

TREC & KREC of twins: Decomposing the covariance matrix

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

September, 2022

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

Keywords:

Introduction

Data

We have the TREC and KREC measures of 55 pairs of twins. From these 110 twins, we have five covariates, as follow:

- Birth weight (grams);
- Gestational age (weeks);
- Type of childbirth (normal or cesarean);
- Sex (male or 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

Zygoty is a key term since we need to inform how many twin pairs are monozygoty and dizygoty. Besides these five covariates, we may also consider the twin covariate (1 or 2), to verify a randomness supposition.

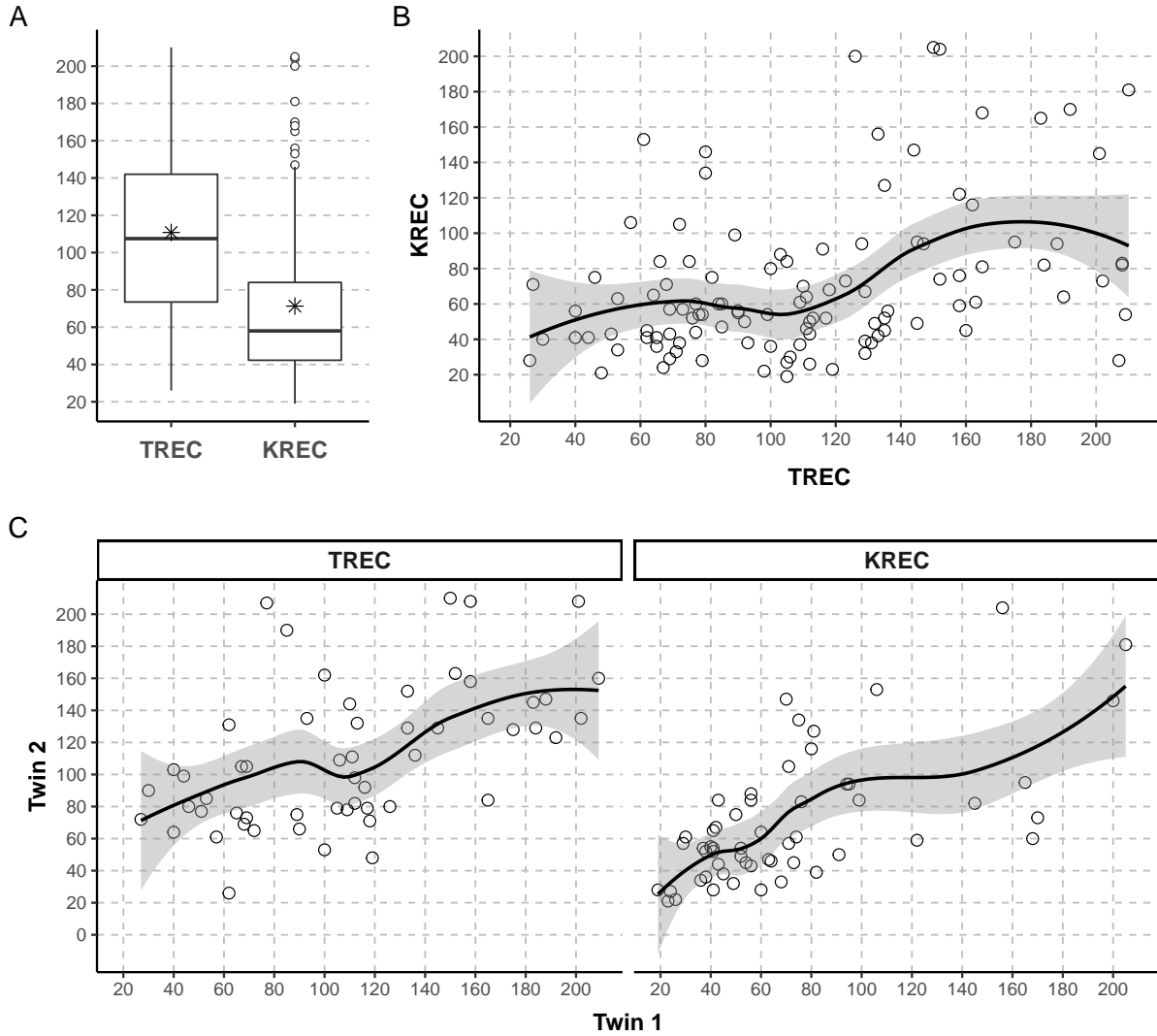


Figure 1: Graph A: TREC and KREC boxplots with their means in asterisks; Graph B: TREC and KREC scatterplot with tendency curve in solid black and 95% confidence band in gray; Graph C: Twins scatterplots per TREC and KREC with tendency curves in solid black and 95% confidence bands in gray.

