

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-79971

(P2007-79971A)

(43) 公開日 平成19年3月29日(2007.3.29)

(51) Int. Cl.	F I	テーマコード (参考)
G06Q 50/00 (2006.01)	G06F 17/60 106	3C100
G06F 19/00 (2006.01)	G06F 19/00 120	
G05B 19/418 (2006.01)	G05B 19/418 Z	

審査請求 有 請求項の数 10 O L (全 25 頁)

(21) 出願番号	特願2005-267348 (P2005-267348)	(71) 出願人	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(22) 出願日	平成17年9月14日 (2005.9.14)	(74) 代理人	100074099 弁理士 大菅 義之
		(74) 代理人	100067987 弁理士 久木元 彰
		(72) 発明者	白井 英大 神奈川県川崎市高津区坂戸3丁目2番1号 富士通エルエスアイテクノロジー株式会社 内

最終頁に続く

(54) 【発明の名称】 製造データ解析方法及び製造データ解析装置

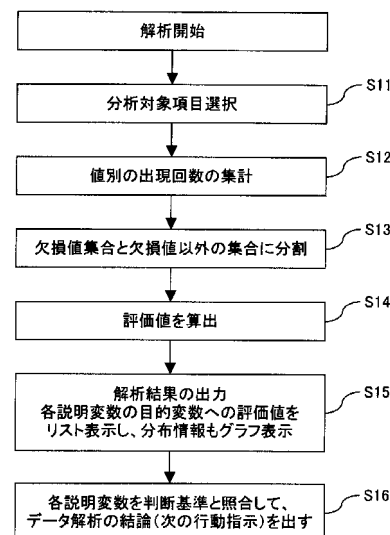
(57) 【要約】

【課題】 本発明の目的は、欠損値のある製造データの解析精度を高めることである。

【解決手段】 分析対象の説明変数を選択する(図3, S11)。各工程から取得した製造データを欠損値のある集合と欠損値の無い集合に分割する(S13)。欠損値のある集合と無い集合の製造データから説明変数が目的変数に与える影響を評価するための評価値を算出する(S14)。算出した評価値を判断基準と照合して製造データの解析を行う(S16)。

【選択図】 図3

データ解析処理の概略を示すフローチャート



## 【特許請求の範囲】

## 【請求項 1】

複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得して記憶手段に記憶し、

前記記憶手段に記憶した複数の製造データの中で説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合に分割し、

前記説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合とを分析して前記説明変数の前記目的変数に対する影響の度合いを評価する製造データ解析方法。

## 【請求項 2】

前記分析ステップは、前記説明変数の値が欠損している集合の前記目的変数の値と、前記説明変数の値が欠損していない集合の前記目的変数の値との間に有意な差があるか否かを分析して、前記説明変数の欠損値が解析結果に影響を与えるか否かを評価する請求項 1 記載の製造データ解析方法。

## 【請求項 3】

前記分析ステップは、各説明変数の値に基づいて少なくとも複数の製造データを第 1 及び第 2 の 2 つの集合に分割し、前記第 1 の集合の前記目的変数の値と前記第 2 の集合の前記目的変数の値とに基づいて各説明変数の前記目的変数に対する影響の強さを示す影響力データを第 1 の評価値として算出する請求項 1 または 2 記載の製造データ解析方法。

## 【請求項 4】

前記分析ステップは、前記説明変数の値が欠損している製造データの集合と前記第 1 の集合の前記目的変数の平均値に有意な差があるか否かを示す第 1 の確率と、前記説明変数の値が欠損している製造データの集合と前記第 2 の集合の前記目的変数の平均値に有意な差があるか否かを示す第 2 の確率とに基づいて前記説明変数の値が欠損している製造データが解析結果に与える影響を評価するための欠損影響率を第 2 の評価値として算出する請求項 3 記載の製造データ解析方法。

## 【請求項 5】

前記分析ステップは、複数の説明変数における同一の値の出現回数を計数し、計数した出現回数を新たな説明変数として追加し、前記出現回数が前記目的変数に与える影響の強さを示す影響力データを算出する請求項 1 または 2 記載の製造データ解析方法。

## 【請求項 6】

前記分析ステップは、前記説明変数の値である文字列または数値の欠損数を計数し、計数した欠損数を新たな説明変数として追加し、前記欠損数が前記目的変数に与える影響の強さを示す影響力データを算出する請求項 1 または 2 記載の製造データ解析方法。

## 【請求項 7】

複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得するデータ取得手段と、

前記データ取得手段により取得した製造データの中で前記説明変数が欠損している製造データの集合と前記説明変数が欠損していない製造データの集合に分割する分割手段と、

前記説明変数の値が欠損している製造データの集合と前記説明変数が欠損していない製造データの集合を分析して前記説明変数の前記目的変数に対する影響の度合いを評価する分析手段とを備える製造データ解析装置。

## 【請求項 8】

前記分析手段は、前記説明変数の値が欠損している集合の前記目的変数の値と、前記説明変数の値が欠損していない集合の前記目的変数の値との間に有意な差があるか否かを分析して、前記説明変数の欠損値が解析結果に影響を与えるか否かを評価する請求項 7 記載の製造データ解析装置。

## 【請求項 9】

前記分析手段は、各説明変数の値に対応する前記目的変数の値に基づいて少なくとも第 1 及び第 2 の 2 つの集合に分割し、前記第 1 の集合と前記第 2 の集合の前記目的変数の値に

10

20

30

40

50

基づいて前記各説明変数の前記目的変数に対する影響の強さを示す影響力データを第1の評価値として算出する請求項7または8記載の製造データ解析装置。

【請求項10】

前記分析手段は、複数の説明変数における同一の値の出現回数を計数し、計数した出現回数を新たな説明変数として追加し、前記出現回数が前記目的変数に与える影響の強さを示す影響力データを算出する請求項7または8記載の製造データ解析装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、製造工程から得られる製造データの解析方法及びその解析装置に関する。

10

【背景技術】

【0002】

複数の工程を経て製品を製造する場合に、要求する品質を満たし、かつ製品の生産効率を高めるために製造工程のどのような要因が製品の品質に影響するかを統計的及び物理的に解析する必要がある。複雑な製造工程を経て製品が製造される場合には、製品の品質に影響する要因は多数を有するので全ての要因を物理的に解析して影響の強い要因を抽出するには膨大な手間を必要とする。そのため、各工程の製造データを収集してデータ解析を行って要因の絞り込みを行った後、物理解析を行うことが一般的である。その際、解析を効率良く行うためにデータの自動解析や、大量のデータの中の隠された因果関係やパターンを探索するデータマイニングが行われている。

20

【0003】

しかしながら実際に得られる製造データには欠損値が存在することが多いのでそのままでは解析することが難しい。

例えば、図19(A)に示すように項目Aと項目Yの関連性を解析する場合、項目Aと項目Yの両方のデータが存在するサンプル01は解析対象となるが、どちらか一方のデータが欠損しているサンプル02、03は解析対象とはならない。

【0004】

また、図19(B)に示すように独立変数(説明変数)の項目A、B、Cと従属変数(目的変数)の項目Yの因果関係を解析する場合、サンプルLot0は、項目A、B、Cと、独立変数の項目Yの値が得られているので関連性を調べることができるが、サンプルLot01は独立変数の項目Cのデータが欠損し、サンプルLot02は独立変数の項目Bと従属変数の項目Yのデータが欠損し、サンプルLot03は、項目Aと項目Cのデータが欠落しているため、項目A、B、Cと項目Yとの関連性は不明である。

30

【0005】

図19(C)は、図19(B)のようにサンプルに欠損値がある場合に、装置名の項目Cが欠損しているサンプルに対しては「不明」(欠損値を示す文字列)で代替し、数値項目Bが欠損しているサンプルに対しては、各サンプルの項目Bの値の平均値「3」で代替した場合を示している。

【0006】

図19(D)は、図19(C)のサンプルから欠損値の多い項目Cを除いた場合の各項目の値と解析結果を示している。サンプルLot02は、従属変数の項目Yの値が不明であるため解析対象とならない。

40

【0007】

図20は、図19(D)のサンプルデータを従属変数(項目Y)への影響の強さを独立変数(項目A、B、C)で比較した場合の比較結果を示している。

項目Aの値による項目Yの値の差は「1」、項目Bの値による項目Yの値の差は「3.5」、項目Cの値によるYの差は不明である。

【0008】

従来、複数の製造データの解析を行う場合、欠損値が有るときには欠損値を正常値と同等に扱うために欠損値に代替値を挿入したり、欠損値の多いサンプルを除いたり、欠損値

50

の多い項目を除いたりしてデータを解析する必要があった。

【0009】

特許文献1には、サンプルデータの特徴値に欠損がある場合に、欠損データと正常データのマンハッタン距離を求め、マンハッタン距離が最小となる正常データを欠損データの代替値として補完することが記載されている。

【0010】

特許文献2には、複数の特徴量とイベントを組み合わせて相互に関連性のある部分状況データを抽出する場合に、入力イベントがどの選択領域内に存在するか否かを判定すること及び入力イベントに欠損値があるときに、欠損値以外の特徴量に対応する選択領域内のイベントに基づいて欠損値の補完値を算出することが記載されている。

10

【0011】

上述したように従来は、解析対象のデータに欠損値がある場合、何らかの方法で欠損値の代替値を求めその代替値を使用してデータ解析を行うか、欠損値の多いサンプルデータを除いてデータ解析を行っていた。

