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

Methods

Data

- Weight;
- Gestational age;
- Type of childbirth;
- Sex;
- Zygosity;
- Twin (1 or 2, to verify the randomness supposition).

[*]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.
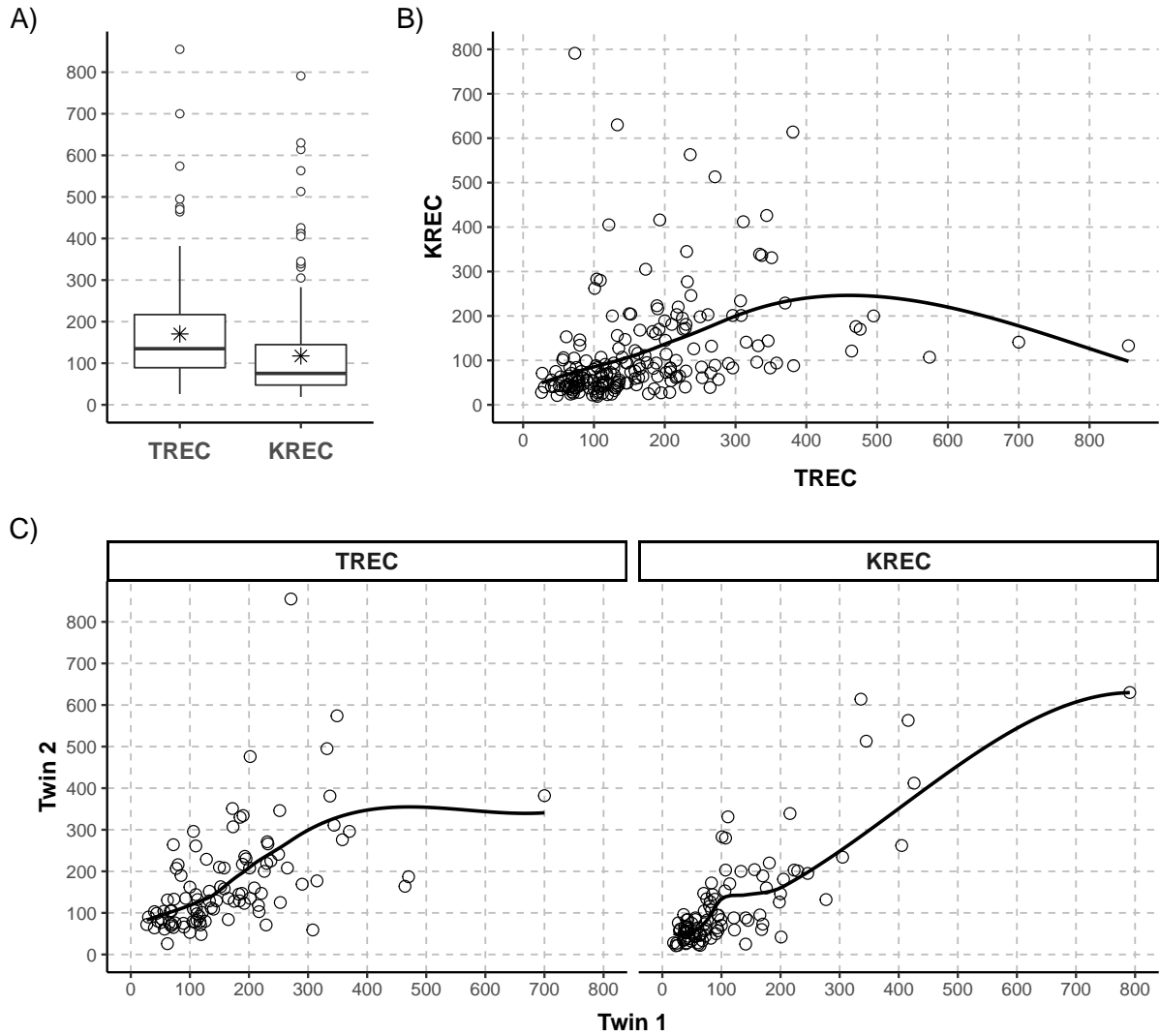

Figure 1: Graph A) TREC and KREC boxplots with their means in asterisks; Graph B) TREC and KREC scatterplot with tendency curve in solid black; Graph C) Twins scatterplots per TREC and KREC with tendency curves in solid black.

A)

Weight

B)

Gestational age

C)   D)   E)

Type of childbirth   **Sex**   **Zygosity**

Normal - 14 (7.1%)   Male - 108 (54.5%)   MZ - 46 (23.2%)

