

Estatística e calculadoras gráficas

Ana Paula Canavarro

*Universidade de Évora
Projecto DIF - CIEFCUL
Projecto T³ da APM*

Esta comunicação incide em três tópicos: 1. A abordagem da Estatística na escola; 2. Contributos das calculadoras gráficas ; 3. Propostas de investigações sobre os conceitos estatísticos.

1. A abordagem da Estatística na escola

A Estatística está presente no dia-a-dia da nossa sociedade e são diversos os seus usos:

- Veicular informação nos diversos órgãos de comunicação (jornais, revistas, televisão, rádio, internet, ...)
- Sondar a opinião pública (preferências, hábitos, situações concretas,...)
- Fundamentar previsões (evolução da população, do mercado, de recordes, ...)
- Justificar tomadas de decisão (construção de escolas, incentivo ao consumo, retirada de brinquedos do mercado,...)

Para que um aluno possa ser matematicamente literado, é indispensável que ele consiga lidar com a Estatística de forma esclarecida. A este propósito, numa publicação do Ministério da Educação, pode ler-se: “*Uma das finalidades da escola é preparar os alunos para as necessidades do mundo real onde vivemos, necessidades e problemas esses que todos os dias aparecem nos meios de comunicação social, televisão, rádio e jornais. Alfabetizar estatisticamente os alunos de modo a perceberem as notícias que ouvem e lêem, é desenvolver-lhes o sentido crítico, a capacidade de argumentar sobre elas e inclusivamente serem capazes de intervir e tomar decisões.*” (Martins, M^a Eugénia; Monteiro, C., Viana, J. P. e Turkman, *Estatística, 10º ano de escolaridade*, ME - DES, 1997, p.8).

A alfabetização estatística dos alunos deve incidir sobre um conjunto de aspectos que de seguida refiro e que constituem aquilo que penso poder chamar-*se literacia estatística*:

- Dominar a linguagem estatística — não só ao nível dos conceitos associados às medidas como também aos gráficos;
- Compreender os procedimentos estatísticos — desde a escolha da amostra até à divulgação de resultados;
- Interpretar as situações e os resultados — o que dizem os resultados sobre a situação em questão;
- Avaliar a validade das conclusões — análise crítica sobre a razoabilidade das conclusões tendo em conta os contextos;
- Ponderar eventuais distorções da realidade — considerar possibilidades de enviesamentos ou manipulações tendenciosas.

2. Contributos das calculadoras gráficas

A partir da introdução em tabela dos valores das variáveis, as calculadoras gráficas, agora disponíveis a professores e alunos, oferecem, rapidamente e com rigor:

- Medidas de tendência central, de localização e de dispersão;
- Representações gráficas diversas;
- Coeficientes de correlação linear e outras;
- Regressões diversas.

Estas possibilidades fazem com que os alunos, em qualquer estudo estatístico, possam concentrar a sua atenção nos aspectos mais elaborados do trabalho, como o interpretar, organizar, discutir, argumentar, e não sobre os aspectos mais mecânicos associados à sua realização.

Além disso, permitem ainda que os alunos possam lidar com os conceitos estatísticos de uma forma mais rica, explorando o seu significado, percebendo o que cada um representa, a que corresponde, o que esconde e como se manipula. Só um tal domínio dos conceitos estatísticos permite aos alunos uma utilização crítica da Estatística. Para desenvolver este domínio, o aluno poderá ele mesmo investigar sobre os diversos conceitos, desenvolvendo assim uma compreensão esclarecida. Estas investigações podem ser realizadas em quaisquer aulas de Matemática graças à existência das calculadoras gráficas.

3. Propostas de investigações sobre conceitos estatísticos

De seguida apresentam-se exemplos de actividades de investigação que incidem sobre três temas particularmente importantes. Muitas outras poderiam ser propostas, e incidindo sobre outros tópicos estatísticos dos programas.

Organização e Representação

Este tema é tratado tanto no 3º Ciclo - 7º e 8º anos - como no 10º ano. É um tema particularmente importante porque a linguagem gráfica tem um grande impacto, é lida com muita rapidez e facilidade, e por isso é muito usada pelos órgãos de comunicação. Os dados podem ser organizados e representados de muitas formas, umas mais adequadas do que outras, dependendo muito do objectivo do estudo. A forma como se organizam os dados influencia os resultados que se mostram.

