**Coletando e analisando dados de experimentos[[1]](#footnote-1)**

1. **Os objetivos da atividade**

* Coletar dados de um experimento e inseri-lo no Excel
* Usar medidas estatísticas, por exemplo, média e desvio-padrão, e gráficos de linhas e de colunas para descrever e comparar dados
* Definir e explicar o que significa significância estatística (e o que isso não significa)
* Calcular e interpretar o valor p
* Avaliar a utilidade dos experimentos para determinar a causalidade e as limitações desses experimentos

1. **Contextualização**

Assim como os cientistas usam experimentos para investigar como os processos físicos funcionam sob certas condições, os cientistas sociais usam experimentos para investigar como as pessoas podem se comportar em determinadas situações. Embora não possamos prever perfeitamente como as pessoas realmente se comportarão, o ambiente controlado de experimentos nos permite isolar os efeitos de uma determinada mudança e identificar razões específicas para o comportamento observado. Se mantivermos todas as condições iguais e só mudarmos uma coisa, então podemos estar mais certos de que quaisquer diferenças que observamos são devidas a essa única mudança.

Os economistas usam experimentos para analisar as interações sociais em que a decisão de uma pessoa afeta o resultado desse indivíduo e os resultados para os outros. Alguns bens e serviços são chamados de bens públicos, porque quando uma pessoa arcar com o custo de prover o bem, todos podem desfrutar dele. Projetos de irrigação e produção de novos conhecimentos são exemplos de bens públicos. O problema com o fornecimento de bens públicos é que pessoas completamente interessadas em si mesmas preferem se beneficiar do bem sem pagar nada por isso - isso é conhecido como "*free-riding*" ou “pegar carona”.

No entanto, existem exemplos reais da provisão bem-sucedida de bens públicos, como projetos comuns de irrigação na Índia e no Nepal. O que poderia explicar tais contribuições sustentadas a um bem público?

Uma explicação é que as pessoas contribuem porque se preocupam com o bem-estar dos outros, ou porque respeitam as normas de que "pegar carona é ruim" (*free-riding*). Ou as pessoas podem contribuir para a vergonha que sentiriam (ou pior) se fossem publicamente punidos. Se os outros membros da sua comunidade sabem que você não contribuiu e que poderiam puni-lo (por exemplo, fofocando sobre você, retendo ajuda no futuro, ou o excluindo), então você pode contribuir tanto por interesse próprio quanto porque você quer ser capaz de pensar em si mesmo como uma boa pessoa. Para ver se a punição poderia resultar em contribuições sustentadas para um bem público, os pesquisadores [Herrmann, Thöni e Gächter (2008)](http://tinyco.re/3969457) fizeram um estudo em que diferentes grupos de pessoas, em vários países, participaram de dois experimentos de bens públicos.

O primeiro experimento teve dez rodadas. Em cada rodada:

* Cada pessoa do experimento (denominados sujeitos) recebe US$ 20.
* Os sujeitos são aleatoriamente classificados em pequenos grupos, tipicamente de quatro pessoas que não se conhecem.
* Eles são convidados a decidir sobre uma contribuição de seus US$ 20 para um conjunto comum de dinheiro.
* O dinheiro neste experimento é um bem público. Para cada dólar contribuído, cada pessoa do grupo recebe US $ 0,40, incluindo o contribuinte.
* Após cada rodada, os participantes são informados sobre o quanto os outros membros do grupo contribuíram.

O segundo experimento foi o mesmo do primeiro, exceto com uma opção de punição adicional. Depois de observar as contribuições de seu grupo, os jogadores individuais poderiam pagar para punir outros jogadores fazendo com que pagassem uma multa de US$ 3. O punidor permaneceu anônimo, mas teve que pagar US$ 1 por jogador punido.

Nesta atividade, primeiro aprenderemos mais sobre como os dados experimentais podem ser coletados com base em um experimento com bens públicos, coletando nossos próprios dados. Em seguida, procuraremos maneiras de descrever e analisar os dados experimentais dos dois experimentos descritos acima, a fim de responder às seguintes questões de pesquisa:

* Houve diferenças de comportamento (contribuições médias) entre os experimentos?
* Podemos atribuir as diferenças observadas no comportamento à mudança nas condições, e não ao acaso ou coincidência?

