

# 数字・英字・かな文字を用いた 暗証番号における安全性評価

グループ演習第1班

石田若菜 岡宏樹 品川和雅 野貴泰

アドバイザー教員 金山直樹

# 背景

多くの認証システムで、パスワードが用いられている。  
一方、パスワードは必ずしも安全に運用されていない。



本発表では、広く普及している**4桁の暗証番号**の  
安全性について調べる。

# 暗証番号の安全性とは？



本発表では、暗証番号の安全性を「推測の難しさ」としてモデル化する

# 目的

4桁の暗証番号で、使用できる文字の総数を増やしたときの安全性について、評価を行う

具体的には、

- 数字(10種類)
- 英字(26種類)
- 英字＋数字(36種類)
- かな文字(46種類)

を用いた暗証番号における安全性の評価を行う

# 評価方法(ランダム性)

- $N$ 個のパスワード  $x_1, \dots, x_N$  の出現確率がそれぞれ  $p_1 \geq \dots \geq p_N$  であるとする。
- 情報量比を次のように定義する

$$H = \frac{\sum_{i=1}^N p_i \times \log_2 p_i}{\log_2 N}$$

直観:  $H$ が1に近いほど、ランダム性が高い

# 評価方法(安全性)

- $N$ 個のパスワード  $x_1, \dots, x_N$  の出現確率がそれぞれ  $p_1 \geq \dots \geq p_N$  であるとする。

本発表では、ランダム性の評価を情報量比で、  
安全性の評価をGuess Workで行う。

$$G = \sum_{i=1}^{\infty} p_i \times i$$

直観:  $p_1, \dots, p_N$  を知っている攻撃者が、  
推測に成功するのに必要な試行回数の期待値

# 実験方法

- 調査方法：アンケート用紙による調査
- 対象者：筑波大学学群生
- 実施場所：各教室
- 回答数：218名
- 有効回答数：212名 (97.2%)

# アンケート内容

次のような設問 1 ～ 1 2 を行う。

設問 1 ～ 4 **擬似的な暗証番号**の設問

- (1) 数字 (0～9)
- (2) かな文字 (あ～ん)
- (3) 英字 (a～z)
- (4) 英字と数字 (a～z, 0～9)

## 設問 1

覚えておくことのできる 4 桁の数字 (0～9) で、なるべく他人に推測されないものを書いてください。

(記入例) 4284、5031

--	--	--	--



# 結果・考察

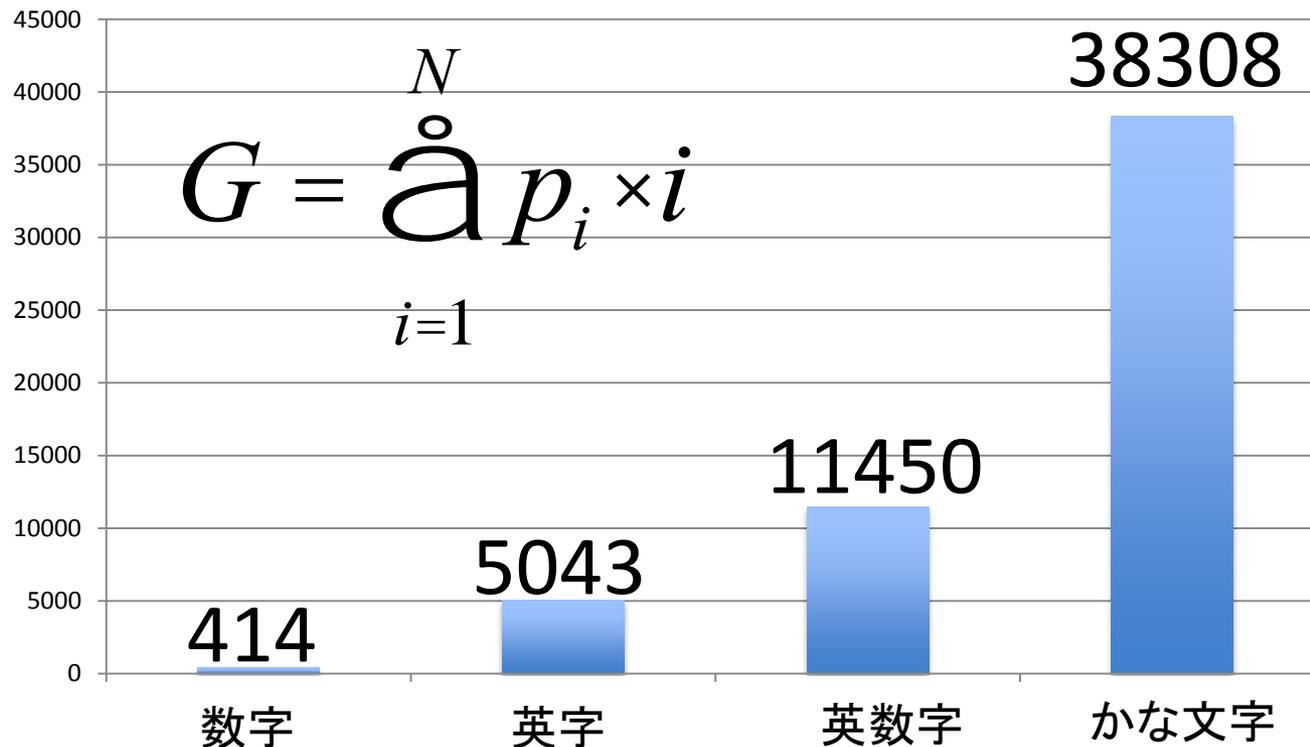
- 安全性の考察
- ランダム性の考察
- 各文字の特徴
  - 数字・英字・英数・かな

