 \dots M$ ) and  $\sum_{j=1 \dots M} x_j = 1$ .

We consider the returns on investment of individual assets  $r_1, r_2, \dots, r_M$  as random variables with a joint probability distribution, where the following notation is adopted. The expectation of  $r_j$  is denoted by  $\mu_j \equiv E(r_j)$ . The *covariance* between  $r_i$  e  $r_j$  is defined by

$$\sigma_{ij}^2 = E((r_i - \mu_i)(r_j - \mu_j)).$$

In particular,

$$\sigma_{jj}^2 = E((r_j - \mu_j)^2) \text{ and } \sigma_{jj} = \sqrt{\sigma_{jj}^2}$$

are the *variance* and *standard deviation* of  $r_j$ .

The *correlation coefficient* between  $r_i$  e  $r_j$  is

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ii} \sigma_{jj}}.$$

Our main interest is on the *portfolio return*,

$$R(x) = \sum_{j=1}^M r_j x_j,$$

a random variable whose mean, variance and standard deviation can be computed from  $\mu_j$  and  $\sigma_{ij}^2$ . The mean *expected return* of the portfolio is

$$e(x) \equiv E(R(x)) = \sum_{j=1}^M \mu_j x_j.$$

The *variance* and *standard deviation* of the return on the investment are

$$\sigma^2(x) \equiv \text{Var}(R(x)) = \sum_{i=1}^M \sum_{j=1}^M \sigma_{ij}^2 x_i x_j \text{ and } \sigma(x) = \sqrt{\sigma^2(x)} .$$

In matrix notation we have:

$$x = [x_1, x_2, \dots, x_M]^T ,$$

$$\mu = [\mu_1, \mu_2, \dots, \mu_M]^T , \text{ Cov} = \begin{bmatrix} \sigma_{11}^2 & \dots & \sigma_{1M}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{M1}^2 & \dots & \sigma_{MM}^2 \end{bmatrix} ,$$

$$e(x) = \mu^T x ,$$

$$\sigma^2(x) = x^T \text{Cov} x ,$$

where vectors are represented as column matrices and the symbol  $^T$  indicates matrix/vector transpose.

As presented in next section, we use as estimators for  $\mu$  and  $\text{Cov}$  the sample mean and covariance. For a general convergence analysis and estimation errors, see Lauretto *et al.* (2003). As usual in this class of models, we consider a stationary stochastic process. In a more general case, stabilization techniques, like those described in Brockwell (1991), can be used.

Usually, portfolio must respect some constraints, which are imposed by existing laws and regulation or simply demanded by the investors to their agents (Alexander & Francis, 1986). Markowitz model is capable of dealing with equality and inequality linear constraints. Some examples include:

- Pension funds managers are obliged by law to respect investment upper bounds in certain classes of assets, like junk bonds and common stocks.
- Investors could require that a certain minimum percentage of their money should be allocated to more conservative assets or specific sectors of the economy, such as govern bonds.

The linear equality constraints have the form:

$$H x = h$$

where  $H$  is a  $K \times M$  matrix of constraints coefficients,  $h$  is a  $K \times 1$  vector of bounds and  $K$  is the number of equality constraints.

The linear inequality constraints are formulated as:

$$G x \leq g$$

where  $G$  is a  $L \times M$  matrix of constraints coefficients,  $g$  is a  $L \times 1$  vector of bounds and  $L$  is the number of inequality constraints.

## 2.1 The objective function and efficient portfolios

It is well known that portfolios with higher expected returns have large risk (variances), see Sharpe & Alexander (1990). So, portfolio optimization based only on the expected return brings solutions with high risks and low effectiveness. In the Markowitz model, the objective function to be maximized is a trade-off between the expected return of the portfolio and its variance.

A portfolio  $x(\alpha)$  is optimal if it maximizes the objective function

$$\begin{aligned} x(\alpha) \in \arg \max_x \varphi(x) &= \alpha e(x) - \sigma^2(x) \\ &= \alpha \mu' x - x' Cov x \end{aligned}$$

subject to the established linear constraints

$$x \geq 0 \mid \mathbf{1}' x = 1, H x = h, G x \leq g,$$

where  $\mathbf{1}$  denotes a (column) unitary vector and the symbol  $'$  denotes matrix or vector transposition.

The parameter  $\alpha$  can be interpreted as a measure of the investor's risk tolerance. In other words,  $\alpha$  is a measure of how much maximizing the expected return is preferred in contrast to minimizing the investment risk. The higher the  $\alpha$  value, the higher the importance (for the investor) of the expected return as compared with the portfolio risk.

Each value of the parameter  $\alpha$  determines a portfolio  $x(\alpha)$  that maximizes  $\varphi(x)$ . An optimal solution  $x(\alpha)$  is called an *efficient solution* or *efficient portfolio*. A portfolio  $x$  is efficient if:

- There are not portfolios with expected return equal to the one of  $x$  and with smaller variance;
- There are not portfolios with variance equal to the one of  $x$  and with a higher expected return.

For simplicity, we denote the expected return, variance and standard deviation of the efficient portfolio  $x(\alpha)$  by  $e(\alpha) \equiv e(x(\alpha))$ ,  $\sigma^2(\alpha) \equiv \sigma^2(x(\alpha))$  and  $\sigma(\alpha) \equiv \sigma(x(\alpha))$ . The curve  $e(\alpha)$  versus  $\sigma(\alpha)$  is called *efficient frontier*.

Figure 1 shows the graph of expected returns ( $e$ ) vs. risk ( $\sigma$ ) and the efficient frontier for a specific portfolio optimization problem. Points A, B, C and D on the efficient frontier represent efficient portfolios with expected return and variance ( $e(\alpha), \sigma(\alpha)$ ). Point A is determined by a high  $\alpha$  value and point D is determined by a low  $\alpha$  value.

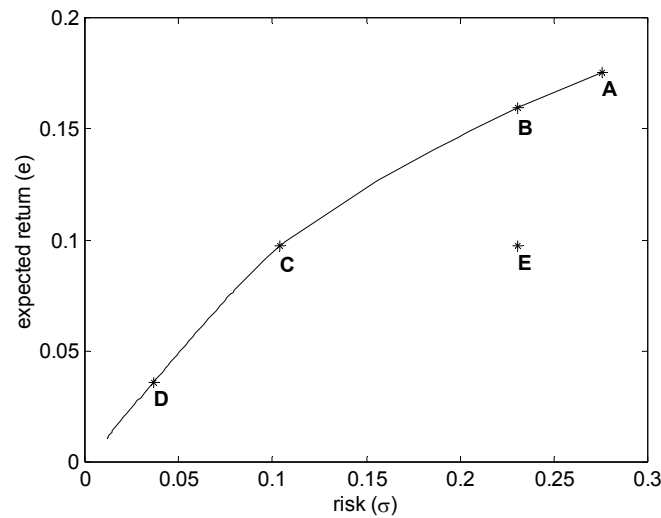


Figure 1 – Example of efficient frontier and some points of interest.

Any portfolio  $x$  whose pair  $(e(x), \sigma(x))$  is below the efficient frontier is *dominated* by portfolios lying on the frontier. For example, portfolio E is dominated by portfolio B (which has the same risk but a higher expected return) and portfolio C (which has the same expected return but a lower expected risk).

By varying the value of  $\alpha$ , an optimization solver based on Markowitz model offers to the investor a set of efficient portfolios, letting him choose the most suitable for his purposes (for example, the investor may elect the efficient portfolio which achieves the desired expected return). Furthermore, analyzing the efficient frontier, the investor may evaluate the return/risk rate of his choice and revise his demands. Let us take the chart on Figure 1 and consider, for example, an investor who wants an expected return ( $e$ ) about 0.18. Then the chosen efficient portfolio should be near the point A and achieve a standard deviation ( $\sigma$ ) about 0.28. However, if the investor accepts a return slightly lower as 0.16, the corresponding efficient portfolio will be near point C and achieve a standard deviation about 0.23. He would conclude that by reducing his expected return demand by 11% he gets a portfolio with a 18% lower standard deviation.

## 2.2 Risk protection through diversification

In the financial markets, investors look for diversification as a means of reducing risk, and devise portfolios containing assets which tend to have opposite price behaviors. As we briefly discuss below, the main differential aspect of the mean-variance model is that diversification is attained naturally, as a consequence of the trade-off between expected return and risk, without the need of setting up *ad hoc* constraints to enforce diversification (Alexander & Francis, 1986).

The objective function  $\varphi(x)$  seeks portfolios with high expected return and low variance. Using the definition of  $\sigma(x)$ , it is easy to see that the standard deviation is lower when the

correlation coefficients  $\rho_{ij}$  of the selected assets are negative. Two assets  $i$  and  $j$  have a negative correlation when  $r_i$  and  $r_j$  behave in opposite directions, i.e., an increase in  $r_i$  suggests a simultaneous decrease in  $r_j$  and conversely.

### 3. The Markowitz model and the optimization problem in media planning

The main topic of this section is the application of the general approach described in section 2 to media optimization.

We shall define a “vehicle” as the smallest unit on which one can take a media purchase decision. It could be an issue of a magazine, newspaper or a particular program in a TV channel. The set of vehicles (“decision units”) will be denoted by:

$$U = \{1, 2, \dots, M\},$$

where  $M$  is the total number of vehicles considered. Two different vehicles will be denoted by  $u$  and  $v$ .

We assume that the data comes from a sample of interviewees belonging to the *campaign target*, i.e. the subgroup of the population which constitutes the consumer market of the product to be announced. This set is denoted by  $I$ .

$$I = \{1, \dots, N\},$$

where  $N$  is the number of interviewees in the sample. We can assume also that to each interviewee we associate a “weight”  $w_i$ . This can be viewed as if each individual in the sample were a representative of a sub-population with weight  $w_i$ . We will denote by  $Pop$  the total size of the target, and then  $Pop = \sum_{i \in I} w_i$ .

