COPPEAD/UFRJ

RELATÓRIO COPPEAD Nº 69

UMA NOVA COMPROVAÇÃO DE QUE O USO DE AMOSTRAGEM ALEATÓRIA SIMPLES EM SIMULAÇÃO E INCORRETO: ESTUDO DO CASO PARTICULAR DO PROBLEMA DO JORNALEIRO

Eduardo Saliby*

Outubro 1981

* Professor da COPPEAD/UFRJ. O autor agradece o apoio financeiro da FINEP.

I. INTRODUÇÃO

Em publicações recentes (Saliby, 1980 e 1981a) o autor criticou o uso tradicional de amostragem aleatória simples em simulação que, segundo argumentou, tratava-se de uma abordagem incorreta, por levar a resultados com maior imprecisão do que a necessária.

Os motivos que o levaram a formular tal crítica derivaram de um estudo em que foram identificadas duas fontes de variação dos resultados de uma simulação: uma delas associada ao conjunto de valores amostrais utilizados numa corrida e, a outra, associada à ordem ou seqüência em que tais valores são empregados.

Estas duas fontes de variação foram identificadas respectivamente como o efeito de conjunto e o efeito de seqüência, sendo a sua importância relativa função do particular problema em estudo.

Numa simulação, o erro associado ao efeito de seqüência não é controlável, no entanto o efeito de conjunto pode e deve ser controlado, o que se consegue usando-se o método de amostragem descritiva (Saliby, 1980). A principal característica da amostragem descritiva reside na seleção determinística dos valores amostrais, em total conformidade com a distribuição de probabilidade que se quer representar, a qual é sempre suposta conhecida de antemão num estudo de simulação; assim, a única fonte de aleatoridade do método de amostragem descritiva decorre da seqüência na qual tais valores são empregados, a qual é obtida por uma permutação aleatória destes valores.

Usando-se um exemplo simples e claro, o objetivo do presente trabalho foi o de comprovar, a validade da crítica feita pelo autor. Para isso, foi estudado o problema clássico do jornaleiro, que representa um tipo particular de problema onde inexiste qualquer variabilidade de resultado decorrente do efeito de seqüên cia, ou seja, o resultado da simulação é o mesmo qualquer que seja a ordem em que os valores amostrais são utilizados.

A versão estudada do problema foi a seguinte:

"Um jornaleiro compra diariamente Q exemplares de um jornal a um custo unitário C=3. Cada jornal é vendido por V=8, sendo que as sobras, ao final do dia, são perdidas, isto é, tem um valor residual R=0. A procura diária de jornais pode ser considerada como uma variável aleatória independente e identicamente distribuída de acordo com a distribuição de probabilidades abaixo:

Demanda	Probabilidade (%)
50	10
60	12
70	15
80	20
90	18
100	15
110	10

O problema do jornaleiro consiste na determinação do valor ótimo de Q, maximizando seu lucro esperado".

Embora pudesse ter sido resolvido analiticamente, este problema foi estudado através da técnica de simulação, dedicando-se particular atenção à variabilidade associada aos resultados assim obtidos.

Inicialmente, foi feita uma investigação sobre o efeito da duração de uma corrida na precisão dos resultados, confirmando um fato já conhecido: à medida em que esta duração aumenta, a precisão torna-se maior. Ilustrou-se também a existência de uma equivalência entre o método analítico de solução e a utilização de corridas de simulação com duração infinita.

Posteriormente, foi estudada a aplicação da teoria do Mo delo Linear de Resposta (Saliby, 1980) na explicação da variabilidade dos resultados obtidos, relacionando-os com os respectivos momentos amostrais das variáveis de entrada.

Segundo a crítica feita o uso de amostragem aleatória em simulação (e também em aplicações do Método de Monte Carlo) resulta na introdução de uma parcela de erro no processo de solução, associada ao efeito de conjunto, que é totalmente desnecessária. Utilizando-se amostragem descritiva, esta fonte de erro é controlada sem que se prejudique, em qualquer aspecto, a validade do estudo; neste caso, a única fonte de variação dos resultados remanescente e que não é controlável, é aquela associada ao efeito de seqüência.

Porém, no caso particular do problema do jornaleiro, o resultado é invariante com a seqüência na qual os valores são utilizados, e portanto o uso de amostragem descritiva conduz à sua solução exata, fato este que indiretamente confirma a crítica feita ao uso de amostragem aleatória simples em simulação.

II. RESULTADOS OBTIDOS NA SIMULAÇÃO DO PROBLEMA DO JORNALEIRO E SUA RELAÇÃO COM OS RESPECTIVOS VALORES TEÓRICOS

Fixada a quantidade comprada de jornais em Q=80 unidades/dia, foram realizados três conjuntos de corridas de simulação do problema do jornaleiro, cada um constituído de M=100 corridas independentes. A única diferença entre cada um destes conjuntos residiu na duração de suas corridas:

N= 100 dias para o 1º conjunto
N= 500 dias para o 2º conjunto
N=1000 dias para o 3º conjunto

Variou-se a duração das corridas com o objetivo de se ilustrar um fato, que apesar de óbvio, é de fundamental importância: à medida em que a duração de uma corrida de simulação aumenta, também aumenta a precisão das estimativas que são objeto do estudo, ou, equivalentemente, a variância associada aos resultados torna-se cada vez menor.

