Risk of Heart Disease and Attack based on Health, Wealth and Education.

Andriani Christanty, Johanna Fernandez & Tristan Carlisle
GROUP 1 - PROJECT 1 UWA Data Analytics Boot Camp

Motivation

- Although on the decline, Cardiovascular conditions still account for the largest proportion of deaths in Australia (27%) (DoH, 2021; Leeder, Gibberd, Dobson, Lloyd, & medicine, 1984)

- More than 1 million Australians live with cardiovascular conditions such as heart disease, stroke or vascular conditions (DoH, 2021).

- As these conditions account for such a large proportion of mortality, our group seek to assess what factors may be intertwined with thearts disease and heart attack. Thus, potentially develop mitigation strategies that can be implemented to reduce risk factors and ultimately save lives.

Limitations in finding Medical AU Datasets

Some data collections cannot be made publicly available. Access to others needs special approval, which in some cases must be given by the data providers. In some instances, access is available through tables prepared by the AIHW for specific requests, (for which charges may apply).

Researcher access to these collections may be restricted due to a range of factors including but not limited to:

- conditions imposed by the AIHW Ethics Committee with respect to use of the collection
- agreements with the suppliers of data to the AIHW
- practicalities in providing access to older collections
- data quality limitations.

(Australian Institute of Health and Welfare, 2021)

BRFSS 2015 Survey CSV

- Found on Kaggle
- Land-Line and Cell-Phone Survey Result
- Included data for 50 American states, the District of Columbia Guam & Puerto Rico.
- Originally 441, 456 records down to **253,680 records** after the clean-up

(CDC & BRFSS, 2016)

Hypothesis

Overarching Hypothesis

Lifestyle and health metrics influence a person's risk of developing heart attack/heart disease.

Refined Hypotheses

- *Ho: The level of income in a household does not impact a person's risk to having a heart disease/ heart attack*
- Ha: As the person's income goes up, it reduces the risk or exposure to having a heart disease / heart attack

- *Ho: There is no relationship between education level , mental/physical health and heart attack/ disease occurrence.*
- Ha: Education level, mental/ Physical health influence the likelihood of Heart attack/ disease occurrence.

Metrics as provided by BRFSS 2015 Survey

High Blood Pressure

High Cholesterol

Body Mass Index (BMI)

History of Stroke

Diabetes

Physical Activity

Daily Consumption of Fruits

Daily Consumption of Vegetables

Income

Alcohol Consumption

Access to Healthcare

General Health

Mental Health

Physical Health

Difficulty Walking

Sex

Age

Education

Cigarette Smoking

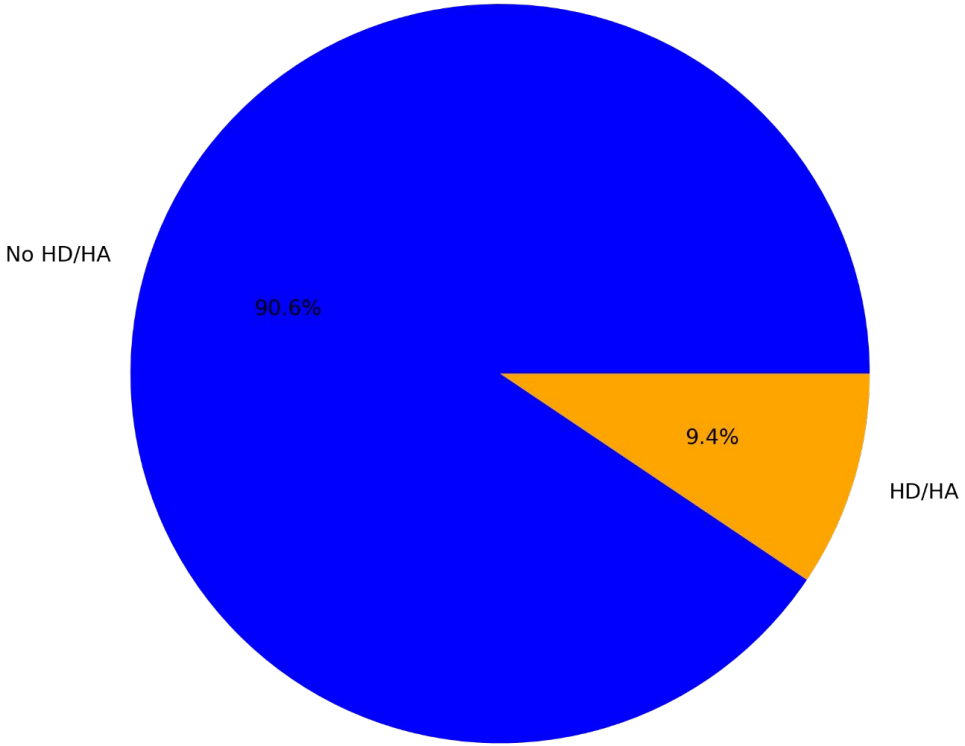
Data Cleanup & Exploration

We went through each of the column in the CSV that we obtained, and made sure the values in each column were within the expected values. For this, we referred back to the CDC Code Guideline book (BRFSS Codebook, 2015).

As the CSV was already modified on Kaggle, some of the headers and the values in the columns did not correspond to those of the Codebook. We explored further the author's website and were able to find some notes about these changes.

Data Analysis and Discussion: Health metrics

Prevalence of HD/HA in Sampled American Population - BRFSS 2015 Survey

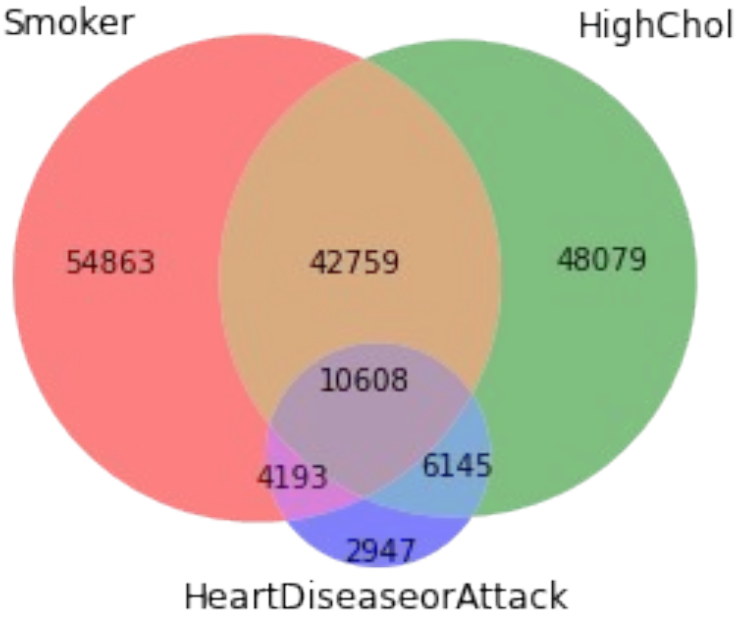


Cigarette Smoking, High Cholesterol & HD/HA

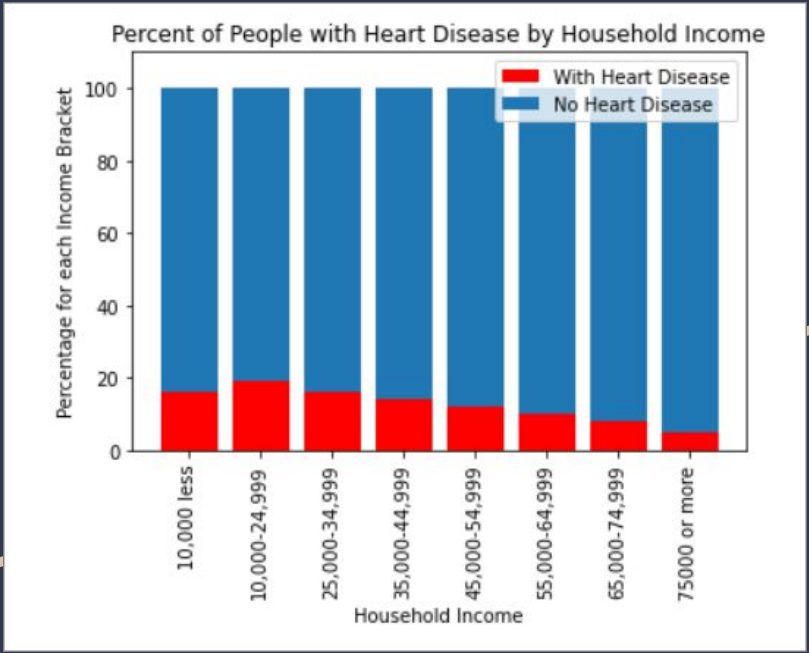
How common are overlaps among sampled population with HD/HA, smoking, high Cholesterol?

We are interested in seeing whether individuals who have HD/HA are more likely to have two risk factors

Overlaps can be seen among the three, with - Most people with HD/HA (44%) have 2 risk factors (smoking cigarettes & High Cholesterol), while 17% are smokers without high cholesterol, and 26% are people with high cholesterol who are not smokers

