EMPIRICAL COMPARISON OF U.S. CENSUS BUREAU POPULATION ESTIMATES USED IN MORTALITY AND POPULATION DATA SYSTEM OF THE UNIVERSITY OF PITTSBURGH, DEPARTMENT OF BIOSTATISTICS

by

Weilian Sang

BS, Nanjing University, 1997

MS, Nanjing University, 2000

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis is submitted

by

Weilian Sang

It was defended on

December 7, 2004

and approved by

Essay Advisor: Gary M. Marsh, Ph.D
Professor
Department of Biostatisticics
Graduate School of Public Health
University of Pittsburgh

Committee Member: Ada Youk, Ph.D
Research Assistant Professor
Department of Biostatisticics
Graduate School of Public Health
University of Pittsburgh

Committee Member: Ravi Sharma, PhD
Assistant Professor
Department of Behavioral & Community Health Sciences
Graduate School of Public Health
University of Pittsburgh

Gary M. Marsh, PhD

EMPIRICAL COMPARISON OF U.S. CENSUS BUREAU POPULATION ESTIMATES USED IN MORTALITY AND POPULATION DATA SYSTEM OF THE UNIVERSITY OF PITTSBURGH, DEPARTMENT OF BIOSTATISTICS

Weilian Sang, MS

University of Pittsburgh, 2004

The current Mortality and Population Data System (MPDS) database contains the cause of death data and population data from 1950 to 2001, and it was designed to provide data for public health related studies. The cause of death data in the MPDS are provided by the National Center for Health Statistics and are updated annually as new cause of death data from NCHS are released. Since the actual annual population data is not available, the intercensal population figures have to be estimated based on the census population data. The population figures used in the MPDS were estimated by using year-based linear interpolations and extrapolations at the county level, while the census bureau used the cohort-component method to estimate the population. The purpose of this thesis is to compare these two population estimates using two approaches, 1) determine if there are any important differences between them, and 2) evaluate the effect of the difference on the calculation of the mortality rates. The results showed that at national, state and county level, sex was a factor that contributes to the difference between the two populations, while other factors such as year, race, and age group did not affect the difference greatly. The difference between the two population estimates mainly comes from the difference between the female groups of the two populations. The effect of the difference on the calculation of the standardized mortality ratios (SMRs) was analyzed by using data from an occupational cohort study. The results from the analysis of the occupational cohort data showed that the significance of the SMRs for each cause of death was not different when using different rates from the two population estimates. The 95% confidence intervals for the SMRs for the major categories of cause of death overlap. The SMRs calculated with new and old population estimates as reference populations were not significantly different.

Key words: population estimation, MPDS, Standard Mortality Rate (SMR), difference ratio (DR), OCMAP

TABLE OF CONTENTS

LIST OF ACRONYMS

DR: Difference Ratio

FIPS: Federal Information Processing Standards Publications

IRS: Internal Revenue Service

MPDS: Mortality and Population Data System

NCHS: National Center for Health Statistics

PDB: Population Database

SMR: Standardized Mortality Ratio

LIST OF TABLES

LIST OF FIGURES

PREFACE

I cannot thank my thesis advisor Professor Gary Marsh sufficiently for his patience, guidance, and providing motivation. I would also like to thank Dr. Ada Youk for all the guidance through all the time I am studying here. I am grateful to Dr. Ravi Sharma for being my committee member and giving me guidance through my completion of this thesis. I would like to thank the faulty in the Graduate School of Public Health for their teaching.

I also want to thank all the members of Dr. Marsh's group, who have been very nice to me and gave me a lot of helps during my stay here. I wish to thank all my friends for their support and help.

Last, but not the least, I dedicate this thesis to my parents and my siblings for their unwavering dedication and support. I want to acknowledge the love and support of my husband, Wei Hu, who has given me not only cares to my life, but also invaluable comments to my work these years.

1. INTRODUCTION

The current Mortality and Population Data System (MPDS) database, maintained by Department of Biostatistics at University of Pittsburgh, was constructed during the period of 1982-1984, and it contains the cause of death data and population data from 1950 to 2001 (Marsh et al, 1998). The Mortality and population data system (MPDS) was designed to accomplish the following goals (the following section is taken from the MPDS manual):

1. Assemble a well-organized and easily accessible database containing detailed U.S. mortality and population data in a standardized format.
2. Provide a user-oriented retrieval mechanism to facilitate the extraction of information from the database
3. Provide an interface to the occupational cohort mortality analysis program (OCMAP) via death rate generation.

The cause of death data in the MPDS are provided by the National Center for Health Statistics (NCHS) and are updated annually as new cause of death data from NCHS are released. The individual death records include codes for sex, race, age of death, year of death, and geographic location (county and state of residence at the time of death). In MPDS, the death records are categorized and linked with the corresponding population data to form death rates specific for five-year age groups, five-year time periods, race, sex, geographic location and cause of death.

However, the annual population data is not available because the population census is not conducted every year, but every ten years. So the intercensal population figures have to be estimated based on the census population data. For years the population figures used in the MPDS were estimated by using year-based linear interpolations and extrapolations at the county level, while the census bureau used different methods to estimate the population.

1.1. Population estimation methods used in MPDS

The current population data used in the MPDS database were derived from census year-based linear interpolations and extrapolations at the county level, and the state and national population figures were from the aggregation of the county level population. The formula that was used to extrapolate the population is as follow, for any given race/sex/age group in each county, where:

POP (I) is the population at year I,

POP (I-1) is the population at previous year,

Then, the extrapolated population=POP (I) + [POP (I)-POP (I-1)]

1.2. Population estimation methods used by United States Census Bureau (http://www.census.gov)

The estimated postcensual population figures from the Census Bureau were derived by using different methods from the census year-based linear interpolations and extrapolations used for the current MPDS population data. Different levels of population figures (national, state, county, cities and town) were estimated independently. One of the guiding principles in the Census Bureau's subnational population estimates methodology is that all of the population estimates are consistent, which means that the sum of the county estimates must be equal to the independently produced state characteristics population estimates, the sum of the state estimates must be equal to the independently produced national characteristics population estimates.

1.2.1. National population estimation

Estimates of the United States resident population include persons resident in the 50 states and the District of Columbia, and the estimates were derived quarterly by updating the resident population enumerated in the census year through the components of the population change. The changes of the population include: (1) +births to U.S. resident women; (2) -deaths to the U.S. residents; (3) +net international migration; (4) +net movement of U.S. Armed Forces and civilian citizens. Births and deaths estimates are from birth certificates and death certificates data supplied by the National Center for Health Statistics (NCHS). International migration data come from the survey data and the data supplied by the Immigration and Naturalization Service (INS).

1.2.2. State population estimation

The U.S. Census Bureau produces estimates of the resident population by age, sex, race and Hispanic origin for each state in the United States on an annual basis. The demographic procedure used to do the state population estimation is called a cohort-component method and it is based on the traditional demographic accounting system. This method follows each birth cohort across time according to its exposure to mortality, fertility and migration. Each component of the population change was estimated separately. Starting with a base population, deaths are subtracted from the population and births are added to the population, forming new cohorts. Estimates of net international migration and net internal migration are added to or subtracted from the population, and the components of change are measured separately by age, sex, race and Hispanic origin for each state and added to the base population as: P1=P0+B-D+NDM+NIM, where :

P1=population at the end of the period

P0=population at the beginning of the period

B=births during the period

D=deaths during the period

NDM=net internal migration during the period

NIM=net international migration during the period

A major assumption of this approach is that the components change can be closely approximated by administrative data in a demographic change model. The administrative data used by the Census Bureau for the population estimation include birth certificates, death certificates, Internal Revenue Service (IRS) data, Medical enrollment records, Armed Forced data, group-quarters population data, and data derived from the America Community Survey, Social Security files, Census data, and other internal Census Bureau data and so on.

1.2.3. County population estimation

The resident population estimates for the 3141 counties in the United States was produced by age, sex, race and Hispanic organic on an annual basis. The Census Bureau develops the county population estimates with the cohort-component population method, as described in state population estimation.

2. STATEMENT OF THE PROBLEM

One of the many goals of the current MPDS is to provide the specific mortality rates. The calculation of the specific mortality statistics requires the cause of death data and the estimates of the population at risk. Because the intercensual population data come from estimation, the accuracy of the population estimation could affect the accuracy the mortality rates calculated from the estimated population (Anderson, RN et al, 2003). In past years MPDS database has been using its own population estimates, which were estimated using extrapolation method, not the population estimates from the Census Bureau. The main questions are: 1) are the two sets of the population estimates close enough to ignore the difference? And 2) does the difference between the two population estimates affect the calculation of the death rates?

The purpose of this thesis is to compare the two sets of estimated populations at the national level, state level and the county level. The one was estimated by using the linear interpolations and extrapolations method used in the current MPDS and the other one was estimated by the cohort-component population method used by the Census Bureau. The comparison will be used to see whether the two sets of population data estimated by the two different approaches have a very large difference and to see whether this difference would greatly affect the calculation of the standardized mortality ratios (SMRs).

3. MATERIALS AND METHODS

3.1. Population Data

The new census 2000 data along with the captured intercensal data from 1970 through 1999 estimated by Census Bureau were taken and assembled into a master update file for the Population Database (PDB). And then the current PDB was converted into ASCII and merged with the master update file with the population data from Census Bureau. Multiple output files were generated by this process, the matched/unmatched key datasets. The output of the unmatched key datasets is due to FIPS changes over years. For example, effective 01/01/1983, La Paz AZ 04012 was spawned from Yuma 04047. Because these new FIPS codes are not in the MPDS, there is no estimation for these places. In this case, the unmatched data will show the missing populations in MPDS. All these matched/unmatched datasets share the same space-delimited format as: Geold Sex Race Year Age_Grp Cur_Popn New_Popn Cur_Popn-New_Popn, where Geold is the geo entities for each county, and Sex is code as 1(Male) and 2(Female), Race is coded as 1(White) and 2(Nonwhite), Year is from 1970 through 2000, indicating which year the population is, Age_Grp is the five-year age groups, and there are totally 18 age groups. Cur_Popn is the population figure currently used by MPDS for each group, New_Popn is the population figure from Census Bureau for each group, and finally Cur_Popn-New_Popn is the population difference for each group of these two sets of population. Another variable called diff_ratio was generated by: diff_ratio= (New_Popn -Cur_Popn)/ New_Popn.

Because the MPDS contains death data and population data from 1950 though 2001, to complete the update of the MPDS with the population data from the Census Bureau, the availability of the county level population data from 1950 through 1969 by Census Bureau was investigated. Several people from the office of Census Bureau were contacted and it was found that they did not do estimate on the county level at that period of time.

The two sets of population data were checked and it was found that the there are slight differences for the population data from 1970 through 1989. The only existing difference from that period of time is from the change on the FIPS coding. The reason is because the MPDS population data before 1990 was updated by using the population estimates from the Census

Bureau. Therefore this thesis will focus on the analysis of the two sets of population from 1990 though 1999, to see whether the difference between the two populations is large and whether that difference will impact the mortality calculations. Because the mortality rates in MPDS are calculated at the county levels, the difference at the county level will be the focus of the analysis. The difference ratio (DR) in the analysis is the difference of the two population figures (New_Popn -Cur_Popn) divided by the population estimated by the Census Bureau (New_Popn).

SAS version 8.2 was used for all descriptive statistics and test statistics. S-PLUS 6.1 and Excel 2000 were used for the graphic displays. Microsoft Access 2000 was used for the database management. OCMAP+ was used for the performance of the SMR analysis.

3.2. The cohort data

Data from an occupational cohort study was used to determine the effects of the different population estimates on the calculation of morality rates. The cohort contained 39,868 employees from 13 different plants who were employed at least one day between January 1, 1980 and December 31, 1998. These individuals accumulated n=574,478 person-years at risk.

4. DATA ANALYSIS AND RESULTS

The two sets of population data contain the national, state and county population figures. The dataset was separated into three datasets by using the database program Access query function, one is the dataset with all county population figures, one is the state population figures and the other one is the national population figures.

A preliminary analysis showed that the difference between the two populations among age group 6 to age group10 (age 25-49) is more significant than other age groups, so the age groups were combined into two age groups, agegroup1 with the agegrp6 to agegrp10, and agegroup2 from the originally agegrp1-5 and agegrp11-18.

4.1. Difference of the two populations nationally

The DRs at the national level by race, sex and age group for each year nationally were summarized, to show if the DRs marginally for different sexes, races, and age groups are different. (diff=new_pop-cur_pop, diff_ratio=diff/new_pop).

4.1.1. Total difference of the two populations by year

Table 1 and Figure 1 below indicate that the DRs for the two populations have an increasing trend from year 1990 to year 1999. In 1990, the difference is about 5.5%, while in year 1999, the difference increases to about 7.8%.

Table 1 Total difference between the two population estimates at the national level (1990-1999)

Year	New Population	Current Population	Difference	DR
1990	261214108	246802996	14411112	0.0552
1991	264634923	249465633	15169290	0.0573
1992	268310916	252328856	15982060	0.0596
1993	271932272	255108201	16824071	0.0619
1994	275319485	257653629	17665856	0.0642
1995	278641113	260133442	18507671	0.0664
1996	281945295	262491075	19454220	0.0690
1997	285479981	264849180	20630801	0.0723
1998	288921211	267208008	21713203	0.0752
1999	292342786	269567552	22775234	0.0779

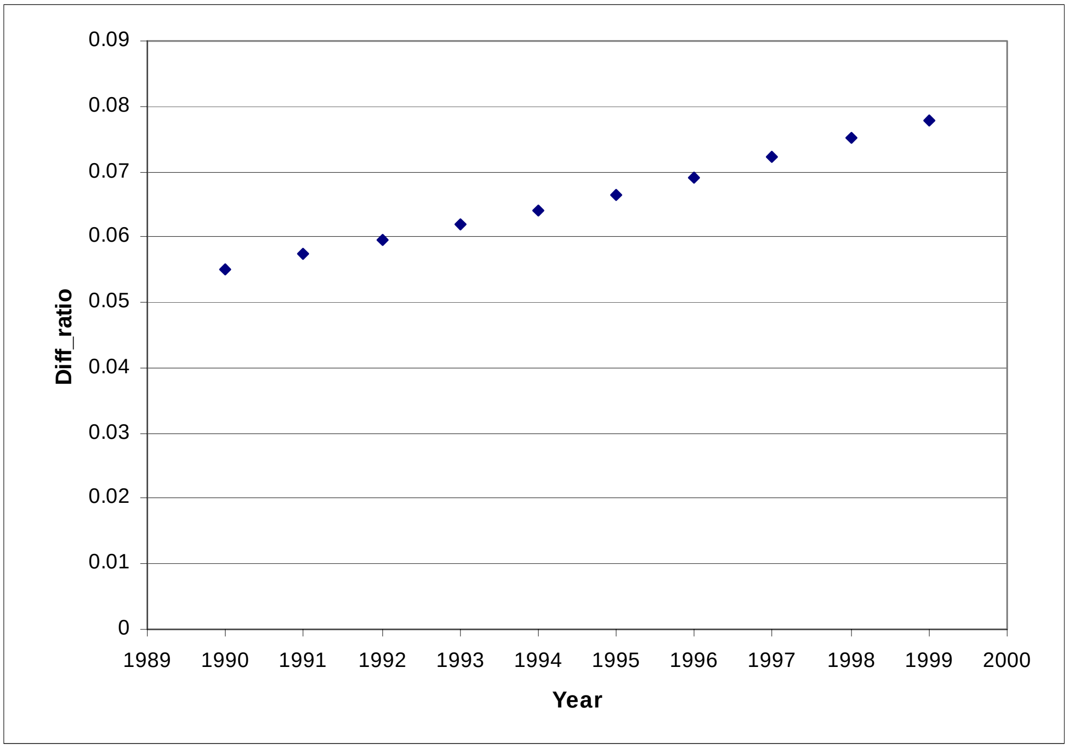


Figure 1 Total difference between the two population estimates at the national level (1990-1999)

4.1.2. Difference of the two populations by sex for each year

Table 2 and Figure 2 below show that the DR for female is much larger than that for male for all of these 10 years. The DRs for male are around zero (about -0.016) for all 10 years, while the DRs for female are all larger than 30%. Therefore, it appears that sex is an important factor that affects the difference of the two estimates of the national population. We can also see that the DRs for the female group increase from 1990 to 1999 slightly (31% to 36%), while for male, the DRs are stable, with the value a bit below 0. So the difference of the two groups comes from the difference between the female groups and the year trend of the differences is also due to the increase of the difference among the female group of the two populations.

Table 2 The difference ratio of the two population estimates at the national level by sex (1990-1999)

Year	Diff_ratio	
	Male	Female
1990	-0.0170	0.3107
1991	-0.0168	0.3148
1992	-0.0167	0.3191
1993	-0.0167	0.3238
1994	-0.0166	0.3285
1995	-0.0165	0.3330
1996	-0.0163	0.3383
1997	-0.0157	0.3446
1998	-0.0153	0.3502
1999	-0.0150	0.3555

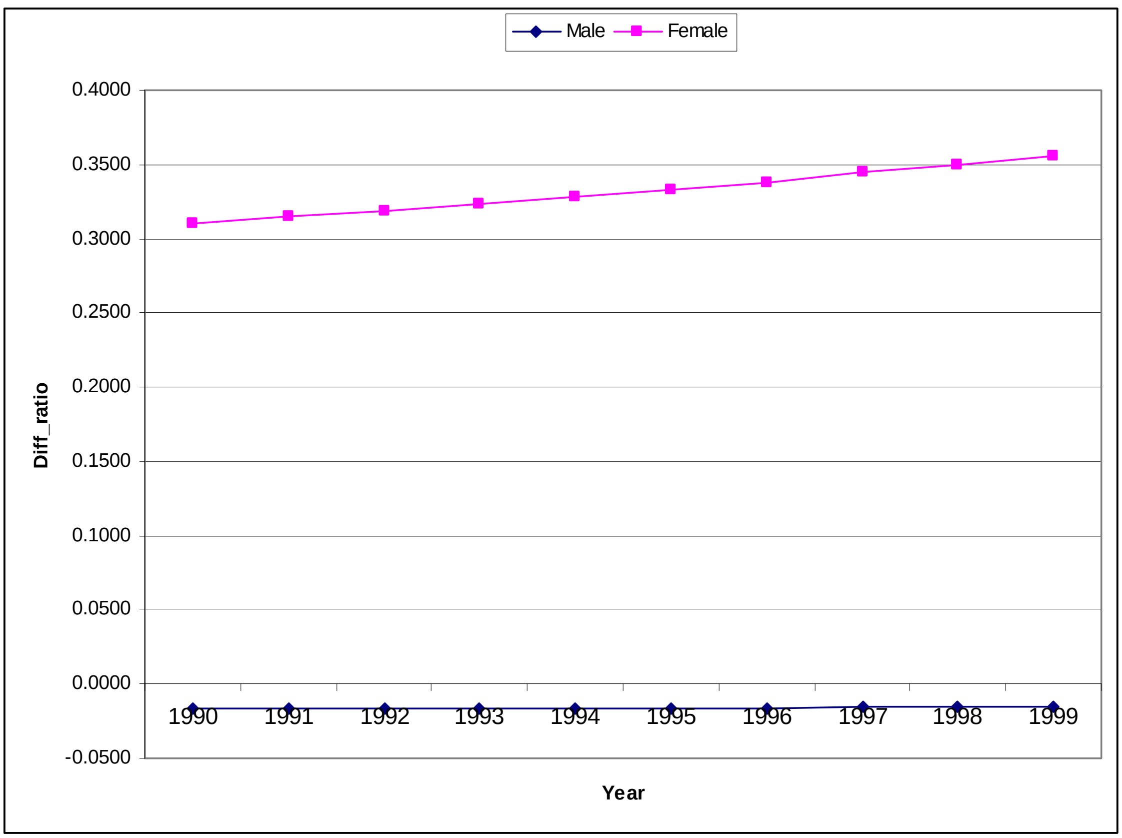


Figure 2 The difference ratio of the two population estimates at the national level by sex (1990-1999)

4.1.3. Difference by race for each year

From the figure of the DR vs. year by race below (Figure 3), it can be seen that the DRs for both Non-white and White are around 0.05 to 0.08 for these 10 years, with White slightly larger than that of Non_white. The differences are increasing for both Whites and Non_whites from 1990 to 1999. Because the differences for White and Non_white are very close for each year, race might not be an important factor to the difference of the two population estimates.

Table 3 The difference ratio of the two population estimates at the national level by race (1990-1999)

	DR	
Year	White	Non_white
1990	0.0597	0.0508
1991	0.0615	0.0533
1992	0.0633	0.0560
1993	0.0652	0.0587
1994	0.0671	0.0614
1995	0.0690	0.0639
1996	0.0712	0.0669
1997	0.0741	0.0705
1998	0.0767	0.0737
1999	0.0793	0.0766

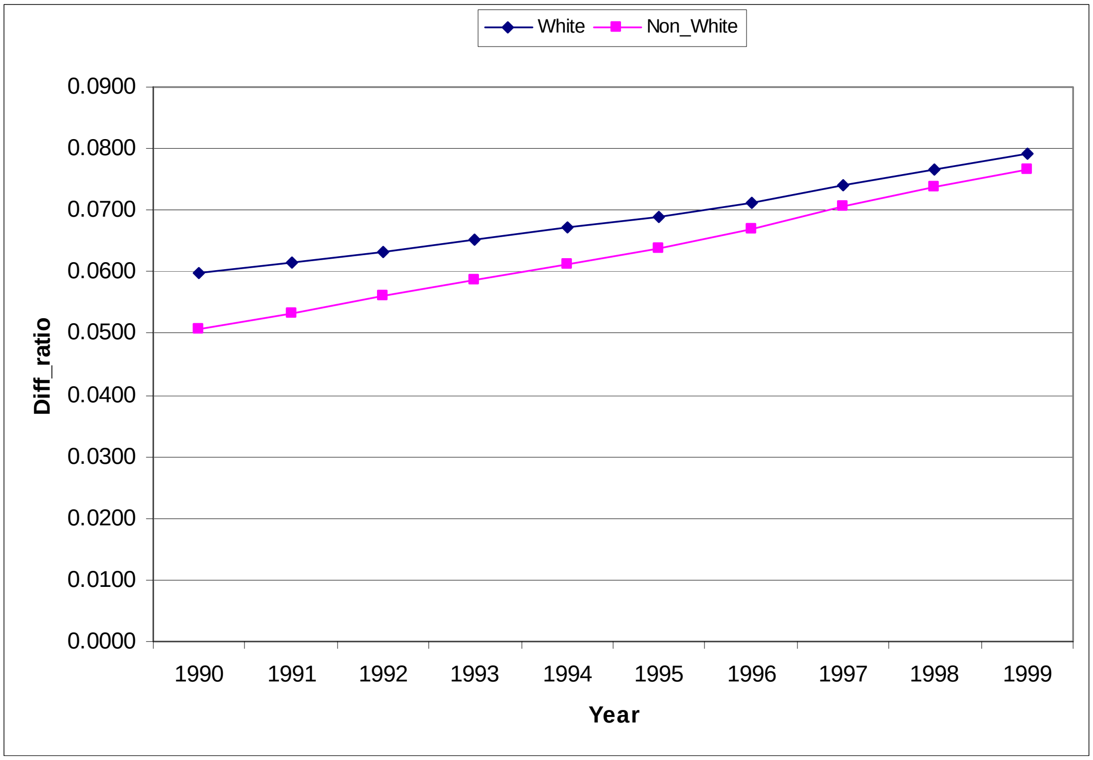


Figure 3 The difference ratio of the two population estimates at the national level by race (1990-1999)

4.1.4. Difference by 18 age groups for each year

Table 4 and Figure 4 below show the difference of the two populations for the 18 age groups. From 1990 to 1996, the DRs for all age groups are not very large (0-0.1), with younger age groups having big differences. From year 1996 to 1999, the DRs vary among the age groups. Different age groups are a potential factor for the estimation of the populations after 1996.

