



Dados reais e tecnologia

José Duarte

O acesso a dados reais através das bases de dados

Desde o início da escolaridade que o programa de Matemática do ensino básico, no que diz respeito a OTD, refere a importância de recolher, ler, organizar e interpretar dados (1.º ciclo), compreender e produzir informação estatística com vista à tomada de decisões (2.º ciclo) e conduzir estudos estatísticos, formulando questões, planeando a recolha e a representação de dados, com vista à tomada de decisões informada e argumentada (3.º ciclo). No ensino secundário, o trabalho em Estatística aprofunda o que foi realizado anteriormente, acentuando o desenvolvimento da capacidade de análise crítica dos dados e suas representações, com vista a fornecer aos alunos ferramentas de análise e desenvolver capacidades de argumentação informada face a afirmações e informações com que se confronta no quotidiano, nomeadamente através dos *media*.

Nos últimos anos desenvolveram-se, principalmente através de *sites* na Internet, conjuntos de bases de dados nacionais e internacionais que constituem um enorme manancial de informação colocada à disposição de todos os utilizadores. Essa informação é hoje um capital de grande importância na vida das

empresas para identificarem tendências e desenharem cenários de evolução e desenvolvimento. Também a educação, com o acesso progressivo a computadores e à Internet, pode beneficiar destes recursos em situação de sala de aula, permitindo que os alunos sejam confrontados com dados reais, suscetíveis de fornecer contextos significativos para trabalhos e projetos, ponto de partida para boas e desafiantes tarefas matemáticas.

O portal do Instituto Nacional de Estatística (INE)

O portal do INE (www.ine.pt) apresenta um extenso repositório de dados estatísticos em permanente atualização, relativos a diferentes variáveis e indicadores como a população residente, o número de alojamentos ou a taxa de desemprego. Também oferece o acesso a *Publicações*, como o *Boletim Mensal de Estatística*, onde pode encontrar dados estatísticos mensais e trimestrais organizados por temas como a população e condições sociais, a agricultura, a indústria, o comércio ou os serviços.

Recentemente ficaram disponíveis os resultados preliminares do Recenseamento da População e Habitação de 2011 (Censos 2011), em todas as unidades estatísticas (indivíduos, famílias,

> Incluir/retirar indicadores > Alterar condições de seleção > Alterar formato do quadro > Visualizar quadro > Visualizar gráfico

Mi [ícones]

Local de residência	População residente (N.º) por Local de residência e Sexo; Decenal (1)		
	Período de referência dos dados		
	2011		
	Sexo		
	HM	H	M
	N.º	N.º	N.º
Portugal	10 561 614	5 047 387	5 514 227
Continente	10 047 083	4 799 593	5 247 490
Região Autónoma dos Açores	246 746	121 533	125 213
Região Autónoma da Madeira	267 785	126 261	141 524

População residente (N.º) por Local de residência e Sexo; Decenal - INE, Censos - séries históricas
 Nota(s):
 (1) Dados Provisórios

Última atualização destes dados: 07 de dezembro de 2011

Figura 1. Quadro da população residente (INE, 2011)

alojamentos e edifícios) das diferentes zonas geográficas em que se encontra dividido o país (NTUS).

Por exemplo, em População residente pode visualizar um quadro com dados por região e por sexo (figura 1).

No entanto esta informação não é estática, pois nela pode incluir outros indicadores (por exemplo, o número de alojamentos), alterar as condições de seleção (por exemplo, selecionando algumas regiões ou algum período de tempo, em particular) ou visualizar um gráfico que ainda pode personalizar (formato, eixos e legenda).

Mas não fica por aqui: estes dados podem ser exportados diretamente para formato Excel e, em seguida, editados na folha de cálculo. Da mesma forma, os gráficos podem ser também exportados por este processo ou em formato de imagem (JPG ou PDF). Note que, quando edita na folha de cálculo um ficheiro de dados importado do INE, ele integra duas folhas: a folha com o quadro de dados e uma folha de metainformação que contém informação específica acerca dos dados (conceitos envolvidos, amostra, processo de recolha, etc.).

