An Experiment on the Life Expectancy of Men and Women across Continents (2018)

Submitted by Heet Shah, Sheau Kwan and Yu Ping Chuang

Email : shah_heet@bentley.edu, kwan_shea@bentley.edu, chang_yupi@bentley.edu

INTRODUCTION

According to Our World in Data (2017), life expectancy refers to the number of years a person is expected to live[1]. A nation's life expectancy reflects its social and economic conditions and the quality of its public health and healthcare infrastructure, among other factors[2] (BMJ, 2018). While deciding which topic to choose for our experiment, we came across many topics but one that stood out for us was the life expectancy rates across the world. In almost every study, life expectancy was the outcome which was being affected, and the variables affecting it differed from study to study.

Many studies were done a while ago where there were factors like epidemic diseases or wars which quite frankly were very rampant 50 years back and highly impacted the life expectancy around the globe. However, in this day and age with modern medicines and advancements in science, epidemic diseases have drastically reduced mortality rates.

We were left wondering that have we reached a stage where the life expectancy across all continents are the same? Also, are there differences in the life expectancy of men and women across continents? We came across multiple studies, and one of them Scientific American[3] shows a research where women have a longer life expectancy compared to men, we wanted to check this for ourselves in our paper.

After discussing various methods to use, we decided on using a two factor anova and orthogonal matrix as these could help us.

[1] https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted

[2] https://www.bmj.com/content/362/bmj.k2562

[3] https://www.scientificamerican.com/article/why-is-life-expectancy-lo/

LITERATURE REVIEW

We came across an article[4] from Inde (2011) where the life expectancy in Europe of Males and Females were studied. This piqued our curiosity and made us wonder if there are any differences between the life expectancy across all continents or across developed and developing continents or if there a difference between the life expectancy between males and females within continents.

We then dug a little deeper and studied the United Nations report on World Mortality 2017 [5]. This report shows that there has been an increase in life expectancy between 1990 and 2015, however the difference between the life expectancy between Men and Women were not explored. Since this report was published in 2017, we have chosen to take data from the most recent year ie 2018 and to check whether we get similar results to this report and to deep dive into the difference in the life expectancy between Men and Women. Also we were interested in knowing whether Gender and Continent are important factors that contribute to a higher life expectancy or not.

One thing to note is that in the absence of wars, new epidemics, or substantial economic reforms, lack of improvement or stagnation in life expectancy gains are viewed as a cause for concern, and actual declines in life expectancy are particularly alarming. A research of BMJ (2018) shows stagnation or declines in life expectancy may signal a decline in the health profile of the population driven by adverse socio-economic trends, a deterioration in the provision or quality of healthcare services, or worsening behavioral factors.

This experiment is the first step in understanding if there is a difference in life expectancy across the world and across genders, and if there is a further analysis can be conducted to find out why.

[4] https://www.ined.fr/en/everything_about_population/demographic-facts-sheets/focus-on/life-expectancy-europe/

[5] http://www.un.org/en/development/desa/population/publications/pdf/mortality/World-Mortality-2017-Data-Booklet.pdf

METHODOLOGY

Data source:

We got our data - Life expectancy by gender and continent 2018 from the database Statista[6] which is sourced from the Population Reference Bureau.

The sources for Population Reference Bureau are:

National statistical offices' websites, regional organizations' websites, online databases, statistical yearbooks, and bulletins from various countries; demographic surveys such as the Demographic and Health Surveys, Multiple Indicator Cluster Surveys, and Performance and Monitoring Accountability 2020 Surveys; the UN Demographic Yearbook 2016 and Population and Vital Statistics Report of the UN Statistics Division; World Population Prospects: The 2017 Revision, World Contraceptive Use 2018, and World Urbanization Prospects: The 2018 Revision of the UN Population Division; the International Data Base of the International Programs Center, U.S. Census Bureau; World Development Indicators online database of the World Bank; AIDS info online database of the UNAIDS; and FAOSTAT online database of the Food and Agricultural Organization of the United Nations.

