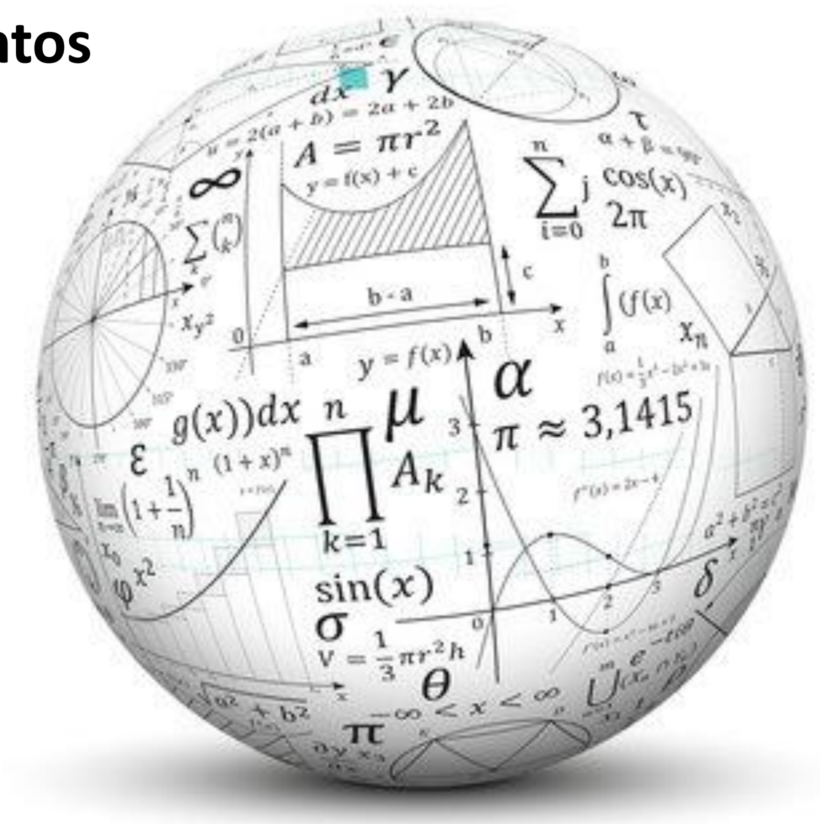
MAP 2110 – Modelagem e Matemática

1º Semestre - 2023

Prof. Dr. Luis Carlos de Castro Santos

lsantos@ime.usp.br



Revisitando o Problema da Aula Passada

A função de ajuste proposta:

$$f(x) = \beta x^{\alpha}$$

Para os dados:

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

Usando a transformação de variáveis:

$$g(x) = \ln(y) = \alpha \ \ln(x) + \ln(\beta)$$

$$f(\hat{x}) = \hat{y} = \ a\hat{x} + b$$

$$\alpha \rightarrow a$$

$$\ln(\beta) \rightarrow b$$

$\Rightarrow$

$$\alpha = a$$

$$\beta = e^{b}$$

Usando organização sugerida:

x	y		x=ln(x)	y=ln(y)	yx	x^2	g	y-g	(y-g)^2		x	y	f	y-f	(y-f)^2
0,5	0,7		-0,6931	-0,3567	0,2472	0,4805	-0,3032	-0,0535	0,0029		0,5	0,7	0,7384	-0,0384	0,0015
1	3,4		0,0000	1,2238	0,0000	0,0000	1,1266	0,0972	0,0094		1	3,4	3,0852	0,3148	0,0991
1,5	7,2		0,4055	1,9741	0,8004	0,1644	1,9630	0,0111	0,0001		1,5	7,2	7,1207	0,0793	0,0063
2	12,4		0,6931	2,5177	1,7451	0,4805	2,5564	-0,0387	0,0015		2	12,4	12,8899	-0,4899	0,2400
2,5	20,1		0,9163	3,0007	2,7495	0,8396	3,0167	-0,0160	0,0003		2,5	20,1	20,4247	-0,3247	0,1055
			1,3218	8,3596	5,5423	1,9649			0,0142						0,4523
			sx	sy	syx	sx2			S						S
			1,7470												
			(sx)^2												
			den	8,0774		a	2,0628		beta	3,0852					
			num_a	16,6622		b	1,1266		alpha	2,0628					
			num_b	9,1002											

Norma do resíduo do modelo nas variáveis originais

Norma do resíduo do modelo nas variáveis transformadas

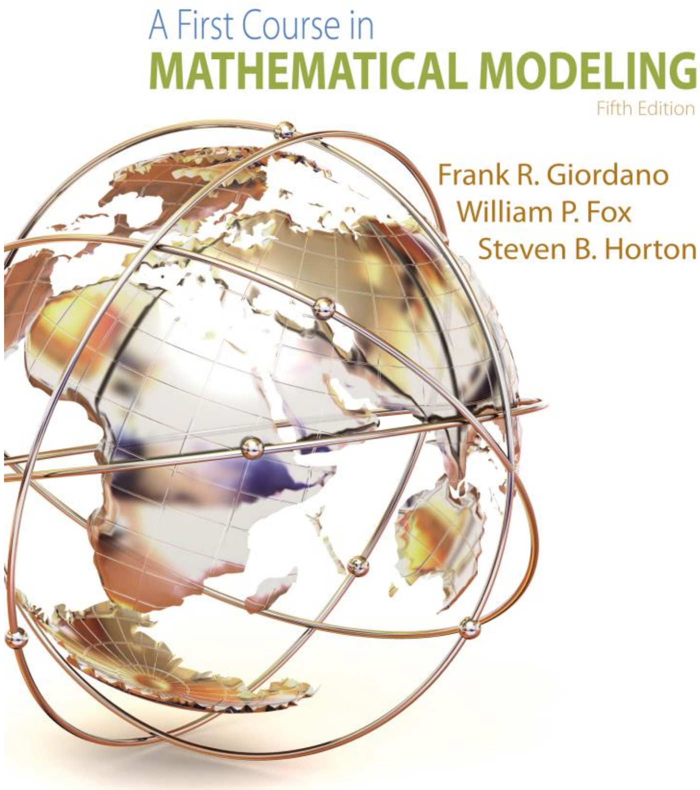
O resíduo das variáveis transformadas foi minimizado, o resíduo nas variáveis originais é uma consequência apenas.

