On some assumptions of the null hypothesis statistical testing

Alexandre Galvão Patriota*

Departamento de Estatística, IME, Universidade de São Paulo
Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil

Abstract

It is widely known that Bayesian and classical statistical approaches depend on different logical principles. In order to avoid mistaken inferences and misguided interpretations, the practitioner must respect the inference rules embedded into each statistical method. This paper intends to discuss and to explicit some important assumptions inherent to classical statistical models and null statistical hypotheses. Furthermore, the p-value definition and its limitations are analyzed and an alternative measure of evidence, the s-value, is discussed. It is offered the steps to compute s-values and also standard examples them with comparing p-values. The examples denunciate that p-values, as opposed to s-values, fail to hold the logical consequence.

Key-words: Classical statistics, Inference, Logical Principles, P-value, Statistical hypothesis

1 Introduction

In social sciences, the majority of the events are contingent, full of uncertainties and permeate by nuisance variables. For instance, cognitive skills are affected by a number of factors such as education, culture, age, tiredness, genetics, etc. It is impractical to contemplate all factors that influence a specific cognitive skill. Probability and statistical models are mathematical tools utilized to handle with contingent and uncertain events (Fisher, 1955; McCullagh, 2002; Kadane, 2011). These tools are defined in terms of sets and functions, which are fully consistent with the modern formulation of mathematics. Statistical models are employed to make inferences about unknown quantities and to test the consistency of scientific statements (Fisher, 1955). However, as all models in mathematics, statistical models have their own domains of applicability, their own internal rules, their own principles, their own limitations and so on (Fisher, 1922; Hájek, 2008; Dempster, 1968). It is important to comprehend those internal features in order to avoid inadequate interpretations obtained from prohibited inferential rules (Fisher, 1955; Kempthorne, 1976; Berger and Sellke, 1987; Lavine and Schervish, 1999). This paper discusses some hidden assumptions of statistical model and null hypotheses. Thereafter, the formal definition of a p-value and its limitations are analyzed. Finally, it is revisited a new measure of evidence, called s-value, that overcomes some limitations of the p-value. The steps to compute s-values are provided as well as some standard examples. The examples illustrate a feature of p-values: they do not respect the reasoning of the logical consequence (if one hypothesis $H_{01}$ implies another one $H_{02}$, then, by the logical consequence, *email: patriota@ime.usp.br; fax: (+55 11) 3091-6130

\*
we would expect more evidence against $H_{02}$ than that against $H_{01}$). S-values are much easier to compute than p-values and they do respect the reasoning of the logical consequence. Therefore, the general advice of this paper is to employ s-values as additional measures of evidence for testing null statistical hypothesis or as support measures over the parameter space (or over the space of hypotheses).

2 Statistical Models

It is extremely difficult to introduce probability and statistical models by adopting an easy language without ambiguity. This paper avoids the set-theoretic notation and will not introduce the primary probability space where all quantities are well defined (random variables, statistics, estimators, induced spaces, etc.). The reader should be aware that the language used here is informal, and to avoid ambiguities it will be required to make many textual caveats. The reader is referred to Cox and Hinkley (1974), Schervish (1995), Lehmann and Casella (1998) and McCullagh (2002) for a detailed discussion on statistical models.

Roughly speaking, the steps before choosing a statistical model are:

1. Define the objectives of the study;
2. Define the population of interest;
3. Define the quantities of interest;
4. Define an adequate experiment to collect the sample.

The practitioner must have vast prior knowledge to construct an appropriate experiment to access the quantities of interest, for each field has its idiosyncrasies that must be taken into account. The experiment may be randomized in specific strata or layers or clusters (different treatments, genders, groups of risk and so on). All these features guide us to determine the class of probability distributions that will be considered in the statistical model. Typically, in scientific experiments, there are direct observable quantities (age, gender, measured height and weight, etc.) and unobservable quantities (intelligence, “feelings of morale”, “sense of belonging”, etc.). These quantities might be random and non-random and are ingredients of a statistical model.

In this paper, random observable quantities are denoted by upper-case Latin letters, say $X$ or $T$, and their observed counterparts are denoted by lower-case Latin letters, say $x$ or $t$. Random and non-random unobservable quantities are denoted by the Greek letters $\gamma$ and $\theta$, respectively. The unobservable random quantities are called latent random variables (Bollen, 2002). Let us informally represent a statistical model by the triplet

$$ (X, \gamma, M), $$

where $X$ represents the observable random variables, $\gamma$ represents the latent random variables and $M$ is a family containing joint probability (density) functions of the random variables, that is, $M = \{g_\theta : \theta \in \Theta \subseteq \mathbb{R}^p, p < \infty \}$ where $g_\theta$ is a possible joint probability (density) function of $(X, \gamma)$, for each $\theta \in \Theta$. It should be clear that $\theta \in \Theta$ is an indexer of possible probability distributions, here it is not a random variable. Through residual analyses, one can verify if the family $M$ is adequate or inadequate to be used. Clearly, it is not possible to assure that the family $M$ contains the
true generator mechanism of the data, that is, the mechanism that effectively generates the data. Furthermore, the true generator mechanism can hardly exist in terms of probability distributions. In simulation studies, we know the data true generator mechanism, since we generate them by using computational devices. In practice, however, this expression has many conceptual problems and it is beyond the scope of this paper to discuss the ontological meanings of the word true.

When it is known the probability distribution the governs the random quantities, then \( M \) contains only one element, namely \( M = \{ g \} \), where \( g(x, \gamma) \equiv f_\gamma(x)f_0(\gamma) \) is the joint probability (density) function of the observable and unobservable random variables, with

\[
X|\gamma \sim f_\gamma \quad \text{and} \quad \gamma \sim f_0,
\]

where \( f_\gamma \) is the probability (density) function of the random variable \( X \) given \( \gamma \) and \( f_0 \) is the probability (density) function of the random variable \( \gamma \). In this latter example, it is assumed that the joint probability distribution that governs the random quantities is known. In this context, it is possible to provide full probabilistic descriptions of the random quantities (mean, variance, quantiles, probabilities, etc.). As aforementioned, in practice it is difficult to known the true generator (if it exists) of the random quantities and the family \( M \) typically has more than one element.

