Journal of Social Sciences and Humanities
Vol. 5, No. 3, 2019, pp. 165-170
http://www.aiscience.org/journal/jssh
ISSN: 2381-7763 (Print); ISSN: 2381-7771 (Online)

"Statistics Gone on Holiday": Misinterpretations of Hypothesis Tests Propagated by Internet Resources

Christopher Thron[*], Nancy Hernandez

Department of Science and Mathematics, Texas A&M University-Central Texas, Killeen, Texas, USA

Abstract

"Type I error" is a basic concept in statistical hypothesis testing. However, the term is used in two subtly different senses in statistics texts and other statistical literature. Specifically, type I error can be construed either as a conditional event (i.e. presuming that the null hypothesis is true) or an unconditional event. We explain the distinctions between the different usages of type I error, and we conduct a logical analysis of popular statistics web sites to determine their usage of the terminology. Our analysis shows that ambiguous and inconsistent usage of this terminology leads to wrong interpretations of significance level in many web pages, leading in turn to faulty interpretations of the results of experiments. We discuss the reasons for this long-standing lack of consensus in the definition of type I error. The unconditional-event definition is more intuitive and agrees with the original formulation Neyman and Pearson in 1933, but professional statisticians favor the conditional-event definition. The fact that users of statistics come from widely different fields makes it difficult to arrive at a single agreed-upon definition. We conclude that even in a rigorous technical subject like statistics, ambiguous terminology can go unrecognized and can continue to produce errors in reasoning.

Keywords

Social Sciences, Hypothesis Testing, Type I Error, Significance Level, Conditional Probability, Internet, Terminology

Received: March 17, 2019 / Accepted: May 9, 2019 / Published online: May 31, 2019

1. Introduction

It is well known that there are several common errors that trip up students of statistics and even occur in statistics textbooks and published research [1-4]. Are these misunderstandings preventable? Or are they simply due to the fact that the subject is difficult to understand? In this paper, we show that at least one common conceptual error in statistics may be traced back to inconsistencies in the use of the term, 'type I error' which, as we shall show, is commonly assigned different meanings in different contexts by different authors. These differences though widespread are equally widely unacknowledged.

In statistical hypothesis testing, two closely related concepts are *significance level* (which is typically denoted by the letter α) and *type I error*. The following incorrect argument involving these two concepts is not uncommon in the popular statistics literature, and even occurs frequently in statistics textbooks [1]:

(a) In a statistical experiment, a type I error occurs whenever a null hypothesis (*H0*) is rejected even though it is true.

(b) The significance level α is defined as the probability of a type I error.

(c) For any statistical experiment, the probability that *H0* will be falsely rejected is α.

* Corresponding author
E-mail address: thron@tamuct.edu (C. Thron)

As an example of this erroneous argument, consider a survey which asks respondents to state their political party. Suppose that (for various reasons) members of party *A* falsely identify themselves as belonging to a different party. Suppose also that based on previous data, it may be determined that 10% of respondents of party *A* will falsely identify themselves: in other words, the probability of Type I error is 0.1. According to the argument (a)-(c) above, this should be mean that 10% of respondents are *A* presenting themselves as non-*A*.

This conclusion is false. Suppose that in actuality, only 30% of the population is party *A*. Of these, 10% misidentify themselves. It follows that the actual proportion of misidentified *A*'s in the survey is 10% of 30%, or only 3%.

By using the terminology of conditional probability, we may characterize the argument's error more precisely. The correct definition of significance level α is as a conditional probability:

$$\alpha = P\,(H0 \text{ is rejected given that } H0 \text{ is true for a given experiment}) \quad (1)$$

but in statement (c) above, the significance level has been identified with an unconditional probability:

$$P\,(H0 \text{ is rejected and } H0 \text{ is true for a given experiment}) \quad (2)$$

The error arises because two inconsistent notions of type I error have been used in the same argument. In statement (a), type I error is characterized as an unconditional event, namely the false rejection of *H0*. But in the definition of α in statement (b), type I error is a *conditional event,* which applies only to experiments in which *H0* is presumed to be true.

