

USO DA REGRESSÃO LINEAR SIMPLES NA PREVISÃO DA PRODUÇÃO DE LEITE DO ESTADO DE SÃO PAULO (1)

José F. Noronha (2)

1 — INTRODUÇÃO

Em 1962 foi feita a primeira tentativa de previsão da produção de leite no Estado de São Paulo com "Aplicação de Amostragem e Equações de Regressão", de GOMES Jr. e PIVA (4). Usou-se naquela pesquisa uma série de dados de sete anos e foi estimada uma equação de regressão da produção do mês de janeiro, no qual era feito o levantamento da produção do Estado por meio de amostragem, sobre a média das produções mensais. Com a modificação da estrutura do questionário da previsão de safras, os autores do primeiro estudo bem como a Divisão de Levantamentos e Análises Estatísticas sugeriram uma reformulação daquele estudo, tendo como objetivo central identificar o melhor mês para o levantamento dos dados usando uma série maior de dados estatísticos

sobre a produção de leite do Estado.

Destarte, o presente estudo abrange o período de 1950 a 1967, num total de 18 anos, com dados fornecidos pelo Departamento de Produção Animal (PDA) da Secretaria da Agricultura de São Paulo.

Estes dados representam a parcela da produção, nos vários meses de cada ano, fiscalizada pelos técnicos do PDA e representam a maior série de dados mensais existentes sobre o assunto.

1.1 — OBJETIVOS

O objetivo principal deste estudo é identificar o melhor mês para o levantamento dos dados sobre a produção de leite no Estado de São Paulo e estimar nova

(1) O autor agradece aos Eng.ºs Agr.ºs Fernando S. Gomes Jr., Salomão Schattan, ao Economista Rural Harry W. Ayer, pelas sugestões apresentadas, bem como ao Dr. Pedro Treu, pela gentileza de ter facilitado a coleta de dados.

(2) Engenheiro agrônomo, M.S.

equação de regressão para fins de previsão da produção total anual.

Pretende-se, também: a) identificar os meses de produções baixas, médias e altas, durante o ano; b) formular um modelo simples que permita prever a produção mensal com antecedência de pelo menos um ano.

1.2 — METODOLOGIA

Com base nos dados de produção mensal fornecidos pelo PDA as análises se fundamentaram nos modelos Estatísticos de Análises de Variância e de Regressão Linear Simples.

Usou-se na Análise de Variância o Modelo Paramétrico de Blocos Casualizados, em que os vários Blocos são representados pelos Anos e os tratamentos pelos Meses dentro de cada ano, SCHEFFE (5).

Na análise de regressão estimou-se uma equação de regressão linear simples da produção de cada mês sobre a média das produções mensais. E, também, uma equação de tendência da produção dentro de cada mês, ao longo da série, fazendo-se 1950 = 1 ano, EZEKIEL e FOX (2).

1.3 — PRESSUPOSIÇÕES

Ao efetuar-se as computações, entretanto, tornou-se necessário fazer as seguintes pressuposições acerca dos dados utilizados e das estimativas de produção:

a) a produção do dia da en-

trevista representa a média diária do mês,

- b) a produção controlada pelo PDA representa, em toda a série analisada, aproximadamente um terço da produção do Estado,
- c) a proporção da produção total anual controlada pelo PDA é constante ao longo de cada ano em toda a série, isto é, em cada mês do ano as estatísticas do PDA representam uma terça parte da produção total daquele mês.

Dentro deste quadro de pressuposições é que as previsões poderão ser tomadas como de maior ou menor fidedignidade.

2 — RESULTADOS E DISCUSSÃO

2.1 — RAZÃO ENTRE AS ESTIMATIVAS DO SEP E DO PDA

Como primeiro passo da análise dos dados procurou-se relacionar a produção controlada pelo PDA com a estimativa do Serviço de Estatística da Produção (SEP) do Ministério da Agricultura, como teste da segunda pressuposição feita (quadro 1).

Verifica-se que a produção controlada, tem variado entre 32% a 41% da produção estimada pelo SEP, com a média de 36% para todo o período. Portanto, ao considerar-se como sendo um terço da produção total, muito provavelmente não se está cometendo um erro de estimativa por excesso.

