



## Non-parametric analysis of the effects of $\alpha$ S1-casein genotype and parturition non-genetic factors on milk yield and composition in Murciano-Granadina goats

M. G. Pizarro, V. Landi, F. J. Navas González, J. M. León & J. V. Delgado

To cite this article: M. G. Pizarro, V. Landi, F. J. Navas González, J. M. León & J. V. Delgado (2019) Non-parametric analysis of the effects of  $\alpha$ S1-casein genotype and parturition non-genetic factors on milk yield and composition in Murciano-Granadina goats, Italian Journal of Animal Science, 18:1, 1021-1034, DOI: [10.1080/1828051X.2019.1611388](https://doi.org/10.1080/1828051X.2019.1611388)

To link to this article: <https://doi.org/10.1080/1828051X.2019.1611388>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 22 May 2019.



[Submit your article to this journal](#)



Article views: 404



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)

## Non-parametric analysis of the effects of $\alpha$ S1-casein genotype and parturition non-genetic factors on milk yield and composition in Murciano-Granadina goats

M. G. Pizarro<sup>a</sup> , V. Landi<sup>a</sup> , F. J. Navas González<sup>a</sup> , J. M. León<sup>b</sup>  and J. V. Delgado<sup>a</sup> 

<sup>a</sup>Departamento de Genética, University of Cordoba, Córdoba, Spain; <sup>b</sup>Centro Agropecuario Provincial de Córdoba, Diputación Provincial de Córdoba, Córdoba, Spain

### ABSTRACT

Limited sample sizes imply parametric assumptions could be violated, even if traits have been reported to fulfil parametric assumptions. Parametric studies have addressed a non-significant influence of CSN1S1 genes on Murciano-Granadina milk yield, fat, protein and dry extract. We used non-parametric categorical tests to find alternative statistical methods to analyse the power to explain the variability found in the population regarding milk yield and its components. We analysed 2090 records for milk yield, and its components from 710 Murciano-Granadina CSN1S1-genotyped goats. Categorical regression equations were issued to predict which and at what level these factors may determine milk yield (kg), fat (kg), protein (kg) and dry extract (kg). All environmental (farm and parturition year) and animal-inherent factors (genotype, birth type and age) resulted statistically significant ( $p < .05$ ) except for birth season and month. CSN1S1 genotype was highly statistically significant and explained from 8.3% to 9.2% of protein and fat content variability, resembling the values for highly selected French breeds. Seasonal peaks and lows resembled other breeds'. Heterozygote advantage of certain combinations of E allele with those alleles strongly or weakly influencing milk components and yield such as A, B, B2, F and homozygote BB genotype reported the highest statistically significant effects on milk components and yield. Our results suggest that non-parametric tests may report contextually valid results when having a large sample size is not possible. Selecting for certain CSN1S1 genotypes may promote the efficient production of better-quality milk in greater amounts, improving the international competitiveness and profitability of local breeds.

### HIGHLIGHTS

- Non-parametric tests are crucial if normality and heteroskedasticity analyses fail.
- Murciano-Granadina milk traits compared with highly selected international breeds'.
- E allele combinations and BB reported highest effects on milk components and yield.

### ARTICLE HISTORY

Received 22 November 2018  
Revised 1 April 2019  
Accepted 17 April 2019

### KEYWORDS

Dairy goats; environmental factor; categorical regression; casein genotype; parturition

## Introduction

Among non-genetic factors, those linked to environment strongly impact on the survival and productivity of dairy goats. These factors not only affect productivity at a quantitative level but also affect the composition and, therefore, the quality of goat milk-derived products (milk, cheese, among others) (Samson and Olajumoke 2017). Contrastingly, genetic factors account for a high proportion of the high variability between individuals, within and between breeds. Such property of internal and external heterogeneity of milk yield and composition is the basis for the improvement of milk

traits in goats. This improvement is carried out selecting the does and bucks accounting for the highest yields and producing the milk with the composition of the best quality (Samson and Olajumoke 2017). However, both sets of factors are so intertwined that confounding effects can appear. Thus, the importance of determining the exact impact that can be attributed to each group of factors separately.

Among genetic factors, the influence of genetic variants for  $\alpha$ S1-casein ( $\alpha$ S1-CN) on the traits related to milk components aimed at cosmetic and cheese production is currently studied in common and new dairy

**CONTACT** Dr F. J. Navas González  [fjng87@hotmail.com](mailto:fjng87@hotmail.com)  Departamento de Genética, University of Cordoba, Córdoba, Spain

 Supplemental data for this article can be accessed [here](#).

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

species. Its relation with the profitability of their milk still concerns the experts, despite the first advances date back to decades ago (Hristov et al. 2018).

Research on the polymorphisms of the genes involved in  $\alpha_{S1}$ -CN production has traditionally suggested that the 18 alleles possibly comprising the  $\alpha_{S1}$ -CN loci (CSN1S1) can strongly (A, B, C, H, L, M), intermediately (E and I) or weakly (D, F and G) affect the components present in the milk of small ruminants, especially goats (Giorgio et al. 2018). By contrast, other alleles, such as O1, O2 and N have been reported to be related to no amount of  $\alpha_{S1}$ -CN (Bonanno et al. 2013). Strong alleles are associated with higher protein, fat and casein contents, whereas no effect has been reported on milk production. This provides a basis for studies stating that cheese yield of goat milk is promoted by the combination of certain alleles within CSN1S1, such as AA, in a range from 8.2% to 17.7%, when compared to the rest of possible allelic combinations (Vassal et al. 1994).

According to Delgado et al. (2017), Murciano-Grandina bucks are routinely tested for A, B, E, F and N  $\alpha_{S1}$ -CN alleles aiming at selecting for better quality milk producers. Goats with strong alleles have a greater ability to synthesise  $\alpha_{S1}$ -CN than goats with weak alleles. They also produce milk higher in casein, fat, calcium, and phosphorus, providing goat milk-based products with better properties (smaller casein micelles and higher coagulation time ( $r$ ) and curd firmness ( $a_{30}$ ) (Pal et al. 2017). The outstanding performance of the allelic combinations mentioned above may stem on physiological better capabilities to utilise energy and dietary protein more efficiently when feeding the animals on highly nutritive diets (Bonanno et al. 2013). However, the study of non-genetic and genetic factors exerting their influence on the yield and properties of goat milk is a constantly developing topic as still the intrinsic mechanisms that drive them have not been deeply isolated and are not completely known yet.

The use of parametric tests is routine practice when studying traits related to milk production. However, the assumptions for such tests are not met or at least, not reported in most of the studies. This study aims to assess the influence of non-genetic factors, such as farm, parturition year, parturition month, birth season and birth type on the milk yield and milk components of Murciano-Granadina goats, isolating and overcoming the physiological response of the animals to be expected considering their  $\alpha_{S1}$ -casein genotype. We used non-parametric categorical tests to analyse the power to explain the variability found in the

population of a model comprising all of them and issued several specific regression equations to predict which and at what level these factors may determine goat milk yield (kg), fat (kg), protein (kg) and dry extract content (kg).

## Material and methods

### First stage: study premises

At a first stage, we assessed the statistical properties of the dependent variables of milk yield, fat, protein and dry extract components in kg from the whole routine milk recording of Murciano-Granadina goat breed until 2017 ( $n=2,359,479$  records from 151,997 goats). We submitted this whole set to a depuration process and deleted the observations from the animals whose productive registries were outside the reference values for milk yield and content of protein, fat and dry extract for the Murciano-Granadina goat breed and tested for the assumptions of non-normality, heteroskedasticity, sphericity, multicollinearity and existence of outliers with the raw data (phenotypes). Then, we adjusted the phenotypes found in the population grouping the females at the same parturition number considering parturition number as the grouping criterion for each of the fixed effects considered in our model (farm, birth year, birth month, birth season and birth type) and tested for the assumption of normality again as a normal distribution of data could be expected when assessing the data properties from the goats of the same lactational stage. To fulfil this aim, we used the split file routine of SPSS Statistics for Windows, Version 24.0, IBM Corp. (2016) (SPSS, Chicago, IL). The two procedures (on the whole and on the adjusted data, respectively) reported that all assumptions had been violated, suggesting our data should follow a non-parametric approach.

