# Causal inference in the interface ML / OR

Claus Gwiggner

Capgemini invent

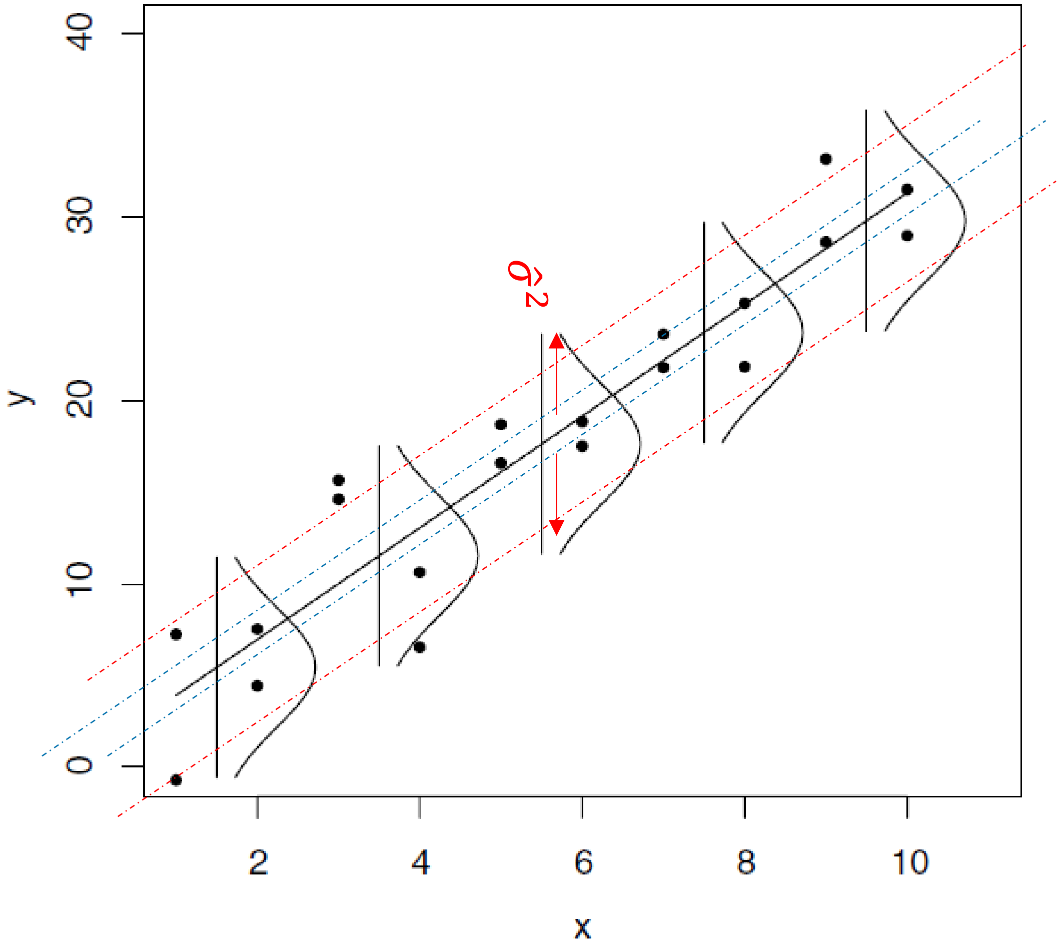
Company confidential

# Agenda



Public © Capgemini 2025. All rights reserved | 

# Classical Regression re-visited




Aim: Forecast of Y with data from X

- Regression function: $E[Y \mid X = x] \coloneqq \hat{\mu}(x) = \hat{\alpha}_0 + \hat{\alpha}_1 x$
- Probabilistic Forecast: $Y_{X=x} \sim \mathcal{N}(\hat{\alpha}_0 + \hat{\alpha}_1 x, \hat{\sigma}^2)$
  - Parameter uncertainty: $V_\alpha(\hat{\mu}(x)) = V(x^T \hat{\alpha})$
  - Model uncertainty: $V_m(\hat{\mu}(x)) = V_\alpha(\hat{\mu}(x)) + \hat{\sigma}^2$
- Statistical inference: $\mathcal{H}_0 : \alpha_1 = 0$
- More relevant for discrete variables
  - Poisson regression
    - $E[Y \mid X = x] \coloneqq \lambda(x) = e^{\hat{\alpha}_0 + \hat{\alpha}_1 x}$
    - $Y_{X=x} \sim \mathcal{P}(\lambda)$

