

Qual a proporção de casos de Covid-19 no RS?

Prevalência de Covid-19 no RS com amostra aleatória e informação populacional.

Idéia geral

Considere duas quantidades de interesse:

- N_y : o número de casos já confirmados no RS.
- N_x : o número real de casos no RS.

Autoridades e pesquisadores gostariam de conhecer o número verdadeiro de casos da doença N_x , para uma população de tamanho N . Geralmente governos e órgãos responsáveis por testes conseguem divulgar somente o valor de N_y . Suponha que $N_y \leq N_x$ e que N_y seja conhecido, algumas razões para isso são discutidas em (ref.).

Se todas as quantidades definidas acima fossem observadas poderíamos apresentar os dados na forma de uma tabela 2×2 como essa abaixo.

Casos Reais	Não confirmados	Casos confirmados	\sum
Não	???	???	$N - N_x$???
Sim	???	???	N_x ???
\sum	$N - N_y$	N_y	N

Proporções de casos em populações finitas podem ser representadas por probabilidades. Primeiramente defina as variáveis aleatórias:

- Y : indicadora de caso já confirmado ($Y = 1$).
- X : indicadora de caso real ($X = 1$).

Assim a proporção de casos reais da doença, N_x/N , pode ser escrita como a probabilidade de uma pessoa na população ser um caso real, $\pi = P(X = 1)$. A proporção de casos já confirmados N_y/N , representa a probabilidade de uma pessoa ser um caso confirmado, $\rho = P(Y = 1)$. Suponha que $\rho \leq \pi$ (ref.).

Como calcular π ?

1. Estatísticas oficiais?

Devido a quantidade limitada de testes (em situações de emergência), geralmente os casos são detectados apenas em unidades da população em estado grave que procuram atendimento, pessoas que pertencem a grupos de risco, profissionais de saúde, ou algum outro critério de prioridade. O que provavelmente não representará a proporção real de casos na região de interesse.

2. Testar toda a população do RS?

Inviável! (Impossível seja pelo custo ou tempo, não eficiente, ...)

3. Estimar π através de uma amostra aleatória?

Se coletarmos uma amostra aleatória de tamanho n da variável X , $\mathbf{x} = (x_1, \dots, x_n)$, podemos calcular a proporção amostral

$$\hat{\pi} = \frac{\sum_{u=1}^n x_u}{n} = \frac{n_x}{n}.$$

E se também observamos Y para cada unidade?

- Utilizando informações amostrais que representem as contagens no interior da tabela acima conseguimos estimar π com maior eficiência (ref.). Como?

Na amostra teríamos a tabela

Casos Reais	Não confirmados	Casos confirmados	\sum
Não	$n - n_x$	0	$n - n_x$
Sim	$n_x - n_y$	n_y	n_x
\sum	$n - n_y$	n_y	n

Obs.: Pesquisadores ainda podem ter interesse em quantificar a proporção da diferença entre as duas probabilidades, ou seja, desejam estimar uma quantidade μ tal que $\pi = \mu \times \rho$.

Obs. 2: Nesse caso precisaremos conhecer ρ . Sabemos o valor de ρ ? Se assumirmos ρ conhecido (quando o número de casos detectados são precisamente considerados corretos. . .

Obs. 3: Conhecendo ρ , um estimador para μ pode ser dado por $\hat{\mu} = \frac{\hat{\pi}}{\rho} = \frac{n_x}{n\rho}$, se estimarmos $\hat{\pi} = n_x/n$, a proporção amostral de casos reais.

No caso do RS

Estimativas do governo indicam $\pi = ?$. Qual o estimador utilizado? (media amostral?) Assumindo uma população de $N = 113.000.00$ habitantes no RS (ref.). . . Informacoes oficiais dizem que existem $Y = 348$ casos confirmados na população do RS. . .