The sources also include direct communication with demographers and country experts from around the world. Specific data sources may be obtained by contacting the authors of the 2018 World Population Data Sheet. For countries with complete registration of births and deaths, rates are those most recently reported. For more developed countries, vital rates refer to 2017 or 2016.

6 www.statista.com/statistics/270861/life-expectancy-by-continent/

About the dataset:

We have 2 independent variables i.e. Males and Females and a two way anova would help us calculate the mean differences between each group. We decided to conduct a two factor anova first.

We chose to compare the life expectancy rates for males and females across the four regions between the continents of Asia, North America, Africa and Europe.

The data we have will require us to do two factor ANOVA with no replication and no interaction. Since the data has no replication, the error term is zero. There is no "pure" way to estimate error since error is measured by considering more than one observation (replication) at the same treatment combination. Without replication, a cell mean equals the individual data value.

$TSS = SSBr + SSBc + SSIr,c$

With allocation of degrees of freedom of $RC - 1 = (R - 1) + (C - 1) + (R - 1)(C - 1)$

Since we cannot perform ANOVA without error, we will have to assume that there is no interaction between continent and gender. We will have to assume SSIr,c and MSIr,c represents the role of SSW and MSW respectively.

Once we completed that, we wanted to check for differences between the mean life expectancy in different continents. For that, we can use an orthogonality matrix. Since we are only interested in difference in one factor, we disregard the other factor and use that to assume we have 2 replications for each continent. This is one of the limitations of orthogonality matrix that have have developed but if there is a sex effect it will apply to all continents. Since we are just interested in finding the differences in continents, we developed an orthogonal matrix to check if there was a difference in the life expectancy between the following:

1. Developed (NA, Europe, Oceania) v/s Developing (Latin America & Caribbean, Asia, Africa) continents
2. Latin America & Caribbean v/s Africa
3. Europe v/s North America

ANALYSIS AND DISCUSSION OF RESULTS

Males are represented by 1 and Females by 2. The levels for each continent are as follows: North America - 1, Oceania - 2, Europe - 3, Latin America - 4, Asia - 5, Africa – 6

The output is as follows:

Two way Anova:

Between-Subjects Factors		
	N	
Gender	1.00	6
	2.00	6
Continent	1.00	2
	2.00	2
	3.00	2
	4.00	2
	5.00	2
	6.00	2

Tests of Between-Subjects Effects					
Dependent Variable: LifeExpectancy					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	458.167[a]	6	76.361	56.564	.000
Intercept	66454.083	1	66454.083	49225.247	.000
Gender	60.750	1	60.750	45.000	.001
Continent	397.417	5	79.483	58.877	.000
Error	6.750	5	1.350		
Total	66919.000	12			
Corrected Total	464.917	11			
a. R Squared = .985 (Adjusted R Squared = .968)					

We can see that gender and continent are both significant factors in determining life expectancy. Since both factors are significant, even if there was interaction, it would only solidify our findings as the F_{calc} would be smaller than it is calculated right now and it would make the factors remain significant.

Ranks	
	Mean Rank
VAR00001	5.50
VAR00002	4.50
VAR00003	5.00
VAR00004	3.00
VAR00005	2.00
VAR00006	1.00

Test Statistics[a]	
N	2
Chi-Square	9.143
df	5
Asymp. Sig.	.103
a. Friedman Test	

Friedman test shows there is no difference between life expectancy of different continents. This is different from the results of our two way ANOVA. The possible reason could be that Friedman test doesn't assume normality of the factor (continent) under consideration and allows for non-parametric testing of the factor. This puts our assumption of normality into question.

There has been some competing research that question the validity of Friedman test since both the technique test in different ways. Zimmerman and Zumbo (1993) discussed this difference in procedures and explained that the Friedman test (devised by the noted economist Milton Friedman) is not really a form of ANOVA but an extension of the sign test. Sign test tend to have lower power compared to ANOVA or F-test.

