N O R M A, S.A.R.L.

Sociedade de Estudos para o
Desenvolvimento de Empresas

V

1966

A INVESTIGAÇÃO OPERACIONAL NA EMPRESA

Documento nº. 2

Í N D I C E

Capítulo II

ESTUDO ELEMENTAR DE ALGUNS MODELOS E TÉCNICAS UTILIZADAS NA INVESTIGAÇÃO OPERACIONAL

Capítulo II

ESTUDO ELEMENTAR DE ALGUNS MODELOS E TÉCNICAS UTILIZADAS NA INVESTIGAÇÃO OPERACIONAL

1. GENERALIDADES

Como se disse no Capítulo I nº. 5.1, um trabalho de investigação operacional comporta as seguintes fases:

a) Formular o problema

b) Construir um modelo matemático que represente a operação

c) Deduzir uma solução a partir do modelo

d) Testar o modelo e a solução dele derivada

e) Controlar a solução

f) Pôr a solução em prática

Perante elas não é difícil reconhecer o método de investigação adoptado pelas ciências que tratam dos diversos domínios ou aspectos do mundo empírico. Em que se distingue pois dessas ciências a investigação operacional?

Como se viu, é o carácter interdisciplinar e o estudo dos sistemas como um todo que conferem à investigação operacional um "espírito" que está presente em todas as fases do processo de investigação, diferenciando-a assim das ciências cujo objecto de estudo é constituído por um domínio particular do mundo empírico.

Posto isto vamo-nos referir a cada uma das referidas fases.

Na formulação do problema há a distinguir dois aspectos: o problema da entidade que tem de tomar decisões e o problema de investigação pròpriamente dito. Este é uma transformação do primeiro, envolvendo inicialmente a definição de uma base científica para encontrar uma política (1) como solução.

O problema da entidade que toma as decisões raramente é apresentado com clareza à equipa de investigação operacional e, portanto, na maior parte dos casos, é esta que tem de observar o sistema considerado, seus fins e as diferentes possibilidades existentes. É necessário também identificar outros participantes susceptíveis de serem afectados pelas decisões, assim como determinar as operações em que eles intervêm e os seus objectivos.

O problema para a equipa de investigação operacional consiste depois, fundamentalmente, em determinar qual a política mais eficiente (política óptima) para a entidade que tem de tomar decisões. Consequentemente, a equipa tem de definir a medida ou quantificação da eficiência que vai utilizar e o critério de optimização que vai tomar.

O modelo que representa a operação já foi caracterizado em termos gerais no nº. 2 do Capítulo I. No entanto, dado que os modelos matemáticos ocupam uma posição chave na metodologia da investigação operacional, interessa analisar mais detalhadamente a natureza dos modelos, o seu papel na investigação científica em geral e na investigação operacional em particular.

O modelo é um esquema simplificado para a interpretação da realidade. Em consequência da complexidade do mundo real, é necessário formular hipóteses simplificadoras que levem à compreensão de um certo fenómeno. A mera acumulação de observações não pode fornecer explicação satisfatória do fenómeno e, portanto, o investigador tem necessidade de sistematizar e racionalizar os factos conhecidos, seleccionando os aspectos mais importantes e desprezando os que considera irrelevantes.

É este processo de abstracção, acompanhado de uma generalização, que conduz à construção do modelo (físico ou conceptual) que vai permitir representar, com certo grau de aderência, o fenómeno real.

(1) Entende-se por política uma sucessão de decisões.

Desde meados do século XIX, a matemática tem vindo a assumir importân cia crescente na construção dos modelos que se utilizam em diversas ciências. São os chamados modelos matemáticos.

A equipa de investigação operacional constrói e utiliza modelos matemáticos - modelos operacionais - no estudo das operações das organizações. Estes modelos são modelos de explicação-previsão e modelos de decisão. De facto, os modelos operacionais são de explicação-previsão não só porque mostram como as coisas se passam, como os diversos factores reagem entre eles, como a medida de eficiência varia em função das variáveis de acção, mas também porque não se limitam a verificar o passado, devendo aplicar-se ao futuro para permitirem a tomada de uma decisão. Por esta última razão são também modelos de decisão.

