TREC & KREC of twins: Decomposing the covariance matrix

Henrique Aparecido Laureano* Wagner Hugo Bonat†
Stéfanne Maria Jeha Bortoletto‡
Carolina Cardoso de Mello Prando§

August, 2022

Multivariate linear *regression* models built to handle multiple responses and with genetic and environmental interrelations of twins data (Bonat and Hjelmborg 2022) were applied to understand the dynamics from the TREC and KREC measures.

Keywords:

Introduction

Data

We have the TREC and KREC measures of 100 pairs of twins. From these 200 twins, we have five covariates, as follow:

- Birth weight (grams);
- Gestational age (weeks);
- Type of childbirth (normal or cesarean);
- Sex (male of female);
- Zygosity (MZ: monozygotic or DZ: dizygotic).

*Instituto de Pesquisa Pelé Pequeno Príncipe, Curitiba, Paraná, Brazil
†Laboratório de Estatística e Geoinformação, Universidade Federal do Paraná, Curitiba, Paraná, Brazil
‡Faculdades Pequeno Príncipe & Instituto de Pesquisa Pelé Pequeno Príncipe, Curitiba, Paraná, Brazil
§Faculdades Pequeno Príncipe & Instituto de Pesquisa Pelé Pequeno Príncipe, Curitiba, Paraná, Brazil

Zygosity is a key term in our modeling framework since we need to inform how many twin pairs are monozygotic and dizygotic. Therefore, we drop a pair that do not present this information. We stay with 198 twins. Besides these five covariates, we may also consider the twin covariate (1 or 2), to verify a randomness supposition.

In Figure 1 we have some plots to see how our two responses behave. We can see that they are asymmetric with both having some upper outliers. For the TREC we have a mean of 101.3 and a median of 99.5. For the KREC we have a mean of 66.43 and a median of 56. In the scatterplots we see a clear positive association, linear until the outliers started to act.

