1978

Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII

Elaine W. Shoben
University of Nevada, Las Vegas -- William S. Boyd School of Law

COMMENT

DIFFERENTIAL PASS-FAIL RATES IN EMPLOYMENT TESTING: STATISTICAL PROOF UNDER TITLE VII

Elaine W. Shoben *

In this Comment, Professor Shoben advocates the use of a statistical technique — a test of the difference between independent proportions — to assess the substantiality of differences in pass rates among various groups on employment tests, in order to facilitate determination of disproportionate impact under title VII of the Civil Rights Act of 1964. She then compares this method with the procedure adopted in the Federal Executive Agency Guidelines on Employee Selection Procedures and suggests several flaws in the latter approach.

EMPLOYMENT selection procedures based upon test scores are subject to scrutiny under title VII of the Civil Rights Act of 1964 [1] if the pass-fail rates for groups defined by sex, race, or ethnicity are substantially different. A plaintiff may establish a prima facie violation of the statute by showing that a disproportionate impact upon these groups results from differences in test pass rates.[2] If this showing is made, the burden then shifts to the defendant to demonstrate the necessity of the test as a valid predictor of job performance.[3] The use of an un-

* Assistant Professor of Law, University of Illinois. A.B., Barnard, 1970; J.D., Univ. of California, Hastings College of Law, 1974.

[1] 42 U.S.C. §§ 2000e to 2000e–17 (1970 & Supp. V 1975). Title VII provides, in § 703(a)(1), that it is an unlawful employment practice for an employer to "fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin." 42 U.S.C. § 2000e–2(a)(1) (1970).

[2] Dothard v. Rawlinson, 97 S. Ct. 2720, 2726 (1977); Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975); Griggs v. Duke Power Co., 401 U.S. 424, 431–32 (1971). Title VII provides that employment tests may be administered to job applicants if they are not "used" to discriminate against a group protected by the Act. Civil Rights Act of 1964, § 703(h), 42 U.S.C. § 2000e–2(h) (Supp. V 1975). A test which has the effect of disproportionately excluding one racial group is a test "used" to discriminate within the meaning of § 703(h) unless the user can demonstrate the business necessity of the test requirement. Griggs v. Duke Power Co., 401 U.S. at 433–36; *cf.* Washington v. Davis, 426 U.S. 229 (1976) (constitutional challenge on equal protection grounds to the use of a test with discriminatory impact requires proof of intent).

[3] *See* cases cited note 2 *supra*. Requirements for test validation are set forth in the EEOC guidelines, 29 C.F.R. §§ 1607.4–.9 (1976). Test validation is also addressed by guidelines prepared by the Departments of Justice and Labor and the Civil Service Commission, Federal Executive Agency Guidelines on Em-

validated test which has the effect of disproportionately excluding a group protected by title VII is an unlawful employment practice under the Act.

The practical problem which has emerged with respect to the plaintiff's evidentiary burden in testing cases is how to determine whether a "substantial" difference exists in the pass rates of two groups. In order to resolve this dilemma, some federal administrative agencies [4] and various courts [5] have looked to numerical disparities in test results. However, they have typically relied solely on intuitive assessments of the substantiality of these disparities in particular cases.[6] This approach is of questionable validity. A single group of black and white applicants who take a test is only a sample of the relevant population of blacks and whites.[7] Title VII, however, requires a determination of whether a test would have a discriminatory impact [8] on the

---

ployee Selection Procedures, 41 C.F.R. §§ 60-3.5 to .7 (1977) [hereinafter cited as Agency Guidelines].

[4] The Agency Guidelines, *supra* note 3, have adopted a crude rule of thumb for determining whether adverse impact exists in a test or other selection process. *See* p. 805 *infra*.

[5] *See, e.g.*, Firefighters Inst. v. City of St. Louis, 549 F.2d 506, 510 n.4 (8th Cir. 1977); United States v. City of Chicago, 549 F.2d 415, 429 (7th Cir. 1977); Chicano Police Officers Ass'n v. Stover, 526 F.2d 431, 434-35 (10th Cir. 1975); Smith v. Troyan, 520 F.2d 492, 498 (6th Cir. 1975); United States v. Georgia Power Co., 474 F.2d 906, 912 n.5 (5th Cir. 1973); Bridgeport Guardians, Inc. v. Bridgeport Civil Serv. Comm'n, 482 F.2d 1333, 1335 (2d Cir. 1973); Chance v. Board of Examiners, 458 F.2d 1167, 1171-73 (2d Cir. 1972); Castro v. Beecher, 459 F.2d 725, 729 (1st Cir. 1972); Hill v. Western Elec. Co., 12 Fair Empl. Prac. Cas. 1175, 1180 (E.D. Va. 1976).

[6] *See* cases cited note 5 *supra*.

[7] The relevant population in employment discrimination cases consists of those potential job applicants residing in the geographic community surrounding the employer. *See* Hazelwood School Dist. v. United States, 97 S. Ct. 2736, 2743-44 (1977); Green v. Missouri Pac. R.R. Co., 523 F.2d 1290, 1294 (8th Cir. 1975); United States v. Ironworkers Local 87, 443 F.2d 544, 551 n.19 (9th Cir. 1971). *But see* League of United Latin American Citizens v. City of Santa Ana, 12 Fair Empl. Prac. Cas. 651, 668 (C.D. Cal. 1976) (employer's active recruitment policy outside city would not push out the limits of the relevant geographic area to dilute the percentage of minorities in the pool). *See also* Note, *Employment Discrimination: Statistics and Preferences Under Title VII*, 59 VA. L. REV. 463, 469-70 (1973). This group may be limited in some cases to those persons in the surrounding community who possess the prerequisite skills for the job in question. The relevant market for a school teaching job may be limited, for example, to those people in the community who have teaching certificates. *See* Hazelwood School Dist. v. United States, 97 S. Ct. 2736, 2742 n.13 (1977); International Bhd. of Teamsters v. United States, 431 U.S. 324, 339 n.20 (1977). *See also* Mayor of Philadelphia v. Educational Equality League, 415 U.S. 605 (1974).

[8] The terms "discriminatory impact," "disparate impact," "adverse impact," and "adverse effect" are used as synonyms in title VII cases. *See* B. SCHLEI & P. GROSSMAN, EMPLOYMENT DISCRIMINATION LAW 73 n.42 (1976). They are all based on the *Griggs* concept of a requirement which "operate[s] to disqualify

population from which the sample is drawn.[9] Statistical analysis provides a method whereby reliable inferences can be made about this population, based on the performance of the particular sample.