Fase 1 - retrato de 01/04 (pesquisadores acreditam que testes retratem situação em 01/04)

População	Não era	Já era caso	\sum	Amostra	Não era	Já era caso	\sum
Não caso	???	0	???	Não caso	4.187	0	4.187
Caso real	???	390	???	Caso real	2	0?	2
\sum	111.299.610	390	11.300.000	\sum	4.189	0?	4.189

*Pop. data updated in 30/04/2020 from <http://ti.saude.rs.gov.br/covid19/>

A proporção estimada de casos no RS usando somente os dados da pesquisa Estimativas do governo dizem que

$$\hat{\pi} = 2/4189 \approx 0.00048,$$

O relatório do governo diz que a porporção de pessoas com anticorpos no RS seja aproximadamente igual a 0,05%. O estimador calibrado $\hat{\pi}_A = 0,0005119542$.

Numeros pequenos!!! . . .

Fase 2 - retrato de 14/04 (* mesma idéia da fase 1)

População	Não era	Já era caso	\sum	Amostra	Não era	Já era caso	\sum
Não caso	???	0	???	Não caso	4.994	0	4.994
Caso real	???	777	???	Caso real	6	0?	6

População	Não era	Já era caso	Σ	Amostra	Não era	Já era caso	Σ
Σ	11.299.223	777	11.300.000?	Σ	4.500	0?	4.500

*Pop. data in 30/04/2020 from <http://ti.saude.rs.gov.br/covid19/> **There is information about tests on some relatives from the same Government's study. They are not randomly selected in this case... how to use such information???

- Comparar estimativas brutas do governo... com estimador ajustado...
- Como juntar as informacoes temporais? E espaciais, quando houverem... no caso do RS os nnumeros sao muitissimo pequenos...
- E quando houver testes na rede de esgoto de uma regioao, como combinar diferentes niveis de informacao?

Amostragem + Dados populacionais

Ao invés de utilizar apenas o número de casos na amostra podemos tornar mais eficiente...

Além do teste, se soubermos o número de unidades na amostra que já foram consideradas caso anteriormente...

Um estimador ajustado (calibrado) para π e dado por $\hat{\pi}_A = q + \frac{n_x - n_y}{n}$... como obter? Próxima seção...

Obs. 4: Um estimador ajustado (calibrado) para μ pode ser dado $\mu_A = 1 + \frac{n_x - n_y}{nq}$... com obter?

Função de verossimilhança

Para o problema acima temos $X \sim Binomial(n, \pi)$ e $Y|X \sim Binomial\left(X, \frac{q}{p}\right)$, assim a distribuição conjunta das variáveis é dada por

$$p(Y, X) = \pi^X (1 - \pi)^{1-X} \times (\rho/\pi)^Y [1 - (\rho/\pi)]^{X-Y}$$

Agora sponha que observamos uma amostra aleatória (simples) de tamanho n da população de interesse, $(\mathbf{y}, \mathbf{x}) = \{(y_1, x_1), \dots, (y_n, x_n)\}$. A distribuição conjunta de probabilidade pode ser dada por

$$L(\pi; \mathbf{y}, \mathbf{x}) = \binom{n}{n_x} \pi^{n_x} (1 - \pi)^{n - n_x} \times \binom{n_x}{n_y} (\rho/\pi)^{n_y} [1 - (\rho/\pi)]^{n_x - n_y}$$

Verossimilhança em termos de μ ...

$$L(\mu; \mathbf{y}, \mathbf{x}) = \binom{n}{n_x} \pi^{n_x} (1 - \pi)^{n - n_x} \times \binom{n_x}{n_y} (\rho/\pi)^{n_y} [1 - (\rho/\pi)]^{n_x - n_y}$$

- Isso é o mesmo que pós estratificação, raking e calibragem...

Referências

Estimating the proportion of Corona cases with a random sample

- <https://grazeconomics.wordpress.com/2020/04/22/estimating-the-proportion-of-corona-cases-with-a-random-sample/>

Disclosure limitation for data release??? researchers need data!!!

- <https://www.statschat.org.nz/2020/04/18/prevalence-estimation-is-it-out-there/>
- <https://www.statschat.org.nz/2020/04/19/counting-rare-things-is-hard/>