

Estatísticas

Há mentiras, mentiras malditas e estatísticas.

Samuel Langhorne Clemens
mais conhecido como Mark Twain
Escritor

Introdução

Números são poderosos. Aparentemente livres de toda a ambigüidade e carga emocional das palavras, os números e as estatísticas são poderosas evidências que podem efetivamente fortalecer qualquer argumento. Mas as estatísticas não são uma panacéia. Aliás, tão simples e direto quanto estes números prometem ser, as estatísticas, se não forem usadas cuidadosamente, podem criar mais problemas do que soluções.

Por quê? Bem, 75% de todos os usuários de estatísticas não possuem uma compreensão adequada das estatísticas que estão usando. Surpreendentemente, este número tem aumentado na alarmante taxa de mais de 5% ao ano. Os leitores não são melhores; o leitor médio não sabe como avaliar e interpretar apropriadamente as estatísticas que lê. A razão principal para este uso inadequado das estatísticas é o desconhecimento do que a estatística pode e do que ela não pode fazer. Quatro de cinco usuários acham que as estatísticas falam por si mesmas, e deixam a interpretação inteiramente a cargo do leitor. Mas isso é pura besteira. O uso dos números pode ser tão ambíguo quanto as palavras e precisa igualmente de explicações.

Este problema é muito parecido de vários modos com as citações literais ou transcrições. Frequentemente, se espera das transcrições que façam sozinhas seu trabalho; são usadas como parte do argumento, ao invés de uma evidência que requer interpretação. Se deixarmos a interpretação a cargo do leitor, como saber que tipo de interpretação inesperada pode aparecer? O único modo de evitar este risco é fornecer a interpretação você mesmo.

Mas antes de começar a escrever estatísticas, nós temos de realmente ler algumas...

Lendo estatísticas

Como ditto anteriormente, os números podem ser poderosos. Esta é uma das razões das estatísticas serem evidências tão persuasivas. Entretanto, este mesmo poder pode também tornar os números e as estatísticas um pouco (se não muito) intimidantes. Isto é, usualmente as aceitamos como a verdade, sem nunca questionarmos a sua veracidade e adequação. Embora possa parecer um traço positivo colocar estatísticas na sua dissertação e rezar para que o leitor se submeta ao seu poder, lembre-se que, antes de escrevermos estatísticas, devemos ser leitores de estatísticas. E para ser um leitor eficaz significa perguntar as questões difíceis. Abaixo, listamos um conjunto de questões difíceis que devemos perguntar aos números que lemos.

1) A sua evidência provém de fontes confiáveis?

Isto é obviamente importante não só para as estatísticas, mas para qualquer evidência que usar em sua dissertação. Para as estatísticas, isto é particularmente importante, pois, como veremos neste capítulo, há diversos modos de manipulá-las para produzir o resultado esperado. Portanto, você tem de tirar suas estatísticas de fontes confiáveis. Isto não significa que as fontes confiáveis sejam infalíveis, mas sim que seja menos provável de que estejam utilizando práticas enganadoras. Com uma fonte confiável, você não precisará se preocupar muito com as questões a seguir. Ainda assim, lembre-se que ler estatísticas é um pouco como estar no meio de uma guerra. Não confie em ninguém, suspeite de todos.

2) Qual é a origem dos dados?

Os dados e as estatísticas não caem completas do céu. São sempre um produto da pesquisa. Portanto, para entender as estatísticas, você deve também conhecer de onde elas provêm. Por exemplo, se as estatísticas vieram de uma pesquisa de opinião, algumas das questões incluem:

- Quem estava perguntando?
- Quais eram exatamente as questões?
- Quem está interpretando os dados?
- Qual a questão que suscitou a pesquisa?
- O que (políticas ou procedimentos) depende potencialmente dos resultados da pesquisa?

- Quem se beneficia de alguma interpretação específica dos dados?

Estas questões são uma maneira de detectar vieses potenciais ou fraquezas nos dados que estiver lendo. O objetivo deste exercício não é encontrar dados “puros”, objetivos, mas tornar quaisquer vieses explícitos, para interpretar mais acuradamente a evidência.

3) Todos os dados foram relatados?

