

Chapter 1

Examples

The following chapter presents the Bayesian LMG implementation on two examples. Simulated data is used for the first example. Empirical data is used for the second example.

1.1 Simulated Data

We assume a simple model for the first example:

$$\begin{aligned} Y_i &\sim \mathcal{N}(\beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4, \sigma^2), \\ \beta_1 &= 0.5, \beta_2 = 1, \beta_3 = 2, \beta_4 = 0, \sigma^2 = 1 \\ \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4 &\sim \mathcal{N}(0, 1) \end{aligned}$$

The values of the four predictors are sampled from a standard normal distribution. These values are then multiplied by the regression coefficients. A standard normal distributed error is added to obtain the dependent variable. Fifty observations were sampled.

The following Code was used to simulate the data :

```
x1 <- rnorm(50, 0, 1); x2 <- rnorm(50, 0, 1)
x3 <- rnorm(50, 0, 1); x4 <- rnorm(50, 0, 1)
b1 <- 0.5; b2 <- 1; b3 <- 2; b4 <- 0.8
#b1 <- 1; b2 <- 1; b3 <- 1; b4 <- 1

y <- b1*x1 + x2*b2 + b3*x3 + b4*x4 + rnorm(50, 0, 1)

df <- data.frame(y = y, x1 = x1, x2 = x2, x3 = x3, x4 = x4)
```

The model is fitted using the `rstanarm` package ([Stan Development Team, 2016](#)) with the default priors for the regression and σ^2 parameters. These default priors are called weakly informative priors, because they take into account the order of magnitude of the variables by using variance of the observed data. Information about these priors can be found in [Stan Development Team \(2017\)](#). A burn-in period of 1000, a sample size of 20000 and a thinning of 20 were

Table 1.1: Samples from the posterior distributions of the regression parameters

	x1	x2	x3	x4	sigma
sample 1	0.555	1.023	1.99	0.835	1.103
sample 2	0.283	1.153	2.19	0.679	1.143
sample 3	0.571	1.131	2.03	0.877	1.026
sample 4	0.455	1.231	2.00	0.925	1.266
sample 5	0.600	0.908	1.90	0.700	0.991
sample 6	0.730	1.018	2.03	0.401	1.266
sample 7	0.363	1.029	1.84	0.619	1.145
sample 8	0.296	0.805	1.77	0.599	1.290
sample 9	0.177	0.812	1.61	0.376	1.406
sample 10	0.694	1.018	1.72	1.028	1.192

chosen, resulting in a posterior sample size of 1000. The first few posterior samples are shown in Table 1.1.

For each posterior sample of the parameters the R^2 value is calculated. The R^2 of the submodels is then calculated by the conditional variance formula for each posterior sample. The resulting R^2 values of the first few posterior samples is shown in Table 1.2. The thinning makes sense in this case to reduce the computational burden and still obtain an appropriate posterior of the R^2 values (Link and Eaton, 2012).

The `hier.part` package is used to calculate the LMG value for each posterior sample. The LMG posteriors are shown in Table 1.3. The independent component (I) represents the LMG value. The joint contribution (J) represents the difference from the independent component to the explained variance of the predictor-only model (T). Assuming stochastic or non-stochastic regressors has an influence on the uncertainty. First, non-stochastic regressors are assumed. The resulting LMG values and joint contributions with a 95% credible interval are shown in Table 1.3. An option to display the resulting LMG distribution is shown in Figure 1.1. Using the default weakly informative priors, the LMG distributions obtained from the Bayesian framework are very similar to the bootstrap confidence intervals assuming non-stochastic predictors of the LMG estimates obtained from the `relaimpo` package, as shown in Table 1.4.