# 結果・考察

安全性とランダム性について

# 実験結果：安全性の考察

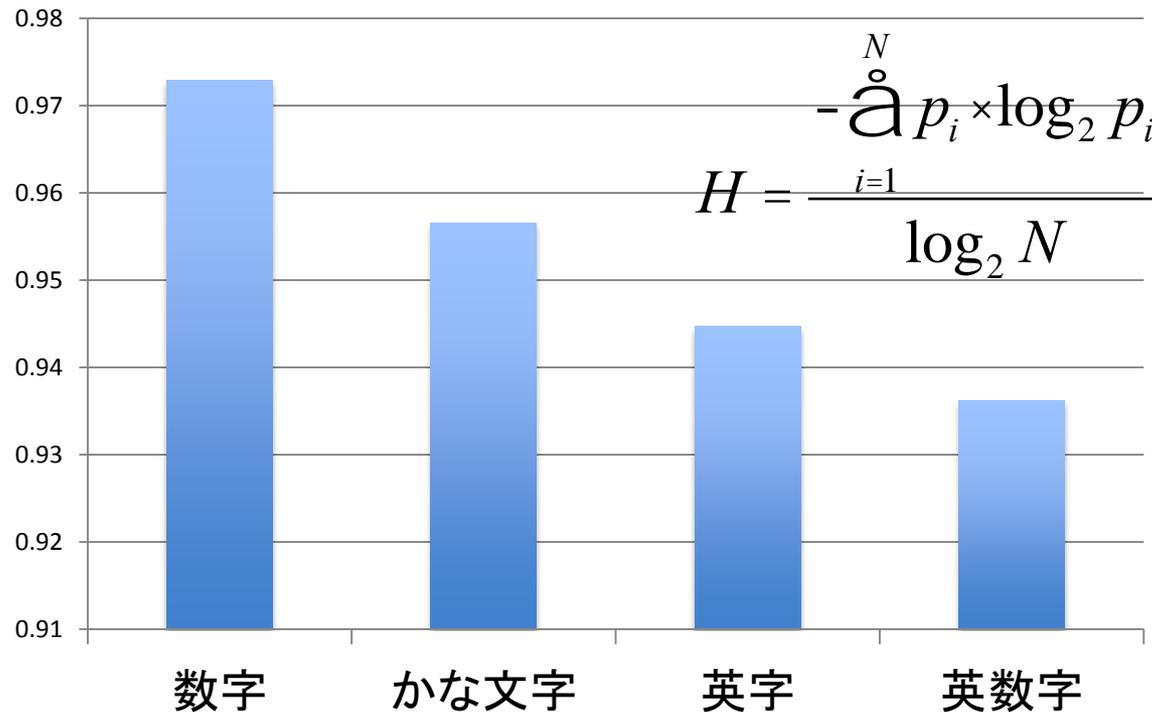
- Guess Work (試行回数の期待値)の比較



(注) 数字とかな文字ではそもそもパスワードの分布が異なるはずなので、このような結果になることは決して自明ではない。

# 実験結果：ランダム性の考察

- 情報量比の比較(高いほどランダム)



- 数字 > かな文字 > 英字 > 英数字