Then, a smaller set of animals ( $n=2090$  records from 710 goats) was selected and genotyped for  $\alpha_{S1}$ -Casein to assess for the effect of genotype. To select the animals comprising our sample, we carried out univariate and bivariate analyses and mixed model procedures using an Animal Model (BLUP) by Restricted Maximum Likelihood, with the MTDFREML software package (Boldman et al. 1995), iterating until a convergence criterion of  $10^{-12}$  was obtained. The analyses were run including the relationship matrix of animals with direct records related through at least one known ancestor. This matrix comprised the 151,997 goats in the historical pedigree updated until 2017. After convergence was reached, we directly estimated predicted breeding values using the MTDFREML software.

To select the sample of goats that would be genotyped for milk production and components, we assessed the combination of three of the four traits measured through combined selection index (ICO) procedures as suggested by Van Vleck (1993). Dry extract was excluded from the combined index (ICO) given it provides an indirect measurement of fat and protein content. We based on the estimated phenotypic relationship between each of the three traits to quantify their weight when the breeding goal was milk production and quality. Most milk marketing orders employ a multiple component pricing system that pays producers on the basis of milk fat, protein and other dairy solids. This pricing method derives component values from prices for manufactured dairy products (cheese, butter, non-fat dry milk and dry whey), which rise and fall with changing market conditions. As a result, milk component levels are rather more important factors in herd management than milk yield itself. In matrix notation, the weights to be applied on the selection index combining the partial scores of each modality were obtained as  $b = P^{-1}g$ , where  $b$  is the vector of weights to be applied to each production or content trait,  $P$  is the phenotypic (co) variance matrix and  $g$  is the vector of genetic (co)variances of every trait with each other. MatLab r2015a (The MathWorks, Inc. 2015) was used to compute all selection index combinations. As a result and considering the market demands, the weights for milk yield, fat and protein followed the proportion of 1: 1: 1, respectively. The combined index used (ICO) could be represented by the following equation:

$$ICO = \frac{PBV_{\text{milk yield}}W_1}{\mu_{\text{milk yield}}} + \frac{PBV_{\text{fat}}W_2}{\mu_{\text{fat}}} + \frac{PBV_{\text{protein}}W_3}{\mu_{\text{protein}}},$$

where PBV is the predicted breeding value for each of the traits and animals included in the matrix;  $W_1$  is the weight for milk yield,  $W_2$  for fat and  $W_3$  for protein in kg and normalised at 210 d; and  $\mu$  the mean for each of the traits included in the ICO computed in kg and at 210 d. After ICO was computed for each of the animals included in the matrix, we sorted a total of 710 animals from the whole routine milk recording of Murciano-Granadina goat breed in a ranking considering their ICO value obtained at the previous genetic evaluation. The quest for the animals producing the highest milk quantity of the highest possible quality and the genotypes responsible for such characteristics may be the sacrifice of other important traits which may be necessary for profit and/or productivity. Animals with extreme PBVs may be less efficient and less balanced than we could expect at first. Furthermore, not all traits are

affected to the same degree by selection for these extremes. For these reasons, we chose 236 females presenting the lowest ICO values in the rank, 238 females with values around percentile 50, and the 236 females presenting the highest ICO values in the rank, so as to perform an adjusted representative sampling of the genotype distribution in the population. These goats belonged to 59 different farms and all the records were collected at different periods from January 2005 to October 2016. To design the model used at the previous genetic evaluation, the authors chose a routine process including the effects commonly reported by literature concerning milk yield, and protein, fat and dry extract content, and using the same common parametric statistical tools to analyse them (Brito et al. 2011; Samson and Olajumoke 2017). This way, and as the model and methods had been tested relying on information provided by parametric tools, we could compare and contrast the information derived from the application of non-parametric tests on our data comprising commonly parametrically studied variables.

### **Animal sample**

After sample selection, our study considered direct observations from 710 Murciano-Granadina studbook registered entire lactating goats, from 14.03 to 130.13 months of age. The age range was not normally distributed ( $p < .001$  for Shapiro–Wilk Francia's tests for normality of a sample of up to 5000 cases). The minimum age in the range was 14.03 months, the Q1 age was 17.70 months, the median age was 28.77 months, the Q3 age was 41.10 months, and the maximum age was 130.13 months. The distribution of parturition age within the same parturition number was checked as well using Shapiro–Wilk Francia's W test, Kurtosis and Skewness what revealed a normal distribution, with a  $p > .05$  for Shapiro Wilk Francia's W test, a value around 3 for Kurtosis or a value in the range from  $-0.5$  to  $0.5$  for skewness.

### **Standardising milk yield: lactation typification**

In general terms, the management policies applied by Murciano-Granadina farmers base on two kidding seasons per year supported in the polyestric capacity of the breed, which gives interesting profits for lactation periods that do not last longer than 210–240 d (Delgado et al. 2017). In this study, the total milk production, milk yield, and composition were estimated

until 210 days of lactation adapting the procedures in Brito et al. (2011) and expressed in kg.

To quantify the milk yield of each goat, we computed real production ( $RP_j$ ) after the following expression:

$$RP_j = d_1 P_1 + 30 \sum_{i=n}^{n_j-1} P_i + [d_2 - 30(n_j - 2)] P_{n_j}$$

where  $RP_j$  is the real production of goat  $j$ ;  $P_1$  is milk production in the first control;  $n$  is the total number of controls;  $P_i$  is milk production in control  $i$ ,  $P_n$  is milk production in the last control.

Official controls were implemented each 4 weeks (30 d) alternating morning and afternoon controls following the AT4 procedure described by the Royal Decree Law 368/2005, of 8 April 2005, which regulates the official control of the milk yield for the genetic evaluation in the bovine, ovine and caprine species of the Spanish Ministry of Agriculture (2005), except for the first control and the last which were assessed individually for each goat. We used this information to compute the days ( $d_1$ ) between birthdate (BD) and the date when the first control (FC) took place, through the following formula:

$$d_1 = FC - BD$$

and the days between the penultimate control (PC) and the last control (LC):

$$d_2 = LC - PC$$

Then, in an attempt to save the differences between goats that may presumably be owed to differences in milking period among other factors, we considered birthdate records and the date at which several controls took place until the goats reached 210 d of lactation to normalise the milk production for each goat.

We determined normalised milk production per goat at 210 d through the following formula:

$$NP_j = d_1 P_1 + A + B$$

where  $NP_j$  is the normalised production for goat  $j$ :

$$A = 30 \sum_{i=1}^{n_j-2} \frac{P_i P_j + 1}{2}$$

$$B = [d_2 - 30(n_j - 2)] \frac{P_{n_j-1} + P_{n_j}}{2}$$

The model to compute normalised productions at 210 d could be represented by the following formula:

$$MP_{210} = \sum_{i=1}^{n-1} \left[ \left( \frac{pldc_i + pldc_{i+1}}{2} \right) \cdot l_{i \ i+1} \right]$$

where  $MP_{210}$  is the accumulated milk production until 210 d of lactation;  $pldc_i$  is the milk production under milk control  $i$ ;  $pldc_{i+1}$  is the milk production in the next milk control and  $l_{i \ i+1}$  is the interval in days between two consecutive controls.

### **Milk components analyses**

The total volume of milk collected at one milking was determined for each of the 710 goats after a typified lactation period of 210 d, at seven controls after kidding with a periodicity of 30 d. The collection of milk samples was performed monthly and sent to an official laboratory (Milk Quality Laboratory, Córdoba, Spain) for quantification of contents of total protein, fat and dry extract using a MilkoScan<sup>TM</sup> FT1 analyser. After purging data, we retained 2090 records from 710 goats for the statistical analyses. [Supplementary Table S1](#) shows the number of goats and observations of each CSN1S1 genotype used in this study.

