MODELAGEM ESTATÍSTICA

AULA 5

Prof. Guilherme Augusto Pianezzer

CONVERSA INICIAL

Vejamos como expandir os conceitos de regressão linear simples para a regressão linear múltipla. A diferença entre cada um desses modelos é que na simples utilizamos apenas uma variável de controle, enquanto na múltipla podemos utilizar duas ou mais variáveis explicativas.

Como motivação, usaremos o seguinte exemplo nessa aula: considere um indivíduo interessado em aumentar sua massa muscular a partir de um incremento na quantidade de comida ingerida; ele decide participar de uma pesquisa científica de um novo remédio que objetiva aumentar seu apetite. Após tomar determinada dose de remédio e ficar algumas horas em jejum, o indivíduo se alimenta e tem sua quantidade de comida ingerida registrada. Vejamos se existe relação com o tempo em jejum após ingerir o medicamento e a dose de remédio administrada com a quantidade de comida ingerida.

TEMA 1 – MODELO ESTATÍSTICO

O modelo que estamos tentando tratar nesse momento refere-se à regressão linear múltipla com duas variáveis explicativas (tempo em jejum após ingerir medicamento e dose de remédio administrada) para analisar uma variável resposta (a quantidade de comida ingerida).

1.1 Modelo de Regressão Linear Múltipla

Como desejamos determinar o comportamento da variável resposta em função das duas variáveis explicativas, podemos definir o modelo de regressão linear múltipla como da forma

$$y = \alpha + \beta x_1 + \gamma x_2 + \epsilon$$

Note que, nesse modelo, y representa a variável resposta, x_1, x_2 representam as duas variáveis explicativas e ϵ representa a variabilidade do modelo resultado das variáveis que não foram consideradas na explicação do fenômeno. Nosso objetivo é determinar os parâmetros α, β, γ que configuram a linearidade do modelo.

Note também que, considerando $x_1 = x_2 = 0$ percebemos que o parâmetro α representa a resposta média obtida sem a influência dos parâmetros de controle. Se mantivermos a variável x_2 fixa, β representa o quanto

cada incremento na variável x_1 causa na variável y para aquele nível de x_2. De forma similar, se mantivermos a variável x_1 fixa, γ representa quanto cada incremento na variável x_2 causa na variável y para aquele nível de x_1.

Veja que $\alpha + \beta x_1 + \gamma x_2$ representa um plano, denominado *superfície de resposta*. Observe que o modelo de regressão linear múltipla pode ser facilmente generalizado para mais de duas variáveis de forma que:

$$y = \alpha + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n + \epsilon$$

1.2 Suposições para o modelo

Para aplicar o modelo de regressão linear múltipla de forma adequada é necessário supor algumas condições necessárias. Em primeiro lugar, é importante confirmar que o erro possui uma distribuição $N(0, \sigma^2)$, ou seja, que a variável erro não possui viés.

Essa suposição é importante, pois tendo viés na variável não considerada podemos afirmar que não estamos controlando todas as variáveis que são, de fato, adequadas para a explicação da variável resposta. Nesse caso, é importante realizar um novo planejamento do experimento. Em segundo lugar, é importante que as variáveis x_1, x_2 assumam valores fixos.

1.3 Exemplo

Como discutido no início, desejamos discutir o efeito de diferentes doses de medicamento administrada e o tempo em jejum após ingeri-lo na de comida. Dessa forma, consideramos a variável x_1 como o tempo em jejum após ingerir o medicamento e a medimos em minutos. Consideremos a variável x_2 como a dose de medicamento administrada e a medimos em mililitro. Por fim, consideramos a variável resposta y como a quantidade de comida ingerida e a medimos em gramas.

Para essa análise, fizemos um acompanhamento durante duas semanas, com o total de 14 observações que estão descritas na Tabela 1, a seguir.