## 結果・考察

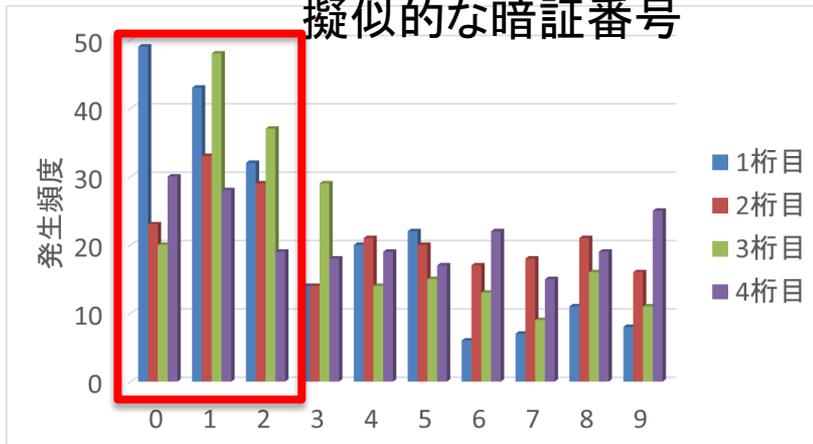
### 数字パスワード(設問1、5)の特徴

# 実験結果：数字の特徴

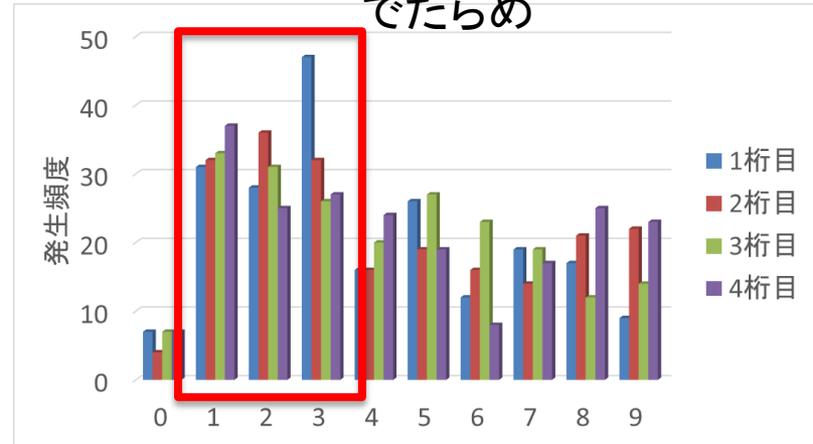
0 1 2 1

(例)0から2が多い

擬似的な暗証番号



でたらめ



1 0 2 3

10月23日

(例)日付(1~31)の範囲が多い

	擬似的な暗証番号	でたらめ
日付	1~31の範囲が多い(58%)	1~31の範囲が少ない(36%)