1. **Coletando dados – um experimento com bens públicos**

Antes de dar uma olhada mais de perto nos dados experimentais, você realizará um experimento com bens públicos, para saber como os dados experimentais podem ser coletados. Siga as instruções abaixo para formatar o experimento.

* 1. **Instruções para formatar o experimento**

Primeiro, forme um grupo de pelo menos quatro pessoas. Escolha uma pessoa para ser o **administrador** do experimento. O administrador irá monitorar o experimento, enquanto as outras pessoas participam do experimento.

**Funções do Administrador**

* **Preparando o experimento:** acesse o website [“Economics Game”](http://tinyco.re/2938847) (Economia dos Jogos), na parte inferior da página clique em “*Create a Multiplayer Game and Get Logins*” (Criar um jogo multijogador e obter logins). Em seguida, clique em “*Externalities and public goods*” (Externalidades e bens públicos). Sob o título "*Voluntary contribution to a public good*” (Contribuição voluntária para um bem público), clique em "*Choose this Game*” (Escolher este jogo). Digite o número de pessoas que participam do experimento e selecione "1" para o número de universos. Em seguida, clique em "*Get logins*” (Obter logins). Um pop-up aparecerá, mostrando os IDs de login e senhas para cada participante do experimento e para o administrador.
* **Comece o experimento:** dê a cada participante um ID de login diferente. O experimento deve ser realizado anonimamente, por isso certifique-se de que os participantes não conhecem os IDs de login dos outros participantes. Agora você está pronto para começar a primeira rodada do experimento. Existem dez rodadas no total.
* **Confirme se todas as rodadas estão concluídas:** no canto superior direito da página, clique em "Login", insira seu ID de login e senha e, em seguida, clique no botão verde "Login". Você será levado à página de administração do experimento, que mostrará a contribuição média em cada rodada e os resultados da rodada que acabou de ser realizada. Espere até que todos os participantes tenham terminado as dez rodadas antes de atualizar esta página.
* **Colete os resultados do experimento:** quando os participantes terminarem as dez rodadas, atualize a página. A tabela no topo da página mostrará a contribuição média (em Euros) para cada uma das dez rodadas. Selecione a tabela inteira, copie e cole em uma nova planilha no Excel.

**Funções dos participantes**

* **Login:** assim que o administrador tiver criado o experimento, acesse o *website* ["Economics Games"](http://tinyco.re/2938847). No canto superior direito, clique em "Login", insira o ID de login e a senha que seu administrador indicou e, em seguida, clique no botão verde "Login". Você será levado ao experimento de bens públicos que seu administrador configurou.
* **Execute a primeira rodada do experimento:** leia atentamente as instruções no topo da página antes de começar o experimento. Em cada rodada, você deve decidir quanto contribuir para o bem público. Digite sua escolha para cada universo (grupo de participantes) do qual você faz parte (se os mesmos participantes estiverem em dois universos, depois faça a mesma contribuição em ambos), depois clique em "*Validate*” (Validar).
* **Veja os resultados da primeira rodada:** você verá os resultados da primeira rodada, incluindo quanto cada participante (incluindo você) contribuiu, os pagamentos e os lucros. Clique em "Next” (Próximo) para começar a próxima rodada.
* **Complete todas as rodadas do experimento:** repita os dois passos anteriores até que você tenha executado dez rodadas no total, e então colete os resultados do experimento com seu administrador.

Os resultados do seu experimento serão parecidos com a Figura 1.

**Figura 1 – Tabela formatada com cada rodada "Round" e contribuição média "Average contribution" como as variáveis de linha**

| **Rodada (Round)** | **10** | **9** | **8** | **7** | **6** | **5** | **4** | **3** | **2** | **1** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Contribuição média (Average contribution) |  |  |  |  |  |  |  |  |  |  |

Nas perguntas abaixo, você irá comparar seus resultados com os da [Figura 3 de Herrmann et al. (2008)](http://tinyco.re/3969457), mas primeiro você precisa reformatar sua tabela para se parecer com a Figura 2.

