PWS Cup 2023 競技参加の手引き

PWS 2023 実行委員会

2023 年 8 月 29 日

はじめに

本文書では，PWS Cup 2023 (本コンテスト) の競技ルールや参加方法について説明する．本コンテストは本文書に記載した内容に従って実施される予定である．本文書は合理的で公平なコンテストの実施の為に予告なく変更することがある．

本文書の推奨する使い方は，以下の通りである．まず，第 1 章を読み，本コンテストの流れを理解する．続いて，第 6 章のルールと 7.1 節の注意事項をを確認し，第 2 章 に従ってコンテストに取り組む．その際，第 3 章と第 4 章，第 5 章 に詳細な情報が記述されているので，必要に応じて参照する．

目 次

第1章　コンテストの基本設計

1.1　概要

本コンテストでは，表形式で与えられる個人データの集合を対象に，有用性と (属性暴露攻撃を想定した) 安全性をともに高く保ったデータに加工する技術を競う．

1.1.1　加工対象のデータ

本コンテストで加工対象とするデータは，各行が 1 個人の特徴を表す表形式のデータである．データの各行には 1 個人の情報が，各列には居住エリアや年代などの列ごとに特定の特徴の情報が, それぞれ格納されているものと想定する．

1.1.2　加工後のデータに求められる性質

加工後のデータは，有用性と安全性がともに高いことが求められる．

1.1.3　コンテストの流れ

本コンテストは**審判** J と複数の**チーム** $T_1, T_2, \ldots \in \mathcal{T}$ により行われる．コンテストは，以下の 2 つの段階からなり，順に実施される．

1. **加工段階**: 各チーム T_i は自身の**元データ** O_i を加工して**加工データ** A_i を作り，審判 J に提出する．加工の際には，加工後のデータが元のデータに近く (有用性が高く)，かつ，元データの一部である**秘密データ** R_i が他のチームから推測しにくく (安全性が高く) なることを目指す．
2. **攻撃段階**: 各チーム T_j は全チームが加工した加工後のデータを J から受け取り，他の各チームの秘密データ R_i の推定を行う．推定結果を J に提出する．

さらに，J は各チームから受け取った加工後のデータと推定結果から，各加工後のデータの評価と各推定結果の評価を行う．

1.2　データセット

本コンテストでは，加工対象のデータとして，ランダムに生成された表形式のデータを用いる．すべての値は c 種類のカテゴリ値であるものとし，c 種類のカテゴリ値を便宜的に 0 以上 c 未満の整数で代用する．

各チーム T_i は n 行 m 列 $(n, m_P, m_R$ を自然数とし，$m := m_P + m_R)$ の行列である元データ O_i を加工する．O_i の左側 m_P 列からなる部分行列を**公開データ** P，右側 m_R 列からなる部分行列を**秘密データ** R_i と呼び，$m = m_P + m_R$ である．P は J により用意され，攻撃段階の開始時に全チームに公開される．R_i は T_i が生成し，他のチームには秘密にして保持する．R_i は**生成器** G に T_i が任意に選択した非負整数である**シード** S_i を入力して生成される．G は J により攻撃段階の開始時に全チームに公開される．G は入力として受け取ったシードを元にして，擬似乱数生成器により各要素が 0 以上 c 未満の整数である n 行 m_R 列の行列を一様ランダムに生成する，決定的な関数である．

1.3　加工段階

加工段階は以下のように実施される．

1. (事前準備) 審判 J は公開データ P，生成器 G を作成し，公開する
2. 各加工チーム T_i はシード S_i を任意に選択し，秘密データ $R_i := G(S_i)$ を計算する．
3. 各 T_i は P と R_i を横方向に連結し，元データ O_i を作る．
4. 各 T_i は O_i を加工し，加工データ A_i を作る．
5. 各 T_i は A_i と S_i，R_i を J に提出する．[1]
6. (事後処理) J は各 T_i に対して $R_i = G(S_i)$ となることを確認する．

1.3.1　加工方法

O_i を加工して A_i を作る際に許可される加工は，以下の 2 種類である．

1. 任意の要素 (P の要素でも R_i の要素でもよい) の値の別の値 (0 以上 c 未満の整数値) への書き換え
2. 任意の要素の値の削除 (削除済みを表す特別な値 “$*$” に書き換え)