Algum cuidado deve ser empregado aqui para que não sejam confundidos dois problemas distintos, muito embora possam levar a um mesmo resultado numérico:

i. Tal como apresentado, o problema do jornaleiro consiste no estudo da distribuição do seu lucro diário, através de uma amostra de valores simulados deste lucro e, neste caso, calculando-se $\bar{L}$ e S_L, a média e desvio-padrão do lucro simulado, como estimativas de μ_L e σ_L, a média e o desvio-padrão desta distribuição.

ii. O problema acima descrito, que foi aqui estudado, não deve ser confundido com o estudo da distribuição do lucro médio do jornaleiro num certo período de tempo, como por exemplo em 100 dias.

Em ambos os casos, as distribuições estudadas têm um mesmo valor esperado, e algumas outras características em comum; além disso, permitem o uso de praticamente um mesmo programa para sua simulação. No entanto, são dois problemas diferentes.

Situações como esta não são peculiares ao problema do jornaleiro, mas comuns à grande maioria dos estudos de simulação. A nosso ver, uma das principais fontes de erro na interpretação de resultados de simulação reside na comum confusão entre a duração de uma corrida de simulação e o período de tempo que se quer simular.

No caso do jornaleiro esta distinção fica clara se observarmos que, para o estudo da distribuição do lucro diário, o período simulado foi sempre o mesmo para todos os conjuntos de corridas, e igual a <u>1 dia</u>. Já com relação à duração da corrida, ela variou em função do particular conjunto de corridas considerado.

Para cada um destes conjuntos de corridas, foram calculados os seguintes valores, que são apresentados na Tabela I:

i. $\bar{\bar{L}} = \dfrac{\sum_{i=1}^{M} \bar{L}_i}{M}$, onde

$\bar{L}_i$ é o lucro médio diário relativo à corrida i deste conjunto.

ii. $\bar{S}_L = \dfrac{\sum_{i=1}^{M} S_{L_i}}{M}$, onde

S_{L_i} é o desvio-padrão do lucro diário relativo à corrida i deste conjunto.

iii. $$\mathrm{Var}(\bar{L}) = \frac{\sum_{i=1}^{M} (\bar{L}_i - \bar{\bar{L}})^2}{M-1}$$

Var($\bar{L}$) é uma medida da variabilidade associada à resposta $\bar{L}$.

iv. $$\mathrm{Var}(S_L) = \frac{\sum_{i=1}^{M} (S_{L_i} - \bar{S}_L)^2}{M-1}$$

Var(S_L) é uma medida da variabilidade associada à resposta S_L.

Os valores da Tabela I mostram que tanto μ_L como σ_L, respectivamente o lucro esperado e seu desvio-padrão, estão sendo corretamente estimados, isto é, livre de tendências, uma vez que $\bar{\bar{L}}$ e S_L situaram-se bastante próximos de seus correspondentes valores teóricos, que são:

$$\mu_L = 344.80$$

e

$$\sigma_L = 82.13$$

Já os valores de Var($\bar{L}$) e Var(S_L) mostram que, de fato, a precisão das estimativas melhora consideravelmente com o aumento de N, a duração da corrida.

Mais ainda, uma vez que para um particular dia simulado, o lucro independe do que ocorre em outros dias, valem as relações:

$$\sigma^2_{\bar{L}} = \frac{\sigma^2_L}{N} \tag{1}$$

e

Tabela I

Sumário dos Resultados Obtidos nos 3 Conjuntos de 100 Corridas de Simulação do Problema do Jornaleiro

CONJUNTO	DURAÇÃO DA CORRIDA (N)	$\bar{\bar{L}}$	$\bar{S}_L$	VAR($\bar{L}$)	VAR(S_L)
1	100	344.76	82.245	59.790	38.689
2	500	345.42	81.716	11.192	6.329
3	1000	345.03	81.979	6.513	3.922

$$\sigma^2_{S_L} \simeq \frac{\sigma^2_L}{2N} (1 + \frac{G_2}{2}) \qquad (2)$$

onde

σ^2_L = 6744.96 é a variância do lucro diário.

G_2 = 0.013 é a medida do excesso de curtose, dada por

$$G_2 = \frac{E(L - \mu_L)^4}{\sigma^4_L} - 3$$

Usando-se as expressões (1) e (2) foram obtidos os valores teóricos de $\sigma^2_{\bar{L}}$ e o $\sigma^2_{S_L}$ que foram estimados por Var($\bar{L}$) e Var (S_L).

Os valores de σ^2_L e $\sigma^2_{S_L}$ correspondentes a cada conjunto de corridas são fornecidos na Tabela II, juntamente com os de μ_L = E($\bar{L}$) e de σ_L = E(S_L) que, naturalmente, não variam com a duração da corrida.