Os modelos operacionais também se classificam em exactos (ou deterministas) e estocásticos. Um modelo exacto utiliza-se quando o acaso desempenha um papel pouco importante; o modelo estocástico contém explìcitamente variáveis aleatórias.

Quanto à forma das relações que os integram os modelos podem ser lineares ou não lineares. No primeiro caso todas as relações são lineares, no segundo pelo menos uma não é linear.

A consideração da variável tempo conduz a uma nova classificação dos modelos em estáticos e dinâmicos. O modelo é estático quando todas as variáveis que nele intervêm se referem ao mesmo instante $\underline{t}$; é dinâmico quando as relações compreendem variáveis desfasadas.

A gama diversíssima de modelos matemáticos de que dispõe a investigação operacional é constituída fundamentalmente pelos modelos aplicáveis ao estudo dos problmeas básicos que se põem à entidade executiva de uma organização e que foram descritos sucintamente no nº. 5 do Capítulo I: "Stocks", repartição, filas de espera, ordenação, escolha do itinerário, substituição, concorrência e pesquisa.

Construído o modelo que representa a operação, segue-se a terceira fase do processo de investigação: derivação de soluções a partir do modelo.

Existem dois métodos principais para deduzir uma solução óptima (ou vizinha da solução óptima) de um modelo: o método analítico e o método numérico.

O método analítico consiste em empregar a dedução matemática que comporta a utilização de diversos ramos das matemáticas tais como a análise e a álgebra linear. O método numérico consiste fundamentalmente na utilização de técnicas que permitem determinar aproximadamente a solução óptima, quer ensaiando diversos valores nas variáveis de acção quer adoptando métodos iterativos. Quando certos elementos do modelo não podem ser calculados exactamente ou por razões matemáticas ou por razões práticas utilizam-se frequentemente métodos de amostragem aleatória conhecidos por métodos de Monte-Carlo.

Para obter soluções do modelo é preciso utilizar técnicas de cálculo mas existe uma grande diferença entre as técnicas e o modelo porque as primeiras não pressupõem modelo particular, são mais gerais e têm existência própria. No entanto, o desenvolvimento dos modelos e das técnicas é simultâneo, um abrindo a via ao outro; assim, por exemplo, a utilização de calculadores electrónicos é uma técnica mas o seu aperfeiçoamento permite conceber modelos novos cuja construção não seria útil com a ausência de tais meios.

O corpo deste capítulo é constituído por uma descrição elementar dos modelos e das técnicas para a sua resolução na investigação operacional.

Na quarta fase do processo de investigação põe-se o problema de <u>testar o modelo e a solução dele derivada</u>.

Dado que o modelo é uma representação simplificada da operação, ele será bom se possui suficiente aderência à realidade, isto é, se, apesar do seu carácter incompleto, serve para explicar a realidade, fazer previsões e tomar decisões com certo grau de precisão. Utilizam-se largamente em investigação operacional métodos estatísticos para testar os modelos e as suas soluções.

Na fase de <u>controle da solução de um modelo</u>, há que atender ao facto de ela só permanecer como solução quando as variáveis incontroláveis mantêm os seus valores e não há modificações das relações entre as variáveis. O controle pode, segundo os casos, operar-se por experiência directa, por analogias com situações comparáveis ou ainda fazendo uma previsão "a posteriori": neste último caso, para verificar um novo método de gestão de "stocks", por exemplo, tomar-se-á o funcionamento do sistema nos dois últimos anos e far-se-á uma gestão fictícia que a cada instante utilize o novo método em vez do antigo; podem assim comparar-se os benefícios calculados "a priori" pelo modelo com os que são dados por esta comparação.