A base de dados PORDATA

Outro site de referência que oferece um conjunto significativo de dados estatísticos reais é a PORDATA (www.pordata.pt), uma base de dados sobre Portugal, organizada pela Fundação Francisco Manuel dos Santos (Figura 2). Aqui pode encontrar dados globais sobre Portugal e cada um dos seus municípios, mas também a possibilidade de os comparar com dados de outros países europeus, organizados por temas como população, educação, saúde, emprego, ambiente e território e ciência e tecnologia.

Em cada um desses temas, pode ainda aceder a um conjunto de subtemas. Por exemplo, em Educação pode escolher *Escolaridade da população*, *Alunos matriculados no ensino superior* ou *Despesas com a Educação*, só para referir alguns exemplos. Se no tema *População*, escolher *Óbitos* e *Esperança de vida*, tem ainda acesso

a um conjunto e categorias de informação como *Esperança de vida à nascença por sexo*, *Óbitos por algumas causas de morte por 100.000 habitantes* ou *Taxa bruta de mortalidade e taxa de mortalidade infantil*.

Se aceder a qualquer destas categorias como, por exemplo, *Esperança de vida à nascença por sexo*, surge um quadro como na figura 3.

Esse quadro pode ser personalizado de acordo com as suas necessidades, introduzindo mais anos ou removendo anos que não pretende, introduzindo as variações em valor absoluto ou percentual, de ano para ano ou a partir de um ano que considere como base de comparação (índice 0), e visualizando a informação sob a forma de gráficos estáticos ou dinâmicos. Nestes últimos pode selecionar diferentes tipos de gráficos e escalas, assim como introduzir variação na variável independente (por exemplo, ano) e visualizar o correspondente efeito. E tal como já acontecia com o portal do INE, os dados podem ser exportados em formato *xlsx* (para a folha de cálculo) ou *pdf*.

Mas, para além dos dois grandes portais de dados acima referidos (INE e PORDATA), existem hoje inúmeros sites que disponibilizam dados, focados sobre temáticas específicas como, por exemplo, as bases de dados dos Jogos Olímpicos (em <http://www.databaseolympics.com> ou <http://jogos-olimpicos.terra.com.br>), os sites sobre a liga americana de basquetebol NBA (em <http://www.nba.com/> ou <http://pt.global.nba.com/>) ou a Agência Portuguesa do Ambiente (<http://www.qualar.org>).

Em resumo, longe vão os tempos em que tínhamos de inventar valores ou recolhê-los de um manual escolar, quantas vezes desatualizados ou despropositados, para os usar em tarefas de estatística na sala de aula. Hoje, as grandes bases de dados ou os sites de natureza temática permitem-nos aceder e trabalhar com dados reais e em tempo quase real, criando oportunidades ao professor para criar tarefas com sentido, suscetíveis de envolver os alunos em atividades significativas. Além do acesso aos dados em bruto, a própria tecnologia tem vindo a disponibilizar ferramentas e opções que permitem organizar, comparar

Figura 2. Página de entrada do site PORDATA (Portugal)