Confrontando-se os valores teóricos da Tabela II com as respectivas estimativas da Tabela I, nota-se a existência de uma boa concordância entre eles.

Antes de prosseguirmos no estudo da variação dos resultados, serão abordados dois importantes aspectos relacionados com este:

i. Inicialmente, a duração de uma corrida define também o tamanho da amostra de demanda nela empregada. À medida em que a duração aumenta, esta amostra passa a ser cada vez mais representativa da população de demandas, ou em outras palavras, os erros amostrais

Tabela II

Valores Teóricos Associados às Distribuições

Amostrais de $\bar{L}$ e S_L

N	μ_L	σ_L	$\sigma^2_{\bar{L}}$	$\sigma^2_{S_L}$
100	344.80	82.13	67.450	33.94
500	344.80	82.13	13.49	6.79
1000	344.80	82.13	6.745	3.394

tornam-se cada vez menores.

O efeito desta redução do erro amostral manifesta-se de forma direta na precisão do resultado da simulação, já que, conforme notamos, a variabilidade associada aos resultados decresce com N.

Mais ainda, se se pudesse imaginar a possibilidade de se realizar corridas com duração infinita, tal erro amostral inexistiria e, em conseqüência, o resultado da simulação seria exato, igual àquele analiticamente obtido.

Pode ser argumentado que uma duração infinita não tenha significado algum do ponto de vista prático por se tratar de uma situação irreal.

Nossa opinião não é esta!

De fato, uma vez feita a distinção entre a duração de uma corrida e o período de tempo está sendo simulado, vê-se que os resultados do estudo continuariam válidos ainda que uma corrida tivesse duração infinita, visto que a duração da corrida é função dos recursos disponíveis (tempo de máquina) para a resolução do problema e como já foi dito, não deve ser confundida com o período de tempo que se está simulando que, neste caso, é sempre de um dia.

Em termos conceituais, uma corrida com duração infinita representa uma situação ideal em que os resultados seriam livres de erro, isto é, onde seria possível representar toda a informação contida num modelo probabilístico (no caso a distribuição de demanda) por meio de uma amostra.

Supondo-se a possibilidade de realização desta corrida infinita, os valores das variáveis de respostas $\bar{L}$ e S_L seriam exatamente iguais aos valores procurados de μ_L e σ_L.

Assim, vê-se que a melhor forma de se interpretar uma corrida de simulação com determinada duração não é como N dias de trabalho de jornaleiro mas sim como N observações de apenas um dia de trabalho.

ii. O segundo aspecto de importância deste problema em particular consiste no fato de que o resultado de uma corrida independe da seqüência na qual os valores amostrais são utilizados, pois, em geral o resultado de uma simulação depende de forma conjunta das duas fontes de erro amostral: o conjunto de valores gerados e sua seqüência (Saliby, 1980).

O problema do jornaleiro representa portanto um caso particular em que apenas um destes fatores de erro está presente, associado ao conjunto de valores de entrada.

É fácil de se verificar tal propriedade pois, uma vez fixada a composição da amostra de demanda, isto é, a freqüência de ocorrência de cada valor, o resultado de uma corrida será sempre o mesmo, qualquer que seja a ordem na qual tais valores sejam utilizados.

Exemplificando, numa corrida com duração de N=100 dias uma amostra aleatória de demandas geradas foi a seguinte:

Valor da Demanda	Freqüência
50	9
60	10
70	9
80	26
90	20
100	18
110	8

Independente da seqüência na qual estas 100 demandas diárias ocorrerem, os resultados de uma corrida serão sempre os mesmos:

$$\bar{L} = 355.20$$

e

$$S_L = 79.85$$

Uma vez que a ordem na qual os dias são simulados é irrelevante para a obtenção do resultado final, torna-se justificável a suposição de que toda a simulação ocorre num único lapso de tempo. Fazendo uma analogia, uma corrida de simulação não seria vista como um filme mas simplesmente como uma única fotografia.

Completando o raciocínio: uma vez que a seqüência de ocorrência dos eventos não é relevante, toda variabilidade associada aos resultados desta simulação é devida ao efeito de conjunto. Por sua vez, esta variabilidade associada ao efeito de conjunto tem duas características:

i. Ela pode ser explicada, em grande parte, através do Modelo Linear de Resposta (Saliby, 1980). Isto será visto na seção seguinte.

ii. De maior importância ainda é que, numa simulação, o erro amostral associado ao efeito de conjunto pode ser totalmente controlado com o uso de amostragem descritiva, levando à solução exata do problema.

III. ESTUDO DA VARIAÇÃO DOS RESULTADOS ATRAVÉS DO MODELO LINEAR DE RESPOSTA

De acordo com a teoria do Modelo Linear de Resposta (MLR) os resultados de uma corrida de simulação relacionam-se com os momentos amostrais das variáveis de entrada segundo um modelo de regressão linear.

A parcela de variação explicada pelo MLR depende de particular problema em estudo; no caso do jornaleiro, era esperado que este poder de explicação fosse bastante alto, uma vez que, como já foi visto anteriormente, a variabilidade dos resultados depende unicamente do efeito de conjunto, justamente aquele controlado através do MLR.