Um dos problemas mais delicados que finalmente se põe à equipa de investigação operacional é o de <u>pôr a solução em prática</u>.

As soluções são geralmente executadas por pessoal cujos conhecimentos matemáticos são muito reduzidos, e por consequência, se a equipa de investigação operacional quer assegurar o cumprimento das suas recomendações deve apresentar as soluções em termos muito simples.

Na prática, estas diferentes fases de investigação raramente se sucedem na ordem indicada. Muitas podem ser simultâneas e em numerosos estudos, por exemplo, a fase que consiste em enunciar o problema só fica terminada quando o próprio estudo está virtualmente terminado. Normalmente as diferentes fases influenciam-se mutuamente durante o trabalho de investigação.

2. INTRODUÇÃO MATEMÁTICA

2.1 ÁLGEBRA LINEAR

Dá-se o nome de <u>vector</u> a um énuplo ordenado de números reais $x_1, x_2, \ldots x_n$ sujeito a determinadas regras de cálculo. Aos números x_i $(i=1,2,\ldots n)$ dá-se o nome de <u>componentes</u> do vector.

É necessário distinguir entre o conjunto $x_1, x_2, \ldots x_n$ escrito como uma linha - $\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}$ - e escrito como uma coluna $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

Seja X o vector escrito em coluna e X^* o mesmo vector escrito em linha (X^* é o transposto de X).

Uma matriz A(m×n) é um quadro de mn números, chamados elementos de A, que se dispõem em m linhas e em n colunas:

$$A = \begin{bmatrix} a_1^1 & a_1^2 \dots a_1^n \\ a_2^1 & a_2^2 \dots a_2^n \\ \vdots & \vdots \quad \vdots \\ a_m^1 & a_m^2 \dots a_m^n \end{bmatrix} \quad \text{ou} \quad A = \left\{ a_i^j \right\}^{(1)} \quad (i=1,2,\dots m;\ j=1,2,\dots n)$$

Notemos que i se refere à linha e j á coluna em que está o elemento a_i^j. É claro que A pode interpretar-se como um conjunto ordenado de m vectores-linha $\begin{bmatrix} a_i^1 & a_i^2 \dots a_i^n \end{bmatrix}$ (i=1,2,...m) ou um conjunto ordenado de n vectores-coluna $\begin{bmatrix} a_1^j \\ a_2^j \\ \vdots \\ a_m^j \end{bmatrix}$ (j=1,2,...n).

Quando m=n, diz-se que a matriz A é quadrada e de ordem n.

Sendo A e B matrizes m×n, a adição e subtração definem-se do modo seguinte:

$$C = \left\{ c_i^j \right\} = A \pm B \quad \text{onde} \quad c_i^j = a_i^j \pm b_i^j.$$

A multiplicação de A(m×n) por B(n×p) é a matriz C(m×p) tal que $c_i^j = \sum_{k=1}^{n} a_i^k \; b_k^j$.

Em alguns modelos matemáticos da investigação operacional intervêm equações lineares que são da forma

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = b$$

(1) Também se designam os elementos por a_{ij} e a matriz por $\left\{ a_{ij} \right\}$ (i=1,2,...m; 1,2,...n).

onde $a_1, \ldots a_n$, b são números dados e $x_1, x_2, \ldots x_n$ são as incógnitas.

O <u>sistema</u> de <u>m</u> equações lineares a <u>n</u> incógnitas é

$$a_1^1 x_1 + \ldots + a_1^n x_n = b_1$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_m^1 x_1 + \ldots + a_m^n x_n = b_m$$

que se pode escrever condensadamente na forma matricial

$$A X = B$$

onde $A = \{ a_i^j \}$, $X = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$ e $B = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$

Sejam X_1, $X_2 \ldots X_r$ <u>r</u> vectores com <u>n</u> componentes. Diz-se que estes vectores são <u>linearmente dependentes</u> se existe um conjunto de números (<u>escalares</u>), não todos nulos, tais que

$$c_1 X_1 + \ldots + c_r X_r = 0$$

em que <u>0</u> designa o <u>vector nulo</u>, isto é, o vector que tem todas as componentes nulas.