Assuming that the ANOVA F-test is more robust than the Friedman test and there is sufficient evidence showing that there is a significant difference in life expectancy due to continent (based on our previous research), we went ahead with the orthogonality matrix and augmented ANOVA to decompose the sum of square by continent contrasts defined in the end of the Methodology Section.

Gender	**North America**	**Latin America & the Caribbean**	**Europe**	**Asia**	**Oceania**	**Africa**
Male	77	73	75	71	76	61
Female	81	79	82	74	80	64
Mean	**79**	**76**	**78.5**	**72.5**	**78**	**62.5**

Next, we tried one-way ANOVA using only continent as a factor and gender as two replicates.

Below is the SPSS output:

ANOVA

LifeExpectancy

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	397.417	5	79.483	7.065	.017
Within Groups	67.500	6	11.250		
Total	464.917	11			

Orthogonal matrix

	North America	**Latin America & the Caribbean**	**Europe**	**Asia**	**Oceania**	**Africa**
Developed v/s Developing	-1	1	-1	1	-1	1
Latin America v/s Africa	0	-1	0	0	0	1
Europe v/s North America	1	0	-1	0	0	0

Divide 1st row by: $\sqrt{(-1)^2 + (-1)^2 + (-1)^2 + (1)^2 + (1)^2 + (1)^2 = 6}$ = 2.449

Divide 2nd row by: $\sqrt{(0)^2 + (-1)^2 + (0)^2 + (0)^2 + (0)^2 + (-1)^2 = 2}$ = 1.4142

Divide 3rd row by: $\sqrt{(1)^2 + (0)^2 + (-1)^2 + (0)^2 + (0)^2 + (0)^2 = 2}$ = 1.4142

Orthonormal matrix for Men and Women

	North America	Latin America & the Caribbean	Europe	Asia	Oceania	Africa
Developed v/s Developing	-0.4083	0.4083	-0.4083	0.4083	-0.4083	0.4083
Latin America & Caribbean v/s Africa	0	-0.7071	0	0	0	0.7071
Europe v/s North America	0.7071	0	-0.7071	0	0	0

	Y1	Y2	Y3	Y4	Y5	Y6			
	79	76	78.5	72.5	78	62.5	z	z^2	(z^2)*2(row)
Contrast 1	-32.2557	31.0308	-32.0515	29.60175	-31.8474	25.51875	-10	100	200
Contrast 2	0	-53.7396	0	0	0	44.19375	-9.54585	91.489	182.978
Contrast 3	55.8609	0	-55.50735	0	0	0	0.35	0.1225	0.245
							SSBc=191.4115*2=382.82		

ANOVA					
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	397.42	5	79.48	7.065	0.017
Developed v/s Developing	200	1	200	17.78	
Latin America & Caribbean v/s Africa	182.978	1	182.978	16.26	
Europe v/s North America	0.245	1	0.245	0.02	
Other Differences	14.597	2	7.2985	0.649	
Within Groups	67.5	6	11.25		
Total	464.92	11			

$F(1,6) = 5.99$

Contrast 1 = Developed vs Developing continents, $F_{crit} = 5.99$ and $F = 17.78$, $F > Fcrit$ which indicates that there is a statistical difference in the life expectancy between Developed v/s Developing continents.

Contrast 2 = Latin America v/s Africa, $F_{crit} = 5.99$ and $F = 16.26$, $F > Fcrit$ which indicates that there is a statistical difference in the life expectancy between Latin America and the Caribbean v/s Africa.

Contrast 3 = Europe v/s North America, $F_{crit} = 5.99$ and $F = 0.02$, $Fcrit > F$ which indicates that there is no statistical difference in the life expectancy between Europe v/s North America.