Public © Capgemini 2025. All rights reserved | 

# Case Study Introduction

**Data Description**

Dataset with 5 variables and 700 rows. Each row represents a purchase. Attributes per row:

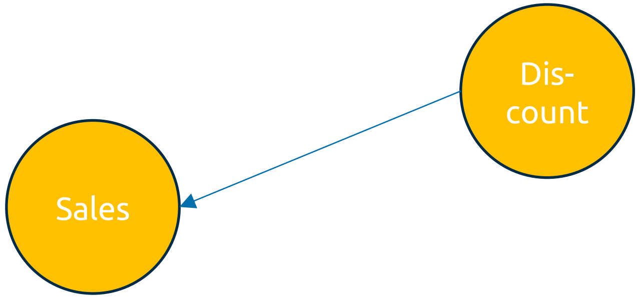
- **Discount**: Indicates if a €1 discount was granted (Binary / Categorical)
- **Visits**: Number of visits to the webshop (Numeric)
- **Sales**: Purchase amount (Numeric)
- **Sales per Visit**: Sales / Visits (Numeric)
- **Is_Loyal**: Indicates loyalty if Discount was applied or if Sales were very high (Binary / Categorical)

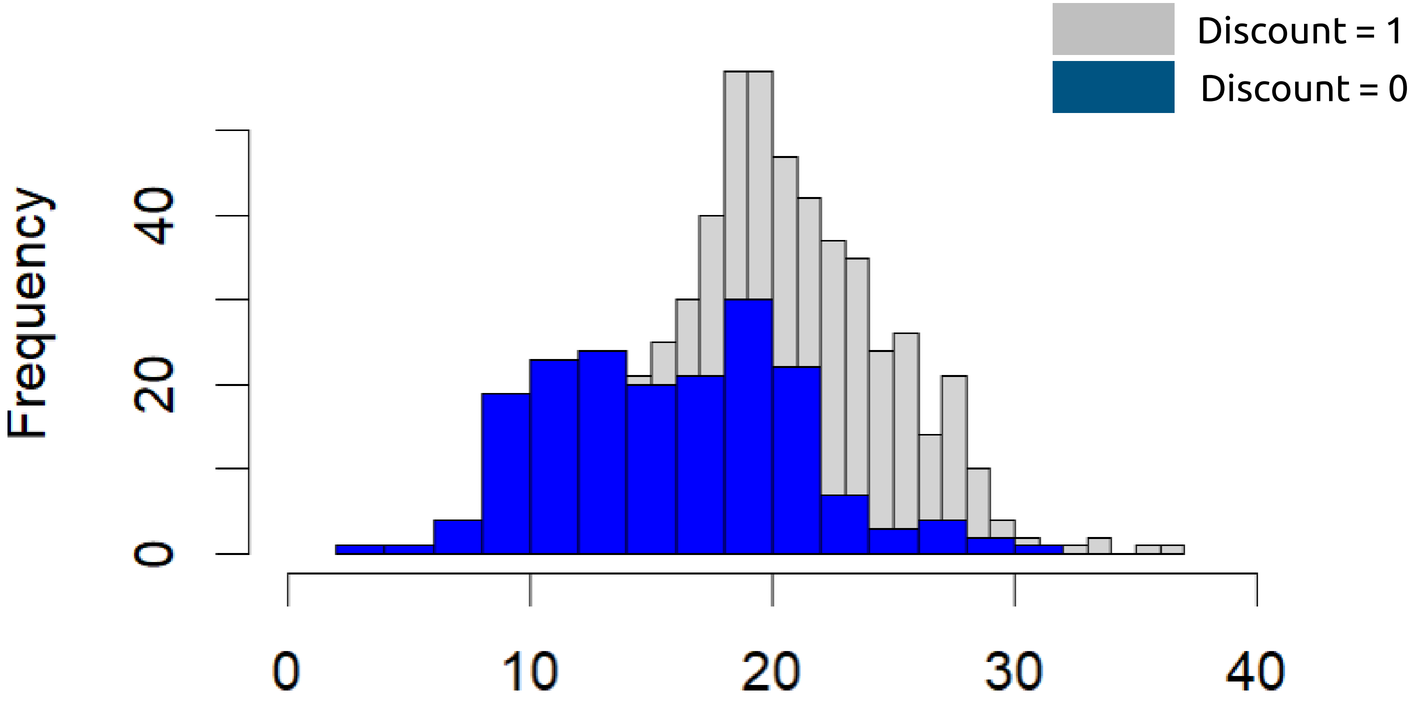
**Hypothesis**
"The average purchase amount increases when the customer received a discount."

