AULA

Dependência bivariada,

regressão e correlação linear.

Ministrante Prof. Dr. Vladimir Belitsky,
IME-USP

14 de novembro de 2017

Distr. bivariada discreta. Lembrete.

Observação: No que segue-se, consideraremos somente as variáveis aleatórias discretas cujo conjunto de valores é finito (as variáveis aleatórias, como, por exemplo, geométrica e a de Poisson estão, então, fora da presente apresentação).

No nosso curso, a *distribuição bivariada* representa-se por tabela e/ou por função (com o domínio em $\mathbb{R}^2$ e os valores em $\mathbb{R}$).

Recordamos que a tabela e a função supramencionadas as vezes chamam-se também por *distribuição bivariada*, o que é uma nomenclatura cómoda quando não causa confusão.

Distr. bivariada discreta. Lembrete.

Abaixo, há um exemplo de tabela representando uma distribuição bivariada.

Y \ X	0	1	2	3
0	1/8	0	0	1/8
1	0	2/8	2/8	0
2	0	1/8	1/8	0

Usando o presente exemplo, recordaremos a notação genérica:
n – quantidade de valores que X pode assumir (aquí, $n = 4$);
$x_1, \ldots, x_n$ – os valores que X pode assumir (aqui, $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3$);
m – quantidade de valores que Y pode assumir (aqui, $m = 3$);
$y_1, \ldots, y_m$ – os valores que Y pode assumir (aquí, $y_1 = 0, y_2 = 1, y_3 = 2$);
$p(x_i, y_j)$ – a probabilidade de X assumir o valor x_i e, simultaneamente, Y assumir o valor y_j (aqui, por exemplo, 1/8 na primeira linha e última coluna é o valor de $p(x_4, y_1)$).

Distr. bivariada discreta. Lembrete.

Eis a cara genérica de Tabela de Distribuição Bivariada (ou, em outras palavras, Tabela que representa uma Distribuição Bivariada):

$Y \backslash X$	x_1	x_2	$\cdots$	x_n
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	$\cdots$	$p(x_n, y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$	$\cdots$	$p(x_n, y_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	$\cdots$	$p(x_n, y_m)$

Observação: Se X e/ou Y fosse variável aleatória geométrica ou de Poisson (ou qq outra variável aleatória cujo conjunto de valores não é finito mas é infinito contável), então teriamos distribuição bivariáda infinita; as tabelas de tais distribuições são difíceis para desenho e é por isto (e por outras razões que não revelarei) trataremos no nosso curso só distribuições finitas.

Distr. bivariada discreta. Lembrete.

Aqui, recordo – via um exemplo – o conceito *distribuição marginal* e sua construção a partir de distribuição bivariada (a mesma construção vale para caso de distribuição conjunta de mais que duas variáveis).

Y \ X	0	1
3	1/10	2/10
5	3/10	4/10
$\boldsymbol{P}[X = x_i]$	4/10	6/10

$\Rightarrow$

$\boldsymbol{P}[Y = y_j]$
3/10
7/10

Distr. bivariada discreta. Lembrete.

O exemplo abaixo recorda a construção da distribuição bivariada com as distribuições marginais dadas, e que seja tal que as variáveis tormen-se independentes.

Y \ X	0	1	$\boldsymbol{P}[Y = y_j]$
3	12/100	18/100	3/10
5	28/100	42/100	7/10
$\boldsymbol{P}[X = x_i]$	4/10	6/10	

Distr. bivariada discreta. Dependência.

As duas tabelas acima correspondem às mesmas distribuições marginais. Isto motiva a pergunta:
Dadas duas distribuições univariadas, quantas tabelas bivariádas com tais distribuições como suas marginais podem ser construidas?
A resposta é:
Em geral, são infinitas (exceto alguns casos "patológicos").

Mostraremos isto no exemplo, onde as distribuições univariadas assumem 3 valores; o caso de 2 valores, por ser mais simples, foi deixado como exercício para casa.

Distr. bivariada discreta. Dependência.

Y \ X	1	2	3
1	4/18	1/18	1/18
2	2/18	0	4/18
3	0	5/18	1/18

Y \ X	1	2	3
1	4/18	1/18	1/18
2	2/18	1/18	4/18
3	1/18	4/18	1/18

Y \ X	1	2	3
1	6/36	4/36	2/36
2	1/36	2/36	9/36
3	5/36	6/36	1/36

Distr. bivariada discreta. Dependência.

Cada tabela tem seu *carater de dependência*.

Observe que a *independência* foi definida rigorosamente. Tudo que não seja independência, chama-se *dependência*. Esta, infelizmente, é a única definição de "dependência" que podemos ter, pois as dependências são infinitas assim como as tabelas com as marginais dadas. Pior ainda: em geral, a dependência não pode ser medida por um valor só.

Se procurarmos por experessão da dependência, acharemos que esta carateriza-se por distribuições condicionais. Por exemplo, na distribuição da primeira das três tabelas acima, o valor 3 da variável aleatória Y "puxa" a variável aleatória X a assumir o valor 2. No nível intuitivo, isto significa que se você precisasse apostar na X e estivesse sabendo que Y assumiu valor 3, então apostaria no que X assuma valor 2. No nível formal, tem-se:

$$\boldsymbol{P}\left[X=1\,\middle|\,Y=3\right]=0,\ \boldsymbol{P}\left[X=2\,\middle|\,Y=3\right]=5/6,$$
$$\boldsymbol{P}\left[X=3\,\middle|\,Y=3\right]=1/6.$$

Distr. bivariada discreta. Dependência absoluta (ou, em outros termos, completa, perfeita, funcional.

Y \ X	0	1	2	3
0	1/8	0	0	1/8
1	0	2/8	2/8	0
2	0	1/8	1/8	0

Y \ X	0	1	2	3
0	1/10	0	0	2/10
1	0	3/10	0	0
2	0	0	4/10	0

Distr. bivariada discreta. Dependência absoluta.

No primeiro dos exemplos acima, não há dependência absoluta. No segundo exemplo, Y depende completamente de X (ao saber o valor que X assumiu, podemos dizer definitivamente o valor de Y), mas X não depende de Y (sabendo o valor 0 de Y, não sabemos se X assumiu 0 ou 3). Tal "assimetria" ocorre devido ao fato que X tem mais valores a assumir que Y. Apesar desta assimetria, neste caso, dizemos que há dependência absoluta.

Distr. bivariada discreta. Dependência absoluta.

Y \ X	0	1	2	3
2	1/10	0	0	0
5	0	3/10	0	0
6	0	0	0	2/10
12	0	0	4/10	0

Neste exemplo, Y depende completamente de X, e X depende completamente de Y; diz-se que há dependência absoluta.

Distr. bivariada discreta. Dependência linear.

Y \ X	0	1	2	3
1	1/10	0	0	0
3	0	3/10	0	0
5	0	0	2/10	0
7	0	0	0	4/10

Neste exemplo, há dependência absoluta de carater *linear* (a ser chamada simplesmente por *dependência linear*):

$$Y = 2X + 1$$

Esta chama-se *positiva* (o valor de Y cresce com o crescimento do valor de X; a indicação disto é a positividade do coeficiente junto a X na equação; observe, porém, que se reescrevermos a equação na forma $X = \frac{1}{2}Y - \frac{1}{2}$, o coeficiente junto a Y tb será positivo).

Distribuição bivariada; dependência linear.

Y \ X	0	1	2	3
1	0	0	0	2/10
3	0	0	4/10	0
5	0	3/10	0	0
7	1/10	0	0	0

Neste exemplo, há dependência absoluta de carater *linear* (a ser chamada simplesmente por *dependência linear*):

$$Y = 7 - 2X$$

Esta chama-se *negativa* (o valor de Y decresce com o crescimento do valor de X; a indicação disto é a negatividade do coeficiente junto a X na equação; observe, porém, que se reescrevermos a equação na forma $X = -\frac{1}{2}Y + \frac{7}{2}$, o coeficiente junto a Y tb será negativo).

Distribuição bivariada; dependência linear.

