

ESTIMATION OF GENETIC AND ENVIRONMENTAL PARAMETERS FOR TWO BINARY TRAITS WITH AN UNDERLYING CONTINUOUS SCALE- A SIMULATION STUDY

Reda Elsaid¹ and Manal Elsayed²

1- Environmental Studies and Research Institute (ESRI), Department of Sustainable Development, Menofiya University, Sadat Branch, Egypt 2- Department of Animal Production, Faculty of Agriculture, Ain Shams University, Shoubra El-Khaima, Cairo, Egypt

SUMMARY

The study objective was to compare heritability and correlations estimated from linear model with those estimated from threshold model. Two simulation programs were used, one to simulate two continuous traits with the same level of heritability (h^2) of 0.17 and another to modify these traits into binary traits. Fifteen replicates were generated with two classes of fixed effect. Each replicate was analyzed twice, once by GS algorithm and another by MTDFREML algorithm using animal model. The concerned results from each analysis were heritability and genetic and residual correlations. Bias and mean squared errors (MSE) of h^2 and genetic and residual correlations estimates were used to assess the quality of h^2 and genetic and residual correlations estimates obtained by different algorithms. For the two traits, the statistical model used included type of algorithm (MTDFREML vs Gibbs sampling), type of trait (continuous trait vs binary trait) on the bias of heritability and genetic and residual estimates plus the 2-way interaction. For bias of heritability, all these effects were significant ($p < 0.01$). For genetic correlation estimates, none of the effects and interaction were significant ($p > 0.05$) while for residual correlation estimates, all these effects were significant ($p < 0.0001$). In estimation of heritabilities for continuous traits, the GS algorithm is equivalent to MTDFREML algorithm. For binary traits, the GS algorithm is the best. GS estimates of genetic correlations were similar to MTDFREML estimates while for the residual correlation, the marginal maximum likelihood (GS) estimates are superior to the estimates from the linear model (MTDFREML).

Keywords: *Continuous traits, binary traits, heritability estimates quality, Gibbs sampling, bias, mean squared errors, genetic and environmental correlation*

INTRODUCTION

For the estimation of genetic merit for normally distributed traits, best linear unbiased prediction (BLUP) has been extensively used, because it yields the maximum likelihood estimator of the best linear predictor (Wang *et al.*, (1993). However many studies (Meijering, 1985; Sorensen *et al.*, 1995) have also used linear models to predict the genetic merit of animals for traits that are recorded with

Issued by The Egyptian Society of Animal Production

discrete categories although the nature of such data violate the assumption of normality (Luo *et al.*, 2001).

In the last decade, Templeman, (1993) reported that nonlinear approaches for analysis of discrete traits in animal breeding have been proposed as means to obtain more accurate estimates of genetic parameters and better predictors of genetic merit of candidates for selection. Mantysaari *et al.* (1991) reported that in predicting breeding values, knowledge of genetic and phenotypic parameters, i.e., heritabilities and genetic and phenotypic correlations, is essential. When the estimation is based on discrete observations using a threshold model, similar parameters on an underlying "liability" scale are required. Simulation studies have proposed that the genetic correlation estimates are not much affected by the loss of information due to the discreteness of the data (Gianola, 1982). Mantysaari *et al.* (1991) reported that in studies, where the goal is estimation of environmental effects, the estimate of the residual correlation might be more important than the heritability and the genetic correlation. They also reported that the threshold model did not show significant improvement over the much simpler method of estimation with the linear model, except in the estimation of residual correlation.

The objective of the present study was to compare estimates of heritabilities, genetic and environmental correlations estimated from a linear model to those obtained from a threshold model.

MATERIAL AND METHODS

Simulation procedure:

Two methods of simulation were used to generate samples for the present study. One is concerned with two underlying continuous response variables generation and another with changing these two continuous variables into two binary variables each with two categories 0 and 1. The first method is the Monte Carlo simulation technique using SAS (1996) with assumed mean (0) and variance (1). Analla *et al.* (1995) reported that this technique also assumed that the expected additive genetic value of progeny G_k is equal to the average genetic values of the parents [sire (S_i) and dam (D_j)] plus a deviation due to the Mendelian sampling as follows:

$$G_k = 0.5 (S_i + D_j) + X \cdot 0.5h^2\sigma_p^2 \quad \text{Model 1}$$

Where:

G_k is equal to the genetic value of an individual k , a progeny of sire (S_i) and dam (D_j),

X is a random number taken from normal distribution with mean 0 and variance 1,

h^2 is the heritability and

σ_p^2 is the phenotypic variance.

