

Atividade Prática no RStudio n. 1 – Valor: 5 pontos

Providências preliminares

Instale e importe o pacote AER no Rstudio. Para isso, utilize o comando `install.packages("AER")`. Em seguida, abra o pacote com `library(AER)`.

Descrição do caso.

Você foi contratado por um colégio para realizar um estudo sobre a relação entre o número de alunos em sala de aula e a pontuação média dos alunos de cada turma no ENEM. O colégio lhe forneceu a seguinte tabela de dados:

Número de alunos na turma	Pontuação Média
23	430
19	430
30	333
22	410
23	390
29	377
35	325
36	310
33	328
25	375

Para gerar um relatório que atenda às demandas do colégio, faça no RStudio as seguintes análises, e gere um arquivo RMarkdown em formato HTML, contendo:

1. Um gráfico de dispersão cujo eixo x (horizontal) é o número de alunos na turma e o eixo y (vertical) é a pontuação média.
2. A média, a variância, e o desvio padrão da pontuação média dos alunos.
3. A covariância e a correlação entre o número de alunos e a pontuação média.
4. Os resultados de um modelo linear em que o número de alunos na turma é a variável que explica as notas médias por turma no ENEM;

$$MediaENEM_i = a + b \cdot TamTurma_i + \varepsilon_i$$

5. O gráfico de dispersão feito anteriormente, mas agora com a incorporação da reta da regressão do modelo na cor azul.
6. O R^2 do modelo com 2 casas decimais.

7. Uma tabela com os coeficientes estimados (intercepto e coeficiente angular), erros padrão, estatísticas t , e p -valores da variável explicativa e do intercepto do modelo.
8. Apresente a matriz de variâncias e covariâncias desses coeficientes. Dica nº 1: sempre que tiver dúvidas sobre uma função do R, a procure na documentação do software: <https://www.rdocumentation.org/>. Dica nº 2: a função “name” apresenta quais informações podem ser extraídas de outras funções (como a função summary por exemplo).

```
> names(summary(mod))
[1] "call"          "terms"          "residuals"      "coefficients"
[5] "aliased"        "sigma"          "df"             "r.squared"
[9] "adj.r.squared" "fstatistic"     "cov.unscaled"
> |
```

9. Com base no valor estimado do intercepto e do seu erro padrão, mostre a relação dos mesmos com a fórmula da estatística t e efetue um teste de hipóteses (lembre-se da primeira lista de exercícios) investigando se o valor do intercepto é estatisticamente igual a zero ou se ele é diferente de zero. Apresente o p -valor desse teste de hipóteses. É possível rejeitar a hipótese nula a um nível de significância de 5%? Justifique.
10. Repita o procedimento da questão acima para a variável explicativa “Número de Alunos na Turma”. Apresente o p -valor do teste de hipóteses. É possível rejeitar a hipótese nula a um nível de significância de 5%?

O modelo elaborado acima utiliza uma variável explicativa (número de alunos em sala de aula) e um intercepto (isto é, o valor da variável dependente quando a variável independente é igual a zero). Em alguns casos pode ser necessário elaborar um modelo sem intercepto para garantir que a variável dependente assumirá valor zero quando a variável independente também for zero.

11. Parta da mesma base de dados utilizada acima e elabore um modelo sem intercepto e apresente as principais informações do modelo (função **summary**).
12. Insira, na cor vermelha e no mesmo gráfico feito anteriormente, a reta de regressão do modelo sem intercepto.
13. Compare os coeficientes angulares dos dois modelos. Qual é a diferença entre os modelos? E qual modelo você utilizaria para fazer inferências sobre a relação entre n. de alunos e nota no ENEM? Justifique sua resposta em termos estatísticos. (Dica: pesquise na bibliografia sobre modelos de regressão sem intercepto ou modelos de regressão que passam pela origem).