SUBMISSION TYPE
Poster

TITLE
Small Sample? No Problem! Using Permutation Testing for Adverse Impact Analysis

ABSTRACT
Small sample sizes are a common issue in I-O psychology, limiting statistical power and increasing the risk of errors. This study examines permutation testing for adverse impact analysis in small sample settings. Permutation testing, a non-parametric method, is well-suited for small samples due to fewer assumptions. We compared it with other methods of adverse impact analysis across three organizational datasets. Results show permutation testing outperforms the 4/5ths rule by avoiding false positives in small samples, making it a practical tool for small sample analysis in I-O psychology.

WORD COUNT
2999

Small Sample? No Problem! Using Permutation Testing for Adverse Impact Analysis

One of the most common problems in the field of I-O psychology is small samples sizes. Small sample sizes can make it difficult to detect meaningful differences between variables due to a lack of statistical power, thereby increasing the risk of Type II errors (Van de Schoot & Miocević, 2020). In addition, small samples are more prone to random variations in the data, increasing the risk of Type I errors as well (Van de Schoot & Miocević, 2020). These issues can make it difficult to replicate the results of studies with small samples, likely contributing to the replication crisis. Small sample sizes also oftentimes violate the assumptions underlying the use of traditional methods of analysis, calling into question the results from these methods (Siegel, 1957). The problem of small sample sizes is especially prevalent in the applied setting (though still common in academia), where practitioners are often limited to a small convenience sample due to limitations in resources (Brawley & Pury, 2017). While many methods have been suggested to handle small sample sizes, each have their drawbacks, and practitioners have largely conceded to the fact that small sample sizes are here to stay. With this in mind, this study examines the use of permutation testing, also called randomization testing, as a method to handle small sample sizes in the field of I-O psychology, specifically examining its use for adverse impact analysis. Organizations are required by law to conduct adverse impact analysis on selection procedures, even when the sample size available is small. According to the BLS, small businesses (defined as having fewer than 500 employees) make up the majority of establishments in the United States. This prevalence means smaller samples are likely a common problem when analyzing adverse impact (Bureau of Labor Statistics, 2020). Permutation testing is common in other statistical fields as a way of doing hypothesis testing for small samples but has only been used rarely in the field of I-O psychology (Huo, 2014). As a statistical technique it is fairly similar to bootstrapping, with one important difference, described in more detail below. The goal of this study is to introduce permutation testing to a wider audience within I-O psychology, and to examine the use of permutation in adverse impact analysis, a common analysis done by practitioners in the field.

Permutation Testing

Permutation testing consists of three distinct stages. As permutation testing is used in hypothesis testing, a null and alternative hypothesis comparing two or more groups must first be determined. In the first stage, the test statistic is determined. This test statistic can be a variety of different statistics, including mean difference, median difference, correlations, F-statistics (ANOVA), and r-squared (Good, 2005). The main qualifier is that the test statistic needs to demonstrate the relationship between two or more groups (Good, 2005). Next, each value in the dependent variable is randomly assigned to a value in the independent variable, “shuffling” the data around. This random assignment, called a permute, is done without replacement, meaning each value occurs the same number of times as in the observed data. This process is then repeated thousands of times to create a distribution of the test statistic. Lastly, the p-value is calculated by locating where on the distribution the observed statistic falls. What this process does is create a scenario where the null hypothesis is true, and the observed statistic can then be compared to that distribution to find the likelihood of getting the same result if the null hypothesis is in fact true.

Traditional tests are parametric in nature, meaning they make certain assumptions about the underlying population distribution from which the sample data is drawn to calculate what the test statistic is and the significance level. These assumptions include a normal distribution, homogeneity of variance, and independence of observations (Glass et al., 1972). When the sample size is large enough, traditional parametric tests have been shown to be quite robust even when assumptions are violated (Lumley et al., 2002). However, when the sample size is small, it becomes very important to observe these assumptions carefully. Permutation testing is a non-parametric test, meaning it does not rely on assumptions about the underlying population distribution (Janssen & Pauls, 2003). This makes permutation testing more suitable for small sample sizes (Hothorn et al., 2006). The main assumption for using permutation data is the assumption of exchangeability, meaning that any combination of the data is equally as likely under the null hypothesis (Good, 2005). In comparison to traditional parametric tests, permutation testing does not have a high bar to have its assumptions met.

