



Universidade Federal de São Paulo
Escola Paulista de Medicina
Programa de Pós-Graduação em
Ciência Cirúrgica Interdisciplinar



MANUAL DE ESTUDO

Bioestatística aplicada à pesquisa clínica

Apostila de apoio para alunos de pós-graduação

Autor: **Rafael L. Pacheco**

São Paulo • outubro de 2026

Estatística descritiva

A **figura 1** mostra os primeiros 15 pacientes de uma base de dados de um ensaio clínico randomizado (ECR) que avaliou o efeito de uma intervenção hipotética em comparação ao placebo em pacientes hospitalizados com Covid-19.

Figura 1. Base de dados de um ensaio clínico randomizado hipotético

	id	grupo	idade	gravidade	peso_kg	tempo_hosp	mortalidade
1	1	Placebo	18	Moderada	69	18	Vivo
2	2	Placebo	34	Moderada	73	13	Vivo
3	3	Intervenção	18	Leve	64	21	Morte
4	4	Intervenção	41	Moderada	92	22	Vivo
5	5	Intervenção	37	Moderada	77	14	Vivo
6	6	Intervenção	39	Leve	79	18	Morte
7	7	Intervenção	23	Grave	64	23	Morte
8	8	Placebo	26	Leve	74	15	Vivo
9	9	Placebo	36	Moderada	63	15	Vivo
10	10	Intervenção	39	Moderada	62	21	Vivo
11	11	Intervenção	31	Grave	73	28	Vivo
12	12	Placebo	32	Grave	74	16	Vivo
13	13	Placebo	54	Moderada	75	9	Vivo
14	14	Placebo	18	Grave	63	13	Vivo
15	15	Intervenção	28	Moderada	59	24	Morte

A amostra total do ECR foi de 400 participantes ($N_{\text{intervenção}} = 200 / N_{\text{placebo}} = 200$). A base possui as seguintes variáveis:

- id: identificação, marcador único que distingue cada participante;
- grupo: grupo de alocação no ECR (Placebo/Intervenção);
- idade: idade em anos na linha de base (antes do início da intervenção);
- gravidade: gravidade da doença na linha de base (Leve, Moderada, Grave);
- peso_kg: peso em kgs na linha de base;
- tempo_hosp: tempo total da hospitalização em dias (desfecho secundário do ECR);
- mortalidade: mortalidade no final do estudo, avaliado pela morte durante a hospitalização ou alta hospitalar com vida (desfecho primário do ECR).

O que é uma variável?

Uma variável é uma característica mensurável que pode variar entre as observações, no caso, entre os participantes deste estudo.

Existem diferentes tipos de variáveis. Elas podem ser categóricas, como no caso da variável Grupo e Gravidade. Uma variável categórica apenas categoriza as observações.

As variáveis categóricas podem ser nominais, ou seja, apenas nomeiam as observações ou ordinais caso atribuam alguma ordem ou sentido lógico às observações. No exemplo, Grupo é uma variável categórica nominal e Gravidade é uma variável categórica ordinal, pois há uma ordem lógica nas suas categorias (leve -> moderada -> grave).

Outro tipo especial das variáveis categóricas ocorre quando só existem duas categorias possíveis. Por este motivo, Grupo e Mortalidade são consideradas como variáveis dicotômicas (“duas categorias”).

Idade, Peso e Tempo de Hospitalização são variáveis numéricas. Não é fácil definir variável numérica. Uma variável numérica representa uma quantidade, um valor que pode ser medido ou contado. Uma boa abordagem para definir se a variável é numérica é checar se a média dessa variável faz sentido.

Observar-se que é possível pensar na média de idade, peso e dias hospitalizados. Ou seja, estas variáveis são numéricas. Em contraste, não se pode classificar a variável Id como numérica, apesar de ser codificada com números. Não faz sentido pensar no número médio de Id. O Id está categorizando, identificando, nomeando, as observações. Portanto, Id é uma variável categórica.

Interpretando as principais medidas da estatística descritiva

A **tabela 1** apresenta a descrição dos grupos do estudo na linha de base.

Característica		Placebo (n = 200)	Intervenção (n = 200)
Idade (anos)		30,2 (9)	31,6 (8,4)
Gravidade da doença	Leve	35,5 % (71)	32 % (64)
	Moderada	28,5 % (57)	32 % (64)
	Grave	36 % (72)	36 % (72)
Peso (kg)		69,5 [64 – 77 / 50 – 90]	69,5 [64 – 77 / 50 – 93]

Média (Desvio-padrão) ou Mediana [Quartil 1 – Quartil 3 / Min – Máximo].

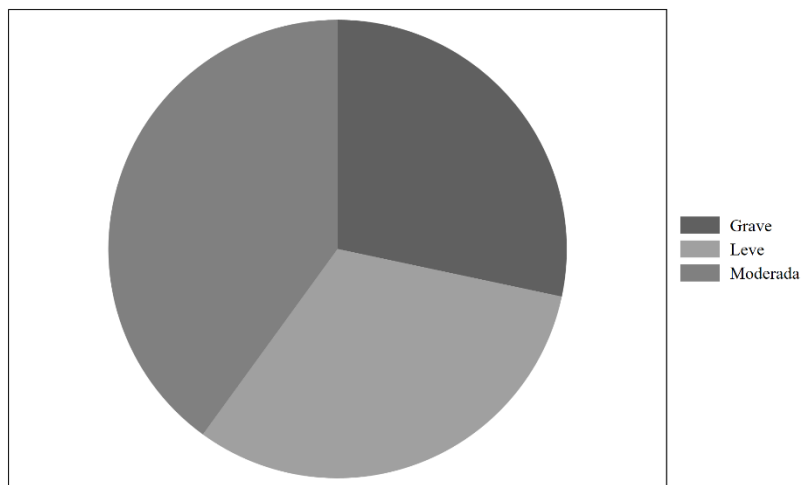
A **tabela 1** apresenta diferentes medidas da estatística descritiva que auxiliam na descrição dos grupos. A Idade foi descrita pela média e o desvio padrão, a Gravidade com porcentagem e o Peso pela mediana, quartis 1 e 3 e mínimo e máximo.

Variáveis categóricas

O modo mais frequente de apresentar uma variável categórica é por meio de porcentagens. De modo geral, é mais eficiente a apresentação de porcentagens por meio de uma tabela, como a **tabela 1**.

