ST344 Group Coursework

subject to moderation

October 2023

An outline of the task

The coursework will be on Sustainable Development Goals (SDGs) set up by the United Nations (see https://sdgs.un.org/goals). Data sets for some specific indicators for the first three SDGs (SDG1: 'No Poverty'; SDG2: 'Zero Hunger'; SDG3: 'Good Health and Well-being') are provided in the Moodle. You will be expected to use data for Brazil, Japan and Russia. Your team must also choose seven other countries, and ensure that you have at least one country from the eight Regional Groups defined by the United Nations:
Sub-Saharan Africa, Northern Africa and Western Asia, Central and Southern Asia, Eastern and South-Eastern Asia, Latin America and the Caribbean, Australia and New Zealand, Oceania, Europe and Northern America. You do not need to consider the subgroups within these eight regions.

Begin by discussing how your team will divide the work. All teams must consider SDG1, SDG2 and SDG3. Teams with only 5 members also consider SDG1 and SDG2, but may choose to consider only two of the three indicators under SDG3. Teams with 7 members are expected cover more breadth[1] (with an additional 500 words, 1 table, 1 figure allowed - as needed - in addition to the specifications given below).

You are expected to create sensible summary tables and plots. Your exploratory data analysis should inform a discussion of the quality of the data, and the challenges in obtaining data.

You are expected to read the following article provided on the Moodle (2023Nemerimana.pdf):
M. Nemerimana, S. Havugarurema et al. (2023), Factors associated with recovery from stunting at 24 months of age among infants and young children enrolled in the Pediatric Development Clinic (PDC): A retrospective cohort study in rural Rwanda. *PLOS ONE*, 18: 1-17.

You are expected to relate the article to the indicators you summarise, and to discuss the role of detailed country level studies in progress towards achieving the SDGs. You are expected to interpret Table 4, to comment the choice of measure for stunting, and whether alternative statistics models would be useful, with reasons. You might consider whether randomised trials would be better than the use of routine medical data. Optional additional reading on Code of Professional Ethics, ethics of research in developing countries, and ethics of cluster randomised trials will be provided.

Variation between groups

The full data sets will be available from the Moodle page. It is unlikely that you will all choose the same countries from the Regional Groups, but this will be checked on your project plans.

[1] For example, by identifying, discussing or reviewing additional relevant research literature, or further developing statistical modelling ideas, or broadening the analysis to include more countries.

Data and material provided

Goal	File Name	Type
SDG 1	poverty-surveys-per-decade.csv	Metadata
SDG 1	GDP-povertyPercentages.csv	Data for countries
SDG 1	share-extreme-poverty3.65dollarDay.csv	Data for countries, groups
SDG 1	total-population-in-extreme-poverty.csv	Data for countries, groups
SDG 2	Metadata-Country-Stunting.csv	Metadata
SDG 2	Metadata-Country-UnderFed.csv	Metadata
SDG 2	Metadata-Regions-Stunting.csv	Metadata
SDG 2	StuntingU5yrs2023.csv	Data for countries
SDG 2	StuntingCountryOview.Rda	Summary by country as R file
SDG 2	Underfed.csv	Data for countries
SDG 3	Metadata-Country-MaternalMortality.csv	Metadata
SDG 3	Metadata-Regions-MaternalMortality.csv	Metadata
SDG 3	Metadata-NeonatalDeaths.csv	Metadata
SDG 3	MaternalMortalityRate.csv	Data for Countries 1998-2018
SDG 3	MaternalMortOview.Rda	Summary by Entities as R file
SDG 3	NeonatalDeathRates.csv	Neonatal mortality rate 1998-2021
SDG 3	Under5Drate.Rda	
SDG 3	Under5DrateOview.Rda	Metadata
	M. Nemerimana et al (2023) *PLOS one*	Article - Stunting
	ISI Code of Professional Ethics	Official statistics ethics statement
	Hutton (2000) Stat. Meth. Med. Res	Article – country & professional ethics

Documents, deliverables and marking

General advice

You should document your choices, reasoning and approaches, results found when applying your approaches, and your conclusions about the effectiveness (or otherwise) of the methods you chose. Comment on whether you can see an effect of Covid-19 on the statistics. Based on your work, you should also include recommendations about what data might be useful in the future, and about how one might best tackle these challenges if there were no constraints on time or budget.