1.4　攻撃段階

攻撃段階は，以下のように実施される．

1. (事前準備) 審判 J は各 A_i を整列した**整列加工データ** B_i を作り，全チームの整列加工データ $\{B_i\}_i$ を公開する．
2. 各 T_j は他の各 T_i に対し，B_i と公開データ P から秘密データ R_i を推定した**推定秘密データ** $F_{i,j}$ を作る．
3. 各 T_j は他の全チームに対する推定秘密データ $\{F_{i,j}\}_{i\neq j}$ を J に提出する．

1.5　評価

提出されたデータから，J により各評価値が計算される．

1.5.1　有用性評価値

加工データ A_i の有用性評価値 u_i は，O_i から A_i を作る際に値を変更しなかった要素の割合とし，

$$u_i := 1 - \mathsf{HammingDist}(O_i, A_i)/|O_i|$$

とする．ここで，$\mathsf{HammingDist}(\cdot,\cdot)$ は 2 つの大きさが同じ行列を受け取り，対応する位置にある異なった要素の数を計算する関数である．$|\cdot|$ は行列の要素数である．A_i が不正 (未提出，形式違反など) の場合は $u_i := 0$ とする．

1.5.2　攻撃評価値

推定秘密データ $F_{i,j}$ の攻撃評価値 $a_{i,j}$ は，R_i に対する推定 $F_{i,j}$ の正答率とし，

$$a_{i,j} := 1 - \mathsf{HammingDist}(R_i, F_{i,j})/|R_i|$$

とする．A_i が不正の場合は $a_{i,j} := 1$ とする．A_i が不正でなく，かつ，$F_{i,j}$ が不正の場合は $a_{i,j} := 0$ とする．

[1] R_i は J が G と S_i から生成できるため本来は提出不要であるが，次ステップでの意図せぬ誤り (プログラムのバグや環境の違い，提出データの出し間違いなど) の検出のために提出する．

1.5.3 安全性評価値

加工データ A_i の安全性評価値 s_i は，他の全チームからの推定のうち最悪の (最も正答率が高い) 推定の誤答率とし，

$$s_i := 1 - \max_j a_{i,j}$$

とする．A_i が不正 (未提出，形式違反など) の場合は $s_i := 0$ とする．

1.5.4 加工段階評価値

チーム T_i の加工段階評価値 x_i は，A_i の有用性評価値 u_i と安全性評価値 s_i の平均値とし，

$$x_i := (u_i + s_i)/2$$

とする．

1.5.5 攻撃段階評価値

チーム T_j の攻撃段階評価値 y_j は，他の全チームに対する推定の正答率の平均値とし，

$$y_j := \frac{1}{|\mathcal{T}| - 1} \sum_{i \neq j} a_{i,j}$$

とする．$|\cdot|$ は集合の要素数である．

1.6 順位

J は計算された評価値に基づき，予備戦，本戦，全体のそれぞれについて加工段階と攻撃段階の順位を，さらに，全体についての総合順位を決定する．具体的な順位の計算方法は，以下の通り．

1.6.1 加工段階の順位 (予備戦，本戦)

各チーム T_i の加工段階評価値 x_i で決定する (大きいほど上位)．同順位の場合は，同順位のチーム間で自身の加工データ A_i の安全性評価値 s_i により決定する (大きいほど上位)．

1.6.2 攻撃段階の順位 (予備戦，本戦)

各チーム T_j の攻撃段階評価値 y_j で決定する (大きいほど上位)．同順位の場合は，同順位のチーム間で自身の全攻撃評価値 (T_j の場合は $\{a_{i,j}\}_{i \neq j}$) のうち 1 番目に小さい値により決定する (大きいほど上位)．さらに同順位の場合は，同順位のチーム間で自身の全攻撃評価値 (T_j の場合は $\{a_{i,j}\}_{i \neq j}$) のうち 2 番目に小さい値により決定する (大きいほど上位)．以下同様にして同順位を解消する．