It may seem that the two definitions are virtually the same, but in fact they are utterly incompatible. In statement (a), type I error is one of four possible experimental outcomes, namely: *H0* is true and rejected; H*0* is true and not rejected; *H0* is false and rejected; *H0* is false and not rejected. On the other hand, in statement (b) type I error is one of *two* possible outcomes in a hypothetical experiment in which *H0* is true: either *H0* is not rejected, or type I error occurs.

Which is the correct definition of type I error? The issue boils down to question of whether type I error is a *conditional* or *unconditional* event. Our research (which we shall present in the following sections) shows that no consensus among users of these statistical concepts. For internet sources in particular, some sources say one thing, others say the other, some are ambiguous, and a fair number are logically inconsistent. It's no wonder that faulty arguments keep cropping up regularly—and will continue to do so until the inconsistency is resolved, or at least recognized.

2. Method

The internet serves as a barometer of current opinion, as well as an important source for many individuals seeking technical information. In order to determine the prevailing understandings of type I error, we conducted two simple searches on Google ("type I Error" and '"probability of committing a type I error"), and took the top 13 and 5 hits, respectively. Related web pages at these sites were also examined. Some of these pages were dedicated to education (e.g. Khan Academy); others were sites for general information (e.g. about.com); others were technical articles; others were college course material that was posted online.

We examined these sites for their usage of type I error terminology, and subjected these texts to rigorous logical analysis. A summary of our findings is given in Table 1 at the end of Section 3.

3. Result

3.1. Definitions of Type I Error in Internet Sources

As mentioned in the Introduction, the key difference between the two notions of type I error lies in whether a type I error is specified as an unconditional or conditional event. In the first case the truth of *H0* is considered as part of the outcome, while in the second it is taken as a given. If a definition uses the conjunctions "and", "even though", "when", or "while", it is impossible to tell which of these two possibilities is meant. In most of our Internet sources, we found definitions to be ambiguous. Some examples are:

"…This is the error of rejecting *H0* even though it is true" [5-7].

"When the null hypothesis is true and you reject it, you make a type I error" [8].

"Rejecting the null hypothesis when it is in fact true is called a Type I error" [9].

"A type I error, also known as an error of the first kind, occurs when the null hypothesis (*H0*) is true, but is rejected" [10].

Other sources identify type I as a "false positive", indicating that it is one of four possible outcomes (false positive, false negative, true positive, true negative). In this case, type I error is represented as an unconditional event. Some examples are as follows:

"A Type I error is often referred to as a 'false positive', and is the process of incorrectly rejecting the null hypothesis in favor of the alternative." [11].

"The first kind of error that is possible involves the rejection of a null hypothesis that is actually true. Type I errors are equivalent to false positives." [12-13].

"Type I error occurs when a significance test results in the rejection of a true null hypothesis." [14].

"A type I error is the incorrect rejection of a true null hypothesis (a "false positive")." [10].

"Type I error, also known as a "false positive": the error of rejecting a null hypothesis when it is actually true. In other words, this is the error of accepting an alternative hypothesis (the real hypothesis of interest) when the results can be attributed to chance. Plainly speaking, it occurs when we are observing a difference when in truth there is none (or more specifically - no statistically significant difference)." [15].

In reference [17], type I error was defined as a probability, so that type I error was conflated with α. This imprecise use of language only muddies the waters.

3.2. Uses and Illustrations of Type I Error in Internet Sources

Many sources endeavor to explain type I error by presenting a 2-by-2 table with columns labeled "Null hypothesis is true", and "Null hypothesis is false", and rows labeled "Null hypothesis is rejected" and "Null hypothesis is not rejected". The top left entry in the table is always labeled "type I error", and the bottom right entry "type II error". Unfortunately, such tables are entirely ambiguous, unless the column labels are clarified. The four table entries could be interpreted as portraying four possible experimental outcomes. Or, the table could be read column by column, where the entries in each column represent two possible outcomes under the condition represented by the column heading. The interpretation of the table thus depends on whether the reader identifies the column headings as givens for the experiment, or as descriptions of the outcome. Only one of our sources [16] provided a suitable explanation of the table.

