Lactose Intolerance

Evidence for Short Stature or Vitamin D Deficiency in Prepubertal Children.

By Edward Muñoz Garro

1. Introduction

1.1. Background

This work is carried out based on the work of Setty-Shah, et al. (2013), in which it is analyzed whether some variables such as height, weight or vitamin D levels are significant or not in lactose intolerance in preschooler children, what is done in this work is a reanalysis of the data and corrections regarding the way the data is analyzed.

1.2. Problem

In this work, 87 children of similar ages and who were not taking vitamin supplements were studied to determine if overweight, short stature or vitamin D deficiency are significant in lactose intolerance, for which measures of vitamin D levels were taken weight, height and other variables, the body mass index was calculated, classified the children in normal weight and overweight, for the performance of the test. The result of the experiment was that neither weight nor vitamin D levels are significant in lactose intolerance.

1.3. Interest

Lactose intolerance is a serious problem for some people, since these people deprive themselves of eating many foods, including much of the pastry, if deterministic factors are found, treatment could be obtained from an early age.

2. Data acquisition and cleaning

2.1. Data sources

The data was obtained from the following link:

https://escholarship.umassmed.edu/datasets/1/

2.2. Feature selection

For this study the following database variables were used: Individual, Group, Vitamin D, Height, BMI, Weight, Height z score and Weight z score

3. Exploratory Data Analysis

3.1. Weigth and Vitamin D in the intolerance

In the Setty-Shah, et al. (2013) article, based on the body mass index, which is continuous, the individuals were classified as "Overweight" and "Normal Weight", which is clearly a discrete variable. Vitamin D level was first tested as the continuous variable on the “y” axis and two discrete variables on the “x” axis (Weight and Group), with this they carried out the factorial ANOVA test; but before normality was tested (apparently visually) and it is not said whether homoscedasticity was tested.

If we verify their data in R with a histogram, we can see that there is a bell shape, but very displaced to the left, if we perform the Shapiro test we can see that there is no normality in the data (p = 9.414E-05), but If we test with leveneTest, it can be seen that if there is homoscedasticity (p = 0.6079), in any case, a parametric test cannot be performed, since both assumptions must be met.

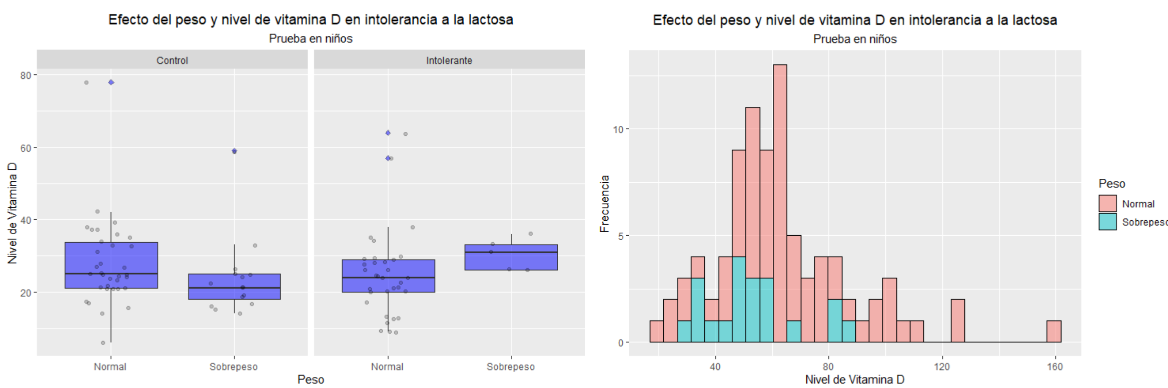

Figure 1: Visualization of points distribution at left, histogram at right.

Then they performed the ANOVA Factorial test, the same finding that there are no significant differences, but these results are not valid because there is no normality. After seeing that there are no significant differences, it is stated that a posthoc test was performed, which does not make any sense, since this test is used to see more clearly where the differences are, but the ANOVA said that there are no significant differences.

3.2. Weight and height z-score on intolerance

The next thing to do was to repeat the experiment, but with the z-score value of the height of the individuals, in this case if there was normality (p = 0.6099) and homoskedasticity (p = 0.2345), in the article the test is carried out ANOVA and we verified in R that indeed, there are no significant differences between groups, nor in the interaction of groups and weight, but if in weight, when doing a post-hoc analysis, the difference noted is that overweight children in the control group are taller than children with normal weight (somewhat normal), so it can be said that there are no differences.

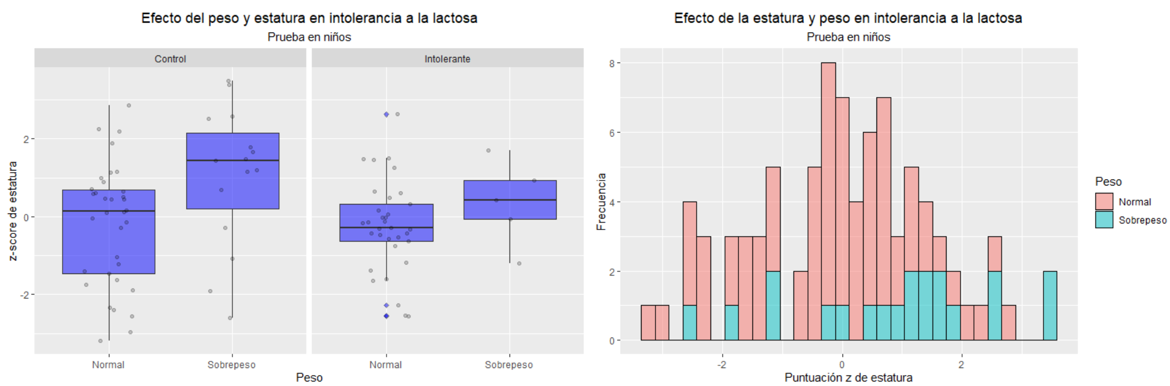

Figure 2: Visualization of points distribution at left, histogram at right.