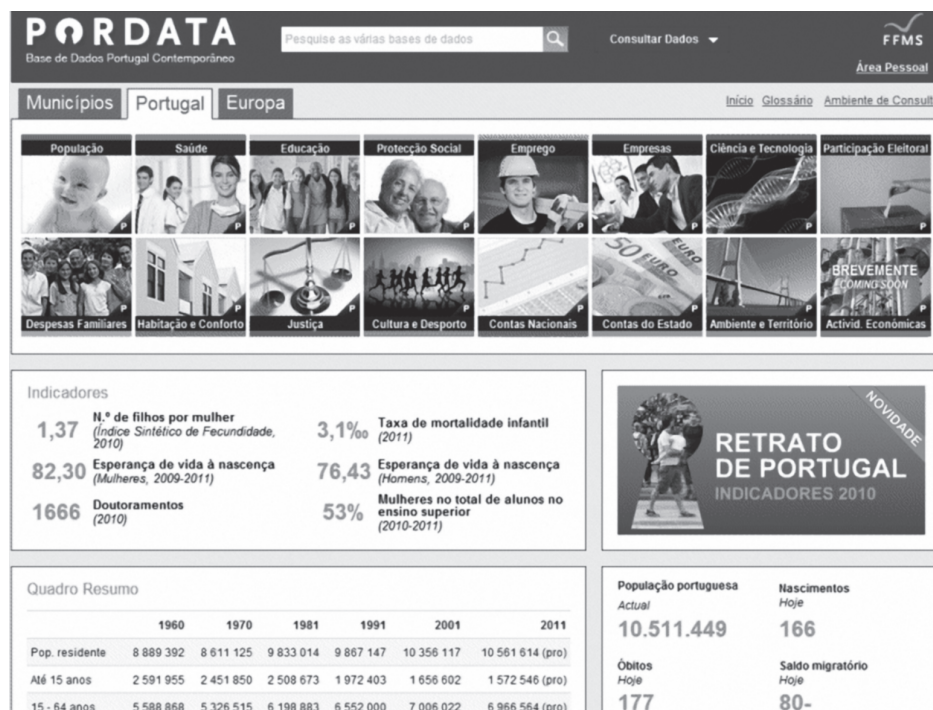
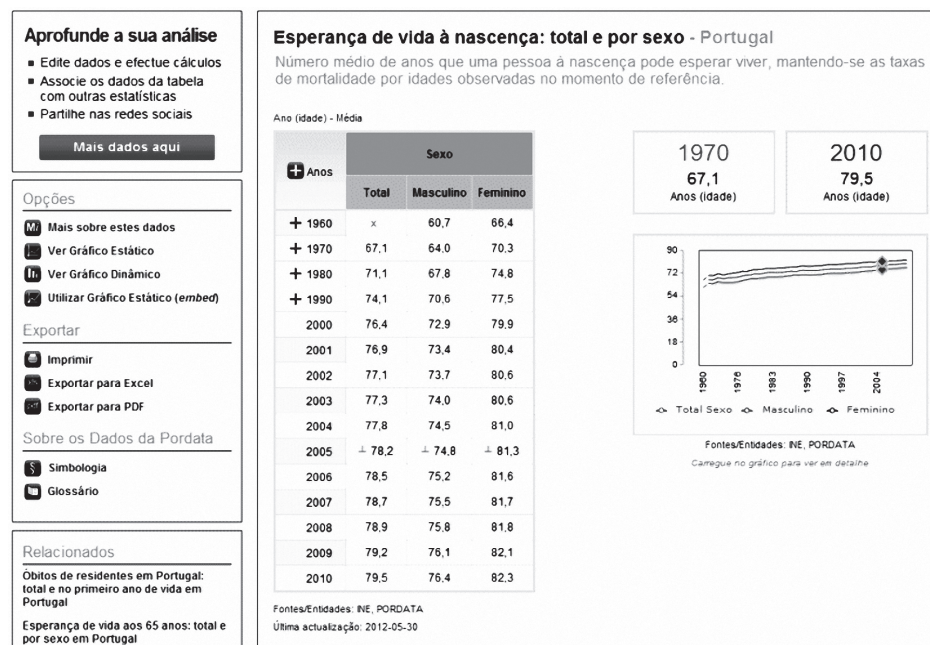


Figura 3. Quadro de dados do PORDATA que pode ser personalizado



e personalizar a informação apresentada, adequando a oferta às necessidades da procura e facilitando a análise e interpretação dessa mesma informação.

O acesso a dados reais através de questionários a distância

A própria turma pode oferecer um excelente contexto para aceder a dados reais dos alunos. Quantas vezes já solicitámos, através de recolha realizada em papel ou de um simples braço no ar, dados como idade, peso, opiniões sobre eventos, ou preferências culturais ou desportivas? Mas se quisermos alargar a amostra a todas as turmas de um ano de escolaridade ou a toda

a escola, este processo é de todo desaconselhado. O mesmo se passa se quisermos aceder a dados de amostras não directamente acessíveis, como os elementos das equipas que participam num torneio, por exemplo, ou os alunos participantes numa excursão de finalistas.

