

OS GRÁFICOS — COMO SE FAZEM E PARA QUE SERVEM

Frederico Pimentel Gomes

Assistente da Cadeira de Matemática
E. S. A. «Luiz de Queiroz»

1 — INTRODUÇÃO — Os agrônomos, como tantos outros técnicos, recorrem frequentemente aos gráficos para representar os resultados de suas observações e experiências, e se vêem obrigados a consultá-los com frequência. Parece-nos, porém, que o papel dos gráficos é geralmente mal compreendido e a sua confecção é muitas vezes descuidada.

Este artigo tem por fim esclarecer o leitor em palavras simples e acessíveis sobre os conhecimentos básicos necessários a quem lida com os gráficos.

2 — O PAPEL DOS GRÁFICOS EM ESTATÍSTICA — Ao contrário do que geralmente se pensa, os gráficos não provam nada. Assim diz o estatístico FISHER (1) e o mesmo afirma o matemático WOODS (2) e suas opiniões não são contestadas por ninguém que se dedique ao assunto.

O papel dos gráficos não é, portanto, **provar** as nossas conclusões, mas apenas **sugerir** os testes estatísticos a serem aplicados e os pontos a serem analisados, e **apresentar** as conclusões indicadas pela Estatística Analítica.

3 — COMO SE FAZEM OS GRÁFICOS — Em geral utiliza-se na confecção dos gráficos o sistema cartesiano ortogonal da Geometria Analítica Plana. Nas abcissas toma-se uma das variáveis, por exemplo: os anos agrícolas, e nas ordenadas a outra, por exemplo: a produção brasileira de trigo. Os pontos obtidos são ligados por meio de segmentos de retas e obtemos uma linha quebrada (fig. 4).

As grandezas representadas são, em geral, inteiramente diversas, como no exemplo anterior. Por isso a escala vertical é, muitas vezes, inteiramente independente da horizontal. Como cada escala dá um aspecto diferente ao gráfico, o leitor compreenderá facilmente que não é difícil obter gráficos inteiramente diversos feitos com os mesmos dados. É o que acontece, por exemplo, com os gráficos das figuras 1 e 2.

