

Quantidade de tanques na Segunda Guerra Mundial

VINÍCIUS DANDARO SANTOS*

Universidade de São Paulo

vdandaro11@gmail.com

Resumo

Se você é uma pessoa que sempre achou a matemática difícil e prefere história, este artigo é perfeito para mostrar a você que matemática é atraente enquanto nós conversamos um pouco sobre história. O que faremos aqui é analisar como a matemática (ou a estatística, mais precisamente) foi usada durante a Segunda Guerra Mundial para auxiliar a vitória dos Aliados. Veremos, também, que esse problema pode ser resolvido através de duas formas: pela estatística clássica e pela estatística bayesiana.

1. VAMOS COMEÇAR PELA HISTÓRIA?

EM 1939, Adolf Hitler já havia remilitarizado a Alemanha e comandava a expansão das fronteiras alemãs pelos territórios vizinhos, violando assim, o Tratado de Versalhes, assinado em 1919. Apesar disso, o estopim da Segunda Guerra Mundial só veio com a invasão alemã da Polônia. Após essa invasão, Reino Unido e França declararam guerra à Alemanha nazista e, com a posterior adesão dos Estados Unidos, União Soviética e China, formaram o grupo dos Aliados. Opondo-se aos Aliados estava o Eixo, grupo formado por Alemanha, Itália e Japão.

Em uma guerra, a vitória depende de diversos fatores, mas é fundamental que se consiga obter informações (e essas informações devem ser bastante precisas) sobre o adversário. Você já assistiu ao filme “O Jogo da Imitação”? Ele conta a história de Alan Turing, que desenvolveu uma máquina, também durante a Segunda Guerra Mundial, capaz de decifrar a Enigma, máquina usada pelos nazistas para informar ordens e planos de guerra aos seus combatentes, com a qual a comunicação era feita por códigos. Com as informações obtidas a partir desses códigos decifrados, os Aliados estavam à frente dos alemães em algumas situações e poderiam preparar melhor o ataque ou a defesa.

Neste artigo, mostraremos um outro meio pelo qual foi obtida informação do adversário: a estatística. O problema em questão era estimar, com uma precisão considerável, a quantidade de tanques produzidos pelos alemães. Durante uma batalha, se você tiver mais equipamento que o seu adversário, a vitória é mais fácil de ser conquistada e um dos maiores equipamentos de guerra daquela época eram os tanques. Portanto, saber quantos tanques o adversário tinha era uma informação de grande valor.

Para estimar a quantidade de tanques alemães, os matemáticos tinham apenas os números de série dos tanques capturados durante os combates. Na verdade, eles tinham uma outra coisa também. Você consegue imaginar o quê? Isso mesmo, eles tinham muita sorte! Esse detalhe foi fundamental, porque os alemães organizavam os números de séries de seus produtos sempre em ordem crescente partindo do 1 e isso era fundamental para que os cálculos apresentados abaixo tivessem utilidade.

*Agradecimentos a Prof. Dra. Juliana Cobre, minha orientadora, por me acompanhar nesse estudo.

Mas então, quantos tanques você acha que os alemães estavam produzindo? Algo em torno de 100, 1000 ou 10000?

Entre 1940 e 1942, o serviço de inteligência dos Aliados (sem estatística) estimou que os alemães estavam produzindo em torno de 1400 tanques por mês. Já os estatísticos estimaram que a produção era de 246 tanques por mês. Após a guerra, os registros alemães mostraram que a produção real era de 245 tanques por mês. Isso mesmo, os estatísticos erraram por apenas um tanque! Vamos ver como eles fizeram isso?

2. AGORA É A VEZ DA ESTATÍSTICA

Não se irrite, já vamos começar os cálculos. Antes disso você precisa conhecer alguns termos e ferramentas que serão usados mais à frente para a solução do nosso problema. Nesta seção, explicaremos os seguintes termos: população, amostra, variável aleatória, distribuição uniforme, valor esperado, variância, função de probabilidade e função de verossimilhança.