QUADRO 1. — Razão entre a Produção de Leite Controlada pelo PDA-S.A. e a Estimativa do SEP-M.A., de 1950 a 1966

Ano	Razão: PDA/SEP %	Ano	Razão: PDA/SEP %
1950	39	1960	41
1951	36	1961	41
1952	35	1962	39
1953	33	1963	36
1954	32	1964	34
1955	32	1965	36
1956	33	1966	36
1957	34	—	—
1958	38	—	—
1959	39	—	—

2.2 — COMPARAÇÃO DAS MÉDIAS MENSAS PELA ANÁLISE DE VARIÂNCIA

Usou-se a análise de variância para testar a hipótese nula $H_0: \bar{P}_j = \bar{P}_f = \bar{P}_m = \dots = \bar{P}_d$, ou seja, que as médias de janeiro, fevereiro, etc., até dezembro, são iguais entre si. Formouse o seguinte quadro de análise, segundo modelo usado (quadro 2).

Os resultados obtidos, sendo ambos altamente significantes, levam à rejeição da hipótese nula ou seja, à aceitação de sua alternativa, que admite haver diferenças significativas entre as produções dos vários meses do ano. Dêste modo, fica evidenciada a estacionalidade da produção, tanto quanto uma evolução significativa da produção anual ao longo da série.

QUADRO 2. — Análise de Variância da Produção Mensal de Leite Controlada pelo PDA, 1950-67

Fonte de variação	Graus de liberdade	Soma dos quadrados	Quadrados Médios	«F»
Anos	17	18.178.723.193	1.069.336.658	161**
Meses	11	1.888.800.836	171.709.167	26**
Resíduo	187	1.243.587.742	6.650.202	—
Total	215	21.311.111.771	—	—

** Estatisticamente significante, ao nível de 1% de probabilidade.

Usou-se a seguir o teste de DUNCAN para identificar as diferenças mínimas significativas entre as médias mensais que fo-

ram classificadas por ordem decrescente para fins de análise, PIMENTEL GOMES (3) (quadro 3).

QUADRO 3. — Médias da Produção Controlada pelo PDA, em cada Mês, no Período de 1950-67

Mês	Média dos 18 anos
Dezembro	39.871
Janeiro	37.492
Novembro	37.095
Março	35.167
Outubro	34.978
	Média dos 12 meses
	33.932
Fevereiro	33.894
Abril	32.537
Setembro	32.063
Maiο	31.898
Agosto	31.717
Julho	31.365
Junho	29.106

Nota: Os traços unem as médias que não apresentam diferenças significativas, ao nível de 5% de probabilidade.

A produção mensal atinge seu valor máximo em dezembro e o mínimo em junho, que diferem estatisticamente dos demais meses. Não há diferença significativa entre as produções dos meses de novembro e janeiro. Estes três meses podem ser considerados como de produção alta em relação aos demais.

Encontram-se em uma faixa de produção aproximadamente igual à média mensal do ano os meses de fevereiro, março, abril e outubro. Em uma faixa que pode ser

considerada de baixa produção encontram-se os meses de maio, junho, julho, agosto e setembro, que sofrem os efeitos prolongados do período da seca.

Esta variação acentuada da produção durante o ano tem sido um desafio constante à modernização da tecnologia da produção do leite não só em São Paulo mas em todo o País. As flutuações estacionais têm causado sérios problemas tanto na fase de produção como na comercialização do produto.

2.3 — IDENTIFICAÇÃO DO MÊS REPRESENTATIVO DA PRODUÇÃO MÉDIA MENSAL DO ANO PELA ANÁLISE DE REGRESSÃO

Para identificar o mês que pu-

desse representar a produção média mensal do ano, procurou-se estimar uma equação de regressão entre a produção de cada mês e a média dos doze meses no período estudado, cujos resultados estão resumidos no quadro 4.

QUADRO 4. — Características das Equações de Regressão Linear Simples Entre a Produção Controlada de Cada Mês e a Média dos Meses, no Período Estudado

Mês	a	b	r ² (%)	Syx	t _b
Janeiro	1.793	0,8572	93	2.528	15
Fevereiro	718	0,9790	96	1.880	20
Março	723	0,9443	97	1.795	22
Abril	2.153	0,9767	96	2.048	19
Maiο	31	1,0628	95	2.134	18
Junho	2.361	1,0847	95	2.235	17
Julho	299	1,0723	96	1.896	20
Agosto	4.698	0,9217	92	2.788	14
Setembro	2.658	0,9754	94	2.370	16
Outubro	3.019	0,8838	93	2.544	15
Novembro	1.477	0,8749	97	1.692	23
Dezembro	- 178	0,8555	97	1.746	22

Verifica-se em tôdas as equações um elevado coeficiente de determinação (r^2), destacando-se os meses de março, novembro e dezembro, cujas produções explicam individualmente 97% da produção média do ano. Tôdas as equações apresentam o valor de t significativo ao nível de 1% de probabilidade, ao se testar a hipótese de $\beta = 0$.

