Statistics as Evidence of Age Discrimination

Gregory L. Harper

Follow this and additional works at: https://repository.uchastings.edu/hastings_law_journal

Part of the Law Commons

Recommended Citation
Available at: https://repository.uchastings.edu/hastings_law_journal/vol32/iss5/11

This Note is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Law Journal by an authorized editor of UC Hastings Scholarship Repository.
Statistics as Evidence of Age Discrimination

By Gregory L. Harper*

The use of statistics as circumstantial evidence of both disparate impact and disparate intent has become a mainstay of employment discrimination litigation. This is a result of the judicial recognition that unlawful discrimination may be subtle and lacking in direct evidence. The diversity of employment discrimination cases brought under Title VII of the Civil Rights Act of 1964 has resulted in considerable refinement of and reliance upon statistical evidence.

In 1967, Congress chose to enact the Age Discrimination in Employment Act (ADEA) rather than merely amend Title VII to include age as a protected classification. Nevertheless, the courts

* B.S., B.A., 1972, University of Illinois, Champaign-Urbana. Member, Second Year Class.
1. See text accompanying notes 24-26 infra.
2. "Since the passage of the Civil Rights Act of 1964, the courts have frequently relied upon statistical evidence to prove a violation." United States v. Ironworkers Local 86, 443 F.2d 544, 551 (9th Cir.), cert. denied, 404 U.S. 984 (1971) (footnote omitted).
3. See Id.
7. In January, 1967, President Johnson's annual message to the Senate called for and outlined the legislation that would later be enacted as the ADEA. His call did not surprise the Congress. He had already issued Executive Order No. 11,141, 3 C.F.R. 179 (1964-1965 Comp.), banning age discrimination by government contractors. In 1965, the Secretary of Labor, pursuant to a statutory mandate in § 715 of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-14 (1976), had already issued a report to Congress on age discrimination. Some members of Congress believed that an amendment of the Civil Rights Act to include age as a protected classification would be the best course, but others believed that enforcement would be more efficient through the procedures of the Fair Labor Standards Act, 29 U.S.C. §§ 201-219 (1976 & Supp. III 1979), rather than under the supervision of the Equal Employment Opportunity Commission (EEOC). The latter view prevailed, and separate legislation was passed. Nevertheless, the EEOC did take over enforcement responsibility from the Labor Department in July, 1979. See generally Note, The Age Discrimination in Employment
were quick to analogize the ADEA to Title VII, and to utilize their Title VII experience in the judicial enforcement of the ADEA. In one of the earliest ADEA cases, *Hodgson v. First Federal Savings,* the Fifth Circuit stated that “[w]ith a few minor exceptions the prohibitions of [the ADEA] are in terms identical to those of Title VII of the Civil Rights Act of 1964 except that ‘age’ has been substituted for ‘race, color, religion, sex or national origin.’” The First Circuit recently reiterated this view in *Loeb v. Textron, Inc.*, stating:

[O]ne naturally might expect to use the same methods and burdens of proof under the ADEA as under Title VII. Nothing in either the ADEA or its legislative history indicates a different conclusion.

The mere fact that Congress chose to pass a separate statute rather than to amend Title VII does not imply that age discrimination was intended to be subject to different standards and methods of proof than race or sex discrimination.

Courts faced with ADEA claims thus have used Title VII rulings in determining whether a prima facie case, or causation, has been established, evaluating motions for a directed verdict, and resolving disputes relating to notice and time limitations. Moreover, the use of statistics to prove unlawful age discrimination has been expressly allowed based on its analogous use in Title VII cases.

This Note first reviews the use of statistical evidence in Title VII cases and its applicability to ADEA cases. It next examines the statistical characteristics peculiar to age discrimination as compared to race or sex discrimination, and proposes methods of sta-

---


8. 455 F.2d 818 (5th Cir. 1972).
9. Id. at 820 (footnote omitted).
10. 600 F.2d 1003 (1st Cir. 1979).
11. Id. at 1015.
14. Id. at 311-12.
statistical analysis specifically applicable to age discrimination litigation. The Note concludes that statistical evidence is valid in ADEA cases, but because the characteristic of age is different from that of race or sex, different statistical tests should be employed in ADEA as compared to Title VII litigation.

The Use of Statistics in Title VII Cases

In 1970, the United States Supreme Court, in *Griggs v. Duke Power Co.*, 17 held: “Congress directed the thrust of the Act to the consequences of employment practices, not simply the motivation.” 18 The consequences of discrimination are most effectively demonstrated by statistics—numerical analysis that illustrates how and to what extent a protected class 19 of individuals is adversely affected by a particular employment practice. 20 *Griggs* itself illustrates this point in that only statistical evidence was presented, and on its strength the Court found unlawful discrimination. 21

Five years after *Griggs*, in *International Brotherhood of Teamsters v. United States*, 22 the Supreme Court expressly addressed the use of statistical evidence in Title VII cases, while also distinguishing between the two types of Title VII claims: “disparate treatment” and “disparate impact.” 23 In a disparate treatment case, the plaintiff must allege that, because of the plaintiff’s race, color, religion, national origin, or sex, the employer has intentionally treated him or her differently from other employees. 24 Hence, a disparate treatment case ultimately requires proof of a discriminatory motive (intent). 25 In a disparate impact case, the plaintiff...

---

18. *Id.* at 432 (emphasis in original).
21. To have been hired at Duke Power, an applicant must either have had a high school diploma or have passed an achievement exam. Statistical evidence demonstrated that, statewide, 34% of the white males had completed high school, while only 12% of the black males had done so, and that 58% of the white applicants were able to pass the achievement exam compared to 6% of the blacks. 401 U.S. at 430 n.6.
23. *Id.* at 335 n.15.
24. *Id.*
25. *Id.*
must allege that the employer has unnecessarily utilized a facially neutral employment practice that unintentionally has had an adverse effect on a protected class.26

Whereas Griggs showed the power of statistics in disparate impact cases,27 the Teamsters Court commented on the use of statistics in disparate treatment cases.28 The Court held that although proof of discriminatory motive is critical in disparate treatment cases, it could "in some situations be inferred from the mere fact of [statistical] differences in treatment."29

Statistically, the distinction between disparate impact and disparate treatment cases is important because, although statistical evidence alone will often be determinative in a disparate impact case as it was in Griggs, it can only raise an inference of discriminatory intent in a disparate treatment case such as Teamsters.30

As the Court in Teamsters noted: "[T]his was not a case in which the government relied on 'statistics alone.' The individuals who testified about their personal experiences with the company brought the cold numbers convincingly to life."31 The Court then reasoned that, in a disparate treatment case, the most important function of statistical evidence is to help establish the prima facie case.32

---

26. Id. See text accompanying note 18 supra.


28. 431 U.S. at 339 n.20.

29. Id. at 335 n.15. Regardless of whether disparate impact or disparate treatment is at issue, statistical evidence should never be ignored, because "[o]ur cases make it unmistakably clear that '[s]tatistical analyses have served and will continue to serve an important 'role' in cases in which the existence of discrimination is a disputed issue." Id. at 339 (quoting Mayor of Philadelphia v. Educational Equality League, 415 U.S. 605, 620 (1974)).

30. "Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force [representative of the community]." 431 U.S. at 339 n.20.