#### 3.1 Some basic statistics

For each person in the sample of interviewees, several social-demographic and media consumption data are collected. Several questions are included in order to get a better understanding of the chance that one particular individual consumes a given vehicle. To know for example that a person saw/read the last edition of a magazine is not enough to ascertain the frequency with which this person actually does it. In this respect, questions related to habits are relevant. So, the media consumption data usually consist of several questions like:

- Consumption of the last issue.
- Time since last consumption.
- Number of saw/read/listened issues among the last ten, for example.

From these questions we compute a single measure of how much a person sees/listens/reads a particular vehicle, or in other words the probability that person  $i$  consumes vehicle  $v$ . Here we denote this measure by  $F(i,v) \in [0,1]$ . This index will represent a fraction of the total number of opportunities to see/listen/read the vehicle  $v$  by the individual  $i$ :  $F(i,v) = 0$  means that the individual  $i$  never sees/listens/reads the vehicle  $v$ .  $F(i,v) = 1$  means the opposite, i.e., the individual  $i$  always see/listen/reads vehicle  $v$ . In our work,  $F(i,v)$  may be interpreted as an

estimate of the probability that the individual  $i$  in the population would see/listen/read an advertisement exhibited in vehicle  $v$ . The computation of  $F(i,v)$  is an *ad-hoc* procedure, which depends on the available data and the importance assigned by the media planner for each question. Section 4 presents a particular computation of  $F(i,v)$  for the case study of this work.

The estimated return of a vehicle  $v$  is the mean of the values of  $F(i,v)$ , weighted by  $w_i$ , over all members of the sample. We denote this mean by  $\hat{\mu}_v$ .

$$\hat{\mu}_v = \frac{1}{Pop} \sum_{i=1}^N w_i F(i,v).$$

The value  $\hat{\mu}_v \in [0,1]$  is an estimate of the *coverage* (also called *rating*, see appendix) of vehicle  $v$ , or the probability that an individual from the population will be exposed to an advertisement placed in vehicle  $v$ . If  $x_v$  advertising insertions are purchased in vehicle  $v$ , the average expected number of exposures per individual in that vehicle is

$$\hat{\mu}_v x_v.$$

The estimated covariance between two vehicles  $u$  and  $v$  is computed by:

$$\hat{\sigma}_{u,v}^2 = \frac{1}{Pop} \sum_{i=1}^N w_i (F(i,u) - \hat{\mu}_u) (F(i,v) - \hat{\mu}_v).$$

In particular,

$$\hat{\sigma}_{v,v}^2 = \frac{1}{Pop} \sum_{i=1}^N w_i (F(i,v) - \hat{\mu}_v)^2 \quad \text{and} \quad \hat{\sigma}_{v,v} = \sqrt{\hat{\sigma}_{v,v}^2}$$

are the estimated variance and standard deviation.

The correlation coefficient,

$$\hat{\rho}_{u,v} = \frac{\hat{\sigma}_{u,v}}{\hat{\sigma}_{u,u} \hat{\sigma}_{v,v}},$$

varies within the interval  $[-1,+1]$ . When two vehicles  $u$  and  $v$  have a high positive correlation, that means that both reach approximately the same universe (an individual who likes vehicle  $u$  also likes vehicle  $v$ , and *vice versa*). On the other hand, when  $u$  and  $v$  have high negative correlation coefficient, that means that the vehicles have a tendency to reach different or even exclusive targets (an individual who likes vehicle  $u$  doesn't like vehicle  $v$  and *vice versa*).

Similarly to what happens in the financial markets, if we base our optimization procedure exclusively on the expected return (which, in our case is represented by  $\hat{\mu}_v$ ) we will produce schedules with lower reach if the selected vehicles have positive correlation coefficients among them. In this sense, a rational optimization procedure will naturally select vehicles with negative correlation, increasing in this way the probability of obtaining a higher reach.

The *Target Rating Point* (or simply TRP), another common measure of a schedule's efficiency, represents the cumulative rating points, on a specific target, of all vehicles



scheduled, i.e., the sum of the expected rating of each insertion. Its estimation is quite similar to the average expected number of exposures per individual, except the fact it is presented in a percentage scale:

$$TRP = 100 \hat{\mu}' x . \tag{1}$$

### 3.2 Objective functions

In our model, the decision variable is a vector  $x = [x_1, x_2, \dots, x_M]'$ , where  $x_v$  is the number of ads to be inserted in vehicle  $v$ .

We denote by  $c_v$  the cost of each advertisement insertion in vehicle  $v$  and by  $c$  the vector  $c = [c_1, c_2, \dots, c_M]'$ . The *total cost* of schedule  $x$  is:

$$\sum_{v=1}^M c_v x_v \text{ or, in matrix notation, } c' x .$$

The *expected return* of a schedule  $x$  is the average number of exposures per individual in the population, and this is obtained by:

$$\hat{e}(x) = \hat{\mu}' x . \tag{2}$$

The variance and standard deviation (in hits/individual) of the return are obtained by:

$$\hat{\sigma}^2(x) = x' Cov x \text{ and } \hat{\sigma}(x) = \sqrt{\hat{\sigma}^2(x)} ,$$

where *Cov* is the matrix of estimated covariances among vehicles,

$$Cov = \begin{bmatrix} \hat{\sigma}_{11}^2 & \dots & \hat{\sigma}_{1M}^2 \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{M1}^2 & \dots & \hat{\sigma}_{MM}^2 \end{bmatrix} .$$

The interpretation of standard deviation in the present context is the following. Suppose a schedule  $x$  with  $e(x) = m$  and  $\hat{\sigma}(x) = sd$ . The average number of impressions per individual is in the interval  $[m - sd, m + sd]$  with 67% of confidence (assuming a normal approximation). So, when two distinct schedules give the same expected return and same cost but different standard deviations, the one with smaller standard deviation (and consequently, smaller variance) is preferable, because its probability of success (achieving the desired average of impressions per individual) is larger.

It is worthwhile to discuss the importance of trade-off between cost-effectiveness and variance. Here we call cost-effectiveness the amount of exposures per monetary unit paid in a vehicle or a schedule. Most of media optimizers are oriented by cost-effectiveness, i.e. they seek schedules composed by vehicles cheaper and with more expected exposures. Since practice shows that these vehicles have in general a large variance, the built schedules variance is also large.

The Markowitz model allows several distinct optimization formulations. Currently, our media optimization tool under development has two optimization modes, described below. Both of them are based on the equilibrium between the cost-effectiveness (which has to be maximized) and the variance (to be minimized).

### 3.2.1 TRP maximization at a fixed budget:

In this optimization mode, the planner sets up the available budget and the optimizer goal is to return a set of schedules with optimal trade-off between TRP (to be maximized) and variance (to be minimized).

From equations (1) and (2), we see that maximizing expected  $TRP(x)$  is equivalent to maximizing  $\hat{e}(x)$ . So, a portfolio  $x(\alpha)$  is optimal if it maximizes the objective function

$$\begin{aligned} x(\alpha) \in \arg \max_x \varphi(x) &= \alpha \hat{e}(x) - \hat{\sigma}^2(x) \\ &= \alpha \hat{\mu}' x - x' Cov x \end{aligned}$$

subject to the linear constraints

$$x \geq 0 \mid c' x = B, H x = h, G x \leq g,$$

where  $\alpha$  is a “risk tolerance” parameter,  $B$  is the advertisement budget and  $H, h, G, g$  are matrices and vectors for equality and inequality constraints presented in section 2.

The parameter  $\alpha$ , already discussed in section 2, plays a similar role in the present objective function. The higher the  $\alpha$  value, the higher the importance of  $\hat{e}(x)$  component as compared with the  $\hat{\sigma}^2(x)$  component. The solver automatically varies  $\alpha$ , returning not just a unique schedule, but a set of schedules with different expected TRP and variances, i.e. the efficient frontier.

### 3.2.2 Cost minimization at fixed TRP:

In this optimization mode, the planner sets up the desired TRP and the optimizer goal is to return a schedule set with optimal trade-off between cost and variance (both to be minimized).

A portfolio  $x(\alpha)$  is optimal if it maximizes the objective function

$$x(\alpha) \in \arg \max_x \varphi(x) = -\alpha c' x - x' Cov x$$

subject to the linear constraints

$$x \geq 0 \mid 100 \hat{\mu}' x = R, H x = h, G x \leq g,$$

where  $\alpha$  is a “risk tolerance”,  $R$  is the desired TRP and  $H, h, G, g$  are matrices and vectors for equality and inequality constraints presented in section 2.

Analogously to the previous optimization mode, the solver automatically varies  $\alpha$ , returning a set of schedules with different costs and variances.

### 3.3 Linear constraints in media optimization

The constraints  $H x = h, G x \leq g$  are general enough for covering a large set of restrictions found in problems of media optimization. One kind of restriction has been already shown:

the reader may notice that the available budget or the desired TRP (see previous section) are particular cases of linear equality constraints. At least two other classes of constraints commonly found in media planning also can be formulated as linear constraints through the appropriated setup of  $H$ ,  $h$ ,  $G$  and  $g$ , as we show in the sequence.

### 3.3.1 Upper and lower bounds

The media planner may set up the minimum ( $x_v \geq m_v$ ), maximum ( $x_v \leq M_v$ ) or exact ( $x_v = E_v$ ) number of advertisements to be placed in some vehicles. The media planner could define a set of vehicles and specify the bound to be satisfied for each vehicle belonging to this set.