Bootstrapping is another non-parametric test that is similar to permutation testing and is commonly used in the field of I-O psychology. The main mathematical difference between the two is that in bootstrapping the randomization is done with replacement, meaning the same value can be selected multiple times (Janssen & Pauls, 2003). As a result, bootstrapping has the potential to be more variable with small samples sizes (Good, 2005), making permutation testing more appropriate for dealing with small sample sizes. In addition, bootstrapping does not directly test distributions between groups, only parameters. To describe the difference, "permutation testing tests hypotheses about distributions and bootstrapping test hypotheses about parameters" (Good, 2005).

Given the limitations of these other methods, there is potential for the use of permutation testing to add additional benefit beyond other methods in certain contexts. In particular, we see value in using permutation testing to assist with adverse impact analysis, a common analysis conducted in applied settings. Due to the nature of adverse impact analysis evaluating selection rates of minority and majority groups, small sample sizes are commonly an issue, especially for the minority groups. In addition, the underlying population distribution is rarely known, making a non-parametric test more appropriate (Morris & Lobsenz, 2000). Lastly, adverse impact analysis begins with the null hypothesis stating that there is no differences between the majority and minority groups, meaning that the assumption of exchangeability required for permutation testing is met.

To our knowledge, permutation testing has only been used rarely in I-O psychology. In a systematic literature review of the use permutation testing in educational and behavioral sciences, only papers related to the RIASEC test of vocational interests (Holland, 1985), which uses permutation testing in the scoring of the test, were relevant to the field of I-O psychology (Huo, 2014). A review of more recent articles only resulted in three articles in the field that reported the results of permutation testing but did not explain its use in any detail (Tavacıoğlu et al., 2019; Sinniah et al., 2022; Ji et al., 2024). Based on a literature review, there remains a need in the research to explore the use and benefits of permutation testing for hypothesis testing in the field of I-O psychology. For adverse impact analysis, a common analysis in the field that often face small sample sizes, the potential benefits of permutation testing are worth exploring. Therefore, we propose the following hypotheses:

Hypothesis 1: Permutation Testing will demonstrate similar results to traditional tests of adverse impact analysis.

Hypothesis 2: Permutation Testing will outperform other methods of hypothesis testing in small samples.

Methods

To test our hypotheses, we conducted 2 separate studies. All data for these studies used real organizational data. In Study 1 we analyzed data from a previously validated personality instrument, in order to evaluate how permutation testing performed as a method of adverse impact analysis relative to other methods. In Study 2, we analyzed two separate datasets on a previously validated personality instrument. The personality instrument was the same for both datasets, though different than the instrument used in Study 1. One dataset had a small sample (N = 28) and the other had a large sample (N = 645). In this study, the goal was to examine how permutation testing performed relative to other methods in a small sample, and then compare those results to the larger sample size to check for Type 1 and Type II errors. Details on the methods for these two studies are as follows

Participants and Procedure

Study 1:

The first dataset was a sample of 353 participants on a validated personality instrument, which included ethnicity information and self-reported performance data. In a previous study, this performance data was used to create an algorithm from the personality data, based on regression models between performance and the various personality traits. This algorithmic score was put in percentiles for each participant, and cutoffs were tested at various points to check for both predictive validity and adverse impact.

Study 2:

Two of the datasets had data on the same personality instrument, though different from the first dataset. One of these two datasets was a small sample of managers (N = 28) at a national retail organization. This dataset included criterion data such as sales data, employee engagement numbers by region, and supervisory ratings of performance. In a previous study, these criterion items were combined to create an overall measure of performance and then a stepwise regression model was run to determine which personality factors were most important to performance and their overall weight. From here, an algorithm was developed to determine an overall percentile score for each participant. Based on analysis of the data, a cutoff was set at the 66th percentile, where everybody who scored about the cutoff was recommended to hire and everybody below was not. The other dataset from this personality instrument was a larger sample size ($N = 645$) taken from a variety of different sources. The algorithm from the previous dataset was recreated and the cutoff set at the 66th percentile, in order to gather selection rate data from this sample. This cutoff had a basis in earlier analysis, but for this particular study we just needed a set cutoff.