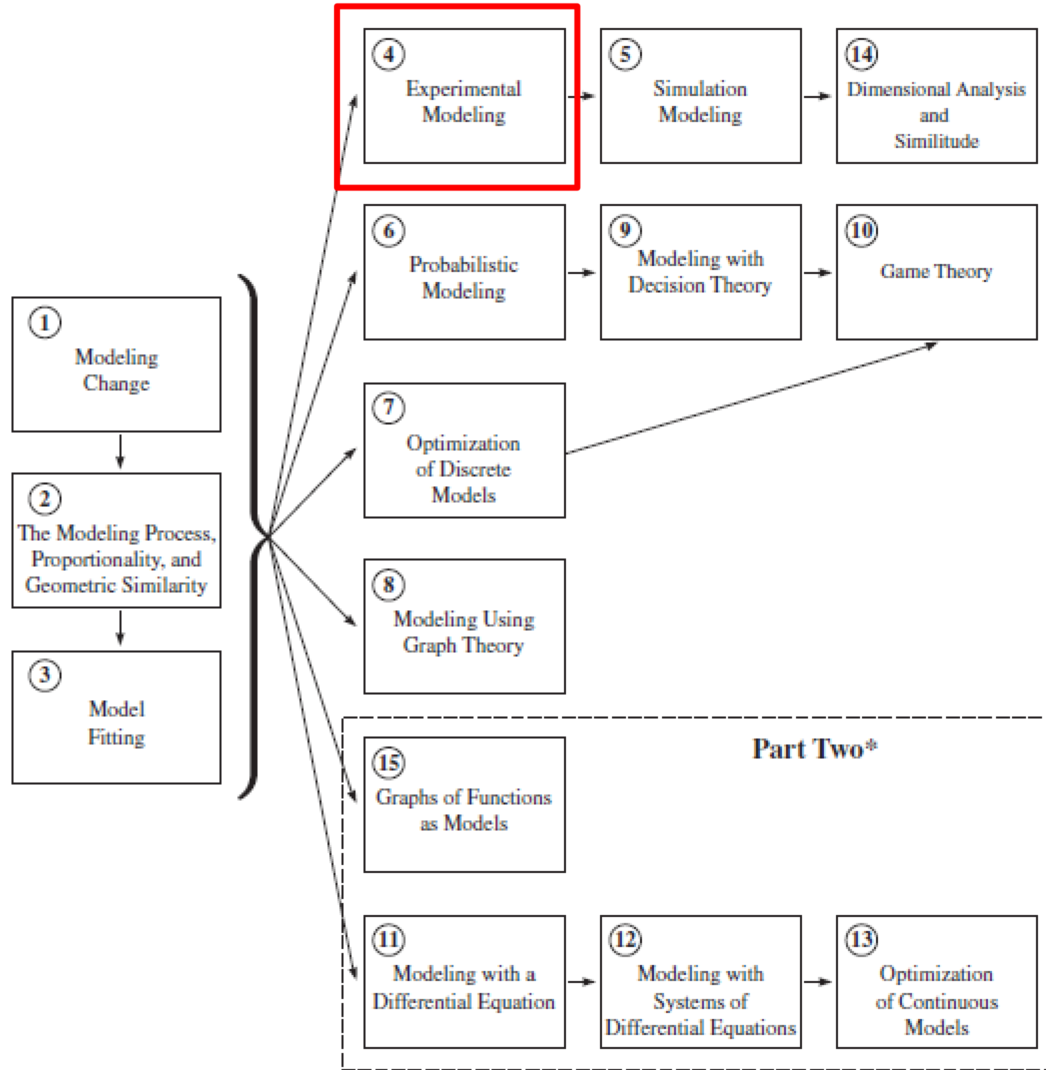
Comparando agora os modelos aplicados aos mesmos dados:

x	**y**	y=beta*(x^alpha)	y=beta*(x^2)	y=beta*(x^2) transf	r1^2	r2^2	r3^2
0,5	0,7	0,7384	0,7967	0,7842	0,0015	0,0094	0,0071
1	3,4	3,0852	3,1869	3,1368	0,0991	0,0454	0,0693
1,5	7,2	7,1207	7,1705	7,0578	0,0063	0,0009	0,0202
2	12,4	12,8899	12,7476	12,5472	0,2400	0,1208	0,0217
2,5	20,1	20,4247	19,9181	19,605	0,1055	0,0331	0,2450
	alpha	2,0628			S1	S2	S3
	beta	3,0852	3,1869	3,1368	0,4523	0,2095	0,3633

Nesse modelo o ajuste foi feito sem os dados serem transformados

MAP2110 – aula 13





*Part Two requires single-variable calculus as a corequisite.