Table 1.2: R^2 for all submodels for the first six posterior samples

	sample 1	sample 2	sample 3	sample 4	sample 5	sample 6
none	0.000	0.000	0.000	0.000	0.000	0.000
x1	0.055	0.023	0.057	0.041	0.067	0.081
x2	0.055	0.059	0.069	0.080	0.048	0.055
x3	0.639	0.687	0.640	0.580	0.667	0.637
x4	0.068	0.040	0.072	0.076	0.053	0.014
x1 x2	0.099	0.075	0.114	0.109	0.104	0.122
x1 x3	0.681	0.702	0.684	0.610	0.719	0.702
x1 x4	0.125	0.064	0.131	0.118	0.123	0.096
x2 x3	0.750	0.806	0.770	0.722	0.769	0.747
x2 x4	0.126	0.102	0.144	0.159	0.104	0.071
x3 x4	0.705	0.725	0.709	0.653	0.718	0.650
x1 x2 x3	0.778	0.812	0.798	0.738	0.805	0.794
x1 x2 x4	0.171	0.118	0.190	0.190	0.161	0.139
x1 x3 x4	0.748	0.741	0.755	0.685	0.772	0.716
x2 x3 x4	0.819	0.847	0.843	0.799	0.823	0.762
all	0.848	0.854	0.873	0.817	0.861	0.809

Table 1.3: Variance decomposition for non-stochastic predictors. I = LMG values, J = joint contribution, Total = total explained variance in one-predictor only model

Variable	I	J	Total
x1	0.039 (0.007, 0.094)	0.012 (0.004, 0.019)	0.051 (0.012, 0.113)
x2	0.083 (0.03, 0.164)	-0.022 (-0.032, -0.013)	0.06 (0.015, 0.133)
x3	0.618 (0.497, 0.714)	-0.02 (-0.031, -0.008)	0.598 (0.472, 0.697)
x4	0.064 (0.014, 0.134)	-0.001 (-0.002, -3.301×10^{-4})	0.063 (0.014, 0.133)

Table 1.4: Variance decomposition for non-stochastic predictors. I = LMG values, J = joint contribution, Total = total explained variance in one-predictor only model

Variable	LMG value (95%-CI)	
	Relaimpo	Bayesian framework
x1	0.092 (0.046, 0.157)	0.039 (0.007, 0.094)
x2	0.109 (0.062, 0.177)	0.083 (0.03, 0.164)
x3	0.636 (0.548, 0.723)	0.618 (0.497, 0.714)
x4	0.03 (0.007, 0.072)	0.064 (0.014, 0.134)