Analysis

To evaluate the use of permutation testing in adverse impact analysis, we compared to 4 other methods of adverse impact analysis: bootstrapping, the 4/5ths rule, chi-square test of independence, and Fisher's exact test. The 4/5ths rule is the most commonly used method in adverse impact analysis and is considered important for legal standards (EEOC, 1978). However, the 4/5ths rule has been criticized for being an arbitrary threshold that does not consider external factors and is sensitive to small sample sizes (Roth et al., 2006). A preferred method of testing for adverse impact analysis is evaluating if there is a statistically significant difference between the majority and minority group (Roth et al., 2006). For this research, the commonly used chi-square test of independence was used to evaluate the significance of these differences. As at least one of our datasets dealt with small samples, we used Yate's continuity correction in calculating Chi-square (Yates, 1934). As Fisher's exact test has been recommended for use when small samples are a concern in adverse impact analysis (Morris & Lobsenz, 2000), we also included this method in our comparison.

Bootstrapping is not commonly used for adverse impact analysis but was run as a comparison for permutation testing due to the similarity with permutation testing. 10,000 iterations were bootstrapped for each adverse impact analyses, using the selection ratio as the test statistic. In order to evaluate for adverse impact using bootstrapping, a confidence interval was calculated for the selection ratio. If the confidence interval did not fall within the acceptable range for a selection ratio (0.8-1.2), it was considered evidence for adverse impact.

For permutation testing, the absolute difference in the selection rate was chosen as the test statistic, as this calculation would evaluate if there was a significant difference between the selection ratio of the majority and minority groups regardless of the direction of the relationship. 10,000 permutation were run, each time randomly assigning the “Hire” and “Not Hire” labels to the participants and then calculating the selection rates for the minority and majority groups. Lastly, the absolute value of the difference between the m minority and majority groups’ selection rates was calculated. The results of these 10,000 permutations were then plotted to create a distribution from which the observed value could be compared. A p-value was calculated by counting the number of permutations that had a value greater than or equal to the observed value.

Study 1

For Study 1, we conducted adverse impact analysis between all unique values of ethnicity collected and the “Caucasian” majority group and each of the potential cutoff points. The methods used to conduct adverse impact analysis were permutation testing, bootstrapping, 4/5ths rule, chi-square test of independence, and Fisher’s exact test. The main goal with this dataset was to evaluate how permutation testing performed relative to other methods of adverse impact analysis on a “typical” dataset. By looking at ethnicity and multiple cutoff points, we were able to get more datapoints for how permutation testing performed relative to the other methods.

Study 2

For Study 2, we ran all the methods used in Study 1 for both the small and large dataset. Results between the 2 datasets were then compared to check for Type I and Type II errors, with the assumption that the larger dataset would be closer to the "true" value As computational constraints were a concern for permutation testing and bootstrapping with the larger sample size, we focused on gender rather than ethnicity for this comparison, in order to limit the number of tests that needed to be run.

Results

Study 1

Results for Study 1 can be found in Tables 1-6. Using the 4/5ths rule, potential adverse impact was detected in 6 different cases. Four of these cases were at various cutoffs for the Native American minority group. Of note, the Native American minority group only had eight people, which may have impacted results. The other two cases were for the African American and Asian minority groups at the 50th percentile cutoff. In all cases, the minority group was selected at a higher rate than the majority group. For both chi-square test of independence and Fisher's exact test, there was only evidence of adverse impact in the Native American minority group at the 50th percentile cutoff. In the bootstrap analysis, there was evidence of adverse impact at the 50th, 30th, and 20th cutoff points for the Native American minority group. For the permutation test, there was only evidence of adverse impact in the Native American minority group at the 50th percentile cutoff, coming to the same conclusions as the chi-square test of independence and Fisher's exact test.

As chi-square test of independence and Fisher's exact test have been recommended as the preferred to tests for adverse impact analysis, the fact that the permutation tests came to the same conclusions is significant. Therefore, we found support for Hypothesis 1.

Study 2

Results for Study 2 can be found in Tables 6-12. For the small dataset, there was evidence of adverse impact toward the female minority group using the 4/5ths rule and bootstrap analysis. However, the differences between the minority and majority group were not found to be significant in the chi-square test of independence, Fisher's exact test, and the permutation test. Given this discrepancy, the question arises: which test is produces the most accurate results?