4 Experimental Modeling

O nome mais adequado para o capítulo talvez fosse <u>Modelagem Empírica</u>

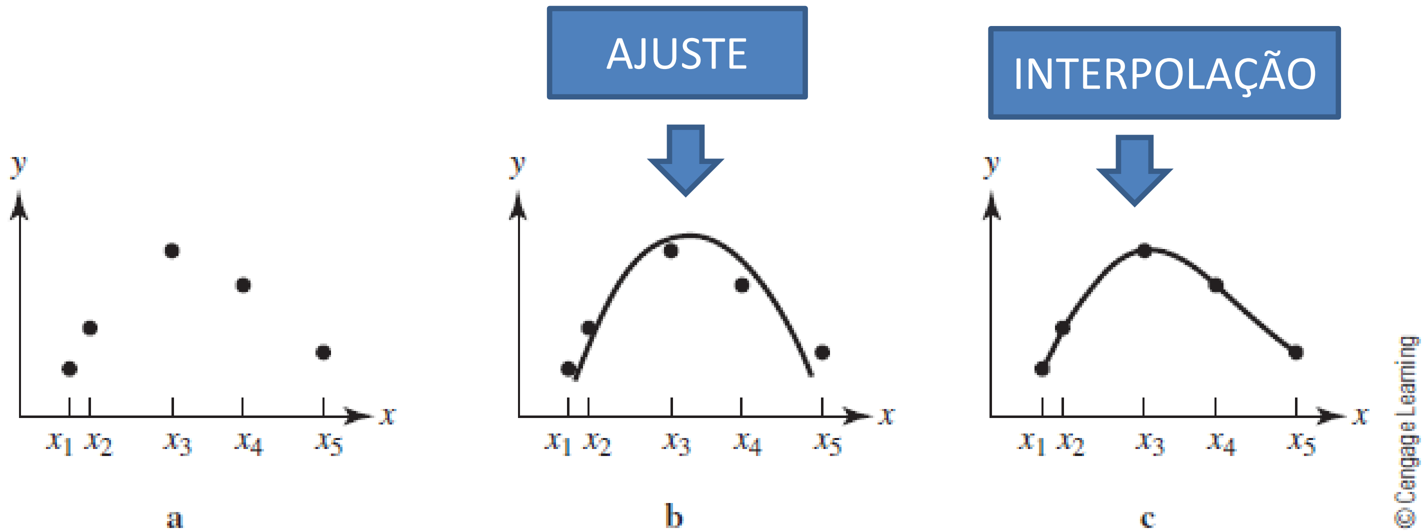


Figure 4.1

If the modeler expects a quadratic relationship, a parabola may be fit to the data, as in b. Otherwise, a smooth curve may be passed through the points, as in c.

Analisando os gráficos o pesquisador pode decidir utilizar alguma abordagem.

4.1 Harvesting in the Chesapeake Bay and Other One-Term Models

"Explorando a Baia de Chesapeake e outro modelos de termo único"

PAUL HORN / Inside Climate News

Bluefish = anchova

Bluecrab

Table 4.1 Harvesting the bay, 1940–1990

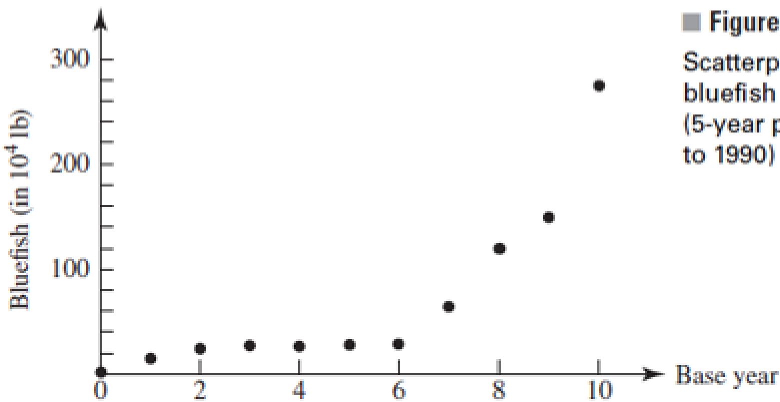
Year	Bluefish (lb)	Blue crabs (lb)
1940	15,000	100,000
1945	150,000	850,000
1950	250,000	1,330,000
1955	275,000	2,500,000
1960	270,000	3,000,000
1965	280,000	3,700,000
1970	290,000	4,400,000
1975	650,000	4,660,000
1980	1,200,000	4,800,000
1985	1,500,000	4,420,000
1990	2,750,000	5,000,000

Observando os dados aparentemente por volta de 1970 a tendência de captura de anchova tem um crescimento acelerado e a de caranguejos azuis parece atingir um platô.

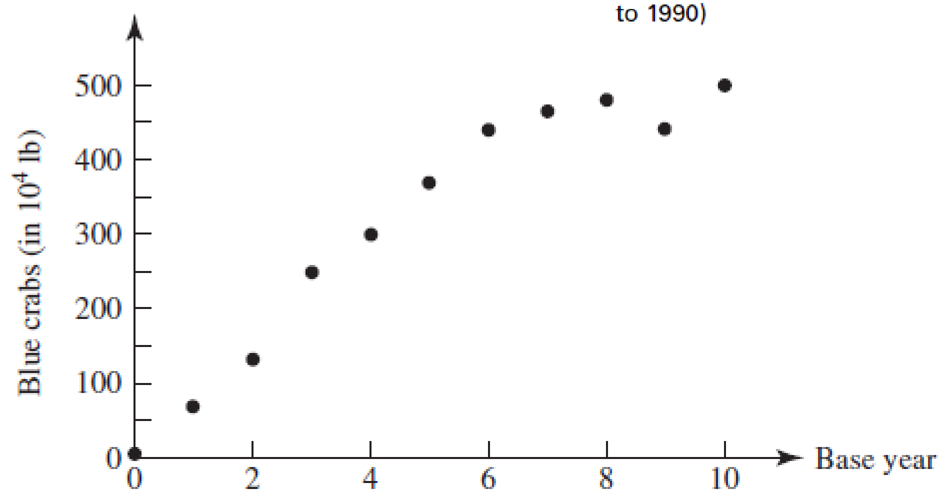
Em ambos os casos se procurar uma relação direta entre o tempo e massa total capturada:

$$y \propto f(x)$$

Como obter essa relação ?



■ **Figure 4.2**
Scatterplot of harvesting bluefish versus base year (5-year periods from 1940 to 1990)



■ **Figure 4.3**
Scatterplot of harvesting blue crabs versus base year (5-year periods from 1940 to 1990)

Uma estratégia é propor uma relação funcional conhecida. Alguns exemplos comuns são de "escada de potências":