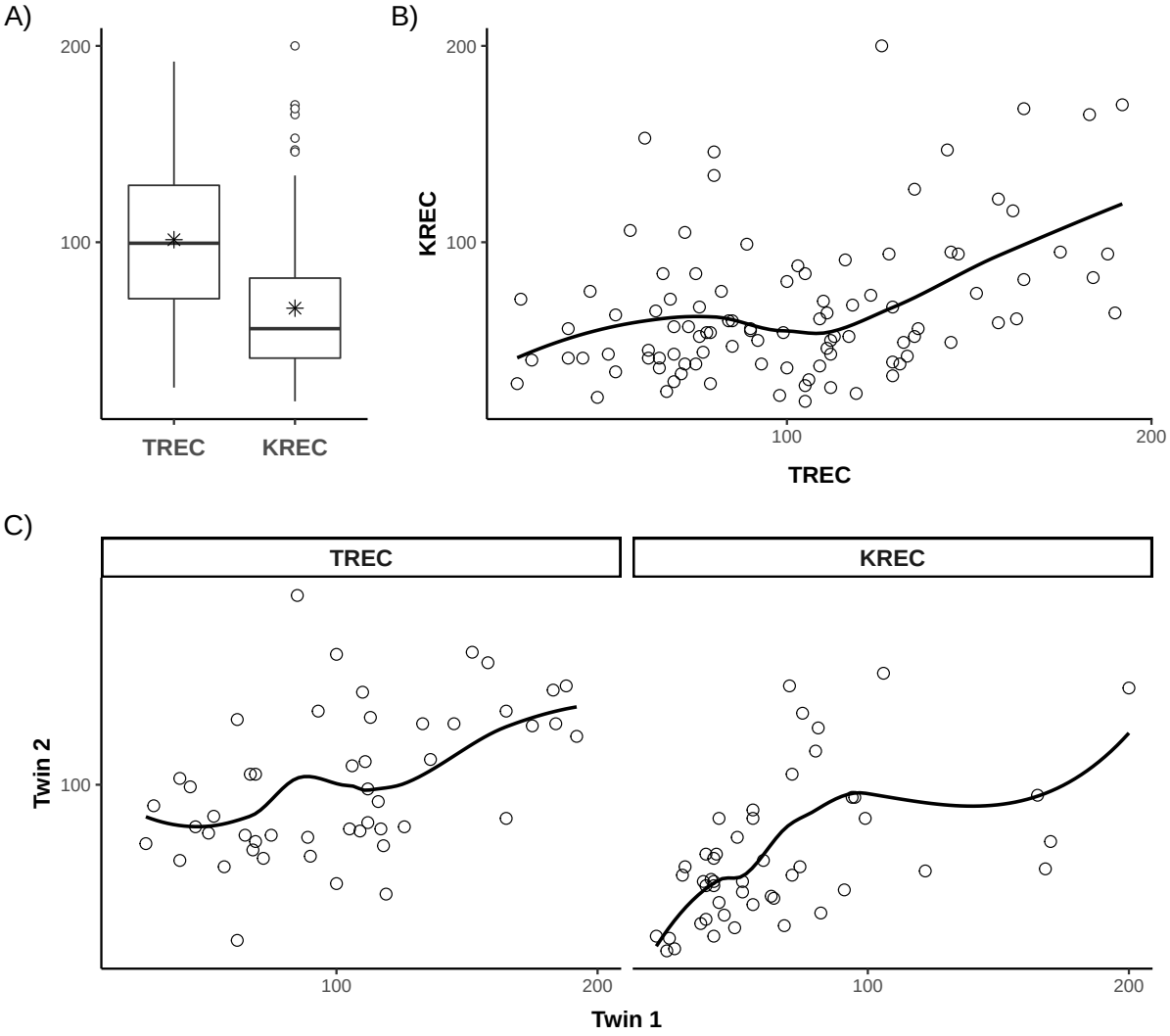

Figure 1: Graph A) TREC and KREC boxplots with their means in asterisks; Graph B) TREC and KREC scatterplot with tendency curve in solid black; Graph C) Twins scatterplots per TREC and KREC with tendency curves in solid black.

In Figure 2 we have the plots of the covariates. We have a mean birth weight of 2318.76 grams and a mean gestational age of 35.18 weeks. 92.9% of the twins were born from cesarean childbirth and 76.8% are zygotes. In terms of sex, the sample is balanced.