1.6.3 加工段階と攻撃段階の順位 (全体)

加工段階と攻撃段階のそれぞれについて，以下のように順位を決定する．本戦と予備戦の順位の 9:1 の重み付き和で決定する (小さいほど上位)．同順位の場合は，同順位のチーム間で本戦の順位により決定する (小さいほど上位)．

1.6.4 全体の総合順位

全体での加工段階の順位と全体での攻撃段階の順位の総和で決定する (小さいほど上位)．同順位の場合は，同順位のチーム間で全体での加工段階の順位により決定する (小さいほど上位)．

1.7　パラメータ設定

1.7.1　予備戦

予備戦は以下のパラメータ設定で実施する．

- $n = 10^5$
- $c = 10$
- $m_P = 6$
- $m_R = 10$

予備戦で使用する公開データ P は，n 行 m_P 列の行列を，要素ごとに独立に $\min\{\mathsf{Rand}(c), \mathsf{Rand}(c)\}$ を計算することにより作成した．ここで，$\mathsf{Rand}(x)$ は 0 以上 x 未満の整数を一様ランダムに返す関数である．

1.7.2　本戦

本戦で使用する各パラメータは後日決定し，公開する．

1.8 参考: 本コンテストで使う主な記号

- J: 審判
- $\mathcal{T}$: チームの集合
- $T_i \in \mathcal{T}$: チーム
- n: 行列行数．自然数
- m_P, m_R: 行列列数．自然数
- c: 値のカテゴリ数．自然数
- P: 公開データ．n 行 m_P 列の行列
- R_i: 加工チーム T_i の秘密データ．n 行 m_R 列の行列
- O_i: 加工チーム T_i の元データ．n 行 m 列の行列 ($m := m_P + m_R$)
- A_i: 加工チーム T_i の加工データ．n 行 m 列の行列
- B_i: 加工チーム T_i の整列加工データ．n 行 m 列の行列
- $F_{i,j}$: 加工データ A_i に対する攻撃チーム T_j の推定秘密データ．n 行 m_R 列の行列
- S_i: 加工チーム T_i のシード．非負整数
- G: 生成器．$G(S_i) = R_i$ となる決定的な関数
- u_i: 加工データ A_i の有用性評価値．0 以上 1 以下の実数値．大きいほどよい
- s_i: 加工データ A_i の安全性評価値．0 以上 1 以下の実数値．大きいほどよい
- $a_{i,j}$: 推定秘密データ $F_{i,j}$ の攻撃評価値．0 以上 1 以下の実数値．0 以上 1 以下の実数値．大きいほどよい
- x_i: 加工チーム T_i の加工評価値．0 以上 1 以下の実数値．大きいほどよい
- y_i: 攻撃チーム T_j の攻撃評価値．0 以上 1 以下の実数値．大きいほどよい

第2章　システムの使い方

本コンテストは，CodaLab 上に用意されたシステム上で実施する．本章では，本コンテストのシステムの使い方を説明する．

2.1　参加登録

コンテストのシステムを利用するために，最初に以下の手順により参加登録を実施してください．

1. CodaLab (`https://codalab.lisn.upsaclay.fr/`) でアカウントを登録する．コンテスト中は，このアカウントを使ってデータの提出を行う
2. PWS Cup 2023 のページ (`https://codalab.lisn.upsaclay.fr/competitions/15334`) へアクセスする
3. “Participate” タブを選択する
4. 参加規定に同意し，“I accept the terms and conditions of the competition” をチェックし，“Register” ボタンを押す
5. `https://forms.gle/JrLg6R7B7WZZt6K77` より，ステップ 1 で登録した CodaLab のアカウントの username を審判に報告する
6. 審判が CodaLab の username を確認すると，CodaLab から参加許可の通知が届き，システムを使えるようになる

2.2　チーム名の登録

参加登録の終了後，以下の手順でチーム名の登録を行ってください．

1. システムの “Learn the Details” タブから “参加チーム一覧” を選択する
2. 表から自身のチームのチーム ID (左端の列の数字 2 文字) を確認する
3. 画面右上のアカウント名にマウスポインタを重ね，“Settings” をクリックする
4. “Team name” の欄に，“(チーム ID):(チーム名)” の書式でチーム名を記入する．例えば，チーム ID が “99” でチーム名が “ほげほげ” の場合は，“99: ほげほげ” と記入する
5. “Save Changes” をクリックして記入したチーム名を保存する