Source: https://github.com/juanitorduz

Public © Capgemini 2025. All rights reserved | 

# Causal Analysis (I/III)

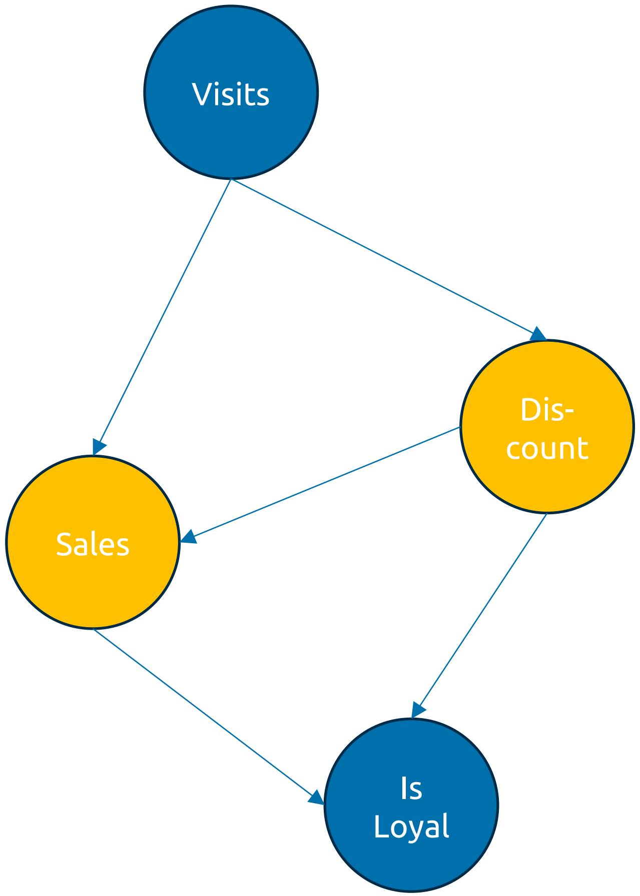



```
> # Compute the difference in means
> #attach(data)
> mean(sales[discount==1])-mean(sales[discount==0])
[1] 4.79915

mean(sales_per_visit[discount ==0])-mean(sales_per_visit
[discount ==1])
[1] -0.07705032
```