Table 4 The difference ratio of the two population estimates at the national level by age group (1990-1999)

Age Group	Year									
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
0-4	0.0903	0.094	0.0989	0.1036	0.1087	0.1133	0.1179	0.1244	0.1348	0.1473
5-9	0.0837	0.0851	0.087	0.0891	0.0912	0.0942	0.0987	0.1031	0.1011	0.0924
10-14	0.0793	0.0818	0.0841	0.0864	0.0887	0.0904	0.0921	0.0915	0.0937	0.1033
15-19	0.079	0.0826	0.086	0.0888	0.0913	0.0936	0.0962	0.0958	0.0934	0.0824
20-24	0.0803	0.0804	0.0807	0.0826	0.085	0.0886	0.0942	0.1229	0.1603	0.2032
25-29	0.0701	0.0736	0.0773	0.0801	0.0824	0.083	0.0838	0.0786	0.0678	0.0511
30-34	0.0586	0.0616	0.0645	0.0677	0.071	0.0741	0.0775	0.0788	0.0792	0.0856
35-39	0.0498	0.0526	0.0551	0.0575	0.0599	0.0623	0.0648	0.0604	0.0522	0.0422
40-44	0.0409	0.0424	0.0455	0.0484	0.051	0.0536	0.0559	0.059	0.0594	0.0549
45-49	0.0353	0.0377	0.0383	0.0403	0.042	0.0437	0.0449	-0.0002	-0.026	-0.0443
50-54	0.0315	0.0325	0.0335	0.0343	0.0356	0.0372	0.0396	0.1004	0.1172	0.141
55-59	0.0269	0.0286	0.0302	0.0315	0.0327	0.0338	0.0351	0.047	0.0768	0.0919
60-64	0.0197	0.0217	0.0236	0.0256	0.0275	0.0294	0.0314	0.0444	0.0697	0.098
65-69	0.0132	0.0153	0.0175	0.0194	0.0215	0.0238	0.0264	0.0191	0.0053	-0.0047
70-74	0.0067	0.0086	0.0109	0.0132	0.0156	0.0179	0.0204	0.0232	0.0355	0.0397
75-79	0.0039	0.0054	0.0068	0.0082	0.0099	0.012	0.0144	0.0171	0.0116	0.0041
80-84	0.0023	0.0047	0.0068	0.0086	0.0104	0.0124	0.0141	0.0132	0.0094	0.0091
85+	0.0005	0.0033	0.0066	0.0101	0.0132	0.0175	0.0207	0.0229	0.0305	0.0353

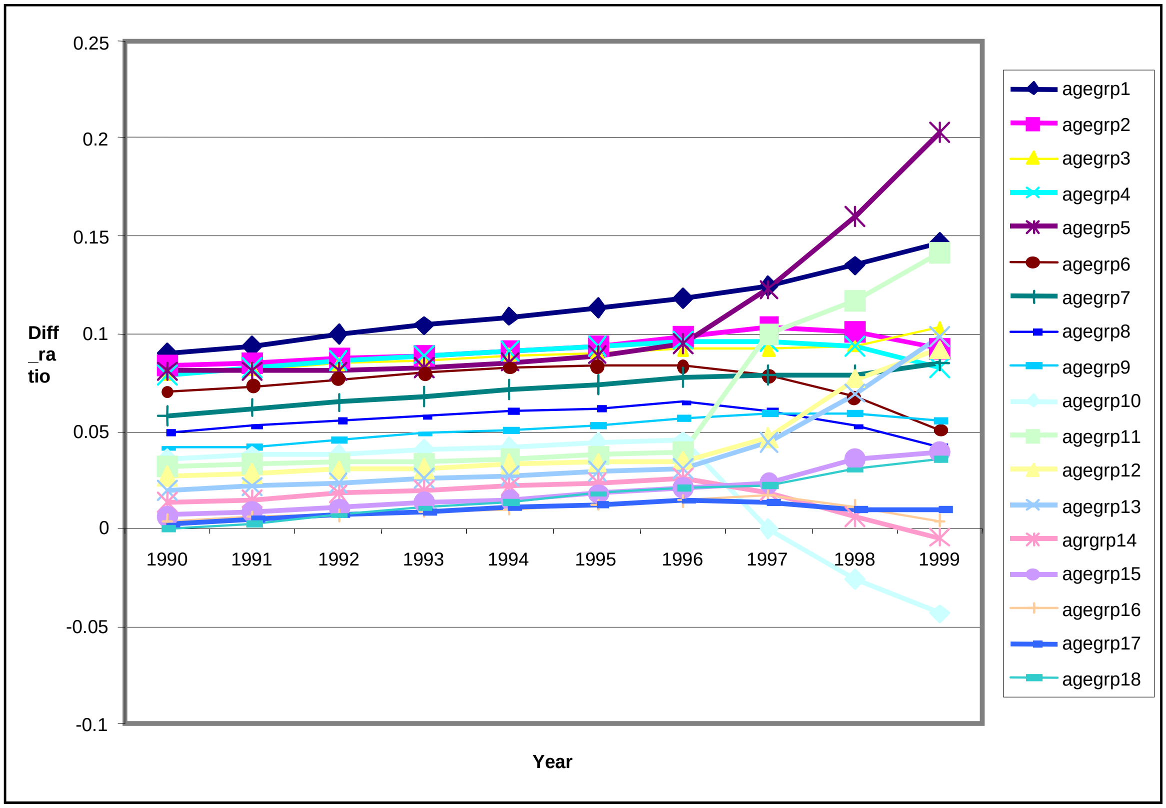


Figure 4 The difference ratio of the two population estimates at the national level by age group (1990-1999)

4.1.5. Difference by the combined two age groups for each year

Figure 5 shows the difference of the two population figures by the two combined age groups, age group1 from the originally agegrp6-10 and age group2 for agegrp1-5 and agegrp11-18. It shows that for all 10 years, age group2 has higher DRs than that of age group1. From year 1996, the difference between the DRs for the two age groups is increasing, with the DRs for age group 2 increasing and decreasing for age group 1. Age group may be a factor that affects the difference for year 1997-1999.

Figure 5 The difference ratio of the two population estimates at the national level by the combined two age groups (1990-1999)

4.2. Graphical displays of the population difference by States

To analyze the population data at the state level the difference of the two populations by year, by sex, by age group, and by sex and age group are graphically displayed to show the contribution of each factor to the difference.

4.2.1. Difference by year

From Figure 6 it can be seen that the DRs have the same pattern for every state across the 10 years (1990-1999). Most states have low DRs (below 0.1), and several states have high DRs, as high as 0.28 for NM. States with DRs higher than 0.1 include AZ,CA,CO,FL,NY,NV,NJ,IL.

Because it seems that year is not an influencing factor for the difference, the population for all these years were combined together for further analysis.

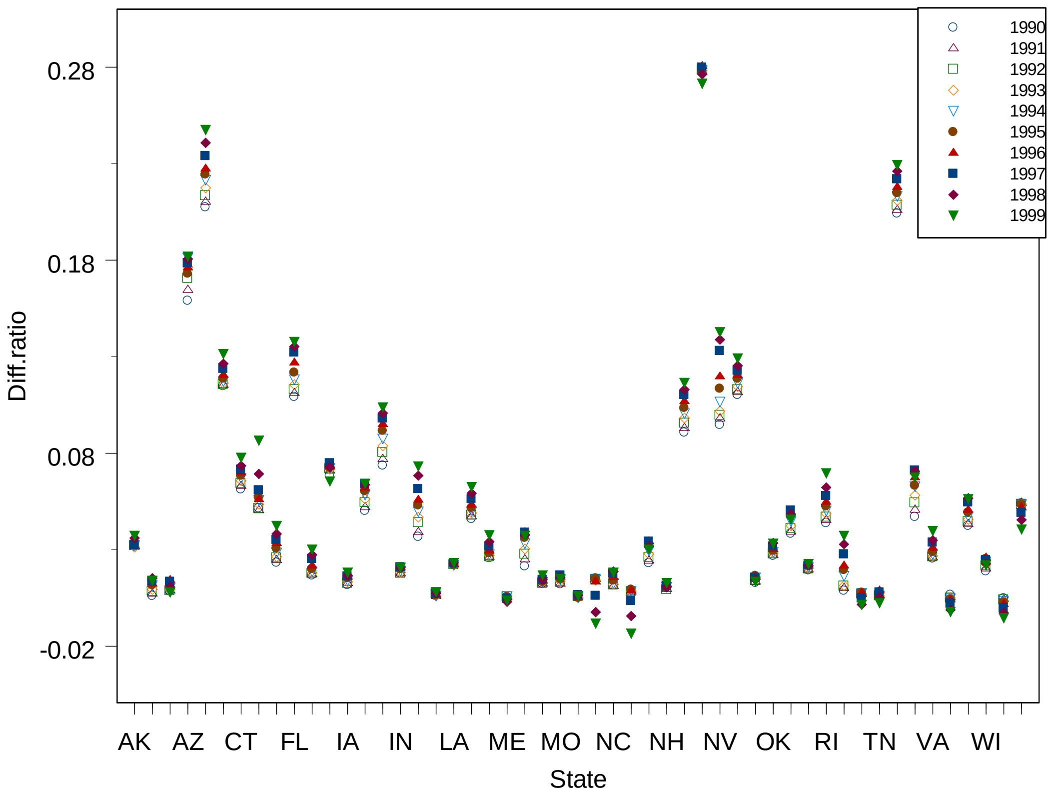


Figure 6 The difference of the two populations marginally by state for each year from 1990 to 1999

(The states are ordered alphabetically)

4.2.2. Difference of State total

Figure 7 shows that the DRs for most states are below 0.1, with several states have DR higher than 0.1, as high as 0.278 for NM, 0.225 for CA, 0.216 for TX, 0.175 for AZ, 0.123 for FL, 0.120 for CO, 0.118 for NY, 0.117 for NV and 0.103 for NJ.

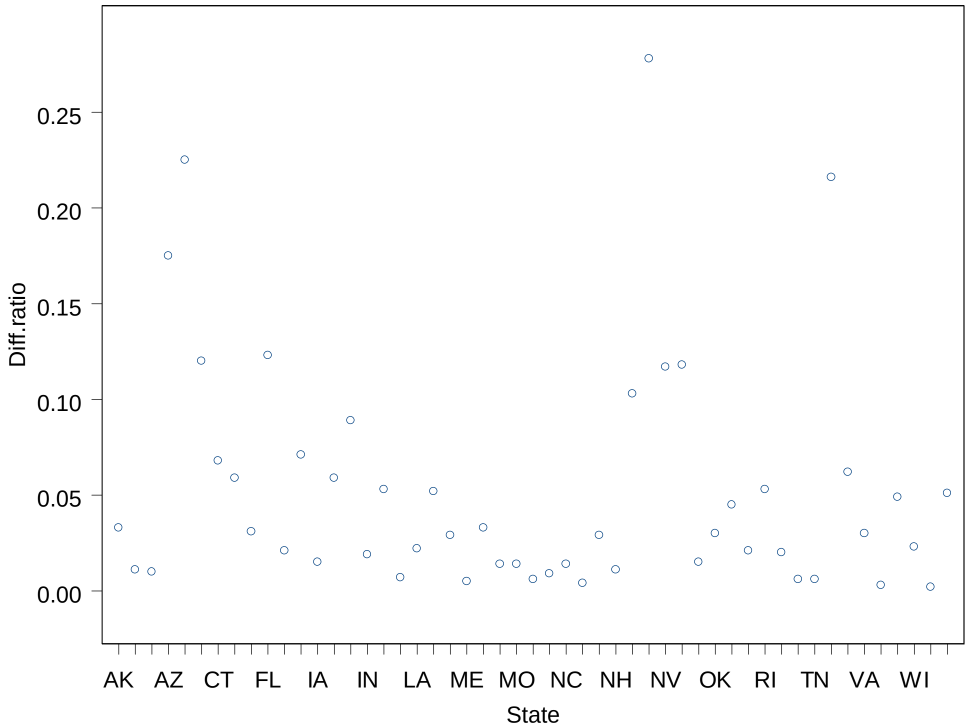


Figure 7 The difference of the two populations marginally by state for all 10 years combined together (1990-1999)

(The states are ordered alphabetically)

4.2.3. Difference by sex

The difference of the two populations at the state level for male and female is shown in figure 8. For the male subgroup, the DRs for all states are around 0, while for females, the DRs vary significantly from state to state. For the female group, some states have very high DRs, as high as 0.752 for NM, and about half of the states have DRs higher than 0.3. From this figure it appears that the differences of the two populations at the state level come from the female group. There is no apparent difference for the male group between the two populations.

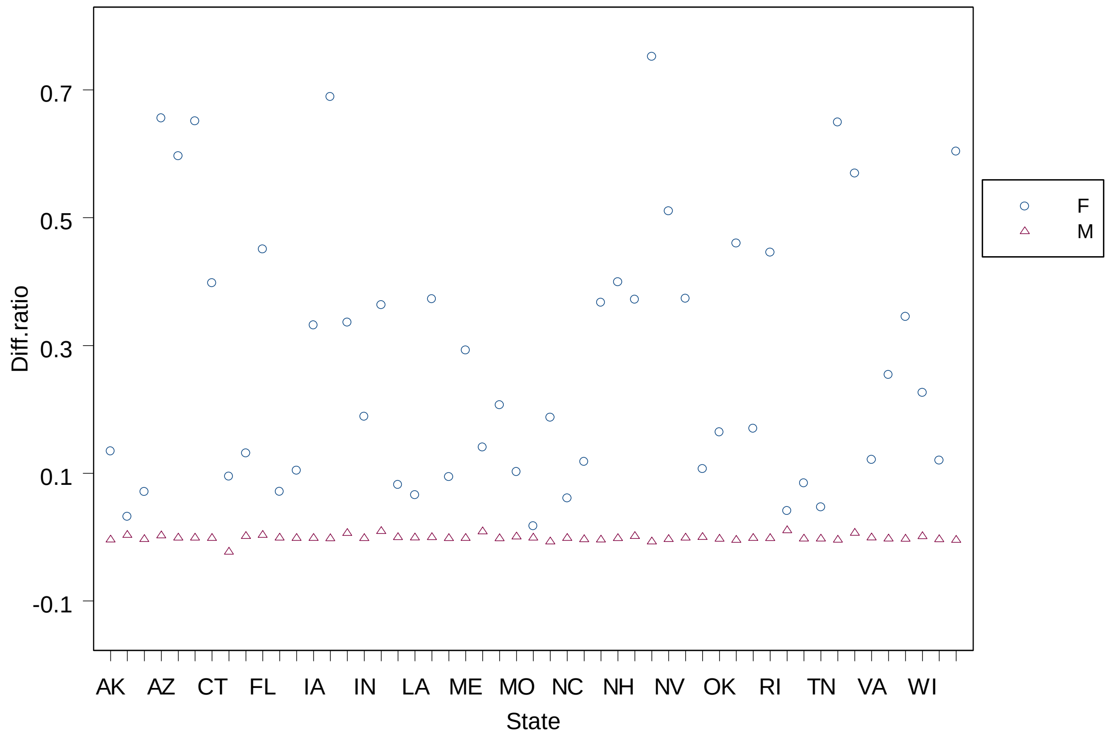


Figure 8 The difference of the two populations marginally by state for each sex for all 10 years combined together (1990-1999)

(The states are ordered alphabetically)

4.2.4. Difference by age group

Figure 8 shows the difference between the two populations at the state level for two age groups (agegroup1 is from original agegrp6 to agegrp10 and agegroup2 from the original agegrp1-5 and agegrp11-18).

From the figure we can see that for all states, the DRs for age group2 are consistently higher than that of agegroup1 for all states. Because all states have the same trend for the two age groups, age group is not considered as a factor that contributes to the differences among states.

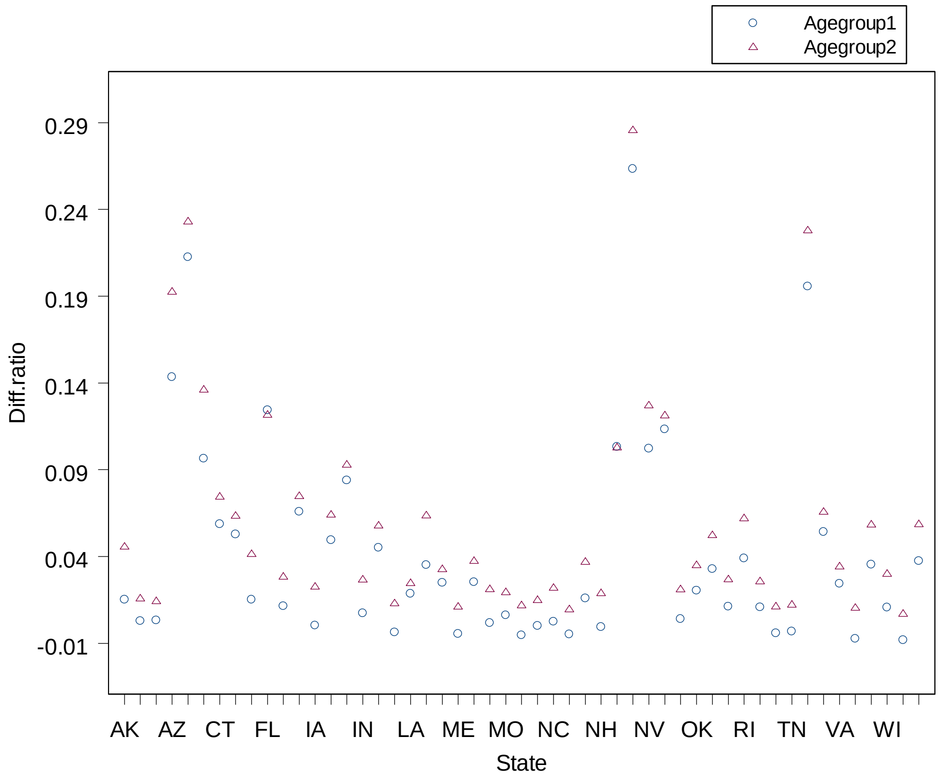


Figure 9 The difference of the two populations marginally by state for the two combined age groups for all 10 years combined(1990-1999)

(The states are ordered alphabetically; agegroup1 is from original agegrp6 to agegrp10, and agegroup2 from the originally agegrp1-5 and agegrp11-18)

4.2.5. Difference by sex and age group

Figure 10 shows the difference between two populations by age group and sex. The figure shows that for the male group, the differences for both age groups are small (around 0), while for the female group, agegroup2 has a higher DR than agegroup1. It appears that the difference between the age groups from figure 8 comes from the female group. There is no interaction between the age group and the sex for the DRs of the two populations.

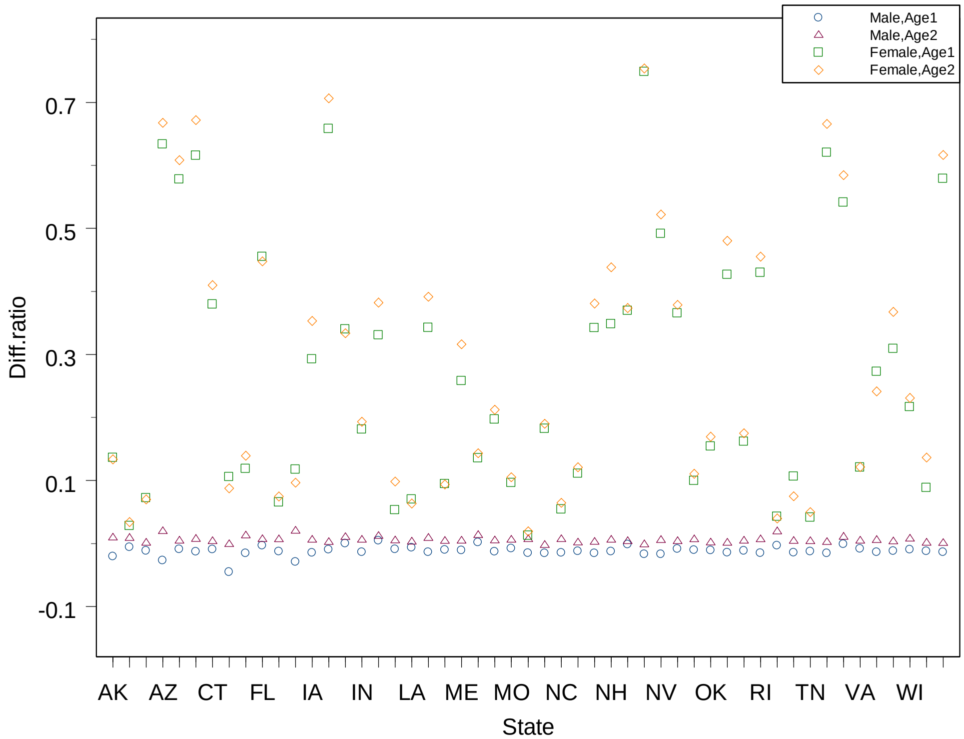


Figure 10 The difference of the two populations marginally by state for the two combined age groups and two sex groups for all 10 years (1990-1999)

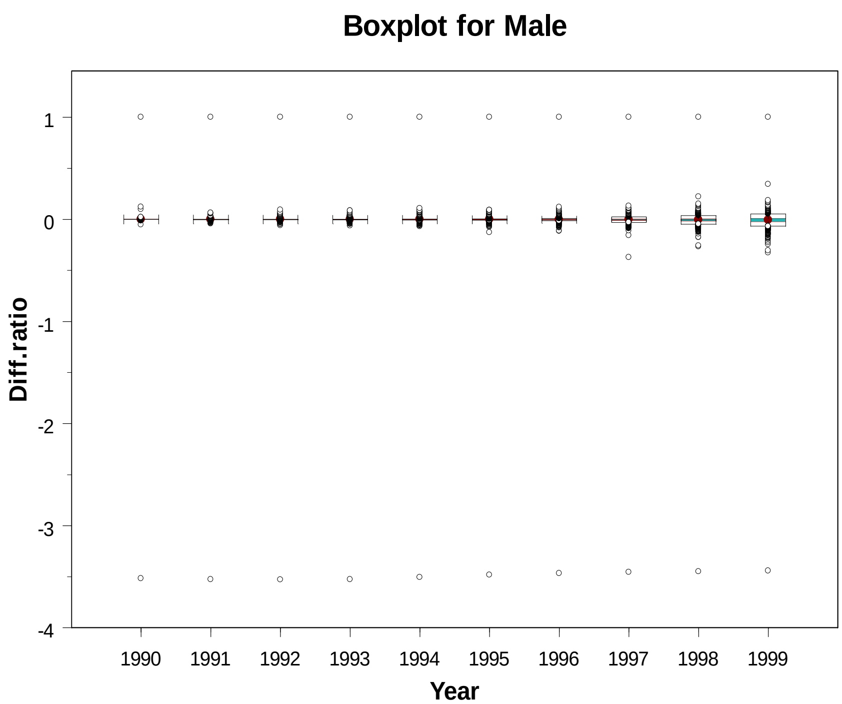
(The states are ordered alphabetically; agegroup1 is from original agegrp6 to agegrp10, and agegroup2 from the originally agegrp1-5 and agegrp11-18)

4.3. Graphical displays of the population difference by counties

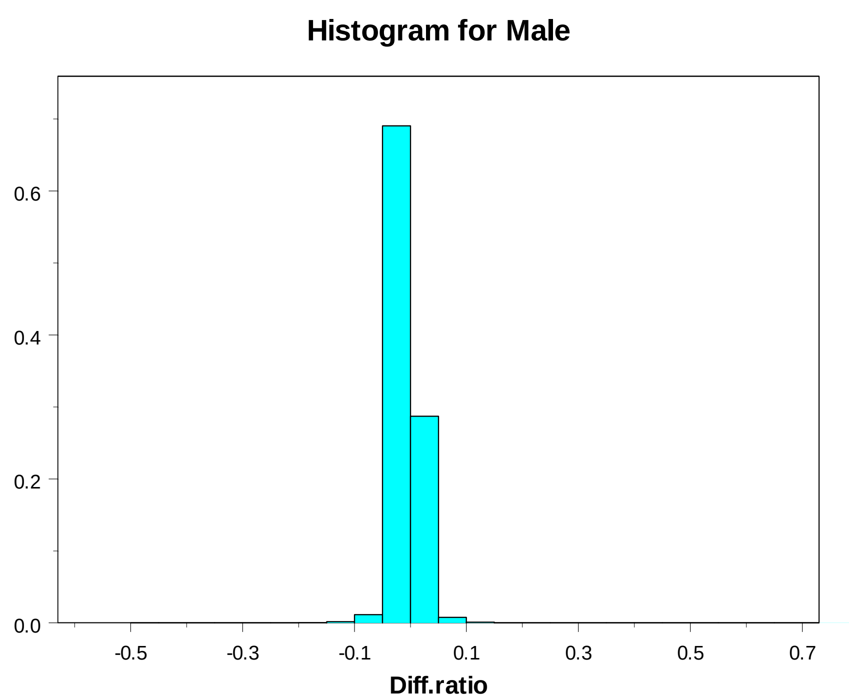
Histograms of the DRs for the 10 years combined show the distribution of the DRs by sex, race and age group. Boxplots of the DRs by year marginally for each sex, race and age group were displayed to show the median, percentiles and outliers of the differences for counties for each year.