### **Second stage: genotype effect determination**

At a second stage, we 'adjusted' our phenotypes and tested for the differences between genotypes, following a non-parametric approach, given the nature of the data in our smaller sample.

### **DNA bank and blood samples**

We used the information from 117,225 blood samples from Murciano-Granadina breeding billies, and goats maintained at the DNA Bank of Animal Breeding Consulting, Inc. DNA samples were the property of the National Association of Breeders of Goats of Murciano-Granadina Breed (CAPRIGRAN). For our study sample, we did not consider those animals from the studbook from which there were not blood samples available.

### **$\alpha$ S1-casein (CNS1) genotyping: exons 9 and 19**

We used the microsatellite markers that have commonly been addressed in literature for allelic type determination in different goat breeds worldwide (Caravaca et al. 2009). This allelic type corresponds to changes in the protein structure determined by different combinations of mutations of the exons 9 and 19 of the CSN1 gene. The deletions of a cytosine in the position 9888 in exon 9 and an insertion of 11 bases in position 9981 are considered to be the determinants for alleles A, B, F and N. In the same way, the

**Table 1.** Possible combinations found when determining for CSN1 allelic type.

Allelic type	9981-2		457 bp line element
	9888-90 C deletion	CCGTAATGTTT insertion	
A	No	No	N/A
F	Yes	Yes	N/A
M <sup>a</sup>	Yes	Yes	N/A
N	Yes	No	N/A
B	No	Yes	N/A
E	No	Yes	Yes <sup>b</sup>

N/A: non-applicable as we did not consider it for allelic denomination.

<sup>a</sup>Pérez MJ, Leroux C, Bonastre AS, Martin P. 1994. Occurrence of a LINE sequence in the 3' UTR of the goat  $\alpha$ s1-casein E-encoding allele associated with reduced protein synthesis level. *Gene*. 147(2):179–187.

<sup>b</sup>Bevilacqua C, Ferranti P, Garro G, Veltri C, Lagonigro R, Leroux C, Pietrola E, Addeo F, Pilla F, Chianese L. 2002. Interallelic recombination is probably responsible for the occurrence of a new  $\alpha$ s1-casein variant found in the goat species. *FEBS J*. 269(4):1293–1303.

inclusion of a LINE element of 457 bases in exon 19 determines the presence of allele E (Pérez et al. 1994). The sequences coding for the four allelic types of the CSN1 gene present in Genbank are AJ504710, AJ504711, AJ504712 and X56462, and they correspond to the four alleles that we identified F, N, A and B, respectively (Table 1). Previous statistical analyses revealed that the presence of the 457bp LINE element determined for E allele, with and without the presence of the inclusion of 11 bases described in Table 1.

### Mutation identification

We used a PCR test to seek for the alleles described by Maga et al. (2009) identifying the mutations present in our samples. We used a series of primers surrounding the variations in exons 9 and 19. Specifically, the pair of primers FF (GTATGGAAGTGTGGAATAGTTT-6FAM) and FR (TGGGGGTTGATAGCCTTGTA) amplify a fragment of 257 bases in exon 9 for the B allele and a fragment of 246 for the A allele, respectively. The N allele produces a fragment of 256 bases, while the N allele one of 245. The second set of primers, EF (CTATCATGTCAAACCATTCTATCC-6FAM), ER (CAATTTCACTTAAGGATGTTACAC) and E-EF(TCCCATTCTCCCAATCATC-HEX) produce a fragment of 133 bases if the element LINE is not present, while this fragment presents 225 bases otherwise.

We found two additional DNA combinations in exon 9 that had not been considered at first. According to our results, some individuals presented a 265 base-pairs DNA fragment and then the possibility of presenting 244 or 255 base pairs in the same exon. Bevilacqua et al. (2002), reported the same 265 base pair fragments at exon 9 and the same ten base pairs difference

between alleles (238 bp and 248 bp for M and F alleles, respectively).

### Statistical analysis

The variables in our study were numeric and continuous (milk yield (kg), fat (kg), protein (kg) and dry extract (kg)), while the factors considered in our model were categorical (farm, parturition year, parturition month, birth season, birth type and genotype). No transformation was carried out over normalised milk yield production in kg. Then, the percentage of fat, protein and dry extract were determined at the laboratory and computed by multiplying such percentages by the total number of kg of normalised milk yield at 210 d. We discarded the possibility of including inbreeding as a measure of the relatedness among animals to account for the polygenic effect on milk yield and components traits. Although additive genetic component is not a negligible component in the traits analysed, the value for mean inbreeding in the whole routine milk recording of Murciano-Granadina goat breed until 2017 ( $n=2,359,479$ ) is 0.29%, hence not relevant. We also considered that a previous study in the same breed (Deroide et al. 2016) had reported no statistically significant effect of both linear and quadratic inbreeding coefficient, hence, our decision not to include such effect in our model. Production data, i.e. milk production and components, among others, have normally been assumed to follow a normal distribution in literature. However, scientists often face challenges regarding the properties of the data obtained from field observations what promotes the assessment of new alternatives to process such data (Young 2017). All the data available in the routine milk recording of Murciano-Granadina goat breed ( $n=2,359,479$ ) were tested to check whether data violated the assumptions for regular parametric tests to report valid results. The data distribution showed to be non-normally distributed (Kolmogorov–Smirnov,  $p < .001$ ), highly positively skewed and platykurtic for milk yield, fat, protein and dry extract content expressed in kg. That is, compared to a normal distribution, its central peak was lower and broader and shifted to the right and its tails were shorter and thinner (Supplementary Table S2). Levene's test for equality of error variance reported that the error variance around predicted scores was not the same for all predicted values ( $p < .001$ ), thus, there was not homogeneity of variances for each combination of the levels of the independent variables (farm, birth year, birth month, birth season and birth type), and the

assumption of homoscedasticity was violated. Mauchly's *W* Test of Sphericity (Mauchly's  $W=0.29$ ),  $\chi^2(5) = 8,327,656.071$ ,  $p < .001$ , indicated that the variances of the differences were not equal, hence, the assumption of sphericity had been violated as well.

Despite this finding in the global population and before deciding on the nature of the tests to follow, the assumptions of normality were tested again as a normal distribution of data could be expected when assessing the data properties from the goats of the same lactational stage and all the statistical parametric assumptions were tested on our sample as well. Except for punctual evidences of normality (Kolmogorov–Smirnov,  $p < .05$ ), for goats at their 4th to 6th lactation that did not affected all the variables considered (milk yield, fat, protein and dry extract content), almost all the traits resulted non-normally distributed at any of the lactational stages considered from the 1st to the 6th lactation of the goats. Then, when our sample was considered, Mauchly's test of sphericity,  $\chi^2(5) = 27,543.024$ ,  $p < .001$ , indicated that the variances of the differences were not equal, hence, the assumption of sphericity had been violated. Shapiro–Wilk Francia's tests for normality performed with StataCorp Stata version 14.2 (StataCorp, College Station, TX) showed all the variables, highly significantly deviated from a normal distribution ( $p < .001$ ). As the limiting factor in this study was the number of animals genotyped, the reason for the small sample used, we tested for homoscedasticity between the levels of the genotype variable as well in this smaller set. Levene's test for equality of error variance reported a significant value for almost all variables ( $p < .05$ ) thus, the variances of the samples considering all the factor computed were unequal.

The existence of outliers could be a cause for the lack of normality of the variables studied in our population. However, almost all these significant univariate outliers fell within the range of values found in the literature for the breed. The high productivity of the Murciano-Granadina dairy breed reaches 600–700 l of milk in the livestock registered in the studbook at 210 d of lactation (Delgado et al. 2017). However, it is usual to

register individuals who reach 1000–1300 litres of milk produced in 210 days of lactation in routine performance checks. Hence, we decided to retain these outliers in the analysis, as long as they did not exceed the values found in nature for the breed, to assess what influence they may exert in the estimation process of the models considered in our study. As the data were non-normally distributed, we used Spearman rho to test for multicollinearity. The Spearman rho ( $\rho$ ) correlation reported highly statistical multicollinearity among all the variables ( $p < .01$ ). Hence, we used non-parametric tests to statistically assess the information recorded. As the factors and variables in the model did not fit to a normal distribution ( $p < .001$ ), a Kruskal–Wallis H was performed to study the potentially existing differences between levels of the same factor. We show a summary of the levels for all the factors included in the model in Table 2.