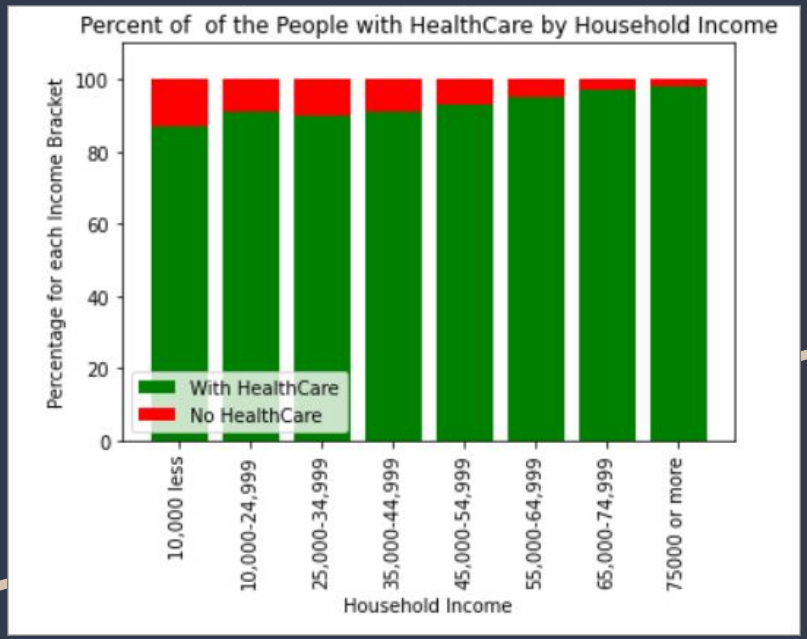


Analysis & Discussion: Income



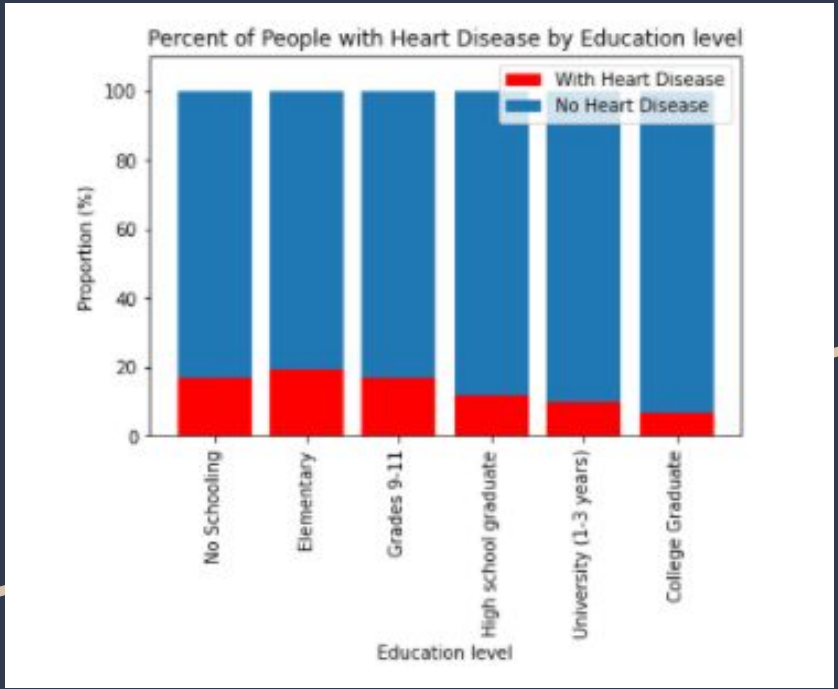
- Is there a relationship between income and HD/HA?
- From $10,000 onwards we see a steady decrease in the proportion of people with HD/HA over the total population for each income bracket
- This could have an interaction with accessibility to medical care and other fitness and lifestyle facilities.

Analysis & Discussion: Income & HealthCare Access



- Does the person's income impact their access to Health Care
- We see an increase to the percentage of people having access to HealthCare as their income increases

Analysis & Discussion: Education



Education and heart disease/attack

Ho: Education and heart disease/attack are independent from each other.

Ha: Education and heart disease/attack are dependent on one another.

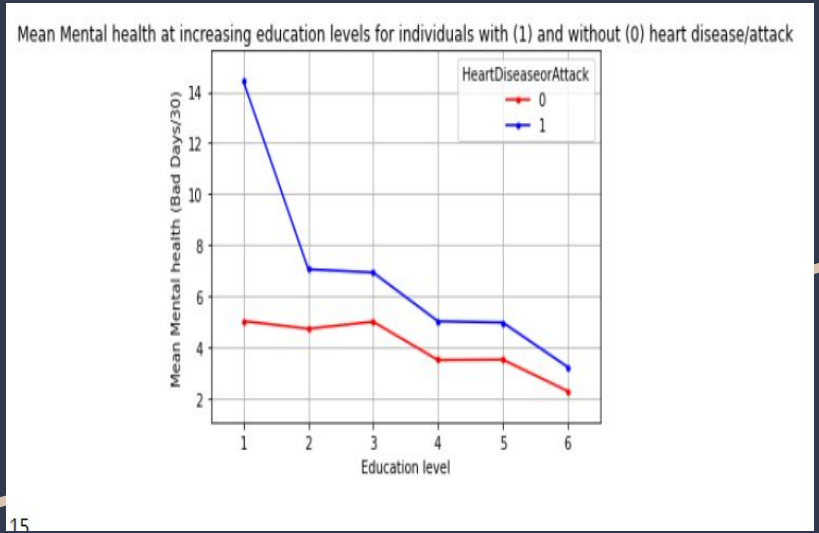
- A Chi-squared test was completed as we have categorical data
- P-values < 0.05 thus, dependency established