A organização e representação dos dados pode ser manipulada de modo a destacar aspectos pretendidos. Pode mesmo dizer-se que a partir dos mesmo dados e com representações diferentes se podem mostrar várias verdades.

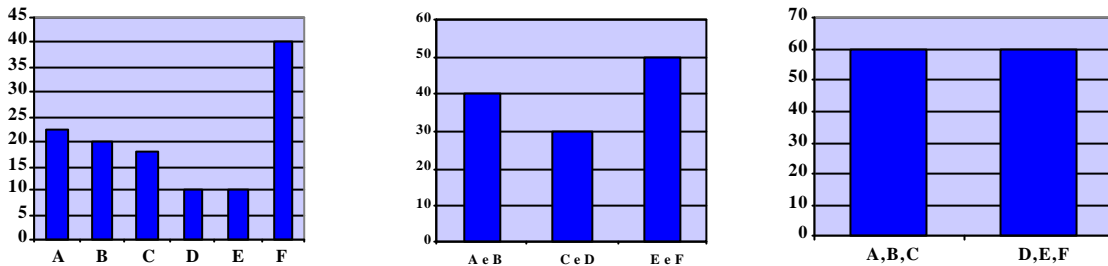
Uma tarefa de investigação que poderá ter interesse para os alunos consiste em analisar as diferenças que se observam quando se constroem gráficos de barras relativos a uma mesma situação mas com barras de amplitudes diferentes.

-
1. Representa os dados através de gráficos de barras

A	B	C	D	E	F
22	18	20	10	10	40

- De modo que se vejam as seis barras.
- Associando A e B, C e D, e E e F.
- Associando as três primeiras barras e as três últimas.
- Compara os três gráficos e discute as consequências da variação das amplitudes das barras.

A partir da comparação destes gráficos pode observar-se como o agrupamento dos dados em classes de amplitudes diferentes sugere informações distintas sobre a mesma situação.



Uma outra tarefa de investigação importante relativamente a representações consiste em analisar as diferenças que se observam quando se visualiza o mesmo gráfico num sistema de eixos e noutro com o eixo dos yy truncado.

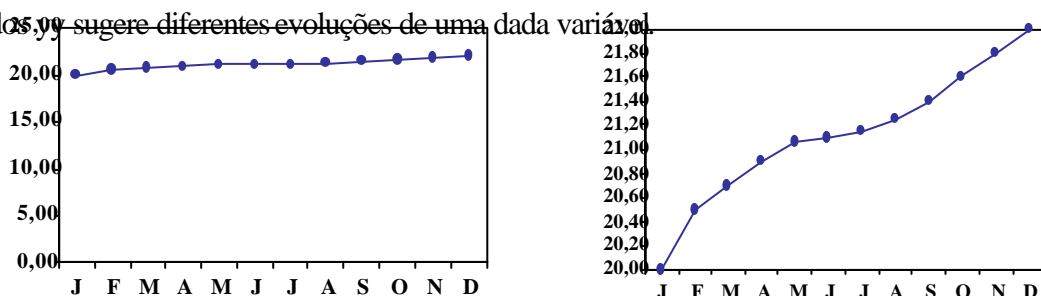
1. Representa os dados usando um polígono.

J	F	M	A	M	J	J	A	S	O	N	D
20	20,5	20,7	20,9	21,1	21,1	21,2	21,2	21,4	21,6	21,8	22

- Observa o gráfico que a calculadora te fornece automaticamente. Como caracteriza o comportamento desta variável?
- Visualiza o gráfico obrigando a que o mínimo do eixo dos yy seja 0 e que o seu máximo seja 25. O que te sugere agora o gráfico?
- Compara os dois gráficos e discute as consequências da truncagem do eixo dos yy.

A comparação dos gráficos possibilita perceber como a simples truncagem do eixo

dos yy sugere diferentes evoluções de uma dada variável.