**Figura 2 – Tabela formatada com cada rodada "Round" e contribuição média "Average contribution" como as variáveis nas colunas**

|  |  |
| --- | --- |
| **Rodada (Round)** | **Contribuição média (Average contribution)** |
| 1 |  |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |
| 6 |  |
| 7 |  |
| 8 |  |
| 9 |  |
| 10 |  |

* 1. **Reformatando a tabela**

Use os resultados do experimento realizado para responder às seguintes perguntas.

1. Prepare um gráfico de linhas com a contribuição média no eixo vertical e o período (de 1 a 10) no eixo horizontal. Descreva como as contribuições médias mudaram ao longo do experimento.
   1. **Preparando um gráfico de linhas com múltiplas variáveis**
2. Compare seu gráfico de linhas com a Figura 3 de Herrmann et al. (2008). Comente quaisquer semelhanças ou diferenças entre os resultados (por exemplo, o montante contribuído no início e no final, ou a mudança na média das contribuições ao longo do experimento)[[2]](#footnote-2).
3. Você pode pensar em alguma razão pela qual seus resultados são semelhantes (ou diferentes) aos da Figura 3? Talvez você ache útil ler a seção "*Experiments*" de [Herrmann et al. (2008)](http://tinyco.re/3969457) para uma descrição mais detalhada de como os experimentos foram conduzidos.
4. **Descrevendo os dados (análise descritiva dos dados)**

Vamos agora utilizar os dados das Figuras 2A e 3 de [Herrmann et al. (2008)](http://tinyco.re/3969457), e avaliar o efeito da opção de punição nas contribuições médias. Em vez de comparar dois gráficos mostrando todos os dados de cada experimento, como fizeram os autores do estudo, usaremos medidas estatísticas para comparar os dados e mostrar os dados de ambos os experimentos (com e sem punição) no mesmo gráfico.

Primeiro, baixe e salve os [dados](https://tinyco.re/6011816). A planilha contém duas tabelas:

* A primeira tabela mostra as contribuições médias em um experimento de bens públicos sem punição (Figura 3).
* A segunda mostra contribuições médias em um experimento de bens públicos com punição (Figura 2A).

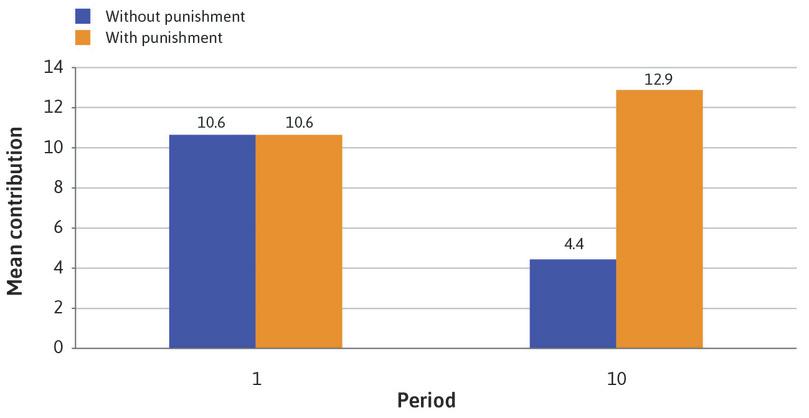
Você pode ver que em cada período (linha), a contribuição média varia entre as cidades, ou seja, há uma distribuição de contribuições médias em cada período.

A média e a variância[[3]](#footnote-3) são duas maneiras de analisar as distribuições. Agora usaremos essas medidas, juntamente com outras medidas (intervalo e desvio-padrão) para analisar a distribuição de contribuições em ambos os experimentos.

