

VOLUME DE MADEIRA EM ÁREA REFLORESTADA: UM ESTUDO SIMULADO DE INFERÊNCIA ESTATÍSTICA

PAUL GERHARD KINAS*
JUARENZE CARDOSO NEVES**

RESUMO

O entendimento de conceitos de inferência estatística clássica como distribuição amostral e intervalos de confiança pode ser facilitado e ampliado recorrendo-se a estudos simulados. Apresenta-se um exemplo fictício em que o volume médio de madeira por hectare precisa ser estimado a partir de informações amostrais sobre a quantidade, perímetros e alturas de árvores. A simulação é implementada no Statistica©. Análise descritiva de 2000 réplicas permite conhecer as características do estimador não-convencional utilizado. Verifica-se que sua distribuição amostral é aproximadamente lognormal e que o estimador tem viés. A cobertura do intervalo de confiança e a probabilidade do erro II em um teste de hipótese são avaliadas, e os efeitos resultantes de suposições não atendidas, verificadas. Conclui-se que o processo de simulação é simples, e que sua utilização em cursos introdutórios de inferência é viável. Além da oportunidade de vivenciar concretamente os conceitos, abre-se ao estudante a possibilidade para explorar situações não-convencionais.

PALAVRAS-CHAVE: estimação, teste de hipóteses, simulação estocástica.

ABSTRACT

Wood volume in a forested area: a simulated study in statistical inference

Classical concepts of statistical inference like sampling distribution and confidence intervals can be simplified and made more effective with simulation studies. We present a fictitious example where the volume of wood per hectare needs to be estimated from sampling information about quantity, perimeter and height of trees. Simulations are implemented in Statistica©. A basic descriptive analysis of 2000 replicates lets us know the characteristics of our unconventional estimator. We verify that the sampling distribution is approximately lognormal and that the estimator is biased. The coverage of the confidence interval and the type II error probability in a given hypothesis test are evaluated and the effects of unmatched suppositions verified. We conclude that the simulation procedure is simple

* Professor do Dep. de Matemática – FURG. dmtkinas@super.furg.br

** Professor do Dep. de Matemática – FURG. dmtjcn@super.furg.br

and that its use in introductory inference courses feasible. It can give the student an opportunity to experience abstract concepts and simultaneously go beyond standard situations.

KEY WORDS: inference, test of hypotheses, stochastic simulation.

1 – INTRODUÇÃO

A distribuição amostral de um estimador e as probabilidades dos erros tipo I e II (em testes de hipótese) às vezes são de difícil compreensão para estudantes em cursos introdutórios de inferência estatística. Além disso, em vista do reduzido leque de estimadores abordados nestes cursos (médias, proporções, desvio-padrão), cria-se uma percepção equivocada de que o leque de utilizações práticas é limitado para aqueles que não dispõem de profundo conhecimento do ferramental de estatística matemática.

A possibilidade de utilizar-se dos recursos de simulação, além de propiciar ao estudante um meio novo de aprender esses conceitos, vivenciando-os em estudos empíricos, amplia muito as possibilidades da sua aplicação a problemas não-convencionais. Por exemplo, experimentar o surgimento de uma distribuição amostral através da simulação de um grande número de réplicas de um experimento facilita a compreensão dos conceitos sobre a forma da distribuição, erro amostral, esperança e viés.

Com a finalidade de ilustrar essa possibilidade, apresenta-se um exemplo cuja solução por procedimentos convencionais seria complexa. Demonstra-se que, com a aquisição de habilidades mínimas para simulação de dados em planilhas eletrônicas, sua solução torna-se simples e perfeitamente manejável num curso introdutório de inferência. Espera-se que o reconhecimento da simplicidade na análise motive o estudante a explorar novas possibilidades em sua área de interesse, libertando-se das suas aparentes limitações.

