Feedback for ADS2 Group Project

Your report is knittable and nicely formatted.

You used diverse tools for formatting.

Some general issues:

-   Figure and table captions should have been written differently. You wrote them in the text directly. It led to the situation that some captions got disjoint from the respective figure or table. It is especially clear in Tables 3.1 and 3.2. You should have specified captions in another way: in the code chunk options (`fog.cap`), `kable(…, caption = "Table title")`, MD table caption (if you prefer using the in-built tables), or anyhow else.

-   For the sake of saving space, it is advised to not show outputs of all code chunks or show just trimmed results. But it is better to show \*at least some\* intermediate results!

-   I feel like in general there could have been a bit more discussion of the results (as opposed to just describing them).

-   It would be reasonable to get rid of the non-essential ending of the region names: `Europe & Central Asia - WB` to `Europe & Central Asia`.

-   Inconsistent approaches for hypothesis tests: sometimes, you use unpaired tests, sometimes -- paired.

I like the way you diagnosed your data (at the beginning of the paper).

But it would be better to leave more detailed comments about the results of your diagnostics.

For instance, the results of your `str(substance\_use)` is somewhat unclear.

It would be better to code `age` as a numeric covariate or as an ordered factor.

Later, it will affect analysis negatively by creating multiple dummy variables that will make the results difficult to interpret.

I would disagree with your claims about Figure 0 (BTW, why not Figure 1?).

Many pieces of data show a non-normal behavior.

Which is itself something interesting.

Overall, using a boxplot is a controversial choice for this case.

You answered question 1 correctly.

Your data preparation code is really nice and apt.

But I would prefer to get a plot and some other regions included for a better contrast.

您的报告可编织，格式规范。

你使用了不同的工具进行格式化。

一些一般性问题：

- 图和表的标题应该换一种写法。您直接将其写入了正文。这导致一些标题与相应的图或表脱节。这在表 3.1 和 3.2 中尤为明显。您应该以另一种方式指定标题：在代码块选项（`fog.cap`）、`kable(..., caption = "Table title")`、MD 表格标题（如果您喜欢使用内置表格）或其他任何方式。

- 为了节省空间，建议不要显示所有代码块的输出结果，或者只显示经过修剪的结果。但最好至少显示一些中间结果！

- 我觉得总的来说，可以对结果进行更多的讨论（而不仅仅是描述）。

- 去掉地区名称中的非必要结尾是合理的："欧洲和中亚-世行 "改为 "欧洲和中亚"。

- 假设检验的方法不一致：您有时使用非配对检验，有时使用配对检验。

我喜欢你诊断数据的方式（在论文开头）。

但最好能就诊断结果留下更详细的评论。

例如，您的 "str(substance\_use) "结果有些不清楚。

最好将 "年龄 "作为一个数字协变量或有序因子来编码。

之后，它会产生多个虚拟变量，使结果难以解释，从而对分析产生负面影响。

我不同意你关于图 0 的说法（顺便说一下，为什么不是图 1？）

很多数据都显示了非正态行为。

这本身就很有趣。

总的来说，在这种情况下使用方框图是一个有争议的选择。

你正确回答了问题 1。

你的数据准备代码非常漂亮和贴切。

但我更希望能得到一个包含一些其他区域的曲线图，以形成更好的对比。

You answered question 2 correctly.

Your plots are nice.

But there are several important issues:

-   The figures are valid and it is nice that you used available CIs (BTW, which CI? 95%? 99%?), but there is a problem: different graphs have different Y-scales. This visualization is suboptimal, as it makes it difficult to compare different age cohorts. It would be better to plot all the groups on one graph using different colors and line types. Similar to what was done in Figure 2.2.

-   From this, we come to another point: Figure 2.1 is redundant.

-   The conclusions about males are incomplete. But still reasonable.

-   It would be more reasonable to use unpaired tests for hypothesis testing.\

Data preparation was done nicely.

The figures are nice and illustrative, but they suffer the same issues as Figures 2.1 and 2.2.