1. Utilizando os dados para as Figuras 2A e 3 de Herrmann et al. (2008):
   1. Calcule a contribuição média em cada período (linha) de maneira separada para ambos os experimentos, usando a função AVERAGE do Excel.
   2. Prepare um gráfico de linhas com a contribuição média no eixo vertical e período (de 1 a 10) no eixo horizontal (com uma linha separada para cada experimento). Certifique-se de que a legenda das linhas esteja claramente rotulada de acordo com o experimento (com punição ou sem punição).
   3. Descreva as diferenças e semelhanças que você observa na contribuição média ao longo do tempo em ambos os experimentos.

Em vez de olhar para todos os períodos, podemos nos concentrar nas contribuições no primeiro e no último período. Prepare um gráfico de colunas mostrando a contribuição média no primeiro e no último período para ambos os experimentos. Seu gráfico deve se parecer com a Figura 3 abaixo.

Figura 3 – Contribuições médias nos Períodos 1 e 10, com e sem punição



* 1. **Preparando um gráfico de barras para comparar dois grupos**

A média é uma medida útil de uma distribuição, mas não é uma descrição completa de como se comportam os dados. Precisamos também conhecer como estão "dispersos" os dados para obter uma imagem mais clara e para fazer comparações entre as distribuições. A variância é uma maneira de medir a dispersão - quanto maior a variância, mais dispersos são os dados.

Uma medida semelhante à variância é o desvio padrão[[4]](#footnote-4), o qual é a raiz quadrada da variância. O desvio padrão é comumente utilizado porque fornece uma prática para analisar grandes conjuntos de dados, por exemplo, a maioria dos dados (95%, se houver muitas observações) está dentro de dois desvios-padrão da média.

1. Utilizando os dados para as Figuras 2A e 3 de Herrmann et al. (2008):
   1. Calcule o desvio padrão para os Períodos 1 e 10, de maneira separada, para os dois experimentos. A regra geral se aplica? Em outras palavras, a maioria dos valores está dentro de dois desvios padrão da média?
   2. Como mostrado na Figura 3, a contribuição média para ambos os experimentos foi de 10,6 no Período 1. Com referência aos seus cálculos de desvio padrão, explique se isso significa que os dois conjuntos de dados são os mesmos.
   3. **Calculando o desvio-padrão**

Outra medida de dispersão é a amplitude[[5]](#footnote-5), o intervalo formado pelos valores menor (mínimo) e maior (máximo) de uma variável específica. Por exemplo, podemos dizer que o número de períodos no experimento de bens públicos varia de 1 a 10. Quando conhecemos os valores mais extremos em nosso conjunto de dados, temos uma imagem melhor da aparência de nossos dados.

1. Calcule o valor máximo e mínimo para os períodos 1 e 10 separadamente, para os dois experimentos.
   1. **Encontrar o mínimo, o máximo e o intervalo de uma variável**
2. Uma maneira concisa de descrever os dados é quadro com a estatística descritiva. Com apenas quatro medidas (média, desvio padrão, valor mínimo e valor máximo), podemos ter uma ideia geral da aparência dos dados.
   1. No Excel, prepare um quadro com a estatística descritiva conforme mostrado na figura abaixo. Prepare mais três quadros com a estatística descritiva, para o Período 10 (sem punição), Período 1 (com punição) e Período 10 (com punição). Use suas respostas para as perguntas II a IV para preencher os quadros com as estatísticas descritivas.

Figura – Quadro com estatísticas descritivas

| **Variável** | **Média** | **Desvio-padrão** | **Mínimo** | **Máximo** |
| --- | --- | --- | --- | --- |
| Contribuição (Período 1, sem punição) |  |  |  |  |

* 1. Comente quaisquer semelhanças e diferenças nas distribuições, ao longo do tempo e entre experimentos.

1. **Mudar as regras do experimento teve um efeito significativo no comportamento?**

A opção de punição foi introduzida no experimento de bens públicos para ver se poderia ajudar a sustentar as contribuições, em comparação com o experimento sem uma opção de punição. Agora, usaremos um método chamado teste de hipótese para comparar os resultados de ambos os experimentos mais formalmente.