Tabela 1 – Quantidade de comida ingerida em função do tempo em jejum e da dose de medicamento administrada

Observação	Tempo em jejum após ingerir medicamento $(min.)$	Dose de medicamento administrada $(ml.)$	Quantidade de comida ingerida $(g.)$
1	195	4	1.004
2	255	4	1.636
3	195	4,6	852
4	255	4,6	1.506
5	225	4,2	1.272
6	225	4,1	1.270
7	225	4,6	1.269
8	195	4,3	903
9	255	4,3	1.555
10	225	4	1.260
11	225	4,7	1.146
12	225	4,3	1.276
13	225	4,72	1.225
14	230	4,3	1.321

Usaremos esses dados como exemplo para compreender o modelo de regressão linear múltipla.

TEMA 2 – ESTIMAÇÃO DOS PARÂMETROS DO MODELO

Assim como no modelo de regressão linear simples, nosso objetivo é minimizar o quadrado do erro realizado pela estimativa. Lembre-se que escolhemos utilizar o termo quadrático, caso contrário, o somatório resultaria em zero e não conseguiríamos analisá-lo adequadamente. O método que analisamos aqui também é chamado de *Método dos Mínimos Quadrados*. Entretanto, aqui consideramos uma generalização que se torna mais realizável quando levamos em conta a forma matricial de resolução.

2.1 Método dos Mínimos Quadrados

Para uma determinada observação, podemos utilizar o modelo desenvolvido para encontrar uma estimativa para esse resultado. Dessa forma,

$$\hat{y}_i = \alpha + \beta x_{1i} + \gamma x_{2i}$$

Por exemplo, espera-se que, no caso em que $i = 1$, $x_{11} = 195$ e $x_{21} = 4$, temos que $y_1 = 1.004$. Entretanto, por se comportar de forma linear, o modelo estimado não consegue prever de forma adequada os próprios valores observados. Caso isso fosse possível, só precisaríamos determinar a reta que passa por todos os pontos pelos métodos simples de matemática básica. Como

isso não é possível, desejamos minimizar a variação (do quadrado) dos valores observados com o valor previsto de forma que possamos encontrar a melhor reta que descreve esses pontos. Assim, a variação é dada por

$$e_i = y_i - \widehat{y_i}$$
$$e_i = y_i - \alpha - \beta x_{1i} - \gamma x_{2i}$$

Note que o somatório dos e_i ao quadrado é dado por

$$L = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_{1i} - \gamma x_{2i})^2$$

Como gostaríamos de encontrar quais parâmetros α, β e γ minimizam L, devemos encontrar os pontos críticos dados por

$$\begin{cases} \dfrac{\partial L}{\partial \alpha} = 0 \\ \dfrac{\partial L}{\partial \beta} = 0 \\ \dfrac{\partial L}{\partial \gamma} = 0 \end{cases}$$

Assim, derivando cada uma das expressões, obtemos

$$\begin{cases} \dfrac{\partial L}{\partial \alpha} = -2 \displaystyle\sum_{i=1}^{n} (y_i - \alpha - \beta x_{1i} - \gamma x_{2i}) = 0 \\ \dfrac{\partial L}{\partial \beta} = -2 \displaystyle\sum_{i=1}^{n} (y_i - \alpha - \beta x_{1i} - \gamma x_{2i}).x_{1i} = 0 \\ \dfrac{\partial L}{\partial \gamma} = -2 \displaystyle\sum_{i=1}^{n} (y_i - \alpha - \beta x_{1i} - \gamma x_{2i}).x_{2i} = 0 \end{cases}$$

Ou seja,

$$\begin{cases} \displaystyle\sum_{i=1}^{n} y_i = n\alpha + \beta . \sum_{i=1}^{n} x_{1i} + \gamma \sum_{i=1}^{n} x_{2i} \\ \displaystyle\sum_{i=1}^{n} y_i . x_{1i} = \alpha \sum_{i=1}^{n} x_{1i} + \beta \sum_{i=1}^{n} x_{1i}^2 + \gamma \sum_{i=1}^{n} x_{1i} . x_{2i} \\ \displaystyle\sum_{i=1}^{n} y_i . x_{2i} = \alpha \sum_{i=1}^{n} x_{2i} + \beta \sum_{i=1}^{n} x_{1i} . x_{2i} + \gamma \sum_{i=1}^{n} x_{2i}^2 \end{cases}$$

Assim como fizemos na aula de regressão linear simples, podemos resolver esse sistema de equações para encontrar uma expressão para os valores de α, β e γ. Entretanto, quando desejar generalizar esse método para mais de duas variáveis, perceberemos que será uma tarefa muito trabalhosa escrever a forma explícita da solução. Nesse caso, partiremos para a representação matricial desse sistema de equações.