In Figure 1 and the C panel of Figure 3 we have some plots to see how our two responses behave. We can see that the KREC measures are asymmetric with a stronger left tail. Both are positively correlated with an almost linear correspondence. By crossing twins, we do not

see any clear leakage. For the TREC we have a mean (standard deviation, sd) of 110.78 (46.56) and for the KREC a mean (sd) of 71.29 (42.67).

In Figure 2 we have the plots of the covariates. We have a mean (sd) birth weight of 2266.28 (566.45) grams and a mean (sd) gestational age of 34.98 (3.03) weeks. 88.2% of the twins were born from cesarean childbirth, 74.5% are zygotes, and 60.9% are from the male sex.

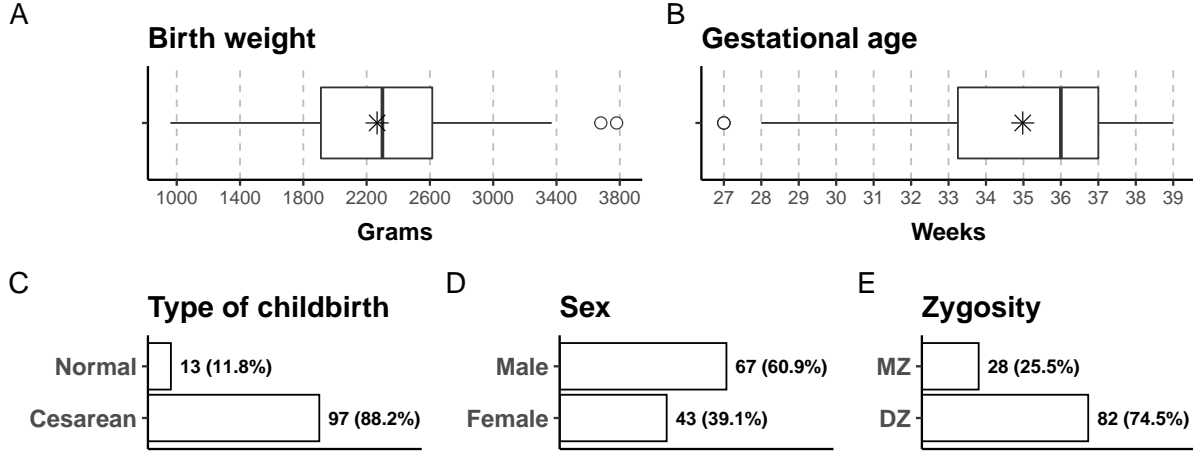


Figure 2: Graphs A and B are the boxplots of the numerical variables with their means in asterisks; Graphs C, D and E are the barplots of the categorial variables with their levels frequencies and respective sample percentages.

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 Hjelmberg 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), `{Matrix}` (Bates, Maechler, and Jagan 2022) and `{ggfortify}` (Tang, Horikoshi, and Li 2016).

Results

We model the TREC and KREC measures in a bivariate fashion. 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.

In the mean structure 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				
Cesarean childbirth (intercept)	106.9038	5.1533	20.7446	0.0000
Normal childbirth (add to intercept)	36.6523	14.5258	2.5233	0.0116
KREC				
Cesarean childbirth (intercept)	67.8584	4.8132	14.0984	0.0000
Normal childbirth (add to intercept)	28.3342	13.2053	2.1457	0.0319

From the six tested covariates we end up with two. For both responses, only the type of childbirth presented to be as statistically significant. In Table 1 we have a summary of the estimated effects. About the covariance structure, what we do is decompose the TREC variance (2167.713), KREC variance (1820.538), and their covariance (777.2108) 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	817.7266	132.5380	37.7230	6.1698	0.0000
KREC	487.8804	660.9370	26.7987	0.7382	0.4604
TREC & KREC	265.9359	159.3236	34.2167	1.6692	0.0951
Genetic component (A)					
TREC	1115.7542	317.5889	51.4715	3.5132	0.0004
KREC	1124.9519	626.5544	61.7923	1.7955	0.0726
TREC & KREC	358.6664	241.5068	46.1479	1.4851	0.1375

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

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

	Estimate	SD	z-value	Pr(> z)
Environmentability				
TREC	0.4229298	0.1009032	4.1914398	0.00002772
KREC	0.3024992	0.3955448	0.7647659	0.44441095
TREC & KREC	0.4257684	0.2795338	1.5231374	0.12772433
Correlation	0.4210331	0.3306534	1.2733360	0.20289870
Heritability				
TREC	0.5770702	0.1009032	5.7190463	0.00000001
KREC	0.6975008	0.3955448	1.7633928	0.07783426
TREC & KREC	0.5742316	0.2795338	2.0542475	0.03995174
Correlation	0.3201396	0.1929394	1.6592760	0.09706021

