Universidad Tecnológica de Pereira

MAESTRÍA EN INVESTIGACIÓN OPERATIVA Y ESTADÍSTICA

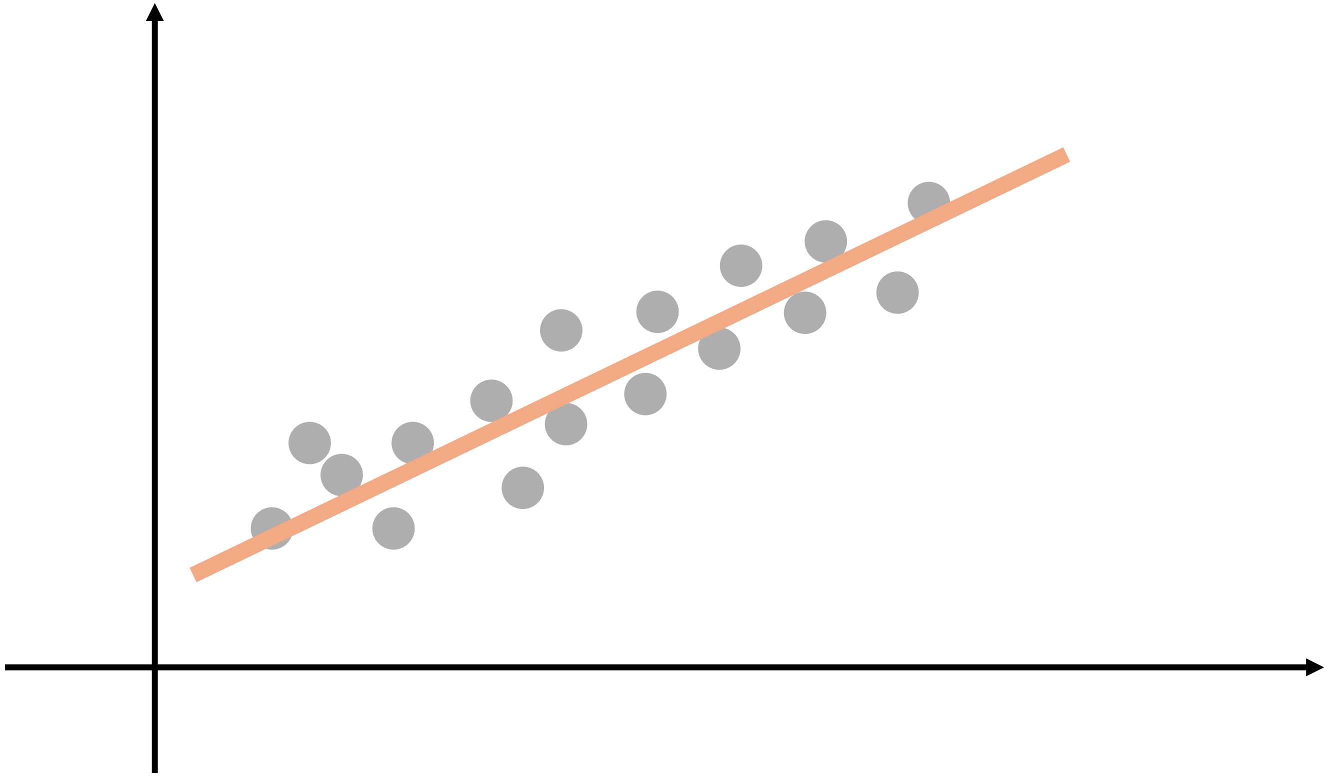
Multivariate statistics

Structural Equation Models SEM

June 07, 2024

Orlando Sabogal-Cardona
@Antonio Sabogal
orlando.sabogal.20@ucl.ac.uk

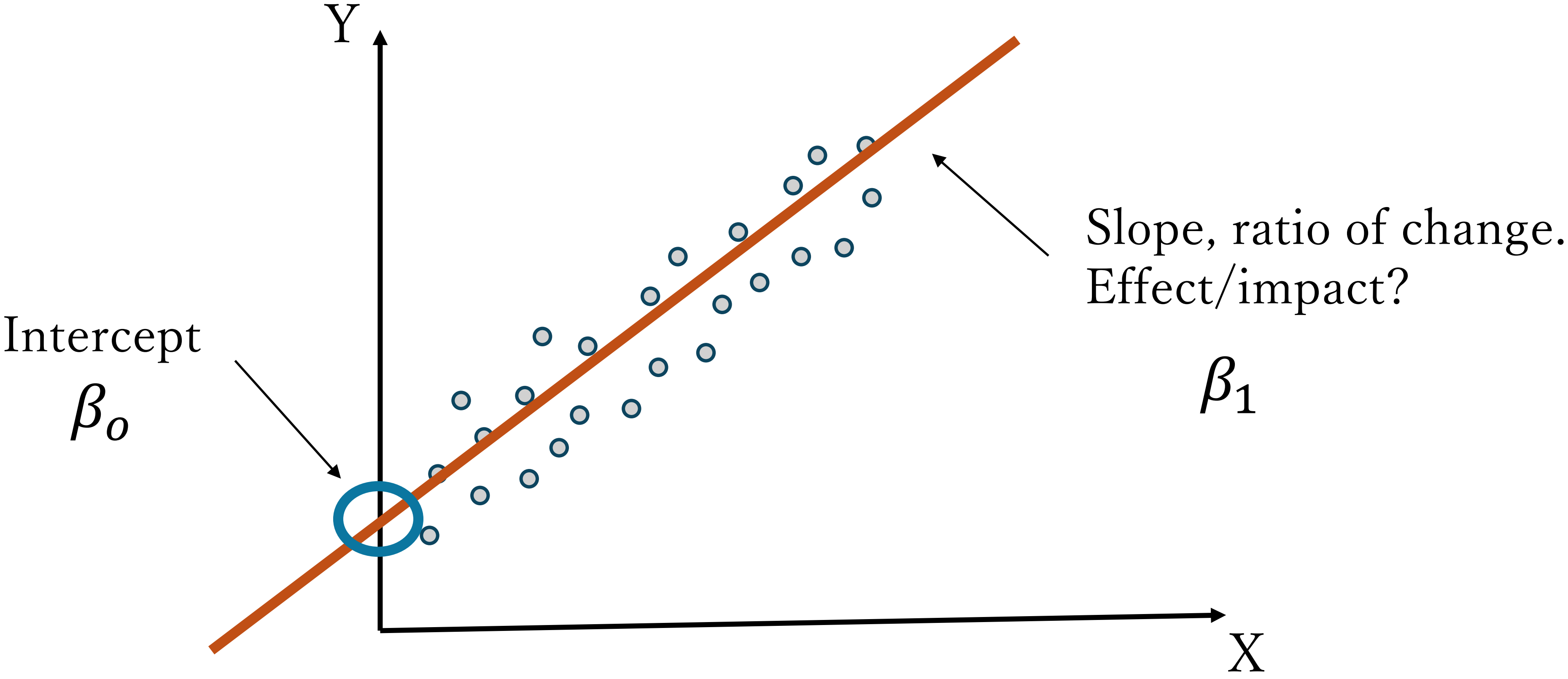
A note on linear regression



- The line of "best fit"
- "Explain" Y given X
- An abstraction of how the real-world works

An easy way to think about LR: You are trying to figure out the ingredients in your food.

Master LR and will conquer the world (of statistics)



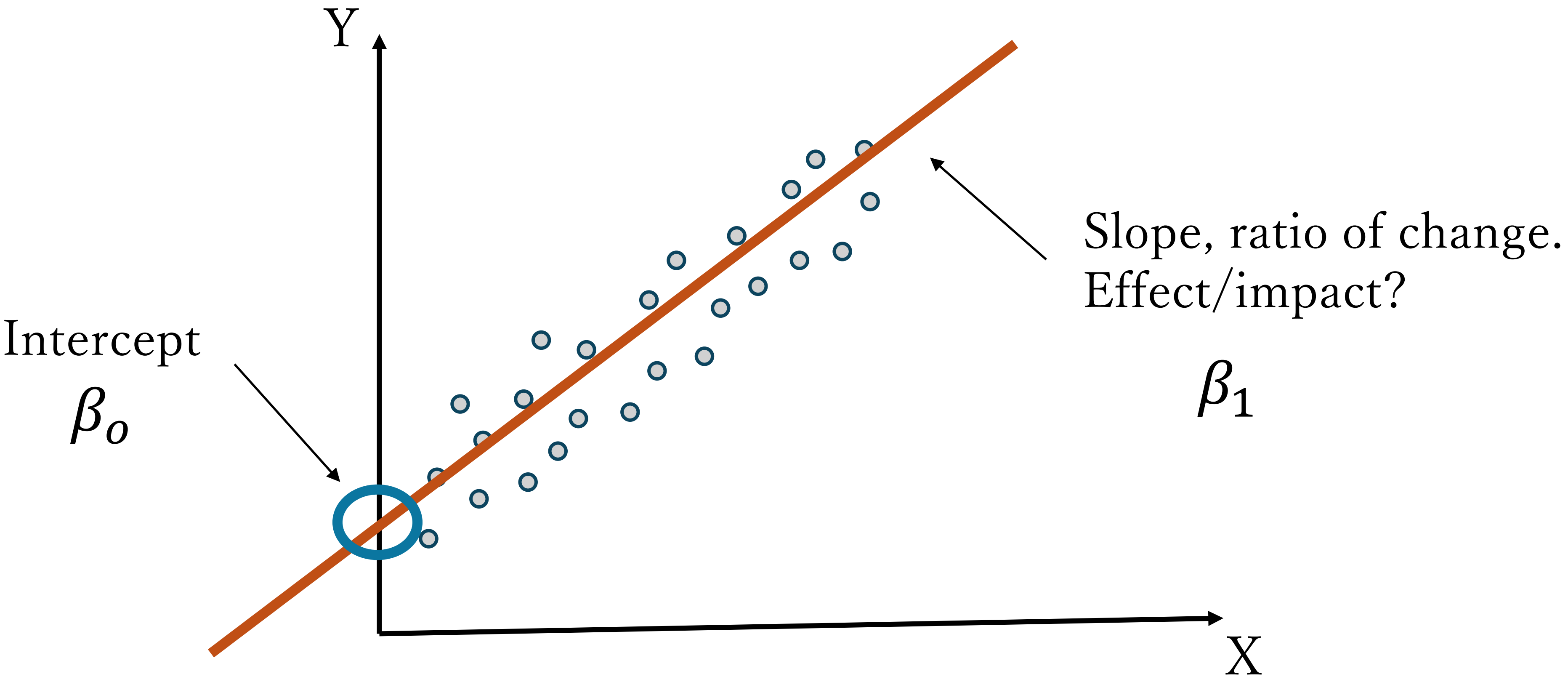
Intercept

β_o

Slope, ratio of change.
Effect/impact?

