

INCOME INEQUALITY AMONG RACES: A STATISTICAL ANALYSIS

*Tyler T. Yu, Mercer University
Miranda M. Zhang, Mercer University*

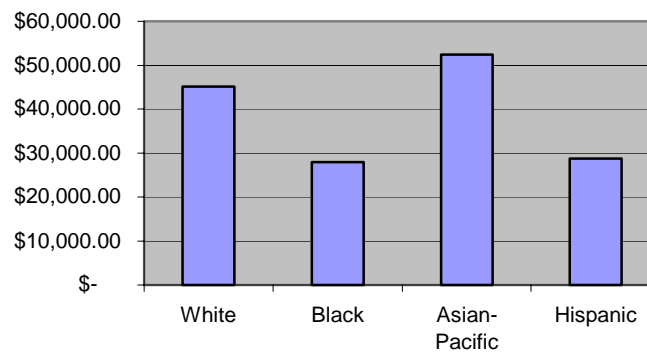
ABSTRACT

This paper examines income distribution among races in the United States. Using median income data from the U. S. census statistics for Caucasian, Black, Asian-Pacific Islander, and Hispanic families for the years 1990-1999, the paper first determines whether the income diversity among race groups is statistically significant. Then the correlation of the income diversity and education and marital status, two most commonly cited factors contributing to this diversity. Finally, some conclusions are drawn based on the empirical results of the study.

INTRODUCTION

The economic benefits of the boom in the last decade have been spread to all parts of the society but in an extremely uneven pattern. According to a survey conducted by the Federal Reserve at the end of 2001 in the midst of the economic recession, economic inequality grew markedly even as incomes increased at almost all levels (Andrews [1]). There have been numerous studies with reference to income inequality among different racial groups over the past several years. These studies have attempted to explain how various socioeconomic factors or conditions affect income distribution. Until the dawn of the 21st century, it was evident that there had been an apparent gap among Caucasians, Blacks, Asian-Pacific Islanders, and Hispanics in the United States, as shown in Figure 1. The crucial questions are 1) whether or not the difference is statistically significant and 2) the possible causes of the difference.

Figure 1
Average of Income



Lerman [6] examined the relationship between marriage and household income. The author argued that marriage improved the income or economic status of households. The study identified and discussed several reasons why households with married couples tended to have higher income. According to the research, married couples could have dual incomes and increased ability to share economic risks. The research also suggested that married men tended to work additional hours to meet their family needs if necessary.

In the paper, "Education, Income Inequality, and Mortality: A Multiple Regression Analysis," Muller [7] studied the cause and effect relationship among education, income inequality, and mortality. Using data gathered from the US Census Statistics and the Gini coefficient (a measure of equality of income distribution), the study concluded that there was a strong causal relationship between education and income inequality. This causal relationship resulted in different levels of mortality. For example, people who did not have high school education were more likely to have a lower level of income and work on jobs that were high risk to life and/or health.

In "Black White wage Inequality in the 1990's: A Decade of Progress," Kenneth Couch and Mary Daly studied the wage gap between black and white males and found the gap to have declined during the 1990s at a rate of about 0.60 percentage points per year (Couch and Daly [5]). The research used the standard decomposition methods to study the gap trend concluded that "although overall male wage inequality became less of a drag on the relative wages of blacks during the 1990's, it continued to temper the convergence in the black-white wage gap."

The purpose of this paper is to determine, by using of hypothesis testing, whether the income diversity among between Caucasians, Blacks, Asian-Pacific Islanders, and Hispanics is statistically significant. If so, we will then estimate the correlation of the income diversity and education and marital status, two most commonly cited factors contributing to this diversity.

DATA AND MODELS

This research uses median income data from the U. S. census statistics for Caucasian, Black, Asian-Pacific Islander, and Hispanic families for the years 1990-1999. We will first determine whether the income diversity among race groups is statistically significant. Table 1 displays the median income per household of each race group.

As indicated in the table, there are disparities among the race groups for the study period. However, to determine whether the differences are statistically significant, an ANOVA analysis will be performed. An important part of the analysis involves the ANOVA and Kruskal-Wallis tests (Anderson, Sweeney, and Williams [2]). One can use these tests to determine whether the income are different among the four races: Caucasian, Black, Asian-Pacific Islander, and Hispanic families for the years 1990-1999. Whether the differences are statistically significant or not is a question answered with the aid of the above mentioned tests. Like other hypothesis-testing procedures, these tests compare sample results with those that are expected when the null hypothesis is true. The hypotheses for the ANOVA test are:

Ho: $\mu_w = \mu_b = \mu_h = \mu_a$ (where μ_x represents income of each ethnic group)
 Ha: $\mu_w \neq \mu_b \neq \mu_h \neq \mu_a$ (the U's are not all equal)