This functionality is desirable in several cases: for example, when there are pre-purchased spaces in some vehicles, or when the advertiser wants concentrate focus in some specific magazines.

### 3.3.2 Constraints over entire sets

Instead of specifying bounds for *each* vehicle in a set, the media planner frequently may specify bounds to be satisfied for the *group* of vehicles. For sake of clarity, we give two hypothetical examples:

- At least 40% of the available budget must be applied in women magazines;
- No more than 20% of the budget may be applied in tabloids.

## 3.4 Linear $\times$ Integer Programming

The Markowitz model belongs to a so-called Quadratic Programming class – since  $\sigma^2(\alpha)$  is quadratic. However, Wolfe (1959) demonstrates how to solve the quadratic programming problem using an adaptation of the Simplex algorithm for linear programming problems, namely the LC-Simplex (Linear-Complementarity Simplex). The average running time of LC-Simplex method is polynomial (Schrijver, 1986), what in practice means that LC-Simplex is fast enough for solving a large family of linear optimization problems. An available (improved) implementation of Wolfe's method is the Critical Point, see Stern (1995).

One of the main mathematical differences between portfolio and media optimization problems is that in portfolio optimization the decision variables are fractions of a total amount to be invested in the several available assets, and in practical terms these decision variables can be seen as rational numbers. On the other hand, in media optimization the decision variables are the number of advertisements in each vehicle, i.e., integer numbers. So, media optimization is an integer programming problem.

Since integer programming algorithms are much more computer intensive than LC-Simplex (Schrijver, 1986), in some cases their use in media optimization could be unfeasible. Salkin & Mathur (1989) discuss the possibility of solving first the continuous formulation of the problem via linear programming and then rounding the result. In our approach, we solve the optimization problem using a solver based on LC-Simplex and apply the following rounding method for each efficient solution  $x$ :

- a) Round all elements of  $x(\alpha)$  using the traditional nearest integer criteria, (Graham *et al.*, 1994, chap. 3); let  $y$  denote the vector of rounded elements of  $x(\alpha)$ ;
- b) Supposing the optimization problem is that described in subsection 3.2.1, i.e. optimization of trade-off between TRP and variance at a fixed budget, we verify three (mutually exclusive) conditions:
  - If the relative difference between the cost achieved by  $y$  and the budget is less than 2%, i.e. if  $\text{abs}(c'y - B)/B < 0.02$ , then we use  $y$  as the final schedule integer approximation for  $x(\alpha)$ .
  - If  $(B - c'y)/B > 0.02$ , we set a vector  $r = [r_1 \dots r_M]$ , where  $r_i = \lceil x_i(\alpha) \rceil - x_i(\alpha)$  if  $y_i = \lfloor x_i(\alpha) \rfloor$  or  $r_i = 1$  otherwise; we sort this vector in increasing order, getting the indexes  $i^{(1)} \dots i^{(M)}$  such that  $r_{i^{(1)}} \leq r_{i^{(2)}} \leq \dots \leq r_{i^{(M)}}$ ; beginning from  $j=1$ , if  $(c'y + c_{i^{(j)}} - B)/B < 0.02$ , then increment  $y_{i^{(j)}}$  by 1; if  $c'y > B$  then stop, otherwise increment  $j$  and repeat the process.
  - If  $(c'y - B)/B > 0.02$ , we set a vector  $r = [r_1 \dots r_M]$ , where  $r_i = x_i(\alpha) - \lfloor x_i(\alpha) \rfloor$  if  $y_i = \lceil x_i(\alpha) \rceil$  or  $r_i = 1$  otherwise; we sort this vector in increasing order, getting the indexes  $i^{(1)} \dots i^{(M)}$  such that  $r_{i^{(1)}} \leq r_{i^{(2)}} \leq \dots \leq r_{i^{(M)}}$ ; beginning from  $j=1$ , if  $(B - c'y - c_{i^{(j)}})/B < 0.02$ , then decrement  $y_{i^{(j)}}$  by 1; if  $c'y < B$  then stop, otherwise increment  $j$  and repeat the process.

If the optimization problem is optimization of trade-off between cost and variance at a fixed TRP (described in subsection 3.2.2), the rationale is the same as above, simply replacing  $c$  by  $\hat{\mu}$  and  $B$  by  $R$ .

There is no theoretical guarantee that after applying this rounding method the final solution is optimal (and frequently it is not), but as discussed in Section 4, in our practical experiments, this approach has given very good approximations for the exact quadratic integer programming problem.

### 3.5 Clustering vehicles with high similarities in consumption and price

In media planning, it is usual to find groups of vehicles with very similar patterns of consumption and prices. A typical case occurs in Radio and Television media: in contrast to specialized institutes of TV audience research which install “people meters” in the sample households and maintain an updated database of programs schedules, the media research institutes based on interviews do not collect data of audience by programs, but audience by broadcast station, day of week and intervals of time in the day (e.g. half hour intervals). In these cases, each vehicle is defined in terms of a channel, interval of time of the day and day of week (e.g. Channel 5/Monday/18:00-18:30).

If a program at Channel 5 is exhibited on Mondays from 18:00 to 19:00, then almost everyone who watches this channel on Monday from 18:00 to 18:30 also watches it from 18:30 to 19:00, and conversely. Therefore, these two vehicles (Channel 5/Monday/18:00-18:30 and Channel 5/Monday/18:30-19:00) tend to have the same cost per insertion and a high coefficient of correlation.

In order to improve numerical stability at optimization phase (Stern *et al.*, 2006, chap. 3-4), we adopt the following strategy:

1. cluster the vehicles with very high correlation coefficients, creating a new set of “composite vehicles”;
2. solve the optimization problem over the composite vehicles;
3. after optimization phase, spread the insertions of each composite vehicle among its original vehicles.

For the vehicles clustering, we define an undirected graph whose vertices represent the vehicles, and each edge  $(u, v)$  exists if and only if all of the conditions below hold:

- the vehicles  $u$  and  $v$  pertain to the media group (i.e. the same broadcast station, same journal, etc);
- $\frac{abs(\hat{\mu}_u - \hat{\mu}_v)}{\min(\hat{\mu}_u, \hat{\mu}_v)} < 0.05$  ,  $\frac{abs(c_u - c_v)}{\min(c_u, c_v)} < 0.05$  and  $\sigma_{u,v}^2 > 0.95$  .

On this graph we perform a breadth-first search algorithm to find connected components (Cormen *et al.*, 1999, chap. 23). From the connected components, we define new “composite vehicles”, with the following attributes. Let  $A$  be a connected component containing  $p$  vehicles. We assume that the index of individual consumption for the composite vehicle  $A$ ,  $F(i, A)$ , is the average of values for the indexes  $F(i, u)$  for the vehicles pertaining to  $A$ :

$$F(i, A) = \frac{1}{p} \sum_{v \in A} F(i, v), \forall i \in \{1 \dots I\} .$$

In this case, the estimated coverage of  $A$ ,  $\hat{\mu}_A$ , can be computed by:

$$\hat{\mu}_A = \frac{1}{p} \sum_{v \in A} \hat{\mu}_v .$$

If  $B$  is a second connected component containing  $q$  vehicles, the estimated covariance between A and B,  $\hat{\sigma}_{A,B}^2$ , can be easily estimated (DeGroot, 1986, chap. 4):

$$\hat{\sigma}_{A,B}^2 = \frac{1}{pq} \sum_{v \in A} \sum_{u \in B} \hat{\sigma}_{u,v}^2 .$$

The cost per insertion,  $c_A$ , is taken as the average between the costs of original vehicles pertaining to  $A$ :

$$c_A = \frac{1}{p} \sum_{v \in A} c_v$$

Finally, the minimum and maximum number of allowed insertions,  $m_A$  and  $M_A$ , are computed by:

$$m_A = \sum_{v \in A} m_v ; \quad M_A = \begin{cases} \sum_{v \in A} M_v , & \text{if } M_v < \infty \forall v \in A \\ \infty , & \text{if } \exists v \in A | M_v = \infty \end{cases} .$$

The next step is to solve the optimization problem defined in previous sections, on the composite vehicles. After the optimization, the next step is to spread the insertions allocated to each composite vehicle into its corresponding original vehicles. Given a composite vehicle  $A$  with  $x_A$  insertions, we compute the difference between  $x_A$  and the sum of the minimum numbers of insertions required vehicles belonging to  $A$ :

$$y_A = x_A - \sum_{v \in A} m_v$$

The quantity  $y_A$  is then spread among the original vehicles by a Multinomial distribution (DeGroot, 1986, chap. 5), where the probability of one insertion be allocated to vehicle  $v \in A$  is proportional to the size of interval between the maximum and minimum numbers of insertions allowed:

$$p_v = \frac{\min(y_A, M_v - m_v)}{\sum_{v \in A} \min(y_A, M_v - m_v)}, \quad \forall v \in A.$$

Let  $y_v$  be the quantity generated from the Multinomial process described above, for all vehicles  $v \in A$ . Then the final number of insertions to be allocated to each vehicle  $v$  is given by  $x_v = m_v + y_v$ .

### 3.6 Estimation of reach-frequency distribution

In this paper, the schedule's reach-frequency distribution is computed by estimating the expected number of times each individual will be exposed to the advertising insertions. Consider a schedule  $x = [x_1, x_2, \dots, x_M]^t$ , where  $x_v$  is the number of advertisements in vehicle  $v$ . If  $F(i, v)$  is the estimated probability of interviewee  $i$  be exposed to one advertisement in vehicle  $v$ , then the expected number of advertisements seen/listened by interviewee  $i$  in vehicle  $v$  is  $F(i, v) x_v$ . Therefore, the expected total number of advertisements seen/listened by interviewee  $i$  in the schedule is

$$f_i = \sum_{v=1}^M F(i, v) x_v.$$

The reach-frequency distribution is computed by a histogram (weighted by  $w_i$ ) of  $f_i$ ,  $i = 1, \dots, N$ .