β_1

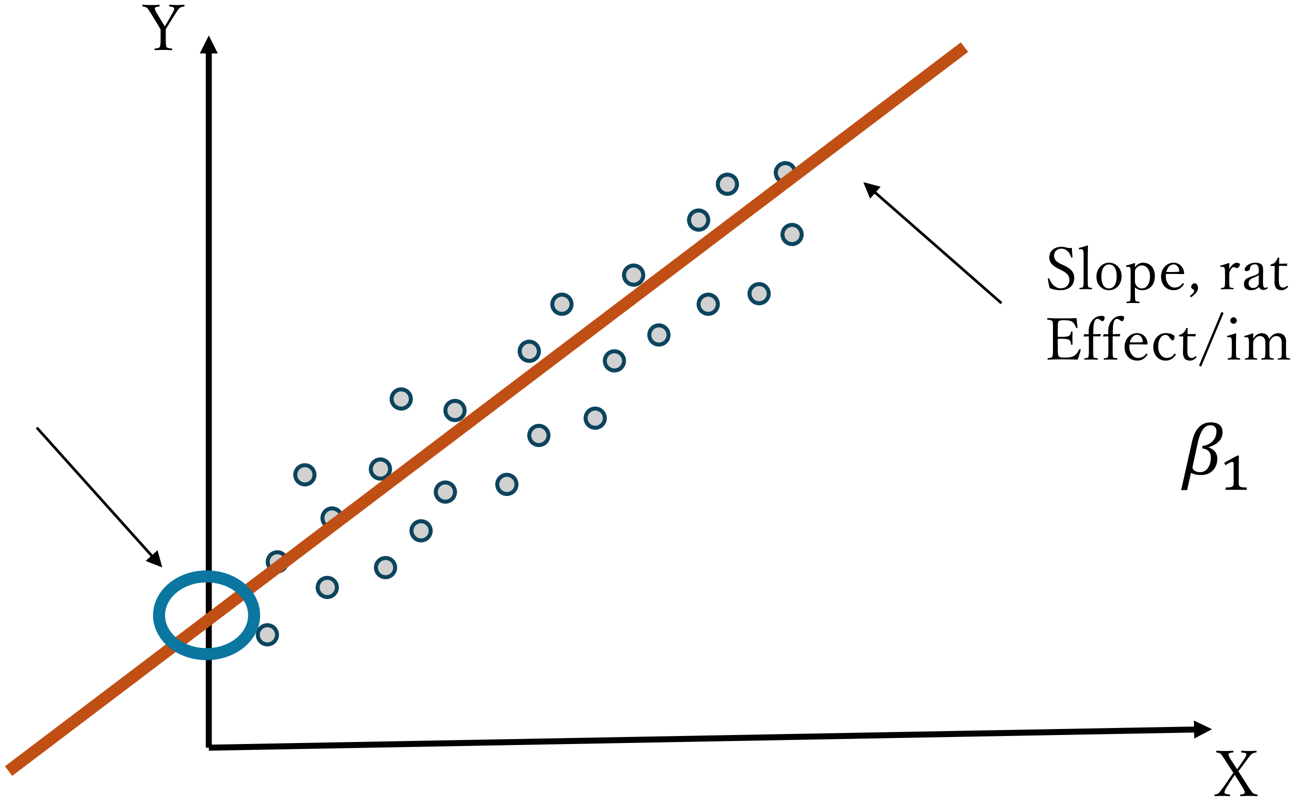
$$Y = \beta_0 + \beta_1 X$$



$$Y = \beta_0 + \beta_1 X$$

But remember: we have a sample and we do not know the parameters β_0 and β_1

A useful way to think about this equation is a the "data generator process"



Intercept

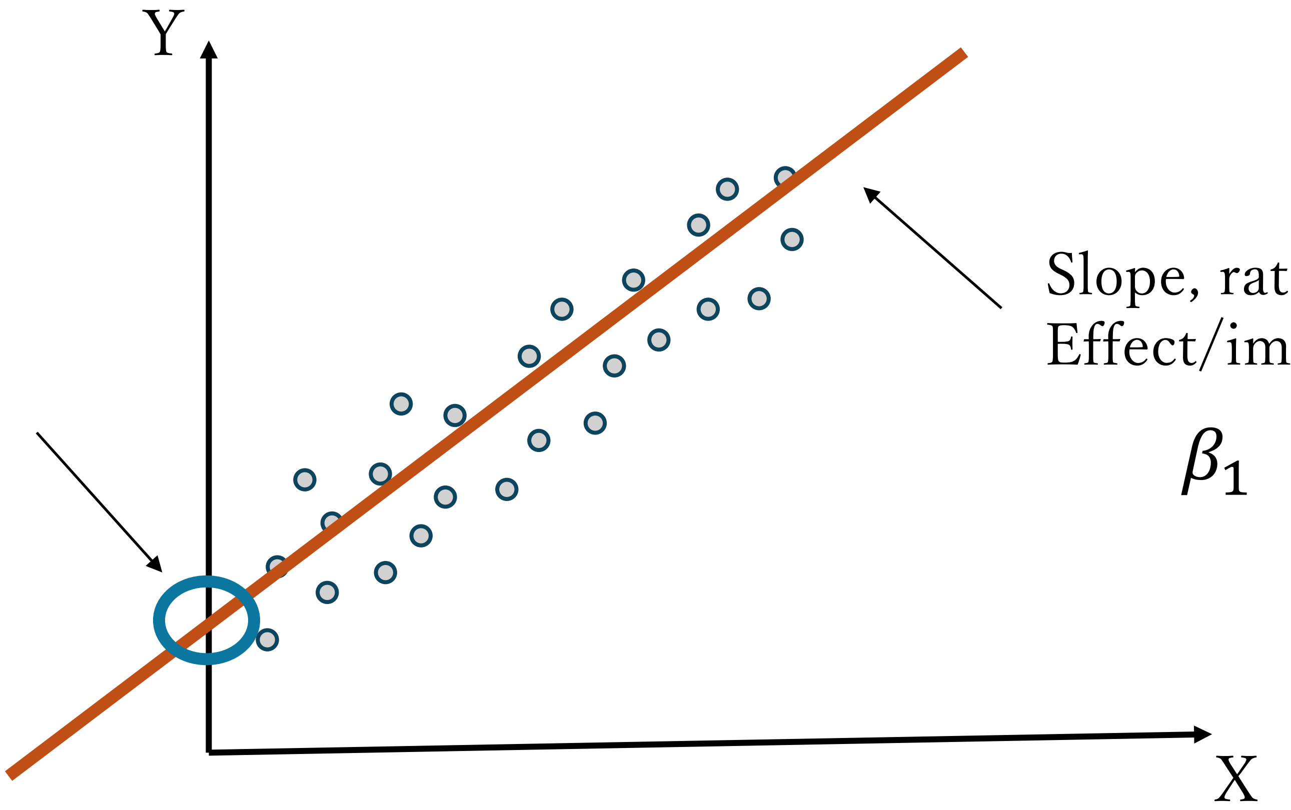
β_o

Slope, ratio of change.
Effect/impact?

β_1

$$\left.\begin{aligned} Y &= \beta_0 + \beta_1 X \\[2em] \hat{Y} &= \widehat{\beta_0} + \widehat{\beta_1} X + e \end{aligned}\right\}$$

Take a minute here to
remember the Central
Limit Theorem



Intercept

β_o

Slope, ratio of change.
Effect/impact?

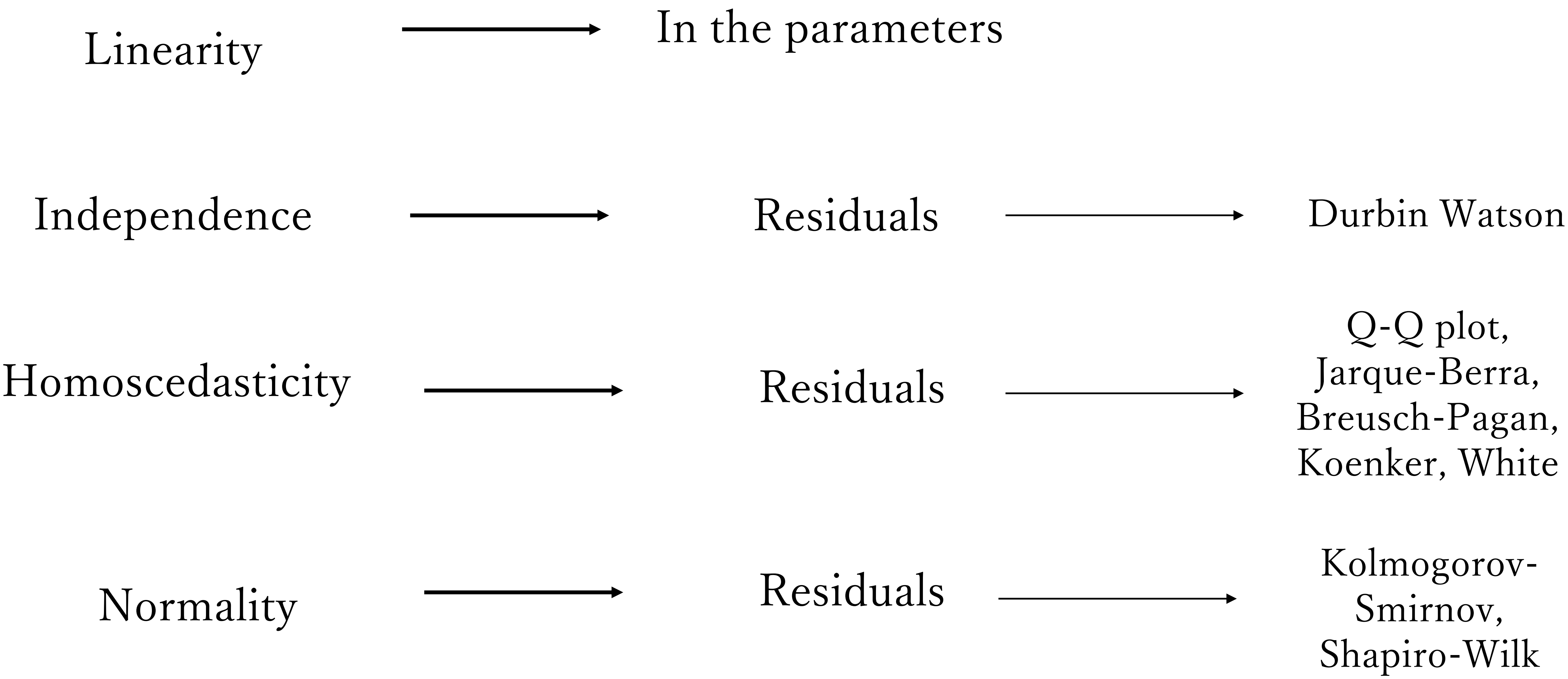
β_1

$$Y = \beta_0 + \beta_1 X$$

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X + e$$

$e = Y - \hat{Y}$ ⟶ Error/residual

The assumptions



Linearity → In the parameters

Independence → Residuals → Durbin Watson

Homoscedasticity → Residuals → Q-Q plot, Jarque-Berra, Breusch-Pagan, Koenker, White

Normality → Residuals → Kolmogorov-Smirnov, Shapiro-Wilk

You should also check for multicollinearity → Variation Inflation Factor VIF

Linearity → In the parameters

Independence → Residuals → Durbin Watson

Homoscedasticity → Residuals → Q-Q plot, Jarque-Berra, Breusch-Pagan, Koenker, White

Normality → Residuals → Kolmogorov-Smirnov, Shapiro-Wilk

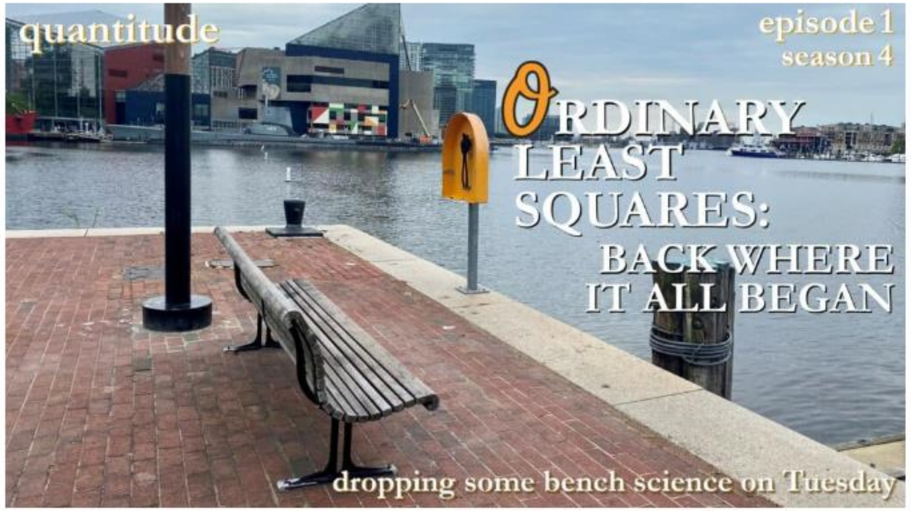
So far, how does the output look like?

