In: Praticando Geografia: Técnicas de campo e laboratório. Organizador: Luis Antonio Bittar Venturi. São Paulo: Oficina de Textos, 2005, pp.175-186

Sistematização de dados quantitativos

Prof. Dr. Emerson Galvani

> É possível mentir usando estatísticas, mas se mente mais, e melhor, sem estatísticas. É preciso entender que as amostras podem levar a conclusões erradas. Contudo, as opiniões pessoais, sem base de dados, levam, em geral, a conclusões muito mais erradas.
>
> Frederick Mosteller (Vieira, 1999).

1. Apresentação

A graduação em Geografia exige, pela própria natureza do curso, um número significativo de trabalhos de campo. Essas *saídas* realizadas pelas diferentes áreas/disciplinas, cada qual com seu instrumental apropriado, produzem em cada trabalho um volume de informações específicas e, quando retornamos para a sala de aula, a grande questão que se apresenta é "o que fazer com os dados quali-quantitativos coletados no trabalho de campo?" Tradicionalmente, os alunos de graduação em Geografia não são muito afeitos à área de exatas o que, em certas ocasiões, limita a análise e interpretação dos dados observados. Por vezes, os resultados finais obtidos ficam prejudicados por falta de uma análise mais numérica (estatística) dos dados observados.

O que se pretende com este capítulo não é formar especialistas em estatística, mesmo porque não há espaço nem necessidade para isto, mas sim, fornecer os princípios básicos de estatística descritiva permitindo uma melhor análise dos dados obtidos nos trabalhos de campo e também, desmistificar o trauma que é imposto aos nossos alunos com relação às ciências exatas. Vale lembrar que essa sistematização de dados quantitativos, como chamaremos a partir deste momento, aplica-se a qualquer tipo de informação, seja ela produto de questionários ou medições específicas em cada área/disciplina do conhecimento.

2. Medidas de Tendência Central

A análise de um conjunto de dados com uso de tendência central nos permite avaliar para onde caminha nosso dado. Uma espécie de *raio-X* inicial. Esse *raio-X* pode ser determinado com a utilização dos indicadores descritos a seguir.

2.1 Média Aritmética ($\overline{X}$)

A média aritmética é o procedimento mais simples e comum passível de ser aplicado a um conjunto de dados. Esta medida de tendência central expressa o somatório de todos os elementos da série dividido pelo número total de elementos. Numericamente, a média aritmética é expressa por:

$$\overline{X} = \frac{\sum_{1}^{n} X_i}{n}$$

Em que X_i é cada elemento da série, e *i* varia de *1* a *n*; *n* é o número de elementos e o símbolo Σ significa somatório de todos os elementos da série. Resumindo, soma-se todos os elementos e divide pelo número total de elementos da série

2.2 Moda (MO)

A moda ou modo (MO) é o valor presente na série que ocorre com maior freqüência. Existem séries em que nenhum dado se repete, nesses casos não existe a moda da série. Isso geralmente ocorre em séries reduzidas (menos que cinqüenta elementos amostrados). De forma análoga, podem ocorrer séries com duas (bi-modal) ou mais modas. Nesses casos prevalece o valor de maior freqüência de ocorrência ou, em caso de empate a série poder apresentar mais de uma moda.

2.3 Mediana (ME)

A mediana é aplicável em séries extensas de dados (mais de mil informações) nas quais existem extremos que possam *contaminar* a média, ou seja, alguns dados que *fogem* da *tendência central*, podendo sub ou superestimar as análises. A mediana é determinada ordenando-se os dados de forma crescente ou decrescente e identificando a posição central da série. Em caso de séries com número impar de elementos, a mediana estará na posição central da série. Para séries com número par de elementos, a mediana será a média dos elementos que ocupam a posição central da série. O conceito de Mediana gera algumas confusões: a Mediana é simplesmente o valor que se situa na posição central do conjunto de dados ordenados. *Assim, tem que haver uma relação de ordem nos valores*.

