R Project

**Introduction**

In the following pages we will describe our findings based on gathered statistical data from IPUMS USA, regarding household incomes from 2000 to 2014.

We analyzed data considering both gender and race. In particular, we divided the race in Caucasian, African American, Hispanic and Asian (Chinese and Japanese) and studied males and females respectively.

In particular, our thesis is based on proving the existence of some fundamental differences between both gender and race caused by the financial crisis of 2008.

Finally, we tried to estimate a possible trend of recovery for each demographic, considering the future growth rate.

**Sample and Variables**

In order to have a most impartial and objective result as possible, we made few adjustments at our sample data and variables. First of all, every value is expressed in terms of purchasing power in 2015. Since the data are available up to 2014, we wanted to compare every measurement in the same way. Moreover, we decided to exclude from our sample every household income which exceeded 2 million dollars, defining those as outliers. This is due to the fact that the sample was constructed by assigning to not available data a value of 9999999, which it would be kept real values far away from the average.

Furthermore, since household income could be formed by more than a single individual, for instance, children can potentially contribute, we decided to standardize our sample and consider only the data of couples without children. By doing this, we wanted to create a more representative and equal income base.

**Hypothesis**

In gathering our data, we decided to tests to main hypothesis:

1. Is it really true that females’ household income is systematically lower than males’ one?
2. How significant is the difference between income of white and black people?

For both questions we developed two null hypothesis (Ho) and their respective alternative hypothesis.

1. A) Ho: Average Male Income = Average Female Income

B) Ha: Average Male Income > Average Female Income

1. A) Ho: Average White Income = Average Black Income

B) Ha: Average White Income ≠ Average Black Income

It is important to outline to facts: it is a reasonable assumption to treat these variables as independent for our purposes and in the first case we will conduct a one-sided test, whereas in the second we will consider both possibilities, either higher or lower.

Our test statistics in both cases will be represented by the difference between the two variables. Furthermore, since we are considering only a difference between averages across 15 years, our sample size will be not large enough to take advantage of the central limit theorem. For this reason, we are going to use a t-distribution.

**Findings:**

Graphs:

Here is a visual representation of inflation adjusted household income between 2000 and 2014, considering separately race and gender.

*Gender*

*Race:*

It is important to note that regardless any intrinsic difference between the variables, the sample data as a whole outlines the great impact of the financial crisis in household income. This is impressive since the difference is really demarked even considering only one of the multiple variables which were influenced by this sudden event.

Summary Statistics: