Causal Inference for Beginners

00 A Bigger Picture

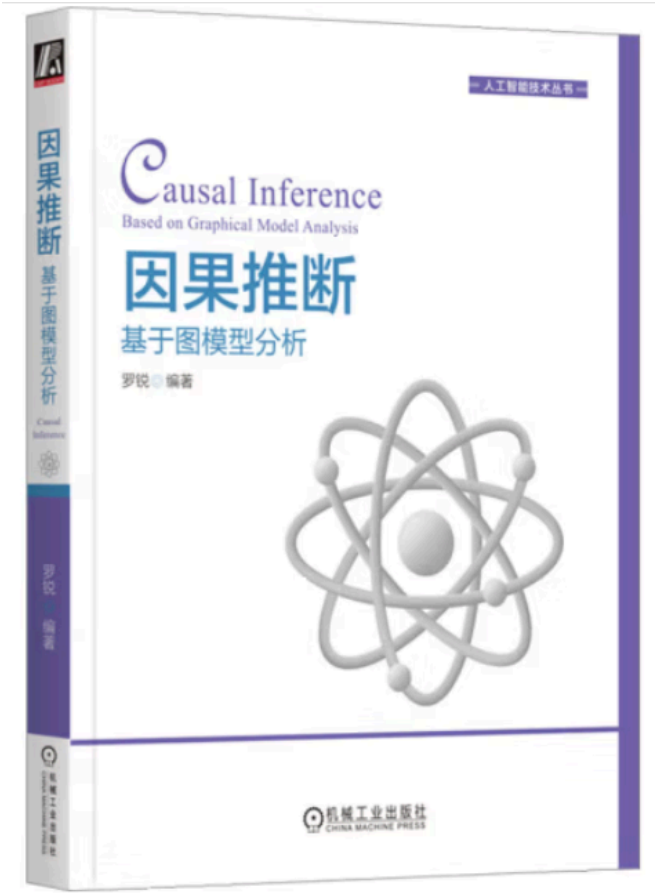
Be a God with Scientific Theory

Jiahao Zhu

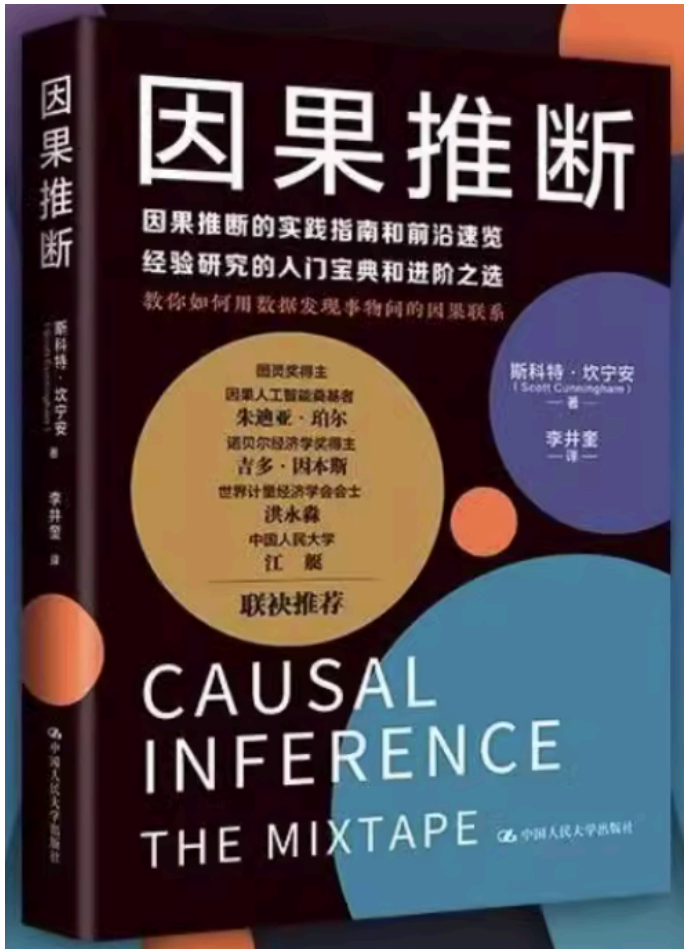
Mars, International School of Business & Finance, Sun Yat-sen University

2023-09-12

Preface



《因果推断：基于图模型分析》罗锐



《因果推断》斯科特·坎宁安

Outline

1. Warm-up
2. What is theory?
3. the Productive Explanation framework
4. A fictitious case
5. Conclusion

Warm-up

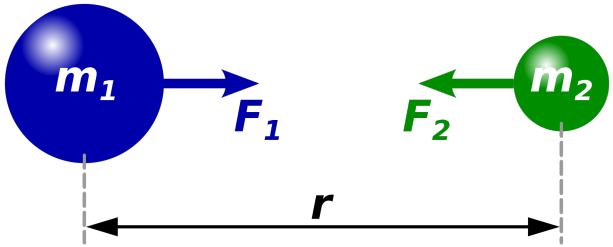
Playing the God

And God said, Let there be light: and there was light.

