1.0 Introduction

Happiness is something everyone wants in their entire life. It is an emotional state defined by the feeling of satisfaction and enjoyment. In this project, we decided to make some tests on the happiness of people who live in five of the European countries, which are the United Kingdom, Spain, Germany, Italy and France. The reason why we are interested in this project is because Europe is different from our country, Malaysia which is in Asia. This is because European countries are famous for their development compared to other regions. A fact said that some of the countries in Europe had the most visitors from all around the world. Thus, this shows that European countries really have the catalyst to become one of the happiest regions in the world. Therefore, by studying the happiness of people living in these countries, we may know a little bit about their daily lives, environment and their thoughts towards their countries.

2.0 Dataset

The dataset that we have chosen were some samples of people from these five countries from Europe with their mean of rating of happiness. These are the variables gathered from the dataset:

Country	Number of respondents, n	Mean of happiness, $\bar{x}$
Germany	2003	7.5
Spain	1013	7.6
France	1529	7.8
Italy	1506	7.0
United Kingdom	1499	7.8
Total	7550	37.7

From these data, we can calculate the value of average happiness of respondents from all countries, standard deviation and variance to make data analysis from these values.

1. Average, $\bar{x} = \frac{37.7}{5} = 7.5$

2. Standard deviation, $\sigma =$
$$\sqrt{(7.5 - 7.5)^2 + (7.6 - 7.5)^2 + (7.8 - 7.5)^2 + (7.0 - 7.5)^2 + (7.8 - 7.5)^2/(5-1)}$$
$$= 0.33$$

3. Variance, $\sigma^2 = (0.33)^2 = 0.11$

3.0 Data Analysis

3.1 Hypothesis sample test

Claim: All of the countries' mean of Happiness should be equal to 7.8 as the m=7.8 is identified as the ideal mean of happiness for the country to achieve.

1) Hypothesis

$H_0 : m = 7.8$

$H_1 : m \neq 7.8$

The hypothesis is based on the claim where all of the countries' mean of Happiness should be equal to 7.8 as the m=7.8 is identified as the ideal mean of happiness for the country to achieve.

2) Calculate value of Z statistics

$$z = \frac{7.5 - 7.8}{(0.33/\sqrt{5})} = -2.0328$$
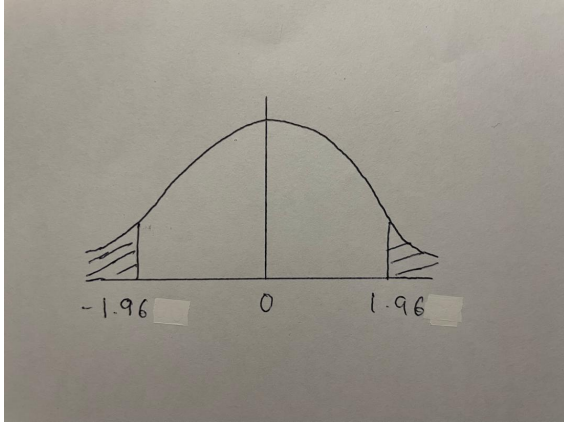
3) Calculate value of critical value

$\alpha = 0.05$,

CV= $Z_{0.025}$=1.96

4) Calculate p-value of Z

P(Z>2.0328) x2 = 0.02104x2

= 0.04208

Since this is a two-tailed test, we divide the alpha because we need to do a statistical difference testing in two different directions.

5) Draw graph

6) Decision and Conclusion

Since 0.04208<0.05, we reject H_0. Therefore, we do not accept the claim that the average mean of happiness should equal 7.8 in order to be considered as an ideal mean of happiness.

3.2 Correlation test

Claim: Test if there is correlation between two data, number of people and the mean of happiness. This claim will show us if the number of people will affect the mean of happiness between the five European countries. If there is effect between both variables, this means that there exists correlation. However, if the number of people does not affect the mean of happiness, this shows that there is no correlation between them.

x	y	xy	y^2	x^2
2003	7.5	15022.5	56.25	4012009
1013	7.6	7698.8	57.76	1026169
1529	7.8	11926.2	60.84	2337841
1506	7.0	10542	49.00	2268036
1499	7.8	11692.2	60.84	2247001
$\Sigma = 7550$	$\Sigma = 37.7$	$\Sigma = 56881.7$	$\Sigma = 284.69$	$\Sigma = 11891056$

Scatter plot:
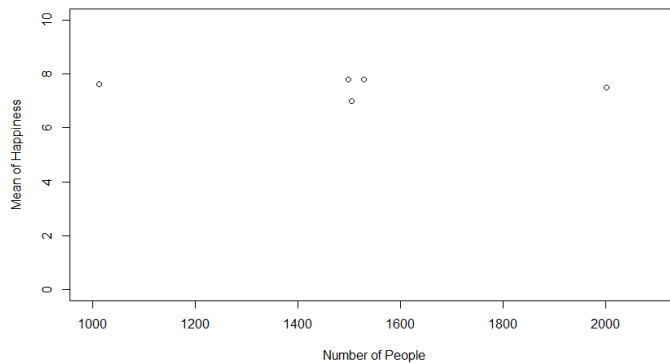
From the above graph, we can say that it is a weak linear relationship between the number of people and their means of happiness. Therefore, we need to conduct a test to make a conclusion if there is correlation or not.