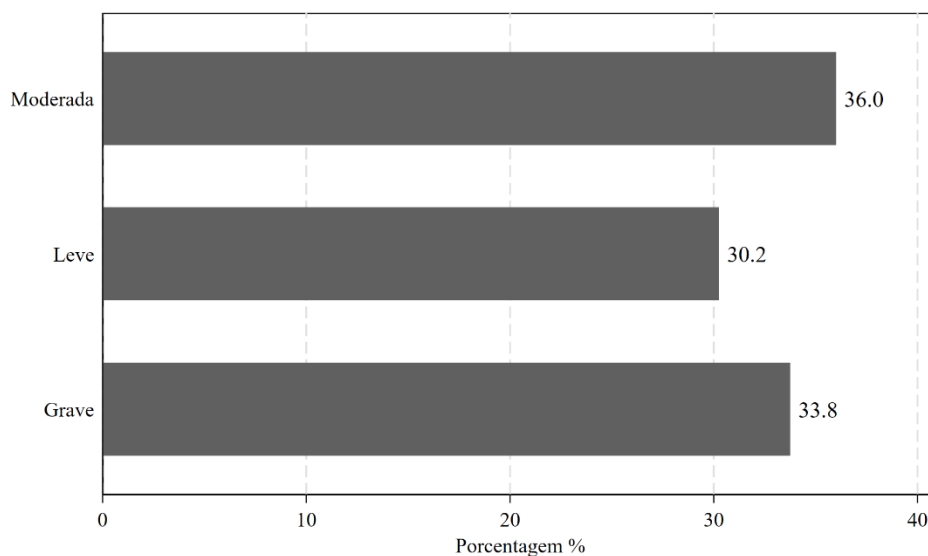
A **figura 2** mostra um modo muito triste de se apresentar variáveis categóricas. Os gráficos de pizza são muito ineficientes e quase nunca há valor em se adicionar este tipo de gráfico. A mesma informação da **figura 2** pode ser resumida em poucos números que estão indicados na **tabela 1** (35,5 % (71) / 28,5 % (57) / 36 % (72)). Além dos gráficos de pizza ocuparem muito espaço para pouca informação, visualmente podem enganar e levar a conclusões errôneas. Apresentar a porcentagem de forma numérica é muito mais preciso.

Figura 2. Gravidade da doença no grupo intervenção (um triste gráfico de pizza)



Os gráficos de barra (**figura 3**) são mais eficientes e menos enganosos que os de pizza, mas ainda assim podem ser facilmente substituídos por uma tabela.

Figura 3. Gravidade da doença no grupo intervenção (um menos triste gráfico de barras)



Variáveis numéricas

A média aritmética é a medida mais intuitiva e conhecida para se descrever uma variável numérica. Pela **tabela 1**, observa-se que as médias de idade no grupo placebo e intervenção são, de aproximadamente 30 e 31 anos, respectivamente.

Os valores mínimos e máximos também são autoexplicativos. A **tabela 1** mostra que dentre os 200 participantes do grupo placebo, o com menor peso tinha 50kg e o com maior peso, 90kg.

Alguns livros de estatística descrevem a amplitude, ou seja, a diferença entre o máximo e o mínimo. No exemplo, a amplitude do peso no grupo intervenção foi de 40kg (90-50). De modo geral, é mais informativo apresentar os valores mínimos e máximos. Quem quiser que faça a subtração.

A mediana, também conhecida como quartil 2, representa o valor máximo que “a primeira metade possui”, ordenados do menor para o maior valor. O quartil 1 e 3, por sua vez, representam o valor máximo que 25% e 75% da amostra possuem.

Pela informação da **tabela 1**, a mediana do peso do grupo placebo foi de 69,5 kg (Q1 = 64 kg, Q3 = 77 kg, Mínimo = 50 kg, Máximo = 90 kg). Como o grupo placebo possui 200 participantes, deriva-se, entre outras coisas, que:

- os 100 participantes menos pesados tinham entre 50 a 69,5 kg;
- os 50 participantes menos pesados tinham entre 50 a 64 kg;
- os 100 participantes mais pesados tinham entre 69,5 a 90 kg;
- os 50 participantes mais pesados tinham entre 77 a 90 kg.

Os quartis são medidas posicionais, que nos ajudam a dividir a amostra de maneira estratégica. Podia-se relatar percentis ao invés de quartis. De fato, a mediana nada mais é do que o quartil 2 ou o percentil 50%. O quartil 1 é o percentil 25% e o quartil 3 o percentil 75%.

O que significaria relatar que o percentil 1% do estudo foi de 51kg? Significaria que o 1% de participantes menos pesados tinham de 50 a 51kg.

Nota-se que esta interpretação destas medidas leva em conta o ordenamento sequencial dos participantes do menor ao maior valor.

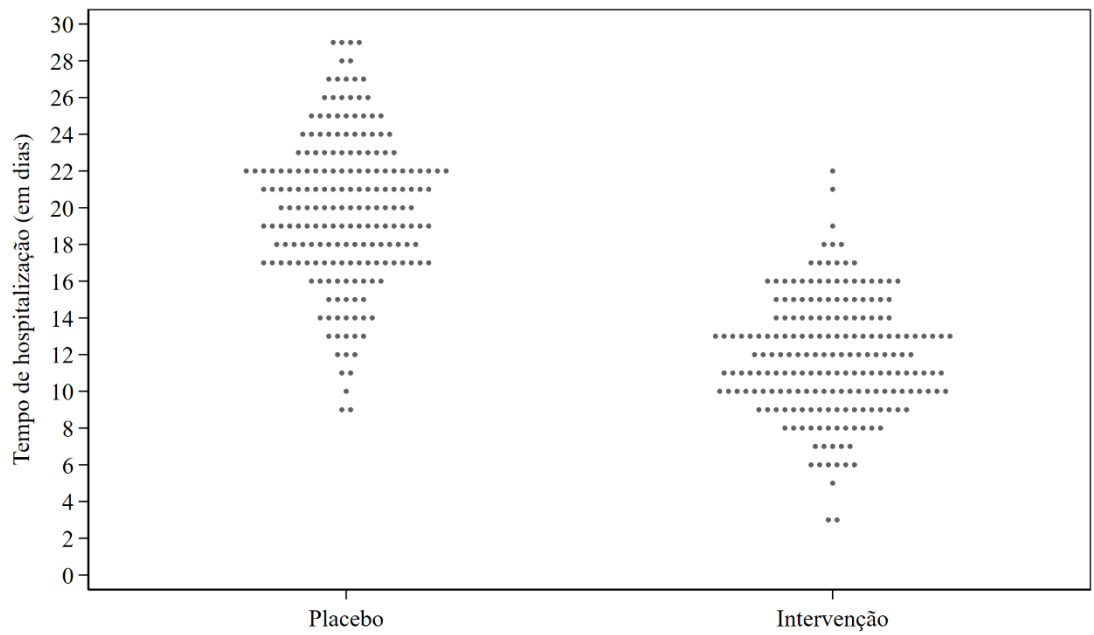
E o desvio-padrão?

O desvio-padrão é uma medida que descreve a variabilidade de dados apenas indiretamente. A regra geral de interpretação do desvio-padrão é:

- 68% da amostra estará contida entre 1 desvio-padrão da média.
- 95% da amostra estará contida entre 2 desvio-padrões da média.
- 99,97% da amostra estará contida entre 3 desvio-padrões da média.