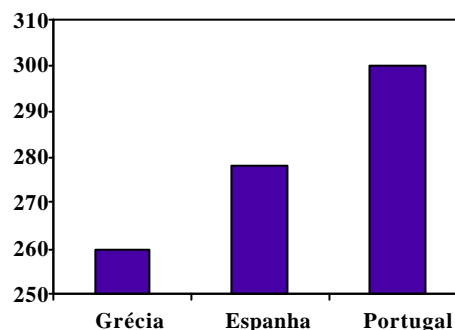
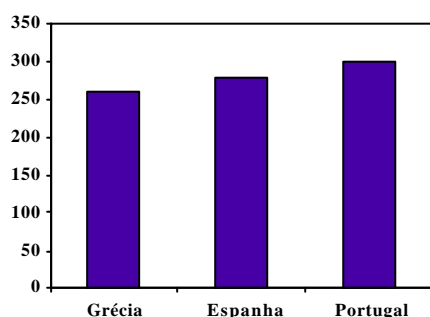
Ainda sobre o tema representações, é bastante interessante analisar as diferenças que se observam quando se visualiza o mesmo gráfico em sistemas de eixos com diferentes valores mínimos para o eixo dos yy.

1. Representa os seguintes dados usando gráficos de barras:

A	B	C
260	278	300

- Observa o gráfico que a calculadora te fornece automaticamente. O que sugere?
- Visualiza agora o gráfico obrigando a que o mínimo do eixo dos yy seja 250. O que te sugere agora?
- Compara os dois gráficos e discute as consequências da variação do valor mínimo do eixo dos yy.

Como os gráficos permitem observar, a subtração de uma faixa horizontal sugere comparações enganosas entre as diversas barras. Enquanto que o primeiro gráfico mostra situações pouco distintas, o segundo gráfico mostra grandes diferenças.



Medidas

Embora os programas prevejam a abordagem de diversas medidas, aqui refiro apenas duas delas, a média e o desvio-padrão, pela grande importância que têm. A média é a medida estatística mais utilizada e conhecida do grande público. É uma medida que, pretendendo representar todos, muitas vezes não representa nenhum. É um bom indicador para amostras bastante homogêneas e “normais”, mas pode ser muito enganosa para amostras “reais”. Além disso, pode ser usada e manipulada com o objectivo de legitimar conclusões incorrectas.

O desvio padrão é uma medida de dispersão muito informativa mas é raramente usada na veiculação de informação estatística nos órgãos de comunicação. É um excelente complemento da média para conhecer uma amostra porque informa acerca das diferenças entre os diversos dados e a sua média, ou seja, acerca da sua dispersão, e por isso pode ser muito útil na comparação de amostras com médias próximas.

Para um bom conhecimento do significado destas medidas, os alunos podem realizar diversas actividades. Uma delas será analisar as características de amostras distintas com a mesma média mas diferentes desvios padrão.

1. Determina a média e o desvio padrão das seguintes amostras:

- a) 5 5 5 5 5 6 6 6 6 6
- b) 1 2 3 4 5 6 7 8 9 10
- c) 1 1 5.5 5.5 5.5 5.5 5.5 5.5 10 10

2. Compara-os e tenta justificar os valores obtidos.

3. Consegues arranjar uma amostra cuja média seja igual às anteriores?

Com esta actividade, os alunos podem observar que amostras muito diferentes podem ter exactamente a mesma média. O desvio padrão ajuda a exibir essas diferenças:

Para 5 5 5 5 5 6 6 6 6 6, $M = 5.5$ e $\sigma = 0.5$

Para 1 2 3 4 5 6 7 8 9 10, $M = 5.5$ e $\sigma = 2.87$

Para 1 1 5.5 5.5 5.5 5.5 5.5 5.5 10 10 $M = 5.5$ e $\sigma = 3,67$

Outra actividade com interesse consiste em analisar o que acontece à média e ao desvio padrão de determinada amostra quando se produzem alterações nos dados.

1. Escolhe uma amostra de números ao teu gosto.

a) Determina a sua média e o seu desvio padrão

b) O que acontece a estes valores se acrescenta um novo dado igual à média?

- c) E se acrescentar um valor dez vezes superior ao maior da amostra escolhida?
 d) E se o valor acrescentado for dez vezes inferior ao menor da amostra escolhida?

Com esta actividade é possível compreender que a média pode ser facilmente alterada por pequenas alterações na amostra, sobretudo se for por um dado que se afasta muito do valor da média.

2	3	5	7	11	13	Média: 6,8	Desvio padrão: 4	
2	3	5	6,8	7	11	13	Média: 6,8	Desvio padrão: 3,7
2	3	5	7	11	13	130	Média: 24,4	Desvio padrão: 43,2
0,2	2	3	5	7	11	13	Média: 5,8	Desvio padrão: 4,3

Outra hipótese de trabalho consiste em analisar amostras onde existem *outliers* e comparar o que acontece às medidas e representações gráficas quando estes são considerados e quando não são.

