1.1 Using data to answer statistical questions

Luis Talavera

2022-11-11

Exercises

1.1 Aspirin the wonder drug An analysis by Professor Peter M Rothwell and his colleagues (Nuffield Department of Clinical Neuroscience, University of Oxford, UK) published in 2012 in the medical journal The Lancet (http://www. thelancet.com) assessed the effects of daily aspirin intake on cancer mortality. They looked at individual patient data from 51 randomized trials (77,000 participants) of daily intake of aspirin versus no aspirin or other anti-platelet agents. According to the authors, aspirin reduced the incidence of cancer, with maximum benefit seen when the scheduled duration of trial treatment was five years or more and resulted in a relative reduction in cancer deaths of about 15% (562 cancer deaths in the aspirin group versus 664 cancer deaths in the Control group). Specify the aspect of this study that pertains to (a) design, (b) description, and (c) inference.

Ans.

a. Design: They looked at individual patient data from 51 randomized trials (77,000 participants) of daily intake of aspirin versus no aspirin or other anti-platelet agents.
b. Description: duration of trial treatment was five years or more and resulted in a relative reduction in cancer deaths of about 15% (562 cancer deaths in the aspirin group versus 664 cancer deaths in the Control group)
c. Inference: According to the authors, aspirin reduced the incidence of cancer, with maximum benefit seen when the scheduled duration of trial treatment was five years or more

1.2 Poverty and age The Current Population Survey (CPS) is a survey conducted by the U.S. Census Bureau for the Bureau of Labor Statistics. It provides a comprehensive body of data on the labor force, unemployment, wealth, poverty, and so on. The data can be found online at www.census.gov/cps/. The 2014 CPS ASEC (Annual Social and Economic Supplement) had redesigned questions for income that were implemented to a sample of approximately 30,000 addresses that were eligible to receive these. The report indicated that 21.1% of children under 18 years, 13.5% of people between 18 to 64 years, and 10.0% of people 65 years and older were below the poverty line. Based on these results, the report concluded that the percentage of all people between the ages of 18 to 64 in poverty lies between 13.2% and 13.8%. Specify the aspect of this study that pertains to (a) description and (b) inference.

Ans.

a. Description: The report indicated that 21.1% of children under 18 years, 13.5% of people between 18 to 64 years, and 10.0% of people 65 years and older were below the poverty line.
b. Inference: Based on these results, the report concluded that the percentage of all people between the ages of 18 to 64 in poverty lies between 13.2% and 13.8%.

1.3 GSS and heaven Go to the General Social Survey website, http://sda.berkeley.edu/GSS. Enter HEAVEN as the row variable and then click Run the Table. When asked whether they believed in heaven, what percentage of those surveyed said yes, definitely; yes, probably; no, probably not; and no, definitely not?

Ans.

- yes, definitely 64.6%
- yes, probably 20.8%
- no, probably 8.7%
- no, definitely 5.9%

HEAVEN **BELIEF IN HEAVEN**

Description of the Variable

1382. Do you believe in. . . c. Heaven

Percent	N	Value	Label
64.6	2,399	1	YES, DEFINITELY
20.8	774	2	YES, PROBABLY
8.7	322	3	NO, PROBABLY NOT
5.9	218	4	NO, DEFINITELY NOT
	55,591	0	IAP
	172	8	DK
	123	9	NA
100.0	**59,599**		**Total**

Properties

Data type:	numeric
Missing-data codes:	0,8,9
Mean:	1.56
Std Dev:	.88
Record/column:	1/6189

Figure 1: Do you believe in. . . c. Heaven?

1.4 GSS and heaven and hell Refer to the previous exercise. You can obtain data for a particular survey year such as 2008 by entering YEAR(2008) in the Selection Filter option box before you click Run the Table.

- **a.** Do this for HEAVEN in 2008, giving the percentages for the four possible outcomes. Note that HEAVEN is not available for the 2014 data because the question wasn't asked that year.
- **b.**Summarize opinions in 2008 about belief in hell (row variable HELL). Was the percentage of "yes, definitely" responses higher for belief in heaven or in hell?

Ans.

- **a.**

Variables					
Role	**Name**	**Label**	**Range**	**MD**	**Dataset**
Row	**HEAVEN**	BELIEF IN HEAVEN	1-4	0,8,9	1
Filter	**YEAR(2008)**	GSS YEAR FOR THIS RESPONDENT	1972-2014		1

Frequency Distribution		
Cells contain: -**Column percent** -N of cases		**Distribution**
HEAVEN	1: YES, DEFINITELY	**64.3** 853
	2: YES, PROBABLY	**20.8** 276
	3: NO, PROBABLY NOT	**8.8** 117
	4: NO, DEFINITELY NOT	**6.0** 80
	COL TOTAL	***100.0*** *1,326*

Figure 2: Do you believe in. . . c. Heaven? 2008

- **b.**

Variables					
Role	**Name**	**Label**	**Range**	**MD**	**Dataset**
Row	**HELL**	BELIEF IN HELL	1-4	0,8,9	1
Filter	**YEAR(2008)**	GSS YEAR FOR THIS RESPONDENT	1972-2014		1

Frequency Distribution		
Cells contain: -**Column percent** -N of cases		**Distribution**
HELL	1: YES, DEFINITELY	**52.6** 698
	2: YES, PROBABLY	**20.3** 270
	3: NO, PROBABLY NOT	**14.8** 196
	4: NO, DEFINITELY NOT	**12.3** 163
	COL TOTAL	***100.0*** *1,327*

Figure 3: Do you believe in. . . c. hell? 2008

The answer "yes, definitely" has higher percentage of responses for belief in heaven than belief in hell

1.5 GSS for subject you pick At the GSS website, click Standard Codebook under Codebooks and then click Sequential Variable List. Find a subject that interests you and look up a relevant GSS code name to enter as the row variable. Summarize the results that you obtain.

Ans. The variable of my interest was "HELPFUL", this variable is measured by the following question " Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?"

HELPFUL **PEOPLE HELPFUL OR LOOKING OUT FOR SELVES**

Text of this Question or Item

161. Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

Percent	N	Value	Label
48.8	18,320	**1**	HELPFUL
44.5	16,674	**2**	LOOKOUT FOR SELF
6.7	2,516	**3**	DEPENDS
	21,840	**0**	IAP
	190	**8**	DK
	59	**9**	NA
100.0	**59,599**		**Total**

Properties

Data type: numeric
Missing-data codes: 0,8,9
Record/column: 1/670

Figure 4: HELPFUL summarize