**World Happiness Report**

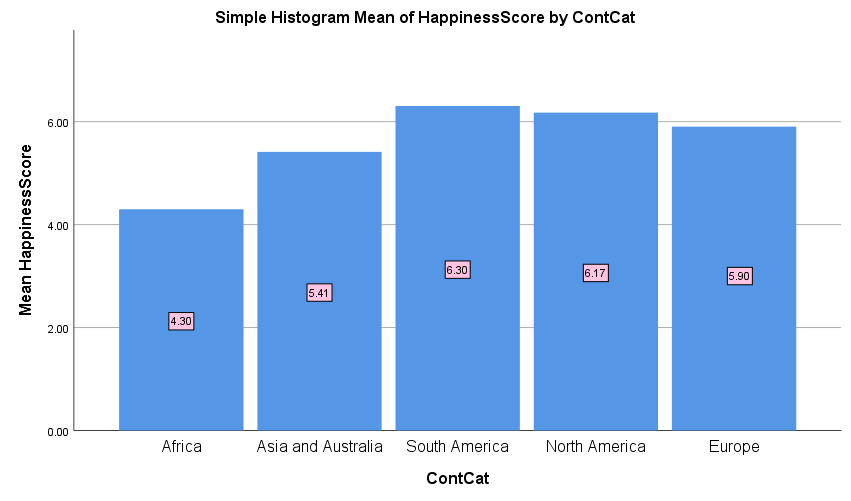
**The Dataset is from Kaggle site:** https://www.kaggle.com/unsdsn/world-happiness

The data is from 2013-2016 and consists of a world happiness report. The happiness report creates happiness score from 155 countries, which we translated into five continents- Europe, North America, South America, Asia-Australia, and Africa. To increase the sample size, we merged Asia and Australia together. The variables that investigated the extent to which each variable contributes to a country’s happiness score in this study are Economy, Life Expectancy, Continents, Freedom, and Happiness Score. All of the variables are quantitative, except for continents and countries which are our only categorical variables.

**Variables:**

* **The happiness score** was based on a metric measurement by asking the question, “How happy are you on a scale from 1 to 10 with 10 being the happiest?”
* **Economy** describes the GDP per capita
* **Life Expectancy** describes the extent to which a person’s health contributes to the calculation of the happiness score
* **Freedom** is based on the perceived ability to make life decisions and to which the extent that freedom contributed to the calculation of the Happiness Score.
* **Continents** are the countries grouped together by continent, including the merge of Asia and Australia.

World Happiness Score Visualization



Q1: What continents have a higher happiness score? Compare the means of different continents. ANOVA-test

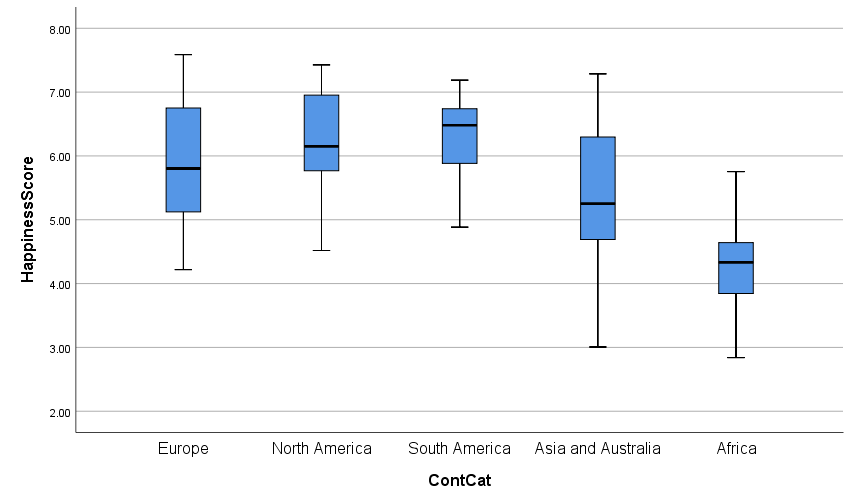
Assumptions

1) Each population is an SRS

2) Distribution is normal, but N is less than 40 for North America and South America

3) Assume all SD is equal or 1.07/ .637 = 1.67 which is less than 2 so assumptions are satisfied

Graph



**Comment:**

- We can see in the boxplot an overlap between North and South America.  Asia- Australia, and Africa have the lowest happiness mean value.