Ladder of powers

⋮

z^2

z

$\sqrt{z}$

$\log z$

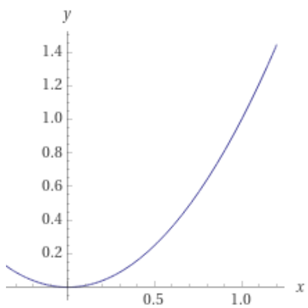
$\frac{1}{\sqrt{z}}$

$\frac{1}{z}$

$\frac{1}{z^2}$

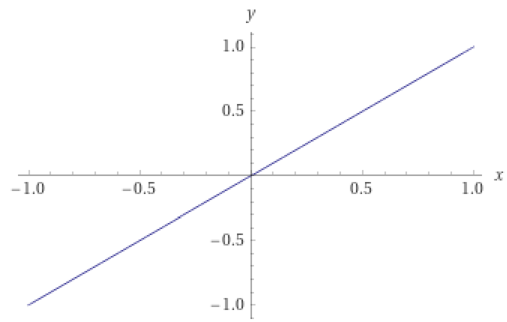
⋮

plot x^2



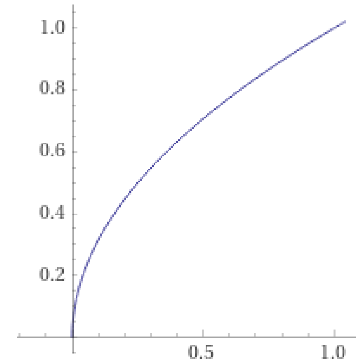
plot x

Plot



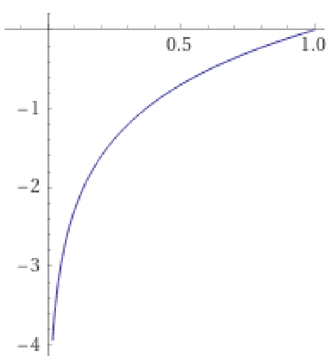
plot $\sqrt{x}$

Plots



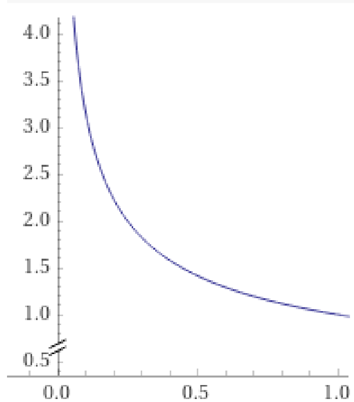
plot $\log(x)$

Plots



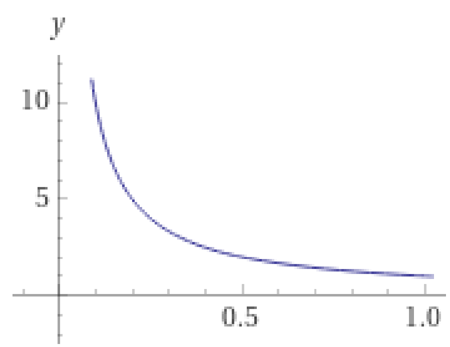
plot $\frac{1}{\sqrt{x}}$

Plots



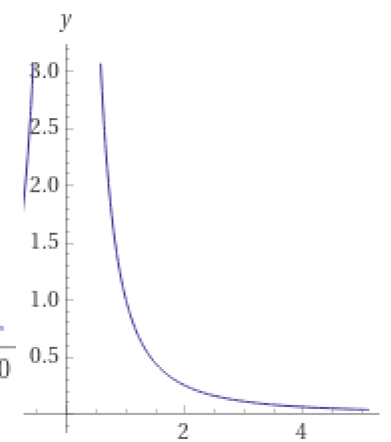
plot $\frac{1}{x}$

Plots



plot $\frac{1}{x^2}$

Plots



Essa busca é realizada testando uma transformação de variáveis que produza uma reta:

A transformação pode ser aplicada tanto de um lado quanto do outro da equação na busca do melhor aspecto geral da curva.

Table 4.2 Ladder of Transformations

$\vdots$

z^3

z^2

z (no change)

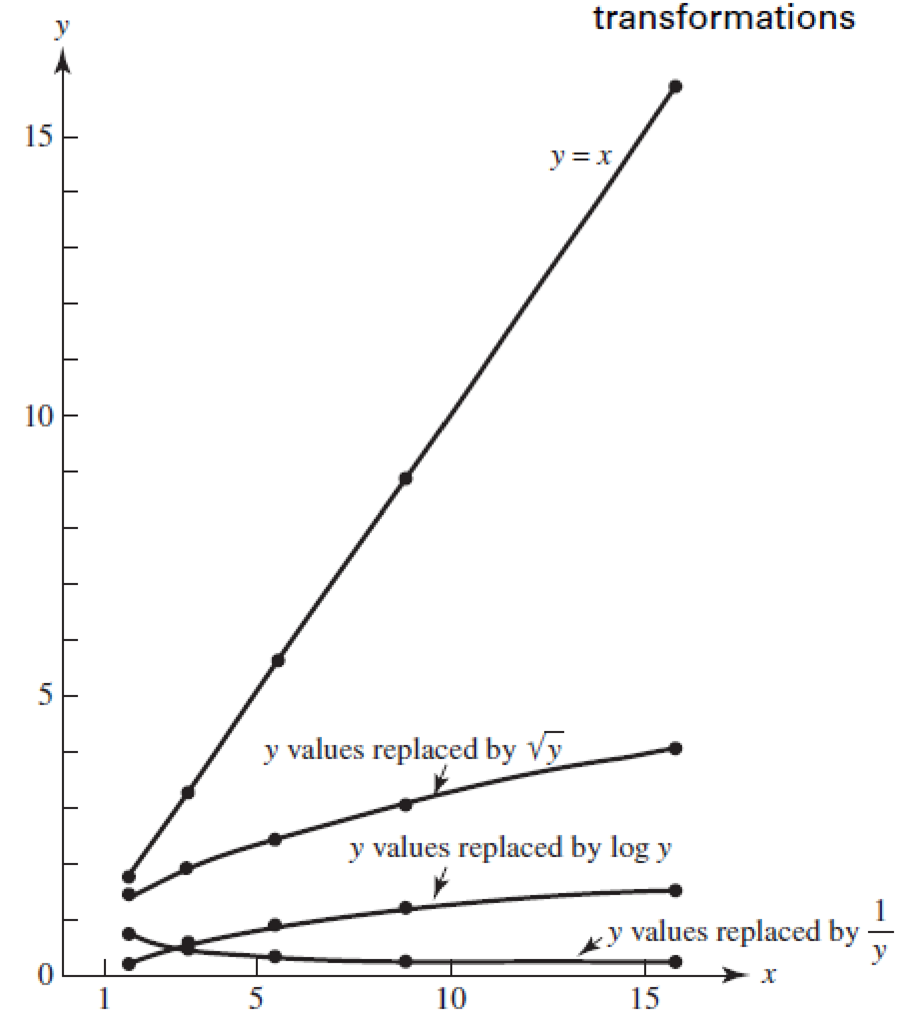
* $\sqrt{z}$, $\log z$, $\frac{-1}{\sqrt{z}}$, $\frac{-1}{z}$

$\frac{-1}{z^2}$

$\vdots$

*The transformations most often used.

