



Analytics para área da saúde

# Estatística Descritiva



Variável	Média	DP	Assimetria	Curtose
Sexo	-	-	-,870	-1,243
Tempo de serviço	-	-	-,259	-,589
Formação acadêmica	-	-	1,196	2,173
Percepção emocional	67,7	10,6	-,943	3,578
Expressão emocional	64,4	8,94	-,776	3,493
Capacid. p/lidar c/emoções	72,1	10,5	-1,309	5,202
Gestão disciplina sala aula	37,7	15,1	,171	-,612

Potência Anaeróbia						
	Potência Máxima (W/Kg)			Índice de fadiga (W/s)		
	M1	M2	Δ%	M1	M2	Δ%
GO	9,47 ± 0,06	9,71 ± 0,17	2,5	7,26 ± 0,10	6,77 ± 0,43	7,2
LA	9,30 ± 0,27	9,37 ± 0,37*	0,8	7,56 ± 0,22	6,66 ± 0,53	13,5
ZA	9,49 ± 0,44	9,76 ± 0,54*	2,8	7,83 ± 0,76	6,96 ± 0,88	12,5
VO	8,84 ± 0,48	9,16 ± 0,37*	3,6	7,24 ± 0,19	6,38 ± 0,45	13,5
ME	9,07 ± 0,21	9,22 ± 0,14	1,7	7,94 ± 0,55	6,87 ± 0,54	15,6
AT	9,63 ± 0,51	9,60 ± 0,42*	0,3	7,40 ± 0,58	6,63 ± 0,88	11,6

Legenda: M1=Momento 1; Início do período preparatório, M2=Momento 2, Final do Período Preparatório; GO goleiros; LA laterais; ZA zagueiros; VO volantes; ME meias; AT atacantes; diferença estatisticamente significativa entre M1 e M2\* (p<0,05).

A estatística descritiva é um ramo da estatística que aplica várias técnicas para descrever e sumarizar um conjunto de dados. Algumas medidas que são normalmente usadas para descrever um conjunto de dados são medidas de tendência central e medidas de variabilidade ou dispersão.

Medidas de tendência central incluem média, mediana e moda. Medidas de variabilidade incluem desvio padrão, variância, valor máximo e mínimo, obliquidade e curtose.

# Variáveis quantitativas

**Variáveis Quantitativas:** são informações em uma tabela que podem ser medidas experimentalmente/empiricamente e possuem valores numéricos. Essas variáveis podem ser divididas em dois grupos, as variáveis quantitativas discretas e as contínuas. Abaixo seguem as definições de cada uma.

**Variáveis discretas:** são variáveis inteiras e positivas que em geral são resultados de alguma contagem, como por exemplo a quantidade de filhos da pessoa, ou quantidade de indivíduos em um experimento, etc. Também podem ser utilizadas de forma quantitativa, por exemplo dando graus sobre uma observação, por exemplo notas onde 1 = ruim, 2 = bom, 3 = muito bom, 4 = excelente.

**Variáveis contínuas:** são variáveis no conjunto dos números reais dentro de uma escala contínua. Por exemplo peso de um objeto em quilogramas, tamanho de um objeto em centímetros, tempo de duração de um experimento em segundos, etc.

# Variáveis qualitativas

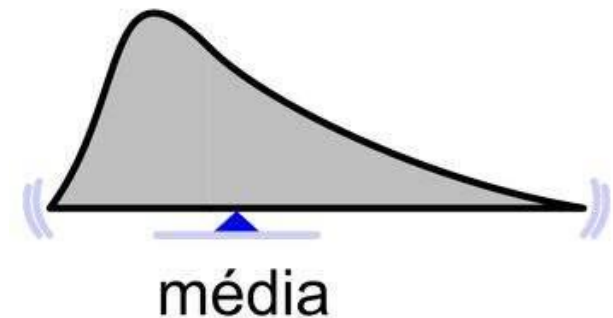
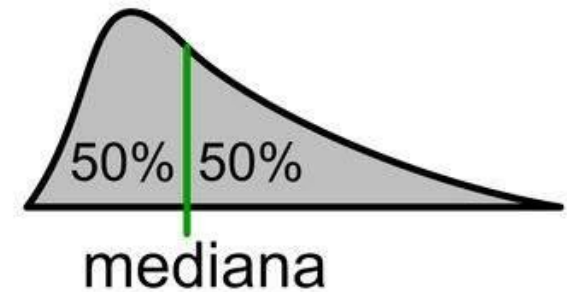
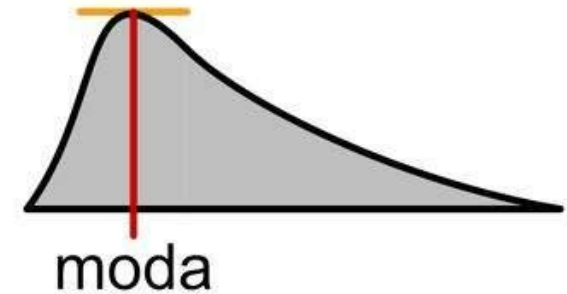
**Variáveis Qualitativas (ou categóricas):** são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Podem ser nominais ou ordinais.

**Variáveis nominais:** são variáveis apresentadas em grupos e categorias, sem nenhum tipo de ordenação. Exemplos: sexo biológico da pessoa, religião, cor da pele, etc.

**Variáveis ordinais:** são variáveis apresentadas em grupos e categorias nos quais existe uma ordenação. Exemplos: escolaridade (1o, 2o, 3o graus), estágio da doença (inicial, intermediário, terminal), mês de observação (janeiro, fevereiro,..., dezembro).