2.2 Representação matricial do Método dos Mínimos Quadrados

Para realizar a representação matricial, definimos as seguintes matrizes:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}, A = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}, \mathrm{E} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Note que a matriz Y representa a matriz contendo a resposta observada em cada situação. A matriz X guarda os valores medidos das variáveis explicativas. A representa a matriz com os parâmetros que precisam ser determinados no modelo, enquanto E representa a matriz contendo os erros de cada observação. Assim, o mesmo modelo discutido na seção anterior pode ser reescrito como

$$Y = XA + \mathrm{E}$$

Também desejamos aqui minimizar o erro quadrático de todas as observações. Assim, redefinimos L como

$$L = \sum_{i=1}^{n} e_i^2 = \mathrm{E}^T.\mathrm{E} = (Y - XA)^T.(Y - XA)$$

Desenvolvendo as propriedades de matriz transposta, obtemos

$$L = Y^TY - Y^TXA - A^TX^TY + A^TX^TXA = Y^TY - 2AX^TY + AX^TXA$$

Na sequência, novamente derivamos em relação a cada um dos parâmetros. Na forma matricial, escrevemos resumidamente esse resultado como

$$\frac{\partial L}{\partial A} = 0$$

Ou seja,

$$\frac{\partial L}{\partial A} = -2X^TY + 2X^TXA = 0$$

Logo,

$$X^TXA = X^TY$$

Tendo inversa, podemos escrever a matriz dos parâmetros assim:

$$A = (X^TX)^{-1}X^TY$$

2.3 Análise do exemplo na forma algébrica

Para os dados da Tabela 1, vamos determinar os parâmetros α, β e γ do modelo a partir da resolução do sistema de equações. Tanto na forma algébrica, quanto na forma matricial é essencial ter auxílio para os cálculos manuais considerando as grandezas que precisam ser calculadas. Veja a Tabela 2.

Tabela 2 – Tabela de auxílio para os cálculos manuais do exemplo dado

i	x_{1i}	x_{2i}	y_i	x_{1i}^2	x_{2i}^2	$x_{1i}.x_{2i}$	$x_{1i}.y_i$	$x_{2i}.y_i$
1	195	4	1.004	38.025	16	780	195.780	4.016
2	255	4	1.636	65.025	16	1.020	417.180	6.544
3	195	4,6	852	38.025	21,16	897	166.140	3.919,2
4	255	4,6	1.506	65.025	21,16	1.173	384.030	6.927,6
5	225	4,2	1.272	50.625	17,64	945	286.200	5.342,4
6	225	4,1	1.270	50.625	16,81	922,5	285.750	5.207
7	225	4,6	1.269	50.625	21,16	103,5	285.525	5.837,4
8	195	4,3	903	38.025	18,49	838,5	176.085	3.882,9
9	255	4,3	1.555	65.025	18,49	1.096,5	369.525	6.686,5
10	225	4	1.260	50.625	16	900	283.500	5.040
11	225	4,7	1.146	50.625	22,09	1.057,5	257.850	5.386,2
12	225	4,3	1.276	50.625	18,49	967,5	287.100	5.486,8
13	225	4,72	1.225	50.625	22,2784	1.062	275.625	5.782
14	230	4,3	1.321	52.900	18,49	989	303.830	5.680,3
$\sum$	3.155	60,72	17.495	716.425	264,2584	13.683,5	4.001.120	75.738,3