Os dois primeiros termos são a base de qualquer estatística. População é todo um universo de eventos que será analisado, enquanto amostra é uma parte desse universo. Ficou um pouco difícil de entender? Vamos ver um exemplo. Imagine que a diretora de uma escola com 3000 alunos deseja saber quantos desses alunos têm mais de 18 anos. Levaria muito tempo para perguntar a idade de cada aluno, então ela resolve pedir ajuda a um estatístico. O estatístico, então, seleciona um grupo com 100 alunos e pergunta a idade de cada um. A partir dos dados obtidos com esse grupo, ele estima quantas pessoas na escola inteira têm mais de 18 anos. Você consegue dizer o que é a população e o que é a amostra nesse exemplo? Exatamente, os 3000 alunos da escola representam a população e o grupo de 100 alunos representa a amostra, ou seja, uma amostra é escolhida de forma aleatória para que se estime alguma informação da população. Conseguiu entender?

Outro termo muito importante é a variável aleatória. Ela associa a cada elemento de um espaço amostral um número real. Você pediu um exemplo? Vamos lá! Ao lançarmos dois dados honestos, o espaço amostral S é formado por todos os resultados possíveis. Assim, $S = \{(1, 1); (1, 2); \dots; (5, 6); (6, 6)\}$. Podemos ter uma variável aleatória X definida como “a quantidade de números menores que 3 em cada rodada de lançamentos”. Dessa forma, se o resultado dos dois lançamentos for (2,2), por exemplo, X vale 2, pois os dois números são menores que 3; se o resultado for (2,4), X vale 1; e se o resultado for (4,5), X vale 0. Além disso, também é importante você saber que se uma variável aleatória pode assumir apenas valores contáveis, damos a ela o nome de variável aleatória discreta, por exemplo, quando ela pode assumir apenas números naturais.

Na sequência temos a distribuição uniforme de uma variável aleatória. Quando uma variável aleatória segue uma distribuição uniforme, significa que a probabilidade de qualquer um dos eventos acontecer é a mesma. No exemplo anterior, a probabilidade de cada um dos resultados acontecer é a mesma e vale $\frac{1}{36}$. Dessa forma, dizemos que X apresenta uma distribuição uniforme.

Já avançamos bem, mas vamos continuar. O valor esperado de uma variável aleatória é o valor médio que essa variável terá caso uma experiência seja realizada muitas vezes. Suponhamos que uma variável aleatória discreta X assuma os valores x_1, x_2, x_3, \dots , e a probabilidade dela assumir cada um desses valores é $p(x_1), p(x_2), p(x_3), \dots$ (representamos todos esses resultados por $p(x_i)$, $i = 1, 2, 3, \dots$). Então o valor esperado de X é a soma de todos os valores assumidos pela variável aleatória multiplicados pela probabilidade de observarmos aquele valor. Esta definição é resumida pela fórmula abaixo:

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i). \quad (1)$$

Começaram a aparecer os símbolos estranhos, mas não se assuste, o uso é mais fácil do que parece. Que tal avançarmos mais um pouco? Está acabando essa parte.

A partir do conceito de valor esperado, fica fácil compreender o que é variância. Variância indica a “distância” dos dados em relação ao valor esperado, ou seja, indica se os dados da distribuição estão mais concentrados ou se estão mais dispersos.

Agora temos a função de probabilidade. Ela associa uma probabilidade a cada possível ocorrência de uma variável aleatória. A função $p(x_i)$ citada no parágrafo anterior é uma função de probabilidade. Mas vamos a um exemplo. Você gosta de baralho? Se tomarmos a variável aleatória X como o naipe de uma carta retirada aleatoriamente de um baralho com 52 cartas, as possíveis ocorrências são: ouros, copas, paus ou espadas. A função de probabilidade associará a cada uma dessas ocorrências uma probabilidade igual a $\frac{1}{4}$.

Por fim, temos a função de verossimilhança. Essa função é utilizada para estimar uma quantidade de interesse a partir de dados obtidos da amostra. Vamos imaginar uma situação contrária a anterior: nós escolhemos 10 cartas de um baralho qualquer (amostra), calculamos a proporção de cada naipe nessa amostra e, por fim, utilizamos uma certa ferramenta para estimarmos a proporção de cada naipe no baralho todo, verificando, assim, se ele é “honesto”. Essa “certa ferramenta” que nós utilizamos é exatamente a função de verossimilhança.

Você consegue perceber que as aplicações da função de probabilidade ou da função de verossimilhança acontecem em sentidos contrários? Pois é, a expressão das duas é a mesma, o que muda é a quantidade de interesse. Por exemplo, a função de probabilidade é dada por $P(x|\theta)$ (lê-se função de x dado θ). Nessa função, θ é conhecido e nós queremos alguma informação sobre x . Já na função de verossimilhança correspondente, $L(\theta|x)$ (lê-se função de θ dado x), a expressão é a mesma, mas nós conhecemos x e queremos alguma informação sobre θ .

Acabamos a parte maçante e agora temos tudo para continuar. Vamos ver como essas definições entram no nosso problema?

3. O PROBLEMA

O nosso objetivo é estimar quantidade de tanques que os alemães estavam produzindo com base nos tanques capturados. O que nós temos disponível, portanto, para realizar esse cálculo é a quantidade de tanques capturados e os números de série de cada tanque. Vamos pensar juntos para já relacionar algumas das definições que nós vimos na seção anterior com esse problema?

Qual seria a nossa população? Isso mesmo! A população é composta por todos os tanques produzidos pelos alemães durante a Segunda Guerra. E como você já deve ter imaginado, a amostra são os tanques capturados. É dessa amostra que retiraremos o maior número de série e, falando no assunto, já aproveitaremos para definir a variável aleatória X como sendo o maior número de série de uma amostra, ela será importante logo mais. Portanto, nós obtemos informações de uma amostra e estimamos uma quantidade de interesse da população. Você se lembra o que nós devemos utilizar para fazer essa transição? É exatamente nesse ponto que entra a função de verossimilhança, tal qual será explicado nas próximas seções.

Para realizarmos os cálculos, podemos proceder de duas maneiras: utilizando a estatística clássica ou a estatística bayesiana. É importante ter em mente que, independente da técnica adotada, é fundamental que os números de série dos tanques sejam ordenados de forma crescente partindo do 1, pois os cálculos foram desenvolvidos a partir dessa premissa. Mas, para a sorte dos Aliados, já se sabia que os alemães ordenavam seus produtos exatamente dessa forma. Dito isso, nós faremos o uso das duas técnicas para decidirmos qual das duas nos traz resultados mais precisos. Quer arriscar um palpite? Qual você acha que vai se mostrar mais precisa? Descobriremos dentro de alguns instantes.

Você pode ter imaginado que a função de verossimilhança terá um papel fundamental no nosso problema. E você tem razão! Ela será utilizada tanto na solução clássica quanto na bayesiana. Dada essa importância, vamos iniciar nossos cálculos determinando essa função.

Precisamos nomear algumas quantidades antes de prosseguirmos: o tamanho da amostra (quantidade de tanques capturados), chamaremos de n ; o tamanho da população (o que nós queremos saber), chamaremos de N ; e o maior número de série da amostra, chamaremos de m .

Para encontrarmos a função de verossimilhança, precisamos calcular a probabilidade de “retirar” uma amostra com maior número de série m e tamanho n entre todos os N tanques produzidos.

Vamos calcular, primeiramente, quantas amostras de tamanho n existem entre os N tanques, valor que representa o número de casos possíveis. Esse cálculo é feito utilizando combinação linear: precisamos encontrar a quantidade de grupos com tamanho n entre os N tanques. Aplicando os valores n e N na fórmula de combinação, temos

$$C_{N,n} = \binom{N}{n}. \quad (2)$$

Agora, precisamos calcular quantas amostras existem com tamanho n e com maior número de série m , valor que representa o número de casos favoráveis, ou seja, os casos que nos interessam. Como o maior número de série foi fixado, ou seja, ele obrigatoriamente tem que valer m , então deveremos ter uma combinação de $m - 1$ tomados em grupos de $n - 1$. Assim, temos

$$C_{m-1,n-1} = \binom{m-1}{n-1}. \quad (3)$$

Segundo os cálculos de probabilidade, devemos dividir o número de casos favoráveis pelo número de casos possíveis para encontrarmos a probabilidade de um evento acontecer. Portanto, a probabilidade que procuramos é dada pela função abaixo:

$$P(m|N) = \frac{\binom{m-1}{n-1}}{\binom{N}{n}}, \quad (4)$$

caso $n \leq m \leq N$, ou 0, caso m esteja fora desse intervalo.