Where: Ho is the null hypothesis
 Ha is the alternative hypothesis
 U's are the treatment means

Table 1
Median Income, 1990 - 1999 By Ethnic Group

	Caucasian	Black	Asian-Pacific	Hispanic
1990	\$ 45,064	\$ 27,307	\$ 53,850	\$ 29,867
1991	\$ 43,961	\$ 26,538	\$ 50,119	\$ 29,228
1992	\$ 43,429	\$ 25,059	\$ 50,176	\$ 27,971
1993	\$ 42,612	\$ 24,837	\$ 51,255	\$ 27,272
1994	\$ 43,597	\$ 27,764	\$ 51,848	\$ 27,337
1995	\$ 44,395	\$ 28,390	\$ 50,675	\$ 26,859
1996	\$ 44,915	\$ 28,162	\$ 52,141	\$ 27,797
1997	\$ 46,262	\$ 29,689	\$ 53,821	\$ 29,212
1998	\$ 47,769	\$ 30,052	\$ 53,993	\$ 30,262
1999	\$ 48,950	\$ 31,778	\$ 56,316	\$ 31,663

The U. S. Census statistics

The null hypothesis states that the means for the income of the four categories of races are equal. The research hypothesis charges that at least one of the four categories is significant different from others.

The ANOVA model was constructed using time-series data for the four categories of races. These data (10 observations for each category) were collected for the years 1990-1999 from the U.S. Census statistics.

The F-test at a 1% level of significance led to the rejection of the null hypothesis and to the conclusion that the difference among the four categories in family income is statistically significant. The result of the ANOVA test is summarized and reported in Table 2.

Table 2
Anova Results

Source	d.f.	Sum of Squares	Mean Square	F-ratio	Sig.
Between	3	4.44E+09	1478346978	386.4	0.000
Within	36	1.38E+08	3825648		
Total	39	4.57E+09			

The Critical F [1%, 3, 36] = 4.31

Since the calculated F-Ratio is greater than the critical F value at a 99% confidence level, i.e., 386.4 is greater than 4.31, *i.e.*, $p < 0.01$, the null hypothesis must be rejected. Rejecting the null in this case implies that the difference among population means is statistically significant.

The primary underlying assumption of the ANOVA analysis is that the populations under study are distributed normally with equal variance. While there is no indication that the populations are not normally distributed, it is nonetheless scientifically sound to also run the Kruskal-Wallis test, the nonparametric alternative to ANOVA, which does not require the assumptions of normality and equal variance. The hypotheses of the Kruskal-Wallis test are:

Ho: $\mu_w = \mu_b = \mu_h = \mu_a$ (where μ_x represents income of each ethnic group)
 Ha: $\mu_w \neq \mu_b \neq \mu_h \neq \mu_a$ (the U's are not all equal)

Where: Ho is the null hypothesis
 Ha is the alternative hypothesis and
 μ 's are the population means.

We can now perform the Kruskal-Wallis test, using the same data for the four categories of races. The test, once again, will determine whether the difference among four population means is statistically significant. The result of the Kruskal-Wallis test is summarized and reported in Table 3.

Table 3
Kruskal-Wallis Test Results

Levels	Observations	Mean Rank	Chi-Square	Asymptotic Significance
Caucasian	10	25.50		
Black	10	9.50		
Hispanic	10	11.50	33.073	0.000
Asian	10	35.50		$p < 0.01$
Total	40			

The critical H value [(x squared), 3, 0.01] = 11.35

The Chi-square test at a 1% level of significance led to the rejection of the null hypothesis and to the conclusion that the difference is significant. That is, since the calculated Chi-square, 33.073, is greater than the critical Chi-square, 11.35, therefore, we must reject the null hypothesis.

Tukey HSD and Bonferroni methods were used to conduct the multiple comparisons of the mean differences among the four categories. That is, these two methods were used to determine “where” the differences come from. As a result of the comparison, the differences among all pairs of races are statistically significant except the difference between African Americans and Hispanics. The results of both methods are reported in Table 4.