Then after conducting a Kruskal–Wallis H with three or more groups ( $k$ ), we computed the strength effect of the factors on the variables tested. Kruskal–Wallis H produces Chi-square values with  $k - 1$  degrees of freedom. We can transform  $\chi^2$  into an  $F$  value with  $k - 1$  numerator degrees of freedom (dfn) and  $N - k$  denominator degrees of freedom (dfd) using the expression  $F(\text{dfn}, \text{dfd}) = \chi^2 / (k - 1)$ , modified from Murphy et al. (2014). From  $F(\text{dfn}, \text{dfd})$ , we calculate partial eta squared (Lakens 2013) from the following equation  $\eta^2 = (F \cdot \text{dfn}) / (F \cdot \text{dfn} + \text{dfd})$ . Afterwards, we studied the pairwise comparisons for any dependent variables for which the Kruskal–Wallis test is significant aiming at assessing whether there are the differences between groups of the same factor affecting them using Dunn's test. In multiple comparisons, the risk of obtaining Statistical hypothesis testing bases on rejecting the null hypothesis if the likelihood of the observed data under the null hypotheses is low. If we test for multiple comparisons (hypotheses), the likelihood of incorrectly rejecting a null hypothesis increases, that is rejecting the existence of statistically significant differences between two or more groups (i.e. making a Type I error). The Bonferroni correction compensates for that increase by testing each hypothesis (comparison) at a significance

**Table 2.** Category description for environmental factors considered.

Factor	Type	Levels
Farm	Nominal	59 farms
Parturition year	Ordinal	2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016.
Parturition month	Ordinal	January, February, March, April, May, June, July, August, September, October, November, and December
Birth season	Ordinal	Autumn, Winter, Summer, and Spring
Birth type	Nominal	1–5 kids
Genotype	Nominal	AA, BA, BB, AE, AE2, BE, EM, EF
Age	Numeric	Minimum age in the range was 14.03 months, $Q_1$ age was 17.70 months, median age was 28.77 months, $Q_3$ age was 41.10 months, and maximum age was 130.13 months

level of  $\alpha/m$ , where  $\alpha$  is the desired overall alpha level, and  $m$  is the number of hypotheses.

As all the variables had been previously reported to be non-normally distributed (Shapiro–Wilk Francia's tests ( $p < .001$ ), an independent-sample median test was carried out to assess the differences in the median between categories (levels) within the same factor. Shapiro–Wilk Francia's tests were carried out with the .sfrancia routine of StataCorp Stata version 14.2 (StataCorp, College Station, TX). All non-parametric tests were carried out using the independent samples package from the non-parametrical task of SPSS Statistics for Windows, Version 24.0, IBM Corp. (2016) (IBM SPSS Statistics, Armonk, NY). Categorical regression (CATREG) on the data was used to describe how our variables linearly depended on the predictors (factors) considered. The resulting regression equation can be used to predict milk yield (kg), fat (kg), protein (kg) and dry extract (kg) for any combination of the seven independent factors (Table 2). We performed a categorical regression with the Optimal Scaling procedure from the Regression task from SPSS Statistics for Windows, Version 24.0, IBM Corp. (2016) (IBM SPSS Statistics, Armonk, NY).

### Ethical approval statement

The study was conducted in accordance with the Declaration of Helsinki. The Spanish Ministry of Economy and Competitiveness through the Royal Decree-Law

53/2013 and its credited entity the Ethics Committee of Animal Experimentation from the University of Córdoba permitted the application of the protocols present in this study as cited in the 5th section of its 2nd article, as the animals assessed were used for credited zootecnical use. This national Decree follows the European Union Directive 2010/63/UE, from the 22 September 2010.

### Results

The percentage ( $\mu \pm SD$ ) of fat in our milk samples was of  $5.213\% \pm 1.021\%$ ; for protein percentage,  $3.544\% \pm 0.557\%$ ; and dry extract percentage  $14.308\% \pm 2.042\%$ , respectively. We report the results from Kruskal–Wallis H for all the variables and levels considered in the study and partial eta squared ( $\eta^2$ ) as a measure of the strength of the factors the variables tested (Table 3).

Our model design is a MANOVA, which by definition involves non-independent or repeated measures. When reporting a certain partial eta<sup>2</sup> value, we numerically express that the effect for group differences in our MANOVA accounts for the values of such differences plus their associated error variance, to then turn it into a percentage. Then, univariate follow-up tests must be carried to further isolate where we can find the significant and interesting mean differences exactly. Literature recommends the use of partial eta square instead of classical eta square when using a

**Table 3.** Summary of the results of the Kruskal–Wallis  $H$  test and the determinative coefficient (through partial eta squared ( $\eta^2$ )) on milk yield (kg), fat (kg), protein (kg) and dry extract (kg) in Murciano-Granadina goats.

Factor	Variable	$\chi^2$	dfn	Sig.	dfd	F	Partial eta squared
Farm	Milk yield (kg)	537.742	58	0.000	2031	9.27	0.206
	Fat (kg)	449.630	58	0.000	2031	7.75	0.178
	Protein (kg)	457.883	58	0.000	2031	7.90	0.181
	Dry extract (kg)	455.046	58	0.000	2031	7.85	0.180
Parturition year	Milk yield (kg)	115.111	11	0.000	2078	10.47	0.052
	Fat (kg)	89.131	11	0.000	2078	8.10	0.040
	Protein (kg)	139.802	11	0.000	2078	12.71	0.062
	Dry extract (kg)	111.832	11	0.000	2078	10.17	0.050
Parturition month	Milk yield (kg)	33.540	11	0.000	2078	3.05	0.016
	Fat (kg)	40.011	11	0.000	2078	3.64	0.019
	Protein (kg)	26.656	11	0.005	2078	2.42	0.012
	Dry extract (kg)	27.409	11	0.004	2078	2.49	0.013
Birth season	Milk yield (kg)	20.104	3	0.000	2086	6.70	0.009
	Fat (kg)	13.210	3	0.004	2086	4.40	0.006
	Protein (kg)	10.377	3	0.016	2086	3.46	0.005
	Dry extract (kg)	9.104	3	0.028	2086	3.04	0.004
Birth type	Milk yield (kg)	122.248	4	0.000	2085	30.56	0.054
	Fat (kg)	104.782	4	0.000	2085	26.20	0.047
	Protein (kg)	116.294	4	0.000	2085	29.07	0.052
	Dry extract (kg)	110.702	4	0.000	2085	27.68	0.050
Genotype	Milk yield (kg)	16.745	7	0.019	2082	2.39	0.008
	Fat (kg)	19.574	7	0.007	2082	2.80	0.009
	Protein (kg)	17.120	7	0.017	2082	2.45	0.008
	Dry extract (kg)	17.437	7	0.015	2082	2.49	0.008
Age	Milk yield (kg)	330.966	92	0.000	1997	3.60	0.140
	Fat (kg)	323.539	92	0.000	1997	3.52	0.137
	Protein (kg)	326.967	92	0.000	1997	3.55	0.138
	Dry extract (kg)	325.757	92	0.000	1997	3.54	0.138

dfn: degrees of freedom numerator; Sig.: Significance; dfd: degrees of freedom denominator.

multifactor design. The reason for this is that, through the use of partial eta square, we report an index of the strength of association between an independent variable and a dependent variable that excludes the variance produced by other variables (Brown 2008). Because the Kruskal–Wallis test statistic bases on a single independent factor entered into analysis and consequently, no other variable accounts for variance in the dependent variable,  $\eta p^2$  equals  $\eta^2$ . This property is especially relevant when testing empirical variables among which there is not any possibility of overlapping, such as those in our study. We show Kruskal–Wallis H frequencies of animals over the median for all the levels of the factors affecting milk yield (kg), fat (kg), protein (kg) and dry extract (kg) in [Supplementary Table S3](#). [Supplementary Table S4](#) shows the results for Dunn tests' pairwise comparison between the different levels of the factors and variables. [Supplementary Table S5](#) shows a summary of the results for independent samples median test, cross-comparing the significant results reported by the Dunn's Test for each pairwise comparison of two levels within a certain factor with the medians obtained for each of the levels considered.