Na maior parte dos casos, a resposta a esta questão é fácil: Não. Portanto, uma melhor maneira de pensar sobre isto é se todos os dados foram apresentados em contexto. Mas isso é muito mais complicado quando você está considerando uma questão maior, que seria se o texto apresentou evidências suficientes para você formar sua própria conclusão. Se for esse o caso, isto significa que o texto não excluiu dados que contradizem ou enfraquecem as informações ou as afirmações apresentadas.

Um exemplo pode ser encontrado no noticiário da noite. Se você pensar nas tempestades repentinas, que dificultam tanto a vida no verão, você certamente lembrará dos repórteres avisando às pessoas para não sair de casa por causa do risco de inundações. Para sustentar este ponto, eles dizem que a Defesa Civil já relatou o aumento de 30% no nível de rios, represas e açudes. A intenção deles é usar este número para assustar-nos para que fiquemos em casa. Embora este número pareça alto, a notícia não considera que antes desta chuva, houve um período de estiagem de mais de três semanas sem chuva. A lição aqui é que mesmo quando a interpretação é “acurada” de modo geral, os dados podem não ser evidências reais para a interpretação oferecida. Isto significa que você não tem como verificar se é de fato correta.

Como podemos saber se todos os dados foram incluídos? A maneira mais importante para pensar sobre isso é que usualmente há uma comparação implícita no uso das estatísticas. Como podemos fazer uma comparação válida sem ter todos os fatos? Boa pergunta. Você talvez terá que procurar os outros dados em outro lugar.

4) Os dados foram interpretados corretamente?

Se o autor forneceu-lhe as suas estatísticas, é sempre uma boa idéia você mesmo interpretá-las. Isto é, embora seja útil ler e entender a interpretação do autor, ela ainda é meramente *uma interpretação*. Não é a palavra final sobre o tema. Além do mais, algumas vezes os autores (inclusive você, portanto tome cuidado) pode perfeitamente usar boas estatísticas e terminar com interpretações perfeitamente incorretas. A seguir, dois erros comuns a evitar:

- *Confundir correlação com causalidade*: Só porque duas coisas variam juntas não significa que você possa pressupor causalidade. Pode ser nada mais do que coincidência, ou ambas são causadas por um terceiro fator comum. Tais relações são chamadas de espúrias.

Um exemplo pode ser um estudo que “descobriu” que quanto mais bombeiros são enviados ao combate às chamas de qualquer incêndio, mais danos o fogo causará. Ei! Eu pensei que os bombeiros deveriam melhorar as coisas e não piorar. Mas antes de começarmos a fechar os corpos de bombeiros, devemos procurar explicações alternativas. Esta “descoberta” contraditória pode ser facilmente explicada apontando um terceiro fator que causa ambas: o tamanho do incêndio. Qual é a lição aqui? A correlação não é igual à causação. Então, é importante não somente mostrar que duas variáveis covariam, mas também qual é o mecanismo causal.

- *Ignorar a margem de erro*: Quando os resultados de uma pesquisa de opinião são relatados, freqüentemente incluem-se margens de erro. Você já deve ter visto algo como “uma margem de erro de mais ou menos 5 pontos percentuais”. O que isso significa? A estória mais simples é que as pesquisas de opinião são normalmente realizadas com amostras de uma população maior e portanto, nunca são perfeitamente exatas. Há sempre um intervalo de confiança dentro do qual espera-se que a população da qual a amostra foi retirada esteja. Então, se eu disser que o número de alunos da UFF que encontram dificuldades em usar estatísticas nos seus escritos é de 60% mais ou menos 4%, isto significa que, presumindo um intervalo de confiança de 95%, que tenho 95% de certeza de que o número de alunos está entre 56% e 64%.

Mas por que isso é importante? Porque se, depois de apresentar este capítulo para os estudantes da UFF, um novo levantamento de opinião for feito e encontrar que somente 56% dos alunos, mais ou menos 3%, estão tendo dificuldades com estatísticas, eu poderia ir falar com o reitor e pedir uma gratificação ou um prêmio, já que eu produzi uma contribuição significativa para as habilidades de escrita dos estudantes da UFF. Entretanto, ele poderia sem dúvida apontar que a) isto pode ser uma relação espúria e b) a mudança real não é

significativa, já que cai dentro da margem de erro dos resultados originais. Qual é a lição aqui? As margens de erro importam, portanto você não pode só comparar simplesmente as porcentagens.

