

Não Resposta

Airlane P. Alencar

13 de junho de 2023



Índice

- 1 Referências
- 2 Objetivo
- 3 Roteiro
- 4 Efeitos da Não Resposta
- 5 Missing at random?
- 6 Viés
- 7 Tipos de Missing
- 8 Ponderação



Referências

- Lohr, S. Sampling.
- Aula de Pamela Campanelli.



Objetivo

- O que fazer para evitar não resposta?
- O que fazer quando temos não resposta?

Exemplo: National Crime Victimization Survey (NCVS)

- Entrevistador pode não conseguir contactar o domicílio;
- Pessoa está doente e não pode responder;
- Recusa a responder
- Pesquisador tenta conseguir informações demográficas sobre os não respondentes (sexo, idade, raça, rural/urbano) e depois tenta levar em conta essas informações na análise.
- Falta de dados só em alguns itens: renda.



Roteiro

- Tentar evitar a não resposta.
- Tentar contactar subamostra dos não respondentes e usar essa subamostra para inferir sobre os não respondentes.
- Usar modelo para predizer os valores dos não respondentes e propor pesos para corrigir os resultados do estudo e/ou usar imputação.
- Ignorar não resposta (não recomendado mas muito usado).



Efeitos da Não Resposta (Missing)

Exemplo: Thomsen e Siring (1983).

- Pesquisa de voto em 1969 na Noruega.
- Três ligações seguida por envio por correio.
- 9,9% de não resposta ao final do estudo.
- Na eleição era possível identificar se a pessoa votou ou não na eleição.

Porcentagem de pessoas que votaram segundo faixa etária para todos da amostra e para os que não responderam

	Todos	20-24	25-29	30-49	50-69	70-79
Não respondentes	71	59	56	72	78	74
Todos da amostra	88	81	84	90	91	87



Holt e Elliot, 1991

Estudo no Reino Unido tem mais não resposta para

- residentes em Londres;
- domicílios sem carro;
- pessoas solteiras;
- casais sem filhos;
- pessoas mais velhas;
- divorciados e viúvos;
- new Commonwealth origin;
- menos escolaridade;
- self-employed (autônomo).



Tamanho de amostra grande não importa

- Literacy Digest Survey de 1936 do exemplo 1.1 de Lohr tem 2,4 milhões de respondentes mas a taxa de resposta é só de 25%!
- Censo decenal americano não cobre toda a população e a taxa de cobertura varia de acordo com diferentes grupos demográficos. Detalhes em Mulry (2005) sobre censo em 2000.

Manda carta, questionário e cartão para lembrar de responderem. Só 65% dos questionários voltam preenchidos.

Das casas não desocupadas, são respondidos 74% dos quest.

Das casas não desocupadas que não responderam, seguem pesquisadores para entrevista (42 milhões de domicílios).

Alternativas: ligação gratuita por telefone; internet; be counted form espalhados em bibliotecas, detrans e estações de metrô.



Ignora não resposta

A maioria dos estudos ignora a não resposta e apresenta seus resultados como o estudo de Hite no cap. 1 de Lohr sobre insatisfação nas relações.

- Livro Women and Love: A cultural revolution in progress (1987) de Hite.
- 84% das mulheres são “não satisfeitas emocionalmente em seus relacionamentos”.
- 70% das casadas por 5 anos ou mais tiveram relações fora do casamento.
- Recebeu muitas críticas como a Time.
- Hite mandou 100 mil questionários e só recebeu 4,5% preenchidos.
- Os questionários foram enviados para associações como de defesa das mulheres, grupos de aconselhamento,..., em geral de mulheres que participam do grupo all-women dos EUA.
- O questionário tem 127 itens (longo).



Missing at random?