Table 1 shows values of parametric phenotypic, genetic and residual variances, genetic and residual covariances, genetic and residual correlations and heritability.

With the same level of h^2 0.17 for the two traits, 15 replicates were simulated with two levels of one fixed effect.

Each sample was categorized using a random variety from a binomial distribution (RANBIN function) with SAS (1996) to obtain the binary response variables studies. Two copies of each generated sample were obtained, the first contained the two underlying continuous variables and the second contained the two binary responses.

Table 1. The assumed parametric values of phenotypic, genetic and residual variances, genetic and residual covariances, genetic and residual correlations and heritability for the two traits

Parameter	Value of the two traits
Phenotypic variance	1.21052
Additive genetic variance	0.21052
Residual variance	1.0
Additive genetic covariance	0.10528
Residual covariance	0.30
Genetic correlation	0.50
Residual correlation	0.30
Heritability	0.17

The mean of the simulated variables was constant at 0.5 (the best mean value of the simulated traits to maintain the average of the binary traits as it is in the continuous traits). All the parameters above as indicated by Mantysaari *et al.* (1991).

Statistical analysis

Heritability estimates of the studied variables, genetic and environmental correlations were estimated for each sample obtained from the animal model using two algorithms (Multiple trait animal model program (MTDFREML) proposed by Boldman *et al.* (1995) and Gibbs Sampling program (GS) proposed by Van Tassell and Van Vleck (1995).

The linear animal model used for two continuous and binary traits was:

$$y = X\beta + Z_a a + e$$

where,

y is a vector of observations;

X is an incidence matrix for fixed effects;

β is a vector of an overall mean and parity (2 classes);

Z is an incidence matrix for random effect;

a is a vector of direct genetic effect of the animal; and

e is a vector of random errors normally and independently distributed with zero mean and variance $\sigma_e^2 I$.

Bias and mean squared errors (MSE) were calculated to estimate the correspondence between assumed and estimated values. Estimates of bias were calculated considering the sign.

$$\text{Bias} = [E(b^R) - B] \quad (\text{Neter } et al., 1985)$$

Where,

b^R is the expected value of the deviation of the biased estimator from the true parameter B.

$$\text{MSE} = E(b^R - B) = \sigma^2(b^R) + [E(b^R) - B]^2 \quad (\text{Neter } et al., 1985)$$

Where,

b^R is the expected value of the squared deviation of the biased estimator from the true parameter B.

Analysis of variance

Analysis of variance was performed to study the effect of type of algorithm (MTDFREML or GS) and type of trait (continuous or binary trait) on the bias in h^2 and correlation estimates obtained from the study samples. To analyze the correlations obtained in this study, correlation coefficients were transformed using Fisher's formula (1921) to transform r to a normal deviate Z where, $z=0.5\ln[(1+y)/(1-y)]$

The following model was adopted using SAS (1996) to analyze the bias of heritability estimates and genetic and residual correlations:

$$Y_{ijk} = \mu + a_i + t_j + (at)_{ij} + e_{ijk}$$

Where,

Y_{ijk} is the dependent variable (studied trait) of the n^{th} record in the i^{th} type of algorithm, j^{th} type of trait;

μ the overall mean of bias;

a_i the effect of the i^{th} type of algorithm, $I = 1$ (MTDFREML) to 2 (GS);

t_j the effect of the j^{th} type of trait, $j = 1$ (continuous trait) to 2 (binary trait);

$(at)_{ij}$ the effect of the interaction between the i^{th} type of algorithm and the j^{th} type of trait; and

e_{ijk} the effect of random error, associated with each observation assumed to be normally and independently distributed with 0 mean and variance σ_e^2 .

RESULTS AND DISCUSSION

Analysis of variance of bias of h^2 estimates

Table (2) shows the analysis of variance for bias of h^2 estimates for the two traits. This table shows that all main effects and interactions were significant ($p \leq 0.001$).

Table 2. Analysis of variance of bias for heritability estimates for the two traits

Source of variation	D.F.	Bias for trait no. 1		Bias for trait no. 2	
		M.S.	Pr>F	M.S.	Pr>F
Type of algorithm	1	0.0313	0.0001	0.0395	0.0001
Type of trait	1	0.0269	0.0001	0.0273	0.0002
Algorithm*Trait	1	0.0163	0.0002	0.0256	0.0003
Error	56	0.0010		0.0018	
C.V.%		599.441		904.046	
R^2		0.484		0.571	

C.V=coefficient of variation, R^2 =coefficient of determination, Pr=probability of type I error

Figures (1-4) illustrate the 2-way interaction. Figures 1 and 2 show that the GS had smaller magnitude of bias than MTDFREML whatever type of trait is and the difference between binary and continuous traits increased by using MTDFREML.