Many sources also give examples of type I error to illustrate its meaning. In many cases, the examples indicate an unconditional conception of type I error in which the truth of *H0* is not a precondition for the event in question. For example, reference [9] gives the example of two drugs being compared for effectiveness in treating the same condition, and states: "a Type I error would be deciding that Drug 2 is more effective, when in fact it is no better than Drug 1". Clearly, this is a situation where it is possible that *H0* is false, that is Drug 2 is actually better than Drug 1. Similarly, reference [17] gives the example of an inspector deciding whether building is safe, and identifies the outcome "Reject 'building is not safe' when it is not safe" as a type I error.

3.3. Treatment of α Under the Two Definitions

The common definition of α as the 'probability of type I error' is only consistent with a conditional-event definition of type I error, which restricts type I error to experiments where the truth of *H0* is fixed as a precondition. This definition was found in 19 of the 25 sources examined.

On the other hand, if type I error is understood as a possible outcome in an experiment in which *H0* is not necessarily true, then this definition is incorrect as we explained in the Introduction. Instead, one should define α in this case as the probability of type I error given that *H0* is true. Of the 25 sources we examined, 4 defined type I error in this way ([10] gave both this definition and the conditional-event definition, indicating that the article had multiple authors). Another possibility is to define α as an upper bound or maximum value for the probability of type I error. This follows from the usual equation for conditional probabilities:

$$\begin{aligned} \alpha &= P(H0 | H0 \text{ is rejected}) \\ &= P(H0 \text{ is true and rejected}) / P(H0 \text{ is true}) \quad (3) \\ &\geq P(H0 \text{ is true and rejected}). \end{aligned}$$

Among the internet sources that we examined, 9 referred to α at some point as the maximum probability of type I error. However, all 9 of these sources defined α as the probability of type I error (with no mention of 'maximum'), indicating some ambiguity in the authors' conception.

3.4. Faulty Arguments Resulting from Inconsistent Definitions

Our assertion that definitional ambiguities lead to mistakes is strongly supported by the fact that 11 out of the 25 sources actually make incorrect statements involving type I error. Here are some examples, along with our analyses of their mistakes:

"An α of 0.05 indicates that you are willing to accept a 5% chance that you are wrong when you reject the null hypothesis." [8]

Besides mistaking an unconditional for a conditional event, the author of [8] has also reversed the condition.

"For a 95% confidence level, the value of α is 0.05. This means that there is a 5% probability that we will reject a true null hypothesis. In the long run, one out of every twenty hypothesis tests that we perform at this level will result in a type I error." [12-13].

The authors of [12] and [13] should have said, 'In the long run, one out every twenty hypothesis tests *in which H0 is actually true* will result in a type I error.'

"What if I said the probability of committing a type I error was 20%? A more common way to express this would be that we stand a 20% chance of putting an innocent man in jail." [18].

Source [18] unambiguously identified type I error as "one of four possible outcomes". If the significance level is 20%, then the chance of obtaining a false conviction is actually somewhat less than 20%.

"An example of α risk in finance would be if one wanted to test the hypothesis that the average yearly return on a group of equities was greater than 10%. So the null hypothesis would be if the returns were equal to or less that 10%.... If, after statistically looking at the sample, you determine that the average yearly return is higher than 10%, you would reject the null hypothesis. But in reality, the average return was 6% so you have made a type I error. The probability that you have made this error in your test is the α risk." [19-20].

Sources [19] and [20] repeats the error of [8] above by reversing the condition.

3.5. Summary of Internet Sources' Usage of Type I Error Terminology

Table 1 summarizes our findings related to type I error usage in top-ranked web sites. Out of 25 references, 15 give a mixed message by including statements supporting both of the two mutually exclusive definitions of type I error. Only 6 references mention conditional probability, which is a key concept in understanding the difference between the two definitions. No references gave any mention of possible confusions between definitions.