O texto é estruturado em várias seções. Na seção dois, apresenta-se o exemplo a ser analisado. Na seção três, apresenta-se a análise da distribuição amostral do estimador. Na seção quatro, examina-se como falhas nas suposições podem afetar o processo de estimação. Na quinta seção, aborda-se o cálculo de erros em testes de hipótese. Finalmente, na sexta e última seção, comentam-se os ensinamentos extraídos do exercício. Todas as simulações, cálculos estatísticos e elaboração de gráficos foram realizados no Statistica© (StatSoft Inc., 1998), utilizando apenas o módulo de estatística básica (*Basic Statistics*).

2 – O CASO EM ESTUDO

O volume de madeira disponível numa grande área reflorestada define o preço que uma fábrica de celulose está disposta a pagar para adquiri-la. O volume total médio de madeira por hectare (μ_T) pode ser obtido em função de medidas mais simples: o perímetro médio (μ_P), a altura média (μ_H) e o número médio de árvores por hectare (μ_N).

Para obter informação sobre esses parâmetros dispõe-se de três amostras (supostamente) independentes que fornecem os valores médios de perímetro (P), altura (H) e do número de árvores por hectare (N). Com base nessas estimativas, propõe-se estimar o volume total médio de madeira por hectare (μ_T), por meio da expressão:

$$T = N \frac{P^2}{4\pi} H. \quad (1)$$

O estimador T foi obtido pelo método dos momentos (Hoel, 1981). Este não é o único método para encontrar um estimador, nem tampouco o melhor. Sua escolha deve-se à simplicidade e ao apelo intuitivo adequados num enfoque introdutório. É importante notar que as médias amostrais N , P e H estão sujeitas a variabilidades probabilísticas (os erros amostrais), que são modeladas pelas respectivas distribuições amostrais. Para amostras suficientemente grandes ($n > 30$), poderíamos recorrer ao Teorema Central do Limite (t.c.l.) para justificar distribuições amostrais aproximadamente normais para cada uma dessas variáveis. No entanto, para o estimador de interesse T , que também é uma variável aleatória, a distribuição amostral não é óbvia.

Para alguém com formação sólida em probabilidade e em cálculo, um estudo teórico dessa distribuição é factível. No entanto, aqui os objetivos são outros. Primeiramente, quer-se demonstrar que, contando apenas com conhecimento introdutório de probabilidade e estatística, pode-se obter inferência útil sobre μ_T com um estudo simulado bastante simples.

O segundo objetivo está na ilustração de conceitos essenciais da estatística freqüentista: intervalos de confiança, nível de significância e poder de um teste. Finalmente, o terceiro objetivo está na avaliação dos possíveis efeitos de suposições não satisfeitas sobre a inferência. Propõe-se que a utilização de réplicas simuladas facilita o entendimento desses conceitos que, não raro, parecem obscuros ao usuário eventual.

Fixando arbitrariamente os parâmetros $\mu_N = 400$ árvores/hectare, $\mu_P = 0,9\text{m}$ e $\mu_H = 7,7\text{m}$, tem-se que $\mu_T = 198,5\text{m}^3$. Privilegiados pelo

conhecimento do volume médio real de madeira por hectare, pode-se passar a estudar por meio de simulações a distribuição amostral do estimador T , assim como o viés e a precisão.

Para simular as réplicas, parte-se do pressuposto de que, a partir de estudos de campo, seja adequado modelar as distribuições amostrais dos diversos componentes como segue:

$P \Rightarrow$ Distribuição *Normal* (0,90; 0,22);

Ou seja, para uma distribuição *Normal* (a , b), a densidade de probabilidade é dada por:

$$f(P) = \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2b^2}\right), \quad (2)$$

sendo a esperança $E(P) = a$ e o desvio-padrão $DP(P) = b$. Quando uma distribuição de probabilidade surge em consequência de variabilidade entre amostras, ela recebe o nome de distribuição amostral e o seu desvio-padrão é denominado erro-padrão no jargão estatístico.

$H \Rightarrow$ Distribuição *Lognormal* (2; 0, 3);

Para uma distribuição *Lognormal* (a , b), a densidade de probabilidade é:

$$f(H) = \frac{1}{xb\sqrt{2\pi}} \exp\left(-\frac{(\ln(H)-a)^2}{2b^2}\right), \quad (3)$$

sendo $E(H) = \exp(a + b^2/2)$ e $DP(H) = \sqrt{e^{2a}e^{b^2}(e^{b^2} - 1)}$.