For the first trait, Figure 3 and 4 shows that the binary trait had higher MSE than continuous trait using MTDFREML while by using GS, a smaller difference in MSE between continuous and binary traits was observed. This result confirms those obtained by Mousa and Elsayed (2001) and Elsaid (2004) indicating that GS had

consistently smaller MSE than MTDFREML, due to the influence of the prior distribution of the variance components on the posterior distribution.

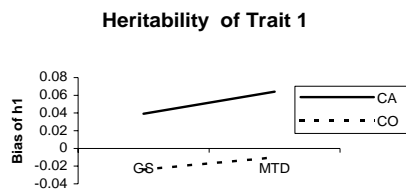


Figure 1. Algorithm type-trait type interaction, trait 1

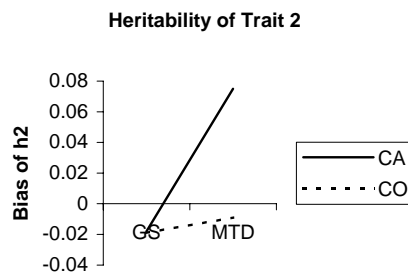


Figure 2. Algorithm type - trait type interaction, trait 2

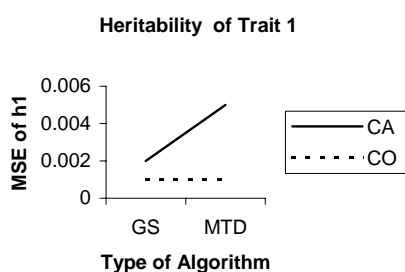


Figure 3. Algorithm type-trait type interaction, trait 1

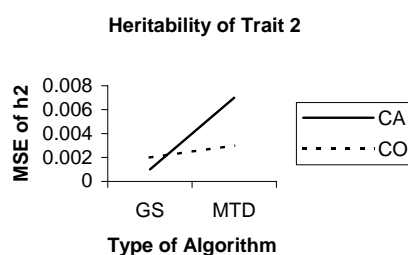


Figure 4. Algorithm type-trait type interaction, trait 2

Average h^2 estimates

The average values for the estimates of the 15 samples resulting from MTDFREML and GS algorithms for continuous and binary traits are shown in Table 3. The estimates calculated from the continuous variable values are close to the true parameters. This result is in agreement with those reported by Mousa and Elsayed (2001) and Elsaid, (2004). Standard deviations of estimates are small (0.034 to 0.039); and the values of t statistic ranged from -0.714 to -3.485. When parameter estimation is performed from the discrete scores, but assuming normality (MTDFREML) the values of h^2 clearly are underestimated (0.106 and 0.095) and the t statistics are large (6.9 to 9.8). When the binary trait was analyzed by GS, the value of t statistics approached levels close to continuous variable estimates (-1.45).

Also, Table 3 shows that for continuous trait, the magnitude of average of bias of h^2 resulting from MTDFREML was smaller than corresponding ones resulting from GS (-0.01 vs -0.02). For binary traits, the magnitude of bias for the estimates resulting from GS was smaller than those resulting from MTDFREML (-0.015 vs 0.064). This result is in line with previous works (Mantysaari *et al.*, 1991; Matos *et al.*, 1997; Boettcher *et al.*, 1999; Luo *et al.*, 2001 and Elsaid, 2004) explaining the

reason for this difference between MTDFREML and GS as the heritabilities from MTDFREML (linear model) are expressed on observed scale while heritabilities from GS (threshold model) are on an underlying liability scale. Therefore, threshold model was statistically more appropriate than linear model for binary trait (either single binary trait or multiple binary traits), yields greater estimates of h^2 and closer to the real value.

