Statistical Literacy

Problem Set 7

Rémi Viné
Due November 13th, 2023

1. **Some understanding on foundations.**

    (a) The point estimate $\hat{p}$ can be lower than the lower bound of the confidence interval (regardless on the confidence level)

        A True

        B False

    (b) The point estimate $\hat{p}$ is exactly in the middle (midpoint) of the confidence interval (regardless on the confidence level)

        A True

        B False

    (c) The larger the standard error the larger the Margin of Error (for a given α)

        A True

        B False

    (d) Having the maximum level of confidence is always a good choice, *i.e.,* a 99.999% confidence interval will be preferred over a 95% confidence interval.

        A True

        B False

2. **Proportions.**

(a) When constructing a 99% CI for proportions, what critical value will you use?

(b) Using a sample of 55 observations, observing a proportion of 0.44, and looking at a 95% confidence interval (of a given variable), what is the Margin of Error in such a case?

(c) The UNHCR is interested at the share of children (age lower than 18 years old) entering its camp. Assume that they do not have census data and conduct a survey on a *representative* sample of 100 asylum seekers. The proportion of children is 38%.

    i. You are asked to construct a 99% confidence interval for the population proportion of the share of children requesting asylum in UNHCR camps.

    ii. Assume that you consider three groups: (i) children, (ii) women, (iii) men, and assume that you consider that each group is equally represented in camps (children, women, men, are each accounting for a third of the population). Does the confidence interval calculate contest this *prior* assumption?

    iii. Now, what if the confidence level is not 99% but 38.3% *(Note that such a low level of confidence is very unlikely to ever be used in statistics)*. Would this change your answer to the previous question?

    iv. Based on the previous example and your understanding of the Margin of Error, how does the Margin of Error change with the confidence level $(1-\alpha)$?

    v. Assuming the UNHCR wants to make sure that the 33.33% bound is not included in the confidence interval for the reasons mentioned above. **Rounding down** (to make sure one does not exceed the 33.33% with this maximum MoE) at the **sixth** decimal number, what would then be the maximum Margin of Error?

    vi. Using this maximum Margin of Error and the point estimate found in the current study, $\hat{p} = 0.38$, what is the required sample size for a 95% confidence interval whose interval does not cross the 33.33% threshold?

(d) Assume that you want to investigate the share of battles using radioactive weapons. You take a sample of 550 battles in history and observe a proportion of 1 battle (out of 550) using radioactive components. Can you construct a 90% confidence interval of the proportion of battle using radioactive components? Can you give a general comment on the issue raised here?

3. **Biases.** *Difficult*

Assume poll agencies are interested in knowing the proportion of people who purposefully declare their income wrongly to the tax authority. Suppose poll agencies in fact made so many polls that all possible configurations of samples of 10000 people have been made. The following proportions are obtained:

Polls	Proportion of polls	Share of people wrongly stating
Online	53%	12%
Phone	12%	9%
On-street	20%	5%
Knocking the door	15%	16%

Table 1: Polls

(a) The statistical agency of the country, gathering all income tax declarations and cross-comparing them with different sources of information (employers, financial markets, *etc.*) found that the proportion of wrong statements is in fact 22%. What can you conclude on the polls?

(b) If one online poll (of 10000 people) got a proportion of 15% of people wrongly reporting, do you think that increasing the share of the surveyed would lead the pool result to be closer to 22%? Comment on this.