Public © Capgemini 2025. All rights reserved |

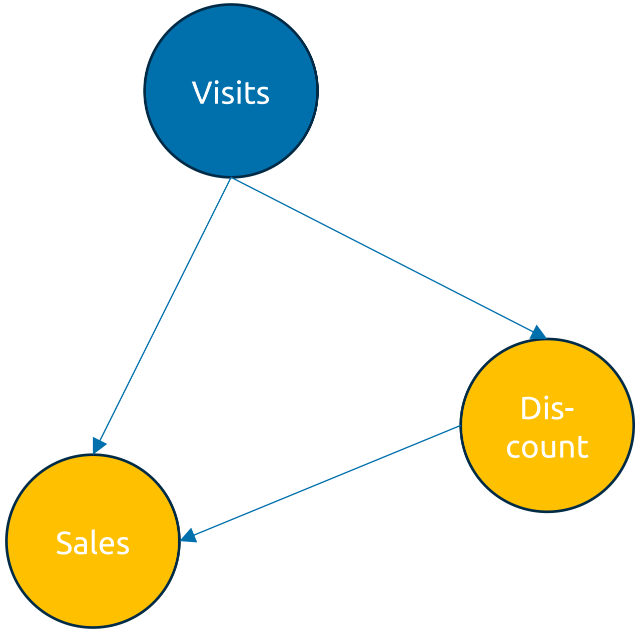
# Causal Analysis (II/III)




```
> # Full model
> m_full<-lm(sales~visits+discount+is_loyal-1,data=data)
> summary(m_full)

Call:
lm(formula = sales ~ visits + discount + is_loyal - 1, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5188 -1.2831  0.0545  1.3581  6.1058 

Coefficients:
         Estimate Std. Error t value Pr(>|t|)    
visits   0.840656   0.007082 118.703   <2e-16 ***
discount 0.097974   0.728280   0.135   0.8930    
is_loyal 1.933808   0.753418   2.567   0.0105 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.907 on 697 degrees of freedom
Multiple R-squared:  0.991,	Adjusted R-squared:  0.9909 
F-statistic: 2.545e+04 on 3 and 697 DF,  p-value: < 2.2e-16
```

Public © Capgemini 2025. All rights reserved | 

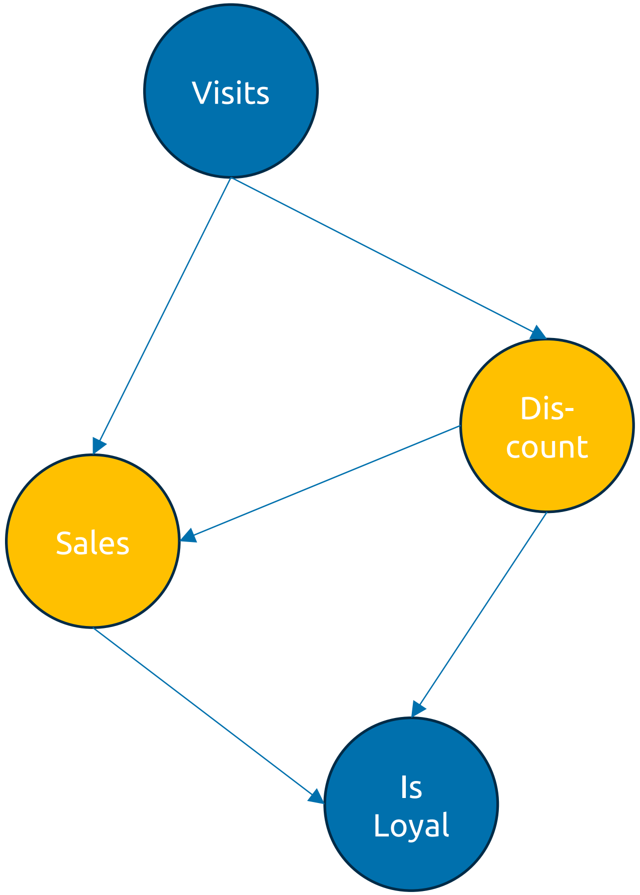
# Causal Analysis (III/III)




```
> m_causal <- lm(sales~visits+discount-1,data=data)
> summary(m_causal)

Call:
lm(formula = sales ~ visits + discount - 1, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5392 -1.2576  0.0537  1.3508  6.1330 

Coefficients:
         Estimate Std. Error t value Pr(>|t|)    
visits   0.845953   0.006802   124.4   <2e-16 ***
discount 1.914486   0.172519    11.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.914 on 698 degrees of freedom
Multiple R-squared:  0.9909,	Adjusted R-squared:  0.9908 
F-statistic: 3.786e+04 on 2 and 698 DF,  p-value: < 2.2e-16
```

Public © Capgemini 2025. All rights reserved | 

# Causal Diagram (DAG)




Confounder Variable (influences discount & sales)
→ Keep in model
→ „Control for it“