# Medidas de posição

As medidas de posição são valores que representam a tendência de concentração dos dados observados em experimentos. As mais importantes são as medidas de tendência central. As três medidas de tendência central mais utilizadas são: **média**, **moda** e **mediana**. Outras medidas de posição são os quartis, máximos, mínimos, percentis e outliers.



A média é uma das medidas de posição mais conhecidas e utilizadas. Existem diversos tipos de médias e a mais utilizada é a média aritmética amostral. Outros tipos de médias também utilizadas são média geométrica, média ponderada e média harmônica. Vamos começar os exemplos de médias com a média aritmética amostral. Supondo um conjunto de medidas  $(x_1, x_2, x_3, \dots, x_n)$  então a média desses valores é

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

# Média ponderada



Supondo um conjunto de medidas  $x$

$$x = x_1, x_2, x_3, \dots, x_n$$

Sendo  $n$  a quantidade de medidas  $x_i$ , e seja um conjunto de pesos  $w$

$$w = w_1, w_2, w_3, \dots, w_n$$

então a média ponderada amostral  $\bar{x}$  é dada pela seguinte equação.

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + w_n x_n}{w_1 + w_2 + w_3 + \dots + w_n} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

# Moda



**Moda** é o valor ou valores que **aparecem mais frequentemente**, ou seja, o “valor mais comum”. O termo moda refere-se ao **uso popular**, significando o objeto que se está usando muito no tempo presente. A moda não é necessariamente única, ao contrário da média ou da mediana. Ela é bastante útil principalmente quando os **valores não são numéricos**, e a média e a mediana podem não ser bem definidas.



# Exemplo

## Conjunto de dados

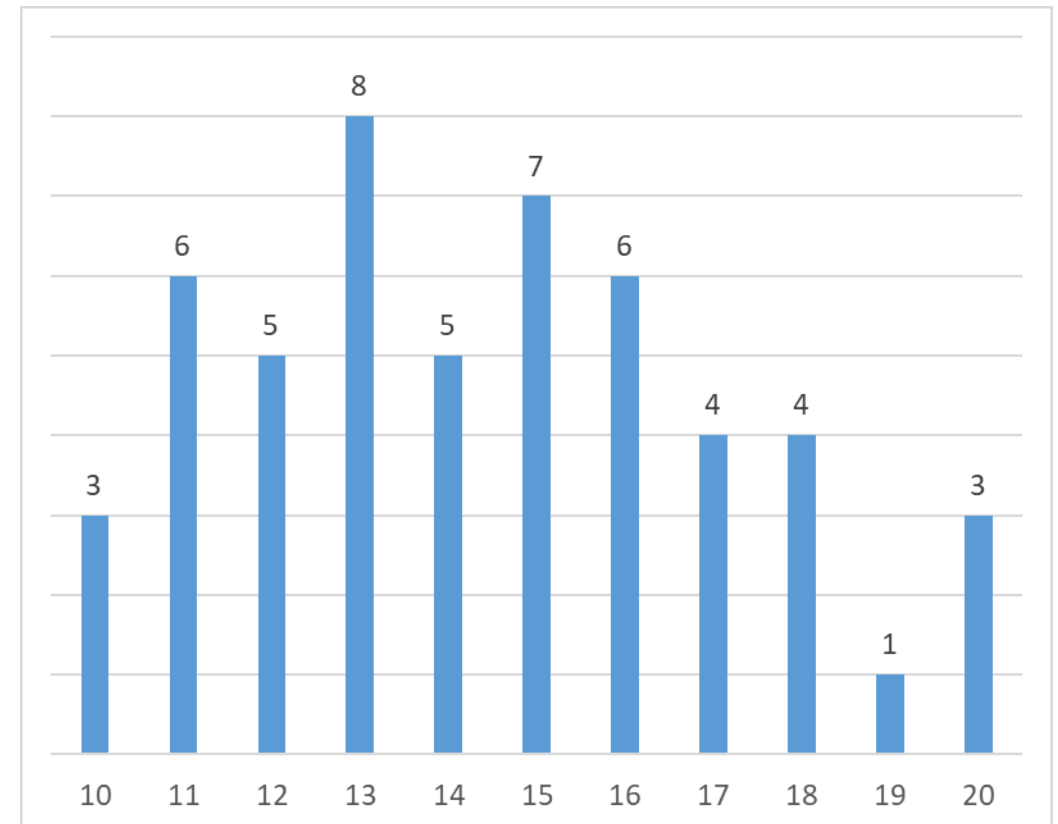
16, 14, 11, 15, 11, 20, 15, 11, 16, 14, 14, 12, 16,  
19, 18, 12, 12, 18, 10, 10, 13, 16, 11, 12, 17, 14,  
16, 16, 14, 15, 18, 18, 15, 17, 13, 13, 20, 11, 12,  
15, 15, 17, 10, 13, 13, 15, 17, 13, 13, 13, 20, 11