31. Id. at 339.

32. "We have repeatedly approved the use of statistical proof, where it reached proportions comparable to those in this case [9% of the lower level employees were of a minority race, but only 0.7% of the upper level employees were of a minority race], to establish a prima facie case of racial discrimination in jury selection cases. Statistics are equally competent in proving employment discrimination." Id. (citations and footnotes omitted). See generally Shoben, Probing the Discriminatory Effects of Employee Selection Procedures with Disparate Impact Analysis Under Title VII, 56 Tex. L. Rev. 1, 42-43 (1977).
The Prima Facie Case

The plaintiff in a Title VII action bears the critical burden of establishing a prima facie case of employment discrimination. In a disparate impact case this consists of establishing the plaintiff as a member of a protected class of individuals, describing the facially neutral employment practice, and demonstrating, usually by means of statistics, the extent of the adverse impact on the plaintiff's class.

A prima facie case of disparate treatment may be established in either of two ways. Under the approach generally used in suits by an individual, the plaintiff must show that he or she was a member of a protected class of individuals, that he or she was denied an actual, existing employment opportunity, that he or she was qualified for that opportunity, and that the opportunity was given to an individual who was not a member of the protected class. Under the approach generally used in class actions, the plaintiffs must show that they are members of a protected group, that they were denied employment opportunities, and that the denials came during a period when the defendant was exhibiting a "pattern or practice" of discrimination against the protected group. In none of the models is it necessary to show discriminatory motive or intent at the prima facie stage.

35. "Opportunity" includes hiring, promotion, or job retention possibilities.
37. Although the phrase "pattern or practice" appears verbatim in § 707(a) of Title VII, it is not a term of art, but is intended to reflect its usual meaning. See International Bhd. of Teamsters v. United States, 431 U.S. 324, 336 n.16 (1977); United States v. Ironworkers Local 86, 443 F.2d 544, 552 (9th Cir.), cert. denied, 404 U.S. 984 (1971).
Issues in Title VII Statistics

The Model

Statistical evidence of discrimination consists of a comparison between an actual employment situation or series of events and a model of what the situation would be, or how the events would have transpired, if the situation or events were determined solely by chance. The existence of the model is theoretical and therefore usually not overtly stated. In Teamsters, for example, when the Court disparaged the fact that eighty percent of the nonwhite employees held lower paying jobs while only thirty-nine percent of the white employees held such jobs, the Court must implicitly have considered a model situation in which the two percentages were closer together. To formulate such a model the Court must have made two assumptions: (1) that the nonwhite employees were as competent to hold the higher paying jobs as the white employees and (2) that racial discrimination was not a factor in determining which employees would receive the higher paying jobs. These two assumptions defined the model to which the actual occurrence was compared.

A common misconception is that the model only suggests a single “ideal” guideline to be tempered by common sense allowances. In Teamsters, for example, although absolute equality between the two percentages would be ideal, common sense dictates that there probably would be some “accidental” disparity between whites and nonwhites even if equal competency existed and no racial discrimination were present. Statisticians term this likely disparity “chance variability.”


40. A comparison of an employment situation at some instant in time, like a snapshot, is termed “static.” A comparison of a job history consisting of several events over an interval of time, like a motion picture, is termed “flow.” See generally Schliem & Grossman, supra note 5, at 1161-65.


42. 431 U.S. at 337-38.


44. Id. at 244-45.
While chance variability indicates that a particular outcome of any system cannot be predicted with certainty, it also indicates that, over many operations, the overall performance of a system becomes predictable if the performance of each operation is determined solely by chance. The "ideal" result ends up as only the most frequent result, the mode within a range of possible results, each with its own probability of occurrence. By adding together the probabilities of successive possible results, it can be determined how often resultant values will be either greater or less than some chosen limit. By choosing the actual result as the limit, it can be determined how often the chance model would have predicted a result greater or less than the actual occurrence.

At some point, termed the "level of significance," it is said that because the model would have produced that high or that low a result only a small percentage of the time, then the actual occurrence was probably determined by something other than pure

45. Id. at 255-60; ORKIN & DROGIN, supra note 41, at 38.
46. See J. FREUND & F. WILLIAMS, ELEMENTARY BUSINESS STATISTICS—THE MODERN APPROACH 42 (3d ed. 1977) [hereinafter cited as FREUND & WILLIAMS]. In a normal distribution, see note 121 infra, the mode is also the average. See FREEDMAN, supra note 43, at 70-73.
47. For example, suppose that the resultant distribution of the number of times a fair coin landed "heads" when tossed 100 times in each of 100 trials was:

<table>
<thead>
<tr>
<th>number of trials</th>
<th>number of heads</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>x x</td>
</tr>
<tr>
<td>9</td>
<td>x x x</td>
</tr>
<tr>
<td>8</td>
<td>x x x x</td>
</tr>
<tr>
<td>7</td>
<td>x x x x x</td>
</tr>
<tr>
<td>6</td>
<td>x x x x x x</td>
</tr>
<tr>
<td>5</td>
<td>x x x x x x x</td>
</tr>
<tr>
<td>4</td>
<td>x x x x x x x x</td>
</tr>
<tr>
<td>3</td>
<td>x x x x x x x x</td>
</tr>
<tr>
<td>2</td>
<td>x x x x x x x x</td>
</tr>
<tr>
<td>1</td>
<td>x x x x x x x x</td>
</tr>
</tbody>
</table>

This distribution shows, for example, that in 6 out of the 100 trials, the number of heads tossed was 55.
48. The distribution, see note 47 supra, also shows that 41 or fewer heads were tossed in 4 out of the 100 trials.
49. If another coin had been tossed 100 times and came up heads 41 times, it could be said that if the coin were fair it would have come up with a greater number of heads 96% of the time.
50. See generally FREEDMAN, supra note 43, at 442; ORKIN & DROGIN, supra note 41, at 115.
chance. The model is then rejected as not being descriptive of the actual system. If a court's assumption of equal competency was not shown to be erroneous, the court would regard the assumption of nondiscrimination as incorrect. Statisticians term the necessary abandonment of the model's nondiscrimination assumption as a "rejection of the null hypothesis." The rejection of the null hypothesis constitutes evidence of discrimination.

The choice of a particular level of significance is subjective. Scientists, to compare the quality of each other's data, have set some standard levels. One standard level is a five percent level of significance. Some commentators suggest that the five percent level of significance was chosen because the early abbreviated statistical tables included that level. A more generous view is that the five percent level was chosen because, being two standard deviations from the average, it is relatively easy to calculate. Most modern statisticians are critical of using five percent or any other level as an absolute standard of significance. Nevertheless, many courts have followed the scientist's path and used the five percent level as a determination of valid statistical evidence, and no court has yet challenged the scientific standard as being inapplicable to legal determinations.

A somewhat greater amount of legal attention has been focused on the subjective decision whether a one-tailed or a two-tailed test is to be used with a particular level of significance. A one-tailed test focuses on how often the model produces a value either greater or lesser than the actual occurrence, while a two-tailed test focuses on how often the model produces a result more

51. See Freedman, supra note 43, at 442.
52. Id. at 444.
53. Id. at 491-92.
55. See text accompanying notes 50-51 supra.
57. The 5% level of significance may also be referred to as its 95% counterpart.
59. Id. at 61.
60. Id. at 494.
62. See note 48 supra.
extreme from the average than the actual occurrence. In discrimination practice, a two-tailed test considers the occurrence of "reverse" discrimination to be as possible as the discrimination at issue, while the one-tailed test does not. From the plaintiff's point of view, because the possibility of reverse discrimination in most cases is remote or is not at issue, one-tailed tests are sufficient.