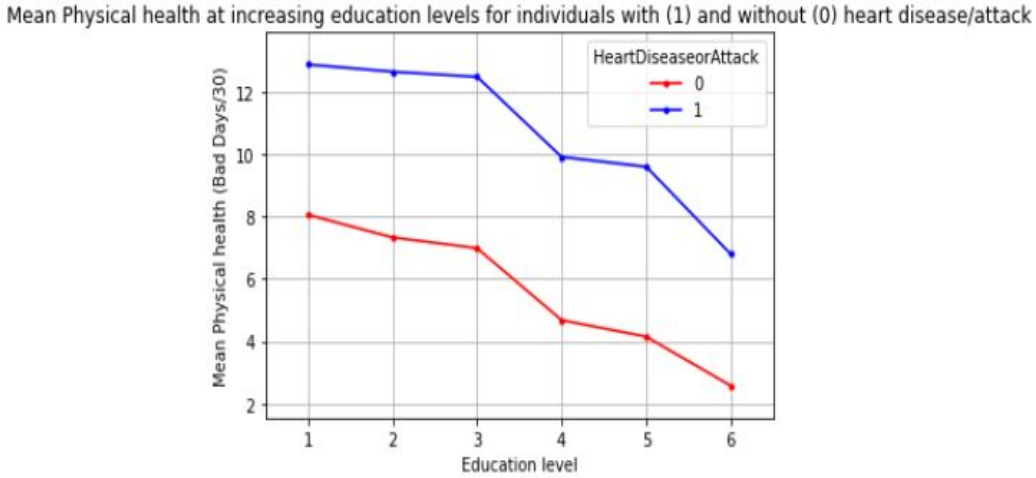
```
chisqt = pd.crosstab(original_df["Education"],original_df["HeartDiseaseorAttack"], margins=True)
value = np.array([chisqt.iloc[0][0:].values,
                  chisqt.iloc[1][0:].values,
                  chisqt.iloc[2][0:].values,
                  chisqt.iloc[3][0:].values,
                  chisqt.iloc[4][0:].values,
                  chisqt.iloc[5][0:].values])

print(chi2_contingency(value)[0:3])
```

Analysis & Discussion

- The interplay between Education and Health
  - Two way ANOVA with interaction: P-value (<0.01)
  - Tukey post hoc
    - tukey_hsd
  - Shapiro wilks and levenes (normality and homogeneity of variance)





Post Mortem

Post Mortem

Challenges:

- Matplotlib-Venn not in the original Matplotlib package
- We had a mixture of count and continuous data integrating them in order for statistical tests was a challenge.

Further Research:

- Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks?

- It may be useful if we are able to find more BRFSS Survey datasets from subsequent years so that we can analyze the trend

- Developing more charts and diagrams composed of various different metrics may provide us with valuable insights

- Assessing how various categories of education may impact health metrics and HA/HD occurrence

- Linking access to Health care, and fitness facilities with income to assess how income influences HA/HD occurrence
- Looking into accessing Australian datasets

Summary & Implications

1. Most people with HD/HA (44%) have 2 risk factors (smoking cigarettes & High Cholesterol), while 17% are smokers without high cholesterol, and 26% are people with high cholesterol who are not smokers

   Can explore developing a program alerting HCP when their patients possess the two risk factors, so that a plan aimed at eliminating or reducing at least one of the two metrics can be initiated

2. Yearly wage influences HA/HD occurrence. Although this may not be a direct influence the trend is present. A majority of the population has a health care plan however, we see a slight increase in proportion as income increases.

   Raising the minimum wage may reduce the occurrence of HA/HD. Furthermore, introduction of more comprehensive health care plans may be adequate. However, further research is required to assess which factors should be addressed in this plan.

3. Education and HD/HA are dependent variables ($p < 0.01$). As education level increases we see a drop in HD/HA. Furthermore, as education levels increased the number of bad days in last 30 days dropped. Education may give individuals the ability to cope with hardships and manage physical health.

   We recommend government intervention to increase accessibility to higher education. Addition to Education budget could reduce cost associated with medical facilities to house patients with HA/HD as well as facilitate economic growth.


Q&A

References

(DoH), A. G. D. o. H. (2021). What we're doing about cardiovascular conditions. Retrieved from https://www.health.gov.au/health-topics/chronic-conditions/what-were-doing-about-chronic-conditions/what-were-doing-about-cardiovascular-conditions

Australian Institute of Health and Welfare. (2021, September 16). *Reports & data*. AIHW. Retrieved February 6, 2022, from https://www.aihw.gov.au/reports-data

CDC & BRFSS. (2016). *Behavioral Risk Factor Surveillance System 2015 Codebook Report Land-Line and Cell-Phone data* [E-book].