Perceba que encontramos os seguintes parâmetros:

$$n = 14, \sum_{i=1}^{n} y_i = 17.495, \sum_{i=1}^{n} x_{i1} = 3.155, \sum_{i=1}^{n} x_{2i} = 60,72$$

$$\sum_{i=1}^{n} x_{i1}^2 = 716,426, \sum_{i=1}^{n} x_{2i}^2 = 264,2584, \sum_{i=1}^{n} x_{1i}.x_i = 13.683,5$$

$$\sum_{i=1}^{n} x_{1i}y_i = 1.001.120, \sum_{i=1}^{n} x_{2i}y_2 = 75.738,3$$

Podemos utilizar esses dados para escrever o sistema de equações:

$$\begin{cases} \sum_{i=1}^{n} y_i = n\alpha + \beta.\sum_{i=1}^{n} x_{1i} + \gamma \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} y_i.x_{1i} = \alpha \sum_{i=1}^{n} x_{1i} + \beta \sum_{i=1}^{n} x_{1i}^2 + \gamma \sum_{i=1}^{n} x_{1i}.x_{2i} \\ \sum_{i=1}^{n} y_i.x_{2i} = \alpha \sum_{i=1}^{n} x_{2i} + \beta \sum_{i=1}^{n} x_{1i}.x_{2i} + \gamma \sum_{i=1}^{n} x_{2i}^2 \end{cases}$$

$$\begin{cases} 17.495 = 14\alpha + 3.155\beta + 60,72\gamma \\ 4.001.120 = 3.155\alpha + 716.425\beta + 13.683,5\gamma \\ 75.738,30 = 60,72\alpha + 13.683,50\beta + 264,2584\gamma \end{cases}$$

Na maior parte dos casos é possível determinar esse sistema. Ao resolvê-lo, podemos encontrar

$$\alpha = -520{,}08$$
$$\beta = 10{,}78$$
$$\gamma = -152{,}15$$

2.4 Análise do exemplo na forma matricial

Na forma matricial, devemos escrever as matrizes X e Y e determinar A a partir de $A = (X^T X)^{-1} X^T Y$. Temos, no exemplo que

$$X^T.X = \begin{bmatrix} 14 & 3.155 & 60{,}72 \\ 3{,}155 & 716.425 & 13.683{,}5 \\ 60{,}72 & 13.683{,}5 & 264{,}26 \end{bmatrix}$$

Podemos encontrar a matriz inversa. Nesse caso, temos que

$$(X^T.X)^{-1} = \begin{bmatrix} 30{,}24760 & -0{,}04172 & -4{,}78994 \\ -0{,}04172 & 0{,}00018 & 0{,}00004 \\ -4{,}78994 & 0{,}00004 & 1{,}10244 \end{bmatrix}$$

Além disso, precisamos calcular a matriz $X^T Y$ que resulta em

$$X^T Y = \begin{bmatrix} 17.495 \\ 4.001.120 \\ 75.738{,}3 \end{bmatrix}$$

Assim, encontramos parâmetros dados pela matriz A como $(X^T X)^{-1} X^T Y$

$$A = \begin{bmatrix} -520{,}08 \\ 10{,}78 \\ -152{,}15 \end{bmatrix}$$

Note que nosso modelo se torna

$$y = -520{,}08 + 10{,}78x_1 - 152{,}15x_2$$

TEMA 3 – ANÁLISE DE VARIÂNCIA E MEDIDAS DE ASSOCIAÇÃO

Vejamos como realizar a análise de variância para verificar se, de fato, existe relação entre as variáveis explicativas e a variável resposta. Também vejamos duas medidas de associação chamadas de *coeficiente de determinação múltiplo* e *coeficiente de determinação ajustado*, similares ao analisado no método de regressão linear simples.

3.1 Análise de variância – teste F

A análise de variância a partir do uso do teste F permite verificar se existe relação, de fato, entre as variáveis explicativas e a variável resposta. Nesse caso, precisamos encontrar SQT, SQE e SQR. Pode-se mostrar que

$$SQT = Y^T Y - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$SQR = A^T X^T Y - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$SQE = SQT - SQR$$

Nesse caso pode-se usar a ANOVA, como apresentada a tabela a seguir.