Y \ X	0	1	2	3
1	0	0	0	1/4
3	0	0	1/4	0
5	0	1/4	0	0
7	1/4	0	0	0

Observe que são as posições dos zeros que determinam a dependência; os valores não nulos podem mudar (sem que sejam anulados). Neste exemplo, a dependência é linear negativa dada pela relação

$$Y = 7 - 2X$$

assim como era no exemplo anterior, embora as probabilidades não nulas divergem de tabela para tabela.

Distribuição bivariada; dependência linear.

Y \ X	0	1	2	3
1	0	0	0	1/4
12	0	0	1/4	0
-5	0	1/4	0	0
7	1/4	0	0	0

Observe que para a presença da dependência linear não é suficiente que (na tabela da distribuição bivariada) haja um e só um valor naõ numo em cada linha e cada coluna da tabela. Além desta condição, é preciso que os valores estejam em relação linear:

$$y_j = bx_j + a \text{ para cada } j$$

Na tabela acima, por exemplo, não há relação linear.

Distribuição bivariada; dependência linear.

Nosso objetivo corrente é apresentar uma coeficiente que sente a presença da dependência linear; o curioso é que a sensibilidade é perfeita: o coeficiente assume valor 1 ou -1 caso haver a dependência linear, e assume valores estritamente dentro do intervalo $(-1, 1)$ caso não haver. Ainda mais, quando haver independência, o coeficiente assume valor 0 (a recípoca não é válida).

O maravilhoso e poderos coeficiente chama-se *o coeficiente de correlação linear*. Quem sugeriu este termo, pensou que "correlação" soa mais bonito que "dependência".

Vale avisar: o funcionamento do coeficiente é tal milagroso, que ele está sendo empregado em quase tudo, o que, por sua vez, faz com que os usuários esquecem que existem dependências (correlações) não lineares (esquecem por que o coeficiente não é apropriado para identificar tais correlações).

Covariânica.

Recorde a definição de variância (para uma variável aleatória):

$$\mathrm{Var}\,[X] = \boldsymbol{E}\Big[\,(X - \boldsymbol{E}(X))^2\,\Big] = \boldsymbol{E}\Big[\,(X - \boldsymbol{E}(X))\,(X - \boldsymbol{E}(X))\,\Big]$$

A Covariância entre X e Y defina-se por analogia:

$$\mathrm{Cov}\,(X,\ Y) = \boldsymbol{E}\Big[\,(X - \boldsymbol{E}(X))\ \times\ (Y - \boldsymbol{E}(Y))\,\Big] \qquad (1)$$

É a fórmula que serve tanto para o caso discreto quanto para o continuo. O segundo caso será tratado adiante, e no momento, vamos expressar a fórmula nos termos mais poupáveis relacionados à distribuição bivariáda discreta:

$$\mathrm{Cov}\,(X,\ Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\Big[\,(x_i - \boldsymbol{E}(X))\ \times\ (y_j - \boldsymbol{E}(Y))\,\Big]$$

Covariância.

A fórmula primordial que definiu a covariância será logo substituida pela outra que é mais simples no cálculo do valor da covariância e mais intuitiva na interpretação da covariância como medida de dependência. A fórmula primorial serve bem às demostrações das seguintes propriedades da covariância:

PROPRIEDADE 1:
$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,\, Y)$.
COROLÁRIO DA PROPR. 1: $\mathrm{Cov}(X,\, X) = \mathrm{Var}(X)$.

PROPRIEDADE 2. Se b for uma constante e X uma variável aleatória , então $\mathrm{Cov}(b, X) = 0$.

PROPRIEDADE 3. Se a e b foram constantes e X e Y variáveis aleatórias, então $\mathrm{Cov}(aX + b,\, Y) = a\mathrm{Cov}(X,\, Y)$.

PROPRIEDADE 4. $\mathrm{Cov}(X,\, Y) = \boldsymbol{E}\,[XY] - \boldsymbol{E}\,[X]\,\boldsymbol{E}\,[Y]$

Covariância.

Demonstração de $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,\ Y)$:

$$
\begin{aligned}
\mathrm{Var}\,(X+Y) &= \boldsymbol{E}\left[\Big((X+Y) - \boldsymbol{E}(X+Y)\Big)^2\right] \\
&= \boldsymbol{E}\left[((X+Y) - \boldsymbol{E}(X) - \boldsymbol{E}(Y))^2\right] \\
&= \boldsymbol{E}\left[(\{X - \boldsymbol{E}(X)\} + \{Y - \boldsymbol{E}(Y)\})^2\right] \\
&= \boldsymbol{E}\Big[\{X - \boldsymbol{E}(X)\}^2 + \{Y - \boldsymbol{E}(Y)\}^2 + \\
&\qquad + 2\{X - \boldsymbol{E}(X)\}\{Y - \boldsymbol{E}(Y)\}\Big] \\
&= \boldsymbol{E}\Big[\{X - \boldsymbol{E}(X)\}^2\Big] + \boldsymbol{E}\Big[\{Y - \boldsymbol{E}(Y)\}^2\Big] + \\
&\qquad + 2\boldsymbol{E}\Big[\{X - \boldsymbol{E}(X)\}\{Y - \boldsymbol{E}(Y)\}\Big] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,\ Y)
\end{aligned}
$$

Covariância.

A PROPRIEDADE 4 nos dá a fórmula alternativa para a covariância:

$$\mathrm{Cov}(X,\, Y) = \boldsymbol{E}\,[XY] - \boldsymbol{E}\,[X]\,\boldsymbol{E}\,[Y]$$

Se você for usar essa fórmula para calcular a covariância, eis o detalhamento:

$$\mathrm{Cov}(X,\, Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) x_i\, y_j \;-\; \boldsymbol{E}\,[X]\,\boldsymbol{E}\,[Y]$$

(acredito que detalhar o cálculo de $\boldsymbol{E}\,[X]$ e de $\boldsymbol{E}\,[Y]$ seja disnecessário).

Covariância.

Recorde a propriedade: Se X e Y são independentes, então $\boldsymbol{E}[XY] = \boldsymbol{E}[X]\,\boldsymbol{E}[Y]$.

Isto, junto com a fórmula

$$\mathrm{Cov}(X,\, Y) = \boldsymbol{E}[XY] - \boldsymbol{E}[X]\,\boldsymbol{E}[Y]$$

dá mais uma propriedade da covariância:
PROPRIEDADE 5. Se X e Y foram independentes, então $\mathrm{Cov}(X,\, Y) = 0$.

E isto nos faz torcer para que seja válida a seguinte "PROPRIEDADE" Quanto mais perto ao zero for $\mathrm{Cov}(X,\, Y)$, mais próxima à independência está a distribuição de X e Y.

Covariância.

Infelizmente, a última propriedade não se vinga, como mostra o seguinte

EXEMPLO de distribuição bivariada de variáveis aleatórias dependentes mas com a covariância nula (livro de Bussab e Morettin, Seção 8.4)

Y \ X	0	1	2
1	3/20	3/20	2/20
2	1/20	1/20	1/20
3	4/20	1/20	3/20

Covariância.

Apesar da decepção causada pelo último exemplo, há uma grande tentação no sentido de criar da covariância uma medida de dependência:
seria um valor que mede a diferença entre a dependência e independência (e ainda indica a presença da dependência completa com caráter linear).

Por exemplo, $\mathrm{Cov}(X, Y)$ para X, Y da tabela à esquerda seria a distância entre sua distribuição real e a distribuição que X e Y teriam se fossem independentes mas preservassem suas distribuições marginais (a distribuição da tabela à direita).

Y \ X	0	1
3	1/10	2/10
5	3/10	4/10

Y \ X	0	1
3	12/100	18/100
5	28/100	42/100

Da covariância para a correlação linear.

Acontece que se a covariância for normalizada adequadamente, o resultado herderá as boas propriedades da covariância, mas adquirirá outras, as quais, em conjunto com as herdadas fará deste resultado um bom candidado para o medidor entre a independência e a dependência linear.