【特許文献1】特許第3654103号公報

【特許文献2】特開2001-184329号公報

【発明の開示】

【発明が解決しようとする課題】

【0012】

しかしながら、欠損値の代わりに入れる代替値としてどのような値を使用するか、あるいは欠損値の割合の高いサンプルを除く基準を何パーセントに設定するか等により解析結果が変わってしまうので解析精度にばらつきが生じてしまうという問題点がある。さらに、項目毎に欠損していないサンプルを使用すると、サンプル数が項目毎に異なることになるので項目間の解析精度にばらつきが生じることになる。

20

【0013】

例えば、回帰木分析では、独立変数である項目の値により従属変数である項目の値が高い集合と低い集合の2つの集合に分けて独立変数の従属変数に対する影響の強さを判定する必要がある。サンプルに欠損値があると分析できなので上述したように代替値を入れて分析することになるが、代替値の値により従属変数の値が高い集合か低い集合の何れかの集合に含まれてしまうことになり、それにより解析誤差が生じてしまう。

30

【0014】

本発明の課題は、製造データに欠損値がある場合のデータの解析精度を高めることである。

【課題を解決するための手段】

【0015】

本発明の製造データ解析方法は、複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得して記憶手段に記憶し、前記記憶手段に記憶した複数の製造データの中で説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合に分割し、前記説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合とを分析して前記説明変数の前記目的変数に対する影響の度合いを評価する。

40

【0016】

この発明によれば、製造データの一部に欠損がある場合でも、欠損値のある集合を解析することで製造データの解析精度を高めることができる。また、説明変数の欠損値が目的変数に与える影響を解析することで解析の信頼性を高めることができる。

【0017】

本発明の製造データ解析方法において、前記分析ステップは、前記説明変数の値が欠損している集合の前記目的変数の値と、前記説明変数の値が欠損していない集合の前記目的変数の値との間に有意な差があるか否かを分析して、前記説明変数の欠損値が解析結果に

50

影響を与えるか否かを評価する。

【0018】

このように構成することで説明変数が欠損している製造データが説明変数の値に影響を与えるか否かを分析することで解析結果の信頼性を高めることができる。

本発明の製造データ解析方法において、前記分析ステップは、各説明変数の値に基づいて少なくとも複数の製造データを第1及び第2の2つの集合に分割し、前記第1の集合の前記目的変数の値と前記第2の集合の前記目的変数の値とに基づいて各説明変数の前記目的変数に対する影響の強さを示す影響力データを第1の評価値として算出する。

【0019】

このように構成することで、どの説明変数が目的変数に影響する度合いが強いかを評価することができる。

10

本発明の製造データ解析方法において、前記分析ステップは、前記説明変数の値が欠損している製造データの集合と前記第1の集合の前記目的変数の平均値に有意な差があるか否かを示す第1の確率と、前記説明変数の値が欠損している製造データの集合と前記第2の集合の前記目的変数の平均値に有意な差があるか否かを示す第2の確率とに基づいて前記説明変数の値が欠損している製造データが解析結果に与える影響を評価するための欠損影響率を第2の評価値として算出する。

【0020】

このように構成することで説明変数の欠損値が解析結果にどの程度影響するかを評価することができるので解析結果の信頼性を高めることができる。

20

本発明の製造データ解析方法において、前記分析ステップは、複数の説明変数における同一の値の出現回数を計数し、計数した出現回数を新たな説明変数として追加し、前記出現回数が前記目的変数に与える影響の強さを示す影響力データを算出する。

【0021】

このように構成することで説明変数の特定の値が解析結果に与える影響を評価することができる。

本発明の製造データ解析方法において、前記分析ステップは、前記説明変数の値である文字列または数値の欠損数を計数し、計数した欠損数を新たな説明変数として追加し、前記欠損数が前記目的変数に与える影響の強さを示す影響力データを算出する。

【0022】

30

このように構成することで説明変数の欠損値が解析結果に与える影響を評価することができるのでデータ解析の信頼性を高めることができる。

図1は本発明の原理説明図である。図1において、製造データ解析装置は、複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得するデータ取得手段1と、前記データ取得手段1により取得した製造データの中で前記説明変数が欠損している製造データの集合と前記説明変数が欠損していない製造データの集合に分割する分割手段2と、前記説明変数の値が欠損している製造データの集合と前記説明変数が欠損していない製造データの集合を分析して前記説明変数の前記目的変数に対する影響の度合いを評価する分析手段3とを備える。

【0023】

40

この発明によれば、製造データの一部に欠損がある場合でも欠損値のある集合を解析することで製造データの解析精度を高めることができる。また、説明変数の欠損値が目的変数の値に与える影響を解析することで解析の信頼性を高めることができる。

【発明の効果】

【0024】

本発明によれば、製造データの一部に欠損がある場合でも欠損値のある集合を解析することで製造データの解析精度を高めることができる。また、説明変数の欠損値が解析結果に与える影響を評価することで解析の信頼性を高めることができる。

【発明を実施するための最良の形態】

【0025】

50

以下、本発明の実施の形態を図面を参照しながら説明する。図2は、実施の形態の製造データ解析装置11の構成を示す図である。

図2において、入力装置12はキーボード等からなり、製造ラインの各工程の製造データを入力する。演算処理装置13はパーソナルコンピュータなどからなり、後述する製造データ解析処理を実行し、解析結果をディスプレイ等からなる出力装置14に表示する。

【0026】

図3は、実施の形態の製造データ解析処理の概略を示すフローチャートである。以下の処理は演算処理装置13の図示しないCPUにより実行され、演算途中のデータ及び解析結果のデータはメモリ等の記憶手段に格納される。

【0027】

各製造工程から得られる製造データ(説明変数の値と目的変数の値)の中で分析対象の説明変数の項目を選択する(図3, S11)。説明変数とは、例えば、各製造工程で使用される装置名、製造条件、処理日時などの製品の出来映え(品質)に影響を与える要因であり、目的変数とは説明変数の影響を受けて値が決定される変数であり、製品の品質を示す不良率、歩留まり、製品の電気的特性値などである。

【0028】

次に、説明変数の項目に無関係に同じ値の出現回数を計数して出現回数を集計する(S12)。説明変数の値別の出現回数とは、例えば、各製造工程で同一の装置が使用されている場合の装置番号、製造条件、計測値、欠損値等の値の出現回数である。

【0029】

次に、各製造工程から収集した製造データの中で説明変数の値が欠損している製造データの集合と、説明変数の値が欠損していない製造データの集合に分割する(S13)。

次に、各説明変数の評価値を算出する(S14)。評価値とは、例えば、説明変数の目的変数に対する影響の強さを示す値、欠損のある製造データを含めて解析した場合と欠損のある製造データを除いて解析した場合で解析結果が異なる確率などである。

【0030】

次に、各説明変数の評価値をリスト表示し、同時に説明変数の値(装置番号、工程条件など)と目的変数の値の関係を示すグラフを表示する(S15)。

最後に、各説明変数の評価値を予め定めてある判断基準と照合してデータ解析を行い、解析結果に基づいて次にどのような解析を行うかを定める(S16)。

【0031】

上記の製造データ解析処理によれば、説明変数の欠損値のある製造データの集合と欠損値の無い製造データの集合に分割し、説明変数の値が欠損している製造データが解析結果に及ぼす影響を評価することでデータの解析精度及び解析の信頼性を高めることができる。

【0032】

次に、半導体の製造工程から得られる製造データを解析する場合を例に取り製造データ解析方法を具体的に説明する。図4は、半導体の製造工程と各製造工程で使用される装置を示す図である。

【0033】

半導体製造工程は、シリコン基板に絶縁膜を形成する工程、パターンを形成する工程、エッチングする工程等を何回も繰り返し、合計で1000以上の工程を経て製品が完成する。

【0034】

図4は、半導体製造工程の一部を示しており、前処理工程と、膜を形成するA膜工程と、露光するためのB露光工程と、膜を形成するC膜工程と、製品の最終検査を行い不良率を測定する不良率測定工程からなる。

【0035】

前処理工程では装置No66の装置が使用され、次のA膜工程では装置No1~5の装置が使用され、次のB露光工程では装置No1とNo7の装置が使用されている。また、

10

20

30

40

50

次のC膜工程では装置No1とNo8の装置が使用され、不良率測定工程では装置No109～113の装置が使用されている。

【0036】

半導体の製造工程では異なる工程で同じ装置が繰り返し使用されており、例えば、装置No1の装置はA膜工程とB露光工程とC膜工程で使用されている。

次に、上記の半導体製造工程の説明変数と目的変数の因果関係の強さを解析する製造データ解析処理を図5のフローチャートを参照して説明する。

【0037】

最初に各製造工程の製造データを取得してメモリ、外部記憶装置等に記憶すると共に、製造データの中の異常値と外れ値を除外する(図5, S21)。製造データが異常値または外れ値か否かの判断は、例えば、解析対象の製造工程に存在しない装置名を異常値と判断し、説明変数の値がサンプルデータの平均値の $\pm 4 \times$ 標準偏差の範囲に入らないとき、その値を外れ値と判断する。

10

【0038】

図6は、図4の半導体の製造工程から得られる製造データの一例を示しており、サンプルデータのLot45のB露光条件の説明変数の値「100」は、他のサンプルの値の平均値 $\pm 4 \times$ 標準偏差の範囲に入らないので外れ値と判断され、同じサンプルデータのC膜工程の説明変数である装置名「sh???'については、そのような装置名は登録されていないので異常値と判断され、それらのデータは欠損値として処理される。