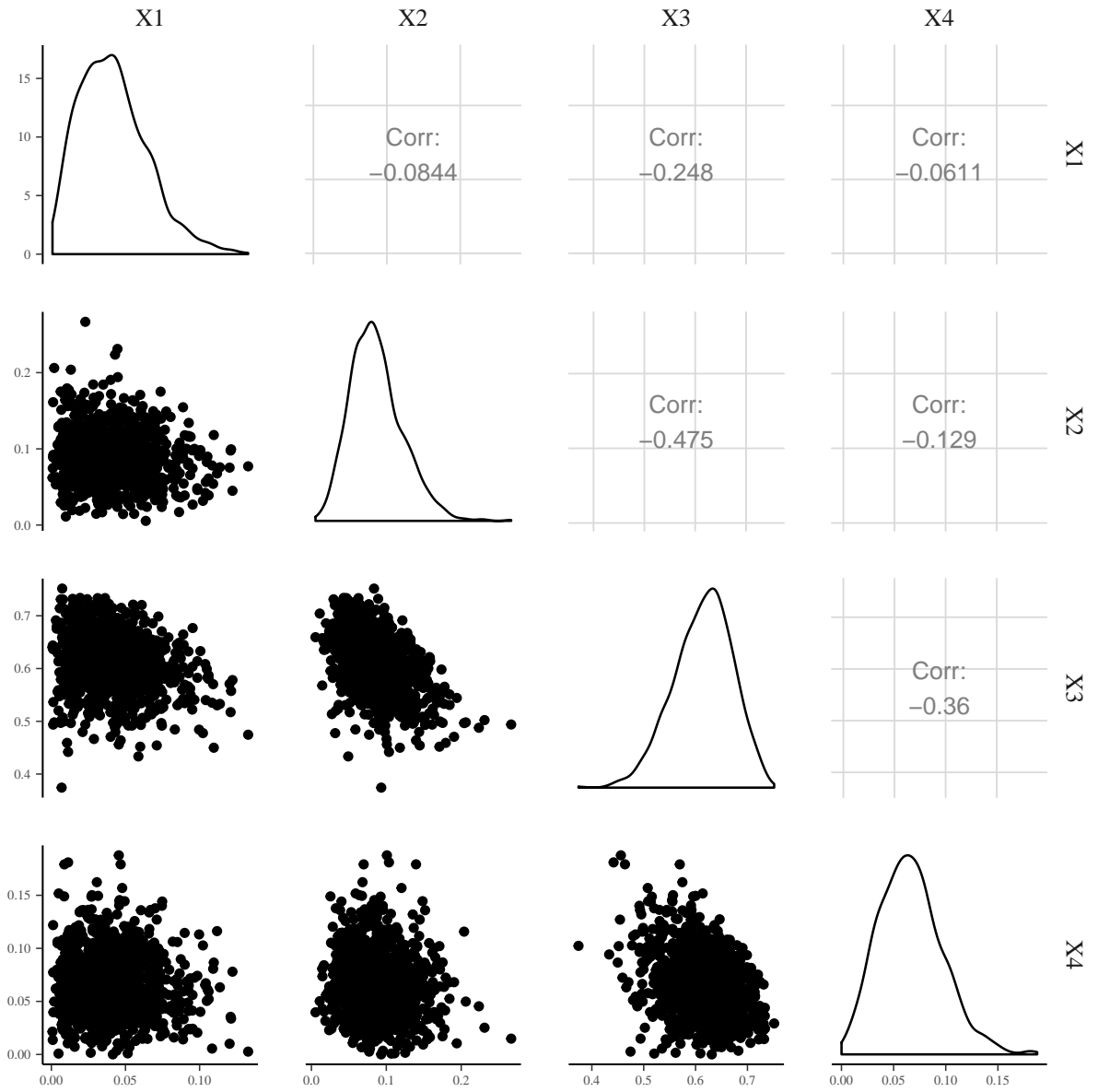


Figure 1.1: Posterior distribution of LMG values.

In this example we know that the predictor values were sampled from a normal distribution. It would therefore be more reasonable to assume stochastic predictors. Under the assumption of weak exogeneity and conditional independence, the posterior distributions of the regression parameters β are valid for non-stochastic and stochastic predictors. However, the uncertainty about the LMG values needs to include the uncertainty about the covariance matrix. If we know the distribution of the predictors we can incorporate this information and obtain the posterior distribution of the covariance matrix. The package **Jags** was used for inference about the covariance matrix in a Bayesian way. As an alternative non-parametric bootstrap was used for inference about the covariance matrix. Both approaches resulted in very similar LMG values. Using the default priors further resulted in very similar LMG values as the non-parametric bootstrap option of the **relaimpo** package. Table 1.5 shows the LMG values of these approaches. Using the bootstrap samples of the covariance matrix or samples from the posterior covariance matrix resulted in very similar LMG values. Bootstrap seems to be a valuable option for stochastic predictors when the distribution of the predictors is unknown. Even when the distribution is known, the difference seems to be tiny. A benefit of going the full Bayesian way is that we can also include prior knowledge about the covariance matrix. For stochastic predictors, compared to non-stochastic predictors, the uncertainty about the covariance matrix is reflected in the larger credible intervals. Even when we would know the exact regression parameters, there is a lot of uncertainty in the LMG values caused by the uncertainty about the covariance matrix.

Table 1.5: LMG values assuming stochastic predictors with 95% CI.

