**Seminário**

**Andrea Vieira Gonçalves**

Os dados são referentes a mortalidade infantil nos anos de 2000 e 2007 nas subprefeituras do município de São Paulo.

Sugestão para o seminário:

Sugere-se que na apresentação do seminário sejam executados os comandos um a um, com comentários e interpretação das saídas de cada comando.

Descrição do banco de dados

|  |  |
| --- | --- |
| Nome da variável | Descrição |
| regiao | Região geográfica da cidade |
| da | Distrito administrativo |
| subprefeitura | Subprefeitura |
| dacodigo | Código do distrito administrativo |
| sub | Código da subprefeitura |
| n | Sequência numérica para identificar o registro |
| mi2000 | Mortalidade infantil no ano de 2000 |
| mi2007 | Mortalidade infantil no ano de 2007 |
| neotardia2007 | Mortalidade neonatal tardia no ano de 2007 |
| neoprecoce2007 | Mortalidade neonatal precoce no ano de 2007 |
| posneo2007 | Mortalidade posneonatal no ano de 2007 |
| ado2007 | % de mães adolescentes no ano de 2007 |
| baixopeso2007 | % de baixo peso ao nascer no ano de 2007 |
| tmcausasexternas2007 | Taxa de mortalidade por causas externas em 2007 |
| tmcausasexternas2000 | Taxa de mortalidade por causas externas em 2007 |
| neoprecoce2000 | Mortalidade neonatal precoce no ano de 2000 |
| neotardia2000 | Mortalidade neonatal tardia no ano de 2000 |
| posneo2000 | Mortalidade posneonatal no ano de 2000 |
| consprenatal2007 | % mulheres que realizaram 7 ou mais consultas de pré-natal |

Sugestão de análise

1. Descreva o banco de dados e a natureza de cada variável

Comando: desc

1. Identifique e recodifique os valores faltantes (“missings”)

Comandos: tab, recode, list

1. Crie uma nova variável numérica para a variável *string* regiao

A nova variável (reg) deverá ser codificada como 1 – Sul, 2 – Norte, 3 – Leste, 4 – Oeste e 5 - Centro-oeste

1. Renomeie as variáveis:
   1. ano2000 🡪 mortalidade2000
   2. ano2007 🡪 mortalidade2007
   3. consprenatal2007 🡪prenat2007
   4. ado2007 🡪 madolescente2007
2. Descreva as variáveis ado2007, baixopeso2007 e consprenatal2007 por meio de tabelas e apresente as estimativas das estatísticas: valor médio, mediano, desvio padrão, quartis e valores mínimo e máximo

Comandos: sum, d

1. Faça o box plot para as variáveis das variáveis mortalidade, posneo2007, ado2007, baixopeso2007 e consprenatal2007. Investigue a existência de valores outlier
2. Faça o histograma para as variáveis posneo2007, ado2007, baixopeso2007 e consprenatal2007 investigando a normalidade das distribuições.
3. Investigue a existência de normalidade para as variáveis posneo2007, ado2007, baixopeso2007 e consprenatal2007 e indique quais transformações dos dados seriam úteis para normalizar a distribuição.

Comandos: ladder

1. Faça o gráfico de dispersão entre as variáveis
   1. ano2007 x prenat2007
   2. ano2007 x ado2007
   3. ano2007 x madolescente2007
   4. ano2007 x baixopeso2007

Comando: graph, save graph

1. Apresente o coeficiente de correlação de Pearson/Spearman para as variáveis
   1. ano2007 x prenat2007
   2. ano2007 x ado2007
   3. ano2007 x madolescente2007
   4. ano2007 x baixopeso2007

Comando: pwcorr

1. feche o arquivo log