week 6 day 2

"Exact testing for model/data fit for log-linear models"
"Algebraic & Geometric Methods in Statistics"

Sonja Petrović

Created for Math/Stat 561

Feb 15, 2023.

Agenda

- Chapter 9 from our textbook: Fisher's exact test
- Part of chapter 8, as we may need the cone of sufficient statistics.

Goals

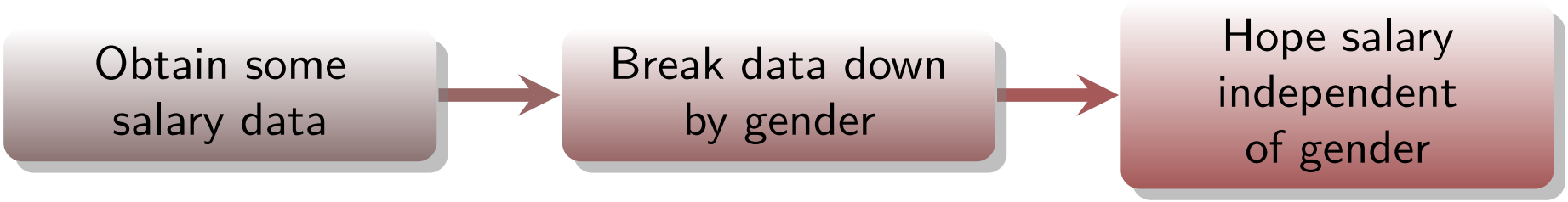
- Understand hypotheses testing for model/data fit
- What is a p-value?
- Asymptotic vs. exact tests
- Fisher's test and example
- General goodness of fit test for log-linear models
- Open problems and relation to projects!

At the heart of statistical reasoning

- Given: data, find out if it is usual/expected? surprising/outlier? quantify??

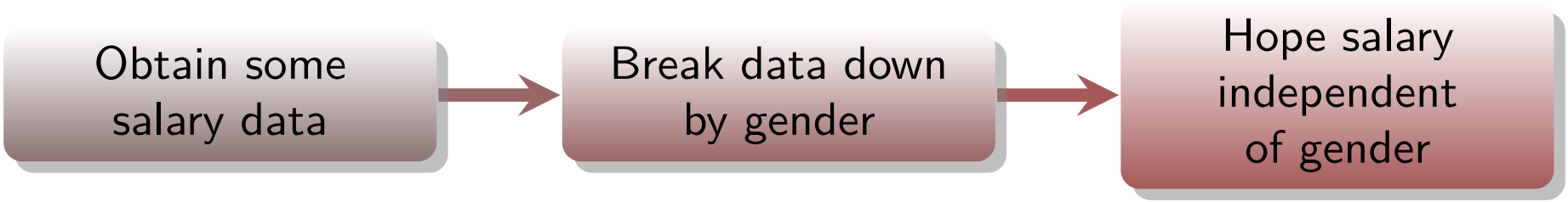
At the heart of statistical reasoning

- Given: data, find out if it is usual/expected? surprising/outlier? quantify??
- Do all genders get fair salary in Chicago for data science jobs?



At the heart of statistical reasoning

- Given: data, find out if it is usual/expected? surprising/outlier? quantify??
- Do all genders get fair salary in Chicago for data science jobs?