The formal statistical model is defined with sigma-fields and a family of probability measures (see, for instance, Lehmann and Casella, 1998; McCullagh, 2002; Lehmann and Romano, 2005; Patriota, 2013), the reader must keep in mind that model (2.1) is a simplified version that shall help us to understand some important features of the classical statistical model and the null hypothesis statistical testing.

3 Scientific and Statistical Hypotheses

In science, it is common to formulate statistical hypotheses to test scientific statements. A non-trivial step is to translate a scientific statement into statistical language. In the classical paradigm, a statistical hypothesis is a statement about probability distributions that potentially govern the experimental data. That is, in order to create a statistical hypothesis, one must be able to transform a scientific statement in terms of probability distributions. For instance, the statement “this coin is not biased” is typically transformed into “\( P(\text{this coin turns up head}) = 0.5 \)”, that is, the following is taken as a hidden principle\(^1\):

\[
\text{“This coin is not biased”} \land \text{“theoretical assumptions”} \iff \text{“}P(\text{this coin turns up head}) = 0.5\text{“}.
\]

The theoretical assumptions are attained from the chosen experiment. One experiment may be performed by independently throwing \( n \) times the coin over a smooth surface. The observable random variable is the number of times the coin turned up head. In this simplified version, no

\[^1\text{There are two principles used in standard logic and in statistics to allow inferences, namely: modus ponens and modus tollens. Let} \, T \, \text{and} \, H \, \text{two logical statements (e.g.,} \, T \, \text{is the scientific statement in conjunction with statistical assumptions and} \, H \, \text{is the statistical statement). The modus ponens states that: “if} \, T \to H \, \text{is true and} \, T \, \text{is true, then} \, H \, \text{is true”, it is represented by} \, (T \to H) \land T \to H \). \, \text{The modus tollens states that: “if} \, T \to H \, \text{is true and} \, H \, \text{is false, then} \, T \, \text{is false”, it is represented by} \, (T \to H) \land \neg H) \to \neg T \). \text{Here, a true statement is an axiom or a consequence derived from an axiom by using strictly the rules of a consistent formal system. In a practical perspective, a true statement can be interpreted as a statement that corresponds with the observed facts (Popper, 1989), the application of modus ponens and modus tollens allow us to make inferences.}
latent variables are considered. Assuming that the coin cannot land on edge in the experiment, one statistical model that can represent this experiment is the binomial model \((X, \mathcal{M})\), where \(\mathcal{M} = \{g_\theta : \theta \in (0, 1)\}\) with
\[
g_\theta(k) = \frac{n!}{k!(n-k)!} \theta^k (1 - \theta)^{n-k}, \quad \text{for } k = 1, \ldots, n,
\]
where \(n!\) is the usual factorial notation, \(\theta\) is the probability of the studied coin turns up head and \(g_\theta(k)\) is the probability of the coin turns up head exactly \(k\) times in the performed experiment.

The scientific statement and its statistical counterpart are related by
\[
\text{"This coin is not biased" \land \text{"Theoretical assumptions" } \iff \theta = 0.5}.
\]
The null hypothesis is then represented by \(H_0 : \theta = 0.5\), that is \(H_0\) is a statement about probabilities: “if the coin is not biased, then [by the above principle and model assumptions] the probability of the coin turns up head is 0.5”.

Under this null hypothesis, the statistical model reduces to \((X, \mathcal{M}_0)\), where \(\mathcal{M}_0 = \{g_{0.5}\}\). In general, the alternative hypothesis is defined to be \(H_1 : \theta \neq 0.5\) and under this alternative hypothesis the statistical model is \((X, \mathcal{M}_1)\), where \(\mathcal{M}_1 = \{g_\theta : \theta \neq 0.5\}\). Notice that the union of both restricted families under \(H_0\) and \(H_1\) must be the original family, that is, \(\mathcal{M}_0 \cup \mathcal{M}_1 = \mathcal{M}\).

In the binomial model, it is implicitly assumed in the “Theoretical assumptions” that “\(P(\text{this coin turns up head})\)” does not change over all throws. Of course, this assumption is oversimplified for actual processes, since in each throwing the coin is submitted to impacts causing microscopic cracks, warps and, consequently, modifications in “\(P(\text{this coin turns up head})\)” over time. Other statistical models can be implemented relaxing some the imposed suppositions: 1) latent random variables can be incorporated to model dependence among the throwing and 2) covariates may be inserted to model variations in \(\theta\). That is, by changing some “Theoretical assumptions”, many statistical models could be used to model the outcomes of the very same experiment.

The concept of coin biasness can be further elaborated. One may prefer to relate the statement “this coin is not biased” with the structural topology of the coin, e.g., types of symmetries around the mass center of the coin, etc. Under this latter definition, it is possible to defined degrees of biasness based on a measure of symmetry and another completely different statistical model will emerge. This simple example can illustrate the complexity of statistical models and the problem of translating a simple scientific hypothesis into a statistical language. The presented example is applied in problems whose outcomes are either “success” or “failure”; for instance, the random variable \(X\) may be defined to be the number of allergic patients, out of \(n\), who react positively to a specific treatment.

### 3.1 Logical relations between the null and alternative hypotheses

In general, a full statistical model is initially specified \((X, \gamma, \mathcal{M})\). After establishing the null and alternative hypothesis \(H_0\) and \(H_1\), reduced statistical models emerge \((X, \gamma, \mathcal{M}_0)\) and \((X, \gamma, \mathcal{M}_1)\) under these hypotheses, respectively, where \(\mathcal{M}_0 \cup \mathcal{M}_1 = \mathcal{M}\). The null hypothesis states “at least one marginal probability distribution listed in \(\mathcal{M}_0\) generates the observable random variable”.