Comparando os resultados no Período 10 de ambos os experimentos, podemos ver que a contribuição média no experimento com punição é 8,5 unidades maior do que no experimento sem punição (veja a Figura 3). É mais provável que esse comportamento seja devido ao acaso ou é mais provável que seja devido à diferença nas condições experimentais?

1. Você pode realizar outro experimento para entender por que podemos ver diferenças no comportamento devido ao acaso.
   1. Primeiro, jogue uma moeda seis vezes, usando apenas uma mão e registre o número de vezes que você obtém “caras”. Em seguida, usando a mesma mão, jogue uma moeda seis vezes e registre o número de vezes que obtém “caras”.
   2. Compare os resultados da Questão I (a). Você obteve o mesmo número de “caras” nos dois casos? Mesmo se você soubesse, a sequência dos resultados (por exemplo: cara, coroa, coroa ...) é a mesma nos dois casos?

O ponto importante a ser observado é que, mesmo quando realizamos experimentos sob as mesmas condições controladas, devido a um elemento aleatório, podemos não observar exatamente o mesmo comportamento cada vez que realizamos o experimento.

Se é provável que as diferenças observadas não sejam devidas ao acaso, dizemos que as diferenças são estatisticamente significativas[[6]](#footnote-6). Para determinar se as diferenças de médias são estatisticamente significativas ou não, precisamos considerar o tamanho da diferença em relação ao desvio padrão de ambas as distribuições (como os dados estão espalhados).

O fato de que a significância estatística depende de uma comparação relativa é muito importante. O tamanho da diferença por si só não pode nos dizer se algo é estatisticamente significativo ou não. De fato, mesmo que as diferenças observadas sejam grandes, não é uma garantia de que as diferenças sejam estatisticamente significativas. As Figuras 5 e 6 mostram a pontuação média no exame de dois grupos (representada pela altura das colunas e relatada nas caixas acima das colunas), com os pontos representando os dados subjacentes.

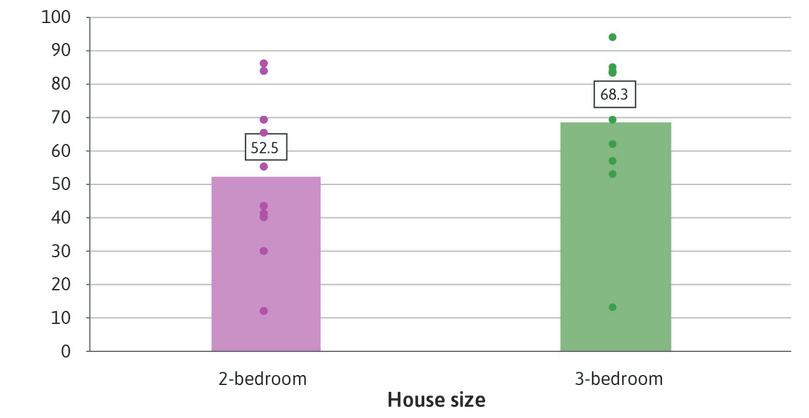
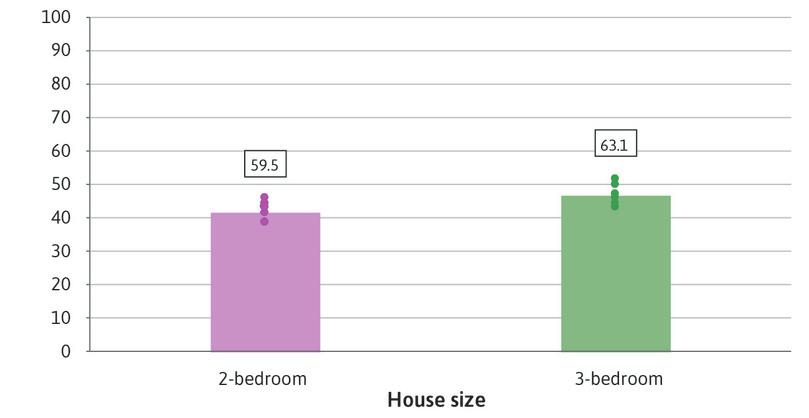
Figura – Exemplo de uma grande diferença de meios que não é estatisticamente significativo 

