

Regressão Linear

Jaqueline V Moreira

12/04/2025

O exercício

Em um estudo de produção utilizado por uma fábrica, dois indicadores de controle de umidade estão sendo avaliados quanto à qualidade e permanência na cartela de medidas de controle. Durante 10 dias, mediu-se um indicador de umidade de uma amostra de material produzido e da umidade relativa no local de produção do mesmo material. As medidas estão apresentadas, em percentuais, na tabela a seguir:

No local	46	53	37	42	34	29	60	44	41	48
No material	12	14	11	13	10	8	17	12	10	15

Com base em seus conhecimentos, proceda no ajuste, nos testes relativos e na interpretação de um modelo de regressão linear simples, considerando $\alpha = 5$.

Correlação

O coeficiente de correlação r - ou correlação de Pearson - é uma métrica estatística que quantifica a “força” da relação linear entre duas variáveis numéricas. o coeficiente r possui domínio de -1 a 1, sendo:

- 0 = Nenhuma correlação linear
- 1 = correlação linear positiva perfeita, onde ambas as variáveis seguem a mesma direção.
- -1 = correlação linear negativa perfeita, onde as variáveis seguem direções opostas. (enquanto uma cresce a outra decresce)

Cálculo do coeficiente r

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

sendo:

x_i, y_i = dados observados

\bar{x}, \bar{y} = média amostral de X e Y

De acordo com o exercício proposto, a variável independente X é “umidade relativa no local” e a variável dependente Y é a “umidade do material”.

Cálculo manual através do excel

Cálculo com o RStudio

```
x <- c(46, 53, 37, 42, 34, 29, 60, 44, 41, 48)
y <- c(12, 14, 11, 13, 10, 8, 17, 12, 10, 15)

cor(x, y)
```

```
## [1] 0.9317929
```

De acordo com o coeficiente r obtido, considera-se que há uma relação linear positiva forte entre a umidade relativa no local e a umidade do material.

gráfico

Regressão Linear

A regressão linear é um modelo estatístico que tem como objetivo quantificar o poder preditivo de uma variável explicativa sobre uma variável resposta. Em outras palavras, ela busca medir o quanto uma variável independente é capaz de explicar o comportamento observada na variável dependente.

Como dados de saída de um modelo de regressão linear têm-se:

- a formula da equação da reta que melhor se ajusta aos dados (construída através do método dos minimos quadrados)
- valor de r^2 Coeficiente que indica o quão bem alinhada aos dados é a reta construída - fitting line ratio.

$$\text{Equação básica da reta} = Y = b_0 + b_1x + e$$

onde: b_1 = coeficiente angular da reta, indica quanto a variável Y varia em relação ao aumento de uma unidade da variável X . No caso concreto, quanto a umidade no material varia a cada 1% de aumento na umidade do local.

e = erro ou residuo do modelo. Representa a parte de Y que não é explicada por X .

$$e = Y_i - Y_{\text{predito pela reta}}$$

Cálculo com o RStudio

```
x <- c(46, 53, 37, 42, 34, 29, 60, 44, 41, 48)
y <- c(12, 14, 11, 13, 10, 8, 17, 12, 10, 15)

reg_linear = lm(y~x)

summary(reg_linear)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.54457 -0.70725 -0.00041  0.50263  1.54376
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.34765    1.66421   0.209    0.84
## x           0.27310    0.03761   7.261 8.72e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.023 on 8 degrees of freedom
## Multiple R-squared:  0.8682, Adjusted R-squared:  0.8518
## F-statistic: 52.72 on 1 and 8 DF,  p-value: 8.716e-05
```

De acordo com os coeficientes encontrados pelo RStudio, a formula da reta que melhor descreve os dados é:

$$Y = 0,34765 + 0,27310x$$

O valor do coeficiente r^2 (Multiple R-squared) é 0,8682, dessa forma, considera-se que mais de 86% da variação da unidade do material (Y) é explicada pela unidade do local (x).