Definição: O coeficiente de correlação linear entre variáveis aleatórias X e Y denota-se por $\rho\Big(X, Y\Big)$ e defina-se via a seguinte relação:

$$\rho\Big(X, Y\Big) = \frac{\mathrm{Cov}\,(X,\ Y)}{\sigma\,(X)\ \sigma\,(Y)}$$

Correlação linear.

Podemos provar rigorosamente que:

(1) Se X e Y são independentes então $\rho(X,\, Y) = 0$

(2) $\rho(X,\, Y)$ nunca é maior que 1 e nunca é menor que -1.

(3) $\rho(X,\, Y) = 1$ se e somente se X e Y têm relação linear positiva.

(4) $\rho(X,\, Y) = -1$ se e somente se X e Y têm relação linear negativa.

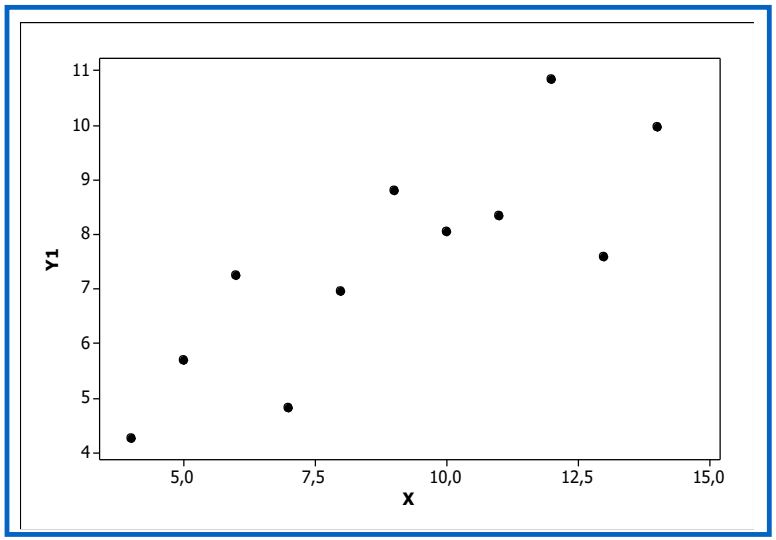
Correlação linear.

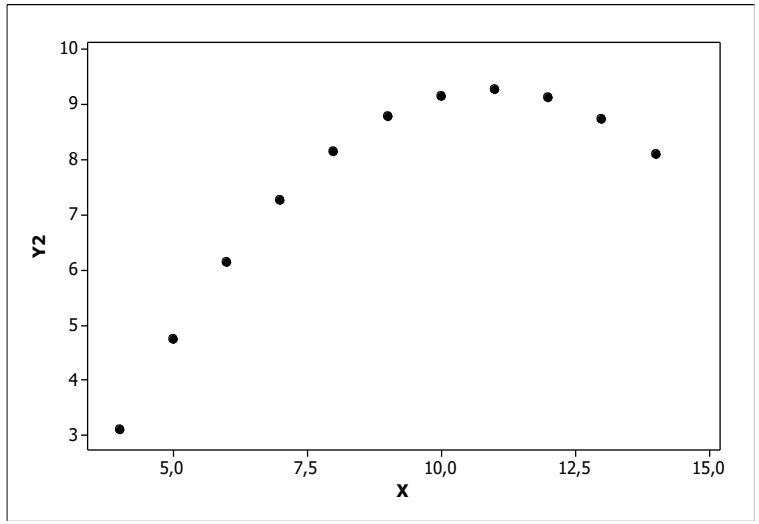
O coeficiente de correlação linear não pode ser uma medida perfeita da dependência. As razões para tal são muitas. Eis as duas que são pesadas e que merecem ser mencionadas porque refletem propriedades do coeficiente que são frequentemente ignoradas (em aplicações):

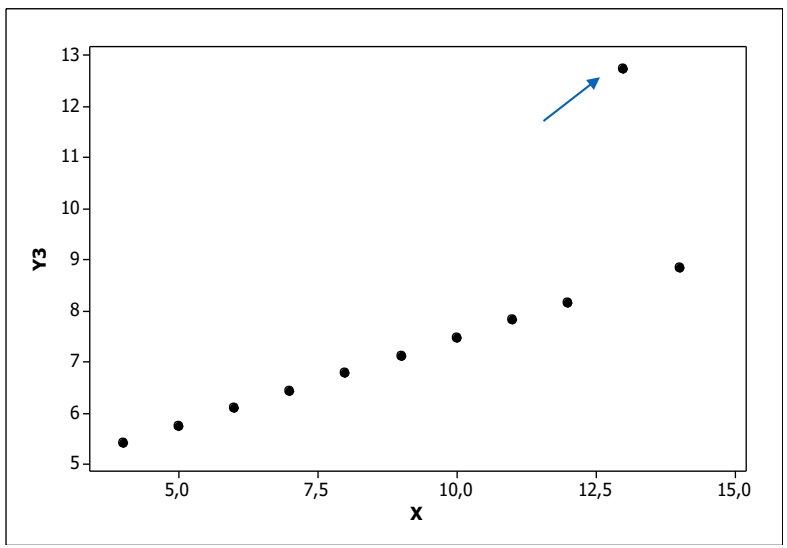
(I) Se o valor do coeficiente de correlação linear for 0 isto não garante que as variáveis sejam independentes.

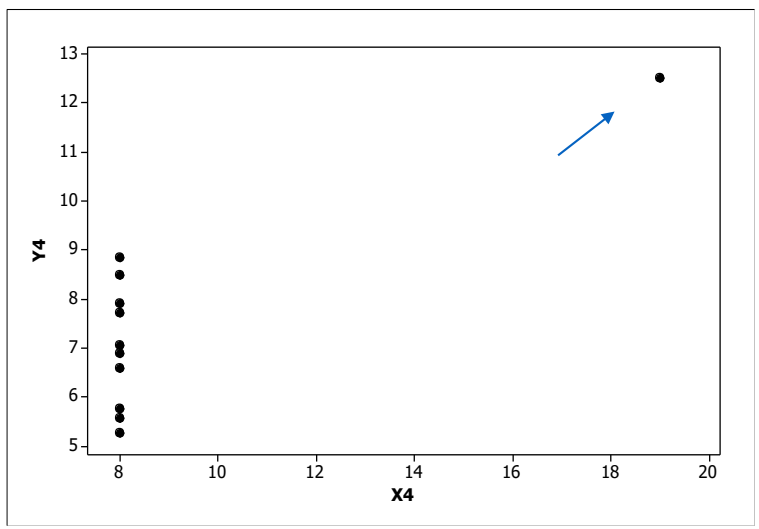
(II) Para cada valor do coeficiente de correlação linear, que não seja 1 ou -1 existe uma infinitude de dependências correspondentes a este valor. Isso está ilustrado na transparência seguinte, onde você vê 4 distribuições bivariadas diferentes que têm o mesmo valor do coeficiente de correlação linear $(0,816)$.

Correlação linear.









Cada desenho representa uma distribuição bivariada: o ponto com coordenadas x e y significa que $\boldsymbol{P}[X = x, Y = y] = 1/11$.

Distribuição bivariada; dependência linear levemente "estragada".

Y \ X	0	1	2	3	4	5	6
1	1/20-	0+	0	0	0	0	0
3	0+	2/20-	0+	0	0	0	0
5	0	0+	3/20-	0+	0	0	0
7	0	0	0+	5/20-	0+	0	0
9	0	0	0	0+	1/20-	0+	0
11	0	0	0	0	0+	2/20-	0+
13	0	0	0	0	0	0+	6/20-

Distribuição bivariada; dependência linear "estragada".