Table 4
Multiple Comparisons Of The Income Differences:
Tukey Hsd And Bonferroni Methods

Category	Caucasian	Black	Hispanic	Asian
Caucasian	N/A	Significantly different	Sign. different	Sign. different
Black	Significantly different	N/A	Not sign. difference	Sign. different
Hispanic	Significantly different	Not Sign. difference	N/A	Sign. different
Asian	Significantly different	Sign. different	Sign. different	N/A

CONTRIBUTING FACTORS

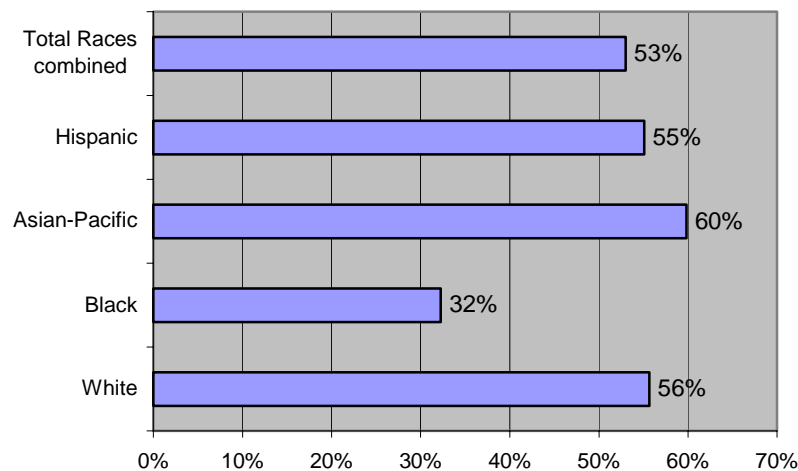
In the previous section, we have shown that the income disparity among the races is statistically significant. The next step is to identify the contributing factors causing the disparity. Based on the review of literature, we have identified two factors commonly believed as the causes of uneven distribution of income: education and marital status. In this study, these two factors will be tested to see if they are indeed correlated with the income disparity among races.

Marital Status

According to Gary Becker, marriage makes families better off partly by allowing individuals with families to specialize, which yields greater productivity, hence income (Becker [3], Lerman [6]). In addition to specialization, the sharing of economic and social resources in marriage yields economies of scale and provides a risk-sharing protection against unexpected events (Oppenheimer [7], Lerman [6]). The economies of scales arise because the costs of maintaining a household are nearly fixed (ex. Housing, heating, transportation, etc.)

During the period from 1990 – 1999, there was a large dispersion in married households among the races. As shown in Figure 2, in 2000, 52% of all households consisted of married couples. It also displays that the majority of White and Asian households consisted of married couples, while nearly half of Hispanics households were married. In contrast, only about one-third of black households consisted of married couples, a rate well below the national average.

Figure 2
Marital Status by Race, 2000



Education

It is commonly postulated that education is also a relevant factor that affects income inequality. As pointed out by Muller [7], the cause and effect relationship between education and income inequality is statistically significant. The question here is that whether or not education is also correlated to the income inequality among the races? Table 5 provides information on high school completion ratios and college graduation ratios for the four races included in the study.

Table 5 shows that as a general trend, education levels of all races have increased in the last decade. The percentage increase in the proportion of a race completed high school is 6.5% for white, 16.7% for black, 5.3% for Asians, and 10.4% for Hispanics, respectively. The percentage increase in the proportion of a race completed college is 17.7% for white, 3.6% for black, 6.2% for Asians, and 18.5% for Hispanics, respectively.

Table 5
Completion Of High School And College By Races (%), 1990-1999

Year	Caucasian		Blacks		Asians		Hispanics	
	Hi. School	College	Hi. School	College	Hi. School	College	Hi. School	College
1990	79.1%	22%	66.2%	11.3%	80.4%	39.9%	50.8%	9.2%
1995	83%	24%	73.8%	13.2%			53.4%	9.3%
1997	83%	24.6%	74.9%	13.3%	84.9%	42.2%	54.7%	10.3%
1998	83.7%	25%	76%	14.7%			55.5%	11%
1999	84.3%	25.9%	77%	15.4%	84.7%	42.4%	56.1%	10.9%

Source: the U.S. Census Bureau Statistical Abstract

The relationships between income inequality and marital status as well as education were investigated using the Pearson Product-Moment Correlation coefficients as well as Spearman's Rank Correlation coefficients. The Pearson product moment correlation coefficient, R_{xy} , is defined as follows:

$$R_{xy} = S_{xy} / (S_x S_y) \text{ or } R_{xy} = \frac{[\sum X_i Y_i - (\sum X_i \sum Y_i) / n]}{[(\sum X_i^2 - (\sum X_i)^2 / n) \sum Y_i^2 - (\sum Y_i)^2 / n]} \quad (1)$$

where

S_{xy} = sample covariance

S_x = sample standard deviation of x

S_y = sample standard deviation of y

X_i = the i th observation of x

Y_i = the i th observation of y

The Pearson Product-Moment Correlation method requires that population has a normal distribution. Although there is no reason to believe that the population does not have normal distribution, we want to make certain about the result by applying the Spearman rank-correlation method, the nonparametric alternative to the Pearson Product-Moment Correlation method (Anderson, Sweeney, and Williams [2]. The Spearman rank-correlation coefficient, R_{xy} , is as follows:

$$R_{xy} = 1 - (6 \sum d_i^2) / n(n^2 - 1) \quad (2)$$

where

n = the number of items or individuals being ranked

X_i = the number of item i with respect to one variable

Y_i = the rank of item i with respect to a second variable

d_i = X_i - Y_i

Using SPSS, these two methods were performed on two sets of data: 1) income of races vs. marital status; and 2) income of races vs. percentage of high school completion rate; and 3) income of races vs. percentage of college graduation rate. The results were reported in Table 6.