Variables	Parameter	Standard Error	t value	p value	
Intercept	---	---	---	---	*
X1	---	---	---	---	
X2	---	---	---	---	*
X2	---	---	---	---	
X4	---	---	---	---	***
X5	---	---	---	---	**

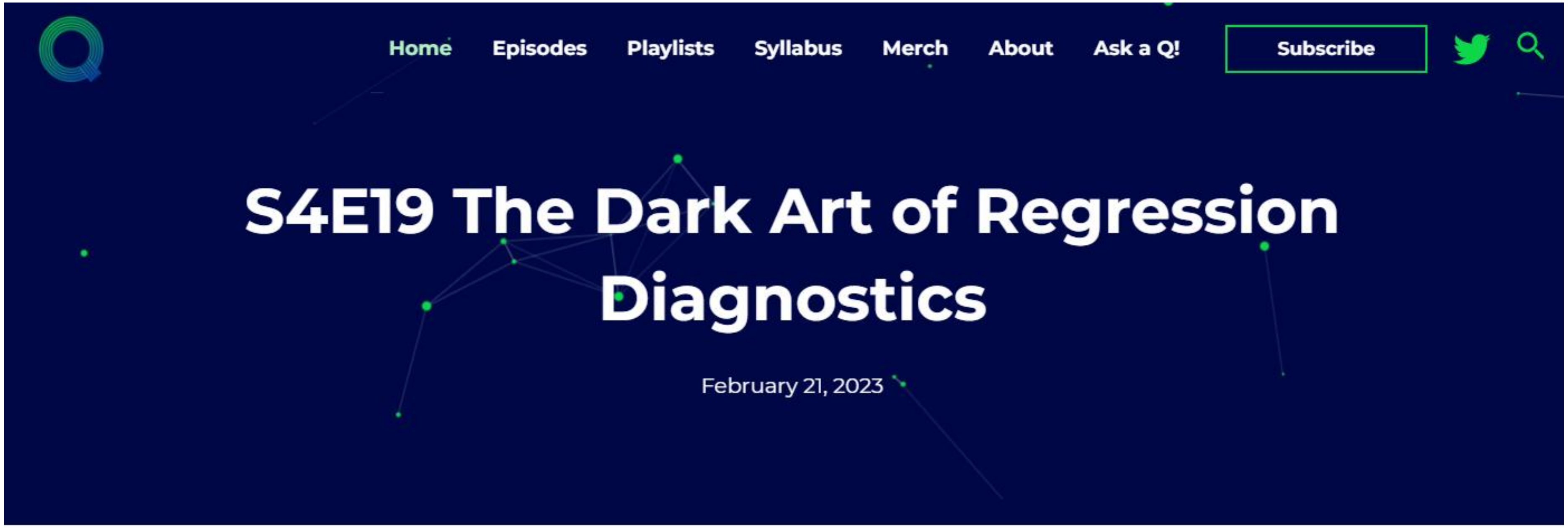
R-squared and adjust R-squared are also presented
Significance of parameter: t-statistic
Significance of parameter (with reference to a value): Wald test
Overall significance of the regression (all parameters = 0): F-test

S4E01 Ordinary Least Squares: Back Where It All Began

September 13, 2022

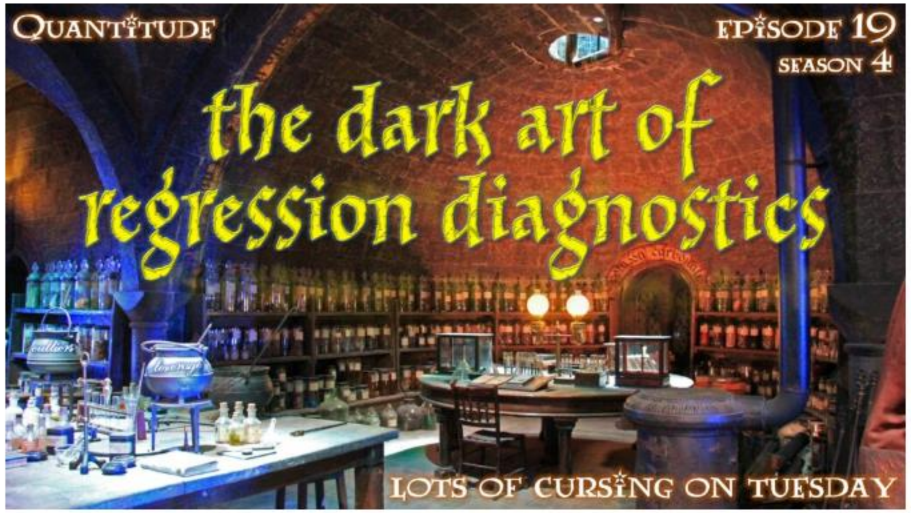


https://quantitudepod.org/s4e01-ols/



S4E19 The Dark Art of Regression Diagnostics

February 21, 2023



https://quantitudepod.org/s4e19-regression-diagnostics/

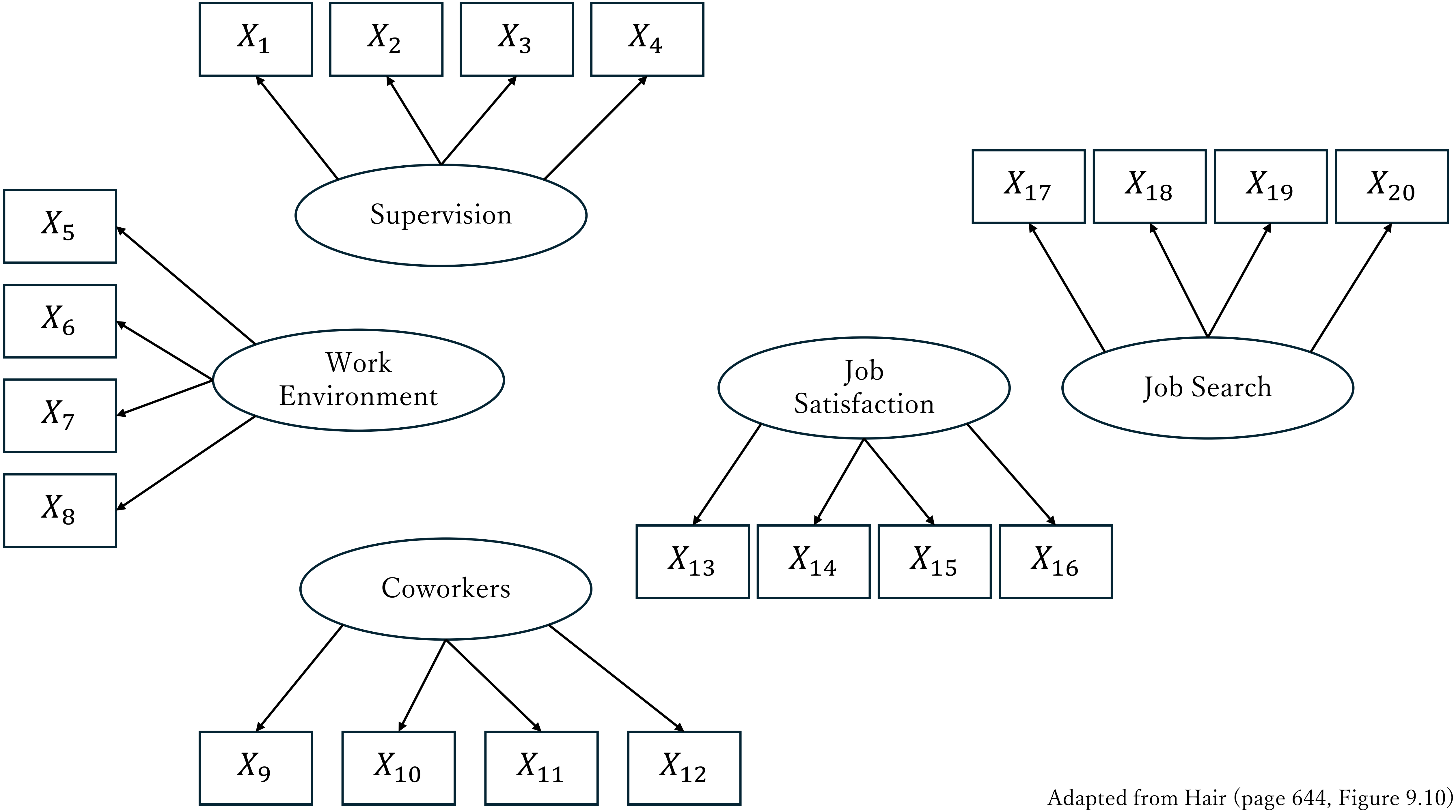
S5E02 Multicollinearity: The Usual Suspect

September 19, 2023

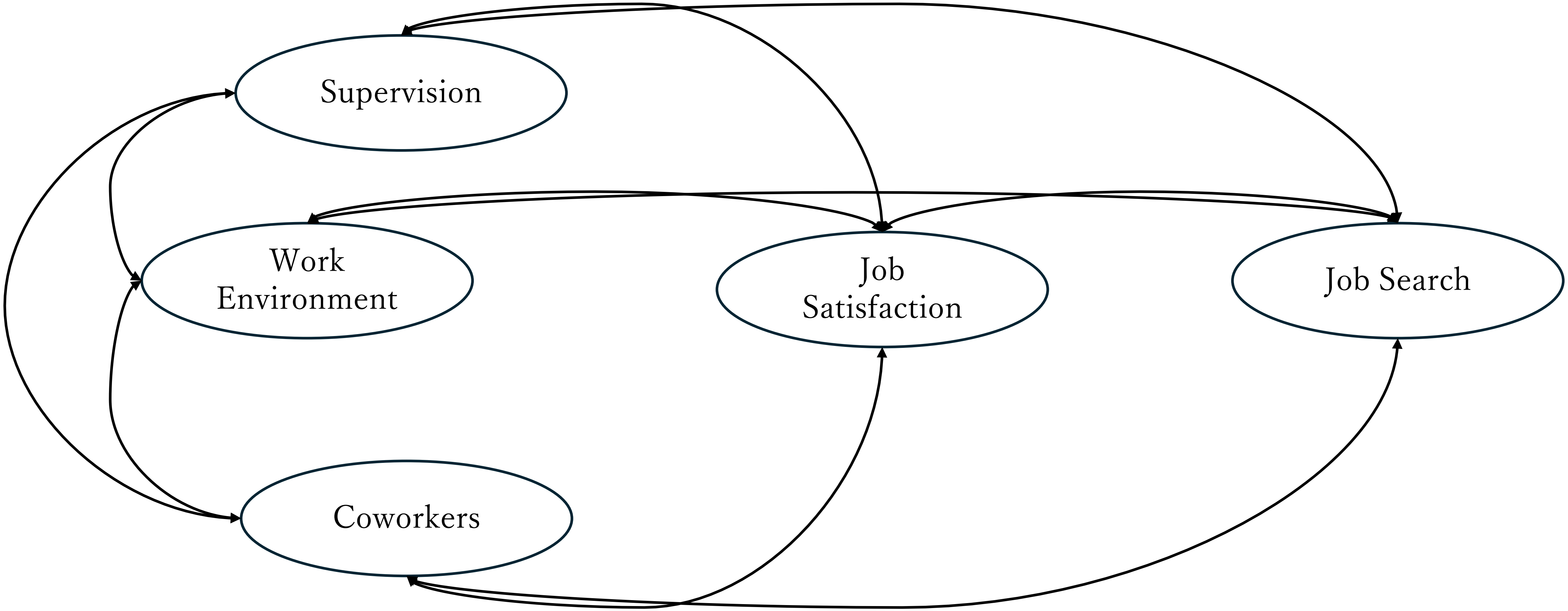
https://quantitudepod.org/s5e02-multicollinearity/