## Background: „Back-door paths“ (Pearl, 1993)

„A back-door path is a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable”

Collider Variable (no logical impact on sales)
→ Leave it out
→ „Reverse causality“

“A variable is a collider along a particular path if it has two arrows pointing directly at it.“

Public © Capgemini 2025. All rights reserved |

# Conclusions and Discussion

AI / ML

- Empirical models

OR

- Mechanistic models

- Scenario-based optimization
- Stochastic optimization

AI Forecasting

- “Understand the data”

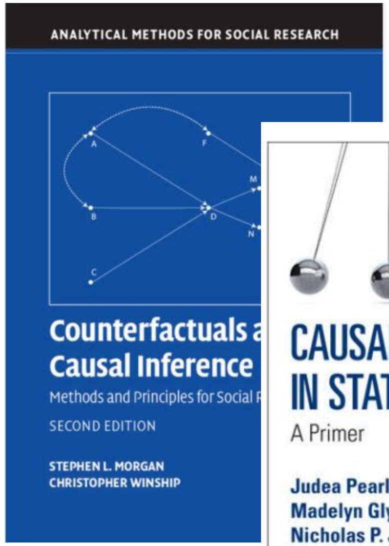
 © Capgemini 2025. All rights reserved 

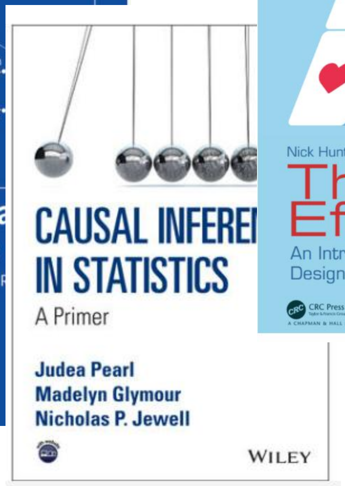
# References

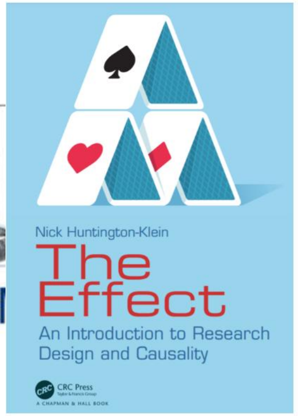

J. Pearl, D. Mackanzie, The Book of Why: The New Science of Cause and Effect, Basic Books, 2018

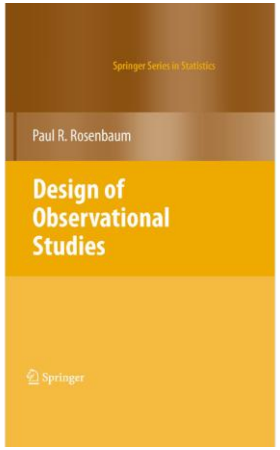
Williamson et al., Introduction to causal diagrams for confounder selection. *Respirology* (2014) **19**, 303–311

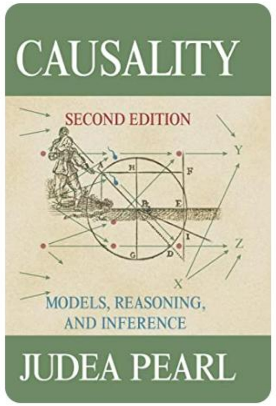
Hernán et al., Causal Knowledge as a Prerequisite for Confounding Evaluation:
An Application to Birth Defects Epidemiology. American Journal of Epidemiology (2002), Vol. 155, No. 2

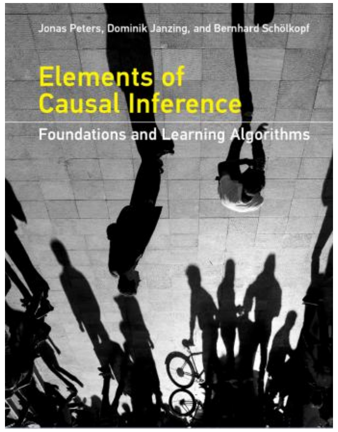








Public © Capgemini 2025. All rights reserved | 