UTM

UNIVERSITI TEKNOLOGI MALAYSIA

Project 2 Report

PSDA

SEC 3

No	Name	Matric number
1.	Muhammad Kevin Rezza Ramadan	A19EC0264

Lecturer:
DR. ARYATI

Date: 26/06/2020

Introduction

Obesity in the United States is a major health issue resulting in numerous diseases, specifically increased risk of certain types of cancer, coronary artery disease, type 2 diabetes, stroke, as well as significant increases in early mortality and economic costs. Numerous studies have attempted to identify contributing factors for obesity in the United States. Common factors include an overconsumption of food, in this case, simple breakfast like cereal can also contribute to the increased rate of obesity.

In this project, I decided to find the conditions from factors inside cesuch as Calories, Protein, Fat, sodium, carbs. These aspects could be the key to figuring out the pattern of the increased of obesity conditions for the united states. My target population is the name of the cereal. I hope that I can find out the pattern of the increased of obesity conditions for the united states.

Methodology

To carry out inferential statistics, secondary data (nutritional value and other variables of cereals) is obtained from https://www.kaggle.com/crawford/80-cereals. The target population is the name of the cereals. Variables such as calories is identified. Inferential statistics are carried out by using hypothesis testing two sample, correlation, regression, and chi square test of independence.

	name	mfr	type	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight	cups	rating
2	100% Brar	N	C	70	4	1	130	10	5	6	280	25	3	1	0.33	68.403
3	100% Natı	Q	C	120	3	5	15	2	8	8	135	0	3	1	1	33.9837
4	All-Bran	K	C	70	4	1	260	9	7	5	320	25	3	1	0.33	59.4255
5	All-Bran w	K	C	50	4	0	140	14	8	0	330	25	3	1	0.5	93.7049
6	Almond Dı	R	C	110	2	2	200	1	14	8	-1	25	3	1	0.75	34.3848
7	Apple Cinr	G	C	110	2	2	180	1.5	10.5	10	70	25	1	1	0.75	29.5095
8	Apple Jack	K	C	110	2	0	125	1	11	14	30	25	2	1	1	33.1741
9	Basic 4	G	C	130	3	2	210	2	18	8	100	25	3	1.33	0.75	37.0386
10	Bran Chex	R	C	90	2	1	200	4	15	6	125	25	1	1	0.67	49.1203
11	Bran Flake	P	C	90	3	0	210	5	13	5	190	25	3	1	0.67	53.3138
12	Cap'n'Crui	Q	C	120	1	2	220	0	12	12	35	25	2	1	0.75	18.0429
13	Cheerios	G	C	110	6	2	290	2	17	1	105	25	1	1	1.25	50.765
14	Cinnamon	G	C	120	1	3	210	0	13	9	45	25	2	1	0.75	19.8236
15	Clusters	G	C	110	3	2	140	2	13	7	105	25	3	1	0.5	40.4002

Table1: a section of secondary data of cereals nutritional value

Hypothesis Testing 1-Sample

One sample test on calories contained on cereal 1 to cereal 15

I want to determine whether the calories mean of 15 cereals is greater than 150 calories under the t-test of 5% significance level. I assumed that the data are normally distributed.

H0 : μ = 150

H1 : μ > 150

For the test, I perform the test statistic and critical value by using R programming, using z = (xbar-mu)/(sd/sqrt(n)) for test statistics and z.alpha = qnorm(1-alpha) for critical value

```
> z = (xbar-mu)/(sd/sqrt(n))
> n = 15
> sd = 21.8
> xbar = 101.3
> mu = 150
> alpha = 0.05
>
> z = (xbar-mu)/(sd/sqrt(n))
> z
[1] -8.652032
>
> z.alpha = qnorm(1-alpha)
> z.alpha
[1] 1.644854
>
```

Code1: Hypothesis test 1 by R programming

From the code i got -8.652032 for the test statistic and 1.644854 for critical value, Since the statistic value is less than the critical value, the null hypothesis is not rejected.

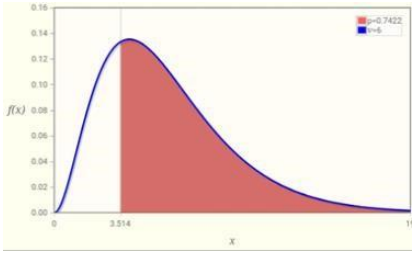
Therefore, there is not enough evidence to claim that the calories mean for 15 cereals is greater than 150 calories under 5% significance level

Independence Chi Square Test

In this test, I would like to find out whether obesity increasing rate is dependent for the calories. I assumed that the data are normally distributed.

H0: obesity increasing rate is independent for the calories

H1: obesity increasing rate is dependent for the calories



Graph 1: Graph of Chi-squared test

```
> x = c(70,120,70,50,110,110,110,130,90,90,120,110,120,110,110)
> d <- data.frame(x)
> chisq.test(d, correct=FALSE)

        Chi-squared test for given probabilities

data:  d
X-squared = 70.789, df = 14, p-value = 1.388e-09

> alpha <- 0.05
> x2.alpha <- qchisq(alpha,
+                    df=14,
+                    lower.tail=FALSE)
>
> x2.alpha
[1] 23.68479
>
```

Code 2: Chi-squared test by R programming

For the test, I used the formula of Chi-squared test statistics: chisq.test(d, correct=FALSE) and x2.alpha <- qchisq(alpha,+ df=14,+ lower.tail=FALSE) to perform the Chi-squared critical value by using R programming.