A expressão acima representa a função de probabilidade. Podemos, a partir dela, encontrar a expressão da função de verossimilhança $L(N|m)$:

$$L(N|m) = \begin{cases} \frac{\binom{m-1}{n-1}}{\binom{N}{n}}, & n \leq m \leq N, \\ 0, & \text{caso contrário.} \end{cases} \quad (5)$$

Uma rápida observação: essa função toda tem o mesmo significado do que foi escrito abaixo da função (4), porém agora está condensado em uma única expressão.

Com essas ferramentas em mãos, vamos avançar para cada uma das análises. Mas antes, é importante perceber que, para que os cálculos sejam feitos, apenas o maior número de série é necessário. Dessa forma, não importam quais são os outros números: precisamos saber apenas qual o maior deles e quantos números são ao todo.

4. ANÁLISE CLÁSSICA

A primeira técnica que nós usaremos para solucionar o problema será a análise clássica. Essa solução pode ser encontrada em [1]. Como o valor esperado de uma variável aleatória representa

o valor médio que se obtém ao realizar uma experiência muitas vezes, é interessante termos esse valor em mãos. Portanto, precisamos calcular o valor esperado da variável aleatória X (definida anteriormente como sendo o maior número de série de uma amostra), isto é, $E(X)$. Da expressão (4), sabemos qual o valor de $P(m|N)$, devemos, então, aplicar esse valor na expressão (1). Assim, temos:

$$E(X) = \sum_{m=n}^N mP(m|N) = \sum_{m=n}^N m \frac{\binom{m-1}{n-1}}{\binom{N}{n}}. \quad (6)$$

Antes de escrevermos o resultado, vamos entender algumas notações. O símbolo \hat{N} significa um estimador para N , ou seja, um estimador para o valor de N na população. É esse estimador que procuramos. Outra notação importante é \bar{m} , que representa a média amostral. Como você pode encontrar demonstrado em diversos materiais de estudo sobre estimadores, a média amostral é um bom estimador para o valor esperado. Portanto, como estamos calculando o valor esperado da variável aleatória X , então um bom estimador para esse valor esperado é a média amostral \bar{m} e, assim, podemos dizer que $E(X) = \bar{m}$. Continuando a partir de (6), fazendo as operações matemáticas necessárias, temos o seguinte resultado:

$$E(X) = \sum_{m=n}^N m \frac{\binom{m-1}{n-1}}{\binom{N}{n}} \Rightarrow \bar{m} = n + n \frac{\hat{N} - n}{n + 1}. \quad (7)$$

Isolando \hat{N} , encontramos um estimador para o número total de tanques por meio da análise clássica:

$$\hat{N} = \bar{m} \frac{n + 1}{n} - 1. \quad (8)$$

Todo resultado de estimação precisa ter um intervalo de confiança, isto é, uma margem de erro para o resultado. É a partir do intervalo de confiança que verificaremos a precisão do estimador: caso o intervalo seja grande, o estimador é pouco preciso, e caso o intervalo seja pequeno, o estimador é bastante preciso. Portanto, o próximo passo é encontrar o intervalo de confiança para o valor de N estimado pela expressão (8).

Quando se constrói um intervalo de confiança, a pessoa que está fazendo o cálculo precisa definir qual a probabilidade de o verdadeiro valor da quantidade estimada estar dentro daquele intervalo. Vamos para mais um exemplo. Suponha que alguém escolheu uma confiança de 90% para um intervalo. Isso significa que se fossem avaliadas 100 amostras, em 10 delas o valor real da quantidade analisada não seria coberto pelo intervalo definido. Isso mesmo, estaria errado! Mas então por que não escolhemos sempre uma confiança de 100%? Porque quanto maior a confiança escolhida, maior é o intervalo obtido. Portanto é necessário ter um equilíbrio.