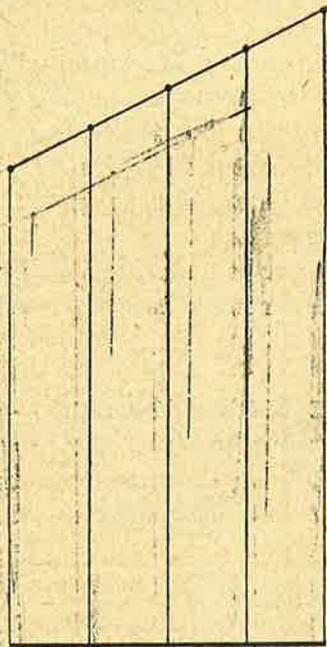


FIG. 1



FIG. 2

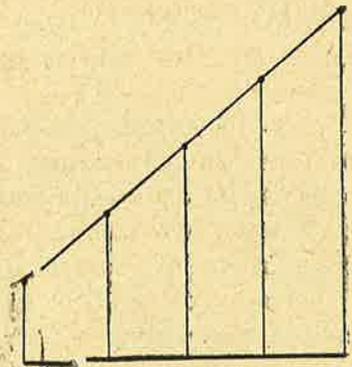


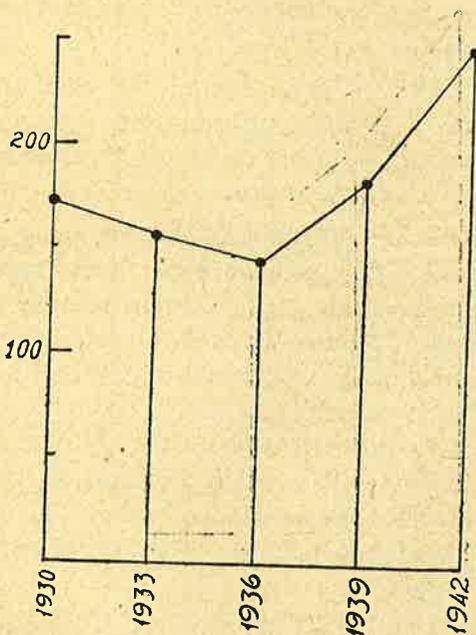
FIG. 3

O mais racional, quando possível, é tomar como unidade, tanto na escala vertical como na horizontal, o erro standard da grandeza aí representada. Se, por exemplo, no eixo das abscissas temos a produção de leite diária de cada vaca de um rebanho com um erro standard de ± 3 litros por dia, e nas ordenadas a porcentagem de gordura desse leite com um erro de

$\pm 0,5\%$ de matéria gorda, podemos tomar nas abcissas 1 cm para cada 3 litros de leite e nas ordenadas 1 cm para 0,5% de gordura.

Não é raro acontecer que as variações em estudo sejam muito pequenas em relação ao valor dos dados analisados. Por exemplo, em um trabalho que publicámos em 1943 (3) a uma média geral de 3,756% de matéria graxa no leite das vacas holandêsas da Escola Superior de Agricultura “Luiz de Queircz” correspondiam 12 médias parciais, das quais 7 se afastavam

Fig. 4 — Produção brasileira de trigo em milhares de toneladas



menos de 0,1% dessa média geral. Para tornar bem visível essa variação no gráfico deveríamos representar 0,1% de gordura por 2 cm. Mas assim sendo o gráfico todo teria 80 cm. de comprimento, o que é demais para uma página da “Revista de Agricultura”. Por outro lado, se reduzíssemos a escala, desapareceriam todos os detalhes em estudo.

Que fazer?

Para um caso como esse o recurso é manter a escala conveniente e suprimir a parte do gráfico que não interessa. Foi o que fizemos no gráfico que publicamos nessa ocasião, e que reproduzimos agora (fig. 5). Nêle só aparecem as ordenadas que vão de 3,55% a 4,00% de gordura. E assim os 80 cm. originais ficaram reduzidos a apenas 9, o que é mais razoável.

Surge, porém, uma nova dificuldade: a supressão da parte inferior do gráfico exagera e dá uma idéia falsa da variação das médias, que parece excessiva. É o que se vê, por exemplo, no gráfico da fig. 3, feito com os mesmos dados que foram usados para as figuras 1 e 2, mas com supressão de sua parte inferior.

Isto se evita facilmente chamando a atenção do leitor para o artifício utilizado, ou representando no gráfico, como fizemos, os limites de 5% e 1% de probabilidades indicados pela Estatística Analítica.

Os outros tipos de gráficos, como por exemplo esses que representam a população por uma figura humana, e o valor da exportação de cacau por um saco de dinheiro, são mais de interesse artístico do que científico. Na realidade êles trazem complicações inúteis que dificultam a sua interpretação exata.

4 — OS GRÁFICOS E A CORRELAÇÃO — A noção de correlação em Estatística corresponde à de função em Matemática. Existe correlação entre duas ou mais variáveis quando há uma correspondência entre os valores assumidos por essas variáveis.

Tomemos o caso de duas variáveis apenas.

A dependência entre essas variáveis pode ser expressa por uma função $y = f(x)$ que representa analítica e aproximadamente a correspondência existente entre elas. Essa função pode ser, por exemplo, do tipo $y = ax + b$ (equação de uma linha reta), que é o caso mais simples e mais comum. Temos então uma **correlação linear**. Em geral, quando se fala simplesmente em correlação subentende-se a correlação linear.

Um curioso exemplo de correlação linear é o que nos dá o

livro "Food and Life" (4). Lá encontrámos na pag. 465 e seguinte a igualdade:

$$E = 0,48774 - 0,008456W,$$

em que E é a eficiência dos animais lá analisados na utilização dos alimentos para o crescimento, e W é o peso vivo desses animais.

