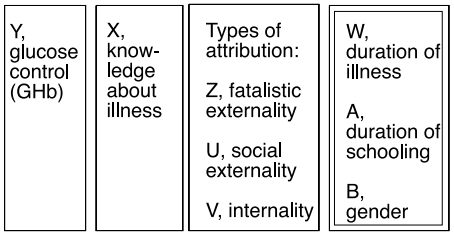
Deriving graphs from data

- Two examples
- General strategy
- Checking for nonlinear relations

Slide 1

What determines the well-being of diabetic patients?



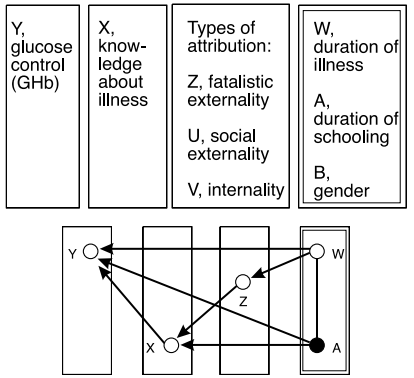
primary response | intermediate variables | purely explanatory v.

6 quantitative variables: X, Y, Z, U, V, W, 2 binary variables: A, B;

Cross-sectional study of 68 diabetic patients; 1990 Mainz

Slide 2

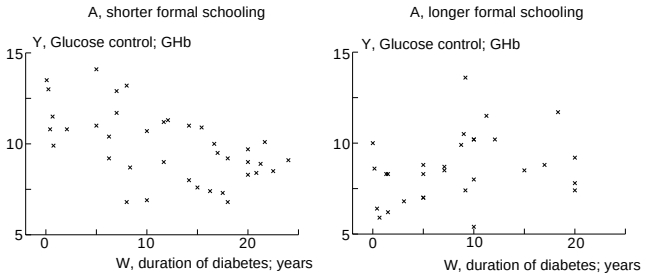
The graphical summary after analysis



$$Y: \ X + A * W, \quad X: \ Z + A, \quad Z: \ W.$$

Slide 3

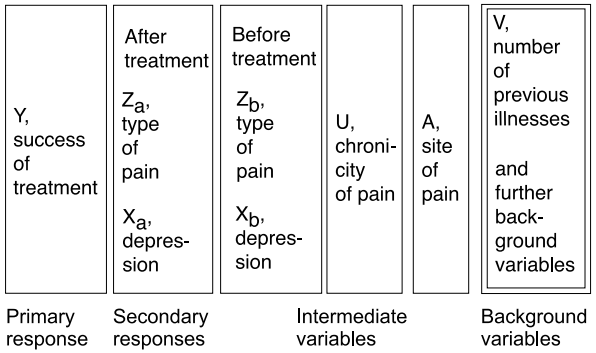
The interactive effect $A * W$



Slide 4

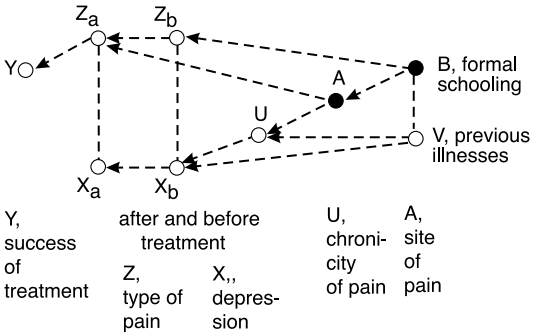
What determines successful treatment of chronic pain patients?

– a first ordering of the variables



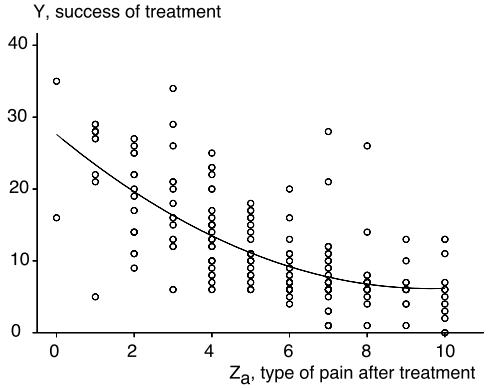
201 patients, before and after a three-week stationary treatment, 1996; Mainz

Slide 5



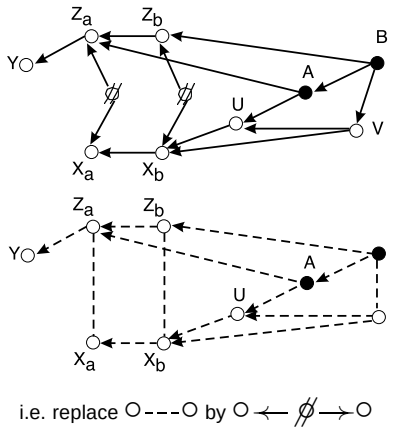
Slide 6

A nonlinear dependence of success of treatment on intensity of pain



Slide 7

A fully directed graph which could generate the joint response graph



Slide 8

General concepts

Types of observation on individuals

- responses
- intermediate outcomes
- treatments, quasi-treatments, risk factors; intrinsic variables
- baseline, background or context variables

