UTM UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

UTM BACHELOR OF COMPUTING SCIENCE (DATA ENGINEERING) COMPUTING PROGRAMME

PROBABILITY AND STATISTICAL DATA ANALYSIS

(SECI2143-02)

REPORT: PROJECT 2

(MALNUTRITION ACROSS THE GLOBE)

NAME OF STUDENTS: NUR HADIRAH MUNAWARAH BINTI ROZMIZAN

MATRICS NUMBER: A19EC0201

LECTURER: DR CHAN WENG HOWE

SUBMISSION: 27TH JUNE 2020
DATE

SIGNATURE &:
SCORE

INTRODUCTION

Malnutrition continues to be the reason for making children much more vulnerable to disease and death. There are 4 broad types of malnutrition: wasting, stunting, underweight and overweight. The dataset is created by Ruchi Bhatia. The purpose of the study is to know the number of malnutrition cases country-wise. In addition, this data determines which countries bear the greatest contributions of all forms of malnutrition. This data is focused on children under 5 years, which is 0-59 months. The main purpose is to know which country has the highest number of malnutrition cases. There are about 5 cases that I am going to study which are hypothesis testing, correlation, regression, anova and goodness of fit. I hope that I will able able to have correct proposed analysis and determine the number of malnutrition cases country-wise as well as the greatest contributions of all forms of malnutrition.

Statistical Analysis

Hypothesis testing on 1 sample test (Variable: Underweight)

a) Hypothesis Statement

H_0: μ = 8

H_1: μ ≠ 8

b) Execution of test

df=151

α = 0.05

$z = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ =6.205744

cv: $Z_{0.05}$= ±1.645

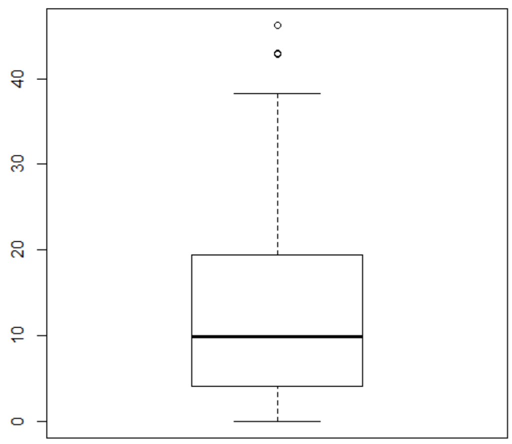
p-value = 1.959964

$\bar{x}$ = 13.32538

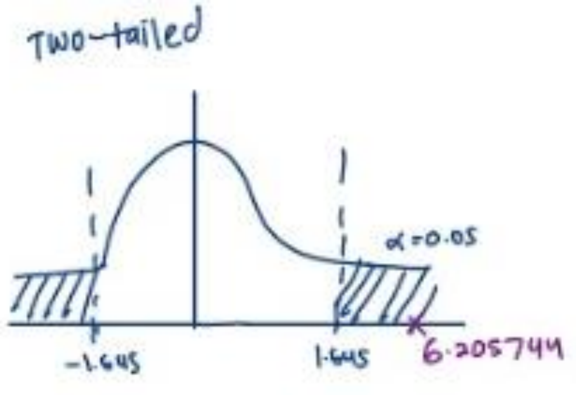
s=10.93298

R studio #recheck

```
One Sample t-test

data:  Underweight
t = 6.0053, df = 151, p-value = 1.367e-08
alternative hypothesis: true mean is not equal to 8
99 percent confidence interval:
 11.01196 15.63879
sample estimates:
mean of x
 13.32538
```



Boxplot of underweight



c) Interpretation of Results

Conclusion = Reject Ho at α = 0.05. There is sufficient evidence that the mean of underweight is not equal to 8.

d) Conclusion and Discussion

Since c-value = 1.645 < test statistic = 6.205744 at α = 0.05, reject Ho. This result suggests that the average underweight is not good average.