## 4. A case study: results and discussion

To exemplify the methodology and evaluate its results, in this section we show the application of the proposed model to a set of real data. We present some cases where the objective is to find optimal schedules for advertising in four media: TV, radio, newspapers and magazines. The information source is a database containing media consumption habits and related information collected through a sample in São Paulo metropolitan area (the largest one of Brazil), between January and March of 2005. The original sample consists of 2,062 interviewees of both sexes and 10 years old or more.

Beyond media consumption habits (to be detailed below), other social-demographics characteristics like sex, age, income and others were collected for each interviewee. Brazilian

research institutes employ some default criteria – based on some of these features – for grouping population in five social class: A (the highest), B, C, D and E (the lowest). The target for our numerical experiments is composed by adults (20 or more years old) of both sexes from social classes A or B, which corresponds to 36.2% of the total population of São Paulo metropolitan area. The filtered sample contains 759 interviewees.

#### 4.1 Estimating individual consumption of vehicles

In section 3.1 we introduced the vehicle individual consumption function,  $F(i,v)$ , as an estimator for the probability that an interviewee  $i$  will see/listen an advertisement insertion on vehicle  $v$ . In this section we show the criteria adopted to define  $F(i,v)$  in our case study. We notice that the chosen criteria is just an example created by the authors, only for purpose of illustrating the practical application of the proposed optimization model. Many other criteria satisfying the conditions described in 3.1 could be used.

In our dataset, each question about a vehicle's consumption has two or more alternative answers, generally in ascending order to degree of consumption, and the interviewee can choose only one answer. In our criteria, the first step is to assign values to these possible alternatives, in order to get a quantitative measure of the vehicle consumption. For convenience, these values are set in the  $[0,1]$  interval. The second step is to define  $F(i,v)$  as a combination of the questions related to the vehicles – more specifically, a convex combination.

In the sequence we show the detailed definition of  $F(i,v)$  for each media. For sake of clarity, we take as example one specific vehicle in each media, thus avoiding unnecessary formalization.

##### a) Magazines:

Example: *Veja Magazine*

The computation of  $F(i,v)$  is:

$$F(i,v) = 0.3 \textit{LastIssueRead} + 0.5 \textit{Frequency} + 0.1 \textit{BoughtQuantity} + 0.1 \textit{SpentTimeReading}, \quad (2)$$

where each variable's value is set accordingly to the answers:

*LastIssueRead*: Have you read the last issue of *Veja*?

- a) No:  $\textit{LastIssueRead} = 0.0$
- b) Yes:  $\textit{LastIssueRead} = 1.0$

*Frequency*: Of the last 5 issues, how many issues did you read?

- a) 0:  $\textit{Frequency} = 0.0$
- b) 1:  $\textit{Frequency} = 0.2$
- c) 2:  $\textit{Frequency} = 0.4$
- d) 3:  $\textit{Frequency} = 0.6$
- e) 4:  $\textit{Frequency} = 0.8$
- f) 5:  $\textit{Frequency} = 1.0$

*BoughtQuantity*: Of the last 5 issues, how many issues did you buy?

- a) 0: *BoughtQuantity* = 0.0
- b) 1: *BoughtQuantity* = 0.2
- c) 2: *BoughtQuantity* = 0.4
- d) 3: *BoughtQuantity* = 0.6
- e) 4: *BoughtQuantity* = 0.8
- f) 5: *BoughtQuantity* = 1.0

*SpentTimeReading*: On average, how much time do you spend reading a Veja's issue?

- a) I don't read it: *SpentTimeReading* = 0.00
- b) Less than 15 min: *SpentTimeReading* = 0.15
- c) 16 to 30 min: *SpentTimeReading* = 0.30
- d) 31 to 45 min: *SpentTimeReading* = 0.45
- e) 46 min to 1 hour: *SpentTimeReading* = 0.60
- f) 1 to 2 hours: *SpentTimeReading* = 0.80
- g) More than 2 hours: *SpentTimeReading* = 1.00
- h) I don't know: *SpentTimeReading* = 0.30

**b) Newspapers:**

Example: Folha de São Paulo, Mondays.

$$F(i,v) = 0.3 \textit{LastIssueRead} + 0.5 \textit{Frequency} + 0.2 \textit{SpentTimeReading}, \quad (2)$$

where each variable's value is set accordingly to the answers:

*LastIssueRead*: In the last 7 days, have you read Folha de São Paulo on Monday?

- a) No: *LastIssueRead* = 0.0
- b) Yes: *LastIssueRead* = 1.0

*Frequency*: Of the last 5 issues, how many issues did you read?

- a) 0: *Frequency* = 0.0
- b) 1: *Frequency* = 0.2
- c) 2: *Frequency* = 0.4
- d) 3: *Frequency* = 0.6
- e) 4: *Frequency* = 0.8
- f) 5: *Frequency* = 1.0

*SpentTimeReading*: On average, how much time do you spend reading an issue of this newspaper on Mondays?

- a) I don't read it: *SpentTimeReading* = 0.00
- b) Less than 15 min: *SpentTimeReading* = 0.15
- c) 16 to 30 min: *SpentTimeReading* = 0.30
- d) 31 to 45 min: *SpentTimeReading* = 0.45
- e) 46 min to 1 hour: *SpentTimeReading* = 0.60
- f) 1 to 2 hours: *SpentTimeReading* = 0.80
- g) More than 2 hours: *SpentTimeReading* = 1.00
- h) I don't know: *SpentTimeReading* = 0.30



**c) Radio:**

Example: Station CBN FM Mondays-Fridays 6:00-6:30

$$F(i,v) = 0.4 \textit{ListenedYesterday} + 0.6 \textit{Frequency} \quad (4)$$

*ListenedYesterday*: Have you listened CBN FM yesterday from 6:00 to 6:30?

- a) No: *ListenedYesterday* = 0.0
- b) Yes: *ListenedYesterday* = 1.0

*Frequency*: On the last 30 days, how frequently have you listened CBN FM from 5:00 to 9:00 on Mondays-Fridays?

- a) Never: *Frequency* = 0.0
- b) Rarely: *Frequency* = 0.3
- c) Sometimes: *Frequency* = 0.7
- d) Very often: *Frequency* = 1.0

**d) TV:**

Example: Station Globo Mondays-Fridays 6:00-6:30

$$F(i,v) = 0.4 \textit{SawYesterday} + 0.6 \textit{Frequency} \quad (4)$$

*SawYesterday*: Have you watched Globo Station yesterday from 6:00 to 6:30?

- a) No: *SawYesterday* = 0.0
- b) Yes: *SawYesterday* = 1.0

*Frequency*: On the last 30 days, how frequently have you watched Globo Station from 6:00 to 7:00 on Mondays-Fridays?

- a) Never: *Frequency* = 0.0
- b) Rarely: *Frequency* = 0.3
- c) Sometimes: *Frequency* = 0.7
- d) Very often: *Frequency* = 1.0

**4.2 Pre-selection of vehicles**

The dataset contains 4681 vehicles. In this study, we ignore those vehicles whose coverage rates on the target are below minimum thresholds: 4% for TV and Radio, and 0.5% for newspapers and magazines. This pre-selection results in a set of 718 vehicles, distributed among the media in the following way: 58 newspapers (considering distinct days of week), 244 radio programs (distinct broadcast stations, days of week and half-hour intervals), 72 magazines and 344 TV programs.

Figure 2 shows the histograms of pre-selected vehicles according to coverage rates over the selected target, by each kind of medium. In this set of vehicles, newspapers and radio programs seem to be better distributed in terms of coverage than other media. Magazines and radio programs have the smallest coverage rates (below 8%, except by one magazine with coverage rate about 18%). Some TV programs have large coverage rates, some of them greater than 50%.

Figure 3 shows the histograms of pre-selected vehicles according to the cost per insertion. The symbol R\$ represents the Real, Brazilian monetary unit. Among the four media, radio programs are the cheaper vehicles, with costs below R\$ 1,100. TV medium has the largest range of variation, from R\$ 1,000 to R\$ 200,000. Magazines costs are more concentrated below R\$ 60,000, although one of them reaches about R\$ 120,000.

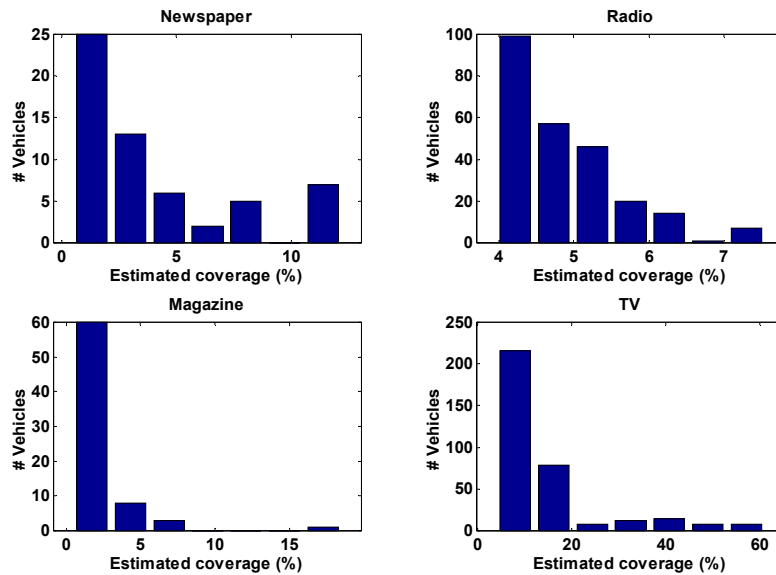


Figure 2 – Histograms of vehicles in each medium, according to coverage rates.