Tabela 3 – Tabela da ANOVA ampliada com o cálculo de F_0

Variação	$\boldsymbol{SQ}$	$\boldsymbol{gl}$	$\boldsymbol{MQ}$	$\boldsymbol{F_0}$
Regressão	SQR	2	MQR	$\frac{MQR}{MQE}$
Erro	SQE	$n-2$	MQE	
Total	SQT	$n-1$		

O teste de hipótese será o teste F. Note que teremos

$$\begin{cases} H_0: \alpha = \beta = \gamma = 0 \\ H_1: \alpha \neq \beta \neq \gamma \end{cases}$$

E rejeitamos H_0 se

$$F_0 > F_{(1-\alpha,2,n-2)}$$

3.2 Teste ANOVA para o exemplo dado

No caso discutido, encontramos os seguintes valores necessários:

$$SQT = Y^T Y - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 22.527.889 - \frac{(17.495)^2}{14} = 665.387{,}2$$

$$SQR = A^T X^T Y - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 22.514.467{,}9 - \frac{(17.495)^2}{14} = 651.966{,}1$$

$$SQE = 665.387{,}2 - 651.966{,}1 = 13.421{,}1$$

$$MQR = \frac{SQT}{2} = \frac{651.966{,}1}{2} = 325.983$$

$$MQE = \frac{SQE}{n-2} = \frac{13.421{,}2}{11} = 1.220{,}1$$

$$F_0 = \frac{MQR}{MQE} = \frac{325.983}{1.220{,}1} = 267{,}18$$

E construímos a tabela da ANOVA, conforme segue.

Tabela 4 – Tabela da ANOVA ampliada com o cálculo de F_0

Variação	SQ	gl	MQ	F_0
Regressão	651.966,1	2	325.983	
Erro	13.421,2	11	1.220,1	267,18
Total	655.387,2	13		

Consultando a Tabela F para um nível de significância de 95%, verificamos que $F_{(0.95,2,11)} = 3{,}98$. Como $F_0 > F$ (i.e., $267{,}18 > 3{,}98$) podemos rejeitar H_0, confirmando que as variáveis explicativas, de fato, explicam a variável resposta de forma adequada.

3.3 Medidas de associação – coeficiente de determinação

Podemos definir dois coeficientes de determinação que também validam se as variáveis explicativas, de fato, explicam a variável resposta. Nesse caso, o coeficiente de determinação indica se o modelo escolhido é realmente linear. Assim, definimos o coeficiente de determinação múltiplo como

$$R = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}$$

Para o exemplo dado, temos

$$R = 1 - \frac{13.421{,}2}{655.387{,}2} = 0{,}9798$$

Como $R \to 1$ podemos afirmar que o comportamento das variáveis é linear. Podemos também utilizar o coeficiente de determinação ajustado

$$R_a = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2)$$

Para o exemplo dado, temos

$$R_a = 1 - \frac{13}{11}(1 - 0{,}9798) = 0{,}9762$$

TEMA 4 – TESTES INDIVIDUAIS E INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS

Podemos determinar o intervalo de confiança para os parâmetros α, β, γ discutidos.

4.1 Intervalo de Confiança para α

Podemos mostrar que o intervalo de confiança para α é dado por

$$\hat{\alpha} - t_{\left(\frac{\alpha}{2},n-p-1\right)}\sqrt{MQE.(X^T.X)^{-1}_{11}} \leq \alpha \leq \hat{\alpha} + t_{\left(\frac{\alpha}{2},n-p-1\right)}.\sqrt{MQE.(X^T.X)^{-1}_{11}}$$

Dessa forma, sendo $t_{(0,05/2,11)} = 2{,}201$ temos

$$-520{,}08 - 2{,}201.\sqrt{1.220{,}1\,.\,30{,}24760} \leq \alpha \leq -520{,}08 + 2{,}201.\sqrt{1.220{,}1\,.\;30{,}24760}$$

Então,

$$-942{,}91 \leq \alpha \leq -97{,}25$$

4.2 Intervalo de Confiança para β

Podemos mostrar que o intervalo de confiança para β é dado por

$$\hat{\beta} - t_{\left(\frac{\alpha}{2},n-p-1\right)}\sqrt{MQE.(X^T.X)^{-1}_{22}} \leq \beta \leq \hat{\beta} + t_{\left(\frac{\alpha}{2},n-p-1\right)}.\sqrt{MQE.(X^T.X)^{-1}_{22}}$$