Some statisticians, however, believe that most uses of one-tailed tests are "data mining" per se. Although the practical difference between the two tests at the five percent level of significance is slight, it can affect the outcome in "borderline" cases. This Note uses both one-tailed and two-tailed tests to illustrate the difference that the use of a particular test can make in close cases.

The Population

The most common factual dispute in Title VII cases involves the question of who should be included in the actual experience that is to be compared to the model, or, more technically, how is the relevant population to be determined. The relevant population is usually defined as all individuals who were available and qualified for the job opportunity.

In discrimination cases involving hiring practices, the relevant population usually comes from the surrounding community and is therefore external to the employer's work force. Issues that can

---

63. Most statisticians would consider the test of significance formulated in note 49, supra, as improper. The proper test in that instance is two-tailed because bias can be in the form of too many heads as well as too few heads. Tosses of 59 or more heads are just as unlikely to occur as tosses of 41 or fewer, and therefore are just as indicative of possible bias. If the five "59 or more" occurrences are added to the four "41 or fewer" occurrences, then the model would have produced a result less extreme than 41 heads only 91% of the time.

64. Braun, supra note 54, at 68-69.


66. "Population" is a term of art in statistics and refers to the entire universe of individuals possessing one or more characteristics in common. Freedman, supra note 43, at 301. See generally Braun, supra note 54, at 62-67.

68. See generally Schles & Grossman, supra note 5, at 1162-65. In ADEA cases, because a large and indeterminable portion of the older persons in a community are not available for work, the requirement of availability usually precludes comparisons between the work force and the community. But cf. Polsteroff v. Fletcher, 452 F. Supp. 17, 22 (N.D. Ala. 1978) (comparing the percentage of executives over the age of 55 in a particular federal
arise in these cases include determining the geographic boundaries and other characteristics of the community. In addition, if the hiring opportunity requires special skills, education, or experience, the relevant population must be restricted further to include only those members of the community who possess these requisites.

If the issue concerns promotions or firings, the relevant population is internal to the employer's work force, consisting of that portion of the work force qualified and available to perform the job. In such cases, determination of the necessary qualifications and which employees possess those qualifications is often at issue.

**The Sample**

The employer's selection of individuals from the population to be either hired, promoted, or fired, statistically is termed the "sample" and is to be compared with the spectrum of possible samples that the chance model generates. The possible samples generated by the model are solely determined by chance, and therefore termed "simple random samples." Certain assumptions must be made regarding the nature of the actual sample in order that it too may be considered random and, as such, comparable to the model's samples.

Of course, actual samples are not random. Employers make each employment decision by considering many factors of competence in addition to the minimal qualifications that determine the population. Most of these additional factors are legitimate. If it is assumed, moreover, that these legitimate factors are independent of race or sex, a selection based solely on legitimate factors would be random with respect to race or sex.

In a Title VII disparate treatment case, the plaintiff contends that one of the factors the employer considered in choosing the

---

71. "Firing" includes "forced retirements," "accepted resignations," and other such euphemisms.
73. See FREZAMAN, supra note 43, at 308; ORKIN & DROGIN, supra note 41, at 7; Braun, supra note 54, at 70 n.64.
sample was race, color, religion, national origin, or sex. These are illegitimate factors per se; the conscious use of any one of them is unlawful discrimination. In a disparate impact case, the plaintiff does not maintain that one of these factors was used consciously by the employer, but that a seemingly neutral factor that the employer did use was irrelevant to competency and unintentionally biased members of the protected class from obtaining the employment opportunity.74 The use of that factor is unlawful discrimination.75 For both disparate impact and disparate treatment cases, by assuming that the use of legitimate factors will not affect the random nature of the sample with respect to race or sex, statistical analysis can evaluate the actual result in terms of the ideal model.

**Title VII Percentage Analysis**

Most Title VII statistics compare the percentage of protected individuals in the actual sample—those individuals hired, promoted, or fired by the employer—with the ideal or expected percentage as predicted by the model, which is the percentage of protected individuals in the population from which the employer chose.76 The employer will maintain that any difference in percentage results solely from chance variability.77 Conversely, the plaintiff will argue that the difference in percentages is attributable to unlawful discrimination.78 In both disparate impact and disparate treatment cases, if the difference is dramatic on its face, complete statistical analysis may not be necessary.79 Many courts, in such instances, have chosen to interpret statistics in their “rough” form.80 But if the difference, while favoring the plaintiff, is not dra-

---

74. See note 21 supra.
75. See text accompanying note 18 supra.
76. For example, if the population consisted of 20% women, the sample would be expected to include approximately 20% women. Although the sample percentage may not be exactly 20%, if the sample is relatively large, see FREEDMAN, supra note 43, at 329-33, it will be close to that figure, and large deviations rarely will occur. See generally Braun, supra note 54, at 76-80.
77. This is the “null hypothesis.” See Braun, supra note 54, at 68-69.
78. This is the alternative hypothesis. Id. In discrimination cases the presumption (null hypothesis) is always the same—that the observed difference is the result of chance and that the defendant did not engage in discriminatory practices. The alternative hypothesis is the plaintiff’s allegation—that the difference was the result of the defendant’s discriminatory practices.
80. See generally SCHLEI & GROSSMAN, supra note 5, at 1184-86.
matic, the statistical model can generate the percentage of samples that would have been less discriminatory if chance were the only determinant. If ninety-five percent or more of the model's samples would have been less discriminatory, most courts will hold that unlawful discrimination is evident, and the prima facie case established.

The Application of Title VII to Age Discrimination

Although Title VII has often been used as a guideline in age discrimination litigation, several courts have cautioned against strained applications of Title VII law in ADEA cases. In Laugesen v. Anaconda Co., the Sixth Circuit used Title VII for guidance in evaluating motions for a directed verdict and in determining causation, yet emphasized:

We do not here decide whether Congress intended that actions under the [ADEA] must invariably be guided by the law applicable to Title VII cases. That the law is embodied in a separate act and has its own unique history [should] at least counsel the examiner to consider the particular problems sought to be reached by the statute.

In Rodriguez v. Taylor, the Third Circuit stated that although Title VII experience was a useful guideline for most age issues, equal competency should not be as automatically presumed in ADEA cases as in Title VII cases. Mastie v. Great Lakes Steel Corp. broadly attacked the use of statistical evidence in ADEA


82. See note 57 supra.
83. See text accompanying note 61 supra.
84. See notes 8-16 & accompanying text supra.
85. 510 F.2d 307 (6th Cir. 1975).
86. Id. at 311-12.
87. Id. at 316.
88. Id. at 312 (footnote omitted).
90. 569 F.2d at 1236-37 (there may be statistically significant correlations between bona fide necessary qualifications and age).
cases, holding that "statistics in age discrimination litigation cannot be used exclusively to establish a prima facie case . . . . [T]here are persuasive reasons, in this court's opinion, for placing less weight on statistical evidence in age discrimination cases."92 It is important, therefore, to recognize the differences between Title VII and ADEA cases that affect the use of statistical evidence in each.

The Impact-Treatment Dichotomy

Until recently, the courts considering ADEA cases have been hesitant to recognize disparate impact fact situations. In Coates v. National Cash Register Co.,93 for example, the defendant employer had introduced a new line of electronic machines to replace its older mechanical models. A training program was instituted in which younger, newly hired employees, rather than existing field service employees, were trained on the new machines. Eventually, the field service employees who could only service the mechanical models were laid off because they did not have training in electronics.94 Although the facts of the case were classically fitted to the disparate impact model,95 the court ruled that, while the layoffs were not intentionally discriminatory, the training program was,96 thereby placing the case within the disparate treatment model.