Para o nosso problema, vamos escolher uma confiança de 95% e definir S como sendo o limitante superior do intervalo. Como o $N \geq m$, então basta calcularmos o valor de S para termos um intervalo de confiança para N da forma $IC = [m, S]$, ou seja, N deve ser maior ou igual a m e menor ou igual a S .

Para calcularmos o valor de S , precisamos da função de probabilidade dada em (4). Além disso, temos que o somatório de todos os valores de N possíveis entre m e S deve ser igual à confiança escolhida, 95%, neste caso. Assim, temos

$$\sum_{N=m}^S P(N|m) = \sum_{N=m}^S \frac{\binom{m-1}{n-1}}{\binom{N}{n}} = 0,95. \quad (9)$$

Efetuando os cálculos, concluímos que

$$S = \frac{m}{0,05^{\frac{1}{n}}}. \quad (10)$$

Dessa forma, o intervalo de confiança para N é dado por

$$IC = \left[m, \frac{m}{0,05^{\frac{1}{n}}} \right]. \quad (11)$$

5. ANÁLISE BAYESIANA

Já encontramos uma primeira solução através da análise clássica. Agora é a vez da análise bayesiana, que é apresentada em [2]. Essa análise é feita combinando uma informação a priori com a função de verossimilhança através do Teorema de Bayes. A informação a priori, como o próprio nome sugere, é uma informação que já é conhecida antes de se fazerem os cálculos. Ela pode ser obtida através de estudos prévios ou da observação de eventos. No nosso caso, essa informação virá do serviço de inteligência dos Aliados na época.

A informação é que a população seguia uma distribuição uniforme, ou seja, se escolhermos um tanque ao acaso, a probabilidade de cada um ser sorteado é a mesma. A distribuição uniforme é dada por

$$P(N) = \begin{cases} \frac{1}{N^*}, & m \leq N \leq N^*, \\ 0, & \text{caso contrário.} \end{cases} \quad (12)$$

em que N é o número total de tanques na amostra e N^* é um valor maior que N .

Para encontrar um possível valor para N^* , precisamos recorrer ao serviço de inteligência dos Aliados para que N^* não seja menor que N . A inteligência estimou que os alemães possuíam em torno de 1400 tanques, portanto, para não correr riscos, vamos escolher $N^* = 2000$. Dessa forma, temos

$$P(N) = \begin{cases} \frac{1}{2000}, & m \leq N \leq 2000, \\ 0, & \text{caso contrário.} \end{cases} \quad (13)$$

Agora já temos a informação a priori dada em (13) e a função de verossimilhança dada em (5). O próximo passo é encontrar a informação a posteriori. Para isso, utilizaremos a versão do Teorema de Bayes que relaciona a função de verossimilhança com a informação a priori para obter uma informação a posteriori. Sua fórmula é dada por:

$$P(N|m) = \frac{L(N|m)P(N)}{\sum_{N=m}^{\infty} (L(N|m)P(N))}. \quad (14)$$

Note que $L(N|m)$ é a função de verossimilhança; $P(N)$ é a informação a priori; e $P(N|m)$ é a informação a posteriori, aquela que precisamos para prosseguir com os cálculos. Substituindo a função de verossimilhança e a informação a priori em (14), temos o seguinte:

$$P(N|m) = \frac{\binom{m-1}{n-1} \binom{N}{n}^{-1} \frac{1}{2000}}{\sum_{N=m}^{\infty} \binom{m-1}{n-1} \binom{N}{n}^{-1} \frac{1}{2000}} = \frac{\binom{N}{n}^{-1}}{\sum_{N=m}^{\infty} \binom{N}{n}^{-1}}. \quad (15)$$

A simplificação para esses cálculos utiliza séries binomiais infinitas e é um pouco complicada, por isso omitiremos aqui, mas você pode encontrá-la em sites de computação algébrica como, por exemplo, em [3]. A informação a posteriori (15), após as simplificações, pode ser reescrita como

$$P(N|m) = (n-1) \frac{(m-1)! (N-m)!}{(m-n)! N!}. \quad (16)$$

Por fim, precisamos calcular o valor esperado de N e sua variância a posteriori, utilizando a expressão (16). Mas, novamente, o cálculo é extenso, então usaremos o resultado encontrado em [4]. Este resultado nos mostra que:

$$E(N|m) = \mu = \frac{n-1}{n-2}(m-1). \quad (17)$$

$$Var(N|m) = \sigma^2 = \frac{(n-1)(m-1)(m-n-1)}{(n-2)^2(n-3)}. \quad (18)$$

Um intervalo para N é dado por $N = \mu \pm \sigma$. Portanto, na análise bayesiana temos que

$$N = \mu \pm \sigma = \frac{n-1}{n-2}(m-1) \pm \sqrt{\frac{(n-1)(m-1)(m-n-1)}{(n-2)^2(n-3)}}. \quad (19)$$