CATREG was performed to the six qualitative independent factors (farm, parturition year, parturition month, birth season, birth type and genotype) and the numerical independent covariable of the age with the four continuous variables related to milk production and composition (milk yield, fat, protein and dry extract, all expressed in kg) as dependent variables. Then, stepwise linear regression to the data with the resulted quantifications was applied, and we present the summary results with the significant variables in [Table 4](#). The standardised coefficients ( $\beta$ ) are listed in [Table 4](#) as well. CATREG reported all of the independent variables except for parturition month and birth season to be significant for all the variables tested.

Farm factor explained a moderate percentage of variance (43.8–48.1%) for the four variables tested according to CATREG standardised coefficients. This percentage ranged from 15.4 to 17.3% for age, from

12.6 to 14.5% for birth type, from 8.3 to 9.1% for genotype and from 5.7 to 10.9% for parturition year. All the coefficients for the factors were significant and positive. CATREG R squared coefficient ranged from 28.7 to 33.3% for the fat (kg) and milk yield (kg) variables, respectively ([Table 5](#)). We used the stepwise method to avoid multicollinearity. The standardised solutions for the regression equations can be found in [Table 6](#).

## Discussion

As the necessary assumptions to run parametric tests were violated (non-normality, heteroskedasticity, sphericity, multicollinearity and existence of outliers), we implemented non-parametric tests. These data properties could have their base on the fact that we work with a population under selection. Selection pressure promotes the erosion of the left tail of the distribution, as goats presenting a low performance are discarded even before being considered in the database, what alters the properties of the distribution of the data from the start. Based on this selective context, we could address the imbalance in favour of between-farms variability when compared to within-farm variability as one of the reasons for the distortion of distribution properties. Our sample presents a between-farm variability that ranges from 17.8% to 20.6% for fat and milk yield, respectively. A similar effect power of farms on the yield and quality of the milk of 6 breeds of goats has been reported to range from around 20% to 45% (Vacca et al. 2018). As reported by Nolan and Heinzen (2017), as the sample size increases, a normal distribution of scores will more closely resemble a normal curve, hence reducing the effects of violating the normality

**Table 5.** Model summary of stepwise linear regression with transformed variables.

Variable	$R^2$	Adjusted $R^2$	Sig.
Milk yield (kg)	0.333	0.308	0.000
Fat (kg)	0.287	0.261	0.000
Protein (kg)	0.295	0.268	0.000
Dry extract (kg)	0.289	0.263	0.000

Sig.: Significance.

**Table 4.** Standardised coefficients and significance of CATREG model.

Variable Parameter factor	Milk yield (kg)		Fat (kg)		Protein (kg)		Dry extract (kg)	
	Standardised Coefficients ( $\beta$ )	Significance						
Farm	0.481	0.000	0.442	0.000	0.438	0.000	0.442	0.000
Parturition year	0.057	0.000	0.069	0.000	0.109	0.000	0.083	0.000
Parturition month	-0.027	0.951	0.078	0.354	0.051	0.539	0.041	0.951
Birth season	-0.058	0.365	0.023	0.737	0.032	0.825	0.038	0.365
Birth type	0.145	0.000	0.126	0.000	0.133	0.000	0.127	0.000
Genotype	0.091	0.000	0.092	0.000	0.083	0.000	0.091	0.000
Age	0.173	0.000	0.159	0.000	0.154	0.000	0.164	0.000

**Table 6.** Regression equations for the milk production and milk content variables tested.

General model regression equation		Legend
$Zy_{mf\text{pde}} = \beta_{\text{farm}} * Z_{\text{farm}} + \beta_{\text{parturitionyear}} * Z_{\text{parturitionyear}} + \beta_{\text{parturitionmonth}} * Z_{\text{parturitionmonth}}^{(a)} + \beta_{\text{birthseason}} * Z_{\text{birthseason}}^{(a)} + \beta_{\text{birthtype}} * Z_{\text{birthtype}} + \beta_{\text{genotype}} * Z_{\text{genotype}} + \beta_{\text{age}} * Z_{\text{age}}$		$Zy_{mf\text{pde}}$ = Z score for each milk production or content variable (milk yield, fat, protein and dry extract) $\beta$ = standardised coefficient for each of the factors appearing in the subindex. $Z$ = Z score for each of the factors appearing in the subindex Legend $Zy_m$ = Z score for milk yield variable. $\beta_{\text{Farm}}Z_{\text{Farm}}=0.481(Z_{\text{Farm}})$ $\beta_{\text{Parturitionyear}}Z_{\text{Parturitionyear}}=0.057(Z_{\text{Parturitionyear}})$ $\beta_{\text{Birthtype}}Z_{\text{Birthtype}}=0.145(Z_{\text{Birthtype}})$ $\beta_{\text{Genotype}}Z_{\text{Genotype}}=0.091(Z_{\text{Genotype}})$ $\beta_{\text{Age}}Z_{\text{Age}}=0.173(Z_{\text{Age}})$
Specific regression equations		
Milk yield (kg)	$Z'y_m = 0.481(Z_{\text{farm}}) + 0.057(Z_{\text{parturitionyear}}) + 0.145(Z_{\text{birthtype}}) + 0.091(Z_{\text{genotype}}) + 0.173(Z_{\text{age}})$	$Zy_f$ = Z score for fat content variable. $\beta_{\text{Farm}}Z_{\text{Farm}}=0.442(Z_{\text{Farm}})$ $\beta_{\text{Parturitionyear}}Z_{\text{Parturitionyear}}=0.069(Z_{\text{Parturitionyear}})$ $\beta_{\text{Birthtype}}Z_{\text{Birthtype}}=0.126(Z_{\text{Birthtype}})$ $\beta_{\text{Genotype}}Z_{\text{Genotype}}=0.092(Z_{\text{Genotype}})$ $\beta_{\text{Age}}Z_{\text{Age}}=0.159(Z_{\text{Age}})$
Fat (kg)	$Z'y_f = 0.442(Z_{\text{farm}}) + 0.069(Z_{\text{parturitionyear}}) + 0.126(Z_{\text{birthtype}}) + 0.092(Z_{\text{genotype}}) + 0.159(Z_{\text{age}})$	$Zy_p$ = Z score for protein content variable. $\beta_{\text{Farm}}Z_{\text{Farm}}=0.438(Z_{\text{Farm}})$ $\beta_{\text{Parturitionyear}}Z_{\text{Parturitionyear}}=0.109(Z_{\text{Parturitionyear}})$ $\beta_{\text{Birthtype}}Z_{\text{Birthtype}}=0.133(Z_{\text{Birthtype}})$ $\beta_{\text{Genotype}}Z_{\text{Genotype}}=0.083(Z_{\text{Genotype}})$ $\beta_{\text{Age}}Z_{\text{Age}}=0.154(Z_{\text{Age}})$
Protein (kg)	$Z'y_p = 0.438(Z_{\text{farm}}) + 0.109(Z_{\text{parturitionyear}}) + 0.133(Z_{\text{birthtype}}) + 0.083(Z_{\text{genotype}}) + 0.154(Z_{\text{age}})$	$Zy_{\text{de}}$ = Z score for dry extract variable. $\beta_{\text{Farm}}Z_{\text{Farm}}=0.442(Z_{\text{Farm}})$ $\beta_{\text{Parturitionyear}}Z_{\text{Parturitionyear}}=0.083(Z_{\text{Parturitionyear}})$ $\beta_{\text{Birthtype}}Z_{\text{Birthtype}}=0.127(Z_{\text{Birthtype}})$ $\beta_{\text{Genotype}}Z_{\text{Genotype}}=0.091(Z_{\text{Genotype}})$ $\beta_{\text{Age}}Z_{\text{Age}}=0.164(Z_{\text{Age}})$
Dry extract (kg)	$Z'y_{\text{de}} = 0.442(Z_{\text{farm}}) + 0.083(Z_{\text{parturitionyear}}) + 0.127(Z_{\text{birthtype}}) + 0.091(Z_{\text{genotype}}) + 0.164(Z_{\text{age}})$	