4.3.1. Difference by Sex

Figure 11 (b) and 12(b) show that for sex=1(male), the DR has the highest frequencies around 0, while for female, the DR varies more (from -0.1 to 1.0). Therefore for most counties, the differences between the two male subgroups are small (close to 0) and the differences for female populations vary among counties. The boxplots below (figure 11(a) and figure 12(a)) also give the same results. For each year, the male population estimates are close, with DRs close to 0 for most counties, the DRs for the female population estimates vary greatly among counties for all 10 years.

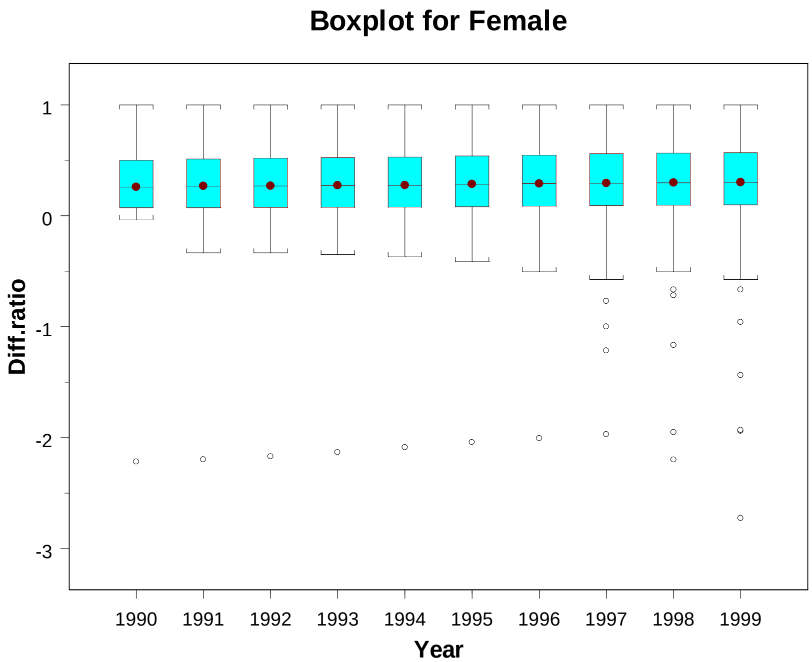


11(a) Boxplot of the difference ratios for each year

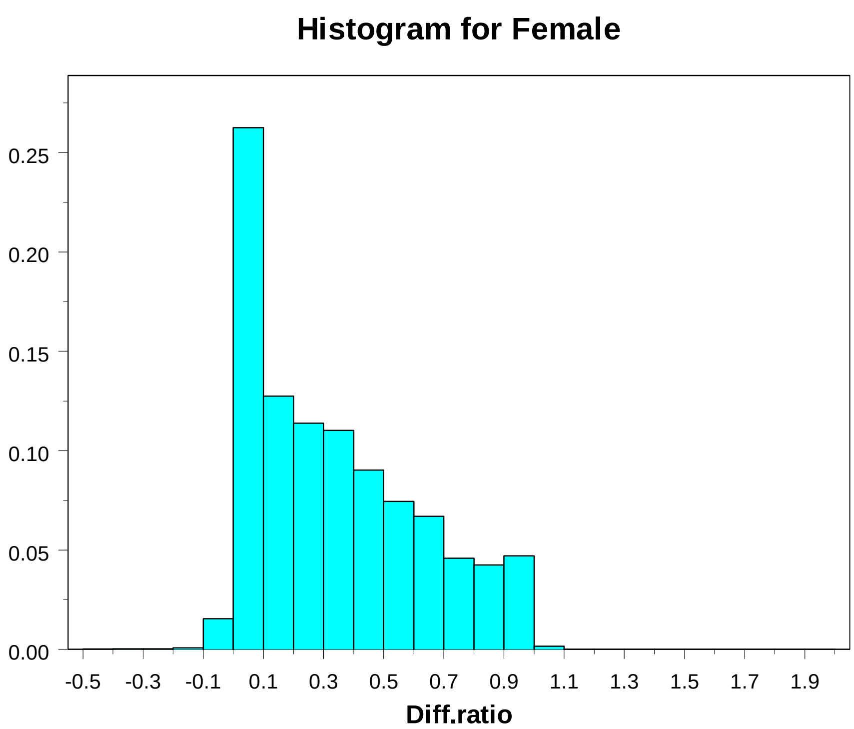


11(b) Histogram of the difference ratios for all years

Figure 11 The difference of the two population estimates at the county level for Male (1990-1999)



12(a) Boxplot of the difference ratios for each year



12(b) Histogram of the difference ratios for all years

Figure 12 The difference of the two population estimates at the county level for Female (1990-1999)

4.3.2. Difference by Race

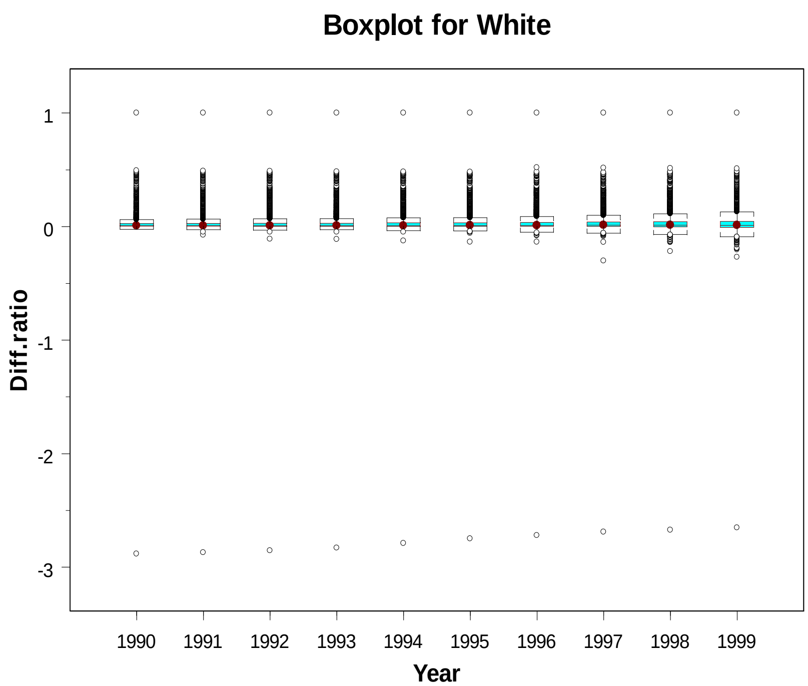
The histograms (Figure 13(b) and 14(b)) show that the distributions of the DR for White and Non_white are similar, clustered around 0, which means that for most counties, the differences of the two populations for both races are small. The boxplots (Figure 13(a) and 14(a)) below give the same results, with the same pattern for all 10 years.

4.3.3. Difference by original 18 age groups

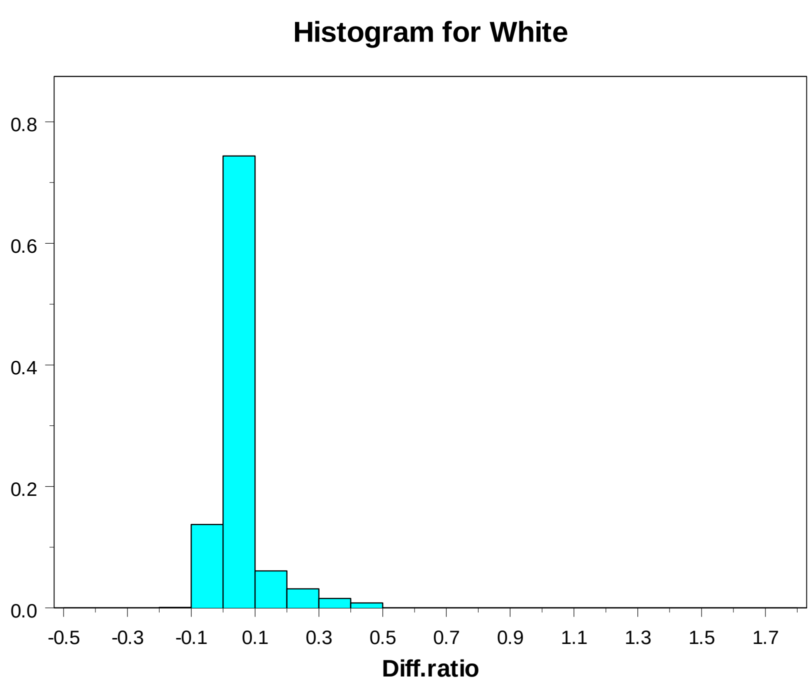
The following histograms and boxplots (Figure 15 to Figure 32) indicate that age groups do not appear to be an important factor for the difference of the two populations at the county level.

4.3.4. Difference by 2 combined age groups

Figure 33 and Figure 34 show the DRs of the two populations for the two combined age groups (Figure 32 for agegroup 1 and Figure 33 for agegroup2). The boxplots show that for both age groups, there is the same pattern across the 10 years (1990-1999). The histograms (32(b) and 33(b)) show a slight difference between two age groups. Although the DRs for both age groups are clustered around zero, agegroup1 has more negative DRs than agegroup 2.

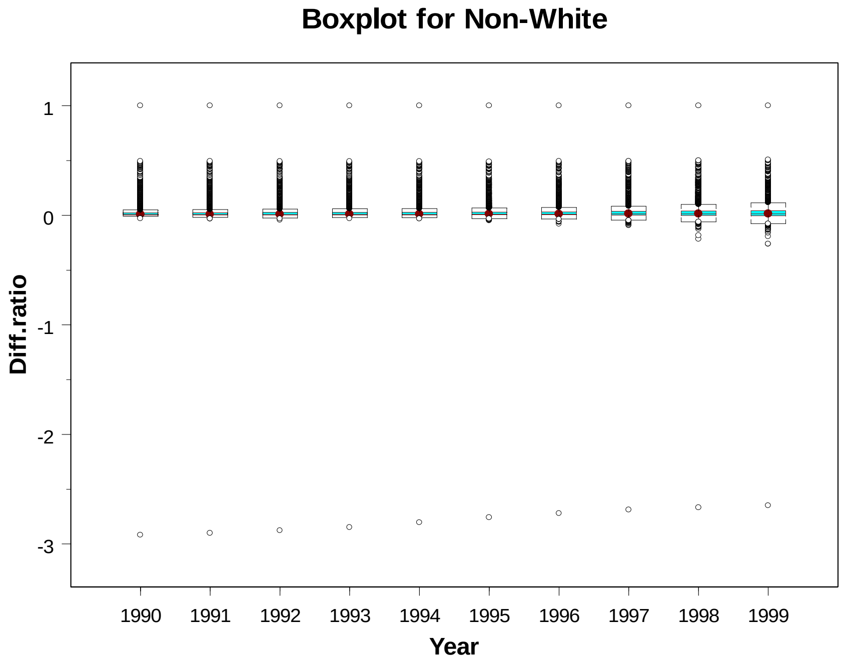


13(a) Boxplot of the difference ratios for each year

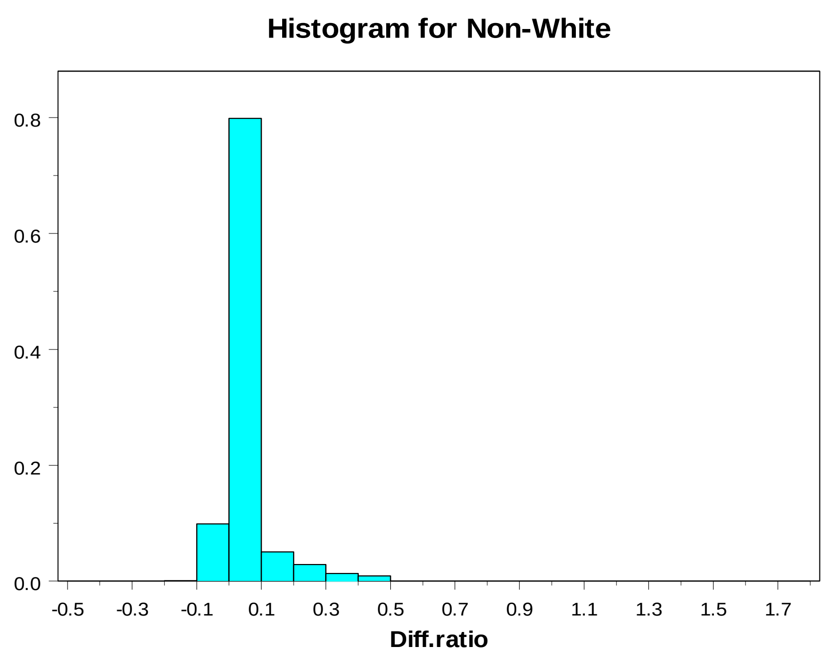


13(b) Histogram of the difference ratios for all years

Figure 13 The difference of the two population estimates at the county level for White (1990-1999).

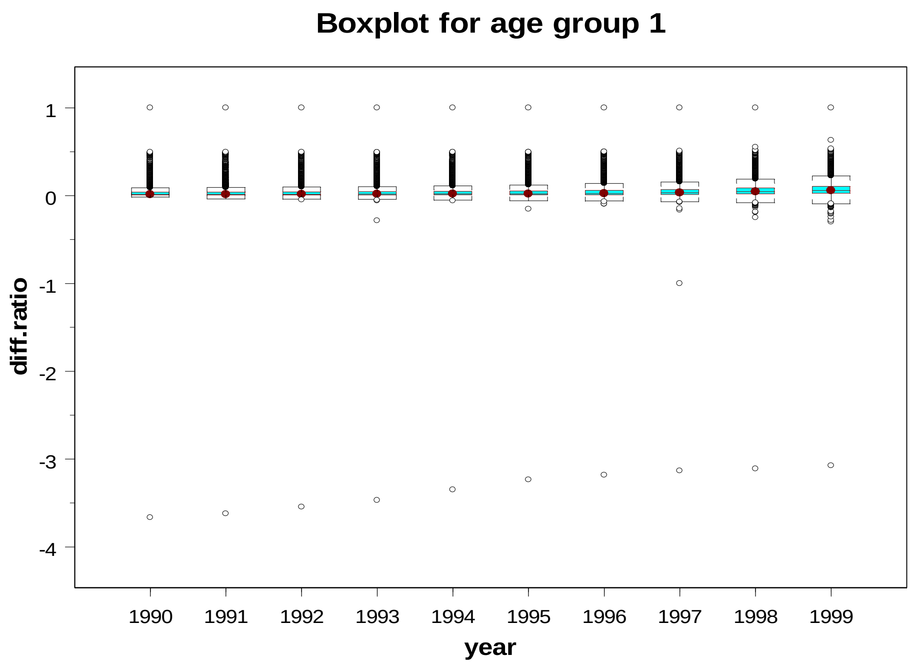


14 (a) Boxplot of the difference ratios for each year

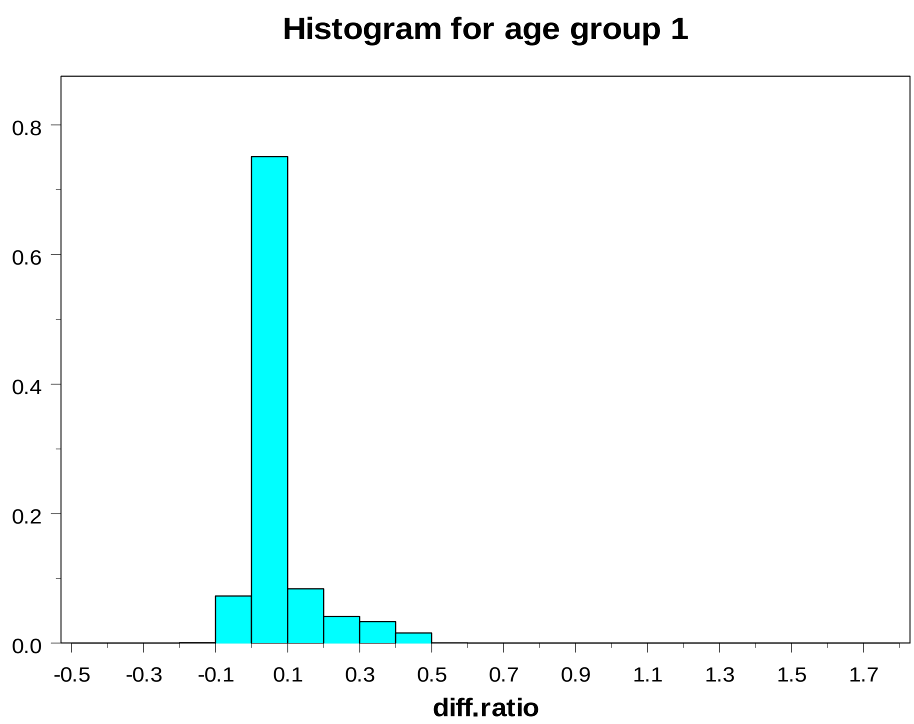


14(b) Histogram of the difference ratios for all years

Figure 14 The difference of the two population estimates at the county level for Non-White (1990-1999).

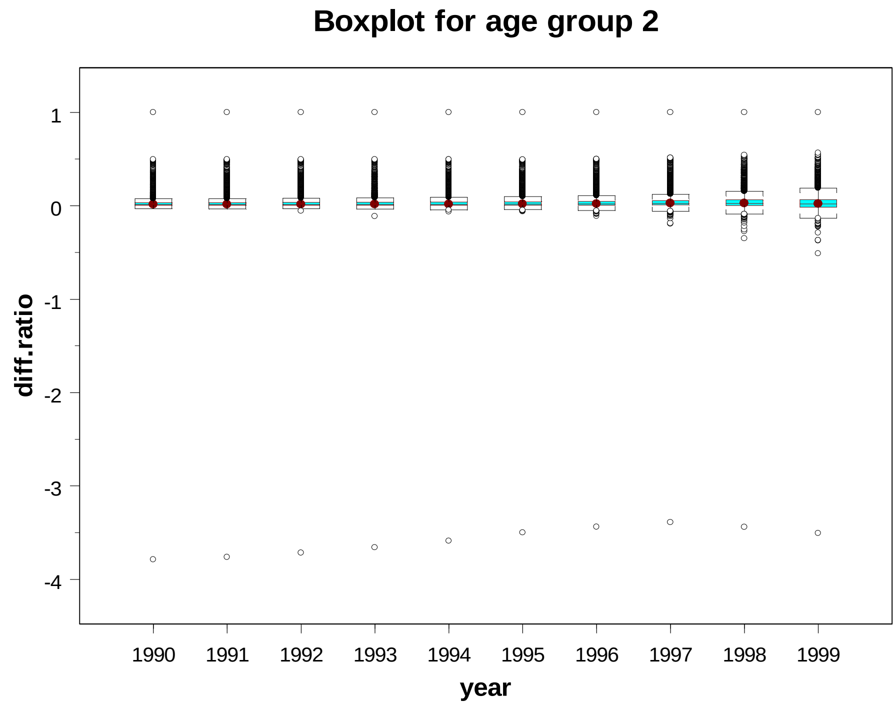


15(a) Boxplot of the difference ratios for each year

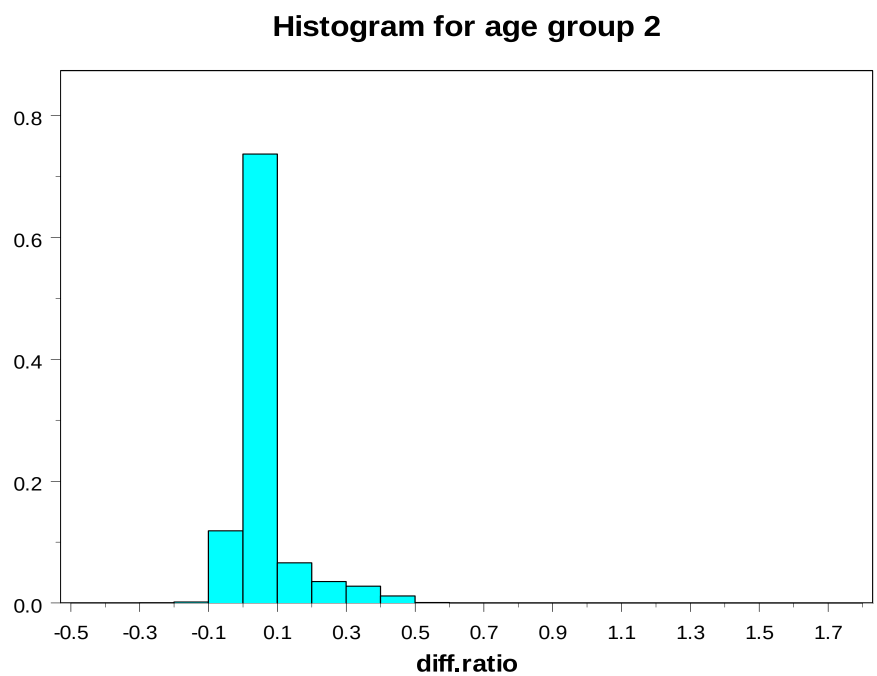


15(b) Histogram of the difference ratios for all years

Figure 15 The difference of the two population estimates at the county level for age group 1 (1990-1999)

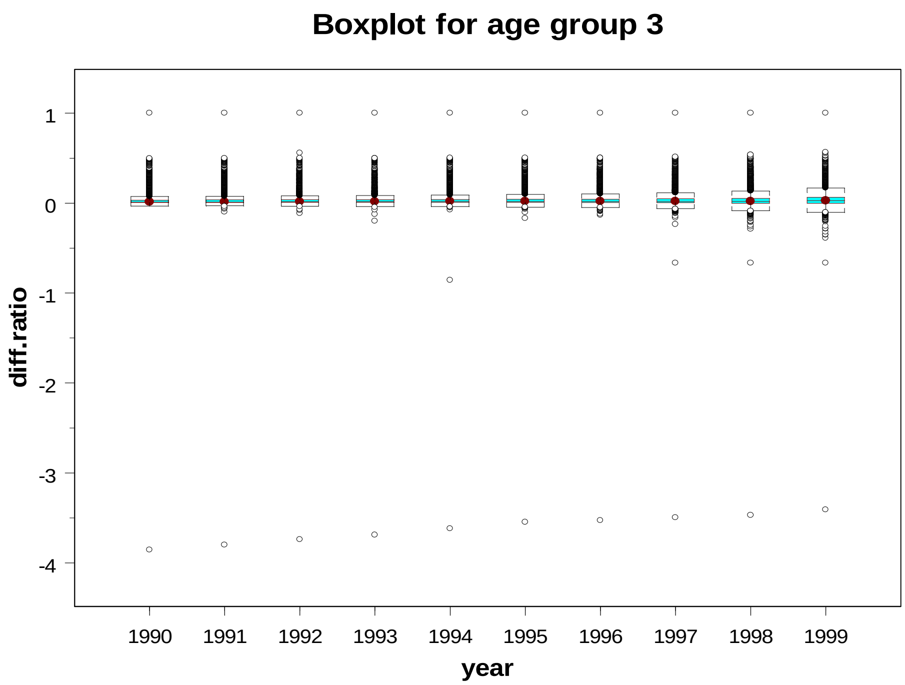


16(a) Boxplot of the difference ratios for each year

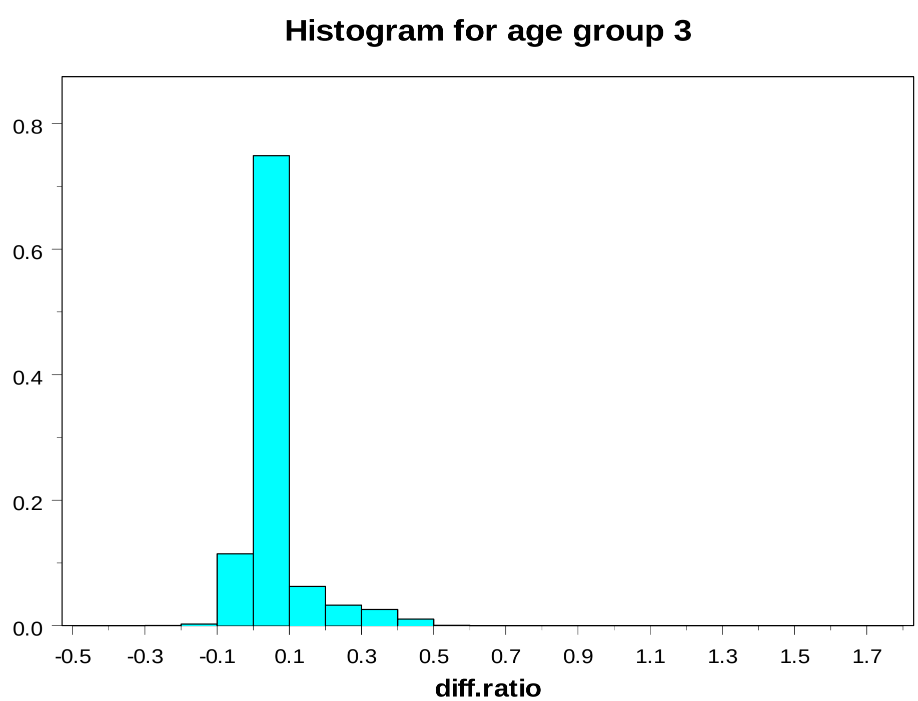


16(b) Histogram of the difference ratios for all years

Figure 16 The difference of the two population estimates at the county level for age group 2 (1990-1999)

