

Law and Statistical Disorder: Statistical Hypothesis Test Procedures
And the Criminal Trial Analogy

Tung Liu
Associate Professor of Economics
Phone: (765) 285-5360
Fax: (765) 285-4313
tliu@bsu.edu

and

Courtenay C. Stone
Professor of Economics
Phone: (765) 285-2857
Fax: (765) 285-4313
00ccstone@bsu.edu

Department of Economics
Ball State University
Muncie, IN 47306

Law and Statistical Disorder: Statistical Hypothesis Test Procedures
And the Criminal Trial Analogy

Abstract

Virtually all business and economics statistics texts start their discussion of hypothesis tests with some more-or-less detailed reference to criminal trials. Apparently, these authors believe that students are better able to understand the relevance and usefulness of hypothesis test procedures by introducing them first via the dramatic analogy of the criminal justice system. In this paper, we argue that using the criminal trial analogy to motivate and introduce hypothesis test procedures represents bad statistics and bad pedagogy. First, we show that statistical hypothesis test procedures can not be applied to criminal trials. Thus, the criminal trial analogy is invalid. Second, we propose that students can better understand the simplicity and validity of statistical hypothesis test procedures if these procedures are carefully *contrasted* with the difficulties of decision-making in the context of criminal trials. The criminal trial discussion provides a bad analogy but an excellent counter-example for teaching statistical hypothesis procedures and the nature of statistical decision-making.

Key words: hypothesis tests, criminal trials, Neyman-Pearson hypothesis test procedures.
JEL codes: A22, C12, K14.

Law and Statistical Disorder: Statistical Hypothesis Test Procedures
And the Criminal Trial Analogy¹

Tung Liu and Courtenay C. Stone²

A few years ago, we examined a large number of business and economics statistics texts and published a critical review of the approach used by most of them in their specification of the null hypothesis for one-tailed hypothesis tests.³ In our survey, we were surprised to discover that virtually all textbook authors began their hypothesis test discussions with a more-or-less detailed description of the similarities between criminal trials and hypothesis tests. Apparently, these authors believed that students are better able to understand the relevance and usefulness of hypothesis test procedures by introducing them via the dramatic analogy of the criminal justice system. In our view, however, using the criminal trial analogy to motivate and introduce hypothesis test procedures represents bad statistics and bad pedagogy. In this paper we demonstrate the nature of these statistical and pedagogical crimes.

The Crime Scene

“Perhaps the most commonly known example [of the use of analogies in teaching statistical concepts] is the likening of a statistical hypothesis test to the process of a criminal trial in which the ‘presumption of innocence’ plays the role of

¹ We would like to thank seminar participants at Ball State University and conference participants and discussants at the 2005 Eastern Economics Association and the Midwestern Economics Association Meetings for their comments and interest in this paper.

² The authors are Associate Professor and Professor of Economics, respectively, at Ball State University, Muncie, IN.

³ Liu and Stone (1999).

assuming the truth of the null hypothesis. ... [This] analogy ... has found common usage in a large number of introductory statistics texts”⁴

The use of the criminal trial analogy varies greatly across business and economics statistics texts. In some cases, the criminal trial analogy appears and disappears in an almost “hit and run” fashion. For example, one statistics text simply notes:

“Determining the null and alternative hypotheses is often a difficult task for students. The null hypothesis represents the situation that is assumed to be true unless the evidence is strong enough to convince the decision maker that it is not true. A common analogy is with the legal system, in which a defendant is assumed innocent unless the evidence convinces a jury that the person is guilty. All little bit of evidence is not sufficient.”⁵

...

“Math statisticians argue that you should never use the phrase ‘accept the null hypothesis.’ Instead you should use ‘do not reject the null hypothesis.’ ... This is why in a jury verdict to acquit a defendant, the verdict is ‘not guilty’ rather than innocent. Just because the evidence is insufficient to convict does not necessarily mean that the defendant is innocent.”⁶

In contrast, other texts use extended discussions of the criminal trial analogy to introduce hypothesis test procedures. For example,

“You will discover that hypothesis testing has a wide variety of applications in business and economics, as well as many other fields.