A relação entre X e Y do tipo

$$Y = a + bX + \mathcal{E}$$

onde as variáveis aleatórias X e $\mathcal{E}$ são independentes, dá a dependência linear levemente estragada (pelo $\mathcal{E}$, é claro), se a amplitude dos valores de $\mathcal{E}$ for muito menos que a de X.
O "ruido" $\mathcal{E}$ não precisa ser necessareamente independente de X.
Suponha que X pode assumir n valores $x_1, \ldots, x_n$, e suponha que há n variáveis aleatórias $\mathcal{E}_1, \ldots, \mathcal{E}_n$, e suponha que dado que X assumiu valor x_i, o correspondente valor de Y obtem-se via a fórmula

$$a + bx_i + \mathcal{E}_i$$

Então, se as amplitudes dos valores de todos os ruidos forem significativamente menores que a amplitude dos valores de X, haverá relação linear levemente estragada entre Y e X.

Distribuição bivariada; dependência linear "estragada".

Vale notar que na presença da relação

$$Y = a + bX + \mathcal{E}$$

ou da relação

$$a + bx_i + \mathcal{E}_i$$

mas com as amplitudes de valores de ruido compatíveis com a amplitude dos valores de X, podemos ainda dizer que Y depende linearmente de X com ruido. Só que neste caso, ao analizar a tabela da distribuição bivariada de X e Y será dificil adivinhar a presença desta dependência.

O funcionamento do coeficiente de correlação linear para relações lineares levemente estragadas.

Tipicamente acontece que

(1*) quando X e Y têm relação linear levemente estragada então $\rho(X, Y)$ está próximo ao 1 (ou ao -1, dependendo se a relação linear for positiva ou negativa.

(2*) Com o aumento da amplitude do ruido, que estraga a relação linear, o valor de $\rho(X, Y)$ afasta-se de 1 (ou -1) na direção do 0 (mas não tem obrigação a se aproximar ao 0).

(3*) Com o aumento da amplitude do ruido, que estraga a relação de independência, o valor de $\rho(X, Y)$ "saia" do zero.

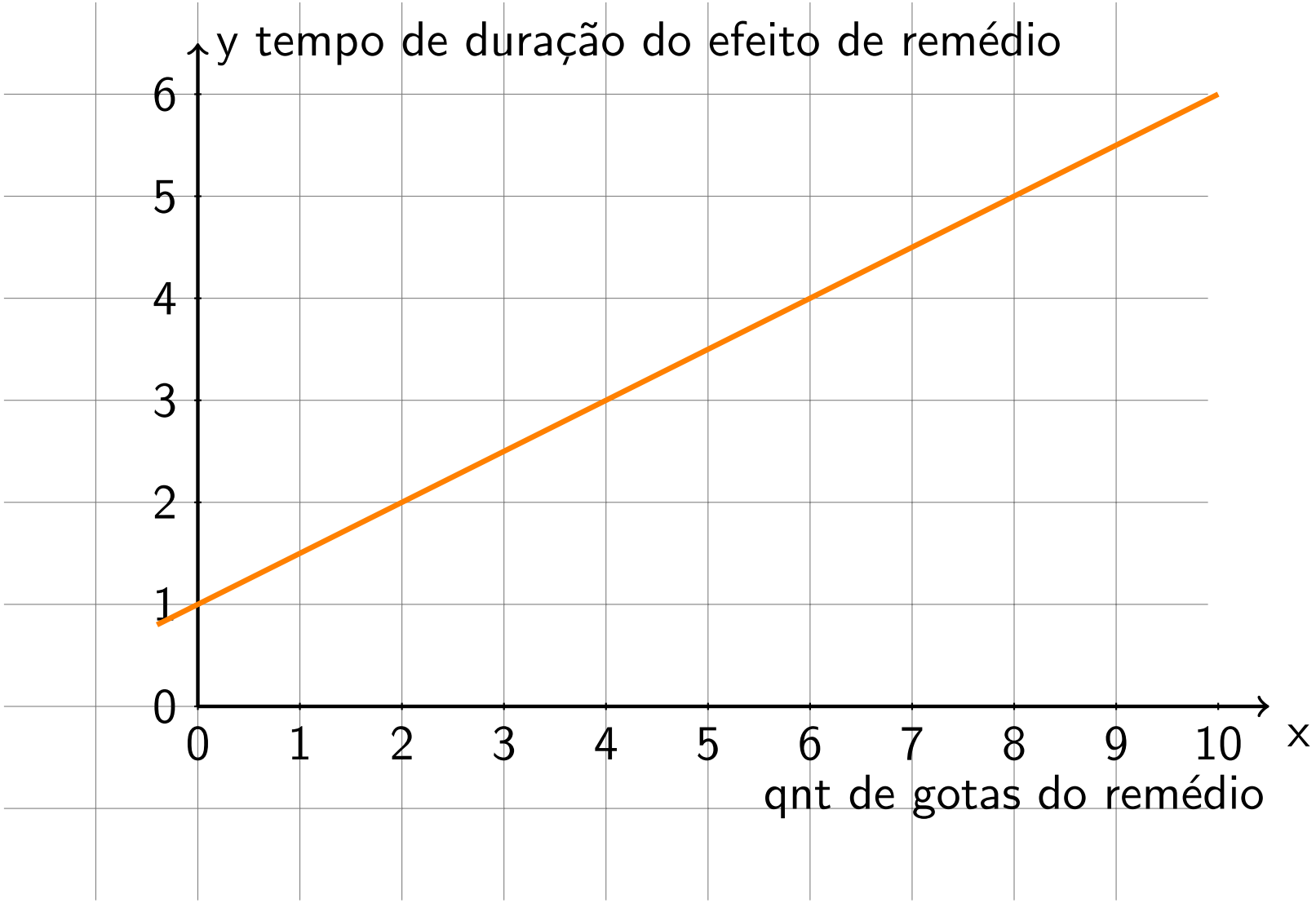
O funcionamento do coeficiente de correlação linear para relações lineares levemente estragadas.

