2025年 応用心理学 Ⅰ ワークシート

学籍番号：

氏名：

（　　　班）

以下の課題を実施してください。

それぞれの課題に使ったRスクリプト（コード）も下に入力してください。

課題１

データのクリーニングを行って、聞き書き全体（４班分）での上位頻出単語30件を確認しましょう。

1. 内容と関係ない記号や単語（ストップワード）を除外してください。
2. 上位頻出単語30件を棒グラフで出力した結果を貼り付けてください。

ポイント：以下の処理が必要です

・ExcelまたはCSV/TSVデータを読み込む

・読み込んだデータを結合する

・単語を集計する

・棒グラフで図示する

課題２

クリーニングしたデータを使って、あなたの班の聞き書きデータの上位頻出語30件を確認しましょう。棒グラフで出力した結果を貼り付けてください。

課題３

クリーニングしたデータにTF-IDFを使って、各班の特徴的な語彙を30件ずつ抜き出しましょう。

1. 棒グラフで出力した結果を貼り付けてください。
2. 課題２および課題３の結果は、どの程度あなたの班の聞き書きの内容を反映していましたか。最も適切に反映されていた話題、最も反映されていなかった話題はなんだったかそれぞれ説明してください。

課題４

聞き書きデータにはそれぞれ話題に関するタグがつけられています\*。

1. TF-IDFを使って、各話題の特徴的な語彙を30件ずつ抜き出しましょう。棒グラフで出力した結果を貼り付けてください。
2. 各話題について発言を単位として共起分析を行いましょう。Bag of Wordsを用いた共起分析を行い、話題ごとに共起ネットワークのグラフを出力し、貼り付けてください。
3. ①と②の結果から読み取れる各話題の特徴について説明してください。

\* ひとつの発言に複数のタグが付いている場合がありますが、重複は無視してタグの有無だけで判断してください。