In the case of Geller v. Markham,97 however, the court did confront the issue of whether a disparate impact claim was valid in age discrimination cases. The Geller court found that the defendant employer's practice of giving preference in hiring to teacher applicants of lesser experience, and hence lesser salary requirements, was facially neutral and not intentionally discriminatory, yet nonetheless had a discriminatory impact on the older applicants.98 The court expressly noted that the distinction between disparate impact and disparate treatment applies to age discrimination as well as to race and sex discrimination,99 and held that the

92. Id. at 1320. The Mastie case generally has not been followed on this point.
94. Id. at 659.
95. Laying off employees who do not possess necessary skills would be the neutral employment practice. See text accompanying note 34 supra.
96. 433 F. Supp. at 661.
98. Id. at 839.
99. Id. at 837.
disparate impact shown in this case made the employer’s “neutral” practice unlawful.\textsuperscript{100} Thus, the distinction between disparate impact and disparate treatment cases and the corresponding notions of necessary discriminatory motive have now been applied to ADEA cases.

The ADEA Prima Facie Case

\textit{Geller} indicates that prima facie cases of disparate impact can be established in ADEA actions in the same manner as in Title VII actions.\textsuperscript{101} Class action prima facie cases of disparate treatment may also be established in the same manner in ADEA actions as in Title VII actions.\textsuperscript{102} There is, however, a very important factor in the disparate treatment model used for most individual claims under Title VII\textsuperscript{103} that, in some jurisdictions, is not necessary for individual prima facie cases under the ADEA.

In suits by individuals under Title VII, the term “protected class” does more than just define the plaintiff. Individual claims of disparate treatment cannot arise under Title VII if the employer has awarded the job to another member of the protected class.\textsuperscript{104} For example, an individual female complainant cannot claim sex discrimination if the opportunity in question was awarded to another female. This is the necessary legal result of regarding all members of the protected class equally, which presents no problem under Title VII because race, sex, and the other Title VII classifications are fixed and discontinuous. Age, however, is a continuous variable.\textsuperscript{105} The question thus becomes: “Should sixty year old employees be regarded as equal to forty year old employees?” This is not to ask whether sixty year old plaintiffs should receive a differ-

\textsuperscript{100} Id. at 839.
\textsuperscript{101} See note 34 & accompanying text \textit{supra}.
\textsuperscript{103} See note 36 & accompanying text \textit{supra}.
\textsuperscript{104} See note 36 & accompanying text \textit{supra}.
\textsuperscript{105} “Continuous and discontinuous” are used here in their mathematical sense. A light switch is an example of a discontinuous function; it is either on or off. Race, color, religion, sex, and national origin are similarly discontinuous; one either belongs or does not belong to the particular class. A light dimmer, however, is an example of a continuous function. It has a theoretically infinite number of settings between off and full brightness. So too, with age, there exists a theoretically infinite number of ages from birth to death.
ent measure of protection than forty year old plaintiffs, but rather, whether it is unlawful to replace, on the basis of age, a sixty year old protected worker with a forty year old protected worker. Following Title VII precedent, some courts have held that this practice is not unlawful.\textsuperscript{108} A number of opinions, however, have recognized that such blind adherence to Title VII is contrary to the spirit of the ADEA.\textsuperscript{107} As a result, in contrast to Title VII, a prima facie case of individual disparate treatment under the ADEA should not require that the job be given to an unprotected individual. One court has stated that even replacement by an individual of equal age may not preclude a claim.\textsuperscript{108} Thus, in ADEA cases the term "protected class" functions only as a determinant of proper plaintiffs.

Statistical Analysis in ADEA Cases

The technique of percentage comparisons of groups as used in Title VII\textsuperscript{109} cases also has been used effectively in ADEA cases.\textsuperscript{110} If discrimination within the protected class is unlawful, however, a comparison of protected to unprotected workers will not reveal whether or not discrimination within the protected class is present.\textsuperscript{111} Recognizing this failing, some courts have drawn lines of comparison at ages tailored to the age groups in which the discrimination is alleged to have occurred in a particular case.\textsuperscript{112} To be


\textsuperscript{109} See notes 76-80 & accompanying text supra.


\textsuperscript{111} For example, a defendant could replace all workers over the age of 60 with workers in their forties, yet the percentage of protected workers employed would remain constant.

\textsuperscript{112} See Marshall v. Sun Oil Co., 605 F.2d 1331, 1333 (5th Cir. 1979) (two comparisons done, one with the employees subdivided at age 50 and the other with the employees subdivided at age 60); Polstorff v. Fletcher, 452 F. Supp. 17, 22 (D. Ala. 1978) (comparison with the employees subdivided at age 55); Mistretta v. Sandia Corp., 15 Empl. Prac. Dec. ¶ 7902, at 6499, 6504-05 (D.N.M. 1977), aff’d sub. nom. EEOC v. Sandia Corp., 23 Empl. Prac. Dec. ¶ 31,175 (10th Cir. 1980) (two comparisons done, one with the employees subdivided at
statistically valid, such group comparisons must involve large sample sizes.\textsuperscript{113} If the disparity between the groups is dramatic,\textsuperscript{114} the harm of missing "within group" discrimination is not noticed.\textsuperscript{115} If the disparity is not dramatic, however, this deficiency may be critical. In addition, the validity of the results is questionable because grouping data for the purpose of analysis may overly dramatize the results. Such arbitrary grouping should be considered data mining.\textsuperscript{116} Any line of comparison in a statistical test must be justifiable independently of the data,\textsuperscript{117} but the only independently justifiable age delineation is the legislative separation of protected and unprotected individuals. Therefore, the best statistical tests for ADEA cases are those that do not employ grouping.

**Statistical Analysis in Age Discrimination Cases**

**Mean Analysis**

The most common method of statistical analysis applicable to age 50 and the other with the employees subdivided at age 60).

\textsuperscript{113} Percentage comparisons of groups are dependent upon the central limit theorem which is valid only for large samples and populations. See notes 122, 131 & accompanying text infra. See also Marshall v. Sun Oil Co., 605 F.2d 1331 (5th Cir. 1979) (sample of 156); Polstorff v. Fletcher, 452 F. Supp. 17 (D. Ala. 1978) (sample of 350); Mistretta v. Sandia Corp., 15 Empl. Prac. Dec. ¶ 7902 (D.N.M. 1977), aff’d sub. nom. EEOC v. Sandia Corp., 23 Empl. Prac. Dec. ¶ 31,175 (10th Cir. 1980) (sample of 306).

\textsuperscript{114} In Sun Oil, 24% of the employees laid off were over 60 years old, but only 3% of the work force was over 60. Marshall v. Sun Oil Co., 605 F.2d 1331, 1333 (5th Cir. 1979). In Polstorff, 30% of the employees over 55 were adversely affected by the reorganization, while only 3.5% of the employees under 55 were so affected. Polstorff v. Fletcher, 452 F. Supp. 17, 22 (D. Ala. 1978). In Mistretta, 81% of the employees over the age of 60 were adversely affected by the reorganization, although only 5.9% of all employees were adversely affected. Mistretta v. Sandia Corp., 15 Empl. Prac. Dec. ¶ 7902, at 6499 (D.N.M. 1979), aff’d sub. nom., EEOC v. Sandia Corp., 23 Empl. Prac. Dec. ¶ 31,175 (10th Cir. 1980).