.....

There are a variety of *nonstatistical* (our italics) applications of hypothesis testing, the best known of which is a criminal trial.

When a person is accused of a crime, he or she faces a trial. The prosecution presents its case and a jury must make a decision on the basis of the evidence presented. In fact, the jury conducts a test of hypothesis. There are actually two hypotheses that are tested. The first is call the **null hypothesis** and is represented by H_0 ... It is

H_0 : The defendant is innocent.

The second is called the **alternative** or **research hypothesis** and is denoted H_1 . In a criminal trial it is

⁴ Martin (2003), p. 5-6.

⁵ Groebner et al. (2006), p. 305.

⁶ Ibid, page 307.

H_1 : The defendant is guilty.

Of course, the jury does not know which hypothesis is correct. They must make a decision on the basis of evidence presented by both the prosecution and the defense. There are only two possible decisions—convict or acquit the defendant.

In statistical parlance, convicting the defendant is equivalent to *rejecting the null hypothesis in favor of the alternative*. Acquitting a defendant is phrased as *not rejecting the null hypothesis in favor of the alternative*. Notice that we do *not* say that we accept the null hypothesis. In a criminal trial, that would be interpreted as finding the defendant innocent. Our justice system does not allow this decision.

When testing hypotheses there are two possible errors. ... In a criminal trial, a Type I error is made when an innocent person is wrongly convicted. A Type II error occurs when a guilty defendant is acquitted. ...

In our justice system, Type I errors are regarded as more serious. As a consequence, the system is set up so that the probability of a Type I error is small. This is arranged by placing the burden of proof on the prosecution (the prosecution must prove guilt—the defense need not prove anything) and by having the judges instruct the jury to find the defendant guilty only if there is ‘evidence beyond a reasonable doubt.’ In the absence of enough evidence, the jury must acquit even though there may be some evidence of guilt. The consequence of this arrangement is that the probability of acquitting guilty people is relatively large. Oliver Wendell Holmes, a United States Supreme Court justice, once phrased the relationship between the probabilities of a Type I and Type II errors in the following way: ‘Better to acquit 100 guilty men than convict one innocent one.’ In Justice Holmes’s opinion, the probability of a Type I error should be 1/100 of the probability of a Type II error.

...

Let’s extend these concepts to *statistical* (our italics) hypothesis testing.”⁷

Table 1 emphasizes one author’s view of the extensive similarities between criminal trials and statistical hypothesis tests. According to this author

“... [the] judicial analogy for hypothesis testing is a particularly powerful one, as many of the facets of the legal process have a direct counterpart (map) in the formal statistical procedure. ... Moreover, the process of a criminal trial mirrors that of a statistical hypothesis test to a large degree.”⁸

Criminal Charges

⁷ Keller and Warrack (2004), pp. 320-1.

⁸ Martin (2003), p. 6.

The texts that use the criminal trial analogy as described above are actually using a *nonstatistical* problem disguised as an example of statistical hypothesis test procedures in an attempt to make it easier and, perhaps, more interesting for students to understand the rationale behind these procedures.⁹ In our view, this masquerade is a serious felony, not simply a well-intentioned misdemeanor.

In a series of articles written between 1924 and 1934, Neyman and Pearson developed the statistical hypothesis test procedures utilized by statisticians everywhere and described and used in every statistics textbook today.¹⁰ In perhaps their best known article, written more than 70 years ago, they present their view of the basic nature of hypothesis test procedures:

“In general terms that problem is this: Is it possible that there are any efficient tests of hypotheses based upon the theory of probability, and if so, what is their nature? Before trying to answer this question, we must attempt to get closer to its exact meaning. In the first place, it is evident that the hypotheses to be tested by means of the theory of probability must concern in some way the probabilities of the different kinds of results of certain trials. That is to say, they must be of a statistical nature, or as we shall say later on, they must be statistical hypotheses.”¹¹