Types of study

- randomized experiments
- cross-sectional studies
- longitudinal studies
  prospective (cohort), retrospective

Slide 9

Steps in the planning of analyses

- Understand how data were collected
- Get a first ordering of the variables relevant for the research question
  - Arrange variables in boxes for
    - dependencies (arrows) between boxes
    - variables on equal footing (lines) within boxes
  - Variables in one box are to be considered conditionally on all variables to the right (and sometimes also on variables in same box)
- Remember objectives
  - simplification via conditional independencies
  - suggestion of potential generating processes
  - tracing of paths of possible development

Slide 10

Steps in analysis

- Compare the observed one-way and two-way distributions for complete and for incomplete data
- Check for nonlinear and interactive effects

- Fit sequences of univariate conditional distributions
- Summarize results with the help of an independence graph

Slide 11

Checks for nonlinearities

Only quantitative measurements

- For all bivariate regressions compute t statistics for squared terms in:

$$Y \text{ on } (X - \bar{X}) \text{ and } (X - \bar{X})^2$$

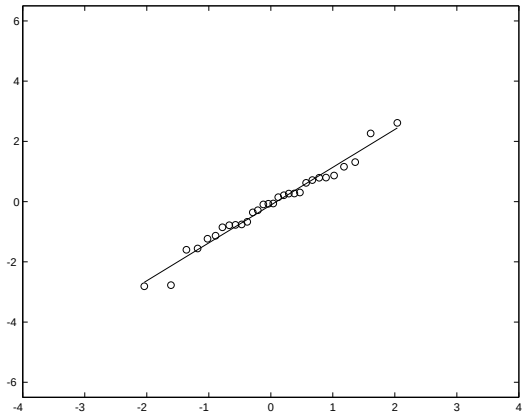
- For all trivariate regressions compute t statistics for cross-product terms in:

$$Y \text{ on } (X - \bar{X}), (V - \bar{V}) \text{ and } (X - \bar{X}) \times (V - \bar{V})$$

Plot t-values against expected normal order statistics
(see Cox and Wermuth, 1994)

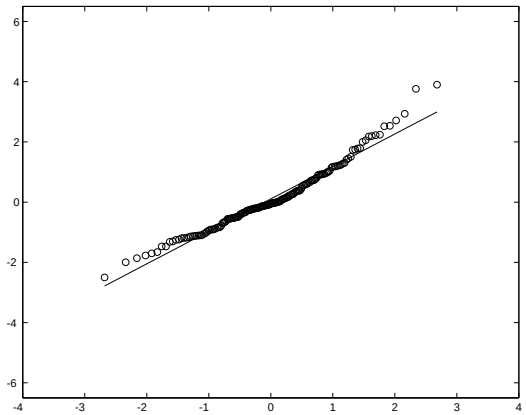
Slide 12

No important quadratic effects in the diabetes data



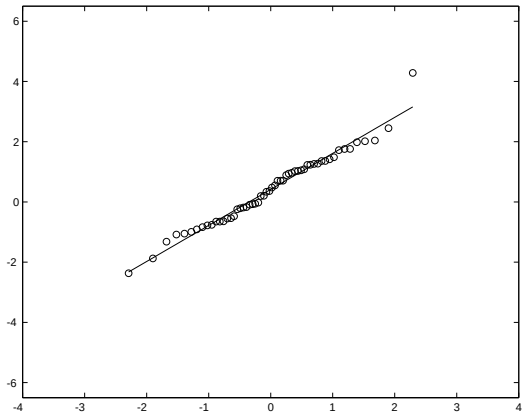
Slide 13

Some important interactive effects in the diabetes data?



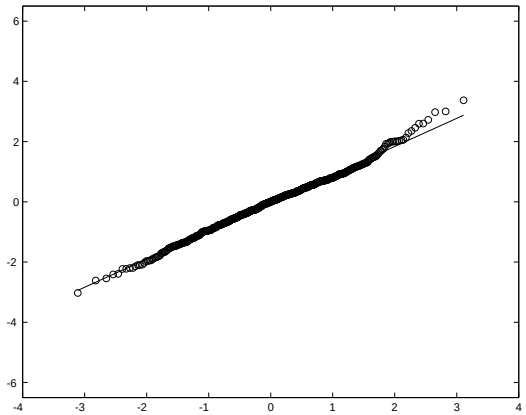
Slide 14

One important quadratic effect in the chronic pain data



Slide 15

No important interactive effects in the chronic pain data



Slide 16