---

2See footnote # 1
It is simple to see that the alternative hypothesis $H_1$ is not the negation of $H_0$. In addition, the negation of the null hypothesis cannot be written in statistical terms, since $\neg H_0$ includes all possible mechanisms, not necessarily probabilistic ones, that could generate the observable variables $X$. The negation of $H_0$ is

$$\neg H_0 : \text{"It is not the case that \textquoteleft at least one marginal probability distribution listed in } \mathcal{M}_0 \text{ generates the observable random variable } X\textquoteright."$$

Therefore, by the usual logical relations, $H_1$ does imply $\neg H_0$, but $\neg H_0$ does not imply $H_1$. Therefore, the practitioner should be aware that a decision between $H_0$ and $H_1$ is very limited, since there are much more possibilities beyond the disjunction $H_0 \lor H_1$. These logical relations lie in the core of many controversies about null hypothesis statistical testing. For instance, Bayesian procedures typically define a prior probability such that $\pi(H_0 \lor H_1) = 1$. The problem with this latter procedure is that the universe of possibilities is assumed to be closed, giving the impression that the alternative hypothesis is the negation of the null hypothesis, since by the probability properties the following is a consequence: $\pi(H_0) = 1 - \pi(H_1)$.

The statistical hypotheses $H_0$ and $H_1$ are not necessarily exhaustive, because, as said previously, the family $\mathcal{M}$ might not contain the true generator mechanism. Even after making post-data analyses to verify if the model assumptions are adequate (residual analyses, simulated envelopes and so on), it is not possible to guarantee that the disjunction $H_0 \lor H_1$ is true (which is equivalent to say that $\mathcal{M}$ contains the true probability (density) function). Let us assume that $H_0 \lor H_1$ is necessarily true, then following inference rules are valid\(^3\):

- Empirical evidence to reject $H_0$ is empirical evidence to accept $H_1$: $\neg H_0 \Rightarrow H_1$.
- Empirical evidence to reject $H_1$ is empirical evidence to accept $H_0$: $\neg H_1 \Rightarrow H_0$.

However, if the disjunction $H_0 \lor H_1$ is not necessarily true, the preceding inference rules are not valid anymore, rather we have the following

- Empirical evidence to reject $H_0$ is not necessarily empirical evidence to accept $H_1$: $\neg H_0 \not\Rightarrow H_1$.
- Empirical evidence to reject $H_1$ is not necessarily empirical evidence to accept $H_0$: $\neg H_1 \not\Rightarrow H_0$.

That is the main difference between uncertain inference and decision theory as professor Sir Ronald Fisher argued in some of his papers (Fisher, 1935, 1955). On the one hand, if the disjunction $H_0 \lor H_1$ is not necessarily true, we have uncertain inference and more difficulties arise, for the universe of possibilities are not closed. Under this latter context, the practitioner must not use the inferential rules “$\neg H_1 \Rightarrow H_0$” and “$\neg H_0 \Rightarrow H_1$”. On the other hand, if the disjunction $H_0 \lor H_1$ is (assumed to be) necessarily true, we have decision theory and the space of decisions becomes well defined, for the inferential rules “$\neg H_1 \Rightarrow H_0$” and “$\neg H_0 \Rightarrow H_1$” are valid. It is important to note that the classical statistical model is sufficient general to allow these two situations discussed above:

1. **The Fisherian procedure** considers that $H_0 \lor H_1$ is not necessarily true. P-values were initially defined to hold this situation, they were designed to detect discrepancies between the null hypothesis and the observed data. It is not required even to define an alternative

\(^3\)It is a simple logical consequence of exhaustive and mutually exclusive hypotheses.
hypothesis; in this context, as aforementioned, some inference rules should not be employed. A very small p-value indicates a large discordance between the postulated null hypothesis and the observed data, however, a not very small p-value does not indicate evidence in favor of the null hypothesis.

2. The Neyman-Personian procedure considers that $H_0 \lor H_1$ is true. That is the case for the statistical tests developed by Neyman and Pearson. They developed the most powerful test for a fixed significance level (the probability of rejecting the null when it is false). A rejection region is build based on this procedure and a decision is taken by verifying if the observed sample lies or not in the rejection region. The Bayesian procedure is more aligned with the Neyman-Personian procedure than with the Fisherian, for at least some logical principle are shared between them. Naturally, regarding $H_0 \lor H_1$ as true might be only an artificial assumption to resolve a statistical problem, the statistician may not consider this as True in an ontological sense.

The above two perspectives lead to different types of statistical inferences, moreover one cannot be used to invalided the other, since they use different logical principles. Many papers in the scientific literature confound these two intrinsically different perspectives (see Hubbard et al., 2003, and the references therein).

Recently, Trafimow (2003) explicitly assumed that $H_0 \lor H_1$ is true, defined p-values by conditional probabilities and employed the rules of conditional probabilities to show that p-values are internally flawed; he wrote “the Bayesian analyses presented earlier not only suggest possible problems with NHSTP but also demonstrate when these potential problems become actual problems and when they do not”. Trafimow (2003) deliberately applied the Bayesian reasoning to analyze the p-values’ behavior and to conclude that they are flawed. In a recent Editorial note published by “Basic and Applied Social Psychology” (BASP), Trafimow and Marks (2015) communicated that the null hypothesis significance testing procedure (NHSTP) was banned from BASP. The Editorial note said that

“prior to publication, authors will have to remove all vestiges of the NHSTP (p-values, t-values, F-values, statements about significant differences or lack thereof, and so on).”

(Trafimow and Marks, 2015)

The attempts of writing classical statistics with Bayesian notation is a strong source of misinterpretations and controversies. One of the reasons is that some conditional statements are not probabilistic statements. The p-value is formally defined in the next section; as the reader shall see it has nothing to do with the formal definition of conditional probabilities and it is not connected with any Bayesian interpretations.

4 Definition of P-values

A p-value is built with the purpose of capturing a disagreement between the observed data and the postulated null hypothesis. In this context, it is the general advice to define a positive real statistic $T$ (it is a function of the random sample $X$) such that: the larger is its observed value $t$, the stronger is the disagreement between the observed data and the null hypothesis $H_0$ (Cox, 1977; Mayo and Cox, 2006; Patriota, 2013). The set $C_{H_0}(t) = \{x : T(x) \geq t\}$ describes all sample values
which have stronger disagreements with the postulated null hypothesis $H_0$ than the observed one $t$. This set has three important elements, namely: the null hypothesis of interest $H_0$, the random statistic $T$ and the observed statistic $t$. It is needless to say that $T$ strongly depends on $H_0$.

If $C_{H_0}(t)$ is small compared to the total set $C_{H_0}(0)$, then the observed experiment provides strong evidence against $H_0$; this happens when the observed $t$ is large enough to lie in the extreme right tail of the statistics $T$’s distribution. One way to measure the size of $C_{H_0}(t)$ is through probabilities. As the null hypothesis states probability distributions that represent the scientific statement of interest, the p-value is computed for the case with the highest probability in $H_0$. Let us consider the model without latent variables $(X, M)$, where $M_0 = \{g_0 : \theta \in \Theta_0\}$ is the set of probability (density) functions restricted under the specifications of $H_0$. Let $P_0$ be the probability measure associated with $g_0$, that is, if $g_0$ is a probability function, then $P_0(A) = \sum_{x \in A} g_0(x)$ and if $g_0$ is a probability density function then $P_0(A) = \int_A g_0(x) dx$. The p-value is formally defined by

$$p(H_0, t) = \sup_{\theta \in \Theta_0} P_0(C_{H_0}(t)). \quad (4.2)$$

Therefore, as $p(H_0, t)$ is (greater than or equal to) the case with the highest probability in $H_0$, the smaller the value of $p(H_0, t)$, the larger is the evidence against $H_0$. The formula (4.2) explicitly says that the classical p-value is not a conditional measure in the probabilistic sense, it is instead a conditional measure in the possibilistic sense. The reader should noticed that the usual representation p-value $= P(T \geq t | H_0)$ is inadequate, since (a) the probability $P$ is meaningless in the context of classical statistical models and (b) the conditional probability is being misused, since its formal definition is being ignored. The conditional probability is defined by $P(A|B) = \frac{P(A \cap B)}{P(B)}$, where $P(B) > 0$ and $A$ and $B$ are events of the same type (they must be listed in the same sigma-field). As for random variables, the conditional probability is defined analogously for the probability (density) function $g_0$. In classical statistics, it is easy to see that the events $\{x : T(x) \geq t\}$ and $H_0$ are not of the same type, for they are not listed in the same sigma-field; otherwise it is a Bayesian analysis. In classical statistics, there is not a probability distribution defined over the subsets of $\mathcal{M} = \{g_0 : \theta \in \Theta\}$ and as $\mathcal{M}$ cannot (even ideally) list all possible measures, a probability measure over the subsets of $\mathcal{M}$ will be conceptually ill-defined.

**Technical remark:** for each observed statistic $t$, the quantity $P_0(C_{H_0}(t))$ is fixed while $P_0(C_{H_0}(T))$ is random for each $\theta \in \Theta$. If, for each fixed $t$, $P_0(C_{H_0}(t)) = P_0(C_{H_0}(t))$ for all $\theta_1, \theta_2 \in \Theta_0$, then the statistic $T$ will be (informally) said to be ancillary to $\Theta_0$, and then the “sup” operation in (4.2) vanishes. This happens in many problems under normal distributions when the interest is centered in testing population means and/or variances. In this context, if $T$ is a continuous random variable and it is ancillary to $\Theta_0$, the distribution of $p(H_0, T)$ is uniform between 0 and 1. This allows the practitioner to interpret a p-value in terms of ideal replications of the performed experiment:

“If the performed experiment were repeated $N$ times, then it is estimated that $p(H_0, t) \times N$ of those experiments would produce p-values smaller than the observed one.”

This interpretation of repeating sampling from the same population is criticized by Fisher (1955). Fisher systematically argued that p-values should be interpreted as measures of discrepancy between the observed data and the postulated null hypothesis.
5 Problems of p-values and an alternative measure of evidence

The p-value is a coherent measure to verify a discrepancy of a specific null hypothesis with the observed data. Nevertheless, there is a serious limitation in the use of p-values in nested hypotheses. Consider that the p-value’s computation under $H_0^{(1)}$ is extremely complicated. Let $H_0^{(2)}$ be an auxiliary hypothesis such that $H_0^{(1)} \Rightarrow H_0^{(2)}$, i.e., if $H_0^{(1)}$ is true, then $H_0^{(2)}$ is true, because all statements in $H_0$ are also statements in $H_0^{(2)}$. By the logical reasoning: if $H_0^{(2)}$ is false, then $H_0^{(1)}$ must also be false, since if all statements in $H_0^{(2)}$ are false then, in particular, all statements in $H_0^{(1)}$ must be false. The practitioner, led by this logical reasoning, would compute the p-value under $H_0^{(2)}$ and conclude that if there is evidence to reject $H_0^{(2)}$, i.e., the p-value computed under $H_0^{(2)}$ is small, then there must be evidence to reject $H_0^{(1)}$, for the p-value computed under $H_0$ must be even smaller. However, p-values do not met this latter logical reasoning. That is, it is not guaranteed that $p(H_0^{(1)}, t) \leq p(H_0^{(2)}, t)$ even for the very same data; for more details, see, for instance, Schervish (1996) and Patriota (2013). This happens because the test statistic $T$ is built for a specific null hypothesis, therefore, the respective p-value is valid only for this specific null hypothesis. In previous work, Patriota (2013) proposed an alternative classical measure of evidence that meets the above logical reasoning; it is called s-value and will be presented in what follows.

The general purpose of the s-value is almost the same of that of the p-value: to verify a discrepancy of null hypotheses with the observed data, but maintaining all logical consequence among null hypotheses. In order to define s-values, let us consider the simplest statistical model without latent variables $(X, \mathcal{M})$, where $\mathcal{M} = \{g_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$ and let $P_\theta$ be the probability measure associated with the probability (density) function $g_\theta$. The likelihood-ratio function is

$$\lambda(\theta; x) = \frac{g_\theta(x)}{\sup_{\theta \in \Theta} g_\theta(x)},$$

provided that $\sup_{\theta \in \Theta} g_\theta(x) > 0$. Notice that, $0 \leq \lambda(\theta; x) \leq 1$ for all $\theta \in \Theta$. The likelihood-ratio confidence region with significance level $\alpha$ is defined by

$$\Lambda_\alpha(x) = \{\theta \in \Theta : \lambda(\theta; x) \geq c_\alpha(\theta)\},$$

where

$$P_\theta(\lambda(\theta; X) \geq c_\alpha(\theta)) \geq 1 - \alpha, \quad \inf_{\theta \in \Theta} P_\theta(\lambda(\theta; X) \geq c_\alpha(\theta)) = 1 - \alpha$$

and $0 \leq c_\alpha(\theta) \leq 1$. The following equivalent notation may be used

$$P_\theta(\lambda(\theta; X) \geq c_\alpha(\theta)) = P_\theta(\Lambda_\alpha(X) \ni \theta).$$

The quantity $P_\theta(\Lambda_\alpha(X) \ni \theta)$ is the probability of the confidence region $\Lambda_\alpha(X)$ contains $\theta$. It is the formal definition of a general confidence region for the parameter $\theta$ (Schervish, 1995). The confidence region does not have any ontological meaning of True lurked in its interpretation.