- The spread of some data values is very distinct like Africa has the fewest data comparing to others like Asia and Australia

- Can see some clear difference in means of happiness score in some continents like South America, Asia and Australia and Africa. However, there can still be some overlaps like the means of America continents and Europe can be overlap => can possibly group two American continents together

**Hypothesis:**

Null- All means of happiness scores are equal

Alternative- Not all means are equal,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | |
| Happiness Score | | | | | |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 84.165 | 4 | 21.041 | 26.460 | .000 |
| Within Groups | 121.669 | 153 | .795 |  |  |
| Total | 205.835 | 157 |  |  |  |

Compute test statistic:

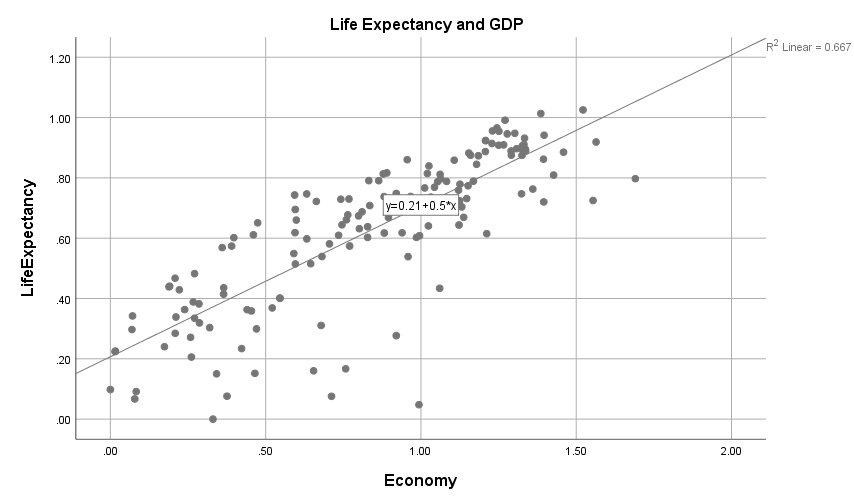
F= 26.46

P-value= <.0001

**Conclusion:**

Because P-value is much smaller than 0.01.  We can conclude that there is some clear difference between the happiness core of different continents.

**Q2: Is there a relationship between Economy and life expectancy?  Linear regression**



There is a positive relationship between life expectancy and economy. From the graph we can see a strong relationship with few possible outliers since the data points are pretty close-clustered to each other.

Regression line equation-

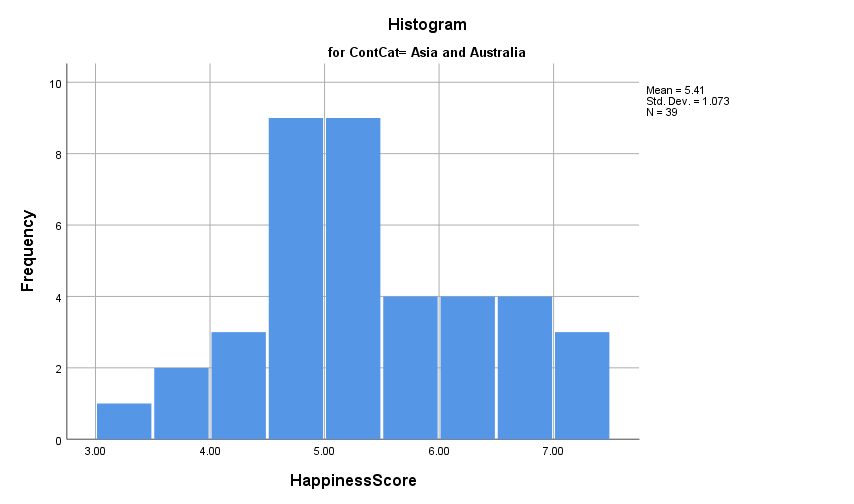
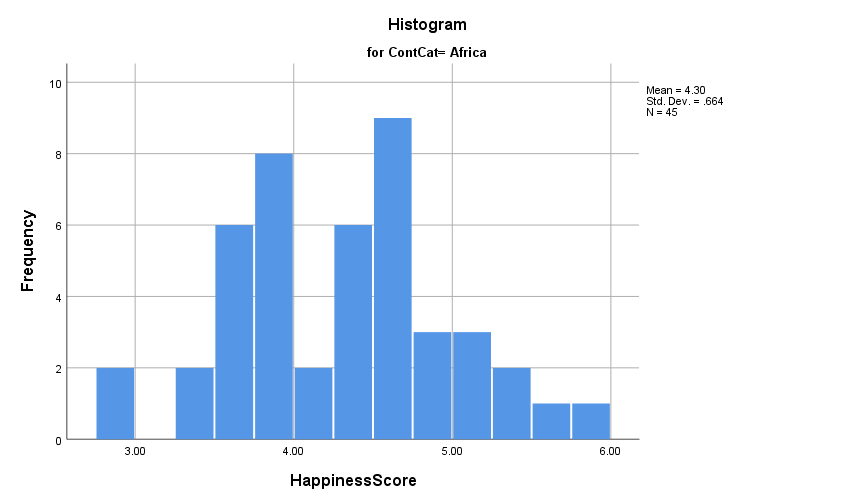
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .816a | .667 | .664 | .14311 |
| a. Predictors: (Constant), Economy | | | | |