Esse trabalho pode agora ser desenvolvido de forma eficaz com o recurso a tecnologia que permite a construção de formulários *on-line*. O *Google Docs* (em docs.google.com) é um dos serviços da denominada *Web 2.0*. que oferece essa possibilidade, desde que crie uma conta (grátis). Não há nada para instalar. Das várias opções que surgem, desde permitir carregar ou criar documentos e apresentações, apenas nos interessa, neste momento, uma: criar um formulário. Posso dar-lhe um

Sexo

Masculino

Feminino

Idade

Até 10 anos

Entre 11 e 15 anos

Entre 16 e 20 anos

Entre 21 e 30 anos

Mais de 30 anos

Considera as redes sociais apropriadas para

Crianças do 1º ciclo

Jovens adolescentes do 2º/3º ciclos

Jovens adultos no ensino secundário

Adultos

A escola do ensino básico é uma instituição onde
(assinale as opções que considere adequadas)

se divulga a cultura da humanidade

se ensinam e aprendem conteúdos

se prepara para uma profissão

se prepara para a vida

se treinam habilidades

Outro:

Figura 4. Questões do tipo escolha múltipla e seleção a partir de uma lista

Figura 5. Questões do tipo caixa de verificação

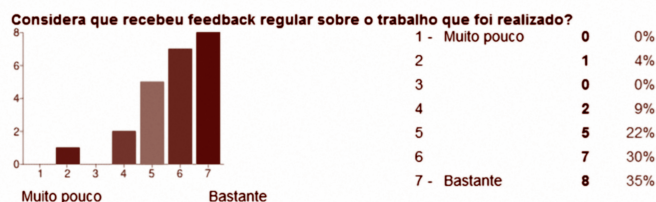


Figura 6. Exemplo de resumo da resposta a uma questão

ar mais profissional, escolhendo um tema (formato e fundo) e os dados recolhidos através das perguntas que formulo são simultaneamente adicionados a uma folha de cálculo (lida pelo Excel) e sobre os quais são gerados, de forma automática, gráficos e tabelas síntese da informação recolhida.

Como em todos os questionários, a dificuldade está nas perguntas a colocar e na forma de o fazer, face ao objetivo pretendido. O *Google Docs*, para a criação de formulários, permite sete diferentes tipos de perguntas. No entanto, por razões de espaço e de interesse, foco-me aqui apenas em três deles: *escolha múltipla*, *selecionar a partir de uma lista* e *caixas de verificação*.

Os tipos escolha múltipla e selecionar a partir de uma lista permitem que o respondente se posicione relativamente a uma e só uma das opções, como acontece por exemplo com a variável sexo ou com a variável idade (figura 4).

Como resultado, a soma dos totais parciais, em valor absoluto e em percentagem, dão respetivamente o total de respondentes (se a pergunta for de resposta obrigatória) e a percentagem de 100%.

Já no que respeita a uma pergunta do tipo *caixa de verificação*, o respondente pode optar por assinalar mais do que uma opção, o que tem implicações na análise dos resultados. É o que acontece, por exemplo, quando construímos perguntas que admitem mais do que uma escolha para a mesma pergunta, como é o caso dos dois exemplos abaixo (figura 5). As implicações são que provavelmente a soma do total de respostas, em todas as categorias, ultrapassa o número total das pessoas inquiridas (100%), o que requer cuidados nos comentários interpretativos relativos aos dados recolhidos.

Após todos os alunos da turma terem preenchido o formulário, no instante imediato pode aceder-se a todos os dados disponíveis numa folha de cálculo do *Google Docs* e a um resumo (*Formulário — Mostrar resumo das respostas*), em tabela e na forma de gráficos (figura 6), que a ferramenta gera de forma automática, permitindo uma primeira análise e discussão.

No entanto, os dados podem ser transferidos para o formato Excel (*Ficheiro — Transferir — Formato Excel*) e, na folha de cálculo, serem organizados de forma personalizada em tabelas e gráficos, de modo a melhor corresponderem aos objetivos do professor.

Se bem que esta recolha de dados tenha bastante valor, quando a amostra se encontra a distância e sempre que se quer preservar o anonimato, ela também pode ser usada para recolher dados na aula, quando exista acesso a computadores com ligação à Internet, pois permite uma rápida introdução dos dados e proporciona, de imediato, um primeiro tratamento da informação que até pode ser suficiente para os objetivos que o professor tenha em vista.

Representar e tratar dados reais com a tecnologia

Discutido o acesso a dados reais, na era da tecnologia, o professor decide o resto, talvez mesmo o essencial: o que fazer aos dados e como utilizar a tecnologia na organização, representação e apresentação da informação?