Quando aquela relação só tem lugar com $c_1 = c_2 = \ldots = c_r = 0$ os vectores dizem-se <u>linearmente independentes</u>.

Teorema I. <u>O sistema AX=B, onde m=n (isto é, A é quadrada de ordem n), tem uma solução única se e só se as colunas (ou linhas) de A são linearmente independentes.</u>

Teorema II. <u>O sistema AX=B, onde m>n, tem solução se e só se o número máximo de colunas (ou linhas) independentes das matrizes A e $[A \vdots B]$ é o mesmo, onde $[A \vdots B]$ é a matriz que se obtém de A ampliando-a com a coluna B.</u>

No caso em que $m < n$, suponhamos que as primeiras <u>m</u> colunas de <u>A</u> são linearmente independentes. Escrevendo o sistema AX=B na forma

$$x_1 P_1 + x_2 P_2 + \ldots + x_n P_n = B,$$

onde $P_1, P_2, \ldots P_n$ são as colunas de $\underline{A}$, e $P_1, \ldots P_m$ são as colunas independentes, tem-se

$$x_1 P_1 + \ldots + x_m P_m = B - x_{m+1} P_{m+1} - \ldots - x_n P_n$$

e, utilizando o teorema I, vê-se que há uma família de soluções do sistema AX=B, dependendo do $x_{m+1}, \ldots x_n$.

2.2 PROBABILIDADES E ESTATÍSTICA

2.2.1. ACONTECIMENTOS, CONCEITO DE PROBABILIDADE.

Dadas as dificuldades de ordem filosófica que surgem na definição do conceito de probabilidade, e porque elas não interessam no nosso curso, utilizaremos o cálculo de probabilidades baseado no conceito frequencista de probabilidade.

Entenderemos por acontecimento a ocorrência ou não ocorrência de certo fenómeno. Pela palavra prova designaremos alguma acção ou experiência numa dada classe de acções possíveis. Suponhamos, por exemplo, que observamos os carros que passam por um determinado ponto de uma estrada e que esses veículos podem ser de n tipos possíveis $T_1, T_2, \ldots T_n$. Considere-se a prova que consiste em observar o tipo de cada carro que passa por esse ponto da estrada. Se são observados N carros, designe N_i o número dos que pertencem ao tipo T_i, de modo que $N_1 + \ldots + N_n = N$. Então, se C é um carro, a probabilidade p_i do acontecimento "C é do tipo T_i" define-se como

$$p_i = \lim_{N = \infty} N_i/N,$$

supondo que este limite quase sempre existe. É óbvio que $0 \leqq p_i \leqq 1$ e $p_1 + p_2 + \ldots + p_n = 1$.

Designando por E_i o acontecimento "C é do tipo T_i", é costume escrever $P(E_i)$ em vez de p_i e então

$$P(E_1) + P(E_2) + \ldots + P(E_n) = 1.$$

Vamos introduzir seguidamente o conceito de <u>acontecimentos compostos</u>.

O acontecimento "A ou B" que se designa por A+B é chamado <u>reunião</u> de A e B e o acontecimento "A e B", simbolizado por AB, é denominado <u>intersecção</u> de A e B. Calculemos P(A+B) e P(AB).

