1. hypothesis
   The number of textual provisions in the moderation guidelines of a Facebook group has a significant impact on the average number of posts and comments per person over a specific period.

2. null hypothesis
   The number of textual provisions in the moderation guidelines of a Facebook group has no significant impact on the average number of posts and comments per person over a specific period.

3.
   - independent variable
     The number of textual provisions in the Facebook group's moderation guidelines (measured by the number of distinct rules).
   - dependent variable
     The average number of posts and comments per person over a specific period
   - confounding variables
     (a) group topics
     (b) user demographics
     (c) real-world events

4. data collection and management

   (a) Sample Selection (10 hours)
   - Identify 20 public and 20 private Facebook groups.
   - 10 AUD per group, as incentives for admins to gain access to data.

   (b) Moderation Guidelines Collection (20 hours)
   - Manually code the number of provisions across all groups.
   - 20 AUD per hour

   (c) Posts Data Collection(30 hours)
   - Use Facebook's Graph API to pull data on posts and comments over 1-month period, as well as the number of active members.
   - API and Data Extraction Tools: 100 AUD.

   (d) Confounding Variables Collection (10 hours)
   - Record or analyze group category and post topics, etc.
   - 10 AUD per hour

5. statistical test
   I choose Spearman's rank correlation coefficient to do the test.

   Assumptions:

(a) The moderation guidelines and activity metrics are measured on an ordinal scale.

(b) The data is likely to be skewed.

(c) Each Facebook group's data is independent.

(d) The relationship between the two variables is monotonic.

Since I have assumed that the independent and dependent variables are **ordinal** rather than continuous, and the data are highly likely to be **skewed** rather than normally distributed, a **non-parametric test** is required. Among non-parametric tests, Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described, which aligns perfectly with my hypothesis of exploring the relationship between variables. Additionally, I have assumed a **monotonic relationship** rather than a linear relationship between the two variables, so Spearman's coefficient, which determines the strength and direction of the monotonic relationship is more appropriate than Pearson's coefficient.

6. limitations and investigations
One major limitation is the potential sample bias: by only analyzing Facebook groups with accessible and clearly defined guidelines, the sample may not represent all types of groups, especially less active ones. Additionally, the coding of textual provisions is inherently subjective — differences in language complexity, context, or enforcement may not be fully captured by a simple count. What's more, confounding variables such as group size, topic, or external events may also influence user engagement, making it difficult to isolate the effect of moderation guidelines.

Future research could mitigate these issues by employing a longitudinal design to track changes within a same group over time. The quantitative analysis of guidelines can also be accompanied by more nuanced qualitative coding, such as a questionnaire of satisfactory, to better assess guideline quality, and incorporating additional control variables like group size and topic, to better isolate the effects of moderation practices.

7. alternative narrative
The relationship between the number of provisions and average posts per person may be confounded by group topics. Fewer provisions might indicate lighter topics, thus increasing activity. Testing the monotonic relationship between topics and posts collected from datasets could reveal associations, suggesting provisions are not the sole primary factor.