Structural Equation Models SEM

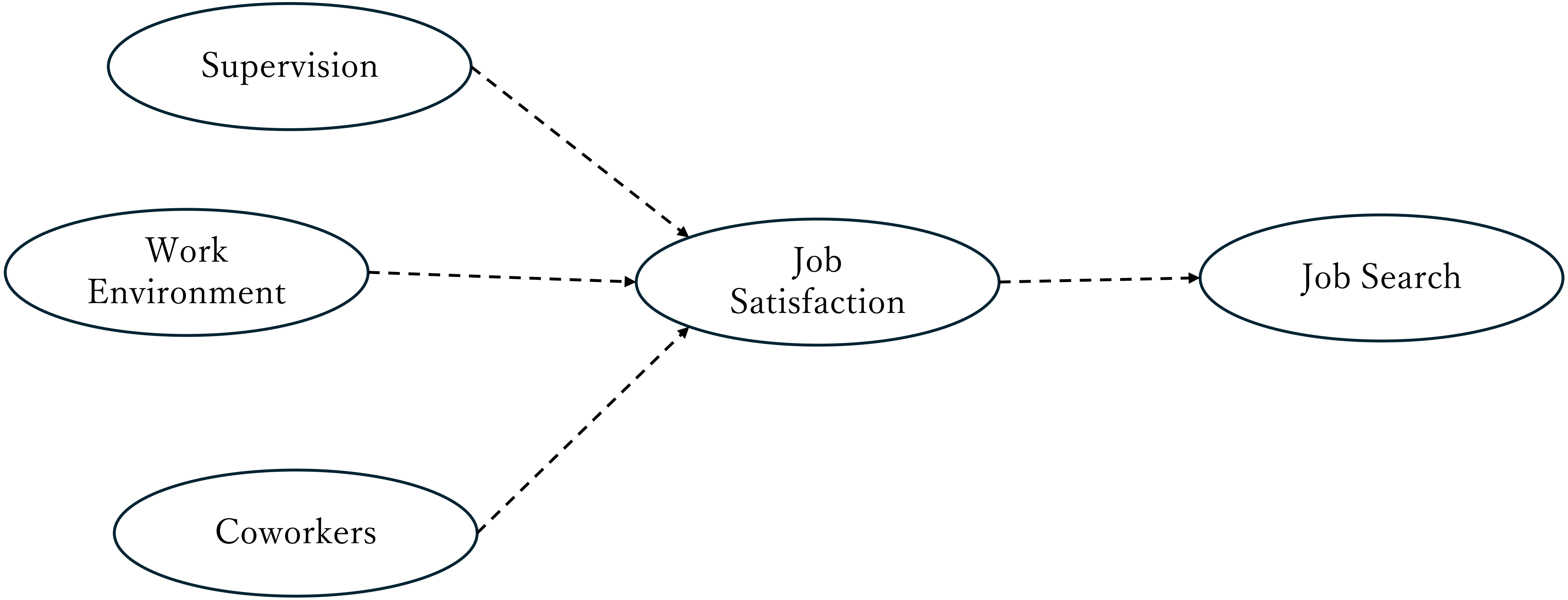
- An extension of factor analysis and multiple regression analysis.
- SEM explores several relationships simultaneously
- Useful to test a theory represented through a system of equations
- Multivariate and multi-equation research problem



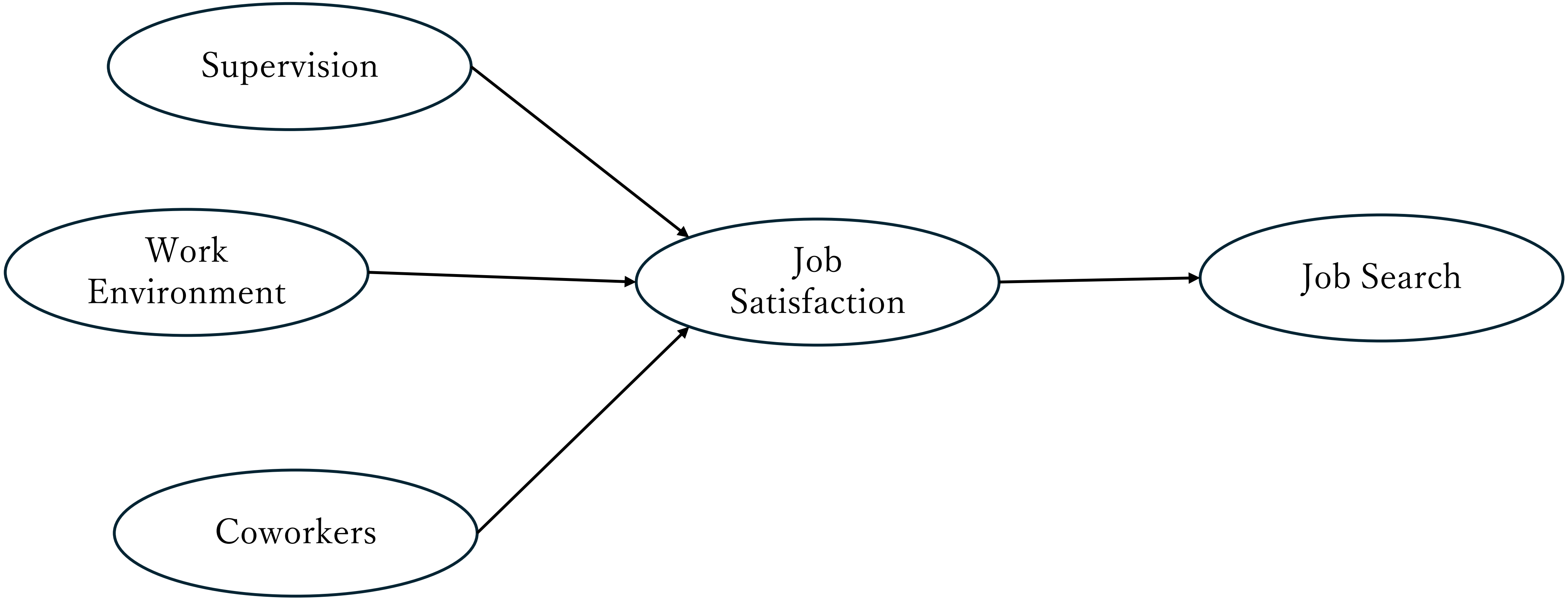
Adapted from Hair (page 644, Figure 9.10)



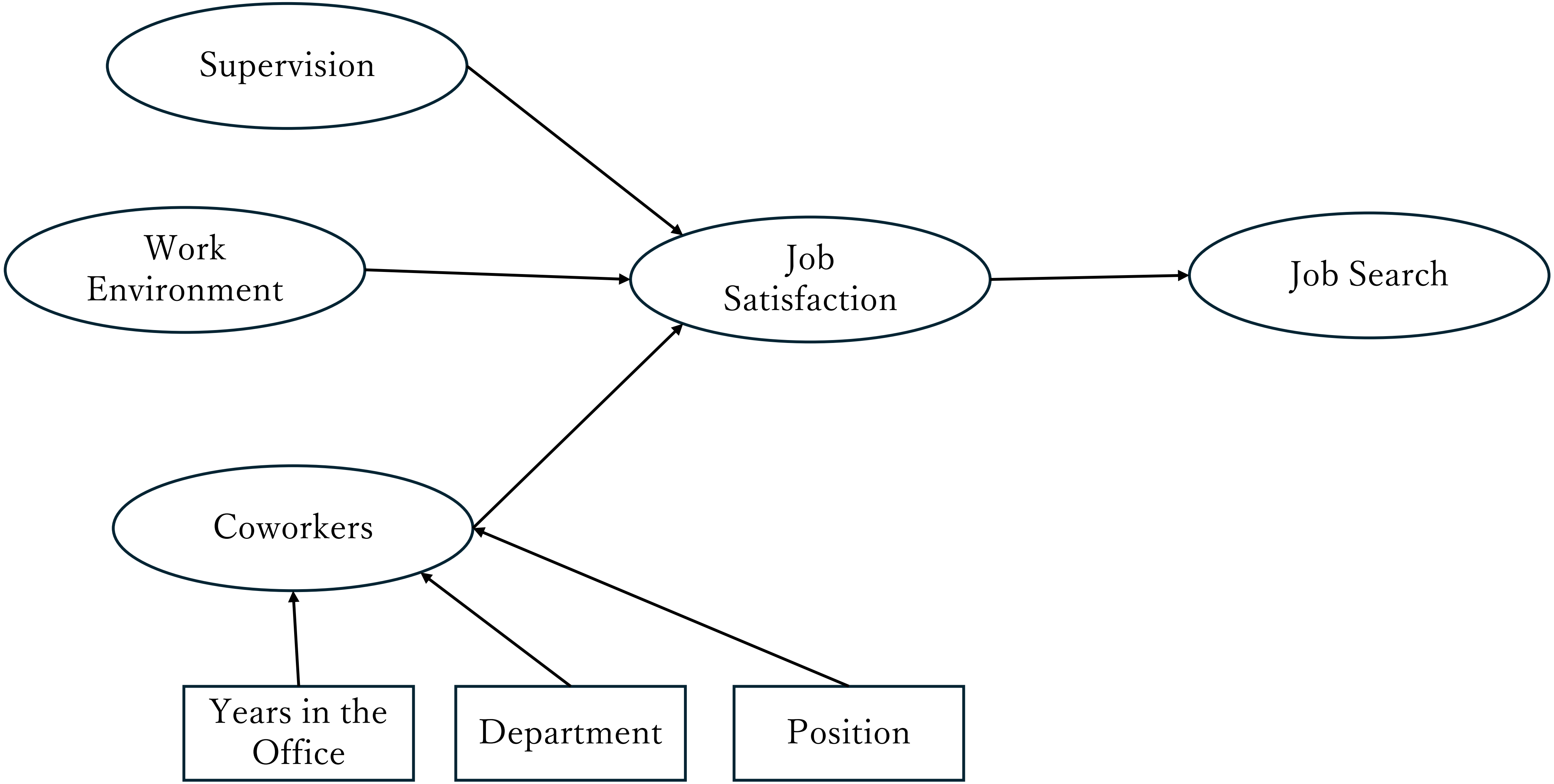
Adapted from Hair (page 621, Figure 9.5)



