

Estimativa de uma proporção

Eduardo Cunha
Raul Aparício Gonçalves

RESUMO E OBJETIVOS

A Inferência Estatística é um tema da disciplina de MACS que permite otimizar o uso da tecnologia na criação de experiências estatísticas, em particular na obtenção de amostras e cálculo imediato de medidas estatísticas.

Nesta atividade pretende-se que os alunos, perante uma situação problema, executem simples sondagens eleitorais sobre o um ato eleitoral já realizado, e que, portanto, são já conhecidos os resultados. Desta forma temos os dados da população em estudo e, embora os alunos não os observem, poderão estudar estatisticamente obtendo vários parâmetros. O parâmetro a estimar será a proporção populacional de votos num dado partido, em branco e nulo. Com recurso à aplicação Listas e Folha de Cálculo, da TI-Nspire CX, e numa página já predefinida, o aluno começará por obter vários dados da população, passando se seguida para outra página onde obterá uma amostra aleatória e introduzirá as fórmulas que lhe permitam calcular medidas estatísticas da sua amostra, no caso a proporção amostral, e, ainda, o cálculo do intervalo de confiança a 95%. Por fim, irá explorar o significado de nível de confiança a 95%.

Em conclusão, com esta atividade pretende-se:

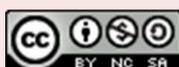
- Ver e rever os procedimentos num processo de Inferência Estatística.
- Recordar e reforçar os termos usado em Inferência Estatística: população, parâmetros, amostra, estatísticas, estimativa, intervalo de confiança.
- Consolidar os conceitos e interpretação de estimativa, intervalo de confiança e nível de confiança.
- Estudar a importância da dimensão da amostra no processo de Inferência Estatística.
- Promover a discussão em sala de aula e a capacidade de análise.

MATERIAIS E PREPARAÇÃO

- TI-Nspire CX ou CX II-T
- Ficha do Aluno (uma por aluno)
- Ficheiro inferencia_sondagem.tns

Esta tarefa deve ser realizada, preferencialmente, em grupo devendo cada grupo ter acesso a pelo menos uma unidade portátil (calculadora) TI-Nspire CX II ou ao respetivo software.

A tarefa pressupõe que os alunos já tenham conhecimento sobre Inferência Estatística, em particular, cálculo da estimativa da proporção populacional, podendo, no entanto, também ser



Estimativa de uma proporção

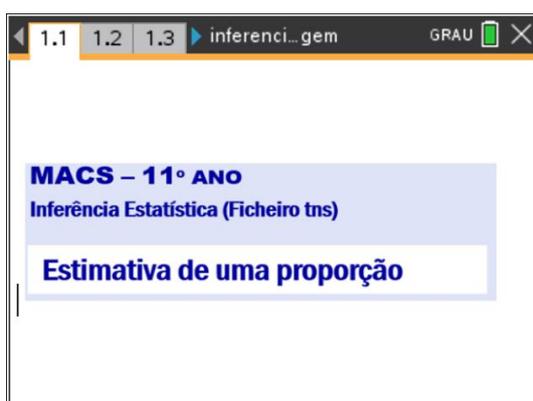
Eduardo Cunha
Raul Aparício Gonçalves

utilizada como introdução ao tema, devendo neste caso o professor reservar algum tempo inicial para expor o processo de Inferência Estatística. Sugere-se que, num ou noutro caso, os alunos possam consultar o seu manual.

TAREFAS E INVESTIGAÇÕES PARA OS ALUNOS

Considera-se que o contexto dado a esta tarefa, situação problema, merece uma leitura cuidada por parte dos alunos e uma discussão em grande grupo turma, em particular, a importância de se estar a realizar sondagens para uma eleição que já ocorreu. Na verdade, este processo tem interesse quase puramente científico, é uma tarefa de investigação académica. Poderão ser levantadas aqui várias questões, como, por exemplo, o voto é uma decisão pessoal e tomada num dado momento, como seria possível garantir que no momento de realização de uma dada sondagem o eleitor tenha a mesma decisão que irá tomar aquando do voto efetivo?! Como garantir que o eleitor não responderá diferente do que pensa votar?! Estas questões não poderão ser respondidas perentoriamente, talvez por isso tenhamos sondagens a falhar!