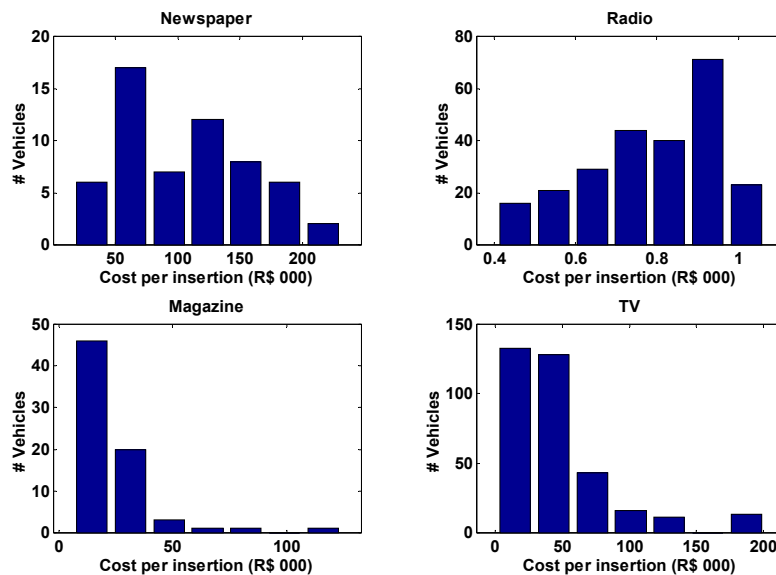


Figure 3 – Histograms of vehicles in each medium, according to cost per insertion.

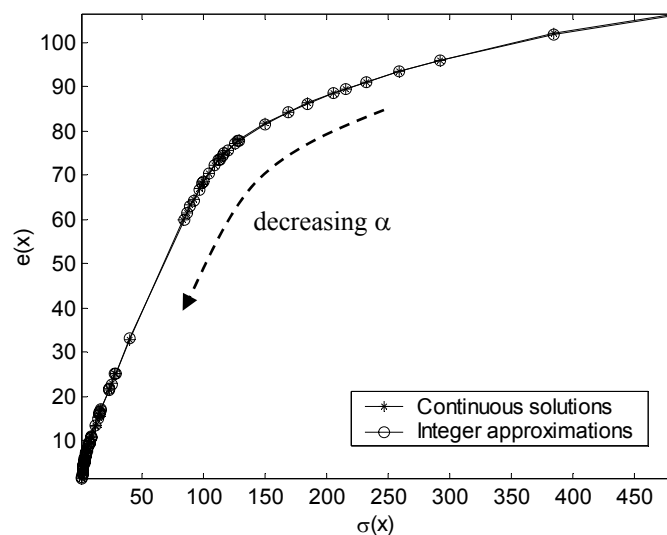
### 4.3 Numerical Results

#### 4.3.1 Case 1

In order to illustrate the features of the model and show its effectiveness as a decision-support tool in media planning, we run two hypothetical case studies. In the first case, the specified budget is R\$ 1 million and the goal is to find schedules with optimal TRP–variance balance. No additional constraints or bounds are established, in order to show the diversification provided by the mean-variance model as a natural consequence of balance between expected return and risk.

The algorithm returned a set of 107 distinct efficient schedules. As discussed in section 2, the solver automatically varies  $\alpha$  and finds the solutions that maximize the objective function  $\varphi(x) = \alpha \hat{e}(x) - \hat{\sigma}^2(x)$ . So, each efficient schedule corresponds to a distinct trade-off between the expected return  $\hat{e}(x)$  and the variance  $\hat{\sigma}^2(x)$ , controlled by the value of the parameter  $\alpha$ .

Figure 4 presents the efficient frontier for this case study, where the schedules expected returns  $\hat{e}(x)$  are plotted against the standard deviations  $\hat{\sigma}(x)$ . More specifically, the vertical axis represents the average frequency of exposures per individual of target, and the horizontal axis represents the standard deviation of number of exposures per individual. Each marker represents an efficient schedule returned by the solver. Two curves are plotted: the first representing the continuous solutions provided by the optimization solver, and the second representing the integer approximations provided by the rounding method described in section 3.4. We notice that in this case study the integer approximations are very close the continuous solutions in terms of mean-variance balance.



**Figure 4** – Efficient frontier for case 1 (continuous dotted curve), and sequence of schedules in decreasing order of  $\alpha$  (dashed arrow).

The points plotted in the top right corner of the graph represent efficient schedules where the values of the “risk tolerance” parameter  $\alpha$  in the objective function are high, and therefore these are schedules with high average number of exposures per individual and (very) high standard deviation in the number of exposures per individual. As the  $\alpha$  value becomes lower, the average number of exposures (and consequently the TRP) becomes gradually less important as compared with the standard deviation. So the efficient schedules become gradually more “conservative”, i.e., they tend to minimize the standard deviations in number of exposures, at the cost of achieve lower average numbers of exposures. The dashed arrow indicates the order of schedules, as  $\alpha$  decreases.

In the sequence, we analyse the efficient schedules by means of some of the most common indicators used in media planning, and discuss some possible criteria to select the most appropriate schedules.

Table 1 shows some indicators computed for some efficient schedules: the schedule number, its total cost (in R\$ thousands), the average number of exposures per individual (analogous to TRP, as discussed in Section 3.1), the standard deviation in number of exposures per individual, the total number of purchased insertions, the reach 1+, average frequency 1+, Reach 3-10 (percentage of target exposed between 3 and 10 times to the schedule), cost per thousand exposures and cost per thousand exposed people. As shown in Figure 4, many schedules are almost superposed, suggesting high similarities. Due to these similarities, we choose to show only a subset of the efficient schedules, taken in steps of 10, including the last one (naturally, in an optimization system, all efficient schedules must be shown so that the media planner can choose the best ones, but for our purposes in this paper, we consider this subset good enough).

**Table 1** – Some efficient schedules for fixed cost of R\$ 1 million.

Sched. number	Cost (R\$000)	Average exposures per indiv.	Std Dev exposures per indiv.	Purchased Insertions	Reach 1+	Freq. 1+	Reach 3-10	Cost per thousand exposures	Cost per thousand people
1	1,000	106.2	482.7	2,470	5%	2,129.3	0%	2,279	4,852,064
11	1,000	77.9	128.9	1,533	39%	198.2	0%	3,104	615,412
21	1,001	68.5	99.8	1,378	52%	132.5	0%	3,535	468,505
31	999	25.1	27.9	387	70%	36.0	5%	9,646	347,178
41	991	15.1	13.7	184	86%	17.4	21%	15,915	277,708
51	1,002	8.9	6.1	81	95%	9.4	46%	27,158	253,927
61	1,001	7.3	4.4	62	97%	7.5	64%	33,289	250,615
71	998	5.4	2.8	32	97%	5.6	79%	44,743	249,123
81	1,013	4.1	2.0	29	97%	4.2	76%	59,967	252,238
91	1,007	3.2	1.7	30	96%	3.4	67%	75,091	252,378
101	1,018	2.5	1.4	26	94%	2.6	44%	100,225	261,518
107	1,013	1.6	1.2	22	90%	1.7	14%	156,884	271,210

Figure 5 presents the frequency distribution for the schedules presented in Table 1. In each graph, the horizontal line represents the number of exposures, and each bar in position  $k$  represents the expected percentage of people in the target exposed  $k$  times to the advertisements of the schedule. The last bar represents the percentage of people exposed 30 or more times.

We notice that all schedules costs are close the specified budget (R\$ 1 million). The small observed differences are below the tolerance limit of 2% established by the rounding method adopted.