The use of statistical inference in discrimination cases is not unprecedented. The Supreme Court has recently recognized that statistical analysis is a helpful tool for examining discriminatory effect in at least two circumstances. In *Castaneda v. Partida*,[10] the Court acknowledged the probative value of probability theory in determining the significance of the particular underrepresentation of Mexican-Americans on Texas grand juries.[11] The Court also approved the application of statistical analysis in *Hazelwood School District v. United States*,[12] a title VII employment discrimination action. In that case, Justice Stewart recognized the efficacy of probability analysis in determining the likelihood that the racial composition of the employer's work force would deviate to a certain degree from that of the relevant population.[13]

Unfortunately, the particular statistical technique employed by the Court in *Castaneda* and *Hazelwood* cannot be applied to cases concerning the significance of differences in pass-fail rates

---

Negroes at a substantially higher rate" than whites. Griggs v. Duke Power Co., 401 U.S. 424, 426 (1971).

Title VII claims based upon "discriminatory impact" should be distinguished from claims of "unequal" or "disparate" treatment. "Disparate treatment" refers to an employer's unequal policy or practice which differentiates solely on the basis of race, color, religion, sex, or national origin. On the other hand, "disparate impact" refers to the discriminatory effect of facially neutral employment practices such as tests. *See* International Bhd. of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977). *See generally* Blumrosen, *Strangers in Paradise:* Griggs v. Duke Power Co. *and the Concept of Employment Discrimination*, 71 MICH. L. REV. 59 (1972). If a title VII plaintiff can prove disparate treatment in a particular employment practice, it is unnecessary to use statistics to demonstrate the effect of that practice.

[9] *See* Dothard v. Rawlinson, 97 S. Ct. 2720, 2727 (1977).

[10] 430 U.S. 482 (1977).

[11] *Id.* at 496 n.17.

[12] 97 S. Ct. 2736 (1977).

[13] *Id.* at 2743 n.17. *See also id.* at 2736, 2747 n.5 (Stevens, J., dissenting).

The *Castaneda* decision was inspired, in part, by a 1966 article in the *Harvard Law Review* in which Michael Finkelstein urged that statistical inferences based on the binomial distribution would assist the evaluation of claims that racial minorities are unconstitutionally excluded from grand juries. Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases*, 80 HARV. L. REV. 338 (1966). This analysis was extended, in *Hazelwood*, to employment discrimination cases in which the composition of the employer's work force is compared with that of the population in the surrounding geographic area. Hazelwood School Dist. v. United States, 97 S. Ct. 2736, 2743 n.17 (1977).

in employment tests. The use of the binomial test is appropriate when evaluating the likelihood of a result composed of a series of events, each with only two possible outcomes, such as the selection of either a black or a white from the relevant population of school teachers.[14] In such a case the binomial allows a comparison to be made between the racial composition of the sample selected and the known racial composition of the population. When a sample of this dichotomous population is subjected to a test or other employment selection procedure which itself produces a second dichotomy, there are now four relevant categories — *e.g.*, two racial groups each composed of passers and failers. The relevant legal question in such a case is whether there is a substantial difference in the population pass rates of the two groups. A statistical test is needed which can be used to evaluate the significance of the sample pass rates when information about population pass rates is not available.

This Comment will propose that the discriminatory effect of employment tests in title VII cases be measured by the use of a related statistical procedure known as testing the difference between independent proportions. Through such an analysis, reliable inferences can be drawn from sample data about the effect of employment tests upon the groups constituting the relevant pool of potential job applicants. This statistical procedure will then be compared with the rule recently set forth in the Federal Executive Agency Guidelines on Employee Selection Procedures.[15] The results obtained under these two approaches will be contrasted to demonstrate the disutility of those Guidelines.

---

[14] *See* F. MOSTELLER, R. ROURKE & G. THOMAS, PROBABILITY WITH STATISTICAL APPLICATIONS 130–37 (2d ed. 1970) [hereinafter cited as MOSTELLER]. The use of the binomial test requires the assumption that the probability of selecting a member of either group remains unchanged throughout the selection process. This will be the case when selection is made from a theoretically infinite population or from a finite population with replacement. For example, the probability of selecting a woman from an infinite population which is 50% female will remain .5 with each selection of either a male or a female. When a population is very large, it can safely be assumed that this requirement is satisfied even though the population is finite. *See* Finkelstein, *supra* note 13, at 353–54. The assumption of an unchanging probability of selection between the groups in question over time is considered satisfied when dealing with a geographic population, such as a county or city, whose racial composition is deemed constant between census counts. *See* Castaneda v. Partida, 430 U.S. 482, 486 & n.6 (1977) (ratio of Mexican-Americans in county taken from census and assumed constant); Hazelwood School Dist. v. United States, 97 S. Ct. 2736, 2743–44 (1977) (census data used to determine racial ratio of teachers, subject to appropriate geographic limitations).

[15] *See* Agency Guidelines, *supra* note 3, § 60–3.4.

I. Testing the Difference Between Independent Proportions

A. A Hypothetical Problem

An example of the typical problem posed in a title VII test requirement case is as follows:

Jean Taylor applies for an assembly line job with the Acme Company. The only requirement for the job is that the applicant score above 100 on a test designed to measure general intelligence and personality characteristics desirable for worker cooperation. No experience or other qualifications are necessary; the company provides on-the-job training for all new employees.

Taylor, who is black, scores below 100 on the required test and is rejected for the job. After exhausting administrative remedies, she sues Acme under title VII on the ground that the test has a discriminatory impact on blacks. The complaint alleges that only 43% of the one hundred blacks taking the test scored above 100, whereas 51% of the two hundred whites scored above that figure.

The company's answer admits that 43% of the blacks and 51% of the whites scored above 100 on this test the last time it was administered, but denies that the test has a discriminatory impact on blacks. Acme also pleads the job-relatedness of the test.