Leeder, S. R., Gibberd, R., Dobson, A., Lloyd, D. J. A., & medicine, N. Z. j. o. (1984). Declining mortality rates from ischemic heart disease in Australia. *14*(4), 388-394.

Teboul, Alex. "Heart Disease Health Indicators Dataset Notebook." *Kaggle*, Nov. 2021, www.kaggle.com/alexteboul/heart-disease-health-indicators-dataset-notebook. Accessed 5 Feb. 2022.

Appendix

Physical health

ANOVA

	PR(>F)
C(Education)	0.000000e+00
C(HeartDiseaseorAttack)	0.000000e+00
C(Education):C(HeartDiseaseorAttack)	1.269650e-16
Residual	NaN

TUKEY (education)

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	4	6	2.445824	2.324361	2.567287	81.151009	0.001
1	4	3	2.619076	2.352716	2.885436	39.627078	0.001
2	4	5	0.613247	0.480332	0.746162	18.594027	0.001
3	4	2	3.050079	2.657897	3.442262	31.342667	0.001
4	4	1	3.553634	1.718754	5.388515	7.805085	0.001
5	6	3	5.064900	4.805901	5.323900	78.810533	0.001
6	6	5	1.832578	1.715105	1.950050	62.869395	0.001
7	6	2	5.495904	5.108683	5.883125	57.199574	0.001
8	6	1	5.999459	4.165632	7.833286	13.184586	0.001
9	3	5	3.232323	2.967759	3.496887	49.237597	0.001
12	5	2	3.663326	3.272362	4.054291	37.761634	0.001
13	5	1	4.166881	2.332261	6.001502	9.153297	0.001

1 non significant

TUKEY (HA/HD)

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	0	1	5.42314	5.31014	5.53614	133.026234	0.001

TUKEY (interaction)

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(0, 4)	(0, 6)	2.102125	1.955279	2.248971	66.162300	0.001000
1	(0, 4)	(0, 3)	2.303645	1.969498	2.637792	31.863362	0.001000
2	(0, 4)	(0, 5)	0.530280	0.368736	0.691824	15.171516	0.001000
3	(0, 4)	(0, 2)	2.650466	2.151236	3.149697	24.537776	0.001000
4	(0, 4)	(0, 1)	3.373168	1.068188	5.678148	6.763706	0.001000
5	(0, 4)	(1, 4)	5.225655	4.883896	5.567415	70.669876	0.001000
6	(0, 4)	(1, 6)	2.112270	1.762444	2.462096	27.906894	0.001000
7	(0, 4)	(1, 3)	7.793990	7.094861	8.493118	51.524821	0.001000
8	(0, 4)	(1, 5)	4.915236	4.561731	5.268740	64.263345	0.001000
9	(0, 4)	(1, 2)	7.958568	6.957814	8.959322	36.755444	0.001000
10	(0, 4)	(1, 1)	8.186961	3.038265	13.335657	7.349189	0.001000
11	(0, 6)	(0, 3)	4.405770	4.081084	4.730455	62.715174	0.001000
12	(0, 6)	(0, 5)	1.571845	1.430909	1.712780	51.547001	0.001000
13	(0, 6)	(0, 2)	4.752591	4.259643	5.245538	44.559836	0.001000
14	(0, 6)	(0, 1)	5.475292	3.171665	7.778920	10.985226	0.001000
15	(0, 6)	(1, 4)	7.327780	6.995266	7.660294	101.853492	0.001000
16	(0, 6)	(1, 6)	4.214395	3.873595	4.555195	57.154420	0.001000
17	(0, 6)	(1, 3)	9.896114	9.201458	10.590771	65.842839	0.001000
18	(0, 6)	(1, 5)	7.017361	6.672786	7.361935	94.124818	0.001000
19	(0, 6)	(1, 2)	10.060693	9.063058	11.058327	46.609069	0.001000
20	(0, 6)	(1, 1)	10.289086	5.140995	15.437176	9.237289	0.001000
21	(0, 3)	(0, 5)	2.833925	2.502333	3.165518	39.500070	0.001000
24	(0, 3)	(1, 4)	2.922010	2.474063	3.369957	30.148732	0.001000
26	(0, 3)	(1, 3)	5.490344	4.733616	6.247073	33.533072	0.001000
27	(0, 3)	(1, 5)	2.611591	2.154620	3.068562	26.413772	0.001000
28	(0, 3)	(1, 2)	5.654923	4.613114	6.696731	25.087231	0.001000
29	(0, 3)	(1, 1)	5.883316	0.726483	11.040149	5.272942	0.010466
30	(0, 5)	(0, 2)	3.180746	2.683222	3.678270	29.548062	0.001000
31	(0, 5)	(0, 1)	3.903448	1.598837	6.208059	7.828250	0.001000
32	(0, 5)	(1, 4)	5.755935	5.416674	6.095197	78.414254	0.001000
33	(0, 5)	(1, 6)	2.642550	2.295164	2.989937	35.158035	0.001000