Variable	Relaimpo	Bayesian framework	
		nonparameteric bootstrap	covariance inference
x1	0.092 (0.02, 0.221)	0.047 (0.005, 0.144)	0.078 (0.018, 0.17)
x2	0.109 (0.026, 0.241)	0.15 (0.055, 0.279)	0.156 (0.052, 0.273)
x3	0.636 (0.452, 0.757)	0.554 (0.395, 0.71)	0.561 (0.442, 0.66)
x4	0.03 (0.008, 0.11)	0.034 (0.008, 0.097)	0.033 (0.011, 0.103)

1.2 Empirical Data

In the following section, the Bayesian LMG implementation is applied on an empirical dataset containing test scores of pupils ($N = 301$) from a study by [Holzinger and Swineford \(1939\)](#) available in the R package `MBESS` ([Kelley, 2017](#)). This dataset was used in [Nimon *et al.* \(2008\)](#) to present commonality analysis, which is another variance decomposition technique. Scores from a paragraph comprehension test (paragrap) was predicted by four verbal tests: general-information (general), sentence-comprehension (sentence), word-classification (wordc), and word-meaning (wordm) (Table 1.6).

The aim of the regression analysis was to determine the strength of association between verbal ability and paragraph comprehension. An overview of the data is shown in Figure 1.3. The regression results of the paragraph comprehension scores on the verbal test scores are shown Table 1.7). A novice researcher may wrongly conclude, that there is little association between the *non-significant* predictors (general information and word classification) and paragraph comprehension. Given the other predictors are already included in the model, the predictors seem not to provide us much information about the expected paragraph comprehension ability. However, we can not conclude out of this regression table, that the association between any of these *non-significant* predictors and the dependent variable is unimportant. We see in Figure 1.3 that the correlations between the predictors is rather high. The LMG metric may therefore provide us some new information about the importance of each predictor.

The Bayesian regression model was fitted in `rstanarm`. The default priors were used for the regression coefficients and the σ^2 parameter. A burn-in period of 20000, a sample size of 20000 and a thinning of 20 resulted in a posterior sample size of 1000. The first few posterior samples are shown in Table 1.8. The resulting R^2 of of these posterior samples is shown in Table 1.9. The LMG values were calculated by using `hier.part`. The independent component (I), joint contribution (J) and total explained variance in a predictor-only model (T) are shown in Table 1.10. Sentence-comprehension and word-meaning seem to be the most important predictors by the LMG metric. However, none of the predictors seem to be unimportant. The joint contributions of each predictor is quite large.

For comparison purposes, the LMG metric was additionally calculated with the `relaimpo` package using parametric bootstrapping. The confidence intervals of `relaimpo` are almost identical to the credible intervals of the Bayesian framework (Table 1.11). Assuming stochastic or non-stochastic predictors would also result in almost identical uncertainty estimates with such a large sample size (Table 1.12).

Table 1.6: Variable description

Variable Name	Description
paragrap	scores on paragraph comprehension test
general	scores on general information test
sentence	scores on sentence completion test
wordc	scores on word classification test
wordm	scores on word meaning test