Table 3. Estimates of heritabilities from data with two continuous and binary variables

Traits	Algorithm			
	MTDFREML Continuous	MTDFREML Binary	GS Continuous	GS Binary
Trait 1				
Heritability	0.181	0.106	0.194	0.185
SD	0.034	0.025	0.027	0.039
Bias	-0.01	0.064	-0.024	-0.015
t	-1.308	9.790	-3.485	-1.454
MSE	0.001	0.005	0.001	0.002
Trait 2				
Heritability	0.179	0.095	0.189	0.188
SD	0.034	0.042	0.044	0.027
Bias	-0.009	0.075	-0.019	-0.018
t	-0.714	6.902	-1.684	-2.553
MSE	0.003	0.007	0.002	0.001

Also, the MSE for continuous trait resulting from MTDFREML is equivalent to those resulting from GS for first trait (0.001 vs 0.001) while for the second trait, the MSE resulting from GS were smaller than those resulting from MTDFREML (0.002 vs 0.003). For binary trait, the MSE resulting from GS were smaller than those resulting from MTDFREML for each trait (0.002 vs 0.005 and 0.001 vs 0.007).

Analysis of variance of genetic and residual correlations

Table 4 shows the analysis of variance of genetic and residual correlations. This table shows that for genetic correlation, all main effects and interaction were not significant ($P>0.05$) while for the residual correlation, all main effects and the interaction were significant ($P\leq 0.0001$).

Table 4. Analysis of variance of genetic and residual correlation estimates for two traits

Source of variation	D.F.	Genetic correlation		Residual correlation	
		M.S.	Pr	M.S.	Pr
Type of algorithm	1	0.00170	0.756	0.1045	0.0001
Type of trait	1	0.00002	0.974	0.1081	0.0001
Algorithm*Trait	1	0.00020	0.917	0.0783	0.0001
Error	56	0.01714		0.0015	
C.V.%		22.9		14.163	
R ²		0.002		0.78	

C.V=coefficient of variation. R²=coefficient of determination

Pr=probability of type I error

Figure (5) shows that the genetic correlation estimates of continuous traits are equivalent to those of the binary traits with MTDFREML while a small difference between continuous and binary traits was observed with GS. Also, this table shows small difference in genetic correlation estimates between MTDFREML and GS. This is in agreement with Mantysaari *et al.* (1991) who reported that marginal maximum likelihood estimates of genetic correlations were similar to linear model estimates.

Figure (6) shows that with GS, the estimates of residual correlation for continuous traits were close to corresponding ones resulting for binary traits while with MTDFREML, the estimates of residual correlation for binary traits were lower than those for continuous traits. This result is in agreement with Mantysaari *et al.* (1991) who reported that when an accurate estimate of residual correlation is needed, the marginal maximum likelihood estimates are superior to the estimates with the linear model. Figure 7 shows the decision chart for estimating genetic and environmental correlations for continuous and binary traits based on bias and MSE.