2.4 Valor Máximo (Vmax) e mínimo (Vmin)

O valor máximo da série é aquele de maior magnitude, ou seja, o maior valor encontrado na série. O valor mínimo, por sua vez, é o menor valor encontrado na série. Em princípio, parece ser uma informação sem importância, contudo nos permite visualizar em que intervalo de medidas encontra-se distribuído o meu conjunto de dados. Serve para evidenciar o *tamanho* dos dados que serão trabalhados. Em séries climatológicas de temperatura do ar, por exemplo, o Vmax equivale à temperatura máxima do ar e o Vmin à temperatura mínima do ar.

2.5 Amplitude (Δ)

A amplitude em um conjunto de dados expressa a diferença entre o Vmax e o Vmin. Essa medida de tendência central expressa a variação máxima dos valores constituintes do conjunto de dados. Dois ou mais conjuntos de dados poderão ter mesma média, porém diferentes Vmax, Vmin e Δ, evidenciando-se tratar de séries distintas.
A seguir será apresentado, para um conjunto simples de dados exemplo de cálculo das medidas de tendência central: média, moda, mediana, valor máximo, valor mínimo e amplitude.

Tabela 1. Valores arbitrários para duas variáveis A e B. Dados brutos (à esquerda) e a dados ordenados (à direita) em forma crescente.

Dados Brutos		**Dados Ordenados**	
A	**B**	**A**	**B**

121	171	121	152
171	152	157	168
158	170	158	169
173	168	163	170
184	169	171	171
163	171	173	171
157	190	184	190

Tabela 2. Resultados da análise de tendência central para o conjunto de dados da tabela 1.

Medida de tendência	Variável A	Variável B
Média ($\overline{X}$)	161	170
Moda (MO)	?	171
Mediana (ME)	163	170
Valor máximo (Vmax)	184	190
Valor mínimo (Vmin)	121	152
Amplitude (Δ)	63	38

Esses procedimentos podem ser efetuados *facilmente* com o programa *Excel da Microsoft*[1] pelos seguintes passos: com o conjunto de dados dispostos em duas colunas; entrar na barra de ferramentas no atalho *fx*; em seguida em *estatística* e selecionar a análise de tendência desejada; selecionar o intervalo de dados. O resultado é mostrado. Caso a barra de ferramentas não disponibilize o atalho *fx* clique em *inserir* e em seguida em *fx,* seguindo os mesmos procedimentos descritos acima.

Qual a diferença de interpretação entre a Mediana e a Média?

Embora a Média seja um valor mais fácil de ser entendido, tem restrições, na medida em que pode nos induzir a um erro de tendência se a amostra analisada apresentar valores de amplitude elevados. Por exemplo, na distribuição dos dados da tabela 1 a Média da variável A é 161 e a mediana é 163. Caso uma amostra tivesse apresentado valor de 300 e não 121. Isto faria com que a Média saltasse para 187, ou seja, seria superior a todos os valores individuais, mas a Mediana continuaria a ser 163. Se olharmos para todos os 7 valores individuais da nossa amostra, verificamos que o número 163 é o melhor representante da distribuição desse conjunto de dados. **Assim, no caso das variáveis quantitativas, quando o valor da Mediana é muito diferente da Média, é aconselhável considerar sempre a Mediana como valor de referência mais importante.**

3. Medidas de dispersão

As medidas de dispersão são úteis quando diferentes conjuntos de dados apresentam a mesma média e mediana, porém variabilidades distintas. Este tipo de análise pode ser utilizado para comparar quantos conjuntos de dados forem necessários, pois os cálculos são efetuados individualmente para cada conjunto.

3.1 Desvio em relação à média (DM)

[1] A citação de marca comercial não implica na recomendação por parte do autor do referido programa. Existem outros programas estatísticos como *Origin, MatLab, Estatística, SAS*, entre outros que efetuam tais procedimentos, contudo com a massificação de uso do Office da Microsoft, o Excel é facilmente encontrado em qualquer computador.

