Biostat 212 Final Project

Hubert Jenq

3/22/2017

OECD Data set

The organization for Economic Co-operation and Development (OECD) is a 35-member group that advocates policies that will help the economic and social climate of people around the world. As a result, the growth of gdp per capita (growth), per capita GDP (initgdp), population growth (popgro), average investment to GDP ratio (inv), and average secondary school enrollment rate (humancap) are of interest to compare between OECD countries and those not in OECD. The measurements in the dataset are for each 5-year period. In this analysis, we will not account for the temporal effects and assume that each measurement is a independent sample.

To compare the difference in the above parameters between the OECD and non-OECD countries, a wilcoxon-rank sum test and ansari-bradley test will be used to detect differences in their median, and their dispersion. The null hypothesis for the wilcoxon-rank sum test is that the medians are equal. The null hypothesis for the Ansari-Bradley test is that they have the same dispersion parameter. The p-values for the tests between OECD countries and non-OECD countries are in table 1.

Table 1: Wilcoxon Rank Sum test between OECD and non-OECD countries

	P-value	Lower 95% CI value	Upper 95% CI value
growth	0.0001508	-0.0137754	-0.0044850
initgdp	0.0000000	-1.8080934	-1.5462678
popgro	0.0000000	0.2355169	0.2737499
inv	0.0000000	-0.7185427	-0.5650619
humancap	0.0000000	-0.9238556	-0.7241845

Table 2: Ansari-Bradley test between OECD and non-OECD countries

	P-value
growth	0e+00
initgdp	0e+00
popgro	0e+00
inv	9e-07
humancap	0e+00

Here we see that for all of the variables of interest they are significantly different in their disperson and mean between the OECD and non-OECD countries. Correcting for multiple hypotheses will still yield significant p-values for all of the tests performed. From this, we can conclude that there are significant differneces in population growth, GDP per capita, GDP growth, investment to gdp ratio, and secondary school enrollment between OECD and non-OECD countries.

The 95% confidence intervals can be examined for direction. We see that the OECD countries have significantly higher GDP growth, initial GDP, investment per GDP, and secondary education. Non-OECD countries have higher population growth rates.

To further visualize the difference, a kernel density estimate of popgro is constructed below in figure 1 for OCED countries and non-OECD countries. Here we see the population growth for OECD members is shifted to the right, showing that there are higher average annual population growth rates for OECD members compared to non-OECD members.