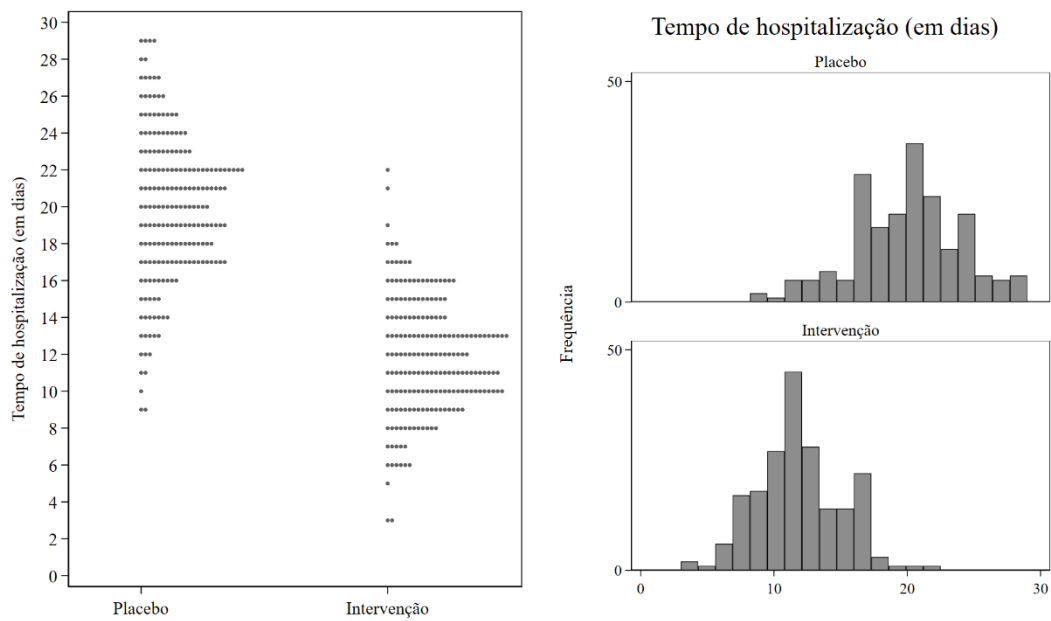
A **figura 4** apresenta um *dot plot* com a distribuição do tempo de hospitalização de cada participante por grupo. Nota-se que o tempo de hospitalização é, em média, maior no grupo placebo do que no intervenção.

Figura 4. Tempo de hospitalização por grupo.



Outra forma de apresentar a distribuição dos dados é por meio de histogramas. A **figura 5** apresenta o *dot plot* (figura 4 não centralizada) e os histogramas por grupo.

Figura 5. Tempo de hospitalização por grupo (*dot plots* e histogramas).



A **figura 6** mostra a visualização da regra geral de interpretação do desvio-padrão.

A média de tempo de hospitalização no grupo placebo foi de 20 dias (desvio-padrão = 4 dias) em comparação com 11,8 dias no grupo intervenção (desvio-padrão = 3,2 dias).

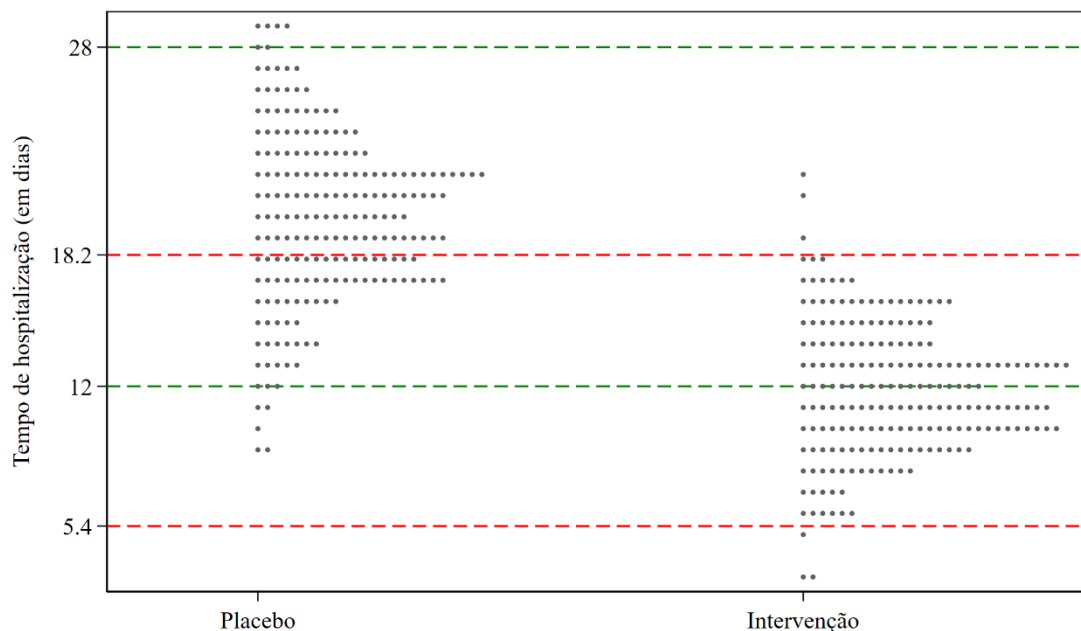
Seguindo a regra geral de interpretação do desvio-padrão, espera-se que:

- 68% dos participantes do grupo placebo tiveram entre 16 e 24 dias de hospitalização;
- 95% dos participantes do grupo placebo tiveram entre 12 e 28 dias de hospitalização (**linha verde**);
- 99,97% dos participantes do grupo placebo tiveram entre 8 e 32 dias de hospitalização;

O mesmo se aplica ao grupo intervenção:

- 68% dos participantes do grupo intervenção tiveram entre 8,6 e 15 dias de hospitalização;
- 95% dos participantes do grupo intervenção tiveram entre 5,4 e 18,2 dias de hospitalização (**linha vermelha**);
- 99,97% dos participantes do grupo intervenção tiveram entre 2,2 e 21,4 dias de hospitalização;

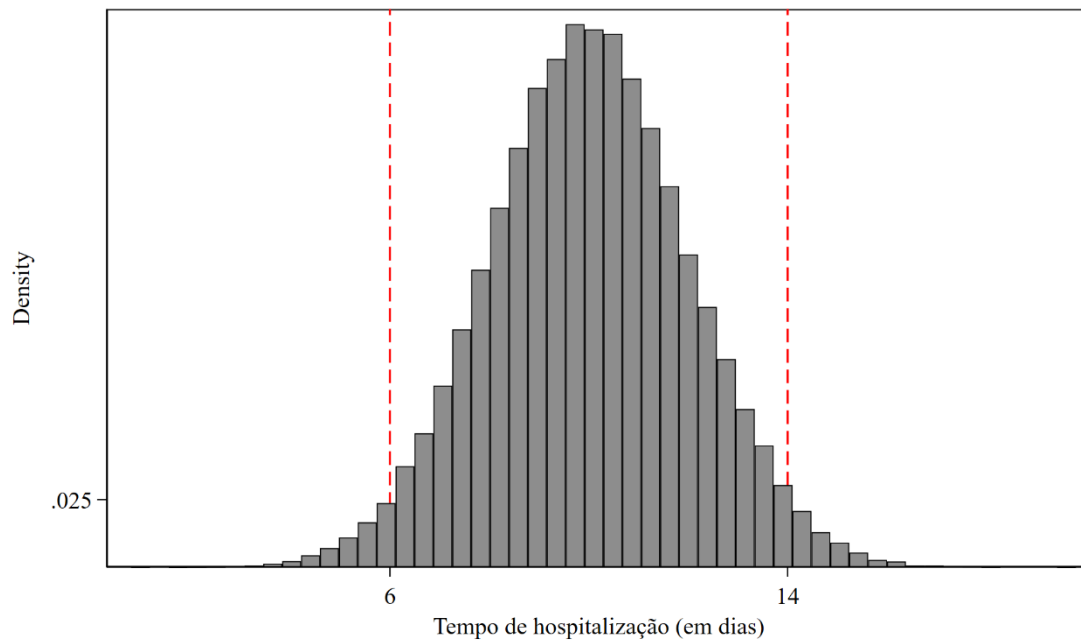
Figura 6. Tempo de hospitalização por grupo com intervalo delimitado por 2 desvio-padrões da média



A regra geral de interpretação do desvio-padrão funciona melhor quando a distribuição é normal. De fato, ela só é exata quando o tamanho amostral for grande e os dados possuírem distribuição normal.

A **figura 7** mostra o resultado de uma simulação de uma amostra de 10.000 participantes randomizada de uma população com média 10 e desvio-padrão 2. Pela regra geral de interpretação, obtém-se que aproximadamente 95% da distribuição está entre 6 e 14.

Figura 7. Regra geral de interpretação do desvio-padrão em distribuição normal.



Exercícios

Exercício 1

A tabela abaixo representa as características da população de um ensaio clínico randomizado que avaliou o efeito da **intervenção A** comparada a **intervenção B** para o tratamento de dor em metástases de pacientes com **dor lombar crônica**.

	Intervenção A (n = 1000)	Intervenção B (n = 1000)
Idade (anos)	65 [60 – 95] / A = 48 – 100	65 [55 - 75] / A = 36 - 82
Sexo (feminino)	60% (600)	60% (600)
Dor no início do estudo (escala de 0-10)	6 (2)	3 (0,5)

Dados são relatados como: Mediana [Q1 – Q3], Média (Desvio-padrão), A (amplitude) = Min - Max, Porcentagem%(total).

- Compare as populações do estudo quanto as características apresentadas na tabela.
- O que você pode concluir sobre os dados da **tabela**?

Exercício 2

Descreva o que você espera dos três grupos de pacientes a seguir que receberam Hidroxicloroquina, Ivermectina ou Nenhum tratamento em um estudo observacional:

	Hidroxicloroquina (200 Pacientes)	Ivermectina (200 Pacientes)	Nenhum tratamento (200 Pacientes)
Idade	Média = 39	Média = 39	Média = 39
	DP = 2	DP = 3	DP = 10
Sexo	50% Homens	50% Homens	90% Homens
Gravidade (0 a 10), sendo que 10 são pacientes mais graves	Mediana = 5 Q1 = 4 Q3 = 6	Mediana = 5 Q1 = 5 Q3 = 5	Mediana = 5 Q1 = 5 Q3 = 10

Exercício 3

Discuta as seguintes informações sobre o número de filhos na sala de aula ($n = 100$ alunos).

Mediana = 0 ; $Q1 = 0$; $Q3 = 0$. Percentil 99 = 0. Amplitude = 0 - 3 .

Exercício 4

A tabela abaixo apresenta as características da população de um ensaio clínico randomizado que avaliou o efeito da **intervenção A** comparada a **intervenção B** para o tratamento de asma em crianças.

	Intervenção A ($n = 1000$)	Intervenção B ($n = 1000$)
Idade (anos)	6 [3 - 9] / A = 1 - 11	1 [1 - 1] / A = 0 - 18
Sexo (feminino)	30% (300)	70% (700)
Gravidade no início do estudo (escala de 0-10, sendo que 10 são pacientes mais graves)	6 (2)	7 (0,5)

Dados são relatados como: Mediana [$Q1 - Q3$], Média (Desvio-padrão), A (amplitude) = Min - Max, Porcentagem%(total).

- Compare as populações do estudo quanto as características apresentadas na tabela.
- O que você pode concluir sobre os dados da **tabela**?

Exercício 5

Descreva, em **poucas palavras**, os três grupos de pacientes com Covid-19 que foram tratados em um hospital.

	Intervenção A (400 Pacientes)	Intervenção B (400 Pacientes)	Intervenção C (400 Pacientes)
Idade	Média = 29 DP = 5	Média = 39 DP = 3	Média = 39 DP = 20
Gravidade (0 a 10), sendo que 10 são pacientes mais graves	Mediana = 5 Q1 = 3 Q3 = 5 Percentil 99 = 5 Min = 0 Máx = 5	Mediana = 0 Q1 = 0 Q3 = 0 Percentil 99 = 0 Min = 0 Máx = 10	Mediana = 9 Q1 = 8 Q3 = 10 Percentil 5 = 8 Min = 0 Max = 10

Exercício 6

A tabela abaixo representa as características da população de um ensaio clínico randomizado que avaliou o efeito da intervenção A comparada a intervenção B para o tratamento de dor em metástases de pacientes com dor lombar crônica.

	Intervenção A (n = 1000)	Intervenção B (n = 1000)
Idade (anos)	50 [18-62] A = 18 - 99	50 [50 - 62] A = 45 - 99
Sexo (feminino)	60% (600)	60% (600)
Dor no início do estudo (escala de 0-10)	3 (1)	9 (0,5)

Dados são relatados como: Mediana [Q1 – Q3], Média (Desvio-padrão), A (amplitude) = Min - Max, Porcentagem%(total).

- Compare as populações do estudo quanto as características apresentadas na tabela.
- O que você pode concluir sobre os dados da **tabela**?

Medidas de tamanho de efeito

A **tabela 2** apresenta o resultado da avaliação dos desfechos do ensaio clínico cuja base de dados apresentada na **figura 1**.

Tabela 2. Resultado dos desfechos

Desfecho	Placebo (n = 200)	Intervenção (n = 200)
Tempo de hospitalização (em dias)	20 (4)	11,8 (3,2)
Mortalidade	50 % (100)	25 % (50)

Média (Desvio-padrão).

A maneira mais intuitiva de descrever a mortalidade é pela proporção de participantes que morreram em cada grupo. Esta proporção é denominada “risco absoluto”. O risco absoluto de morte no grupo placebo foi de 50% (100/200) em comparação a 25% (50/200) no grupo intervenção.

O *odds* (chance) é uma forma alternativa de descrever o acontecimento de eventos. Apesar de corriqueiramente se utilizar os termos risco/chance/probabilidade como sinônimos, o *odds* na estatística possui uma definição específica.

Enquanto o risco é a divisão entre quem teve o evento por quem poderia ter o evento (todos os expostos a intervenção), o *odds* é a divisão entre quem teve o evento por quem não teve o evento.

No caso, o *odds* da mortalidade no grupo placebo foi de 1 (100/100), enquanto no grupo intervenção o *odds* foi de 1/3 (50/150).