Table 6
The Results Of The Pearson Product-Moment Correlation Coefficient And The Spearman Rank-Correlation Coefficient

	Pearson Product-Moment Coefficient	Spearman's Rank Correlation Coefficient Correlation
Rx1y	0.669	0.80
Rx2y	0.754	1.00
Rx3y	0.958	1.00

where

Rx1y ⇒ income of the races vs. marital status of the races

Rx2y ⇒ income of the races vs. high school education of the races

Rx3y ⇒ income of the races vs. college education of the races

As shown in Table 7, the correlation coefficients generated by both methods are positive, indicating direct relationship between income level of the races and marital status, high school completion ratios, and college graduation ratios. Also, the coefficients have magnitudes large enough to indicate a strong relationship.

Hypothesis Test

The hypotheses can be formulated as follows:

H₀: P_s = 0

H_a: P_s ≠ 0

The null hypothesis, H₀, states that there is no rank correlation (P_s = 0) between income level and marital status and education, whereas the alternative hypothesis, H_a, says that there is a rank correlation between the two pairs, (P_s ≠ 0).

Under the hypothesis of no rank correlation ($\rho_s = 0$), the rankings are independent, and the sampling distribution of P_s is as follows

$$\begin{aligned} U_{rxy} &= 0 \\ T_{rxy} &= \frac{U_{rxy}}{\sqrt{1/n-1}} \end{aligned}$$

Where $n \geq 10$, for normal distribution

The empirical results of the tests indicate that there is a significant positive correlation between education levels (both high school and college) and income of the races, but no significant correlation between income and marital status.

CONCLUSION

The distribution of income and the extent of inequality are two of the controversial issues of our times concerned by policy makers, economists and the society as a whole. This paper examined the issues from race/ethnicity perspective. We first applied ANOVA test and the Kruskal-Wallis test to the incomes of Caucasian, Black, Asian-Pacific Islander, and Hispanic families to determine the significance of the differences. The results of both tests indicated that the differences were statistically significant. Specifically, the significant differences were found, by using Tukey HSD and Bonferroni methods, among incomes of Caucasian, Black, Asian-Pacific Islander, and Hispanic families except for the difference between incomes of Black and Hispanic, which was not significant.

We then measured the correlation between the incomes of races and two most commonly cited contributing factors for income inequality, marital status education levels. To do so, we applied both The Pearson Product-Moment Correlation Coefficient and its nonparametric alternative - Spearman Rank-Correlation Coefficient to estimate the correlation coefficients and to conduct hypothesis testing. The results revealed that there was a significant correlation between income and education, and contrary to the popular wisdom, the same was not true for income and marital status.

REFERENCES

- Andrews, Edmund L., "Economic Inequality Grew in 90's Boom, Fed Reports," *New York Times*, January 23, 2003.
- Anderson, David R.; Sweeney Dennis J; and Thomas A Williams, *Statistics for Business and Economics*, Seventh Edition, South-Western College Publishing, 1999.
- Becker, Gary, *A Treatise on the Family*, Harvard University Press, Cambridge, Massachusetts, 1981.
- Center on Budget and Policy Priorities 2000, "Poverty Rate Hits Lowest Level Since 1979 As Unemployment Reaches a 30-Year Low," <http://www.cbpp.org/9-26-00pov.htm> (8/17/2003) .
- Couch, Kenneth and Mary C. Daly, "Black-White Wage Inequality in the 1990s: A Decade of Progress," *Economic Inquiry*, January, 2002, 31-41.

- Lerman, Robert I. "Marriage and Economic Well being of Families with Children: A Review of the Literature," July 2002. <http://www.urban.org> (8/22/2003).
- Muller, Andreas, "Education, Income Mortality: A Multiple Regression Analysis," *the British Medical Journal*, January, 2002. <http://proquest.umi.com> (8/22/2003).
- Oppenheimer, Valerie, "The Continuing Importance of Men's Economic Positions in Marriage Formation," *The Ties That Bind*, edited by Linda Waite, 283-301, New York: Aldine De Gruyter, 2000.
- The U.S. Census Bureau Statistical Abstract.