ここで登録したチーム名は，審判や他の参加チームがリーダーボードを確認する際に使われます．チーム名の登録を終えた後は，コンテスト期間中はチーム名を変更しないでください．

2.3　加工段階の操作方法

加工段階では，各チームは加工データと関連するデータを作成し，これらを含む zip ファイルを提出する．zip ファイルを提出すると，競技システムは提出されたデータの書式の確認と，有用性の評価を行う．

2.3.1　zip ファイルの作成

1. 3.1.1 節に従い，必要なファイルを入手する
2. 1.3 節を参考に，加工を実施する
3. 3.1.2 節に従い，提出用の zip ファイルを作成する

2.3.2　zip ファイルの提出

1. “Participate” タブをクリックし，“’Submit / View Results’ を選択する
2. 予備戦では “予備戦 - 加工段階” ボタンを，本戦では “本戦 - 加工段階” ボタンをクリックする．
3. 必要に応じメモを入力し，“Submit” ボタンをクリックする
4. 提出する zip ファイルを選択する
5. 選択した zip ファイルがシステムにアップロードされる．しばらく待つと，提出した zip ファイルの STATUS 列の値が “Submitting” として表示される
6. 1 分ほど待ち，“Refresh status” ボタンをクリック．うまくいった場合は，STATUS 列の値が “Finished” に変わる．もし STATUS 列の値が “Failed” となった場合は，提出が何らかの要因によりうまくいっていない．提出データに問題がないか確認いただきたい．画面内のメッセージや “View scoring error log” から見られるエラーログが参考になるかもしれない．

   投稿が失敗する理由がわからない場合は，7.2 節の問い合わせ先より相談いただきたい
7. ブラウザの更新ボタンからページを更新し，再度手順 2 までを行うと，投稿データの SCORE 列に有用性評価値が表示される

2.3.3　リーダーボードへの結果の登録

加工段階では，任意の数の加工データを提出することができる．各参加チームは，加工段階の期間内に提出済みのデータの中から一つを選択し，リーダーボードに登録しなくてはならない．リーダーボードに登録した加工データの有用性評価値は公開される．リーダーボードに登録するデータは加工段階の期間内に何度でも変更することができる．**加工段階の終了時に登録されていた加工データが加工段階での最終的な提出データとして扱われる．**リーダーボードへの登録の手順は以下の通りである．

1. 投稿ページを開き，リーダーボードへ登録したい提出データの “+” ボタンをクリックする
2. “Submit to Leaderboard” ボタンが出現する
3. 提出データのチェックマークの列にチェックマークが表示される．これは，この提出データがリーダーボードに登録されていることを意味する

2.3.4　リーダーボードの確認

リーダーボードは以下の手順により確認する．

1. “Results” タブを選択.
2. 予備戦では “予備戦 - 加工段階” ボタンを，本戦では “本戦 - 加工段階” ボタンをクリックする
3. “加工段階の結果” の表が加工段階のリーダーボードである．“攻撃段階の結果” の表は無関係であるがシステムの制約により表示されてしまうものであるので，無視する

加工段階のリーダーボードの見方 リーダーボードの各行が各チームがリーダーボードに登録したデータの結果を表す．各列の意味は以下の通りである．

- Entries: 投稿済みの zip ファイルの数
- Date of Last Entry: 最終投稿日
- Team Name: チーム ID とチーム名
- U0: 有用性評価値．括弧内は順位

その他の列は今後の拡張のために確保しているものであり，現時点では未使用である．

2.4 攻撃段階の操作方法

攻撃段階では，各チームは推定秘密データと関連するデータを作成し，これらを含む zip ファイルを提出する．zip ファイルを提出すると，競技システムは提出されたデータの書式と内容の確認を行う．(攻撃評価値の計算は，攻撃段階終了後に審判が計算する．)