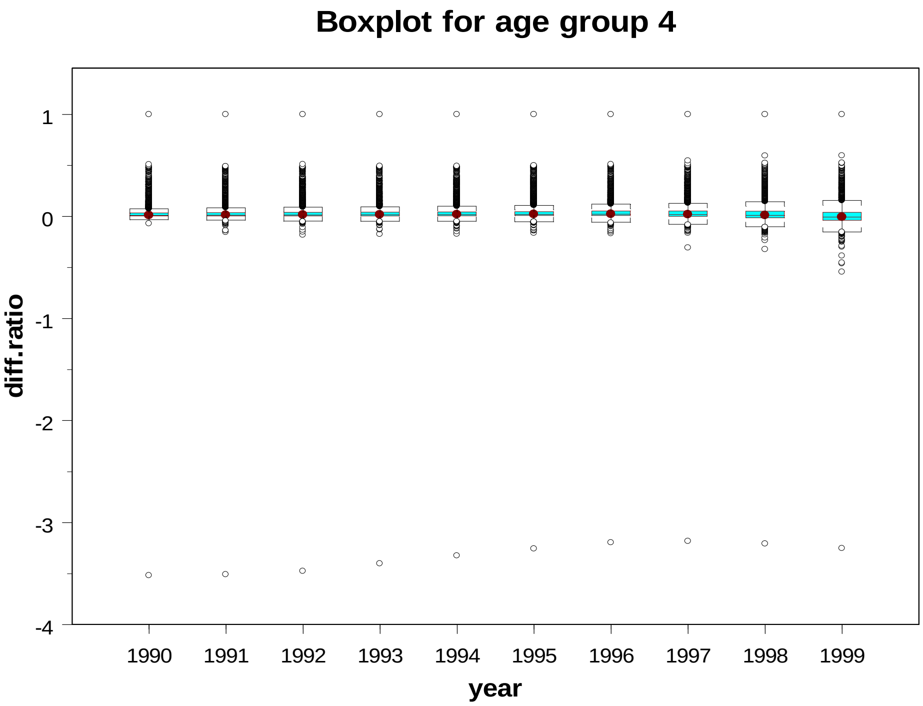


17(a) Boxplot of the difference ratios for each year

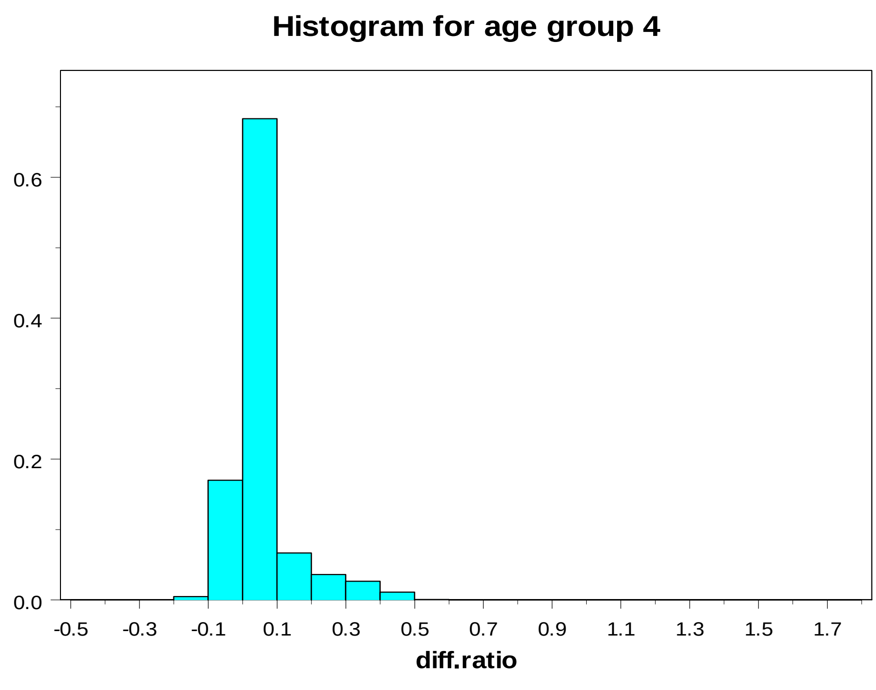


17(b) Histogram of the difference ratios for all years

Figure 17 The difference of the two population estimates at the county level for age group 3 (1990-1999)

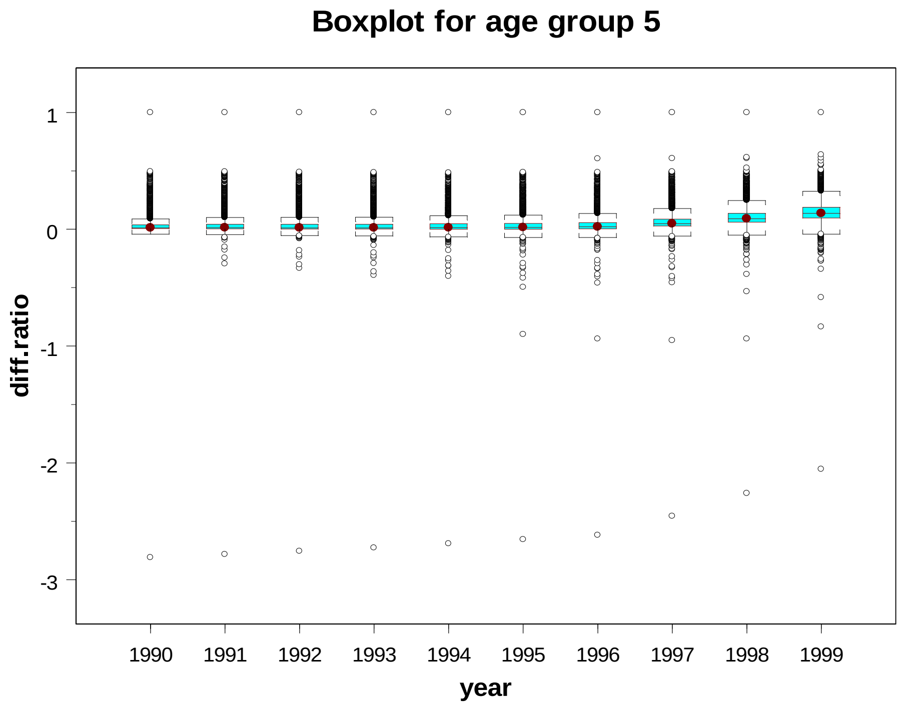


18(a) Boxplot of the difference ratios for each year

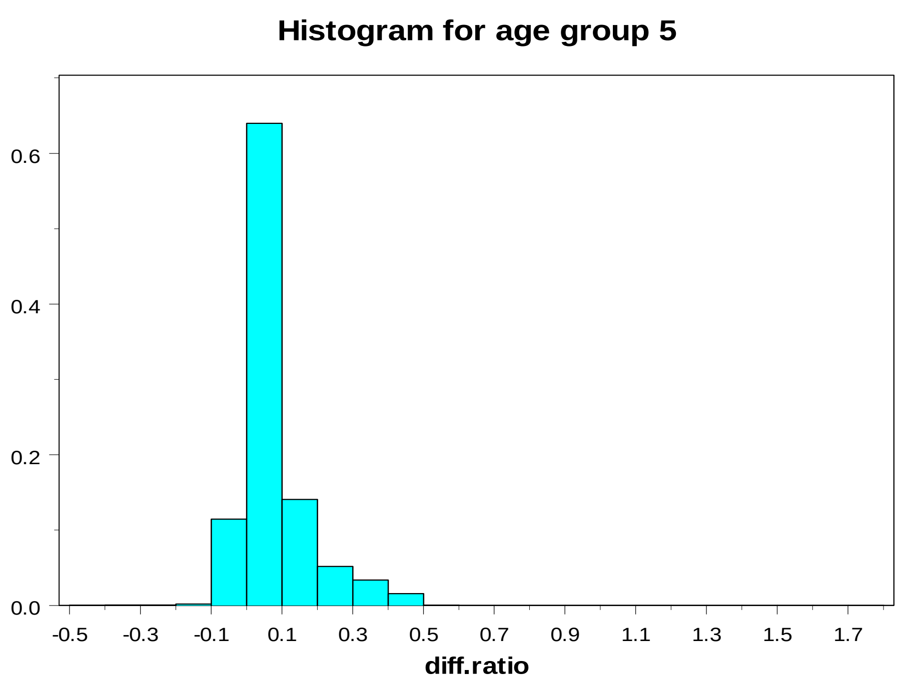


18(b) Histogram of the difference ratios for all years

Figure 18 The difference of the two population estimates at the county level for age group 4 (1990-1999)

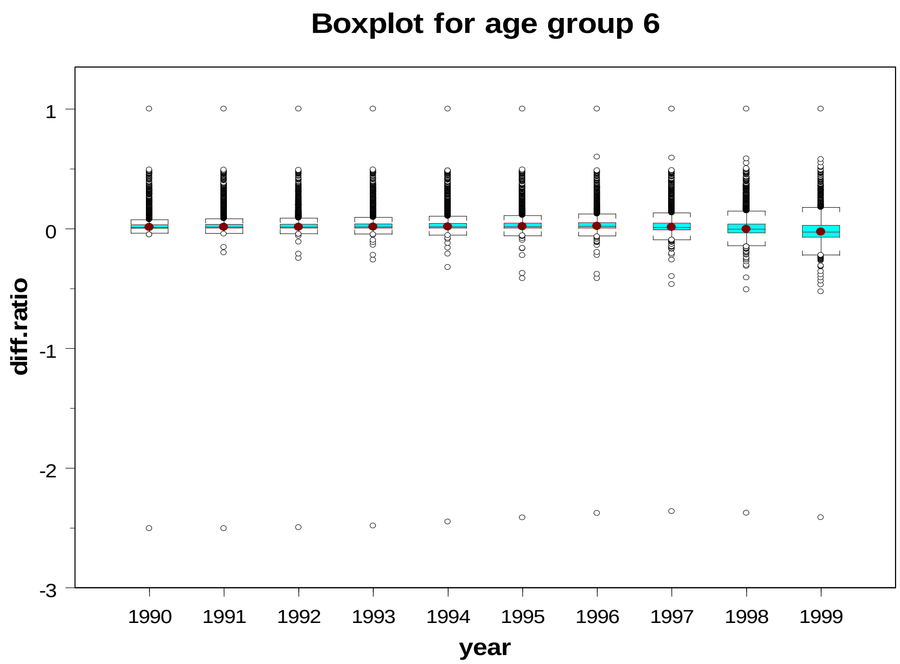


19(a) Boxplot of the difference ratios for each year

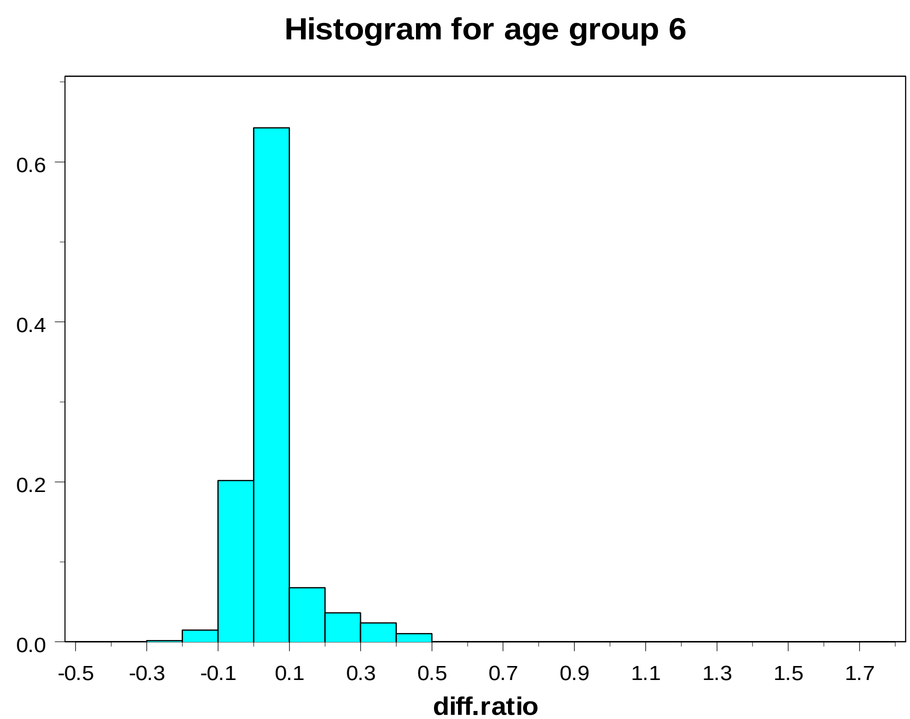


19(b) Histogram of the difference ratios for all years

Figure 19 The difference of the two population estimates at the county level for age group 5 (1990-1999)

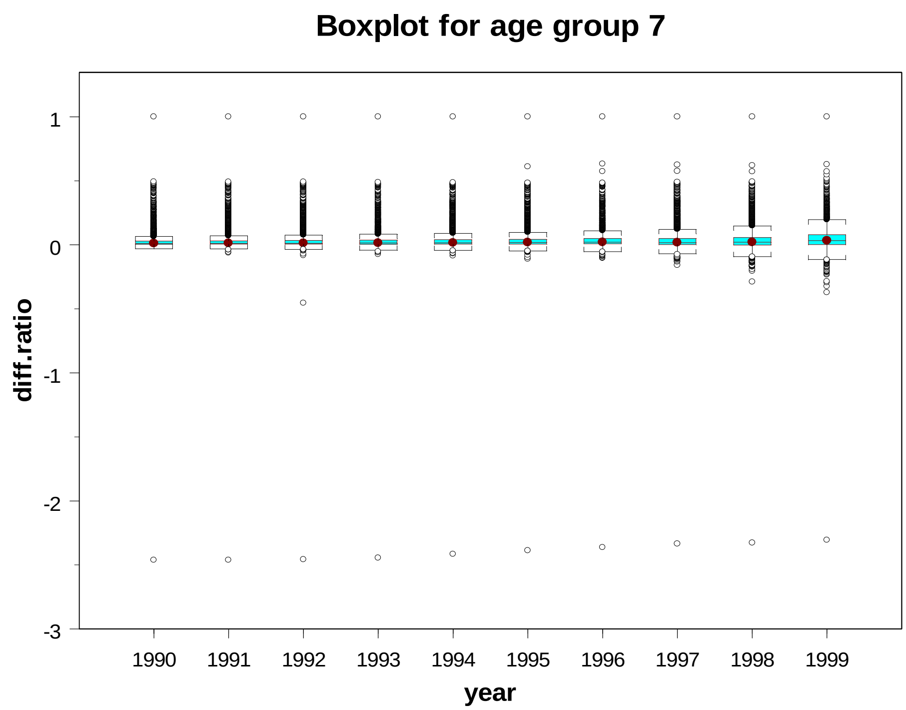


20(a) Boxplot of the difference ratios for each year

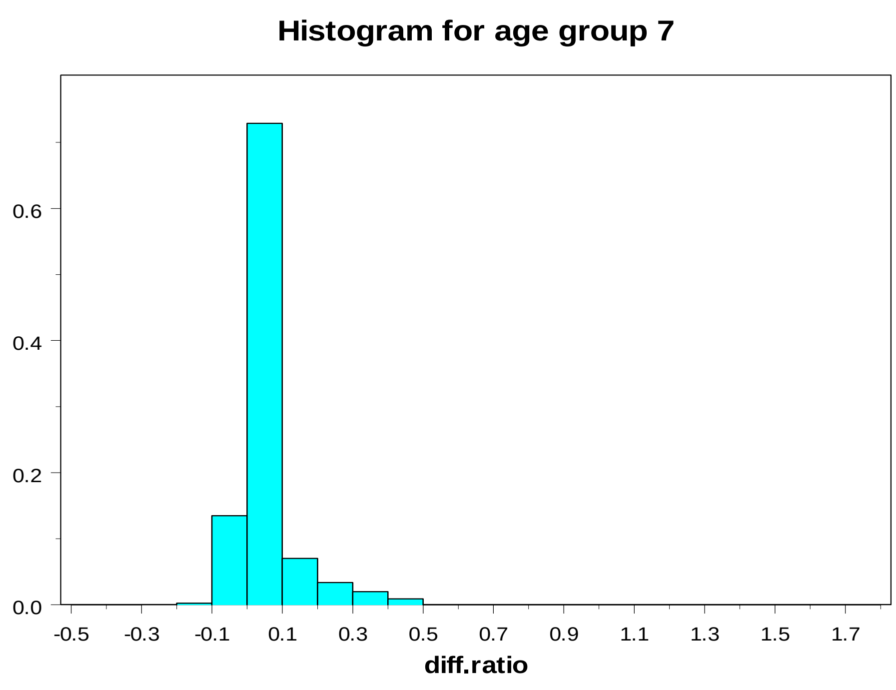


20(b) Histogram of the difference ratios for all years

Figure 20 The difference of the two population estimates at the county level for age group 6 (1990-1999)

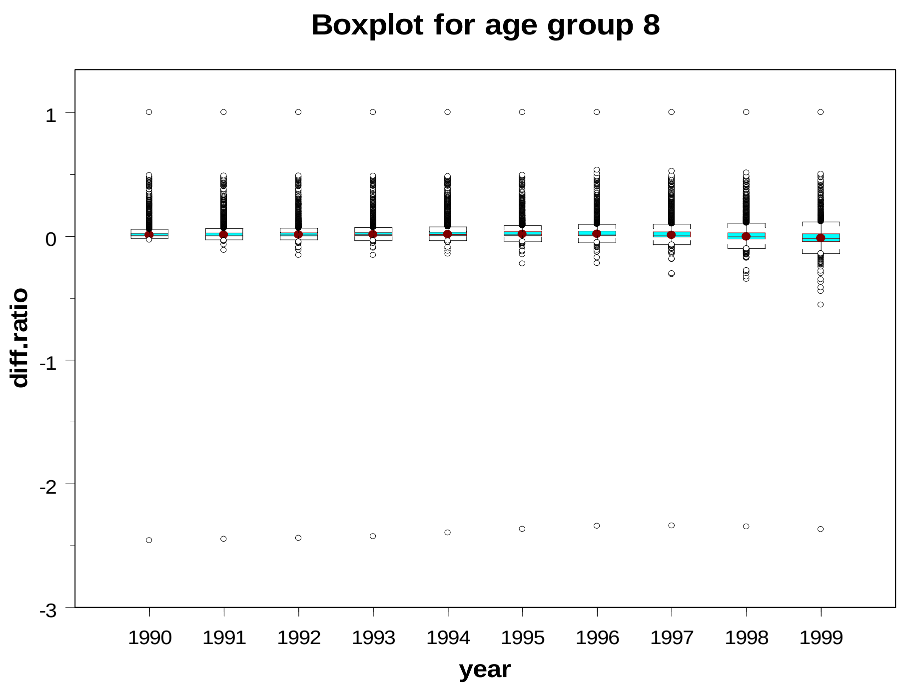


21(a) Boxplot of the difference ratios for each year

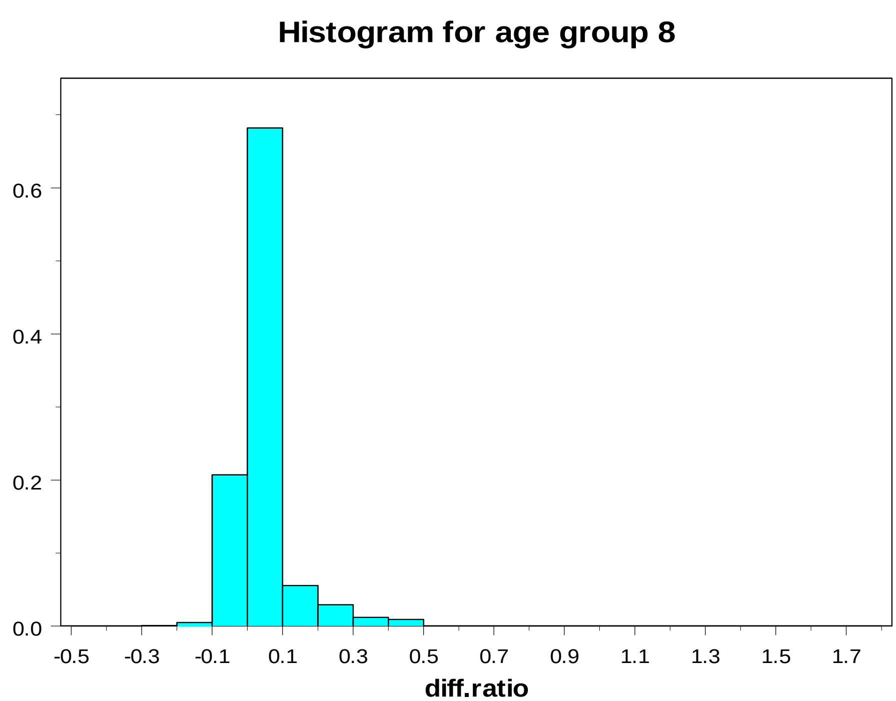


21(b) Histogram of the difference ratios for all years

Figure 21 The difference of the two population estimates at the county level for age group 7 (1990-1999)

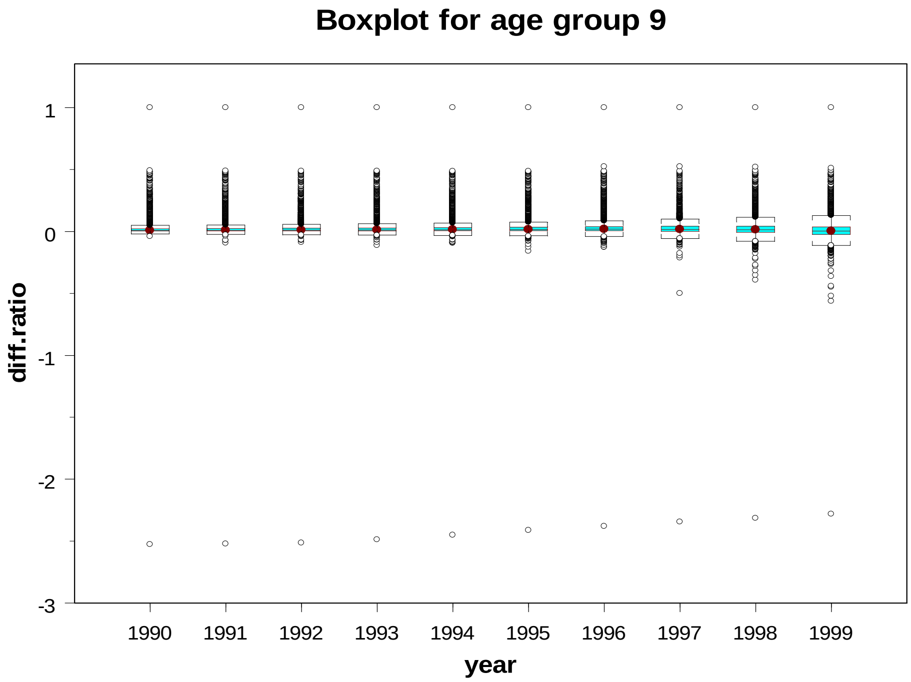


22(a) Boxplot of the difference ratios for each year

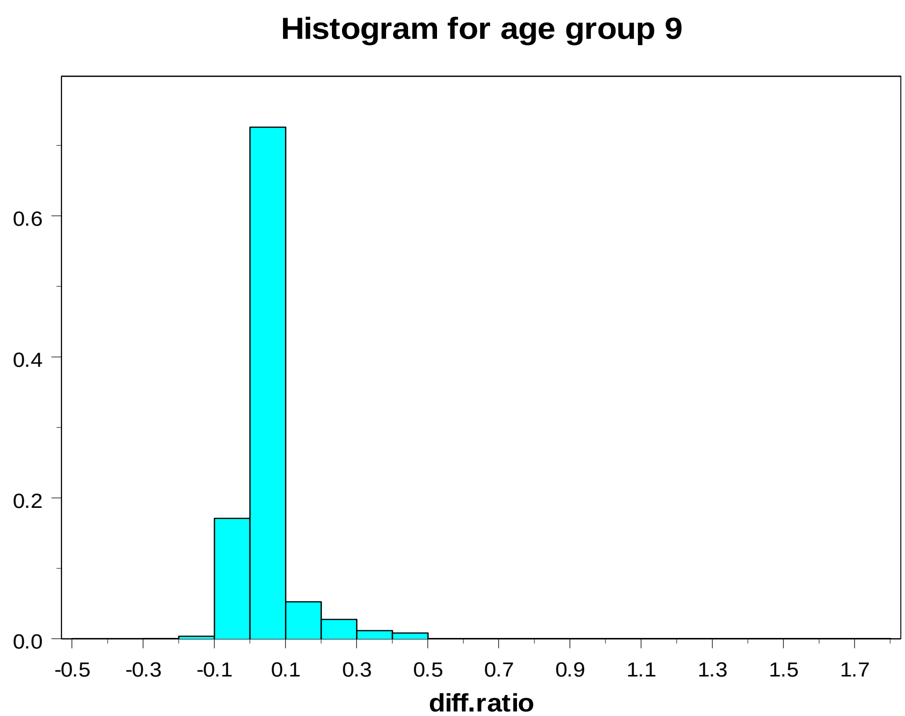


22(b) Histogram of the difference ratios for all years

Figure 22 The difference of the two population estimates at the county level for age group 8 (1990-1999)