The court must decide on the basis of the admitted difference in sample pass rates whether the test has a discriminatory impact on the black population and thus whether the burden of validation should shift to the employer.[16]

In order to evaluate what the evidence of percentage pass rates reveals about the impact of the test, it is first necessary to identify the groups upon which an impact is legally relevant. The pass rates are known, in this hypothetical, only for those blacks and whites who actually took the test. But the focus in title VII cases is on the discriminatory impact the test would have on all blacks and whites in the relevant geographic area who could have taken the test.[17] It is this group which constitutes

---

[16] On current issues concerning test validation, see Reiter, *Compensating for Race or National Origin in Employment Testing*, 8 Loy. Chi. L.J. 687 (1977); White & Francis, *Title VII and the Masters of Reality: Eliminating Credentialism in the American Labor Market*, 64 Geo. L.J. 1213 (1976); Johnson, *Albemarle Paper Company v. Moody: The Aftermath of Griggs and the Death of Employee Testing*, 27 Hastings L.J. 1239 (1976); Note, *Testing for Special Skills in Employment: A New Approach to Judicial Review*, 1976 Duke L.J. 596; Note, *Employment Testing and Proof of Job-Relatedness: A Tale of Unreasonable Constraints*, 52 Notre Dame Law. 95 (1976).

[17] *See* note 7 *supra*.

the relevant *population* for purposes of assessing the discriminatory effect of the test; those who actually took the test form only a *sample* of this group.

If population figures are available, then no further inquiry into the impact of the test is necessary, because the legally relevant impact is defined in terms of that population.[18] This was the situation in *Dothard v. Rawlinson*,[19] where the Supreme Court invalidated a minimum height requirement for prison guards which was found to discriminate illegally against women. National height statistics showed that the five-foot, two-inch requirement would exclude one-third of the women in the United States, but only about one percent of the men.[20] The Court shifted the burden of proof to the defendant to demonstrate job-relatedness because of the disproportionate impact evidenced by the known population statistics; it did not consider what percentage of women who had actually applied for the job of prison guard had been excluded.[21]

In many cases, including the Jean Taylor hypothetical, only *sample* data will be available. Disparate performances in a sample drawn from the relevant population may not, however, justify the conclusion that the test has a discriminatory impact on that population as a whole. Although blacks had a 43% and whites a 51% pass rate in the hypothetical test, it is possible that there is no difference in pass rates between blacks and whites in the relevant population. In another sample, blacks may pass at a higher rate than whites. When only sample data are available, it is necessary to employ some process of inference — in other words, a statistical test — in order to determine whether the difference that appeared in the sample would occur by chance if in fact there would be no difference in the pass rates of the two races in the relevant population.[22]

---

[18] *See* pp. 794–95 *supra*.

[19] 97 S. Ct. 2720 (1977).

[20] *Id.* at 2727.

[21] *Id.* at 2727–28; *see* Griggs v. Duke Power Co., 401 U.S. 424, 430 n.6 (1971) (population statistics with respect to high school diploma requirement determined from statewide census data on black and white high school graduates). *Compare* Johnson v. Goodyear Tire & Rubber Co., 491 F.2d 1364, 1371 (5th Cir. 1974) (statewide rather than citywide census data on high school education by racial groups used in employment discrimination case on ground that relevant population of potential job applicants could not be limited to those already living in the city), *with* Pettway v. American Cast Iron Pipe Co., 494 F.2d 211, 237 (5th Cir. 1974) (relevant population limited to employees in assessing impact of requirement of high school diploma for promotion).

[22] Inductive statistics proceeds by hypothesis testing in the manner described in the text. A hypothesis is stated, and then the evidence is examined to see whether it is consistent with the stated hypothesis. A sample result is "not consistent" with the hypothesis when the probability of obtaining the result by

B. A Statistical Test of the Significance of Differences Between Independent Proportions

There would be two possible outcomes if all blacks and all whites (the relevant population) were to be given the Acme test. First, there might be no difference between the proportions of the two groups passing the test. If this is true, the proportion of whites in the population who pass minus the proportion of blacks in the population who pass will equal zero.[23] This may be written algebraically as

$$P_w(\text{pop}) - P_b(\text{pop}) = 0.$$

Second, if blacks as a group perform better than whites, or if whites outperform blacks, then the difference between the proportion of blacks and the proportion of whites passing the test will not equal zero. Denoted algebraically,

$$P_w(\text{pop}) - P_b(\text{pop}) \neq 0.$$

Conclusions about the discriminatory effect of a test may be made directly on the basis of differences in population pass rates, since they represent the performance of all of those persons who would be affected by the test.

In order to evaluate the difference between the proportions found in a sample, $P_w(\text{sam}) - P_b(\text{sam})$, it is necessary to decide whether the sample result is reasonably probable given the hypothesis, equivalent to the first outcome above, that there is no difference in population proportions. This hypothesis is called the "null hypothesis." [24] The analysis is premised on this hypothesis because of the title VII requirement that unvalidated employment tests not have the effect of discriminating between specified groups in the relevant population of job applicants.[25] A large discrepancy in sample proportions will be highly unlikely if, in fact, there is no difference in performance between the proportions in the population. Alternatively, if the hypothesis of

chance alone is very low if the hypothesis is true. The hypothesis is then "rejected." On the nature of hypothesis testing, see H. BLALOCK, SOCIAL STATISTICS 109–16 (2d ed. 1972).

[23] Although the discussion in text is cast in terms of a *racial* impact, the same type of analysis applies with respect to any of the classifications within the scope of title VII. If the test in question is graded numerically rather than on a simple pass-fail basis, then the appropriate statistical procedure is to test the difference between the average scores or *means* of the two groups. For a description of this procedure, which is similar to the one described in this Comment, see H. BLALOCK, *supra* note 22, at 220–28; L. HOROWITZ, ELEMENTS OF STATISTICS FOR PSYCHOLOGY AND EDUCATION 220–36, 260–77 (1974).

[24] *See* H. BLALOCK, *supra* note 22, at 109–16; L. HOROWITZ, *supra* note 23, at 163–66.

[25] *See* notes 2 & 16 *supra*.

equivalent population proportions is confirmed by the sample data, it is reasonable to conclude that the test does not have a discriminatory impact on a particular group.