【0039】

20

次に、分析対象項目を選択する(S22)。この分析対象項目としては、例えば、図6に示すようなA膜工程の装置名、B露光工程の装置名、B露光工程条件、C膜工程装置名、不良率測定工程で測定される不良率が選択される。

【0040】

次に、複数の説明変数で同一の値が出現する回数を計数して値別の出現回数を集計する(S23)。この値別の出現回数には装置名等の文字項目の欠損数、数値項目の欠損数も含まれる。

【0041】

次に、説明変数毎に説明変数の値が欠損している製造データの集合と値が欠損していない製造データの集合に分割する(S24)。

30

次に、各説明変数の目的変数に対する影響の強さを示す影響力データ(第1の評価値)と、欠損値を含む製造データが説明変数と目的変数の関係に影響を与える度合いを評価するための欠損影響率(第2の評価値)と、解析の信頼性を示す解析信頼率(第3の評価値)を算出する(S25)。

【0042】

影響力データは、説明変数毎にその値(装置名あるいは数値)により2つの集合に分割し、2つの集合の目的変数の値の平均値の差を2つの集合の標準偏差の差で標準化した値として求める。なお、説明変数の値が3つ以上ある場合には、影響力データの値が最も大きくなるように2つの集合に分割する。

【0043】

40

例えば、t検定のt値を用いて影響力データを以下のように算出する。

分割した2つの集合の目的変数の値の平均値を $X_1$ 、 $X_2$ 、目的変数の平方和を $S_1$ 、 $S_2$ 、データ数を $N_1$ 、 $N_2$ とすると、分割された集合の分散に有意差がない場合には、t値は以下の式で表せる。

【0044】

$$t = (X_1 - X_2)$$

$$\times \{ (1/N_1 + 1/N_2) \cdot (S_1 + S_2) / (N_1 + N_2 - 2) \}^{-1/2}$$

欠損影響率は、例えば、以下のようにして算出する。

【0045】

同じ説明変数で欠損のあるデータの集合の目的変数の値の平均値と分割した2つの集合

50

の中で平均値の高い方の集合（第1の集合）の目的変数の平均値に有意な差があるか否かを示す確率と、同じ説明変数で欠損のあるデータの集合の目的変数の平均値と分割した2つの集合の中で平均値の低い方の集合（第2の集合）の目的変数の平均値に有意な差があるか否かを示す確率の和の平均値として算出する。

【0046】

解析信頼率（第3の評価値）は、例えば、解析信頼率 = (1 - 欠損影響率) × (分割した2つの集合の目的変数の平均値に有意な差があるか否かを示す確率)、として算出する。

【0047】

次に、全ての説明変数と目的変数について上記の影響力データ、欠損影響率、解析信頼率の計算が終了したか否かを判別する（S26）。分析が終了していなければ（S26, NO）、ステップS24に戻り、残りの説明変数の値を欠損値の集合と欠損値以外の集合に分割して上記の処理を行う。

【0048】

他方、全ての説明変数の項目と目的変数の関係の分析が終了したときには（S26, YES）、ステップS27に進み、算出した影響力データに基づいて説明変数を順位付ける。

【0049】

次に、各説明変数と目的変数の因果関係の強さを示す影響力データの順位、影響力データ、欠損影響率、解析信頼率をリスト表示し、同時に説明変数の値と目的変数の値の分布状態をグラフ表示する（S28）。

【0050】

次に、各説明変数の影響力の順位と影響力データと欠損影響率を予め定めてある判断基準と照合してデータ解析の結論を出す（S29）。具体的には、影響力の順位が5位以内で、条件（a）影響力データが2以上で、条件（b）欠損影響が20%以下を満たすか否かを判断する。

【0051】

解析対象の説明変数の影響力データの順位が5以内で、判断条件（a）と（b）を全て満たす場合には、ステップS30に進み、この説明変数が目的変数（例えば、不良率）を左右する要因であると判断して確証を得るための詳細な調査を行う。

【0052】

また、解析対象の説明変数の影響力データが「2」以上で条件（a）を満たすが、欠損影響率が20%より大きく条件（b）を満たさない場合には、ステップS31に進み、この説明変数の欠損値が重要であると判断し、欠損値の真の値を調査する（サンプルデータの再取得などにより）。

【0053】

また、解析対象の説明変数の影響力データが「2」未満で条件（a）を満たさないときには、ステップS32に進み、解析対象のサンプルデータからは信頼できる結果が得られないと判断し、サンプルデータの追加や他の説明変数の解析などを検討する。

【0054】

次に、上述した製造データ解析処理の内容を、図6と図7の各工程の製造データを用いて具体的に説明する。

図7は、図6の製造データに装置No1の使用回数と文字項目の欠損数と数値項目の欠損数を説明変数の項目として追加した場合の製造データを示している。

【0055】

図7の説明変数の「A膜工程装置名」に関して欠損のある集合と無い集合の分割について説明する。サンプルデータの中でA膜工程の装置名が欠損しているのは、45個のサンプルデータの中で、Lot03、Lot07、Lot11、Lot15、Lot19、Lot23、Lot27、Lot31、Lot35、Lot39、Lot43、Lot45である。従って、それらのサンプルデータがA膜工程装置名の値が欠損している製造デー

10

20

30

40

50



タの集合となり、残りのサンプルデータが欠損の無い製造データの集合となる。

【0056】

他の説明変数に関しても同様に欠損のある製造データの集合と欠損の無い製造データの集合に分割する。

次に、各サンプルデータについて同一の値の出現回数と文字項目の欠損回数と数値項目の欠損回数を計数する。

【0057】

例えば、サンプルデータLot02は、工程全体で装置No1の装置が2回使用されているので装置No1の使用回数が「2」となり、B露光工程の装置名が欠損し、B露光工程条件が欠損しているため文字項目欠損数と数値項目欠損数がそれぞれ「1」となり、それらの値が新たな説明変数の値として追加される。

10

【0058】

この結果、図7に示すようにサンプルデータの説明変数として「装置No1の使用回数」と「文字項目欠損数」と「数値項目欠損数」が追加される。

以上のようにして製造データの説明変数として「装置No1の使用回数」と「文字項目欠損数」と、「数値項目欠損数」を追加したなら、それらのデータに基づいて各説明変数の影響力データ、影響力の順位、欠損影響率、解析信頼率を計算する。

【0059】

図8は、図7の製造データの分析結果の評価値リストを示しており、この評価値リストは、影響力データにより各説明変数が目的変数に与える影響の強さを順位づけたものである。

20

【0060】

図8から「装置No1使用回数」が目的変数である不良率に与える影響が最も大きく、「B露光工程条件」が不良率に与える影響が2番目に大きく、「文字項目欠損数」が不良率に与える影響が3番目に大きいことが分かる。

【0061】

説明変数「装置No1使用回数」の値は、サンプルデータ全体で0, 1, 2, 3の4個であるので、それらの値で2つの集合に分割し、分割した集合の影響力データの値が最も大きくなるように集合を選定する。この例では、装置No1の使用回数が2以下の集合と、使用回数が3以上の2つの集合に分割したときに影響力データの値が最も大きくなった。

30

【0062】

説明変数の「装置No1使用回数」に関しては、使用回数が2以上か、3以上かによりサンプルデータを2つの集合に分割した。図8に示すように、対象値（装置No1使用回数）が2以下の集合（目的変数の値が低い方の集合）の目的変数の平均値meanは「mean = 5.79」、標準偏差sは「s = 1.599」、データ数Nは「N = 37」となった。また、対象値が3以上の集合（目的変数の値が高い方の集合）の目的変数の平均値は「mean = 9.10」、標準偏差は「s = 2.229」、データ数は「N = 4」となった。

【0063】

これら2つの集合の目的変数の平均値の差を標準分散の差で標準化し、t検定によりt値を計算すると、影響力データの値は「3.7929」となる。全ての説明変数の影響力データを計算し、欠損影響率も考慮して影響力の順位づけを行ったところ、説明変数「装置No1使用回数」の不良率に対する影響力の順位が1番高かった。

40

【0064】

実施の形態では、装置No1の使用回数の欠損値は存在しないものとして計算したので、欠損影響率は「0」、解析信頼率は「99.951%」となった。

図9は、装置No1の使用回数と製品の不良率の関係を示すグラフであり、グラフの横軸は装置No1の使用回数、縦軸は不良率（パーセント）を示している。

【0065】

50

図9のグラフから、装置No1の使用回数が2以下のデータの不良率の平均値は「5.53%」の近傍に分布しており、使用回数が3以上のデータとは不良率の平均値が異なることが分かる。

【0066】

上記の結果から装置No1の装置の使用回数が不良率に与える影響が最も大きく、使用回数が0~2回では不良率はそれほど高くないが、使用回数が3回以上となると不良率が高くなることが分かる。

【0067】

次に、説明変数の「B露光工程条件」に関しては、対象値(露光工程条件の値)が5.6以下か、それとも5.6より大きいかにより2つの集合に分割した。図8に示すように、露光工程条件の値が5.6以下のデータの集合の不良率の平均値は「5.53」、標準偏差は「1.801」、データ数は「29」であった。また、露光工程条件の値が5.6より大きいデータの集合の不良率の平均値は「7.925」、標準偏差は「1.286」、データ数は「8」であった。

10

【0068】

上記の2つの集合の不良率の平均値と標準偏差とから影響力データを算出すると、その値は「3.5047」である。従って、説明変数「B露光工程条件」の目的変数に対する影響力の大きさの順位は2番目になる。