2.4.1 zip ファイルの作成

1. 3.2.1 節に従い，必要なファイルを入手する
2. 1.4 節を参考に，攻撃を実施する
3. 3.2.2 節に従い，提出用の zip ファイルを作成する

2.4.2 zip ファイルの提出

1. “Participate” タブをクリックし，“’Submit / View Results’ を選択する
2. 予備戦では “予備戦 - 攻撃段階” ボタンを，本戦では “本戦 - 攻撃段階” ボタンをクリックする
3. 以後の操作は 2.3.2 節と同様である

2.4.3 リーダーボードへの結果の登録

リーダーボードへの結果の登録方法は加工段階とほとんど同じであるので，2.3.3 節 を攻撃段階に読み替えていただきたい．**攻撃段階でも攻撃段階の期間内に提出済みのデータの中から一つを選択し，リーダーボードに登録しなくてはならない**ので，ご注意いただきたい．

2.4.4 リーダーボードの確認

リーダーボードは以下の手順により確認する．

1. “Results” タブを選択．
2. 予備戦では “予備戦 - 攻撃段階” ボタンを，本戦では “本戦 - 攻撃段階” ボタンをクリックする
3. “攻撃段階の結果” の表が攻撃段階のリーダーボードである．“加工段階の結果” の表は無関係であるがシステムの制約により表示されてしまうものであるので，無視する

攻撃段階のリーダーボードの見方　リーダーボードの各行が各チームがリーダーボードに登録したデータの結果を表す．各列の意味は以下の通りである．

- Entries: 投稿済みの zip ファイルの数
- Date of Last Entry: 最終投稿日
- Team Name: チーム ID とチーム名
- 検査結果: チーム ID ファイルの書式確認結果．1.0 は書式確認が成功したことを表す
  - **重要*: 最終的な提出データでは，この値が *1.0* となっていることを必ず確認してください．*1.0* となっていない場合は，書式違反となり，攻撃段階は未提出扱いとなります**
- 提出した値: 提出したチーム ID ファイルに書かれたチーム ID を 10 進数に変換した値．例えば，チーム ID が 05 の場合は 5.0，チーム ID が 10 の場合は 10.0 となる
  - **重要*: 最終的な提出データでは，この値が自身のチーム *ID* を *10* 進数に変換した値となっていることを必ず確認してください．そうなっていない場合は，書式違反となり，攻撃段階は未提出扱いとなります**
- FXX (XX は数字 2 文字): チーム ID が XX であるチームに対する推定秘密データの書式確認結果．自身のチーム ID が YY である場合，`b_XX_YY` の書式確認結果が FXX 列に表示される．zip ファイルに対応する推定秘密データが正しいファイル名で含まれていて，かつ，書式確認が成功したときに 1.0 となる
  - **重要*: 1.0* となっていない場合は，*(*書式違反やファイル名間違いなどの要因により，*)* 対応する推定秘密データは受理されていません．提出したつもりのすべての推定秘密データについて，対応する列の値が *1.0* となっていることを必ず確認してください．*1.0* となっていない列に対応する推定秘密データは未提出扱いとなります**

その他の列は今後の拡張のために確保しているものであり，現時点では未使用である．

第3章　入出力

PWS Cup 2023 の各段階で参加チームが受け取るデータ，システムに提出するデータを説明する．

3.1　加工段階

3.1.1　受け取るもの

各チームは，システム (“Participate” タブ内 “Files” の Public Data (対応する段階のもの)) から以下のファイルを含む zip ファイルを入手する．

- `p.csv`: 公開データ P．4.2.1 節の書式に従っている

各チームは，5.1.1 節 に記載の場所から以下を入手する．

- `g.rb`: 生成器 G．5.2.1 節の手順に従って使用する
- `common.rb`: 生成器 G の補助ファイル．`g.rb` と同じディレクトリに配置する

3.1.2　提出するもの

各チームは以下の全ファイルを直下に含む zip ファイルを提出する．

- `a.csv`: 加工データ A_i．4.3.1 節の書式に従って作成する
- `r.csv`: 秘密データ R_i．4.3.2 節の書式に従って作成する
- `seed.txt`: シード S_i．4.3.4 節の書式に従って作成する

zip ファイルのファイル名は，任意であるが，使える文字は半角英数字および一部の半角記号 (`._-`) である．各ファイルは第 4 章の書式に従って作成すること．