$N \Rightarrow$ Distribuição *Uniforme discreta* entre 300 e 500.

Uma distribuição de probabilidade uniforme discreta entre os números naturais a e b (sendo $a < b$) tem massa de probabilidade dada por:

$$P(N = x) = \frac{1}{b-a+1} I_{\{x=a, a+1, \dots, b\}}, \quad (4)$$

onde $I_{\{A\}}$ é função indicadora que vale 1 para o conjunto A , e zero em caso contrário. Tem-se ainda que $E(N) = (a + b)/2$ e $DP(N) = \{[(b(b-1)(2b+1) - (a-1)a(2a-1))/6(b-a+1) - (a+b)^2/4]^{1/2}$.

Note-se que os estimadores, embora com distribuições amostrais diferentes, são todos não-viesados, uma vez que suas esperanças

coincidem com os respectivos parâmetros. Isto é, $E(P) = \mu_P$, $E(H) = \mu_H$ e $E(N) = \mu_N$.

3 – DISTRIBUIÇÃO AMOSTRAL DE T

No Statistica©, criou-se um arquivo de dados com 2000 linhas (casos) e com três colunas para variáveis H , N , e P (Figura 1). Para cada uma das colunas, gerou-se uma amostra de variáveis aleatórias em conformidade com as distribuições especificadas anteriormente. Acrescentou-se uma quarta coluna com valores calculados de T em cada uma das 2000 linhas. As demais colunas foram acrescentadas posteriormente em conformidade com a necessidade da análise. A sintaxe utilizada para a criação das colunas está resumida em um anexo.

	1 H	2 N	3 P	4 T	5 N_LI95	6 N_LS95	7 N_IC95	8 LI95	9 LS95	10 IC95	11 DIF_AMP1
1	8,590	457,000	,500	78,092	-175,748	331,932	1	22,803	267,438	1	263,044
2	6,667	496,000	,755	150,193	-103,646	404,033	1	43,857	514,360	1	37,176
3	6,424	323,000	,736	89,451	-164,389	343,290	1	26,120	306,337	1	227,462
4	6,588	443,000	1,199	333,918	80,079	587,758	1	97,504	1143,554	1	-538,370
5	8,042	487,000	,464	67,124	-186,716	320,964	1	19,600	229,876	1	297,403
6	5,557	435,000	1,068	219,583	-34,257	473,423	1	64,118	751,996	1	-180,198
7	9,340	355,000	,332	29,082	-224,758	282,922	1	8,492	99,595	0	416,576
8	8,210	383,000	,627	98,213	-155,627	352,052	1	28,678	336,345	1	200,013
9	9,661	396,000	,714	154,995	-98,845	408,835	1	45,259	530,804	1	22,134
10	5,315	402,000	1,470	367,612	113,772	621,451	1	107,343	1258,943	1	-643,921
11	10,591	345,000	,620	111,812	-142,027	365,652	1	32,649	382,919	1	157,410
12	5,595	438,000	,990	191,306	-62,533	445,146	1	55,862	655,158	1	-91,617
13	5,358	478,000	1,047	223,502	-30,338	477,342	1	65,263	765,417	1	-192,475

FIGURA 1 – As 11 primeiras colunas (variáveis) da planilha. Cada linha representa uma simulação. Aparecem as primeiras 13 linhas do total de 2000 simulações (ver detalhes no texto).

A distribuição desses 2000 valores de T pode ser interpretada como uma construção empírica da sua distribuição amostral. Algumas estatísticas descritivas desta distribuição estão listadas na Tabela 1, identificadas como caso ' H e P independentes'. Consta-se que a média da distribuição amostral (isto é, o valor aproximado de $E(T)$) resultou acima do valor esperado ($\mu_T = 198,5m^3$), indicando que o estimador T é viesado, tendendo a superestimar o volume médio de

madeira por hectare. O viés, definido como a diferença $E(T) - \mu_T$, é de $11,88m^3$.