From the code, I got 70.789 and 23.68479 for the statistic and critical values. Since the statistic value is not less than the critical value, the null hypothesis is rejected.

From the test, I conclude that there is enough evidence to claim that obesity increasing rate is dependent for the calories. This might be caused of not only calories determine the increase in obesity rate, but also other factor.

Correlation

Calories and Ratings are two of the variables that maybe can make the obesity rate rising. I would like to find if there any sufficient evidence of linear relationship between calories and ratings of the cereal for the obesity increase rate. In this test, I took sample from 15 different cereal from different manufacturer. The type of correlation coefficients that used in this report is Pearson's product-moment correlation coefficient.

H0: There is no linear correlation between calories and ratings of the cereal for the obesity increase rate

H1: The linear correlation exists between calories and ratings of the cereal for the obesity increase rate

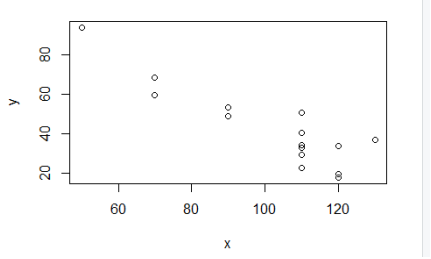
```
> x <- c(70,120,70,50,110,110,110,130,90,90,120,110,120,110,110)
> y <- c(68.4,33.9,59.4,93.7,34.3,29.5,33.1,37.0,49.1,53.3,18.0,50.7,19.8,40.4,22.7)
> #calculate corr. coefficient
> cor(x,y)
[1] -0.895747
>
> cor.test(x,y)

        Pearson's product-moment correlation

data:  x and y
t = -7.2648, df = 13, p-value = 6.32e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9651453 -0.7086571
sample estimates:
       cor
-0.895747
```

Code 3: The calculation of Correlation by R Programming

For the test, I used the formula of Correlation: cor.test(x,y) to perform the correlation coefficient and the P-value by using R programming. From the code, I got 6.32^{-6} for the p-value and -0.895747 for the correlation coefficient. Since the P-value is less than the significance level alpha=0.05, the null hypothesis is rejected. Therefore, there is sufficient evidence of linear relationship between calories and ratings of the cereal for the obesity increase rate with the correlation coefficient of -0.895747 at the 5% level of significance.



Graph 2: Scatter plot of the relationship between exports and imports (Correlation)

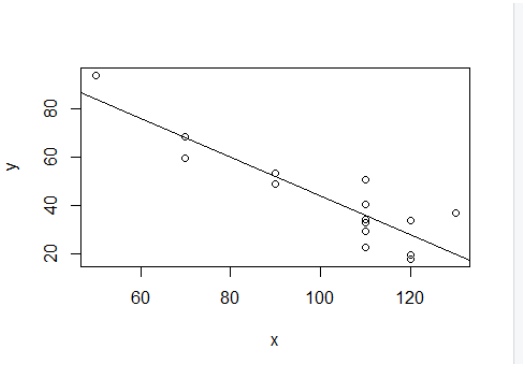
From the scatter plot and the correlation coefficient value, r = -0.895747 significant at the level of 0.05, shows that there is negative linear relationship between x (calories) and y (rating) values of the 15 different cereal from different manufacturer. In short, the calories. values will increases as the rating value of the cereal.

Regression

I took sample from 15 different cereal from different manufacturer, The dependent variable (y) is the collected rating values whereas the independent variable (x) is the calories of 15 different cereal from different manufacturer. This analysis aims to test the existence of a linear relationship between the variable x and y.

The null hypothesis, H0 : β1 = 0

The alternative hypothesis, H1 : β1 ≠ 0



Graph 3: Scatter plot of the relationship between exports and imports (Rregression)

```
data:  x and y
t = -7.2648, df = 13, p-value = 6.32e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9651453 -0.7086571
sample estimates:
      cor
-0.895747

> model <- lm(y~x)
> model

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
   124.1097      -0.8015
```

Code 4: R calculation for regression

For the test, I used the formula lm(y~x), to perform the linear regression. From the regression graph, this model is shown in a negative linear relationship with an equation of y = 124.1097 - 0.8015(x). The value of intersection coefficient, β0 of 124.1097 indicates that the range of the calories values that not being affected with the rating. The value of slope coefficient, β1 of – 0.8015 just indicates the average value which increasing the calories values that being affected with the rating of the cereal.

Therefore, there is a relationship between the x and y variables based on the calculated value of β1. The coefficient of determination in this relationship is -0.895747.

Conclusion

Among americans citizen, mostly on its own, or as part of oat-based breakfast cereals, and All-Bran, but no other breakfast cereals, were associated with a reduction in obesity risk. This effect may be due to particular characteristics of these cereal nutritional value such as calorie content, but these relationships needs further investigation and case study. Generally perceived as part of a balanced diet, and as a healthier option compared to a cooked breakfast, this research would seem to suggest that cereal is not as healthy as it pretends to be, and might just making you unhealthy instead.

However, there are still hopes for american citizen whose still categorized as obese or overweight. Things that possible to do start from the simplest thing by choosing wisely what cereal to consume, paying more attention the quality and quantity of the cereal..Those things hopefully can make american citizen more healthy thus making american citizen to be prosperous and live a happy life in the future.

Appendix

https://en.wikipedia.org/wiki/Obesity_in_the_United_States#:~:text=the%20CDC%20figures.-,Sex,of%20American%20adults%20were%20obese.

https://www.kaggle.com/crawford/80-cereals

https://www.esquireme.com/content/22359-are-cereals-making-us-obese

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5618177/