To answer this question, we reran the analysis on the larger dataset. In the larger dataset, there was no significant difference between the minority and majority group on any of the methods tested. Therefore, it was concluded that the significant differences found using the 4/5ths rule and the bootstrap analysis in the small sample was a "false positive". Thus, permutation testing outperformed both the 4/5th rule and bootstrapping in the small sample, providing evidence to support Hypothesis 2.

Discussion

Our findings suggest that permutation testing is a viable method in evaluating for adverse impact, particularly in small samples. Specifically, the results support the use of permutation testing in this context by showing that it performed similarly to the widely accepted Fisher's exact test and the chi-square test of independence. Notably, permutation testing provided consistent results across both large and small datasets, avoiding the false positives observed with the 4/5ths rule and bootstrapping. This consistency indicates that permutation testing may help mitigate the issue of spurious findings due to sampling variability, a problem often exacerbated in small sample analyses.

Small sample sizes are a prevalent concern in both research and practice, especially in applied settings where data availability is limited. Traditional methods often struggle to meet their underlying assumptions in these situations, potentially resulting in increased Type I or Type II error rates. Permutation testing, which makes fewer assumptions regarding the population distribution, is better equipped to handle these challenges, thereby providing a more reliable method for hypothesis testing when sample sizes are insufficient to meet parametric assumptions.

Furthermore, this study has implications for the broader use of permutation testing in other areas of I-O psychology beyond adverse impact analysis. For instance, examining group differences in employee engagement, leadership effectiveness, or job satisfaction could benefit from the flexibility and robustness of permutation testing, particularly in studies with small or specialized populations. By raising awareness of permutation testing to the I-O psychology community, this research opens new avenues for addressing the small sample issue that has plagued many applied settings.

Limitations & Future Research Directions

While using real organization data can be considered a strength of our study, it also limited the number of scenarios under which we could compare permutation testing to other methods of analysis. This is particularly true when it comes to conducting sensitivity analysis to evaluate the number of Type I and Type II errors with each method. Future studies should use simulated data to supplement the findings from this real-world dataset in an efficient manner. Additionally, future studies should further explore the potential benefits of permutation testing above and beyond current methods, as the current study only partially demonstrated these benefits. Lastly, future studies should explore the use of permutation testing in other areas of I-O psychology limited by small sample sizes, such as employee engagement and team research.

References

Biddle, D. (2017). Adverse impact and test validation: A practitioner's guide to valid and defensible employment testing. Routledge.

Brawley, A. M., & Pury, C. L. (2017). Little things that count: A call for organizational research on microbusinesses. Journal of Organizational Behavior, 38(6), 917-920.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43, 38290–39315.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of educational research, 42(3), 237-288.

Good, P. I. (2005). Resampling Methods: A Practical Guide to Data Analysis (3rd ed.). Birkhäuser. ISBN 978-0817643867.

Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2006). A lego system for conditional inference. The American Statistician, 60(3), 257-263.

Huo, M., Heyvaert, M., Van den Noortgate, W., & Onghena, P. (2014). Permutation tests in the educational and behavioral sciences. Methodology.

Janssen, A., & Pauls, T. (2003). How do bootstrap and permutation tests work?. The Annals of statistics, 31(3), 768-806.

Ji, J., He, B., Gong, S., Sheng, M., & Ruan, X. (2024). Network analysis of occupational stress and job satisfaction among radiologists. Frontiers in Public Health, 12, 1411688.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. Annual review of public health, 23(1), 151-169.

Morris, S. B., & Lobsenz, R. E. (2000). Significance tests and confidence intervals for the adverse impact ratio. Personnel Psychology, 53(1), 89-111.

Roth, P. L., Bobko, P., & Switzer III, F. S. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. Journal of Applied Psychology, 91(3), 507.

Siegel, S. (1957). Nonparametric statistics. The American Statistician, 11(3), 13-19.

Sinniah, S., Al Mamun, A., Md Salleh, M. F., Makhbul, Z. K. M., & Hayat, N. (2022). Modeling the significance of motivation on job satisfaction and performance among the academicians: the use of hybrid structural equation modeling-artificial neural network analysis. Frontiers in Psychology, 13, 935822.

Tavacıoğlu, L., Taç, U., Eski, Ö., & Gökmen, N. (2019). Burnout and job satisfaction among Turkish oceangoing seafarers. International maritime health, 70(4), 232-238.

U.S. Bureau of Labor Statistics. (2020). 2020 Results of the Business Response Survey. https://www.bls.gov/brs/.

Van de Schoot, R., & Miocević, M. (2020). *Small sample size solutions: A guide for applied researchers and practitioners* (p. 284). Taylor & Francis.

Yates, F. (1934). Contingency tables involving small numbers and the χ 2 test. Supplement to the Journal of the Royal Statistical Society, 1(2), 217-235.

Tables

Table 1. *Study 1 Adverse Impact Analysis Ethnicity: 4/5ths rule*

Ethnicity	Cutoff	Selection Ratio	Evidence of Adverse Impact?
African-American (n=31)	10th percentile	1.042	No
Hispanic (n=18)	10th percentile	0.928	No
Asian (n=38)	10th percentile	0.997	No
Native American (n=8)	10th percentile	1.114	No
African-American (n=31)	20th percentile	1.117	No
Hispanic (n=18)	20th percentile	1.069	No
Asian (n=38)	20th percentile	1.047	No
Native American (n=8)	20th percentile	1.283	Yes
African-American (n=31)	30th percentile	1.164	No
Hispanic (n=18)	30th percentile	1.169	No
Asian (n=38)	30th percentile	1.147	No
Native American (n=8)	30th percentile	1.503	Yes
African-American (n=31)	40th percentile	1.187	No
Hispanic (n=18)	40th percentile	1.070	No
Asian (n=38)	40th percentile	1.152	No
Native American (n=8)	40th percentile	1.533	Yes
African-American (n=31)	50th percentile	1.425	Yes
Hispanic (n=18)	50th percentile	1.104	No
Asian (n=38)	50th percentile	1.337	Yes
Native American (n=8)	50th percentile	1.933	Yes

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the selection ratio was < 0.8 or greater than 1.2.

Table 2. *Study 1 Adverse Impact Analysis Ethnicity: Chi-square test of Independence*

Ethnicity	Cutoff	p-value	Evidence of Adverse Impact?
African-American (n=31)	10th percentile	0.727	No
Hispanic (n=18)	10th percentile	0.646	No
Asian (n=38)	10th percentile	1.000	No
Native American (n=8)	10th percentile	0.724	No
African-American (n=31)	20th percentile	0.344	No
Hispanic (n=18)	20th percentile	0.811	No
Asian (n=38)	20th percentile	0.768	No
Native American (n=8)	20th percentile	0.289	No
African-American (n=31)	30th percentile	0.308	No
Hispanic (n=18)	30th percentile	0.470	No
Asian (n=38)	30th percentile	0.309	No
Native American (n=8)	30th percentile	0.108	No
African-American (n=31)	40th percentile	0.346	No
Hispanic (n=18)	40th percentile	0.931	No
Asian (n=38)	40th percentile	0.402	No
Native American (n=8)	40th percentile	0.176	No
African-American (n=31)	50th percentile	0.067	No
Hispanic (n=18)	50th percentile	0.885	No
Asian (n=38)	50th percentile	0.114	No
Native American (n=8)	50th percentile	0.046	Yes

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the p-value was < 0.05.

Table 3. *Study 1 Adverse Impact Analysis Ethnicity: Fisher's Exact Test*

Ethnicity	Cutoff	p-value	Evidence of Adverse Impact?
African-American (n=31)	10th percentile	0.751	No
Hispanic (n=18)	10th percentile	0.420	No
Asian (n=38)	10th percentile	1.000	No
Native American (n=8)	10th percentile	1.000	No
African-American (n=31)	20th percentile	0.350	No
Hispanic (n=18)	20th percentile	0.771	No
Asian (n=38)	20th percentile	0.833	No
Native American (n=8)	20th percentile	0.209	No
African-American (n=31)	30th percentile	0.309	No
Hispanic (n=18)	30th percentile	0.439	No
Asian (n=38)	30th percentile	0.267	No
Native American (n=8)	30th percentile	0.057	No
African-American (n=31)	40th percentile	0.335	No
Hispanic (n=18)	40th percentile	0.809	No
Asian (n=38)	40th percentile	0.379	No
Native American (n=8)	40th percentile	0.144	No
African-American (n=31)	50th percentile	0.056	No
Hispanic (n=18)	50th percentile	0.808	No
Asian (n=38)	50th percentile	0.084	No
Native American (n=8)	50th percentile	0.027	Yes

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the p-value was < 0.05.