Essa medida de dispersão nos fornece uma idéia da variabilidade dos dados em torno da média, sendo portanto a diferença entre o valor observado (X_i) e a média do conjunto ($\overline{X}$), representado numericamente por:

$$DM = X_i - \overline{X}$$

Determinados conjuntos de dados podem apresentar médias iguais, contudo com acentuados desvios em relação à média, veja o exemplo.

A	B	DM "A"	DM "B"
4	9	-1	4
6	1	1	-4
4	5	-1	0
6	5	1	0
5	1	0	-4
5	9	0	4
$\overline{X}$ **=5**	$\overline{X}$ **=5**	**Σ=0**	**Σ=0**

Tabela 3. Valores de A e B e desvio em relação à média.

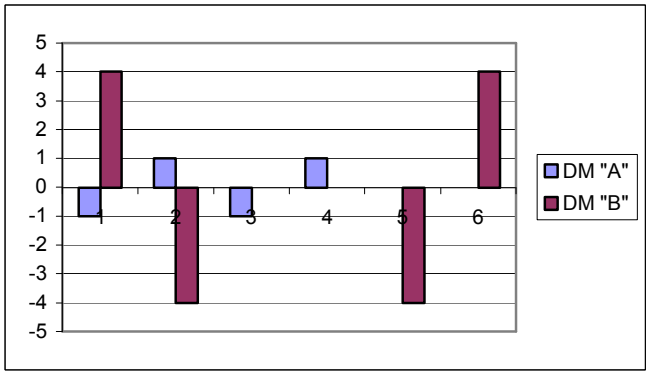

Figura 1. Desvio em relação à média para o conjunto de dados da tabela 3.

No exemplo da tabela 3, representado pela figura 1, observa-se que, embora os conjuntos de dados A e B apresentem a mesma média e mediana, a variabilidade do conjunto A é menos acentuada que em B. Desta forma, a análise somente da média e mediana pode levar a conclusões não satisfatórias. Vale lembrar que o somatório (Σ) dos desvios em relação à média deve ser igual a zero. O desvio em relação à média tem a desvantagem de não fornecer um único indicador da variabilidade dos dados, ficando restrito a uma análise visual dos dados, para tanto existem outros índices, como o que veremos a seguir.

3.2 Variância da Amostra (S^2)

Para se avaliar a variabilidade da amostra faz-se uso da noção de variância. Numericamente é determinada pela somatória do quadrado do desvio em relação à média, dividida pela quantidade de elementos da série.

$$S^2 = \frac{\sum \left(X_i - \overline{X}\right)^2}{n-1}$$

Tabela 4. Exemplo de cálculo de variância.

X	X-$\overline{X}$	(X-$\overline{X}$)^2
4	-1	1
6	1	1
4	-1	1
6	1	1
5	0	0
5	0	0
$\overline{X}$ =5		Σ (X-$\overline{X}$)^2 = 4

Então, a variância será calculada assim:

$$S^2 = \frac{4}{6-1}$$

$$S^2 = 0{,}8$$

Um dos problemas que prejudicam a análise por meio da variância da amostra é justamente o fato de o resultado ser expresso na unidade de medida dos dados elevado ao quadrado. Por exemplo, se a unidade dos dados da tabela 4 for metro, a variância será expressa em metros ao quadrado (m^2), se for quilograma a variância será expressa em kg^2, o que causa dificuldade na interpretação da variância da variável.