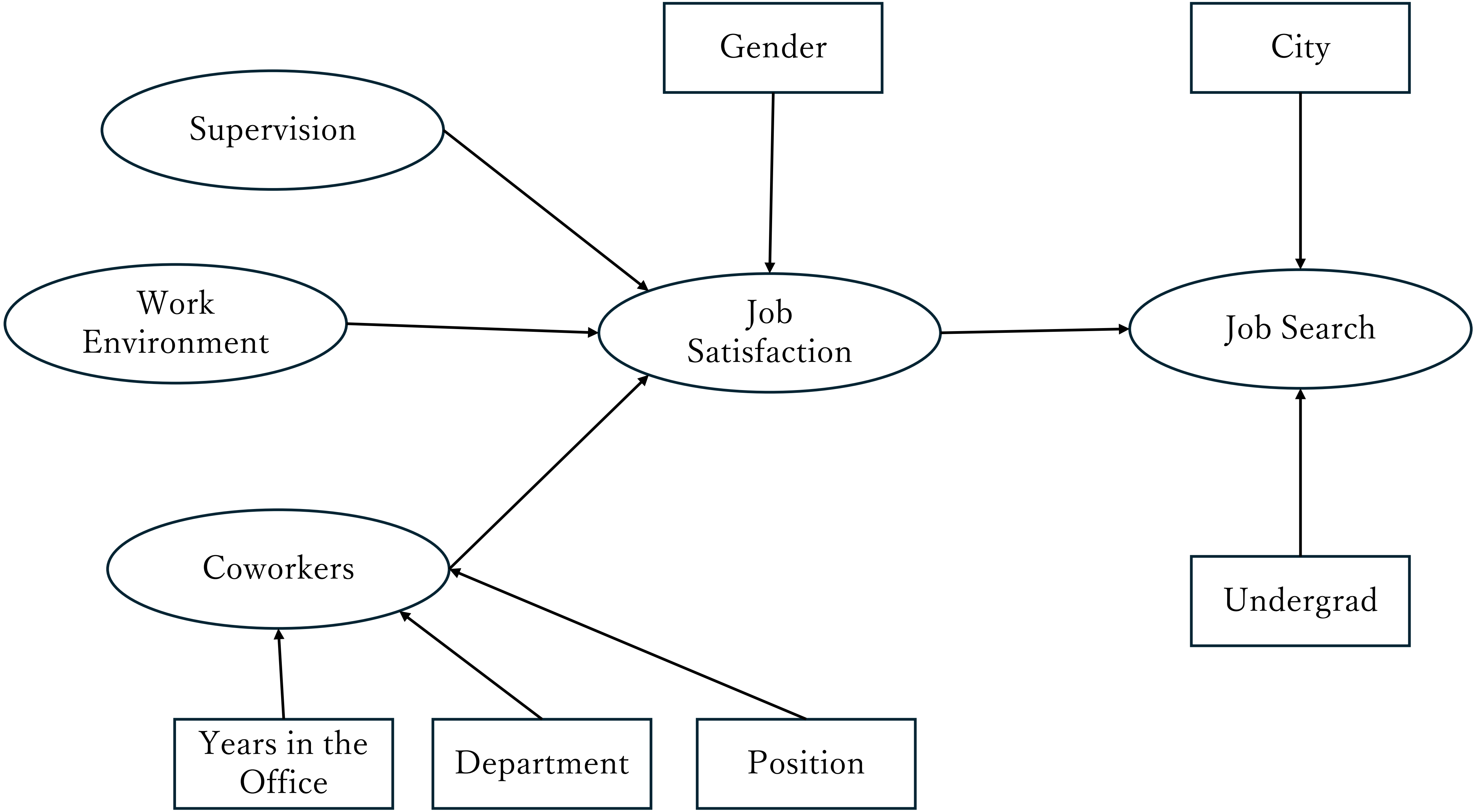
Adapted from Hair (page 621, Figure 9.5)



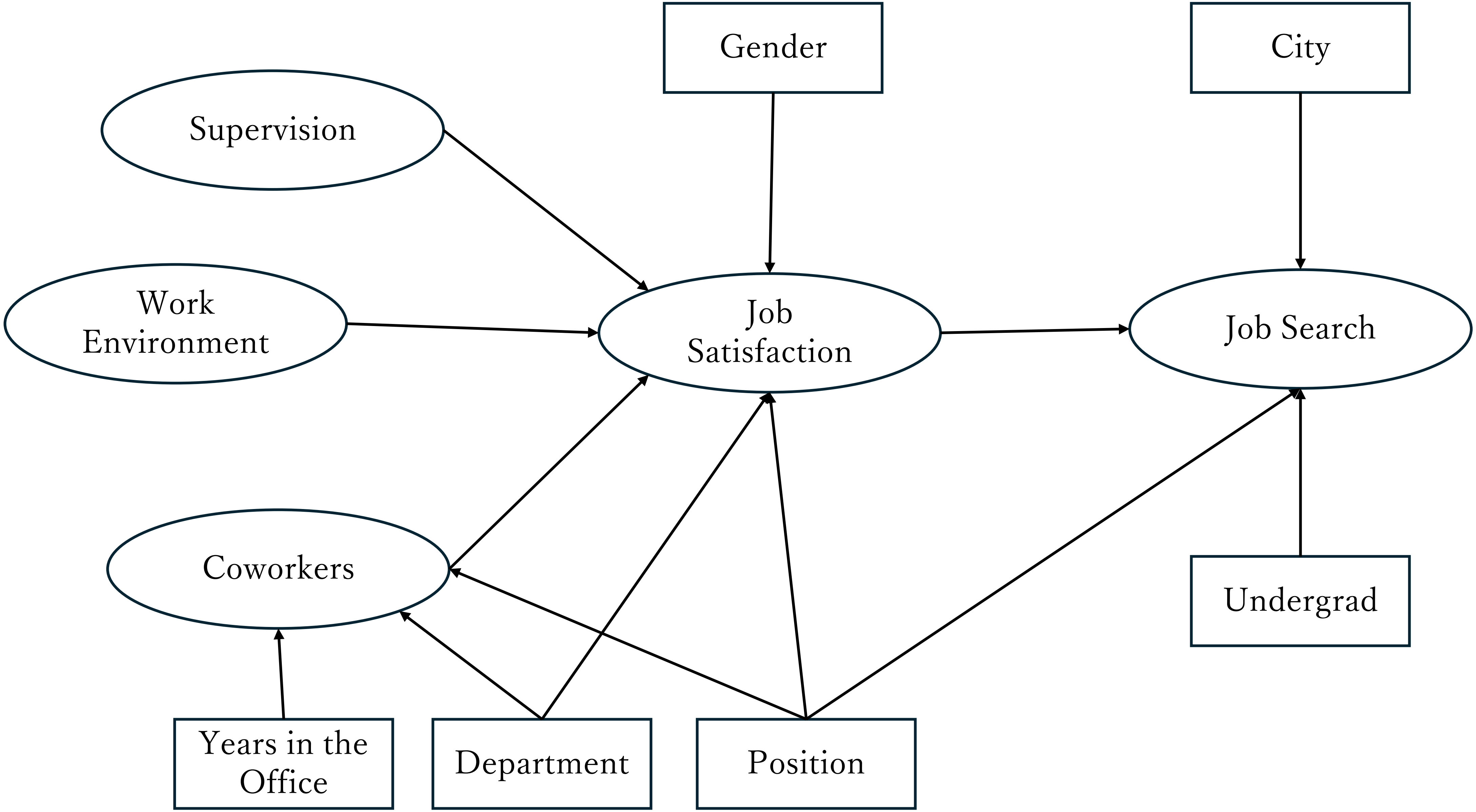
Adapted from Hair (page 621, Figure 9.5)



Adapted from Hair (page 621, Figure 9.5)



Adapted from Hair (page 621, Figure 9.5)



Adapted from Hair (page 621, Figure 9.5)

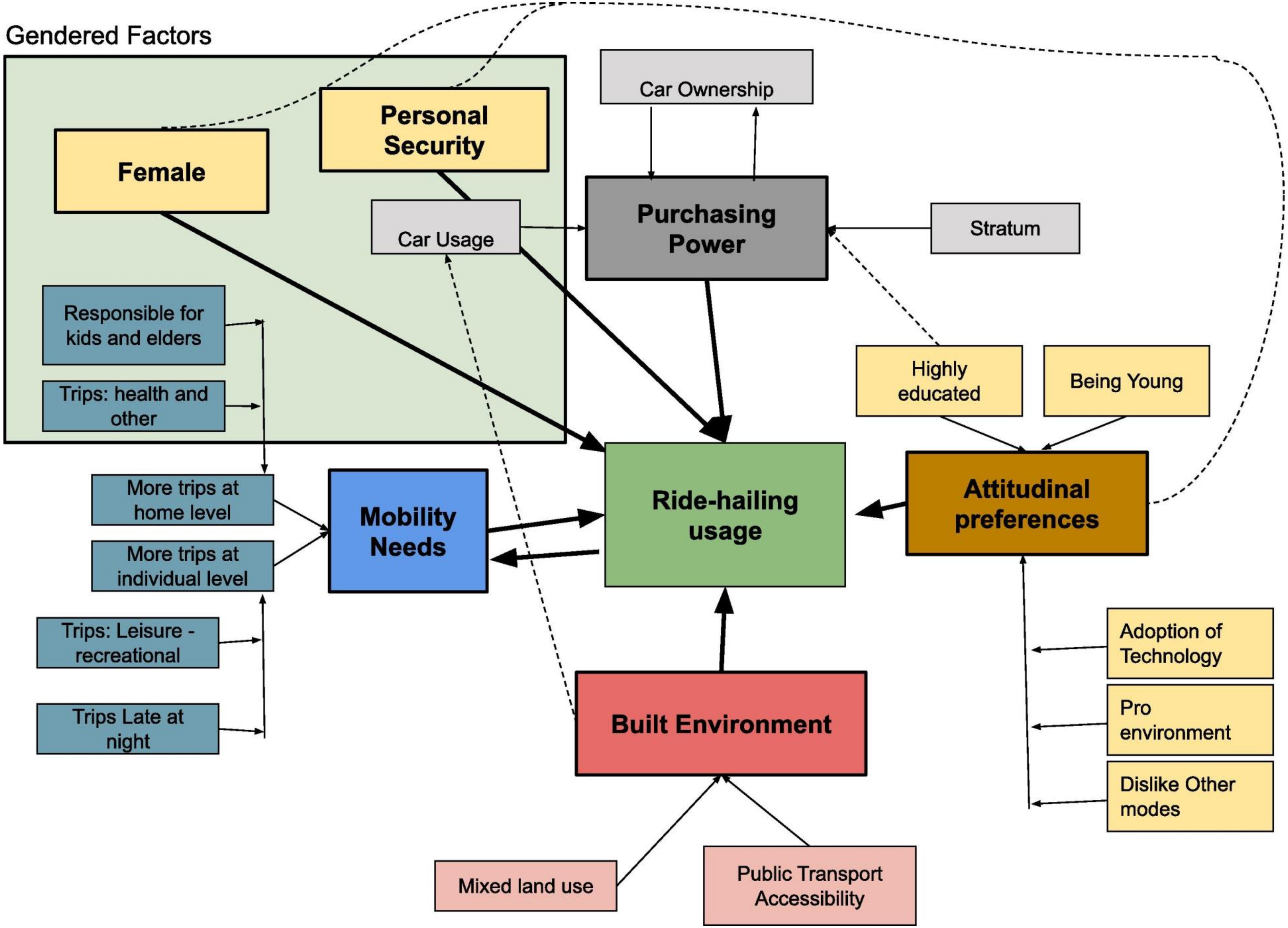
Structural Equation Models SEM

- It is a family of models
- Constructs (latent variables) are included
- Dependent variables in one regression can become independent variables for other regressions
- Theory-based (a model represents a theory)