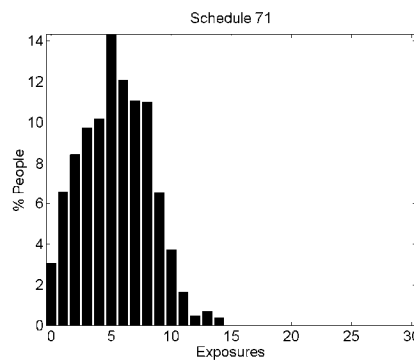
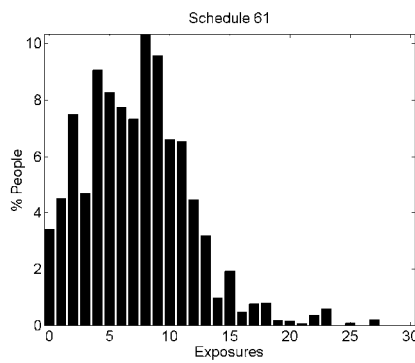
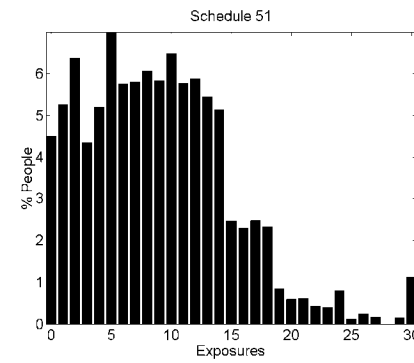
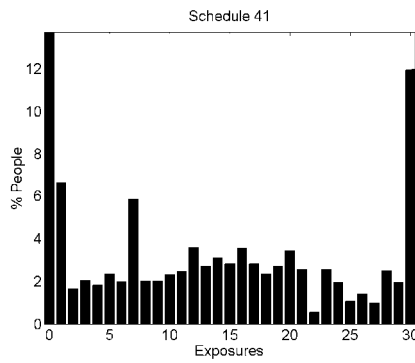
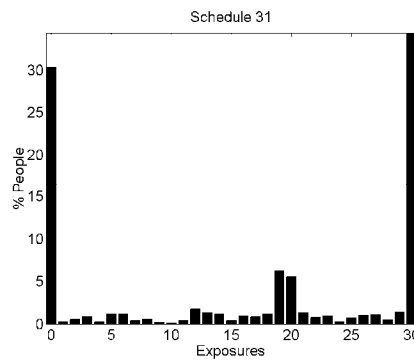
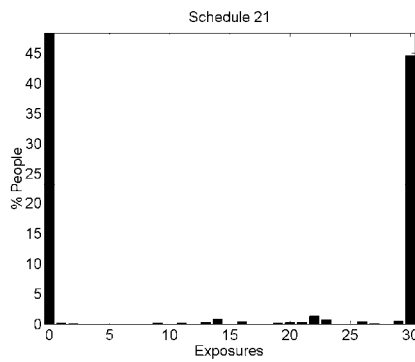
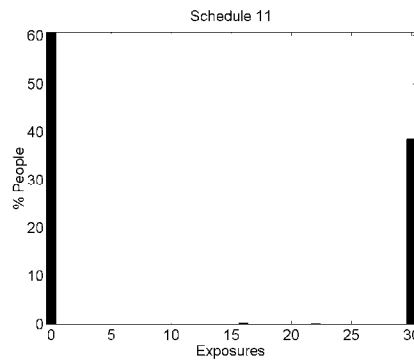
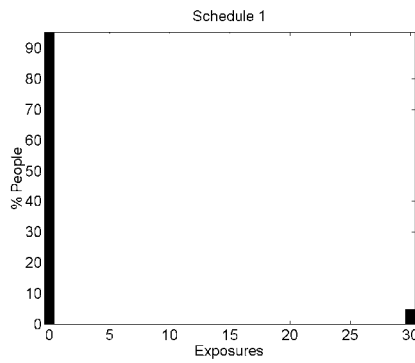
The first schedule in Table 1 corresponds to the right top point in Figure 4, discussed above. In this schedule, the component  $\hat{e}(x)$  is overweighted and therefore the efficient schedule is that with maximum TRP (despite the variance). All advertisements are concentrated only in vehicles with low cost per thousand exposures, as shown in column 9. Columns 6 and 7 (reach 1+ and frequency 1+) show that a few people (5%) are overexposed to the ads (2129 exposures on average), while the remaining (95%) are not exposed at all. This schedule exemplifies the behavior of linear optimizers based just on the cost-effectiveness. When using those optimizers, the media planner must set up several constraints and bounds in order to attain diversified schedules.

It can be observed that the expected number of exposures and the standard deviation tend to decrease as the value of  $\alpha$  decreases, as discussed. The number of insertions also decreases, since the schedules start incorporating more expensive vehicles with higher coverage rates or vehicles negatively correlated (e.g. vehicles of different media) in order to decrease the standard deviation. As a consequence, the reach increases and the frequency 1+ decreases – more people are exposed, less people are overexposed. The extreme cases are the last schedules, where  $\alpha$  is so small that the variance component  $\sigma^2(\alpha)$  dominates the objective function. In these cases, the solver return schedules with minimum variance, although with small TRP. So, the last schedules are too conservative. Therefore, the most suitable schedules are those located in the medium region of the table, as we discuss below.

From the media planner's point of view, the concept of standard deviation in exposures per individual is directly connected to the shape of frequency distribution. Schedules with high standard deviations are those where the frequency distribution is concentrated on the extremes, i.e. many people are not exposed, while others are exposed many times. Conversely, schedules with low standard deviations are those where the frequency distribution is concentrated around its average, what means to expose people more homogeneously to the campaign. Figure 5 illustrate this fact: each schedule provides a frequency distribution progressively more concentrated than its predecessor, what explains the decreasing in standard deviations observed in Table 1.

Since the model provides many efficient schedules, one natural question is how to choose the best one. Many different criteria could be adopted. For example, a possible criterion is choosing the schedule with largest percentage of people in a given interval of exposures. In our case study, if we establish the interval of 3 to 10 exposures, the schedule 71 should be chosen, since its reach in this interval is the largest (79%).

Finally, we discuss briefly the diversification provided by the mean-variance model in media planning, as a natural consequence of trade-off between expected return and variance in exposures per individual. Figure 6 shows, for each schedule of Table 1, the relative proportions of each media in the total number of TRPs provided by the schedule. Each column corresponds to a schedule and is consisted by four areas, relative to the four medium (newspapers, radio, magazines, TV). The height of each area represents the relative participation of the corresponding medium in the total TRP of schedule. We notice that the first schedule is composed only by insertions in radio, the most cost-effective medium; then, other media are being gradually incorporated, in reverse order of cost-effectiveness: TV, magazines and newspapers.



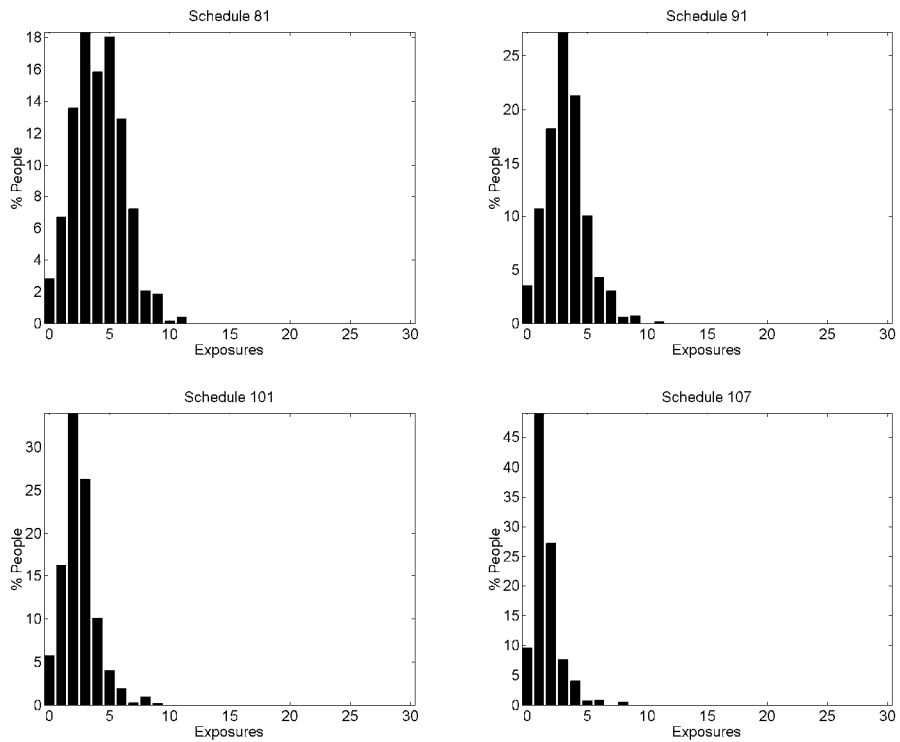


Figure 5 – Frequency distributions for efficient schedules of Table 1.

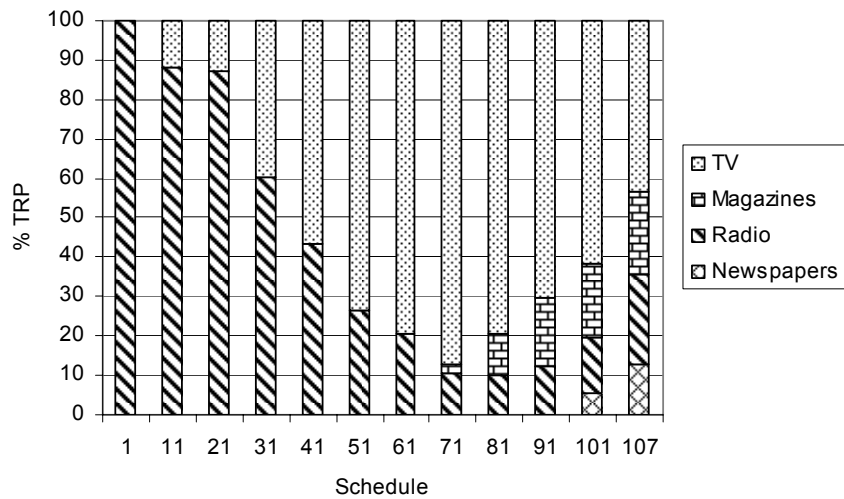


Figure 6 – Relative participation of each medium in total TRP for schedules of Table 1.

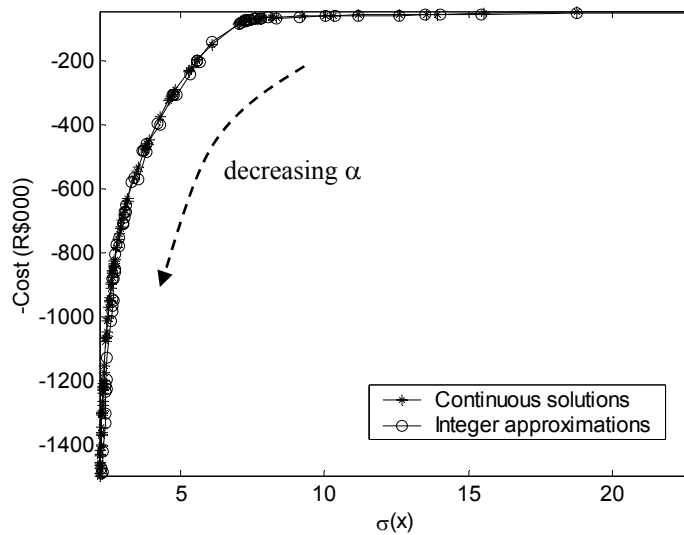
### 4.3.2 Case 2

In this case, the goal is to find schedules with optimal cost-variance balance, such that the schedules attain a fixed amount of 500 TRPs, which means an average of 5 exposures per individual in the target. The only additional constraint is that the schedules can not exceed the cost of R\$ 1.5 millions.