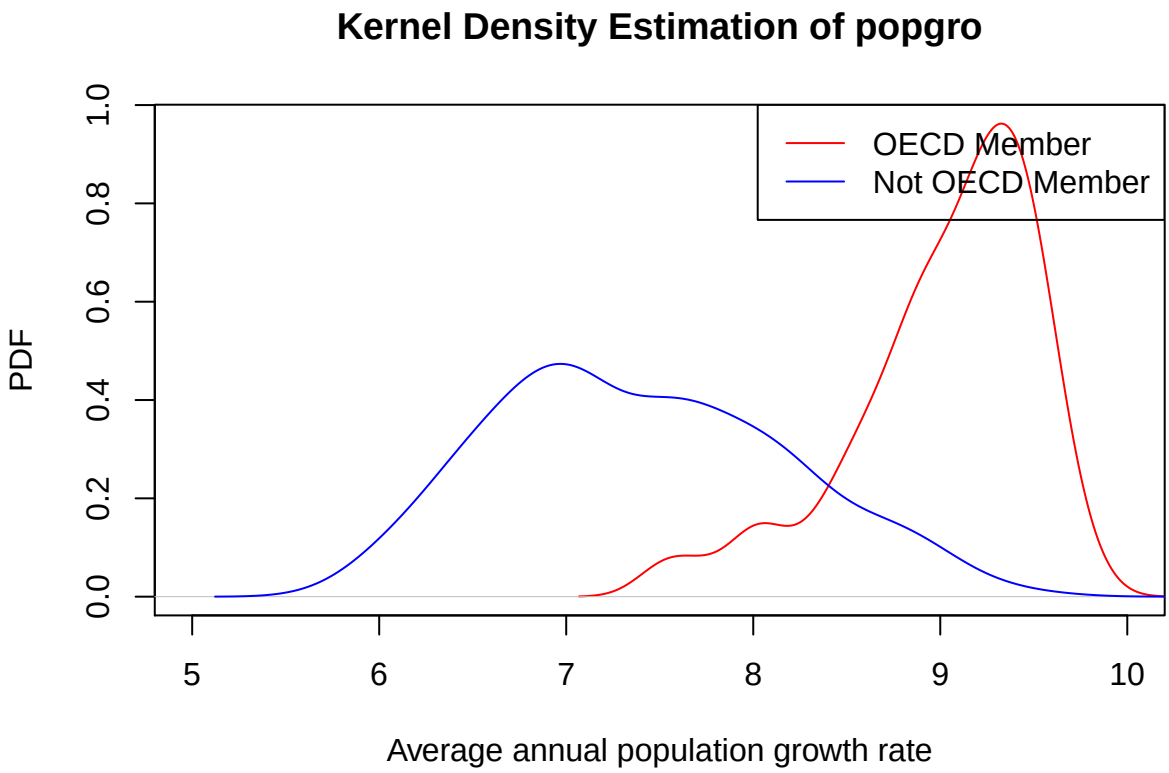
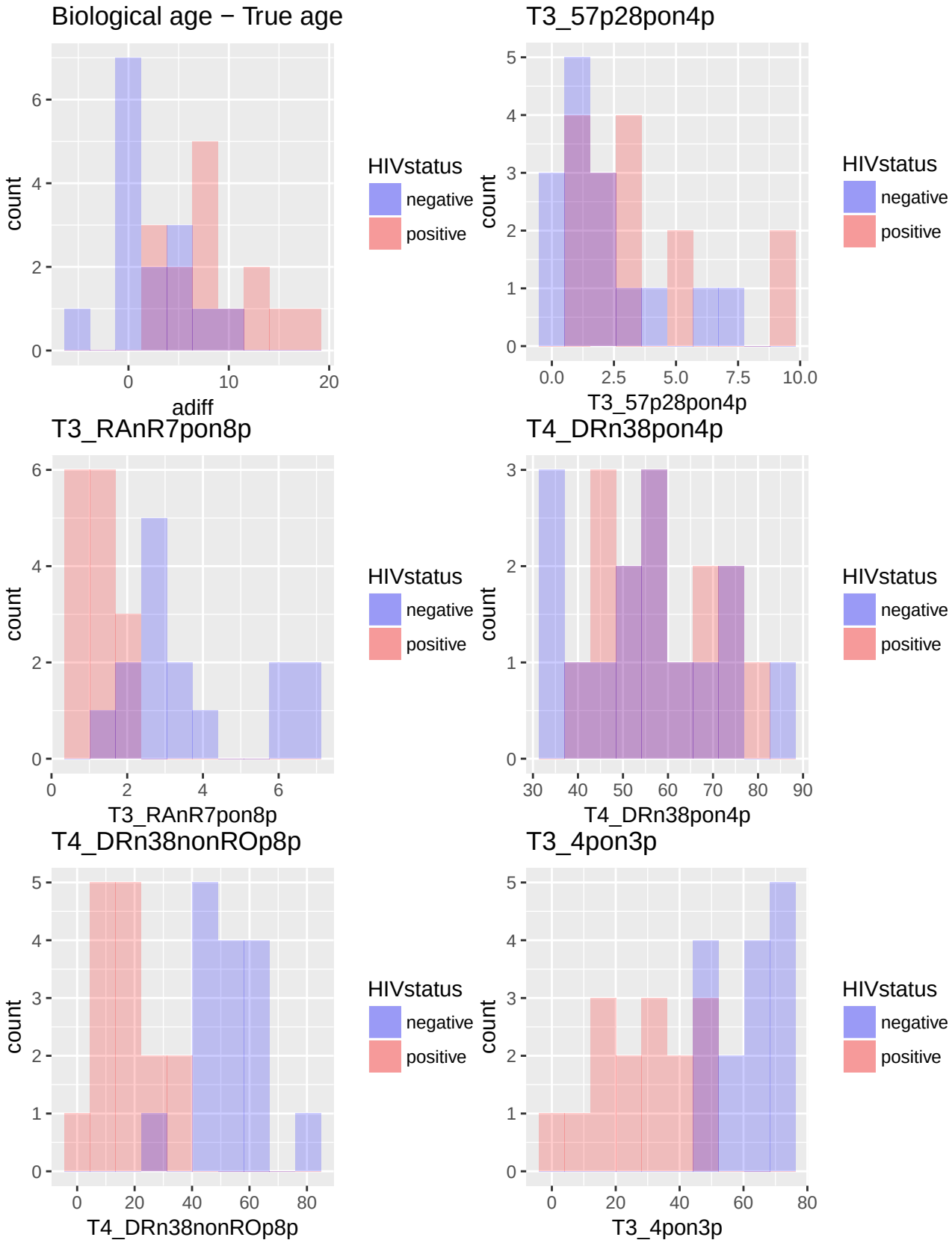