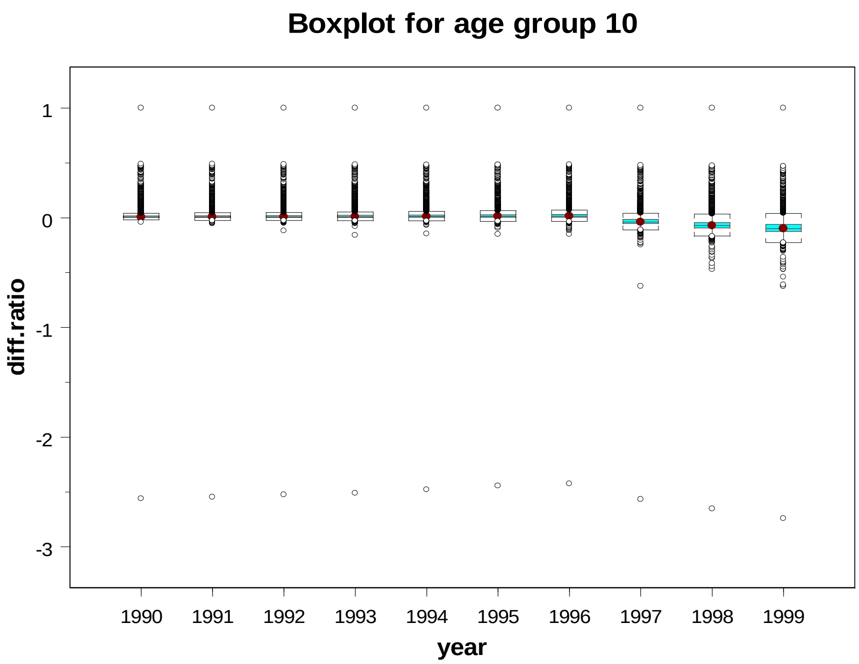


23(a) Boxplot of the difference ratios for each year

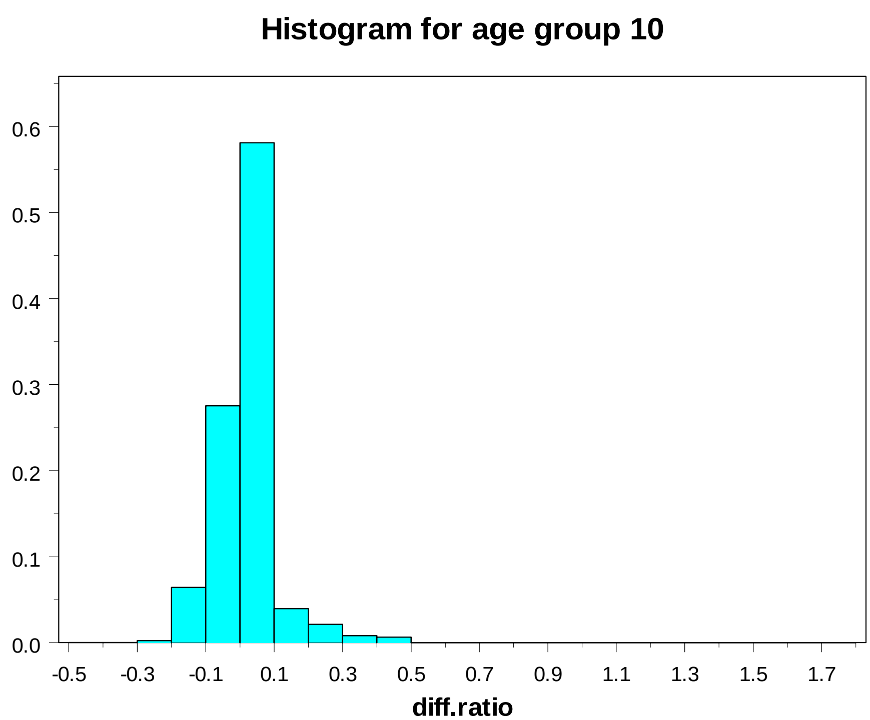


23(b) Histogram of the difference ratios for all years

Figure 23 The difference of the two population estimates at the county level for age group 9 (1990-1999)

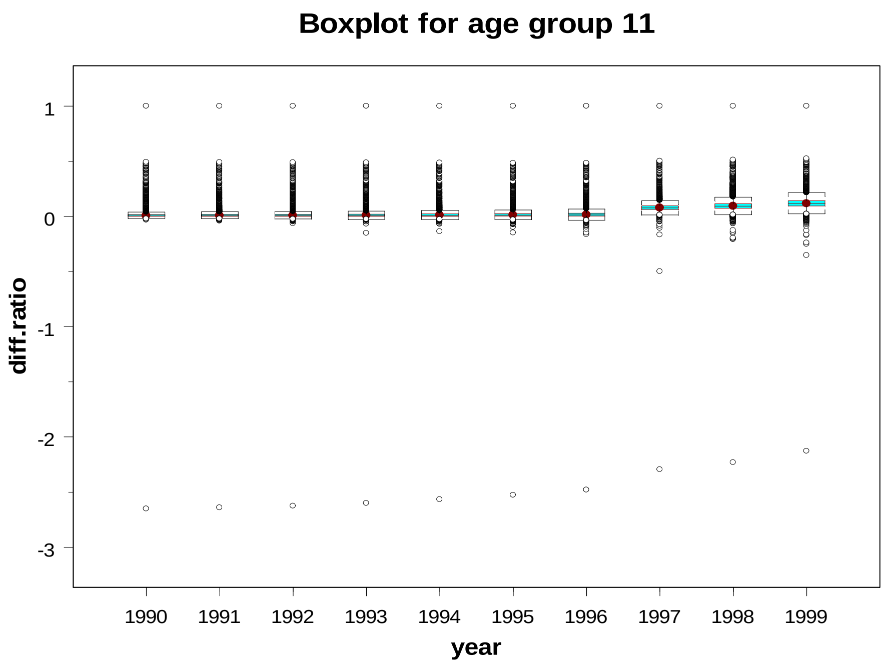


24(a) Boxplot of the difference ratios for each year

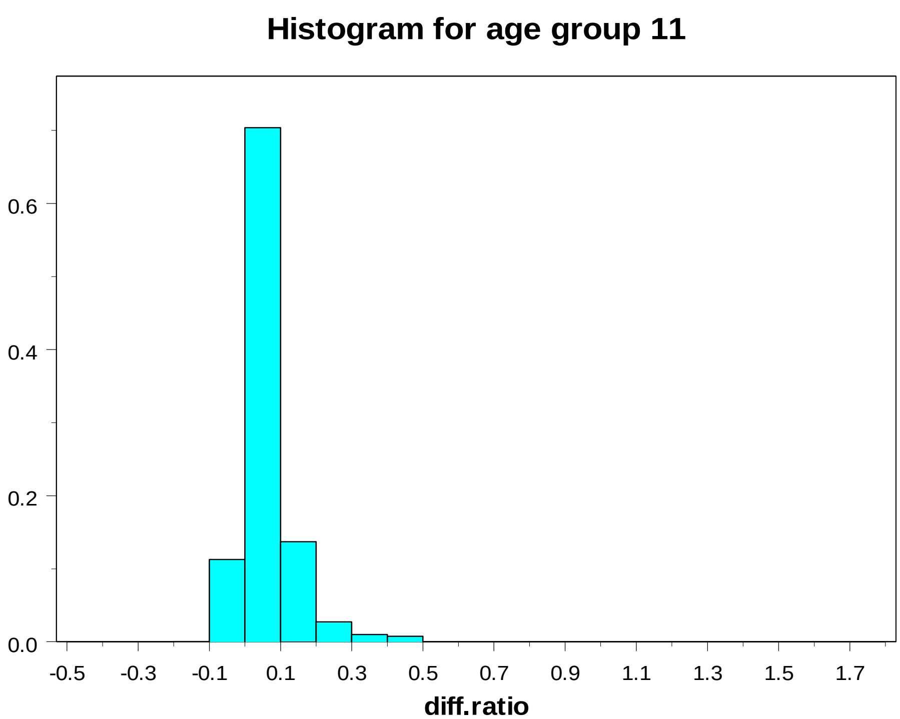


24(b) Histogram of the difference ratios for all years

Figure 24 The difference of the two population estimates at the county level for age group 10 (1990-1999)

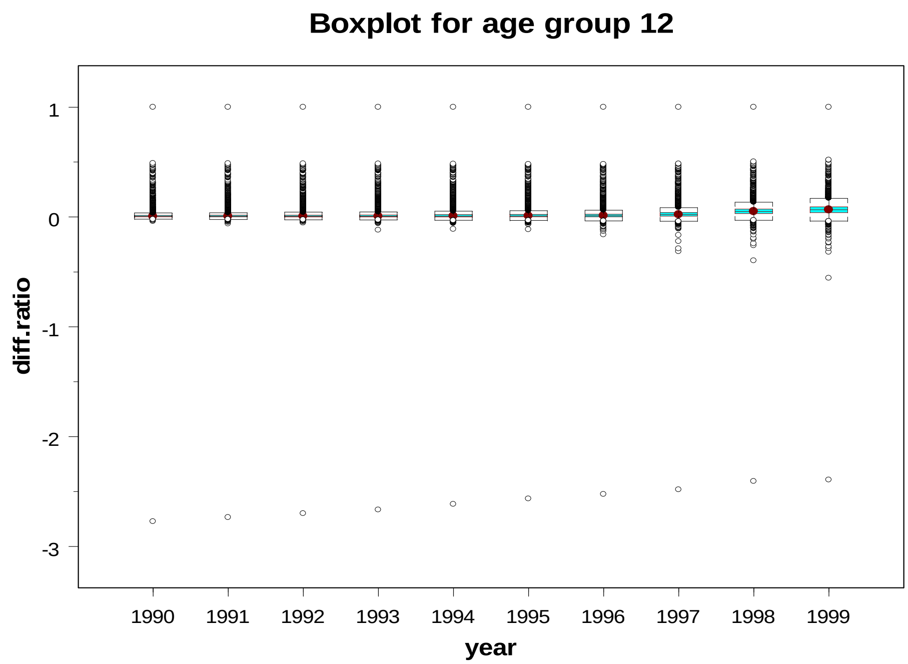


25(a) Boxplot of the difference ratios for each year

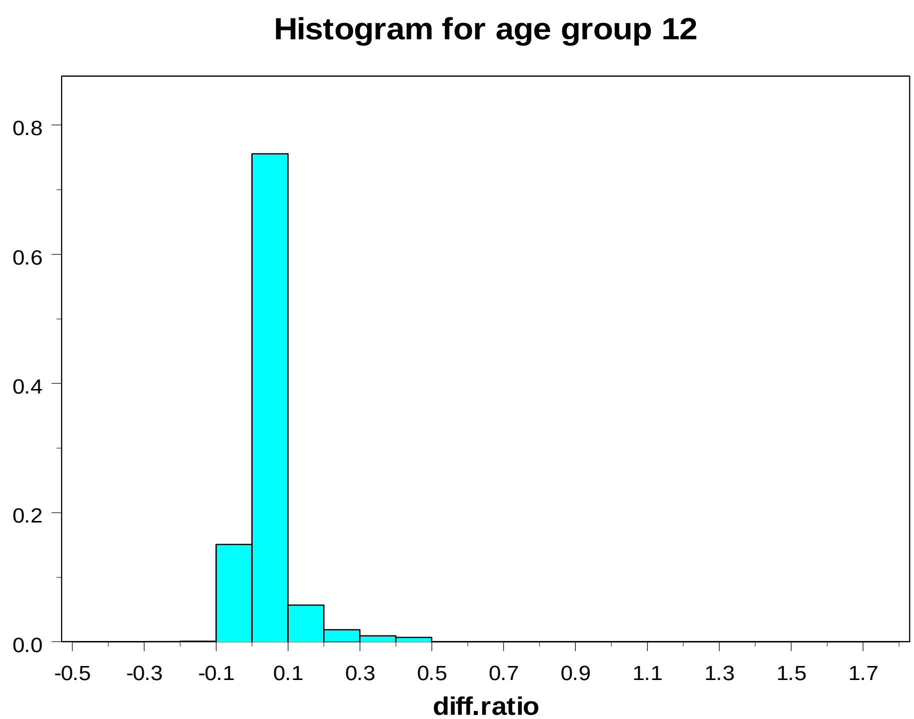


25(b) Histogram of the difference ratios for all years

Figure 25 The difference of the two population estimates at the county level for age group 11 (1990-1999)

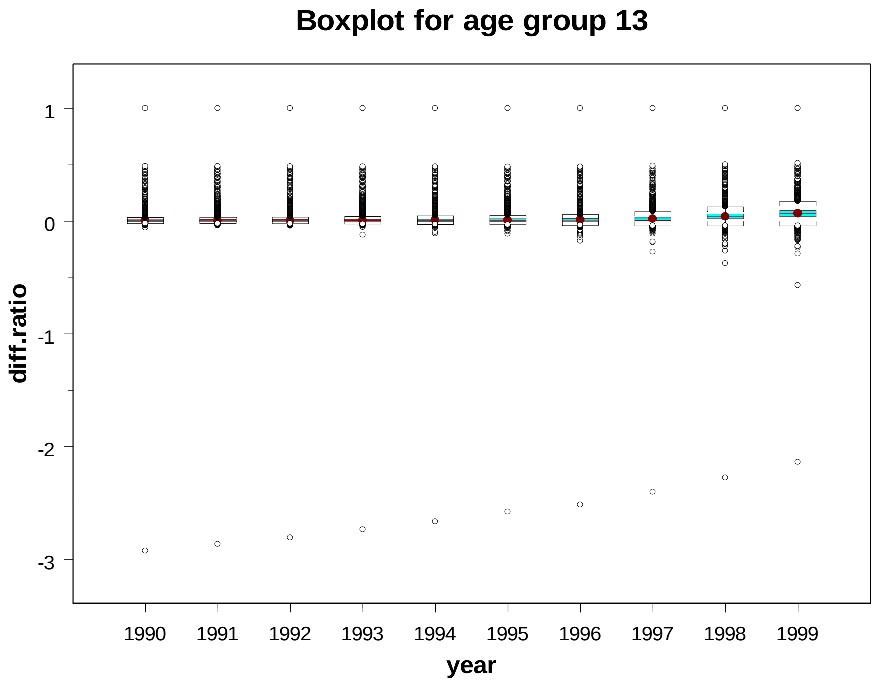


26(a) Boxplot of the difference ratios for each year

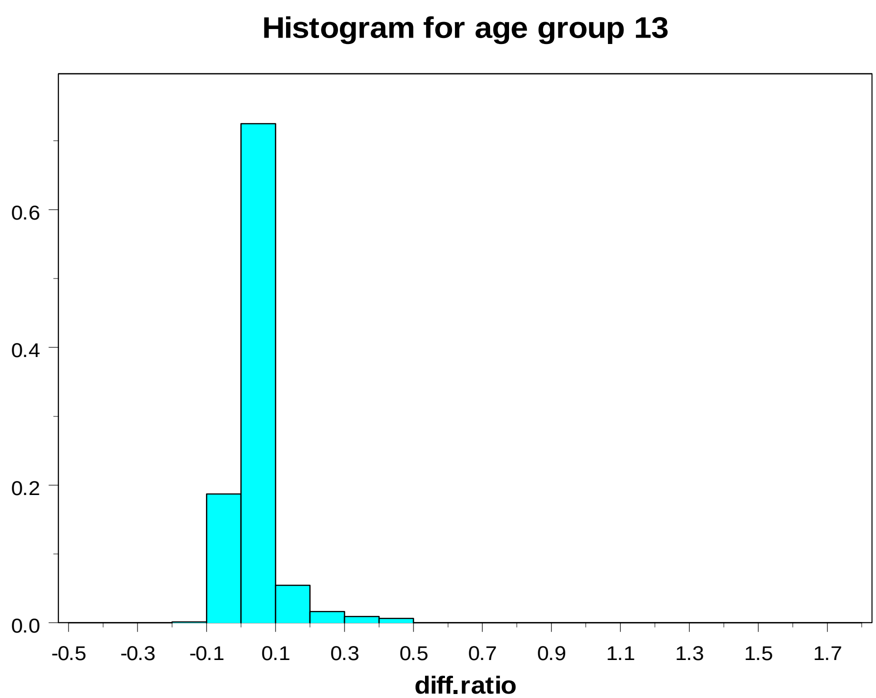


26(b) Histogram of the difference ratios for all years

Figure 26 The difference of the two population estimates at the county level for age group 12(1990-1999)

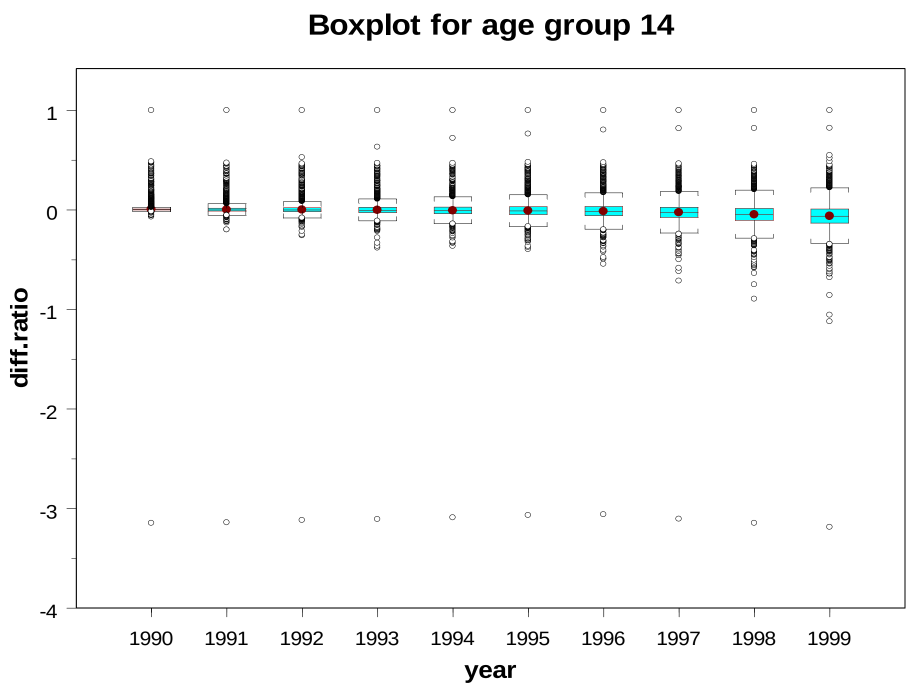


27(a) Boxplot of the difference ratios for each year

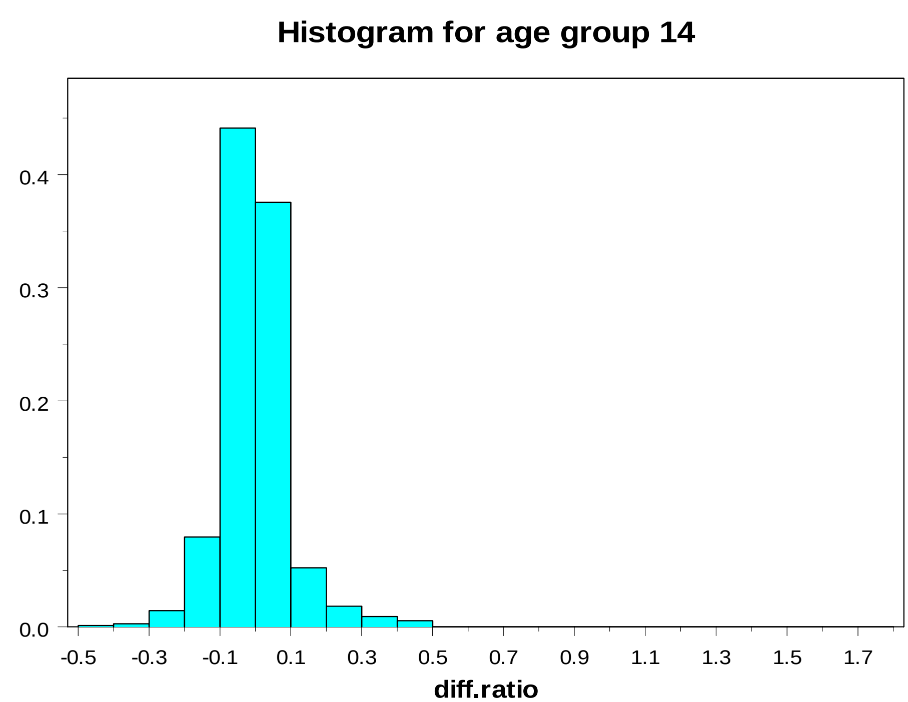


27(b) Histogram of the difference ratios for all years

Figure 27 The difference of the two population estimates at the county level for age group 13(1990-1999)

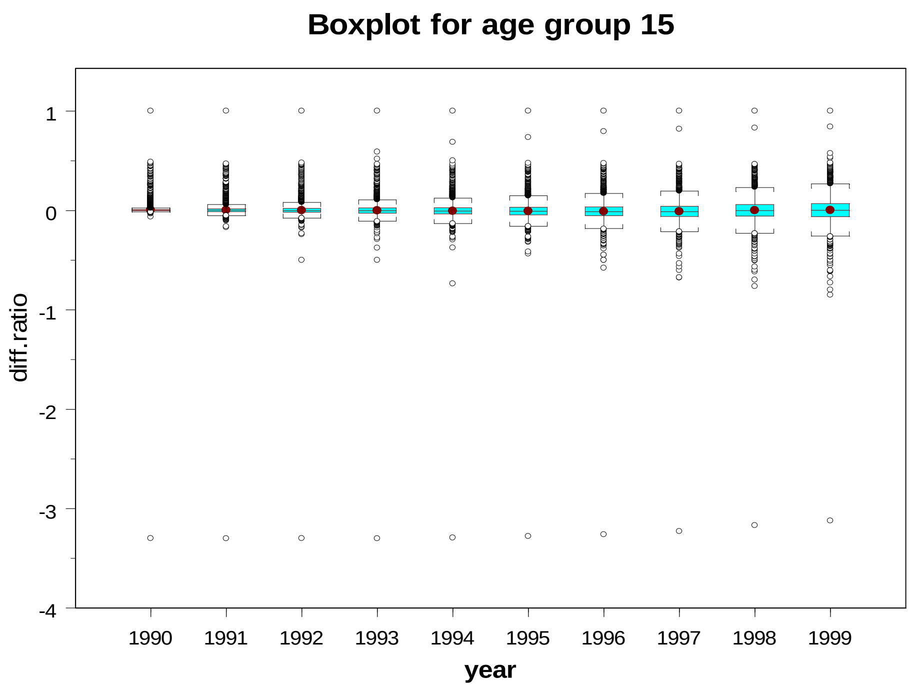


28(a) Boxplot of the difference ratios for each year

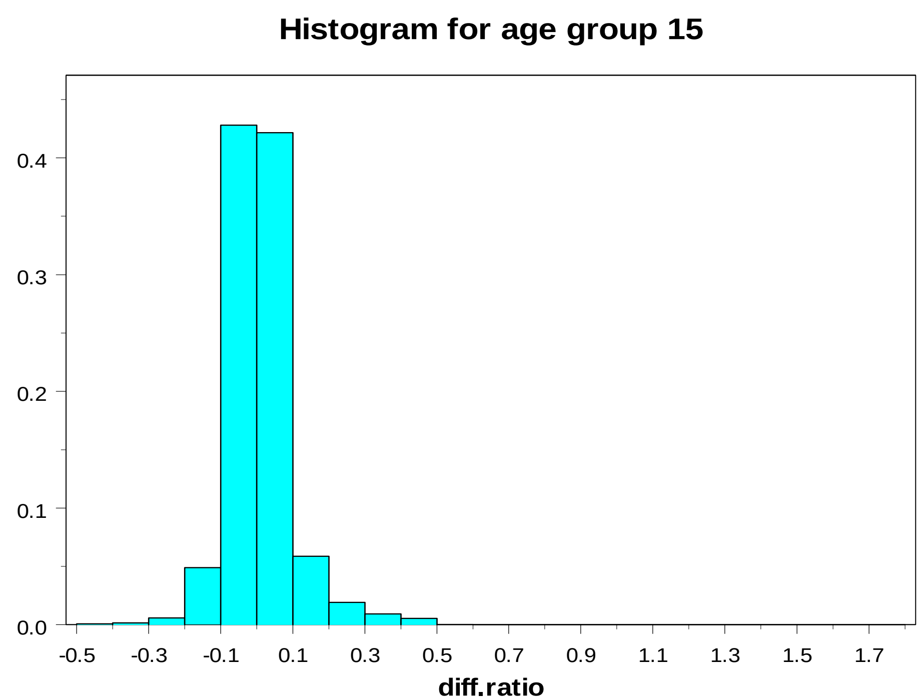


28(b) Histogram of the difference ratios for all years

Figure 28 The difference of the two population estimates at the county level for age group 14 (1990-1999)