Dêste modo, qualquer dos meses poderá ser usado para estimar a média mensal do ano com razoável grau de precisão, mas os meses de novembro, dezembro e

março são, em ordem de prioridade decrescente, os mais indicados por apresentarem os menores erros-padrão das estimativas e maiores coeficientes de determinação.

Resta, portanto, decidir sôbre qual dêstes três meses deveria ser escolhido para a realização da estimativa da produção anual. Do ponto de vista estatístico há, certamente, maior precisão na escolha do mês de novembro. Porém, considerando-se que a diferença entre os erros-padrão de novembro e março é relativamente pe-

quena (quadro 3) e, também, que uma estimativa realizada no primeiro trimestre do ano dá margem a uma tomada de decisão com efeitos dentro do mesmo ano, é possível argumentar a fa-

vor do levantamento dos dados em março ao invés de novembro.

As melhores alternativas são, portanto, as seguintes:

Mês da previsão	Equação estimativa
Março	$Y/3 = 723 + 0,9443 X/3$
Novembro	$Y/3 = 1.477 + 0,8749 X/3$
Dezembro	$Y/3 = -178 + 0,8555 X/3$

onde X representa a produção estimada para determinado mês (em 1.000 litros) e Y representa a produção média mensal do ano (em 1.000 litros).

Estimada a produção do mês, através do levantamento por amostragem, calcula-se pela equação estimativa a média mensal do ano que, multiplicada por 12, resulta no valôr da produção total anual.

2.4 — ESTIMATIVA PARA 1968

Não tendo sido feito o levantamento dos dados em março de 1968, procurou-se fazer uma estimativa com base na amostragem de novembro dêste mesmo ano. Assim, a produção total anual foi calculada como se segue:

A produção do dia da entrevista, estimada através da amostragem feita pela Seção de Previsão de Safras foi de 3.912.273 litros, com um êrro de amostragem de 6,21%. Considerando-se esta estimativa como a média dos trinta dias dêste mês chega-se a um total de 117.368.190 litros para o mês em consideração. Substituiu-

do-se êste valor na equação estimativa e multiplicando-se por 12 o resultado obtido, encontra-se 1.285 milhões de litros para o ano de 1968, em todo o Estado de São Paulo.

Esta estimativa, entretanto, deve ser tomada com reserva, uma vez que o ano de 1968 sofreu as conseqüências de uma das grandes sêcas dos últimos tempos, o que pode ter provocado uma redução irregular na produção de novembro.

Portanto, sômente a repetição dêste procedimento a cada nôvo ano, introduzindo os valôres observados na série, corrigindo, assim, os parâmetros da equação, poderá consolidar a fidedignidade das estimativas obtidas.

2.5 — PREVISÃO DA PRODUÇÃO MENSAL

Com o objetivo de prever a produção mensal do ano seguinte, procurou-se ajustar uma equação de tendência da produção para cada mês, com base na série estuda, considerando-se 1950 = 1 ano (quadro 5).

QUADRO 5. — Características das Equações de Tendência da Produção Mensal Controlada pelo PDA, para 1950 = 1 ano

Mês	Valôres de				
	a	b	r ² (%)	S _{yx}	t _b (¹)
Janeiro	19.705	1.872,33	87	3.910	10,5
Fevereiro	18.315	1.639,91	85	3.797	9,5
Março	19.437	1.655,81	80	4.535	8,0
Abril	17.242	1.610,00	82	4.161	8,5
Maió	18.069	1.455,69	80	4.053	7,9
Junho	15.297	1.453,52	83	3.608	8,9
Julho	16.784	1.534,84	89	2.943	11,5
Agôsto	15.150	1.743,83	89	3.358	11,4
Setembro	15.962	1.694,83	92	2.761	13,5
Outubro	17.560	1.833,37	89	3.506	11,5
Novembro	19.077	1.896,62	90	3.494	12,0
Dezembro	21.910	1.890,57	86	4.267	9,8

(¹) Todos os valôres de t são significantes ao nível de 1% de probabilidade.

Vale lembrar que as equações de tendência foram ajustadas a partir dos dados do PDA, razão pela qual se supõe que representam, aproximadamente, um têrço da produção do Estado em cada mês.