It is noteworthy saying that for some statistical models (normal distribution in general), the following occurs:

$$P_\theta(\lambda(\theta; X) \geq c_\alpha(\theta)) = 1 - \alpha \quad \text{for all } \theta \in \Theta,$$

\footnote{This second true was attained by using strictly the rules of classical logic.}

The general purpose of the s-value is almost the same of that of the p-value: to verify a discrepancy of null hypotheses with the observed data, but maintaining all logical consequence among null hypotheses. In previous work, Patriota (2013) proposed an alternative classical measure of evidence that meets the above logical reasoning; it is called s-value and will be presented in what follows.
in this case, the confidence region is said to be exact. For exact confidence regions, the value $c_\alpha(\theta)$ is the $(1 - \alpha) \times 100\%$ quantile of the random variable $\lambda(\theta, X)$. Observe that $\Lambda_\alpha(x)$ contains all $\theta$’s that generate likelihood values greater than (or equal to) $c_\alpha(\theta)$ times the largest likelihood value, namely, $\sup_{\theta \in \Theta} g_\theta(x)$. This set is very intuitive, for it contains the optimal values for $\theta \in \Theta$ according to the likelihood function. The definition of s-values follows.

**Definition 5.1.** Let $\Theta_0$ be a non-empty parameter subset related with $H_0$ and let $\Lambda_\alpha(x)$ be the likelihood-ratio confidence region with significance level $\alpha$. Then, the s-value is defined by

$$s(H_0; x) \equiv s(\Theta_0; x) \equiv \sup_{\alpha \in [0, 1]} \{ \alpha \in [0, 1] : \Lambda_\alpha(x) \cap \Theta_0 \neq \emptyset \}.$$ 

If $\Theta_0 = \emptyset$, define

$$s(\emptyset; x) \equiv 0.$$

This general definition is valid for very general hypotheses. Let $\Theta_{01}$ and $\Theta_{02}$ be two parameter subsets related with the hypotheses $H_0^{(1)}$ and $H_0^{(2)}$, respectively. In this context, if $H_0^{(1)} \Rightarrow H_0^{(2)}$, then $\Theta_{01} \subseteq \Theta_{02}$; Patriota (2013) showed that the following always occurs $s(\Theta_{01}; x) \leq s(\Theta_{02}; x)$. A possible interpretation for the s-value, under regular conditions and assuming that $\Theta_0$ is non-empty, reads

“$s(\Theta_0, x)$ is greater than (or equal to) the maximum significance level $\alpha_M$ such that $\Lambda_{\alpha_M}(x)$ and $\Theta_0$ have at least one element in common”.

The smaller $s(\Theta_0, x)$ is, the more distant $\Theta_0$ is from the maximum likelihood estimate of $\theta$ and, consequently, the more unlikely $H_0$ is according to the likelihood-ratio confidence region. Observe that, if $H_0 : \theta = \theta_0$, where $\theta_0$ is a given vector (or number if $\Theta \subseteq \mathbb{R}$), then $\Theta_0 = \{\theta_0\}$ and the s-value reduces to

$$s(\{\theta_0\}; x) = \max\{\alpha \in [0, 1] : \theta_0 \in \Lambda_\alpha(x)\},$$

and its interpretation reads

“$s(\{\theta_0\}, x)$ is equal to the maximum significance level $\alpha_M$ such that $\Lambda_{\alpha_M}(x)$ contains $\theta_0$”.

Therefore, the farer away $\theta_0$ is from the center of $\Lambda_\alpha(x)$, which in regular conditions is the maximum likelihood estimative, the more is the observed evidence against $H_0$.

5.1 Types of decisions

In this section, some types of decisions are studied. Let $\hat{\theta}$ be the maximum likelihood estimative of $\theta$, then, under regular conditions, we have that $\hat{\theta} \in \Theta$.

**First case:** no alternative hypothesis is defined, then the general advice of this paper is to use the s-value as a thermometer of discrepancy between null hypotheses and the observed data. The smaller is $s(\Theta_0, x)$, the stronger is the evidence against $H_0$. Patriota (2013) showed that if $\hat{\theta} \in \Theta_0$, then $s(\Theta_0, x) = 1$ and the observed data produce no evidence against $H_0$, which does not mean evidence in favor of $H_0$. In a working paper, we are showing that s-values are always greater than p-values (based on the likelihood-ratio statistics) for some specific models. This indicates that if a s-value is small, then the respective p-value must be even smaller. Therefore, one could just compute the s-value to verify discrepancies of the null hypothesis with the observed data. The use
of s-values is justified for general hypotheses, because p-values are much more difficult to compute than s-values and furthermore do not satisfy the logical consequence.

**Second case:** an alternative hypothesis $H_1$ is defined and let $\Theta_1$ be its related parameter space. Patriota (2013) showed that on the one hand if $\hat{\theta} \in \Theta_0$, then $s(\Theta_0, x) = 1$; on the other hand if $\hat{\theta} \in \Theta_1$, then $s(\Theta_1, x) = 1$. If the practitioner wants to decide between $H_0$ or $H_1$, then there are three possibilities

- If $s(\Theta_1, x) = 1$ and $s(\Theta_0, x) = a$, then reject $H_0$ and accept $H_1$ whenever $a$ is sufficiently small.
- If $s(\Theta_1, x) = b$ and $s(\Theta_0, x) = 1$, then accept $H_0$ and reject $H_1$ whenever $b$ is sufficiently small.
- If $s(\Theta_1, x) = s(\Theta_0, x) = 1$, then do not reject neither accept $H_0$. More data are required.