- A maioria dos estudos negligencia a análise da não resposta.
- Nesse caso, está assumindo que os não respondentes são semelhantes aos respondentes.
- Ex: Na pesquisa NCVS (National Crime Victimization Survey), a taxa de vítimas é subestimada se ignora a não resposta (Bider and Cantor, 1984).
- Resultados que só consideram as respostas completas são representativas só para a população de respondentes!
- Isso tem que estar claro no estudo.



Pop Express of the Social Sciences Research Centre of the University of Hong Kong (aula Campanelli)

Telephone Survey antes da anexação de Hong Kong de volta para a China em 1997

Dias antes	Tamanho da amostra	% confiantes
1000	1716	35%
900	554	45%
800	640	43%
700	990	41%
600	550	44%
500	560	44%
400	538	55%
300	515	58%
200	1301	66%
100	502	63%
75	525	65%
50	536	72%
25	544	70%



Pop Express of the Social Sciences Research Centre of the University of Hong Kong

Telephone Survey antes da anexação de Hong Kong de volta para a China em 1997

Dias antes	Tamanho da amostra	% confiantes	Response
1000	1716	35%	76%
900	554	45%	64%
800	640	43%	52%
700	990	41%	58%
600	550	44%	54%
500	560	44%	65%
400	538	55%	51%
300	515	58%	53%
200	1301	66%	65%
100	502	63%	43%
75	525	65%	42%
50	536	72%	44%



Viés

$$\mu = \frac{N_R}{N} \mu_R + \frac{N_M}{N} \mu_M$$

Viés aproximado se uso só respondentes:

$$E(\bar{Y}_R) - \mu \approx \mu_R - \mu = \mu_R - \frac{N_R}{N} \mu_R - \frac{N_M}{N} \mu_M = \frac{N_M}{N} (\mu_R - \mu_M)$$



Para reduzir o viés

- Investir tempo no planejamento da pesquisa e no seguimento da não resposta (non response follow-up).
- Estudos que enviam questionários na área de vendas tem entre 10% a 15% de resposta!
- Estudos para planejar os experimentos (curso MAE0317) Montgomery (2000) e para não resposta Hidiroglou et al. (1993).
- ex: Estudo fatorial sobre a não resposta no censo americano de 1990 (Dillman et al., 1995)
Não resposta de 65%.
Se nem manda carta antes, envelope para resposta e cartão lembrete a resposta é de 50% e se manda tudo isso a resposta é de 60%.



Para reduzir o viés

- Pesquisa por Correio, Telefone, Fax (já era), Internet tem menor resposta do que pesquisa pessoalmente.
- Treinar entrevistadores.
- Pesquisas com questionários autopreenchidos tem menos respostas do que com entrevistador.
- Pesquisas feitas no computador podem ter respostas para todos os itens mas isso não significa que as respostas sejam verdadeiras. Às vezes a resposta é obrigatória e respondem de qualquer jeito.
- Responder à pesquisa já é um favor e o questionário deve ser curto e conciso.

Para diminuir o questionário, pode aplicar subconjuntos de questões para alguns subconjuntos de respondentes.

- Clara e boa introdução ao estudo.
- Incentivos e Desincentivos
- Follow-up é mais caro que mandar lembrete para que se responda



Tipos de Não resposta - Little e Rubin, 2002

$$\phi_i = P(\text{resposta}_i) = P(R_i = 1)$$

- Missing Completely at Random (MCAR): Não resposta completamente aleatória (utopia do pesquisador= devaneio)
Se ϕ_i não depende da resposta y_i , nem das covariáveis x_i , nem do planejamento do experimento.
- Missing at Random given covariates (MAR): Não resposta aleatória condicional às covariáveis
Se ϕ_i depende de x_i mas não de y_i .
Pode propor modelo para ϕ_i em função de x_i .
- Not Missing at Random (NMAR): faltante não aleatória.
Ex: a pessoa vítima de crime menos provavelmente irá responder à pesquisa.
Pesquisa sobre atos ilícitos: drogas, aborto, prostituição,



Pesos - Aula Campanelli(2019)