Figure 7 presents the efficient frontier for this case study, where the schedules costs are plotted against the standard deviations in expositions per individual,  $\hat{\sigma}(x)$ . In order to make the reading of graph more similar to that of Figure 4 and more intuitive, the sign of the costs are inverted. So, the vertical axis represents the schedules costs (with inverted sign) and the horizontal axis represents the standard deviation on number of exposures per individual. As it can be noticed, both curves representing respectively the continuous optimal solutions and the integer approximations are very close, suggesting that the rounding method adopted is a good approximation.

Analogously to Figure 4, the points plotted in the top right corner of the graph represent efficient schedules where the values of “risk tolerance” parameter  $\alpha$  in the objective function are high, and therefore these are schedules with low cost and (very) high standard deviation in the number of exposures per individual. As the  $\alpha$  value becomes lower, the cost becomes gradually less important as compared with the standard deviation. So the efficient schedules become gradually more “conservative”, i.e., they tend to minimize the standard deviations in number of exposures, by purchasing more expensive vehicles with higher coverage rates.

Table 2 shows the indicators computed for some of efficient schedules provided by the system (see Table 1 description for details on these indicators). For this case, a larger subset is shown, since after a certain point the schedules costs increase quickly (see Figure 7). Figure 8 shows the frequency distribution for some selected schedules.



**Figure 7** – Efficient frontier for case 2 (continuous dotted curve), and sequence of schedules in decreasing order of  $\alpha$  (dashed arrow).

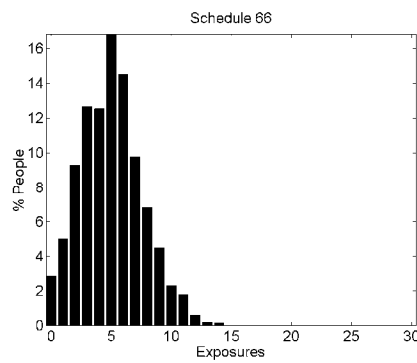
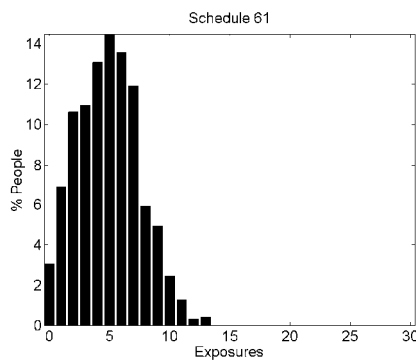
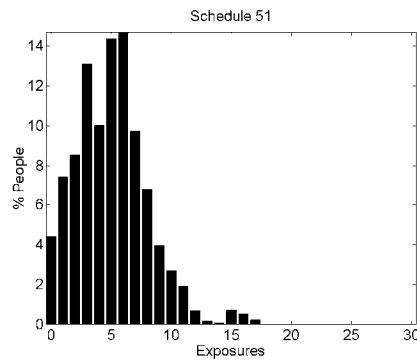
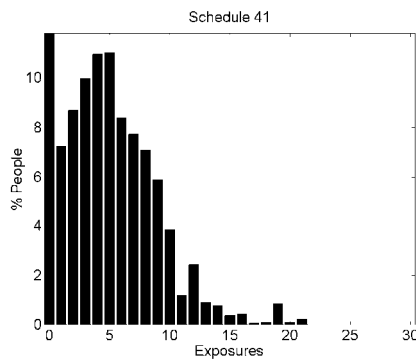
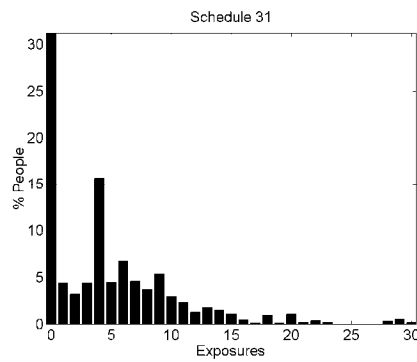
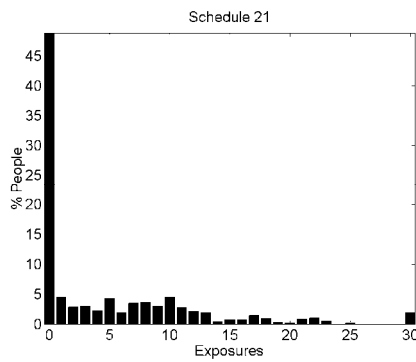
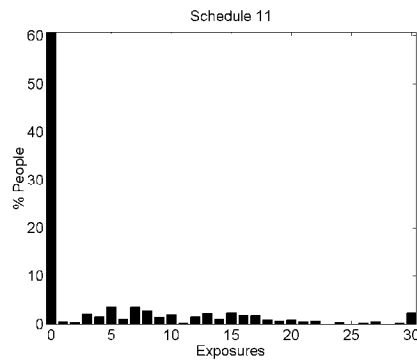
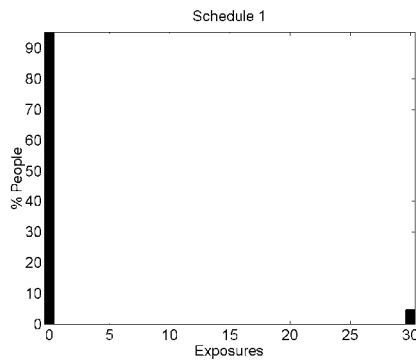


We notice that, for all schedules, the average number of exposures per individual is close to the goal (5 exposures per individual), and the small observed differences are within the tolerance limit established.

The first schedule in Table 2 is the cheapest one, and therefore that with lower cost per thousand exposures. Columns 6 and 7 (reach 1+ and frequency 1+) show that a few people (5%) are overexposed to the ads (100 exposures on average), while the remaining (95%) are not exposed at all. In the subsequent schedules, the cost is approximately increasing and the standard deviation are approximately decreasing, as consequence of decreasing the value of  $\alpha$  (small oscillations in the monotonicity of these curves are due to the rounding process). We also can notice that after schedule 61 the number of insertions, reach and frequency become stable. Incidentally, this fact may suggest a simple criterion for choosing the most appropriate schedule: choose the cheapest schedule which satisfies some indicator threshold. For example, if we are interested in schedules with a minimum 95% of reach, we can choose schedule 51, which is the cheapest one to satisfy this condition. Schedule 65 is the cheapest one to attain more than 80% of reach 3-10.

**Table 2** – Some efficient schedules for 500 TRPs and maximum cost of R\$ 1,5 million.

Sched. number	Cost (R\$000)	Average exposures per indiv.	Std Dev exposures per indiv.	Purchased Insertions	Reach 1+	Freq. 1+	Reach 3-10	Cost per thousand exposures	Cost per thousand people
1	47	5.0	22.7	116	5%	100.0	0%	2,279	227,870
6	55	5.0	12.6	105	25%	19.7	7%	2,672	52,762
11	64	5.0	8.3	99	39%	12.7	19%	3,090	39,202
16	67	5.0	7.8	100	44%	11.3	19%	3,256	36,928
21	72	4.9	7.2	100	51%	9.7	27%	3,510	33,947
26	82	5.0	7.0	100	53%	9.5	32%	3,956	37,640
31	200	5.0	5.6	77	69%	7.3	48%	9,639	70,199
36	396	5.0	4.2	54	86%	5.8	60%	19,159	111,423
41	482	5.1	3.8	50	88%	5.7	65%	23,061	132,235
46	650	5.0	3.1	38	94%	5.3	72%	31,627	166,383
51	705	5.0	3.0	42	96%	5.3	75%	33,985	178,443
56	805	4.9	2.7	33	96%	5.1	76%	39,477	203,099
61	880	4.9	2.6	33	97%	5.1	77%	43,192	219,523
62	947	5.1	2.7	36	97%	5.2	79%	45,007	235,885
63	946	5.1	2.6	35	97%	5.2	78%	45,264	235,679
64	964	5.0	2.6	38	97%	5.2	79%	46,244	240,153
65	983	5.0	2.6	38	97%	5.2	80%	47,132	244,855
66	1,013	5.0	2.6	37	97%	5.2	80%	48,855	252,271
67	1,127	5.0	2.4	32	97%	5.2	83%	54,138	280,899
68	1,213	5.0	2.4	33	97%	5.1	83%	58,690	301,246
69	1,227	5.1	2.4	35	97%	5.2	82%	58,657	304,597
70	1,196	5.0	2.4	35	97%	5.2	82%	57,315	297,013
71	1,234	5.0	2.4	36	97%	5.1	83%	59,481	306,324
72	1,305	5.0	2.4	27	97%	5.2	82%	63,446	326,882
73	1,331	5.0	2.4	35	97%	5.2	83%	64,051	331,903
74	1,422	5.0	2.3	34	98%	5.1	84%	68,581	352,505
75	1,486	5.1	2.3	35	98%	5.2	84%	70,603	368,116
76	1,473	5.0	2.3	34	98%	5.1	84%	70,929	364,964
77	1,489	5.0	2.3	35	98%	5.2	84%	71,647	369,004