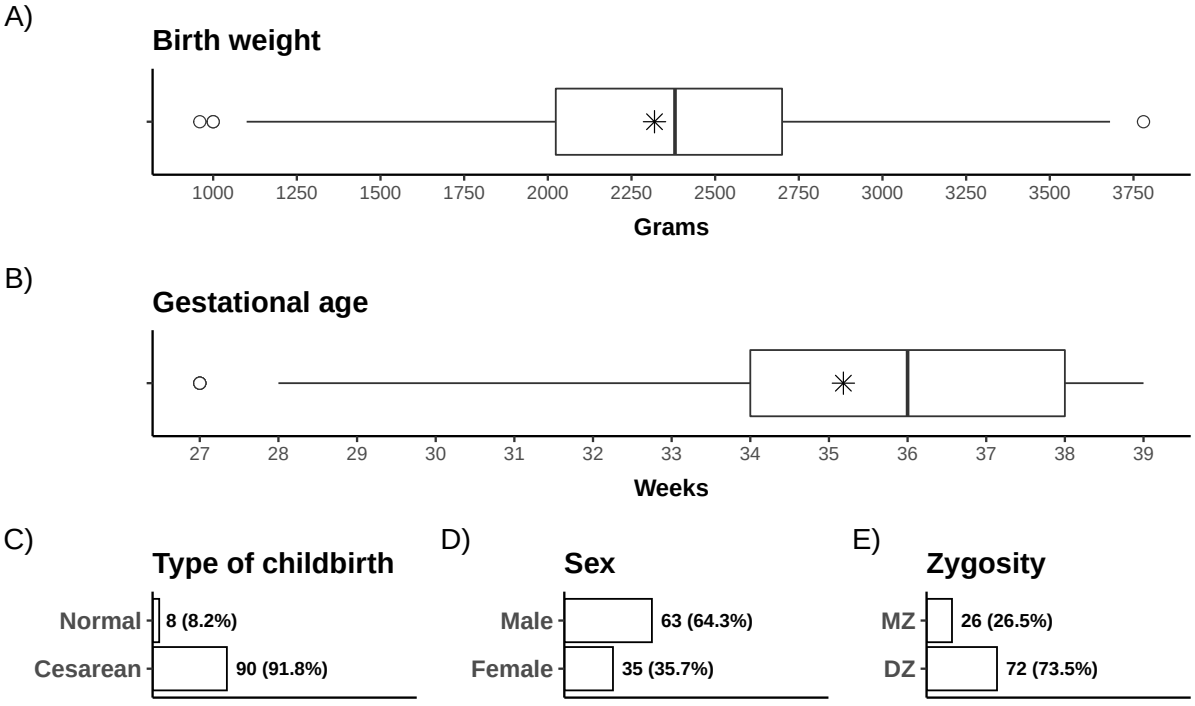

Figure 2: Graphs A) and B) are the boxplots of the numerical variables with their means in asterisks; Graphs C), D) and E) are the barplots of the categorial variables with their level frequencies and respective sample percentages.

Modeling framework

The statistical analysis was performed through the R (R Core Team 2022) language and environment for statistical computing. Besides the model fitting performed through the `{mglm4twin}` (Bonat 2022, 2018; Bonat and Hjelmborg 2022; Bonat and Jorgensen 2016) package, the leading R packages used are: `{dplyr}` (Wickham et al. 2022), `{tidyr}` (Wickham and Girlich 2022), `{ggplot2}` (Wickham 2016), `patchwork` (Pedersen 2020), `{kableExtra}` (Zhu 2021) and `{Matrix}` (Bates, Maechler, and Jagan 2022).

Results

We model the TREC and KREC measures in a bivariate fashion and in two fronts, in the mean and covariance structures. We use a model called ACE to study heritability and the genetic and environmental relationships. Basically, we decompose the covariance matrix into three:

- A: genetic/heritability effect or component;
- C: common environment effect or component;
- E: unique environment effect or component.

Besides the covariance decomposition itself, we can also insert covariates in it. In both mean and covariance structures we tested for the effect of six covariates. With a bivariate model, we are able to handle the TREC and KREC measures jointly. We started with the ACE model, however, the C component presented to be non-significant. Thus, we stayed with an AE model and from this performed a covariates selection procedure.

Table 1: Summary of the mean structure of the AE bivariate model.

	Estimate	SD	z-value	Pr(>|z|)
TREC				
Mean gestational age & female sex	42.9477	55.5558	0.7731	**0.4395**
A week in the gestational age	1.7035	1.5713	1.0841	**0.2783**
Male sex	-1.5849	7.7800	-0.2037	**0.8386**
KREC				
Mean gestational age	52.4270	51.3111	1.0217	**0.3069**
A week in the gestational age	0.4047	1.4513	0.2789	**0.7804**

Starting from the mean structure, from the six covariates we end up with two. The significant ones, for each response, are presented in Table 1. For the TREC measures, the statistically significant covariates are the gestational age and the twin sex. For KREC, only the gestational age is significant.

About the covariance structure, what we do is decompose the TREC variance (13698.107), KREC variance (13787.069), and their covariance (580.0059) in the manner described in Table 2.

At Table 3 we have the environment and heritability measures. The heritability is statistically significant in all scenarios, marginally; jointly; and its correlation.

Given the small sample size and the inferential level of complexity involved in the bivariate model, we were not able to obtain numerical convergence by inserting covariates in the covariance structure. As a solution, we fitted univariate AE models with the same mean structure as in the bivariate scenario.

Table 2: Summary of the covariance structure of the AE bivariate model.

	Estimate	SD	Percentage	z-value	Pr(>|z|)
Environment component (E)					
TREC	705.5144	NaN	43.9445	NaN	**NaN**
KREC	525.5812	574.3538	38.2875	0.9151	**0.3601**
TREC & KREC	380.2592	166.5821	65.5613	2.2827	**0.0224**
Genetic component (A)					
TREC	847.2605	221.5772	52.7734	3.8238	**0.0001**
KREC	772.3264	504.0718	56.2624	1.5322	**0.1255**
TREC & KREC	188.6464	220.0809	32.5249	0.8572	**0.3914**

Table 3: Summary of the genetic components of the AE bivariate model.