Utilizando ainda o exemplo supracitado, seja <u>A</u> o acontecimento "C é do tipo T_i" e <u>B</u> o acontecimento "C é azul". Suponha-se que, observando <u>N</u> carros, N_i são do tipo T_i, N_a são azuis, N_{ai} são azuis <u>e</u> do tipo T_i, por forma que

$$P(A) = N_i/N \ , \quad P(B) = N_a/N$$

onde se admite, aqui e no que se segue, que em qualquer das razões que dá uma probabilidade já se tomou $N \to \infty$. É claro que AB é o acontecimento "C é azul e do tipo T_i" e, portanto, $P(AB) = N_{ai}/N$. O acontecimento A+B é "C é azul ou do tipo T_i" e o número de carros possuindo esta propriedade é $N_a + N_i - N_{ai}$. Então

$$1) \quad P(A+B) = (N_a + N_i - N_{ai})/N = P(A) + P(B) - P(AB).$$

Admita-se agora que não há carros do tipo T_i azuis. Então a ocorrência do acontecimento <u>A</u> (isto é, "C é do tipo T_i) exclui a possibilidade da ocorrência do acontecimento <u>B</u> (isto é, "C é azul"), e vice-versa. Dizemos então que A e B são <u>acontecimentos mutuamente exclusivos</u>. Neste caso, como $N_{ai} = 0$, tem-se $P(AB) = 0$ e

$$2) \quad P(A+B) = P(A) + P(B).$$

Acontece frequentemente que a ocorrência de um acontecimento vai depender da ocorrência de outro. Por exemplo, no caso supracitado, se o carro C é azul, então a probabilidade de que C seja do tipo T_i é N_{ai}/N_a, diferente de $P(A) = N_i/N$. Chamamos a N_{ai}/N probabilidade condicional do acontecimento A, dada a ocorrência do acontecimento B, e escrevemos $P(A/B)$. Então

$$P(A/B) = \frac{N_{ai}}{N_a} = \frac{N_{ai}}{N} \Big/ \frac{N_a}{N} = \frac{P(AB)}{P(B)} ,$$

isto é,

3) $P(AB) = P(A/B).\ P(B).$

Anàlogamente, pode ver-se que

$$P(AB) = P(B/A).\ P(A).$$

Suponhamos agora que

$$N_i/N = N_{ai}/N ,$$

isto é,

$$P(A) = P(A/B).$$

Diz-se então que o acontecimento A, "C é do tipo T_i" é independente do acontecimento B, "C é azul", e então

4) $P(AB) = P(A).\ P(B),$

isto é, no caso de acontecimentos independentes A e B, a probabilidade da ocorrência de ambos é o produto das probabilidades de A e de B.

As fórmulas 1), 2), 3) e 4) podem ser generalizadas para o caso de n acontecimentos. Designando por $A_1, A_2, \ldots A_n$ os acontecimentos, obtém-se:

$$1') \quad P(\sum A_i) = \sum P(A_i) - \sum P(A_i\, A_j) + \sum P(A_i\, A_j\, A_k) + \ldots + (-1)^{n-1} P(A_1\, A_2 \ldots A_n).$$

Se $A_1, A_2, \ldots A_n$ forem mutuamente exclusivos então

$$2') \quad P(\sum A_i) = \sum P(A_i).$$

A fórmula análoga a 3) é

$$3') \quad P(A_1\, A_2 \ldots A_n) = P(A_1).\ P(A_2/A_1).\ P(A_3/A_1\, A_2) \ldots P(A_n/A_1\, A_2 \ldots A_{n-1})$$

e, se $A_1\, A_2 \ldots A_n$ são independentes, vem

$$4') \quad P(A_1\, A_2 \ldots A_n) = P(A_1)\ P(A_2) \ldots P(A_n).$$

2.2.2. <u>VARIÁVEIS ALEATÓRIAS E FUNÇÕES DE DISTRIBUIÇÃO</u>

Se uma variável <u>X</u> pode tomar os valores $x_1\, x_2, \ldots x_n \ldots$ (em número finito ou infinidade numerável) com as probabilidades $p_1\, p_2, \ldots, p_n \ldots$ tais que $\sum_i p_i = 1$, diz-se que <u>X</u> é uma variável <u>casual</u>, <u>aleatória</u> ou <u>estocástica discreta</u>. Chamamos à sucessão $p_1, p_2, \ldots p_n \ldots$ <u>a função de frequências</u> ou <u>função densidade de probabilidade</u>. A <u>função de distribuição</u> será

$$P(X \leqq x) = \sum_{x_i \leqq x} p_i.$$

O <u>valor esperado</u> é

$$\bar{x} = E(X) = \sum_i p_i\, x_i$$

e a <u>variância</u>

$$\sigma^2 = \sum_i (x_i - \bar{x})^2 \; p_i = \sum_i x_i^2 \, p_i - (\bar{x})^2 \,.$$

O desvio-padrão é σ.