Com base nos resultados obtidos nos 3 conjuntos de corridas, foram estudadas as seguintes regressões:

$$\bar{L} = \mu_L + \alpha_L (\bar{D} - \mu_D) + \varepsilon$$

$$\bar{L} = \mu_L + \alpha_L (\bar{D} - \mu_D) + \beta_L (S_D - \sigma_D) + \varepsilon$$

$$S_L = \sigma_L + \alpha_S (\bar{D} - \mu_D) + \varepsilon$$

$$S_L = \sigma_L + \alpha_S (\bar{D} - \mu_D) + \beta_S (S_D - \sigma_D) + \varepsilon,$$

onde:

- $\bar{D}$ e S_D são respectivamente a média e o desvio-padrão da amostra de demandas gerada numa corrida.

- μ_D e σ_D são respectivamente a média e o desvio-padrão da distribuição de demanda.

Identificando-se pelo índice I as regressões utilizando apenas uma variável independente ($\bar{D}$) e pelo índice II as regressões utilizando as duas variáveis independentes ($\bar{D}$ e S_D), os valores obtidos para os coeficientes de correlação múltipla R, e para os coeficientes de determinação R^2, nos 3 conjuntos de corridas, são apresentados na Tabela III para a resposta $\bar{L}$ e na Tabela IV para a resposta S_L.

Nestas tabelas também são fornecidos os valores de $\hat{\mu}_L$ e $\hat{\sigma}_L$, estimativas de μ_L e σ_L, correspondentes ao termo independente de cada uma das regressões obtidas.

Os resultados das Tabelas III e IV mostram que grande parte da variabilidade associada à $\bar{L}$ e S_L é, de fato, explicada através do MLR, e que o termo constante da regressão é, também um bom estimador tanto para μ_L como para σ_L.

Mais ainda, os resultados mostram também que as correlações observadas não dependem da duração da corrida, sendo aproximadamente constantes para cada uma das regressões testadas. Este fato foi, na verdade, confirmado por outros resultados que obtivemos.

Por fim, tanto o efeito do aumento de duração de uma corrida como a dependência existente entre a resposta $\bar{L}$ e a média amostral das demandas geradas $\bar{D}$, pode ser visualizada através das figuras I, II e III, referentes, respectivamente, ao 1º, 2º e 3º conjunto de corridas.

Nestas figuras, obtidas com o auxílio do SPSS, o eixo horizontal representa os valores de $\bar{D}$ enquanto que o eixo vertical representa os respectivos valores de $\bar{L}$.

Um asterisco representa um par de valores ($\bar{D}$, $\bar{L}$) relativo a uma particular corrida, enquanto que os eventuais algarismos também presentes nas figuras indicam o número de observações coincidentes; assim, por exemplo, um <u>2</u> indica a superposição de 2 destes pontos.

Tabela III

Resultados Obtidos Usando-se o Modelo Linear de

Resposta no Estudo da Resposta $\bar{L}$

CONJUNTO DE CORRIDAS	REGRESSÃO USANDO $\bar{D}$ COMO VARIÁVEL INDEPENDENTE			REGRESSÃO USANDO $\bar{D}$ e S_D COMO VARIÁVEIS INDEPENDENTES		
	R_I	R_I^2	$\hat{\mu}_L$	R_{II}	R_{II}^2	$\hat{\mu}_L$
1	.85	.72	344.50	.98	.97	344.66
2	.84	.71	344.91	.99	.98	344.77
3	.84	.71	344.76	.99	.98	344.77

Tabela IV

Resultados Obtidos Usando-se o Modelo Linear de

Resposta no Estudo da Resposta S_L

CONJUNTO DE CORRIDAS	REGRESSÃO USANDO $\bar{D}$ COMO VARIÁVEL INDEPENDENTE			REGRESSÃO USANDO $\bar{D}$ e S_D COMO VARIÁVEIS INDEPENDENTES		
	R_I	R_I^2	$\hat{\sigma}_L$	R_{II}	R_{II}^2	$\hat{\sigma}_L$
1	.68	.46	82.08	.90	.81	82.84
2	.65	.42	81.84	.89	.80	82.13
3	.74	.55	81.99	.92	.85	82.12