Genetic Correlation

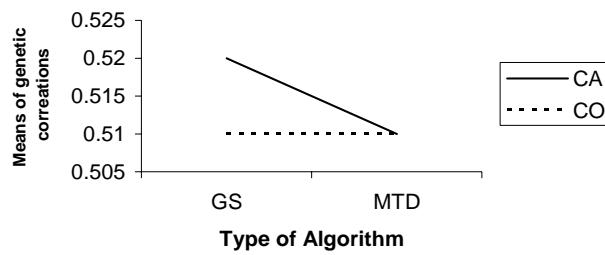


Figure 5. Algorithm type - trait type interaction

Residual Correlation

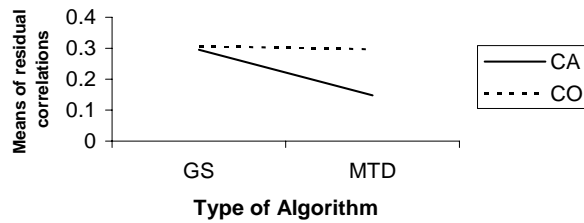


Figure 6. Algorithm type - trait type interaction

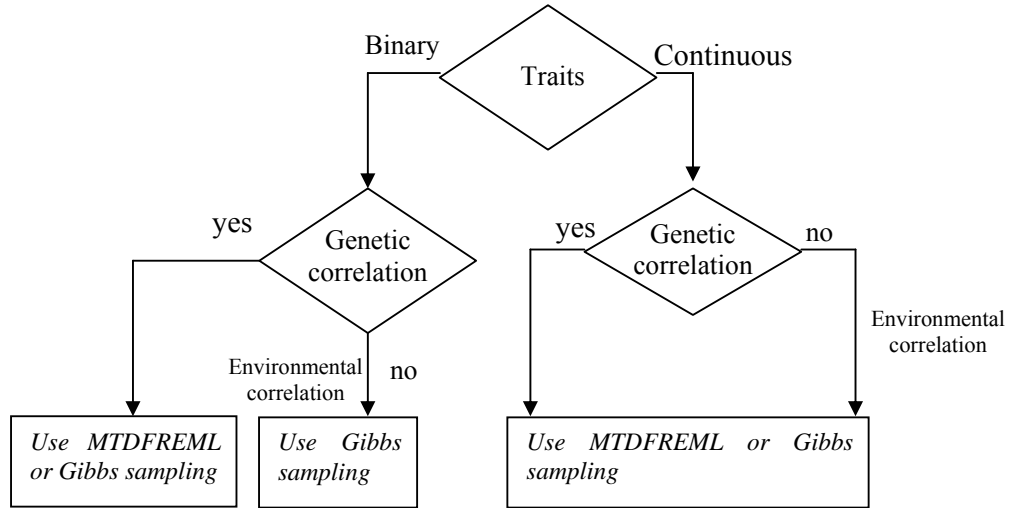


Figure 7. Decision chart for recommending the best algorithm to estimate genetic and environmental correlations based on bias and MSE

Table (5) shows the same trend as that in figure 5, 6 and 7, where the high MSE for binary trait with MTDFREML compare to GS (0.015 vs 0.009). When parameter estimation is performed from the discrete scores, but assuming normality (MTDFREML) the values of residual correlation clearly is an underestimate (0.148) and the t statistics are large (35.546). When the binary trait is analyzed by GS, the value of t statistics approached levels close to continuous variable estimates (0.483).

Table 5. Estimates of genetic and residual correlations from data with two continuous and binary variables

Traits	Algorithm			
	MTDFREML	MTDFREML	GS	GS
	Continuous	Binary	Continuous	Binary
Genetic correlation	0.511	0.513	0.509	0.52
SD	0.027	0.122	0.082	0.090
Bias	-0.011	-0.013	-0.009	-0.020
t	-0.402	-0.401	-0.411	-0.858
MSE	0.011	0.015	0.007	0.009
Residual correlation	0.297	0.148	0.307	0.295
SD	0.009	0.017	0.043	0.037
Bias	0.003	0.152	-0.007	0.005
t	0.361	35.546	-0.598	0.483
MSE	0.001	0.023	0.002	0.001

CONCLUSIONS

Based on bias and MSE, the present results indicated the following:

- 1- For continuous trait, the GS algorithm is equivalent to MTDFREML algorithm in estimating the heritability.
- 2- For binary trait, GS is the best algorithm in estimating the heritability.
- 3- The MTDFREML algorithm is equivalent to GS algorithm in estimating genetic correlation.
- 4- The threshold model (GS) is the best for estimating the residual correlation and the estimate of the residual correlation might be more important than the heritability and the genetic correlation if the aim is estimation of environmental effects.

REFERENCES

- Analla, M., A. Sanchez, A. Munaz and J.M. Serradilla, 1995. Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain. *Small Ruminant Research*, 17:51-55.
- Boettcher, P. J., L. K. Jairath, and J. C. M. Dekkers, 1999. Comparison of methods for genetic evaluation of sires for survival of their daughters in the first three lactations. *J. Dairy. Sci.* 82:1034-1044.
- Boldman, K.G.; L.A. Kriese; L.D. Van Vleck; C.P. Van Tassell and S.D. Kachman, 1995. A Manual for Use of MTDFREML. A Set of Programs To Obtain Estimates of Variances and Covariances [Draft]. U.S. Department of Agriculture, Agricultural Research Service. pp 114.
- Elsaid, R.M., 2004. Simulation study on sire progeny testing. PhD. Thesis, Fac. Agric., Ain Shams Univ., Egypt.
- Fisher, R.A., 1921. On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1:3-32.
- Gianola, D., 1982. Theory and analysis of threshold characters. *J. Anim. Sci.* 54:1079-1096.
- Luo, M.F., P.J. Boettcher; L.R. Schaeffer and J.C.M. Dekkerst, 2001. Bayesian inference for categorical traits with an application to variance component estimation. *J. Dairy. Sci.* 84:694-704.
- Mantysaari, E.A., R.L. Quaas and Y.T. Grohn, 1991. Simulation study on covariance component estimation for two binary traits in an underlying continuous scale. *J. Dairy. Sci.* 74:580-591.
- Matos, C.A.P., D.L. Thomas, D. Gianola, R.J. Tempelman, and L.D. Young, 1997. Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models: I. estimation of genetic parameters. *J. Anim. Sci.* 75:76-87.
- Meijering, A., 1985. Sire evaluation for calving traits by BLUP and nonlinear methodology. *Tierzuchtg. Zuchtgsbiol.* 102:95-105.
- Mousa, E. and M. Elsayed, 2001. Quality of variance components estimated by Gibbs sampling in populations with and without selection and varying heritability. *Egyptian J. Anim. Prod.* 38:97-110.
- Neter, J., W. Wasserman and H. Kutner, 1985. Applied linear statistical model regression, analysis of variance and experimental designs. Second edition Library of Congress Catalog. Card No. 84:395.

