

## Coeficiente de correlação linear de Pearson e regressão linear

Professor Fiore

Atividade 1 – Esta atividade tem por objetivo treinar o cálculo do coeficiente de correlação linear de Pearson e da equação de regressão linear. Para cada tabelas abaixo, calcule o coeficiente de correlação de Pearson fale sobre o resultado, indicando a qualidade da correlação e se ela é positiva ou negativa. Para os casos de correlação forte, encontre a equação de regressão.

Tabela 1					
x	1	2	2	5	7
y	12	10	8	7	5

Tabela 2						
x	1	2	2	5	7	4
y	5	2	8	10	1	2

Tabela 3				
x	10	15	12	7
y	8	10	10	8

Problema 1 – Este problema tem por objetivo mostrar uma aplicação simples e eficaz do coeficiente de correlação linear e da equação de regressão.

Uma empresa de máquinas pesadas, deseja estudar melhor a relação entre quantidade produzida e o custo de produção. A tabela ao lado mostra a quantidade de máquina do tipo A produzidas a cada mês e o custo de produção mensal, em milhares de reais.

Antes de elaborar um relatório sobre a relação entre o custo e a quantidade produzida responda as questões a seguir.

Mês	Quantidade	Custo mensal (milhar de reais)
Janeiro	12	95
Fevereiro	15	115
Março	12	100
Abril	10	84
Mai	13	105
Junho	12	102
Julho	18	142
Agosto	15	120

1. Você acha que existe relação de causa e efeito entre as grandezas? Explique.
2. Monte um diagrama de dispersão para os dados da tabela.
3. Você acha que existe uma boa correlação entre as grandezas envolvidas? Ela pode ser linear? Explique.
4. Calcule o coeficiente de correlação linear de Pearson para os dados e explique o resultado.
5. Encontre a equação da reta interpoladora ou a equação de regressão pelo método dos mínimos quadrados.
6. Estime o custo para a fabricação de 23 máquinas.
7. Estime a quantidade de máquinas possíveis de se fabricar com um investimento de 150 mil.
8. Explique quais são as vantagens em se usar métodos matemáticos de forma adequada, para estimar valores futuros.

Problema 2 – Esta atividade tem por objetivo mostrar que algumas situações possuem correlação não linear. Também mostra que, mesmo quando há um bom coeficiente de correlação, é preciso cuidado.

Em um experimento foram coletados o tempo de queda livre de uma bolinha em intervalos de 0,20 metros e os resultados estão na tabela abaixo.

Altura (metros)	Tempo de queda (segundos)
0	0,00
0,2	0,20
0,4	0,29
0,6	0,35
0,8	0,41
1	0,45

1. Você acha que existe relação de causa e efeito entre as grandezas? Explique.
2. Explique com base em conhecimentos de física básica, porque neste caso a correlação linear não é a mais adequada. E diga qual tipo de correlação melhor descreve este caso. Se necessário, desenhe o diagrama de dispersão.
3. Mesmo sabendo que a correlação linear não é a adequada, calcule o coeficiente de correlação linear de Pearson para o caso e determine a equação de regressão linear. ( $r = 0,9545$  e  $y^* = 0,42x + 0,073$ )
4. A fim de verificar o erro gerado pela equação de regressão para o caso, por não ser uma situação linear, calcule o tempo estimado com a equação de regressão, para queda livre de 10 metros e compare com o tempo real estimado em aproximadamente 1,43 segundos.
5. Explique quais são os riscos ao se usar a equação de regressão linear em um caso não linear?

Atividade estilo ENADE – Esta atividade tem por objetivo ajudar você a aplicar em problema real e mais elaborado os conceitos de coeficiente de correlação linear.

1. Conforme estudado em aula, o coeficiente de correlação linear de Pearson pode ser usado para estudar possíveis relações de causa e efeito entre grandezas, em diversas áreas da ciência. E você como engenheiro, deve estar preparado para atuar nessas áreas, pois os equipamentos projetados e construídos pelos engenheiros são usados em inúmeras áreas, como a medicina.

Sendo assim, considere a situação a seguir retirada da revista brasileira de medicina do esporte, onde há o artigo “Medidas murinométricas e eficiência alimentar em ratos provenientes de ninhadas reduzidas na lactação e submetidos ou não ao exercício de natação”, pelos autores Cybelle da Silva Nery; Isabeli Lins Pinheiro; Gisélia de Santana Muniz; Diogo Antônio Alves de Vasconcelos; Sabrina Pereira de França e Elizabeth do Nascimento.