A correlação linear entre duas variáveis é medida pelo coeficiente de correlação r . Mas o cálculo de r é bastante trabalhoso. Por isso não é raro, antes de calculá-lo, fazer um gráfi-

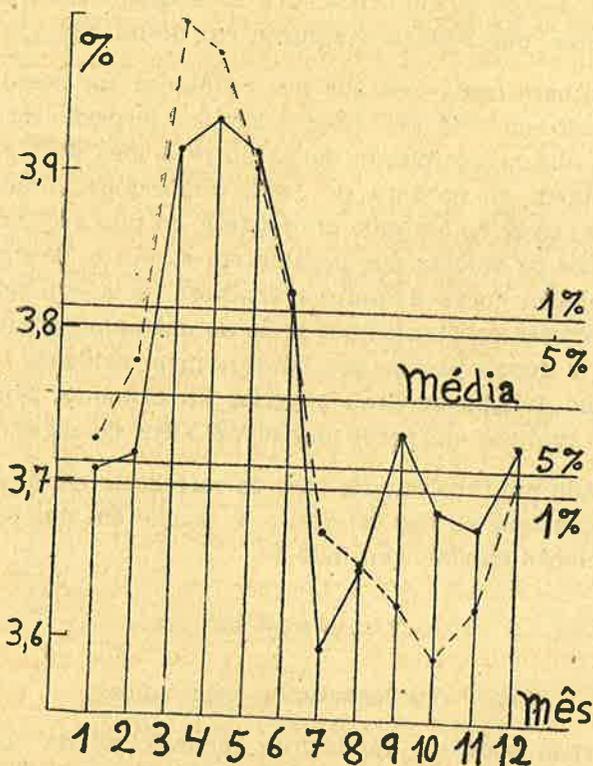


Fig. 5 — Nas abscissas estão os meses do ano e nas ordenadas correspondentes à linha cheia a porcentagem da gordura do leite

co com os dados em estudo, pois o gráfico logo dá uma idéia da existência ou inexistência de correlação e ainda do modo de procurá-la. No gráfico cada par de valores (um de cada variável) é representado por um ponto. O conjunto de pontos deve ter um aspecto mais ou menos alongado para que se espere uma correlação significativa. Quando o conjunto de pontos é aproximadamente circular não deve haver correlação. Mas só o cálculo de r pode dizer a última palavra.

As vezes a simples inspeção do gráfico mostra a possibilidade de existência de uma **correlação curva**, e não linear, isto é, mostra que a equação da função $f(x)$ deve ser de grau superior ao primeiro. Convém então investigar isso, pois nesse, como em todos os outros casos, é a Estatística Analítica, com os seus testes, que resolve o assunto em definitivo.

Foi um caso como esse que nos aconteceu em nosso trabalho publicado em 1944 (5), onde a simples inspeção do gráfico (fig. 6) já sugere correlação curva entre o mês de lactação e a porcentagem de gordura do leite. Supuzemos então que a função $f(x)$ fôsse do segundo grau, isto é, do tipo $a + bx + cx^2$, e calculámos os valores dos parâmetros a , b e c . A análise de uma correlação curva é muito trabalhosa, pois é preciso calcular numerosas constantes por meio de um número enorme de operações, e depois provar que há vantagem evidente em considerar uma função de grau superior ao primeiro. Por isso e por outros motivos em geral não se vai além do segundo grau.

Há ainda um interessante caso de correlação em que os gráficos nos prestam serviço relevante. É aquele em que se espera uma correlação exponencial, isto é:

$$y = e a + bx$$

sendo e a base dos logaritmos neperianos.

