

ビジネスデータ解析特論はビジネス、特に市場調査、製品の開発・生産、組織マネジメント等に関するデータを取りあげ、統計的手法を用いた解析法について学習します。講義構成は演習・討論を繰り返しながらビジネスデータの特徴、統計的手法を適用すること当たりの注意点を学びます。取り上げる具体的なデータとしては、

- 1) 市場販売データ
 - 2) 製品品質の重視度を調べたデータ
 - 3) 試作品の感性評価データ
 - 4) 製品の設計パラメータと性能特性によるデータ
 - 5) 生産条件と製品特性の関係に関するデータ
- があります。これらのデータに対して統計用ソフトを用いて解析を行います。学習することは
- i) ヒストグラム、散布図、層別や欠損値の扱い
 - ii) モデル選択、変数選択の方法
 - iii) 時系列モデル、ARMAモデルの取り扱い
 - iv) 主成分分析、因子分析、クラスター分析の方法等
- があります。

次に講義中に用いたデータを例に挙げます。このデータは『シリコン結晶の精製工程における品質解析を行うもの』です。右の三つの散布図は縦軸：(シリコン単結晶の直径)、横軸：(精製の際の温度)としてあります。図1のように全てのデータをそのままプロットすると何も特徴がないように見えます。次に、データをシリコンを精製する加熱炉1・加熱炉2ごとのグループに分けて考えます。すると、図2ではデータが左側に、図3では右側に集まっていることがわかります。グループごとに分けてプロットをする層別散布図を用いた方が各データの持っている特徴がわかりやすくなります。層別化することによって、各グループに含まれるデータの特徴をおさえたい解析を行うことができ、より正確な解析ができます。このことを念頭において、シリコン単結晶の直径を目的変数とした重回帰を考えます。その際に用いるのが変数選択です。変数選択を行うことによって、より良い回帰式が得られ、目的変数についての様々な解析を行うことができます。

他にも、主成分分析による次元縮小、時系列による予測、クラスター分析による分類なども扱います。

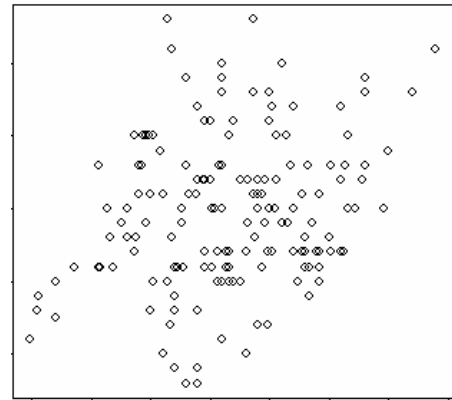


図1：全データの散布図

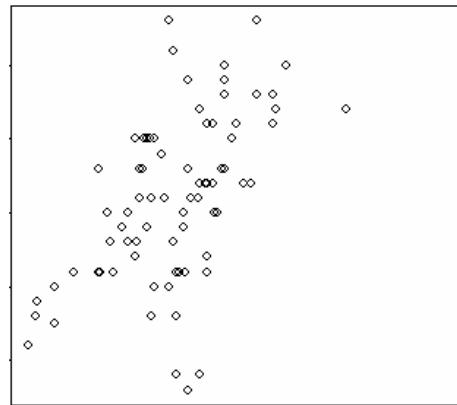


図2：加熱炉1に含まれるデータの散布図

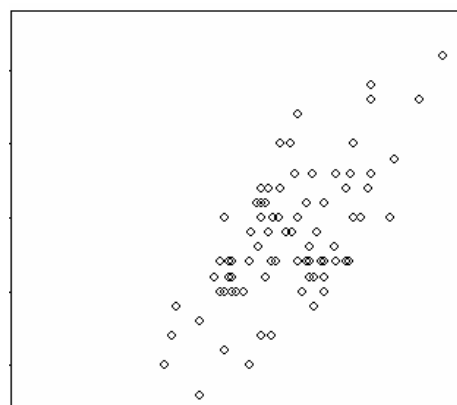


図3：加熱炉2に含まれるデータの散布図