Table 1. Summary of type I error usage in cited web references.

	Reference number (see References for citation)												
	21	**22**	**23**	**11**	**15**	**24**	**19**	**20**	**17**	**5**	**6**	**7**	**25**
Statements supporting conditional event interpretation of type I error:													
Clearly defines type I error as a conditional event		x											
Defines α as a probability of type I error	x	x	x			x	x	x	x	x	x	x	
Statements supporting unconditional event interpretation of type I error													
Defines type I error as unconditional event	x			x	x		x	x	x				
Gives example (s) in which type I error is treated as an unconditional event	x		x	x	x		x	x					x
Defines α as a prob. of type I error given H0 is true			x										x
Refers to α as maximum probability of type I error	x									x	x	x	
Ambiguous or false usage of type I error													
Gives an ambiguous definition of type I error		x	x		x	x	x	x		x	x	x	x
Makes false statement about probability of type I error				x	x		x	x					
Gives clarifying details to further explain the notion of type I error													
Mentions conditional probability		x	x							x	x	x	
Gives 2×2 table (or equivalent)	x												
Mentions potential confusion about type I error	none												

Table 1. Continued.

	Reference number (see References for citation)											
	16	**8**	**26**	**14**	**27**	**28**	**18**	**9**	**12**	**13**	**29**	**10**
Statements supporting conditional event interpretation of type I error:												
Clearly defines type I error as a conditional event												
Defines α as a probability of type I error	x	x	x				x	x	x	x	x	x
Statements supporting unconditional event interpretation of type I error												
Defines type I error as unconditional event				x		x			x	x	x	x
Gives example (s) in which type I error is treated as an unconditional event			x		x	x	x	x			x	x
Defines α as a prob. of type I error given H0 is true				x	x							
Refers to α as maximum probability of type I error								x	x	x	x	x
Ambiguous or false usage of type I error												
Gives an ambiguous definition of type I error	x	x	x			x		x				x
Makes false statement about probability of type I error		x	x				x		x	x	x	x
Gives clarifying details to further explain the notion of type I error												
Mentions conditional probability												x
Gives 2×2 table (or equivalent)	x	x	x				x					x
Mentions potential confusion about type I error	none											

4. Discussion

Our results show that several leading Internet sites give definitions that are either ambiguous or inconsistent, and some sites' definitions contradict the use cases they themselves provide. We have summarized these inconsistencies in Table 1. The ideas and notation of conditional probability, which are well-suited to clarify the confusion, are either downplayed or ignored. Furthermore, none of the sites that we examined gave any indication that two alternative definitions of type I error exist.

It may seem perplexing that these terminological inconsistencies have persisted for so long, and are still widely unrecognized. However, a parallel phenomenon may be found in the field of philosophy. Wittgenstein and other linguistic philosophers have asserted that many long-standing philosophical problems could be attributed to misuse of language. In Aphorism 38 of his *Philosophical Investigations*, Wittgenstein states: “Philosophical problems arise when language goes on holiday.” [31].

The unconditional definition is more natural one in light of ordinary language usage: indeed, the notion of “conditional event” requires a somewhat technical explanation, and is not commonly used in everyday speech. On the other hand, it seems that professional statisticians favor the idea that type I error is a conditional event. For example, an anonymous referee for one of our previous papers commented, “A type I error is a conditional statement. It means ‘rejecting the null hypothesis given that the null hypothesis is true.’” Interestingly, this idea is inconsistent with Neyman and Pearson’s original usage of the terminology. In their 1933 paper [30] which introduces the concept, Neyman and Pearson clearly indicate that type I error is *not* conditioned on the truth of the null hypothesis, and that the probability of type I error depends on the prior probability of the null hypothesis. Thus the original definition of type I error agrees with the simpler, less technical interpretation. Somehow, the meaning of type I error among statisticians has migrated from its original significance.