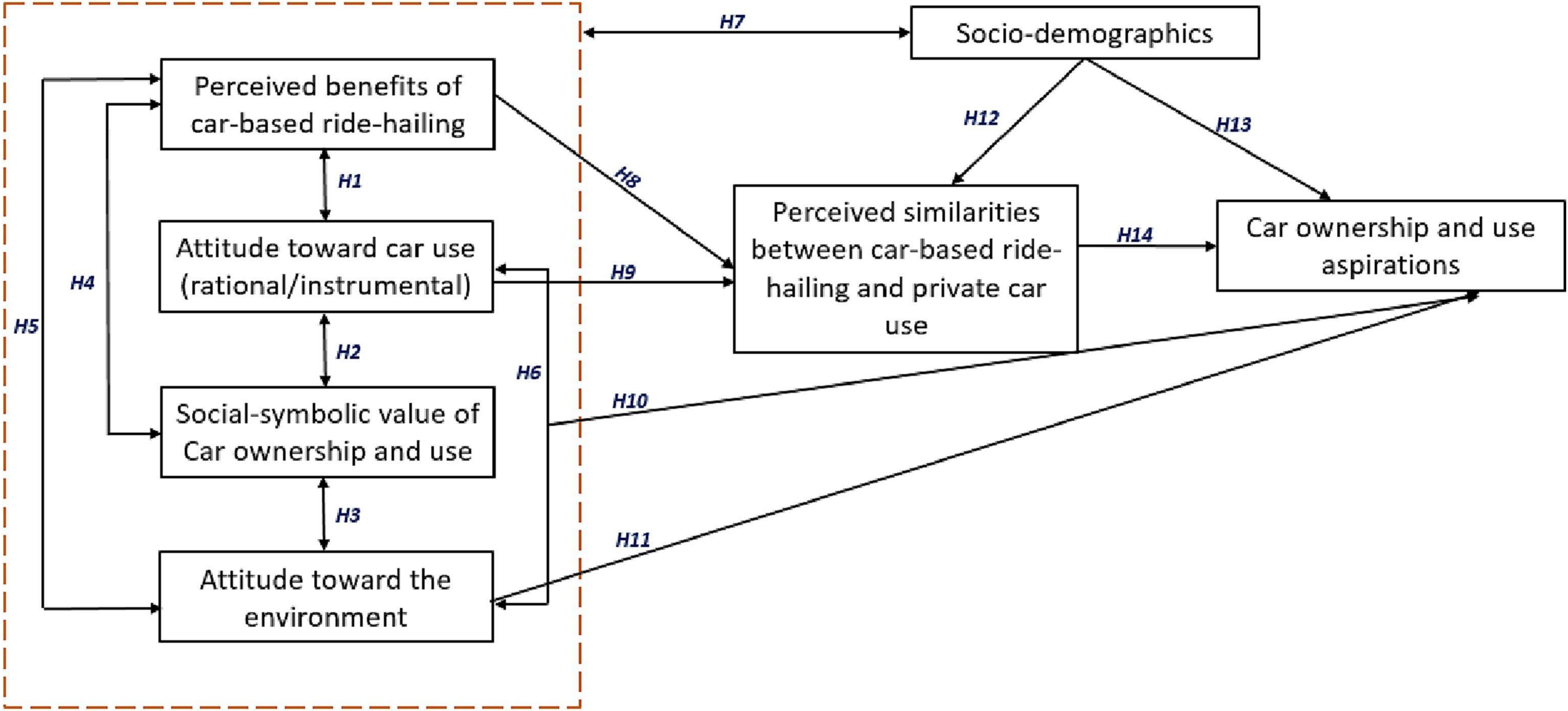
Theory

"**Theory** can be thought of as a systematic set of relationships providing a consistent and comprehensive *explanation* of phenomena. From this definition, we see that theory is not the exclusive domain of academia, but can be rooted in experience and practice obtained by observation of real-world behavior. A conventional model in SEM terminology consists of really two theories, the measurement model (representing how measured variables come together to represent constructs) and the structural model (showing how constructs are associated with each other)."

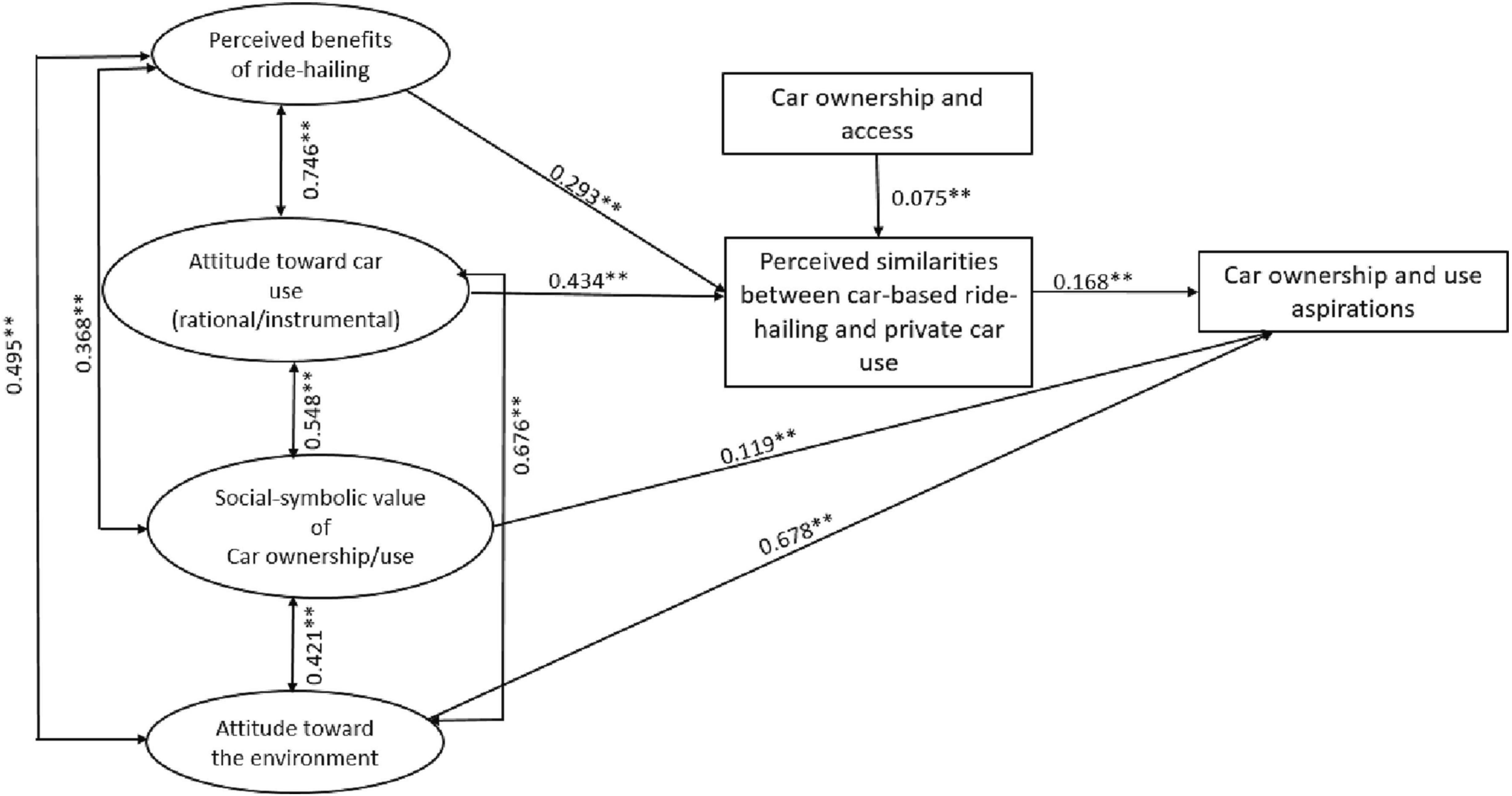
Hair (page, 610)



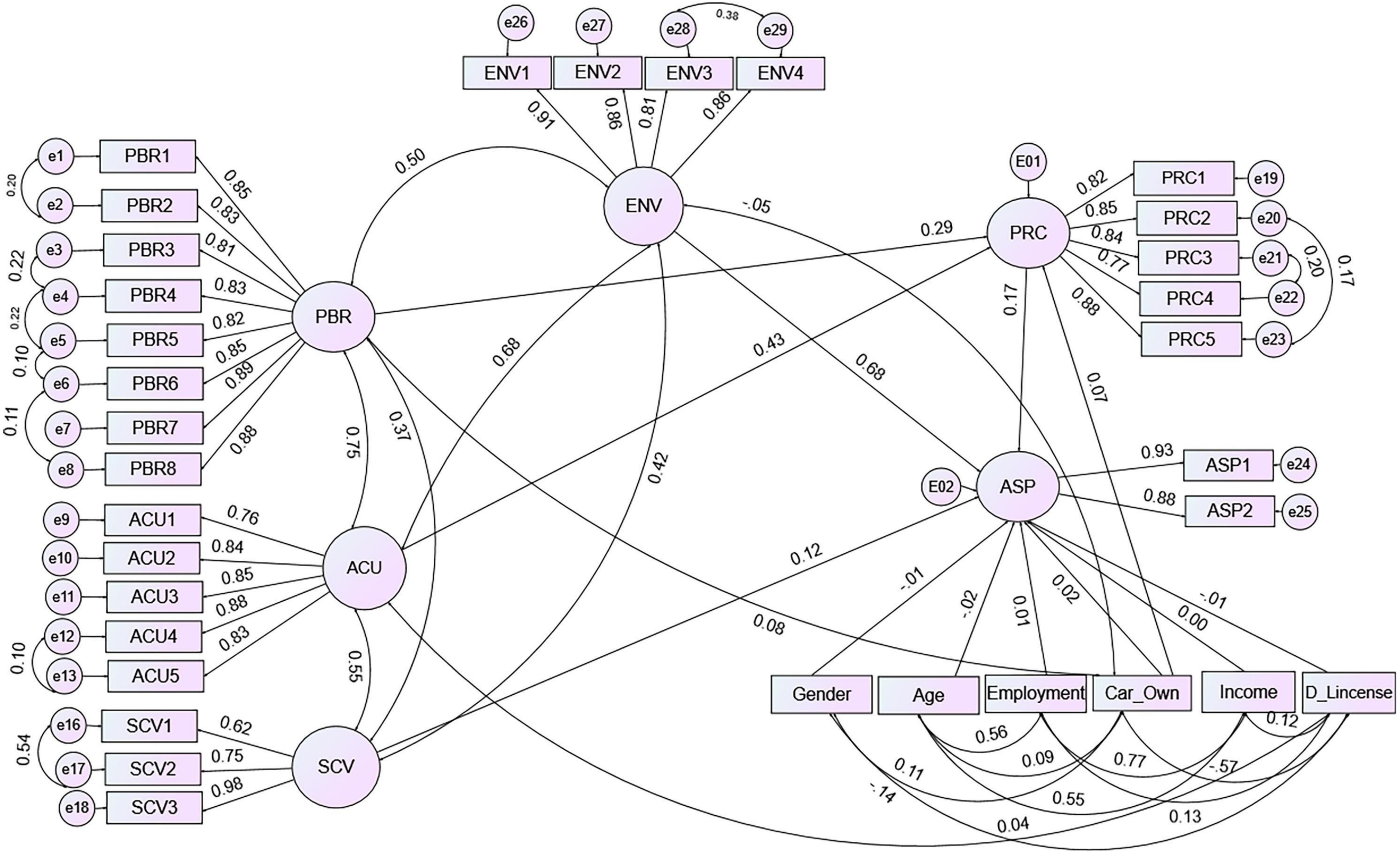
“Not my usual trip” Sabogal-Cardona et al., (2021)


H7
Socio-demographics
Perceived benefits of car-based ride-hailing
H12
H13
H8
H1
Attitude toward car use (rational/instrumental)
Perceived similarities between car-based ride-hailing and private car use
H14
Car ownership and use aspirations
H9
H4
H5
H2
H6
H10
Social-symbolic value of Car ownership and use
H3
H11
Attitude toward the environment