The higher the R square value, the stronger the correlation between the two attributes. Here, R-square is .66 which is greater than .5. We can conclude that there is some decent positive correlation between life expectancy and economy. That means more welfare you are the longer you can possibly live.

**Q3: Are people in Asia-Australia happier than people in Africa? - T-test for two independent samples**

**Graph:**

**Africa:**

****

**Comment:**

Asia and Australia even though has less population than 40, but we can still see in the sample that it has a somewhat normal distribution with a possible left skew, one major peak, graph is unimodal, mean a has value of around 5

Africa has a slight right skew distribution and two major peaks, graph is bimodal, mean a has value of around 4. However, since the population is greater than 40, we can assume that the data is normally distributed by Central Limit Theorem

**Assumptions-**

1. both Africa and Asia-Australia are independent simple random samples
2. Samples are not normally distributed, but the sample sizes (45, 39) for central limit theorem
3. both means of Africa and Asia-Australia and both standard deviations for both variables are *unknown*

**The two-sample t-test:**

**Hypothesis:**

**Null:** Both Africa and Asia-Australia have the same means in happiness score

**Alternative:** Africa and Asia-Australia do not have the same score means

**Compute test statistic:**

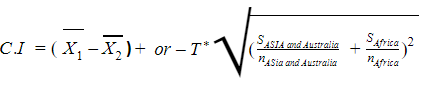
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | ContCat | N | Mean | Std. Deviation | Std. Error Mean |
| HappinessScore | Asia and Australia | 39 | 5.4124 | 1.07255 | .17175 |
| Africa | 45 | 4.2982 | .66398 | .09898 |

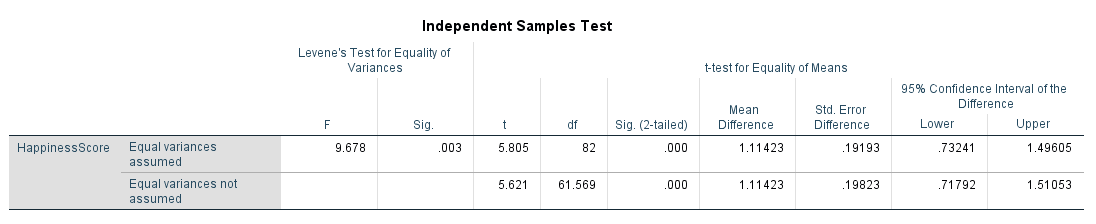
T= 5.804

P-value= 0.003

**Conclusion: The P-value 0.003 is lower than 0.05, therefore we have strong evidence against the Null hypothesis, which proves that both Africa and Asia-Australia do not have an equal means of happiness score. Which is confirmed also in the graph by how distinct these two data sets are.**