Birth month<sup>(a)</sup> and birth season<sup>(a)</sup> were statistically non-significant ( $p > .05$ ).

assumption. Increasing sample size avoids misleading statistics if an outlier is present, reports a more accurate value for the mean of a certain trait in a population, reduces the margin of error, and increase the power to detect differences. However, our results suggest that if we apply the proper non-parametric statistical tests when testing for common parametric assumptions fails, the sample size needed to obtain a similar statistical power is smaller. Despite this could seem contradictory, these results agree with those reported by Yang and Huck (2010), who would state that although when the assumption of normality is unsound, the parametric statistic is robust, and slightly outperforms the non-parametric statistic in terms of type I error rate and power. However, when our data do not meet the assumption of homogeneous covariance matrices, i.e. our sample presents a heteroskedastic nature, the non-parametric approach has a lower type I error rate and higher power than the most robust parametric statistic, what renders the results reported valid for such cases in which having a larger sample size is not possible. Furthermore, for repeated measure variance analyses Yang and Huck (2010) explained violations of the assumption of sphericity can seriously affect results, as the computed  $F$  ratios for the interaction and the main

effect of the repeated measures may be positively biased (too large) when sphericity does not hold.

Supplementary Table S6 shows that the descriptive statistics for all the traits considered were within the standards found for other goat breeds in the species (Brito et al. 2011). The potential of the Murciano-Granadina goat breed based on its very high levels of milk productivity and quality, reaching a maximum of 1124.173 kg of milk in our sample at 210 d of lactation. In routine performance checks, it is usual to register individuals who reach 1091–1291 kg (1 milk litre = 1.030 kg) of milk at 210 d of lactation.

In our sample, we registered average rates in milk of 3.584% protein, 5.279% fat and a dry extract of 14.454%. These are slightly higher percentages than those found for Saanen or Anglo-Nubian dairy goats (Damián et al. 2008), and similar to those for the Jamnapari Koplo breed (Jaafar et al. 2018). Deroide et al. (2016) found slightly lower average values for both milk yield and milk components and moderately lower variation coefficients for Murciano-Granadina goats.

The adjusted  $R^2$  for the categorical regression model ranged from 0.261 to 0.308, for fat and milk yield. This finding implies almost around 30% of the variance of

milk yield and of all the components can be explained by the optimally scaled and transformed factors. The  $F$ -statistic values ( $\alpha=0.05$ ) and the  $p$  values of 0.000, together indicate the adequate performance of the model. Although  $R^2$  reported slightly higher results (0.287 and 0.333, for fat and milk yield, respectively), it is better to report adjusted  $R^2$  as it is an unbiased estimator that corrects for the sample size and numbers of coefficients estimated. Adjusted  $R^2$  is always smaller than  $R^2$ , but the difference is usually very small unless you are trying to estimate too many coefficients from a proportionally small sample. This suggests the number of factors included in the model was correct. Similar results have been reported for highly selected French and Norwegian goat breeds (Carillier-Jacquín et al. 2016). Even though the authors did not report any assumption in the study, the sample size that those authors used was more than four times the sample size used in our study. Kruskal–Wallis  $H$  test and partial eta squared ( $\eta p^2$ ) revealed that the effects of farm, parturition year, parturition month, birth season, birth type, genotype and age were highly statistically significant ( $p < .01$ ) for all the traits except for birth season on protein and dry extract and genotype on milk yield, protein and dry extract, for which the effect was just significant ( $p < .05$ ).

The most powerful effects were those reported for farm and age. In the case of the farm, some authors have ascribed this stronger effect to climate variations at a certain location, the nutritional quality of food provided to animals and herd composition (Jaafar et al. 2018). Unlike in the study by Jaafar et al. (2018), we assess a single breed, so that, the differences in the composition of goat milk samples between farms could be rather determined by factors such as the diet given to goats at first instance (Supplementary Table S5). Simultaneously, our CATREG standardised coefficients ( $\beta$ ) for the effect of farms and age were positive, moderate to high and statistically significant for milk yield and all components. The magnitude and order of magnitude of the standardised coefficients ( $\beta$  coefficients), for each of our predictors, were close to the bivariate correlations (indirectly studied through partial eta squared, see Table 3) for each predictor (independent variables) with the outcome (dependent variables), what confirmed the soundness of our findings.

Samson and Olajumoke (2017) reported that there was a significant effect of age or parity on milk yield in goats, suggesting that milk production tends to increase with parity probably due to the increase in the accumulation of mammary alveoli from the previous lactation until the process gets interrupted

by advances in age. The standardised coefficients obtained in our study (0.154–0.173) support such suggestion, reporting a moderate effect of age on goat milk yield and milk components. Samson and Olajumoke (2017) did not report how the parametric assumptions were approached.

Even though we work on a single breed, other factors inherent to the animal may condition milk yield and milk constituents. For example, genotype was significant for all components, and its effect strength was 0.8% for milk yield and all components, except for fat (0.9%), suggesting the allelic profile of the animals affects milk production and composition, an important factor to be considered in the genetic evaluation of those traits. Caravaca et al. (2009) reported a non-significant effect of  $\alpha S1$ -casein genotype on milk yield and components in Murciano-Granadina goats, strongly contrasting our results and the values found for highly selected French breeds (Carillier-Jacquín et al. 2016). However, such lack of significance found may base on the study having a considerably lower sample (86 goats and 316 observations against 710 goats and 2090 observations). In addition, the fact that the breed was not as selected as it is now, the reduced number of CS1N1 genotypes (3 against 8 genotypes considered) included in the study, and the less appropriate use of parametric statistical tests against the use of non-parametric test for the analysis of categorical data or when our data do not meet multiple of the parametric assumptions.  $\beta$  coefficients for the genotype effect ranged from 0.083 to 0.092, which resembles the values obtained for partial eta squared.

According to Martin and Leroux (2000), polymorphism of the caprine  $\alpha S1$  casein gene is one of the main factors that determine milk's technological properties in terms of cheese yield and curd formation as suggested by our results. This fact may be explained given the implication of the presence of certain genotypes in the production of higher milk yields and dry extract, protein or fat contents. Dunn's test (Table 7) reported significant differences for goat milk yield for BE, AE and BB genotypes, AE and BB for protein and dry extract, and for AE, EF and BB genotypes for fat. Giorgio et al. (2018) reported B and A alleles to strongly influence milk yield and constituents in goats, while E and F alleles intermediately and weakly affected the same traits respectively. Afterwards, the independent samples median test reported the existence of heterozygote advantage of certain combinations of the E allele with those alleles strongly influencing milk components and yield such as A, B and B2 and homozygote BB genotype reported the highest statistically significant effects on

**Table 7.** Summary of the significant ( $p < .05$ ) genotype pairwise comparison through Dunn's test and Bonferroni's significance correction for milk yield (kg), fat (kg), protein (kg) and dry extract (kg) in Murciano-Grandina goats.