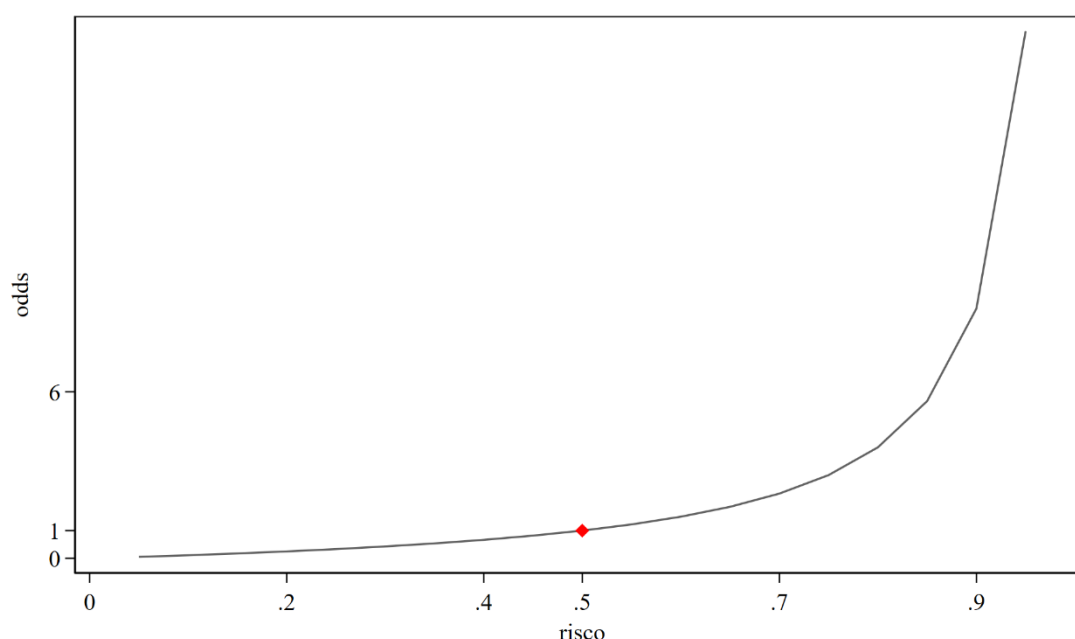
O *odds* geralmente é descrito em uma fração irredutível. No exemplo, o *odds* 1/1 (“um pra um”) do grupo placebo indica que para cada participante que morreu um ficou vivo. No grupo intervenção, para cada participante que morreu, três ficaram vivos.

Qual é o certo ? *Odds* ou riscos? Não existe uma resposta correta. Ambos são formas distintas de expressar a mesma informação. Em casos específicos, faz mais sentido utilizar um ou outro, mas ambos são medidas que transmitem a mesma informação, ou seja, a ocorrência de eventos em um grupo de participantes. Por exemplo, um *odds* de 1, corresponde um risco de 1/2.

O máximo valor de um risco é 100% (1), enquanto o *odds* tende ao infinito. De fato, um risco 1 corresponde a um *odds* de 1/0 (não definido).

Quando o risco absoluto for baixo, o número de não eventos se aproxima do total de participantes, fazendo com que o *odds* se aproxime do risco. Quanto maior o risco, mais distante o *odds* será do risco (**figura 7**).

Figura 7. Relação entre *odds* e risco



Comparando a ocorrência de eventos entre os dois grupos

Risco relativo e diferença de riscos

Ao observar a **tabela 2**, percebe-se que a ocorrência de mortes no grupo intervenção foi a metade do grupo placebo.

Esta comparação intuitiva é o **risco relativo**. O risco relativo é a divisão dos riscos absolutos. No exemplo $\text{Risco relativo} = \text{Risco intervenção} / \text{Risco placebo} = 25\% / 50\% = 1 / 2 = \text{metade}$.

Quão importante é essa redução relativa de 50%? No exemplo, o número de mortes caiu de 100 para 50 (50% para 25%). Ou seja, a intervenção reduziu em 25 a cada 100 (25%) o número absoluto de mortes.

Essa redução absoluta é definida como **diferença de risco**. A diferença de risco no exemplo foi = $25\% - 50\% = -25\% = -25$ em 100 tratados.

Durante a interpretação de um estudo, é importante destacarmos tanto a redução relativa quanto absoluta dos riscos, para melhor percepção da magnitude (tamanho) do efeito encontrado.

Odds ratio

Da mesma forma como podemos dividir os riscos para comparar os grupos, podemos dividir os *odds*. No exemplo, o *odds ratio* é = $\text{Odds intervenção} / \text{Odds placebo} = 1/3 / 1/1 = 1/3$. A interpretação é a de que o *odds* no grupo intervenção é 1/3 (33,33%) em comparação com o *odds* do placebo.

Diferença de médias

Para variáveis contínuas, usualmente se relata a diferença das médias. Segunda a tabela 2, percebe-se que a diferença de média no tempo de hospitalização foi 8,2 dias. Ou seja, em média os participantes do grupo intervenção ficaram hospitalizados 8,2 dias a menos do que os do grupo placebo.

Exercícios

Exercício 7

A tabela a seguir apresenta o resultado de um ensaio clínico randomizado que comparou a alteplase (n = 250) com a tenecteplase (n = 250) na mortalidade de pacientes com acidente vascular cerebral (AVC).

	Morto	Vivo
Alteplase	50	200
Tenecteplase	100	150

- Qual o risco absoluto de morte em cada um dos grupos?
- Qual o risco relativo do desfecho mortalidade?
- Qual é o odds ratio que compara o efeito das duas intervenções?

Exercício 8

Calcule o odds ou risco equivalente das seguintes medidas:

- Risco = $1/2$; Odds = ____
- Risco = 40% ; Odds = ____
- Risco = 1 ; Odds = ____
- Risco = $9/10$; Odds = ____
- Risco = 1%; Odds = ____
- Odds = $5/3$; Risco = ____
- Odds = 67%; Risco = ____
- Odds = $1/100$; Risco = ____
- Odds = $356/294$; Risco = ____

Estatística inferencial

Retornando à hipótese principal do estudo (**figura 1**) de que a intervenção reduz a mortalidade e o tempo de hospitalização de pacientes hospitalizados com Covid-19.

A **tabela 2**, replicada abaixo, apresenta o que aconteceu com os 400 participantes do estudo em questão.

Tabela 2. Resultado dos desfechos

Desfecho	Placebo (n = 200)	Intervenção (n = 200)
Tempo de hospitalização (em dias)	20 (4)	11,8 (3,2)
Mortalidade	50 % (100)	25 % (50)

Média (Desvio-padrão) ou Mediana [Quartil 1 – Quarto 3 / Mín – Máximo].

Como usar a informação do que aconteceu com os participantes do estudo para inferir sobre os pacientes futuros? (“os pacientes que o leitor tratará”).

De modo geral, um estudo científico é feito para generalizar resultados para além da amostra estudada. Como a informação do que aconteceu com os 400 participantes do estudo pode ajudar a inferir sobre o que acontecerá com pacientes futuros em condições similares?