**Média = 14,4**

**Mediana = 14**

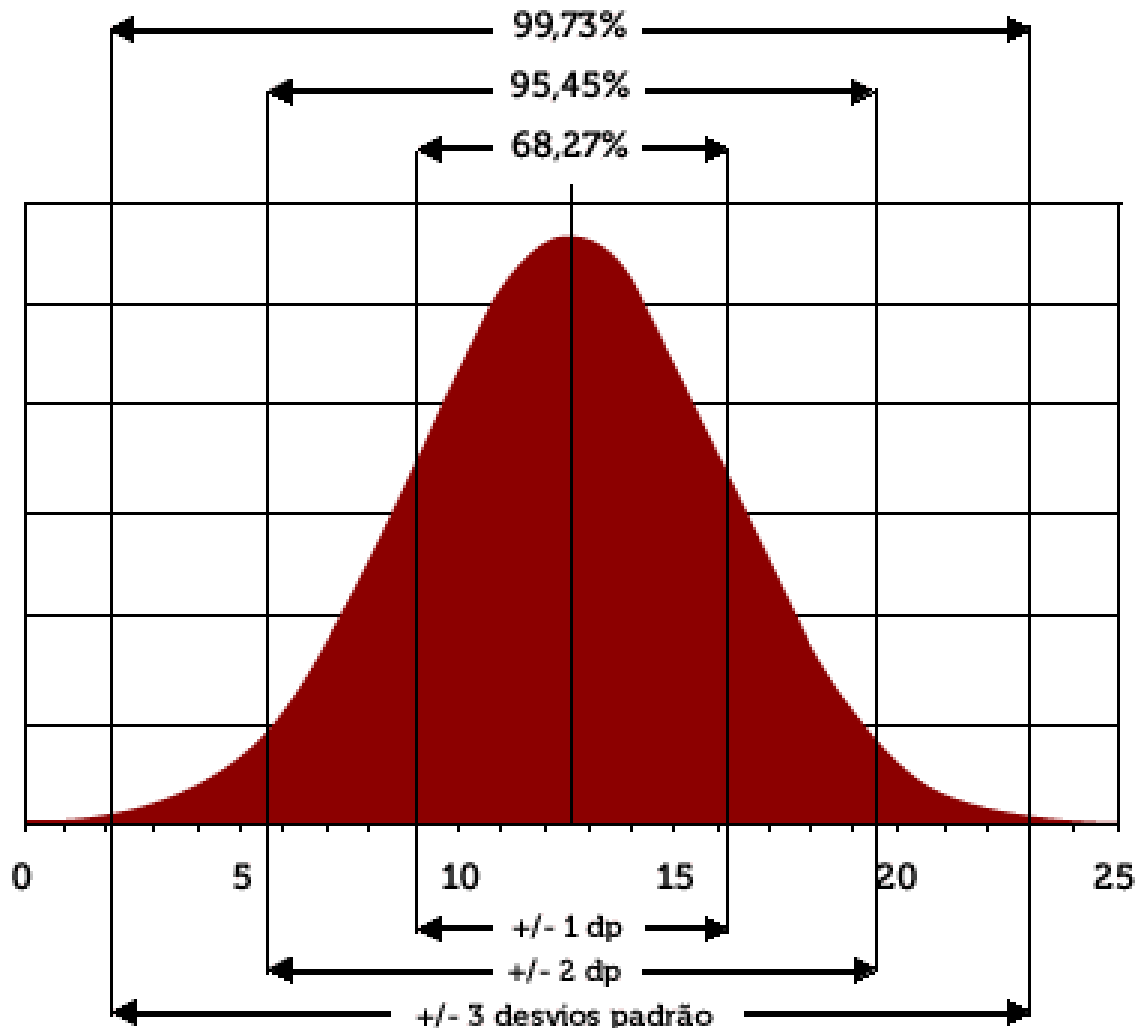
**Moda = 13**

## Distribuição de frequência dos dados



# Desvio padrão

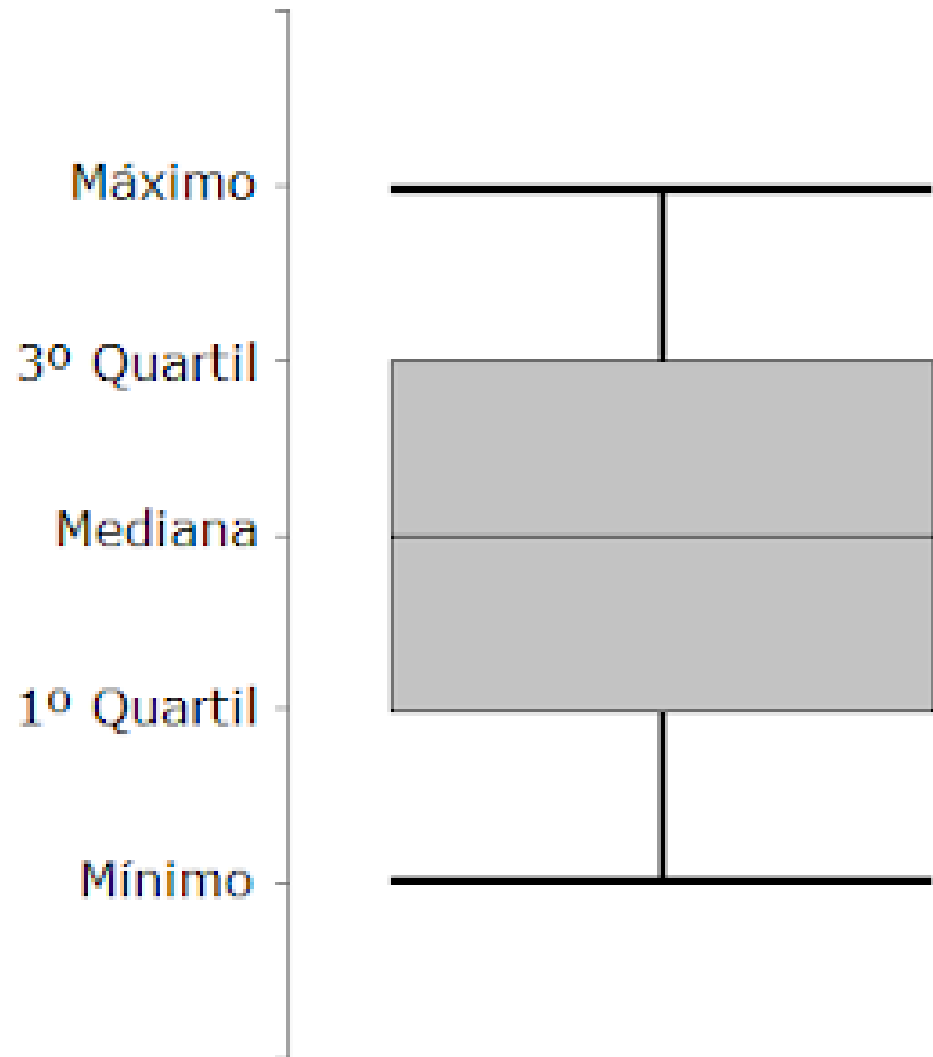
## Distribuição com Desvio 3,5



Em probabilidade, o desvio padrão ou desvio padrão populacional é uma medida de dispersão em torno da média populacional de uma variável aleatória. Também é chamado de desvio padrão amostral e indica uma medida de dispersão dos dados em torno de média amostral. Um baixo desvio padrão indica que os pontos dos dados tendem a estar próximos da média ou do valor esperado. Um alto desvio padrão indica que os pontos dos dados estão espalhados dentre os valores medidos em uma amostra ou população.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2}$$

# Box Plot



O diagrama de caixa também comumente conhecido como box plot (do inglês), é um diagrama de extremos e quartis sendo uma ferramenta gráfica para representar a variação de dados observados de uma variável numérica por meio de quartis. O box plot tem uma