As propriedades (1*), (2*) são empregadas para inferir na presença da relação

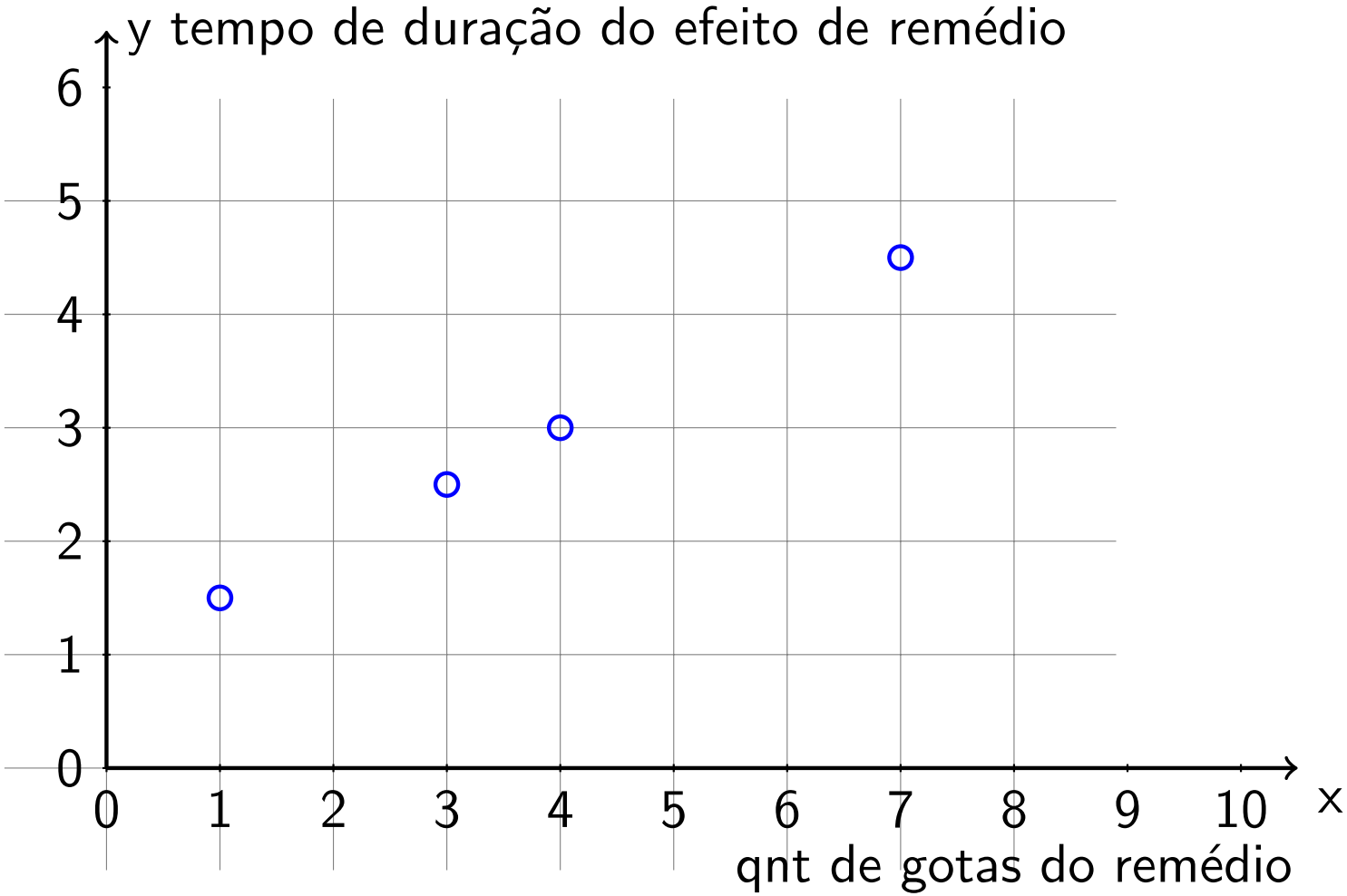
$$Y = a + bX + \mathcal{E}$$

a partir de amostra de observações simultâneas de X e Y. Veja o arquivo `Aula 3 - Descritiva III A2014.pdf`.

Se a Natureza determinou que $y = 1 + \frac{1}{2}x$ como ilustrado no desenho abaixo,

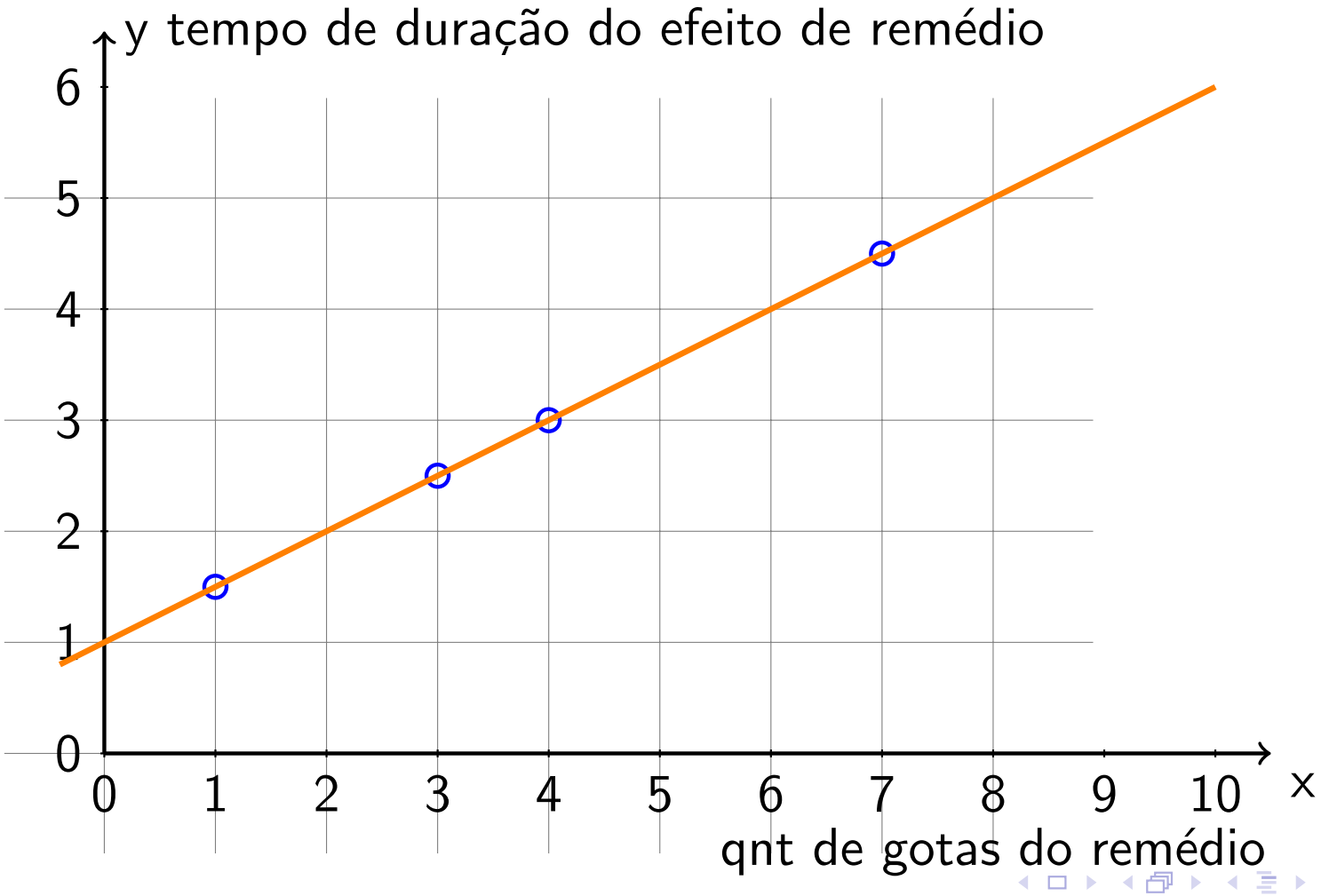


então, ao testar remédio em pessoas, os resultado será do tipo exposto no desenho abaixo:

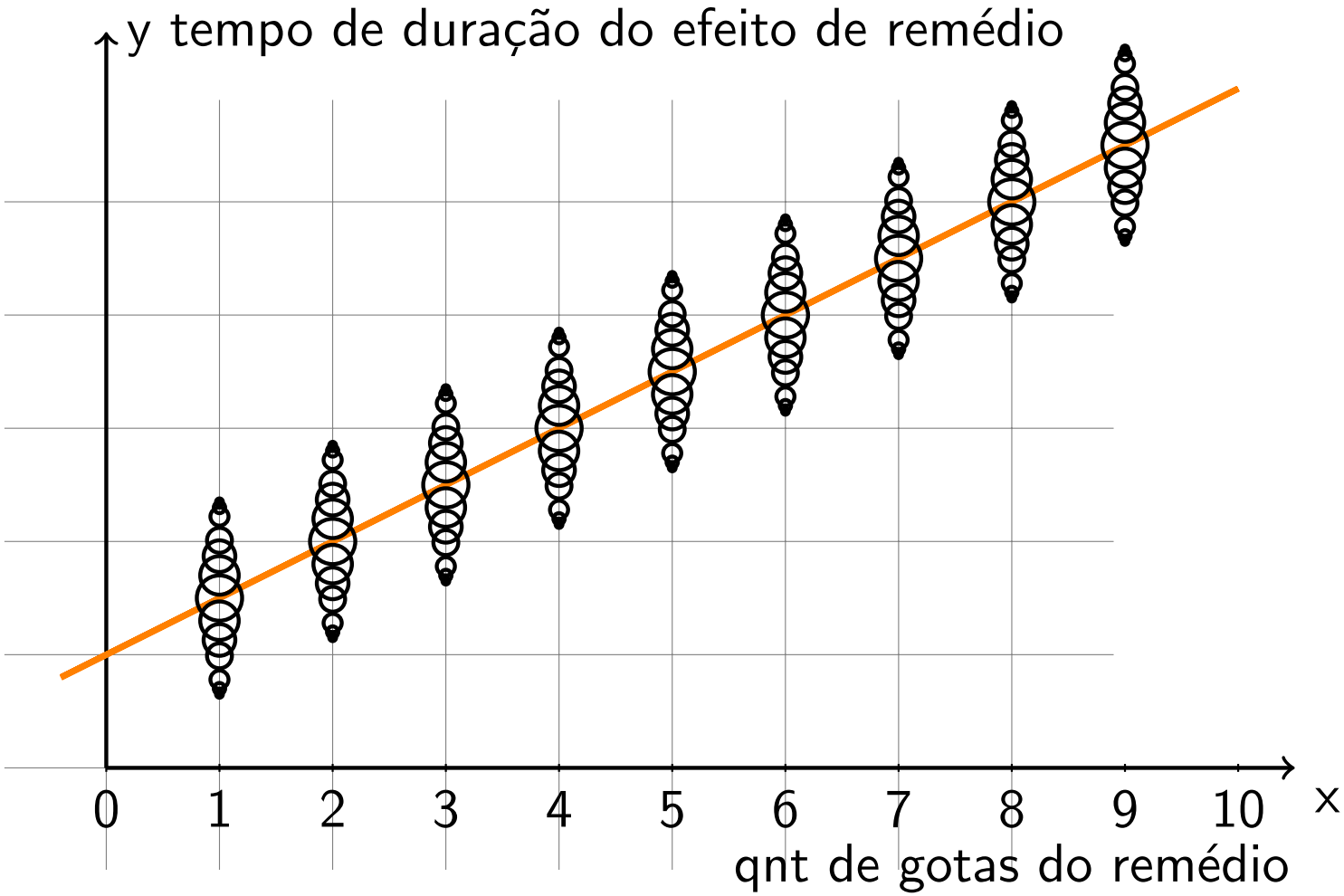


Observe que aqui não é amostra de 4 pontos só. Se derem 3 gotas de remédio a 100 pessoas, todas apresentarão 2,5 horas de duração do efeito deste, de acordo com a Lei da Natureza.

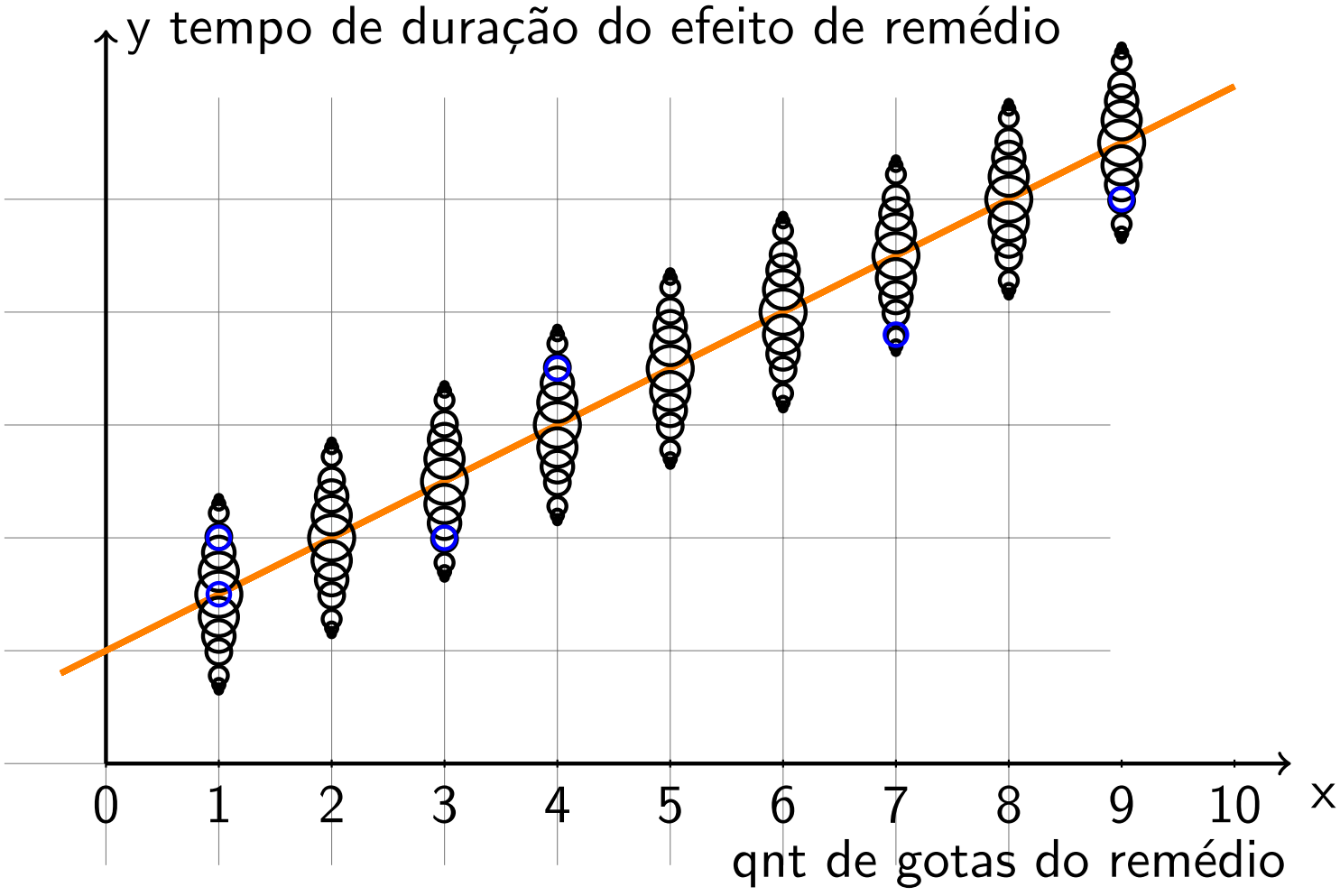
Se a gente suspeitar que a relação (entre qnte de gotas e tempo de duração do efieto) é linear, e se a gente desejar recuperar os parâmetros desta relação (a e b da eqc. $y = a + bx$) então não teremos nem dificuldade na execução desta tarefa, nem a dúvida sobre se acertamos ou não, pois há uma e só uma linha reta que passa por todos os pontas:



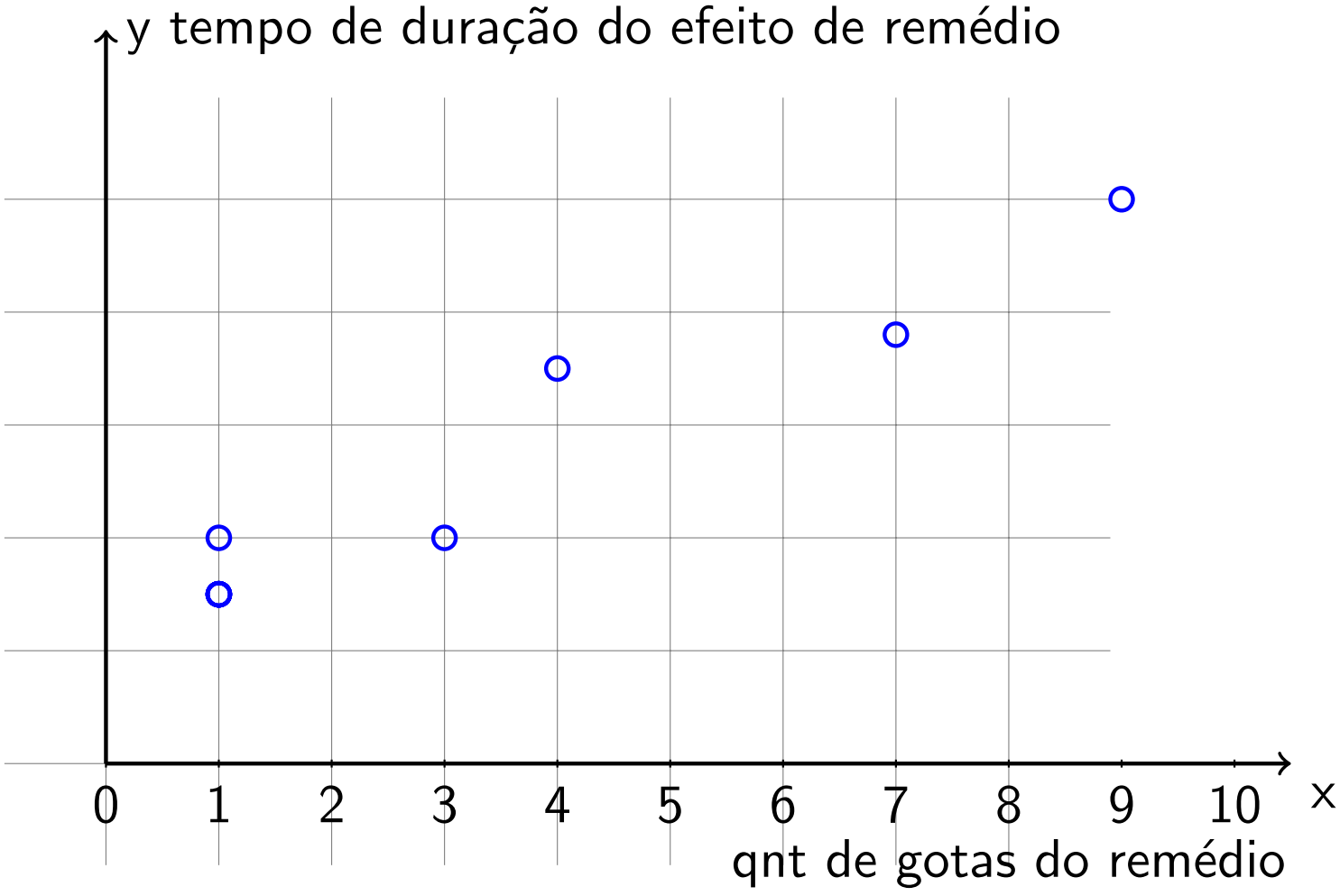
Mas se a Natureza determinou que $y = 1 + \frac{1}{2}x + \mathcal{E}$ como ilustrado no desenho abaixo,