Pesos baseados na amostra

- Para ajustar probabilidades de seleção desiguais
- Para ajustar não resposta

Pesos baseados na população

- Para ajustar não cobertura
- Para ajustar não resposta
- Para reduzir erros amostrais usando pós estratificação
- Para calibrar totais

Lavalée, Beaumont (2015). Why we should put some weight on weights. <https://surveyinsights.org/?p=6255>



Pesos - Aula Campanelli(2019)

- Alguns elementos da população são menos representativos para os dados respondidos.
- Ponderação pode compensar essa subrepresentação.
- Qualquer variável disponível para os dados de respondentes e não respondentes pode ser utilizada, mas quanto mais a variável for relacionada com a resposta do estudo, mais efetiva será para reduzir o viés de não resposta.



Pesos - Aula Campanelli(2019)

	R	NR	Total	Resp	Peso	% pop	Esp	R rate	Peso pop
H	1360	280	1640	0.8293	1.206	47%	1692	0.8038	1.244
M	1757	203	1960	0.8964	1.116	53%	1908	0.9209	1.086
Tot	3117	483	3600	0.8658					

Peso é o inverso da proporção de resposta.

Usando as proporções de homens e mulheres na pop, posso calcular qual deveria ser o tamanho da amostra em cada estrato (Esp=Esperado), e só então calcular a proporção de respondentes ($Rrate = R/Esp$).



Pesos - Aula Campanelli(2019)

	R	NR	Total	Resp	Peso	% pop	Esp	R rate	Peso pop
H	1360	280	1640	0.8293	1.206	47%	1692	0.8038	1.244
M	1757	203	1960	0.8964	1.116	53%	1908	0.9209	1.086
Tot	3117	483	3600	0.8658					

Quem lê tabloides?

	Lê	Não	Total	% Lê
H	1088	272	1360	80.0%
M	176	1581	1757	10.0%
Total	1264	1853	3117	40.6%

Conclusão: 40,6% leem.



Pesos - Aula Campanelli(2019)

	R	NR	Total	Resp	Peso	% pop	Esp	R rate	Peso pop
H	1360	280	1640	0.8293	1.206	47%	1692	0.8038	1.244
M	1757	203	1960	0.8964	1.116	53%	1908	0.9209	1.086
Tot	3117	483	3600	0.8658					

Quem lê tabloides?

	Amostra				Corrigido Peso amostra			
	Lê	Não	Total	%Lê	Lê	Não	Total	%Lê
H	1088	272	1360	80.0%	1312.0	328.0	1640	80.0%
M	176	1581	1757	10.0%	196.3	1763.7	1960	10.0%
Total	1264	1853	3117	40.6%	1508.3	2091.7	3600	41.9%

41,9 % leem.



Pesos - Aula Campanelli(2019)

	R	NR	Total	Resp	Peso	% pop	Esp	R rate	Peso pop
H	1360	280	1640	0.8293	1.206	47%	1692	0.8038	1.244
M	1757	203	1960	0.8964	1.116	53%	1908	0.9209	1.086
Tot	3117	483	3600	0.8658					

Quem lê tabloides?

	Lê	Não	Total	%Lê	Lê	Não	Total	%Lê
H	1088	272	1360	80.0%	1353.6	338.4	1692	80.0%
M	176	1581	1757	10.0%	191.1	1716.9	1908	10.0%
Total	1264	1853	3117	40.6%	1544.7	2055.3	3600	42.9%

42,9 % leem.



Exercício - Aula Campanelli(2019)

	R	NR	Total	%R	Peso
Escola 1	48	102	150		
Escola 2	108	42	150		
Total	156	144	300		

	Respondentes			%Fuma	Corrigido por Pesos			
	Fuma	Não	Total		Fuma	Não	Total	%Fuma
Esc.1	6	42	48					
Esc.2	27	81	108					
Total	33	123	156					