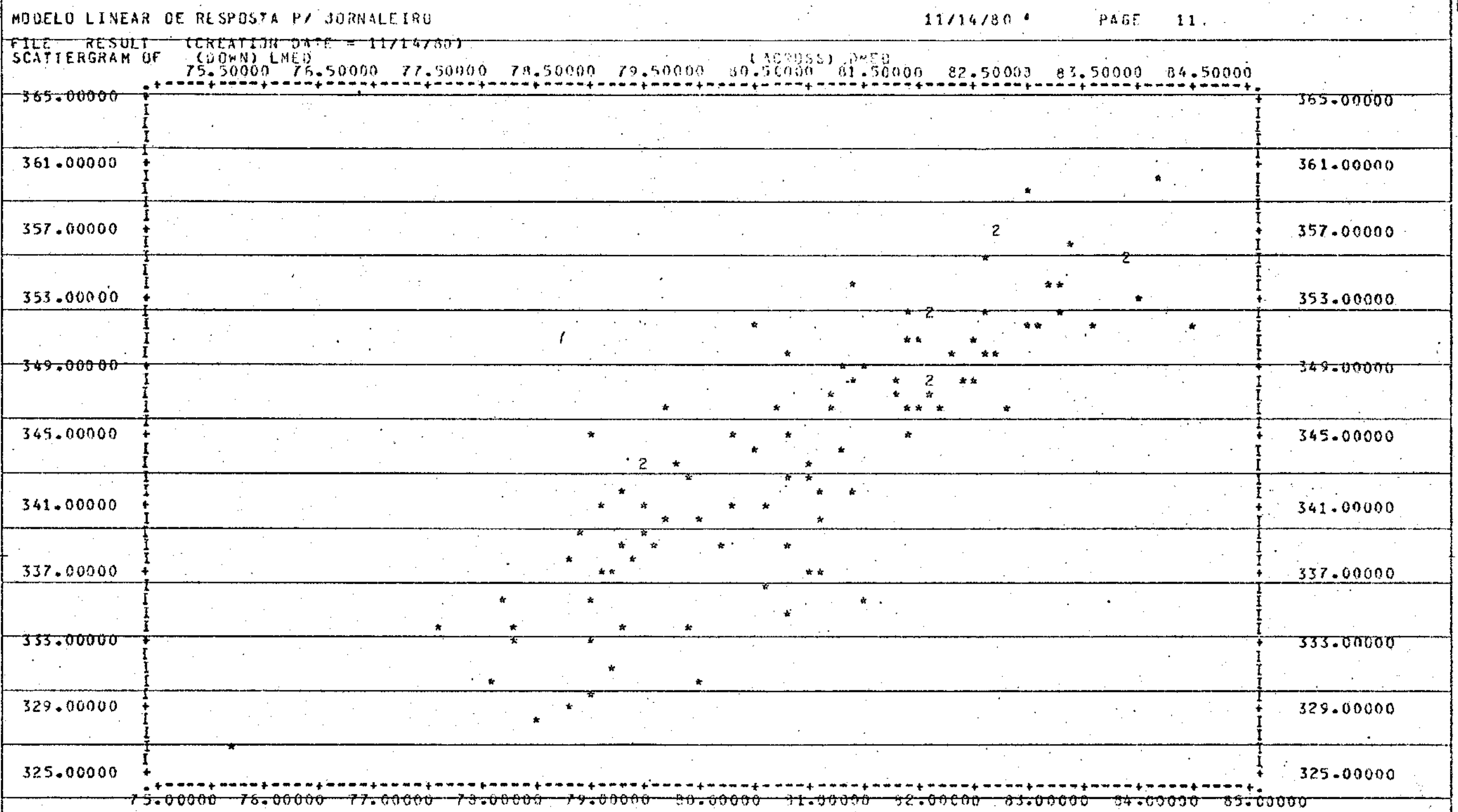

Figura 1

Variação de $\bar{L}$ com $\bar{D}$ para corridas com duração N = 100

MODELO LINEAR DE RESPOSTA P/ JORNALEIRO 11/14/80 PAGE 11

FILE RESULT (CREATION DATE = 11/14/80)
SCATTERGRAM OF (DOWN) LMED (ACROSS) DMED

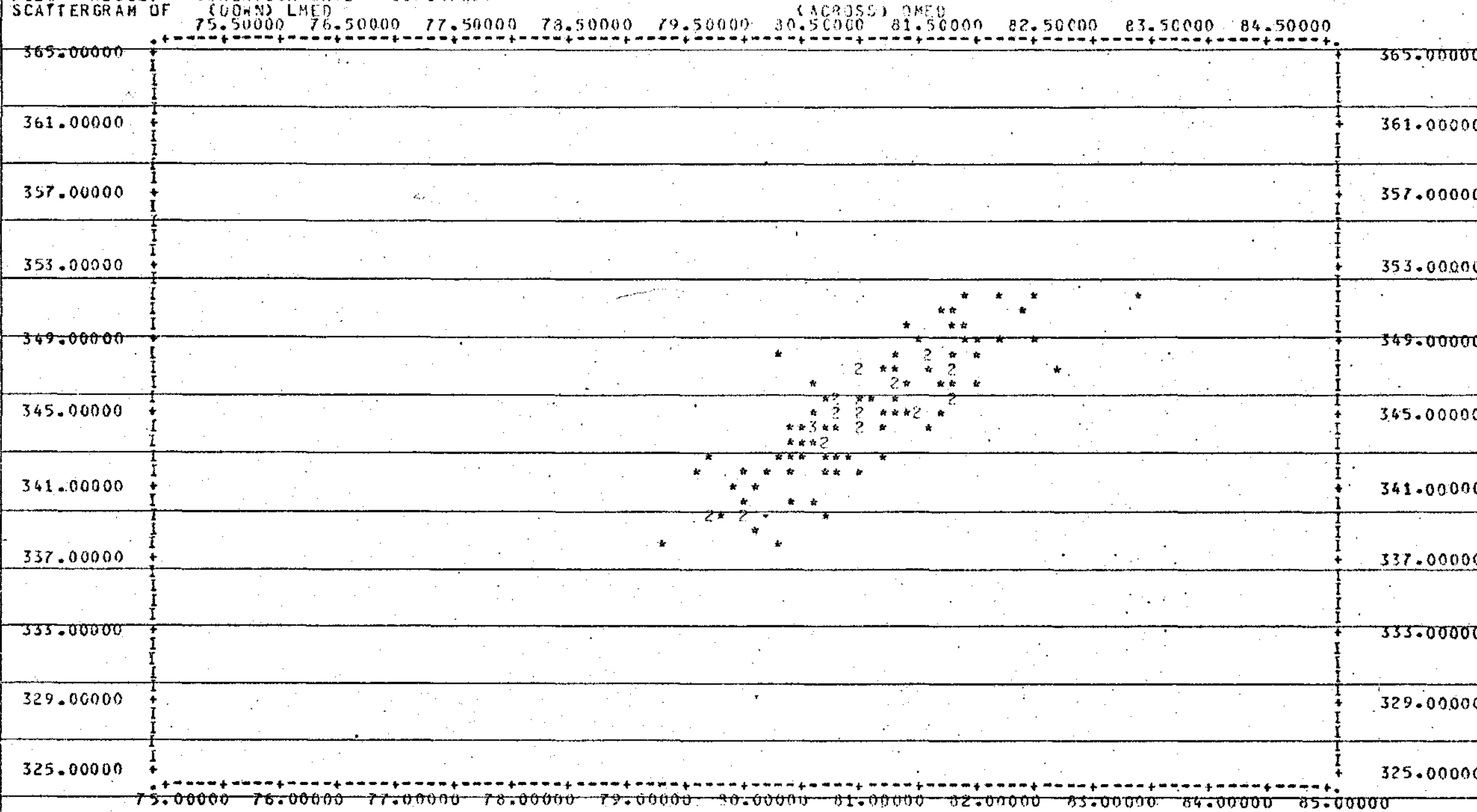

Figura 2

Variação de $\bar{L}$ com $\bar{D}$ para corridas com duração N = 500

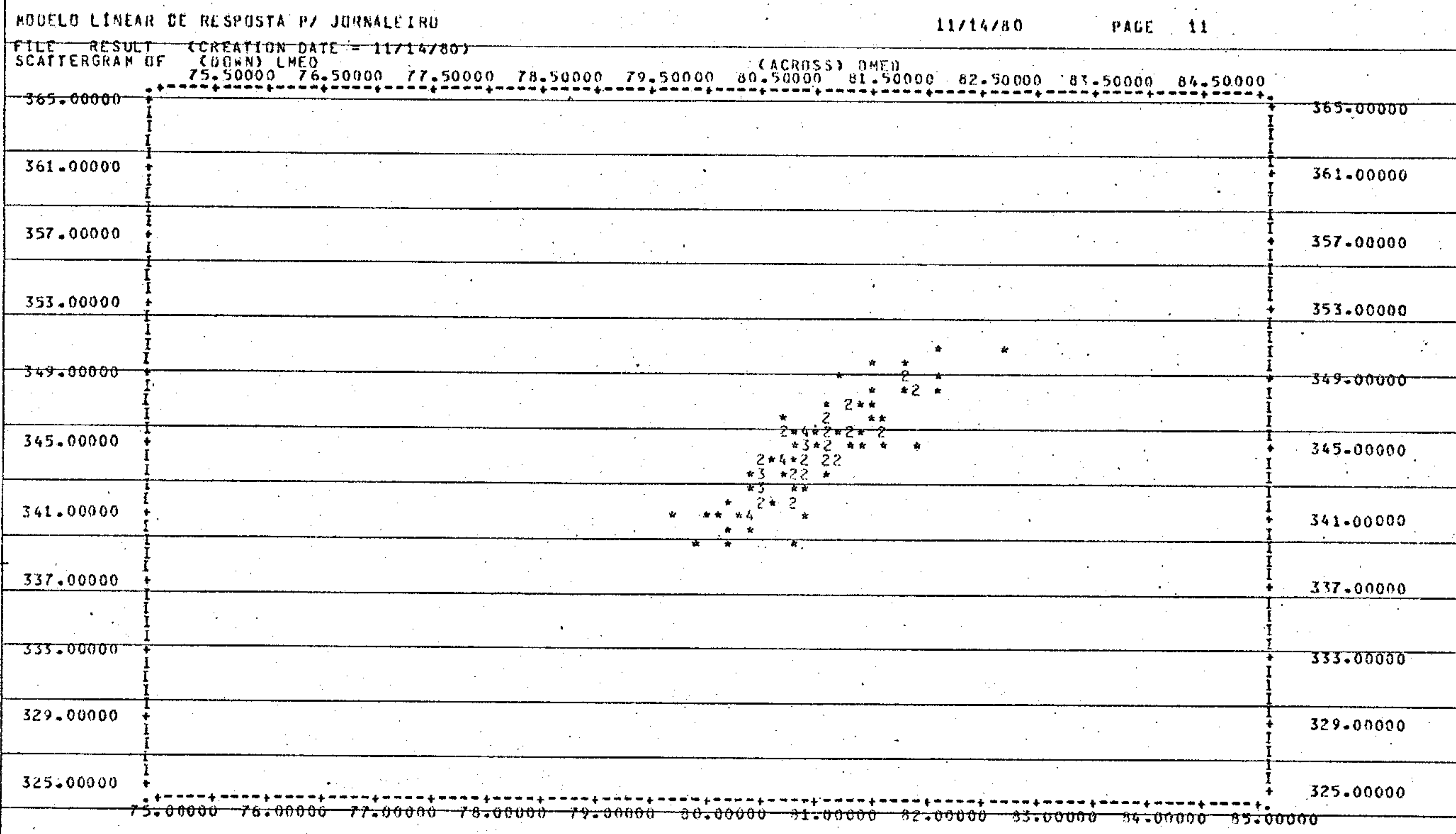

Figura 3

Variação de $\bar{L}$ com $\bar{D}$ para corridas com duração N = 1000

Para permitir uma comparação adequada, as 3 figuras abrangeram um mesmo intervalo de variação tanto para a demanda média (75, 85), como para o lucro médio (325, 365).

Da análise destas figuras pode-se concluir que:

i. A utilização do Modelo Linear de Resposta para os resultados desta simulação é, de fato, aceitável já que a dependência linear de $\bar{L}$ com $\bar{D}$ é clara.

ii. Ao se aumentar a duração de uma corrida, nota-se um efeito de concentração do aglomerado de pontos em torno de um valor central.

Este efeito de concentração deve-se à redução do erro amostral associado à amostra de demanda; os valores de $\bar{D}$ tornam-se cada vez menos dispersos e, em conseqüência, também os valores da resposta $\bar{L}$.