TABELA 1 – Estatísticas descritivas de distribuições amostrais para o estimador T . (Q1 e Q3 denotam 1º e 3º quartil, DP é desvio-padrão)

Simulação	Min	Q1	Média	Mediana	Q3	Max	DP
H e P indep	11,48	120,76	210,37	183,10	275,94	1016,98	123,86
Lognormal(*) (a=5,17 b=0,63)	0	115,34	214,56	176,16	269,07	$+\infty$	149,17
H e P depend	8,07	102,74	240,21	182,22	307,74	2419,53	213,04

(*) valores calculados no Statistica©

A forma da distribuição amostral de T é visualizada pelo histograma (Figura 2). Observa-se uma distribuição assimétrica com maior concentração nos valores mais baixos (assimetria positiva). Solicitou-se ao Statistica© que ajustasse uma curva lognormal à distribuição dos valores de T . Esta distribuição, que tem seus parâmetros listados na Tabela 1, produz um ajuste aceitável conforme alguns parâmetros dessa distribuição listados na linha 'Lognormal' da Tabela 1. Isto também se verifica na inspeção na Figura 2.

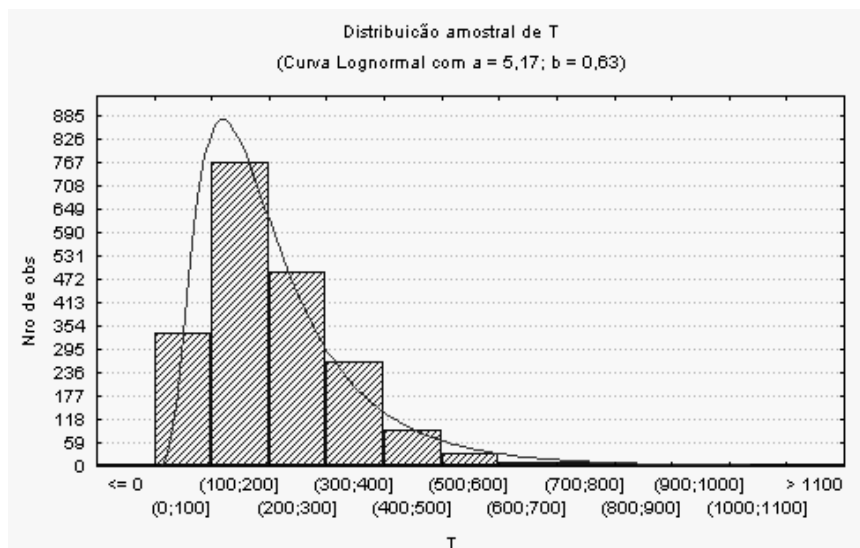


FIGURA 2 – Histograma da distribuição amostral de T baseado em 2000 simulações. A linha contínua é a curva lognormal ajustada aos dados.

A inferência sobre um parâmetro geralmente é apresentado na forma de um *intervalo de confiança* definido como um intervalo de valores que pretende cobrir o parâmetro de interesse com um grau de confiança especificado (Triola, 1999). Os níveis de confiança comumente utilizados são de 90%, 95% ou 99%.

Para um nível de confiança de 95% o intervalo de confiança, pressupondo uma distribuição amostral Normal, seria dado por $T \pm 1,96 DP(T)$. Espera-se, portanto, que, em construindo estes intervalos de confiança para as 2000 linhas, 95% delas contivessem μ_T . Para verificar isso no Statistica© acrescentam-se as colunas 5 e 6, definidas como limites mínimo e máximo do intervalo. A seguir, na coluna 7 com um teste lógico (zero-um), procede-se à identificação dos intervalos que cobriram μ_T . O somatório dos casos bem-sucedidos foi de 1918, que, em relação ao universo de 2000 tentativas, representa 95,9%, sugerindo uma cobertura real um pouco acima do anunciado. No entanto, examinando alguns casos isoladamente, verifica-se que esses intervalos de confiança podem produzir valores absurdos. Por exemplo, tome-se a segunda linha em que $H = 6,7$, $N = 496$ e $P = 0,76$, resultando em $T = 150,2m^3$. O intervalo de confiança correspondente vai de $-103,6m^3$ a $404,0m^3$. Este intervalo, além de incluir um limite inferior negativo, é amplo demais para ter utilidade prática.