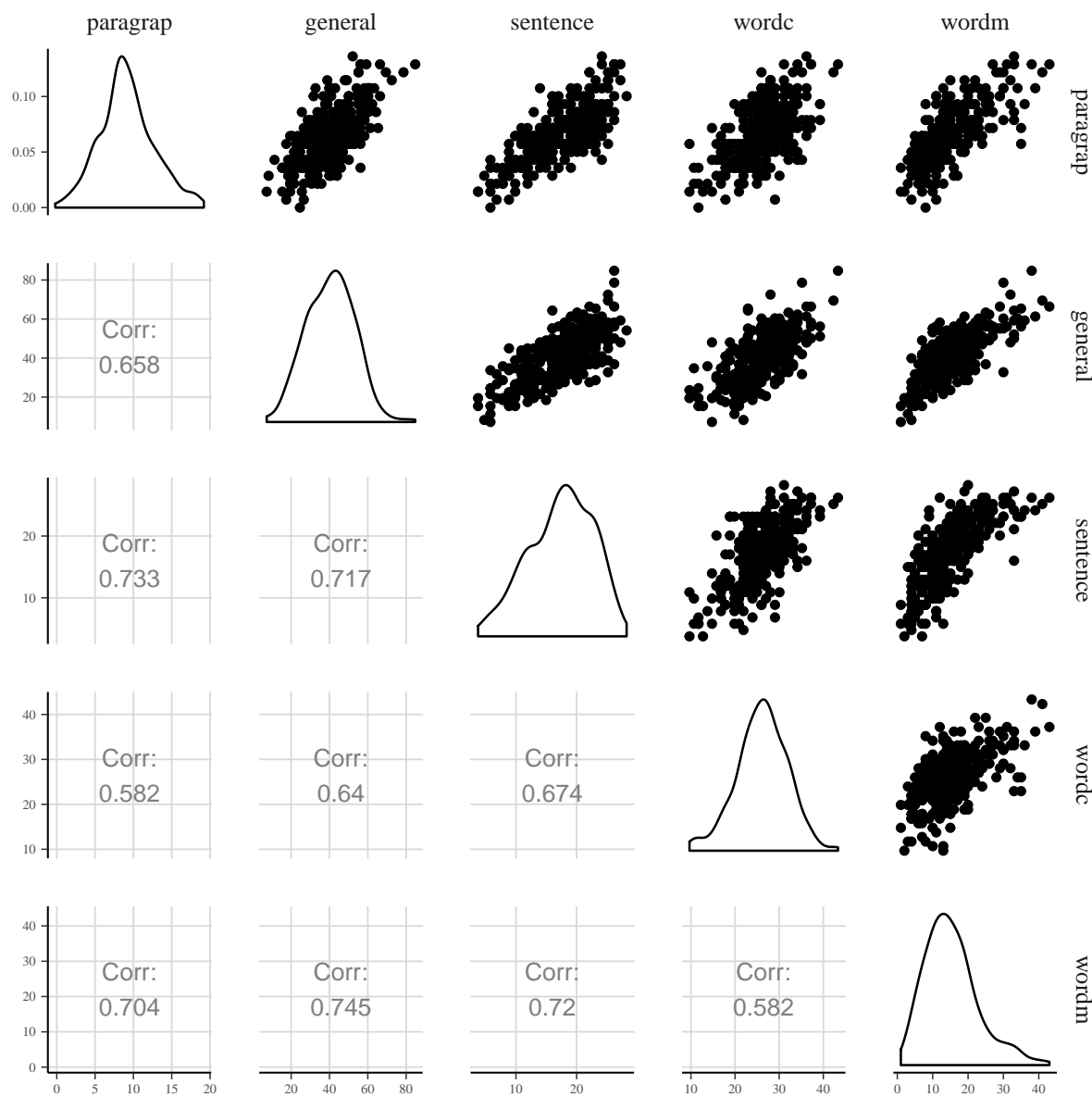


Figure 1.2: Test scores from Holzinger and Swineford’s (1939) Study. N=301

Table 1.7: Regression of paragraph comprehension on verbal tests.

	Coefficient	95%-confidence interval	<i>p</i> -value
Intercept	0.071	from -1.17 to 1.31	0.91
general	0.03	from -0.00 to 0.06	0.084
sentence	0.26	from 0.18 to 0.34	< 0.0001
wordc	0.047	from -0.01 to 0.11	0.14
wordm	0.14	from 0.08 to 0.19	< 0.0001

Table 1.8: Samples from the posterior distributions of the regression parameters

	general	sentence	wordc	wordm	sigma
sample 1	0.041	0.270	0.014	0.113	2.18
sample 2	0.036	0.243	0.010	0.166	2.09
sample 3	0.043	0.295	-0.007	0.103	2.28
sample 4	0.024	0.240	0.051	0.192	2.19
sample 5	0.050	0.297	0.009	0.099	2.28
sample 6	0.027	0.259	0.040	0.157	2.09
sample 7	0.029	0.265	0.023	0.156	2.11
sample 8	0.017	0.259	0.026	0.168	2.18
sample 9	0.036	0.250	0.073	0.123	2.00
sample 10	0.004	0.269	0.059	0.163	2.10