3.3 Desvio Padrão (S)

Uma forma de eliminar o problema da interpretação da variância da amostra é extrair sua raiz quadrada. Tem-se assim o desvio padrão. Esta é uma medida do grau de dispersão dos valores em relação ao valor médio (a média). É um erro dizer que o desvio padrão é a média de todas as diferenças, mas podemos *senti-lo* como algo aproximado. É determinado numericamente pela raiz quadrada da variância:

$$S = \sqrt{\frac{\sum \left(X - \overline{X}\right)^2}{n-1}}$$

3.4 Coeficiente de Variação (CV)

Por vezes, queremos comparar duas variáveis quantitativas quanto ao seu grau de dispersão, por exemplo, o peso (em kg) e a idade (em anos). Esta comparação não pode ser feita comparando-se simplesmente os respectivos desvios padrão, porque estes estão expressos em unidades de medida diferentes, i.e., não se pode comparar a dispersão de massa (kg) com a de idade (anos)! No entanto, é possível fazer esta comparação em termos relativos, se calcularmos o coeficiente de variação de cada conjunto de dados, da seguinte forma:

$$CV = \frac{100.S}{\overline{X}}$$

Onde *CV* é o coeficiente de variação expresso em porcentagem e *S* é o desvio padrão já definido anteriormente.
Veja abaixo um exemplo dos itens 3.3 e 3.4.

Tabela 5. Variáveis A, B e C para cálculo de Desvio Padrão e Coeficiente de Variação.

A	B	C
4	9	9
6	1	1
4	5	1
6	5	2
5	1	8
5	9	9

Desvio padrão de A = 0,9 CV_A = 18,0 %
Desvio padrão de B = 3,6 CV_B = 72,0 %
Desvio padrão de C = 4,0 CV_C= 80,0 %

O Coeficiente de Variação expressa, portanto a variabilidade de cada conjunto de dados normalizada em relação à média, em porcentagem. Assim, a variável *A* oscila, em média, 18,0%.

4. Distribuição de Freqüência

No caso de *variáveis nominais* como o sexo ou a raça, só poderão ser calculadas as freqüências. É totalmente impossível calcular a média ou a mediana do sexo porque a escala destas variáveis não tem sequer uma relação de ordem. Repare-se que por vezes codificam-se as variáveis com números para introdução no computador, torna-se possível determinar, erradamente, médias para variáveis nominais, embora tais resultados, evidentemente, não tenham significado nenhum. No entanto, claro que é também possível calcular as freqüências para todas as outras variáveis ordinais ou quantitativas.

4.1 Freqüência (f)

É o número de vezes que determinado evento ocorreu entre todos os elementos amostrados. Parece ser uma medida banal, contudo auxilia na determinação da freqüência relativa vista a seguir.

4.2 Freqüência Relativa (fr)

A freqüência relativa é o número de vezes que determinado evento ocorreu (*na*) em relação ao número total de elementos da série (*n*). Numericamente é dada por:

$$fr = \frac{na}{n}$$

Tabela 6. Exemplo de cálculo de freqüência e freqüência relativa.

Espécie	(na), f	Fr
A	32	31%
B	17	16%
C	43	41%
D	13	12%
Total (*n*)	105	100%

Pode-se agrupar os dados em intervalos de classes para se ter a freqüência relativa por intervalo do conjunto de dados. O número de intervalos de classes (NIC) depende do total de observações e pode ser dado por:

$$NIC = 1 + 3{,}3\left[Log_{10}(n)\right] \qquad \text{ou} \qquad NIC = 5\left[Log_{10}(n)\right]$$

Em que NIC é o número de intervalo de classes, *n* é o número total de observações e Log_{10} é o identificador de logaritmo normal. O NIC é uma escolha arbitrária que normalmente recai entre cinco e 20. Um número de classes muito pequeno ou excessivo pode ocultar certas propriedades da distribuição de freqüência que seriam evidenciadas com a escolha do número adequado de número de classes. Por exemplo, um conjunto de dados com 250 observações poderia ser re-agrupado em 7 classes. Para cada intervalo de classe poder ser calculada a freqüência, caracterizando-se assim a distribuição de freqüência para a amostra.