Teste de Hipotese

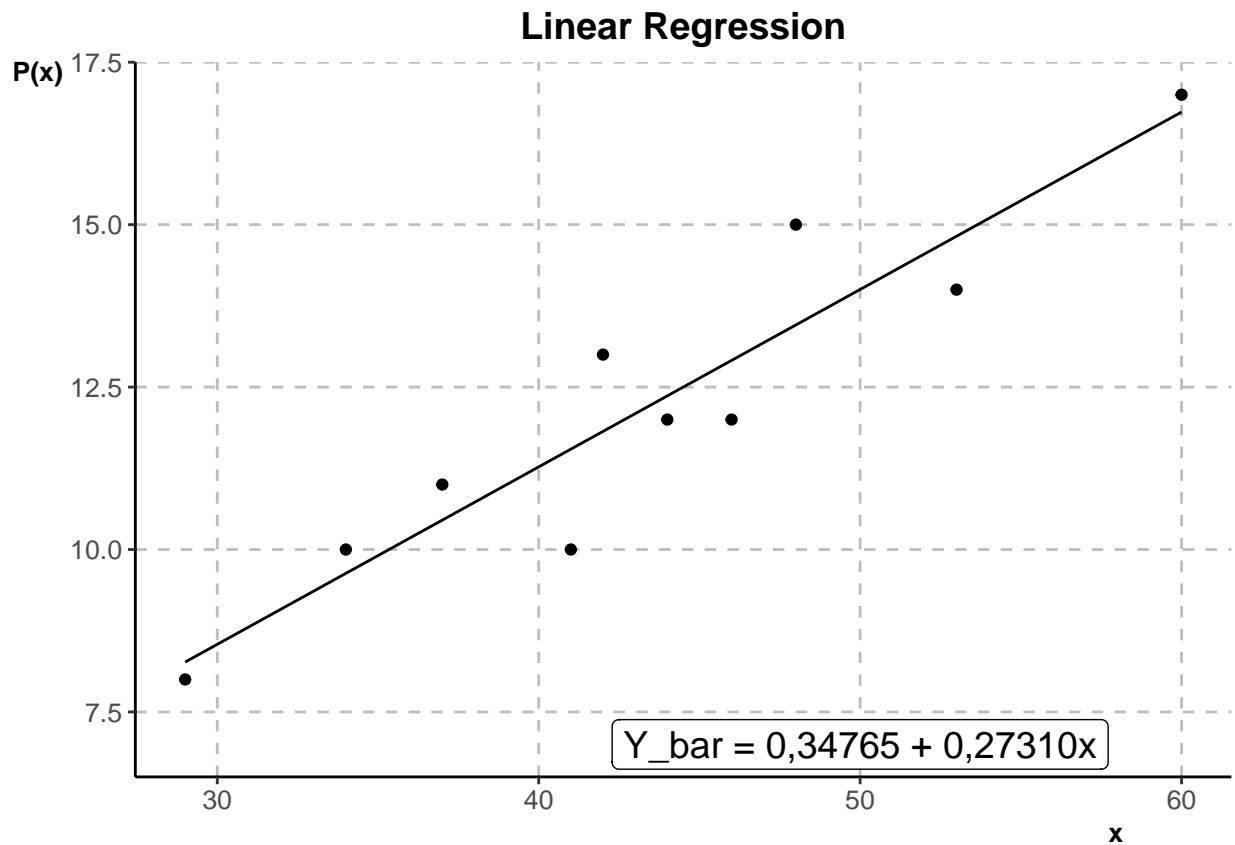
Partindo da premissa mde que H_0 deve ser definida como a hipótese neutra que assume a ausência de efeito ou diferença, considera-se:

$$h_0 = \text{não há relação linear}$$

$$h_1 = \text{há relação linear}$$

Segundo o enunciado, o valor de alfa a ser considerado no teste de significancia é de 5%. Assim, se o p-valor do modelo for inferior a 5% têm-se evidencia suficienete para rejeitar a hipotese nula e assumir a hipotese alternativa.

De acordo com os resultados obtidos pelo RStudio, o p-valor do modelo foi de 8.716e-05. Como $8.716e-05 < 5$, considera-se que há relação linear entre as variáveis X e Y.



Conclusão

Conforme os resultados obtidos, conclui-se que há uma forte correlação entre as variáveis X e Y apresentadas pelo enunciado, sendo possível concluir que a variável explicativa “umidade relativa local” é um bom preditor para a variável resposta “umidade do material” com mais de 86% de Y sendo explicada por X e um p-valor muito inferior ao valor de alfa estabelecido.