

Some Statistical Problems on Employment Discrimination

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/23998

雇用における差別をめぐる統計的諸問題の考察

平 館 道 子

1. はじめに

1964年の公民権法の成立以来、アメリカ合衆国では不当な差別をめぐる問題が多数法廷で争われるようになった。この場合差別を受けていると訴える側も、訴えられた側も統計的な証拠を根拠として提出するようになった。法廷は統計的分析を証拠として実質性をもつものと見なすようになって来たようである。ここでいう統計的分析とは、現代統計学の原理による確率モデルにもとづく推測を意味する。それ以前にも、統計が法廷への証拠として提出され、認められたことはあったが、それは数値としての統計そのものであり、筆跡や指紋の鑑定というような特殊な場合を除いては、定式的な学問としての統計学は認められておらず、小さな役割しか果たして来なかったのである。その点で、我が国でも、二十数年前のサリドマイド問題に関する訴訟における統計分析の証拠としての取扱いを思い出さざるを得ない。

D. A. Conway と H. V. Roberts は雇用における差別の問題に実際に関与し、差別の統計的検証をめぐる問題点を提起した（〔1〕）。1984年には *Journal of Business & Economic Statistics* 誌でこの論文をめぐるコメントが多数発表されている。また、A. P. Dempster は1988年にこの問題に関するベイズ的な観点からのモデルを提起し、現実の科学的な認識における判断の必須性、統計モデルの意味等、推測の基礎に関する見解を述べている（〔2〕）。実際アメリカにおける統計学者達の差別問題における経験は様々な問題を提起している。それは P. Meier（〔5〕）が論じているように、専門家が証人となる時、法廷における争いの渦中であって陥りがちな弊害のような、いわば統

計学からみて外的な問題点も多いが、現実世界の科学的認識に対する統計的推測のあり方という、統計理論の基本が改めて俎上にのせられていると思うのである。

2. 雇用における差別

アメリカにおける雇用の差別問題は、採用、昇進、職務、賃金、賃金の増額等における、性、人種等による差別の有無が法廷で争われて来ているが、ここでは賃金の性による差別の検証をめぐる問題を典型的なものとして取上げよう。

賃金における差別とは、賃金が雇用者の貢献度あるいは生産性とは関係ない何らかの理由から、その他の条件が一定の場合、貢献度の同じ他の雇用者よりも低く決定されていることを意味すると云えよう。従って賃金と貢献度および差別要因と考えられるものとの関連が問題となる。統計分析は、これらの変数の測定データが人事の方針と実際とを、部分的であるにしても、反映しているという前提に立って、データが賃金とこれらの諸要因との関係における不均衡を意味しているかどうかを析出するのである。従って測定データが必要であるが、一般的には貢献度や生産性を各雇用者について測定することは困難であることから、これに代替するものとして、教育程度、職業上の経験、職務などの職業に関する資格が採られることが多い。あるいは企業で実施するテストや研修の成績などが採られたりする。差別要因としては、性や人種など、問題の局面によって採られることは云うまでもない。差別要因は別として、研修の成績などがどの程度厳密に測定できるか疑問の余地がある場合もあり、“誤差を伴う変数”の問題を考慮しなければならないこともある。

これまで争われて来た差別問題で採用された検証方法は、単に上述した資格や成績などと賃金の統計を提示するだけのものから、複雑なモデルによるものまでさまざまであるが、基本的には、分割表によるものと回帰モデルによるものの様である。

分割表による最も単純な図式は次の図に示すもので、以下では資格や成績

等を職業特性と呼ぶことにすれば、 X は1つの特性、あるいは複数の特性から構成される総合的な測度を示し、 Y は賃金を示す。

	Y	低	高
X			
低		f_{111}	f_{121}
高		f_{211}	f_{221}

	Y	低	高
X			
低		f_{112}	f_{122}
高		f_{212}	f_{222}

ここで男性における X と Y との関連性と、女性におけるそれとが異なるかどうかをみて、異なるならば、男性と女性の間不均衡があると考えられる。分析の方法としてはカイ2乗検定、対数線形モデルによるもの等いくつかあることは、よく知られている。