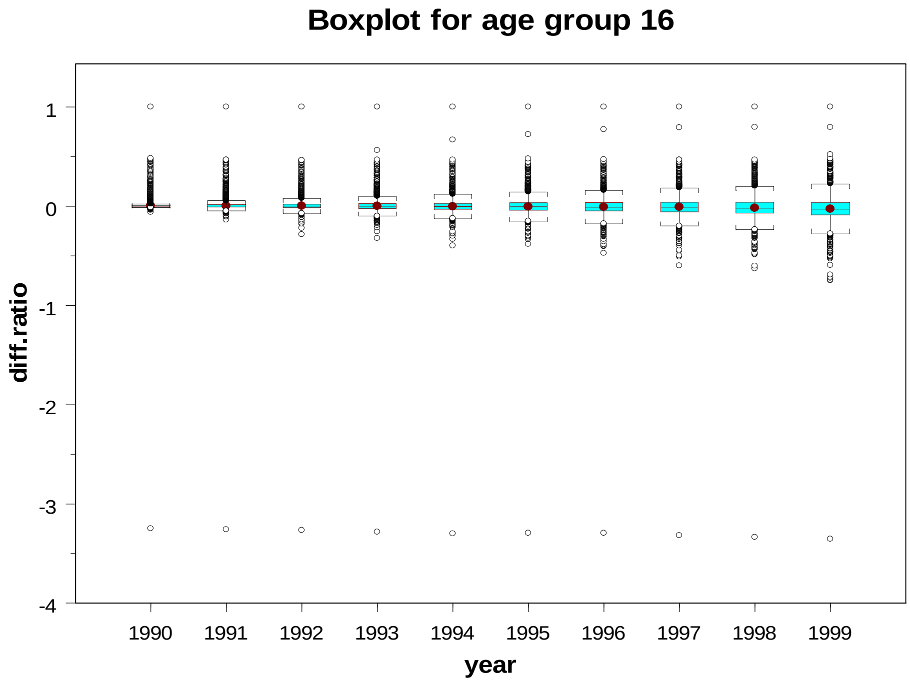


29(a) Boxplot of the difference ratios for each year

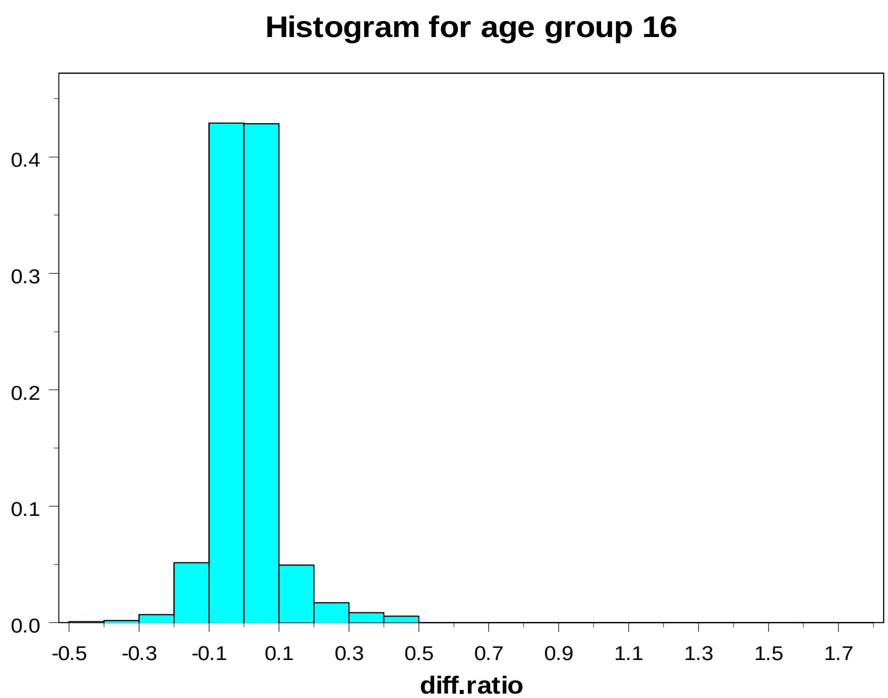


29(b) Histogram of the difference ratios for all years

Figure 29 The difference of the two population estimates at the county level for age group 15 (1990-1999)

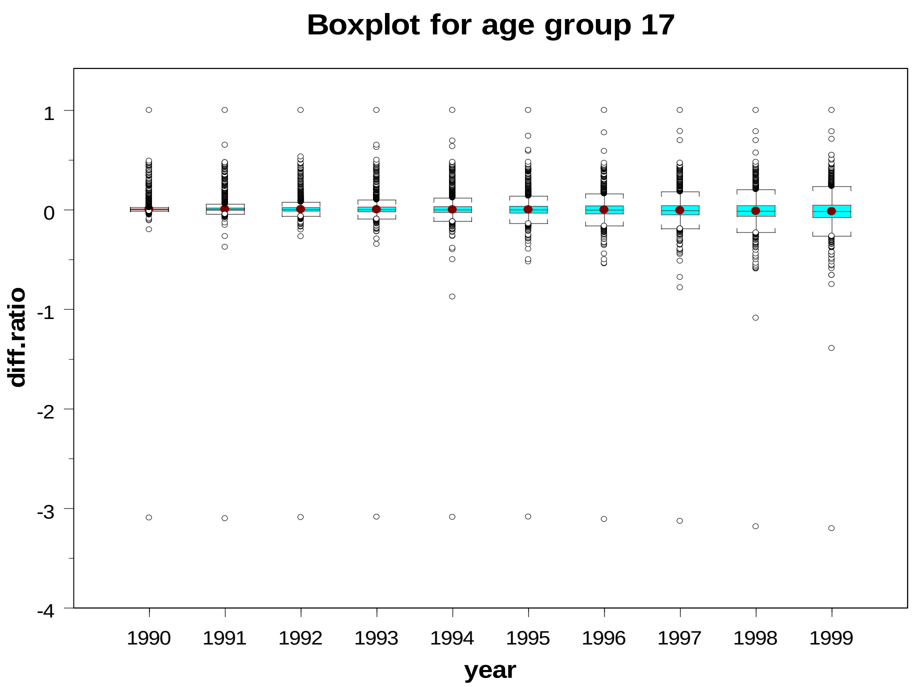


30(a) Boxplot of the difference ratios for each year

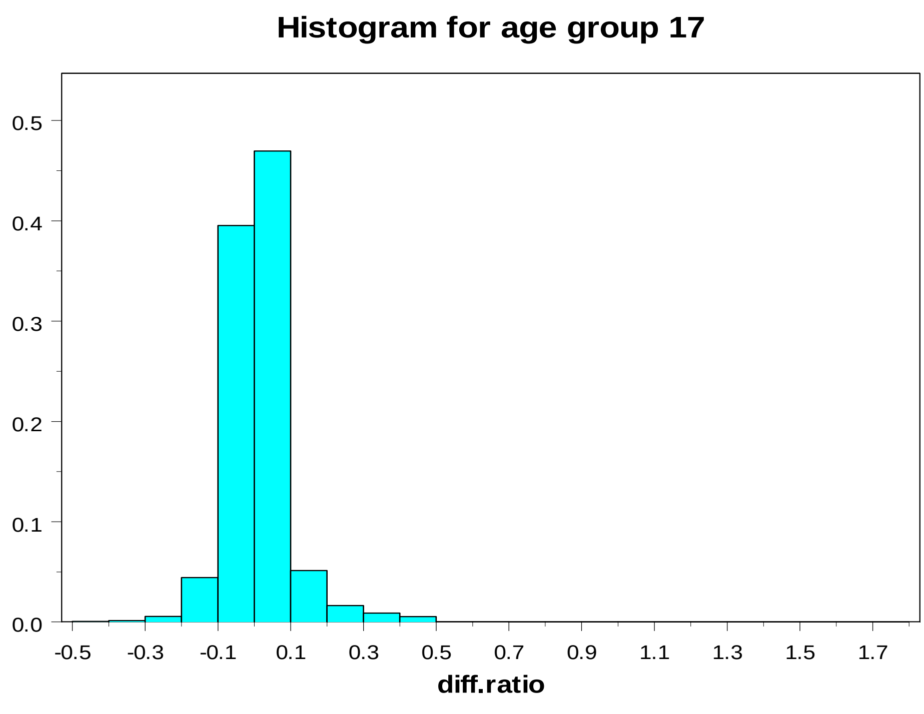


30(b) Histogram of the difference ratios for all years

Figure 30 The difference of the two population estimates at the county level for age group 16(1990-1999)

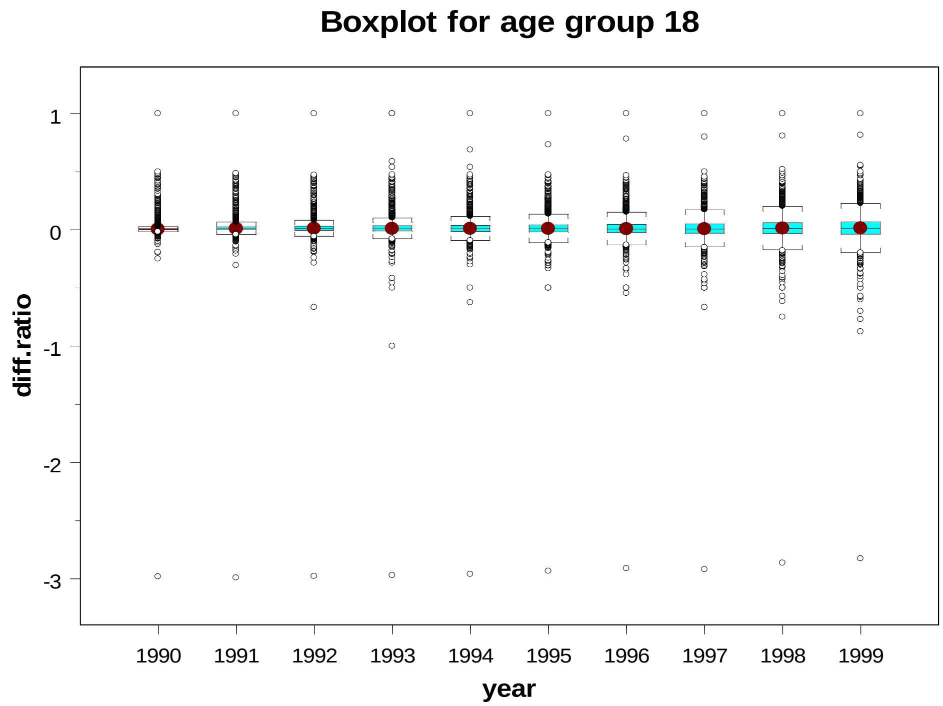


31(a) Boxplot of the difference ratios for each year

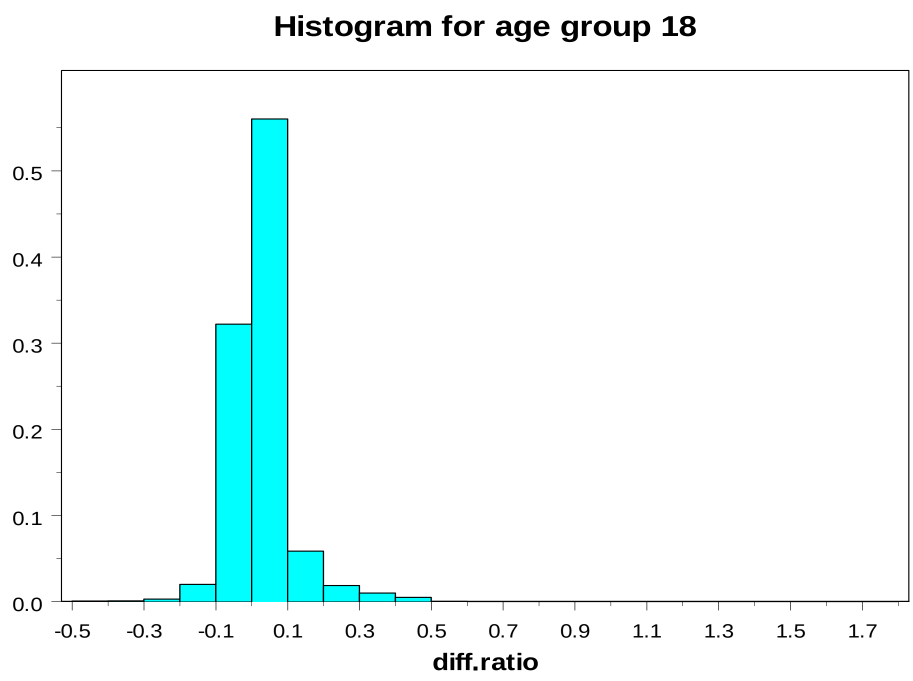


31(b) Histogram of the difference ratios for all years

Figure 31 The difference of the two population estimates at the county level for age group 17(1990-1999)

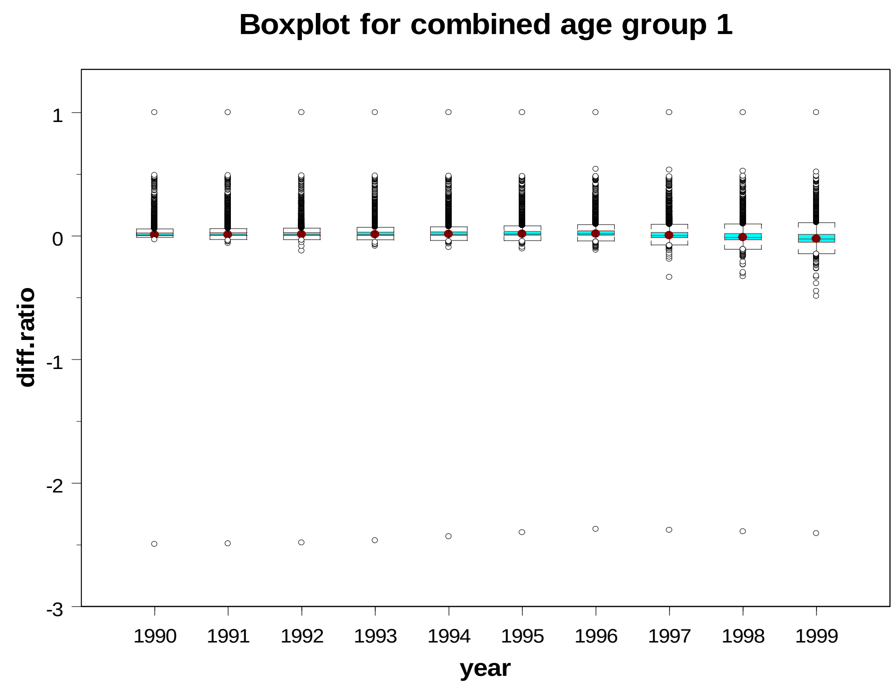


32(a) Boxplot of the difference ratios for each year

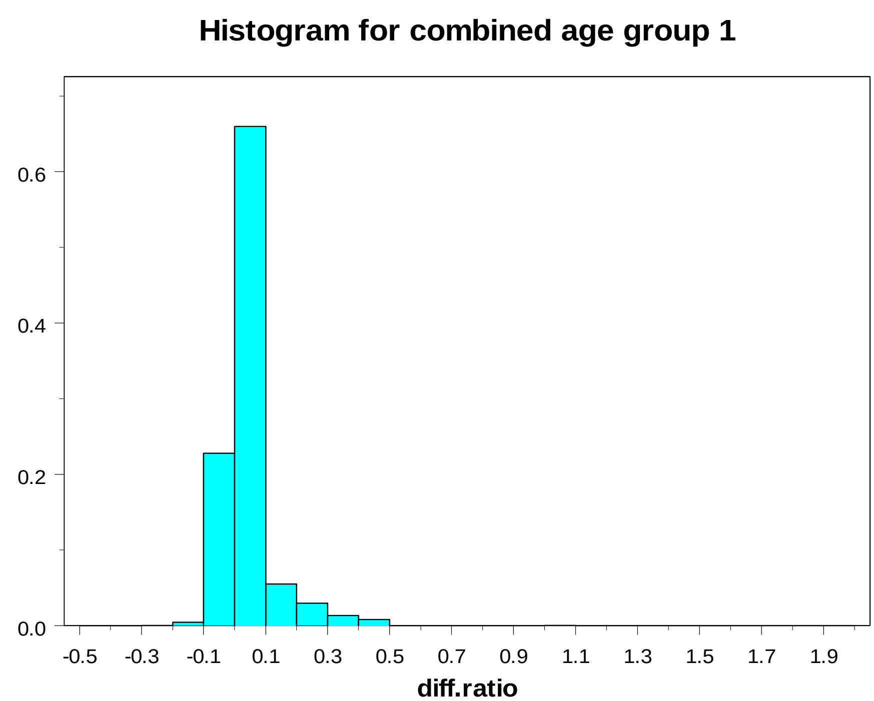


32(b) Histogram of the difference ratios for all years

Figure 32 The difference of the two population estimates at the county level for age group 18 (1990-1999)

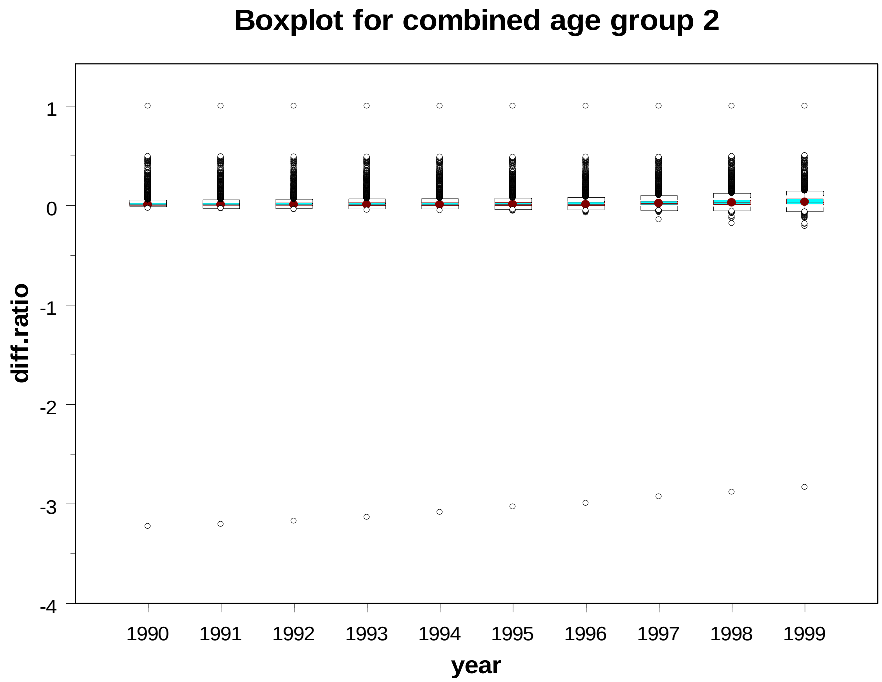


33(a) Boxplot of the difference ratios for each year

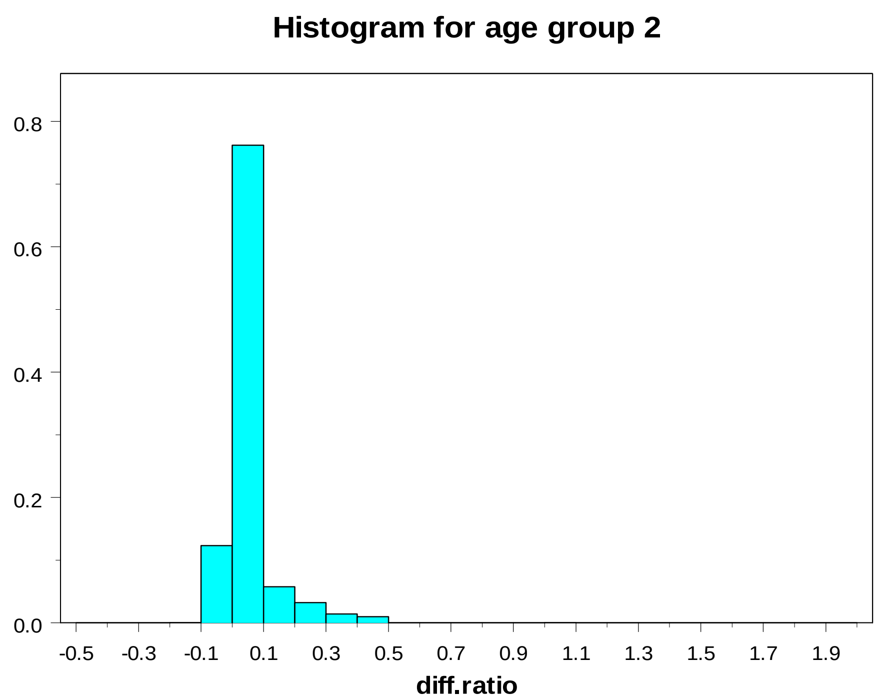


33 (b) Histogram of the difference ratios for all years

Figure 33 Difference of the two population estimates at the county level for combined age group 1 (1990-1999)



34(a) Boxplot of the difference ratios for each year



34(b) Histogram of the difference ratios for all years

Figure 34 Difference of the two population estimates at the county level for combined age group 2 (1990-1999)

4.4. Regression Analysis

Linear regression model was done to test the significance of the factors that may contribute to the difference between two population estimates after controlling for other variables.

4.4.1. Linear regression of difference vs. year, race, sex, age group

The linear regression of the difference between the two populations at the county level vs. year, race, sex, age group and new population estimates was done to identify the most important factors that contribute to the difference of the two populations. The model fits well with adjusted R^2 0.9103. The t statistics (the mean difference divided by its standard error for each variable) and the coefficients for each variable were compared. From the estimates of the parameters and the significance we can see that the variable year is not a significant predictor for the difference. Even with large sample size, the p-values for years are greater than 0.01 in this linear model. Race is also not a significant variable in this model ($p=0.212$). The sex variable has the largest $\beta/se(\beta)$, so it contributes most to the difference of the two population. For the age group, group 6-10 also contributes greatly to the difference, other groups are close.

Table 5 Parameter estimates and test statistics of the linear regression of difference vs. year, race, sex, age group and new population

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	B	179.831	8.121	22.14	<.0001
year90	B	-14.785	6.744	-2.19	0.028
year91	B	-13.881	6.744	-2.06	0.04
year92	B	-12.918	6.744	-1.92	0.055
year93	B	-11.784	6.744	-1.75	0.081
year94	B	-10.481	6.744	-1.55	0.12
year95	B	-9.131	6.744	-1.35	0.176
year96	B	-7.298	6.744	-1.08	0.279
year97	B	-4.604	6.744	-0.68	0.495
year98	B	-2.264	6.744	-0.34	0.737
Age(0-4)	B	-42.426	9.057	-4.68	<.0001
Age(5-9)	B	-74.49	9.056	-8.23	<.0001
Age(10-14)	B	-76.661	9.056	-8.47	<.0001
Age(15-19)	B	-75.49	9.055	-8.34	<.0001
Age(20-24)	B	-46.374	9.056	-5.12	<.0001
Age(25-29)	B	-105.969	9.056	**-11.7**	<.0001
Age(30-34)	B	-125.496	9.058	**-13.85**	<.0001
Age(35-39)	B	-155.524	9.058	**-17.17**	<.0001
Age(40-44)	B	-145.329	9.056	**-16.05**	<.0001
Age(45-49)	B	-159.78	9.053	**-17.65**	<.0001
Age(50-54)	B	-67.684	9.051	-7.48	<.0001
Age(55-59)	B	-66.731	9.05	-7.37	<.0001
Age(60-64)	B	-61.734	9.05	-6.82	<.0001
Age(65-69)	B	-75.863	9.049	-8.38	<.0001
Age(70-74)	B	-57.56	9.049	-6.36	<.0001
Age(75-79)	B	-40.509	9.049	-4.48	<.0001
Age(80-84)	B	-11.922	9.048	-1.32	0.188
Sex	1	-438.265	3.037	**144.29**	<.0001
Race	1	11.402	3.016	3.78	0.212
new_pop	1	0.163	0	593.48	<.0001

4.4.2. Difference vs. year, race, sex and combined age group

The previous analysis showed that the difference between the two populations among age group 6 to age group10 seemed to be more important than other age groups, so the age groups were combined into two age groups, agegroup1 with the agegrp6 to agegrp10, and agegroup2 from the original agegrp1-5 and agegrp11-18. The linear regression was done again for the difference between the two populations at the county level vs. year, race, sex, the combined age group and the new population. The model has R^2=0.9322. From Table 6 we can see that year and race are still not significant in this model (p-values>0.05). Sex has the largest t statistics, which means that it contributes most to the difference.

Table 6 Parameter estimates and test statistics of the linear regression of difference vs. year, race, sex, combined age group and new population

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	942.103	131.725	7.15	<.0001
Year90	1	-160.275	163.149	-0.98	0.3259
Year91	1	-149.147	163.148	-0.91	0.3606
Year92	1	-137.268	163.148	-0.84	0.4001
Year93	1	-123.900	163.147	-0.76	0.4476
Year94	1	-109.213	163.147	-0.67	0.5032
Year95	1	-94.153	163.147	-0.58	0.5639
Year96	1	-74.775	163.146	-0.46	0.6467
Year97	1	-47.438	163.146	-0.29	0.7712
Year98	1	-23.364	163.146	-0.14	0.8861
age	1	227.991	73.090	3.12	0.0018
sex	1	-3688.745	73.585	**-50.13**	<.0001
race	1	93.274	72.962	1.28	0.2011
new_pop	1	0.141	0.001	173.58	<.0001

4.4.3. Linear regression with interaction between age group and sex

The linear regression of the difference between the two populations at the county level vs. year, race, sex, the combined age group and the interaction between age group and sex was performed, with the new population data as an offset, to identify whether an interaction existed between age group and sex. Table 7 below showed that Sex still has a very large t statistic in this regression, the interaction term has a t statistic of 11.15, which is the second largest. Therefore there may be a small interaction between sex and age group.

Table 7 Parameter estimates and the test statistics of the linear regression of difference vs. year, race, sex, age group, new population and interaction

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1346.609	136.595	9.86	<.0001
Year90	1	-159.923	163.109	-0.98	0.3269
Year91	1	-148.833	163.108	-0.91	0.3615
Year92	1	-136.995	163.107	-0.84	0.401
Year93	1	-123.669	163.107	-0.76	0.4483
Year94	1	-109.021	163.106	-0.67	0.5039
Year95	1	-93.998	163.106	-0.58	0.5644
Year96	1	-74.657	163.106	-0.46	0.6472
Year97	1	-47.360	163.106	-0.29	0.7715
Year98	1	-23.326	163.105	-0.14	0.8863
age	1	-584.451	103.178	-5.66	<.0001
sex	1	-4506.002	103.833	**-43.4**	<.0001
race	1	93.395	72.944	1.28	0.2004
Age*sex	1	1627.888	145.956	**11.15**	<.0001
new_pop	1	0.141	0.001	173.88	<.0001

4.4.4. Linear regression without year and race

Because year and race were not significant in the linear model from previous analysis, they were dropped from the model and the regression was performed again, and the model had R^2=0.9340. The results are shown in Table 8. It can be seen that sex still contributes greatly to the difference between two populations.

Table 8 The parameter estimates and the test statistics of the linear regression of difference vs. sex, age group and interaction

(new population as an offset)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1301.495	73.135	17.8	<.0001
age	1	-584.438	103.176	-5.66	<.0001
sex	1	-4506.076	103.832	**-43.4**	<.0001
Age*sex	1	1627.917	145.954	**11.15**	<.0001
new_pop	1	0.141	0.001	173.9	<.0001

4.5. Effects on the calculation of the morality rates

The Standardized Mortality Ratio (SMR) is a measure of mortality in a study population, relative to mortality in a reference population. It is the morality rate with indirect adjustment of the strata (Gordis, 2000) and is calculated as the ratio of the sum of the observed cases in the population based all strata of stratification variables (age, year, sex et al), relative to the sum of the expected number of cases in the population (Checkoway et al, 1989). The expected numbers are based on the reference population. In this thesis, the two population estimates were used as the reference populations to get the standard rates, then the SMRs were calculated based on these rates.

The comparison of the SMRs using data from an occupational study by using new and old rates is shown in tables 9, 10 and 11. Table 9 is the comparison of SMRs from using new rate and old rate for all people in the cohort, while Table 10 and Table 11 are for males and females respectively. By comparing the statistical significance of the SMR for each cause of death, it can be seen from Table 9 for all cohort member, the statistical significant SMRs ($p<0.05$) using new rates remain statistically significant using old rates, except for cancer of large intestine. For males, statistical significance for most causes is the same except for cancer of Ulcer of stomach & Duodenum. For females in Table 11, the statistical significance levels (0.05 or 0.01) are different for several causes.

Figure 35-37 displayed SMRs and the 95% confidence intervals for major categories of cause of death relative to rates from new population estimates and rates for current population estimates for total cohort, males and females respectively. From figure 35 it can be seen that the 95% confidence intervals from all categories overlap and the significance of the two groups of SMRs are consistent. For both male and female in figure 36 and figure 37, the significance of SMRs is consistent and the 95% confidence intervals overlap. Therefore, we can conclude that the difference of the two population estimates has no importatn effect on the calculation of SMRs for the occupational cohort data used in this thesis.

Table 9 SMRs using new rates and old rates for data from an occupational cohort study, total cohort, 1980-1998

Cause of Death	New Rates				Old Rates			
	Obs	SMR	LL	UL	Obs	SMR	LL	UL
All Causes of Death	2929	68.2**	65.7	70.7	2929	66.0**	63.7	68.5
Tuberculosis	0	---*	0.0	84.8	0	--*	0.0	78.5
All Malignant Neoplasms	1039	81.1**	76.9	86.9	1039	79.6**	74.8	84.5
Cancer of Buccal Cavity & Pharynx	17	58.8*	34.3	94.2	17	56.6*	33.0	90.6
Cancer of Digestive Organs & Peritoneum	224	77.3**	67.5	88.1	224	74.9**	65.4	85.3
Cancer of Esophagus	24	65.8*	42.2	98.0	24	63.2*	40.5	94.0
Cancer of Stomach	26	74.9	48.9	109.8	26	71.9	47.0	105.4
Cancer of Large Intestine	81	81.0	64.3	100.6	81	78.8*	62.6	98.0
Cancer of Rectum	13	64.4	34.3	110.2	13	62.8	33.4	107.4
Cancer of Biliary Passages & Liver	19	59.4*	35.8	92.8	19	56.9**	34.2	88.8
Cancer of Pancreas	56	93.6	70.7	121.5	56	91.0	68.8	118.2
Cancer of All Other Digestive Organs	5	75.8	24.6	176.9	5	74.0	24.0	172.7
Cancer of Respiratory System	356	76.4**	68.7	84.8	356	74.5**	66.9	82.6
Cancer of Larynx	6	41.1*	15.1	89.5	6	39.5*	14.5	86.0
Cancer of Bronchus, Trachea, Lung	350	78.3**	70.3	86.9	350	76.3**	68.5	84.7
Cancer of All Other Respiratory	0	-- *	0.0	89.1	0	-- *	0.0	86.9
Cancer of Breast	36	70.2*	49.2	97.2	36	68.0*	47.6	94.1
All Uterine Cancers (Females only)	4	35.2*	9.6	90.2	4	33.8*	9.2	86.5
Cancer of Cervix Uteri (Females only)	1	15.5*	0.4	86.4	1	14.7 *	0.4	82.0
Cancer of Other Female Genital Organs	13	89.5	47.6	153.0	13	87.8	46.8	150.2
Cancer of Prostate (Males only)	63	92.9	71.4	118.9	63	88.9	68.3	113.7
Cancer of Testes and Other Male Genital Organs	2	80.2	9.1	289.6	2	78.8	9.5	284.7
Cancer of Kidney	25	79.8	51.7	117.8	25	78.1	50.5	115.3
Cancer of Bladder and Other Urinary Organs	13	57.9*	30.8	99.0	13	56.9*	30.3	97.3
Malignant Melanoma of Skin	21	93.7	58.0	143.3	21	93.4	57.8	142.7
Cancer of Eye	1	158.5	4.0	883.1	1	157.1	3.9	875.1
Cancer of Central Nervous System	45	122.8	89.5	164.3	45	121.1	88.4	162.1
Cancer of Thyroid & Other Endocrine Glands	5	134.9	43.8	314.9	5	132.3	42.9	308.7
Cancer of Bone	1	38.6	1.0	214.9	1	37.5	0.9	209.0

Table 9 (continued)

Cancer of All Lymphatic, Haematopoietic Tissue	125	107.3	89.3	127.8	125	104.7	87.2	124.8
Lymphosarcoma & Reticulosarcoma	5	84.7	27.5	197.6	5	83.9	27.2	195.9
Hodgkins Disease	4	78.0	21.2	199.6	4	76.4	20.8	195.7
Leukemia & Aleukemia	50	116.7	86.6	153.9	50	114.2	84.7	150.5
Cancer of All Other Lymphopoietic Tissue	66	105.3	81.5	134.0	66	102.5	79.3	130.4
All Other Malignant Neoplasms	88	86.5	69.4	106.6	88	84.1	67.4	103.6
Benign Neoplasms	17	128.3	74.7	205.4	17	124.8	72.7	199.7
Diabetes Mellitus	51	53.0**	39.5	69.7	51	50.7**	37.7	66.6
Cerebrovascular Disease	118	67.2**	55.6	80.4	118	64.4**	53.3	77.2
All Heart Disease	969	72.8**	68.3	77.5	969	70.9**	66.5	75.5
Rheumatic Heart Disease	6	56.4	20.7	122.7	6	55.2	20.3	120.1
Ischemic Heart Disease	694	75.0**	69.5	80.8	694	73.5**	68.1	79.1
Chronic Endocard. Dis.; Other Myocard. Insuff.	38	69.0*	48.8	94.7	38	67.0*	47.4	92.0
Hypertension with Heart Disease	30	68.1*	46.0	97.3	30	63.3**	42.7	90.4
All Other Heart Disease	201	68.0**	58.9	78.1	201	65.3**	56.6	75.0
Hypertension w/o Heart Disease	6	42.5*	15.6	92.6	6	39.6*	14.5	86.2
Non-malignant Respiratory Disease	144	49.0**	41.4	57.7	144	47.8**	40.3	56.2
Influenza & Pneumonia	39	49.5**	35.2	67.7	39	47.7**	33.9	65.1
Bronchitis, Emphysema, Asthma	21	42.1**	26.1	64.4	21	41.1**	25.4	62.8
Bronchitis	3	52.4	10.8	153.2	3	51.5	10.6	150.5
Emphysema	13	37.4**	19.9	64.0	13	36.8**	19.6	63.0
Asthma	5	53.4	17.3	124.6	5	50.0	16.2	116.8
Other Non-malignant Respiratory Disease	84	50.9**	40.6	63.1	84	49.9**	39.8	61.7
Ulcer of Stomach & Duodenum	3	30.4*	6.3	88.9	3	29.5*	6.1	86.3
Cirrhosis of Liver	49	44.6**	33.0	58.9	49	43.4**	32.1	57.4
Nephritis & Nephrosis	12	39.2**	20.3	68.5	12	37.3**	19.3	65.2
All External Causes of Death	234	54.8**	48.0	62.3	234	52.9**	46.3	60.1
Accidents	134	55.5**	46.5	65.7	134	53.8**	45.1	63.7
Motor Vehicle Accidents	82	68.5**	54.5	85.1	82	66.6**	53.0	82.7
All Other Accidents	52	42.7**	31.9	56.0	52	41.3**	30.8	54.1
Suicides	67	60.6**	46.9	76.9	67	59.6**	46.2	75.7
Homicides & Other External Causes	33	44.1**	30.4	62.0	33	40.8**	28.1	57.4
All Other Causes of Death	259	58.0**	51.2	65.5	259	55.9**	49.3	63.1
AIDS	3	3.6**	0.8	10.6	3	3.3**	0.7	9.6
Unknown Causes (In All Causes Category Only)	25				25			

* $p < .05$
** $p < .01$
LL-UL = 95% Confidence Interval

Table 10 SMRs using new rates and old rates for data from an occupational cohort study, Male study members, 1980-1998

Cause of Death	New Rates				Old Rates			
	Obs	SMR	LL	UL	Obs	SMR	LL	UL
All Causes of Death	2476	67.2**	64.5	69.9	2476	65.1**	62.5	67.7
Tuberculosis	0	---*	0.0	95.2	0	---*	0.0	88.4
All Malignant Neoplasms	849	81.4**	76.0	87.1	849	79.1**	73.9	84.6
Cancer of Buccal Cavity & Pharynx	14	53.3*	29.1	89.4	14	51.2**	28.0	86.0
Cancer of Digestive Organs & Peritoneum	189	75.9**	65.5	87.5	189	73.5**	63.4	84.7
Cancer of Esophagus	22	64.0*	40.1	96.9	22	61.5*	38.5	93.0
Cancer of Stomach	22	71.3	44.7	107.9	22	68.4	42.9	103.6
Cancer of Large Intestine	69	82.8	64.4	104.8	69	80.7	62.8	102.1
Cancer of Rectum	12	69.5	35.9	121.4	12	67.7	35.0	118.3
Cancer of Biliary Passages & Liver	17	62.0*	36.1	99.3	17	59.3*	34.5	94.9
Cancer of Pancreas	44	87.4	63.5	117.3	44	85.0	61.8	114.1
Cancer of All Other Digestive Organs	3	55.3	11.4	161.5	3	54.0	11.1	157.7
Cancer of Respiratory System	301	73.5*	65.5	82.3	301	71.6**	63.7	80.1
Cancer of Larynx	6	43.7*	16.0	95.2	6	42.0*	15.4	91.3
Cancer of Bronchus, Trachea, Lung	295	75.3**	66.9	84.4	295	73.3**	65.2	82.2
Cancer of All Other Respiratory	0	---	0.0	100.8	0	---*	0.0	98.4
Cancer of Breast	1	83.0	2.1	462.3	1	80.1	2.0	446.3
All Uterine Cancers (Females only)	0	---	---	---	0	---	---	---
Cancer of Cervix Uteri (Females only)	0	---	---	---	0	---	---	---
Cancer of Other Female Genital Organs	0	---	---	---	0	---	---	---
Cancer of Prostate (Males only)	63	92.9	71.4	118.9	63	88.9	68.3	113.7
Cancer of Testes and Other Male Genital Organs	2	80.2	9.7	289.6	2	78.8	9.5	284.8
Cancer of Kidney	24	86.7	55.5	129.0	24	84.8	54.3	126.2
Cancer of Bladder and Other Urinary Organs	13	62.9	33.5	107.5	13	61.8	32.9	105.7
Malignant Melanoma of Skin	21	108.8	67.3	166.3	21	108.4	67.1	165.7
Cancer of Eye	1	192.1	4.8	1070.3	1	190.5	4.8	1061.5
Cancer of Central Nervous System	39	128.4	91.3	175.5	39	126.8	90.2	173.3
Cancer of Thyroid & Other Endocrine Glands	3	103.4	21.3	302.2	3	101.5	21.0	296.7

Table 10 (continued)

Cancer of Bone	1	46.3	1.2	257.9	1	45.1	1.1	251.4
Cancer of All Lymphatic, Haematopoietic Tissue	105	107.2	87.7	129.8	105	104.8	85.7	126.8
Lymphosarcoma & Reticulosarcoma	3	59.9	12.4	175.0	3	59.4	12.3	173.5
Hodgkins Disease	3	70.2	14.5	205.1	3	68.9	14.2	201.3
Leukemia & Aleukemia	42	116.4	83.9	157.3	42	113.9	82.1	154.0
Cancer of All Other Lymphopoietic Tissue	57	108.5	82.2	140.6	57	105.6	80.0	136.8
All Other Malignant Neoplasms	72	84.3	66.0	106.2	72	81.9	64.1	103.2
Benign Neoplasms	11	103.0	51.4	184.2	11	100.3	50.0	179.4
Diabetes Mellitus	45	57.9**	42.2	77.5	45	55.4**	40.4	74.1
Cerebrovascular Disease	88	61.1**	49.0	75.2	88	58.6**	47.0	72.1
All Heart Disease	876	73.5**	68.7	78.5	876	71.6**	66.9	76.5
Rheumatic Heart Disease	4	54.4	14.8	139.3	4	53.4	14.5	136.7
Ischemic Heart Disease	630	75.1**	69.3	81.2	630	73.6**	67.9	79.6
Chronic Endocard. Dis.; Other Myocard. Insuff.	33	70.3*	48.4	98.8	33	68.3*	47.0	95.9
Hypertension with Heart Disease	29	76.7	51.3	110.1	29	71.4	47.8	102.5
All Other Heart Disease	180	68.9**	59.2	79.8	180	66.3**	57.0	76.8
Hypertension w/o Heart Disease	6	50.9	18.7	110.7	6	47.4	17.4	103.1
Non-malignant Respiratory Disease	118	47.3**	39.2	56.7	118	46.1**	38.1	55.2
Influenza & Pneumonia	33	48.6**	33.5	68.3	33	46.8**	32.2	65.7
Bronchitis, Emphysema, Asthma	15	36.7**	20.5	60.5	15	35.8**	20.0	59.1
Bronchitis	2	41.9	5.1	151.2	2	41.1	5.0	148.5
Emphysema	11	37.1**	18.5	66.3	11	36.5**	18.2	65.2
Asthma	2	31.0	3.8	112.0	2	29.2	3.5	105.3
Other Non-malignant Respiratory Disease	70	49.8**	38.8	62.9	70	48.7**	38.0	61.5
Ulcer of Stomach & Duodenum	3	35.0	7.2	102.3	3	33.9*	7.0	99.2
Cirrhosis of Liver	43	44.4**	32.2	59.9	43	43.3**	31.3	58.4
Nephritis & Nephrosis	11	42.7**	21.3	76.4	11	40.7**	20.3	72.7
All External Causes of Death	193	50.7**	43.8	58.4	193	49.0**	42.4	56.5
Accidents	107	50.0**	41.0	60.5	107	48.6**	39.8	58.7
Motor Vehicle Accidents	65	62.8**	48.4	80.0	65	61.1**	47.1	77.9
All Other Accidents	42	38.1**	27.5	51.5	42	36.9**	26.6	49.8
Suicides	61	60.8**	46.5	78.1	61	59.9**	45.8	76.9
Homicides & Other External Causes	25	37.8**	24.4	55.8	25	35.0**	22.7	51.7
All Other Causes of Death	212	56.8**	49.4	64.9	212	54.8**	47.6	62.7
AIDS	3	3.9 **	0.8	11.4	3	3.5**	0.7	10.3
Unknown Causes (In All Causes Category Only)	18				18			

* $p < .05$
** $p < .01$
LL-UL = 95% Confidence Interval

Table 11 SMRs using new and old rates for data from an occupational cohort study, Female study members, 1980-1998

Cause of Death	New Rates				Old Rates			
	Obs	SMR	LL	UL	Obs	SMR	LL	UL
All Causes of Death	453	74.4**	67.7	81.6	453	71.8**	65.3	78.7
Tuberculosis	0	---	0.0	772.3	0	---	0.0	698.0
All Malignant Neoplasms	190	83.7*	72.2	96.35	190	81.5**	70.3	94.0
Cancer of Buccal Cavity & Pharynx	3	114.3	23.6	334.1	3	110.6	22.8	323.1
Cancer of Digestive Organs & Peritoneum	35	85.7	59.7	119.3	35	83.2	58.0	115.7
Cancer of Esophagus	2	96.5	11.7	348.7	2	91.9	11.1	332.1
Cancer of Stomach	4	104.2	28.4	266.8	4	99.7	27.2	255.3
Cancer of Large Intestine	12	71.7	37.1	125.3	12	69.8	36.1	121.9
Cancer of Rectum	1	34.3	0.9	191.1	1	33.4	0.8	186.3
Cancer of Biliary Passages & Liver	2	43.7	5.3	158.0	2	42.3	5.1	152.9
Cancer of Pancreas	12	126.1	65.2	220.3	12	122.9	63.5	214.8
Cancer of All Other Digestive Organs	2	171.3	20.7	618.9	2	167.1	20.2	603.7
Cancer of Respiratory System	55	97.5	73.4	126.9	55	95.5	71.9	124.3
Cancer of Larynx	0	---	0.0	427.4	0	---	0.0	411.6
Cancer of Bronchus, Trachea, Lung	55	99.9	75.2	130.0	55	97.8	73.7	127.3
Cancer of All Other Respiratory	0	---	0.0	768.1	0	---	0.0	746.6
Cancer of Breast	35	69.9*	48.7	97.2	35	67.7*	47.2	94.1
All Uterine Cancers (Females only)	4	35.2*	9.6	90.2	4	33.8*	9.2	86.5
Cancer of Cervix Uteri (Females only)	1	15.5*	0.4	86.4	1	14.7*	0.4	82.0
Cancer of Other Female Genital Organs	13	89.5	47.6	153.0	13	87.8	46.8	150.8
Cancer of Prostate (Males only)	0	---	0.0	---	0	---	0.0	---
Cancer of Testes and Other Male Genital Organs	0	---	0.0	---	0	---	0.0	---
Cancer of Kidney	1	27.5	0.7	153.5	1	26.9	0.7	150.1
Cancer of Bladder and Other Urinary Organs	0	---	0.0	208.0	0	---	0.0	203.2
Malignant Melanoma of Skin	0	---	0.0	119.2	0	---	0.0	118.5
Cancer of Eye	0	---	0.0	3343.6	0	---	0.0	3300.5
Cancer of Central Nervous System	6	95.4	35.0	207.7	6	93.9	34.5	204.4
Cancer of Thyroid & Other Endocrine Glands	2	248.8	30.1	898.7	2	242.4	29.3	875.6

Table 11 (continued)

Cancer of Bone	0	---	0.0	851.9	0	---	0.0	821.2
Cancer of All Lymphatic, Haematopoietic Tissue	20	107.3	65.5	165.7	20	104.6	63.9	161.5
Lymphosarcoma & Reticulosarcoma	2	223.6	27.1	807.6	2	221.3	26.8	799.5
Hodgkins Disease	1	116.9	2.9	651.2	1	113.8	2.8	634.3
Leukemia & Aleukemia	8	118.5	51.2	233.5	8	115.5	49.9	227.6
Cancer of All Other Lymphopoietic Tissue	9	88.8	40.6	168.5	9	86.4	39.5	164.0
All Other Malignant Neoplasms	16	98.1	56.0	159.2	16	95.3	54.4	154.7
Benign Neoplasms	6	233.2	85.6	507.7	6	226.0	82.9	491.9
Diabetes Mellitus	6	32.5**	11.9	70.6	6	30.9**	11.3	67.3
Cerebrovascular Disease	30	95.0	64.1	135.7	30	91.3	61.6	130.3
All Heart Disease	93	67.1**	54.1	82.2	93	64.9**	52.4	79.6
Rheumatic Heart Disease	2	60.8	7.4	219.5	2	59.2	7.2	213.9
Ischemic Heart Disease	64	74.1*	57.1	94.6	64	72.4**	55.7	92.4
Chronic Endocard. Dis.; Other Myocard. Insuff.	5	61.2	19.9	142.9	5	59.5	19.3	138.8
Hypertension with Heart Disease	1	16.1*	0.4	89.9	1	14.8*	0.4	82.5
All Other Heart Disease	21	60.7*	37.6	92.8	21	58.0**	35.9	88.6
Hypertension w/o Heart Disease	0	---	0.0	159.6	0	---	0.0	148.3
Non-malignant Respiratory Disease	26	58.8**	38.4	86.2	26	57.4**	37.5	84.1
Influenza & Pneumonia	6	55.0	20.2	119.8	6	53.1	19.5	115.6
Bronchitis, Emphysema, Asthma	6	67.1	24.6	146.1	6	65.1	23.9	141.7
Bronchitis	1	105.9	2.6	590.1	1	104.3	2.6	581.1
Emphysema	2	39.4	4.8	142.2	2	39.0	4.7	141.0
Asthma	3	102.9	21.2	300.6	3	95.8	19.8	279.9
Other Non-malignant Respiratory Disease	14	57.5*	31.4	96.4	14	56.5*	30.9	94.8
Ulcer of Stomach & Duodenum	0	---	0.0	286.4	0	---	0.0	279.1
Cirrhosis of Liver	6	45.7*	16.8	99.5	6	44.1*	16.2	95.9
Nephritis & Nephrosis	1	20.6	0.5	114.7	1	19.6	0.5	109.2
All External Causes of Death	41	88.2	63.3	119.7	41	84.2	60.4	114.2
Accidents	27	97.8	64.4	142.2	27	93.7	61.7	136.3
Motor Vehicle Accidents	17	106.0	61.7	169.7	17	101.8	59.3	163.0
All Other Accidents	10	86.4	41.4	158.9	10	82.5	39.6	151.7
Suicides	6	58.3	21.4	126.9	6	57.1	21.0	124.3
Homicides & Other External Causes	8	93.3	40.3	183.8	8	85.4	36.9	168.2
All Other Causes of Death	47	64.4**	47.3	85.6	47	61.6**	45.2	81.9
AIDS	0	---**	0.0	69.0	0	---**	0.0	55.2
Unknown Causes (In All Causes Category Only)	7				7			

