**WEEK 4 REPORT – DESCRIPTIVE STATISTICS**

Group Number:   
Date and Time of Tutorial Session:   
Group Members (First and Last name – no student numbers):

1. Asin, Hanna
2. Fraser, Katarzyna
3. Pustil, Ella
4. Tran, Annika

Graded out of 10 marks.

1. Calculate descriptive statistics: mean, median, and standard deviation, for each combination of time-period and quintile of the fertility dataset and fill in the table below (rounded to two decimal places).

(6 marks, 0.2 per number)

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1990-2005 | | |
| Socioeconomic status | Mean (births/individual) | Median  (births/individual) | Standard deviation (s.d.) |
| Quintile 1 (Poorest) | 6.20 | 6.78 | 1.57 |
| Quintile 2 | 5.21 | 5.49 | 1.45 |
| Quintile 3 | 4.72 | 4.65 | 1.70 |
| Quintile 4 | 4.26 | 4.05 | 1.80 |
| Quintile 5 (Richest) | 2.95 | 2.77 | 1.08 |
|  | 2006-2019 | | |
|  | Mean (births/individual) | Median  (births/individual) | Standard deviation (s.d.) |
| Quintile 1 (Poorest) | 5.54 | 5.73 | 1.82 |
| Quintile 2 | 4.43 | 4.02 | 1.81 |
| Quintile 3 | 4.42 | 3.97 | 1.67 |
| Quintile 4 | 3.84 | 3.69 | 1.67 |
| Quintile 5 (Richest) | 2.56 | 2.45 | 0.94 |

1. What factors are important to consider when choosing between the mean or median when calculating a descriptive measure of central tendency? (1 mark, 0.25 for the criteria, 0.25 for the reason)

Its important to consider that the mean and median are different. Mean is all the values added up and divided by the number of values, therefore it is the average. While the median is the middle value of all the variables when listed in order. You would want to use the mean if there are no outliers in the dataset so the mean would be most accurate. If there were outliners, you would want to use median because then the outliers would not drag/affect the value we would want to use to characterize that data set. Mean would be affected by any outliers while the median would not. The mean would be better for when characterizing datasets is when the sample size is smaller compared to median.

1. Now calculate the *effect size* of the **means** across years **(**i.e., the absolute value of the differences between the means from 1990-2005 and 2006-2019). Round to two decimal places. (1 mark, 0.2 per number)

|  |  |
| --- | --- |
| Subgroup | Effect size (births per individual) |
| Quintile 1 (Poorest) | 0.67 |
| Quintile 2 | 0.78 |
| Quintile 3 | 0.30 |
| Quintile 4 | 0.43 |
| Quintile 5 (Richest) | 0.39 |

1. As a group, discuss the changes you see in both the mean fertility rates (in Question 1 above) and in absolute effect sizes between the two time periods (in Question 3 above), and how this varies across the household wealth quintiles.  
   1. Describe the trend(s) you see in the data, across the time periods, and across quintiles in the table for question 1. (1 mark)

The trend we see is that the period time from 1990-2005 has a higher mean than period 2006-2019. Also, the mean difference of the quintiles between the two time are biggest when its it poorest, so quintile 1 has the biggest mean difference between period 1990-2005 and period 2006-2019 and as the quintile goes up, the difference gets less and less small. Quintile 5, (richest), has the smallest difference of mean between the two time periods.

* 1. Which wealth quintile had the smallest variation (standard deviation from Question 1)? What does this indicate about fertility rates in this group? (1 mark)

For both time periods, it was quintile 5 (the richest) meaning that the fertility rates are consistently close to the mean of each person in that quintile having around the same number of children.

**Notes:**

* **Only one group member submits the report**
* **The report must be a Word .DOC, .DOCX or .PDF file**
* **Make sure everyone in the group has a copy of the report**
* **Double check what you have submitted‼**
  + **view it on OnQ to make sure everything is there and visible**
* **Lastly, everyone in the group needs to submit their own version of their R script.**