回帰モデルによる手法は、 y を同じく賃金、 x を職業特性の観測値のベクトル、 G を男性のとき1、女性の時0をとるものとすれば、最も単純な線形回帰を考えると

$$y = \alpha G + x' \beta + u \quad \dots\dots(1)$$

となり、普通、誤差項と呼ばれる u の確率分布を想定するものである。 u については典型的には互いに独立に正規分布 $N(0, \sigma^2)$ に従うと想定される。 α, β, σ^2 はモデルのパラメータであり、 α の値が正であると判定されれば、特性は同じでも男女の間不均衡が存在し、女性は女性であることのために差別を受けている証拠と考えられよう。Conway - Robertsの分析によれば、 α の推定値は正で多くの場合1%水準で有意となった。

実際の分析では、差別が問題になるのは雇用者の集団の中で中程度以下の人々であり、事務系か現業系かなども問題のあり方が異なるから、例えば職務グループのような比較的等質な集団における分析等、きめ細かな方法が要求されるだろう。しかし例えば事務系といった集団においても、さまざまな特性をもつ人々の間で、男女の差がコンスタントであるという単純な想定は現実的でない、という批判がある。それはモデルが線形であるためだが、データを詳細に観察してもっとよいモデルが考えられるならば、それを採用することがよい決定であろう。しかしこのような検証問題の場合、1つ

の方法によってのみ分析が行われるのではなく、いくつかの方法による結果が整合性をもつ時、結果に対する信頼性と理解が高まるのであろう。その点で、次に述べる逆回帰の問題は興味深い。

Conway - Roberts は前出の論文で逆回帰 (Reverse Regression) 問題を提起している。彼らはさまざまな差別問題に関する多くのケースについて分析している間に、特性の同じグループにおいては男性の方が女性より平均的に賃金が高いが、逆に賃金の方からみると、同じ賃金グループにおいては、男性の方が女性よりは特性が平均的に高い場合が多い、という事実を発見した。前者が女性に対する差別であるとすれば、後者は男性に対する差別を意味すると云えよう。こうした事実にもとづいて、彼らは逆回帰問題を提起したのだが、それは式(1)に対して、次の様に定式化される。すなわち \tilde{x} を例えば教育年数のような特性とすれば、

$$\tilde{x} = \tilde{\alpha} G + \tilde{\beta} y + u \dots\dots\dots(2)$$

である。αが正であれば、同じ賃金でも男性の方が特性のレベルが高いことを意味する。実際に多くのデータにおいて、これら2つの分析が逆方向の差別を示唆したのである。ただし、彼らの分析では、αは多くの場合有意ではない。具体的なデータは持たないが、日本では4年制大学卒の女性を短大卒扱いで採用することがあると聞く。恐らく他の面でもこの種の事が行われているのではないかと思うが、これから考えると逆回帰の結果は驚くべきことと思われる。アメリカではこうした事情はないのであろうか。ただ、アメリカにおいては(少なくとも現在の労働人口については日本でも同じであるが)男性の方が女性より教育その他の職業特性のレベルが平均的に高いという事情があることに注意しなければならない。もしこれらの2種類の回帰から上述したような矛盾するとも言える結果が出るとすれば、女性の特性が男性のそれより平均的に高いことはあり得ない、ということは、数学的に証明できる(〔3〕)。この事はもともと教育や職業上の経験の面で、女性は全体として不利な立場にあるのであって、賃金における差別は二重の構造をもつことを、データが具体的に示していると云えよう。