reta (conhecida como whisker ou fio de bigode) que se estende verticalmente ou horizontalmente a partir da caixa, indicando a variabilidade fora do quartil superior e do quartil inferior.

Os valores atípicos e/ou discrepantes, também conhecidos como outliers podem ser descritos no gráfico como pontos individuais. O diagrama de caixa apresenta a variação em uma amostra da população sem fazer nenhuma suposição sobre a distribuição estatística dos dados.

# Gráfico de Barras

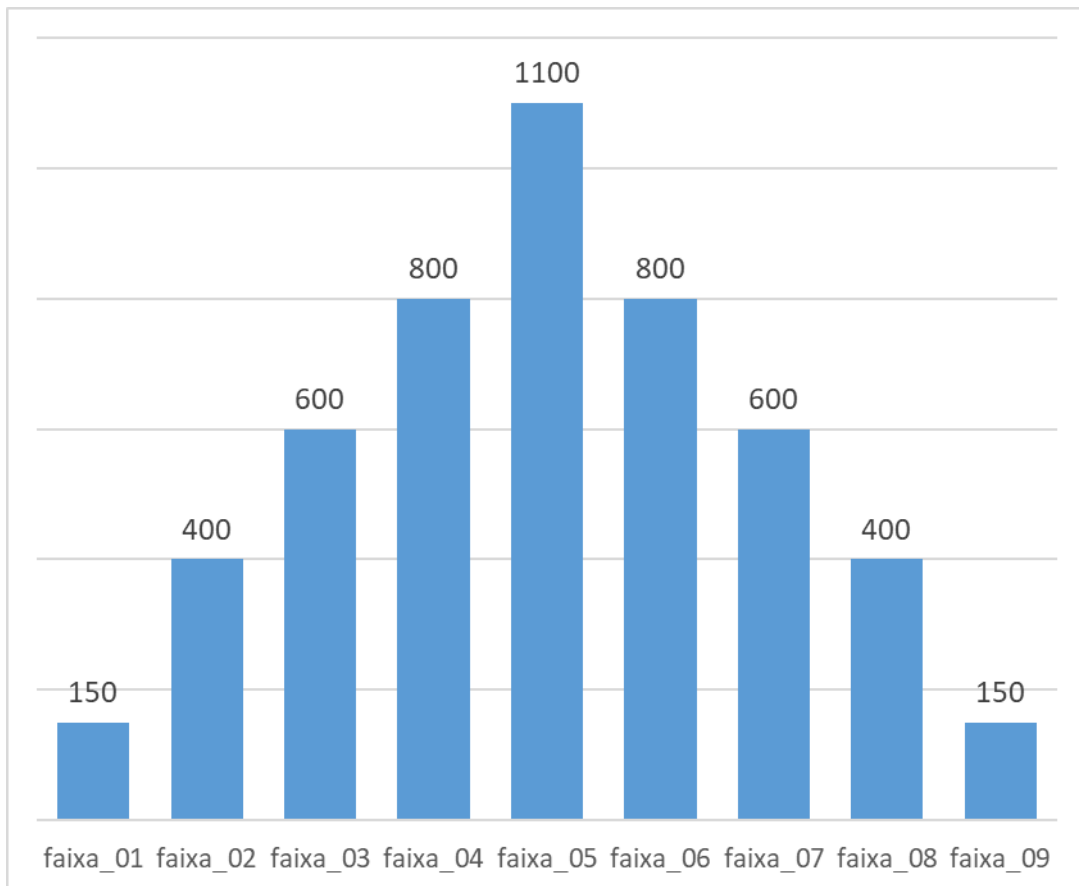


Gráfico de barras ao lado mostra a distribuição de valores de frequência de uma variável qualitativa. Note que os valores são divididos por faixas, como por exemplo faixas de pesos de pacientes.

Vamos supor que a faixa\_01 seja de pacientes com peso dentro do intervalo de 50kg a 60kg, neste caso o gráfico mostra que há 150 pessoas dentro deste intervalo de peso.

Da mesma forma supondo que a faixa\_09 seja de pacientes com peso no intervalo de 120kg à 145kg, neste caso há também 150 pacientes dentro desta faixa de peso.

