The following table shows that there is not equality between the matrices variances covariances of the groups. Indeed at the threshold of significance 5%, the p-value of the statistics of the M Box Test is higher than $0,05$. Therefore, there is no equality of matrices of variances of covariances between the groups. Moreover, with the canonic Correlation coefficient which is of 46,80%, we can conclude that the score function constructed by the Fischer discriminant Analysis is not valid (although the

4.5 Score Function Obtained by Logistic Regression

The constructed model has retained the ratios, R2, R3, R4, R5, R7, R8, R10, R11 and R13 as significant. The estimated coefficients of those ratios are presented in the following table:

Different values of Pr(> |z|), we notice that the coefficients which are statistically null at a level of confidence of 90% are those of the ratios R2, R3, R7, R8, R10. Thus, the Score function obtained by logistic regression is:

$$S(X) = -20,94 + 1,78\ R4 + 2,61\ R5 + 19,57\ R13. \tag{36}$$

The model being constructed, only awaiting validation now. To do so, we have recourse to the residuals Analysis of deviance and ot the hosmer-Lemeshow test.

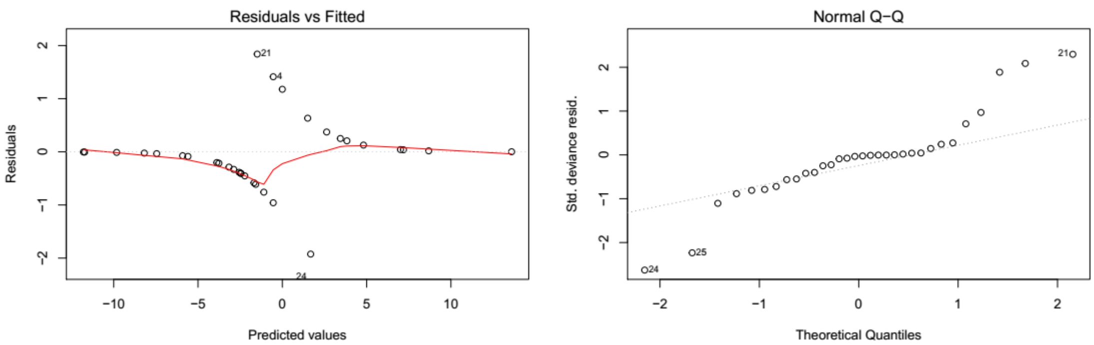

Figure 3. Curve of the residuals of deviance

Residuals of deviance analysis On the Residuals vs Fitted, we notice that no atypical point is found; all the points, are well situated in the interval $[-2, +2]$ moreover, the normal shows that the Residuals of deviance follow a law of Gauss (since the majority of points are situated around the dotted line). Hence, we conclude that the elaborated model is valid.

Hosmer Lemeshow test The result of the Hosmer Lemeshow Test is the following: At the threshold of significance of 5%, the adjustment of the model is good as the p-value of the statistic of the test of chi square to 8 degrees of liberty is lower than 5%. Therefore, the hypothesis $\mathbf{H_1}$: The considered model is not appropriate, is rejected we then conclude that the model is appropriate so valid.
Being given that the Fischer analysis has not been conclusive, it follows that the best model for our database is that obtained by logistic regression.

5 Conclusion

In this article, we have shown how to model a helping tool to the decision making by the Fischer discriminant analysis and the logistic regression, which are techniques founded on statistics and probabilistic methods of the method of Scoring. We have presented the theoretical and the practical construction of the

Table 5. Coefficients of canonic discriminating functions

Fonction 1	
R12	0,75
R7	0,458

Table 6. M Box test of the equality of matrices of covariances

	M de Box	2.987
F	Approximativement	0.92
	ddl1	3.00
	ddl2	17832.40
	Signification	0.43

Table 7. Canonic correlation and proper values

Function	proper.value	% of the variance	% cumulated	canonic correlation
1	0.28	100.00	100.00	0.47

Table 8. Wilk's Lambda test

Test of the function(s)	lambda of Wilk's	khi-deux	ddl	signification
1	0.78	7.15	2	0.03