4.3 Probabilidade (P)

Expressa a relação entre o número de vezes que determinado evento ocorreu (na) e o número total de eventos observados (n), sendo, portanto, a própria freqüência relativa.

$$P = fr = \frac{na}{n}$$

4.4 Tempo de Retorno (T)

Período ou tempo de retorno é definido como o inverso da probabilidade. Em Climatologia é comum se determinar qual o período de retorno (T) para um evento extremo, tal como precipitação superior a 50 ou 100 mm.

$$T = \frac{1}{P = fr/100}$$

Para os dados apresentados na tabela 6 o tempo de retorno da espécie A é 1/0,32 = 3,1. Ou seja, a cada 3 indivíduos, um é da espécie A.

5. Correlação e Regressão Linear

Até os itens anteriores, nossa preocupação foi no sentido de mostrar a tendência central das variáveis, na seqüência mostrar a variabilidade individual de cada conjunto de dados, num segundo momento; e por fim descrever a freqüência de ocorrência de determinada variável. Contudo, caso fosse necessário mostrar uma relação (correlação) entre duas variáveis, como devemos proceder?

5.1 Diagrama de Dispersão

O primeiro passo é montar o diagrama de dispersão. O diagrama de dispersão é um gráfico que permite visualizar a relação entre duas variáveis A e B. Para elaborar o diagrama de dispersão escolha as escalas de x e de y de forma que o gráfico pareça quadrado. A figura 2 mostra três situações de dispersão de dados evidenciam *perfeita correlação* (esquerda), *boa correlação* (centro) *e ausência de correlação* (direita). No caso da figura da direita pode-se descartar a análise das variáveis com uso de regressão linear simples. A Figura do centro dependerá do nível de significância do coeficiente de correlação entre as duas variáveis (R) que será visto a seguir.

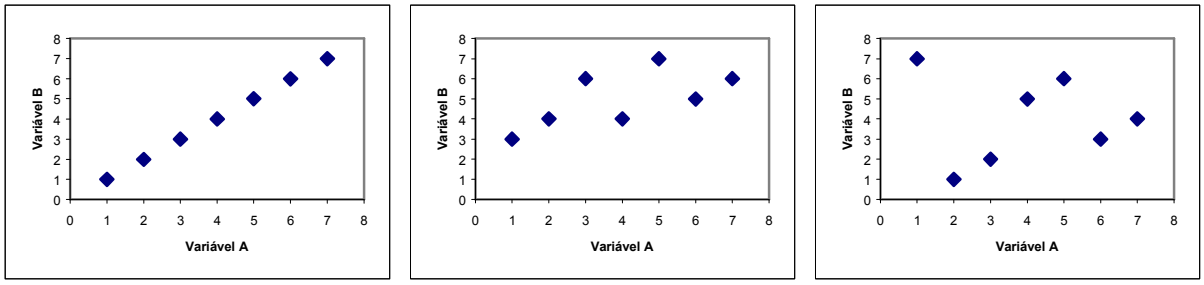

Figura 2. Exemplo de correlação *perfeita* entre duas variáveis A e B (esquerda), *boa* correlação (centro) e *ausência* de correlação (direita).

Percebe-se na figura 2, à esquerda que o aumento de uma unidade em A acarreta igual acréscimo em B, significa que os dados apresentam uma correlação positiva. Poderia ser da mesma magnitude, contudo com decréscimos constantes de B com aumento de A; neste caso a correlação seria negativa.

Para eliminar a subjetividade visual entre as variáveis determina-se numericamente o coeficiente de correlação entre as duas variáveis. O coeficiente de correlação fornece uma medida média da correlação entre duas variáveis.

5.2 Coeficiente de Correlação (R)

O coeficiente de correlação (R) é uma medida do grau de associação linear entre duas variáveis. O R varia de –1,0 a 1,0. Quando zero significa correlação nula e quando 1,0 ou –1,0, correlação perfeita entre as variáveis. Para uma dada amostra com *n* pares de valores das variáveis X e Y, para medir o grau de correlação entre elas, determina-se o coeficiente de Pearson numericamente por:

$$R = \frac{\sum X.Y - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right].\left[\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right]}}$$

Em princípio parece ser algo muito complicado, contudo uma calculadora com funções de estatística fornece os principais indicadores. Também no *Excel* é possível determinar o R clicando em *fx*; em seguida em estatística; depois em *CORREL* selecionar a matriz X (intervalo onde estão os dados a serem relacionados) e a matriz Y e OK. O exemplo abaixo mostra a relação entre temperatura do ar e altitude.