Foram feitas estimativas para todos os meses de 1968 e 1969 com as equações de tendência. Por outro lado, já estão disponíveis os dados do PDA referentes

ao ano de 1968, até o mês de outubro, o que permite a comparação que se segue (quadro 6).

Dêste modo, pode-se obter uma estimativa aproximada da produção mensal do ano seguinte através da extrapolação de um ano nas equações de tendência (³), o que poderá servir de base para tomada de decisão com referência à política de regularização da produção durante o ano.

(³) É relevante lembrar, contudo, que êste procedimento está sujeito a sérias restrições de ordem metodológica. Sabe-se que a tendência é apenas uma das componentes de uma série temporal. Variações cíclicas, sazonais e aleatórias são outras componentes que devem ser levadas em consideração. Por outro lado, a matemática e a estatística não oferecem fundamentos seguros para que se identifique a verdadeira tendência de séries temporais sujeitas aos processos econômicos. Apenas a observação cuidadosa da evolução de fenômeno, durante longo tempo, associada ao bom-senso do pesquisador, torna válido êste tipo de análise. Ministério da Agricultura (³).

QUADRO 6. — Comparação entre a Produção Mensal de Leite Fiscalizado pelo PDA e a Estimada com as Equações de Tendência. São Paulo, 1968 e 1969

Mês	Produção Observada (1.000 litros) 1968	Produção Estimada (1.000 litros)	
		1968	1969
Janeiro	50.656	55.279	57.152
Fevereiro	45.640	49.473	51.113
Março	49.862	50.897	52.553
Abril	47.427	47.832	49.442
Maiο	48.685	45.727	47.183
Junho	47.321	42.914	44.367
Julho	48.168	45.946	47.481
Agosto	52.599	48.283	50.027
Setembro	51.968	48.164	49.859
Outubro	52.960	52.394	54.227
Novembro	—	55.113	57.009
Dezembro	—	57.831	59.721

3 — SUMÁRIO E CONCLUSÕES

O presente estudo é uma segunda tentativa feita para estabelecer uma metodologia de previsão da produção de leite no Estado de São Paulo através dos Levantamentos por Amostragem e Equações de Regressão. A primeira foi realizada em 1962 e se baseou no levantamento da produção diária, por amostragem, realizada no mês de janeiro.

Esta, por outro lado, teve por objetivo central identificar o mês em que deveria ser feito o levantamento da produção, por amostragem, a fim de que se pudesse fazer uma estimativa que, atendendo aos interesses da política agrícola, tivesse a menor margem

possível de êrro. Como objetivo secundário, procurou-se estudar a tendência da produção dentro de cada mês, ao longo da série, como meio de prever a produção mensal.

Usou-se as análises de variância e de regressão linear simples, com dados de produção mensal, de 1950 a 1967, fornecidos pelo Departamento da Produção Animal da Secretaria da Agricultura de São Paulo.

As análises dos dados disponíveis permitiram concluir que: 1) os meses de março, novembro e dezembro são os mais indicados para representar a produção média mensal do ano. Dentre êstes o mês de março parece ser o que

melhor atende aos objetivos da estimativa anual, com vistas a uma política de incentivo à produção, de efeitos a curto prazo; 2) a produção mensal do ano seguinte poderá ser estimada com pequena margem de erro, através das equações de tendência, apesar das inúmeras limitações impostas a esta técnica de estimação; 3) os me-

ses podem ser classificados, segundo o nível de produção, em:

- a) meses de produção alta — novembro, dezembro e janeiro;
- b) meses de produção próxima da média mensal do ano — fevereiro, março, abril e outubro;
- c) meses de baixa produção — maio, junho, julho, agosto e setembro.

LITERATURA CITADA

1. BRASIL. MINISTÉRIO da AGRICULTURA. SERVIÇO de PREVISÃO de SAFRAS. — Safras; extrapolações da área e produção dos principais produtos agrícolas por unidade da Federação até o ano de 1970... Brasília, 1967. (Publicação n.º 23).
2. EZEKIEL, M. & FOX, K. A. — Methods of correlation and regression analysis. New York, John Wiley, 1961. 548p.
3. GOMES, F. P. — Curso de estatística experimental. 2.a ed. Piracicaba, Escola Superior de Agricultura «Luiz de Queiroz», 1963. 384p.
4. GOMES JÚNIOR, F. S. & PIVA, L. H. O. — Previsão da produção de leite no Estado de São Paulo com aplicação do método de amostragem e equações de regressão. Agricultura em São Paulo. 9 (7): 27-31. julho. 1962.
5. SCHEFFE, Henry — The analysis of variance. New York, John Wiley, 1963. 477p.