O tratamento de um grande volume de dados reais não dispensa o uso da tecnologia, pela diversidade de representações que proporciona, pela versatilidade na transição entre elas e pela facilidade no cálculo de medidas estatísticas. E essa tecnologia pode ser bem diversa, desde o *software* mais clássico, como a folha de cálculo ou as calculadoras gráficas, às *applets*, acessíveis mesmo a alunos bastante jovens.

Diagramas para caracterizar conjuntos de dados

As *applets* são ferramentas simples e adequadas para apoiarem objetivos específicos de aprendizagem e para ilustrarem conceitos relacionados com a caracterização de distribuições estatísticas. Os diagramas de caule e folhas e os diagramas de extremos e quartis são representações particularmente importantes para comparar duas distribuições, o que pode acontecer quando

	A	B	C
1	Esperança de vida à nascença, por ano e por sexo		
2	Idade		
3	Sexo		
4	Ano	Homens	Mulheres
5	1990	70,6	77,5
6	1991	70,6	77,6
7	1992	70,8	78
8	1993	71	78,2
9	1994	71,5	78,5
10	1995	71,8	79
11	1996	71,7	79
12	1997	71,9	79,1
13	1998	72,2	79,4
14	1999	72,5	79,6
15	2000	72,9	79,9
16	2001	73,4	80,4
17	2002	73,7	80,6
18	2003	74	80,6
19	2004	74,5	81
20	2005	74,8	81,3
21	2006	75,2	81,6
22	2007	75,5	81,7
23	2008	75,8	81,8
24	2009	76,1	82,1
25	2010	76,4	82,3
26	Dados obtidos de www.pordata.pt em 2012-09-17		

Figura 7. Dados importado do PORDATA

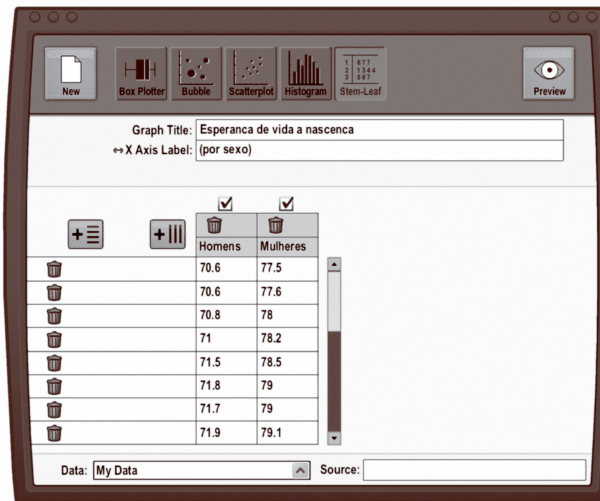


Figura 8. Tabelas com os dados na applet

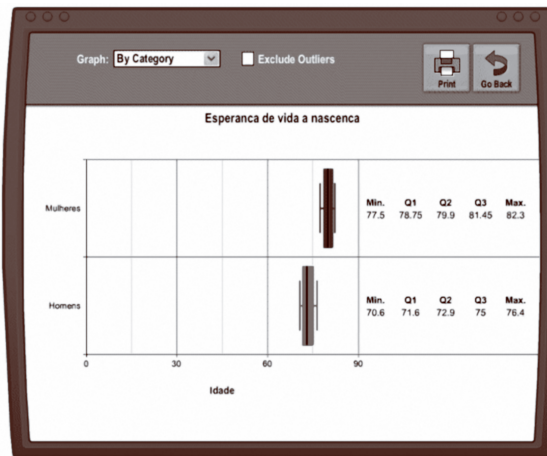
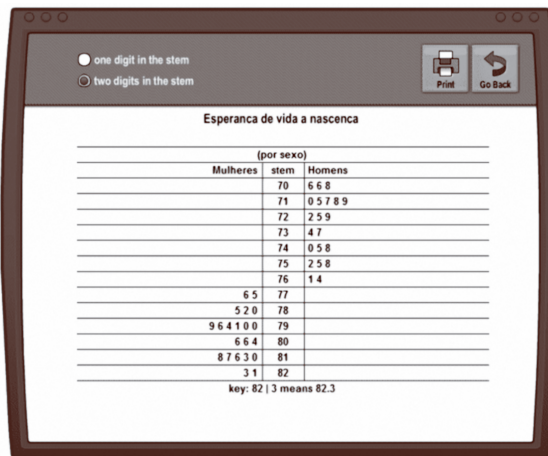


Figura 9. Diagrama de caule e folhas e diagrama de extremos e quartis

queremos comparar a esperança média de vida de alguns países europeus e africanos ou a mesma variável, em Portugal, ao longo de duas décadas, mas analisando a diferença entre homens e mulheres.