Variable	Genotype pairwise comparison	Contrast statistic	SE	Contrast statistic deviation	Sig.	Bonferroni's adjusted sig.
Milk yield (kg)	BB-AE	-438.662	128.192	-3.422	0.001	0.017
Milk yield (kg)	BB-BE	-395.930	126.405	-3.132	0.002	0.049
Fat (kg)	BB-AE	-435.742	128.192	-3.399	0.001	0.019
Fat (kg)	BB-EF	-598.401	166.983	-3.584	0.000	0.009
Protein (kg)	BB-AE	-405.644	128.192	-3.164	0.002	0.044
Dry extract (kg)	BB-AE	-429.784	128.192	-3.353	0.001	0.022

The rest of pairwise comparisons between all possible combinations of levels included in the genotype factor were non-significant ( $p > .05$ ).

protein and fat. Concretely, the highest frequencies of animals over the median for protein content were reported by AE and BB genotypes in increasing order, while AE, EF and BB were those reporting the highest frequencies of animals over the median for fat content (Supplementary Table S5). Among the examples of heterozygote advantage reported by Hedrick (2015), milk yield was reported to be significantly promoted by heterozygote genotypes in dairy cattle, which may support our results. Kadri et al. (2014) found that a 660-kb deletion possibly including the gene RNASEH2B had substantial positive effects on milk yield and milk composition in heterozygotes. The deletion is present as a heterozygote in 13–32% of the Danish, Swedish and Finnish Red cattle, and it may be an example of the potentially superior characteristics of mutants resulting from artificial selection. Verdier-Metz et al. (2001) studied cheese yields computed as fresh yield by dividing the weight of fresh curd by the quantity of milk used for cheese-making and dry matter yield by multiplying fresh yield by the dry matter value of the moulded curd. These authors reported a wide range of variation (55–85 g/kg) in the ratio between fat and protein content of manufactured milk and that fat/protein content ratio linearly accounted for 77% of fresh yield and 87% of dry matter yield variability. This fact implies those milks reporting higher fat/protein ratio values may present a better cheese yield and curd formation performance, which according to our results, may most likely belong to individuals presenting AE hybrid allelic combinations, given their relationship with fat and protein contents over the median (21.580, 14.878 and 59.808 kg for fat, protein, and dry extract content, respectively) as suggested by our results (Supplementary Table S5).

Similarly, Carillier-Jacquelin et al. (2016) found that genotype for  $\alpha$ S1 casein presents a significant effect on milk production, fat content and protein content in French dairy goat breeds. The same authors explain that the inclusion of the effect of  $\alpha$ S1 casein in the genetic and genomic evaluations of male genotypes for the  $\alpha$ S1 casein is known to improve the accuracy

(from 6% to 27%). In the same way, in genomic evaluations performed on the phenotypes of females, as is the case in our study, the precision resulted slightly higher (from 1% to 14%) than in the genomic models that do not include the effect of  $\alpha$ S1 casein, what may translate into slightly higher accuracies and reliabilities of the estimated breeding values derived. Apart from animal inherent characteristics (such as genotype or breed), the environmental effect of factors, such as birth season, have been widely reported to affect milk yield and components in literature. According to Milewski et al. (2018), production season had a minor impact on the chemical composition of goat milk, although it significantly differentiated the chemical composition of rennet cheese produced from that milk and the health quality parameters of both products. Milk from winter feeding had a higher content of protein, whereas the chemical composition of cheese was more beneficial in winter as the content of dry matter, fat and protein were higher. These results support our findings as we only found an effect strength of 0.4–0.9% birth season on dry extract and milk yield, respectively. This strength slightly increased when we considered the month of parturition of the animals, whose effect ranged from 1.2 to 1.9%, for protein and fat content, respectively when we considered partial eta-squared, a parameter that indicates the percentage of variance in the dependent variable attributable to a particular independent variable. By contrast, CATREG standardised coefficients ( $\beta$ ) for birth season and parturition month were statistically non-significant for milk yield and all components. This suggests, these factors not only explained a very low percentage of the variance of milk yield and its components but also that one-unit changes in birth season or parturition month did not produce any significant increase/decrease in milk yield and its components. Dunn's test showed that differences in milk yield and composition between close seasons were not significant. That is to say, the highest differences were reported for seasons such as winter and summer, and spring and summer and autumn (Supplementary Table S5), what relies on

weather oscillations in the area in which the farms are located being especially marked for such seasons.

Similarly, the most significant differences between months existed between July and those months from winter to spring. These increasing/decreasing patterns have similarly been described by milk constituents in Saanen goats (Kljajevic et al. 2018), for whose milk components there was a production peak around January–February–March and a production low around summer months. These changes may have their basis on seasonal changes in blood biochemical and endocrine responses as it has been reported for Indian indigenous goat breeds (Inbaraj et al. 2018). To our knowledge, no study assessing for such differences between seasons through the use of non-parametric tests exists. However, Broucek et al. (2005) found the same differences in milk yield between seasons using a Kruskal–Wallis test for dairy cattle, reporting the negative effect of summer/autumn months against the positive effect of winter–spring months. The same authors would also agree that fat percentage was statistically higher during summer and autumn months as reported by the results of Dunn’s test carried out in our study. These results suggest the fact that when or data do not meet several parametric assumptions, a smaller sample is needed to reach the same conclusions using non-parametric tests.

Kidding type highly statistically influenced all milk components and milk yield with an effect strength that ranged from 4.7% to 5.2% for fat and milk yield, respectively. Some authors have previously verified the influence of the type of birth, simple or multiple, or the number of kids given birth on milk yield and components. These authors explained such differences by the presence of the hormone placental lactogen, progesterone and prolactin during gestation, which are mammary gland stimulants as they differ in quantity according to the type of gestation, simple or multiple. These hormones might affect milk production during lactation and simultaneous pregnancies (Carnicella et al. 2008). For these authors, goats kidding twins yielded more milk and had longer lactation ( $p < .001$ ). The results of Dunn’s test in our study reported that despite the highly statistically significant differences in birth type for almost all pairwise comparisons of the number of kids up to five, these differences were not significant when the goat gave birth to 5 or 1, or when they gave birth 2, 3 or 4 kids.

Galina et al. (1995) reported the increase of the interval of posterior kidding may rely on the increased stress that the goat experiences, as to keep the pregnancy of one or more kids there needs to be a greater

mobilisation of nutrients for gestation and lactation. Thus, when a proper recovery does not follow this special effort, detrimental effects on subsequent lactation start to appear. We could explain the lack of differences between the goats giving birth to one or five kids as normally the exceeding number of kids tends to be hand raised, thus not exerting the effect that this extra number of kids may have exerted.

## Conclusions

When studying large sample sizes is not possible, and variables do not meet critical parametric assumptions such as normality and homoskedasticity, even if we study normally parametrically approached production traits, the use of the appropriate non-parametric tests could render an alternative for the results reported to be contextually valid. The values for milk traits obtained suggest clear similarities between Murciano-Granadina goats and other highly selected international breeds. The control of certain environmental and inherent-to-animal factors such as the selection for certain genotypes may promote the efficient production of better-quality milk in greater amounts, thus becoming a key element to include in breeding programmes aiming at improving the profitability of local breeds.

## Acknowledgements

This work would not have been possible if it had not been for the support and assistance of the National Association of Breeders of Murciano-Granadina Goat Breed, Fuente Vaqueros (Spain) and the PAIDI AGR 218 research group. Especially, we would like to thank Fernández, J. Executive Secretary of the Association. This research did not receive any specific funding.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

M. G. Pizarro  <https://orcid.org/0000-0002-4980-3942>

V. Landi  <https://orcid.org/0000-0003-1385-8439>

F. J. Navas González  <https://orcid.org/0000-0002-0093-5151>

J. M. León  <http://orcid.org/0000-0001-8969-574X>

J. V. Delgado  <http://orcid.org/0000-0003-1657-8838>

## References

- Bevilacqua C, Ferranti P, Garro G, Veltri C, Lagonigro R, Leroux C, Pietrola E, Addeo F, Pilla F, Chianese L. 2002. Interallelic recombination is probably responsible for the