【0069】

また、「B露光工程条件」の値が欠損しているデータの集合の不良率の平均値は「6.75」、標準偏差は「1.4434」、データ数は「4」であった。欠損値の集合の不良率の平均値と2分割された集合の中で高い方の値の集合の不良率の平均値に有意な差があると云える確率と、欠損値の集合の不良率の平均値と2分割された集合の中で値の低い方の集合の不良率の平均値との間に有意な差があると云える確率の平均値を算出する。そして、その確率の平均値を欠損影響率として算出する。なお、平均値に差があると云える確率はt分布を用いて計算している。

20

【0070】

この場合の欠損影響率は「19.364%」となる。その欠損影響率と2分割された集合の不良率の平均値に有意な差があると云える確率から解析信頼率を計算すると、解析信頼率は「80.535%」となる。

30

【0071】

図10は、B露光工程条件と不良率の関係を示すグラフであり、グラフの横軸は露光工程条件の値を、縦軸は不良率をパーセントで示している。

図10のグラフから露光工程条件の値が5.6以下のデータは、不良率が平均値の「5.53%」の近傍に分布しているが、露光工程条件が5.6より大きいデータは、不良率がそれより大きい値となっていることが分かる。

【0072】

図8の評価値リストと図10から説明変数「B露光工程条件」の値が5.6より大きくなると不良率が高くなることが分かる。また、図8の欠損影響率が19.3%であることから、この解析結果に対する説明変数の「B露光工程条件」の欠損値の影響は少ないことが分かる。

40

【0073】

次に、説明変数の「文字項目欠損数」に関しては、対象値(文字項目欠損数)が1以上か、0以下かにより2つの集合に分割する。図8に示すように、文字項目欠損数が1以上の集合の不良率の平均値は「5.709」、標準偏差は「1.672」、データ数は「32」であった。文字項目欠損数が0以下の集合の不良率の平均値は「7.57」、標準偏差は「2.105」、データ数は「9」であった。

【0074】

上記の2つの集合の不良率の平均値と標準偏差とから影響力データを算出すると、その値は「2.7819」である。従って、説明変数「文字項目欠損数」の目的変数に対する

50

影響力の大きさの順位は3番目となる。

【0075】

図11は、文字項目欠損数と不良率の関係を示すグラフであり、グラフの横軸は文字項目欠損数の値を示し、縦軸は不良率をパーセントで示している。

図11のグラフから文字項目の欠損数が1以上のデータは、不良率が平均値の「5.709」の近傍に分布しているが、欠損数が0以下のデータは、不良率がそれより大きな値となっていることが分かる。

【0076】

図8の評価値リストと図11から、説明変数「文字項目欠損数」の値が0のデータは、欠損数が1以上のデータに比べて不良率が高くなることが分かる。

10

次に、説明変数の「B露光工程装置名」に関しては、対象値（B露光工程の装置名）が欠損しているか、それとも装置名が装置No1かにより2つの集合に分割する。図8に示すように、装置No1の装置を使用したデータの集合の不良率の中央値（med）は「6.50」、標準偏差は「1.910」、データ数は「30」であった。

【0077】

上記の値と装置No1の装置を使用したデータの不良率の平均値とから影響力データを算出すると、その値は「2.2837」である。従って、説明変数「B露光工程装置」の不良率に対する影響力の順位は4番目となる。

【0078】

また、「B露光工程装置名」のデータが欠損している製造データの集合の不良率の中央値は「4.70」、標準偏差は「1.528」、データ数は「11」であった。

20

上記の値と欠損値の集合の不良率の平均値から欠損影響率を計算すると、欠損影響率は「97.222」となる。さらに、その欠損影響率と2つの集合間の不良率の平均値に有意な差があると言える確率（t検定値）から解析信頼率を計算すると、解析信頼率は「2.778」となる。

【0079】

図12は、B露光工程装置名と不良率の関係を示すグラフであり、グラフの横軸は装置名を、縦軸は不良率をパーセントで示している。

図12のグラフから装置名の欠損しているデータの不良率が、B露光工程で装置No1の装置を使用したデータの不良率より明らかに低いことが分かる。

30

【0080】

図8の評価値リストと図12から、B露光工程装置名が欠損している場合と装置No1の装置を使用した場合で不良率に差があるが、欠損影響率が高いことから欠損値の真の値を再度調査しないとB露光工程の装置が不良率に与える影響を正確に評価することは難しいことが分かる。

【0081】

次に、説明変数の「C膜工程装置名」に関しては、C膜工程の装置名が欠損しているか、それとも装置名が装置No1かにより2つの集合に分割する。図8に示すように、装置No1を使用したデータの集合の不良率の中央値は「6.10」、標準偏差は「2.349」、データ数は「13」であった。

40

【0082】

上記の値と装置No1を使用したデータの集合の不良率の平均値から「C膜工程装置名」の影響力データを算出すると、その値は「0.9961」となる。従って、説明変数「C膜工程装置名」の不良率に対する影響力の順位は5番目になる。

【0083】

また、「C膜工程装置名」が欠損している製造データの集合の不良率の中央値は「5.65」、標準偏差は「1.683」、データ数は「28」であった。

上記の値と欠損値の集合の不良率の平均値とから欠損影響率を算出すると、その値は「67.481」となる。その欠損影響率から解析信頼率を計算すると解析信頼率は「32.519」となる。

50

## 【 0 0 8 4 】

図 1 3 は、C 膜工程装置名と不良率の関係を示すグラフであり、グラフの横軸は C 膜工程装置名を、縦軸は不良率をパーセントで示している。

図 1 3 のグラフから装置名の欠損しているデータの不良率の方が、装置 N o 1 を使用したときの不良率よりわずかに低いことが分かる。

## 【 0 0 8 5 】

図 8 の評価値リストと図 1 3 から、説明変数「C 膜工程装置名」は欠損影響率が高いことから、欠損値の真の値を再度調査しないと C 膜工程装置名の不良率に対する影響は正確には評価できないことが分かる。

## 【 0 0 8 6 】

次に、説明変数の「数値項目欠損数」に関しては、対象値（数値項目の欠損数）が 0 以下か、それとも 1 以上かにより 2 つの集合に分割する。図 8 に示すように数値項目欠損数が 0 のデータの集合の不良率の平均値は「6.05」、標準偏差は「1.960」、データ数は「37」であった。欠損数が 1 以上のデータの集合の不良率の平均値は「6.75」、標準偏差は「1.433」、データ数は「4」であった。

## 【 0 0 8 7 】

上記の 2 分割した集合の値から「数値項目欠損数」の影響力データを計算すると、その値は「0.6921」となる。従って、説明変数「数値項目欠損数」の不良率に対する影響力の順位は 6 番目となる。この場合、欠損影響率は 0 % となり、解析信頼率は「50.713 %」となる。

## 【 0 0 8 8 】

図 1 4 は、数機項目欠損数と不良率の関係を示すグラフであり、グラフの横軸は数値項目欠損数を、縦軸は不良率をパーセントで示している。

図 1 4 のグラフから数値項目が欠損していないデータの不良率はばらつきが大きく、数値項目の欠損数が 1 のデータの不良率とは大きな違いは無い。

## 【 0 0 8 9 】

図 8 の評価値リストと図 1 4 から数値項目欠損数の不良率に与える影響は小さいことが分かる。

次に、説明変数の「A 膜工程装置名」に関しては、対象値（装置名）が装置 N o 5 か、それとも装置 N o 1 または N o 3 かにより 2 つの集合に分割する。使用する装置名が装置 N o 5 のデータの集合の不良率の中央値（med）は「4.95」、標準偏差は「1.142」、データ数は「10」であった。装置名が装置 N o 1 または N o 3 のデータの不良率の中央値は「7.40」、標準偏差は「1.2237」、データ数は「20」であった。

## 【 0 0 9 0 】

上記の値と 2 分割した集合の不良率の平均値から「A 膜工程装置名」の影響力データを計算すると、その値は「0.4098」となる。

また、「C 膜工程装置名」が欠損しているデータの集合の不良率の中央値は「7.40」、標準偏差は「1.233」、データ数は「11」であった。

## 【 0 0 9 1 】

上記の値と欠損値の集合の不良率の平均値から欠損影響率を計算すると、その値は「98.1397」となる。また、解析信頼率は「0.5861」となる。

図 1 5 は、A 膜工程装置名と不良率の関係を示すグラフであり、グラフの横軸は装置名を、縦軸は不良率のパーセントを示している。

## 【 0 0 9 2 】

図 1 5 のグラフから装置名が欠損しているデータの不良率が、装置名が確定しているデータの不良率より明らかに高いことから欠損値の影響が大ききことが分かる。

図 8 の評価値リストと図 1 5 のグラフから説明変数「A 膜工程装置名」は欠損影響率が高いことから、欠損値の真の値を再調査しないと不良率に与える影響を評価することは難しい。

10

20

30

40

50

## 【0093】

上述した第1の実施の形態によれば、説明変数毎にそれぞれの値が欠損しているサンプルデータ（製造データ）の集合と、欠損していない製造データの集合に分割し、それらの集合を分析することで説明変数の欠損値が解析結果に及ぼす影響を定量的に評価することができる。従って、欠損値が目的変数に与える影響が小さい場合には、欠損値の代替値を使用せずに製造データの解析を行うことができるので代替値を使用する場合に比べてデータの解析精度を高めることができる。

