Method card 7

Course: statistics
Topic: Confidence interval for a parameter 2024-2025

In this card, we explain how to fill in an answer sheet for a confidence interval (CI).

The situation is the following: we wish to get some information about the distribution of a statistical variable (mean, variance, relative frequencies) on a large population but our only information concerns a sample of a smaller **size**. Our problem is to give an interval for this numerical information (mean, variance, relative frequencies), that we call a **parameter**, with a given **level of confidence**.

Step 1. Identifying the context. We have to identify

(a) the population,
(b) the sample,
(c) the variable under study,
(d) the parameter we want information about (mean, variance or proportion), θ.

Step 2. Mathematical setting. We assume that we have a sample $(X_1, ..., X_n)$ of parent variable X. The sample average and variance of the sample are the random variables

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

You have to identify

(a) the parent variable and the parameter: there are two cases.
- If the identified parameter θ is a proportion then the parent variable is a Bernoulli variable and the parameter is the parameter of the Bernoulli variable.
- In other cases (mean, variance) the parent variable is the variable under study and the parameter is its mean or its variance. You have to specify all the information you have on the distribution of this variable in order to proceed to the next steps.

(b) the estimator Θ of the parameter θ: there are two cases.
- The parameter is a mean or a proportion, then the estimator is $\Theta = \overline{X}$.
- The parameter is a variance, then the estimator is

$$\Theta = T_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2,$$

if we know that X has mean m, or $\Theta = S_n^2$ is the mean of X is unknown.

(c) the variable of interest D, (for computing the CI) and its (approximate or exact) distribution. There are several cases.

(i) The parameter θ is a variance. The only case when we can built a CI is when X is a Gaussian variable. Refer to cases A.1.3 and A.1.4 of Appendix A of the lecture.
(ii) The parameter $\theta = m$ is a mean (includes the case of a proportion as the mean of a Bernoulli variable). The cases are summarized in the following table:

X disitribution	sample size	variance	Case (Appendix A)	Comment
Gaussian	all	known as σ^2	A.1.1	
Gaussian	all	unknown	A.1.2	
Non Gaussian	large	known as σ^2	A.1.6	(1)
Non Gaussian	very large	known as a function of θ, $g(\theta)$	A.1.5	(2)
Non Gaussian	very very large	unknown	A.1.6	(3)

Table 6: Confidence interval for a mean

(1) D as in A.A.1 but the distribution is approximative.
(2) A.1.5 is an example with a Bernoulli distribution but also works for other variables (Poisson, exponential, ...) by replacing $\overline{x}_n(1-\overline{x}_n)$ by the proper $g(\overline{x}_n)$.
(3) As for the known variance case but with the actual value s_n^2 of S_n^2 instead of σ^2.

(d) The expression of the confidence interval. Copy the proper interval from the Appendix taking into account whether it should be tw-sided or right/left-sided (this should be clear from the statement).

Step 3. Numerical results. At this stage, you should know exactly the form of the confidence interval. Usually, you are given in the statement everything to compute its actual bounds:

(a) the size of the sample (given),
(b) the confidence level (given),
(c) the actual value of the estimator or so called point estimate (given or computable with the data),
(d) the margin of error (you need a statistical table or a computer to obtain the required quantile),
(e) the confidence interval (as defined in the previous step).