In other words, *statistical* hypothesis tests can not be conducted without the ability to measure the probabilities of the Type I and Type II errors associated with the test. This requirement completely disqualifies the relevance of the criminal trial analogy for statistical hypothesis test purposes. Consider, for example, the difference between the decision matrices associated with statistical hypothesis tests and criminal trials shown in Table 2A and 2B. The decision matrix in Table 2A for the statistical hypothesis test

⁹ Although a few authors (e.g., Keller and Warrack (2004) caution their readers that this example is “nonstatistical”, they do not explain the relevance of the difference between “statistical” and “nonstatistical” examples for hypothesis test procedures.

¹⁰ See, for example, David (1981) and Chiang (website) for further discussion of Neyman and Pearson’s fundamental role in developing statistical hypothesis test procedures.

¹¹ Neyman and Pearson (1933), p. 290.

procedure appears, in one form or another, in virtually all statistics textbooks. The discussion in these texts generally makes the following points:

- 1) The null and alternative hypotheses concern the *numerical* value of a specific population parameter.
- 2) The decision rule is based on the relevant sampling distribution of the test statistic for the population parameter under the null hypothesis and the selected value of the probability of a Type I error.
- 3) The actual decision is based on the selected decision rule and the result of a random sample taken from the relevant population.
- 4) The decision is to reject or fail to reject the null hypothesis.

In contrast, the decision matrix in Table 2B for the criminal trial differs from that in Table 2A in a number of important ways—two of which immediately come to mind. First, there is no underlying sampling probability distribution to describe the decision errors in criminal trials. Second, there are at least six possible outcomes associated with criminal trials, several of which involve “errors” regardless of the guilt or innocence of the person on trial. Clearly, criminal trial procedures and outcomes are fundamentally different from statistical hypothesis test procedures and outcomes.

Table 3 further illustrates this point by “adjusting” Table 1 to show the numerous and crucial differences between criminal trial procedures and statistical hypothesis test procedures. A comparison of Tables 1 and 3 should convince any impartial jury of business and economics statistics textbook authors, faculty and students that the use of the criminal trial analogy to illustrate and illuminate statistical hypothesis test procedures is simply wrong.

Interestingly enough, the first mention of criminal trial outcomes that we have discovered appears in the Neyman and Pearson paper cited above. Their discussion, however, is concerned solely with an emphasis on the errors that occur in statistical hypothesis tests.

“Is it more serious to convict an innocent man or to acquit a guilty? That will depend on the consequences of the error: is the punishment death or fine; what is the danger to the community of released criminals? From the point of view of mathematical theory all that we can do is to show how the risk of errors may be controlled and minimized.”¹²

In this short discussion, Neyman and Pearson clearly distinguish between the considerations involved in a criminal trial and “the point of view of mathematical theory.” They do not use the criminal trial discussion as an analogy for better understanding the nature and relevance of hypothesis testing procedures. Instead, they use it simply to emphasize that the choices of the null hypothesis and the specific significance level for the hypothesis test depend crucially on the relative costs of the alternative errors associated with the hypothesis test.¹³

Recently, Brian Forst discusses, in a somewhat different context, the same issues that Neyman-Pearson covered 70 years ago. Forst, who teaches statistics to criminal justice students, is concerned with both the difficulty of teaching statistical hypothesis test procedures and the fundamental difference between these procedures and criminal justice decisions:

“Requiring students of criminal justice to learn the fundamentals of statistical inference may or may not be good for them, but it surely can enlighten the instructor. In searching for a way to motivate my students to learn about Type I

¹² Neyman and Pearson (1933), p. 296.