Setting the critical probability level, below which the null hypothesis is to be rejected, is more a legal than a statistical question, since it relates to the plaintiff's initial burden of proof.[26] Essentially, it reduces to the question of how certain the trier of fact wants to be that a difference in pass rates might not have occurred by chance alone before the burden is shifted to the employer to validate the test as job-related. This difficult matter is beyond the scope of this Comment. Statisticians often adopt a 5% rule of thumb, rejecting the null hypothesis if the probability of obtaining the sample pass rate difference by chance is less than 5%. This significance level has been adopted in the EEOC Guidelines for test validation[27] and some judicial opinions.[28] The Supreme Court has referred to this rule of thumb in an analogous context,[29] and in the discussion which follows, this Comment will employ the 5% rule.

The particular statistical technique which will be described in this Comment is known as the test for differences between independent proportions.[30] The technique relies on the fact that under certain circumstances, if the employment test is administered many times, the frequency distribution[31] of the sample pass rate differences will approximate a normal distribution.[32]

[26] *See generally* Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1378–93 (1971).

[27] 29 C.F.R. § 1607.5(c)(1) (1976).

[28] *See, e.g.*, Harless v. Duck, 14 Fair Empl. Prac. Cas. 1616, 1623 (N.D. Ohio April 8, 1977) (.05 used as significance level to evaluate difference in employment test pass rates).

[29] Hazelwood School Dist. v. United States, 97 S. Ct. 2736, 2743 n.17 (1977).

[30] *See* H. BLALOCK, *supra* note 22, at 228–32; L. HOROWITZ, *supra* note 23, at 232–39.

[31] A frequency distribution is a measure of the number of times, or frequency, that a particular event will occur if a trial or experiment is repeated many times. *See generally* MOSTELLER, *supra* note 14, at 7–14. The observed frequency distribution can be calculated directly from sample data. For instance, if an employment test is administered to many different groups of blacks and whites, the difference in pass rates between blacks and whites in each group can be computed. The frequency with which particular differences are obtained can be plotted against the differences themselves to get a frequency distribution curve. A theoretical or expected frequency distribution can also be posited. This is the distribution derived from a probability model of the underlying process. A probability model roughly corresponds to the results that would be obtained if the experiment were repeated an infinite number of times. *See id.* at 1–14.

[32] The major characteristics of a normal frequency distribution are as follows: (1) Scores clustered around the average value or "mean" are more frequent than those in the tails of the distribution. (2) Scores grow increasingly less frequent the greater their deviation from the mean. (3) The frequency of scores declines

The circumstances prerequisite for the applicability of this technique are:

1. Independence. — In the employment test context, this means that both the selection and the performance of any test taker must have no bearing on the selection or performance of any other test taker. The requirement of independence in selection will be satisfied so long as the group of test takers is a random sample of the relevant population.[33] Independence in performance can be assumed if no person is allowed to take the test more than once and there are no opportunities for cheating or passing on information about the test to later test takers.

2. Randomness. — In order to satisfy the requirement of randomness, those persons actually taking the test must be representative of the relevant population of potential job applicants. More precisely, the group of test takers must be comparable in composition to that group which would be drawn if each person in the relevant population were equally likely to have been selected.[34] This condition may not be satisfied if, for example, the employer recruits more heavily from some neighborhoods in the vicinity than from others.[35] The pool of test takers might then represent certain educational or economic classes more heavily than those classes are represented in the population as a whole.

3. Sample Size. — The relevant population and the number of test takers in a particular administration must be large enough to overcome certain limitations in the accuracy of the normal approximation. If the relevant population is very large[36] and the smaller of the number of passers and failers in each group is greater than 10,[37] this requirement will be satisfied for the purpose of evaluating the test.[38]

---

symmetrically with their distance from the mean, so that a score 25 points above the mean will be as common as one 25 points below the mean. *See* L. HOROWITZ, *supra* note 23, at 49–50.

[33] *See* H. BLALOCK, *supra* note 22, at 220. Independence must exist not only within groups, but also between groups. For example, if a sample of women is obtained by taking the wives of a previously selected sample of men, there is no independence between the groups of men and women selected. *See id.*

[34] *See id.* at 143.

[35] *See* note 7 *supra.*

[36] *See* note 14 *supra.*

[37] *See* L. HOROWITZ, *supra* note 23, at 147. *See also* H. BLALOCK, *supra* note 22, at 223. When the sample size is small, a test which is not based upon the assumption of normality may be used. One test suitable for evaluating pass-fail rates for a small sample is called Fisher's exact test. *See* H. BLALOCK, *supra* note 26, at 287–91. The application of this test is described in the Appendix, pp. 812–13 *infra.*

[38] It may be difficult in particular cases to say with certainty whether all these assumptions, especially that of randomness, are satisfied. Since departures from

In order to determine the probability of obtaining a particular sample pass rate difference by chance given the null hypothesis, it is necessary to calculate the standard error of the frequency distribution of sample pass rate differences.[39] The calculation proceeds as follows:

A. Calculate the overall pass rate for all those who took the test.

B. Calculate the overall proportion of people in both groups who failed the test.

C. Multiply the proportions from steps A and B. For the sake of convenience, call the product PROD.

D. Divide PROD by the number of blacks in the sample (N_b) and also by the number of whites in the sample (N_w). Add these results and take the square root to yield the standard error.

$$\text{Standard Error} = \sqrt{\frac{\text{PROD}}{N_b} + \frac{\text{PROD}}{N_w}}.$$

Once the standard error is computed, it is possible to locate the sample pass rate difference calculated from the data on the frequency distribution curve. To determine the significance of the location of the particular sample pass rate difference, the "Z statistic" is calculated. The Z statistic is simply a device for measuring distance in a standardized way which makes it possible to attach particular probabilities to particular distances. It is calculated as follows: [40]

---

randomness or independence may come from many sources, it would be impossible for any plaintiff to show the absence of distortion. Consequently, it should be up to the employer to point to a specific source of bias before application of this statistical technique is rejected. This is not unfair to employers, since they will often be in a better position than a title VII plaintiff to describe their recruiting practices and to identify the distortions in selection created by them. Moreover, it should then be open to the plaintiff to show that a demonstrated bias in the data actually operates in the employer's favor. If, for example, as a result of the employer's recruiting, well-educated blacks are more likely to take the test than poorly educated ones, the sample pass rate difference is likely to lead to an underestimation of the population pass rate difference.

