# Notes on Testing Statistical Hypotheses by Lehmann and Romano

compiled by D.Gueorguiev, 4/30/24

## Statistical Inference, Statistical Decisions, and the Purpose of Statistics

Each statistical observation has a set of observations. The values in this set are the values taken by a random variable whose distribution is at least partly unknown. The parameter represents a label of the distribution which can uniquely identify the latter, but the parameter value is unknown; it is known only that lies in a certain set which will be denoted as *parameter space*. Generally, *Statistical inference* is concerned with the methods of using the given observation data to obtain additional information about the distribution of X or the parameter with which it is labeled.

The need for statistical analysis stems from the fact that the distribution of , and hence some aspect of the situation underlying the mathematical model, is not known. The consequence of such lack of knowledge is uncertainty as to the best mode of behavior. To formalize this, suppose that a choice has to be made between a number of alternative actions. The observations, by providing information about the distribution from which they came, also provide guidance as to the best decision. The problem is to determine a rule which, for each set of values of the observations, specifies what decision should be taken. Mathematically, such a rule is a function , which to each possible value of the random variables assigns a decision , that is, a function whose domain is the set of values of and whose range is the set of possible decisions.

In order to see how should be chosen, one must compare the consequences of using different rules. To this end suppose that the consequence of using different rules. To this end, suppose that the consequence of taking decision when the distribution of is is a *loss*, which can be expressed as a nonnegative real number . Then the long-term average loss that would result from the use of in a number of repetitions of the experiment is the expectation evaluated under the assumption that is the true distribution of . This expectation, which depends on the decision rule and the distribution is named as the *risk function* of and will be denoted by . By basing the decision on the observations, the original problem of choosing a decision with loss function is thus replaced by that of choosing , where the loss is now .

The above statement suggests that the aim of statistics is the selection of a decision function which minimizes the resulting risk. As will be seen later, this statement is not sufficiently precise to be meaningful; its proper interpretation is in fact one of the basic problems of the theory. (*Note to myself*: very nice introductory discussion about the goal of Statistics as a theoretical instrument - I typed it one-on-one because I liked it so much).

## Specification of a Decision Problem

The methods required for the solution of a specific statistical problem depend quite strongly on the three elements that define it: the class to which the distribution of X is assumed to belong; the structure of the space of possible decisions

## References

[Testing Statistical Hypotheses, E.L. Lehmann, J.P. Romano, Third Edition, Stanford U., Springer, 2005](https://github.com/dimitarpg13/generalized_synthetic_control_for_testops/blob/main/articles/hypothesis_testing/Lehmann_and_Romano-TestingStatisticalHypotheses.pdf)