Table 1.9: R^2 for all submodels for the first six posterior samples

	sample 1	sample 2	sample 3	sample 4	sample 5	sample 6
none	0.000	0.000	0.000	0.000	0.000	0.000
general	0.426	0.455	0.407	0.449	0.433	0.448
sentence	0.520	0.537	0.508	0.541	0.527	0.554
wordc	0.299	0.312	0.271	0.348	0.298	0.343
wordm	0.462	0.542	0.433	0.559	0.446	0.534
general sentence	0.558	0.583	0.541	0.583	0.566	0.592
general wordc	0.455	0.482	0.428	0.493	0.459	0.490
general wordm	0.511	0.578	0.482	0.588	0.504	0.569
sentence wordc	0.527	0.545	0.511	0.557	0.533	0.567
sentence wordm	0.574	0.627	0.551	0.640	0.571	0.633
wordc wordm	0.497	0.568	0.462	0.595	0.484	0.573
general sentence wordc	0.559	0.583	0.541	0.587	0.566	0.595
general sentence wordm	0.583	0.634	0.560	0.644	0.583	0.638
general wordc wordm	0.523	0.587	0.490	0.606	0.516	0.588
sentence wordc wordm	0.576	0.628	0.552	0.645	0.572	0.636
all	0.584	0.634	0.560	0.647	0.583	0.640