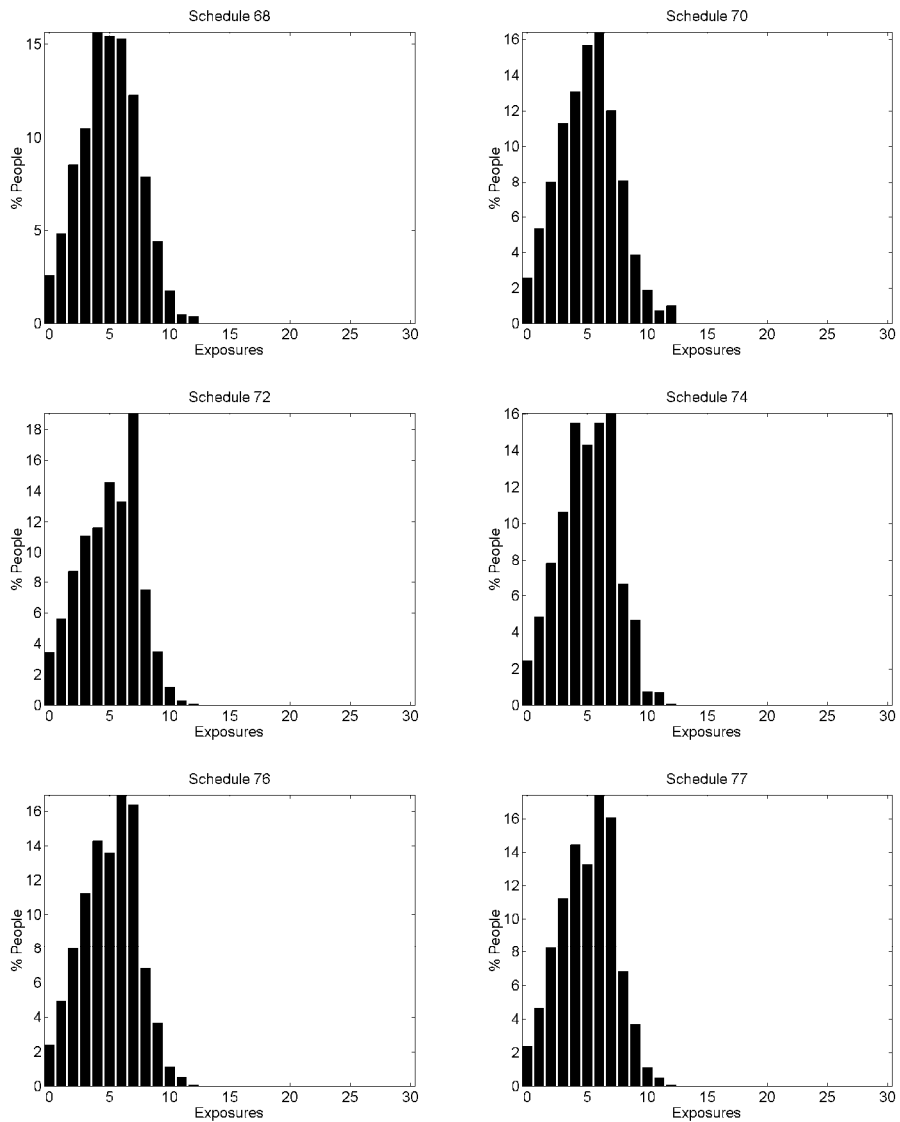


Figure 8 – Frequency distributions for some efficient schedules of Table 2.

## 5. Conclusions and future research

In this paper we establish an analogy between financial and media portfolios. In a way similar to investors, planners must purchase assets (vehicles spots) with high cost effectiveness and high probability of achieving the desired gains (reach-frequency results).

A new media optimization system based on generalized mean-variance model was introduced. We presented several possible advantages of the mean-variance paradigm over other methods like linear programming and greedy algorithms: the flexibility in modeling the

optimization problem, the ability of dealing with many media performance indices – satisfying most of the media plan needs – and, most important, the property of diversifying the media portfolios in a natural way. The case studies presented here show that the real world behavior of the model corresponds to what is expected.

An area of possible improvement lies in the definition of  $F(i,v)$  and its main statistics  $\mu$  (mean vector) and  $Cov$  (covariance matrix). The idea is to estimate  $\mu$  and  $Cov$  as mean and covariance of a multivariate normal distribution. This distribution would be the “latent” continuous distribution that determines (by truncation) the values in the observed sample of respondents.

### Acknowledgements

The authors are grateful to FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and Ipsos Brasil for the support of this optimization tool development (PIPE 02/12864-5), and to Ipsos Brasil for the database used in the examples presented in this paper.

### References

- (1) Alexander, G.J. & Francis, J.C. (1986). *Portfolio Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.
- (2) Brockwell, P.J. & Davis, R.A. (1991). *Time series: Theory and Methods*. Springer, NY.
- (3) Brown, D.B. & Warshaw, M.R. (1965). Media selection by linear programming. *Journal of Marketing Research*, **2**(1), 83-88.
- (4) Cormen, T.H.; Leiserson, C.E. & Rivest, R.L. (1999). *Introduction to Algorithms*. The MIT Press/McGrall-Hill.
- (5) DeGroot, M.H. (1986). *Probability and Statistics*. Addison-Wesley.
- (6) Graham, R.L.; Knuth, D.E. & Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science*. 2<sup>nd</sup> ed. Addison-Wesley, MA.
- (7) Lauretto, M.S.; Pereira, C.A.B.; Stern, J.M. & Zacks, S. (2003). Comparing Parameters of Two Bivariate Normal Distributions Using the Invariant FBST. *Brazilian Journal of Probability and Statistics*, **17**, 147-168.
- (8) Leckenby, J.D. & Ju, K. (1990). Advances in Media Decision Models. *Current Issues and Research in Advertising*, 311-357. Available at [http://www.ciadvertising.org/resource/rtf\\_models/drl\\_CIRA90.pdf](http://www.ciadvertising.org/resource/rtf_models/drl_CIRA90.pdf)
- (9) Markowitz, H.M. (1987). *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Basil Blackwell, Oxford.
- (10) Nie, Z. & Kambhampati, S. (2001). Joint Optimization of cost and coverage of query plans in data integration. Proceedings of the Tenth International Conference on Information and Knowledge Management, Atlanta, 223-230.

- (11) Salkin, H.M. & Mathur, K. (1989). *Foundations of Integer Programming*. North-Holland.
- (12) Schrijver, A. (1986). *Theory of Linear and Integer Programming*. Wiley – Interscience Series in Discrete Mathematics and Optimization.
- (13) Sharpe, W.F. & Alexander, G.J. (1990). *Investments*. Prentice-Hall.
- (14) Stern, J.M. (1995). Critical-Point, a Software for Portfolio Optimization. **In:** ICIAM'95 – Third International Congress on Industrial and Applied Mathematics, Hamburgo.
- (15) Stern, J.M. & Silva, M.E. (1995). Efficient Portfolios at São Paulo Stock Exchange. *Anais do XVII Encontro Brasileiro de Econometria*, **2**, 995-1013, Salvador.
- (16) Stern, J.M.; Pereira, C.A.B.; Ribeiro, C.O.; Dunder, C.; Nakano, F. & Lauretto, M.S. (2006). *Otimização e Processos Estocásticos Aplicados à Economia e Finanças*. IME-USP. Available at: <http://www.ime.usp.br/~jstern/download/publicacoes/otifin.pdf>.
- (17) Surmanek, J. (1993). *Introduction to Advertising Media: Research, Planning and Buying*. NTC Business Books.
- (18) Wolfe, P. (1959). The Simplex Method for Quadratic Programming. *Econometrica*, **27**(3), July.

#### **Appendix – Brief glossary of terms used in media planning**

*Target*: A specific universe of population to which an advertisement campaign is oriented. The target is specified according to affinities between the product and certain personal features like sex, age, social status, etc. For example, the target for a doll campaign could be the universe of girls less than 10 years old.

*Rating / Coverage*: Expected percentage of people in the target who are exposed to a vehicle in a period of time. For example, when 30% of the target is watching a TV program in a certain instant, we say that this program has 30 points of rating. The term *Rating* is commonly used for TV and radio, while *Coverage* is commonly used for magazine and newspapers. In this paper, the reach of a vehicle  $v$  is denoted by  $\mu_v$  (see section 3.1).

*Reach 1+*: Given a media schedule, its *reach* (or *reach 1+*) is the expected percentage of the target exposed one or more times to the scheduled spots. A more general term is *Reach  $s_1 - s_2$* , which means the expected percentage of the target who shall see/hear between  $s_1$  and  $s_2$  spots among the total of scheduled spots ( $s_1 \leq s_2$ ). For example, *Reach 3-6* means the percentage of target expected to see between 3 and 6 of the scheduled spots.

*Frequency 1+*: Is the average number of spots seen by each individual exposed at least once to the scheduled spots.

*TRP (Target Rating Points)*: Cumulative rating points provided by a schedule in a specific target, i.e., the sum of products of vehicle ratings (computed on the target) by their corresponding number of scheduled insertions. Section 3.1 contains the exact formulation of TRP in this work.

*Frequency distribution*: Given a schedule, its *frequency distribution* consists on estimated percentages of people exposed 0, 1, 2, 3... times to the scheduled ads.

*Cost per thousand exposures*: Ratio between one schedule's cost and the expected number of exposures, multiplied by 1000:

$$CPTI = \frac{c'x}{\hat{\mu}'x \cdot Pop/1000}$$

*Cost per thousand exposed people*: Ratio between one schedule's cost and the expected number of exposed people, multiplied by 1000.