Tabela 07. Exemplo hipotético da variação da temperatura do ar (Y) em função da altitude (X).

Altitude (metros) - X	Temperatura do ar (^oC) - Y
1800	14,7
1600	17,6
1400	18,1
1200	18,9
1000	19,6
800	20,8
600	22
400	22,9
200	24,5
0	25,3

Vamos exemplificar os cálculos de maneira mais didática.

Σx e Σy é a soma de todos os valores de x e de y.

$\Sigma x = 9.000$ $\Sigma y = 204,4$

$(\Sigma x)^2$ e $(\Sigma y)^2$ é o somatório de todos os x e y elevados ao quadrado.

$(\Sigma x)^2 = 81.000.000$ $(\Sigma y)^2 = 41.779,4$

Σx^2 e Σy^2 é cada x e y elevado ao quadrado, e em seguida, somam-se todos.

$\Sigma x^2 = 11.400.000$ $\Sigma y^2 = 4.276,2$

Σx.y em primeiro multiplicam-se todos os x vezes os y e, somente em seguida efetua-se, o somatório.

Σx.y=166.140

Na seqüência basta substituir os valores na equação de calculo de R:

$$R = \frac{166.140 - \frac{9.000 * 204,4}{10}}{\sqrt{\left[11.400.000 - \frac{81.000.000}{10}\right].\left[4.276,2 - \frac{41.779,4}{10}\right]}}$$

Efetuando-se os cálculos obtém-se um coeficiente igual a **–0,99**. O sinal negativo de R indica que com a elevação da altitude a temperatura tende a diminuir, fato este de manifestação natural na atmosfera terrestre, principalmente dentro da primeira camada atmosférica[2] (Troposfera). De fato, uma análise da figura 3 permite observar uma correlação negativa elevada entre os dados apresentados.

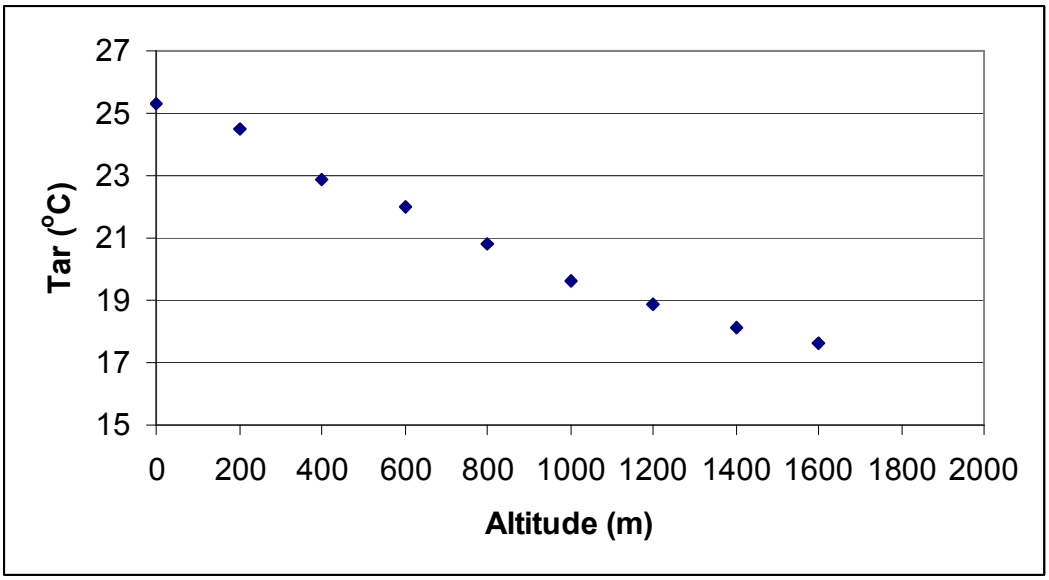

Figura 3. Variação da temperatura do ar em função da altitude (exemplo hipotético, mas próximo da realidade).

