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

Tung Liu
Associate Professor of Economics
Department of Economics
Ball State University
Muncie, IN 47306
Phone: (765) 285-5360
Fax: (765) 285-4313
tliu@bsu.edu

and

Courtenay C. Stone
Professor of Economics
Department of Economics
Ball State University
Muncie, IN 47306
Phone: (765) 285-2857
Fax: (765) 285-4313
00ccstone@bsu.edu

KEY WORDS: hypothesis tests, criminal trials, Neyman-Pearson hypothesis test procedures.

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

Abstract

Virtually all business and economics statistics texts incorporate some more-or-less detailed discussion of the similarities between conducting hypothesis tests and criminal trials. Apparently, the authors of these texts believe that students will be 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 this paper, we show that the criminal trial analogy commonly used by business and economics statistics textbook authors to motivate, illustrate and/or demonstrate hypothesis test procedures represents bad statistics and bad pedagogy. We then show how the criminal trial setting can be used correctly to illustrate some important statistical concepts.

# Law and Statistical Disorder: Statistical Hypothesis Test Procedures And the Criminal Trial Analogy[1]

Several years ago, we published a critical review of the approach used by most business and economics statistics texts in their specification of the null hypothesis for one-tailed hypothesis tests.[2] In conducting our textbook survey, we were surprised to discover that virtually all textbook authors began their hypothesis test discussion with a more-or-less detailed description of the similarities between criminal trials and hypothesis tests. Apparently, these authors believe that students are better able to understand the relevance and usefulness of statistical hypothesis test procedures by introducing them via the dramatic analogy of the criminal justice system. However, use of the criminal trial analogy to motivate and illustrate hypothesis test procedures represents bad statistics and bad pedagogy. In this paper we demonstrate the nature of these errors. We also show that some components of the criminal trial setting, but not the whole trial procedure itself, can be used effectively to illustrate some important statistical concepts.

## 1. The Use of the Criminal Trial Analogy to Explain Hypothesis Test Procedures

> "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 XXXX University and conference participants and discussants at the 2005 Eastern Economics Association and Midwest Economics Association Meetings for their comments and interest in an earlier version of this paper.
[2] XXXX (1999).

> assuming the truth of the null hypothesis. … [This] analogy … has found common usage in a large number of introductory statistics texts ….”[3]

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

> “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. A little bit of evidence is not sufficient.”[4]
>
> …
>
> “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.”[5]

In contrast, other texts use extended discussions of the criminal trial analogy. For example,

> “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
>
> $H_1$: The defendant is guilty.

[3] Martin (2003), p. 5-6.
[4] Groebner et al. (2006), p. 305.
[5] Ibid, page 307.

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 quilt. 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."[6]

Finally, Table 1 provides a detailed list of the alleged extensive similarities between criminal trials and statistical hypothesis tests. In Martin's view,

"… [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."[7]

[6] Keller and Warrack (2004), pp. 320-1.
[7] Martin (2003), p. 6.

Table 1. Martin’s Comparison of Criminal Trials
And Hypothesis Tests[8]

| **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? |

Presumably, these similarities justify the widespread use of this analogy to motivate students and illustrate hypothesis test procedures and applications.

## 2, What’s Wrong With the Criminal Trial Analogy?

As noted above, the criminal trial analogy is widely used in statistics texts as a good example of (or analogy for) the type of problem that can be analyzed and answered using statistical hypothesis test procedures. Unfortunately, the criminal trial analogy is invalid. Statistical hypothesis test procedures are fundamentally different from those for criminal trials. Therefore, students who actually believe the criminal trial analogy to be

[8] Martin (2003), p. 6.

valid are more likely to misunderstand the precise nature of statistical hypothesis test procedures and applications.[9]

Statistical hypothesis test procedures can not be applied in the criminal trial process.[10] In a series of articles written between 1924 and 1934, Neyman and Pearson developed the statistical hypothesis test procedures used in every business and economics statistics textbook today.[11] In perhaps their most well-known article, written more than 70 years ago, they state the basic nature of these 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."[12]

Conducting *statistical* hypothesis tests requires the student to define and measure the probabilities of the Type I and Type II errors related to the alternative decisions associated with these tests. However, it is impossible to do this in criminal trials. Consider the differences in the decision matrices shown in Tables 2A and 2B.

Table 2A. Statistical Hypothesis Test Decision Matrix

| Decision | $H_0$ is "True" | $H_0$ is "False" |
|---|---|---|
| Fail to Reject $H_0$ | Correct | Incorrect (Type II error) |
| Reject $H_0$ | Incorrect (Type I error) | Correct |

[9] Even if students are unlikely to be *overly* misled by this analogy, textbooks, especially statistics textbooks, should not contain material that provides students with such opportunities.
[10] 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.
[11] See, for example, David (1981) and Chiang (website) for further discussion of Neyman and Pearson's fundamental role in developing statistical hypothesis test procedures.
[12] Neyman and Pearson (1933), p. 290.

The decision matrix applicable to statistical hypothesis tests, shown in Table 2A, appears in virtually all statistics textbooks. The discussion in these texts 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.

Table 2B. Criminal Trial Decision Matrix

| Jury's Decision | 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 |

In contrast, the criminal trial decision matrix, shown in Table 2B, differs from that in Table 2A in significant ways—three of which immediately obvious.