	Estimate	SD	z-value	Pr(>|z|)
Environmentability				
TREC	0.454357	NaN	NaN	**NaN**
KREC	0.404945	0.411973	0.982941	**0.325636**
TREC & KREC	0.668405	0.331751	2.014777	**0.043928**
Correlation	0.379516	0.594325	0.638567	**0.523105**
Heritability				
TREC	0.545643	NaN	NaN	**NaN**
KREC	0.595055	0.411973	1.444404	**0.148625**
TREC & KREC	0.331595	0.331751	0.999530	**0.317538**
Correlation	0.282405	0.139745	2.020867	**0.043293**

We started with the same six covariates. In the TREC model, we end up with no significant covariates in the environmental E component and with the type of childbirth as significant in the genetic A component. In the KREC model, the type of childbirth and zygosity are statistically significant in the E component, and in the A component, only the type of childbirth is significant. In **?@tbl-TREC-AE-model-covariance** and Table 4 we have a summary of the covariances decompositions in terms of the significant covariates for the TREC and KREC measures, respectively.

In the TREC measure (**?@tbl-TREC-AE-model-covariance**), 26% of the observed variance has its explanation attributed to the environment component. In a twin born from a cesarean, 72% of the observed variance is explained by the genetic component. Together, both

components explain 98% of the variance. In a twin born from normal childbirth, the explanation by means of the genetic component decreases by 80% - no variance explanation. We can justify that by the low representation in our data, with only 8.16% of the twins being born from normal childbirth.

In the KREC measure (Table 4), in a twin born from a cesarean and dizygotic, nothing of the observed variance is explained by the environment component (-62%). Nevertheless, if we change to a twin born from normal childbirth the attributed explanation grows by 66%, i.e. 4% (-62 + 66) of variance explanation. In a monozygotic twin, the explanation by the environment component is 9% (-62 + 71). In the genetic component, the difference in terms of variance explained by the type of childbirth is abysmal. In a twin born from cesarean, we have 103% (-62 + 165) of variance explained. The effect is so big and uncertain that it blows up. In a twin born from normal childbirth, we have the opposite. The unbalance in the data frequencies ends up generating these unrealistic results.

Table 4: Summary of the covariance structure of the AE KREC model.

	Estimate	SD	Percentage
Environment component (E)			
Cesarean childbirth & DZ zygosity	94.2629	278.5293	6.8669
Normal childbirth	-728.0586	NaN	-53.0376
MZ zygosity	641.8636	681.0078	46.7585
Genetic component (A)			
Cesarean childbirth	1053.4669	338.0661	76.7429
Normal childbirth	1658.3727	NaN	120.8091

References

Bates, Douglas, Martin Maechler, and Mikael Jagan. 2022. *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.4-1. https://CRAN.R-project.org/package=Matrix.

Bonat, Wagner Hugo. 2018. "Multiple Response Variables Regression Models in R: The mcglm Package." *Journal of Statistical Software* 84 (4): 1–30. https://doi.org/10.18637/jss.v084.i04.

———. 2022. *mglm4twin: Multivariate Generalized Linear Models for Twin Data*. R package version 0.3.0. https://github.com/wbonat/mglm4twin.

Bonat, Wagner Hugo, and Jacob V. B. Hjelmborg. 2022. "Multivariate Generalized Linear Models for Twin and Family Data." *Behavior Genetics* 52 (2): 123–40. https://doi.org/10.1007/s10519-021-10095-3.

Bonat, Wagner Hugo, and Bent Jorgensen. 2016. "Multivariate Covariance Generalized Linear Models." *Journal of Royal Statistical Society - Series C* 65: 649–75.

Pedersen, Thomas Lin. 2020. *patchwork: The Composer of Plots*. R package version 1.1.1. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *dplyr: A Grammar of Data Manipulation*. R package version 1.0.9. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Maximilian Girlich. 2022. *tidyr: Tidy Messy Data*. R package version 1.2.0. https://CRAN.R-project.org/package=tidyr.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.3.4. https://CRAN.R-project.org/package=kableExtra.

Appendix

Figure 3: Gestational age and weight scatterplot with tendency curve in solid black.