Indicam-se seguidamente alguns tipos notáveis de distribuições discretas:

a) Distribuição binomial

Domínio de x : 0, 1, 2,...n

Função de frequências : $p_i = C_i^n \, p^i (1 - p)^{n-i}$

$(0 < p < 1)$

Valor esperado : $\bar{x} = np$

Variância : $\sigma^2 = np(1 - p)$

b) Distribuição de Poisson

Domínio de x : 0, 1, 2, 3,...n...

Função de frequências: $p_i = \dfrac{\lambda^i \, e^{-\lambda}}{i!}$ $(\lambda > 0)$

Valor esperado : $\bar{x} = \lambda$

Variância : $\sigma^2 = \lambda$

c) Distribuição uniforme

Domínio de x : 0, 1, 2,...n

Função de frequências : $p_i = \dfrac{1}{n+1}$

Valor esperado : $\bar{x} = \dfrac{n}{2}$

Variância : $\sigma^2 = \dfrac{n^2}{12} + \dfrac{n}{6}$

d) Distribuição geométrica

Domínio : 0, 1, 2,...n...

Função de frequências : $p_i = p^i (1 - p)$ $(0 < p < 1)$

Valor esperado : $\bar{x} = \frac{p}{1-p}$

Variância : $\sigma^2 = \frac{p}{(1-p)^2}$

Quando o domínio de $\underline{x}$ é $]-\infty, +\infty[$, a variável aleatória $\underline{X}$ diz-se contínua quando a <u>função de distribuição</u>

$$F(x) = P(X \leqq x)$$

é contínua e tem derivada $f(x) = F'(x)$ também contínua.
É claro que

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

e a $f(x)$ dá-se o nome de <u>função densidade de probabilidade ou função de frequências</u>. Tem-se

$$P(x < X < x + dx) = f(x)\, dx$$

$$P(a \leqq X \leqq b) = \int_a^b f(x)\, dx = F(b) - F(a)$$

$$P(-\infty < x < +\infty) = \int_{-\infty}^{+\infty} f(x)\, dx = 1$$

Tal como no caso discreto,

$$\bar{x} = E(x) = \int_{-\infty}^{+\infty} xf(x)\, dx$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x)\, dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - (\bar{x})^2$$

Indicam-se seguidamente alguns tipos notáveis de distribuições contínuas:

a) <u>Distribuição rectangular</u>

Domínio de x : $[o, a]$

Função de densidade : $f(x) = \frac{1}{a}$

Valor esperado : $\bar{x} = \frac{a}{2}$

Variância : $\sigma^2 = \frac{a^2}{12}$

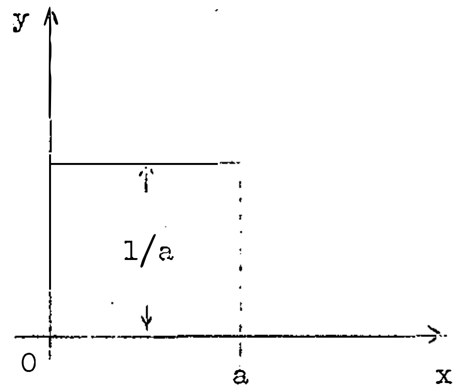

fig.3. Distribuição rectangular

b) <u>Distribuição normal</u>

Domínio de x : $]-\infty, +\infty[$

Função de densidade : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$

Valor esperado : $\bar{x} = m$

Variância : σ^2

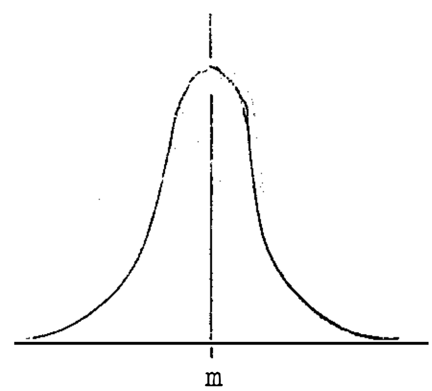

fig.4. Distribuição normal

c) <u>Distribuição exponencial</u>

Domínio de x : $]o, +\infty[$

Função de densidade : $f(x) = a\ e^{-ax}\ (a>o)$

Valor esperado : $\bar{x} = \frac{1}{a}$

Variância : $\sigma^2 = \frac{1}{a^2}$

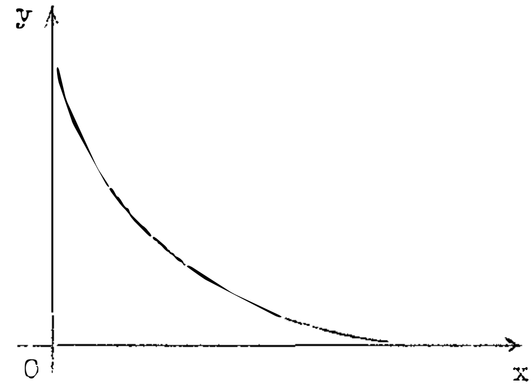

fig.5. Distribuição exponencial

d) Distribuição gama

Domínio de x : $[o, +\infty[$

Função de densidade : $f(x) = \frac{b^{a+1}}{\Gamma(a+1)} x^a \; e^{-bx} (a > -1, b > o)$

onde $\Gamma(v) = \int_o^{+\infty} e^{-x} \; x^{v-1} \; dx$

Valor esperado : $\bar{x} = \frac{a+1}{b}$

Variância : $\sigma^2 = \frac{a+1}{b^2}$

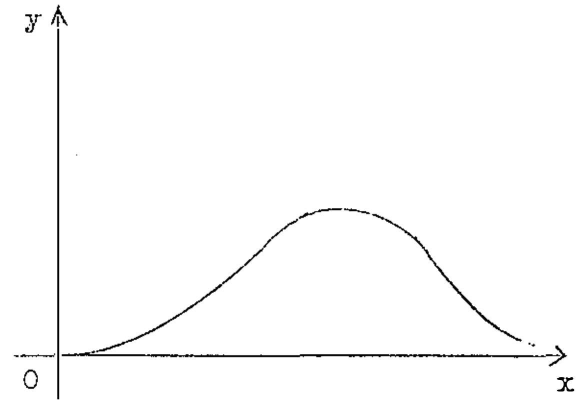

fig.6. Distribuição gama

e) Distribuição beta

Domínio de x: $[o, 1]$

Função de densidade : $f(x) = \frac{(a+b+1)!}{a! \; b!} x^a \; (1-x)^b$

Valor esperado : $\bar{x} = \frac{a+1}{a+b+2}$

Variância : $\sigma^2 = \left(\frac{a+2}{a+b+3}\right) \left(\frac{a+1}{a+b+2}\right)$