# Gráfico de Dispersão

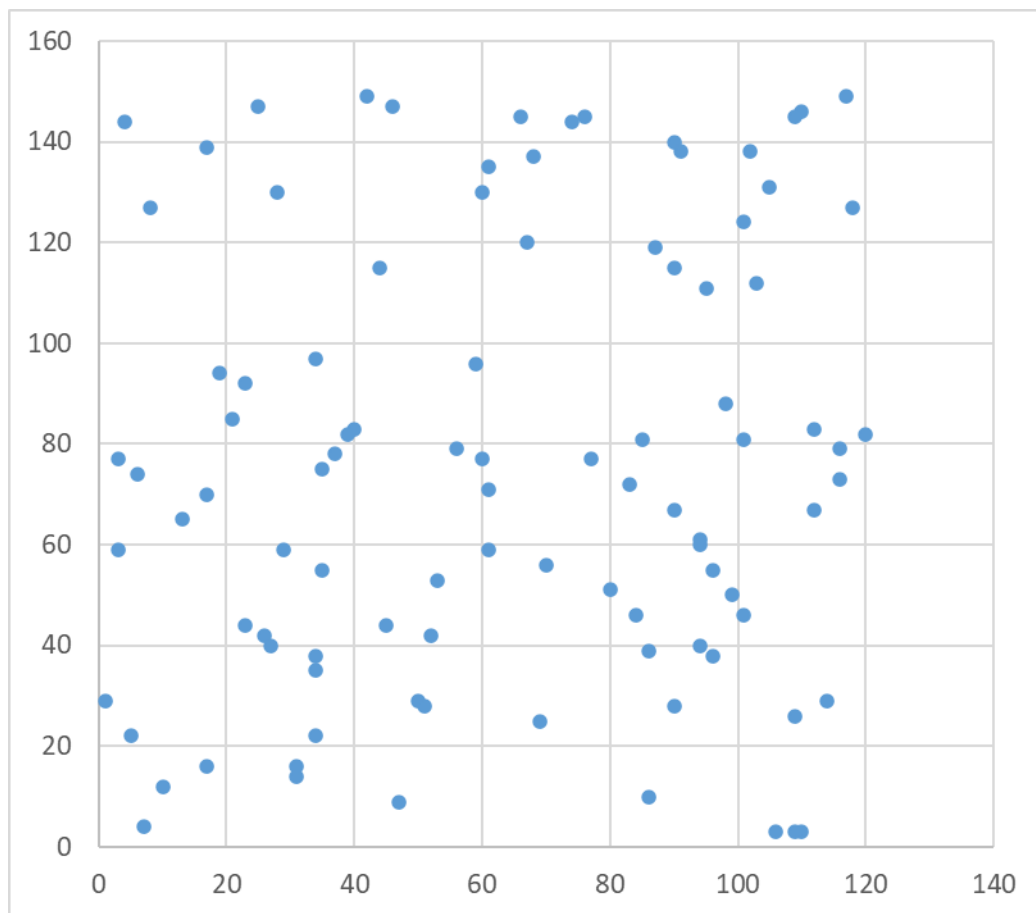
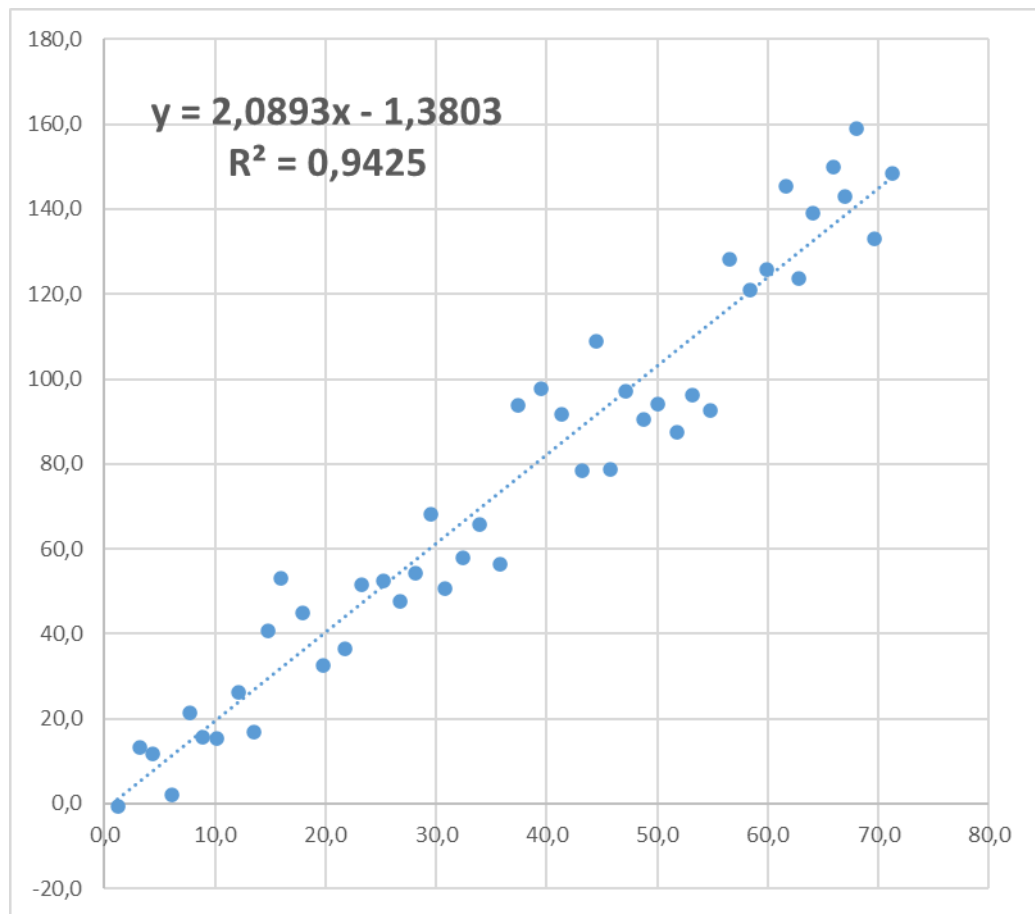


Gráfico de dispersão é muito útil para demonstrar a relação entre duas variáveis quantitativas. A distribuição conjunta dos gráficos de dispersão podem ser colocadas em tabelas de dupla entrada (duas colunas).