Colocam-se de seguida capturas de ecrãs das páginas do documento **inferencia_sondagem.tns**:



Página 1.1 - Tema



Página 1.2 - Problema “Sondagem Eleitoral”

	A	B	C	D
1	Eleitores...		Votantes...	
2				
3	PSD+CD, PS	PCP+PE...	BE	
4				
5				

Página 1.3 – Parâmetros da população

	A amostra	B	C	D
1		Abstencção		
2		PSD+CDS		
3		PS		
4		PCP+PEV		
5		BE		

Página 1.4 – Amostra e estimativas

Estimativa de uma proporção

Eduardo Cunha
Raul Aparício Gonçalves

	E	F	G	H
1	Int Confi...	Lim.Infer...	Lim.Super...	Amplitu...
2	PSD+CD			
3	PS			
4	PCP+PE..			
5	BE			

Página 1.4 - Intervalo de confiança

	A n_a...	B	C lim_i...	D lim_s...
1	PSD+...	Dimensão...	0.407	0.683
2	Branco..	50		
3	Absten...	Partido		
4	PSD+...	PSD+CD...		
5	Branco..	Estimativ...		

Página 2.1 – Vários intervalos de confiança

QUESTÃO 1.

Nesta questão pretende-se que os alunos, recorrendo às funções da TI-Nspire CX, obtenham os principais parâmetros da população, em particular a proporção de votos de cada partido.

Para isso, na página 1.3 do ficheiro tns deverão ser introduzidas as seguintes fórmulas:

- célula A2: =dim(eleitores)
- célula C2: =dim(votacao)
- célula D2: =round(c2/a2*1.0, 4)

(Para se obter um número finito de casas decimais pode ser usada a função round(); o fator 1.0 permite que o resultado surja em forma decimal em vez de escrita natural de fração)

- célula A4: =countif(votantes, a3)
- célula A5: =round(a4/\$c2*1.0, 4)

(As fórmulas destas duas células podem ser arrastadas ou copiadas para as células ao lado, note-se que, para tal, se colocou o símbolo de \$ na referência à célula C2, ficando a coluna com referência absoluta.)

	A	B	C	D	E	F
1	Eleitores...		Votantes...			
2	665		395	0.594		
3	PSD+CD..	PS	PCP+PE...	BE	Branco	Nulo
4	186	173	7	5	15	
5	0.4709	0.438	0.0177	0.0127	0.038	0.022
6						
7						
8						

No ecrã acima, obtido na vista de computador do software TI-Nspire CX, encontra-se a tabela completamente preenchida.

QUESTÃO 2.

Nesta questão os alunos irão realizar a sua primeira sondagem eleitoral, recorrendo às ferramentas da tecnologia TI-Nspire CX, a particular a função randsamp() que permite obter amostras aleatórias de uma conhecida população e às funcionalidade da aplicação Listas e Folha de Cálculo.

O professor poderá levantar, aqui, a questão das tabelas de números aleatórios e o facto de num

Estimativa de uma proporção

Eduardo Cunha
Raul Aparício Gonçalves

processo computacional existir um algoritmo e uma semente, e por isso ser habitual designar números pseudoaleatórios. Poder-se-á aproveitar, aqui, para referir a necessidade de cada aluno ter de atribuir uma semente diferente ao algoritmo gerador de números aleatórios da sua TI-Nspire CX, executando numa página de calculadora a instrução **randseed** seguida de um qualquer número.

Na página 1.4 do ficheiro tns e na coluna A, deve ser inserido, na 2ª linha a cinzento, a expressão geradora da amostra aleatória, de uma dada dimensão, partindo população conhecida:

- **randSamp**(Lista, #Ensaios [,Reposição])
Lista – população, nesta atividade é a lista **eleitores**
#Ensaios – dimensão da amostra, começaremos por dimensão 50
Reposição=0 com reposição (predefinição), eleitores repetidos
Reposição=1 sem reposição, não há repetição de eleitores na amostra.