Figure 1: Kernel Density Estimate

HIV Dataset

Data

In this dataset of 60 observations and 30 covariates, 5 covariates are selected randomly to analyze. T3_57p28pon4p, T3_RAnR7pon8p, T4_DRn38pon4p, T4_DRn38nonROp8p, and T3_4pon3p are the covariates of interest. Histograms of these 5 covariates in the HIV positive and HIV negative group are constructed. The data for the first visit is only used so the treatment effect is not present. From the histograms below, it seems that biological age-true age, T3_4pon3p, T4_DRn38nonROp8p, and T3_RAnR7pon8p have distinct distributions for the HIV positive and HIV negative groups. To further examine this, a wilcoxon rank sum test and ansari-bradley test will be performed to see whether there is a difference in the median or dispersion in the HIV positive vs. negative group.


Biological age − True age
count
adiff
HIVstatus
negative
positive
T3_57p28pon4p
T3_57p28pon4p
T3_RAnR7pon8p
T3_RAnR7pon8p
T4_DRn38pon4p
T4_DRn38pon4p
T4_DRn38nonROp8p
T4_DRn38nonROp8p
T3_4pon3p
T3_4pon3p

The p-values and confidence intervals for the WRST and AB test can be found in table 3 and 4. Here we see that the median shift is significant for biological age-true age, T3_4pon3p, T4_DRn38nonROp8p, and T3_RAnR7pon8p. The above covariates would still be significant with bonferroni correction for multiple hypotheses. The confidence interval for age difference is 3.04 to 8.66 which can be interpreted as there is a 95% chance the true median shift is captured by this interval. A positive median shift indicates that the HIV positive patients have a larger difference between their biological and actual age which points towards accelerated aging. The other three significant covariates have a negative effect, indicating that the HIV negative group has a larger value than the HIV positive group. The Ansari-Bradley tests came up all non-significant indicating that there is no dispersion difference between the covariates of the HIV positive and HIV negative groups.

Table 3: Wilcoxon Rank sum test for HIV effect

	WRST p-val for HIV effect	Lower 95% CI	Upper 95% CI
agediff	0.000	3.047	8.657
T3_57p28pon4p	0.152	-0.500	2.200
T3_RAnR7pon8p	0.000	-3.100	-1.200
T4_DRn38pon4p	0.481	-7.300	14.700
T4_DRn38nonROp8p	0.000	-46.100	-27.900
T3_4pon3p	0.000	-45.700	-26.100

Table 4: Ansari-Bradley test for HIV effect

	Ansari-Bradley P-value for HIV effect
agediff	0.838
T3_57p28pon4p	0.542
T3_RAnR7pon8p	0.901
T4_DRn38pon4p	0.571
T4_DRn38nonROp8p	0.901
T3_4pon3p	1.000

The treatment effect for the HIV positive patients can be examined by comparing the paired data of the 1st and 3rd visit since the treatment is administered after the first visit. Wilcoxon signed rank tests are performed on the same 5 covariates and difference in biological and true age and the values are shown in table 5. Here we see that T3_4pon3p, T4_DRn38nonROp8p, and T3_RAnR7pon8p have significantly different medians for the treated vs. the untreated. By treating the patients, the values of the above covariates decrease. This is consistent with HIV positive/negative patients where the HIV positive patients had higher values of the same three covariates.