A acentuada assimetria da distribuição amostral (Figura 2) fornece a explicação para esse insucesso. Para distribuições lognormais, é preferível construir primeiramente intervalos de confiança para $\log(T)$ transformando apenas os limites obtidos para a variável original. Ou seja, os limites T_{inf} e T_{sup} são assim definidos:

$$T_{inf} = \exp(\ln T - 1,96 \cdot 0,63) \text{ e } T_{sup} = \exp(\ln T + 1,96 \cdot 0,63), \quad (5)$$

onde 0,63 é o valor atribuído a b na distribuição lognormal ajustada para T (Tabela 1).

Para o caso descrito acima (linha 2), o intervalo de confiança agora vai de $43,9m^3$ a $514,4m^3$, conforme indicado nas colunas 8 e 9. Este intervalo, além de evitar valores absurdos, reduziu a amplitude em $37,1m^3$ em comparação ao procedimento anterior (coluna 11). Em média, a redução na amplitude é bem maior, como sugere o intervalo de confiança de 95% para esta diferença, estendendo-se de $134,3m^3$ a $168,3m^3$. Apesar da amplitude média significativamente menor, a cobertura desse intervalo de confiança continua sendo adequada, já que 1889 dos 2000 casos simulados, ou seja 94,5%, cobrem o valor real (coluna 10).

4 – APERFEIÇOANDO O MODELO DE GERAÇÃO DE DADOS

Um aspecto da inferência que é de difícil análise num curso introdutório refere-se aos efeitos sobre estimação e teste de hipóteses, de suposições não atendidas; isto é, a robustez do método. No processo de estimação descrito na seção anterior, a independência entre N , H e P foi tacitamente aceita. Pretende-se analisar nesta seção os possíveis efeitos desta suposição, quando na realidade ela está ausente nos dados.

Supondo independência entre o perímetro e a altura de árvores, criam-se situações pouco realistas como, por exemplo, um valor excessivamente baixo de perímetro associado a um valor muito grande de altura. Para evitar essas distorções, introduziu-se no modelo uma relação linear entre o perímetro e a altura, de forma que a média do perímetro fosse corrigida (para mais ou para menos de $\mu_T = 0,9$) conforme o afastamento (medido em unidades de desvio-padrão) de altura H de sua média observada \bar{H} . Ou seja, modela-se o perímetro médio de P condicionando-o ao valor de H seguindo uma distribuição normal com desvio-padrão fixado em 0,12 e esperança condicional.

$$E(P | H) = 0,9 + 0,2 \frac{H - \bar{H}}{S_H} \quad (6)$$

A relação que resulta entre valores de H e P fica evidenciada na Figura 3. Note-se, no entanto, que este gráfico somente é possível devido à simulação. Na prática da inferência, apenas um ponto estaria disponível ao pesquisador, e a relação entre os estimadores impossível de ser diagnosticada. Nessa nova condição e para os valores de H já disponíveis, geraram-se novas colunas de valores (Figura 4) para perímetro (P_2 ; coluna 12) e, na seqüência, calcularam-se novos valores para o volume (T_2 ; coluna 13).