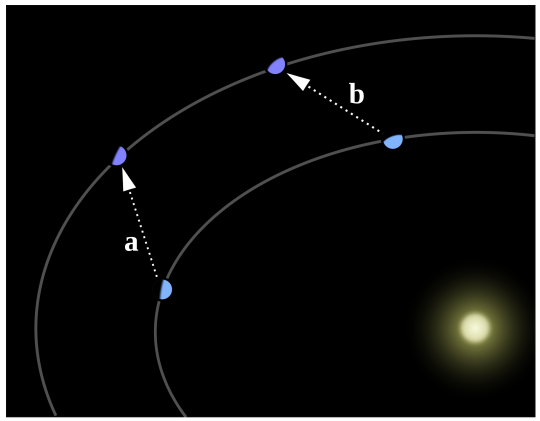
And the authors said, Let there be effects: and there were effects.

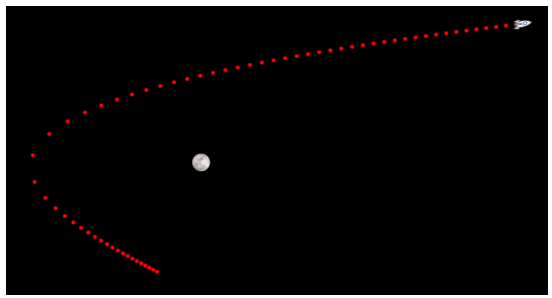
What is theory?

- Scientific theory (Lewin, 1943)
  - Understanding
  - Predicting
  - Controlling
- Example: Newton's law of gravitation
  - Understanding: $F = \frac{G \times m_1 \times m_2}{r^2}$
  - Predicting: the discovery of Neptune (海王星)
  - Controlling: gravitational slingshot (引力弹弓)



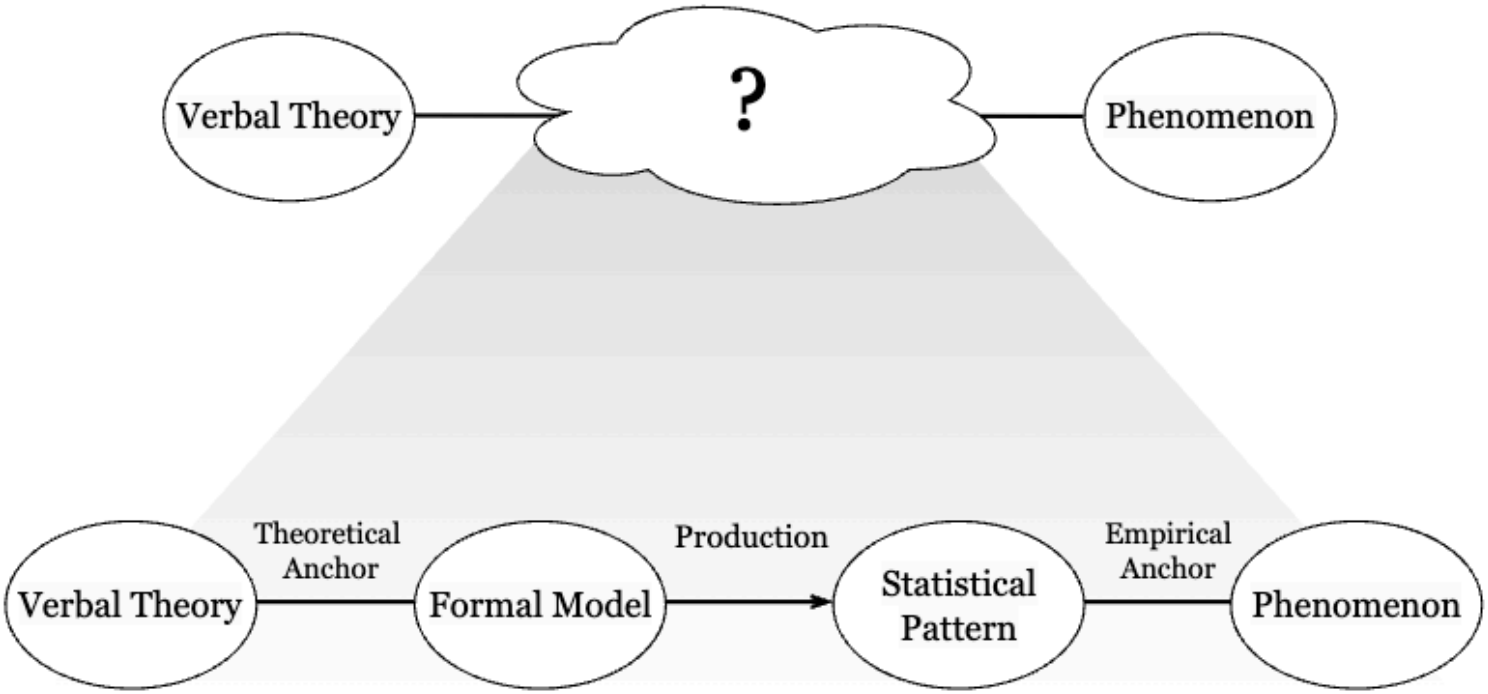
$$F_1 = F_2 = G \frac{m_1 \, x \, m_2}{r^2}$$





Theory is to explain

Figure 1
Explanation as Production.