\textsuperscript{115} Because of the judicial familiarity with percentage comparisons, such group comparisons may even be preferable. See Mistretta v. Sandia Corp., 15 Empl. Prac. Dec. ¶ 7902, at 6501 (D.N.M. 1979), aff’d sub. nom., EEOC v. Sandia Corp., 23 Empl. Prac. Dec. ¶ 31,175 at 17143 (10th Cir. 1980) (judge ignored plaintiff statistician’s Kolmogorov-Smirnov tests in favor of his own percentage analysis; later congratulated on the effort by the appellate court).

\textsuperscript{116} See note 65 supra.

\textsuperscript{117} To illustrate this precept, suppose OXOXXO A OOO B XXX represents twelve employees ordered according to age where line A represents the age of 40, and line B, the age of 50. Further assume that the X's were fired and the O's retained. If the "above 40" group is compared to the "below 40" group (line A), each group suffered 50% layoffs, and no evidence of discrimination is established. If the comparison is made at age 50 (line B), however, then the above 50 group had 100% layoffs, compared to only 33% of the below 50 group, and discrimination seems evident.
age discrimination cases that does not require grouping is a comparison of mean\textsuperscript{118} ages. For example, if an employer randomly selects 50 employees (the sample) out of a possible 500 (the population), the mean age of the selected employees will, to some degree of probability, resemble the mean age of the entire population.\textsuperscript{119} The larger the population and sample size, the closer the resemblance should be.\textsuperscript{120} The various possible means of many random samples taken from the population will always be normally distributed\textsuperscript{121} around the mean of the population.\textsuperscript{122} The expected value of the sample mean is exactly that of the population mean,\textsuperscript{123} but the observed mean (OM) of any sample may not exactly equal the expected mean (EM); the theorem only predicts that it will probably be very close.\textsuperscript{124} The further from the expected mean any observed sample mean is, the more unlikely its occurrence will be.\textsuperscript{125} The likelihood of a particular observed mean occurring can be calculated using the formula

\[
z = \frac{(OM - EM) \times \sqrt{\text{sample size}}}{\text{standard deviation}}\textsuperscript{126}
\]

and the z table.\textsuperscript{127}

\textsuperscript{118} The "mean" is the arithmetic average. See note 159 infra.
\textsuperscript{119} This precept is known as the law of large numbers or the law of averages. See Freund & Williams, supra note 46, at 143; Orkin & Drogin, supra note 41, at 95-98.
\textsuperscript{120} See notes 130-131 & accompanying text infra.
\textsuperscript{121} A "normal" distribution is one in which the frequency of values is highest at the mean and tapers off symmetrically as one moves farther from the mean in either direction. For example, the series 1,2,2,3,3,4,4,5 is roughly a normal distribution about the value "3." The graph of a normal distribution is the familiar "bell-shaped" curve. For another example of a normal distribution, see note 47 supra. See generally Orkin & Drogin, supra note 41, at 67-68; Freedman, supra note 43, at 69-84.
\textsuperscript{122} See Orkin & Drogin, supra note 41, at 85. The statistical theory upon which these precepts are based is known as the central limit theorem. See generally Freund & Williams, supra note 46, at 252; Braun, supra note 54, at 70-71.
\textsuperscript{123} See Freedman, supra note 43, at 255, 373.
\textsuperscript{124} Id.
\textsuperscript{125} See generally Orkin & Drogin, supra note 41 at 85, 95-98. This precept also parallels the example in notes 47-49 supra.
\textsuperscript{126} This is an algebraic simplification of the standard test formula:

\[
z = \frac{\text{observed value} - \text{expected value}}{\text{standard error}}\textsuperscript{126}
\]

\textit{See} Freedman, supra note 43, at 443, where

\[
\text{standard error} = \frac{\sqrt{\text{sample size}} \times \text{standard deviation}}{\text{sample size}}\textsuperscript{126}
\]

\textit{Id.} at 375; Orkin & Drogin, supra note 41, at 129.
\textsuperscript{127} The following is a one-tailed, as opposed to a two-tailed, z table. See notes 63-65.
The application of this formula can be illustrated by the following hypothetical: An employer fires 50 employees (the sample) with a mean age of 38.9 years out of a population of 500 with a mean age of 37.1 years and a standard deviation (SD) of 7 years. If the disparity between the two means results from chance, the probability of its occurrence can be found by substituting these values into the formula, which results in:

& accompanying text supra.

<table>
<thead>
<tr>
<th>z score</th>
<th>%</th>
<th>z score</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>84.14</td>
<td>1.50</td>
<td>93.32</td>
</tr>
<tr>
<td>1.05</td>
<td>86.32</td>
<td>1.55</td>
<td>93.94</td>
</tr>
<tr>
<td>1.10</td>
<td>86.44</td>
<td>1.60</td>
<td>94.52</td>
</tr>
<tr>
<td>1.15</td>
<td>87.50</td>
<td>1.65</td>
<td>95.06</td>
</tr>
<tr>
<td>1.20</td>
<td>88.50</td>
<td>1.70</td>
<td>95.55</td>
</tr>
<tr>
<td>1.25</td>
<td>89.44</td>
<td>1.75</td>
<td>96.00</td>
</tr>
<tr>
<td>1.30</td>
<td>90.32</td>
<td>1.80</td>
<td>96.41</td>
</tr>
<tr>
<td>1.35</td>
<td>91.15</td>
<td>1.85</td>
<td>96.79</td>
</tr>
<tr>
<td>1.40</td>
<td>91.93</td>
<td>1.90</td>
<td>97.13</td>
</tr>
<tr>
<td>1.45</td>
<td>92.65</td>
<td>1.95</td>
<td>97.44</td>
</tr>
</tbody>
</table>

Negative z scores are read as if they were positive.

If the standard deviation of the population is not known and the sample is large enough, the standard deviation of the sample can be used as

\[
SD = \sqrt{\text{average of (deviations from average)}^2}.
\]

For example, to determine the standard deviation of the list 20, 10, 15, and 15, the first step is to calculate the mean:

\[
\text{Mean} = \frac{20 + 10 + 15 + 15}{4} = 15.
\]

The second step is to determine each deviation from the average by subtracting the average from each entry. These deviations are: 5, -5, 0, and 0. Finally, the square root of the average of the squares of the deviations is calculated.

\[
SD = \sqrt{\frac{5^2 + (-5)^2 + 0^2 + 0^2}{4}} = \sqrt{\frac{25 + 25 + 0 + 0}{4}} = 12.5 = 3.54.
\]

\[
SD = \sqrt{\frac{5^2 + (-5)^2 + 0^2 + 0^2}{4}} = \sqrt{\frac{25 + 25 + 0 + 0}{4}} = 12.5 = 3.54.
\]

FREEDMAN, supra note 43, at 63. If the standard deviation of the population is not known and the sample is large enough, the standard deviation of the sample can be used as
From the z table, it is found that, of all possible samples of 50 out of this population of 500, more than 96% would have had a mean age of less than 38.9 years, which surpasses the significance level required by most courts, and thus establishes evidence of discrimination. The important fact to note from this example is that, in cases involving large samples and populations, seemingly slight variances in mean age can be significant.

