1 Categorical Variable and 1 Continuous Variable

Continuous → Categorical

For continuous variables we need to select a specific interval in the histogram which we want to observe in the linked spineplot

In the example we look at:

P(| "Price per m²" $\geq 10 |$ "No. of Rooms" = 3) = 20.7%Note:

the dependent variable

is always set up to look at the conditional distribution



1 Categorical Variable and 1 Continuous Variable

Continuous → Categorical

Often it is not at all clear which subgroup of the continuous variable we might want to look at

Brushing allows for a very flexible and fast way to observe very many intervals in succession just like in an animation



64

Two Continuous Variables

- Anscomb's Quartet (old hat – I know)
- The learning should be always on our minds!
- Tukey: "there is no excuse for failing to plot and look"
- For 2-dim continuous data, there is nothing more efficient to use than a scatterplot; though it might need some modifications



Two Continuous Variables: Scatterplot Smoothers

- If there is a functional relationship between the two variables, a scatterplot smoother can help to visualize the relationship
- If we do not know the exact relationship (linear, quadratic, logarithmic ...) an exploratory estimate is the best we can do
- The general trade-off is between bias ^g and variance
- There are many rules for an optimal parameter choice, but interactivity is often the key for an optimal insight





66

Two Continuous Variables: Degree of Smoothness

- Example Birthweight Data:
 - Smoother: Lowess
 (locally weighted regression, with quadratic base smoother)
 - Parameter: Span (The degree of locality, either defined by an interval or a proportion of the data)
 - Results: left undersmoothed center – suitable smoother right – oversmoothed



67

Two Continuous Variables: Detour

- We should always keep in mind that we usually fit something like y = f(x) + ε
- If it is not clear which variable is dependent and which is not, the results may differ very much
- Principal curves minimize the orthogonal distances towards the smoother, and thus may look quite different



Two Continuous Variables: Density Estimation

- (parametric) 2-dim density estimation is very expensive and often hard to motivate
- With alpha blending (aka alpha transparency) we get an exploratory density estimate
- Areas of high density will be darker, areas of lower density will be lighter
- The two parameters to choose are diameter and base transparency



Two Continuous Variables: Density Estimation

- Example: Italian Olive Oils
- Variation of the parameters size and α-level
- Size 3, 5, 7 points (left to right)
- α-level 0.08, 0.5, 0.92
 (bottom to top)
- Each version of the plot has its own feature which is unique to this plot