Dessa forma, sendo $t_{(0,05/2,11)} = 2{,}201$ temos

$$10{,}78 - 2{,}201.\sqrt{1.220{,}1\,.\;0{,}00018} \leq \beta \leq 10{,}78 + 2{,}201.\sqrt{1.220{,}1\,.0{,}00018}$$

Então,

$$9{,}75 \leq \beta \leq 11{,}81$$

4.3 Intervalo de confiança para γ

Podemos mostrar que o intervalo de confiança para γ é dado por:

$$\hat{\gamma} - t_{\left(\frac{\alpha}{2},n-p-1\right)}\sqrt{MQE.(X^T.X)^{-1}_{33}} \leq \gamma \leq \hat{\gamma} + t_{\left(\frac{\alpha}{2},n-p-1\right)}.\sqrt{MQE.(X^T.X)^{-1}_{33}}$$

Dessa forma, sendo $t_{(0,05/2,11)} = 2{,}201$ temos:

$$-152{,}15 - 2{,}201.\sqrt{1.220{,}1\,.\;1{,}10224} \leq \gamma \leq -152{,}15 + 2{,}201.\sqrt{1.220{,}1\,.1{,}10224}$$

Então,

$$-232{,}87 \leq \gamma \leq -71{,}43$$

TEMA 5 – PREVISÃO PARA A VARIÁVEL RESPOSTA

Agora, com o modelo de regressão linear múltipla determinado, podemos realizar a previsão para a variável de resposta dado uma determinada combinação das variáveis explicativas.

5.1 Valor previsto para os dados do problema

A tabela a seguir apresenta valores previstos para os dados do problema.

Tabela 5 – Valores previstos para os dados do problema

i	x_{1i}	x_{2i}	y_i	$\hat{y}_i$	e_i
1	195	4	1.004	973,42	−30,58
2	255	4	1.636	1620,22	−15,78
3	195	4,6	852	882,13	30,13
4	255	4,6	1.506	1.528,93	22,93
5	225	4,2	1.272	1.266,39	−5,61
6	225	4,1	1.270	1.281,605	11,605
7	225	4,6	1.269	1.205,53	−63,47
8	195	4,3	903	927,775	24,775
9	255	4,3	1.555	1.574,575	19,575
10	225	4	1.260	1.296,82	36,82
11	225	4,7	1.146	1.190,315	44,315
12	225	4,3	1.276	1.251,175	−24,825
13	225	4,72	1.225	1.187,272	−37,728
14	230	4,3	1.321	1.305,075	−15,925

Considera-se que o modelo pode ser utilizado para prever dados em regiões próximas às regiões dos valores analisados.

FINALIZANDO

Finalmente conseguimos generalizar o método de regressão linear para o caso de duas variáveis. O método pode ser facilmente generalizado para mais de duas variáveis desde que se considere o uso da forma matricial. Essa forma é mais adequada para o uso de softwares computacionais.

REFERÊNCIAS

LARSON, R.; FARBER, B. **Estatística aplicada**. 6 ed. São Paulo: Pearson Education do Brasil, 2015.

CASTANHEIRA, N. P. **Estatística aplicada a todos os níveis**. Curitiba: InterSaberes, 2012.

CASTANHEIRA, N. P. **Métodos Quantitativos**. Curitiba: InterSaberes, 2013.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 5 ed. Rio de Janeiro: LTC, 2012.

SIQUEIRA, J. O. **Fundamentos de Métodos Quantitativos**. São Paulo: Saraiva, 2011.

DOWNING, D.; CLARK, J. **Estatística aplicada**. 3 ed. São Paulo: Saraiva, 2010.

MONTGOMERY, D. C.; RUNGER, G. C.; HUBELE, N. F. **Estatística aplicada à engenharia**. 2 ed. Rio de Janeiro: LTC, 2013.

FREUND, J. E. **Estatística aplicada**. 11 ed. Porto Alegre: Bookman, 2007.