# **CA2 - MSc in Data Analytics**

Author: Tony Walsh

e-mail: [sba22238@student.cct.ie](mailto:sba22238@student.cct.ie)

Student ID: sba22238

Abstract

*Butter*

# 

# Inferential Statistics

I wanted to do look at the relationship between the amount of Raw Milk collected at Irish creameries and the selling price of Milk. This looked to be a standard economic relationship.

I developed a simple Linear Regression model to examine this. The independent variable Milk selling price made almost no contribution to the amount of Raw Milk produced. The one variable that made a difference to the amount of Raw Milk collected was month with a training score of 0.85.

I then used the model to examine the amount of Cream, Drinking Milk and Butter produced using Milk Collected as the independent variable. The results for Cream and Drinking Milk were both under a training score of 0.10. However, Milk Collected made a significant contribution to the amount of Butter produced with a training score of 0.95.

The model shows that excess Raw Milk collected is turned into Butter.

I did some investigation and there are a couple of reasons for this. The main ones are Irish butter is a valuable export and it has a long shelf life lasting up to 9 months or longer when frozen. A trip to my local Tesco has Butter with a best before of almost 5 months in the future. This means it is an easy product to store and export on ships. Each kilo of Butter uses up the 25 liters of milk, so it uses up the excess Raw milk collected.

# Butter Supermarket Sample t-test

As part of the inferential statistics section above I was interested in the shelf life of butter. I took samples of butter and cream-based spreads in Tesco and Aldi.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample | 1 | 2 | 3 | 4 | 5 | 6 |
| Tesco | 149 | 116 | 70 | 89 | 138 | 95 |
| Aldi | 109 | 80 | 110 | 116 | 75 | 113 |

Sample for Tesco:

N = 6

Average Self Life: 100.5

Standard Deviation: 18.05

Sample for Aldi:

N = 6

Average: 109.3

Standard Deviation: 30.09

Significance level = 0.05

H0 is that the self life of the two supermarkets is the same

H1 is that the self life of the two supermarkets is different

I will reject H0 if t > 2.22 or if t < -2.22

Using a calculator, I find that t = -0.6143 which is not in the rejection zone therefore we can conclude that the average shelf life of the butter products in the two supermarkets is the same.

# Hypothesis test for the year 2000

I will compare Ireland and Belgium monthly data for the year 2000.

## Raw Milk Deliveries

A Shapiro-Wilk test was performed on the Irish and Belgium data. The Irish data had a test result of 0.11 and he Belgium data a test result of 0.83. In both cases the p value is greater than 0.05 and the null hypothesis is not rejected, and the data is normal.

A t-test with a significance level of 0.05 was performed and the test statistic T equaled 2.5 which was just inside the rejection zone (t > 2.07 or t <-2.07).

I conclude that there is enough evidence to conclude that there are significant differences between the Raw Milk collections between Ireland and Belgium.

## Butter Produced

The Shapiro-Wilk test results were 0.19 for Ireland and 0.15 for Belgium and I conclude that the data is normal.

A t-test with a significance level of 0.05 was performed and the test statistic T equaled 0.889 which is outside the rejection zone (t > 4.37 or t <-4.37) and we can conclude that the sample average between the two samples is not big enough to be statistically significant.

## Milk Produced

The Shapiro-Wilk test results were 0.75 for Ireland and 0.67 for Belgium and I conclude that the data is normal.

A t-test with a significance level of 0.05 was performed and the test statistic T equaled –7.42 which is well within the rejection zone (t > 2.73 or t <-2.73) and I conclude that there is enough evidence to conclude that there are significant differences between the Raw Milk produced between Ireland and Belgium.

Research