- occurrence of a new  $\alpha$ s1-casein variant found in the goat species. *Febs J.* 269:1293–1303.
- Boldman KG, Kriese LA, Vleck LD, Tassell CP, Kachman SD. 1995. A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances. 1st ed. Washington (DC): US Department of Agriculture, Agricultural Research Service.
- Bonanno A, Di Grigoli A, Montalbano M, Bellina V, Mazza F, Todaro M. 2013. Effects of diet on casein and fatty acid profiles of milk from goats differing in genotype for  $\alpha$ S1-casein synthesis. *Eur Food Res Technol.* 237:951–963.
- Brito L, Silva F, Melo A, Caetano G, Torres R, Rodrigues M, Menezes G. 2011. Genetic and environmental factors that influence production and quality of milk of Alpine and Saanen goats. *Genet Mol Res.* 10:3794–3802.
- Broucek J, Mihina S, Kisac P, Hanus A, Uhrincat M, Foltys V, Marencak S, Benc F. 2005. Environmental factors and progeny affecting milk yield and composition during the first lactation. *J Anim Feed Sci.* 14:461.
- Brown JD. 2008. Effect size and eta squared. *JALT Testing & Evaluation SIG News.* 12:38–43.
- Caravaca F, Carrizosa J, Urrutia B, Baena F, Jordana J, Amills M, Badaoui B, Sánchez A, Angiolillo A, Serradilla J. 2009. Short communication: effect of alphaS1-casein (CSN1S1) and kappa-casein (CSN3) genotypes on milk composition in Murciano-Granadina goats. *J Dairy Sci.* 92:2960–2964.
- Carillier-Jacquin C, Larroque H, Robert-Granié C. 2016. Including  $\alpha$  s1 casein gene information in genomic evaluations of French dairy goats. *Genet Sel Evol.* 48:54.
- Carnicella D, Dario M, Ayres MCC, Laudadio V, Dario C. 2008. The effect of diet, parity, year and number of kids on milk yield and milk composition in Maltese goat. *Small Rumin Res.* 77:71–74.
- Damián J, Sacchi I, Reginensi S, De Lima D, Bermúdez J. 2008. Cheese yield, casein fractions and major components of milk of Saanen and Anglo-Nubian dairy goats. *Arq Bras Med Vet Zoo.* 60:1564–1569.
- Delgado JV, Landi V, Barba CJ, Fernández J, Gómez MM, Camacho ME, Martínez MA, Navas FJ, León JM. 2017. Sustainable goat production in adverse environments: Volume II. Murciano-Granadina Goat: A spanish local breed ready for the challenges of the twenty-first century. Cham (Switzerland): Springer. Chapter 18; p. 205–219.
- Deroide CAS, Jacopini LA, Delgado JV, León JM, Brasil LHA, Ribeiro MN. 2016. Inbreeding depression and environmental effect on milk traits of the Murciano-Granadina goat breed. *Small Rumin Res.* 134:44–48.
- Galina MA, Silva E, Morales R, Lopez B. 1995. Reproductive performance of Mexican dairy goats under various management systems. *Small Rumin Res.* 18:249–253.
- Giorgio D, Di Trana A, Claps S. 2018. Oligosaccharides, polyamines and sphingolipids in ruminant milk. *Small Rumin Res.* 160:23–30.
- Hedrick PW. 2015. Heterozygote advantage: the effect of artificial selection in livestock and pets. *J Hered.* 106: 141–154.
- Hristov JP, Teofanova D, Georgieva A, Radoslavov G. 2018. Effect of genetic polymorphism of  $\alpha$ ±S1-casein gene on qualitative and quantitative milk traits in native Bulgarian Rhodopean cattle breed. *Genet Mol Res.* 17:1–5.
- Inbaraj S, Kundu A, De AK, Sunder J, Sejian V. 2018. Seasonal changes in blood biochemical and endocrine responses of different indigenous goat breeds of tropical island agro-ecological environment. *Biol Rhythm Res.* 49:412–421.
- Jaafar SHS, Hashim R, Hassan Z, Arifin N. 2018. A comparative study on physicochemical characteristics of raw goat milk collected from different farms in Malaysia. *Trop Life Sci Res.* 29:195–212.
- Kadri NK, Sahana G, Charlier C, Iso-Touru T, Gulbrandtsen B, Karim L, Nielsen US, Panitz F, Aamand GP, Schulman N, et al. 2014. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet.* 10:e1004049.
- Kljajevic NV, Tomasevic IB, Miloradovic ZN, Nedeljkovic A, Miocinovic JB, Jovanovic ST. 2018. Seasonal variations of Saanen goat milk composition and the impact of climatic conditions. *J Food Sci Technol.* 55:299–303.
- Lakens D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol.* 4:863.
- Maga E, Daftari P, Kültz D, Penedo M. 2009. Prevalence of alphas1-casein genotypes in American dairy goats. *J Anim Sci.* 87:3464–3469.
- Martin P, Leroux C. 2000. Le gène caprin spécifiant la caséine  $\alpha$ s1: un suspect tout désigné aux effets aussi multiples qu'inattendus. *Prod Anim, Special Issue, Molecular genetics: principles and applications in animal populations 125–132* [http://granit.jouy.inra.fr/productions-animales/2000\\_hors-serie/Prod\\_Anim\\_2000\\_hs\\_hs\\_19.pdf](http://granit.jouy.inra.fr/productions-animales/2000_hors-serie/Prod_Anim_2000_hs_hs_19.pdf) (accessed 10 October 2018)
- Milewski S, Ząbek K, Antoszkiewicz Z, Tański Z, Sobczak A. 2018. Impact of production season on the chemical composition and health properties of goat milk and rennet cheese. *Emir J Food Agric.* 30:107–114.
- Murphy KR, Myors B, Wolach A. 2014. Statistical power analysis: a simple and general model for traditional and modern hypothesis tests. London (UK): Routledge; p. 226.
- Nolan S, Heinzen T. 2017. Statistics for the behavioral sciences. New York (NY): Macmillan Learning; p. 783.
- Pal M, Dudhrejija TP, Pinto S, Brahamani D, Vijayageetha V, Reddy YK, Roy P, Chhetri A, Sarkar S, Kate P. 2017. Goat milk products and their significance. *Beverage Food World.* 44:22–25.
- Pérez MJ, Leroux C, Bonastre AS, Martin P. 1994. Occurrence of a LINE sequence in the 3' UTR of the goat  $\alpha$ s1-casein E-encoding allele associated with reduced protein synthesis level. *Gene.* 147:179–187.
- Samson T, Olajumoke O. 2017. Genetic and non-genetic factors affecting yield and milk composition in goats. *J Adv Dairy Res.* 5:175.
- Spanish Ministry of Agriculture. 2005. Real Decreto 368/2005, de 8 de abril, por el que se regula el control oficial del rendimiento lechero para la evaluación genética en las especies bovina, ovina y caprina. [Royal Decree 368/2005, of 8th April, which regulates the official control of the milk yield for the genetic evaluation in the bovine, ovine and caprine species]. Madrid: Ministerio de Agricultura, Pesca y Alimentación. BOE, núm. 97, de 23 de abril de 2005. BOE-A-2005-6564. Spanish.
- The MathWorks, Inc. 2015. 'MATLAB' release R2015a, Natick (MA): The MathWorks, Inc.

- Vacca GM, Stocco G, Dettori ML, Pira E, Bittante G, Pazzola M. 2018. Milk yield, quality, and coagulation properties of 6 breeds of goats: environmental and individual variability. *J Dairy Sci.* 101:7236–7247.
- Van Vleck LD. 1993. Selection Index and introduction to mixed model methods. Boca Raton (FL): CRC Press; p. 481.
- Vassal L, Delacroix-Buchet A, Bouillon J. 1994. Influence des variants AA, EE et FF de la caséine  $\alpha$ s1 caprine sur le rendement fromager et les caractéristiques sensorielles de fromages traditionnels: premières observations. *Le Lait.* 74:89–103.
- Verdier-Metz I, Coulon J-B, Pradel P. 2001. Relationship between milk fat and protein contents and cheese yield. *Anim Res.* 50:365–371.
- Yang H, Huck SW. 2010. The importance of attending to underlying statistical assumptions. *Newborn Infant Nurs Rev.* 10:44–49.
- Young AL. 2017. Analysis of lactation curves using probability distributions and non-linear analysis. Utah: Utah State University, State Agricultural Experiment Station.