1) Hypothesis

H_0 : There is no correlation

H_1 : There exists correlation

2) Calculate value of r

$$r = \frac{(56881.7) - (7550)(37.7)/5}{\sqrt{((11891056) - (7550)^2/5)((284.69) - (37.7)^2/5)}} = -0.098$$

3) Calculate t-test

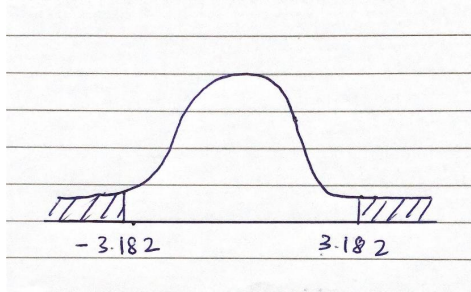$$t = \frac{-0.098}{\sqrt{\frac{1-(-0.098)^2}{5-2}}} = -0.171$$

4) Calculate critical value

$\alpha = 0.05, \ df = 5\text{-}2 = 3$

Reject H_0 if t > $t_{0.025,3}$ = 3.182

$t < - \ t_{0.025,3}$ = -3.182
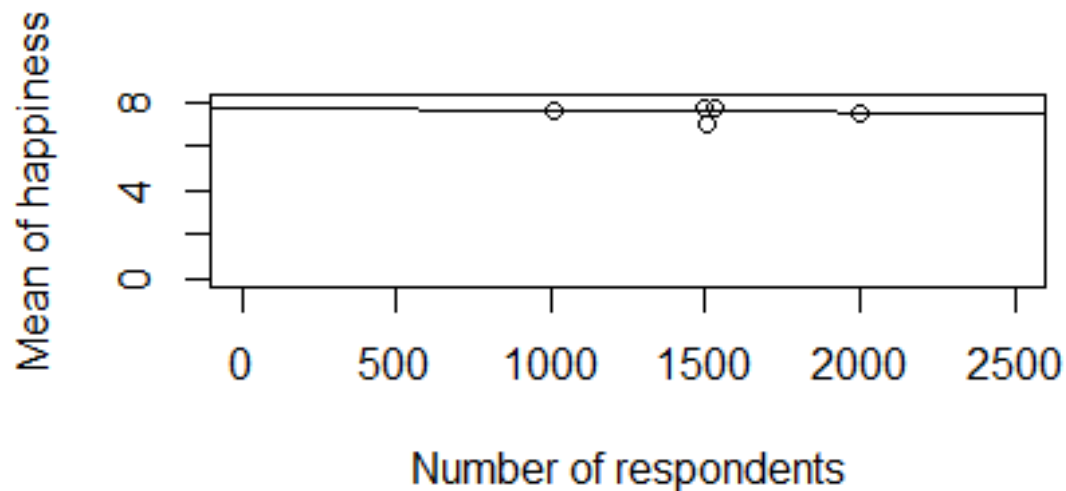
5) Draw graph

-3.182   3.182

6) Decision and Conclusion

Since -3.182 < -0.171 < 3.182, we fail to reject H_0. Therefore, there is no correlation between the number of people and their mean of happiness.

3.3 Regression test

Claim: The number of respondents will have a big impact on the mean of happiness if a strong linear relationship exists between the two variables.

Coefficients:
y estimated = 7.679 - 0.00009234x

Intersection coefficient, b_0
b_0 = 7.679 indicates that for the range of mean happiness observed, 7.679 is the portion of the max heart rate not picked by the respondent.

Slope Coefficient, b_1

b_1= -0.00009234 tell us that the average number of respondent decreases by 0.0009234 on average

Coefficient of Determination R2

H_0 : R2 = 0 (no linear relationship)
H_A : R2 ≠ 0 (linear relationship exists)

Multiple R2 = 0.009683
Adjusted R2 = -0.320

From the calculation, there is a weak linear relationship between the number of respondents and the mean of happiness.