Sim, é uma expressão bem feia. Mas acalme-se! Basta substituir os dados encontrados nas amostras e a conta sai fácil. Abaixo veremos um exemplo de como esses cálculos foram usados na prática. Mas antes, é importante comentarmos sobre o significado de $\pm\sigma$.

Nós vimos que na análise clássica existe o intervalo de confiança para qualquer quantidade estimada. Na análise bayesiana também existe um intervalo desse tipo, porém seu nome é intervalo de credibilidade. A interpretação do intervalo de credibilidade é bem direta: a probabilidade de a variável aleatória, no caso N , estar entre seus limites, é a credibilidade do intervalo. O valor de $\pm\sigma$ representa esse intervalo. Isso significa que N tem valor calculado igual a μ e pode variar entre $\mu - \sigma$ e $\mu + \sigma$. O intervalo de credibilidade, portanto, é dado por $N = \mu \pm \sigma$.

6. EXEMPLO

Vamos, agora, imaginar uma situação em que possamos utilizar os resultados dos cálculos feitos acima. Para isso, vamos supor que 20 tanques alemães foram capturados em algumas batalhas. A simulação dos números de série capturados será feita utilizando um gerador aleatório de números. Vamos programar o gerador para nos gerar aleatoriamente 20 números que podem variar de 1 a 500. O tamanho da população, portanto, é 500. Vamos utilizar os cálculos para ver se eles se aproximam desse número. Os 20 números gerados são:

25	27	35	49	84
120	155	173	215	219
269	317	349	353	355
387	397	465	485	495

Observamos que a amostra tem tamanho 20 e o maior número de série é o 495. Aplicando $n = 20$ e $m = 495$, para estimar o tamanho da população N_{c_1} através da análise clássica, dado em (7), temos que

$$N_{c_1} = 495 \frac{20+1}{20} - 1 \approx 519. \quad (20)$$

O intervalo de confiança com 95% de certeza para N_c , dado em (11), é

$$IC_1 = \left[495, \frac{495}{0,05^{\frac{1}{20}}} \right] \approx [495, 575]. \quad (21)$$

Portanto, o valor de N calculado pela estatística clássica vale 519 e pode variar de 495 até 575.

Agora, aplicando $n = 20$ e $m = 495$ em (19) para estimar o tamanho da população N_{b_1} através da estatística bayesiana, temos

$$N_{b_1} = \frac{20-1}{20-2}(495-1) \pm \sqrt{\frac{(20-1)(495-1)(495-20-1)}{(20-2)^2(20-3)}} \approx 521 \pm 28. \quad (22)$$

O intervalo de confiança na análise clássica tem amplitude de $575 - 495 = 80$, enquanto o intervalo de credibilidade na análise bayesiana tem amplitude de $549 - 493 = 56$. Apesar de termos as duas análises mostrando resultados corretos (uma vez que o valor real de 500 está dentro dos dois intervalos), essa diferença de amplitude evidencia que o cálculo bayesiano é mais preciso do que o clássico. Além disso, deve-se notar que, conforme o tamanho da amostra aumenta, o intervalo de confiança ou de credibilidade diminui, tornando o resultado mais preciso.