1. Considera a seguinte amostra:

2 3 5 7 11 13 100

- a) Determina a média e o desvio padrão.
 b) Representa por diagrama de extremos e quartis e por caixa de bigodes.
 2. Repete agora este estudo tomando todos os dados à excepção do maior.
 3. Compara as medidas e as representações gráficas que obtiveste e discute qual das duas hipóteses (1 ou 2) te parece mais adequada para tratar aquela amostra.

Um *outlier* tem um papel fortemente desestabilizador numa amostra, tanto que pode ser mais adequado não o considerar. Os diagramas de extremos e quartis e os gráficos caixa de bigodes, ambos facilmente representados pelas calculadoras gráficas, ajudam também a perceber visualmente essas diferenças.

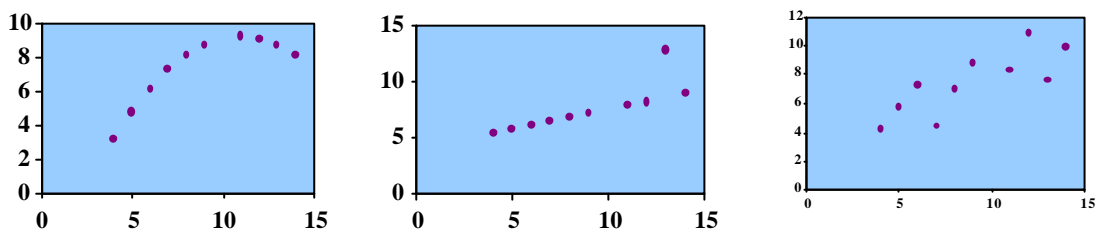
2	3	5	7	11	13	100	média: 20,1	desvio padrão: 32,8
2	3	5	7	11	13		média: 6,8	desvio padrão: 4

Correlação linear

A ideia de correlação entre duas variáveis aparece apenas no 10º anos, mas parece-me que ela poderia com vantagem ser explorada informalmente no 3º Ciclo, onde os alunos já trabalham com funções tipo $y = ax + b$. O coeficiente de correlação é um indicador usado bastantes vezes nos órgãos de comunicação para mostrar a existência de relações de dependência entre duas variáveis que se pretendem associar. No entanto, é um indicador matemático que pode ser usado para relacionar variáveis que nada têm a ver uma com a outra e pode até provar relações fortíssimas entre variáveis completamente independentes — mas que dependem ambas de outros factores, por exemplo.

Para que os alunos possam aperceber-se dos limites e dos perigos do coeficiente de correlação tem bastante interesse analisar diversas nuvens de pontos e estimar o valor do coeficiente de correlação associado.

1. Observa os seguintes gráficos e indica a tua previsão acerca do coeficiente de correlação linear associado a cada caso.



O mais natural é que os alunos indiquem estimativas diferentes para os coeficientes de correlação linear associados aos três casos, mas na realidade, ao determinar o coeficiente na calculadora, observam que ele é exactamente o mesmo: $r = 0,816$.

Este facto comprova rapidamente que o coeficiente é um indicador que pode ser igualmente bom para conjuntos de pontos muito distintos, e portanto, dizer pouco sobre o tipo de relação que existe entre as variáveis que correlaciona.

Outra actividade indicada para perceber as manipulações que se podem realizar de modo a alterar o valor do coeficiente de correlação relativo a duas variáveis

consiste em analisar o efeito de omitir alguns pontos do respectivo conjunto de pares (x,y) .

1. Estuda a correlação linear entre x e y .

x	10	8	13	9	11	14	6
y	8,4	6,9	7,6	8,8	8,3	10	7,3

- a) Qual o valor do coeficiente de correlação?
b) Omite alguns pontos do conjunto e determina novamente o valor do coeficiente. O que aconteceu?
c) Considera agora apenas três pontos. Qual é o valor do coeficiente de correlação linear?
d) Omite todos os pontos à excepção de dois. Qual é agora o valor do coeficiente?
-

Esta actividade permite ganhar consciência de como o valor do coeficiente de correlação pode ser manipulado, em especial, de como pode ser melhorado pela omissão de alguns pontos. Essa manipulação pode eventualmente ser considerada legítima se se omitir um *outlier* mas também pode ser “batota” se se omitirem outros pontos.