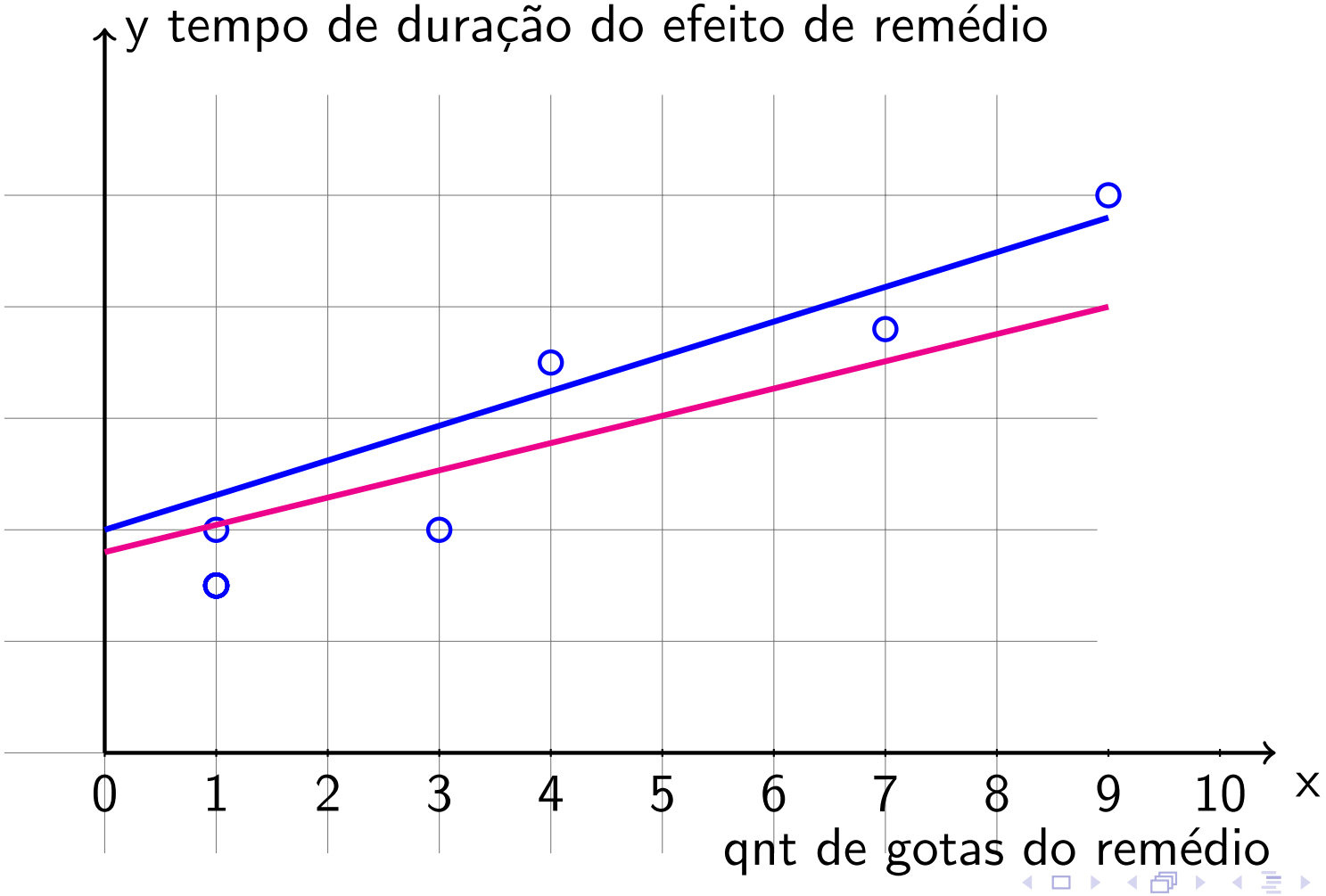
então, ao testar remédio em pessoas, os resultado (em azul) será do tipo exposto no desenho abaixo:



Sea gente desconhecer por completo a regra estabelecida pela Natureza, a gente teria só os pontos da amostra:



Agora, diferentemente, do caso anterior, nem há como suspeitar que a Natureza havia determinado $y = a + bx$ (isto é, a relação linear). Mas podemos suspeitar que haja a relação linear estragada pelo ruido digno e amigável: $y = a + bx + \mathcal{E}$ e colocamos a segunte tarefa: estimar a e b com base na amostra. As estimativas denotar-se-ão por α e β, ou, por $\hat{a}$ e $\hat{b}$. Como achar estes?



Existe um método para a determinação dos valores de α e β. Este chama-se Método de Mínimos Quadrados. Ele manda:

$$\beta = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot (\overline{x})^2}$$