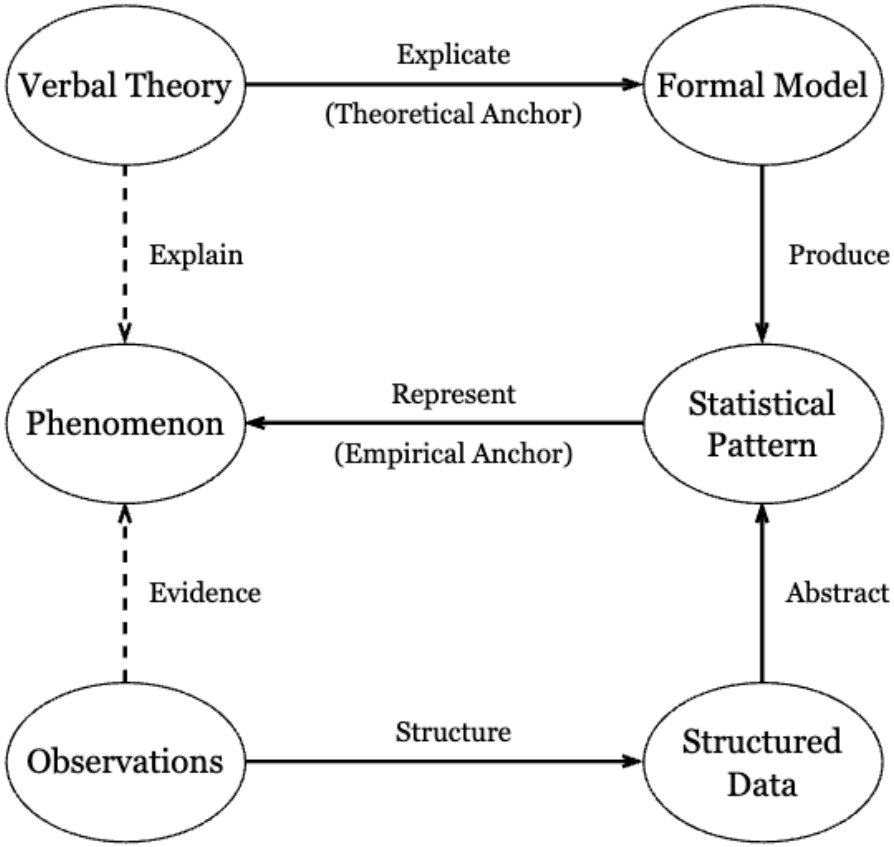
If the world were as theory T says it is, phenomenon P would follow as a matter of course. (Pierce, 1931)

Productive Explanation

Some components

- Observations
  - *unstructured* recordings of sensory information
  - e.g., choices in the MBTI test
- Structured data
  - *structured* recordings of observations
  - encoding/measurement
  - e.g., MBTI scores

Figure 2
The Productive Explanation Model.