[39] A standard error is a measure of the variability of sample means in a sampling distribution. It provides a uniform measure for describing a particular sample pass rate difference's position on the normal curve, and thus supplies information concerning its frequency. For example, one-third of all differences in sample pass rates in a normal distribution fall within one standard error above the mean and one-third fall within one standard error below the mean. If the mean of a particular sampling distribution is 100 and the standard error is 10, then the probability is .67 that an individual sample pass rate difference will fall between 110 and 90. *See* L. HOROWITZ, *supra* note 23, at 38–61.

[40] Dividing the difference in sample pass rates by the standard error makes

$$Z = \frac{P_w(\text{sam}) - P_b(\text{sam})}{\text{Standard Error}} = \frac{\text{Sample Pass Rate Difference}}{\text{Standard Error}}.$$

The probability of obtaining a particular sample pass rate difference by chance alone is less than 5% if the absolute value of the Z score is 1.96 or greater.[41] If the sample data produce such a score, they are regarded as "statistically significant" at the 5% level. This would require rejection of the null hypothesis, that the population pass rates are the same. Such data would, therefore, warrant the conclusion that the test has a discriminatory impact.

C. *Application of the Statistical Test*

In the Jean Taylor hypothetical, 43 of the 100 blacks and 102 of the 200 whites who took the Acme test passed. Since no further information is available regarding the population pass rate, it is necessary to resort to the technique of statistical inference described above. For the application of this technique to be valid, the three assumptions of independence, randomness, and sample size must hold.[42] It is assumed that the employment test is administered in such a way that one test taker's performance does not affect any other's, and that the sample represents a random selection from the relevant population. The sample size requirement is satisfied in the Jean Taylor hypothetical, since well over ten blacks and ten whites both passed and failed the test.[43]

Since the necessary assumptions are satisfied in the hypothetical, the "difference in independent proportions" test can be

---

it possible to "standardize" a normal distribution so that it has a mean of 0 and a variance of 1. *See* note 32 *supra*; H. BLALOCK, *supra* note 22, at 100–01; MOSTELLER, *supra* note 14, at 259–68. It is then possible, by observing the distance of a particular Z score from the standard normal mean of 0, to determine how likely it is that the corresponding result could occur by chance given the null hypothesis. The advantage of standardization is that the computation of the probabilities attaching to particular Z scores has been extensively compiled and tables can be found in any book of mathematical tables or any statistics text.

[41] More precisely, the probability attached to a particular Z score is the likelihood of obtaining a result as extreme as the one obtained from the sample *or more so*.

The 1.96 Z score guide given in the text is the cutoff for the 5% level in a "two-tailed" test, which means that the alternative to the null hypothesis is the hypothesis that either of the groups being tested is equally likely to outperform the other. *See* L. HOROWITZ, *supra* note 23, at 171–74. The absolute value of the Z statistic is used in order to test for significant disparities in favor of either group.

[42] *See* p. 801 & notes 33–38 *supra*.

[43] *See* p. 801 & notes 36–37 *supra*.

used to decide whether the observed difference in sample pass rates is consistent with the initial assumption that there was no difference in the anticipated performance among population groups. It is applied in the following way:

1. State the null hypothesis. The null hypothesis assumes that there is no difference in the proportions of passes between blacks and whites in the relevant population. It can be written as

$$P_w(\text{pop}) - P_b(\text{pop}) = 0.$$

This hypothesis implies that the mean of the frequency distribution of sample pass rates differences is zero.[44]

2. Calculate the standard error of the sample pass rate difference frequency distribution according to the following steps:

a. Calculate the overall proportion of people in both groups who passed the test. For all 300,

$$\frac{43 + 102}{300} = .48.$$

b. Compute the overall proportion of people in both groups who did not pass the test.[45] For the 300 applicants,

$$\frac{57 + 98}{300} = .52.$$

c. Multiply the proportions from steps a and b. Call the product PROD:

$$(.48)\,(.52) = .25 = \text{PROD}.$$

d. Apply the following formula to determine the standard error:

$$\text{Standard Error} = \sqrt{\frac{\text{PROD}}{N_b} + \frac{\text{PROD}}{N_w}},$$

where N_b and N_w are the number of blacks and whites respectively in the sample. Thus,

$$\text{Standard Error} = \sqrt{\frac{.25}{100} + \frac{.25}{200}} = .06.$$

3. Calculate the Z statistic in order to locate the difference in sample pass rates on the frequency distribution. The Z statistic is calculated from the following formula:

$$Z = \frac{P_w(\text{sam}) - P_b(\text{sam})}{\text{Standard Error}}.$$

[44] *See* notes 31–32 *supra*.

[45] Since everybody who takes the test either passes or fails, the sum of the overall pass rate and the overall fail rate must equal 1. Therefore, the overall fail rate can be obtained by subtracting the overall pass rate from 1.

In this example,

$$Z = \frac{.08}{.06} = 1.3.$$

4. Draw a conclusion about the null hypothesis from the Z statistic. The Z statistic based on the proportional pass rates of the sample of blacks and whites who took the test is 1.3. Since this is less than 1.96, the figure used to judge the significance of the sample pass rate difference at the 5% level, it is not possible to reject the null hypothesis. Therefore, there is not a statistically significant difference between the pass rates of blacks and whites in the relevant population. Thus, the evidence is not sufficient to conclude that the test has a discriminatory impact.[46]

II. Proportional Differences Under the Federal Executive Agency Guidelines

The need for a systematic approach in evaluating the substantiality of differences between pass rate proportions has been recognized by the new Federal Executive Agency Guidelines on Employee Selection Procedures.[47] The Agency Guidelines have attempted to fulfill this need by an arbitrary standard known as the four-fifths rule of thumb. This rule purports to assess the substantiality of a difference in pass rates between two racial, sex, or ethnic groups by stating that a difference is not generally considered substantial if the pass rate for one group is at least four-fifths (80%) of the pass rate for the higher group.[48]

The four-fifths rule is an ill-conceived resolution of the problem of assessing the substantiality of pass or acceptance rate differences. It will produce anomalous results in certain cases because it fails to take account of differences in sampling size. It

---

[46] Of course, the plaintiff is not precluded from introducing other evidence tending to show that the employer's employment practices constituted "disparate treatment," *see* note 8 *supra*.

[47] Agency Guidelines, *supra* note 3, §§ 60-3.1 to .14.