¹³ Thus, the choice of which of two alternative claims should be selected as the null hypothesis and whether the significance level of the test should be 0.10 or 0.01 (to pick just two of the typical choices facing investigators) depend, in each case, on the relative costs of the errors in the decision matrix for the test.

and Type II errors and the logic of statistical inference, I have asked them whether they are concerned about errors of inference made by police, prosecutors, juries, and sentencing judges. It has struck me, in discussing these metaphors, that we have a coherent sophisticated, effective framework for managing errors in statistical inference, but no such framework ... in the criminal justice system.”¹⁴

Statistical hypothesis tests are all about the errors—specifically, the consideration of Type I and Type II errors derived from the sampling distributions. Because there is no similar framework of analysis for the criminal justice system, the widespread analogy between statistical hypothesis test procedures and criminal trial procedures is statistically and pedagogically in error.

The Crime Solved: the Verdict and the Sentence.

The criminal trial scenario provides a dramatic and interesting example of the alternative outcomes and the costs involved in the decision-making process in jury trials. Unfortunately, as currently used in numerous business and economics statistics textbooks, it also provides an erroneous and misleading analogy for the decision-making process involved in statistical hypothesis test procedures. However, in our opinion, the criminal trial scenario can be used effectively, in an interesting and dramatic way, to reinforce student’s understanding of the relevance and inherent simplicity of statistical hypothesis test procedures. After students are introduced to the basic concepts of statistical hypothesis test procedures (the choice of the null and alternative hypotheses, the use of the sampling distributions and the probabilities of Type I and Type II errors), the contrast between criminal trial decision-making procedures and statistical hypothesis test procedures can increase the students’ understanding and appreciation of the

¹⁴ Forst (2004), p. xiii.

importance and usefulness of the latter. Discussion of these contrasts would naturally focus on the *differences* shown in Tables 2 and 3.

Summary and Conclusions

Those of us who teach statistics know two important things about statistical hypothesis tests procedures. First, the ability to conduct hypothesis tests and correctly interpret their results is one of the most important skills that business and economics students can acquire. A recent BusinessWeek article emphasized this fact when it reported that statistics and probability "... will become core skills for businesspeople and consumers as we grapple with challenges involving large data sets. Winners will know how to use statistics—and how to spot when others are dissembling."¹⁵

Unfortunately, the ability to appropriately use statistical hypothesis test procedures is one of the most difficult skills for students to learn.¹⁶ Because "criminal trials are inherently dramatic,"¹⁷ it is all too tempting to use the criminal trial analogy in an attempt to make a difficult topic a little easier and more interesting for students. It is obvious that many statistics textbook authors have done so. However, the current use of this analogy in numerous textbooks is problematic on both statistical and pedagogical grounds. We believe that this problem can be corrected if the critical differences between the decision-making procedures that underlie criminal trials and statistical hypothesis tests are emphasized.

¹⁵ Baker (2006), p. 60.

¹⁶ For survey results on the most difficult statistical topics, see Aczel (1995), p. viii.

¹⁷ Advertisement for Law and Order reruns.

References

- Aczel, A. D. 1995. *Statistics: Concepts and Applications* (Irwin Publishers).
- Baker, Stephen. 2006. "Math Will Rock Your World." (*BusinessWeek*, January 23, 2006), pp. 54-62.
- Chiang, Chin Long. "Neyman, Jerzy." *ASA Statisticians in History*, American Statistical Association Web Site (<http://www.amstat.org/about/statisticians/>).
- David, H. A. 1981. "Egon S Pearson, 1895-1980." *American Statistician* 35 (2) 94-95.
- Forst, Brian. 2004. *Errors of Justice* (Cambridge University Press).
- Groebner, David F., Patrick W. Shannon, Phillip C. Fry and Kent D. Smith. 2006. *A Course in Business Statistics, Fourth Edition* (Pearson/Prentice Hall).
- Keller, Gerald and Brian Warrack. 2004. *Statistics for Management and Economics, Abbreviated 6th Edition* (Thomson Brooks/Cole).
- Liu, Tung and Courtenay C. Stone 1999, "A Critique of One-Tailed Hypothesis Test Procedures in Business and Economics Statistics Textbooks," *The Journal of Economic Education* (Winter), pp, 59-63
- Martin, Michael A. 2003. "'It's like ... you know': The Use of Analogies and Heuristics in Teaching Introductory Statistical Methods." *The Journal of Statistics Education*, 11(2) (www.amstat.org/publications/jse/v11n2/martin.html)
- Neyman, Jerzy and Egon S. Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society A*, pp. 289-337.