5.3 Regressão Linear

Observando a figura 3 acima é possível perceber a relação entre as duas variáveis. Se imaginarmos uma reta de tendência passando pelos pontos, observaremos que esta reta passa por quase todos eles (veja figura 4 abaixo). Então, basta ajustar uma reta que teremos a relação entre essas duas variáveis. Lembrando que a equação da reta é descrita por:

$$y = a + b.x$$

[2] Em noites com ocorrência de inversão térmica podem ocorre gradientes de temperatura que indicam aumento da temperatura com a altitude.

Em que neste, caso *y* é a temperatura do ar e *x,* a altitude. O coeficiente linear *a* forncece a posição em que a reta corta o eixo das ordenadas (*y)* e, o coeficiente angular *b* é a tangente trigonométrica do ângulo formado entre a linha da abscissa (x) e a reta ajustada pela regressão linear. A variável *y* é denominada de dependente e a variável *x* é denominada de explanatória ou independente.

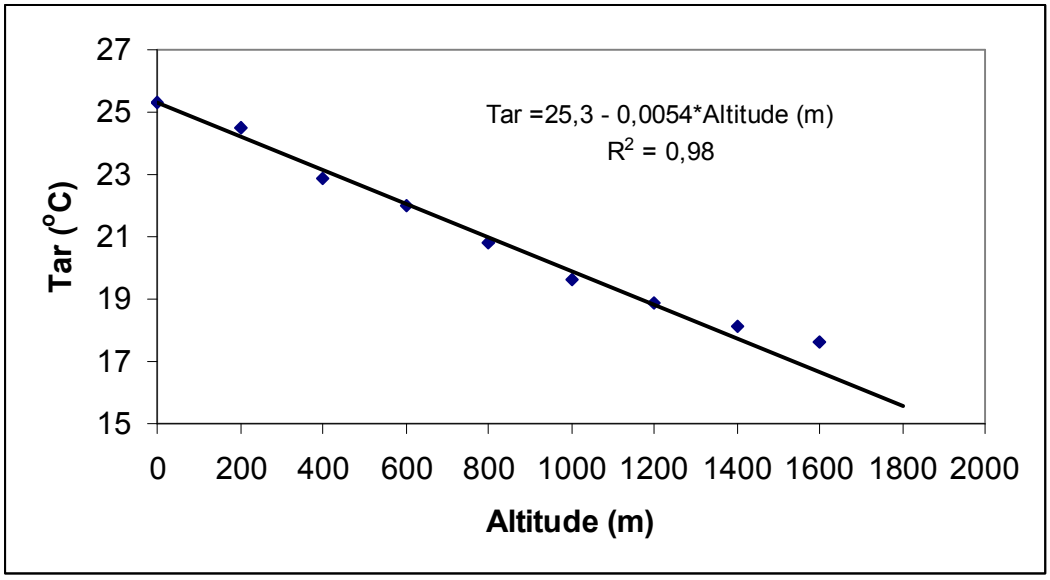

Figura 4. Variação da temperatura do ar em função da variação da altitude (exemplo hipotético, mas próximo da realidade).