Figura – Exemplo de uma pequena diferença de meios que é estatisticamente significativa 

A Figura 5 mostra que uma diferença relativamente grande nas médias não é estatisticamente significativa, porque os dados são amplamente distribuídos (o desvio padrão é grande), enquanto a Figura 6 mostra que uma diferença relativamente pequena é estatisticamente significativa, porque os dados estão fortemente agrupados (o desvio padrão é muito pequeno). Na Figura 5, a diferença das médias provavelmente se deve ao acaso, mas na Figura 6, a diferença das médias provavelmente não se deve ao acaso.

Para determinar a significância estatística, realizamos um teste de hipótese[[7]](#footnote-7), que usa o tamanho da diferença e o desvio padrão para calcular a probabilidade (denominada valor p[[8]](#footnote-8)) de ver os dados que observamos, assumindo que as médias de ambas as distribuições são as mesmas. Como o valor-p é uma probabilidade, varia de 0 a 1 (inclusive). Quanto menor a probabilidade (menor o valor-p), menor a probabilidade de observarmos os dados fornecidos; portanto, maior a probabilidade de nossa suposição ser falsa (em outras palavras, os meios de ambas as distribuições não são os mesmo).

Nossas conclusões dependerão de nossa definição de probabilidade baixa'. Definimos "baixo" escolhendo um ponto de corte (uma porcentagem) também conhecido como nível de significância[[9]](#footnote-9). Qualquer probabilidade baixa que esse ponto de corte seria considerado "baixo". Alguns pontos de corte comumente usados são 1% (valor p de 0,01), 5% (valor p de 0,05) e 10% (valor p de 0,1).

* 1. **Teste de hipóteses**

Para calcular o valor p, usamos uma função no Excel chamada T.TEST.

1. Usando os dados para as Figuras 2A e 3 de [Herrmann et al. (2008)](http://tinyco.re/3969457):
   1. Use a função T.TEST do Excel para calcular o valor de p para a diferença de médias no Período 1 (com e sem punição). Qual é o valor p?
   2. Com um ponto de corte de 5% (valor-p de 0,05), podemos concluir que a diferença de médias é significativa? Por que ou por que não?
   3. **Usando a função T.TEST do Excel**
2. Usando os dados do Período 10:
   1. Use a função T.TEST do Excel para calcular o valor de p para a diferença de médias no período 10 (com e sem punição). Qual é o valor p?
   2. Com um ponto de corte de 5%, podemos concluir que a diferença de médias é significativa? Por que ou por que não?
   3. Com referência às Figuras 5 e 6, explique por que não podemos usar o tamanho da diferença para concluir diretamente se a diferença de médias é significativa ou não.

Um ponto importante a ser observado é que a significância estatística não pode nos dizer nada sobre causalidade. No exemplo do tamanho da casa e das pontuações dos exames mostradas na Figura 6, houve uma relação estatisticamente significante entre as duas variáveis (os estudantes que moram em uma casa de três quartos tiveram pontuações mais altas nos exames, em média, do que os estudantes que moravam em uma casa de dois quartos). No entanto, não podemos dizer que o tamanho maior da casa foi a causa de pontuações mais altas nos exames, porque é improvável que a construção de uma sala extra torne alguém mais inteligente automaticamente. A significância estatística não pode nos ajudar a detectar essas correlações espúrias[[10]](#footnote-10).

No entanto, experimentos podem nos ajudar a determinar se existe um nexo de causalidade entre duas variáveis. Se realizarmos um experimento e encontrarmos uma diferença estatisticamente significante nos resultados, podemos concluir que uma variável é a causa da outra.

1. Consulte os resultados dos experimentos de bens públicos.
   1. Quais características do cenário experimental tornam provável que a opção de punição tenha sido a causa da mudança de comportamento?
   2. Com referência à Figura 3, explique por que precisamos comparar os dois grupos no Período 1 para concluir que existe um nexo de causalidade entre a opção de punição e o comportamento no jogo.