Finalmente, você deve manter em mente que a fonte original de seus dados não será sempre a fonte que estará usando. Isto é, se você encontrar um ensaio que cita um número de estatísticas para sustentar seu argumento, algumas vezes o autor está usando os dados de outra pessoa. Então, você precisa considerar não somente sua fonte, mas também as fontes do autor.

Escrevendo estatísticas

Enquanto estiver escrevendo com estatísticas, lembre-se de sua própria experiência como um leitor de estatísticas. Não se esqueça de quão frustrado ficou quando se deparou com estatísticas obscuras, e quão agradecido ficou por ler estatísticas bem apresentadas. É um sinal de respeito a seu leitor ser claro e direto quanto possível com seus números. Ninguém gosta de ser tratado como idiota. Portanto, mesmo se você achar que obscurecendo seus números ajudará seu argumento, não ceda a esta tentação. Assim que começar a escrever, mantenha em mente o seguinte: primeiro, seu leitor desejará saber as respostas às mesmas questões discutidas acima; segundo, você desejará apresentar suas estatísticas de modo claro e evidente. Abaixo, listamos algumas das armadilhas do mundo da estatística, seguidas de sugestões para evitá-las.

1) *O erro do escritor “mediano”*

Ninguém quer ser mediano. Além disso, ninguém quer somente ver a palavra “média” em um artigo. Por quê? Porque ninguém pode saber exatamente o que isso pode dizer. Há três medidas de tendência central, ou parâmetros médios, portanto, quando você as usa de modo indiscriminado, seu leitor tem apenas 33,3% de chance de compreender corretamente o que você quer dizer.

As seguintes definições vão fazer referência ao seguinte conjunto de números:

5, 5, 5, 8, 12, 14, 21, 33, 38

Média (média aritmética): Esta deve ser a definição mais comum de parâmetro médio. Esta é uma média ponderada – o total de todos os números dividido pela quantidade de números apresentados. Isto significa que a média do conjunto acima é igual a $5+5+5+8+12+14+21+33+38$, tudo isso dividido por 9, que por sua vez é igual a 15,64444444... (Ei! São muitos números após a vírgula - O que devemos fazer sobre isso? A precisão é algo desejável, mas precisão demais é inútil; não torna seu argumento mais forte. Considere o nível adequado de precisão baseado no resultado e arredonde de acordo. Neste caso, 15,6 deve resolver a questão).

Mediana: Dependendo de se você tiver um conjunto par ou ímpar de números, a mediana será ou o número exatamente no meio de um conjunto ímpar de números, ou o valor médio dos dois números no meio do conjunto par de números. Para o conjunto acima (um conjunto ímpar de 9 números), a mediana é 12 (5, 5, 5, 8 < 12 < 14, 21, 33, 38)

Moda: É o número ou valor mais freqüente no conjunto de números. Se, por uma obra do acaso, dois ou mais valores ocorrerem com a mesma freqüência, então você terá de tirar a média aritmética destes valores. Para o nosso conjunto, a moda pode ser 5, já que ocorre três vezes, enquanto que todos os números ocorrem apenas uma vez cada.

Como você pode ver, os números podem variar consideravelmente, assim como sua significância. Portanto, o escritor deve sempre informar qual medida de tendência central está usando. Caso contrário, confusões serão inevitáveis.

2) *Equipare seus fatos com suas questões*

Certifique-se de que suas estatísticas realmente se aplicam ao seu argumento. Se retornarmos à nossa discussão de médias, dependendo da questão em que você estiver interessado, você deve utilizar a estatística apropriada.

Talvez um exemplo ajude a ilustrar esta posição. Seu professor devolveu as notas da primeira prova do semestre. As notas se distribuíram como representado na tabela abaixo.