# 実験考察

- 擬似的な暗証番号、でたらめに関わらず0に近い数字を選ぶ



- パスワード設定で小さい数字を選ぶ傾向がある。

- 擬似的な暗証番号に比べて「でたらめ」は20%以上日付を使わない



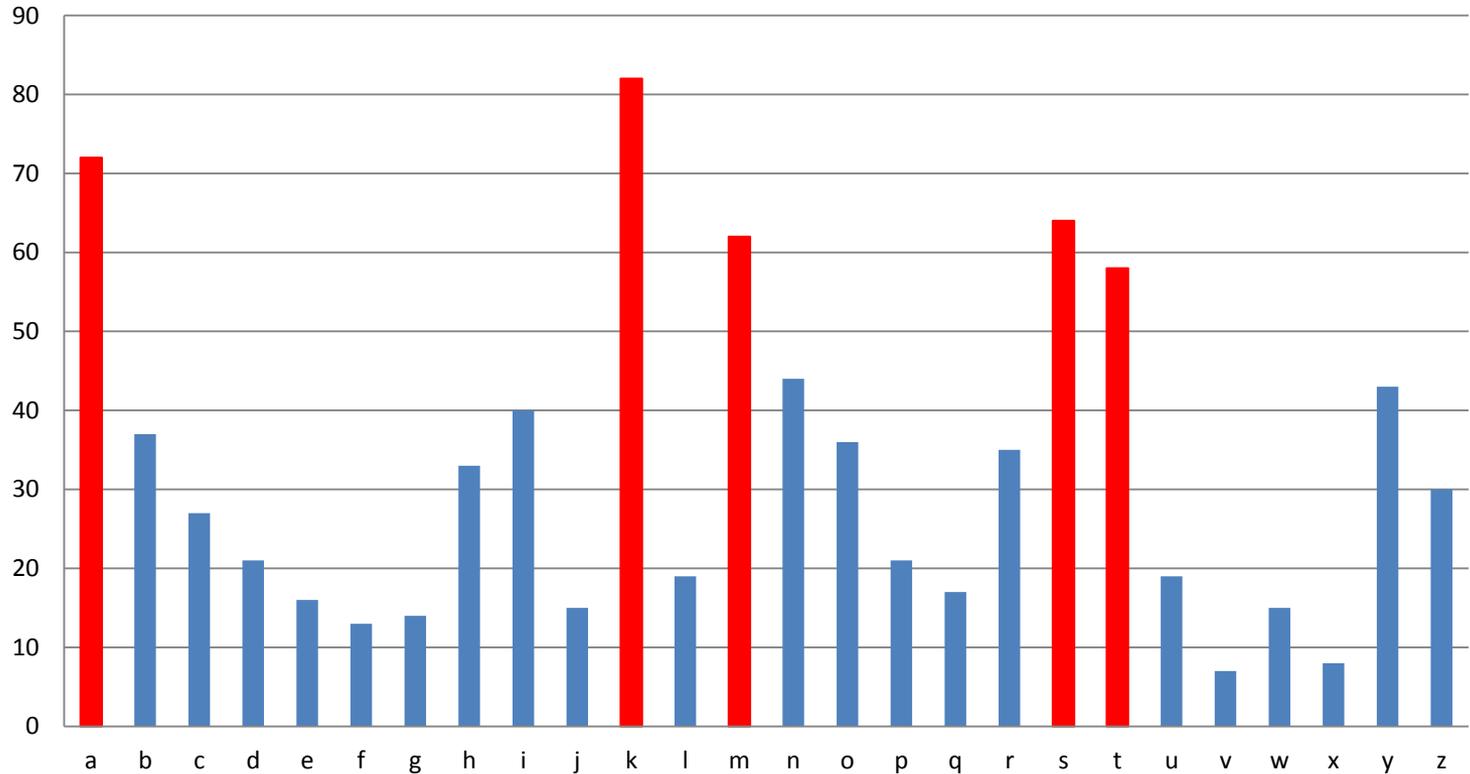
- 「でたらめ」と認知すると誕生日などの日付は使わない傾向がある。

## 結果・考察

### 英字パスワード(設問2、6)の特徴

# 実験結果：英字の特徴

- 「k」「a」「s」「m」「t」の順で出現回数が多い



- 「k」「a」「s」「m」「t」は日本語会話においても出現頻度が高い[2]
- 会話においてよく使われる文字が、パスワードにおいても設定されやすい可能性がある

[2]小寺一興味, 平石光俊: 日本語会話における単音節の出現頻度  
-語音明瞭検査の語表構成の検討-

## 結果・考察

英数字パスワード(設問4、8)の特徴

# 実験結果：英字+数字の特徴

## 文字の並びの解析

英数字の組み合わせは「数数数数」から「英英英英」の全部で16パターン

実験結果は、「英英数数」の組み合わせが圧倒的に多かった(約4分の1)