Table 4. *Study 1 Adverse Impact Analysis Ethnicity: Bootstrapping*

			Confidence Interval		
Ethnicity	Cutoff	Selection Ratio	Lower	Upper	Evidence of Adverse Impact?
African-American (n=31)	10th percentile	1.042	0.924	1.136	No
Hispanic (n=18)	10th percentile	0.928	0.708	1.114	No
Asian (n=38)	10th percentile	0.997	0.873	1.108	No
Native American (n=8)	10th percentile	1.114	1.071	1.164	No
African-American (n=31)	20th percentile	1.117	0.938	1.278	No
Hispanic (n=18)	20th percentile	1.069	0.816	1.294	No
Asian (n=38)	20th percentile	1.047	0.869	1.216	No
Native American (n=8)	20th percentile	1.283	1.205	1.375	Yes
African-American (n=31)	30th percentile	1.164	0.917	1.411	No
Hispanic (n=18)	30th percentile	1.169	0.849	1.470	No
Asian (n=38)	30th percentile	1.147	0.923	1.381	No
Native American (n=8)	30th percentile	1.503	1.385	1.647	Yes
African-American (n=31)	40th percentile	1.187	0.863	1.514	No
Hispanic (n=18)	40th percentile	1.070	0.652	1.486	No
Asian (n=38)	40th percentile	1.152	0.863	1.455	No
Native American (n=8)	40th percentile	1.533	1.011	1.909	No
African-American (n=31)	50th percentile	1.425	1.011	1.858	No
Hispanic (n=18)	50th percentile	1.104	0.585	1.676	No
Asian (n=38)	50th percentile	1.337	0.960	1.762	No
Native American (n=8)	50th percentile	1.933	1.255	2.466	Yes

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the confidence interval selection ratio was < 0.8 or greater than 1.2.

Table 5. *Study 1 Adverse Impact Analysis Ethnicity: Permutation Tests*

Ethnicity	Cutoff	p-value	Evidence of Adverse Impact?
African-American (n=31)	10th percentile	0.563	No
Hispanic (n=18)	10th percentile	0.414	No
Asian (n=38)	10th percentile	1.000	No
Native American (n=8)	10th percentile	0.605	No
African-American (n=31)	20th percentile	0.259	No
Hispanic (n=18)	20th percentile	0.773	No
Asian (n=38)	20th percentile	0.674	No
Native American (n=8)	20th percentile	0.207	No
African-American (n=31)	30th percentile	0.232	No
Hispanic (n=18)	30th percentile	0.439	No
Asian (n=38)	30th percentile	0.267	No
Native American (n=8)	30th percentile	0.057	No
African-American (n=31)	40th percentile	0.336	No
Hispanic (n=18)	40th percentile	0.807	No
Asian (n=38)	40th percentile	0.374	No
Native American (n=8)	40th percentile	0.139	No
African-American (n=31)	50th percentile	0.056	No
Hispanic (n=18)	50th percentile	0.814	No
Asian (n=38)	50th percentile	0.085	No
Native American (n=8)	50th percentile	0.027	Yes

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the p-value was < 0.05.

Table 6. *Study 1 Overall Results of Adverse Impact Analysis Ethnicity*

		Evidence of Adverse Impact?				
Ethnicity	Cutoff	4/5ths Rule	Chi-square test of Independence	Fisher's Exact Test	Bootstrapping	Permutation Tests
African-American (n=31)	10th percentile	No	No	No	No	No
Hispanic (n=18)	10th percentile	No	No	No	No	No
Asian (n=38)	10th percentile	No	No	No	No	No
Native American (n=8)	10th percentile	No	No	No	No	No
African-American (n=31)	20th percentile	No	No	No	No	No
Hispanic (n=18)	20th percentile	No	No	No	No	No
Asian (n=38)	20th percentile	No	No	No	No	No
Native American (n=8)	20th percentile	Yes	No	No	Yes	No
African-American (n=31)	30th percentile	No	No	No	No	No
Hispanic (n=18)	30th percentile	No	No	No	No	No
Asian (n=38)	30th percentile	No	No	No	No	No
Native American (n=8)	30th percentile	Yes	No	No	Yes	No
African-American (n=31)	40th percentile	No	No	No	No	No
Hispanic (n=18)	40th percentile	No	No	No	No	No
Asian (n=38)	40th percentile	No	No	No	No	No
Native American (n=8)	40th percentile	Yes	No	No	No	No
African-American (n=31)	50th percentile	Yes	No	No	No	No
Hispanic (n=18)	50th percentile	No	No	No	No	No
Asian (n=38)	50th percentile	Yes	No	No	No	No
Native American (n=8)	50th percentile	Yes	Yes	Yes	Yes	Yes

Notes. Total sample has an n=353. The majority population had an n=254.