Inference test

1) Hypothesis
H_0 : p = 0 (no linear relationship)
H_A : p ≠ 0 (linear relationship exists)

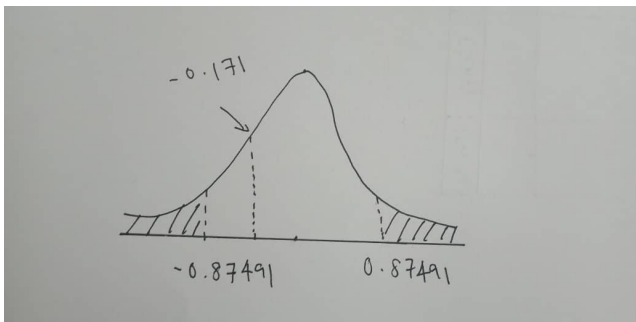2) Standard error
Estimated Standard Error = 0.0005392

3) T-test
t-value = -0.171

4) P value
p = 0.87491

5) Draw graph

6) Decision and Conclusion

Decision = Accept null hypothesis

Conclusion= Since -0.87491 < -0.171 < 0.87491 , we fail to reject H_0.There is no linear relationship between the two variables hence the number of respondent doesn't affect the mean of happiness

3.4 Chi Square test (Equal frequencies/probabilities)

Test the claim that the observed mean of happiness has equal proportions in the 5 countries.

Country	Germany	Spain	France	Italy	United Kingdom
Observed Mean of Happiness	7.5	7.6	7.8	7.0	7.8

1) Hypothesis

$H_0: p_1 = p_2 = p_3 = p_4 = p_5$

H_1: At least one of the proportions is different from others.

2) Calculate Expected frequency

Expected frequency, $E = \frac{n}{k} = \frac{(7.5+7.6+7.8+7.0+7.8)}{5} = 7.5$

Country	Germany	Spain	France	Italy	United Kingdom
Observed Mean of Happiness	7.5	7.6	7.8	7.0	7.8
Expected Mean of Happiness	7.5	7.5	7.5	7.5	7.5
$(O - E)^2/E$	0.0002	0.0005	0.0090	0.0387	0.0090

4) Calculate Chi-Square test
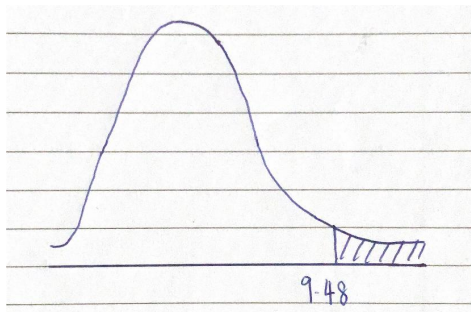
Chi-Square test,

$$X^2 = \Sigma \frac{(O-E)^2}{E} = 0.0002 + 0.0005 + 0.0090 + 0.0387 + 0.0090 = 0.0574$$

5) Calculate critical value

        $\alpha = 0.05$, df = k-1 = 5-1 = 4

        Reject H_0 if $X^2 > X^2_{4,0.05} = 9.48$

6) Draw graph

7) Decision and Conclusion

        Since $0.0574 < 9.48$, we fail to reject H_0. Therefore, we accept the claim that the observed mean of happiness has equal proportions in the 5 countries.

3.5 Anova test

Claim: The mean of happiness is equal in the 5 countries.

1) Hypothesis

        $H0 : \mu1 = \mu2 = \mu3 = \mu4 = \mu5$

        H1 : at least one mean is different
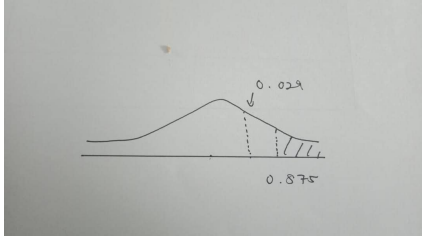
2) F-test

        F value = 0.029

3) Critical value

        Critical value = 0.875

4) Draw graph

5) Decision and Conclusion

Decision = Accept null hypothesis

Conclusion = As $0.029 < 0.875$, we fail to reject H_0. There is sufficient evidence to support the claim that the mean of the population is the same.

4.0 Conclusion

In conclusion, there are lots of ways to carry out analysis of data. Before that, there are some factors that we need to consider when conducting these tests. Firstly, when we are choosing data, we need to make sure that all data must have all the values needed for tests. This is to ensure that we can arrange the data into a table based on the tests that we want to carry out. Besides that, we can calculate some important values such as standard deviation and variance. As for pre-processing, we managed to arrange the data we collected into a table. By doing pre-processing of data, we can change the data from raw data to understandable data. Finally, as for data analysis, we find that if we want to make an analysis, we must find the problem statement first. With the problem statement, it will be easy for us to carry out tests and come out with conclusions from the tests we conduct. In the regression test, we used a y estimate to predict the value of the dependent variable(mean of happiness) for any given independent variable(number of respondents). From these tests, the most interesting thing we found out is that the European countries have a quite high mean of happiness averaging 7.54, which shows that Europeans truly are content and happy with their lives. This could be a benchmark or guide for Malaysians to follow as we pursue a happier life as we only live once, hence we'll have to live it to the fullest.

Appendix

https://perso.telecom-paristech.fr/eagan/class/igr204/data/happiness.csv