Table 1. Martin's Comparison Of Criminal Trials
and Hypothesis Tests¹⁸

Criminal Trial	Hypothesis Test
Defendant is innocent	Null Hypothesis
Defendant is guilty	Alternative Hypothesis
Gathering of evidence	Gathering of data
Summary of evidence	Calculation of the test statistic
Cross-examination	No equivalent
Jury deliberation and decision	Application of decision rule
Verdict	Decision
Verdict is to acquit	Failure to reject the null hypothesis
Verdict is to convict	Rejection of the null hypothesis
Presumption of innocence	Assumption that the null hypothesis is true
Conviction of an innocent person	Type I error
Acquittal of a guilty person	Type II error
Beyond reasonable doubt	Fixed (small) probability of Type I error
High probability of convicting a guilty person	High power
Mistrial	No equivalent – perhaps the role of data snooping?

¹⁸ Martin (2003), p. 6.

Table 2. Decision Matrices for Statistical Hypothesis Tests vs. Criminal Trials

A. Statistical Hypothesis Test

	H_0 is "True"	H_a is "True"
Fail to Reject H_0	Correct	Incorrect (Type II error)
Reject H_0	Incorrect (Type I error)	Correct

B. Criminal Trials

	Defendant is Innocent	Defendant is Guilty
Verdict: Not Guilty	Correct	Incorrect
Verdict: Not Guilty Because Jury Disapproves of the Law (Jury Nullification)	Correct	Incorrect
Verdict: Guilty	Incorrect	Correct
Verdict: Guilty. Convicted of a Lesser Crime Than Committed (Plea Bargain)	Incorrect	Incorrect
Verdict: Guilty. Convicted of a Greater Crime Than Committed	Incorrect	Incorrect
No Verdict: Mistrial	Incorrect	Incorrect