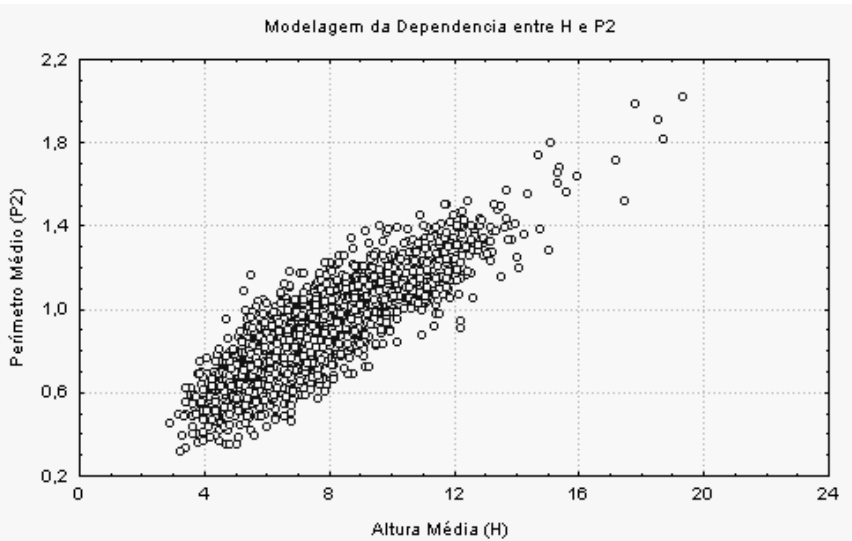


FIGURA 3 – Diagrama de dispersão dos valores de altura média H e perímetro médio P_2 .

STATISTICA: Basic Statistics and Tables

File Edit View Analysis Graphs Options Window Help

22.8028877800955

Vars Cases

Data: MAD_Final.STA 18v * 2000c

NUMER	7	8	9	10	11	12	13	14	15	16	17	18
VALUES	N IC95	LI95	LS95	IC95	DIF AMPL	P2	T2	LLI95	LLS95	IIC95	BETA	BETA2
1	1	22.803	267.438	1	263.044	.941	276.479	80.747	946.666	1	1	1
2	1	43.857	514.360	1	37.176	.738	143.153	41.809	490.155	1	1	1
3	1	26.120	306.337	1	227.462	.844	117.486	34.313	402.273	1	1	1
4	1	97.504	1143.554	1	-538.370	.878	178.876	52.242	612.471	1	1	1
5	1	19.600	229.876	1	297.403	.974	295.895	86.418	1013.144	1	1	1
6	1	64.118	751.996	1	-180.198	.724	100.970	29.489	345.720	1	1	1
7	1	8.492	99.595	0	416.576	1.179	366.788	107.123	1255.881	1	1	0
8	1	28.678	336.345	1	200.013	.981	240.806	70.329	824.521	1	1	1
9	1	45.259	530.804	1	22.134	1.019	316.261	92.366	1082.876	1	1	1
10	1	107.343	1258.943	1	-643.921	.644	70.407	20.563	241.075	1	0	1
11	1	32.649	382.919	1	157.410	1.059	326.146	95.253	1116.725	1	1	1
12	1	55.862	655.158	1	-91.617	.637	79.059	23.090	270.697	1	1	1
13	1	65.263	765.417	1	-192.475	.694	98.283	28.704	336.520	1	1	1

FIGURA 4 – As colunas (variáveis) 7 a 18 da planilha. Cada linha representa uma simulação. Aparecem as primeiras 13 linhas do total de 2000 simulações. (ver detalhes no texto)

Utilizando o procedimento descrito anteriormente, a distribuição amostral torna-se ainda mais assimétrica e dispersa (Tabela 1), com média de $240,21m^3$ e viés positivo de $49,41m^3$. Ou seja, a falha (inconsciente) da suposição de independência entre H e P_2 faz com que

o estimador seja muito mais viesado do que seria se a suposição de independência fosse atendida.

Para os intervalos de confiança construídos pelo processo das transformações (colunas 14 e 15), verifica-se (coluna 16) que 1755 dentre os 2000 casos cobrem o valor real. Ou seja, a cobertura do intervalo de confiança é de apenas 87,8%, enquanto o anunciado é de 95%. Note-se que com a inclusão da dependência entre H e P_2 os intervalos tem amplitude menor. Ilustrando-se com o mesmo caso utilizado anteriormente, os valores passam a ser $N = 496$, $H = 6,7$ e $P_2 = 0,74$, resultando em $T_2 = 143,2\text{m}^3$. O intervalo que é definido pelos limites $41,8\text{m}^3$ e $490,2\text{m}^3$ reduz em $22,1\text{m}^3$ a amplitude em comparação com o caso anterior.