Correlation (Variable: Income Classification, Wasting)

a) Hypothesis Statement

H_0: μ = 0 (No linear correlation)

H_1: μ ≠ 0 (linear correlation exists

```
R studio

        Pearson's product-moment correlation

data:  Income.Classification and Wasting
t = -7.5689, df = 150, p-value = 3.541e-12
alternative hypothesis: true correlation is not equal t
o 0
95 percent confidence interval:
 -0.6320129 -0.3999823
sample estimates:
       cor
 -0.5257073
```

b) Execution of test

α = 0.05

t=-7.5689

df=150

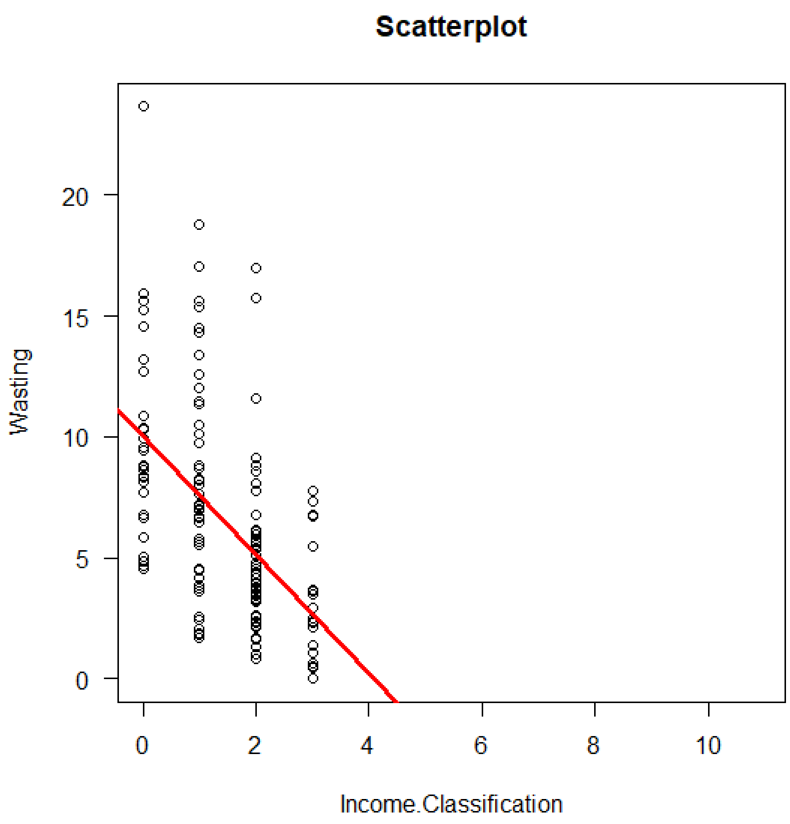
p-value= $3.541e^{-12}$

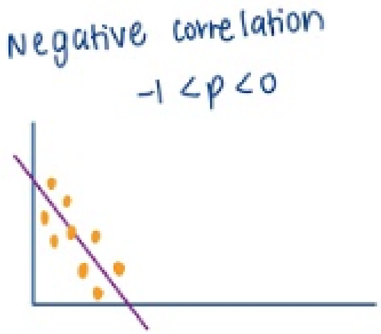
correlation estimate = -0.5257073

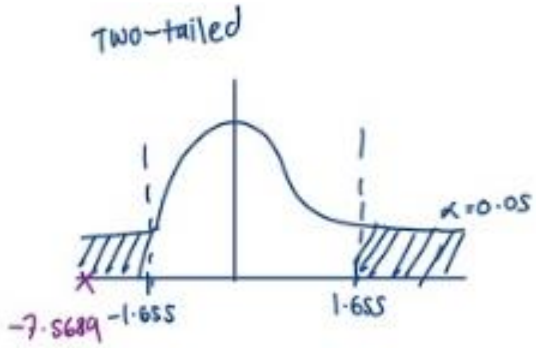
cv : $t_{0.05,148}$ = ±1.655

```
(Intercept) Income.Classification
  10.016904             -2.454751
```

$$\hat{y} = 10.016904 - 2.454751x$$







c) Interpretation of Results

Conclusion = Reject Ho. There is not sufficient evidence of a linear relationship between Income Classification and Wasting at the 5% level of significance.

d) Conclusion and Discussion

Since c-value=-1.655 > test statistic=-7.5689, Reject Ho at α = 0.05. This means that it is negative correlation which is relation shows hat a high score on overweight is related to a low core on income classification. It could be seen that overweight decreases as the income classification increase. A scatter plot and correlation analysis of the data indicates that there is negative relationship between income classification and overweight.

