Plan:
- 11/13: Important talk
- 11/18: Go over for test
- 11/20: Test 2
- 11/25: Go over test 2
- 12/2: Presentations
- 12/4: Projects

Privacy issues:

- Privacy for the data that is used
- Problem: We have data that we actually use, but we need to ensure privacy.

- Our statistical database: You are interested in statistics and no in individual records.
  - Usual database: objective get individual records

E.g.,
+ Census database: good source for correlations
+ Medical database:
  (Need to get data without disclosing sensitive info.)

* Many traditional ways to protect privacy:
  + Limit size of the sample below:

\[
\begin{align*}
X_1, \ldots, X_n

\text{Problem:} \quad & \text{Dev. } 1, \ldots, \text{dev. 1003} \\
& \frac{s_1 + \ldots + s_k}{K} = \bar{s}_1 \\
& \frac{s_2 + \ldots + s_{K-1}}{K-1} = \bar{s}_2 \\
& \bar{s}_1 \cdot K = s_1 + \ldots + s_{K-1} + s_K \\
& \bar{s}_2 (K-1) = \frac{s_1 + \ldots + s_{K-1} + s_K}{\bar{s}_1 \cdot K - \bar{s}_2 (K-1)} = \ldots \bar{s}_K
\end{align*}
\]
\[
\frac{10 + 100 + 80 + 50}{4} = 60
\]

\[
\frac{10 + 100 + 60}{3} = \frac{190}{3} \approx 63.333
\]

\[
S_1 + S_2 + S_3 + S_4 = 510, \quad 41 = 240
\]

\[
S_1 + S_2 + S_3 = 510 - 240 = 190
\]

Another trick:

\[
\begin{align*}
S_1 + \Delta S_1 &= S_1 \\
S_1 + \Delta S_2 &= S_2 \\
\vdots \\
S_1 + \Delta S_n &= S_n
\end{align*}
\]

Save as a random H.

Radically different approach:

- Don't keep exact data
- Keep only bounds: [20, 30], [30, 40], ...

Salary: [10, 30], [20, 180], [100, 800]
Privacy is preserved

- How to compute statistical characteristics.

**If we know** $S_i$

$$S = \frac{S_1 + \ldots + S_n}{n}$$

$$V = \frac{1}{n} \sum_{i=1}^{n} (S_i - S_v)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} S_i^2 - \left( \frac{1}{n} \sum_{i=1}^{n} S_i \right)^2$$

**We only know** $S_i \in [\underline{S_i}, \overline{S_i}]$

$$S_{uv} = \frac{S_1 + \ldots + S_n}{n}$$

$$S_{uv} = \frac{S_1 + \ldots + S_n}{n}$$

In this case, the function is monotonic.

$$Y = f(x_1, \ldots, x_n)$$

We know: $X_i \in [\underline{X_i}, \overline{X_i}]$

We want to test $\{Y = f(x_1, \ldots, x_n) : x_i \in [\underline{x}_i, \overline{x}_i] \}$

$$\Delta_i = \overline{x}_i - \underline{x}_i$$

half width

Use Monte Carlo when $f$ is easy.
Monte Carlo Method

☑ Fewer calls to f
- more computations

\[ \sum_{i=1}^{n} |f(x_1, \ldots, x_i, x_{i+1}, \ldots, x_n) - \bar{y}| \]

⚠️ Question
- Here is the characteristic
- Here is the input
- Find the values.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Weight</th>
<th>Points</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>[10, 20]</td>
<td>15</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>[20, 30]</td>
<td>25</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>[30, 70]</td>
<td>65</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>