	出現回数
英英数数	47
英数英数	27
英数数英	23
英英英数	23
英数英英	17

# 実験結果：英字+数字の特徴

さらに出現回数を詳細に調べてみると・・・

1桁目	2桁目	3桁目	4桁目
a (24)	1 (22)	1 (30)	2 (24)
t (23)	k (19)	2 (18)	1 (16)
k (15)	0 (13)	0 (14)	3 (15)
s (14)	2 (13)	k (13)	4 (14)
j (12)	b (13)	4,8,b,s (9)	0,5,7 (11)

出現頻度

本来なら  $36^4 \doteq 170$ 万通り

英英数数の47個中42個が  
日付になる数字(01~31)!

英字

英字

数字

数字

という並びに絞ると、

英字2桁 × 日付  $\doteq 2$ 万通り

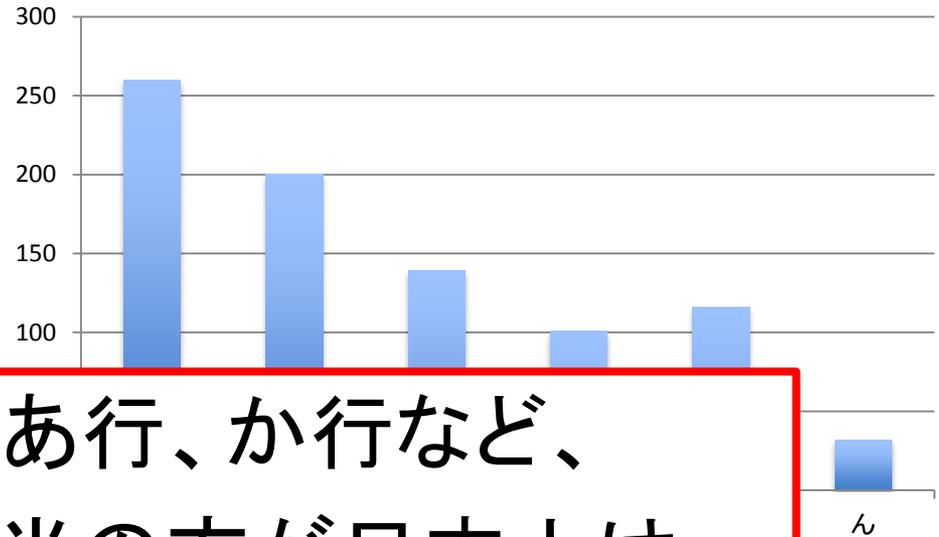
## 結果・考察

かな文字パスワード(設問3、7)の特徴

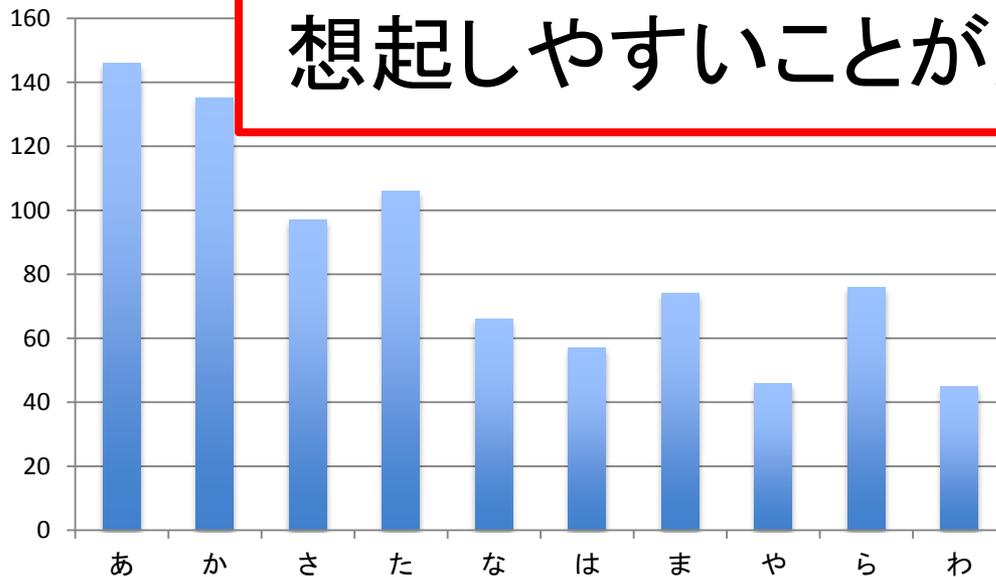
