

## **Table of Content**

<b>1. Formulating Research Question .....</b>	<b>2</b>
<b>2. Collecting Data .....</b>	<b>2</b>
<b>3. Analysing Data.....</b>	<b>2</b>
<b>3.1 Numerical and Graphical Presentation of Data Collected.....</b>	<b>2</b>
<b>3.2 Normality Tests.....</b>	<b>3</b>
<b>3.3 Test on Equality of Variances .....</b>	<b>3</b>
<b>3.4 Two Sample t Test.....</b>	<b>4</b>
<b>4. Interpreting Results.....</b>	<b>4</b>
<b>5. Conclusion.....</b>	<b>4</b>
 <b>Appendix A : Lists of countries in Americas &amp; Asia Regions .....</b>	 <b>6</b>
 <b>References .....</b>	 <b>7</b>

## 1. Formulating Research Question

Research Question:

“Is there a difference in the mean number of deaths due to Alzheimer disease, for the Americas and Asia region?”

Type of test:

- Independent 2 sample t test, with at  $\alpha = 5\%$
- Sample are independent because measurements were taken from different regions

Let  $\mu_1$  be the mean population number of deaths due to Alzheimer disease for Americas region, and  $\mu_2$  be the mean population number of deaths due to Alzheimer disease for the Asia region.

$H_0 : \mu_1 - \mu_2 = 0$  (There is NO difference in mean number of deaths due to Alzheimer, for the Americas and Asia regions)

$H_1 : \mu_1 - \mu_2 \neq 0$  (There is difference in mean number of deaths due to Alzheimer, for the Americas and Asia regions)

## 2. Collecting Data

Data on death statistics by causes is collected from online website “Our World in Data” (Number of deaths by cause, World, 2019)

The countries are classified under the Americas and Asia regions, based on ILO (International Label Organisation) standards. Refer to Appendix A for the list of countries covered under the samples of Americas and Asia regions.

72 observations were collected for the Americas region (N1), and 78 observations were collected for the Asia region (N2). Thus, the 2 samples are slightly unbalanced.

The variables collected for each sample are:

Variable	Type
Region	Qualitative (Nominal)
No. of death die to Alzheimer disease	Quantitative (Discrete)

## 3. Analysing Data

### 3.1 Numerical and Graphical Presentation of Data Collected

Using the descriptive statistics (Diagram 1), the Asia region has a higher mean and median for deaths caused by Alzheimer, than the Americas region. As the median are lesser than the mean, the distribution of data for both the regions are positively skewed. The standard deviation of the Asia sample is wider than the Americas region.

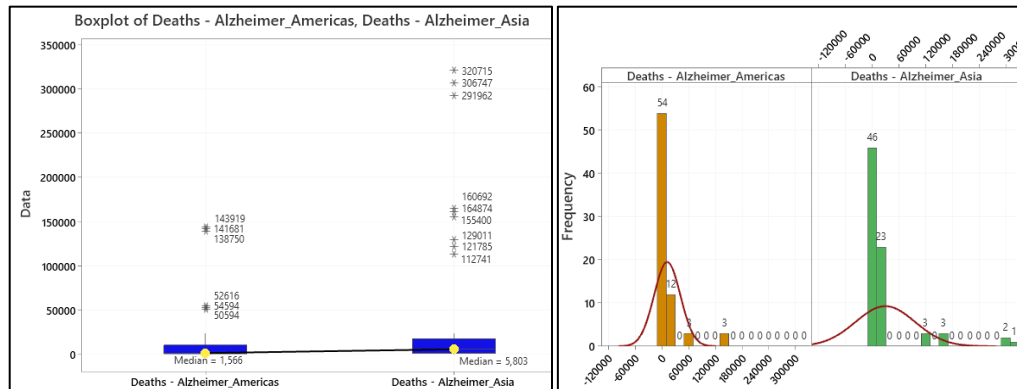
Statistics								
Variable	Total Count	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Deaths - Alzheimer_Americas	72	12091	29435	8	773	1566	10699	143919
Deaths - Alzheimer_Asia	78	28945	66981	21	1361	5803	17778	320715

Diagram 1: Descriptive Statistics from Minitab

In box plot of diagram 2, there are outliers for both the regions, with the sample from Asia having more extreme cases. The median and IQR is larger in the sample from Asia, confirming the Asia sample is more positively skewed, with wider deviation.

From the histogram of diagram 2, most of the frequencies are within the 10,000 number of deaths due to Alzheimer, for both Americas and Asia regions.

Reading from the raw data, the positively skewness in Asia region are from countries such as China, Japan, India. For Americas, the countries that had more deaths due to Alzheimer are USA, Brazil, Mexico. The minimum death due to Alzheimer, is from Greenland at 8 deaths in the Americas, and from Brunei at 21 deaths in the Asia region.