[48] *Id.* § 60-3.4. Subsection (b) of that Guideline reads in part:

> A selection rate for any racial, ethnic or sex group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded as evidence of adverse impact. Smaller differences in selection rate may nevertheless be considered to constitute adverse impact, where they are significant in both statistical and practical terms. Greater differences in selection rate would not necessarily be regarded as constituting adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group.

For a Justice Department commentary on the Agency Guidelines, see Questions and Answers on the Federal Executive Agency Guidelines on Employee Selection Procedures, 42 Fed. Reg. 4052 (1977) [hereinafter cited as Questions and Answers].

also neglects the magnitude of differences in pass rates by considering only the ratio of the two rates.

These flaws in the four-fifths rule can be eliminated by replacing it with a test of the statistical significance of differences in pass rate proportions.[49] The Guidelines have already recognized the usefulness of statistical analysis as an adjunct to the four-fifths rule, especially when the sample size is small.[50] Probability theory should be employed in all cases, however, to evaluate the significance of pass rate differences, and the four-fifths rule should be abandoned altogether. The statistical calculation for testing the difference between proportions, explained in the previous Part of this Comment, can be specified almost as simply as the four-fifths rule. More importantly, as this section will demonstrate, statistical analysis will provide greater consistency and accuracy in determining the discriminatory impact of employment tests. This is because the approach based on statistical significance takes into consideration the size of the sample, and the magnitude of differences in pass rates.

A. Sample Size Problems

The four-fifths rule of the Agency Guidelines applies uniformly to samples of all sizes except very small ones. The shortcoming of this uniform approach is that it fails to reflect the greater precision afforded by large sample sizes.[51] Imagine, for example, two employers, *A* and *B*. Employer *A* has 300 applicants (250 whites and 50 blacks), and employer *B* has 1,200 applicants (1,000 whites and 200 blacks). Assume that the acceptance rates of blacks and whites are identical for both employers. Employer *A* accepts 80% of the whites (200 out of the 250 pass) and 70% of the blacks (35 out of the 50 pass), as does employer *B* (800 out of the 1,000 whites pass as do 140 out of the 200 blacks). Under the four-fifths rule of the Agency Guidelines, these pass rates do not indicate an adverse impact with respect to either employer, since in both cases the ratio of pass rates is 7/8, which is greater than 4/5.[52]

[49] *See* pp. 798–803 *supra*.

[50] Agency Guidelines, *supra* note 3, § 60-3.4; *see* note 48 *supra*; Questions and Answers, *supra* note 48, at 4052–53.

[51] *See* H. Blalock, *supra* note 22, at 292.

[52] Calculation under the Agency Guidelines' four-fifths rule is as follows:

Employer *A*

	Whites	Blacks
Total Applicants	250	50
Number Selected	200	35
Passing Rate	80%	70%
Ratio of Rates	7/8	

Conclusion: Ratio of 7/8 is greater than 4/5, so no adverse impact is shown.

Using the statistical test for independent proportions to ascertain the probability of these outcomes, however, yields a different result. If the test or selection process used by employer *A* would have no adverse racial impact on blacks and whites as a whole, then the probability of obtaining such a difference in pass rates by chance is greater than 5%.[53] This is not considered statistically significant and would not warrant rejection of the null hypothesis.[54] This conclusion with respect to employer *A* is in accord with the result under the four-fifths rule. For employer *B*, however, the probability of obtaining that particular difference in pass rates by chance alone is less than 5%, which is statistically significant.[55] The four-fifths rule of the Agency Guidelines would not compel validation of the selection procedure used by employer *B*, even though statistical analysis demonstrated that a disproportionate impact should be found and that the bur-

---

EMPLOYER *B*

	Whites	Blacks
Total Applicants	1000	200
Number Selected	800	140
Passing Rate	80%	70%
Ratio of Rates	7/8	

Conclusion: Ratio of 7/8 is greater than 4/5, so no adverse impact is shown.

[53] Calculation for testing the difference between independent proportions for employer *A* is as follows:

(1) Overall pass rate: $\frac{200 + 35}{250 + 50} = .7833.$

(2) Overall fail rate: $1 - .7833 = .2167.$

(3) PROD $= (.7833)\ (.2167) = .1697.$

(4) $\sqrt{\frac{.1697}{250} + \frac{.1697}{50}} = .0638$ (the standard error).

Difference between proportions is $80\% - 70\% = 10\%$ (.10).

$$\text{Z score} = \frac{\text{Difference Between Proportions}}{\text{Standard Error}} = \frac{.10}{.0638} = 1.57.$$

Conclusion: The Z score (1.57) is less than the designated cutoff for significance (1.96), so the result is not significant and no discriminatory effect is shown.

[54] *See* p. 800 *supra*.

[55] Calculation for testing the difference between independent proportions for employer *B* is as follows:

(1) Overall pass rate: $\frac{800 + 140}{1000 + 200} = .7833.$

(2) Overall fail rate: $1 - .7833 = .2167.$

(3) PROD $= (.7833)\ (.2167) = .1697.$

(4) $\sqrt{\frac{.1697}{1000} + \frac{.1697}{200}} = .0319$ (the standard error).

Difference between proportions is $80\% - 70\% = 10\%$ (.10).

$$\text{Z score} = \frac{\text{Difference Between Proportions}}{\text{Standard Error}} = \frac{.10}{.0319} = 3.13.$$

Conclusion: The Z score (3.13) is greater than designated cutoff for significance (1.96), so the result is significant and a discriminatory effect is indicated.

den should shift to employer *B* to validate this selection process.

Conversely, the failure of the Agency Guidelines to account for sample size can result in a finding of adverse impact under the four-fifths rule despite a fairly high probability that the difference in pass rates could occur by chance alone. Consider employers *C* and *D*. Employer *C* has 3,000 white applicants and 600 black applicants; 5% of the whites are accepted (150 pass out of the 3,000), and 3% of the blacks are accepted (18 pass out of the 600). Employer D has 1,000 white applicants of whom 5% are accepted (50 pass), and 200 black applicants, of whom 3% are accepted (6 pass). For both employers *C* and *D* the ratio of the black and white acceptance rates is 3/5, evidencing a discriminatory impact under the four-fifths rule.[56] A calculation of the statistical probability of these results, however, again suggests that these employers should not be treated identically. The probability of obtaining by chance alone a result as extreme as employer *C*'s or more so is less than 5%, which is considered statistically significant.[57] The probability of the racial difference in pass rates as extreme as that shown by small employer *D*'s

---

[56] Calculation under the Agency Guidelines' four-fifths rule is as follows:

EMPLOYER *C*

	Whites	Blacks
Total Applicants	3000	600
Number Selected	150	18
Passing Rate	5%	3%
Ratio of Rates	3/5	

Conclusion: Ratio of 3/5 is less than 4/5, so adverse impact is shown.