É possível imaginar-se o que ocorreria se a duração da corrida fosse infinita, ou seja, se o erro amostral não existisse. Neste caso limite, a nuvem de pontos iria se resumir num único ponto central da figura, correspondendo ao par (μ_D, μ_L), que, por sinal, representa a solução exata do problema.

IV. CRÍTICA AO USO DE AMOSTRAGEM ALEATÓRIA DE SIMULAÇÃO

Ora, se a situação ideal é aquela onde não ocorrem erros amostrais, que neste caso derivam unicamente do efeito de conjunto, porque então sujeitarmo-nos a eles, tendo em vista que, ao se realizar a simulação, a distribuição da demanda é suposta conhecida?

O uso de amostragem aleatória simples tem como principal virtude evitar a introdução de tendências na seleção amostral, tendências estas decorrentes da possibilidade de existirem relações desconhecidas entre a variável em estudo e o processo de seleção empregado (Stuart, 1976).

Tal argumento se aplica, no entanto, quando o uso de amostras refere-se a um processo inferencial, isto é, a um processo de obtenção de informação a respeito de uma população; neste caso, por se ter sempre um conhecimento incompleto acerca da realidade em estudo, nada nos resta senão sermos imparciais na seleção amostral e por isso usarmos um procedimento aleatório.

Numa simulação, porém, a situação é totalmente diferente:

Em primeiro lugar, o objetivo de uma amostra numa simulação não é a obtenção de informação mas sim a descrição de uma informação já disponível acerca de uma população. Convém salientar que qualquer estudo de simulação pressupõe a validade de seu modelo, ou seja, pressupõe um conhecimento completo de suas distribuições de probabilidade, sem nenhuma margem de incerteza.

Ora, se não existem incertezas associadas à população em estudo (ao menos a nível teórico), então não há a necessidade de se utilizar um procedimento aleatório na seleção amostral. É bom lembrar, que a margem de erro associada ao uso de amostragem aleatória, é bastante grande.

Supondo-se conhecido *a priori* o tamanho de amostra a ser empregado, o procedimento, neste caso, é a seleção da amostra de forma intencional, de modo que seus valores tenham uma distribuição acumulada de freqüência que seja a mais próxima possível da respectiva função de distribuição de probabilidades, garantindo-se assim uma máxima representatividade amostral, representatividade esta que é a *raison d'être* de qualquer amostra.

O método de amostragem descritiva torna possível a obtenção de amostras que, a menos de eventuais erros de arrendondamento, tenham seus histogramas praticamente iguais aos respectivos gráficos das funções de densidade das distribuições que se quer representar. As principais características do método são:

i. Os valores amostrais são deterministicamente selecionados.

ii. A seqüência ou permutação na qual tais valores amostrais são empregados numa simulação é aleatoriamente obtida.