Cesarean - 184 (92.9%)   Female - 90 (45.5%)   DZ - 152 (76.8%)
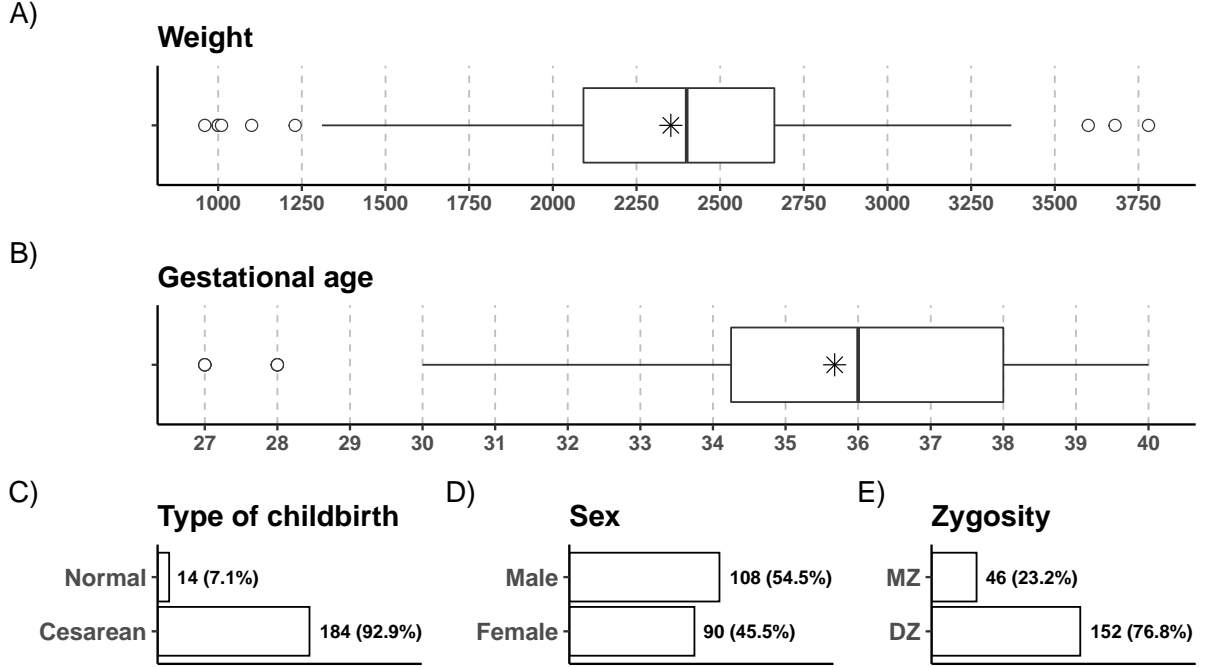
Figure 2: Graphs A) and B) are the boxplots of the numerical variables with their means in asterisks; Graphs C), D) and E) are the barplots of the categorial variables with their level frequencies and respective sample percentages.

Statistical analysis

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

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/regressors in it. In both mean and covariance structures we tested for the effect of six regressors.

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 regressors 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	-97.1912	127.8171	-0.7604	**0.4470**
A week in the gestational age	8.0203	3.5627	2.2512	**0.0244**
Male sex	-34.7133	15.0805	-2.3019	**0.0213**
KREC				
Mean gestational age	-160.7853	126.7882	-1.2681	**0.2047**
A week in the gestational age	7.7597	3.5385	2.1929	**0.0283**

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

About the covariance, what we do is decompose the TREC variance (13698.107), KREC variance (13787.069), and their covariance (4392.4173) in the manner described in Table 2.

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

	Estimate	SD	Percentage	z-value	Pr(>|z|)
Environment component (E)					
TREC	3712.4482	2751.5400	27.1019	1.3492	**0.1773**
KREC	1391.8790	4100.7340	10.0955	0.3394	**0.7343**
TREC & KREC	862.7034	490.8413	19.6407	1.7576	**0.0788**
Genetic component (A)					
TREC	8703.5756	3370.8550	63.5385	2.5820	**0.0098**
KREC	9990.1531	5750.4968	72.4603	1.7373	**0.0823**
TREC & KREC	2633.3467	1090.5257	59.9521	2.4147	**0.0157**

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

4

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

	Estimate	SD	z-value	Pr(>|z|)
Environmentability				
TREC	0.299005	0.209571	1.426747	**0.153653**
KREC	0.122287	0.363368	0.336538	**0.736465**
TREC & KREC	0.246765	0.157990	1.561908	**0.118310**
Correlation	0.379516	0.594325	0.638567	**0.523105**
Heritability				
TREC	0.700995	0.209571	3.344908	**0.000823**
KREC	0.877713	0.363368	2.415491	**0.015714**
TREC & KREC	0.753235	0.157990	4.767623	**0.000002**
Correlation	0.282405	0.139745	2.020867	**0.043293**

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 regressors in the covariance. As a solution, we fitted univariate AE models with the same mean structure as in the bivariate scenario.

We started with the same six regressors. In the TREC model, we end up with no significant regressors 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 Table 4 and Table 5 we have a summary of the covariances decompositions in terms of the significant regressors for the TREC and KREC measures, respectively.

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

	Estimate	SD	Percentage
Environment component (E)	3534.162	2773.739	25.8004
Genetic component (A)			
Cesarean childbirth	9851.598	3647.523	71.9194
Normal childbirth	-10976.948	3140.646	-80.1348

In the TREC measure, 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 7.07% of the twins being born from normal childbirth.

Num gêmeo com parto tipo cesaria e dizigoto, nada da variabilidade do KREC é explicada

pelo componente ambiental (-62%). Contudo, se mudamos para um gêmeo de parto normal a explicação cresce 66%, ou seja, 4% (-62 + 66) da variabilidade é explicada pelo componente ambiental. Num gêmeo monozigoto o percentual explicado é 9% (-62 + 71).

Table 5: Summary of the covariance structure of the AE KREC model.

	Estimate	SD	Percentage
Environment component (E)			
Cesarean childbirth & DZ zygosity	-8562.138	3940.7251	-62.1027
Normal childbirth	9185.825	4003.8748	66.6264
MZ zygosity	9829.275	609.2463	71.2934
Genetic component (A)			
Cesarean childbirth	22781.346	3823.2523	165.2371
Normal childbirth	-22636.617	3909.9002	-164.1873

Quando olhamos para o componente genético a diferença entre tipos de parto é abismal. Num gëmeo dizigoto de parto tipo cesaria temos 103% (-62 + 165) da variabilidade unicamente explicada pelo componente genético. Tal efeito é tão grande mas ao mesmo incerto, que o percentual estoura. No caso de parto normal temos exatamente o contrário. O desbalancemanto das frequências de tipo de parto na base de dados acaba gerando esses resultados "estranhos".

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.