Em resumo o estudo analisa se o “excesso de alimentação no início da vida pode modificar persistentemente consumo e peso corporal.” Também considera “a adoção de exercício físico como uma estratégia útil para evitar excessivo ganho de peso.” Por fim, ele conclui que “a redução da ninhada no aleitamento não alterou o peso corporal ou ingestão alimentar persistentemente. Entretanto, o protocolo de natação foi eficaz em reduzir o ganho de peso em animais controles, mas não naqueles de ninhada reduzida”.

Fonte: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1517-86922011000100010](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-86922011000100010)

Ao longo do artigo há uma tabela que indica a correlação entre o IMC ( $g/cm^2$ ) e o índice de Lee ( $g/cm^3$ ), dois índices usados para verificar a relação entre massa corporal e peso nos ratos da ninhada estudada.

Com base no estudado sobre correlação linear de Pearson e analisando os dados do gráfico, é possível afirmar que:

a. O estudo da correlação (coeficiente de correlação de Pearson) entre os índices (IMC e Lee) revelaram uma **significante e positiva associação entre eles** (figura 4). Isto denota que tanto um quanto o outro pode ser utilizado como determinante da massa corporal e atuar como preditor do excesso de peso em ratos.

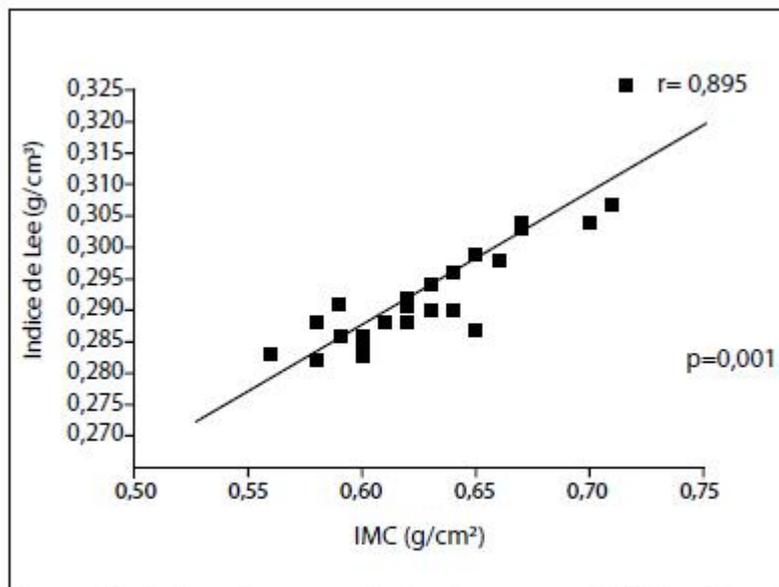


Figura 4. Estudo da correlação entre o índice de massa corporal (IMC) e o índice de Lee de ratos Wistar jovens (n = 25). Teste de correlação de Pearson ( $p = 0,001$ ).

b. O estudo da correlação (coeficiente de correlação de Pearson) entre os índices (IMC e Lee) revelaram uma **fraca associação entre eles** (figura 4), pois o valor é menor que um. Isto denota que tanto um quanto o outro pode ser utilizado como determinante da massa corporal e atuar como preditor do excesso de peso em ratos.

c. O estudo da correlação (coeficiente de correlação de Pearson) entre os índices (IMC e Lee) revelaram uma **significante e negativa associação entre eles** (figura 4). Isto denota que tanto um quanto o outro pode ser utilizado como determinante da massa corporal e atuar como preditor do excesso de peso em ratos.

d. O estudo da correlação (coeficiente de correlação de Pearson) entre os índices (IMC e Lee) revelaram uma **insignificante associação entre eles** (figura 4). Isto denota que tanto um quanto o outro pode ser utilizado como determinante da massa corporal e atuar como preditor do excesso de peso em ratos.

e. O estudo da correlação (coeficiente de correlação de Pearson) entre os índices (IMC e Lee) revelaram uma **associação dispersiva entre eles** (figura 4). Isto denota que tanto um quanto o outro pode ser utilizado como determinante da massa corporal e atuar como preditor do excesso de peso em ratos.

Outros exemplos – O objetivo dos tópicos abaixo é permitir que você reflita em mais exemplos de aplicação para coeficiente de correlação linear e equação de regressão.

A relação entre o tempo de sono e o número de erro na montagem de peças

A relação entre a humidade de papel e o número de vincos por m<sup>2</sup>

A relação entre o tempo semanal de estudo e a nota em uma prova

A relação entre a frequência semanal de estudo e a nota em uma prova.

A relação entre a produção por pés de determinada fruta e a distância entre os pés.

A relação entre o tempo de filtragem da água e a quantidade de resíduos remanescentes