Regression (Variables: Severe wasting, Overweight)

Type of regression = Simple regression

a) Hypothesis Statement and Variables

Ho: $\beta_1 = 0$ (no linear relationship)

H1: $\beta_1 \neq 0$ (linear relationship does exist)

Dependant variable: Overweight

Independent variable: Severe wasting

b) Execution of test

n=150

df=148

$\alpha = 0.05$

Min=-7.0205

Max=19.5701

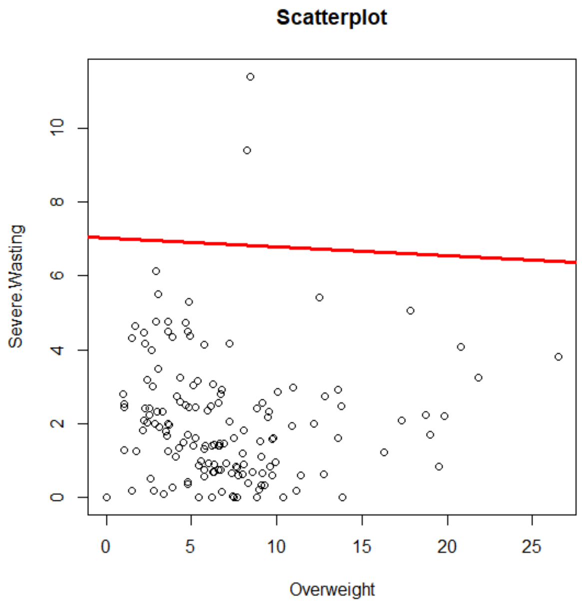
F-statistic = 0.01164

p-value = 0.9142

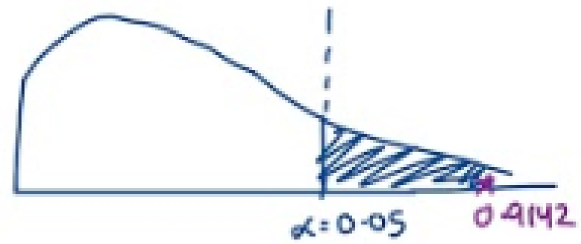
Standard error of estimate = 4.689

R studio

```
Call:
lm(formula = Overweight ~ Severe.Wasting)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.0205 -3.3007 -0.7704  1.9932 19.5701 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     7.02047    0.58889  11.921   <2e-16 ***
Severe.Wasting -0.02385    0.22107  -0.108    0.914    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.689 on 148 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  7.861e-05,	Adjusted R-squared:  -0.006678 
F-statistic: 0.01164 on 1 and 148 DF,  p-value: 0.9142
```



$$\hat{y} = 7.02047 - 0.02385x$$
$$R^2 = 7.861e^5$$



c) Interpretation of Results

reject Ho at α = 0.05. There is sufficient evidence that overweight affect severe wasting.

d) Conclusion and Discussion

Since p-value=0.9142 > α = 0.05, reject Ho. This means that there is no relationship between overweight and severe wasting. The values of severe wasting do not depend on overweight which none of variation in severe wasting is explained by variation in overweight.

Anova (Variable: Overweight, Underweight, Stunting)

a) Hypothesis Statement

H_0: All means are same

H_1: At least one of the means is different

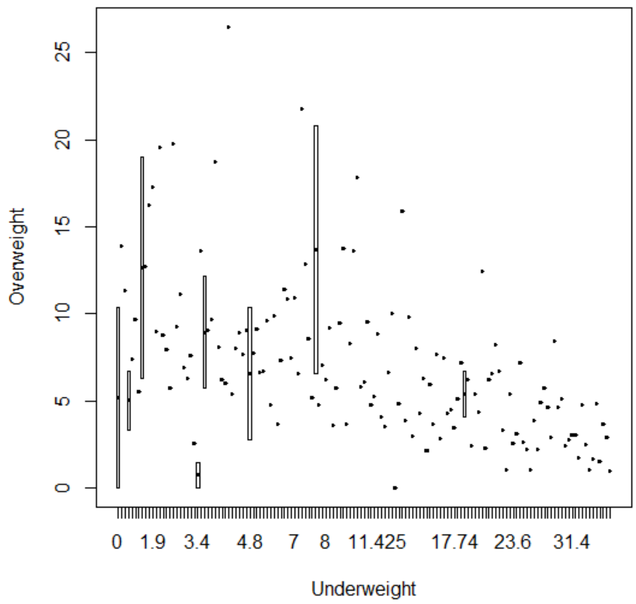
b) Execution of test

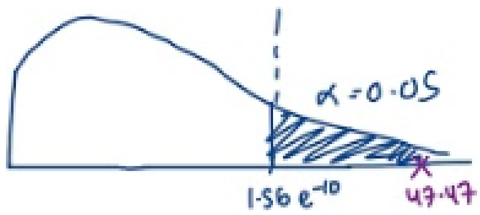
F-statistic = 47.27

P-value=1.56e^{-10}

R studio

```
Df           SumSq Mean     Sq      F value    Pr(>F)
Underweight    1  803.1    803.1    47.27 1.56e-10 ***
Residuals    150 2548.4     17.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0
.1 ‘ ’ 1
```





c) Interpretation of Results

Since F test statistic= 47.47 > F p-value = $1.56e^{-10}$, reject Ho. There is sufficient evidence that not all means are equal.

d) Conclusion and Discussion

There are significant differences between means of underweight and overweight. Therefore, all does not have same mean.