Uma das ferramentas utilizadas para este processo inferencial são os **testes de hipótese** cujo resultados são **valores de p**. Existem diversos tipos de testes de hipótese, que se adequam para o tipo de hipótese e a distribuição da variável estudada.

Entender realmente o que significam os valores de p é um desafio posterior.

No momento, é importante compreender as implicações de um valor de p. Uma das possibilidades de análise do desfecho mortalidade do estudo é por meio de um teste de Chi-quadrado, cujo resultado é um valor de $p = 0,000000241756388112$.

Este valor de p é muito pequeno, e por ser bem menor que 0,05 usualmente seria considerado “estatisticamente significativo” na medicina.

A implicação usual deste resultado é a de que podemos afastar com segurança a hipótese nula. O que isso significa?

A hipótese nula varia de estudo para estudo, portanto o primeiro passo é identificar qual é a hipótese nula da análise em questão. No estudo, a hipótese principal (diz-se alternativa na estatística) é a de que a intervenção reduz o risco de morte. A hipótese nula, portanto, é a de que o risco de morte no grupo placebo e intervenção são idênticos, ou, que o risco relativo de mortalidade é 1.

Como o valor de p é menor que 0,05, rejeita-se a hipótese de que o risco relativo de mortalidade =1, aceitando a hipótese alternativa (a de que há diferença da mortalidade entre os grupos).

A implicação de um valor de $p > 0,05$

O principal erro de interpretação estatística ocorre quando o resultado de um teste de hipótese é “não estatisticamente significativo” ($p > 0,05$).

A implicação de um valor de $p > 0,05$ é a de que não se pode afastar com segurança a hipótese nula.

Isto **não significa que se aceite a hipótese nula como verdade**. Uma das interpretações mais encontradas na literatura é de que : “ $p > 0,05$, não há diferença entre os grupos”. Esta interpretação é danosa pois leva o leitor a aceitar a hipótese nula como verdadeira.

Um $p > 0,05$ apenas significa que não há evidência suficiente para afastar a hipótese nula com segurança. A hipótese nula pode ser verdadeira ou simplesmente pode não existir informação suficiente no estudo para afastá-la com o grau de segurança pré-estabelecido.

A confusão com valores de p é tão grande que muitos estatísticos e periódicos científicos recomendam a não utilização de testes de hipótese – preferindo-se a utilização de **intervalos de confiança**.

Intervalos de confiança

O risco relativo do estudo é $RR = (25/200) / (50/200) = 0,5$. Ou seja, entre os 400 participantes do estudo, houve uma redução relativa de 50% (metade) no risco de morte. O que aconteceu com os 400 participantes é fato e não é preciso de nenhuma ferramenta inferencial para afirmar que, dentre os 400 participantes randomizados, os que receberam a intervenção morreram menos.

A grande questão é: pode se esperar uma redução de 50% na população? Nos pacientes semelhantes que ainda serão tratados?

O intervalo de confiança de 95% do desfecho é $RR\ 0,50$ (IC95% 0,38 a 0,66). O que isso significa?

O intervalo de confiança pode ser interpretado como um intervalo de *compatibilidade*. O resultado do estudo mostra que o efeito da intervenção é compatível com uma redução de 62% a uma redução de 44%.

Ou seja, no melhor cenário (menos conservador) a intervenção pode reduzir 62% a mortalidade na população e no pior cenário (mais conservador) a redução pode ser de 44%.

Mesmo no cenário mais conservador, a redução de mortalidade é importante clinicamente. Deste modo, espera-se que a intervenção reduza de maneira importante a mortalidade de pacientes hospitalizados com Covid-19.

A relação entre intervalo de confiança e valor de p

Não é coincidência que por convenção a medicina adote um intervalo de confiança de 95% e um nível de significância de 5% (0,05).

Em um segundo caso hipotético, o resultado do desfecho foi $RR\ 0,75$ (IC95% 0,26 a 2,16). Na amostra do estudo, observou-se uma redução de 25% no risco de morte com a intervenção.

Qual o efeito compatível na população? O IC95% mostra que o efeito pode ser de uma redução de 74% na mortalidade a um aumento de 116% na mortalidade. Ou seja, não se sabe nem a direção deste efeito (benefício ou dano). De fato, não há informação suficiente nem para afastar um dano importante, uma não alteração ($RR = 1$) ou um aumento importante da mortalidade com a intervenção.

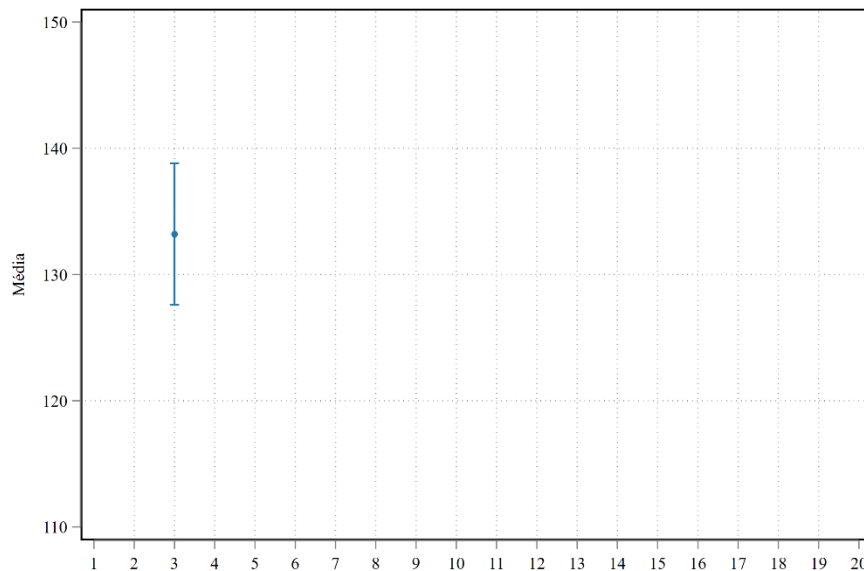
Não é preciso nenhum cálculo para saber que neste segundo exemplo o valor de $p > 0,05$. Ora, se o intervalo de confiança de 95% contempla a hipótese nula ($RR = 1$), é redundante afirmar que não podemos afastar com segurança a hipótese nula, ou, é redundante afirmar que não há significância estatística considerando um nível de significância de 5%.

Algumas simulações para se visualizar intervalos de confiança

O que é um intervalo de confiança? É mais fácil visualizar do que definir.

A **figura 8** mostra o intervalo de confiança de um estudo que mediu a pressão sistólica de 100 brasileiros. A média estimada da pressão sistólica foi de 130,8 mmHg (IC 95% de 127,6 a 138,8).