As experiências podem ser úteis para identificar links causais. No entanto, se o comportamento das pessoas em condições experimentais fosse diferente do comportamento no mundo real, nossos resultados não seriam aplicáveis em nenhum lugar fora do experimento.

1. Discuta algumas limitações dos experimentos e sugira algumas maneiras de abordá-las (ou abordá-las parcialmente). Você pode encontrar as páginas 158–171 do artigo "[What do laboratory experiments measuring social preferences reveal about the real world?](http://tinyco.re/9601240" \t "_blank)"

1. Preparada com base em Doing Economics: empirical projects (http://www.core-econ.org/doing-economics/index.html). [↑](#footnote-ref-1)
2. Benedikt Herrmann, Christian Thöni, and Simon Gächter. 2008. Figure 3 in [‘Antisocial punishment across societies’](http://tinyco.re/3969457" \t "_blank). Science Magazine 319 (5868): p. 1365.  [↑](#footnote-ref-2)
3. A variância é uma medida de dispersão de uma distribuição de frequência, igual à média dos quadrados dos desvios da média aritmética da distribuição. A variância é usada para indicar a "dispersão" dos dados. Uma maior variância significa que os dados estão mais dispersos. Exemplo: o conjunto de números 1, 1, 1 tem variância zero (sem variação), enquanto o conjunto de números 1, 1, 999 tem uma alta variância de 2178 (grande dispersão). [↑](#footnote-ref-3)
4. É uma medida de dispersão de uma distribuição de frequência, sendo igual à raiz quadrada da variância. O desvio padrão tem uma interpretação semelhante à variância. Um desvio padrão maior significa que os dados estão mais dispersos. Exemplo: o conjunto de números 1, 1, 1 tem um desvio padrão de zero (sem variação ou dispersão), enquanto o conjunto de números 1, 1, 999 tem um desvio padrão de 46,7 (grande dispersão). [↑](#footnote-ref-4)
5. O intervalo formado pelo menor (mínimo) e pelo maior (máximo) valor de uma variável específica. O intervalo mostra os dois valores mais extremos na distribuição e pode ser usado para verificar se há algum erro externo (outlier) nos dados. Outliers são algumas observações nos dados que são muito diferentes do restante das observações. [↑](#footnote-ref-5)
6. Quando é improvável que um relacionamento entre duas ou mais variáveis seja devido ao acaso, dadas as suposições feitas sobre as variáveis (por exemplo, ter a mesma média). A significância estatística não nos diz se existe um nexo de causalidade entre as variáveis. [↑](#footnote-ref-6)
7. Um teste no qual uma hipótese nula (padrão) e uma alternativa são apresentadas sobre algumas características da população. Os dados de amostra são então usados para testar a probabilidade desses dados de amostra serem vistos se a hipótese nula for verdadeira. [↑](#footnote-ref-7)
8. A probabilidade de observar os dados coletados, assumindo que dois grupos tenham a mesma média. O valor p varia de 0 a 1, onde valores mais baixos indicam uma probabilidade mais alta de que a suposição subjacente (mesma média) seja falsa. Quanto menor a probabilidade (menor o valor-p), menor a probabilidade de observar os dados fornecidos e, portanto, maior a probabilidade de a suposição ser falsa (os meios de ambas as distribuições não são os mesmos). [↑](#footnote-ref-8)
9. Uma probabilidade de corte que determina se um valor-p é considerado estatisticamente significativo. Se um valor-p for menor que o nível de significância, considera-se improvável que as diferenças observadas sejam devidas ao acaso, dadas as suposições feitas sobre as variáveis (por exemplo, ter a mesma média). Os níveis de significância comuns são 1% (valor p de 0,01), 5% (valor p de 0,05) e 10% (valor p de 0,1). [↑](#footnote-ref-9)
10. Uma forte associação linear entre duas variáveis que não resulta de qualquer relação direta, mas pode ser devido a uma coincidência ou a outro fator não observado. [↑](#footnote-ref-10)