In the course of doing this work you might come across published work by other people, in this same topic area. If you do get ideas from someone else's work, or if you want to mention someone else's findings in your report, be sure to credit the prior publication, and to cite the source accurately. The team might well be interested to know of other such work on the topic. (But the main interest will be in your group's own, new exploration of this data.) Your report will be scanned by the University's Turnitin software system — for general guidance on how to avoid plagiarism, see the one-hour online tutorial at http://warwick.ac.uk/plagiarwise .

Deliverables (from each team):

1. **Team Project Plan (Plan of work and Team Contract).** By Thursday 2nd November, 13:00 (week 5).
   Counts for 5% of ST344 module marks.

2. **Team Report.** By Thursday 30th November, 13:00 (week 9).
   Counts for 30% of ST344 module marks.

3. **Team presentation.** This will be scheduled for Tuesday Week 2 of Term 2 (16 January 2024). Presentation slides (1 file per group) to be submitted by Thursday 11th January 2024, 13:00 (Term 2 Week 1).
   Counts for 10% of ST344 module marks.

In addition, also by Thursday 30th November, 13:00 (week 9), each group should submit its jointly agreed statement – just a short paragraph – about which team members contributed to which parts of the work, and the suggested allocation of marks among the team members (see details below).

Mark Schemes

Team Project Plan: Documents to submit:

1. Plan of Work *on MS Planner.*
   The plan should list the countries you have chosen, give a realistic timeline for the work, the deliverables, with appropriate milestones and meeting dates identified and provisional allocation of tasks to group members **(3 marks)**.

2. Team Contract
   An initial start can be found in a file on the Moodle. The names of each team member must be printed on the document. You must print the contract as 1 or 2 sides of A4, and each team member must sign and date their signature. One team member will need to upload a scan of the document **(2 marks)**.

   A Team Contract that is not signed by all team members will receive a mark of 0 (out of 2), unless mitigating circumstances are reported.

Team report: A formal report, to professional standards. The report should have two parts, both submitted electronically:

1. A PDF document named ST344TeamX.pdf, with 'X' replaced by your team number for the report itself. This document should be written for intelligent readers who do not necessarily have advanced statistical training. The limits are: 3000 words, 7 figures, 8 tables, minimum body-text font size 11pt, and minimum 2cm margins all round. Recommended length: 8 A4 sides including graphs, tables, references and word count.

2. A 'technical appendix' in the form of a .zip archive, which contains a .Rmd file giving full details of the analysis that is reported, along with any other files that are needed in order to allow the .Rmd file to compile, and the analysis to be reproduced.

Marks will be allocated as follows: (Note that this is approximate and may be subject to slight moderation)

- Report structure and presentation: **8 marks**
- Appropriate and well explained analysis, discussion and conclusions: **15 marks**
- Clear, well documented and functioning code in the .Rmd document: **7 marks**

Team presentation: A professional oral presentation of the analysis and main findings, for an audience of fellow statisticians. Presentations are planned to take place in two 2-hour sessions on Tuesday week 2 of Term 2. All teams are attending all presentations. An allocation of the time slot will be published on the Moodle as soon as available. All group members should contribute roughly equally to delivery of the presentation. Marks allocated as follows:

- Structure and pace: **5 marks**
- Clarity of presentation: **5 marks**

Mark allocation: Each team should decide how to distribute the group mark by allocating to each team member a share of $n \times 100\%$ where n is the number of students in the team. This will act as a weighting factor to convert the group mark into an individual mark. For example, suppose the group mark is 70% and a team of 4 students decides to allocate 100% to each team member, then each member receives the mark of 70%. On the other hand if the team decides to allocate 106% to one team member and 98% to the other three members, then the former receives a mark of 74.2% and the other three team members receive the mark 68.6%. The maximum weighting factor that can be awarded is 110%, the minimum weighting factor is 90%. The module leaders reserve the right to moderate the weighting factors, impose equal weighting factors or request further evidence.