Assim, na base de dados PORDATA, procuramos pelo indicador *Esperança de vida à nascença (total e por sexo)* e aprofundamos a nossa análise, personalizando a informação que queremos em *Mais opções aqui*. Dispensamos o *total* e ficamos só com a informação de duas décadas (1990 a 2010), em cada um dos sexos. Exportamos a informação para o Excel e obtemos algo parecido com a figura 7.

Posteriormente, podemos introduzir esses dados na tabela da *applet Advanced Data Grapher*, em <http://illuminations.nctm.org/ActivityDetail.aspx?ID=220> (figura 8).

Esta aplicação, constitui um exemplo de uma ferramenta simples que após introdução dos dados disponibiliza as duas representações (diagrama de caule e folhas e diagrama de extremos e quartis), entre outras, e permite a comparação e análise das distribuições e a sua discussão na sala de aula (figura 9).

Relações e associação entre variáveis

Se pretendermos estudar a existência de associação estatística entre duas variáveis quantitativas e determinar o coeficiente de correlação r (de *Pearson*) que a caracteriza, podemos, por exemplo, usar a folha de cálculo. Consultando uma base de dados brasileira sobre os últimos Jogos Olímpicos, realizados em Londres, em 2012 (<http://jogos-olimpicos.terra.com.br/medalhas/>), podemos aceder, por exemplo, ao número de medalhas

	A	B	C	D	E
1	Os 15 países mais medalhados - Jogos Olímpicos de Londres (2012)				
2					
3	País	Medalhas de ouro	Total de medalhas	Estimativa teórica	Ouro/Total (%)
4	Estados Unidos	46	104	107,5	44%
5	China	38	88	90,5	43%
6	Grã-Bretanha	29	65	71,4	45%
7	Rússia	24	82	60,8	29%
8	Coreia do Sul	13	28	37,4	46%
9	Alemanha	11	44	33,1	25%
10	França	11	34	33,1	32%
11	Itália	8	28	26,8	29%
12	Hungria	8	17	26,8	47%
13	Austrália	7	35	24,6	20%
14	Japão	7	38	24,6	18%
15	Cazaquistão	7	13	24,6	54%
16	Holanda	6	20	22,5	30%
17	Ucrânia	6	20	22,5	30%
18	Nova Zelândia	6	13	22,5	46%

Figura 10. Tabela com as 15 equipas mais medalhadas (Londres, 2012)

DECLIVE	2,13
INTERCEPTAR	9,77
PEARSON	0,94

Figura 12. Valores das funções na folha de cálculo

de ouro e ao número total de medalhas, atribuídas aos países participantes.

A partir dos dados reais registados nas colunas B e C (figura 10), podemos obter um primeiro esboço gráfico da forma como os valores se distribuem, o que nos dá uma primeira leitura intuitiva sobre a existência de uma relação positiva forte, aliás esperada, entre as variáveis, medalhas de ouro e total de medalhas (figura 11), visível na forma como os pontos se alinham em redor de uma hipotética reta (a tracejada).

Do gráfico e da sua confirmação na tabela podemos ainda observar que existem alguns países (Cazaquistão, Hungria, Coreia do Sul e Grã-Bretanha) que se *descolam mais* da zona central da nuvem de pontos, para a parte inferior do gráfico, o que quer dizer que têm, entre os três tipos de medalhas possíveis (ouro, prata ou bronze), um maior rácio ouro — total de medalhas, visível na coluna E da tabela (figura 10). Pelo contrário, países com o Japão ou a Rússia, encontram-se na zona superior, o que revela uma percentagem baixa de medalhas de ouro obtidas, quando comparadas com o número total de medalhas que alcançaram.