Este caso pode ser facilmente reduzido ao da correlação linear pela aplicação de logaritmos. Temos então:

$$\log y = a + bx.$$

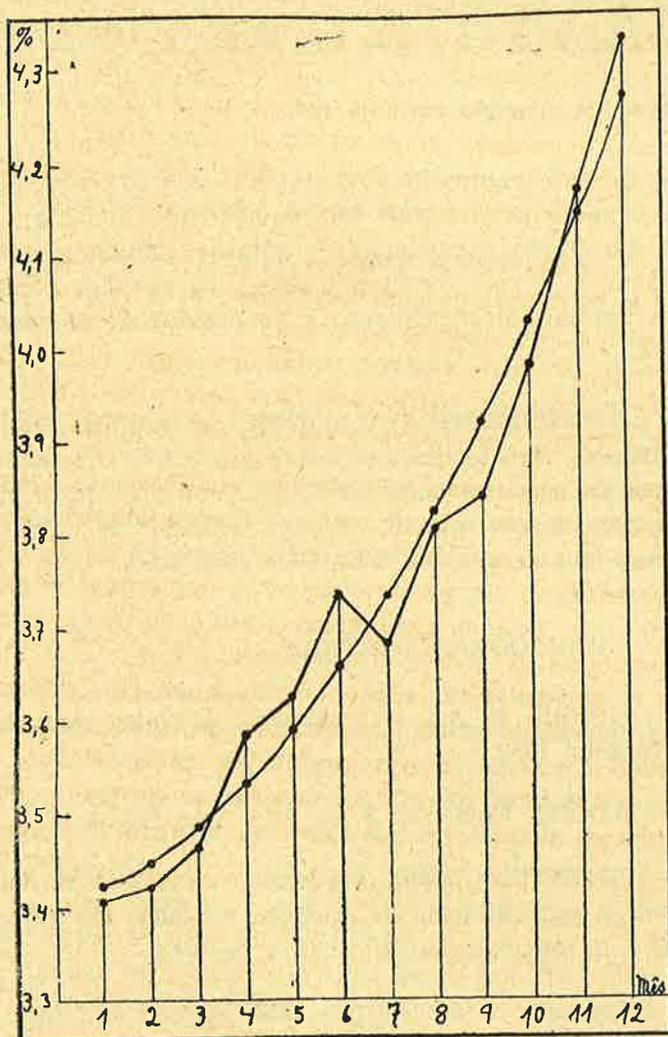


Fig. 6 — A linha poligonal representa as médias mensais observadas em cada um dos meses de lactação. A curva é a representação gráfica da função interpoladora —

$$y = 3,405 + 0,01241 x + 0,00498 x^2$$

Se fizermos $\log y = z$, fica finalmente :

$$z = a + bx ,$$

que é a equação de uma reta.

Neste caso representamos nas abcissas os valores de x e nas ordenadas os de z , que são os logaritmos neperianos de y .

A correlação exponencial é bastante comum em Biologia. Por exemplo, é uma correlação desse tipo que liga o peso vivo de um animal em crescimento à quantidade de alimento consumido.

5 — CONCLUSÃO — A confecção de gráficos requer certos cuidados, sem os quais eles não exprimem a verdade. Gráficos feitos sem esses cuidados devem ser encarados com reserva, principalmente quando desacompanhados de outras informações e dos resultados da análise estatística.

6 — BIBLIOGRAFIA CITADA.

- (1) — FISHER, R. A. — Statistical Methods for Research Workers. 1932.
- 2) — WOODS, Frederick S. — Advanced Calculus. S/ data.
- (3) — PIMENTEL GOMES, Frederico — Como varia durante o ano a porcentagem de gordura do leite em Piracicaba. Rev. de Agricultura, Set. e Out. de 1943.
- (4) — Yearbook of Agriculture 1939 — Food and Life.
- (5) — PIMENTEL GOMES, Frederico — Como varia durante o período de lactação a porcentagem de gordura do leite em Piracicaba. Rev. de Agricultura. Set., Out., Nov. e Dez. de 1944.