Goodness-of-fit test (Variables: Income Classification)

a) Hypothesis Statement

Ho: σ=0.967019

H1: σ>0.967019

R studio

```
Chi-squared test for given probabilities

data:  tab
X-squared = 17.316, df = 3, p-value = 0.0006085
```

b) Execution of test

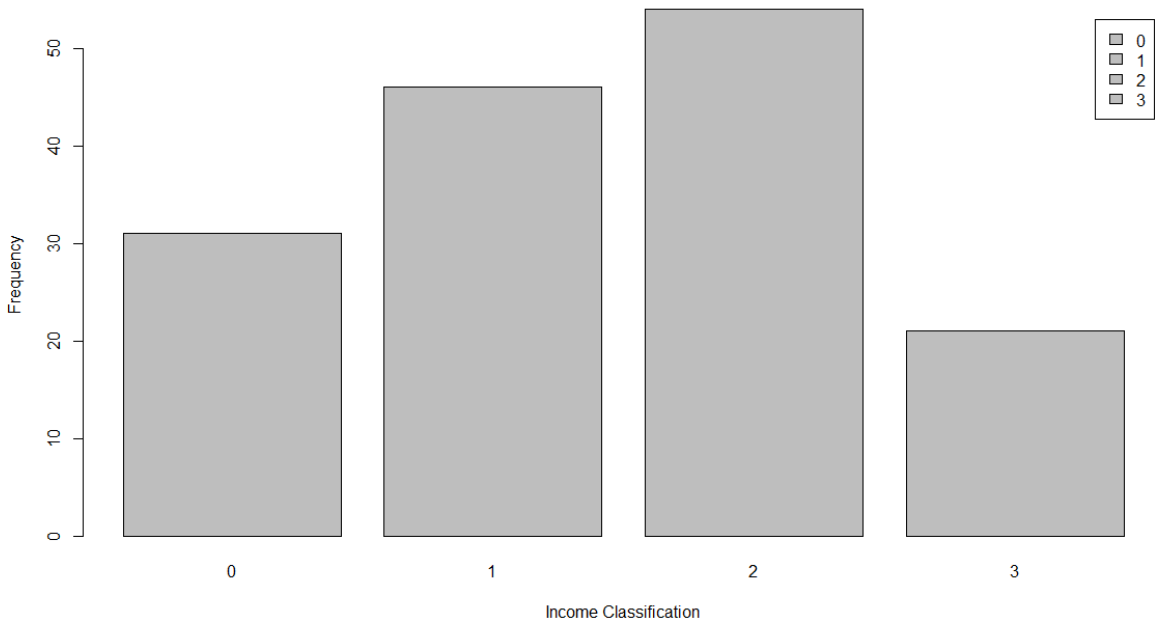
df=3

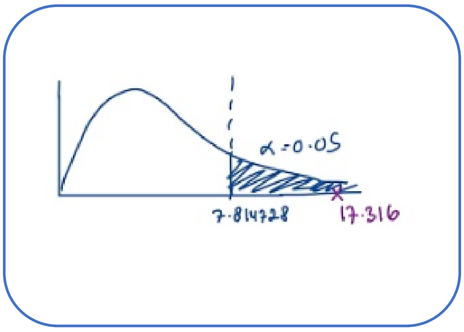
x^2 = 17.316

α = 0.05

$c.v_{0.05,3}$ =7.814728

p-value = 0.0006085





c) Interpretation of Results

Since x^2 = 17.316 > p-value =7.814728 α = 0.05, reject Ho. There is sufficient evidence that support the standard deviation of income classification is 0.967019

d) Conclusion and Discussion

There is evidence to support rejecting the assumption that the standard deviation is 0.967019 at α = 0.05 which is p-value < x^2

CONCLUSION

In a nutshell, all broad types of malnutrition: wasting, stunting, underweight and overweight really does effect malnutrition toward children. It is proved that there are decline or rise in the number of malnutrition cases country-wise. The biggest contribute of malnutrition is that severe wasting and overweight. Severe wasting also causes overweight because the foods are too many and throw their food away. Increasing of severe wasting also increases number of overweight people.