Figure 4.4 Relative effects of three transformations



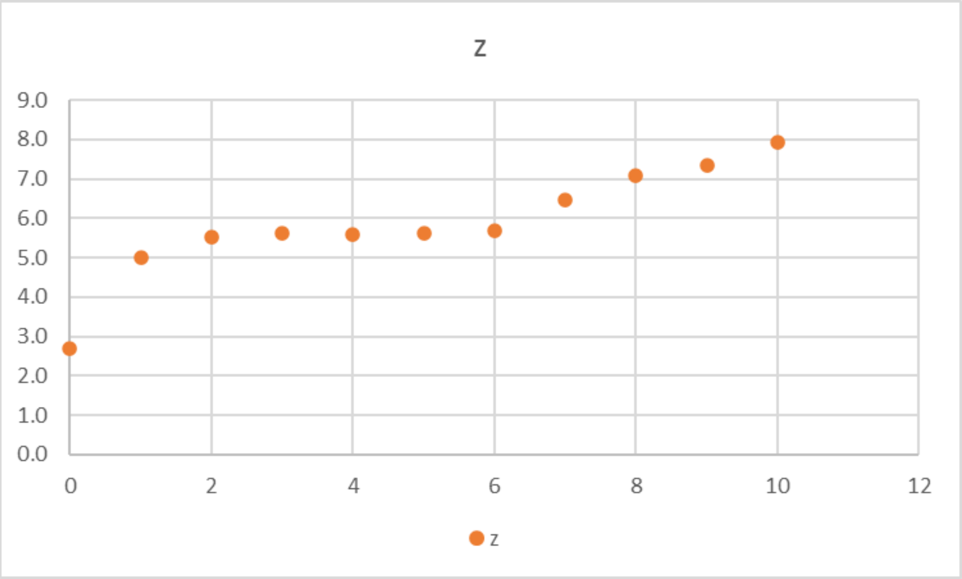
EXAMPLE 1 *Harvesting Bluefish*

Após um processo de tentativa e erro o autor optou pela relação:

$$\log y = ax + b$$

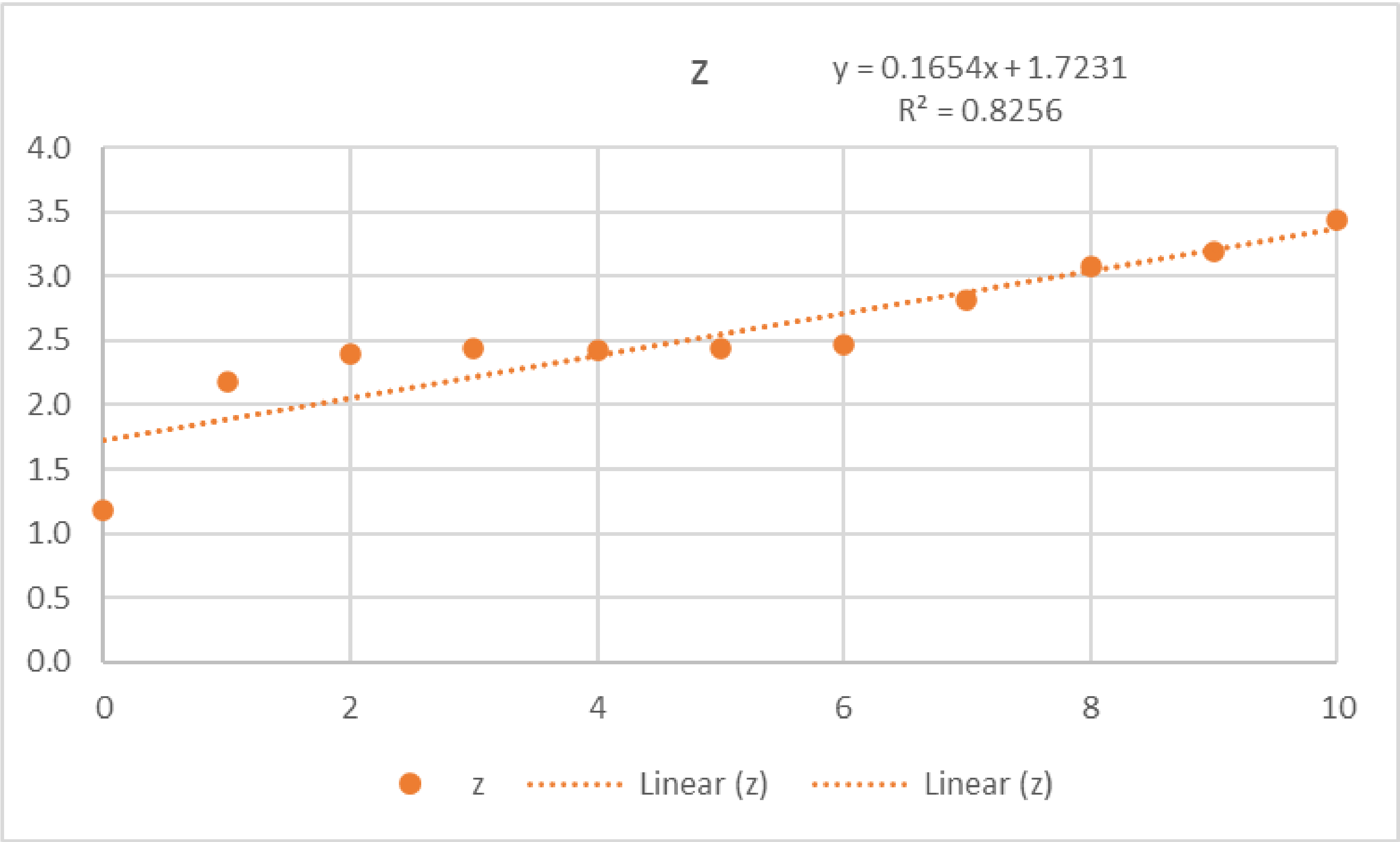
Table 4.3 Harvesting the bay: Bluefish, 1940–1990