15/66 non significant

Appendix

Mental health

ANOVA

	PR(>F)
C(Education)	0.000000e+00
C(HeartDiseaseorAttack)	1.791553e-166
C(Education):C(HeartDiseaseorAttack)	2.909226e-14
Residual	NaN

TUKEY (education)

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	4	6	1.337912	1.232570	1.443254	51.184588	0.001000
1	4	3	1.649576	1.418569	1.880584	28.777925	0.001000
2	4	5	0.022757	-0.092517	0.138032	0.795619	0.900000
3	4	2	1.490835	1.150705	1.830965	17.664343	0.001000
4	4	1	2.910771	1.319424	4.502117	7.371501	0.001000
5	6	3	2.987488	2.762864	3.212112	53.599734	0.001000
6	6	5	1.315154	1.213274	1.417035	52.023160	0.001000
7	6	2	2.828747	2.492920	3.164574	33.946174	0.001000
8	6	1	4.248682	2.658250	5.839115	10.765935	0.001000
9	3	5	1.672334	1.442884	1.901784	29.372991	0.001000
10	3	2	0.158741	-0.235019	0.552502	1.624690	0.845447
11	3	1	1.261194	-0.342471	2.864860	3.169428	0.218931
12	5	2	1.513593	1.174519	1.852667	17.989837	0.001000
13	5	1	2.933528	1.342407	4.524649	7.430186	0.001000
14	2	1	1.419935	-0.203046	3.042916	3.525883	0.125626

TUKEY (HA/HD)

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	0	1	1.640016	1.542013	1.738018	46.384996	0.001

TUKEY (interaction)

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(0, 4)	(0, 6)	1.219678	1.092322	1.347034	44.262944	0.001000
1	(0, 4)	(0, 3)	1.500828	1.211031	1.790626	23.935918	0.001000
2	(0, 4)	(0, 5)	0.014023	-0.126080	0.154126	0.462598	0.900000
3	(0, 4)	(0, 2)	1.222313	0.789343	1.655283	13.047831	0.001000
4	(0, 4)	(0, 1)	1.524927	-0.474125	3.523979	3.525647	0.344969
..	...	...	...	...	...	...	...
61	(1, 3)	(1, 2)	0.126001	-0.922826	1.174827	0.555243	0.900000
62	(1, 3)	(1, 1)	7.493521	2.989527	11.997515	7.689569	0.001000
63	(1, 5)	(1, 2)	2.087758	1.178700	2.996817	10.614570	0.001000
64	(1, 5)	(1, 1)	9.455279	4.981766	13.928792	9.768763	0.001000
65	(1, 2)	(1, 1)	7.367521	2.820915	11.914126	7.489417	0.001000

16/66 non significant

4/15 non significant