Em geral se designa este gráfico como um plano cartesiano com variáveis de coordenadas (x,y). No caso as variáveis no eixo horizontal são os valores da variável x, e o eixo vertical representa os valores da variável y.

No exemplo ao lado há grande dispersão dos dados que nos leva a levantar a hipótese de não haver nenhuma associação entre os dados das variáveis.

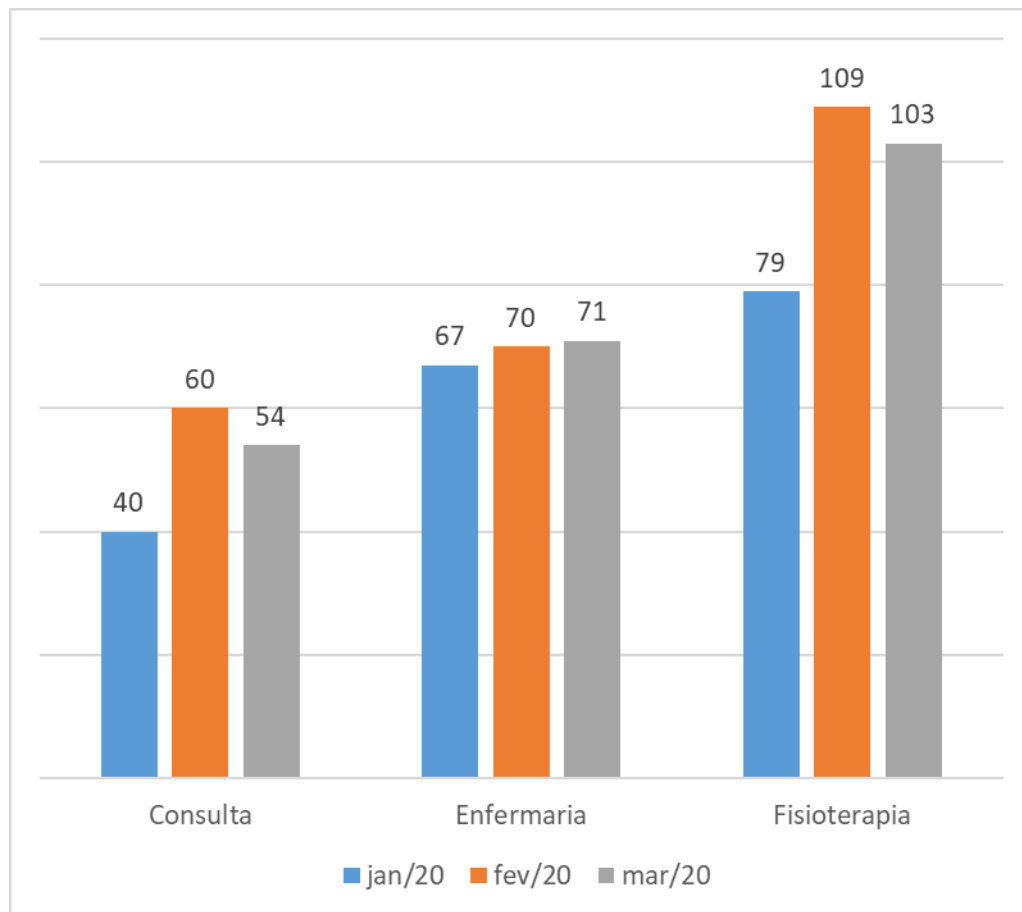
# Gráfico de Dispersão



No exemplo ao lado há uma evidente característica de associação entre os dados. A regressão linear dos dados demonstra que a relação entre as variáveis X e Y é aparentemente linear.

O valor do  $R^2$  de 0,94 indica que a linha de tendência consegue reduzir muito bem o erro médio entre os dados.