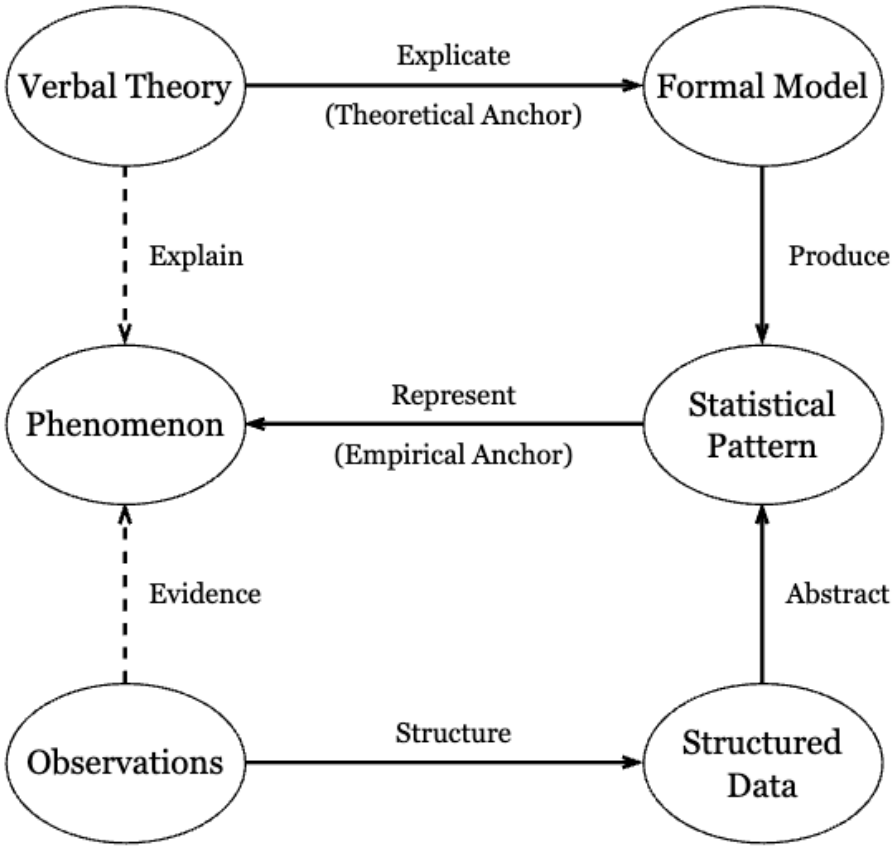


Productive Explanation

Some components

- Phenomenon
  - *general* patterns that need to be explained
  - e.g., "people with certain MBTI scores are more likely to smoke."
- Statistical pattern
  - *general* and *mathematical* relations or patterns in (most) data
  - statistical inference/model
  - e.g., significant correlations between MBTI scores and smoking

Figure 2
The Productive Explanation Model.



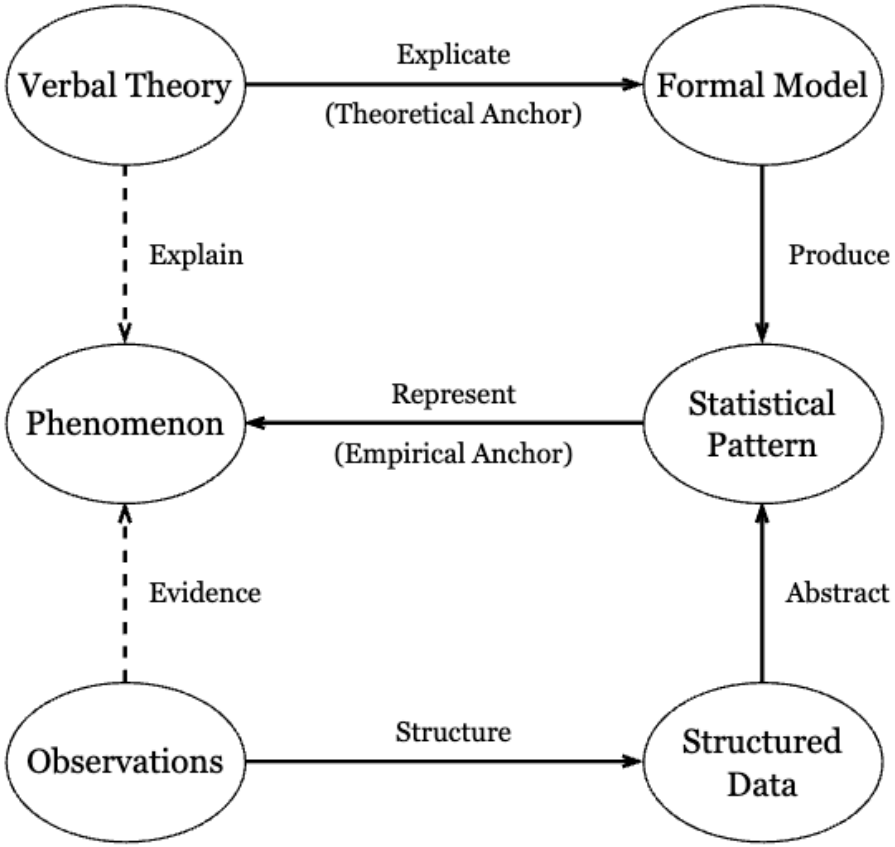
Productive Explanation

Some components

- Verbal theory
  - *assumptions* about the world to explain a phenomenon
    - ★ entities
    - ★ dimensions
    - ★ structures
    - ★ relations (dynamical and causal)
- Formal model
  - *precise* statements of components and relations
  - e.g., utility function, demand/supply function

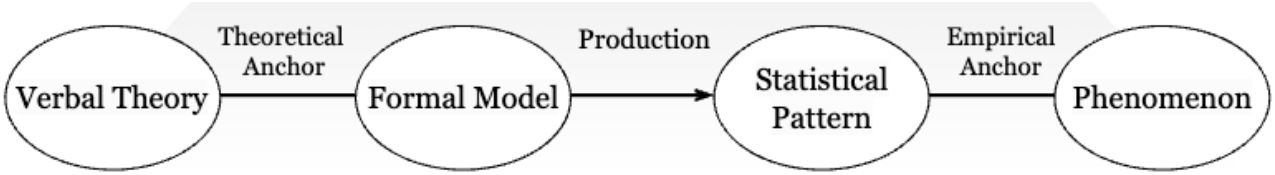
Figure 2

The Productive Explanation Model.