## 【0094】

また、説明変数の値により複数の製造データを2つの集合に分割して、説明変数の目的変数に対する影響の強さを示す影響力データを算出し、欠損値の有る製造データを含めた場合と含めない場合の不良率の平均値が異なる確率を欠損影響率として算出している。これにより、欠損値を含む製造データを用いてデータの解析を行った場合に、欠損値が解析結果にどの程度影響するかを評価することができるので解析結果の信頼性を高めることができる。さらに、解析信頼率を算出することで欠損値が含む製造データを用いた解析結果の信頼性を評価することができる。

10

## 【0095】

また、複数の説明変数に対して同一の値の出現回数、あるいは文字項目または文字項目の欠損数を計数し、計数した値を新たな説明変数としてサンプルデータに追加し、それらの説明変数の解析結果から特定の説明変数の値、あるいは欠損している説明変数の値が目的変数（例えば、不良率）に影響を与えているか否かを解析することができる。

20

## 【0096】

例えば、半導体の製造工程では同じ装置が複数の工程で使用される場合があるので、特定の装置の使用回数と不良率の関係を解析することで、装置の使用回数が不良率に影響を与えているか否かを判断できる。これにより、各製造工程から得られる製造データを解析しただけでは得られない解析結果、つまり複数の工程における説明変数の同一の値（装置名、工程条件）が不良率に及ぼす影響等を解析することができる。製造工程で使用する装置や工程条件の数値の欠損数が不良率、歩留まり、電気特性等の目的変数に影響を与えるか否かは従来の解析手法では全く考慮されていなかったが、本発明の実施の形態の解析方法では特定の装置や特定の工程条件等の要因が不良率に及ぼす影響を解析することができる。

30

## 【0097】

次に、本発明の第2の実施の形態のデータ解析処理を、図16のフローチャートを参照して説明する。

最初に各製造工程の製造データを取得し、取得した製造データの中の異常値と外れ値を除外する（図16，S41）。製造データが異常値または外れ値か否かの判断は、例えば、解析対象の製造工程に存在しない装置名を異常値と判断し、説明変数の値がサンプルデータの平均値の $\pm 4 \times$ 標準偏差の範囲に入らないとき外れ値と判断する。

## 【0098】

次に、分析対象項目を選択する（S42）。次に、分析対象の説明変数の項目を選択する（S22）。分析対象項目としては、例えば、図6に示すようなA膜工程の装置名、B露光工程の装置名、B露光工程条件、C膜工程装置名、不良率測定工程で測定される不良率が選択される。

40

## 【0099】

次に、各サンプルの選択した説明変数の値別の出現回数を集計する（S23）。この値別の出現回数には装置名等の文字項目の欠損数、数値項目の欠損数も含まれる。

次に、各サンプルの説明変数毎にその値が欠損している集合と欠損していない集合に分割する（S24）。

## 【0100】

次に、各説明変数の目的変数に対する影響の強さを示す影響力データ（第1の評価値）と、欠損値を含む製造データが説明変数と目的変数の関係に影響を与える度合いを評価す

50

るための欠損影響率（第2の評価値）と、解析の信頼性を示す解析信頼率（第3の評価値）を算出する（S25）。

【0101】

次に、全ての説明変数と目的変数との関係进行分析したか否かを判別する（S46）。分析が終了していない説明変数が残っている場合には（S46, NO）、ステップS44に戻り、次の説明変数と目的変数の関係の分析を行う。

【0102】

ステップS46において、全ての説明変数と目的変数の関係の分析が終了した場合には（S46, YES）、ステップS47に進み、影響力データの大きさに従って各説明変数を順位付けする。

10

【0103】

次に、説明変数の値が欠損しているデータの集合と、説明変数の特定の値のデータの集合とから欠損値類似率を算出し、欠損値類似率の最も高い説明変数の値を欠損値推定値として選定する（S48）。

【0104】

欠損値類似率は、例えば、説明変数の値が欠損しているデータの集合の目的変数の平均値と、同じ説明変数の特定の値のデータの集合の平均値が一致している確率をt検定により求め、その確率を欠損値類似率として算出する。あるいは、t検定により求めた確率と、尖度による検定を用いた2つの集合が一致している確率と、歪み度による検定を用いた2つの集合が一致している確率との積を欠損値類似率として算出する。

20

【0105】

次に、各説明変数の順位、影響力データの値、欠損影響率、解析信頼率、欠損値類似率をリスト表示し、同時に各説明変数の値と目的変数の値の分布情報をグラフ表示する（S49）。

【0106】

次に、上記のようにして算出した各説明変数の順位、影響力データ、欠損影響率、解析信頼率を予め定めてある判断基準と照合して、影響力データの大きさの順位が5位以内で、(a)影響力データの値が「2」以上か、(b)欠損影響率が20%以下か否かを判断する（S50）。

【0107】

ステップS50で判断基準を全て満たすと判断された場合、すなわち、順位が5位以内で、影響力データが「2」以上で、かつ欠損影響率が20%以下のときには、ステップS51に進み、その説明変数は目的変数（不良率）を左右する大きな要因であると判断して、確証を得るために詳細な調査を行う。

30

【0108】

ステップS50で、条件(a)を満たすが、条件(b)を満たさないと判断された場合、すなわち影響力データの値は「2」以上であるが、欠損影響率が20%より大きいときには、ステップS52に進み、欠損値が解析結果に影響を与えると判断して、欠損値類似率と欠損値推定値を参考して欠損値の真の値を調査する（サンプルデータの再取得を含む）。

40

【0109】

また、ステップS50で、条件(a)を満たさないと判断されたとき、すなわち影響力データの値が2未満のときには、ステップS53に進み、解析対象のサンプルデータからは信頼できる分析結果が得られないと判断し、サンプルデータの追加や他の解析方法を検討する。

【0110】

図17は、上述した製造データ解析処理により作成した第1分岐候補の評価値リストを示す図である。

図17の評価値リストによると、説明変数の「項目D」に関しては、影響力データの値が「13.9181」、欠損影響率が「15.426」であるので、図16のステップS

50

50の判断基準を全て満たすことになる。従って、「項目D」は目的変数（例えば、不良率）への影響が1番大きい要因と判断される。

【0111】

この説明変数「項目D」の値が欠損しているデータの集合と説明変数「項目D」の特定の値の集合との欠損値類似率を、上述した算出方法で算出した場合、「装置No3」との欠損値類似率が「79.456%」で最も高いので説明変数「項目D」の欠損値推定値は「装置No3」となる。

【0112】

なお、図17の「項目D」の「低い集合と分布」の欄に記載されている「No2, No5 = 0.70, s = 0.32」と、「高い集合と分布」の欄に記載されている「No3 = 0.86, s = 0.21」は、サンプルデータを「項目D」の値が「装置No3またはNo5」の集合と「装置No3」の集合の2つに分割したときに、それぞれの集合の目的変数の平均値と標準偏差sを示している。

【0113】

また、「欠損値集合分布」の欄に記載されている「平均 = 0.858, s = 0.25」は、説明変数「項目D」の値が欠損しているデータの集合の目的変数の平均値とその標準偏差sを示している。

【0114】

また、影響力の順位が2番目の説明変数の「項目A」に関しては、「項目A」の値が「11.6」の集合の欠損値類似率が「55.41%」で最大であるので、欠損値推定値は「11.6」となる。

【0115】

同様に、影響力の順位が3番目の説明変数「項目C」に関しては、欠損値の集合と「装置No4」の集合の欠損値類似率が「44.12%」であり、その値が最も大きいので、説明変数「項目C」の欠損値推定値は「装置No4」となる。

【0116】

次に、図18は、説明変数「項目D」の樹木図を作成したときの各集合の目的変数の分布を示す図である。

図18に示すように説明変数「項目D」の全体サンプル数は「100」で目的変数が欠損していないサンプル数は「80」である（この集合を集合1と呼ぶ）。

【0117】

「項目D」の値が欠損しているデータの集合（これを集合2と呼ぶ）の目的変数の平均値は「0.858」、データ数nは「35」、標準偏差sは「0.25」である。

同じ「項目D」の値が「装置No2またはNo5」の集合（これを集合3と呼ぶ）の目的変数の平均値は「0.7」、データ数は「30」、標準偏差は「0.32」である。

【0118】

また、「項目D」の値が「装置No3」の集合（これを集合4と呼ぶ）の目的変数の平均値は「0.86」、データ数は「15」、標準偏差は「0.21」である。なお、集合3は、その下側の分岐線で示すように複数の集合に分割される。

【0119】

上記の集合2と集合3の欠損値類似率と集合2と集合4の欠損値類似率をそれぞれ計算すると、図17に示すように集合4の欠損値類似率が最も高かった。

さらに、集合2と集合3及び集合4の目的変数の値の分布形状を比較すると、集合2と集合4の分布形状が類似している。

【0120】

以上の解析結果から説明変数「項目D」の欠損値として集合4の値「装置No3」が最も可能性が高いと推定される。

上述した第2の実施の形態によれば、説明変数の値の欠損値の集合と、同じ説明変数の他の値の集合が一致する確率を欠損値類似率として算出し、その欠損値類似率が最大となる説明変数の値を欠損値として推定することができる。さらに、その説明変数の欠損値が

10

20

30

40

50

目的変数（例えば、不良率）に及ぼす影響を欠損影響率として求めることができる。

【0121】

また、第2の実施の形態も第1の実施の形態と同様に、説明変数の値が欠損している集合と欠損していない集合に分割し、欠損値の集合が目的変数に影響を与える度合いを欠損影響率等で評価しているため製造データの解析精度及び解析信頼性を高めることができる。