The threshold values $a^*$ and $b^*$ for $a$ and $b$, respectively, are being studied. They shall depend on the sample size, effect sizes, error of type I and II, power of the test, severity (Mayo and Spanos, 2006; Mayo and Cox, 2006), and/or other factors. Notice also that more than one alternative hypotheses $H_1, \ldots, H_k$ can be defined. It is possible to use the s-value in the latter context, but it is beyond the scope of this paper.

5.2 Steps to compute the s-value

The steps to compute the s-value are:

1. Define the statistical model $(X, M)$. Remember that $X$ represents the observable sample and contains $n$ random variables, namely $X = (X_1, \ldots, X_n)$;
2. Define the null hypothesis $H_0$ and its related set $\Theta_0$;
3. If required, define the alternative hypothesis $H_1$ and its related set $\Theta_1$;
4. Compute the likelihood-ratio statistics $\lambda(\theta; x)$;
5. Compute $c_0(\theta)$;
6. Compute $\Lambda_0(x)$;
7. Compute $s(\Theta_0, x)$.
8. If required, compute $s(\Theta_1, x)$.

The step 5 is somewhat difficult to execute for some complex statistical models, since for those models the distribution of $\lambda(\theta, X)$ is not trivial and may depend on $\theta$. In those cases, under regular conditions, the practitioner may apply the limiting distribution of $-2 \log (\lambda(\theta, X))$, which is a chi-squared distribution with $p$ degrees of freedom, where $\dim(\Theta) = p$. Then, step 4 reduces to

$$c_0(\theta) = \exp \left( -\frac{1}{2} \chi^2_{p,1-\alpha} \right), \quad \text{for all } \theta \in \Theta,$$

where $\chi^2_{p,1-\alpha}$ is the $(1-\alpha) \times 100\%$ quantile of a chi-squared distribution with $p$ degrees of freedom. This approximation reduces the complexity, since $c_0(\theta)$ does not depend on $\theta$. Under this
asymptotic approximation, the “asymptotic” s-value then reduces simply to

\[ s_a(\Theta_0, x) = 1 - \inf_{\theta \in \Theta_0} F_{\chi^2_p} \left( -2 \log(\lambda(\theta, x)) \right) = 1 - F_{\chi^2_p} \left( -2 \log \left( \sup_{\theta \in \Theta_0} \lambda(\theta, x) \right) \right), \]

where \( F_{\chi^2_p} \) is the accumulative distribution of a chi-squared distribution with \( p \) degrees of freedom and \( \log \) is the natural logarithm function. If \( \Theta_0 = \{\theta_0\} \), the asymptotic p-value (i.e., the asymptotic approximation for the p-value) based on the likelihood ratio statistics coincides with the above asymptotic s-value. Nevertheless, if \( \dim(\Theta_0) > 0 \) (the Lebesgue dimension), the asymptotic p-value and asymptotic s-value differ from each other. In the asymptotic p-value, the degree of freedom of the chi-squared distribution varies with the dimension of \( \Theta_0 \); more precisely, the asymptotic p-value for the likelihood ratio statistics is

\[ p_a(\Theta_0; x) = 1 - F_{\chi^2_p} \left( -2 \log \left( \sup_{\theta \in \Theta_0} \lambda(\theta, x) \right) \right), \]

where \( q = \dim(\Theta) - \dim(\Theta_0) \), where \( \dim(\Theta) = p \). That is, the cumulative distribution function \( F_{\chi^2_q} \) varies with the chosen null hypothesis, whereas for the s-value \( F_{\chi^2_p} \) does not vary with the chosen null hypothesis. Patriota (2013, 2014) showed that the asymptotic s-value and p-values are connected through the following relation

\[ s_a(\Theta_0, x) = 1 - F_{\chi^2_p} \left( F_{\chi^2_q}^{-1}(1 - p_a(\Theta_0, x)) \right). \]

That is, from a p-value (based on likelihood ratio statistics) we can compute the s-value via the above formulae. If \( p = q \), then \( s(\Theta_0, x) = p(\Theta_0, x) \).

6 Numerical examples

In this section, the s-value is applied for univariate and bivariate normal distributions.

Example 6.1. (Normal distribution, variance known) Let \( X = (X_1, \ldots, X_n) \) be a sample from a normal distribution with population mean \( \theta \) and variance 1. Let \( H_0 : \theta = \theta_0 \) be the null hypothesis of interest. The statistical model is \((X, \mathcal{M})\), where \( \mathcal{M} = \{g_\theta : \theta \in \mathbb{R}\} \) and

\[ g_\theta(x) = \frac{1}{(2\pi)^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta)^2 \right). \]

The likelihood ratio statistic is

\[ \lambda(\theta, x) = \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta)^2 + \frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) = \exp \left( -\frac{n}{2} (\bar{x} - \theta)^2 \right), \]

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) is the maximum likelihood estimate for \( \theta \). It is widely known that

\[ -2 \log(\lambda(\theta, X)) \xrightarrow{P_{\theta_0}} \chi^2_1, \]

where the symbol \( \xrightarrow{P_{\theta_0}} \chi^2_p \) means “follows a chi-squared distribution with \( p \) degrees of freedom, under the law \( P_{\theta_0} \).” Then,

\[ c_\alpha(\theta) = \exp \left( -\frac{1}{2} \lambda_{1,1-\alpha}^2 \right) \]
and
\[
\Lambda_\alpha(x) = \left\{ \theta \in \mathbb{R} : n(\bar{x} - \theta)^2 \leq \chi^2_{1,1-\alpha} \right\} = \left[ \bar{x} - \sqrt{\frac{1}{n} \chi^2_{1,1-\alpha}}, \bar{x} + \sqrt{\frac{1}{n} \chi^2_{1,1-\alpha}} \right].
\]

Let \( H_0 : \theta = \theta_0 \) be the null hypothesis of interest. The s-value is computed by finding the \( \alpha \)-value such that the border of the observed confidence interval \( \Lambda_\alpha(x) \) is \( \theta_0 \). The solution is
\[
s(\{\theta_0\}, x) = 1 - F_{\chi^2_1}(n(\bar{x} - \theta_0)^2).
\]

As aforementioned, for this simple null hypothesis, the s-value is precisely the p-value computed for likelihood ratio statistics. Table 1 depicts numerical s-values to illustrate the univariate normal distribution example for \( n = 10 \) and \( \sigma^2 = 1 \). The null hypothesis is \( H_0 : \theta = \theta_0 \), where \( \theta_0 = -1, 0, 1 \).