Year	Base year	Bluefish (lb)
	x	y
1940	0	15,000
1945	1	150,000
1950	2	250,000
1955	3	275,000
1960	4	270,000
1965	5	280,000
1970	6	290,000
1975	7	650,000
1980	8	1,200,000
1985	9	1,550,000
1990	10	2,750,000



Fazendo $z = \log y$ e propondo um ajuste linear (nesse caso o autor propôs o log base 10).

Usando o ajuste linear nativo do Excel verificamos o valor informado pelo autor.



x model. We fit with least squares the model of the form

$$\log y = mx + b$$

and obtain the following estimated curve:

$$\log y = 0.7231 + 0.1654x$$

where x is the base year and $\log y$ is to the base 10 and y is measured in 10^4 pounds. (See Figure 4.5.)

Using the property that $y = \log n$ if and only if $10^y = n$, we can rewrite this equation (with the aid of a calculator) as

$$y = 5.2857(1.4635)^x \qquad (4.1)$$

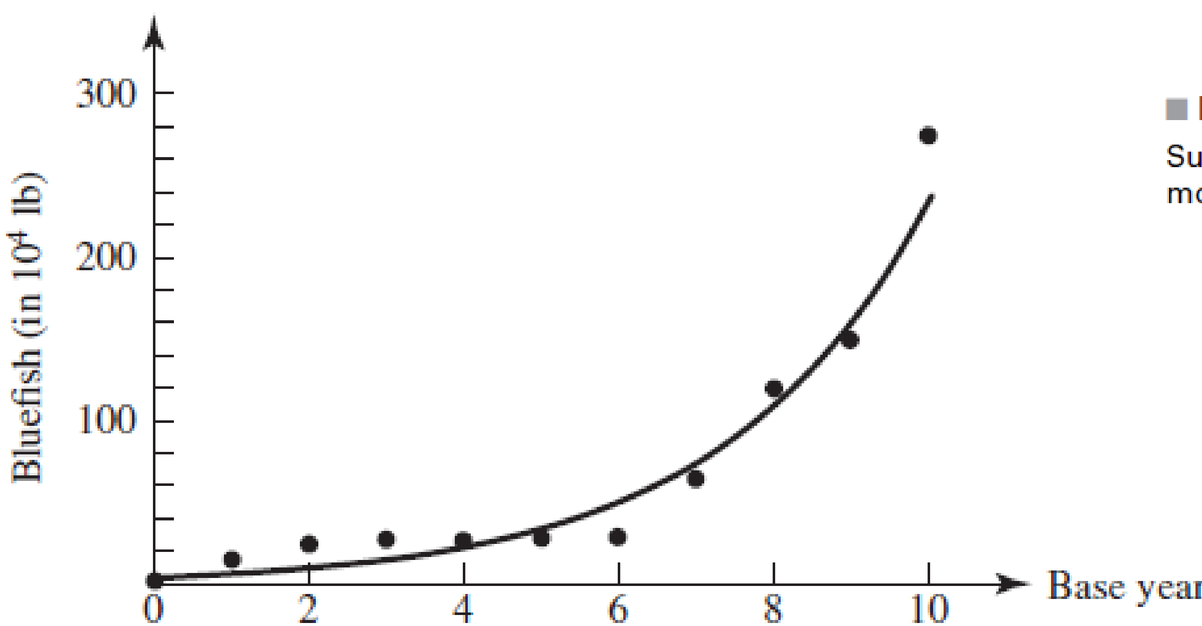


Figure 4.5
Superimposed data and model $y = 5.2857(1.4635)^x$

EXAMPLE 2 *Harvesting Blue Crabs*

Table 4.4 Harvesting the bay: Blue crabs, 1940–1990

Year	Base year	Blue crabs (lb)
	x	y
1940	0	100,000
1945	1	850,000
1950	2	1,330,000
1955	3	2,500,000
1960	4	3,000,000
1965	5	3,700,000
1970	6	4,400,000
1975	7	4,660,000
1980	8	4,800,000
1985	9	4,420,000
1990	10	5,000,000

Pelo aspecto da curva o autor propôs uma relação:

$$y = k\sqrt{x}$$

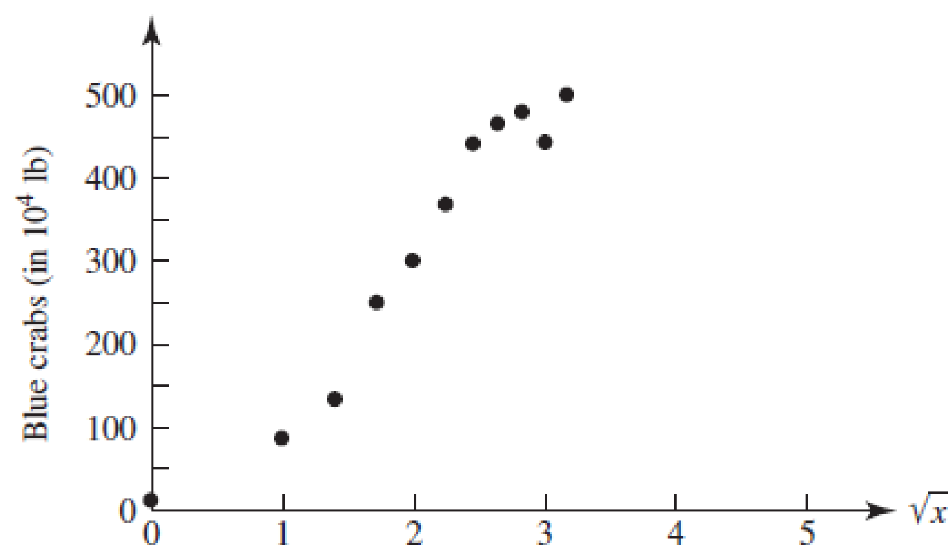
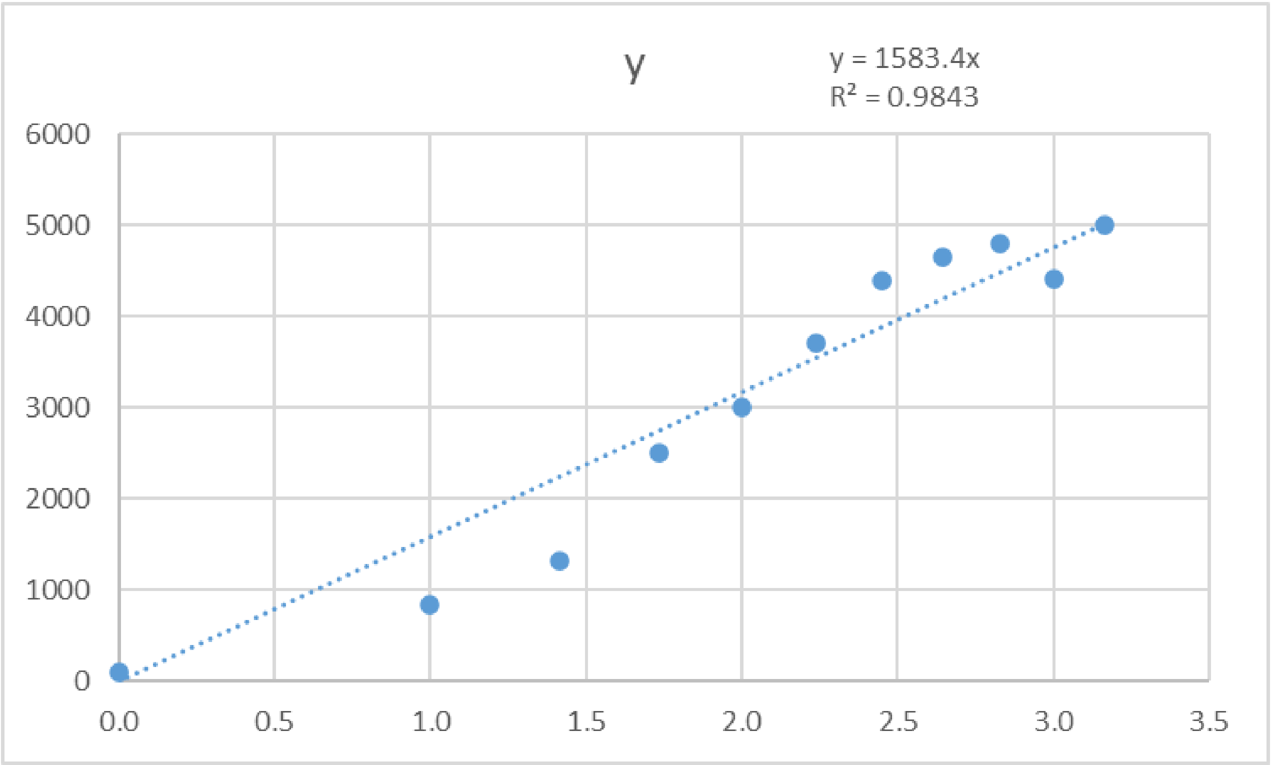
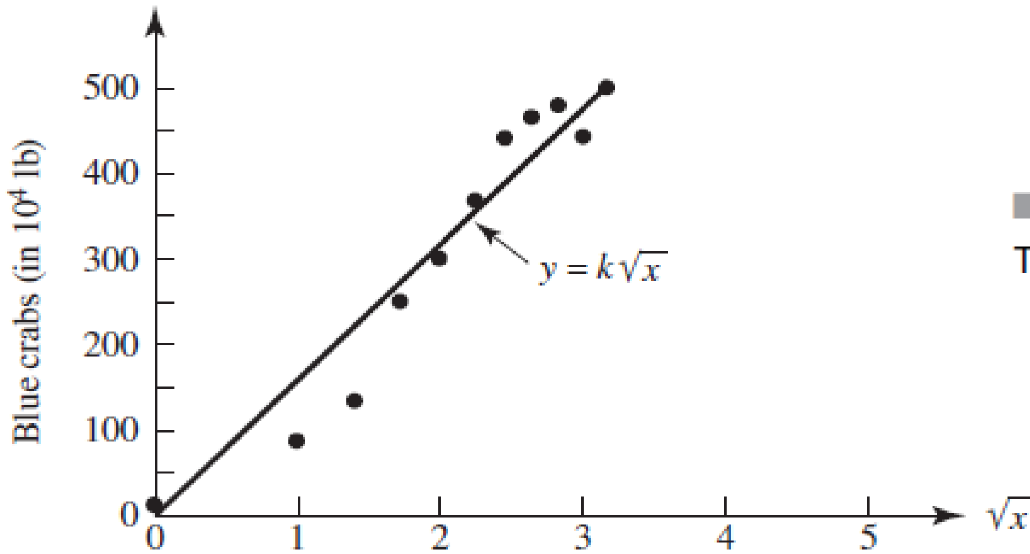