【0122】

本発明は上述した実施の形態に限らず、例えば、以下のように構成しても良い。

(1) 実施の形態は、t検定により影響力データ、欠損影響率等を算出したが、t検定以外の他の検定方法で評価値を算出しても良い。

(2) 実施の形態は半導体製造工程の製造データの解析について説明したが、半導体製造工程に限らず、他の製造工程、あるいは製造工程以外のデータの解析にも本発明は適用できる。

(3) 欠損影響率、解析信頼率及び欠損値類似率は、実施の形態に述べた算出方法に限らず、解析方法に適した他の算出方法で算出して良い。

【0123】

(付記1) 複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得して記憶手段に記憶し、

前記記憶手段に記憶した複数の製造データの中で説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合に分割し、

前記説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合とを分析して前記説明変数の前記目的変数に対する影響の度合いを評価する製造データ解析方法。

(付記2) 前記分析ステップは、前記説明変数の値が欠損している集合の前記目的変数の値と、前記説明変数の値が欠損していない集合の前記目的変数の値との間に有意な差があるか否かを分析して、前記説明変数の欠損値が分析結果に影響するか否かを評価する付記1記載の製造データ解析方法。

(付記3) 前記分析ステップは、各説明変数の値に基づいて少なくとも複数の製造データを第1及び第2の2つの集合に分割し、前記第1の集合の前記目的変数の値と前記第2の集合の前記目的変数の値とに基づいて各説明変数の前記目的変数に対する影響の強さを示す影響力データを第1の評価値として算出する付記1または2記載の製造データ解析方法。

(付記4) 前記分析ステップは、前記説明変数の値が欠損している製造データの集合と前記第1の集合の前記目的変数の平均値に有意な差があるか否かを示す第1の確率と、前記説明変数の値が欠損していない製造データの集合と前記第2の集合の前記目的変数の平均値に有意な差があるか否かを示す第2の確率とに基づいて前記説明変数の値が欠損している製造データが解析結果に与える影響を評価するための欠損影響率を第2の評価値として算出する付記3記載の製造データ解析方法。

(付記5) 前記分析ステップは、複数の説明変数における同一の値の出現回数を計数し、計数した出現回数を新たな説明変数として追加し、前記出現回数が前記目的変数に与える影響の強さを示す影響力データを算出する付記1または2記載の製造データ解析方法。

(付記6) 前記分析ステップは、前記説明変数の値である文字列または数値の欠損数を計数し、計数した欠損数を新たな説明変数として追加し、前記欠損数が前記目的変数に与える影響の強さを示す影響力データを算出する付記1または2記載の製造データ解析方法。

(付記7) 前記分析ステップは、 $(1 - \text{欠損影響率}) / (\text{前記第1の集合と第2の集合の前記目的変数の値の平均値に有意な差があるか否かを示す確率})$ により得られる解析信頼率を第3の評価値として算出する付記4記載の製造データ解析方法。

(付記8) 前記分析ステップは、各製造データの説明変数の値の出現回数を計数し、計数した出現回数を前記製造データに新たな説明変数として追加し、該出現回数に基づいて

10

20

30

40

50



複数の製造データを少なくとも2つの集合に分割し、前記2つの集合の前記目的変数の値に基づいて前記の出現回数が目的変数に与える影響の強さを示す影響力データを算出する付記1または2記載の製造データ解析方法。

(付記9) 前記分析ステップは、各製造データの説明変数の値の欠損数を計数し、計数した欠損数を前記製造データに新たな説明変数として追加し、該欠損数に基づいて複数の製造データを少なくとも2つの集合に分割し、前記2つの集合の前記目的変数の値に基づいて前記説明変数の欠損が目的変数に与える影響の強さを示す影響力データを算出する付記1または2記載の製造データ解析方法。

(付記10) 前記分析ステップは、各説明変数の値に対応する前記目的変数の値により少なくとも第1及び第2の2つの集合に分割し、前記第1の集合の目的変数の値の平均値と第2の集合の前記目的変数の値の平均値とそれぞれの標準偏差とに基づいて各説明変数の前記目的変数に対する影響の強さを示す影響力データを算出し、前記説明変数の値が欠損している製造データの集合の前記目的変数の値の平均値と前記第1の集合の前記目的変数の平均値に有意な差があるか否かを示す第1の確率と、前記説明変数の値が欠損している製造データの集合の前記目的変数の値の平均値と前記第2の集合の前記目的変数の値の平均値に有意な差があるか否かを示す第2の確率とに基づいて欠損影響率を算出し、前記影響力データに基づいて目的変数に対する影響力の強い順に前記説明変数を順位付けし、影響力の強さを示す順位と前記影響力データと前記欠損影響率とに基づいて前記説明変数が前記目的変数に影響を与えるか否かを評価する付記1または2記載の製造データ解析方法。

(付記11) 前記分析ステップは、 $(1 - \text{欠損影響率}) / (\text{前記第1の集合と第2の集合の目的変数の値の平均値に有意な差があるか否かを示す確率})$ により得られる解析信頼率を第3の評価値として算出し、前記第3の評価値が所定値以上か否かにより解析結果の信頼性を評価する付記10記載の製造データ解析方法。

(付記12) 前記分析ステップは、前記説明変数の値が欠損している製造データの集合と前記説明変数の1または複数の値の集合の目的変数の値とに基づいて欠損値類似率を算出し、欠損値類似率が最も高い集合の前記説明変数の値を欠損値の値と推定する付記1乃至11の何れか1項に記載の製造データ解析方法。

(付記13) 複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得するデータ取得手段と、前記データ取得手段により取得した製造データの中で前記説明変数が欠損している製造データの集合と前記説明変数が欠損していない製造データの集合に分割する分割手段と、前記説明変数の値が欠損している製造データの集合と前記説明変数が欠損していない製造データの集合を分析して前記説明変数の前記目的変数に対する影響の度合いを評価する分析手段とを備える製造データ解析装置。

(付記14) 前記分析手段は、前記説明変数の値が欠損している集合の前記目的変数の値と、前記説明変数の値が欠損していない集合の前記目的変数の値との間に有意な差があるか否かを分析して、前記説明変数の欠損値が分析結果に影響するか否かを評価する付記13記載の製造データ解析装置。

(付記15) 前記分析手段は、各説明変数の値に対応する前記目的変数の値に基づいて少なくとも第1及び第2の2つの集合に分割し、前記第1の集合と前記第2の集合の前記目的変数の値に基づいて前記各説明変数の値の前記目的変数に対する影響の強さを示す影響力データを第1の評価値として算出する付記13または14記載の製造データ解析装置。

(付記16) 前記分析手段は、複数の説明変数における同一の値の出現回数を計数し、計数した出現回数を新たな説明変数として追加し、前記出現回数が前記目的変数に与える影響の強さを示す影響力データを算出する付記13または14記載の製造データ解析装置。

(付記17) 前記分析手段、欠損のある製造データの集合と前記第1の集合の前記目的変数の平均値に有意な差があるか否かを示す確率と、前記欠損のある製造データの集合と前記第2の集合の前記目的変数の平均値に有意な差があるか否かを示す確率との平均値を

10

20

30

40

50

欠損影響率を示す第2の評価値として算出する付記13記載の製造データ解析装置。

(付記18) 前記分析手段は、 $(1 - \text{欠損影響率}) / (\text{前記第1の集合と第2の集合の前記目的変数の値の平均値に有意な差があるか否かを示す確率})$ により得られる解析信頼率を第3の評価値として算出する付記13記載の製造データ解析装置。

(付記19) 前記分析手段は、各説明変数の値により少なくとも第1及び第2の2つの集合に分割し、前記第1及び第2の集合の前記目的変数の値に基づいて各説明変数の前記目的変数に対する影響の強さを示す影響力データを第1の評価値として算出し、欠損のある製造データの集合と前記第1の集合の前記目的変数の平均値に有意な差があるか否かを示す確率と、前記欠損のある製造データの集合と前記第2の集合の前記目的変数の値の平均値に有意な差があるか否かを示す確率との平均値を欠損影響率を示す第2の評価値として算出し、前記第1の評価値に基づいて目的変数に対する影響力の強い説明変数の値を順位付けし、影響力の強さを示す順位が所定順位以上で、かつ前記目的変数に対する影響力の強い前記説明変数の値の前記欠損影響率が所定の基準値以下か否かにより、前記説明変数の前記目的変数に及ぼす影響の強さを評価する付記13記載の製造データ解析装置。

(付記20) 前記分析手段は、前記説明変数の値が欠損している製造データの集合と前記説明変数の1または複数の値の集合の目的変数の値に基づいて欠損値類似率を算出し、欠損値類似率が最も高い集合の前記説明変数の値を欠損値の値と推定する付記13乃至19のいずれか1項に記載の製造データ解析装置。

(付記21) 複数の製造工程の複数の説明変数の値と前記複数の製造工程により製造される製品の品質を示す目的変数とからなる製造データを取得して記憶手段に記憶する処理と、

取得した複数の製造データの中で説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合に分割する処理と、

前記説明変数の値が欠損している製造データの集合と前記説明変数の値が欠損していない製造データの集合とを分析して前記説明変数の前記目的変数に対する影響の度合いを評価する処理とをコンピュータに実行させる製造データ解析プログラム。

(付記22) 前記分析処理は、前記説明変数の値が欠損している集合の前記目的変数の値と、前記説明変数の値が欠損していない集合の前記目的変数の値との間に有意な差があるか否かを分析して、前記説明変数の欠損値が分析結果に影響するか否かを評価する付記21記載の製造データ解析プログラム。