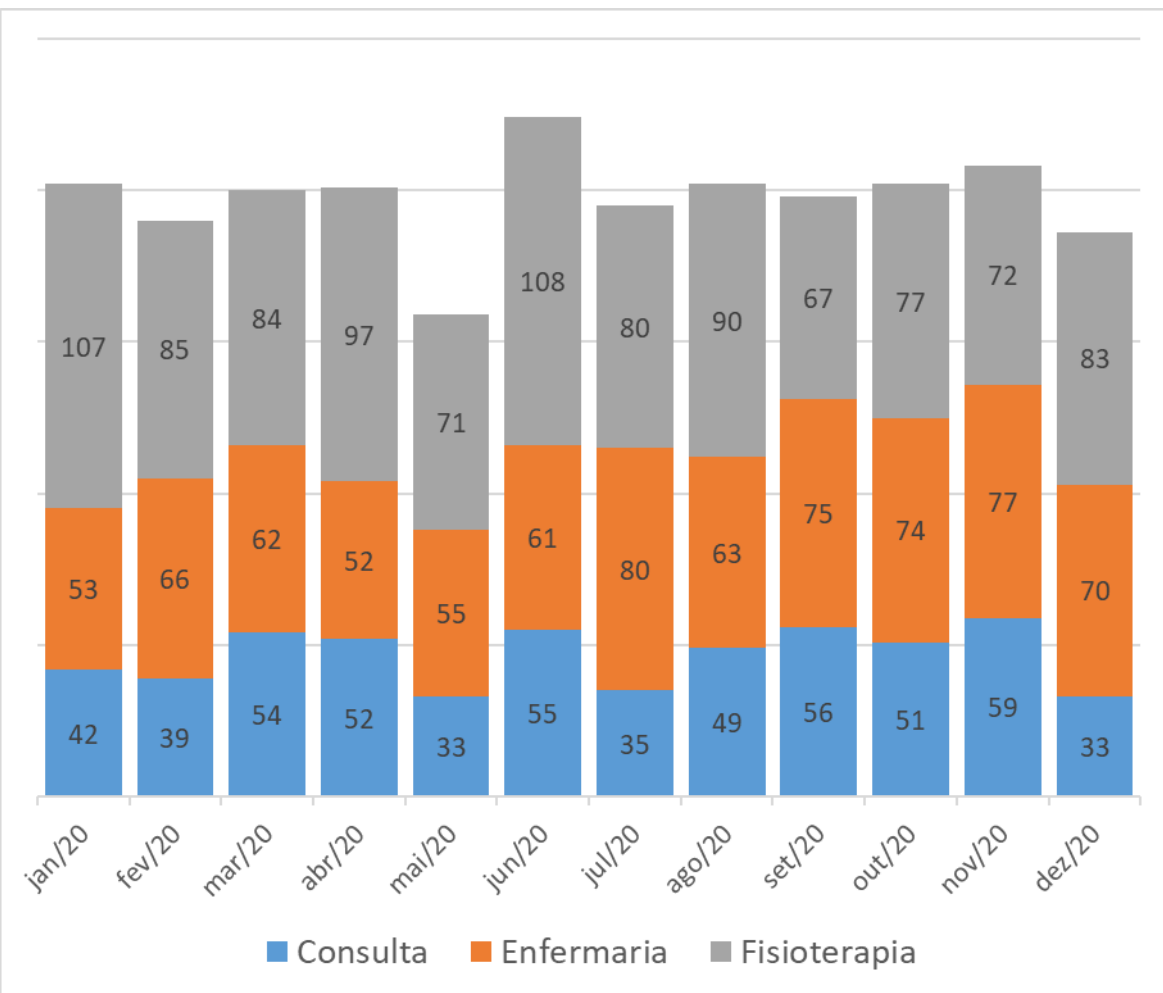
# Gráfico de Barras Comparativo



Os gráficos de barra são bastante versáteis e podem ser utilizados em diferentes tipos de visualizações. No gráfico ao lado, utilizamos os gráficos de barras para representar a quantidade de atendimentos realizados em um ambulatório que são divididos em três tipos: consulta com médico, enfermaria e fisioterapia.

Além da comparação entre os tipos de atendimento também é exibido no gráfico as quantidades de cada um ao longo do tempo, nos meses de janeiro, fevereiro e março de 2020.

# Gráficos de Barra Empilhado



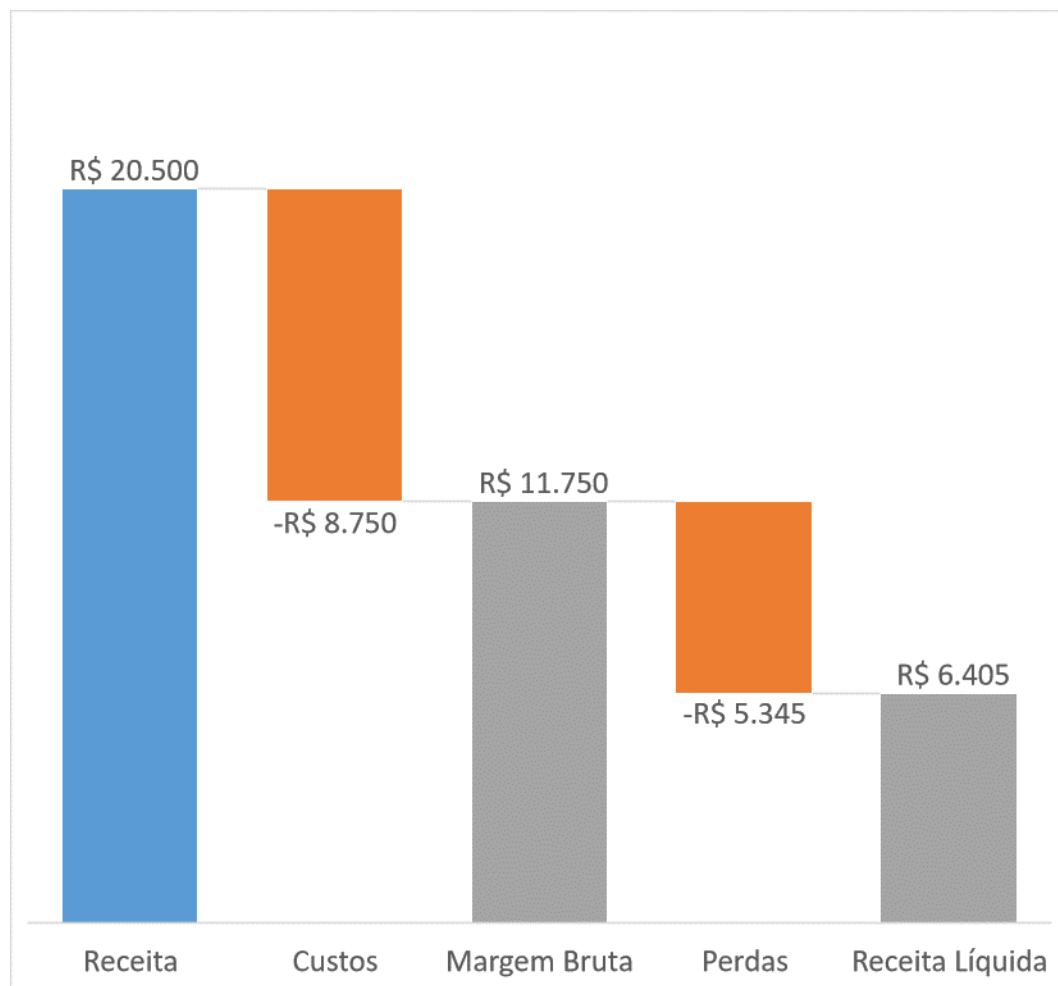
Neste exemplo, utilizamos os gráficos de barra empilhados, no qual a barra é separada nas categorias exibidas, que neste caso são os tipos de atendimento em um ambulatório. As categorias são consulta com médico, enfermaria e fisioterapia.