5 – TESTANDO UMA HIPÓTESE

Nas seções anteriores examinou-se um aspecto da inferência: a estimação. No entanto, em muitos problemas de interesse prático há a necessidade de decidir entre duas possíveis ações em fase da evidência apresentada pelos dados. Este enfoque da inferência é conhecido por “teste de hipóteses” e será examinado nesta seção.

Para exemplificar, considere-se a dúvida inicial da empresa de celulose quanto a compra da área reflorestada. Supõe-se que após estudos mais detalhados estabelece-se como critério para compra um volume médio de madeira μ_T superior a $120\text{m}^3/\text{ha}$. Conseqüentemente, são as seguintes as hipóteses a considerar:

$$H_0: \mu_T \leq 120,0$$

$$H_1: \mu_T > 120,0 \text{ (condições para a compra)}$$

A hipótese H_0 é a hipótese de referência que será avaliada para $\mu_T = 120,0$, podendo ser rejeitada ou não, dependendo das evidências apresentadas pelos dados. Por exemplo, o caso descrito nas seções anteriores apresentou a estimativa $T = 150,2\text{m}^3$, que sugere boas condições para compra. Tecnicamente, a rejeição de H_0 e a conseqüente decisão favorável a compra somente se justifica se esta estimativa dificilmente ocorreria sob a hipótese de referência (Triola, 1999). Na prática, isto equivale a verificar se uma estimativa maior ou igual a $150,2\text{m}^3$ seria comum para o caso em que $\mu_T = 120\text{m}^3$.

Devido a informação incompleta disponível na forma do estimador T , a decisão está sujeita a dois possíveis erros. Comprar quando o volume médio revelar-se menor ou igual a 120m^3 é o erro “tipo I,” cuja probabilidade α denota o *nível de significância* deste teste. Alternativamente, deixar de comprar quando o volume revelar-se acima

de $120m^3$ é o erro “tipo II”, cuja probabilidade é simbolizada por β . O *poder de um teste*, definido como $1-\beta$, é a probabilidade de decidir corretamente a favor da compra.

Na execução de um teste de hipótese, o nível de significância é fixado antecipadamente, geralmente nos níveis de 5% ou 1%. No entanto, em muitas situações práticas, o erro “tipo II” é mais relevante. Por exemplo, para um comprador agressivo, a probabilidade β de perder um bom negócio é de muito interesse. Procedese o cálculo de β seguido da análise de como a falha da suposição de independência afeta esse valor.

Em função dos estudos simulados descritos em seções anteriores, partiu-se da distribuição amostral lognormal como adequada para a estatística T . Isto é, sob a hipótese H_0 , aceita-se a distribuição $\ln T \sim N(\ln(120); 0,63)$.

Com um *nível de significância* $\alpha = 5\%$ e um teste de hipótese unilateral, tem-se em termos de uma variável Normal padronizada

$$z = \frac{\ln T - \ln(120)}{0,628}, \quad (7)$$

a decisão de rejeitar H_0 para $z \geq 1,64$ (Triola, 1999). Isto é equivalente a estabelecer que todo valor de $\ln T$ pertencente ao conjunto $\{\ln T \geq 5,824\}$ seja considerado muito pouco provável de ser observado sob H_0 , justificando a rejeição desta hipótese.

Mas, se de um lado fixou-se a probabilidade α da ocorrência de um erro tipo I, há interesse em conhecer a probabilidade β de ocorrência de um erro tipo II, definido neste caso pela probabilidade condicional

$$\beta = P(\ln T < 5,824 \mid \mu_T = 198,5). \quad (8)$$

Para conhecer a probabilidade β de não rejeitar a hipótese H_0 , quando o valor real de μ_T é igual a 198,5, efetua-se um teste lógico identificando os valores de $\ln T$ inferiores a 5,824 (coluna 17). Entre os 2000 casos simulados, 1731 satisfazem essa condição. Portanto $\beta = 86,7\%$. Alternativamente, para $\ln T_2$ apenas 1569 casos (coluna 18) satisfazem a condição resultando em $\beta = 78,5\%$.