(付記23) 前記分析処理は、各説明変数の値に基づいて少なくとも複数の製造データを第1及び第2の2つの集合に分割し、前記第1の集合の前記目的変数の値と前記第2の集合の前記目的変数の値とに基づいて各説明変数の前記目的変数に対する影響の強さを示す影響力データを第1の評価値として算出する付記21または22記載の製造データ解析プログラム。

(付記24) 前記分析処理は、前記説明変数の値が欠損している製造データの集合と前記第1の集合の前記目的変数の平均値に有意な差があるか否かを示す第1の確率と、前記説明変数の値が欠損している製造データの集合と前記第2の集合の前記目的変数の平均値に有意な差があるか否かを示す第2の確率とに基づいて前記説明変数の値が欠損している製造データが解析結果に与える影響を評価するための欠損影響率を第2の評価値として算出する請求項21記載の製造データ解析プログラム。

【図面の簡単な説明】

【0124】

【図1】本発明の原理説明図である。

【図2】実施の形態の製造データ解析装置の構成図である。

【図3】データ解析処理の概略を示すフローチャートである。

【図4】半導体の製造工程で使用される装置を示す図である。

【図5】第1の実施の形態のデータ解析処理のフローチャートである。

【図6】製造データの一例を示す図である。

【図7】値の出現回数と欠損数を項目として追加した製造データを示す図である。

10

20

30

40

50

- 【図8】説明変数が目的変数へ与える影響の強さを示す図である。
- 【図9】装置No.1の使用回数と不良率の関係を示す図である。
- 【図10】B露光工程条件と不良率の関係を示す図である。
- 【図11】文字項目欠損数と不良率の関係を示す図である。
- 【図12】B露光工程装置名と不良率の関係を示す図である。
- 【図13】C膜工程装置名と不良率の関係を示す図である。
- 【図14】数値項目欠損数と不良率の関係を示す図である。
- 【図15】A膜工程装置名と不良率の関係を示す図である。
- 【図16】第2の実施の形態のデータ解析処理のフローチャートである。
- 【図17】第1分岐候補の評価値リストを示す図である。
- 【図18】各集合の目的変数の分布を示す図である。
- 【図19】独立変数と従属変数と解析結果を示す図である。
- 【図20】独立変数の従属変数への影響の強さを示す図である。

10

【符号の説明】

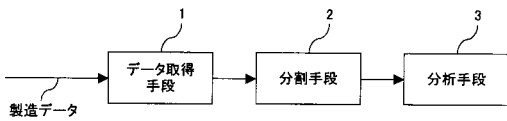
【0125】

- 1           データ取得手段
- 2           分割手段
- 3           分析手段
- 11          製造データ解析装置
- 12          入力装置
- 13          演算処理装置
- 14          出力装置

20

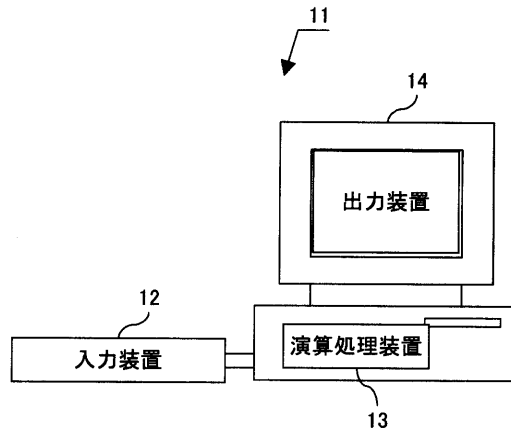
【図1】

発明の原理説明図

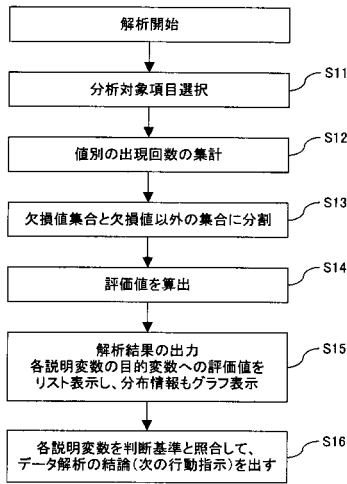


【図2】

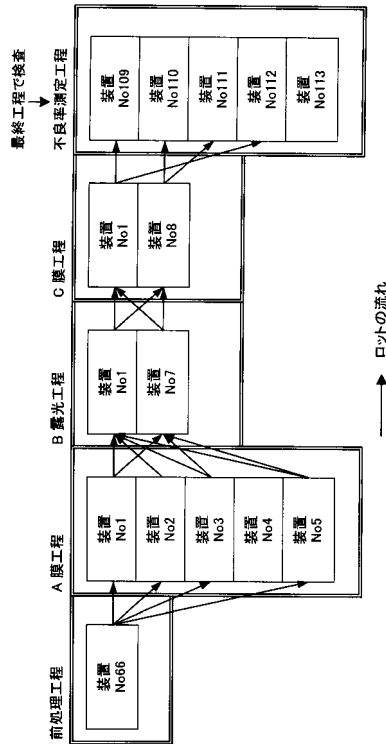
実施の形態の製造データ解析装置を示す図



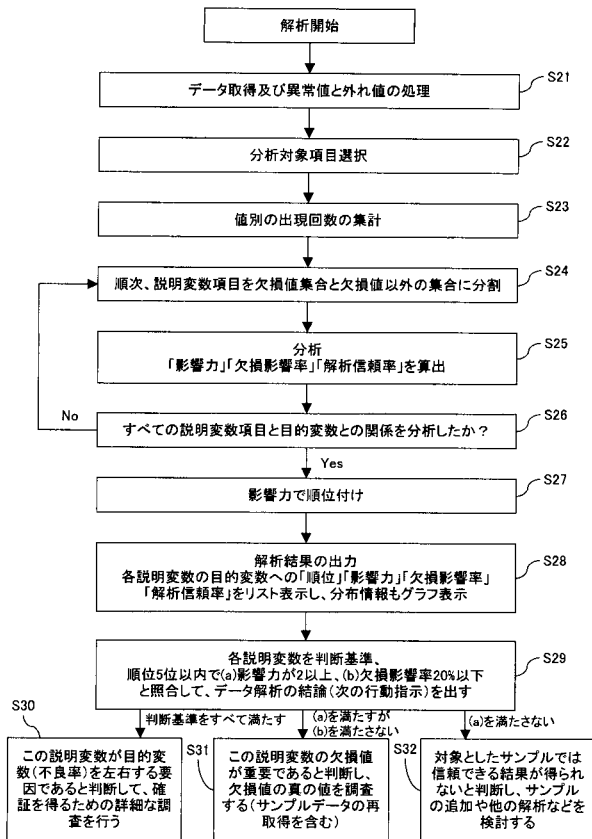
【 図 3 】  
データ解析処理の概略を示すフローチャート



【 図 4 】  
半導体の製造工程で  
使用される装置を示す図



【 図 5 】  
第1の実施形態のデータ解析処理のフローチャート



【 図 6 】  
製造データの一例を示す図

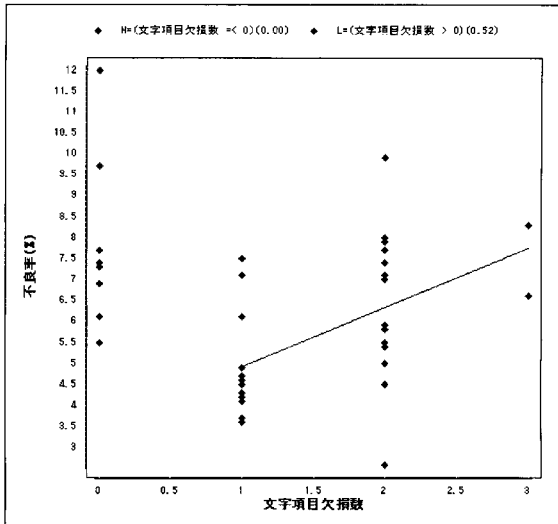
サンプル名	A膜工程装置名	B露光工程装置名	C膜工程装置名	不良率(%)
Lot01	No5			1.9
Lot02	No1		No1	4.6
Lot03		No1		5.4
Lot04	No3	No1	No1	5.5
Lot05	No5	No1		7.1
Lot06	No1	No1	No1	7.4
Lot07		No1		8
Lot08	No3			5.1
Lot09	No5			2.2
Lot10	No1	No1		2.6
Lot11		No1		6
Lot12	No3	No1		5.1
Lot13	No5	No1		2.4
Lot14	No1	No1		3
Lot15	No1	No1		4.1
Lot16	No3	No1	No1	4.6
Lot17	No5	No1		5.7
Lot18	No1	No1	No1	6.4
Lot19		No1		4.8
Lot20	No3	No1		4.4
Lot21	No5	No1		2.9
Lot22	No1		No1	2.1
Lot23		No1		3.3
Lot24	No1	No1	No1	4.5
Lot25	No5	No1		5.4
Lot26	No1	No1	No1	5.8
Lot27		No1		6.5
Lot28	No3			5.9
Lot29	No5			5
Lot30	No1	No1	No1	4.6
Lot31		No1		3.1
Lot32	No3	No1	No1	2.5
Lot33	No5	No1		3.2
Lot34	No1		No1	4.3
Lot35				4.9
Lot36	No3		No1	5.6
Lot37	No5	No1		6.1
Lot38	No1			5.5
Lot39		No1		5
Lot40	No3	No1	No1	4.2
Lot41	No5	No1		2.8
Lot42	No1	No1		2
Lot43				2.7
Lot44	No3	No1	No1	4
Lot45		No1		100

外れ値 (Lot45) and 異常値 (Lot44) are indicated.