Table 3. Criminal Trials vs. Hypothesis Tests: Table 1 Revised

Criminal Trial Components	Differences from Hypothesis Test Components
<u>Defendant is innocent</u> : claim about a nonnumeric characteristic of a specific individual. Also see Table 2B.	NOT EQUIVALENT to the <u>Null Hypothesis</u> : The null hypothesis is a claim about the value of a parameter of a specific population
<u>Defendant is guilty</u> : see above	NOT EQUIVALENT to the <u>Alternative Hypothesis</u> : see above
<u>Gathering of evidence</u> : two highly selective searches for evidence to support innocent verdict (defense attorneys) or guilty verdict (law enforcement/prosecution attorneys). Biases are introduced into the search process by the motivations of the defense and prosecution to strengthen their respective cases and by the laws of evidence.	NOT EQUIVALENT to <u>Sampling</u> : a single random sample from population. The process is designed to avoid bias in data selection.
<u>Summary of evidence</u> : two highly selective summaries intended to support innocent verdict (defense attorneys) or guilty verdict (law enforcement/prosecution attorneys).	NOT EQUIVALENT to the <u>Calculation of the test statistic</u> : the calculation of the test statistic is based on sampling distribution theory and the null hypothesis.
<u>Cross-examination</u> : attempts to discredit the opponents evidence/witnesses	NO EQUIVALENCE : If the data are randomly selected from the specific population, there are no reasons to challenge the data.
<u>Jury deliberation and decision</u> : a process that in which jury members evaluate the evidence and attempt to agree jointly on a unanimous verdict (criminal trial) or a majority verdict (civil trial).	NOT EQUIVALENT to the <u>Application of decision rule</u> : the decision rule is based on the test statistic chosen, the alternative hypothesis (one or two-tailed) and the relevant level of significance.
<u>Verdict</u> : The decision will be to convict, to acquit, or to declare a mistrial because the jury cannot agree on a verdict. In the case of a conviction, the jury's decision may be to convict of a crime that is either less serious or more serious than the one actually committed. It is subject to appeal by the defense, which may or may not result in a new trial. In the case of a mistrial, the prosecution may or may not decide to retry the defendant.	NOT EQUIVALENT to the <u>Decision</u> : the decision depends on the critical value (see above) of the test statistic, the actual value of the test statistic (based on the sample statistic and the null hypothesis) and the decision rule.
<u>Verdict is to acquit</u> : this verdict, based on a unanimous vote of the jury, can occur for several reasons. The jury may not have found the defendant guilty "beyond a shadow of a doubt," they may have believed the defendant to be guilty but thought that the charge was incorrect (e.g., first degree murder vs. manslaughter), or they may have believed the defendant to be guilty but thought that the law was invalid ("jury nullification").	NOT EQUIVALENT to the <u>Failure to reject the null hypothesis</u> : the decision to "fail to reject the null hypothesis" occurs automatically when the test statistic from the sample does not fall into the rejection region (or regions).
<u>Verdict is to convict</u> : this verdict occurs when the jury unanimously votes to convict. Again, there are several possible reasons for this verdict.	NOT EQUIVALENT to the <u>Rejection of the null hypothesis</u> : the decision to reject the null hypothesis occurs automatically when the test statistic from the sample falls into the rejection region (or regions).
<u>Presumption of innocence</u> : The notion that the defendant is presumed innocent until "proven guilty beyond a reasonable shadow of a doubt." Of course, the prosecution does not presume that the defendant is innocent and their job is to convince the jury of the defendant's guilt. Similarly, the defense	NOT EQUIVALENT to the <u>Assumption that the null hypothesis is true</u> : Generally, the null hypothesis is the one that the investigator would like to reject. Hypothesis test procedures are intended to provide valid statistical reasons for doing so.

attorneys do not have to presume that their defendant is innocent—they only have to convince the jury that there is reasonable doubt.	
<u>Conviction of an innocent person</u> : Since there is no way to determine, a priori, whether this will occur (or has occurred) in a specific trial, the probability that it will occur is both unknown and unknowable. Moreover, the impact of additional evidence on this outcome is unclear. Also see Table 2B.	NOT EQUIVALENT to a Type I error: Because the Type I error can be chosen precisely by the investigator, the probability of falsely rejecting the null hypothesis is both controllable and known.
<u>Acquittal of a guilty person</u> : see above	NOT EQUIVALENT to a Type II error: In general, the probability of making a Type II error is unknown (because it requires knowledge of the true value of the population parameter—which, of course, is unknown).
<u>Beyond reasonable doubt</u> : There is no “reasonable” way to determine what this concept means. It differs across defendants, across trials and, in the occurrence of mistrials, clearly differs across jurors in specific trials. The probabilities involved are unobservable and unknown.	NOT EQUIVALENT to a Fixed (small) probability of Type I error: For any specific hypothesis test, the (small) probability of a Type I error is selected explicitly by the investigators. Choices among alternative acceptable values for Type I errors will occur, depending on the alternative costs of making Type I and Type II errors.
<u>High probability of convicting a guilty person</u> : In any specific criminal trial, two things are unknown: whether the defendant is guilty and the actual probability of convicting the defendant in this trial if he is guilty. Thus, there is no way to determine whether the probability of convicting a guilty person is “high” or “low.”	NOT EQUIVALENT to High power: The power of a test is the probability that it will correctly reject the null hypothesis. This can be calculated for different values of the alternative hypothesis and the trade-off between the Type I and Type II errors explicitly determined.
<u>Mistrial</u> : This occurs when the individual jurors agree not agree on a verdict.	NO EQUIVALENCE : Hypothesis tests always yield one of two outcomes. The analysis either “rejects” or “fails to reject” the null hypothesis. There are no “mistrials” in statistical hypothesis tests.