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. *mgglm4twin: Multivariate Generalized Linear Models for Twin Data*. R package version 0.3.0. <https://github.com/wbonat/mgglm4twin>.
- Bonat, Wagner Hugo, and Jacob V. B. Hjelmberg. 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/>.
- Tang, Yuan, Masaaki Horikoshi, and Wenxuan Li. 2016. “ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages.” *The R Journal* 8 (2): 474–85. <https://doi.org/10.32614/RJ-2016-060>.
- 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

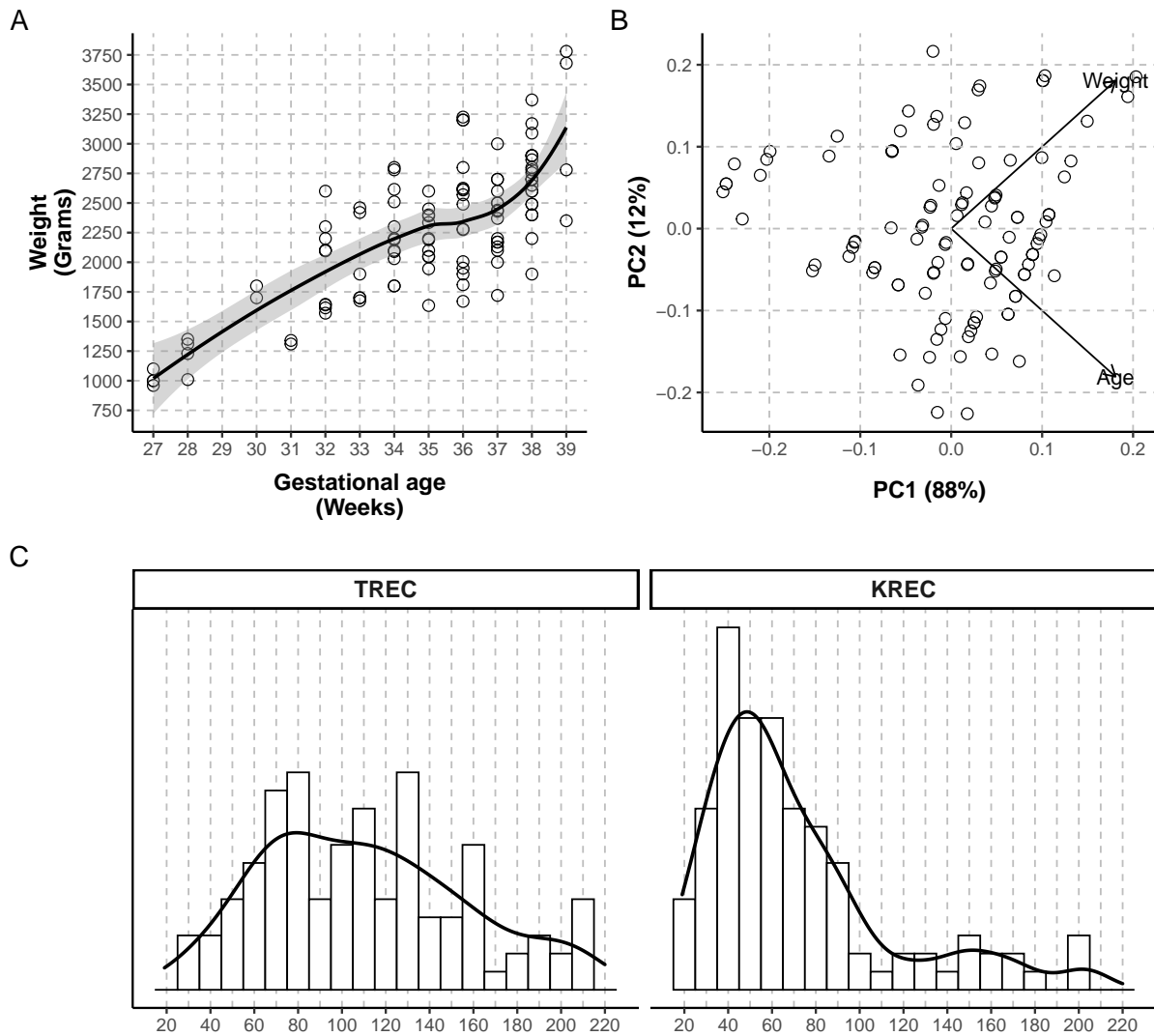


Figure 3: Graph A: Gestational age and weight scatterplot with tendency curve in solid black and 95% confidence band in gray; Graph B: Gestational age and weight biplot from a Principal Components Analysis (PCA); Graph C: TREC and KREC histograms with density curve estimations in solid black.