É muito interessante notar que, no caso de dependência nos dados, o valor efetivo de β é menor que aquele sugerido pelo modelo errado. Ou seja, o poder do teste é, de fato, maior do que ele se propõe ser.

6 – CONCLUSÕES

Ilustrou-se neste trabalho que problemas de inferência que extrapolam os casos clássicos envolvendo médias e desvios-padrão em distribuições normais podem ser tratados em cursos introdutórios de inferência recorrendo a simulações. Essas simulações se resumem na produção automatizada de réplicas. A análise dessas réplicas permite vivenciar concretamente os conceitos freqüentistas clássicos como distribuição amostral, erro-padrão, viés, intervalos de confiança e as probabilidades dos erros I e II.

Outro aspecto instrutivo do processo de simulação aqui proposto é impossível de experimentar de outra forma. Trata-se da análise dos efeitos causados à inferência quando suposições não são satisfeitas. Com a disseminação de pacotes estatísticos que permitam facilidade de acesso a técnicas estatísticas sofisticadas, aumenta também o risco para utilizá-las sem critério. Neste novo contexto, torna-se muito crítico o entendimento da importância de suposições dos modelos e o domínio de ferramentas que permitam experimentar com elas.

Espera-se que a simplicidade do processo de simulação aliada ao potencial que se abre para a análise de situações não-convencionais de inferência sirvam de encorajamento para estudantes e professores, para que lancem mão desses recursos para explorar novas possibilidades em sala de aula.

AGRADECIMENTOS

Os autores agradecem pelas sugestões recebidas de dois revisores anônimos.

REFERÊNCIAS

- HOEL, P. G. *Estatística matemática*. 4. ed. Guanabara Dois, 1981.
- StatSoft, Inc. *Statistica for Windows* [Computer program manual]. Tulsa, OK: StatSoft, Inc., 1998.
- TRIOLA, Mario F. *Introdução à estatística*. Livros Técnicos e Científicos, 1999.

ANEXO

Variável	Descrição	Sintaxe
H	Variável Lognormal com parâmetros $A = 2$ e $b = 0,3$	=VLognorm(rnd(1); a; b)
N	Variável Uniforme discreta entre $A = 300$ e $b = 500$	=(Trunc(rnd(b-a-1))+1)+(a-1)
P	Variável Normal $a = 0,9$ e $b = 0,22$	=VNormal(rnd(1); a; b)
T	Cálculo de T conforme equação (1)	=(N*(P^2)*H)/(4*pi)
N_LI95	Limite inferior do Intervalo de 95% para T	=T - 1,96*129,51
N_LS95	Limite superior do Intervalo de 95% para T	=T + 1,96*129,51
N_IC95	Operador lógico 0/ 1 avaliando se $\mu_T = 198,5$ está contido no IC	=((198,5>=N_LI95) and (198,5<=N_LS95))
LI_95	Limite inferior do IC conforme equação (5)	=exp(log(T) -1,231)
LS_95	Limite superior do IC conforme equação (5)	=exp(log(T) +1,231)
IC_95	Operador lógico 0/1 avaliando cobertura do IC	=((198,5>=LI95) and (198,5<=LS95))
DIF_AMPL	Diferença nas amplitudes dos Intervalos	=(n_ls95 - n_li95) - (ls95 - li95)
P2	Cálculo do perímetro conforme equação (6)	=VNormal(rnd(1); 0,9 + 0,2*((H-7,72)/2,28); 0,12)
T2	Recalculando T conforme equação (1) usando P2	=(N*(P2^2)*H)/(4*pi)
LLI95	Limite inferior do IC 95% usando T2 e equação (5)	=(T2/3,424)
LLS95	Limite superior do IC 95% usando T2 e equação (5)	=(T2*3,424)
IIC95	Operador lógico 0/1 avaliando cobertura do IC	=((198,5>=LLI95) and (198,5<=LLS95))
BETA	Operador lógico 0/1 para avaliar $\ln T < 5,824$	=(log(T) < 5,824)
BETA2	Operador lógico para avaliar $\ln T_2 < 5,824$	=(log(T2) < 5,824)