Nota	Número de alunos
100	4
98	5
95	2
63	4
58	6

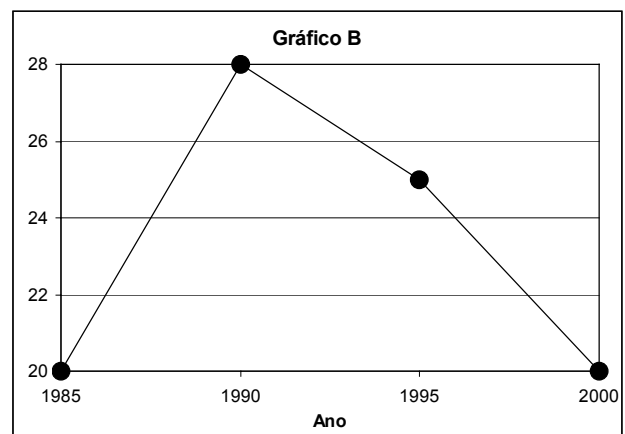
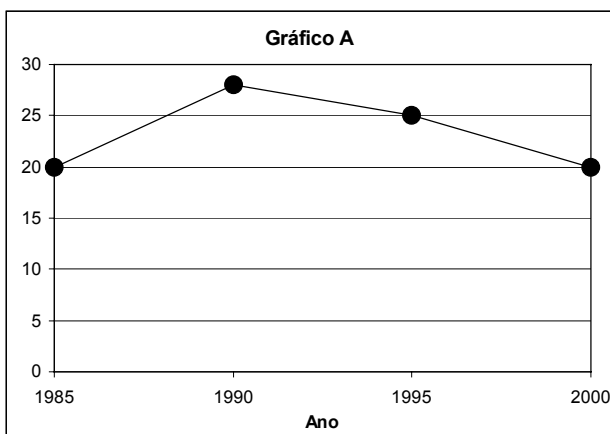
O professor sentiu que a prova deve ter sido muito fácil, porque a nota média (considerando a mediana) foi 95. Quando um outro professor lhe perguntou sobre como foram as notas da prova, ele respondeu, sabendo que suas aulas estavam ficando conhecidas por serem “muito fáceis”, que a nota média (agora considerando a média aritmética) foi 80.

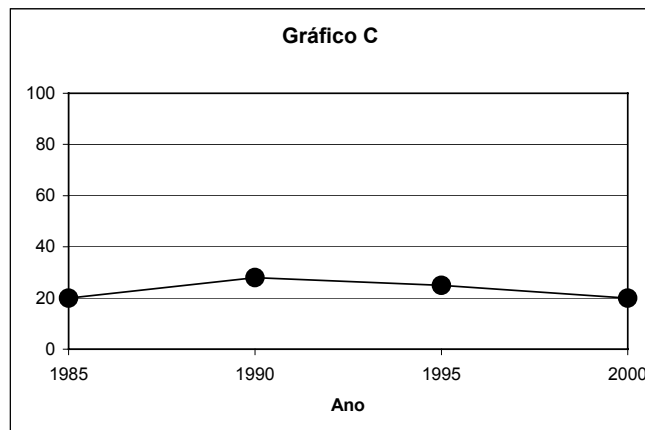
Quando seus pais lhe perguntam como você se justifica indo tão mal nesta disciplina, você responde, “Não se preocupem com o meu 63. Não foi tão ruim quanto parece. A nota média (agora considerando a moda) foi 58”.

Deixarei você decidir qual destas opções é a apropriada. Selecionar o fato ou estatística apropriada ajudará imensamente seu argumento. Não só porque realmente sustentará sua posição, mas não diminuirá a legitimidade dela. Pense em como seus pais reagiriam quando souberem pelo seu professor que a média (mediana) foi 95. O melhor modo de manter a precisão é especificar qual das três formas de “média” usar. Se você mantém isto em mente, você será pelo menos 23% mais persuasivo!

3) *Mostre o quadro inteiro*

Algumas vezes, você pode representar mal suas evidências por acidente ou desconhecimento. Em outras vezes, todavia, estes erros podem ser um pouquinho menos inocentes. Isto pode ser visto facilmente nos maquiavélicos auxílios visuais. Não formate e aplaine a representação para que “sustente melhor” seu argumento. Isto pode ser adquirido através da manipulação de gráficos em inúmeros modos. Tanto a amplitude pode ser diminuída (para retirar os pontos de dados que não se encaixam, p.ex., começando uma série temporal mais tarde ou terminando-a mais cedo), ou a escala pode ser alterada para que pequenas variações pareçam maiores (ou vice-versa). Além disso, não trapaceie nas proporções, sejam horizontais ou verticais. Só porque a *Veja*, a *Istoé*, a *Época*, e os diversos jornais recorrem a esta técnica para tentar eleger seus candidatos a cargos políticos, não quer dizer que este é um recurso lícito para um argumento acadêmico.