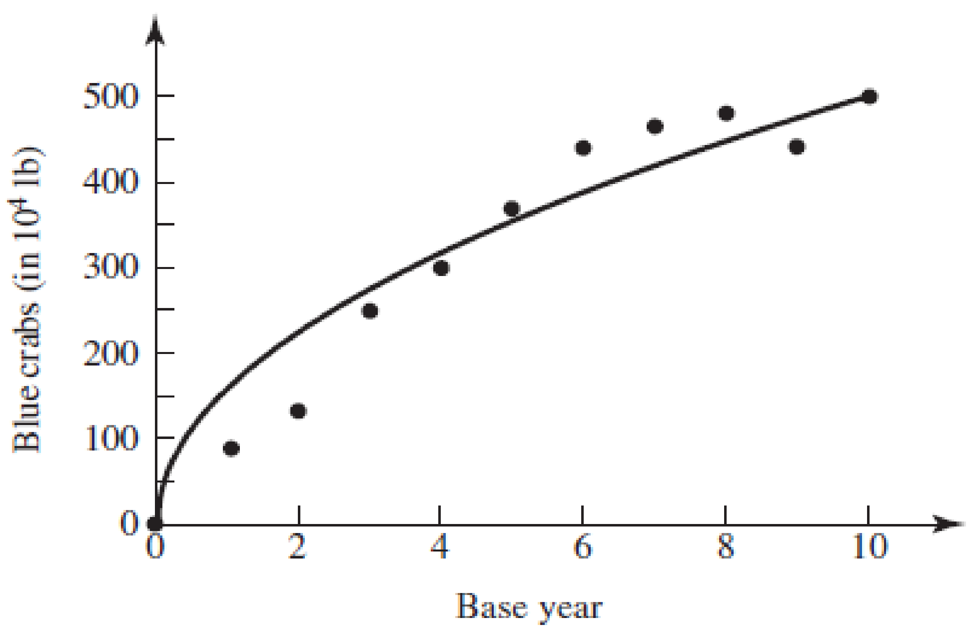
Figure 4.6
Blue crabs (in 10^4 lb) versus $\sqrt{x}$

Usando o ajuste linear nativo do Excel novamente verificamos o valor informado pelo autor.





■ **Figure 4.7**
The line $y = 158.344\sqrt{x}$



Data inferred from a scatterplot in Frederick E. Croxton, Dudley J. Cowden, and Sidney Klein, Applied General Statistics, 3rd ed. (Englewood Cliffs, NJ: Prentice-Hall, 1967), p. 390.

O que é R^2 ?

If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

then the variability of the data set can be measured with two sums of squares formulas:

- The sum of squares of residuals, also called the residual sum of squares:

$$SS_{\text{res}} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2$$

- The total sum of squares (proportional to the variance of the data):

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$$

The most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Quando mais próximos de uma relação linear são os dados mais próximo de 1 é o R^2.

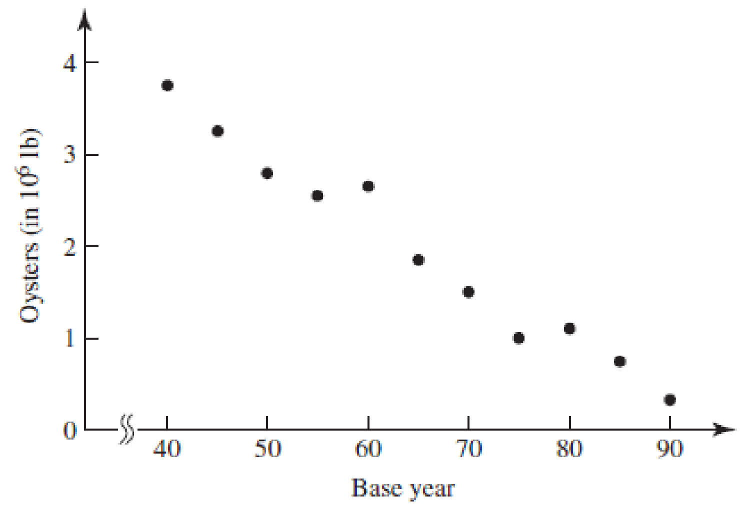
4.1 PROBLEMS

6. Table 4.7 and Figure 4.9 present data representing the commercial harvesting of oysters in Chesapeake Bay. Fit a simple, one-term model to the data. How well does the best one-term model you find fit the data? What is the largest error? The average error?

Table 4.7 Oysters in the bay

Year	Oysters harvested (bushels)
1940	3,750,000
1945	3,250,000
1950	2,800,000
1955	2,550,000
1960	2,650,000
1965	1,850,000
1970	1,500,000
1975	1,000,000
1980	1,100,000
1985	750,000
1990	330,000

Figure 4.9
Oysters (millions of pounds) versus base year



7. In Table 4.8, X is the Fahrenheit temperature, and Y is the number of times a cricket chirps in 1 minute. Fit a model to these data. Analyze how well it fits.

Table 4.8 Temperature and chirps per minute for 20 crickets

Observation number	X	Y	Observation number	X	Y
1	46	40	11	61	96
2	49	50	12	62	88
3	51	55	13	63	99
4	52	63	14	64	110
5	54	72	15	66	113
6	56	70	16	67	120
7	57	77	17	68	127
8	58	73	18	71	137
9	59	90	19	72	132
10	60	93	20	71	137

9. The following data measure two characteristics of a ponderosa pine. The variable X is the diameter of the tree, in inches, measured at breast height; Y is a measure of volume—the number of board feet divided by 10. Fit a model to the data. Then express Y in terms of X.

Diameter and volume for 20 ponderosa pine trees

Observation number	X	Y	Observation number	X	Y
1	36	192	11	31	141
2	28	113	12	20	32
3	28	88	13	25	86
4	41	294	14	19	21
5	19	28	15	39	231
6	32	123	16	33	187
7	22	51	17	17	22
8	38	252	18	37	205
9	25	56	19	23	57
10	17	16	20	39	265

Sugestão: Ordene a tabela


12
1
2
3
4
5
6
7
8
9
10
11

Exemplo:

$n = 2$

Table 3.3 Data collected to fit $y = Ax^2$

x	0.5	1.0	1.5	2.0	2.5
y	0.7	3.4	7.2	12.4	20.1

$$a = \frac{\sum_{i=1}^{m} y_i x_i^n}{\sum_{i=1}^{m} x_i^{2n}}$$

x	y	x^n	x^{2n}	$y\ x^n$	f	$y - f$	$\{y - f\}^2$

$$\sum_{i=1}^{m} x_i^{2n} \quad \sum_{i=1}^{m} y_i x_i^n$$

$$S = \sum_{i=1}^{m} \{(y_i - f(x_i))\}^2$$

Fim Aula 13