4. Statistical Tests

4.1. Weigth and Vitamin D in the intolerance

Since the first test works with non-normal data, personally choose to perform a GLM test, since instead of forcing our data to fit with a specific type of statistical test, what we should do is adapt our statistical test to our data. In them there are two types, the binomial that is designed for data with values of 0 or 1 or between 0 and 1 on its 'y' axis, the second type is the poisson, which is designed for more counting data and is I was able to use for all data that does not meet the binomial test criteria.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	4.26144	0.02037	209.245	< 2e-16	***
Overweight	-0.3025	0.04107	-7.366	1.76E-13	***
IntolerantGroup	-0.1536	0.03022	-5.082	3.74E-07	***
Overweight: IntolerantGroup	0.19467	0.07647	2.546	0.0109	*

Table 1: Results of GLM (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1).

The result of this GLM test is very different from that obtained by the ANOVA test of the study, here we can see a very significant effect between overweight and vitamin D levels ($p < 0.001$), in turn a very significant effect between intolerance to lactose and vitamin D levels ($p < 0.001$), it is also appreciated that the effect between overweight and intolerance is not sufficient to distinguish between a significant effect or product of coincidence ($p = 0.0109$). With this we can affirm that there is a relationship between vitamin D levels and lactose intolerance.

Instead of grouping individuals into the overweight and normal weight groups, I recommend using their weight and height variables directly, since the latter are of a continuous nature, while if we categorize it we are losing an enormous power of analysis and therefore our results will not have the same value. In this case I will directly use the variable BMI (body weight index) in another GLM poisson.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	4.616826	0.058620	78.759	< 2e-16	***
BMI	-0.020401	0.002658	-7.676	1.64e-14	***
IntolerantGroup	-0.405823	0.104365	-3.888	0.000101	***
BMI: IntolerantGroup	0.014293	0.005142	2.779	0.005446	**

Table 2: Results of GLM (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1).

As you can see, the results were not that far from those of the other GLM test initially, however there is now significant evidence that BMI is related to lactose intolerance, all because of the use of a heavier variable statistical. Personally, I do not like to use z-scores since we do not work directly with the data, but rather with abstractions of them, however this type of evidence is totally valid.

It is important to note that if we analyze the dispersion in both GLM we can see that there is no under-dispersion or over-dispersion.

4.2. Weight and height z-score on intolerance

In this case, a z-score is being used, which we can define as the number of standard deviations that a given value takes with respect to the mean of its sample or population and therefore its nature will always be normal. The next step is to proceed with an ANOVA Factorial test, since there are two variables on the x-axis and a continuous y-variable, in addition to complying with the ANOVA assumptions. The results are the following.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)	
Grupo	1	2.58	2.582	1.226	0.27135	
Peso	1	15.81	15.808	7.507	0.00752	**
Grupo:Peso	1	1.18	1.18	0.561	0.45613	
Residuals	83	174.77	2.106			

Table 2: Results of ANOVA (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1).

As can be seen, there is a significant effect with respect to weight, however, in this case, conclusions cannot be rushed, to specify what exactly this effect is, a posthoc test must be carried out, in this case, as it is a subject related to science A Tukey test was used for health.

As a result of the posthoc test, the following can be stated:

- Individuals with normal weight and overweight have significant differences in their vitamin D level.
- Individuals in the normal weight control group had significantly different levels of vitamin D than those in the same overweight group.
- Individuals in the overweight control group had significantly different levels of vitamin D than those in the normal-weight intolerant group.

However, none of this speaks about whether x or y group is determinant for lactose intolerance, so it is stated that there is insufficient evidence for the diagnosis (in this test).

5. Conclusions

Whenever you want to carry out a statistical test, you must first corroborate the nature of the data, since if you take assumptions as valid, you can commit great statistical sins and communicate false results, you should think that our results will be used for the taking such as whether or not a treatment is good for a specific condition. In this case we can see that in some points it coincides with the paper of Setty-Shah, et al. (2013), but in others we obtained the opposite results (as in the GLM tests).

As noted above, overweight is linked to vitamin D levels, also intolerance is linked to vitamin D levels, however, being overweight is not directly related to lactose intolerance, on the other hand the index of body mass is related to both intolerance and vitamin D levels, this difference may be due to the difference in pressure derived from the different types of statistical variable.

A good analysis of the data can lead us to make better decisions when carrying out our tests and communicate the correct results.

6. References

Setty-Shah, N., Maranda, L., Candela, N., Fong, J., Dahod, I., Rogol, A. D., & Nwosu, B. U. (2013). Lactose intolerance: lack of evidence for short stature or vitamin D deficiency in prepubertal children. *PLoS One*, *8*(10), e78653.