Uma vez que a leitura do gráfico nos induz uma forte associação entre as variáveis, visível pela distribuição linear dos pontos em torno da reta que melhor se lhes ajusta, podemos confirmá-la, calculando o coeficiente r de Pearson (função PEARSON, da folha de cálculo) que mede essa associação. Este valor, que pode variar entre -1 e 1 , reflete a força, dada pelo valor absoluto, e o sentido da associação, dado pelo sinal, que indica o sentido da variação. Neste caso concreto, o valor do coeficiente ($+0,94$), confirma a forte associação positiva entre o nº de medalhas de ouro (variável independente) e o total de medalhas (variável dependente).

Com este valor, faz todo o sentido representar a reta de regressão linear que mostra a tendência e permite fazer estimativas para além dos valores registados. A partir dos valores reais

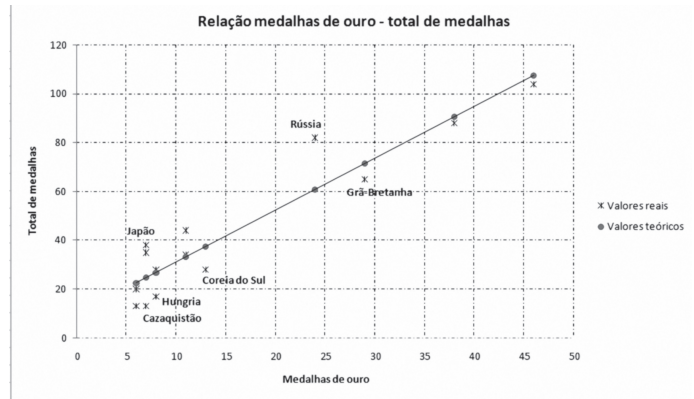


Figura 11. Gráfico que mostra a relação ouro-total

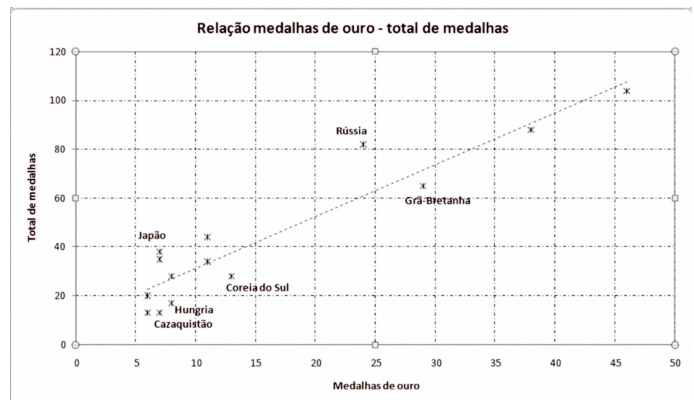


Figura 13. Representação gráfica, com a reta de regressão linear

podemos calcular o declive (função DECLIVE) e a ordenada na origem (função INTERCEPTAR) da reta que melhor se ajusta a esse conjunto de pontos (reta de regressão linear).

Com esses parâmetros e as abcissas reais (coluna B), calculamos os valores teóricos estimados ($y = \text{DECLIVE} * x + \text{INTERCEPTAR}$) na coluna D e, em seguida, acrescentamo-los ao gráfico e pedimos a linha de tendência ou simplesmente unimos os pontos. É mais visível agora que os pontos relativos aos valores reais se distribuem bastante próximos, acima e abaixo da reta que passa pelos valores teóricos estimados (Figura 13).

Em resumo, dados reais e tecnologia constituem hoje recursos do professor de Matemática que representam uma mais-valia para o trabalho de organização e tratamento de dados. As diferentes ferramentas tecnológicas permitem o acesso a grandes volumes de dados, em portais e sites generalistas ou temáticos, facilitam a organização dos dados e a sua representação sob diferentes formas e possibilitam o cálculo de indicadores estatísticos que melhor os caracterizam.

A diversidade e flexibilidade das representações da tecnologia acrescentam compreensão na apropriação dos conceitos, com o professor no papel de orquestrador de boas discussões que permitam estabelecer as pontes necessárias entre os conceitos ‘vistos’ através da tecnologia e os mesmos conceitos quando trabalhados apenas com lápis e papel.

José Duarte
Escola Superior de Educação do Instituto Politécnico de Setúbal