Figure 3.1 is redundant, and Figure 3.2 could have been combined into one item.

In any case, your analysis would be more illustrative if you used the same Y-scales.

Table 3.1 is nice, but there are several important points to consider: 1.

The term effect size is not completely correct here, 2.

Such a hypothesis-testing approach cannot be used in this case.

You are comparing proportions from the different time periods, 3.

The table should have been named in the `kable()` function.

Otherwise, it is hard to attribute to anything.

I assume you used unpaired tests for your hypothesis testing.

Same with Tables 3.2 and 3.3.

In Table 3.3, I like the idea of testing the difference between both slopes.

But the table is incomplete as there are no tests of the statistical significance of this difference.

Also, your analysis would be much better if you coded `age` not as an unordered factor, but as numeric data or an ordered factor.

Currently, Tables 3.1-3.3 are difficult to interpret.

As for the last part of this question, it is nice that you are aware of the need to test assumptions for the linear model.

But you should have done it \*before\* you drew conclusions.

你正确回答了问题 2。

你的情节很好。

但有几个重要问题：

- 数字是有效的，您使用了可用的 CI（顺便说一下，哪个 CI？95%？ 99%？这种可视化效果并不理想，因为它很难比较不同年龄段的人群。最好是使用不同的颜色和线型将所有组别绘制在一张图上。类似于图 2.2 中的做法。

- 由此，我们想到了另一点： 图 2.1 是多余的。

- 关于男性的结论并不完整。但仍然是合理的。

- 使用非配对检验进行假设检验会更合理。

数据准备工作做得很好。

数据准备工作做得很好，图表也很好，很能说明问题，但也存在图 2.1 和图 2.2 的问题。

图 3.1 是多余的，图 3.2 可以合并为一项。

无论如何，如果使用相同的 Y 尺度，您的分析将更能说明问题。

表 3.1 很好，但有几个要点需要考虑： 1.

这里的效应大小一词不完全正确；2.

在这种情况下不能使用假设检验方法。

您是在比较不同时期的比例，3.

表格应该在`kable()`函数中命名。

否则，很难将其归因于任何事物。

我想您在进行假设检验时使用了非配对检验。

表 3.2 和表 3.3 也是如此。

在表 3.3 中，我喜欢测试两个斜率之间差异的想法。

但该表并不完整，因为没有对这种差异的统计显著性进行检验。

另外，如果您不把 "年龄 "作为无序因子，而是作为数值数据或有序因子进行编码，您的分析结果会更好。

目前，表 3.1-3.3 难以解释。

至于问题的最后一部分，很高兴您知道需要检验线性模型的假设。

但你应该在得出结论之前\*\*进行检验。

Also, it would be better to actually show the output of this code.

And there is no object `model` in the previous text.

Thus, which `model` are you talking about?

Unfortunately, the conclusions are not completely clear.

Your research question is not too ambitious but is quite clear.

Your visualizations are nice, but it would be better to plot men and women in Figure 4.2 and remove Figure 4.1 as it would become redundant.

Also, the Y-axis label of Figure 4.1 is wrong.

Figure 4.2 still makes sense, but given the question you stated, it makes sense.

Overall, I like the way how you decided to check your hypothesis.

Hypotheses for your ANOVA model are not formulated clearly.

I like the way you were thinking, but the question was clearly not finished.

另外，最好能实际显示这段代码的输出结果。

而且在前面的文本中没有对象 `model`。

因此，你说的是哪个 `model`？

很遗憾，结论并不完全清楚。

你的研究问题不太宏大，但相当清晰。

你的可视化效果很好，但最好在图 4.2 中绘制男性和女性，去掉图 4.1，因为这会变得多余。

另外，图 4.1 的 Y 轴标签是错误的。

图 4.2 还是有意义的，但考虑到你提出的问题，它也是有意义的。

总的来说，我喜欢你检验假设的方法。

你的方差分析模型的假设表述不清楚。

我喜欢你的思维方式，但问题显然没有说完。