Os Gráficos A, B e C usam todos os mesmos dados, mas as histórias que parecem nos contar são muito diferentes. O Gráfico A apresenta um aumento mediano, seguido por uma diminuição lenta. Por outro lado, o Gráfico B mostra um salto expressivo, com uma ampla queda em seguida. De modo diverso, o Gráfico C parece indicar que praticamente não houve mudança ao longo do tempo. Estas variações são um produto das mudanças nas escalas dos gráficos. Um modo de aliviar este problema é suplementar o gráfico com os números reais no texto, só pela transparência.

Uma das dificuldades dos auxílios visuais é que não há regras fáceis e estritas sobre o que incluir ou excluir. Algum tipo de julgamento estará sempre envolvido. A melhor regra é ter certeza de que apresentou seus gráficos de forma a que seu leitor possa tirar sua própria conclusão a partir dos fatos e verificar suas afirmações. Se o que você for modificar puder afetar a interpretação dos dados pelo leitor, então você deve considerar seriamente manter os gráficos.

4) *Forneça as bases de todas as porcentagens*

Devido ao fato de que as porcentagens são sempre derivadas de uma base específica, elas não significam nada se não estiverem associadas às suas respectivas bases. Portanto, mesmo se eu lhe disser que depois de ler este capítulo você ficaria 23% mais persuasivo (o que eu realmente disse anteriormente), isto não teria significado algum, porque você não tem idéia no quê isto está baseado – 23% mais persuasivo do quê?

Vamos ver a incidência de criminalidade para verificar como isso funciona. Suponha que estejamos falando de duas cidades, Itacatuba-Mirim do Sul e Itacatuba-Mirim do Norte. Em Itacatuba-Mirim do Sul, a taxa de homicídios aumentou para 75%, enquanto que em Itacatuba-Mirim do Norte, a taxa só aumentou 10%. Qual cidade tem maiores problemas de criminalidade? Parece óbvio não é? Tem de ser Itacatuba-Mirim do Sul! 75% é maior que 10%!

Espera um minuto, porque isto é menos óbvio do que parece. Para realmente saber qual destas cidades tem piores problemas, temos de olhar para os números reais. Se eu disser que Itacatuba-Mirim do Sul teve 4 assassinatos no ano passado e 7 neste ano, e que Itacatuba-Mirim do Norte teve 60 no ano passado e 54 este ano, o que você diria? Já que 54 é significativamente menor que 7, você se sentiria mais seguro em Itacatuba-Mirim do Sul, não é?

Não tão rápido, porque ainda faltam alguns dados. Temos de fazer a comparação entre dois padrões equivalentes. Para fazê-lo, devemos olhar para a ocorrência per capita (normalmente dado em casos por 100 mil habitantes). Se Itacatuba-Mirim do Sul tem apenas 70 moradores, enquanto Itacatuba-Mirim do Norte tem 6,6 milhões de habitantes, então a taxa de homicídios de Itacatuba-Mirim do Sul é de 1.000 por 100 mil habitantes e a de Itacatuba-Mirim do Norte é de 1 por 100 mil hab. Que absurdo! Os moradores de Itacatuba-Mirim do Sul estão morrendo feito moscas. Eu acho que vou ficar mesmo na calma e pacífica Itacatuba-Mirim do Norte.

As porcentagens não são realmente diferentes de nenhuma outra forma de estatística: só adquirem significado quando são contextualizadas. Conseqüentemente, as porcentagens devem ser apresentadas em contexto, para que o leitor possa tirar suas próprias conclusões, enquanto você enfatiza os fatos relevantes para seu argumento. Lembre-se, se suas estatísticas realmente sustentam sua posição, então não tenha medo de revelar o contexto que engloba-as

Questões importantes para perguntar (e responder) sobre as estatísticas:

- 1) A questão sendo perguntada é relevante?
- 2) Os dados provêm de fontes confiáveis?
- 3) Margens de erro/Intervalos de confiança – quando uma mudança é realmente uma mudança?
- 4) Todos os dados foram relatados ou só os melhores/piores?
- 5) Os dados foram apresentados em contexto?
- 6) Os dados foram interpretados corretamente?
- 7) O autor confundiu correlação com causalidade?

Conclusão

Agora que você aprendeu umas lições de estatísticas, você tem duas opções. Utilize este conhecimento para manipular seus números para se beneficiar ou utilize este conhecimento para entender e usar melhor as estatísticas para produzir argumentos acurados e honestos. A escolha é sua. Nove de dez autores, entretanto, preferem a última, o outro se arrepende de sua decisão.