Seminário

Kawni Romero Galante

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

Informações sobre o banco de dados

Os dados se referem a 231 mulheres que tiveram seus filhos em determinada maternidade no ano de 2014.

As variáveis que compõem o banco são:

|  |  |  |
| --- | --- | --- |
| Nome da variável | Detalhamento | Códigos |
| id | Número de identificação da participante |  |
| sexo | Sexo do recém-nascido | 1-masculino  2-feminino |
| peso | Peso ao nascer (g) |  |
| tempo | Tempo de gestação (semanas) |  |
| gestprevia | Número de gestações anteriores |  |
| idade | Idade da mãe no momento do nascimento |  |

Sugestão de análise do banco

1. Abrir o arquivo no Excel e utilizar o *software Transfer* para transformar o arquivo xls em dta
2. No Stata abrir o banco dta
3. Criar um arquivo log
4. Descrever as variáveis
5. Transformar a variável sexo de formato **string** para formato **numérico**
6. Construir um label (rótulo) para a nova variável sexo
7. Fazer uma tabela de frequência das variáveis
8. Construir um box plot para a variável tempo e investigar se existem valores aberrantes
9. Resumir as variáveis quantitativas (comando sum) apresentando as medidas de tendência central e dispersão
10. Apresentar as variáveis peso, tempo e idade em gráficos apropriados (histogramas)
11. Avaliar se estas variáveis são distribuídas segundo uma distribuição normal (comando ladder)
12. Construir um rótulo para a nova variável peso ao nascer
13. Apresentar a variável tempo (duração da gestação) em duas categorias: abaixo de 37 semanas (nascimento pré-termo) e 37 semanas e mais (a termo)
14. Construir um rótulo para a nova variável
15. Apresentar a tabela cruzada entre as variáveis sexo e tempo em duas categorias, investigar se existe associação entre as variáveis
16. Construir uma nova variável número de gestações prévias com quatro categorias e construir rótulos para as categorias
17. Apresentar a tabela cruzada entre a variáveis tempo de gestação (em duas categorias) e número de gestações prévias (em quatro categorias) e investigar se existe associação entre as variáveis utilizando o teste qui quadrado de Pearson
18. Investigar a relação entre as variáveis idade da mãe e tempo de gestação (diagrama de dispersão)
19. A relação é linear? Apresentar o coeficiente de correlação de Pearson. Fazer o teste de hipóteses e investigar se os dados são provenientes de uma população na qual o coeficiente de correlação é zero, ou seja, não existe correlação.
20. Investigar a relação entre as variáveis idade da mãe e peso ao nascer (diagrama de dispersão)
21. A relação é linear? Apresentar o coeficiente de correlação de Pearson. Fazer o teste de hipóteses e investigar se os dados são provenientes de uma população na qual o coeficiente de correlação é zero, ou seja, não existe correlação.
22. Fechar o arquivo log