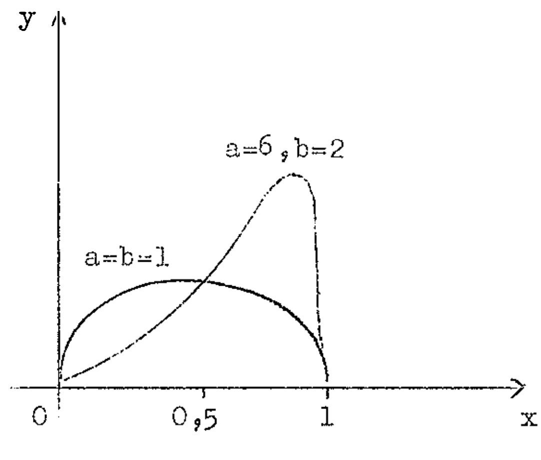

fig.7. Distribuição beta

2.2.3. <u>AMOSTRAGEM</u>

Na maior parte dos estudos de investigação operacional é necessário considerar variáveis aleatórias cujas médias e distribuições não são conhecidas. O papel da teoria da amostragem em investigação operacional é o de sistematizar, tanto quanto possível, as hipóteses que devem ser feitas relativamente às médias ou distribuições de variáveis aleatórias relevantes.

Na maior parte dos casos, há pois que fazer inferências a partir de dados fornecidos por uma <u>amostra</u> da população que se pretende estudar. É claro que essas inferências só serão válidas se a amostra é representativa, isto é, se os elementos que a compõem foram escolhidos sem enviezamentos.

No tipo mais simples de <u>amostragem casual</u> ou <u>aleatória</u>, o processo de selecção é tal que toda a combinação de $\underline{n}$ elementos dos $\underline{N}$ elementos da população tem a mesma probabilidade de ser extraída, e, além disso, a escolha de uma dessas combinações é independente da escolha de outra do mesmo número de elementos (<u>amostragem aleatória simples</u>). O número <u>n</u> de elementos seleccionados designa-se por <u>tamanho da amostra</u>.

Para escolher uma amostra casual pode utilizar-se o processo da lotaria. Este consiste, por exemplo, em numerar cada um dos indivíduos (se a população é finita), pôr o número em fichas ou esferas e tirá-las ao acaso de uma urna. Outro processo mais conveniente consiste em utilizar uma tabela de números aleatórios que contém listas de dígitos tais que cada dígito entre 0 e 9 tem a mesma probabilidade de aparecer numa dada posição numa coluna simples, e cada número de dois algarismos entre 00 e 99 tem a mesma probabilidade de aparecer numa dada posição numa coluna dupla, etc..

Outro tipo de amostragem é a <u>amostragem aleatória estratificada</u>. Considere-se uma população $\underline{C}$ e as subpopulações C_1, $C_2, \ldots C_m$ tais que C_i e C_j $(i \neq j)$ não têm elementos comuns. Se de cada um destes estratos C_i extrairmos uma amostra simples de n_i elementos, obtém-se uma amostra de $\underline{C}$ com $n = \sum_{i=1}^{m} n_i$ elementos.

Designando por N_i o número de indivíduos em C_i, se $\frac{n_1}{N_1} = \frac{n_2}{N_2} = \ldots = \frac{n_m}{N_m} = \frac{n}{N}$ tem-se uma <u>amostra aleatória estratificada proporcional</u>.

Como o objectivo da amostragem é o de fazer inferências acerca da população, tem interesse estimar os parâmetros da distribuição da população a partir da amostra. Os parâmetros mais simples e que se utilizam com maior frequência são a média e a variância.

Se o valor esperado da estimativa sobre todas as possíveis amostras é o parâmetro da população, o método de estimação diz-se <u>não enviezado</u>.

Considere-se uma população (x) de $\underline{N}$ membros com média μ e variância σ^2 e seja $(x_1, x_2, \ldots x_n)$ uma amostra aleatória simples de tal população. Pode mostrar-se que a média da amostra $\bar{x} = \sum x_i/n$ é uma estimativa não enviezada da média da população, isto é, $E(\bar{x}) = \mu$. Para a variância da amostra $s^2 = \sum (x_i - \bar{x})^2/n$, mostra-se que $E(s^2) = \frac{n-1}{n} \sigma^2$ é portanto uma estimativa não enviezada da variância da população é

$$\frac{\sum (x_i - \bar{x})^2}{n-1} .$$