注意　zip ファイルを作成する際は，すべてのファイルが直下に含まれるように作成する必要がある．ファイルがフォルダに入った状態の zip が提出された場合は，システムは該当するファイルが含まれていないと判断し，提出ファイルが受理されない．

例　システムの Starting Kit (“Participate” タブ内 “Files” の Starting Kit (対応する段階のもの)) は正しく作成された zip ファイルの例となっている．

3.2　攻撃段階

3.2.1　受け取るもの

各チームは，システム (“Participate” タブ内 “Files” の Public Data (対応する段階のもの)) から以下のファイルを含む zip ファイルを入手する．

- `b_II.csv`: チーム ID が “`II`” であるチームの整列加工データ (加工データが受理された全チーム分)．4.2.2 節の書式に従っている

3.2.2　提出するもの

各チームは以下の全ファイルを直下に含む zip ファイルを提出する．

- `my_id.txt`: 自身のチームのチーム ID．4.3.5 節の書式に従って作成する
- `f_II_JJ.csv`: 推定秘密データ $F_{i,j}$．ファイル名の `II` は攻撃対象のチーム T_i のチーム ID に，`JJ` は自身のチーム T_j のチーム ID に，それぞれ置換する．複数のチームに攻撃を行う場合は，攻撃を行う全てのチームに対するデータを格納する．4.3.3 節の書式に従って作成する

zip ファイルのファイル名は，任意であるが，使える文字は半角英数字および一部の半角記号 (`._-`) である．各ファイルは第 4 章の書式に従って作成すること．

注意　zip ファイルを作成する際は，すべてのファイルが直下に含まれるように作成する必要がある．ファイルがフォルダに入った状態の zip が提出された場合は，システムは該当するファイルが含まれていないと判断し，提出ファイルが受理されない．

例　システムの Starting Kit (“Participate” タブ内 “Files” の Starting Kit (対応する段階のもの)) は正しく作成された zip ファイルの例となっている．

3.3　注意事項

- ファイル名は大文字と小文字が区別される
- zip ファイルを作成する際は，フォルダを含まず，すべてのファイルが直下に含まれるように作成しなくてはならない
- 各ファイルを作成する際は第 4 章の書式に従うこと

第4章　ファイルの書式

本章では，本コンテストで扱う zip を除いた各ファイルの書式を定める．本コンテスト中に提出する各ファイルは，本章で説明する書式に従う必要がある．従っていないファイルは不正とみなされ，受理されない．

4.1　全体的な書式

全体的な書式は以下の通り．

- ファイル形式: テキストファイル (ASCII)
- 改行文字: LF (\n) または CRLF (\r\n)
- 行数: ファイルごとに指定

最終行の改行はあってもなくても構わない．全角文字は使用できない．

4.1.1　k 行 ℓ 列の csv ファイル

「k 行 ℓ 列の csv ファイル」と指定されたファイルは，全体的な書式に加え，以下の書式にも従う必要がある．
各行は ℓ 列で構成される．すなわち，各行にちょうど $\ell - 1$ 個のカンマが含まれる．空行があってはならない．

セルの書式　各行で行末の改行文字を除去してカンマで区切られた部分文字列 (各行にちょうど ℓ 個含まれます) をセルと呼ぶ．各セルの書式は以下の通り．

- “*” を含まない: 0 以上 c 未満の整数
- “*” を含んでよい: 0 以上 c 未満の整数または “*”

いずれも，値の前後などに不要な空白があるだけでも不正になることに注意いただきたい．また，整数には余分な “0” が先行してはならない．以下は正しい整数の書式の例である．

- “0”
- “3”

以下は不正な整数の書式の例である．

- “00”
- “03”
- “0003”
- “ 3”
- “ 3 ”
- “0 3”

4.2　受け取る各ファイルの書式

4.2.1　公開データ P

- ファイル名: `p.csv`
- 書式: n 行 m_P 列の csv ファイル．セルの書式は「“*” を含まない」 (4.1.1 節を参照)