Table 5: Wilcoxon signed rank test for treatment effect

	WSRT p-val for treatment effect	Lower 95% CI	Upper 95% CI
agediff	0.135	-0.347	3.622
T3_57p28pon4p	0.268	-0.300	1.550
T3_RAnR7pon8p	0.007	-1.850	-0.500
T4_DRn38pon4p	0.561	-9.800	4.150
T4_DRn38nonROp8p	0.003	-21.300	-9.050
T3_4pon3p	0.003	-14.000	-3.150

To further explore the age difference in the HIV positive and negative group, a kernel density estimation is

constructed and graphed in figure 2. Here we see that the HIV positive groups PDF curve is shifted to the right indicating that the age difference is larger. A larger age difference corresponds to accelerated aging which is consistent with our wilcoxon rank sum tests.

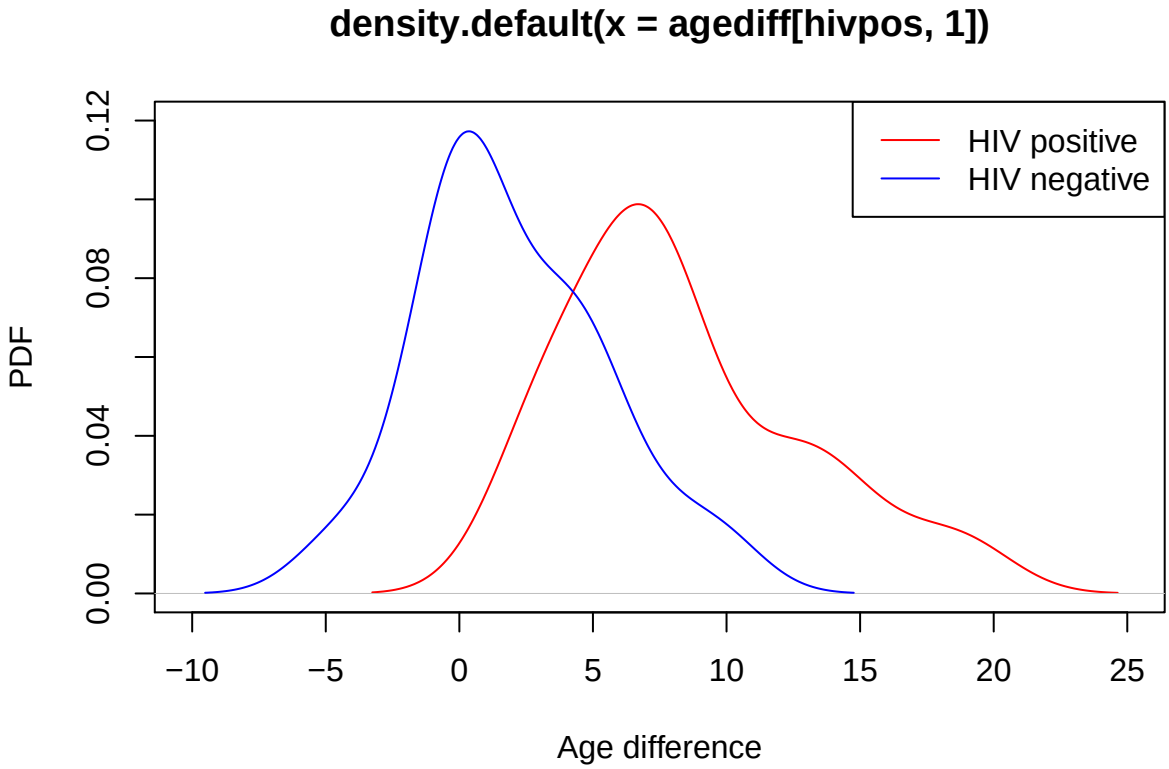
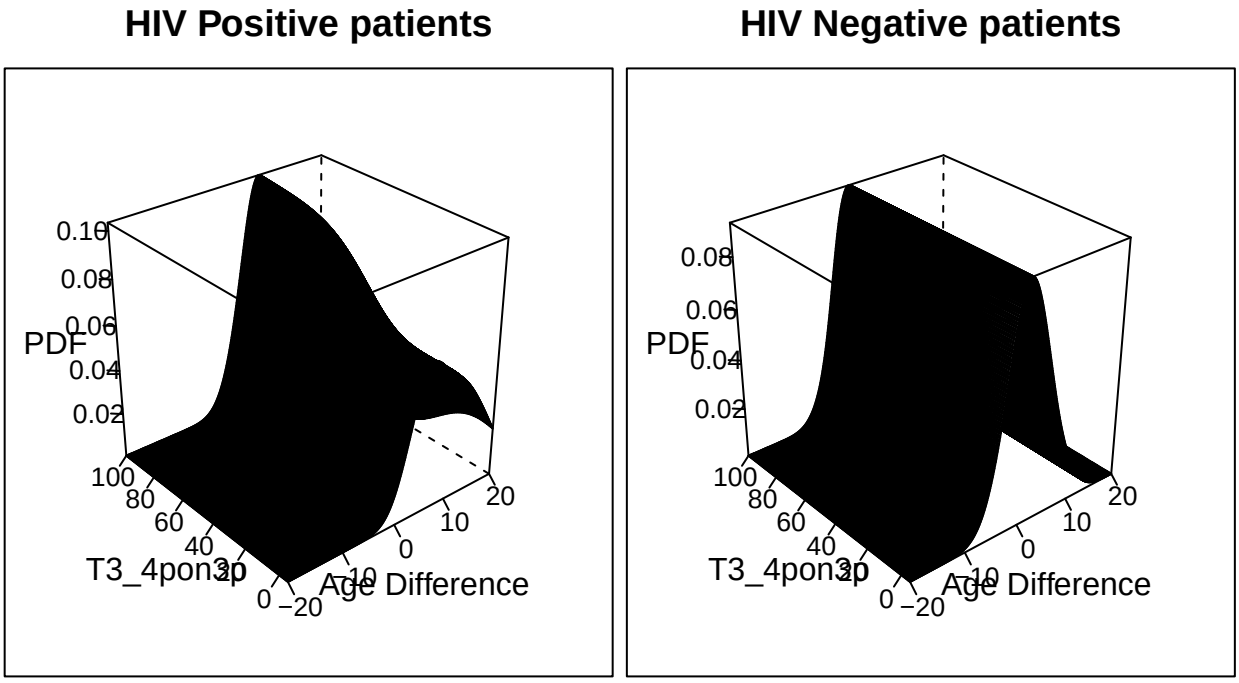


Figure 2: Kernel Density Estimation of Age difference

The two dimensional kernel density estimate of T3_4pon3p and age difference is constructed for the HIV positive and negative groups since T3_4pon3p was shown to be significantly different. Here we see that the bivariate PDF are different between the HIV positive and negative groups. There seems to be no interaction between T3_4pon3p and age difference in the HIV negative group since the PDF of age difference is the despite the value of T3_4pon3p. However, in the HIV positive group T3_4pon3p does affect the PDF of the age difference.



From the above analyses, we see three covariates T3_4pon3p, T4_DRn38nonROp8p, and T3_RAnR7pon8p are significantly different in the HIV positive and negative groups. Additionally, the age difference is significant in the HIV positive and negative groups. Treatment within the HIV positive group will also significantly change the medians of the above three covariates. Finally, we see that there is interaction between T3_4pon3p and age difference from the two-dimensional kernel density estimate that was constructed.