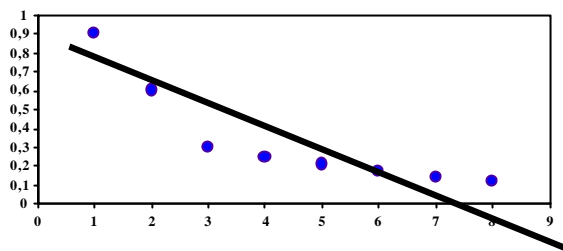
Uma última proposta relacionada com a correlação tem a ver com a exploração de outros tipos de correlação que não apenas a linear, e isto penso que é perfeitamente realizável com alunos do 11º ou 12º ano, que já conhecem muitas famílias de funções. As calculadoras oferecem a possibilidade de trabalhar com correlações diversas e isto abre muito o leque de hipóteses que se podem pôr ao nível da modelação matemática de situações. Uma situação que considero importante explorar com os alunos consiste em avaliar os limites da recta de regressão como modelo da relação entre duas variáveis, mesmo quando o coeficiente de correlação linear é elevado.

1. Estuda a correlação entre x e y .

x	1	2	3	4	5	6	7	8
y	0,9	0,6	0,3	0,25	0,21	0,17	0,14	0,12

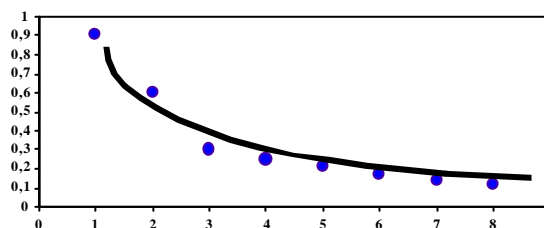
- a) A recta de regressão poderá ser um bom modelo matemático para relacionar as variáveis x e y ?
- b) Sugere outros modelos que te pareçam adequados para estabelecer a relação entre aquelas variáveis.
-

O valor do coeficiente de correlação é elevado, $r = -0,87$, o que nos poderia levar a aceitar a recta de regressão como um bom modelo explicativo da relação entre as



variáveis. No entanto, a nuvem de pontos sugere que os pontos se dispõem sobre uma curva e aparentemente, não será de esperar que se venham a obter valores de y negativos pelo que a recta teria graves limitações ou distorções. Este tipo de relação entre duas variáveis é muito frequente – um exemplo concreto é a evolução de recordes de tempos no atletismo, que vão sendo cada vez menores mas nunca se irão anular.

Explorando com calculadora gráfica outros tipos de correlação mais adequados descobre-se facilmente que uma correlação exponencial é mais apropriada. O respectivo valor de r é $-0,96$.



E como concluir?

A Estatística só tem sentido se for abordada de forma a não desvirtuar a sua natureza, de forma a ser vista e usada como uma ferramenta matemática que permite compreender melhor o mundo que nos rodeia, e em especial, a intervir na

sociedade de uma forma mais informada. Para tal, os alunos devem desenvolver um espírito atento e crítico e a capacidade de compreender que, em qualquer estudo estatístico, as amostras são potencialmente mentirosas, a média pode induzir um grande engano, o desvio padrão procura sempre mostrar a verdade e o coeficiente de correlação pode ser uma fraude descarada.

Bibliografia

Abrantes, P., Serrazina, L., Oliveira, I. (1999). *A Matemática na Educação Básica*.

Lisboa: Ministério da Educação - DEB.

APM, Projecto T³ (1999). *Estatística e Calculadoras Gráficas*. Lisboa: APM.

Martins, E., Monteiro, C., Viana, J., Turkman, A. (1999). *Estatística - 10º ano*.

Lisboa: Ministério da Educação - DES.

NCTM (1991). *Normas para o Currículo e para a Avaliação da Matemática*

Escolar (tradução da APM, do original de 1989, Professional Standards for Teaching Mathematics). Reston: NCTM.

NCTM (1994). *Normas Profissionais para o Ensino da Matemática Escolar*

(tradução da APM, do original de 1991, Professional Standards for Teaching Mathematics). Reston: NCTM.