One of the obstacles to addressing the problem of inconsistent usage is that there are many types of users of statistics. Besides professional statisticians, there are practitioners of statistics in various fields who have developed their own conventions. Indeed, professional statisticians now comprise a minority of statistics users and instructors. For example, in our small university there are separate basic statistics classes in the business, psychology, and criminal justice departments, in addition to the class taught in the mathematics department. These classes are taught by professors in disparate fields who present statistics as a professional tool rather than a mathematical theory.

5. Conclusion

At this stage, it seems difficult if not impossible to achieve agreement over a single definition of type I error, because different types of users have different entrenched usages. The unconditional event interpretation of type I error is preferred for popular consumption, because it is easy to grasp and requires no mathematical sophistication to appreciate. On the other hand, professional statisticians habitually refer to the significance level α as the “probability of type I error”, which requires that type I error be defined as a conditional event.

We suggest that users of statistics should keep this definitional ambiguity in mind as they make use of the hypothesis testing literature. In particular, teachers of statistics should point it out to their students, and to provide them with concepts of conditional probability that will enable them to make proper distinctions and avoid erroneous conclusions.

The significance of our result extends beyond the immediate application we have considered. Our example shows that even in a rigorous technical field such as statistics, widespread confusions may persist due to lack of disciplinary consensus on terminology. There may well be cases in other academic fields where long-standing misconceptions spring from faulty terminology.

References

[1] Bar-Hillel M, Falk R. Some teasers concerning conditional probabilities. Cognition. 1982; 11 (2): 109-122. doi: 10.1016/0010-0277 (82) 90021-x.

[2] Falk R. Misconceptions of statistical significance. Journal of Structural Learning. 1986; 9: 83–96.

[3] Rossmo DK. Failures in Criminal Investigation: Errors of Thinking. The Police Chief 2009: 76 (10).

[4] Thron C, Miller V. Persistent Confusions about Hypothesis Testing in the Social Sciences. Social Sciences. 2014: 4 (2); 361-372. doi: 10.3390/socsci4020361.

[5] Khanacademy.org. Hypothesis test for difference of means [Internet]. n. d. [cited 21 July 2016]. Available from: https://www.khanacademy.org/math/probability/statistics-inferential/hypothesis-testing-two-samples/v/hypothesis-test-for-difference-of-means.

[6] Khanacademy.org. Small sample hypothesis test [Internet]. Khan Academy. n.d. [cited 21 July 2016]. Available from: https://www.khanacademy.org/math/probability/statistics-inferential/hypothesis-testing/v/small-sample-hypothesis-test.

[7] Khanacademy.org. Type 1 errors [Internet]. n. d. [cited 21 July 2016]. Available from: https://www.khanacademy.org/math/probability/statistics-inferential/hypothesis-testing/v/type-1-errors.

[8] Minitab.com. What are type I and type II errors? [Internet]. n. d. [cited 21 July 2016]. Available from: http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/type-i-and-type-ii-error/.

[9] Smith MK. Type I and II Errors [Internet]. n. d. [cited 21 July 2016]. Available from: https://www.ma.utexas.edu/users/mks/statmistakes/errortypes.html.

[10] Wikipedia.org. Type I and type II errors. [Internet]. n. d. [cited 21 July 2016]. Available from: https://en.wikipedia.org/wiki/Type_I_and_type_II_errors.

[11] Explorable.com. Type I Error and Type II Error - Experimental Errors [Internet]. n. d. [cited 21 July 2016]. Available from: https://explorable.com/type-i-error.

[12] Statistics.about.com. Hypothesis Test Example [Internet]. n. d. [cited 21 July 2016]. Available from: http://statistics.about.com/od/HypothesisTests/a/Hypothesis-Test-Example-With-Calculation-Of-Probability-Of-Type-I-And-Type-II-Errors.htm.

[13] Statistics.about.com. What Is the Difference Between Type I and Type II Hypothesis Testing Errors? [Internet]. n.d. [cited 21 July 2016]. Available from: http://statistics.about.com/od/Inferential-Statistics/a/Type-I-And-Type-II-Errors.htm.