Para determinação numérica de *a* e *b* faz-se uso das expressões:

$$b = \frac{\sum X.Y - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{\left(\sum X\right)^2}{n}}$$

$$a = \overline{Y} - b.\overline{X}$$

Em princípio, parece uma operação complicada. Contudo, fazendo uso do *Excel* e dos procedimentos abaixo se tem facilmente os valores de *a* e *b*. Exemplo: produza um gráfico de dispersão entre duas variáveis x e y. Em seguida, com o botão direito do *mouse* dê um clique sob os pontos no gráfico; em seguida dê um clique no botão direito do *mouse* e escolha adicionar linha de tendência. Selecione o tipo de reta de ajuste linear e em opção habilite com um clique a "opção" *exibir equação* e *valor do R-quadrado* e OK. Lembrando que a equação obtida será do tipo y = b.x + a. Para o exemplo da tabela 7 e figura 4 a equação da reta é:

Tar =25,3 - 0,0054.Altitude (m)

Assim, para valores de altitude onde não foram avaliados a temperatura do ar é possível estimar seu valor, por exemplo, na altitude de 1500m o valor de Tar, fazendo uso da equação, será de 17,2 ^{o}C. O uso da regressão linear permite, portanto, uma redução da

amostragem do trabalho de campo, permitindo maior rapidez e menor custo na obtenção dos dados.
Cabe lembrar que a regressão linear só se aplica quando os elementos em análise apresentam entre si uma relação de **dependência natural**.

6. Dígitos significativos e Arredondamento de dados

A precisão de medidas de dados contínuos sempre pode ser aprimorada melhorando-se o instrumento de medida. Por isso, os estatísticos fazem distinção entre dígitos significativos, que representam uma informação precisa, e dígitos que servem apenas para localizar a vírgula. Nos resultados, devem ser apresentados apenas os dígitos significativos, para evitar a falsa impressão de exatidão (VIEIRA, 1999). O resultado de um cálculo estatístico não deve conter mais dígitos significativos que os dados de menor precisão.
Por exemplo, um sensor de temperatura (termômetro de mercúrio) tem uma precisão de 0,2 ^{o}C. Isso significa que as leituras com precisão serão obtidas em intervalos de 0,2 em 0,2 ^{o}C e por extrapolação, poderia se chegar a leituras intermediárias de 0,1. Ao final, teremos uma tabela com intervalos de 0,2 ou 0,1 ^{o}C.

Tabela 8. Exemplo hipotético de valores de temperatura do ar.

hora	Temperatura do ar (^{o}C)
06h00min	18,2
08h00min	19,5
10h00min	20,6
12h00min	22,9
14h00min	24,5
média	21,14 = 21,1

A precisão centesimal não tem significado numérico algum para estas medidas, pois a precisão de medidas é de no máximo 0,1 ^{o}C. Assim, deve-se reduzir a precisão para o número de casas decimais compatíveis com a precisão do sensor que gerou o valor, ou seja, 21,1 ^{o}C.
O arredondamento dos dados deve seguir os seguintes critérios:

- Se você vai cortar dígitos e o resto é menor do que 5, apenas faça o corte;
- Se você vai cortar dígitos e o resto é maior do que 5, aumente o último número em uma unidade;
- Se você vai cortar dígitos e o resto é exatamente igual a 5, a convenção é:
    - Se o dígito anterior ao que vai ser cortado é par, apenas faça o corte;
    - Se dígito anterior ao que vai ser cortado é impar, aumente esse dígito em uma unidade.

Esta prática faz com que, ao longo das operações, os aumentos e reduções devidos aos arredondamentos sejam compensados.

Ao final deste capitulo o que se pretende não é elucidar todos os conceitos em sistematização de dados com uso da estatística, nem tampouco desencorajar o estudo mais aprofundado do tema. O que se espera é fornecer os conceitos mínimos para uma melhor formação de profissionais em áreas em que a “estatística” sempre é vista com preconceito, talvez pelo desconhecimento de suas potencialidades ou pela forma como lhe foi apresentada.

7. Bibliografia de Apoio

ASSIS, F.N. *Aplicações de Estatística à Climatologia: Teoria e Prática.* Pelotas: Ed. Universitária/UFPEL, 1996.

JOHNSTON, R.J. *Multivariate statistical analysis in Geography*. New York: Longman Scientific & Technical, 1986.

VIEIRA, S.. *Princípios de Estatística*. São Paulo: Pioneira, 1999.