4.2.2　整列加工データ B_i

- ファイル名: `b_II.csv` (チーム ID が `II` の場合)
- 書式: n 行 m 列の csv ファイル．セルの書式は「“*” を含んでよい」 (4.1.1 節を参照)

4.3　提出する各ファイルの書式

4.3.1　加工データ A_i

- ファイル名: `a.csv`
- 書式: n 行 m 列の csv ファイル．セルの書式は「“*” を含んでよい」 (4.1.1 節を参照)

4.3.2　秘密データ R_i

- ファイル名: `r.csv`
- 書式: n 行 m_R 列の csv ファイル．セルの書式は「“*” を含まない」 (4.1.1 節を参照)

4.3.3　推定秘密データ $F_{i,j}$

- ファイル名: `f_II_JJ.csv` (`II` には攻撃対象 T_i のチーム ID を，`JJ` には自身 T_j のチーム ID を入れる)
- 書式: n 行，m_R 列の csv ファイル．セルの書式は「“*” を含まない」 (4.1.1 節を参照)

4.3.4　シード S_i

- ファイル名: `seed.txt`
- 書式: 1 行目が 0 以上 10^{100} 未満 (100 桁以下) の整数のみからなる，1 行のテキストファイル

整数の前後に不要な空白があるだけでも不正となる．また，不要な先行の “0” を含む場合も不正になる．

4.3.5　チーム ID

- ファイル名: `my_id.txt`
- 書式: 1 行目が自身のチーム ID(数字 2 文字) のみからなる，1 行のテキストファイル

チーム ID の前後に不要な空白があるだけでも不正となる．また，チーム ID の 1 文字目の “0” を省略した場合も不正になる．例えば，チーム ID が “03” の場合，“3” のように 1 文字目の “0” を省略してはならない．

第5章　スクリプトの使い方

本コンテストで提供する各種スクリプトの使い方，コンテスト中に参加者が作成する必要のあるファイルの作成方法の例を示す．

5.1　準備

5.1.1　スクリプトの入手先

本コンテストで参加者に提供する各スクリプトは`https://github.com/hamadakoki/pwscup2023/tree/master/scripts` から入手可能である．

5.1.2　実行環境

本コンテストで配布するスクリプトが想定する実行環境は，dockerhub の`hamadakoki/codalab:legacy-py37-ruby-01`である．スクリプトは類似の環境でも動作することを目指して作成されているが，保証はない．本コンテストで配布するスクリプトは，主に以下の各コマンドが実行可能であることを前提としている．

- ruby
- bash
- sort
- paste
- zip
- echo

5.1.3　前提

すべての`*.rb` ファイルと`*.sh` ファイル，`*.csv` ファイルを同一のディレクトリに置いておく．以下の説明での各コマンドは，このディレクトリで実行する．このディレクトリで `hoge moge` というコマンドを実行することを，以後は以下のように記述するものとする．

```
$ hoge moge
```

5.2　配布されたスクリプトの使い方

5.2.1　秘密データ R_i (r.csv) の作成

入力: シード S_i (`seed.txt`)

出力: 秘密データ R_i (`r.csv`)

```
$ ruby g.rb seed.txt r.csv
```

5.2.2　元データ O_i (o.csv) の作成

入力: 公開データ P (p.csv)，秘密データ R_i (r.csv)

出力: 元データ O_i (o.csv)

```
$ ./pr2o.sh p.csv r.csv o.csv
```

5.2.3　加工データ A_i (a.csv) の書式確認

入力: 加工データ A_i (a.csv)

出力: 書式に問題がなければ何も出力せずに正常終了．問題があった場合は，標準エラーにエラーメッセージを出力して異常終了

```
$ ruby check_a.rb a.csv
```

5.2.4　有用性評価値の計算

入力: 公開データ P (p.csv)，秘密データ R_i (r.csv)，加工データ A_i (a.csv)

出力: 標準出力に A_i の有用性評価値が出力

```
$ ruby score_utility.rb p.csv r.csv a.csv
```

5.2.5　推定秘密データ $F_{i,j}$ (f.csv) の書式確認

入力: 推定秘密データ $F_{i,j}$ (f.csv)