[14] Onlinestatbook.com. Type I and II Errors [Internet]. n. d. [cited 21 July 2016]. Available from: http://onlinestatbook.com/2/logic_of_hypothesis_testing/errors.html.

[15] Huang H. Multiple Hypothesis Testing and False Discovery Rate [Internet]. n. d. [cited 21 July 2016]. Available from: http://www.stat.berkeley.edu/~hhuang/STAT141/Lecture-FDR.pdf.

[16] Lap.umd.edu. PSYC200_0803.pdf [Internet]. n. d. [cited 21 July 2016]. Available from: http://lap.umd.edu/psyc200/handouts/PSYC200_0803.pdf.

[17] Jones J. Hypothesis Testing [Internet]. n. d. [cited 21 July 2016]. Available from: https://people.richland.edu/james/lecture/m113/def_testing.html.

[18] Sigmazone.com. Roger Clemens and a Hypothesis Test [Internet]. n. d. [cited 21 July 2016]. Available from: http://www.sigmazone.com/Clemens_HypothesisTest.htm.

[19] Investopedia.com. Type I Error Definition | Investopedia [Internet]. 2007. [cited 21 July 2016]. Available from: http://www.investopedia.com/terms/t/type_1_error.asp.

[20] Investopedia.com. Alpha Risk Definition | Investopedia [Internet]. 2010. [cited 21 July 2016]. Available from: http://www.investopedia.com/terms/a/alpha-risk.asp.

[21] Noortgate WV, Onghena P, Sotos AE, Vanhoof S. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. Educational Research Review 2007; 2 (2); 98-113. doi: 10.1016/j.edurev.2007.04.001.

[22] Banerjee A, Bhawalkar JS, Chaudhury S, Chitnis UB, Jadhav SL. Hypothesis testing, type I and type II errors [Internet]. n.d. [cited 21 July 2016]. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2996198/.

[23] Campbell R. Type I and II error. [Internet]. n. d. [cited 21 July 2016]. Available from: http://www.cs.uni.edu/~campbell/stat/inf5.html

[24] Emathzone.com. Type 1 Error and Type 2 Error. [Internet]. n.d. [cited 21 July 2016]. Available from: http://www.emathzone.com/tutorials/basic-statistics/type-1-error-and-type-2-error.html.

[25] Ilvento TP. Values, Type I and Type II Error [Internet]. n. d. [cited 21 July 2016]. Available from: http://www.ilvento.org/Lectures/Hypothesis%203.pdf.

[26] King WB. Hypothesis testing [Internet]. n. d. [cited 21 July 2016]. Available from: http://ww2.coastal.edu/kingw/psycstats/mathreview/reviewfiles/hypothesis.html.

[27] Onlinecourses.science.psu.edu. 7.2 - Terminologies, Type I and Type II Errors for Hypothesis Testing [Internet]. n. d. [cited 21 July 2016]. Available from: https://onlinecourses.science.psu.edu/stat500/node/40.

[28] Quora.com. In statistics, do the probabilities of Type 1 and Type 2 errors have to add up to 1? [online]. n. d. [cited 21 July 2016]. Available from: https://www.quora.com/In-statistics-do-the-probabilities-of-Type-1-and-Type-2-errors-have-to-add-up-to-1.

[29] Quora.com. In statistics, what is a type 1 and type 2 error? [Internet]. n.d. [cited 21 July 2016]. Available from: https://www.quora.com/In-statistics-what-is-a-type-1-and-type-2-error.

[30] Stomponstep1.com. P-Value, Statistical Significance & Types of Error [Internet]. 2014. n. d. [cited 21 July 2016]. Available from: http://www.stomponstep1.com/p-value-null-hypothesis-type-1-error-statistical-significance/.

[31] Neyman, J., & Pearson, E. S. (1933, October). The testing of statistical hypotheses in relation to probabilities a priori. In Mathematical Proceedings of the Cambridge Philosophical Society (Vol. 29, No. 04, pp. 492-510). Cambridge University Press.