Assim, na linha de edição da expressão geradora da lista **amostra** (coluna A da Folha de Cálculo), deve ser colocada a expressão:

- **randsamp**(**eleitores**,50,1)

Nas restantes células devem ser inseridas as seguintes fórmulas:

- célula C2: **=countif**(**amostra**, b2)
(copiar esta fórmula para as células abaixo)
- célula D2: **=round**(c2*1.0/**dim**(**amostra**), 4)

	A amostra	B	C	D
=	=randsamp			
1	Abstencão	Abstencão	21	0.42
2	PSD+CD	PSD+CDS	14	0.3889
3	PSD+CD	PS	12	0.3333
4	Abstencão	PCP+PEV	0	0.
5	Abstencão	BE	0	0.

Nas restantes células da coluna D, as percentagens são calculadas sobre o número de votantes, que se pode obter subtraindo o número de abstenções (célula C2) à dimensão da amostra, isto é:

- célula D3: **=round**(c3*1.0/(**dim**(**amostra**) – \$c\$2), 4)
(copiar esta fórmula para as células abaixo, usando os atalhos **ctrl**+**C** e **ctrl**+**V** ou arrastando)

Finalizada a construção da tabela o aluno deve comparar os valores das suas estimativas com o valor do parâmetro a estimar, fazendo uma análise crítica. Deve, ainda, partilhar com os colegas o resultado que obteve e a sua reflexão, discutindo entre todos os resultados obtidos pela turma.

Por fim, os alunos podem simular outras sondagens com a mesma dimensão e prestar atenção aos valores que vão obtendo, tendo sempre em atenção os parâmetros a estimar. Irão surgir, dada a dimensão da amostra, muitas variações nos valores das estimativas. Os alunos poderão alterar o valor da dimensão da amostra, por exemplo para 100, e observar qual o efeito desta alteração.

Estimativa de uma proporção

Eduardo Cunha
Raul Aparício Gonçalves

QUESTÃO 3.

Nesta questão os alunos, usando a fórmula para o intervalo de confiança a 95% para a proporção populacional (parâmetro), a partir da proporção amostral (estimativa), inserem as fórmulas que permitem calcular os extremos do intervalo, e de seguida a fórmula para obterem a amplitude do intervalo de confiança.

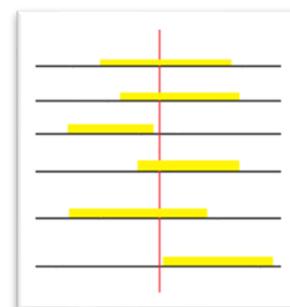
Os alunos terão que usar a referência relativa a outras células nas fórmulas dos extremos do intervalo de confiança, conforme já realizado anteriormente.

Por exemplo, para a coligação “PSD+CDS”, cuja estimativa \hat{p} está calculada na célula d2, devem ser inseridas as seguintes fórmulas:

- célula F2: = d2 – 1.96 * sqrt(d2*(1 – d2) / dim(amostra))
- célula G2: = d2 + 1.96 * sqrt(d2*(1 – d2) / dim(amostra))
- célula H2: = g2 – f2

	E	F	G	H
1	Int Confi...	Lim.Inf...	Lim.Sup...	Amplitude..
2	PSD+CD	0.2538	0.52	0.2662
3	PS	0.2026	0.46	0.2574
4	PCP+PE..	0.	0.	0.
5	BE	0.	0.	0.

Estas fórmulas podem ser copiadas para as células abaixo, usando atalhos **ctrl**+**C** e **ctrl**+**V** ou arrastando-as. Pode ser usada a função **round()** para se obter um certo número de casas decimais. Após o registo dos intervalos obtidos, os alunos, em grupo de trabalho e/ou em grupo turma, deverão analisar os vários intervalos obtidos na turma, realçando-se aqui a amplitude, a variação dos limites inferiores e superiores dos intervalos de confiança e, principalmente, se o intervalo contém os parâmetros a estimar. Poderá ser pertinente representarem-se os vários intervalos na reta real, dispondo-se várias retas horizontais paralelas e centradas no valor do parâmetro a estimar. Esta representação permitirá uma melhor perceção da variabilidade dos intervalos e do número de intervalos que contém o parâmetro.



