Instructions:

- Answer the following questions using SAS and provide the code you used to perform the analysis in a separate file (`.txt` extension, `utf8` encoding).
- Your report must be submitted as a PDF file and should not exceed 15 pages; any additional page will be ignored. Be brief, but precise; only include relevant output.
- Errors are penalized even if they are not directly related to the question.

3.1 **Teaching to read**: the data set `bauman` is drawn from

J. Baumann, N. Seifert-Kessell, L. Jones (1992), *Effect of Think-Aloud Instruction on Elementary Students' Comprehension Monitoring Abilities*, Journal of Reading Behavior, **24** (2), pp. 143–172.

Researchers conducted a study to determine the efficiency of three learning methods for reading. The sample consists of 66 fourth-grade students from an elementary school. The students, 32 girls and 34 boys, were randomly split between three groups. Interest lies in improvement over the default method, directed reading (DR). Two tests were administered before and after the experiment to monitor the effectiveness of the methods; to make these comparable, they were rescaled so that a total of 1 means perfect score.

The `bauman` data contains information about the following variables:

- `group`: experimental group, one of directed reading-thinking activity (`DRTA`), think-aloud (`TA`) and directed reading group (`DR`).
- `mpre`: average pre-test prediction score (standardized) for average of standardized error detection test and comprehension monitoring questionnaire.
- `mpost`: same as `mpre`, but post-test score.
- `dpp`: difference between post-intervention results and pre-intervention results, `mpost-mpre`.

In this first part, we are interested in the improvement in scores and two models are fitted to assess this.

(a) In their paper, Baumann *et al.* run a one-way analysis of variance (ANOVA) for "pre-tests" `mpre` with the `group` factor. Explain what is the purpose of doing such a test in the context of the study.

(b) Fit a one-way ANOVA for `dpp` = `mpost` − `mpre` with `group` as factor (Model 3.1.1). Write down the model equation in terms of `mpost` and show that it is a special case of a linear regression with an offset.

(c) Compare the one-way ANOVA model for `dpp` with `group` to a linear regression model with `mpost` as response and `mpre` and `group` as covariates (Model 3.1.2). Given the output of the latter, is the ANOVA model adequate? Justify your answer.

Transform the dataset from wide to long-format; the latter is more suitable for longitudinal studies. In addition to `group`, your data should contain the following columns

- `id`: unique student identification number.
- `score`: average result for evaluation.
- `test`: categorical variable, one of `mpost` or `mpre` indicating whether the score reported is pre-test or post-test.

Table 1 contains the first 10 lines of the dataset in long format.

We will consider additionally two models for `score` as function of `group` and `test` and an interaction term between the two, but with different covariances for the two scores of students:

- Model 3.1.3 assumes a compound symmetry model;
- Model 3.1.4 assumes an unstructured covariance model.

(d) Explain what is the fundamental difference between Model 3.1.2 and Model 3.1.3–3.1.4.

(e) Write down the covariance matrix implied by Model 3.1.3 and report the estimated correlation between the pre-test and the post-test scores for any student.

(f) Using the output of Models 3.1.3 and 3.1.4, test whether the variability of the mean pre-test and post-tests is the same. Specifically, write down the name of the test, the numerical value of the statistic and the p-value before concluding.

group	test	score	id
DR	mpre	0.23	1
DR	mpost	0.27	1
DR	mpre	0.35	2
DR	mpost	0.42	2
DR	mpre	0.41	3
DR	mpost	0.24	3
DR	mpre	0.57	4
DR	mpost	0.39	4
DR	mpre	0.67	5
DR	mpost	0.56	5

Table 1: First ten lines of the Baumann data in long format

(g) Since the data are longitudinal, one could consider fitting, in addition to Model 3.1.3–3.1.4, a first-order autoregressive covariance model, $\mathsf{AR}(1)$. Would it be useful in this case? Justify your answer.

(h) Up till now, we assumed that the covariance matrix of the pre- and post-intervention scores is the same for all students in Models 3.1.3 and Models 3.1.4. One could however postulate that the parameters of the covariance matrix in Model 3.1.3 differs from one reading teaching method to the next. Is this hypothesis supported by the data?

(i) Use Model 3.1.4 to determine if the teaching methods `DRTA` and `TA` significantly improve over the standard teaching method of directed reading `DR`.

3.2 The dataset `goldstein` originates from the study

> H. Goldstein *et al.* (1993). *A Multilevel Analysis of School Examination Results*, Oxford Review of Education, **19** (4), pp. 425–433.

The authors analyse exam results from inner London schools and student to study study the between-school variation in order to rank schools. The following description, drawn from the OpenBugs examples, describes the data in more detail:

> Standardized mean examination scores were available for 1978 pupils from 38 different schools. Pupil-level covariates included gender plus a standardized London Reading Test (LRT) score and a verbal reasoning (VR) test category (1, 2 or 3, where 1 represents the highest ability group) measured when each child was aged 11. Each school was classified by gender intake (all girls, all boys or mixed) and denomination (Church of England, Roman Catholic, State school or other); these were used as categorical school-level covariates. Both the London reading test score and the verbal reasoning test were performed at the beginning of the year.

The `goldstein` data contains the following variables:

- `score`: standardized end-of-year exam score for each pupil,
- `school`: school id,
- `LRT`: London reading test score,
- `VR`: verbal reasoning (VR) test category (`1`, `2` or `3`, where `1` represents the highest ability group and `3` the lowest),
- `gender`: gender of pupil, either female (`0`) or male (`1`),
- `type`: type of school, either all-girl schools, all-boy school or mixed,
- `denom`: denomination of school, either Church of England, Roman Catholic, state or other.

(a) Give the range of the number of pupils per school and use this information to determine if it is feasible to estimate a fixed (group-)effect for `school`.

(b) Write down the equation of the postulated (theoretical) model for `score` that includes `LRT`, `VR`, `sex`, `type` and `denom` as fixed effects and `school` as random effect, with `VR=3`, mixed for `type` and other for `denom` as baseline categories.
Don't forget to specify the distribution of the errors and random effects.

(c) One could consider adding `VR` as random effect as opposed to fixed effect. Which of the two makes the most sense and how do the models differ conceptually?

(d) Using the fitted model (with a random intercept for `school`),
   i. Report the estimated covariance parameters
   ii. Report obtain the estimated covariance matrix for school 37
   iii. Briefly explain how to obtain the latter given the estimated covariance parameters.
   iv. Write down the proportion of the total variance due to school.

(e) Produce a normal quantile-quantile plot of the predicted random effects for school and hence comment on the model assumption for the random effect.

(f) The goal of Goldstein *et al.* (1993) was to rank schools. What is the benefit of pooling information from schools using random effects in order to estimate their average score?

(g) Plot the predicted school effect as a function of school id, with prediction intervals based on the formula $\widehat{b}_i \pm 1.96\mathrm{se}(\widehat{b}_i)$ (you may need to use the formula to calculate the bounds).

(h) What is the predicted top five school ranking?