Table 9. Results of the logistic regression model

Ratios retenus	Estimate	Std. Error	z value	Pr(> |z|)	$IC_{90\%}$
(Intercept)	-20.94	11.02	-1.90	0.06	-39.06 | -2.81
R2	61.59	40.47	1.52	0.13	-4.99 | 128.16
R3	-16.11	12.21	-1.32	0.19	-36.19 | 3.98
R4	1.78	0.90	1.99	0.05	0.31 | 3.26
R5	2.61	1.25	2.09	0.04	0.55 | 4.66
R7	9.14	5.51	1.66	0.10	0.08 | 18.20
R8	18.10	11.61	1.56	0.12	-1.00 | 37.20
R10	-23.87	18.05	-1.32	0.19	-53.56 | 5.82
R11	-11.61	7.36	-1.58	0.11	-23.72 | 0.49
R13	19.54	9.64	2.03	0.04	3.69 | 35.39

Table 10. Hosmer Lemeshow Test at the level of 5%

X-Squared	df	p-value
32	8	$9{,}3.10^{-5}$

score function for each technique. The validation of the regression model has been carried out by the HosmerLemeshow Test and the Residuals of Deviance Analysis. For the Fischer discriminant Analysis, this latter has not been conclusive at the term of the M Box Test and the Canonic Correlation. The obtained score function by the logistic regression is:

$$S(x) = -20,94 + 1,78\ R4 + 2,61\ R5 + 19,54\ R13. \tag{37}$$

It comes out the ratios R4, R5, R13 are the most discriminating ratios enabling to explain the membership of an individual to a modality of Y. The basic hypothesis according to which the risk of non-repayment depends on the characteristics of the client is found then checked by the logistic regression technique.

Acknowledgements. We thank all our colleagues for the discussions and the suggestions which have helped us to improve the presentation of this work.

Conflict of interests. The authors declare no conflict of interest.

References

1. K. Zaghdoudi, N. Djebali, and M. Mezni, "Credit Scoring and Default Risk Prediction," *University of Jendouba*, 2016.
2. G. Hein, "Le contrôle interne du risque de crédit bancaire," *Université de Nice-Sophia Antipolis*, pp. 18, 2002.
3. T. Roncalli, "La gestion des risques financiers," Economica, pp. 32-325, 390-393, 2004.
4. L. Thomas, C. Edelman, and N. Crook, "Credit Scoring and Its Applications," *Philadelphia, society for Industrial and Applied Mathematics*, 2002.
5. W. Beaver, "Financial ratios as predictors of failure," *Journal of Accounting Research*, vol. 4, pp. 71-111, 1966.
6. E. Altman, "Financial ratios, discriminant analysis and predictor of corporate Bankruptcy," *The Journal of finance*, pp. 589-609, 1968.
7. A. Saunder, L. Allen, "Credit rist measurements: New approach to value at risk and the other paradigms," New York, 2002.
8. M. Schervish, "Theory of statistics," Springer-Verlag, pp. 459, 1995.
9. M. Chavent, "Scoring," 2014.
10. M. Chavent, "Analyse discriminante linéaire et quadratique," 2015.
11. E. Azzouz, "Credit risk management by the Scoring method: Case of the popular bank of Rabat-Kenitra," Remadem, 2009.
12. C. Chesneau, "Modèles de régression," *Université de Caen*, pp. 69 -87, 2017.
13. R. Rakotamolala, "Pratique de la régression logistique: version 2.2," *Université Lumière de Lyon 2*, 2015.
14. S. Menard, "Applied logistic regression analysis," *University of Colorado, Boulder*, pp. 82, 2002.
15. L. Rouviere, *"Régression logistique avec R," Universités Rennes 2*, 2015.
16. H. Calvet, "Méthodologie de l'analyse financière des établissements de crédit: 2^e édition," Economica, Paris, 2002.
17. A. Charbonneau, "La mise en place d'un modèle d'évaluation du risque de crédit dans le cadre de la reforme solvabilité 2," *Université d'Orleans*, page 25-34. 2003.
18. A. Chebill, A. Anderson, "On the existence of maximum likelihood estimates in logistic régression models," Vol. 71, $N°1$, pp. 1-10, 1984.
19. E. Deakin, "A discriminat analysis of predictors of business failure," *Journal of Accounting Research*, Spring, page 167-179, 1972.
20. F. William, M. Carthy, and N.Guo, "The existestence of maximun likelihood estimateds for the logistic regression model," *Maryland Medical Research Institute*, 2009.
21. M. Schreiner, "Scoring the next break through in microcrédit," CGAP, pp. 102, 2002.
22. A. Rencher, "Interpretation of Canonical Discrimant Functions, Canonical variates, and Principal Components," Vol. 4, $N°3$, pp. 217-225, 1992.
23. O. Kindi, and A. Achonu, "A credit Scoring model for the savings and crédit mutual of the patout zone (MECZOP)/ Sénégal," *The Journal of Sustainable Development*, vol. 7, pp. 17-33, 2012.
24. C. Gourieroux, "Courbe de performance, de sélection et de discrimination," *Anales d'économie et de statistique*, $N°28$, page 108-115, 1992.
25. C. Gourieroux, "Économétrie de la finance: L'exemple du risque de crédit," *Actualité de l'économie*, vol. 79, $N°4$, pp. 399-418, 2003.

