# データを隠しながら見せるってどういうこと？

担当者：吉川 正俊

元データ（秘密）

| 番号 | 氏名 | 住所 | 年齢 | 成績 |
|---|---|---|---|---|
| 1 | 江川 | 大阪府 | 19 | 不合格 |
| 2 | 棚橋 | 大阪府 | 20 | 不合格 |
| 3 | 堀江 | 兵庫県 | 20 | 不合格 |
| 4 | 大貫 | 兵庫県 | 19 | 合格 |

集計

統計データ（公開）

| 住所 | 成績 | 人数 |
|---|---|---|
| 大阪府 | 不合格 | 2 |
| 兵庫県 | 不合格 | 1 |
| 兵庫県 | 合格 | 1 |

大阪府の人は不合格だったことがわかってしまう

そこで，**確率的に雑音を加える**

元データ（秘密）に雑音を加える方法

| 番号 | 氏名 | 住所 | 年齢 | 成績 |
|---|---|---|---|---|
| 1 | 江川 | 大阪府 | 19 | 0.8不合格, 0.2合格 |
| 2 | 棚橋 | 大阪府 | 20 | 0.8不合格, 0.2合格 |
| 3 | 堀江 | 兵庫県 | 20 | 0.8不合格, 0.2合格 |
| 4 | 大貫 | 兵庫県 | 19 | 0.8合格, 0.2不合格 |

統計データに雑音を加える方法

| 住所 | 成績 | 人数 |
|---|---|---|
| 大阪府 | 不合格 | 1.7 |
| 兵庫県 | 不合格 | 1.1 |
| 兵庫県 | 合格 | 1.2 |

Google Microsoft Meta LINEヤフー

多くのIT企業が使っている最新のプライバシー保護技術である
**差分プライバシー**(*differential privacy*) の原理を学ぼう