1. The criminal trial decision does not concern a *numerical* value for a specific *population parameter*. Instead, it concerns the *non-numeric* concepts of “guilt” and “innocence” (or perhaps, “guilty” and “not guilty enough”) for a specific individual.
2. The criminal trial decision does not involve just two possible outcomes common to statistical hypothesis tests. Instead, there are at least six possible outcomes, several of which involve “errors” regardless of the guilt or innocence of the person on trial.
3. There is no underling sampling probability distribution to describe the decision errors in criminal trials. The probability of Type I and Type II errors cannot be calculated for selected criminal trial decisions because their actual sampling distributions and the appropriate population distributions underlying them are completely unknown and, most likely, unknowable.

Consequently, criminal trial procedures and outcomes are fundamentally different from statistical hypothesis test procedures and outcomes and, thus, can not be viewed as analogous to them in any sense.

In his recent book, Brian Forst arrives at the same conclusion as he laments both the difficulty of teaching statistical hypothesis test procedures to criminal justice students and the fundamental differences between hypothesis test 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 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." [13]

Statistical hypothesis tests are all about the errors—specifically, consideration of Type I and Type II errors, their probabilities associated with the relevant sampling distributions and the difficult, but necessary, choice of the relevant significance level. There is no similar framework of statistical analysis for the decision process associated with criminal trials. Thus, the widespread textbook analogy between statistical hypothesis test procedures and criminal trial procedures is statistically and pedagogically invalid.

## 3. What's Right With Using the Criminal Trial Setting?

In the previous section, we showed that the use of the criminal trial *analogy* to explicate statistical hypothesis test procedures is statistically and pedagogically invalid. However, there are a number of interesting and valid statistical applications using selected components of the criminal trial setting. These applications can help students better understand statistical analysis and decision-making. For example, as cited above, Forst uses criminal trial outcomes to motivate his students to think about the importance of the Type I and Type II errors that arise in setting up a statistical hypothesis test. Of course, this specific use of criminal trial outcomes in this context did not originate with Forst. Perhaps the first link between criminal trial outcomes and statistical hypothesis test decision errors appears in the classic Neyman and Pearson paper cited previously:

> "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

[13] Forst (2004), p. xiii.

> mathematical theory all that we can do is to show how the risk of errors may be controlled and minimized."[14]

Neither Neyman and Pearson's comments nor Forst's discussion suggests that criminal trials provide a useful or relevant analogy for hypothesis test procedures. Instead, they use the verdicts associated with criminal trials to illustrate and emphasize how the relative costs of the errors associated with the hypothesis test should influence the choice of the null and alternative hypotheses and the significance level selected.

A very different, but also interesting and valid, application of the criminal trial setting uses Bayesian probability analysis. In this case, students are asked to consider the numerical values of the Bayesian probabilities associated with the verdicts of "not guilty enough (innocent)" and "guilty," given the evidence presented in the trial. This approach is illustrated by the following analysis from a mystery novel published 60 years ago.

> "What you are trying to do, of course, is to proceed from probability to certainty, as close as you can get. Say you start, as you see it, with one chance in five that I poisoned Orchard. Assuming that you have no subjective bias, your purpose is to move as rapidly as possible from that position, and you don't care which direction. Anything I say or do will move you one way or the other. If one way, the one-in-five will become one-in-four, one-in-three, and so on, until it becomes … close enough to affirmative certainty so that you will say you know I killed Orchard. If it goes the other way, your one-in-five will become one-in-ten, one-in-one-hundred...; and, when it gets to one-in-ten billion you will be close enough to negative certainty so that you will say you know that I did not kill Orchard. There is a formula---."[15]

There is indeed a formula—the Bayesian probability formula. And, unlike the criminal trials analogy's invalid use for statistical hypothesis test decision-making, its application to Bayesian probability analysis is both entertaining and valid.

[14] Neyman and Pearson (1933), p. 296

[15] Stout (1948), p. 69.

## 4. Summary and Conclusion

Those of us who teach statistics know at least two important things about statistical procedures and decision-making. First, the ability to conduct hypothesis tests and correctly interpret the results is one of the most important skills that business and economics students can acquire. A recent *BusinessWeek* article 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."[16]

Second, the ability to use statistical hypothesis test procedures correctly is one of the most difficult skills for students to learn.[17] Because "criminal trials are inherently dramatic,"[18] 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. Perhaps this is why so many business and economics statistics textbook authors have done so. And yet, the analogy is invalid. Students know that criminal trials are inherently adversarial, their outcomes frequently controversial, the trials are lengthy and the decisions are subject to appeal and, perhaps, reversal. This is precisely the opposite of statistical hypothesis test procedures and decisions. Actually, students might better understand the applicability and usefulness of hypothesis test procedures if they are shown why criminal trial procedures are *not* analogous to statistical hypothesis tests—that is, specifically why these concepts can not and do not apply to criminal trials. Now, that discussion would be inherently dramatic—in a statistical sense, of course.

---

[16] Baker (2006), p. 60.
[17] For survey results on the most difficult statistical topics, see Aczel (1995), p. viii.
[18] 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.

XXXX (1999), "A Critique of One-Tailed Hypothesis Test Procedures in Business and Economics Statistics Textbooks," *XXXX*

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.

Stout, Rex (1948), *And Be a Villain*: The Viking Press.