# 実験結果：かな文字の特徴

## 母音の解析

- 「あ」が多い
- 次に「い」「が」が多い



あ段、い段、あ行、か行など、  
五十音表の前半の方が日本人は  
想起しやすいことが原因かもしれない



- 「あ」「か」「た」「さ」  
の順に多い

# 実験の結果

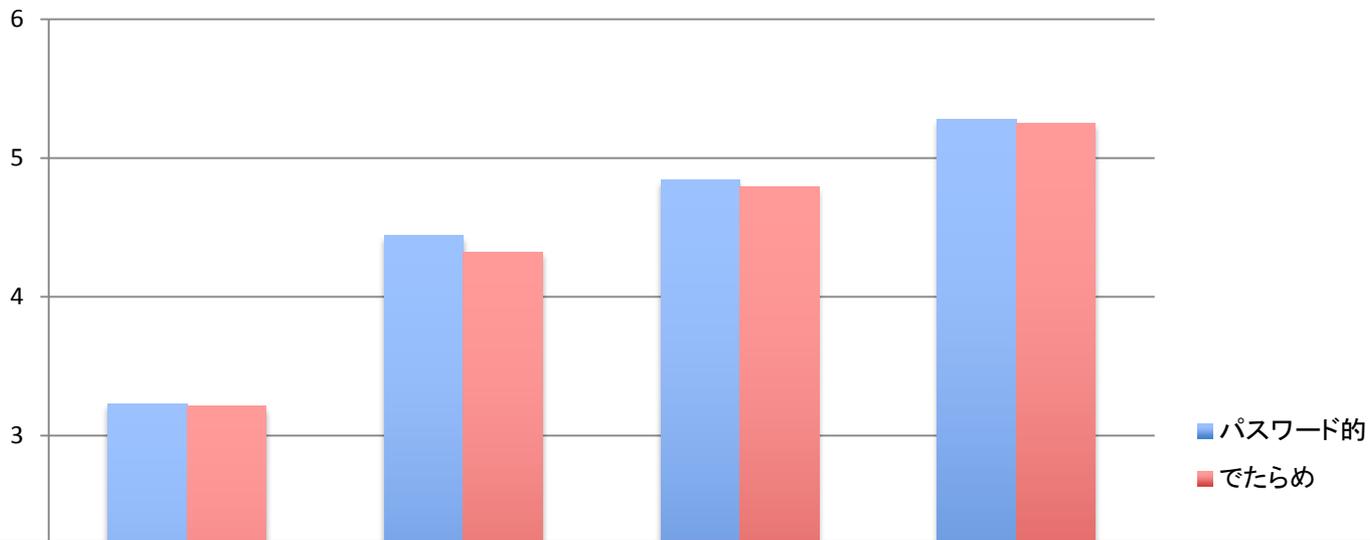
- 安全性 : **かな** > 英数 > 英 > 数
- ランダム性 : 数 > **かな** > 英 > 英数
- 「0」、「a」、「あ」などは出現しやすい

# 結論

- パスワードの出現頻度の特性を明らかにした
- かな文字は英字・英数字よりもパスワードに適している可能性あり
- 数字とかな文字はトレードオフの関係

# 面白い(不思議な)結果

- パスワード的な4桁より、でたらめな4桁の方がランダム性が低いという不思議な結果が出た



人間の「でたらめな列を作る能力」の  
低さを表している？

桁数が少ない場合のみの現象？