Large samples and populations, however, are not commonly present in ADEA cases. If the population size is not at least ten times the sample size, the z score must be multiplied by a correction factor that decreases z, lessening the likelihood of significance. More importantly, from the basic z formula, it is apparent that as the sample size decreases, z decreases, even if the disparity between the mean ages remains the same. Recomputing the previous example with a sample size of twenty instead of fifty demonstrates this effect:

\[
z = \frac{(38.9 - 37.1) \cdot \sqrt{20}}{7} = 1.15.
\]

The corresponding percentage from the z table is only 87.5%—insufficient to reject the null hypothesis that the difference resulted from chance.

As the sample size decreases even further, not only does z continue to decrease, but the central limit theorem begins to break down because the distribution of possible sample means is no longer normal. If the age distribution in the population itself were near normal, a t test could be used for samples smaller than twenty; but there is no reason to expect an age distribution to be normal. Generally, samples of less than twenty are inadequate

129. See note 61 & accompanying text infra.
130. The correction factor for these high sample-to-population ratios is:

\[\sqrt{\frac{\text{population size} - \text{sample size}}{\text{population size} - 1}}\]

FREEDMAN, supra note 43, at 331.
131. See ORKIN & DROGIN, supra note 41, at 85, 98.
132. A t-score is calculated using the same formula as a z-score, but different tables are used depending on the sample size. See FREEDMAN, supra note 43, at 463-65; ORKIN & DROGIN, supra note 41, at 132-36; Braun, supra note 54, at 74 n.76, 89 app. A.
133. See note 138 & accompanying text infra.
for mean analysis in ADEA cases. Instead, the analysis of small samples is dependent upon nonparametric analysis.134

Nonparametric Analysis for Cases of Small Sample Size

*Mastie v. Great Lakes Steel Corporation*135

In 1971, Great Lakes Steel Corporation was forced to shut down one of its mills and absorb the foremen from that mill into its operations at another mill. Two of the foremen could not be absorbed and were laid off. Mastie and Seymour, both age 56, claimed that they were chosen for dismissal because of their age. At trial, it was determined that the relevant population consisted of the individuals in Table 1.136

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petosky</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>Mastie*</td>
<td>56</td>
<td>2</td>
</tr>
<tr>
<td>Sawicki</td>
<td>56</td>
<td>3</td>
</tr>
<tr>
<td>Seymour*</td>
<td>56</td>
<td>4</td>
</tr>
<tr>
<td>Eastman</td>
<td>54</td>
<td>5</td>
</tr>
<tr>
<td>Cason</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td>Haskamp</td>
<td>52</td>
<td>7</td>
</tr>
<tr>
<td>Gardner</td>
<td>48</td>
<td>8</td>
</tr>
<tr>
<td>Hanak</td>
<td>47</td>
<td>9</td>
</tr>
<tr>
<td>Ransom</td>
<td>47</td>
<td>10</td>
</tr>
<tr>
<td>Romanowski</td>
<td>43</td>
<td>11</td>
</tr>
<tr>
<td>Taurence</td>
<td>43</td>
<td>12</td>
</tr>
<tr>
<td>Beard</td>
<td>40</td>
<td>13</td>
</tr>
<tr>
<td>Redman</td>
<td>39</td>
<td>14</td>
</tr>
<tr>
<td>Roman</td>
<td>37</td>
<td>15</td>
</tr>
<tr>
<td>Judd</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>Michaelis</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>Dupier</td>
<td>29</td>
<td>18</td>
</tr>
<tr>
<td>Gregoire</td>
<td>28</td>
<td>19</td>
</tr>
</tbody>
</table>


136. *Id.* at 1303.
The average age of the foremen before the shutdown was 45.05; after the shutdown, the average age was 43.76. The court held that this difference of 1.29 years was not sufficient evidence to indicate discrimination. However, even if the defendant had a blatantly discriminatory policy under which it intentionally released the very oldest employees, thus laying off Petosky rather than Seymour or Mastie, the average age still would have decreased only 1.58 years, demonstrating that even the most blatant discrimination would not have appeared significant in a comparison of means. This example demonstrates the potentially deceptive use of conventional statistical analysis when a small sample size is taken from a population that is not normally distributed.

Ironically, after mishandling these statistics, the Mastie court went on to make general pronouncements on the use of statistics in age discrimination cases.

When the sample size is too small for mean comparisons, a line drawing analysis might be used with some justification. Suppose that a line were drawn in Table 1 just below the four oldest employees (the dashed line). Both released employees came from this subgroup comprised of the four oldest employees. If all of the names were thrown into a hat and then two names were drawn out one at a time, the chances of one of the subgroup being picked are four out of nineteen, or 4/19. For the second draw, there would be only eighteen names in the hat, three of them from the subgroup. The chances of one of the subgroup being picked on the second draw are 3/18, or 1/6. To determine the chance that one of the subgroup would be picked on both draws, the chances are multiplied together: 4/19 x 1/6 = 4/114 = 3.51%. Thus, 96.49% of all possible choices of two would have included at least one person outside the subgroup; this evidence would be statistically signifi-

137. Id. at 1319-20.

138. The most obvious sign that a population is not normally distributed is the incidence of multiple individuals at the extreme ends of the distribution. This situation is demonstrated in Mastie by the bunching of employees at the “extreme” ages of 33, 54, and 56. See note 121 supra.

139. See note 92 & accompanying text supra.

140. Only totally independent chances can be multiplied together. In this instance, the second draw was made independently of the first by subtracting 1 from the pool of 19, but creating independence is not always as simple as this example would indicate. See Freedman, supra note 43, at 212-17; Orkin & Drogin, supra note 41, at 42-46.

141. This test is one-tailed because the samples consisting of both draws coming from a subgroup of the four youngest individuals were considered less, rather than equally, dis-
cant in most courts. The important fact to be noted in this analysis is that it does not use the law of large numbers or the central limit theorem in any way, and therefore the small sample limitation involved with those precepts and their corresponding tests is not present.

The above conclusion, which was reached by multiplying 4/19 by 3/18, works well for simple cases, but for more complex examples, combination theory must be used. Combination theory is used to solve problems such as: How many distinct groups of two can be chosen from a group of four? From a group of four things numbered 1,2,3,4, there are six possible distinct groups of two:

1 and 2, 1 and 3, 1 and 4,
2 and 3, 2 and 4,
3 and 4.

This common sense listing is feasible for small numbers, but is somewhat overwhelming if used to determine, for example, how many groups of two can be chosen from a group of nineteen. Fortunately, mathematicians have proved that the number of groups of size “n” that can be chosen out of a larger group of size “m” is

\[
m! \div n! (m-n)!
\]

Testing this formula on the simple example of two items chosen from a group of four:

\[
\frac{4!}{2! (4-2)!} = \frac{4!}{2! 2!} = \frac{(4 \times 3 \times 2 \times 1)}{(2 \times 1) \times (2 \times 1)} = \frac{24}{4} = 6.
\]

Applying the formula to the number of groups of two that can be chosen from a group of nineteen, which was the size of the group in Mastie, yields:

\[
\frac{19!}{2! (19-2)!} = \frac{19!}{2! 17!} = \frac{19 \times 18}{2 \times 1} = \frac{342}{2} = 171.
\]

criminatory. See notes 64-65 & accompanying text supra.