EMPLOYER *D*

	Whites	Blacks
Total Applicants	1000	200
Number Selected	50	6
Passing Rate	5%	3%
Ratio of Rates	3/5	

Conclusion: Ratio 3/5 is less than 4/5, so adverse impact shown.

[57] Calculation for testing the difference between independent proportions for employer *C* is as follows:

(1) Overall pass rate: $\frac{150 + 18}{3000 + 600} = .0467.$

(2) Overall fail rate: $1 - .0467 = .9533.$

(3) $\text{PROD} = (.0467)\,(.9533) = .0445.$

(4) $\sqrt{\frac{.0445}{3000} + \frac{.0445}{600}} = .0094$ (the standard error).

Difference between proportions is $5\% - 3\% = 2\%$ (.02).

$$\text{Z score} = \frac{\text{Difference Between Proportions}}{\text{Standard Error}} = \frac{.02}{.0094} = 2.13.$$

Conclusion: The Z score (2.13) is greater than designated cutoff for significance (1.96), so the result is significant and a discriminatory effect is indicated.

sample or more so, however, is greater than 5%.[58] This result is not considered statistically significant, and the burden should not shift to employer *D* to validate the selection procedure producing these pass rates.

These two examples of pairs of employers[59] with identical black-white pass rates but different sample sizes demonstrate the importance of accounting for sample size. Discrepancies in pass rates are more likely to be significant as sample size increases. In the extreme case when the sample *equals* the population, *any* discrepancy is significant. The four-fifths rule of the Agency Guidelines ignores sample size, while the test for statistical significance takes this into consideration.[60] A change from the

---

[58] Calculation for testing the difference between independent proportions for employer *D* is as follows:

(1) Overall pass rate: $\frac{50 + 6}{1000 + 200} = .0467$.

(2) Overall fail rate: $1 - .0467 = .9533$.

(3) PROD = (.0467) (.9533) = .0445.

(4) $\sqrt{\frac{.0445}{1000} + \frac{.0445}{200}} = .0163$ (the standard error).

Difference between proportions is 5% − 3% = 2% (.02)

$$\text{Z score} = \frac{\text{Difference Between Proportions}}{\text{Standard Error}} = \frac{.02}{.0163} = 1.23.$$

Conclusion: The Z score (1.23) is less than the designated cutoff for significance (1.96), so the result is not significant and a discriminatory impact has not been shown.

[59] For convenient reference, the four hypothetical situations are summarized below:

	Employers			
	A	*B*	*C*	*D*
Number of White Applicants	250	1000	3000	1000
Number of Black Applicants	50	200	600	200
Black-White Selection Ratio	10%/80%	70%/80%	3%/5%	3%/5%
Agency Guidelines' 4/5 Rule	No Impact	No Impact	Impact	Impact
Test for Statistical Significance	No Impact	Impact	Impact	No Impact

[60] For a recent example of the anomalous results which can stem from the mechanical application of the Agency Guidelines, see Jackson v. Nassau County Civil Serv. Comm'n, 424 F. Supp. 1162 (E.D.N.Y. 1976). That court, using the four-fifths rule, failed to find disparate impact when a test pass rate was 99 out of 113 (87.6%) for whites and only 40 out of 55 (72.7%) for blacks. The ratio of the black pass rate to that of the whites was 72.7/87.6, or .83, which the court noted was "well within" the four-fifths rule of the Agency Guidelines. *Id.* at 1168 n.10. The court observed:

> [T]he present case involves a relatively small number of applicants. . . . The passing percentages would be substantially altered by a shift of several minority candidates from the fail to the pass column. In such a situation,

four-fifths rule to a method based on statistical inference will not change the balance of interests between the government or aggrieved job applicant and the employer, but it will lead to greater accuracy and eliminate certain disparities in disproportionate impact determinations resulting from differences in sample sizes. Specifically, the use of statistical analysis would end the relatively favorable position of large employers compared to small employers under the Agency Guidelines.

B. Magnitude of the Difference Problems

The four-fifths rule of the Agency Guidelines is inadequate not only because it fails to account for sample size but also because it fails to consider the magnitude of the difference in the pass rates. Reconsider employers *B* and *D* from the examples above. The sample size of each of these employers was identical; there were 1,000 white applicants and 200 blacks. As previously observed, employer *B* would satisfy the four-fifths rule, but employer *D* would not.[61] The statistical likelihood that these results would occur by chance alone, however, suggests a contrary conclusion: a disproportionate impact should be found for employer *B*, but not for employer *D*.[62]

The reason for this peculiar phenomenon is that the four-fifths rule is based solely upon the ratio of the pass rates. Employer *B*'s 7/8 (70/80) black-white pass ratio is larger than 4/5, and employer *D*'s 3/5 ratio is smaller. But the difference in magnitude between 10 (80–70 for employer *B*) and 2 (5–3 for employer *D*) is not taken into account under the four-fifths rule.[63]

> some margin of error must be allowed. Plaintiffs bear a heavier burden in demonstrating a disproportionate impact when the number of test takers is small and, in our opinion, a showing of a 4/5 passing rate does not satisfy that burden.

Id. at 1168.

The use of the statistical test for the difference between independent proportions, however, would account for the sample size. Moreover, it would yield an opposite result and would show disparate impact. The Z score can be calculated from this information as 2.40, which is statistically significant. In fact, it can be shown that the probability of obtaining so large a Z score by chance alone is less than 2%.