**Example 6.2.** (Bivariate Normal distribution, with known variances and covariances) Let \( X = (X_1, \ldots, X_n) \) be a sample from a bivariate normal distribution with population mean \( \theta = (\mu_1, \mu_2)^T \) and covariance-variance matrix \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \). The statistical model is \( (X, \mathcal{M}) \), where \( \mathcal{M} = \{g_\theta : \theta \in \mathbb{R}^2\} \) and
\[
g_\theta(x) = \frac{1}{(2\pi)^n} \exp \left( -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^\top (x_i - \theta) \right).
\]

The likelihood ratio statistic is
\[
\lambda(\theta, x) = \exp \left( -\frac{1}{2} (\bar{x} - \theta)^\top (\bar{x} - \theta) \right),
\]
where \( \bar{x} = (\bar{x}_1, \bar{x}_2)^\top \) is the maximum estimate for \( \theta \), where \( \bar{x}_1 \) and \( \bar{x}_2 \) are the sample averages of the bivariate sample. Observe that, here \( p = 2 \). It is also widely known that
\[
-2 \log(\lambda(\theta, X)) \sim \chi^2_2.
\]

Then,
\[
c_\alpha(\theta) = \exp \left( -\frac{1}{2} \chi^2_{2,1-\alpha} \right)
\]
and
\[
\Lambda_\alpha(x) = \left\{ \theta \in \mathbb{R} : n(\bar{x} - \theta)^\top (\bar{x} - \theta) \leq \chi^2_{2,1-\alpha} \right\}.
\]

**Null hypothesis 1:** Let \( H_0^{(1)} : \theta = \theta_0 \) be the null hypothesis of interest, where \( \theta_0 = (\mu_{10}, \mu_{20})^T \) is a given vector; then \( \Theta_{01} = \{\theta_0\} \). The s-value is computed by finding the \( \alpha \)-value such that the border of the observed confidence interval \( \Lambda_\alpha(x) \) is \( \theta_0 \). The solution is (which is also equal to the p-value computed for the likelihood ratio statistics)
\[
s(\{\theta_0\}, x) = 1 - F_{\chi^2_2}(n(\bar{x} - \theta_0)^\top (\bar{x} - \theta_0)).
\]

**Null hypothesis 2:** Let \( H_0^{(2)} : \mu_1 = \mu_2 \) be the null hypothesis of interest, then \( \Theta_{02} = \{\theta \in \mathbb{R}^2 : \mu_1 = \mu_2\} \). The s-value is computed by finding the maximum \( \alpha \)-value such that
\[
\Lambda_\alpha(x) \cap \Theta_{02} = \{ \theta \in \Theta_{02} : n(\bar{x} - \theta)^\top (\bar{x} - \theta) \leq \chi^2_{2,1-\alpha} \}
\]
has at least one element. The solution is (which is not equal to the p-value computed for the
likelihood ratio statistics)

\[ s(\Theta_{02}, x) = 1 - F_{\chi^2} \left( n \min_{\theta \in \Theta_{02}} (\bar{x} - \theta)^\top (\bar{x} - \theta) \right). \]

Notice that

\[ \min_{\theta \in \Theta_{02}} (\bar{x} - \theta)^\top (\bar{x} - \theta) = \min_{\mu \in \mathbb{R}} [(\bar{x}_1 - \mu)^2 + (\bar{x}_2 - \mu)^2] = \frac{n}{2} (\bar{x}_1 - \bar{x}_2)^2. \]

Then,

\[ s(\Theta_{02}, x) = 1 - F_{\chi^2} \left( \frac{n}{2} (\bar{x}_1 - \bar{x}_2)^2 \right). \]

Recall that the p-value computed for the likelihood ratio statistics is

\[ p(\Theta_{02}; x) = 1 - F_{\chi^2} \left( \frac{n}{2} (\bar{x}_1 - \bar{x}_2)^2 \right). \]

Table 2 presents numerical s-values to illustrate the bivariate normal distribution example for \( n = 10 \) and covariance-variance matrix \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \). The null hypotheses considered are \( H_{01} : \mu_1 = \mu_2 = 0 \) and \( H_{02} : \mu_1 = \mu_2 \) for which it is expected to find more evidence against \( H_{01} \) than \( H_{02} \). The s-values were defined to hold this expected behavior, but p-values do not. We purposely choose values for \( \bar{x}_1 \) and \( \bar{x}_2 \) such that p-values are problematic. The figures of Table 2 show that all p-values fail to hold the logical condition for all sample, except for \( \bar{x}_1 = \bar{x}_2 = 0 \).

7 Conclusion

This paper discussed some conceptual and technical problems in the null hypothesis statistical testing. The assumptions of a statistical model are crucial to test adequately scientific hypotheses, since null hypotheses depend on the statistical model (they are probabilistic transcriptions of the scientific statement of interest). Types of decisions in null hypothesis statistical testing strongly depend on important assumptions that are not always made explicit. On the one hand, if the practitioner considers the null and alternative hypotheses are mutually exclusive and exhaustive, then procedures to accept-reject the null are justifiable (e.g., Neyman-Pearsonian and Bayesian procedures). On the other hand, if the practitioner considers that the null and alternative hypotheses are mutually exclusive but not exhaustive, then only procedures to reject the null are justifiable (e.g., Fisherian procedures). A statistical procedure developed under one assumption will certainly fail to be admissible under the other, therefore an extra caution must be taken when comparing different statistical procedures (classical versus Bayesian). By construction, p-values do not respect the following logical reasoning: if \( H_{01} \Rightarrow H_{02} \), then p-value(\( H_{02} \)) \( \geq \) p-value(\( H_{01} \)). That is, the practitioner must not use the p-value to extrapolate the inference made for \( H_{02} \) to \( H_{01} \). This is not a defect in the classic statistical reasoning, because s-values do respect this logical and can be employed in the place of p-values. Asymptotic versions of s-values are extremely simple to compute, simpler even than p-values. S-values can be used as a complementary measure of evidence and, as any other statistical measure, some care is needed when use it to make inferences; rules of thumbs must be avoided, the inferential conclusions must be always complemented with other statistical tools.
8 Acknowledgements

The author gratefully acknowledges grant from FAPESP (2014/25595-0, Brazil).