142. See text accompanying note 119 supra.
143. See notes 122-23 & accompanying text supra.
144. The tests referred to are the z test, see note 126 supra, and the t test, see note 132 supra.
145. “Distinct” means that the order of the group is not considered. For example, the group (2,3) is not distinct from the group (3,2). “Groups,” as used hereafter, will refer to distinct groups.
146. This is also termed the “binomial coefficient.” See generally FREEDMAN, supra note 43, at 231-36. The symbol ! means factorial. By definition, A! = A x (A - 1) x (A - 2) x (A - 3) … 1. Thus 8! = 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1.
Thus, in *Mastie* there were 171 total possible groups of two that could have been chosen out of the nineteen men. Six of those possibilities could have had both choices in a subgroup of the four oldest employees. The percentage of possible samples in which both choices came from the subgroup of four equaled:

\[
\frac{6}{171} = .0351 = \frac{4}{19} \times \frac{3}{18}.
\]

As in any line drawing, subgrouping done for the purpose of statistical analysis should be justified independently of the data.\(^{147}\) It is important to determine how the subgroup was chosen. In this case, the line distinguishes those instances of discrimination that would be equal to or greater than that represented by the dismissal of Mastie and Seymour. Any choice of two within the subgroup would have represented discrimination of at least this magnitude. Therefore, 3.51% of all possible cases would have been equally or more discriminatory. Although this justification is not totally independent of the data, it is not easily attacked.

The necessity for any justification is spared, however, if a probabilistic test that does not require subgrouping is used, such as the Mann-Whitney, or "rank sum," test.\(^{148}\) Suppose the nineteen employees in *Mastie* are sequentially ranked according to age as shown in Table 1.\(^{149}\) These rank numbers, instead of names, could be put on the slips of paper going into the hat. If the rank numbers on the two slips drawn out of the hat were added, their "rank sum" could be as low as 3 (1 + 2), as high as 37 (19 + 18); or any number in between. There could be 171 different possible

---

147. See note 65 & accompanying text supra.


149. This ranking assumes that the exact age of each person is known, as if, for example, Taurence were some months or days older than Romanowski. If the exact age is not known, then those who have equal ages should receive the average rank for their age. Taurence and Romanowski, for instance, would each have a rank of \((11 + 12)/2 = 11\frac{1}{2}\). Seymour, Sawicki, and Mastie would each have a rank of \((2 + 3 + 4)/3 = 3\).
combinations drawn. However, some of the sums would be more likely to arise than others because, although there is only one way to draw either of the extreme values of 3 or 37, there are, for example, four possible ways to draw the rank sum of 10. In fact, the probabilities for the various possible rank sums are normally distributed around a mode of 20, the value exactly midway between the extreme endpoints of 3 and 37. This distribution represents the operation of a chance model of two randomly chosen layoffs.

Table 2 shows, for populations up to size thirty, the critical points for a two-tailed rank sum comparison of two groups for a 5% level of significance. A rank sum value less than or equal to the appropriate critical point surpasses the significance level and establishes a prima facie case of discrimination.

Table 2

<table>
<thead>
<tr>
<th>n₁ (smaller group)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>11</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>12</td>
<td>18</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>21</td>
<td>29</td>
<td>38</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>8</td>
<td>15</td>
<td>22</td>
<td>31</td>
<td>40</td>
<td>51</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>9</td>
<td>15</td>
<td>23</td>
<td>32</td>
<td>42</td>
<td>53</td>
<td>65</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>9</td>
<td>16</td>
<td>24</td>
<td>34</td>
<td>44</td>
<td>55</td>
<td>68</td>
<td>81</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>10</td>
<td>17</td>
<td>26</td>
<td>35</td>
<td>46</td>
<td>58</td>
<td>71</td>
<td>85</td>
<td>99</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>10</td>
<td>18</td>
<td>27</td>
<td>37</td>
<td>48</td>
<td>60</td>
<td>73</td>
<td>88</td>
<td>103</td>
<td>119</td>
<td>137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>11</td>
<td>19</td>
<td>28</td>
<td>38</td>
<td>50</td>
<td>63</td>
<td>76</td>
<td>91</td>
<td>106</td>
<td>123</td>
<td>141</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>11</td>
<td>20</td>
<td>29</td>
<td>40</td>
<td>52</td>
<td>65</td>
<td>79</td>
<td>94</td>
<td>110</td>
<td>127</td>
<td>145</td>
<td>164</td>
<td>185</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>12</td>
<td>21</td>
<td>31</td>
<td>42</td>
<td>54</td>
<td>67</td>
<td>82</td>
<td>97</td>
<td>114</td>
<td>131</td>
<td>150</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>12</td>
<td>21</td>
<td>32</td>
<td>43</td>
<td>56</td>
<td>70</td>
<td>84</td>
<td>100</td>
<td>117</td>
<td>135</td>
<td>154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>13</td>
<td>22</td>
<td>33</td>
<td>45</td>
<td>58</td>
<td>72</td>
<td>87</td>
<td>103</td>
<td>121</td>
<td>139</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>13</td>
<td>23</td>
<td>34</td>
<td>46</td>
<td>60</td>
<td>74</td>
<td>90</td>
<td>107</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>14</td>
<td>24</td>
<td>35</td>
<td>48</td>
<td>62</td>
<td>77</td>
<td>93</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>6</td>
<td>14</td>
<td>25</td>
<td>37</td>
<td>50</td>
<td>64</td>
<td>79</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td>15</td>
<td>26</td>
<td>38</td>
<td>51</td>
<td>65</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>6</td>
<td>15</td>
<td>27</td>
<td>39</td>
<td>53</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>16</td>
<td>28</td>
<td>40</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>16</td>
<td>28</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td>17</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

150. See text following note 144 supra.
151. The four combinations are: 9+1, 8+2, 7+3, and 6+4.
152. Verdooreen, Extended Tables of Critical Values for Wilcoxon's Test Statistic, 50 Biometrika 177 (1963), reprinted in White, The Use of Ranks in a Test of Significance for Comparing Two Treatments, 8 Biometrics 33, 37 (1952). Other Mann-Whitney/Wilcoxon
This table gives the critical points for the lowest possible rank sum of \( n_1 \), which means first that \( n_1 \) must be less than \( n_2 \),\(^{154}\) and second that the rank sum of \( n_1 \) must be the lesser sum based on either older-to-younger ordering or younger-to-older ordering. For *Mastie*, the smaller group (\( n_1 \)) is the group laid off, and the smaller \( n_1 \) rank sum is attained by ordering oldest to youngest, as done in Table 1. Therefore, \( n_1 = 2, n_2 = 19 - 2 = 17 \), and the rank sum of \( n_1 = 2 + 4 = 6 \). According to the table, the 5% critical point for \( n_1 = 2 \) and \( n_2 = 17 \) is 5, which means that the *Mastie* evidence was not quite within 5% significance. This result seems to conflict with the subgrouping method, which placed the result as within 5% significance, unless it is considered that the rank sum table shown is two-tailed, while the other calculation was one-tailed,\(^{155}\) illustrating the difference that the choice of a one-tailed or two-tailed test can make on borderline significance.

*Stringfellow v. Monsanto Co.*

The Mann-Whitney test is clearly superior to the subgrouping methods when the entire sample does not come from the extreme end of the age spectrum. *Stringfellow v. Monsanto Co.*\(^{156}\) illustrates this more common fact situation. Table 3 shows the employees of Monsanto’s manufacturing department.\(^{157}\)


153. For larger samples and populations a z test can be used. See W. Hays & R. Winkler, *Statistics: Probability, Inference and Decision* 829 (2d ed. 1975).

154. Therefore, \( n_1 \) will be either the chosen or the unchosen group, whichever contains the least number of people.