* p < .05
** p < .01
LL-UL = 95% Confidence Interval

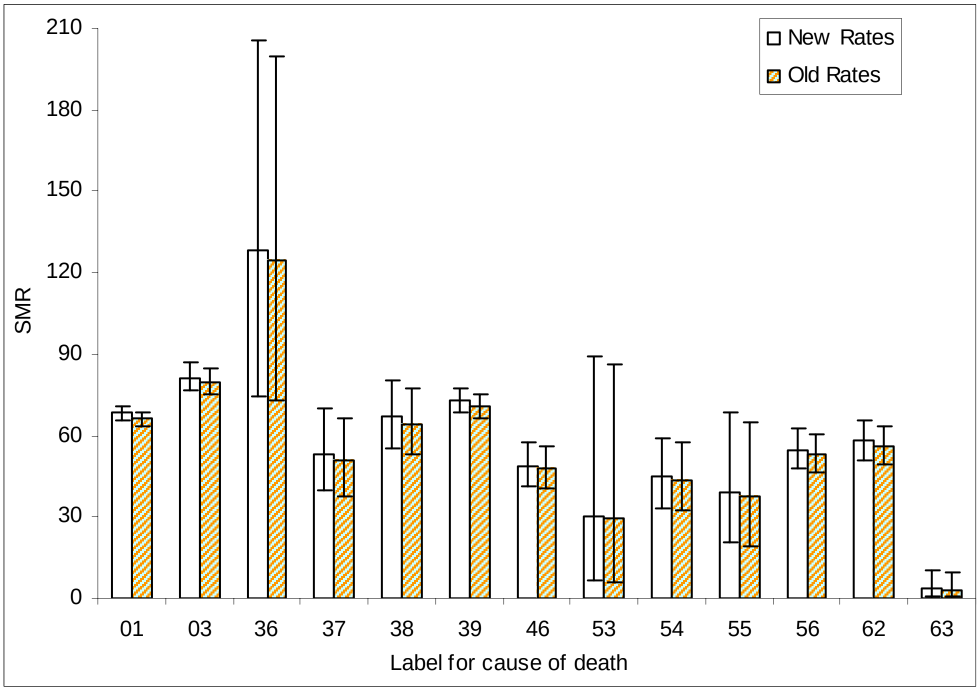


Figure 35 Comparison of SMRs from new and old rates for the whole cohort data

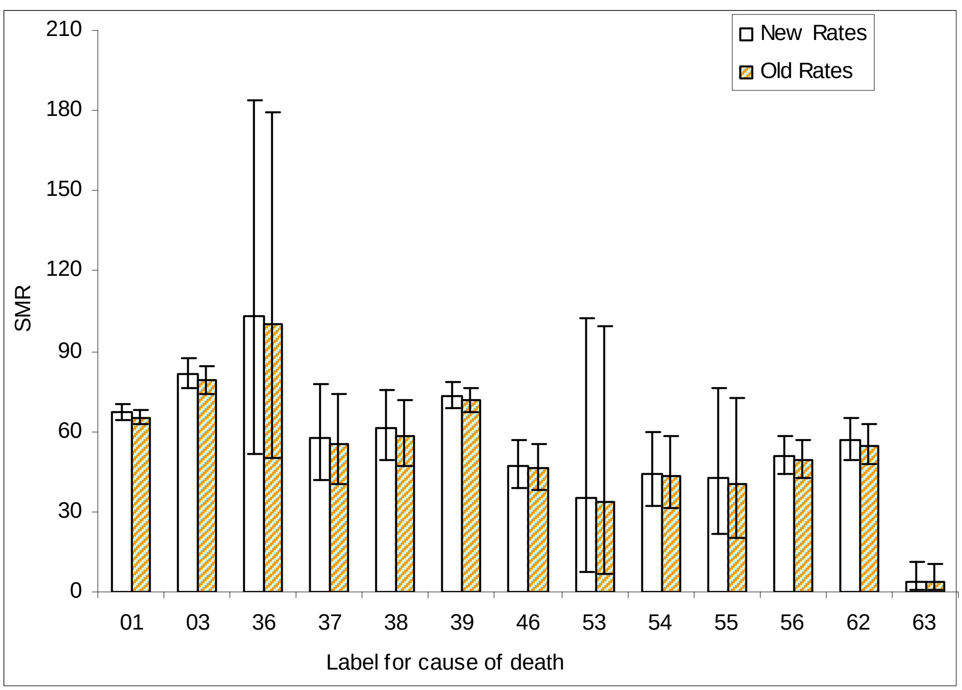


Figure 36 Comparison of SMRs from new and old rates for males

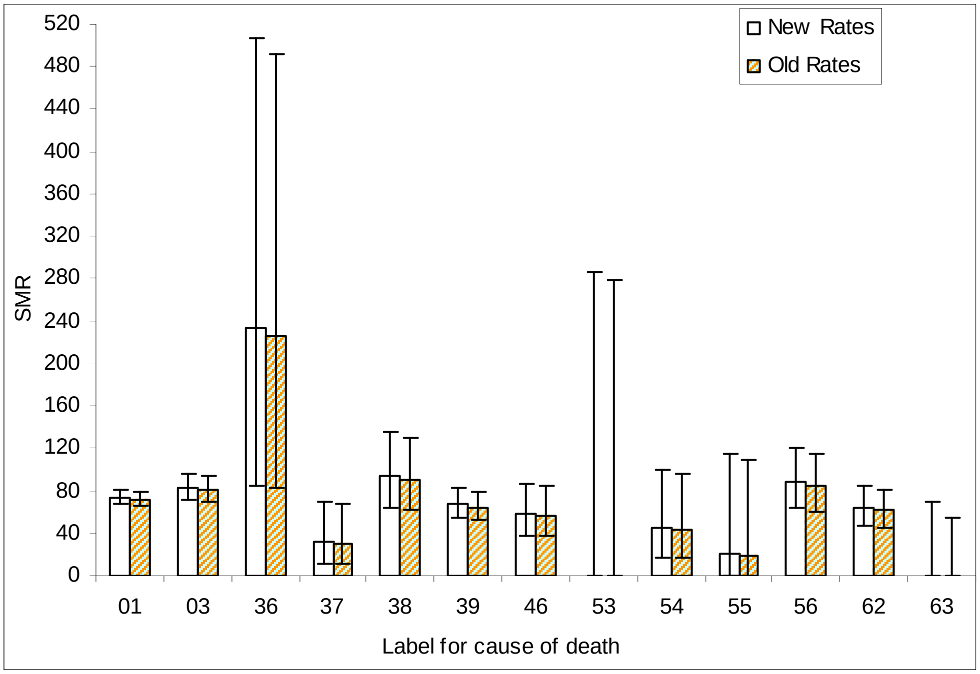


Figure 37 Comparison of SMRs from new and old rates for females

5. CONCLUSIONS

From the results of the descriptive statistics, the graphical displays and the regression analysis, it appears that sex is an important factor that affects the difference of the two population estimates during the period of 1990-1999. The total difference of the two population estimates comes mostly from the difference between the two female groups. The difference of the two population estimates has no important effects on the calculation of SMRs using the occupational cohort data in the example. Because these data are typical of cohort study data, we may infer that the SMRs with these two population estimates as reference populations would not be different using other similar occupational cohort. Therefore, the use of new populations would be sufficient. However, for cohort studies during the period of 1990-1999 on specific subgroups, such as females in some area with big difference between these two population estimates, the effect may not be ignorable.

APPENDIX A: Mortality and Population Data System (2004) ICDA Elements of Default Cause of Death List - 63 Causes

University of Pittsburgh
Mortality and Population Data System (2004)
ICDA Elements of Default Cause of Death List - 63 Causes

Label	Cause of Death	6th & 7th Revision (1950-67)	8th Revision (1968-78)	9th Revision (1979+)	10th Revision* (1999+)
01	All Causes of Death	001-999	000-999	001-999	A00.-Y89.
02	Tuberculosis	001-019	010-019	010-018	A15.-A19.
03	All Malignant Neoplasms	140-205	140-209	140-208	C00.-C97.
04	Buccal Cavity and Pharynx	140-148	140-149	140-149	C00.-C14.
05	Digestive Organs and Peritoneum	150-159	150-159	150-159	C15.-C26., C48.
06	Esophagus	150*	150	150*	C15.
07	Stomach	151	151	151*	C16.
08	Large Intestine	153	153	153*	C18.
09	Rectum	154	154	154*	C20-C21.
10	Biliary Passages and Liver Primary	155	155,156	155,156*	C22.,C24.
11	Pancreas	157*	157	157*	C25.
12	All Other Digestive	152,156,158,159*	152,158,159	152,158,159*	C17.,C19., C.23., C26.,C48.
13	Respiratory System	160-164	160-163	160-165	C30.-C39.
14	Larynx	161*	161*	161*	C32.
15	Bronchus, Trachea, Lung	162,163	162	162*	C33-C34.
16	All Other Respiratory	160,164	160,163	160,163,164,165*	C30.-C31., C37.-C39.
17	Breast	170	174	174,175	C50.
18	All Uterine (female only)	171,172-174	180,181,182.0*, 182.9	179,180,181,182*	C53.-C55.
19	Cervix (female only)	171	180*	180*	C53.

Appendix A (Continued)

20	Other Female Genital Organs (female only)	175,176	183-184*	183-184*	C51.-C52., C56.-C58.
21	Prostate (male only)	177	185	185*	C61
22	Testis and Other Male Genital Organs (male only)	178,179*	172.5,173.5,186* 187	186,187*	C60.,C62.-C63.
23	Kidney	180	189.0,189.1,189.2	189.0,189.1,189.2	C64.-C65
24	Bladder and Other Urinary Organs	181	188,189.9	188,189.3,189.4, 189.8, 189.9	C66.-C68.
25	Malignant Melanoma of Skin	190	172.0-172.4* 172.6-172.9	172*	C43.
26	Eye	192*	190*	190*	C69.
27	Central Nervous System	193*	191,192*	191,192*	C70.-C72.
28	Thyroid Gland and Other Endocrine Glands and Related Structures	194,195*	193,194*	193,194*	C73.-C75.
29	Bone	196	170*	170*	C40.-C41.
30	All Lymphatic and Hematopoietic Tissue	200-205	200-209	200-208*	C81-C96.
31	Hodgkins Disease	201	201	201*	C81.
32	Non-Hodgkins Lymphoma	200, 202, 205*	200, 202*	200, 202.0, 202.1, 202.8, 202.9*	C82, C83.0-C83.8, C84, C85.1-C85.9
33	Leukemia and Aleukemia	204	204-207	204-208	C91.-C95.
34	All Other Lymphopoietic Tissue	203*	203, 208, 209*	202.2, 202.3, 202.4, 202.5, 202.6, ,203*	C88., C90., C96.
35	All Other Malignant Neoplasms	165,191,197-199	171,173.0-173.4* 173.6-173.9 195-199	171,173*, 195-199	C44.-C47., C49.,C76.-C79., C80., C97.
36	Benign Neoplasm	210-239	210-239	210-239	D10.-D36.
37	Diabetes Mellitus	260	250	250	E10.-E14.
38	Cerebrovascular Disease	330-334	430-438	430-438	I60.-I69.
39	All Heart Disease	400-402,410-443	390-398,400.1 400.9,402,404 410-414,420-429	390-398,402,404 410-429	I00.-I02.,I05.-I09., I11., I13.-I14., I20.-I28.,I30.-I52.

Appendix A (continued)

40	Rheumatic	400-402,410-416	390-398	390-398	I00.-I02.,I05.-I09.
41	Ischemic	420,422.1	410-414	410-414	I20.-I25.
42	Chronic Disease of Endocardium and Other Myocardial Insufficiency	421,422.0,422.2	424,428	424,428*	I33.-I41.
43	Hypertension with Heart Disease	440-443	400.1,400.9,402 404	402,404*	I11.,I13.
44	All Other Heart Disease	430-434	420-423 425-427,429	415-417,420-423 425-427,429*	I26.0, I27.-I28., I30.-I32., I42.-I43., I44.-I52.
45	Hypertension w/o Heart Disease	444-447	400.0,400.2 400.3,401,403	401,403,405	I10, I12., I15.
46	Nonmalignant Respiratory Disease	241,470-527*	460-519*	460-519*	J00.-J99.
47	Influenza and Pneumonia	480-483,490-493	470-474,480-486	480-487	J10.-J18.
48	Bronchitis, Emphysema, Asthma	501,502,527.1,241*	490-493	490-493*	J40.-J46
49	Bronchitis	501,502	490,491	490,491	J40.-J42.,J44.
50	Emphysema	527.1	492	492	J43.
51	Asthma	241	493	493	J45.-J46.
52	Other Nonmalignant Respiratory	470-475,500,510-* 527.0,527.2	460-466* 500-519	460-466*,470-478, 494-496,500-519	J00.-J06.,J20.-J22., J30.-J39., J47., J60.-J70.,J80.-J86., J90.-J99.
53	Ulcer of Stomach and Duodenum	540,541	531-533	531-533	K25.-K27.
54	Cirrhosis of Liver	581	571	571	K70.,K74.
55	Nephritis and Nephrosis	590-594*	580-584	580-589	N00.-N29.
56	All External Causes of Death	800-999*	800-999	E800-999*	V01.-Y89.
57	Accidents	800-962	800-949	E800-949	V01.-X59.
58	Motor Vehicle Accidents	810-835	810-823	E810-825	V01.-V99.
59	All Other Accidents	800-802,840-962	800-809,824-949	E800-807,E826-949	W00.-X59.
60	Suicides	963,970-979	950-959	E950-959	X60.-X84.
61	Homicides and Other External	964,965,980-999*	960-978* 980-999	E960-978* E980-999	X85.-Y36., Y40.-Y89.

Appendix A (continued)

62	All Other Causes	020-138,206-207,* 240,242-254,270-326,340-398,450-468,530-539,542-580,582-587,600-795	000-009* 020-136 240-246 251-389 440-458 520-530 534-570 572-577 590-796	001-009* 020-139 240-246 251-389 440-459 520-530 534-570 572-579 590-799	A00.-A09., A20.-B19., B25.-B99., D00.-D09., D37.-D89., E00.-E07., E15.-G99., H00-H99, I70.-I99., K00.-K23., K28.-K67., K71.-K73., K75.-K93., L00.-L99., M00.-M99., N30.-R99.
63	Acquired Immunodeficiency Syndrome (AIDS) (from 1987)	not applicable	not applicable	042-044, 795.8*	B20.-B24.

*Comparability ratios (CR) unavailable at present time. Rates are unadjusted (CR=1.0) for these causes.
Rev: 02/2003

APPENDIX B: SAS Program used for the Analysis

```
data county_all;
  infile "e:\thesis\county\county_pop.txt";
  input geold sex race year agegrp cur_pop new_pop;
  diff=new_pop-cur_pop;
  if new_pop^=0 then diff_ratio=diff/new_pop;
  else diff_ratio=.;

*CREATE DUMMY VARIABLES FOR SEX;
  IF sex=2 then sex_dum=0;
    ELSE sex_dum=1;

*CREATE DUMMY VARIABLES FOR race;
  IF race=2 then race_dum=0;
    ELSE race_dum=1;

* CREATE DUMMY VARIABLES FOR AGEGRP;
ARRAY dummys {*} 3.  age1 - age18;
   DO i=1 TO 18;
    dummys(i) = 0;
  END;
  dummys(agegrp) = 1;
DROP i;

*CREATE DUMMY VARIABLES FOR YEAR;
  IF year=1990 then year1=1;
    ELSE year1=0;
  IF year=1991 then year2=1;
    ELSE year2=0;
  IF year=1992 then year3=1;
    ELSE year3=0;
  IF year=1993 then year4=1;
    ELSE year4=0;
  IF year=1994 then year5=1;
    ELSE year5=0;
  IF year=1995 then year6=1;
    ELSE year6=0;
  IF year=1996 then year7=1;
    ELSE year7=0;
  IF year=1997 then year8=1;
    ELSE year8=0;
  IF year=1998 then year9=1;
    ELSE year9=0;
  IF year=1999 then year10=1;
    ELSE year10=0;
  IF year=2000 then year11=1;
    ELSE year11=0;
run;
```

```
data county_all;
  set county_all;
  if year^=2000;
run;

proc contents data=county_all;
run;

proc tabulate data=county_all;
 class year sex race ;
 var diff_ratio;
 table year,sex ALL,mean*diff_ratio*(race ALL);
RUN;

proc tabulate data=county_all;
 class year agegrp sex  ;
 var diff_ratio;
 table year,agegrp ALL,mean*diff_ratio*(sex ALL);
RUN;

 proc tabulate data=county_all;
 class year agegrp race  ;
 var diff_ratio;
 table year,agegrp ALL,mean*diff_ratio*(race ALL);
RUN;

proc sort data=county_all;
  by year sex;
run;

proc means data=county_all noprint;
 by year sex;
 var new_pop cur_pop diff;
 output   out=total_sex   sum(new_pop   cur_pop   diff)=new_popsum   cur_popsum
diffsum;
run;

data total_sex;
  set total_sex;
  diff=new_popsum-cur_popsum;
  if new_popsum^=0 then diff_ratio=diff/new_popsum;
  else diff_ratio=.;
run;

proc sort data=county_all;
  by year race;
run;

proc means data=county_all noprint;
  by year race;
  var new_pop cur_pop diff;
  output  out=total_race  sum(new_pop  cur_pop  diff)=new_popsum  cur_popsum
diffsum;
run;

data total_race;
set total_race;
```

```
  diff=new_popsum-cur_popsum;
  if new_popsum^=0 then diff_ratio=diff/new_popsum;
  else diff_ratio=.;
run;

proc sort data=county_all;
  by year agegrp;
run;

proc means data=county_all noprint;
 by year agegrp;
 var new_pop cur_pop diff;
 output   out=total_age   sum(new_pop   cur_pop   diff)=new_popsum   cur_popsum
diffsum;
run;

data total_age;
  set total_age;
  diff=new_popsum-cur_popsum;
  if new_popsum^=0 then diff_ratio=diff/new_popsum;
  else diff_ratio=.;
run;


*SUMMARIZE THE DATA marginally;
proc sort data=county_all;
  by year geold sex;
run;

proc means data=county_all noprint;
 by year geold sex;
 var new_pop cur_pop diff;
 output   out=totalbysex   sum(new_pop   cur_pop   diff)=new_popsum   cur_popsum
diffsum;
run;

data totalbysex;
  set totalbysex;
  diff=new_popsum-cur_popsum;
  if new_popsum^=0 then diff_ratio=diff/new_popsum;
  else diff_ratio=.;
run;

proc sort data=county_all;
  by year geold race;
run;

proc means data=county_all noprint;
 by year geold race;
 var new_pop cur_pop diff;
 output   out=totalbyrace   sum(new_pop   cur_pop   diff)=new_popsum   cur_popsum
diffsum;
run;

data totalbyrace;
  set totalbyrace;
  diff=new_popsum-cur_popsum;
```

```
  if new_popsum^=0 then diff_ratio=diff/new_popsum;
  else diff_ratio=.;
run;


proc sort data=county_all;
  by year geold agegrp;
run;

proc means data=county_all noprint;
 by year geold agegrp;
 var new_pop cur_pop diff;
 output  out=totalbyagegrp  sum(new_pop  cur_pop  diff)=new_popsum  cur_popsum
diffsum;
run;


* MATCHED T-TEST FOR DATA IN YEAR 90;
proc ttest data=year00;
   paired cur_pop*new_pop;
run;

proc sort data=county_all;
  by sex year;
run;

proc univariate data=county_all;
  var diff_ratio;
  by sex year;
run;

proc sort data=county_all;
  by race year;
run;

proc univariate data=county_all;
  var diff_ratio;
  by race year;
run;

proc sort data=county_all;
  by agegrp year;
run;

proc univariate data=county_all;
  var diff_rate;
  by agegrp year;
run;

proc sort data=county_all;
  by year;
run;

proc means data=county_all std min max range stderr cv;
  var diff_ratio;
  by year;
run;
```

```
proc ttest data=county_all;
  paired cur_pop*new_pop;
run;

proc ttest data=county_all;
   class sex;
   var diff_ratio;
run;

proc sort data=year90;
 by sex agegrp;
run;

proc univariate data=year90;
  var diff_ratio;
  by sex agegrp;
run;

proc sort data=year90;
  by sex race agegrp;
run;

proc univariate data=year90;
  var diff_ratio;
  by race agegrp;
run;

*PAIRED TTEST IN YEAR 1990 WITHIN EACH SEX;
data year90_m;
   set year90;
   if sex=1;
run;

data year90_f;
   set year90;
   if sex=2;
run;

proc ttest data=year90_m;
   paired cur_pop*new_pop;
run;

proc ttest data=year90_f;
  paired cur_pop*new_pop;
run;

*PAIRED TTEST IN YEAR 1991 WITHIN EACH SEX;
data year91_m;
   set year91;
   if sex=1;
run;

data year91_f;
   set year91;
   if sex=2;
run;
```

```
proc ttest data=year91_m;
   paired cur_pop*new_pop;
run;

proc ttest data=year91_f;
  paired cur_pop*new_pop;
run;

*PAIRED TTEST IN YEAR 2000 WITHIN EACH SEX;
data year00_m;
   set year00;
   if sex=1;
run;

data year00_f;
   set year00;
   if sex=2;
run;

proc ttest data=year00_m;
   paired cur_pop*new_pop;
run;

proc ttest data=year00_f;
  paired cur_pop*new_pop;
run;

* REGRESSION;
proc reg data=county_all;
  model diff=year1-year9 age2-age18 sex_dum race_dum new_pop;
run;

PROC GLM DATA=COUNTY_ALL;
   CLASS YEAR SEX RACE AGEGRP geold;
   MODEL DIFF_RATIO=YEAR SEX RACE AGEGRP geold;
RUN;

PROC GLM DATA=COUNTY_ALL;
   CLASS YEAR SEX RACE AGEGRP geold;
   MODEL  DIFF_RATIO=YEAR  SEX  RACE  geold  sex*race  AGEGRP  sex*agegrp
race*agegrp;
RUN;
```

BIBLIOGRAPHY

Anderson RN, Arias E (2003), The effect of revised population on morality statistics for the United States, 2000. National Vital Statistical Reports, 51(9). Hyattsville, Maryland: National Center for Health Statistics.

Checkoway H, Pearce N and Crawford-Brown DJ, Research methods in occupational epidemiology. New York, Oxford University Press, Inc., 1989.

Delwiche, LD and Slaughter, SJ , Little SAS Book. Second edition, SAS Institute Inc. ,Cary, NC, USA, 1998

Everitt,BS, A handbook of statistical analysis using S-PLUS, 2nd ed. CRC press LLC, 2002

Gordis L, Epidemiology, 2nd ed. W.B. Saunders Company 2000

Mortality and Population Data System User Manual, Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, Sept. 1984

Marsh GM, Youk AO, Stone RA, Sefcik S, Alcorn C (1998). OCMAP-PLUS: A program for the comprehensive analysis of occupational cohort data. JOEM 40(4): 351-362

http://www.census.gov

http://www3.ccps.virginia.edu/demographics/Concepts/estimatesmethods01-09.html

http://docs.lib.duke.edu/federal/guides/us_demographic.html