Table 6. *Study 1 Overall Results of Adverse Impact Analysis Ethnicity* 50th percentile

Ethnicity	Cutoff	Evidence of Adverse Impact?				
		4/5ths Rule	Chi-square test of Independence	Fisher's Exact Test	Bootstrapping	Permutation Tests
African-American (n=31)	10th percentile	No	No	No	No	No
Hispanic (n=18)	10th percentile	No	No	No	No	No
Asian (n=38)	10th percentile	No	No	No	No	No
Native American (n=8)	10th percentile	No	No	No	No	No
African-American (n=31)	20th percentile	No	No	No	No	No
Hispanic (n=18)	20th percentile	No	No	No	No	No
Asian (n=38)	20th percentile	No	No	No	No	No
Native American (n=8)	20th percentile	Yes	No	No	Yes	No
African-American (n=31)	30th percentile	No	No	No	No	No
Hispanic (n=18)	30th percentile	No	No	No	No	No
Asian (n=38)	30th percentile	No	No	No	No	No
Native American (n=8)	30th percentile	Yes	No	No	Yes	No
African-American (n=31)	40th percentile	No	No	No	No	No
Hispanic (n=18)	40th percentile	No	No	No	No	No
Asian (n=38)	40th percentile	No	No	No	No	No
Native American (n=8)	40th percentile	Yes	No	No	No	No
African-American (n=31)		Yes	No	No	No	No
Hispanic (n=18)	50th percentile	No	No	No	No	No
Asian (n=38)	50th percentile	Yes	No	No	No	No
Native American (n=8)	50th percentile	Yes	Yes	Yes	Yes	Yes

Notes. Total sample has an n=353. The majority population had an n=254.

Table 7. *Study 2 Adverse Impact Analysis Gender: 4/5ths rule*

Dataset	Selection Ratio	Evidence of Adverse Impact?
Dataset 1 (n=28)	0.500	Yes
Dataset 2 (n=645)	0.917	No

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the selection ratio was < 0.8 or greater than 1.2.

Table 8. *Study 2 Adverse Impact Analysis Gender: Chi-square test of Independence*

Ethnicity	p-value	Evidence of Adverse Impact?
Dataset 1 (n=28)	0.801	No
Dataset 2 (n=645)	0.486	No

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the p-value was < 0.05.

Table 9. *Study 2 Adverse Impact Analysis Gender: Fisher's Exact Test*

Ethnicity	p-value	Evidence of Adverse Impact?
Dataset 1 (n=28)	0.639	No
Dataset 2 (n=645)	0.453	No

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the p-value was < 0.05.

Table 10. *Study 2 Adverse Impact Analysis Gender: Bootstrapping*

		Confidence Interval		
Ethnicity	Selection Ratio	Lower	Upper	Evidence of Adverse Impact?
Dataset 1 (n=28)	0.500	0	2.4	Yes
Dataset 2 (n=645)	0.917	0.735	1.140	No

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the confidence interval selection ratio was < 0.8 or greater than 1.2.

Table 11. *Study 2 Adverse Impact Analysis Gender: Permutation Tests*

Ethnicity	p-value	Evidence of Adverse Impact?
Dataset 1 (n=28)	0.638	No
Dataset 2 (n=645)	0.452	No

Notes. Total sample has an n=353. The majority population had an n=254. It was considered evidence of adverse impact if the p-value was < 0.05.

Table 12. *Study 2 Overall Results of Adverse Impact Analysis Gender*

	Evidence of Adverse Impact?				
Ethnicity	4/5ths Rule	Chi-squar	Fisher's E	Bootstrap	mutation T
Dataset 1 (n=28)	Yes	No	No	Yes	No
Dataset 2 (n=645)	No	No	No	No	No

Notes. Total sample has an n=353. The majority population had an n=254.