Figura 8. Média (IC95%) de pressão sistólica de uma amostra aleatória de 100 brasileiros

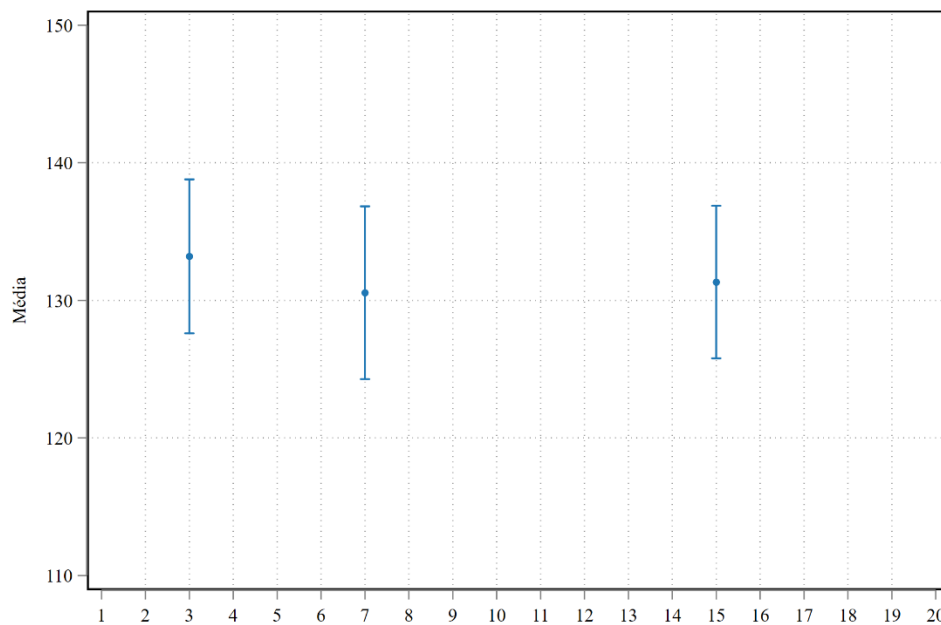


Ou seja, pode-se afirmar que se está confiante de que a média de pressão sistólica entre toda a população do Brasil está entre 127,6 e 138,8 mmHg.

Como saber qual é a exatamente a média? Não se sabe – apenas se se mensurasse a pressão de toda a população.

A figura 9 mostra o mesmo experimento sendo replicado em outras duas amostras aleatórias de 100 participantes (**Figura 9**).

Figura 9. Resultado de 3 estudos distintos com 100 brasileiros

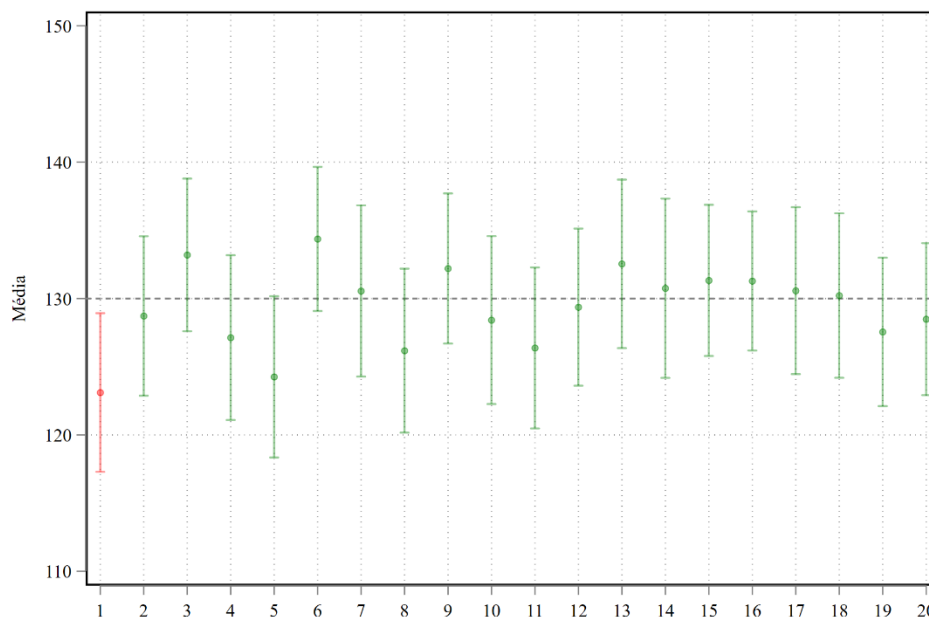


- Estudo 1: 130,8 mmHg (IC 95% de 127,6 a 138,8)
- Estudo 2: 130,5 mmHg (IC 95% de 124,3 a 136,8)
- Estudo 3: 131,3 mmHg (IC 95% de 125,8 a 136,9)

Observe que não só as medidas pontuais são distintas (pois as amostras são distintas) mas os intervalos de confiança são distintos. Ou seja, o intervalo de confiança varia com amostras de mesmo tamanho da mesma população. E ele é calculado de modo que 95% destes variados intervalos de confiança contendam o valor da população (valor real).

Como neste exemplo os dados são simulados, sabe-se o valor real da população (ou seja, o equivalente a medir a pressão de toda população).

Figura 10. Resultado de 20 estudos distintos com 100 participantes



A **figura 10** mostra o resultado de 20 amostras de 100 participantes da mesma população, cuja média foi pré-estabelecida como 130 mmHg. Nota-se que apenas um dos variados intervalos de confiança estudos não contempla o valor real da população.

De fato, 95% dos infinitos diferentes intervalos de confiança de infinitas amostras conterão o valor real da população.

A **figura 11** mostra o resultado de uma simulação com 1.000 amostras de 100 participantes (infinitas amostras seriam exatamente 5%).

Figura 11. Resultado de 1.000 estudos distintos com 100 participantes (média população = 130)

```
. display "{red:The true value is not contained in `r(min)'% of 95% Confidence Intervals}"  
The true value is not contained in 5.300000190734863% of 95% Confidence Intervals
```

A teoria é bela, porém como saber se o intervalo de confiança do estudo que acabou de ser realizado contém o valor real da população?

De fato, usualmente apenas a **figura 8** está disponível. Fora de um ambiente simulado não se sabe o valor real da população. As **figuras 10 e 11** só são possíveis pois numa simulação fixou-se o valor real como 130 mmHg.

O que fazer quando se possui apenas um estudo? Ou apenas 3 estudos (**figura 9**)?

Pode-se apenas confiar que o valor real está contido naquele intervalo de confiança – porém sem afastar a possibilidade de erro, que é controlado pelo nível de confiança 95%.

Não há como obter certeza com estatística. É preciso aceitar uma margem de erro. Usualmente, por convenção, a medicina entende como aceitável que, em média, 1 a cada 20 estudos não captem o valor real em seus intervalos de confiança.

Nenhum estudo é definitivo – em mãos apenas da **figura 8** a única coisa que pode ser feita é agir assumindo que aquele intervalo contém o valor real, porém sem nunca se livrar da angústia de que se pode estar errando.

Intervalos de confiança de medidas de tamanho de efeito

Os intervalos de confiança podem ser calculados para qualquer medida de tamanho de efeito. A **figura 12** mostra a simulação de um ensaio clínico randomizado ($n = 1000$) cujo risco de morte do grupo placebo (risco no baseline foi de 30%) e o risco relativo foi de 0,8.

