

Especialização em Ciências de Dados - Lista 6

06 de fevereiro de 2019

Deverá ser entregue um script com os comandos utilizados e um arquivo (pdf) contendo:

- as hipóteses nula e alternativa;
- o tipo de teste que foi utilizado e o motivo da escolha;
- conclusões das avaliações realizadas.

Parte 1

Vamos trabalhar com o banco de dados **babies** do pacote **UsingR**.

- O banco de dados consiste na observações de diversas variáveis de cada mãe em um pesquisa.
- Contém 1236 observações de 23 variáveis.
- Este banco de dados foi retirado de <http://www.stat.berkeley.edu/users/statlabs/labs.html>.
- Para mais informações use o comando `?babies`.

1. Vamos preparar o banco de dados.

(a) Importe o banco de dados **babies**, não esqueça que o banco de dados está dentro do pacote **UsingR**. Das 23 variáveis, vamos utilizar apenas 4:

- **gestation**: tempo de duração da gestação (em dias, 999 = dado faltante);
- **wt**: peso do recém nascidos (em *ounce*, 999 = dado faltante);
- **age**: idade da mãe no final da gravidez (em anos, 99 = dado faltante);
- **smoke**: se a gestante é fumante (0 = não, 1 = sim, 2 = fumou até a gravidez atual, 3 = ex fumante ou 9 = dado faltante).

(b) Construa um banco novo banco de dados com apenas as informações destas 4 variáveis.

(c) Note que temos algumas observações faltantes no banco de dados. Retire as gestantes com observações faltantes do banco de dados.

2. Primeiramente, vamos estudar o tempo de gestação. Suponha que estamos interessado na média do tempo de gestação.

(a) Estudo prévios indicam que o tempo médio de gestação é de 280 dias. Podemos confirmar esta hipótese?

(b) Agora vamos considerar a média do tempo de gestação de gestantes fumantes. Para isto, crie o vetor **gest_fumantes** contendo as observações do tempo de gestão das gestantes cuja variável **smoke** é igual a 1.

(c) Podemos afirmar que o tempo médio de gestação de gestantes fumantes é menor que 280 dias?

(d) Fumar durante a gestação aumenta as chances do bebê nascer antes do tempo. Compare o tempo médio de gestação entre as duas amostras de gestantes, que fumaram durante a gestação e as que não fumaram. O que podemos concluir?

3. Outra consequência possível do tabagismo durante a gestação é o aumento da chance do bebê nascer com baixo peso. Faça um estudo comparando a média de peso do recém nascidos entre as duas amostra de gestantes (fumantes e não fumantes). O que podemos concluir?

Parte 2

Considere o banco de dados `exec.pay` do pacote `UsingR`. O banco de dados contém o salário (em 10 mil dólares) de CEOs de 199 companhias nos Estados Unidos.

1. Um pesquisa quer avaliar se a mediana dos salários é maior do que 220 mil dólares.
 - (a) Defina as hipóteses do problema.
 - (b) Realize o teste de hipótese adequado.
 - (c) O que podemos concluir?
2. Note que os dados `exec.pay` é assimétrico, realize a transformação `log()` nos dados e verifique que a nova variável aparenta ser simétrica. Repita a avaliação utilizando o teste mais adequado.

Parte 3

Agora considere o banco de dados `shoes` do pacote `MASS`. O banco de dados consiste numa `list` de dois vetores, dando o desgaste de sapatos de materiais A e B para um em cada pé de dez meninos. Queremos avaliar se um material é melhor que o outro usando como critério o desgaste dos sapatos.

1. Defina as hipóteses do problema.
2. Realize o teste de hipótese adequado.
3. O que podemos concluir?