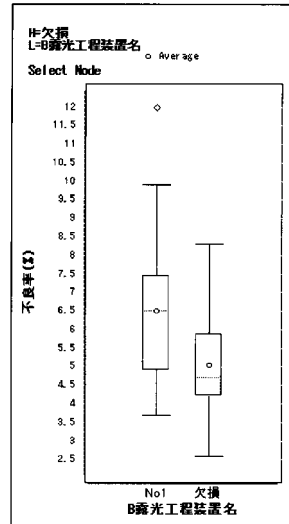
【 図 1 1 】

文字項目欠損数と不良率の  
関係を示す図



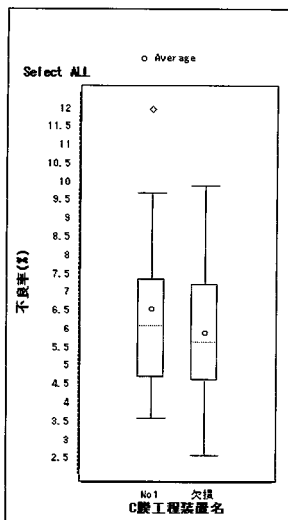
【 図 1 2 】

B露光工程装置名と不良率の  
関係を示す図



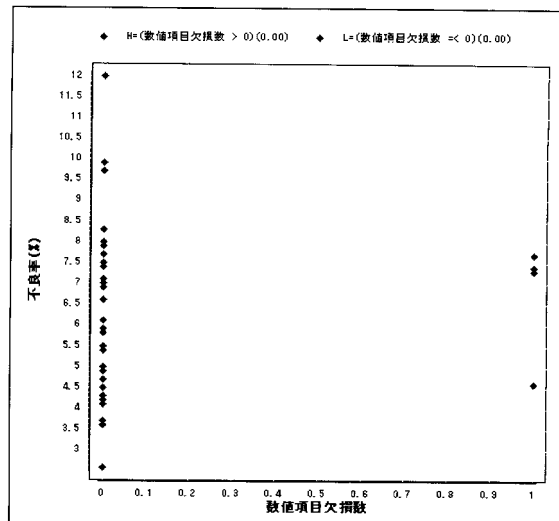
【 図 1 3 】

C膜工程装置名と不良率の  
関係を示す図



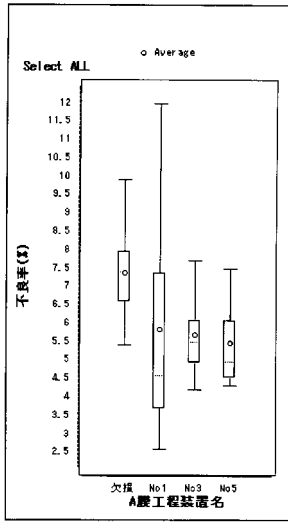
【 図 1 4 】

数値項目欠損数と不良率の  
関係を示す図



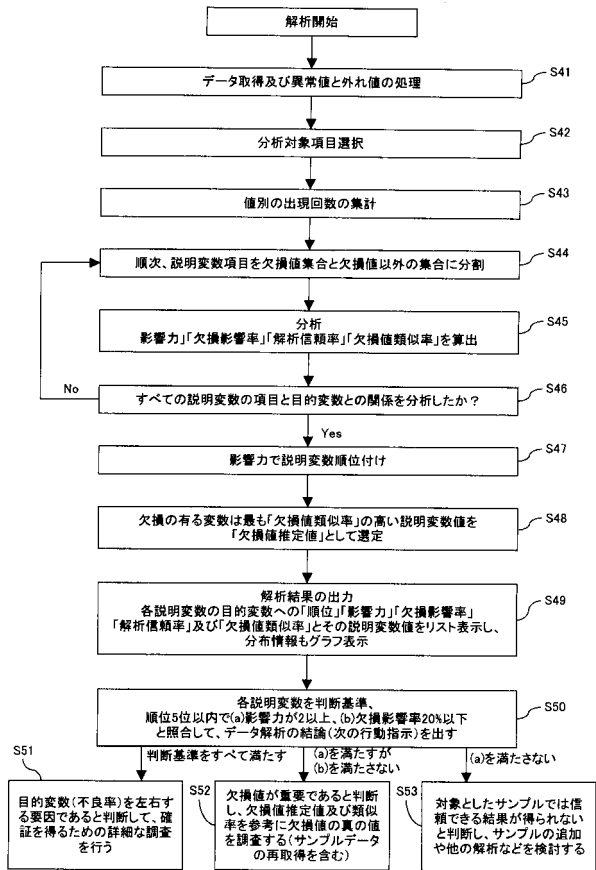
【図15】

A膜工程装置名と不良率の関係を示す図



【図16】

第2の実施形態のデータ解析処理のフローチャート



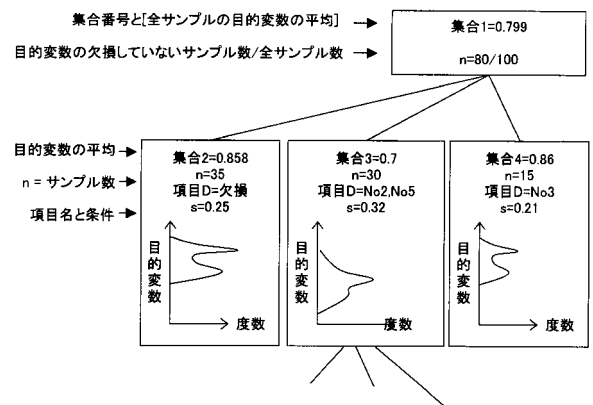
【図17】

第1分岐候補の評価値リストを示す図

Rank	影響力	説明変数名	低い集合と分布	高い集合と分布	欠損値集合と分布	集合のサンプル数	欠損影響率(%)	解析精度(%)	欠損値推定値/欠損値類似率(%)
1	1.9181	項目D	No2, No5 = 0.70, s = 0.32	No3 = 0.85, s = 0.21	平均 = 0.858, s = 0.25	30 / 1535	15.428	84.571	No3 / 79.456
2	11.6222	項目A	(12.4以上) = 0.71, s = 0.25	(12.4未満) = 0.86, s = 0.33	平均 = 0.84, s = 0.32	2633519	9.384	90.635	(11.0) / 75.441
3	9.1815	項目C	No1, No3 = 0.73, s = 0.28	No4, No6 = 0.833, s = 0.34	平均 = 0.88, s = 0.18	283715	10.241	88.178	No4 / 44.12
...	...	...	...	...	...	...	...	...	...
26	0.9554	項目X	No1, No5 = 0.72, s = 0.42	No3, No4 = 0.853, s = 0.28	平均 = 0.83, s = 0.34	601505	7.481	52.519	No3 / 29.77

【図18】

各集合の目的変数の分布を示す図



## 【 図 1 9 】

独立変数と従属変数と解析結果を示す図

項目Aと項目Yの関連性の解析対象となるサンプルと得られる結果

サンプル名	項目A	項目Y	得られる解析結果
Lot01	No5	12	項目A=No5 → 項目Y=12
Lot02	No3		関連性不明(解析対象とならない)
Lot03		15	関連性不明(解析対象とならない)

(A)

独立変数(項目A, B, C)と従属変数(項目Y)のサンプル及び得られる結果

サンプル名	項目A	項目B	項目C	項目Y	得られる解析結果
Lot01	No5	2		12	項目Cが欠損、ABCとYの関連性不明
Lot02	No3		No1		項目BYが欠損、ABCとYの関連性不明
Lot03		5		15	項目ACが欠損、ABCとYの関連性不明
Lot04	No3	2	No1	11	項目A=No3, B=2, C=No1 → 項目Y=11

(B)

独立変数(項目A, B, C)と従属変数(項目Y)のサンプル及び得られる結果

サンプル名	項目A	項目B	項目C	項目Y	得られる解析結果
Lot01	No5	2	不明	12	項目A=No5, B=2, C=不明 → 項目Y=12
Lot02	No3	3(平均値)	No1	不明	項目Y=不明で解析対象とならない
Lot03	不明	5	不明	15	項目A=不明, B=5, C=不明 → 項目Y=15
Lot04	No3	2	No1	11	項目A=No3, B=2, C=No1 → 項目Y=11

(C)

同じサンプルから欠損の多い項目Cを除いた場合に得られる結果

サンプル名	項目A	項目B	項目Y	得られる解析結果
Lot01	No5	2	12	項目A=No5, B=2 → 項目Y=12
Lot02	No3	3(平均値)	不明	項目Y=不明で解析対象とならない
Lot03	不明	5	15	項目A=不明, B=2 → 項目Y=15
Lot04	No3	2	11	項目A=No3, B=2 → 項目Y=11

(D)

## 【 図 2 0 】

独立変数の従属変数への影響の強さを示す図

項目名	対象サンプル	得られる解析結果	影響の強さ
項目A	Lot04 Lot01	A=No3 → Y=11, A=No5 → Y=12	項目Aの違いによるYの差は1
項目B	Lot01, Lot04 Lot03	B=2 → Y=11.5, B=5 → Y=15	項目Bの違いによるYの差は3.5
項目C	Lot04	C=No1 → Y=11	項目Cの違いによるYの差は不明



---

フロントページの続き

(72)発明者 津田 英隆

神奈川県川崎市高津区坂戸3丁目2番1号 富士通エルエスアイテクノロジー株式会社内

Fターム(参考) 3C100 AA56 AA58 BB11