Productive Explanation

Explanation chain



Explanation chain

- Step 1: Represent phenomenon as a statistical pattern (or reversely)
- Step 2: Explicate the verbal theory as a formal model
- Step 3: Evaluate whether the formal model produces the statistical pattern

A fictitious case

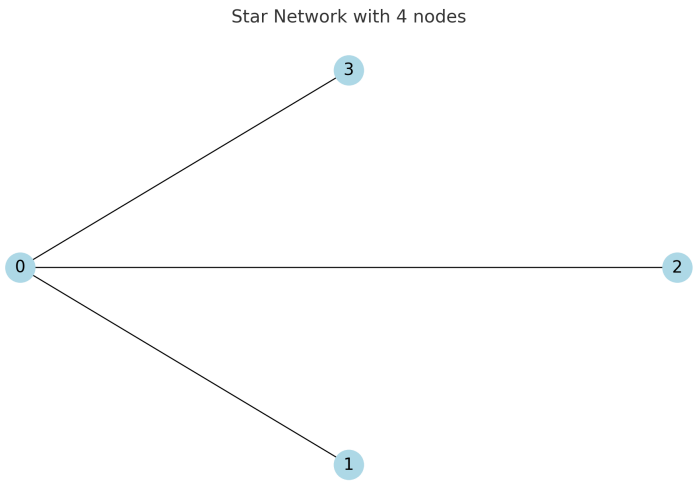
the social influence explanation of smoking

- Phenomenon: introvert people are less likely to smoke
- Statistical pattern: significant negative correlation coefficients between the introversion score and smoking behavior (e.g., r = -0.1, $p < .05$)
  - stable between measurements (e.g., MBTI, Big Five)
  - stable between samples (e.g., country, age, gender)
  - stable between times (e.g., 10-year longitudinal study)

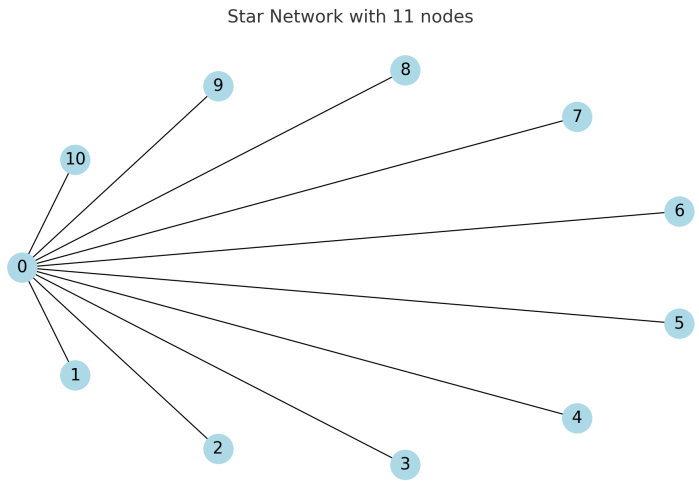
A fictitious case

the social influence explanation of smoking

- Verbal theory: the social influence hypothesis of smoking
  - smoking is a social behavior
  - introvert people have smaller social networks
  - introvert people are less likely to interact with others
  - therefore, introvert people are less likely to smoke



network of the introvert



network of the extrovert

A fictitious case

the social influence explanation of smoking

- Formal model: the social influence model of smoking
  - the introversion score (*intro*) of a person