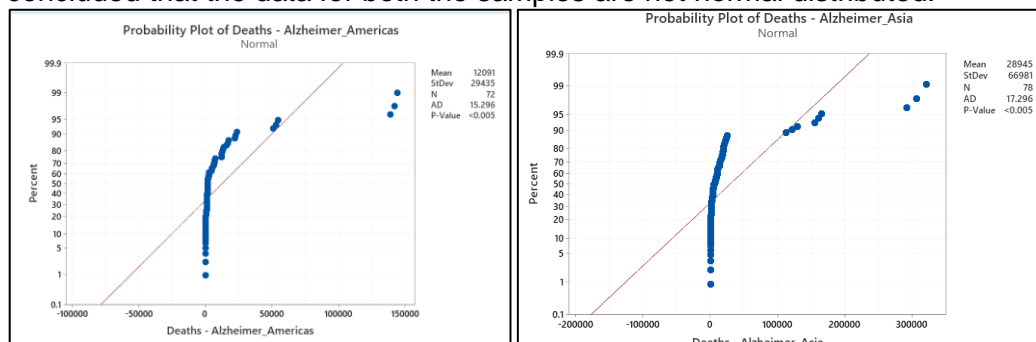


*Diagram 2: Boxplot for Sample 1 (left), and sample 2 (right)*

### 3.2 Normality Tests

Using Minitab to conduct Normality tests for both the samples on Americas and Asia regions, their data points did not follow the fitted distribution lines, as illustrated in diagram 3.

In addition, both their P-values are  $< 0.005$ , which are below the 0.05 threshold. It is therefore concluded that the data for both the samples are not normal distributed.



*Diagram 3: Normality Tests for Sample 1-Americas (left), and Sample 2-Asia (right)*

As the sample sizes are more than 30 ( $N_1=72$ ,  $N_2=78$ ), and applying Central Limit Theorem (CLT), we can still proceed with using t-distribution for hypothesis testing.

### 3.3 Test on Equality of Variances

To check if the 2 independent data sets have equal variances, a hypothesis test for the ratio of two variances was conducted, in place of visual examination via Diagram 2. The below results were obtained:

Method

$\sigma_1$ : standard deviation of Deaths - Alzheimer\_Americas

$\sigma_2$ : standard deviation of Deaths - Alzheimer\_Asia

Ratio:  $\sigma_1/\sigma_2$

The Bonett and Levene's methods are valid for any continuous distribution.

Ratio of Variances

Estimated Ratio	95% CI for Ratio using Bonett	95% CI for Ratio using Levene
0.193117	(0.035, 0.823)	(0.026, 1.091)

Test

Null hypothesis  $H_0: \sigma_1^2 / \sigma_2^2 = 1$

Alternative hypothesis  $H_a: \sigma_1^2 / \sigma_2^2 \neq 1$

Significance level  $\alpha = 0.05$

Method	Statistic	DF1	DF2	P-Value
Bonett	*			0.030
Levene	3.63	1	148	0.059

Diagram 4: Results for Test for Equal Variance

The p-value (Levene method) = 0.059, which is slightly  $> \alpha$  of 0.05. In addition, the 95% confidence interval for the ratio of variance is  $0.026 < \sigma_1^2/\sigma_2^2 < 1.091$ , which include 1. Thus, equal variance shall be used for 2-sample t test.

### 3.4 Two Sample t Test

$\mu_1$ : population mean of Deaths - Alzheimer\_Americas  
 $\mu_2$ : population mean of Deaths - Alzheimer\_Asia  
 Difference:  $\mu_1 - \mu_2$   
*Equal variances are assumed for this analysis.*

### Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Deaths - Alzheimer_Americas	72	12091	29435	3469
Deaths - Alzheimer_Asia	78	28945	66981	7584

### Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
-16854	52439	(-33790, 81)

### Test

Null hypothesis  $H_0: \mu_1 - \mu_2 = 0$   
 Alternative hypothesis  $H_a: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-1.97	148	0.051

Diagram 5: 2 sample t-tests results from Minitab

Based on the results from Minitab (Diagram 5)

- Difference of sample mean = -16,854
- $S_{\text{pooled}} = 52,439$
- The 95% confidence bound :  $-33,790 < \mu_1 - \mu_2 < 81$
- Degree of Freedom,  $DF = 148$  ( $N_1 + N_2 - 2$ )
- P-value = 0.051

### 4. Interpreting Results

#### Using P-Value:

Since P-value = 0.051, which is  $> \alpha = 0.05$ , it is not rare to obtain a difference of sample mean as extreme as -16,584, if the difference of population mean is 0.

#### Using CI:

We are 95% confident that the difference of population mean falls between -33,790 and 81, which includes the claimed difference mean population of 0.