Os tamanhos das barras de cada cor são proporcionais às quantidades de cada tipo de atendimento. Também é possível trabalhar com percentuais que somam 100%.

Outra característica deste tipo de gráfico é acompanhar as quantidades dos tipos de atendimento ao longo do tempo, neste exemplo no ano de 2020 mês a mês.



# Gráficos de Cascata

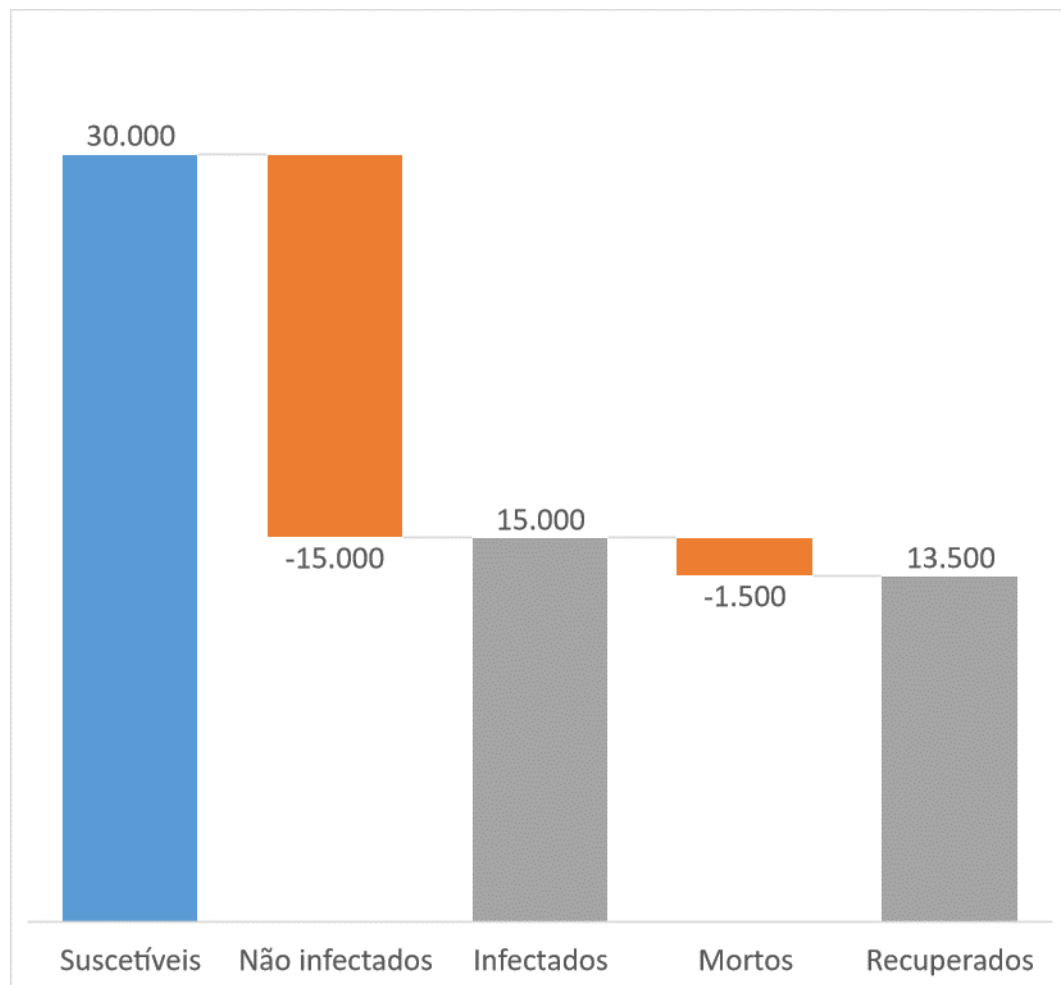


Gráficos de cascata são bastante utilizados em finanças, quando queremos fazer um demonstrativo financeiro e demonstrar visualmente o balanço de uma empresa. Também podem ser utilizados em processos que também tenham o mesmo mecanismo de finanças de entrada e saída de alguma grandeza nesse sistema avaliado. Note que existe as seguintes relações entre as barras:

Margem Bruta = Receita + Custos

Receita líquida = Receita + Custos + Perdas

# Gráficos de Cascata



Neste exemplo mostramos a utilização do gráfico de cascata no contexto de uma pandemia. A primeira barra representa o total de suscetíveis (S) ao vírus em uma dada população, a segunda barra representa o total de não infectados (NI), a terceira barra representa o total de infectados (I), a quarta barra representa o total de mortos (M) e o total de recuperados sem morte (RSM) é a última barra. Lembrando que temos a seguinte relação entre as grandezas (barras) do gráfico:

$$S = NI + I$$

$$I = M + RSM$$