A **figura 13** mostra o resultado de 4,000 ensaios clínicos ($n = 1000$) replicados dessa mesma população.

Figura 12. Resultado de 20 estudos distintos com 1.000 participantes

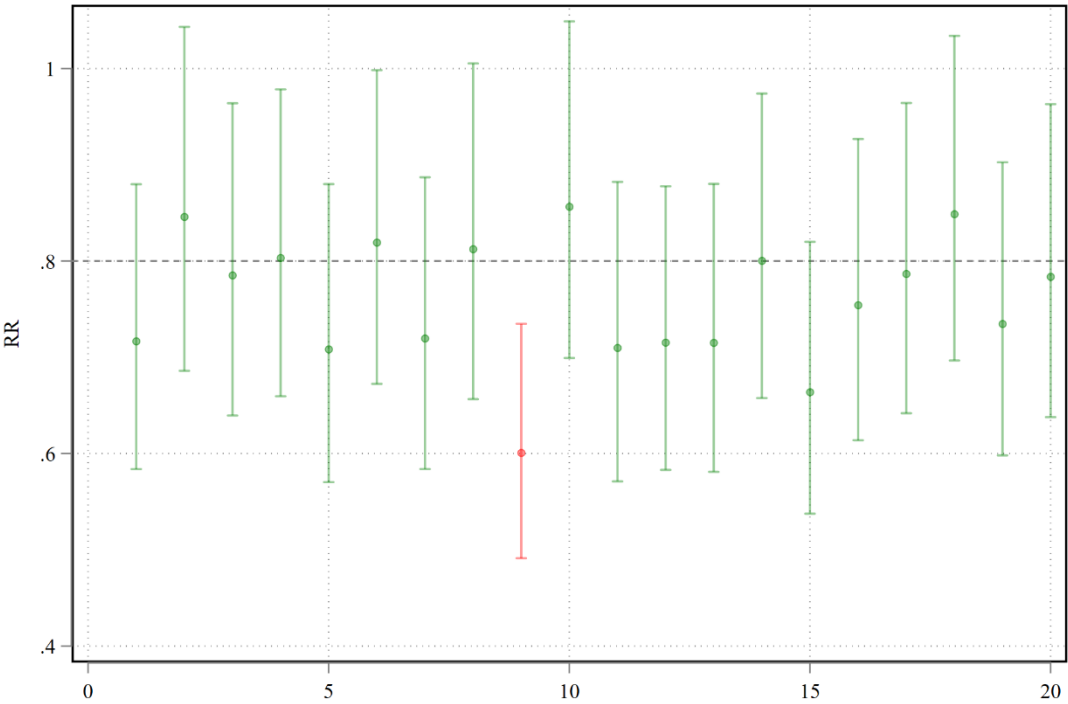


Figura 13. Resultado de 4,000 estudos distintos com 1.000 participantes (risco relativo população = 0,8)

```
. display "{red:The true value is not contained in `r(min)'% of 95% Confidence Intervals}"
The true value is not contained in 5.074999809265137% of 95% Confidence Intervals
```

Exercícios

Exercício 9

Considere um ensaio clínico randomizado hipotético com o seguinte PICO:

- P) Pacientes hospitalizados com neutropenia febril.
- I) Fluconazol + antibióticos usuais.
- C) Placebo + antibióticos usuais.
- O) Mortalidade em 28 dias.

O resultado relatado pelos investigadores foi:

Diferença de risco = - 22% (IC95% = -44% a 0%; $p = 0.06$).

Na conclusão do estudo, os autores apresentam a seguinte informação sobre este resultado numérico.

A redução do risco de morrer em 28 dias de 22% com a utilização do fluconazol (Intervalo de confiança de 95% = redução do risco de morrer em 44% a um aumento do risco de morrer em 0% com a utilização do fluconazol). O valor de p do teste de hipótese realizado foi de 0.06 (sem significância estatística).

Identifique a alternativa **incorreta**:

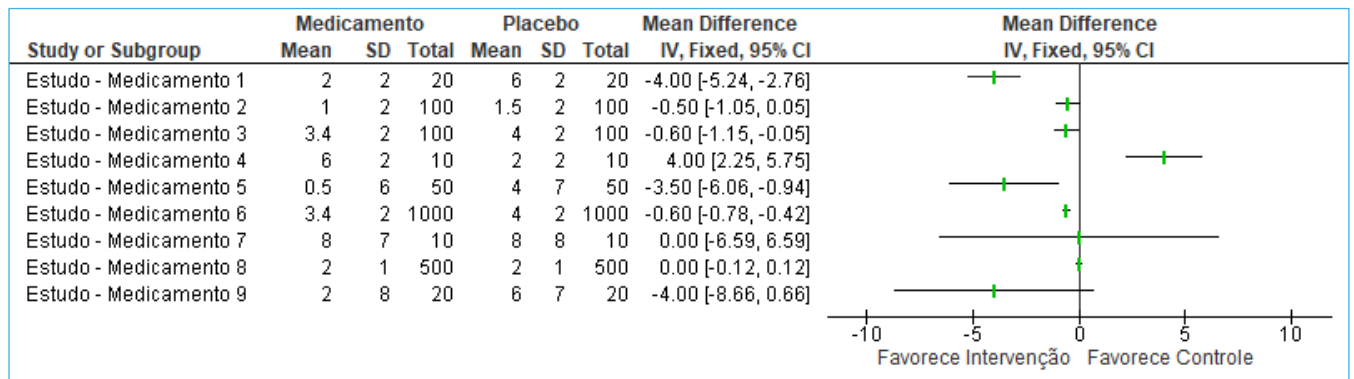
- A) Um valor de p de 0.06 significa que a hipótese nula não pode ser rejeitada em um nível de significância pré-estabelecido de 0.05.
- B) O intervalo de confiança cruza a linha de nulidade, o que significa que o estudo mostrou que a intervenção não tem efeito em pacientes com neutropenia febril.
- C) Apesar do resultado não ser estatisticamente significativo, a interpretação dos limites do intervalo de confiança pode ser a de que existe relevância clínica que suporte a utilização da intervenção.
- D) A qualidade do estudo, o contexto da doença, o custo da intervenção e a disponibilidade de outras alternativas são fundamentais para uma eventual decisão de recomendar o fluconazol.

Exercício 10

O gráfico abaixo apresenta o resultado de nove estudos que avaliaram os medicamentos 1 a 9 para tratamento de pacientes com dor oncológica crônica. A avaliação foi feita utilizando uma escala de 0 a 10, sendo que 0 indica “sem dor” e 10 indica “dor máxima”.

Sabendo que a diferença mínima clinicamente relevante para este score de dor é de **dois** pontos, responda:

- 1) Quais dos nove medicamentos poderiam ser indicados para pacientes com dor oncológica crônica?
- 2) Identifique os valores de p de cada uma das estimativas de 1 a 9. (Possíveis respostas $p > 0,05$ ou $p < 0,05$).



Exercício 11

Qual dos estudos abaixo possui o menor valor de p?