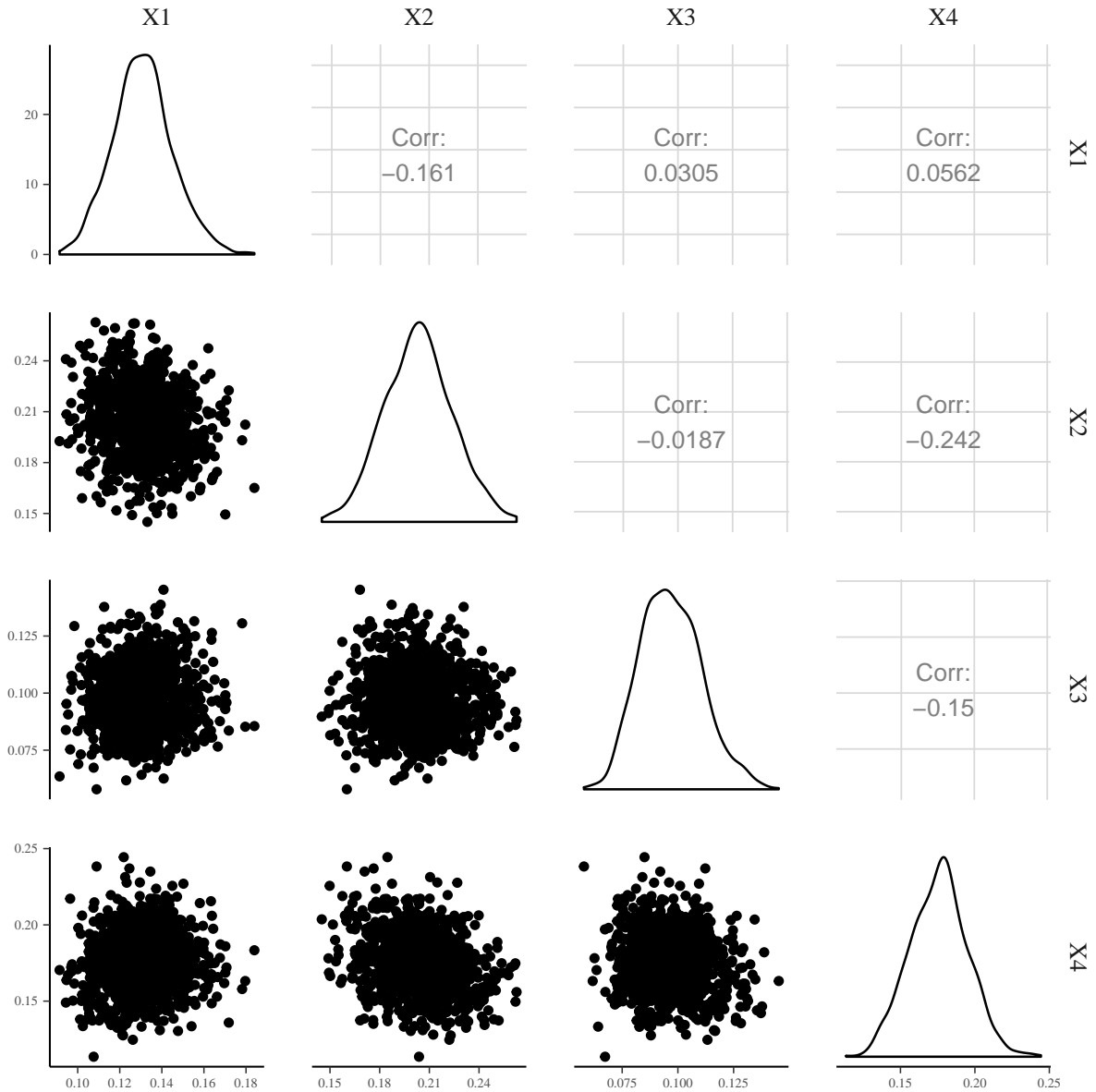


Figure 1.3: LMG posterior distribution of different verbal ability tests

Table 1.10: Variance decomposition for non-stochastic predictors. I = LMG values, J = joint contribution, Total = total explained variance in one-predictor only model

Variable	I	J	Total
general	0.13 (0.105, 0.16)	0.299 (0.258, 0.333)	0.429 (0.364, 0.489)
sentence	0.203 (0.165, 0.245)	0.328 (0.293, 0.355)	0.532 (0.465, 0.588)
wordc	0.096 (0.073, 0.128)	0.24 (0.2, 0.275)	0.336 (0.273, 0.401)
wordm	0.176 (0.138, 0.214)	0.315 (0.278, 0.344)	0.492 (0.421, 0.55)

Table 1.11: Variance decomposition for non-stochastic predictors. I = LMG values, J = joint contribution, Total = total explained variance in one-predictor only model

Variable	LMG value (95%-CI)	
	Relaimpo	Bayesian framework
general	0.131 (0.106, 0.161)	0.13 (0.105, 0.16)
sentence	0.206 (0.168, 0.251)	0.203 (0.165, 0.245)
wordc	0.096 (0.073, 0.125)	0.096 (0.073, 0.128)
wordm	0.178 (0.141, 0.219)	0.176 (0.138, 0.214)

Table 1.12: Variance decomposition for Stochastic predictors. I = LMG values, J = joint contribution, Total = total explained variance in one-predictor only model

Variable	LMG value (95%-CI)	
	Relaimpo	Bayesian framework
general	0.131 (0.106, 0.163)	0.127 (0.092, 0.16)
sentence	0.206 (0.171, 0.246)	0.204 (0.163, 0.249)
wordc	0.096 (0.075, 0.128)	0.097 (0.07, 0.128)
wordm	0.178 (0.144, 0.218)	0.175 (0.136, 0.214)

Bibliography

- Holzinger, K. J. and Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, **48**, 1–91. [6](#)
- Kelley, K. (2017). Mbess (version 4.0.0 and higher) [computer software and manual]. R package version 4.0.0 and higher. [6](#)
- Link, W. A. and Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, **3**, 112–115. [2](#)
- Nimon, K., Lewis, M., Kane, R., and Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, **40**, 457–466. [6](#)
- Stan Development Team (2016). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.13.1. [1](#)
- Stan Development Team (2017). *Stan Modeling Language: User’s Guide and Reference Manual*. [1](#)