$$intro \sim N(50, 15^2), intro \in [0, 100]$$

  - network size

$$N = \theta_1 \times (100 - intro)(\text{set } \theta_1 = 0.1)$$

  - fixed smoking probability of others

$$P_{smoke,others} = \theta_2(\text{set } \theta_2 = 0.2)$$

  - probability to smoke

$$P_{smoke,individual} = \frac{intro}{100} \times \frac{N_{smoke}}{N}$$

A fictitious case

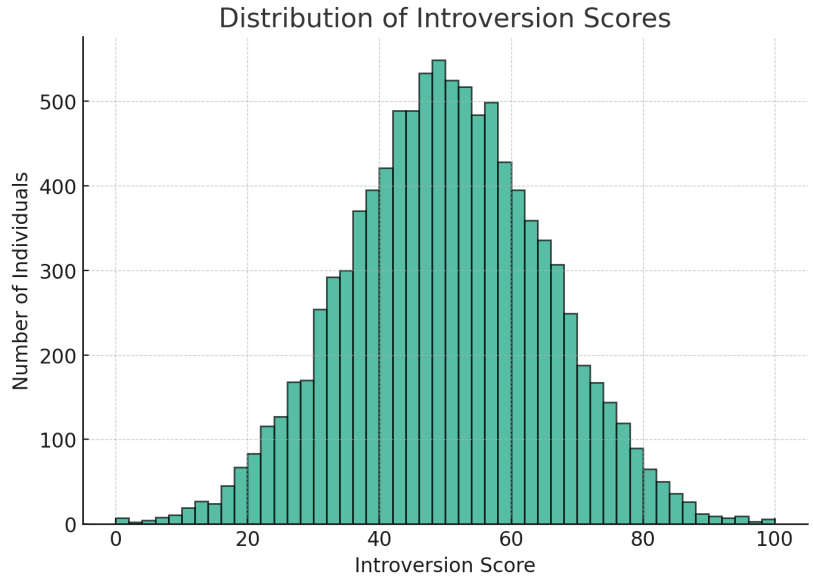
simulation codes

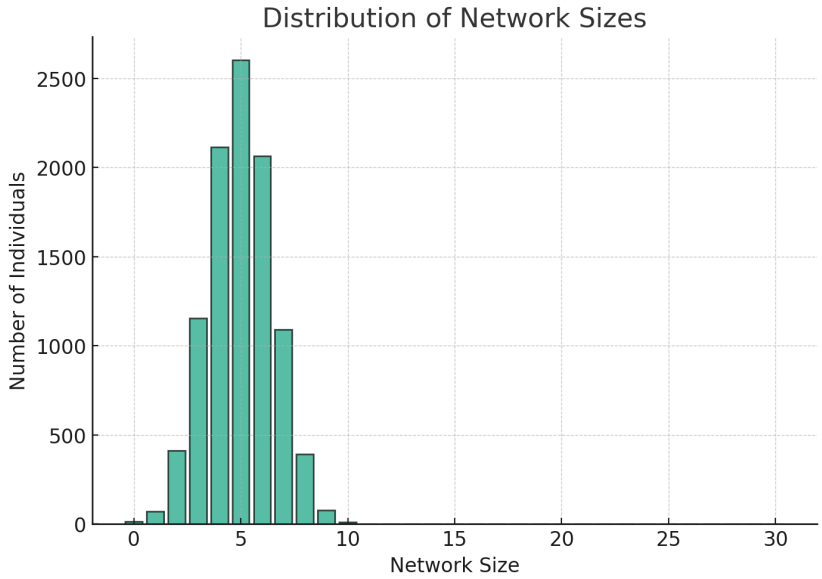
```r
# Simulate introversion scores
intro_scores <- rnorm(num_individuals, mean=50, sd=15)
intro_scores <- pmin(pmax(intro_scores, 0), 100)  # Clipping values to be in [0, 100]

# Calculate network sizes
network_sizes <- 100 * theta1_new - theta1_new * intro_scores
network_sizes_rounded <- round(network_sizes)

# For each individual's social circle, determine the number of smokers
num_smokers_in_network <- rbinom(num_individuals, size=pmax(0, network_sizes_rounded), prob=theta2_new)

# Calculate the probability of each individual smoking
P_individual_smokes <- theta3_new * (num_smokers_in_network / (network_sizes_rounded + 1e-10))  # added a small constant to avoid division by zero

# Determine if the individual smokes based on the calculated probability
individual_smokes <- runif(num_individuals) < P_individual_smokes
individual_smokes_int <- as.integer(individual_smokes)

# Calculate the correlation coefficient between smoking behavior and introversion scores for this simulation
correlation_coefficient <- cor(individual_smokes_int, intro_scores)

# Store the correlation coefficient
correlation_coefficients[i] <- correlation_coefficient
```

