

**Universidade Federal do Paraná – Departamento de Estatística**  
**Disciplina CEO71 – Análise de Regressão Linear**  
**Prof. Cesar Augusto Taconeli**  
**Lista – RLM (Parte 2)**

O arquivo `atletas.csv`, disponível na página da disciplina, contém dados de 200 atletas de determinada modalidade esportiva. As variáveis que compõem a base são as seguintes:

`idade`: Idade da atleta (em anos);  
`imc`: Índice de massa corporal ( $\text{kg}/\text{cm}^2$ );  
`horas`: Horas de treino semanal;  
`massa`: Percentual de massa gorda;  
`gasto`: Gasto semanal com suplementos alimentares (reais);  
`estudo`: Anos de estudo;  
`escore`: Escore de desempenho nas últimas provas (numa escala de 0 a 100).

- a) Faça uma análise exploratória dos dados. Construa gráficos uni e bivariados, calcule medidas de locação e dispersão para cada variável, obtenha as correlações;
- b) Ajuste um modelo de regressão linear múltipla tomando o escore de desempenho como a variável resposta. Escreva a expressão do modelo ajustado;
- c) Qual o valor ajustado pelo modelo para a primeira atleta da base? Qual o resíduo correspondente?

	escore	idade	imc	horas	massa	gasto	estudo
1	79.6	30	20.5	19	21.66	72	15

- d) Qual a probabilidade estimada de um atleta com características semelhantes à 1ª atleta da base ter escore de desempenho superior a 80? E entre 65 e 75?
- e) Quais variáveis são estatisticamente significativas ao nível de 5% de significância?
- f) Qual a variação esperada no escore de desempenho para um aumento de 1% de massa gorda, mantendo fixas as demais variáveis?
- g) Qual a variação esperada no escore de desempenho para 1 ano a mais de idade, mantendo fixas as demais variáveis?
- h) Qual a variação esperada no escore de desempenho para 20 reais semanais a mais gastos com suplementos alimentares?
- i) Obtenha intervalos de confiança (95%) para os parâmetros do modelo;
- j) Com base nos intervalos de confiança obtidos, você rejeitaria, ao nível de significância de 5%, a hipótese de não efeito da massa muscular no desempenho dos atletas? Justifique sua resposta;
- k) Ainda com base nos intervalos de confiança, você rejeitaria a hipótese de que o desempenho caia, em média, dois pontos para 1% a mais de massa gorda? Justifique.
- l) Obtenha a elipse de 95% de confiança para os parâmetros referentes aos efeitos de massa gorda e idade. Adicionalmente, obtenha intervalos de confiança simultâneos, com nível de confiança conjunto igual a 95%. Represente num único gráfico a região de confiança e os intervalos de confiança 95% simultâneos e individuais;
- m) Obtenha o quadro da análise de variância e conduza o teste da hipótese nula de não significância do modelo de regressão;

- n) Qual a estimativa para a variância dos erros aleatórios?
- o) Obtenha o coeficiente de determinação e o coeficiente de determinação ajustado. Interprete-os;
- p) Ajuste um novo modelo de regressão, desta vez desconsiderando a variável massa gorda. Compare os resultados dos dois ajustes. O que houve com o efeito da variável *imc*? Justifique.
- q) Voltando ao modelo original, teste a significância conjunta das variáveis *imc* e *estudo*. De acordo com o resultado do teste, poderíamos excluir ambas as variáveis do modelo ao nível de 5% de significância?
- r) Considere os seguintes perfis de atletas:

Perfil	Idade	IMC	Horas	Massa	Gasto
1	20	20	20	17	100
2	28	23	13	22	60
3	25	21	15	20	85

Para cada um dos perfis apresente um intervalo de confiança (95%) para o escore médio de desempenho e um intervalo de predição (95%) para o escore de um novo atleta com as características apresentadas;

- s) Faça um gráfico de escores observados versus escores ajustados pelo modelo. Acrescente ao gráfico a reta identidade;
- t) Faça um gráfico de resíduos versus escores ajustados. Como os resíduos estão distribuídos?
- u) Padronize cada uma das variáveis explicativas e ajuste um novo modelo de regressão, substituindo as variáveis originais pelas padronizadas. Qual das variáveis apresenta maior efeito no escore de desempenho? Qual a interpretação das estimativas obtidas?