Para o estudo do problema do jornaleiro foi utilizada uma amostra descritiva de tamanho N=100, a qual, tendo em vista a ausência de erros de aproximação, reproduziu fielmente a distribuição de probabilidades associada à demanda diária. A composição desta amostra foi a seguinte:

Valor	Freqüência
50	10
60	12
70	15
80	20
90	18
100	15
110	10

O resultado da simulação assim realizada, e que independeu da ordem na qual os valores amostrais foram utilizados, foi exatamente igual a solução teórica do problema, isto é, μ_L. E isto não foi coincidência!

Convém salientar que em decorrência do uso de amostras descritivas não foi apenas μ_L que se estimou sem erro, mas toda a distribuição associada ao lucro diário, já que, agora, o processo de cálculo utilizado na simulação, não difere em nada do método analítico de resolução!

Note-se também que em se utilizando amostras descritivas de tamanho 100, 500 ou 1000, chega-se sempre a um mesmo resultado, uma vez que é sempre possível representar com exatidão, em qualquer uma destas amostras, a distribuição de freqüências desejada.

É importante frizar que a obtenção de uma solução única e exata para este problema por meio de amostragem descritiva só foi possível por ser um caso particular em que a variação dos resultados depende unicamente do efeito do conjunto. É também oportuno, mencionar que caso o problema fosse outro: ao invés da distribuição do lucro diário, o problema consistisse no estudo da distribuição do lucro médio do jornaleiro, durante um período de K=100 dias de operação, então seria necessária a obtenção de uma amostra de valores deste lucro médio, digamos N=50.

Neste caso, dever-se-ia tomar o cuidado de se definir uma corrida como sendo constituída de um conjunto de N períodos de K dias simulados. As variáveis de resposta forneceriam estimativas relacionadas à distribuição do lucro médio, como por exemplo, seu valor esperado (que é igual ao valor esperado do lucro diário) e seu desvio-padrão.

Desta forma, uma corrida seria baseada num total de N.K= 5000 valores amostrais da demanda, e, ao contrário do problema anteriormente estudado, seu resultado iria também depender da particular seqüência na qual estes valores fossem utilizados.

Finalmente, é importante notar que, neste caso, uma amostra descritiva da distribuição de demanda seria constituída de 5000 valores, os quais deveriam ser agora aleatoriamente permutados.

V. CONCLUSÃO

O objetivo deste estudo foi o de confirmar nosso ponto de vista de que não é correto o uso de amostragem aleatória em simulação. Para isso, estudamos o problema do jornaleiro, um caso particular em que a seqüência na qual os valores amostrais são utilizados, não influencia o resultado final. Portanto, de acordo com nossa teoria, e uma vez controlada a fonte de erro associada ao conjunto de valores gerados, possível através do uso de amostragem descritiva, o resultado da simulação deveria ser igual ao valor teórico procurado, a menos, é claro, de eventuais erros incorridos na aproximação de uma distribuição teórica por um conjunto finito de valores, que aliás não foi o caso do presente estudo.

Ainda mais, as técnicas de redução de variância, como por exemplo variáveis antitéticas, são, em sua quase totalidade, meros, e às vezes até mesmo obscuros controles deste erro amostral associado ao conjunto de valores, erro este que está sendo desnecessariamente introduzido no estudo ao se usar indevidamente o método de amostragem aleatória.

Desta forma, através da amostragem descritiva não só minimizamos os erros, como também praticamente eliminamos a utilidade das técnicas de redução de variância, que aliás, a nosso ver, sempre se constituiram numa incoerência já que, enquanto de um lado era feito o uso de um procedimento aleatório na seleção da amostra, do outro era recomendado o uso de um procedimento sistemático para se atingir uma maior precisão no estudo!

A nosso ver, o estudo do problema do jornaleiro tem como principal vantagem evidenciar a procedência das críticas que fizemos. Deve ser salientado também que a utilização de amostragem descritiva a outros problemas já foi também estudada, tais como a simulação de um sistema de filas e de um outro de estoques (Saliby, 1980); mais recentemente (Saliby, 1981b) foi utilizada a amostragem descritiva numa aplicação do Método de Monte Carlo para o estu do da distribuição da amplitude de uma amostra de valores uniformemente distribuídos. Em todos estes casos foi comprovada a superioridade do método de amostragem descritiva.

Finalizando, tendo em vista que até agora uma das idéias básicas em simulação (e do método de Monte Carlo) reside no uso de amostragem aleatória simples, cremos ser chegada a hora de uma importante revisão do assunto, útil, também, para desmistificar a utilização da técnica de simulação, tornando-a melhor compreendida e aceita por todos.

./sfa.

BIBLIOGRAFIA

SALIBY, E. (1980). A Reappraisal of Some Simulation Fundamentals. Lancaster, U.K. University of Lancaster. Tese de Doutorado.

_________. (1981a). Uma Crítica ao uso de números aleatórios em simulação. Relatório de Pesquisa da COPPEAD (a ser publicado).

_________. (1981b). Sobre o uso incorreto de números aleatórios em simulação: um estudo sobre a distribuição da amplitude de uma amostra com distribuição uniforme. Relatório de Pesquisa nº 24, COPPEAD/UFRJ.

STUART, A. Basic Ideas of Scientific Sampling. 2.ed. London, Charles Griffin, 1976.