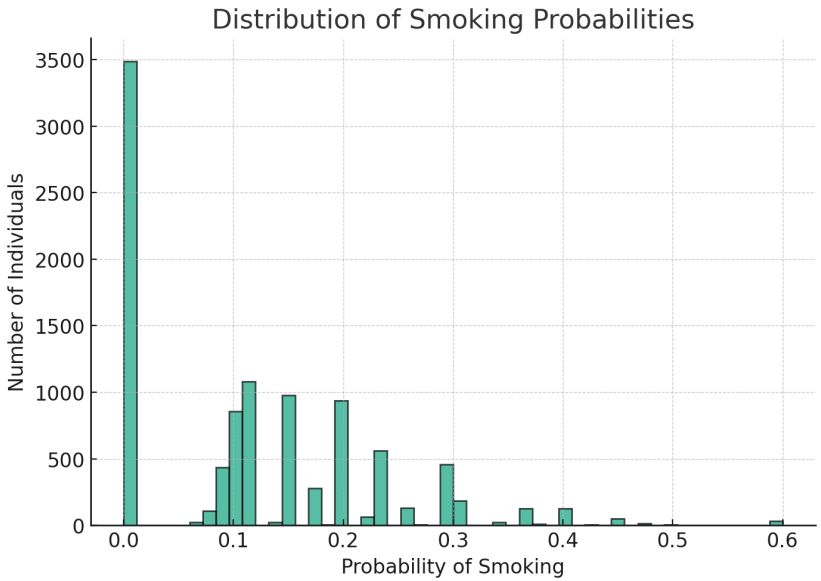
simulation codes in R

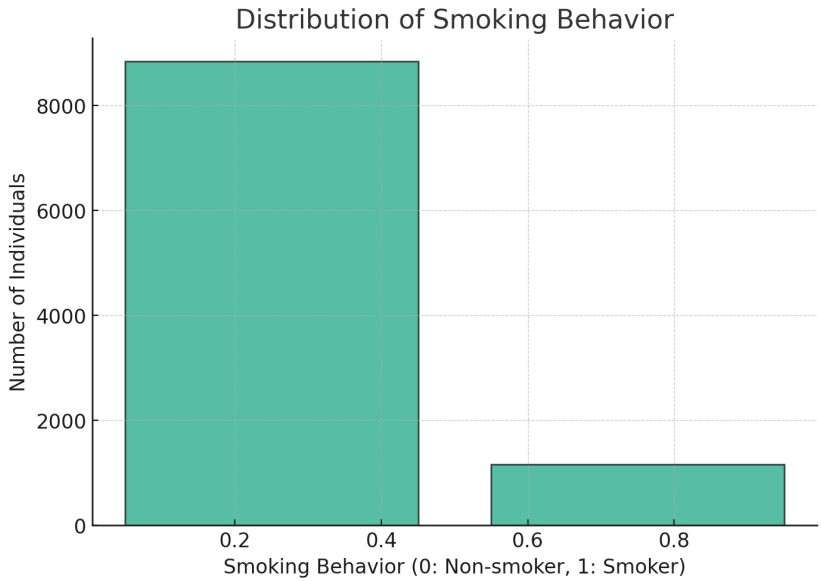
A fictitious case

summary statistics



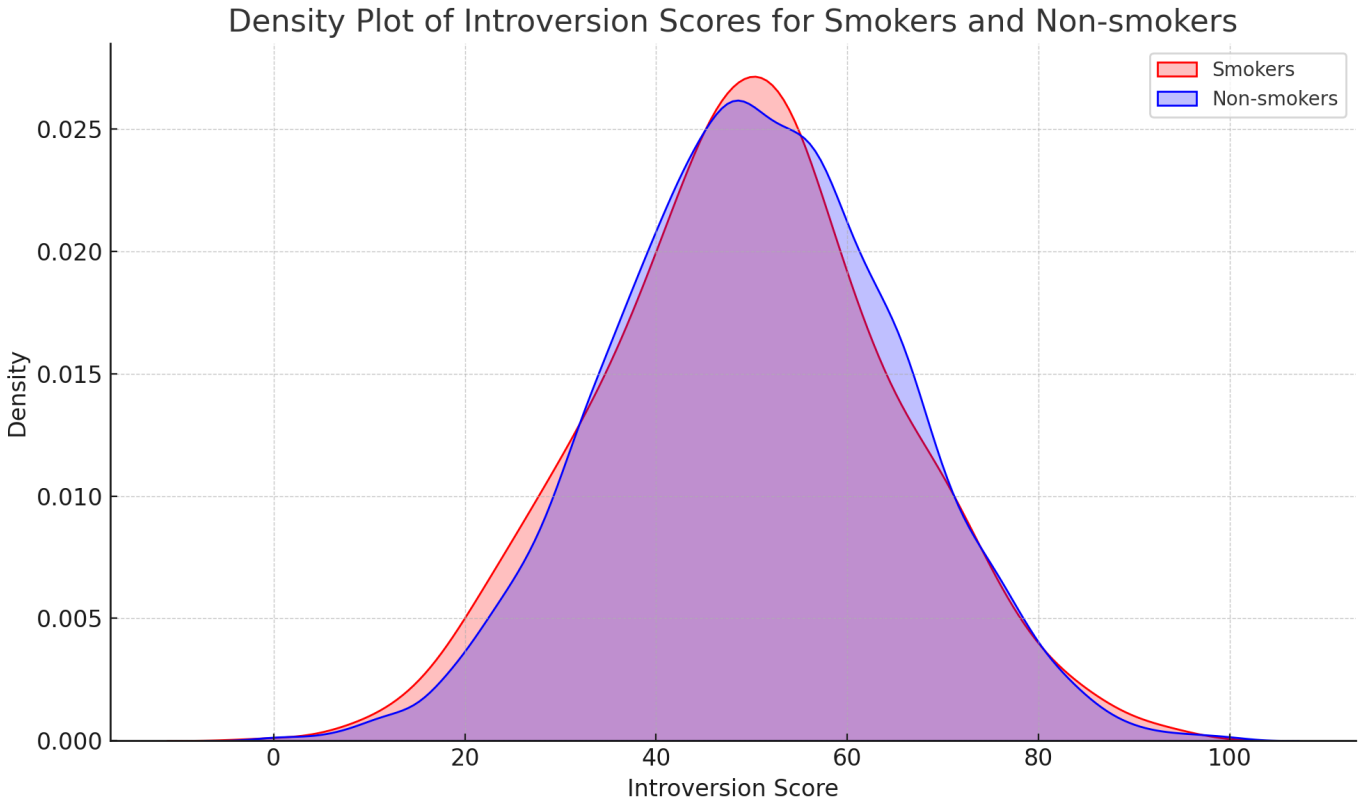






A fictitious case

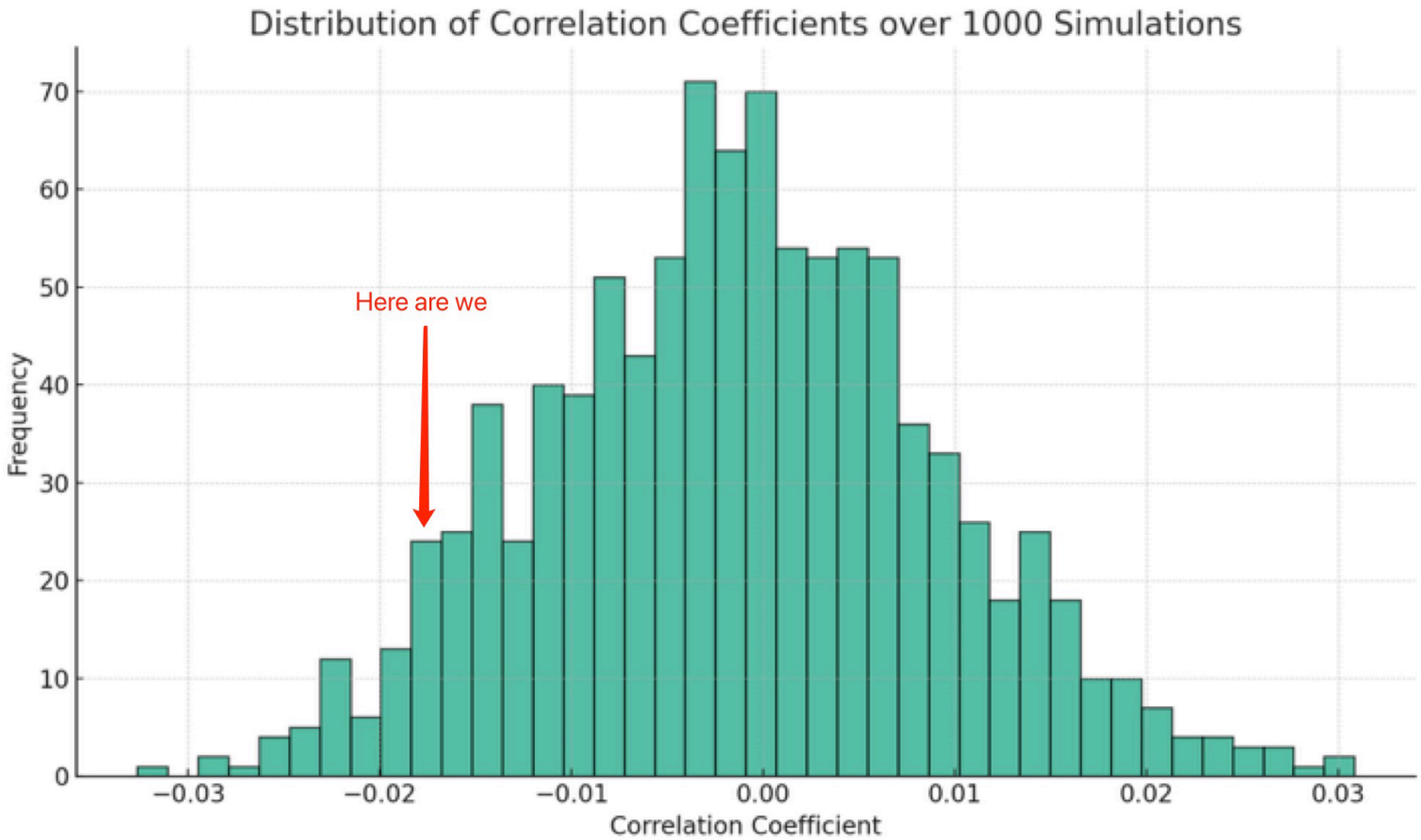
results



$$r_{intro,smoke} = -0.0182, p = .0687$$

A fictitious case

big sample is not everything, multiple sample is



Phenomenon is all about stability!

A fictitious case

the social influence explanation of smoking

- Problems
  - the choice of functions and parameters
  - others' introversion scores
  - network structure (e.g., random network)
  - mutual influence of smoking
  - ...
- Advantages of formal model
  - simple model makes big differences
  - clear assumptions
  - clear validations
  - clear future directions
  - ...

Conclusion

- What is scientific theory?
  - playing God!
- Productive explanation framework
- Simple but illustrative case

References

- Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science, 16*(4), 756-766.
- van Dongen, N., van Bork, R., Finnemann, A., van der Maas, H., Robinaugh, D., Haslbeck, J., ... & Borsboom, D. (2022). Productive Explanation: A Framework for Evaluating Explanations in Psychological Science. *PsyArXiv*.

Thanks for listening

Q&A