155. See note 141 *supra*. The one-tailed critical rank sum for \( n_1 = 2 \) and \( n_2 = 17 \) is 6. C. Kraft & C. Van Eeden, *A Nonparametric Introduction to Statistics* 214 (1968). Thus using a one-tailed Mann-Whitney test, the result would have been found to be significant. However, because the Mann-Whitney test is not based upon line drawing, even if the choice of tails is consistent with that of a line drawing analysis, the resultant probability values will not be exactly the same. This result occurs because the Mann-Whitney test reveals the "within group" discrimination that the line drawing concealed.


Table 3

* = laid off

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns</td>
<td>*</td>
<td>59</td>
</tr>
<tr>
<td>Hatch</td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>McCune</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>Dumas</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>McCroskey</td>
<td>*</td>
<td>57</td>
</tr>
<tr>
<td>Phillips</td>
<td>*</td>
<td>57</td>
</tr>
<tr>
<td>Stringfellow</td>
<td>*</td>
<td>56</td>
</tr>
<tr>
<td>Faulkner</td>
<td>*</td>
<td>56</td>
</tr>
<tr>
<td>Wells</td>
<td>*</td>
<td>55</td>
</tr>
<tr>
<td>McDonald</td>
<td>*</td>
<td>55</td>
</tr>
<tr>
<td>Kundert</td>
<td></td>
<td>54</td>
</tr>
</tbody>
</table>

---------- median ----------

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuberger</td>
<td></td>
<td>54</td>
</tr>
<tr>
<td>Wilkinson</td>
<td>*</td>
<td>53</td>
</tr>
<tr>
<td>Rhoades</td>
<td></td>
<td>53</td>
</tr>
<tr>
<td>Moorhead</td>
<td></td>
<td>51</td>
</tr>
<tr>
<td>Lynn</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>Cunningham</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Kjeldgaard</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>Post</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>Franks</td>
<td>*</td>
<td>45</td>
</tr>
<tr>
<td>Cameron</td>
<td>*</td>
<td>43</td>
</tr>
<tr>
<td>Boldine</td>
<td></td>
<td>39</td>
</tr>
</tbody>
</table>

The average age before the shutdown was 52.5 and the average age after the shutdown was 51.5, a decrease in average age of one year. As in Mastie, the change in average age does not appear significant. Inspection of the list itself, however, gives the impression that older workers bore the brunt of the layoffs. The problem with applying the same line drawing technique utilized in Mastie is also apparent. The laid-off employees are scattered throughout the age spectrum rather than clustered at one end. Any subgrouping would probably be attacked as data mining.

The Mann-Whitney test, however, is still applicable. In this case, as in Mastie, the smaller \( n_1 \) is the group laid off, which makes \( n_2 \) equal to 12 (22 — 10). The smaller rank sum of \( n_2 \) is again attained by ordering from oldest to youngest, with the sum equaling 100. Given the data values in Stringfellow, the two-tailed critical
sum for a 5% level of significance is 85. Because 100 is not less than or equal to the critical sum value of 85, most courts would not consider the statistical evidence as valid.

A line drawing analysis could be achieved in this instance by the use of the "median." Returning to Table 3, the preshutdown median (dashed line) is 54; eleven values are higher and eleven lower. Seven employees from the group older than the median were released, but only three from the group younger than the median lost their jobs.

As in Mastie, the problem can be modeled by assuming that if all of the employees were equally competent, the layoffs could have been determined by drawing the names from a hat. If the names of the eleven younger employees were written on red slips of paper, the names of the eleven older employees written on blue slips of paper, and all the slips placed into the same hat, what is the probability that a random selection of ten names would contain seven or more on blue slips of paper? Because the actual selection had seven names in the older group, the selection of seven or more individuals in the blue group represents age discrimination that would be equal to or greater than that presented in Stringfellow.

The total number of groups of 10 that could have been drawn out of 22 would equal all of the possible samples that could have been drawn. According to the formula for combinations, this would equal:

\[
\frac{22!}{10! \times 12!} = 646,646.
\]

The percentage of those total possibilities that could have split 7 blue to 3 red, plus those that could have split 8 blue to 2 red, plus those that could have split 9 blue to 1 red, plus those that could have split 10 blue to 0 red would equal the percentage of cases in which the discrimination was equal to or greater than that present

158. See text accompanying note 152 supra.

159. It is important not to confuse "median" with "mean." The mean is a weighted average, and the median is the middle value, with 50% of the values lower and 50% higher. For example, the median of the series 1,2,3,4,5, is 3, with 50% of the values lower (1 and 2) and 50% of the values higher (4 and 5). The mean for this series is also 3 (15/5). If the series were 1,2,3,4,10, however, the mean would be 4 (20/5), although the median would still be 3 (1 and 2 lower, 4 and 10 higher).

160. The seven or more names on red slips of paper (the younger half) are considered to be less discriminatory, making this a one-tailed test. See notes 62-65 & accompanying text supra.

161. See note 146 & accompanying text supra.
in the *Stringfellow* fact situation.

To find the number of cases that could have split 7 to 3, the number of groups of 7 that could be chosen from 11 names in the blue half first must be calculated according to the formula for combinations.

\[
\frac{11!}{7! \times 4!} = \frac{11 \times 10 \times 9 \times 8}{4 \times 3 \times 2} = \frac{11 \times 10 \times 3}{1} = 330.
\]

Each of these 330 cases could be matched with any one of the possible groups of three from the eleven in the red half, which equals:

\[
\frac{11!}{8! \times 3!} = 165.
\]

The total number of possible combinations of the 7 to 3 split would be 330 x 165 = 54,450. Similarly, the total number of possible 8-2 splits equals:

\[
\frac{11!}{8! \times 3!} \times \frac{11!}{9! \times 2!} = 9,075.
\]

The number of 9-1 possibilities equals 605 and the number of 10 to 0 possibilities equals 11.\(^{162}\) The total of all splits 7 to 3 or worse is 54,450 + 9,075 + 605 + 11 = 64,141. The ratio of these groups to all of the possible groups of 10 chosen from 22 would be

\[
\frac{64,141}{646,646} = .099 \text{ or } 9.9\%.
\]

Thus, 91.1\% of all possible selections would have been less discriminatory. This result is consistent with the Mann-Whitney analysis insofar as neither shows significance.

The weak point of this example, in addition to its failure to detect discrimination within each side of the median, is its linedrawing. The median is an arbitrary but standard dividing line used throughout statistics. Use of the median is defensible, but only on the basis of convention and simplicity. Other than the Mastie-type subgrouping,\(^ {163}\) the median may be the only type of small sample linedrawing that is defensible against charges of data mining.

\---

162. If application of the formula gives a number that is sufficiently low, it is a good idea to verify the result with a common sense approach. In the case of the 10 to 0 split, it is easy to determine how many groups of 10 can be chosen from the 11 people in the older half if one realizes that each group of 10 leaves out only one of the 11. There would thus be 11 possible groups. Common sense verification gives confidence in the chosen formula and renders the problem less abstract.

163. See text following note 139 *supra.*
Conclusion

Title VII has laid the foundation for the use of statistical evidence in ADEA cases. But the courts have recognized that, unlike race or sex discrimination, age discrimination can take place within the protected class. Hence, the percentage comparisons used in Title VII cases are inapplicable to age cases, and most other types of data grouping are difficult to justify. Statistical analysis in ADEA cases is most appropriately done on an individual age basis. For large samples, this is best accomplished with mean analysis; for small samples, the Mann-Whitney test is particularly useful.