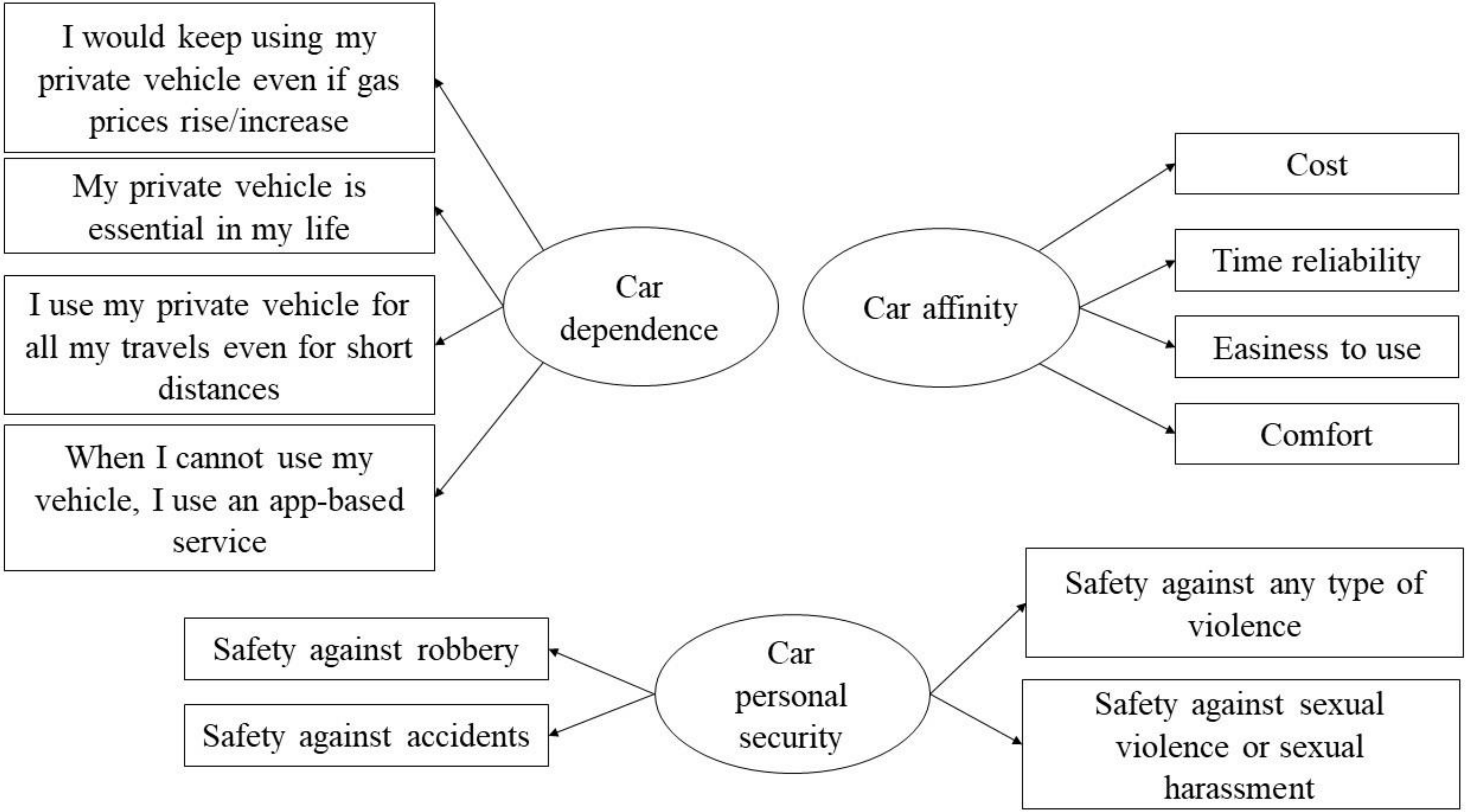
Source: Acheampong, R. A., Agyemang, E., & Yaw Asuah, A. (2023). Is ride-hailing a step closer to personal car use? Exploring associations between car-based ride-hailing and car ownership and use aspirations among young adults. *Travel Behaviour and Society*, *33*. https://doi.org/10.1016/j.tbs.2023.100614


Perceived benefits of ride-hailing
Car ownership and access
0.746**
0.293**
0.075**
Attitude toward car use (rational/instrumental)
0.434**
Perceived similarities between car-based ride-hailing and private car use
0.168**
Car ownership and use aspirations
0.495**
0.368**
0.548**
0.676**
0.119**
Social-symbolic value of Car ownership/use
0.678**
0.421**
Attitude toward the environment

Source: Acheampong, R. A., Agyemang, E., & Yaw Asuah, A. (2023). Is ride-hailing a step closer to personal car use? Exploring associations between car-based ride-hailing and car ownership and use aspirations among young adults. *Travel Behaviour and Society, 33*. https://doi.org/10.1016/j.tbs.2023.100614



Source: Acheampong, R. A., Agyemang, E., & Yaw Asuah, A. (2023). Is ride-hailing a step closer to personal car use? Exploring associations between car-based ride-hailing and car ownership and use aspirations among young adults. *Travel Behaviour and Society, 33*. https://doi.org/10.1016/j.tbs.2023.100614



Car dependence

I would keep using my private vehicle even if gas prices rise/increase

My private vehicle is essential in my life

I use my private vehicle for all my travels even for short distances

When I cannot use my vehicle, I use an app-based service

Car affinity

Cost

Time reliability

Easiness to use

Comfort

Car personal security

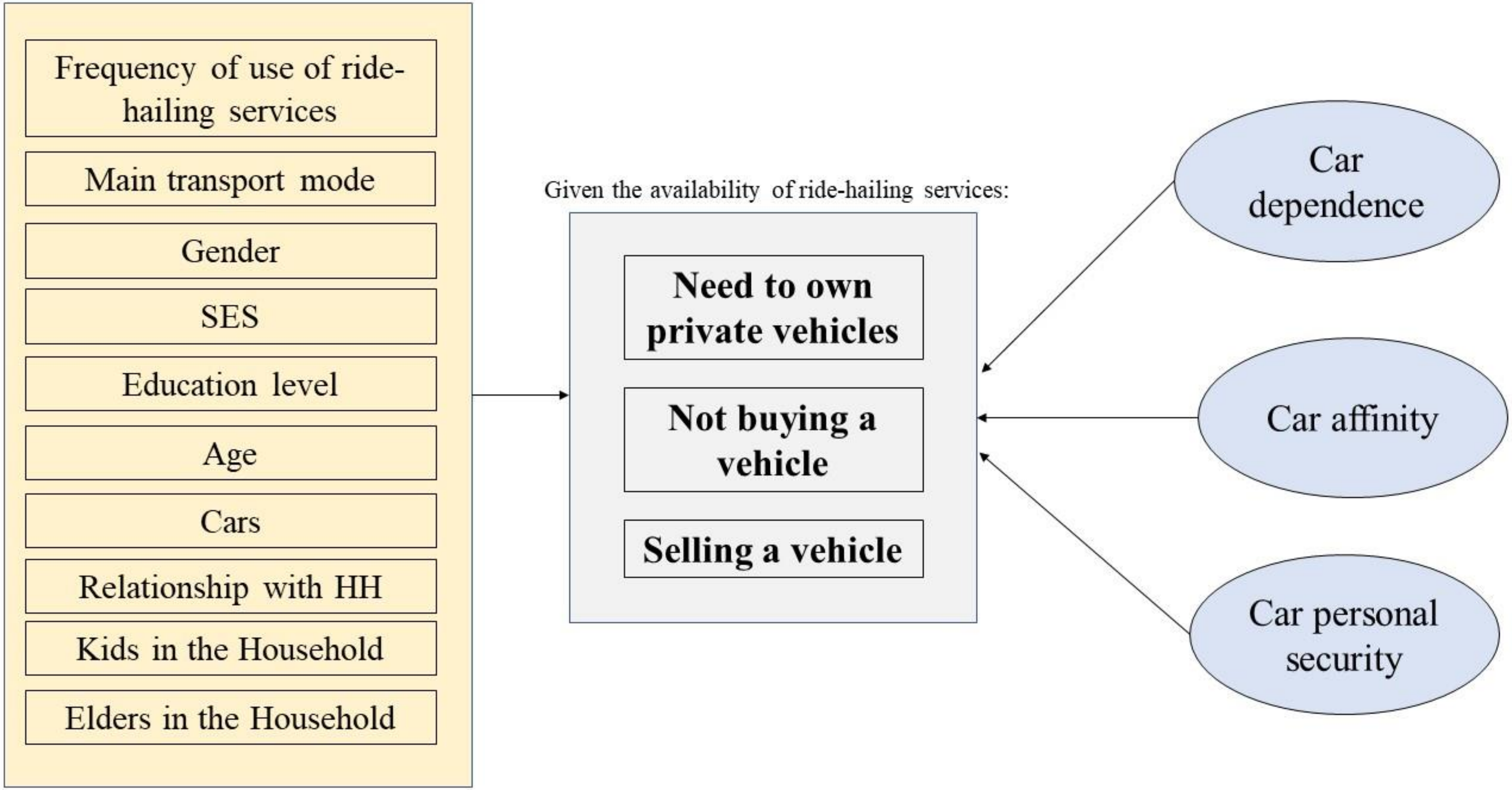
Safety against robbery

Safety against accidents

Safety against any type of violence

Safety against sexual violence or sexual harassment

Sabogal-Cardona, O., Oviedo, D., & Scholl, L. (2023). Can Ride-Hailing Services Reduce Car Ownership?: Lessons from 3 Latin-American Cities.



Frequency of use of ride-hailing services

Main transport mode

Gender

SES

Education level

Age

Cars

Relationship with HH

Kids in the Household

Elders in the Household

Given the availability of ride-hailing services:

Need to own private vehicles

Not buying a vehicle

Selling a vehicle

Car dependence

Car affinity

Car personal security

Sabogal-Cardona, O., Oviedo, D., & Scholl, L. (2023). Can Ride-Hailing Services Reduce Car Ownership?: Lessons from 3 Latin-American Cities.

Theory

- Theory on the measurement model
- Theory on the structural model (regression paths)
- Theory-based approach is necessary: the researchers specify the SEM before estimation. This does not happen with other methods
- Any insight regarding modern data science?

Photo by Rohan Makhecha on Unsplash

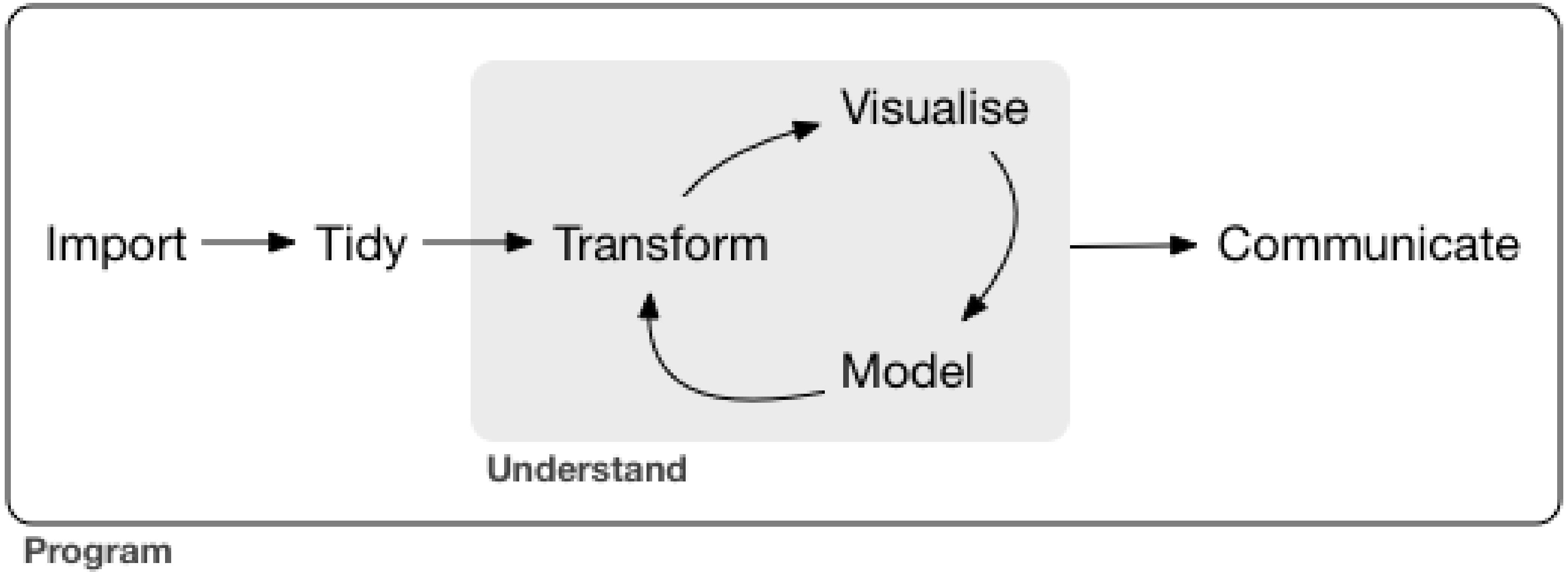
Data science

All the new family: machine learning, data mining, cognitive knowledge (?), business intelligence, analytics, deep learning, Artificial Intelligence.