**Confident interval:**



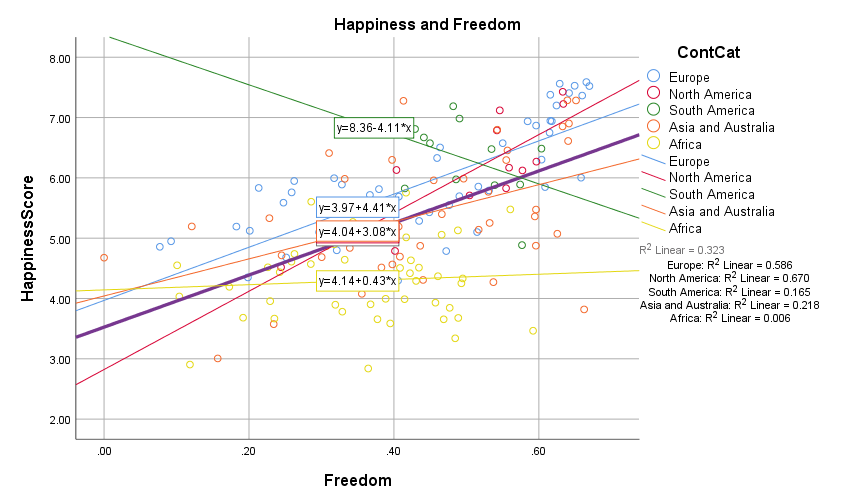


**Conclusion of Confidence Interval:**

**95% confidence Interval is between .73241 and 1.49605**

**That we are 95% confident that the difference between the mean of happiness score in Asia and Australia compare to African is between .73241 and 1.49605**

**Q4: Does how much freedom a citizen have contribute to the overall happiness score?**



By looking at the scatterplot, we can see there is a positive relationship between happiness scores and freedom. There are a few possible outliers

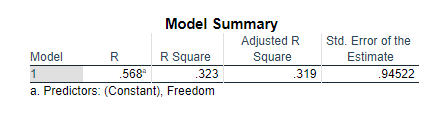
Europe has a positive slope with an R square of .586 indicating a moderate relationship between freedom and happiness score.

North America has a positive slope with an R square of .670 indicating a strong relationship between freedom and happiness.

South America has a negative slope with an R square of .165 indicating the relationship between freedom and happiness score in South America is weak.

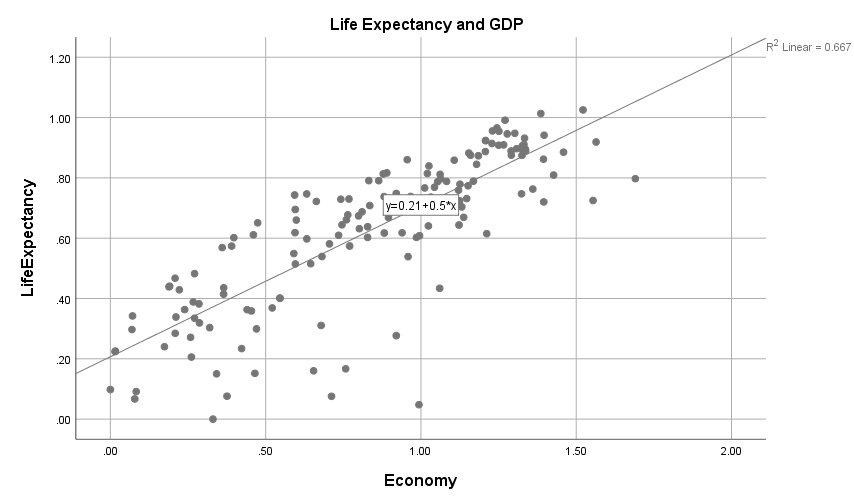
Asia-Australia has a positive slope with an R square of 0.218 indicating a weak relationship between freedom and happiness.

Africa has a positive slope with an R square of 0.006, indicating almost no relationship between freedom and happiness.



The computed R square does not give a strong correlation between the two variables happiness score and freedom status. We can see among the different continents both variables vary.

**Q3: Is there a relationship between Economy and life expectancy?  Linear regression**



There is a positive relationship between life expectancy and economy. From the graph we can see a strong relationship with few possible outliers since the data points are pretty close-clustered to each other.

Regression line equation-

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .816a | .667 | .664 | .14311 |
| a. Predictors: (Constant), Economy | | | | |

The higher the R square value, the stronger the correlation between the two attributes. Here, R-square is .66 which is greater than .5. We can conclude that there is some decent positive correlation between life expectancy and economy. That means more welfare you are the longer you can possibly live.