- SAS, 1996. Statistical Analysis System, SAS User's Guide : Statistics. SAS Institute Inc. Editors, Cary, NC.
- Sorensen, D.A., S. Anderson, D. Gianola, and I. Korsgaard, 1995. Bayesian inference in Neter, J.; W. Wasserman and H. Kutner (1985). Applied linear statistical model regression, analysis of variance and experimental designs. Second edition Library of Congress Catalog. Card No. 84:395.
- threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27:229-249.
- Templeman, R.j., 1993. Poisson mixed models for analysis of counts with application to dairy cattle breeding. Ph.D. Dissertation, Univ., Wisconsin, Madison.
- Van Tassell, C.P. and L. D. Van Vleck, 1995. A Manual for Use of MGSAM. A Set of FORTRAN Programs To Apply Gibbs Sampling to Animal Models for Variance Components Estimation [Draft]. U.S. Department of Agriculture, Agricultural Research Service. pp 86.
- Wang, C.S., J.J. Rutledge and D. Gianola, 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.* 21:41-62.

تقدير المعالم الوراثية والبيئية لصفتين متقطعتين – دراسة محاكاة

رضا السعيد محمد أحمد¹ ، منال محمد أحمد سيد²

1- قسم التنمية المستدامة للبيئة ، معهد الدراسات والبحوث البيئية ، جامعة المنوفية ، فرع السادات، 2- قسم الإنتاج الحيواني ، كلية الزراعة ، جامعة عين شمس

استهدفت هذه الدراسة تحديد نوع البرنامج المستخدم وكذلك نوع الصفة على مدى دقة تقدير المكافئ الوراثي والإرتباطات الوراثية والبيئية . أجريت هذه الدراسة على برامج تقوم بعمل محاكاة لصفتين مستمرتين ذات مكافئ الوراثي قدره 0.17 وأخرتين متقطعتين ذات مستويين (صفر ، 1) ناتجة من تحويل الصفات المستمرة ، وقد كان متوسط الصفة 0.5. خلقت 15 مكررة عشوائية بكل منها موسمان إدرار كتأثير ثابت وحيد. أجرى تحليل لكل من الصفات المستمرة والصفات المتقطعة ببرنامجين هما MTDFREML و Gibbs Sampling واستخدم نموذج الحيوان (Animal model). وقد أستخدم في هذه الدراسة التحيز (Bias) ومتوسط الخطأ مربعاً (Mean squared errors) للمقارنة بين تقديرات المكافئ الوراثي والإرتباطات الوراثية والبيئية المتحصل عليها من البرامج المختلفة. أجرى تحليل تباين لدراسة تأثير نوع البرنامج المستخدم ونوع الصفة (صفة مستمرة وصفة متقطعة) والتداخلات على التحيز بالنسبة لتقديرات المكافئ الوراثي وتقديرات الإرتباطات الوراثية والبيئية وأيضا التداخلات بين هذه العوامل. كان تأثير كل من هذه العوامل معنوياً جداً بالنسبة للتحيز في تقديرات المكافئ الوراثي وكذلك الإرتباط البيئي أما في حالة الإرتباط الوراثي فإن تأثير هذه العوامل لم يكن معنوياً. نستخلص من هذه الدراسة :

- 1- في حالة الصفة المستمرة يكون استخدام برنامج MTDFREML مكافئاً لإستخدام برنامج GS في تقدير المكافئ الوراثي.
- 2- في حالة الصفة المتقطعة يكون برنامج GS أفضل من برنامج MTDFREML في تقدير المكافئ الوراثي.
- 3- سواء بالنسبة للصفات المستمرة أو الصفات المتقطعة فإن إستخدام برنامج الـ MTDFREML يكون مكافئاً لإستخدام برنامج الـ GS في تقدير الإرتباط الوراثي.
- 4- بالنسبة للصفات المتقطعة فإن إستخدام برنامج الـ GS هو الأفضل لتقدير الإرتباط البيئي.