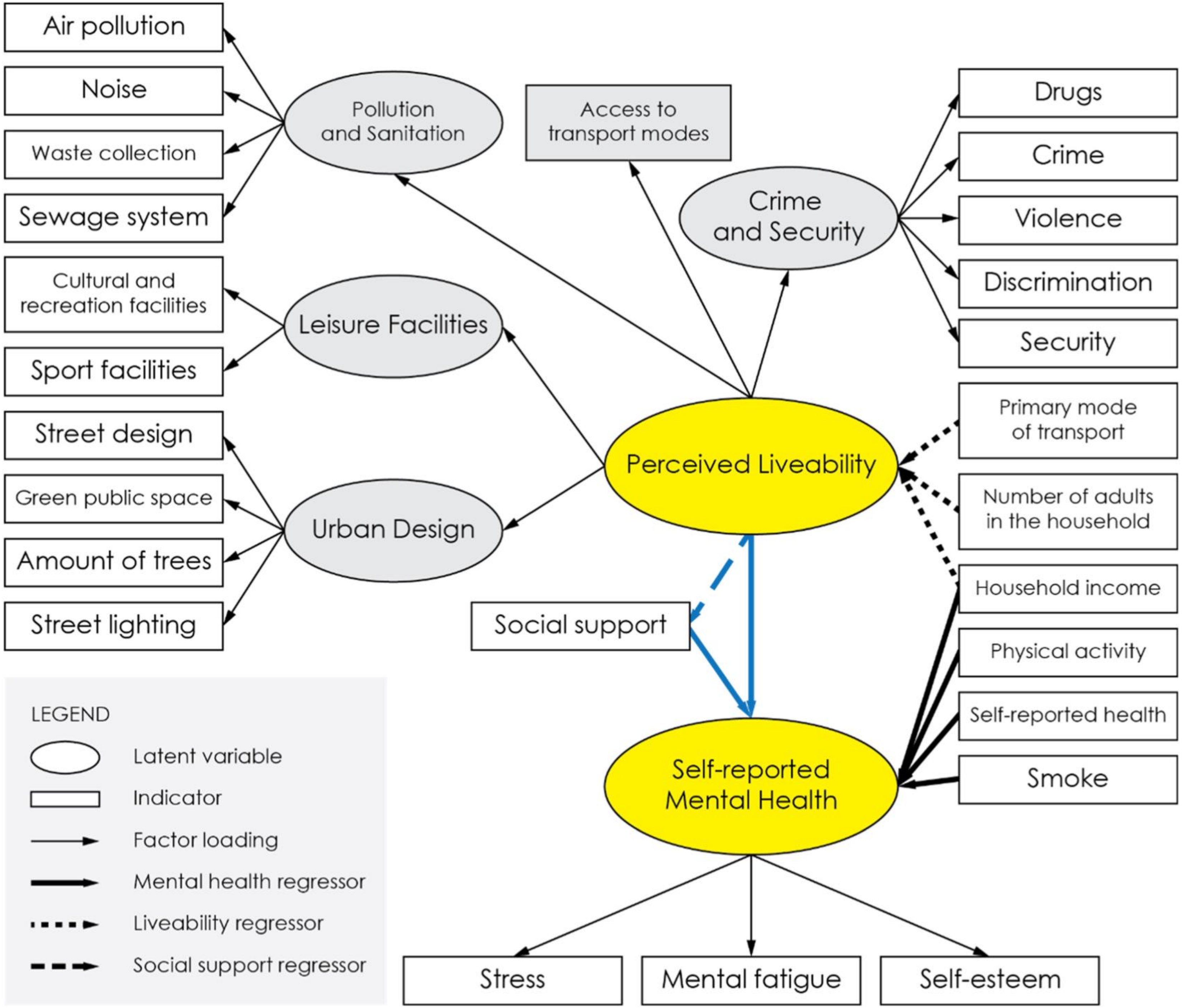
Data science:

- Computational capabilities +
- Strong statistical background +
- Domain knowledge +
- Communication skills

It should be about knowledge production informed by data



Wickham and Grolemund (2017)



Oviedo, D., Sabogal, O., Duarte, N. V., & Chong, A. Z. (2022). Perceived liveability, transport, and mental health: A story of overlying inequalities. *Journal of Transport & Health, 27*, 101513.

Table 4
SEM results (regression paths).

	Estimate	Error	P Value	Standardised estimate
Perceived Liveability				
Adults in the home	−0.046	0.022	0.035	−0.149
Income level				
Low	ref	ref	ref	ref
Medium	0.194	0.077	0.012	0.204
High	0.437	0.119	0	0.326
Main mode of transport				
Car	ref	ref	ref	ref
MIO	−0.253	0.069	0	−0.291
Cycling	−0.3	0.091	0.001	−0.23
Multimodal	−0.162	0.074	0.028	−0.173
Social support				
Perceived liveability	0.31	0.103	0.003	0.21
Self-reported mental health				
Liveability	0.169	0.079	0.033	0.154
Smoker				
No	ref	ref	ref	ref
Yes	−0.206	0.069	0.003	−0.172
Health (Self-assessment)	0.233	0.053	0	0.292
Physical Activity	0.12	0.033	0	0.222
Income level				
Low	ref	ref	ref	ref
Medium	0.249	0.076	0.001	0.238
High	0.079	0.112	0.479	0.054
Social support	0.239	0.042	0	0.321

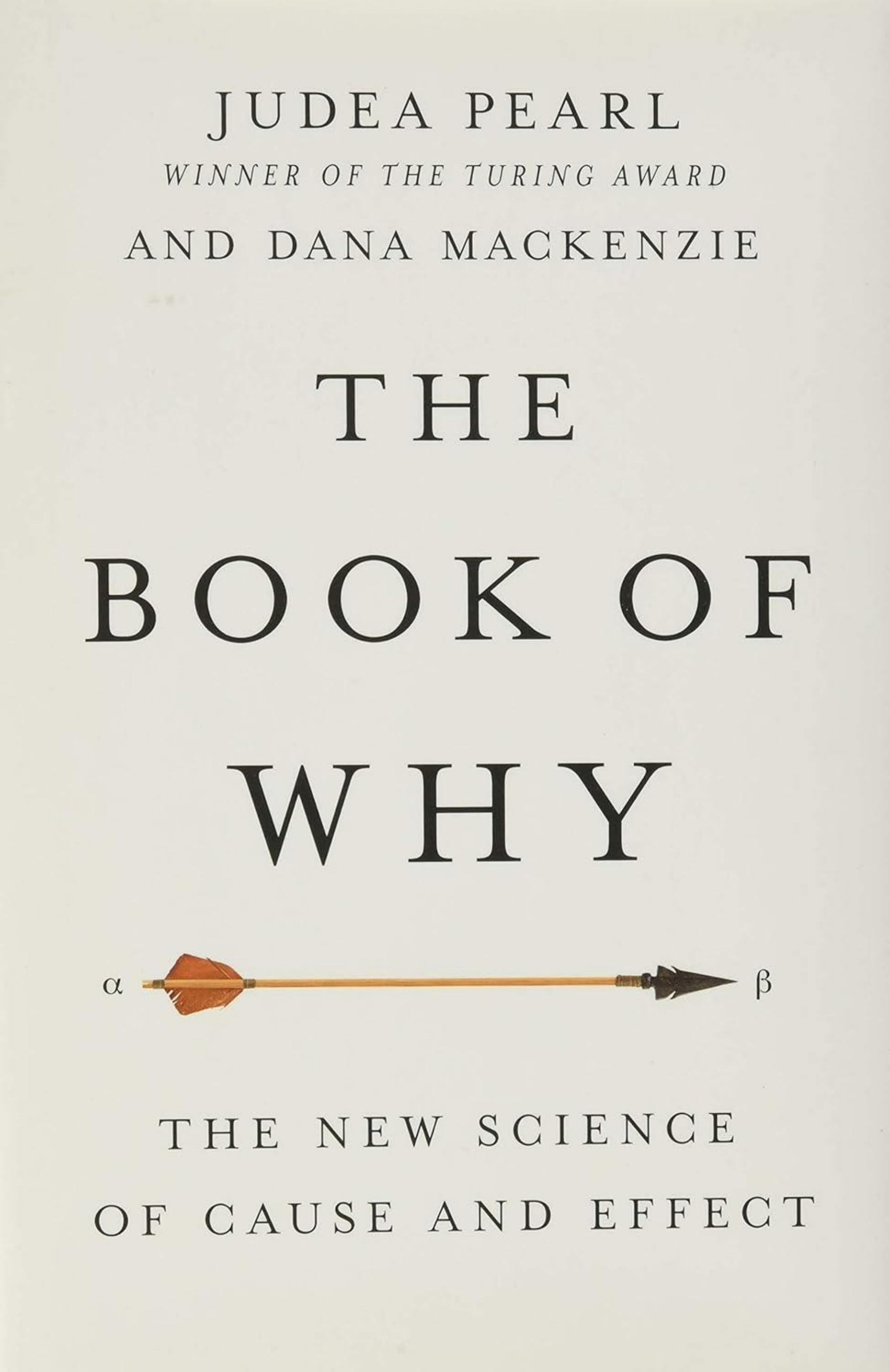
srmr = 0.065; rmsea = 0.032; TLI = 0.982; CFI = 0.968.

Causality

- Cause-and-effect relationship
- Difficult in the context of non-longitudinal data or non-experimental designs (design of experiments, discrete choice models, impact evaluation, even ANOVA/MANOVA)

Causality (in non-experimental settings)

- Covariation (a strong and significant estimate)
- Sequence (temporal): guided by theory on cross-sectional studies.
- Nonspurious covariance
- Theoretical support



JUDEA PEARL

WINNER OF THE TURING AWARD

AND DANA MACKENZIE

THE BOOK OF WHY

THE NEW SCIENCE OF CAUSE AND EFFECT

Model development strategy

- Confirmatory modeling strategy
- Competing models strategy (competing theories)
- Model development strategy: you start with a framework, and you improve it. Model respecification MUST ALWAYS come with theoretical support.

Model Fit

- We still want to specify a model with model-implied variance covariance matrix that is close-enough to the sample variance-covariance matrix.
- How well does the theoretical model fit the observed data?

$$\min(\Sigma - S)$$

Home Episodes Playlists Syllabus Merch About Ask a Q! Subscribe

S4E10 Test Driving Model Identification

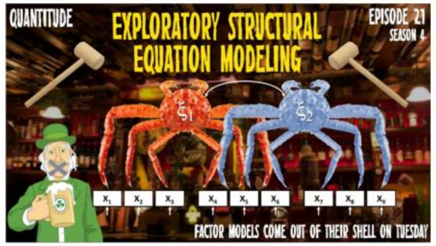
November 29, 2022

https://quantitudepod.org/s4e10-identification/

Home Episodes Playlists Syllabus Merch About Ask a Q! Subscribe

S4E21 Exploratory Structural Equation Modeling

March 14, 2023


QUANTITUDE
EXPLORATORY STRUCTURAL
EQUATION MODELING
EPISODE 21
SEASON 4
ξ1
ξ2
X1 X2 X3 X4 X5 X6 X7 X8 X9
FACTOR MODELS COME OUT OF THEIR SHELL ON TUESDAY

https://quantitudepod.org/s4e21-esem/

From QuantFish:

What is Structual Equation Modeling:

https://www.youtube.com/watch?v=OabNYoXsu2M&list=PL-kVjeOVYChqDCJJVydP4OS8J5Y94q6mM&index=2

SEM Advantages and Limitations:

https://www.youtube.com/watch?v=1GDEabX98xc&list=PL-kVjeOVYChqDCJJVydP4OS8J5Y94q6mM&index=4

4 reasons why your SEM may fail:

https://www.youtube.com/watch?v=pTIgS6pbMjM&list=PL-kVjeOVYChqDCJJVydP4OS8J5Y94q6mM&index=5

Go to the tutorial!

Thank you!

Orlando Sabogal-Cardona
@Antonio Sabogal
orlando.sabogal.20@ucl.ac.uk