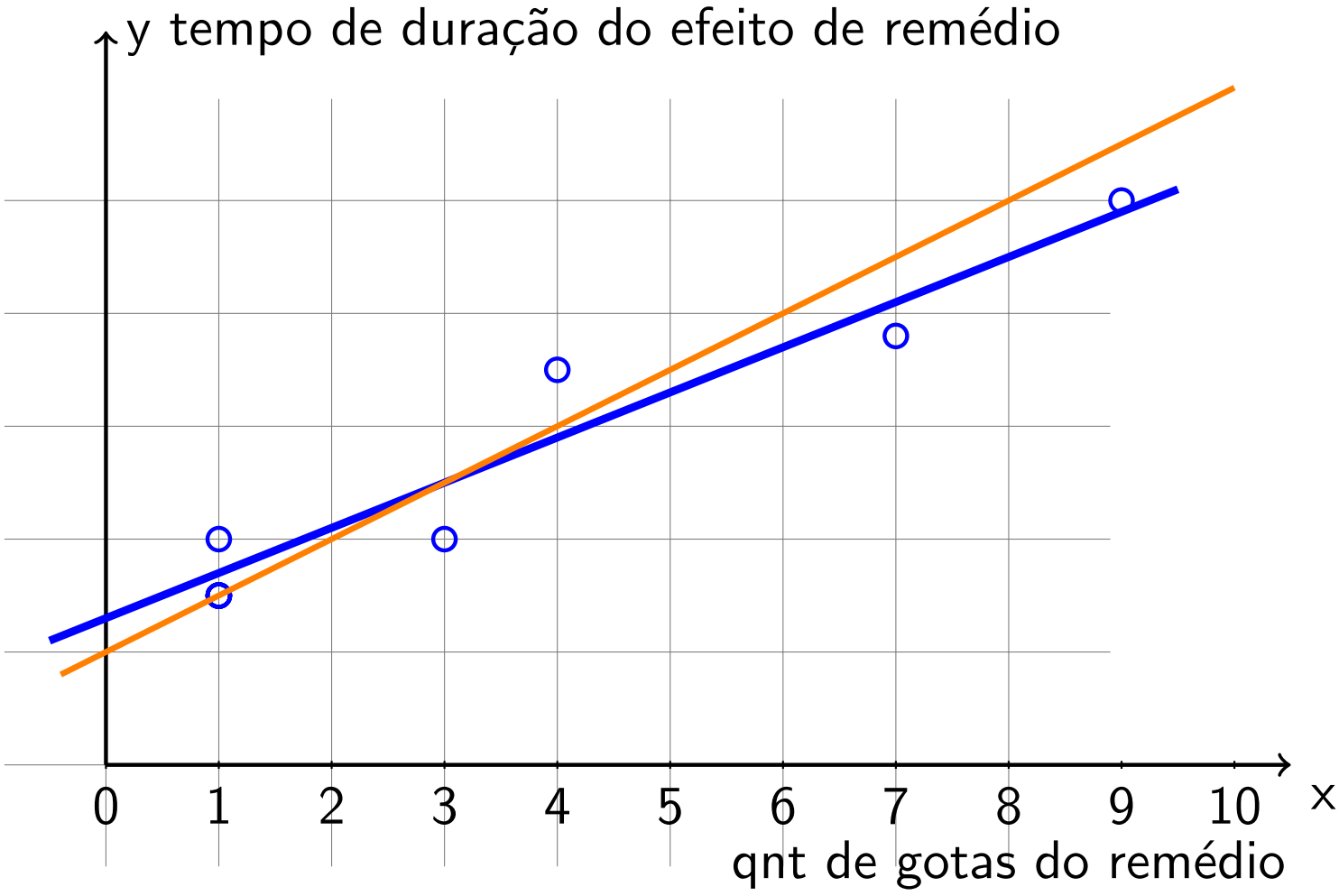
$$\alpha = \overline{y} - \beta \overline{x}$$

Para nossa amostra ($n = 6$)

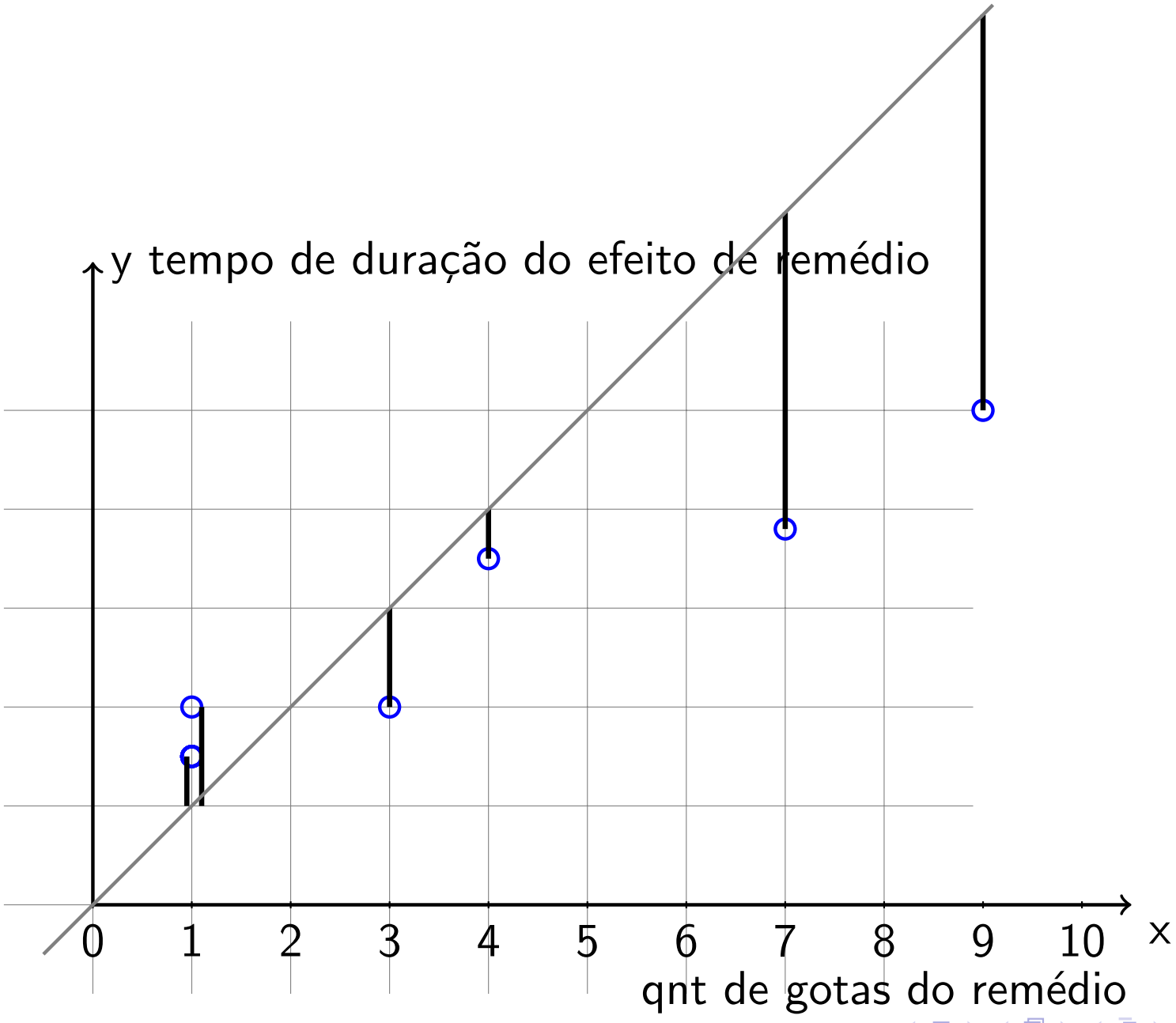
$$(x_1, y_1) = (1, 1.5),\ (x_2, y_2) = (1, 2),\ (x_3, y_3) = (3, 2),$$
$$(x_4, y_4) = (4, 3.5),\ (x_5, y_5) = (7, 3.8),\ (x_6, y_6) = (9, 5)$$

O resultado é $\alpha = 1.3158 \approx 1.3$ e $\beta 0.3962 \approx 0.4$.

Eis o desenho da regua de nossa estimativa (tb coloquei a regua verdadeira, a que tentamos estimar com base na amostra)



Para cada reta possível, calcula-se a soma dos quadrados das distâncias entre os pontos da amostra e a reta:



E escolha-se aquela reta para a qual a soma dos quadrados é a menor possível. Ela chama-se a reta de minimos quadrados.

Este método possui (no mínimo) duas vantagens:
(1) (a mais importante) quando o tamanho de amostra crescer, α calculado pelo método converge a a, e β calculado pelo método converge a b;
(2) (pouco menos importante mas muuuuito cómoda) a solução da tarefa (achar a reta cuja soma dos quadrados das distâncias é a mínima) é analítica (são aquelas fórmulas para α e β que eu apresentei acima).

Importante!

Para aplicar o método de mínimos quadrados para uma amostra $(x_1, y_1), \ldots, (x_n, y_n)$ não é necessário que haja a suspeita de que a amostra veio gerada pela regra $y = a + bx + \mathcal{E}$.

A tarefa pode ser simplesmente *traçar a reta que se ajuste à amostra para representar a parte linear da dependência*.

Ao aplicar o método, você tem que estar ciente que a posição da reta está determinada por penalizações, e que a penalização é **quadrática** em relação da distância entre ponto e reta.

A penalização pode ser de potência 4, ou 8, etc, pode ser logarítmica, pode ser exponencial. No fundo, a questão é como pontos mais afastados da reta afetem a posição desta.