### 5. Conclusion

Thus, we do not reject  $H_0$ .

At 0.05 significance level, there is NO difference in the mean number of deaths due to Alzheimer disease, for the Americas and Asia region.

Alzheimer is a type of brain disease, caused by damaged nerve cells (neurons) in the part of brain that are responsible for memory, language and thinking. There is currently no effective cure for it. (Alzheimer's Association, 2022)

Understanding the mean number of deaths due to Alzheimer in the Asia region (which is equivalent to that of the Americas region) allows better planning and allocation of resources to meet the care for people with Alzheimer disease in the Asia region.

Since  $H_0$  is not rejected, there is a chance of committing Type II error.

A type II error would have occurred if we fail to reject the null hypothesis that there is no difference in the mean number of deaths due to Alzheimer disease, for the Americas and Asia region, when actually there is a different.

## Appendix A : Lists of countries in Americas & Asia Regions

\*ILO Region = Classification based on International Label Organisation

### 1. Countries in America Region

ISO3 Code	Country	ILO Region*	ILO Subregion - Detailed
ARG	Argentina	Americas	South America
BLZ	Belize	Americas	Central America
BOL	Bolivia	Americas	South America
BRA	Brazil	Americas	South America
CHL	Chile	Americas	South America
COL	Colombia	Americas	South America
CRI	Costa Rica	Americas	Central America
ECU	Ecuador	Americas	South America
SLV	El Salvador	Americas	Central America
GUF	French Guiana	Americas	South America
GTM	Guatemala	Americas	Central America
GUY	Guyana	Americas	South America
HND	Honduras	Americas	Central America
MEX	Mexico	Americas	Central America
NIC	Nicaragua	Americas	Central America
PAN	Panama	Americas	Central America
PRY	Paraguay	Americas	South America
PER	Peru	Americas	South America
SUR	Suriname	Americas	South America
URY	Uruguay	Americas	South America
VEN	Venezuela, Bolivarian Republic of	Americas	South America
BMU	Bermuda	Americas	Northern America
CAN	Canada	Americas	Northern America
GRL	Greenland	Americas	Northern America
USA	United States	Americas	Northern America
VIR	United States Virgin Islands	Americas	

### 2. Countries in Asia Region

ISO3 Code	Country	ILO Region*	ILO Subregion - Detailed
CHN	China	Asia and the Pacific	Eastern Asia
HKG	Hong Kong, China	Asia and the Pacific	Eastern Asia
JPN	Japan	Asia and the Pacific	Eastern Asia
PRK	Korea, Democratic People's Republic of	Asia and the Pacific	Eastern Asia
KOR	Korea, Republic of	Asia and the Pacific	Eastern Asia
MAC	Macau, China	Asia and the Pacific	Eastern Asia
MNG	Mongolia	Asia and the Pacific	Eastern Asia
TWN	Taiwan, China	Asia and the Pacific	Eastern Asia
BRN	Brunei Darussalam	Asia and the Pacific	South-Eastern Asia
KHM	Cambodia	Asia and the Pacific	South-Eastern Asia
IDN	Indonesia	Asia and the Pacific	South-Eastern Asia
LAO	Lao People's Democratic Republic	Asia and the Pacific	South-Eastern Asia
MYS	Malaysia	Asia and the Pacific	South-Eastern Asia
MMR	Myanmar	Asia and the Pacific	South-Eastern Asia
PHL	Philippines	Asia and the Pacific	South-Eastern Asia
SGP	Singapore	Asia and the Pacific	South-Eastern Asia
THA	Thailand	Asia and the Pacific	South-Eastern Asia
TLS	Timor-Leste	Asia and the Pacific	South-Eastern Asia
VNM	Viet Nam	Asia and the Pacific	South-Eastern Asia
AFG	Afghanistan	Asia and the Pacific	Southern Asia
BGD	Bangladesh	Asia and the Pacific	Southern Asia
BTN	Bhutan	Asia and the Pacific	Southern Asia
IND	India	Asia and the Pacific	Southern Asia
IRN	Iran, Islamic Republic of	Asia and the Pacific	Southern Asia
MDV	Maldives	Asia and the Pacific	Southern Asia
NPL	Nepal	Asia and the Pacific	Southern Asia
PAK	Pakistan	Asia and the Pacific	Southern Asia
LKA	Sri Lanka	Asia and the Pacific	Southern Asia

## **References**

Alzheimer's Association. (2022). *2022 Alzheimer's Disease Facts and Figures - Special Report: More Than Normal Aging: Understanding Mild Cognitive Impairment*. Retrieved from <https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf>

Number of deaths by cause, World. (2019). Retrieved from Our World in Data: <https://ourworldindata.org/grapher/annual-number-of-deaths-by-cause>