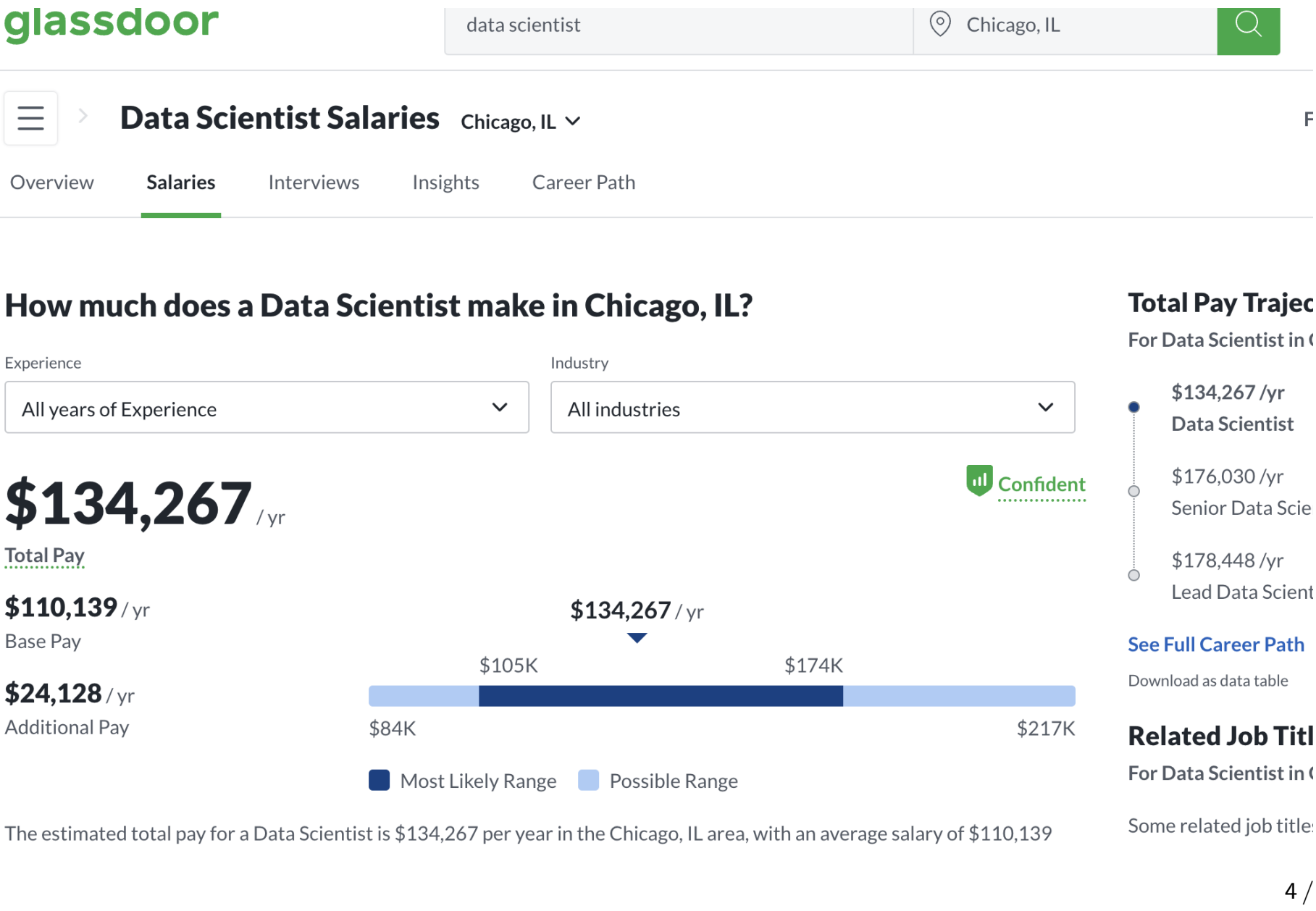


- We expect a certain 'shape' of the data. A certain... distribution!

YOUR everyday intuition $\mapsto$ formal framework.

	M	F	T/Nb	totals
$\leq$ 135K	?	?	?	**13**
$>$ 135K	?	?	?	**13**
totals	**10**	**10**	**6**	**26**

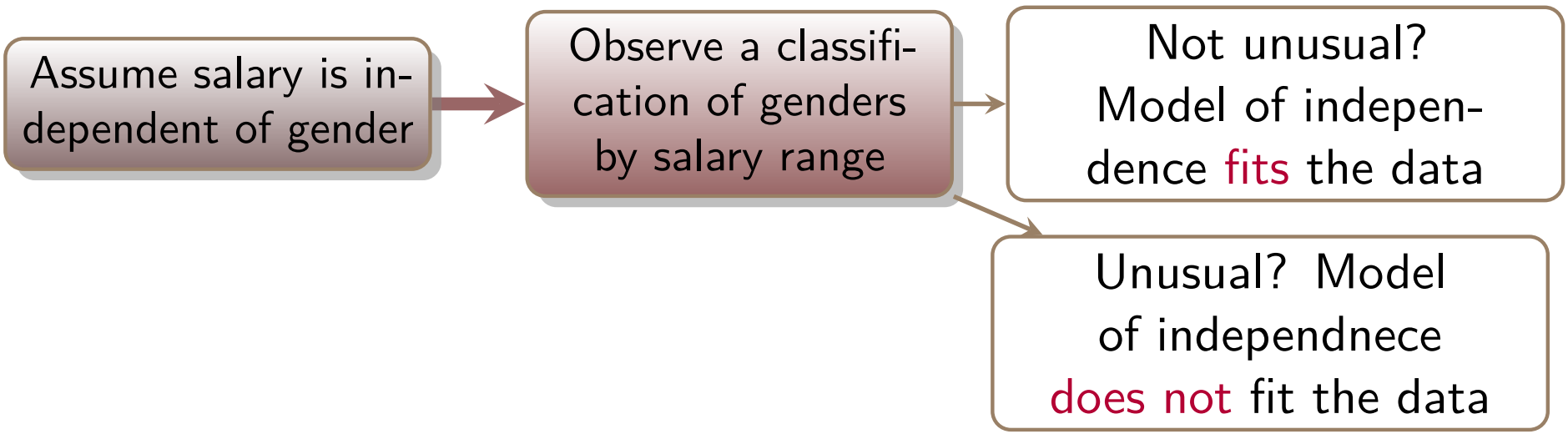
A simple search: Chicago data science salary data



Formal reasoning with data: independence example

- **Modeling**: Construct a statistical model for independence.
- **Question**: Does the model fit the observed set of gender vs. salary ranges?

  (Can it adequately explain how the salary data was generated?)
- **Process**:



Models with a design matrix

- $X_1, \ldots, X_k$ discrete random variables, $X_i \in \{1, \ldots, d_i\}$
- $u =$ a k-way contingency table $u \in \mathbb{Z}_{\geq 0}^{d_1 \times \cdots \times d_k}$ [Draw a table!] Flatten u to vector.

Log-linear model

Sufficient statistics $=$ marginals of u: $P_\theta(U = u) = \exp\{\langle Au, \theta \rangle - \psi(\theta\}$.

Example $X_1 \perp\!\!\!\perp X_2$

$$\left[\begin{array}{cccc|cccc|c|cccc}
1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1 \\
\hline
1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 1
\end{array}\right]_{(d_1+d_2)\times d_1 d_2} \cdot \begin{bmatrix} u_{11} \\ \vdots \\ u_{d_1 d_2} \end{bmatrix} = \begin{bmatrix} u_{1+} & \cdots & u_{+d_2} \end{bmatrix}.$$

License

This document is created for Math/Stat 561, Spring 2023.