Appendix

Proof of the Theorem 2.2

We know that $g_k = argmax\ \mathbb{P}_k f_k(x)$. Maximise $\mathbb{P}_k f_k(x)$ is equivalent to maximise $\ln(\mathbb{P}_k f_k(x))$;

$$\begin{aligned}\ln(\mathbb{P}_k f_k(x)) &= \ln(\mathbb{P}_k) + \ln[f_k(x)] \\ &= \ln(\mathbb{P}_k) - \frac{1}{2}(x-\mu_k)^\top \sum{}^{-1}(x-\mu_k) - \ln\left[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}\right]\end{aligned}$$

Since $\ln\left[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}\right]$ does not depend on k, then maximizing $\ln(\mathbb{P}_k f_k(x))$, is equivalent to maximise $d_k(x) = \ln(\mathbb{P}_k) - \frac{1}{2}(x-\mu_k)^\top \sum^{-1}(x-\mu_k)$. i.e maximizing $\mathbb{P}_k f_k(x)$ is to maximize $d_k(x)$ ■

Proof of the Proposition 2.4

It is important to show that:

$$\mathbb{P}(Y = k \mid X = x) = \frac{exp[d_k(x)]}{\sum_{k=0}^{1} exp[d_k(x)]} \tag{38}$$

with $d_k(x) = \ln(\mathbb{P}_k) - \frac{1}{2}(x-\mu_k)^\top \sum{}^{-1}(x-\mu_k)$.
From the theorem of Bayes, $\mathbb{P}(Y = k \mid X = x) = \dfrac{\mathbb{P}_k f_k(x)}{\sum_{k=0}^{1} \mathbb{P}_k f_k(x)}$ or $\ln(\mathbb{P}_k f_k(x)) = d_k(x) - \ln[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}]$.

Thus

$$\begin{aligned}\mathbb{P}_k f_k(x) &= exp(d_k(x)) exp(-\ln[(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}]) \\ &= exp(d_k(x)) \frac{1}{(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}}\end{aligned}$$

By replacing $\mathbb{P}_k f_k(x))$ by its value in the equation **5** the term $(2\pi)^{\frac{p}{2}} \mid \sum \mid^{\frac{1}{2}}$ is found on the numerator to the denominator and by simplifying we find the equation **38** ■

Proof of the Proposition 3.1

It is important to show that $\mathbb{P}(Y = 1 \mid X = x) = \dfrac{exp(x^\top \beta)}{1 + exp(x^\top \beta)}$.
From the relation 22, we have: $\ln\left(\dfrac{\mathbb{P}(Y = 1 \mid X = x)}{1 - \mathbb{P}(Y = 1 \mid X = x)}\right) = x^\top \beta$. By composing by the exponential function, we obtain $\dfrac{\mathbb{P}(Y = 1 \mid X = x)}{1 - \mathbb{P}(Y = 1 \mid X = x)} = exp(x^\top \beta)$ and by developing, then $\mathbb{P}(Y = 1 \mid X = x) = \dfrac{exp(x^\top \beta)}{1 + exp(x^\top \beta)}$. ■