[61] *See* notes 52, 56 *supra.*

[62] *See* notes 55, 58 *supra.*

[63] An *exclusive* reliance on differences in magnitude, without consideration of sample size, may also lead to anomalous results. An example of the difficulty of using an approach based only on the *magnitude* of a difference between proportions is seen in United States v. Newman, 549 F.2d 240 (2d Cir. 1977). *Newman* was a jury discrimination case where the criminal defendant alleged that blacks were purposely underrepresented on juries in Connecticut by the Government's

Another anomaly which flows from failure to take account of the magnitude of the differences is that different results are obtained if the issue is framed in terms of white-black fail rates rather than black-white pass rates.[64] Although the symmetry of a test is not necessarily an indicium of its utility,[65] the failure of the four-fifths rule to translate from pass to fail rates may prove confusing, since the fact that every person either passes or fails suggests that the determination of discriminatory impact should be the same whether focused on pass rates or fail rates. This potential source of confusion can be eliminated by use of the statistical test described in this Comment. Identical results will thus be obtained regardless of whether the problem is phrased in terms of pass or fail rates.[66]

III. CONCLUSION

The Supreme Court has recognized the probative value of statistical analysis in employment discrimination and jury dis-

---

allegedly excessive use of peremptory challenges to exclude blacks. The evidence revealed that blacks constituted 5% of the population in Connecticut, but only 2.11% of the jurors. The disparity (by subtraction) was thus 2.89 percentage points. *Id.* at 249.

In order to assess the substantiality of this difference, the *Newman* court relied upon a Supreme Court case, Swain v. Alabama, 380 U.S. 202 (1965). *Swain* had held that a difference of 10 points between the percentage of blacks in the relevant population and the percentage of blacks on Alabama juries was insufficient to show discrimination. *Id.* at 208–09. The *Newman* court said that the 2.89-point dispartiy in Connecticut fell far short of the 10-point disparity found inadequate in *Swain*, and so it found the evidence insufficient to show a fourteenth amendment violation. 549 F.2d at 249–50. The fallacy in the court's blind adherence to a subtraction rule, however, is that since only 5% of the Connecticut population is black, *no* number could show a 10-point difference. Even if blacks were totally excluded, the difference would be only 5 percentage points.

[64] For example, employer *A*, *see* note 52 *supra*, has pass rates of 80% for whites and 70% for blacks. Expressed as fail rates, these figures become 20% for whites and 30% for blacks. The white-black fail ratio is 2/3, which is less than 4/5 and would not satisfy a four-fifths rule for fail ratios. The black-white pass ratio, however, is 7/8, which satisfies the Agency Guidelines.

[65] If a method of assessing discriminatory impact which focused only on pass rates is efficient and accurate in all cases, it should not be rejected simply because it could not be applied when cast in terms of fail rates. The four-fifths rule is objectionable because it is not accurate in all cases. The confusion which may be engendered by its asymmetry is an additional concern.

[66] The statistical method for testing the difference between independent proportions involves *multiplying* the overall pass rate by the overall failure rate. Since these numbers are multiplied, their order does not matter and the problem can be approached either in terms of "passing" or "failing" without affecting the outcome of the test. Moreover, the subtraction of the difference in proportions will not be affected by the change since the absolute value of magnitude of the difference is the same for either a difference in passing rates or in failing rates.

crimination cases. It was assisted in these cases by an analysis of probability based upon the binomial distribution. However, the binomial test cannot readily be utilized to assess the discriminatory effect of employment tests or comparable applicant selection procedures. This Comment has therefore suggested an alternative statistical method — a test of the difference between independent proportions — for evaluating a plaintiff's prima facie case of disproportionate impact under title VII.

A comparison of the results obtained by this test with those indicated by the four-fifths rule demonstrates the erratic qualities of the Agency Guidelines. In contrast to the statistical analysis, the four-fifths rule produces anomalous results because it does not account for differences in sample size and does not consider the magnitude of differences in pass rates. Since the advantages of the statistical method outweigh any inconvenience in calculation, it should entirely displace the Agency Guidelines as the means for assessing the substantiality of pass rate differences.

APPENDIX: THE FISHER EXACT TEST [67]

When the sample is small, the Fisher exact test can be used instead of the test for independent proportions. The Fisher exact test requires first that the data be arranged in a 2×2 table:

	PASS	FAIL	
Blacks	A	B	Number of Blacks (N_b)
Whites	C	D	Number of Whites (N_w)
	Total Passing ($A + C$)	Total Failing ($B + D$)	N

In this table, N represents $N_b + N_w$, the total number of people in the sample. A, B, C, and D correspond to the actual number of people in each pass or fail category for each group. These four numbers necessarily total N because everyone in the sample either passes or fails. The exact probability of obtaining a particular outcome by chance given the marginal totals (N_b, N_w, $A + C$, $B + D$) is calculated by means of the following formula:

[67] *See generally* H. BLALOCK, *supra* note 22, at 287–91; MOSTELLER, *supra* note 14, at 315–17.

$$P = \frac{(A+B)!\ (C+D)!\ (A+C)!\ (B+D)!}{N!A!B!C!D!}.$$

By the symbol "!" is meant the factorial of the number which precedes it. The factorial of a number is the product of that number with all the numbers which precede it. For example, $5! = 5 \times 4 \times 3 \times 2 \times 1$, or 120. $0! = 1$, by convention.

This probability formula gives only the probability of obtaining exactly one particular outcome, whereas the relevant statistical question is the probability of obtaining an outcome that extreme or more so. For example, assume that an employer administers a test to 12 applicants, 5 blacks (1 passes and 4 fail) and 7 whites (5 pass and 2 fail). The question for purposes of the legal inquiry is how likely it would be to get either no more than one black pass or no more than one white pass by chance, given a sample of 5 blacks and 7 whites in which there are all together 6 persons passing and 6 persons failing. The null hypothesis is that blacks are no more and no less likely to pass than are whites. It is necessary to commute the probabilities of getting one black pass, no black passes, and one white pass. (Since 6 pass and only 5 blacks take the test, at least one white must pass.)

1	4	5
5	2	7
6	6	12

Probability of one black pass =

$$\frac{5!7!6!6!}{12!4!5!2!1!} = .1136.$$

0	5	5
6	1	7
6	6	12

Probability of no black passes =

$$\frac{5!7!6!6!}{12!0!5!6!1!} = .0076.$$

5	0	5
1	6	7
6	6	12

Probability of one white pass =

$$\frac{5!7!6!6!}{12!5!0!1!6!} = .0076.$$

The total probability for getting a result as extreme as the sample result or more so by chance, given the null hypothesis, is obtained by adding these three numbers. The total probability is $.1136 + .0076 + .0076 = .1288$, which is greater than .05. It is therefore not possible to reject the null hypothesis or conclude that the test has a discriminatory impact under the 5% rule of thumb.