出力: 書式に問題がなければ何も出力せずに正常終了．問題があった場合は，標準エラーにエラーメッセージを出力して異常終了

```
$ ruby check_f.rb f.csv
```

5.2.6　攻撃評価値の計算

入力: 秘密データ R_i (r.csv)，推定秘密データ $F_{i,j}$ (f.csv)

出力: 標準出力に $F_{i,j}$ の攻撃評価値が出力

```
$ ruby score_attack.rb r.csv f.csv
```

5.3　各ファイルの作成例

5.3.1　シード S_i (seed.txt) の作成例

例 1: ランダムなシードを生成する場合

```
$ ./gen_seed.sh seed.txt
```

例 2: シードを “123456789” に指定する場合

```
$ echo '123456789' > seed.txt
```

5.3.2 加工段階の提出ファイルの作成例

例: 提出ファイルのファイル名を ano-submission.zip とするとき

```
$ zip ano-submission.zip a.csv seed.txt r.csv
```

5.3.3 チーム ID ファイル (my_id.txt) の作成例

例: 自身のチーム ID が '03' のとき

```
$ echo '03' > my_id.txt
```

5.3.4 攻撃段階の提出ファイルの作成例

例: 自身のチーム ID が '03' で, チーム '01' とチーム '04' に攻撃し, 提出ファイルのファイル名を atk-submission.zip とするとき

```
$ zip atk-submission.zip my_id.txt f_01_03.csv f_04_03.csv
```

第6章　競技ルール

コンテスト参加者は，以下のルールを遵守すること．

- (不具合の報告) 参加者は，明らかにルールやシステムの不具合と思われる事象を発見した場合は，直ちに審判に知らせること．不具合を故意に利用しないこと．
- (シードの解析の禁止) 他のチームのシードを暴こうとしないこと．(本コンテストの生成器 G は暗号学的に安全ではない方法で作成されている．これは，ボランティアである審判の労力の軽減のためである．コンテストの持続的な実施のために参加者には本ルールを遵守いただきたい．)
- (ソフトウェア，ネットワークなど) 使用するソフトウェアや OS には制限を加えない．
- (チーム間の結託の禁止) 他のチームに秘密データやシード，加工データ，推定秘密データを教えてはならない．ただし，これらが推定できない範囲で，プログラムやモジュールを共有すること等は，禁止しない．
- (リーダーボードへの登録) 各段階では，各段階の終了時にリーダーボードに登録されていた zip ファイルのみが評価対象となる．リーダーボードに登録されていない場合は，未提出扱いとなる．
- (システムへの攻撃の禁止) システムに負荷をかける目的でデータの提出を行ってはならない．
- (表彰) 発表を行ったチームのうち，いくつかのチームを表彰する．

第7章　その他

7.1　注意事項

- 加工段階，攻撃段階の終了前に，忘れずに提出データをリーダーボードに登録すること．登録をしていなかった場合，提出データがなかったものとして扱われる．
- 加工段階が開始したら，即座に自身のシードを定め，秘密データを作成し，システムに投稿してみること．環境によっては，システムと生成器の動作が異なる可能性があり，その場合は加工段階の提出データがシステムに受理されなくなってしまうため，5.1.2 節に記載した環境を作成することを試みていただきたい．どうしても環境構築が難しい場合は，早めに審判に相談すること．
- 本コンテストで利用しているシステムは無償で提供されているコンテストホスティングプラットフォーム上で構築されている．レスポンスが悪いことや，システムが動かないこともあるため，投稿データは早めに提出しておくこと．
- 本コンテストでの提出ファイルの作成ルールは厳密である．第 3 章 や第 4 章 を注意深く読んで作成すること．作成ルールから逸脱している場合は受理されないため，早めに作成から提出までを練習しておくことを推奨する．

7.2　お問い合わせ先

本コンテストに関する不明点などは，以下のメールアドレスより審判にお問い合わせください．

- 審判 (PWS組織委員会 PWS Cup 2023 ワーキンググループ) のメールアドレス: pwscup2023-info(at)iwsec.org (“(at)” をアットマーク “@” に置換ください)