References


Table 1: S-values for testing $H_0 : \theta = \theta_0$, where $\theta_0 = 0, 1$ for some observed values of $\bar{x}$ when $n = 10$.

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>$\theta_0 = 0$</th>
<th>$\theta_0 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td>0.1</td>
<td>0.7518</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5271</td>
<td>0.0114</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3428</td>
<td>0.0269</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2059</td>
<td>0.0578</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1138</td>
<td>0.1138</td>
</tr>
<tr>
<td>0.6</td>
<td>0.0578</td>
<td>0.2059</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0269</td>
<td>0.3428</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0114</td>
<td>0.5271</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0044</td>
<td>0.7518</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0016</td>
<td>1.0000</td>
</tr>
<tr>
<td>1.1</td>
<td>0.0005</td>
<td>0.7518</td>
</tr>
<tr>
<td>1.2</td>
<td>0.0001</td>
<td>0.5271</td>
</tr>
<tr>
<td>1.3</td>
<td>&lt;0.0001</td>
<td>0.3428</td>
</tr>
<tr>
<td>1.4</td>
<td>&lt;0.0001</td>
<td>0.2059</td>
</tr>
<tr>
<td>1.5</td>
<td>&lt;0.0001</td>
<td>0.1138</td>
</tr>
<tr>
<td>1.6</td>
<td>&lt;0.0001</td>
<td>0.0578</td>
</tr>
<tr>
<td>1.7</td>
<td>&lt;0.0001</td>
<td>0.0269</td>
</tr>
<tr>
<td>1.8</td>
<td>&lt;0.0001</td>
<td>0.0114</td>
</tr>
<tr>
<td>1.9</td>
<td>&lt;0.0001</td>
<td>0.0044</td>
</tr>
<tr>
<td>2.0</td>
<td>&lt;0.0001</td>
<td>0.0016</td>
</tr>
</tbody>
</table>
Table 2: S-values and p-values for testing \( H_{01} : \mu_1 = \mu_2 = 0 \) (the s-values and p-values are identical) and \( H_{02} : \mu_1 = \mu_2 \) (the s-values and p-values differ) for some observed values of \((\bar{x}_1, \bar{x}_2)\) that generate problematic p-values (showing that p-values do not respect the logical consequence). The sample size is \( n = 10 \).

<table>
<thead>
<tr>
<th>((\bar{x}_1, \bar{x}_2))</th>
<th>(\bar{x}_1 - \bar{x}_2)</th>
<th>(H_{01} : \mu_1 = \mu_2 = 0)</th>
<th></th>
<th>(H_{02} : \mu_1 = \mu_2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s-value</td>
<td>p-value</td>
<td></td>
<td>s-value</td>
<td>p-value</td>
</tr>
<tr>
<td>(0.00, 0.00)</td>
<td>0.0</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(0.05,-0.05)</td>
<td>0.1</td>
<td>0.9753</td>
<td>0.9753</td>
<td>0.9753</td>
<td>0.8231</td>
</tr>
<tr>
<td>(0.09,-0.11)</td>
<td>0.2</td>
<td>0.9039</td>
<td>0.9039</td>
<td>0.9048</td>
<td>0.6547</td>
</tr>
<tr>
<td>(0.14,-0.16)</td>
<td>0.3</td>
<td>0.7977</td>
<td>0.7977</td>
<td>0.7985</td>
<td>0.5023</td>
</tr>
<tr>
<td>(0.19,-0.21)</td>
<td>0.4</td>
<td>0.6697</td>
<td>0.6697</td>
<td>0.6703</td>
<td>0.3711</td>
</tr>
<tr>
<td>(0.23,-0.27)</td>
<td>0.5</td>
<td>0.5331</td>
<td>0.5331</td>
<td>0.5353</td>
<td>0.2636</td>
</tr>
<tr>
<td>(0.28,-0.32)</td>
<td>0.6</td>
<td>0.4049</td>
<td>0.4049</td>
<td>0.4066</td>
<td>0.1797</td>
</tr>
<tr>
<td>(0.33,-0.37)</td>
<td>0.7</td>
<td>0.2926</td>
<td>0.2926</td>
<td>0.2938</td>
<td>0.1175</td>
</tr>
<tr>
<td>(0.37,-0.43)</td>
<td>0.8</td>
<td>0.2001</td>
<td>0.2001</td>
<td>0.2019</td>
<td>0.0736</td>
</tr>
<tr>
<td>(0.42,-0.48)</td>
<td>0.9</td>
<td>0.1308</td>
<td>0.1308</td>
<td>0.1320</td>
<td>0.0442</td>
</tr>
<tr>
<td>(0.47,-0.53)</td>
<td>1.0</td>
<td>0.0813</td>
<td>0.0813</td>
<td>0.0821</td>
<td>0.0253</td>
</tr>
<tr>
<td>(0.51,-0.59)</td>
<td>1.1</td>
<td>0.0478</td>
<td>0.0478</td>
<td>0.0486</td>
<td>0.0139</td>
</tr>
<tr>
<td>(0.56,-0.64)</td>
<td>1.2</td>
<td>0.0269</td>
<td>0.0269</td>
<td>0.0273</td>
<td>0.0073</td>
</tr>
<tr>
<td>(0.61,-0.69)</td>
<td>1.3</td>
<td>0.0144</td>
<td>0.0144</td>
<td>0.0146</td>
<td>0.0037</td>
</tr>
<tr>
<td>(0.65,-0.75)</td>
<td>1.4</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.0074</td>
<td>0.0017</td>
</tr>
<tr>
<td>(0.70,-0.80)</td>
<td>1.5</td>
<td>0.0035</td>
<td>0.0035</td>
<td>0.0036</td>
<td>0.0008</td>
</tr>
<tr>
<td>(0.75,-0.85)</td>
<td>1.6</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0017</td>
<td>0.0003</td>
</tr>
<tr>
<td>(0.79,-0.91)</td>
<td>1.7</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
<tr>
<td>(0.84,-0.96)</td>
<td>1.8</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
<tr>
<td>(0.89,-1.01)</td>
<td>1.9</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(0.93,-1.07)</td>
<td>2.0</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>