USO DA REGRESSÃO LINEAR SIMPLES NA PREVISÃO DA PRODUÇÃO DE LEITE DO ESTADO DE SÃO PAULO

A N E X O S

ANEXO 1. — Médias das Produções Mensais Controladas pelo PDA. Valores Observados e Estimados e Limites Fiduciais das Estimativas, de 1950 a 1967

Ano	Valores observados (1.000 litros)	Valores estimados ⁽¹⁾ (1.000 litros)	
		Média	Limites fiduciais ⁽²⁾
1950	17.699	18.615	14.109 — 21.289
1951	16.903	17.339	13.313 — 20.493
1952	22.180	21.447	18.590 — 25.770
1953	24.241	24.668	20.651 — 27.831
1954	26.363	26.223	22.773 — 29.953
1955	27.942	26.882	24.352 — 31.532
1956	29.735	30.799	26.145 — 33.325
1957	32.424	33.020	28.834 — 36.014
1958	37.175	37.386	33.585 — 40.765
1959	38.932	40.853	35.342 — 42.522
1960	41.377	42.904	37.787 — 44.967
1961	42.142	43.893	38.552 — 45.732
1962	42.130	44.862	38.540 — 45.720
1963	37.836	38.086	34.246 — 41.426
1964	40.165	38.967	36.575 — 43.755
1965	43.099	40.896	39.509 — 46.689
1966	42.718	38.270	39.128 — 46.308
1967	47.709	45.560	44.119 — 51.299

(¹) Equação estimativa: $Y = 2.169 + 0,9443 X$ referente ao mês de março, 1950-67.

(²) Limites fiduciais ao nível de 95% de probabilidade.

ANEXO 2. — Movimento Mensal da Produção de Leite Fiscalizada pelo Departamento da Produção Animal (PDA) da Secretaria da Agricultura do Estado de São Paulo, 1950-1967, em 1.000 litros

Ano	Mês											
	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Agô.	Set.	Out.	Nov.	Dez.
1950	21.940	18.031	18.947	17.248	16.814	15.228	16.187	16.242	14.950	16.805	19.304	20.692
1951	19.270	17.368	17.596	15.984	15.635	13.690	14.246	14.145	15.896	17.330	19.694	21.985
1952	22.636	21.576	21.946	19.885	19.554	17.675	22.358	21.713	21.781	23.476	25.106	28.462
1953	26.357	24.076	25.357	22.797	24.138	20.976	22.623	22.480	23.429	25.691	25.451	27.522
1954	28.074	24.256	27.004	22.566	25.487	24.029	26.982	26.672	26.669	26.995	27.743	29.879
1955	30.355	27.446	27.702	27.702	28.209	24.444	26.592	26.301	25.598	28.443	28.849	32.668
1956	33.139	30.940	31.850	29.775	29.506	25.750	28.459	22.981	27.998	31.229	31.568	33.627
1957	33.919	32.203	34.202	30.630	31.036	28.397	30.108	31.707	31.235	34.529	34.821	36.306
1958	36.953	36.131	38.826	34.983	35.612	32.738	33.441	35.601	34.850	40.169	41.292	45.510
1959	43.624	42.672	42.497	40.187	39.051	34.694	35.887	33.674	34.665	36.495	38.667	45.069
1960	44.115	42.372	44.669	42.545	38.949	33.845	36.487	36.812	35.857	41.543	48.871	50.462
1961	48.865	41.462	45.716	40.536	40.712	38.342	38.105	38.220	37.676	38.000	45.905	52.163
1962	50.930	43.485	46.743	42.393	39.355	33.945	35.749	35.444	37.628	45.193	46.347	48.352
1963	47.652	41.259	39.567	39.447	36.692	33.301	35.147	31.172	31.463	33.312	40.649	44.365
1964	44.030	40.884	40.500	37.275	32.603	28.746	33.728	40.724	40.556	44.509	47.540	50.872
1965	49.467	40.997	42.543	37.547	36.478	37.177	41.170	44.290	43.814	47.063	47.051	49.590
1966	46.532	40.150	39.762	38.028	38.972	36.643	41.063	43.214	43.705	48.304	48.556	47.690
1967	47.001	44.797	47.588	46.133	45.358	44.284	46.240	49.493	48.356	50.509	50.291	52.457
1968	50.656	45.640	49.862	47.427	48.685	47.321	48.168	52.599	51.968	52.960	—	—

Fonte: Departamento da Produção Animal — S.A.