Para perceber isso, suponhamos que sejam gerados 40 números em vez de 20. Gerando aleatoriamente 40 números entre 1 e 500, temos a matriz abaixo:

21	27	40	44	65
67	78	88	111	128
146	147	152	162	195
201	223	245	251	252
266	274	276	278	296
305	323	326	328	333
269	375	398	406	416
418	448	476	484	497

Agora, o tamanho da amostra é 40 e o maior número de série é 497. Aplicando $n = 40$ e $m = 497$ em (19), vemos o seguinte:

$$N_{b_2} = \frac{40-1}{40-2}(497-1) \pm \sqrt{\frac{(40-1)(497-1)(497-40-1)}{(40-2)^2(40-3)}} \approx 509 \pm 13. \quad (23)$$

Aplicando $n = 40$ e $m = 497$ em (7) e (11), para termos o resultado a partir da análise clássica, temos:

$$N_{c_2} = 497 \frac{40+1}{40} - 1 \approx 508. \quad (24)$$

$$IC_2 = \left[497, \frac{497}{0,05^{\frac{1}{40}}} \right] \approx [497, 536]. \quad (25)$$

Vemos que o intervalo de credibilidade em (23) tem, agora, amplitude de 26. Ou seja, com o aumento de 20 para 40 no tamanho da amostra, a amplitude do intervalo foi reduzida de 56 para 26. No caso do intervalo de confiança, comparando (21) com (25), vemos que a amplitude desse intervalo foi reduzida de 80 para 39. Essas reduções evidenciam que o resultado encontrado está mais preciso.

7. QUAL A NECESSIDADE DISSO HOJE EM DIA?

Chegamos ao fim do artigo, mas depois de tudo isso você deve estar se perguntando qual é, afinal, a utilidade prática de cálculos como estes nos dias atuais. A seguir, está descrita brevemente uma situação em que foram úteis. Uma descrição mais aprofundada pode ser encontrada em [5].

Em 2001, o economista irlandês Cormac Ó Gráda (ou Cormac O'Grada) investigava o colapso do Irish Munster Bank, ocorrido em 1885. Em certo ponto da investigação, Gráda precisava estimar a quantidade de clientes (ou de contas) do banco a partir de dados do escritório central. Basicamente, esses dados informavam a existência de 48 contas, cujos números variavam de 63 até 1812.

Fazendo uma comparação com o nosso problema do tanque alemão, as 48 contas equivalem aos tanques capturados pelos Aliados; o maior número de conta entre todas as encontradas, isto é, o número 1812, equivale ao maior número de série em uma amostra; enquanto o número total de contas equivale ao número total de tanques produzidos pelos alemães.

Utilizando cálculos semelhantes aos apresentados, Gráda estimou a quantidade total de contas existentes no banco. Você consegue dizer quantas são?

Além desse caso, os cálculos são aplicáveis em qualquer outra situação em que há uma organização crescente dos números, partindo do 1. Um outro caso ocorreu em 2008, quando uma pessoa queria estimar o número de iPhones vendidos pela Apple naquele ano. Após perguntar os números de série de alguns usuários de iPhone, a pessoa estimou que, até o fim de setembro de 2008, a Apple havia vendido em torno de 9,1 milhões de iPhones desde o início do ano. Você pode ler uma matéria a respeito desse fato em [6].

Se você se interessou pelo assunto, crie situações com um gerador de números e aplique os cálculos acima desenvolvidos. Pesquise mais exemplos reais em que foram úteis e se aprofunde nos detalhes e conceitos utilizados neste texto.

REFERÊNCIAS

- [1] Goodman, L. A. (1954). "Some Practical Techniques in Serial Number Analysis". *Journal of the American Statistical Association*, 49:97–112.
- [2] Höhle, H. (2006). "Bayesian Estimation of the Size of a Population". *Collaborative Research Center*, 386, Paper 499.
- [3] Wolfram Research (2004). "Hypergeometric Differential Equation". Disponível em: <<http://mathworld.wolfram.com/HypergeometricDifferentialEquation.html>> Acesso em: 03 de março de 2019.
- [4] Marlow, W. (1965). "Factorial Distributions". *The Annals of Mathematical Statistics*, 36:1066–1068.
- [5] Gráda, C. O. (2001). "Should the Munster Bank have been saved?". Tech. Rep. WP01/15, Department of Economics, University of Dublin.
- [6] Arthur, Charles (2008). "Why iPhones are just like German tanks". Disponível em: <<https://www.theguardian.com/technology/blog/2008/oct/08/iphone.apple>> Acesso em: 03 de março de 2019.