Observação: o cálculo da estimativa para a proporção foi realizado com base no número de votantes da amostra, daí se ter usado como divisor a diferença entre a dimensão da amostra e o número de abstenções. No entanto, nesta questão, para o cálculo do intervalo de confiança considerou-se a dimensão da amostra.

QUESTÃO 4.

Nesta questão os alunos são confrontados com a questão: o que significará a expressão “um intervalo de confiança a 95%” e, através da simulação de várias sondagens com base em várias amostras aleatórias de uma dada dimensão, são levados à sua resposta de forma intuitiva.

Estimativa de uma proporção

Eduardo Cunha
Raul Aparício Gonçalves

Para tal, na página 2.1 do ficheiro tns, os alunos têm ao seu dispor um simulador de sondagens, onde colocando apenas a dimensão da amostra e o partido/coligação de que se pretende estimar o resultado eleitoral através da sondagem, são construídos intervalos de confiança.

Os campos a alterar pelos alunos são os constantes nas células sombreadas a cinzento, devendo ter o cuidado de colocar o nome do partido/coligação de forma correta, isto

é, exatamente igual ao que consta na amostra gerada. Para evitar erros o aluno poderá optar por copiar da coluna A (nova amostra – **n_amostra**) o nome do partido/coligação e colar na célula B4. Utilizando o atalho que permite repetir os cálculos, clicando sucessivamente nas teclas **ctrl** + **R**, o aluno vai gerando novas amostras e automaticamente às listas/colunas com os limites inferiores e superiores do intervalo de confiança (colunas C e D da folha de cálculo) vão sendo acrescentados novos intervalos de confiança.

Na célula B6 vão surgindo as diferentes estimativas, em função das diferentes amostras, para a proporção (percentagem) de votos do partido selecionado, e que serve de base ao cálculo dos intervalos de confiança. Os limites do intervalo de confiança são calculados nas células N2 e N4, por forma a estarem “escondidas”, e guardados nas variáveis **lim_inf** e **lim_sup**.

As variáveis são depois capturadas, automaticamente, para as listas **lim_inf_ic** e **lim_sup_ic** que se encontram nas colunas C e D. Gerando-se, assim, vários intervalos de confiança ao se repetir os cálculos: **ctrl** + **R**. Clicar duas a tecla **enter** sobre a expressão para reiniciar a captura dados. O aluno pode usar os atalhos **ctrl** + **1** e **ctrl** + **7** para, respetivamente, ir para a última ou primeira linha de uma dada coluna onde se encontra o cursor.

A estratégia para calcular a percentagem de intervalos de confiança que contêm o parâmetro a estimar, que se espera se aproxime dos 95%, pode basear na análise de 10,20 ou 30 intervalos por cada aluno. Também se poderá definir uma estratégia a partir das ferramentas da TI-Nspire CX, como por exemplo uma contagem de “1s” (verdadeiro – o intervalo contém o parâmetro) colocando-se a seguinte expressão geradora na coluna E: **seq(iffn(lim_inf_ic[i]<pp<lim_sup_ic[i],1,0),i,1,dim(lim_inf_ic))**, sendo **pp** a variável onde se encontra guardado o parâmetro proporção do partido em análise.

	A n_a...	B	C lim_i...	D lim_s...
=	=randsa	=capture	=capture	
1	PSD+...	Dimensão...	0.361	0.639
2	PSD+...	50		
3	PS	Partido		
4	PSD+...	PSD+CD...		
5	Absten...	Estimativ...		
B2	50			

	N	O	P	Q
1	Lim Inf IC.			
2	0.45			
3	Lim Sup...			
		b6.		
N4	$b6 - \text{countif}(n_amostra, "Abstenç")$			

	lim_inf_ic	lim_sup...	E pertence...	F
=	=capture('li	=capture('li	=seq(iffn(lim	
1	Dimensão...	0.369	0.616	1
2	100	0.331	0.575	1
3	Partido	0.39	0.643	1
4	PSD+CDS...	0.384	0.631	1
5	Estimativa...	0.265	0.514	1
6	0.57377	0.425	0.669	1
7		0.331	0.575	1
8		0.254	0.504	1
9	Parâmetr...	0.366	0.617	1
10	0.470886	0.424	0.676	1
11		0.408	0.656	1
C	lim_inf_ic:=capture('lim_inf,1)			