

Lista nº 1 – Exploração de Dados Univariados

Os exercícios abaixo são relativos aos dados de estatísticas vitais de vários países, disponíveis no arquivo *vital.txt*. São projeções para 1996, baseadas em informações do U.S. Census Bureau's International Database.

Variáveis:

Country: país

type: grupo de países (1 = países classificados por agências das Nações Unidas como industrializados, exceto Japão; 2 = seleção de países da África; 3 = países do sudeste da Ásia mais China e Japão; 4 = países da Américas Central e do Sul).

mor.both: taxa de mortalidade infantil – homens e mulheres.

mor.male: taxa de mortalidade infantil – homens.

mor.fem: taxa de mortalidade infantil – mulheres.

life.both: expectativa de vida – homens e mulheres.

life.male: expectativa de vida – homens.

life.fem: expectativa de vida – mulheres.

births: nascimentos, por 1.000 pessoas.

deaths: mortes, por 1.000 pessoas.

migrants: migração líquida, por 1.000 pessoas.

Reveja o conteúdo sobre gráficos, estatísticas descritivas e exploração de dados univariados em qualquer livro de estatística da graduação (por exemplo, o capítulo 2 – Ferramentas para a Exploração de Dados Univariados, do livro *Encontros com o Acaso: um Primeiro Curso de Análise de Dados e Inferência*, de Christopher J. Wild e George A. F. Seber). Espera-se que, sempre que possível, você crie os gráficos apropriados aos dados e comente sobre qualquer coisa de interesse que observe sobre o padrão dos dados. Você é encorajado a sempre olhar os dados e estabelecer conjecturas a serem posteriormente verificadas formalmente como consequência dessa análise exploratória. O professor está disponível para tirar dúvidas no Departamento de Estatística, às terças e quintas de 15h00 às 16h00 ou às quartas-feiras, na FACC, de 13h00 às 14h00. Entregar (em papel) até 11/05/2017, às 16h00, no Departamento de Estatística.

1. Construa um gráfico de ponto rotulado das taxas de natalidade (coluna “ambos”) para os países industrializados. Construa também um gráfico de barras (use uma ordem razoável). O que essa informação nos gráficos lhe sugere, se é que sugere alguma coisa?
2. Repita o exercício anterior para a expectativa de vida dos países africanos.
3. Compare a expectativa de vida de homens e mulheres para os países industrializados, usando:
 - a. Histogramas.
 - b. Box-plots.
 - c. Diagrama de dispersão.
 - i. Plote a função identidade.
 - d. Médias.
4. Compare a expectativa de vida de homens e mulheres para os países industrializados, usando:
 - a. Histogramas.
 - b. Box-plot.
 - c. Diagrama de dispersão.
 - i. Plote a função identidade.

- d. Médias.
5. A partir dos dados, está claro que as expectativas de vida das mulheres tendem a ser maiores do que as expectativas de vida dos homens e você poderá perguntar quão maior.
- Considere a diferença $dif = \text{expectativa de vida das mulheres} - \text{expectativa de vida dos homens}$. Se um país tem valor 3 para dif , o que isso nos diz? O valor 0 é especial. Por quê? A que correspondem os valores maiores e menores do que 0?
 - Considere a razão: $\text{razão} = \text{expectativa de vida das mulheres} / \text{expectativa de vida dos homens}$. Se um país tiver valor 1,2 para razão , o que isso nos diz? O valor 1 é especial. Por quê? A que correspondem valores maiores e menores do que 1?
6. Sejam:
- Mortalidade infantil (ambos).
 - Mortalidade infantil feminina dividida por mortalidade infantil masculina.
 - Expectativa de vida (ambos).
 - A diferença entre expectativa de vida das mulheres e expectativa de vida de homens.
 - Taxa de natalidade.
 - Taxa de mortalidade.
- Aplique os passos seguintes às variáveis listadas acima [(a) – (f)]:
- Construa gráficos adequados que lhe permitam comparar os grupos de países.
 - Anote quaisquer aspectos interessantes que você observar nos gráficos.
 - Procure países que se destacam do seu grupo e rotule-os adequadamente nos gráficos.
 - Tente dar uma explicação possível para o que você observa em (ii) e (iii).
 - Que perguntas os gráficos o estimulam a fazer?
 - De que informação adicional você precisaria para obter as respostas às perguntas em (v)?
7. Baseie-se sempre que possível em gráficos e em estatísticas descritivas adequadas, responda às seguintes perguntas:
- A taxa de natalidade é definida como o número de nascimentos por 100.000 pessoas. Que fatores afetam a taxa de natalidade de um país?
 - As taxas de mortalidade de países industrializados parecem ser maiores em média do que aquelas dos países da Ásia e das Américas Central e do Sul? Você pode sugerir algumas razões para isso?
 - No grupo industrializado, a razão das taxas de mortalidade das mulheres para os homens é altamente incomum? Isso é por que as taxas de mortalidade infantil individuais são altamente incomuns?
 - Calcule a média das taxas de natalidade para os países do sudeste da Ásia (incluindo China e Japão). Que significado, se é que há um, tem esse número? Ele o informa sobre a taxa de natalidade média para toda região? Por quê, ou por que não? Suas respostas mudam se considerarmos a mediana em vez da média?
 - Um relatório da ONU de 1996 afirmou que as rendas médias de 70 países, na sua maioria do Terceiro Mundo, na década de 1990, eram

mais baixas do que em 1980, enquanto que 15 países do Terceiro Mundo, a maioria deles na Ásia, tinham alcançado crescimento econômico sustentado. A inferência direta é que globalmente as coisas pioraram. Um colunista sugeriu que 15 países que tinham progredido incluíam a China e a Índia, que sozinhas respondem por aproximadamente metade da população do planeta, então a imagem global era na verdade muita mais otimista do que havia sido pintada. Que medida você considera a mais relevante nesse caso?

- f. A variável *migrantes* refere-se à migração líquida, isto é, os que entram menos os que saem. Como você construiria uma variável *aumento* que fornecesse o crescimento líquido no tamanho da população (por mil) por ano para cada país?