This is quite interesting to see, especially Contrast 2. One would assume that there would be a statistical difference in the life expectancy of developed and developing continents due to

various external factors, however there is a statistical difference in the life expectancy of Latin America & the Caribbean vs Africa.

One would also assume that there would be a statistically significant difference between life expectancy of North America vs Europe but we can see that there is not. This is probably one of the limitations of the model as it compares the mean life expectancy of North America vs all of Europe.

CONCLUSIONS

Our research has primarily centered around finding if there are any significant differences in life expectancy of different continents and to see if gender plays a role in that.

First, we started off with two factor ANOVA without replication with the data from Statista database and found that gender and continent found are significant factors in life expectancy. Later, we applied Friedman's test to the data to see if the results are the same with respect to continent (the column factor) and found that its actually the opposite as continent is not significant. One of the major differences could be the fact that Friedman's test doesn't assume normality of the data in the sample.

But however, there is competing research questioning the validity of Friedman's test and based on previous research, we could ahead with the fact that continent is a significant factor in predicting life expectancy.

Lastly, we come to main part of our research where we noted differences in the life expectancy of continents looking at different contrasts. We find that there is a significant statistical difference between developed and developing continents as well as Latin America & Caribbean v/s African continents. But, however, there was no difference between North America v/s Europe. I guess we would be to do a detailed comparison year over year or with more replications to come to a more accurate conclusion but based on the data we have, the difference is not statistically significant.

LIMITATIONS AND DIRECTIONS OF FUTURE RESEARCH

While researching on our topic, we found that many times life expectancy was taken into consideration by comparing it with another topic for example changes in life expectancy after an epidemic or effect of life expectancy on Economic growth.
However in our study we were keen on studying the life expectancy for Men and Women across continents and if there were any differences based on the gender or continent.

Some of the limitations are:

1. We have not considered other external factors that could affect the life expectancy, these external factors could be epidemics, medical advance, changes in population policies, etc.

2. We have not compared the life expectancy of countries within each continent. We have taken the average of each continent. There may be cases where some countries may be influencing the life expectancy of the continent at the overall level more than other countries.

For our future research, we can:

1. To support our conclusion, we can conduct a life expectancy study across countries within each continent/choose 2 continents i.e developed and developing and compare the countries from each of them.

2. We could conduct an age wise life expectancy comparison where it would be interesting to see if there are higher life expectancy for certain age ranges or not for each continent.

3. We could study the factors that affect womens' life expectancy and why are they higher than mens' life expectancy. Are there certain factors that make women live longer?

REFERENCES

- American Physical Association. Zimmerman and Zumbo(1993). The relative power of parametric and nonparametric statistical methods.

- http://psycnet.apa.org/record/1992-98949-019 Accessed Nov 29 2018

- BMJ (2018), 362:k2562. Recent trends in life expectancy across high income countries: retrospective observational study. https://www.bmj.com/content/362/bmj.k2562. Accessed Nov 29 2018.

- Ined. Adveev et al. (2011). Life expectancy in Europe. https://www.ined.fr/en/everything_about_population/demographic-facts-sheets/focus-on/life-expectancy-europe/. Accessed Nov 29 2018.

- Our World in Data. Ortiz-Ospina (2017). "Life Expectancy" – What does this actually mean? https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted. Accessed 29 Nov 2018.

- Population Reference Bureau.*Statista - The Statistics Portal*, Statista. Average Life Expectancy at Birth in 2018, by Continent and Gender (in Years). www.statista.com/statistics/270861/life-expectancy-by-continent/. Accessed Nov 29 2018

- Scientific American. The Sciences. Why is life expectancy longer for women than it is for men?.

- https://www.scientificamerican.com/article/why-is-life-expectancy-lo/. Accessed Nov 29 2018.

- United Nations. World Mortality 2017: Data Booklet. http://www.un.org/en/development/desa/population/publications/pdf/mortality/World-Mortality-2017-Data-Booklet.pdf. Accessed Nov 29 2018.