この二重性は、統計的な言葉で言えば、性と特性の間には相関があり、性が直接賃金に与える効果と、性が特性に影響を与え、さらに特性が賃金に影

響を与えることからの間接的な効果があるということであろう。式(1)の α は性の直接的効果を示しており、間接的な効果は含んでいない。

Robertsは逆回帰の問題を考えるに至った主要な理由は、賃金がおおむね厳密に測定されるのに対し、特性の中には厳密に測定されないものもあり、逆回帰の方が因果的な関係の正確な記述を与えるであろう、ということであったと述べている。しかし後に、この二つの型の回帰には、雇用や人事における経営者の行動に関する想定に相違があり、それは例えば1)ある人をどの様な仕事の部署につけるかを決定する場合、と、2)ある仕事を誰にやらせるかを決定する場合の違いである、とも云っている。雇用における差別という問題にはさまざまな側面があり、どの側面が問題になるかによって取扱いが異なるのは当然であろう。しかしモデルの性格から考えると、式(1)の回帰は性や特性が賃金の決定要因となっているという想定に立っており、因果関係を示しているのに対し、式(2)の逆回帰は記述的な性格をもつと云えるだろう。

3. 統計的推測に関する問題点

雇用における差別の検証をめぐる、いくつかの問題が提起されている。それらは、除外された変数の問題、誤差を伴う変数の問題、逆回帰をめぐる問題などである。最初の問題は、実際の賃金決定においてはモデルに明示的に採用されている特性以外のものも考慮されているという問題で、法廷で争う場合にはこの点は焦点の一つになるであろう。第2のものは、特性に測定誤差がある場合の取扱いで、推測に困難が生じることはよく知られている。社会、経済に関するデータはさまざまな源泉からの誤差を含んでいるのが一般的であり、この問題は常に起るであろう。三番目のものについては既に述べた通りである。これらの問題を通して問われているのは、母集団からのランダム・サンプルでも実験によるデータでもない観測データにもとづく統計分析の意味であり、観測データによってそれが生成されるプロセスあるいは構造について何ほどの事がわかるか、という問題であろう。P. Meier ([5])は現在法廷が統計学者に助力を求めている主要な問題の一つとして観測データの分析をあげているが、社会、経済の統計的分析にとっても根本的な問題

である。

法則あるいは構造を分析するために、とくに社会、経済の数量分析でしばしば用いられるのは回帰モデルである。典型的には

$$y = x' \beta^* + u \quad \dots\dots(3)$$

とあらわされ、 β^* 、 x は $k \times 1$ ベクトルである。 u は攪乱項あるいは誤差項と呼ばれ、互いに独立に平均0、分散 σ^2 の同一の分布に従う確率変数であると通常は仮定される。経済分析の分野では、 u は x の中に含まれない除外された変数の総合的效果であると解釈されることが多い。 x が固定される場合、 y は、それが同一の x の値の集団からのランダムサンプルである場合には u の確率法則から導びかれる互いに独立な確率変数である。また x が確率変数である場合、 x の確率分布が、 x を固定した時の y の分布のパラメータを含まなければ、固定した場合と同じ結果が得られることも知られている。しかし差別問題の場合のように経営管理的データである場合についてはどうであろうか。

J. W. Pratt — R. Schlaifer ([6])は、標準的な回帰モデルと非実験データによる確率法則の推測問題を論じ、重要な示唆を与えている。

次の様な線形法則が存在するとしよう。

$$y = x' \alpha^* + w' \delta \dots\dots\dots(4)$$

これを式(3)と同様な形で表現すると、式には明示的に表現されていないが y に対して影響を与える変数は w となり、式(3)の x の係数 β^* は x が y に対して与える直接的な効果 α^* と、 x が w に対して影響を与えることによる y への間接的な効果とを含むものになり、また確率項 u は w から x の影響を除去したものになることを、彼らの結果は示している。このことから、まず、もし式(3)が法則を表現しているのであれば、 u が除外された変数の総合的效果であるという通常の解釈は、 x と w が独立である場合を除いて成立しないことがわかる。また、もし w のうちのどれかが一定に保たれているというようなことがあれば、その変数と x とは無相関であるから β^* はそれに依存して異なるものを表現していることになり、通常の仮定に基いて推定される回帰係数は法則の係数 α^* とはちがうものを推定していることになる。ただし、実験の場合には、 w が識別できる場合はその制御によって、できない場合は

x を確率化することによってこの問題を解決することができる。 w をすべて識別することは実際には不可能であろうが、もし識別できるとしても、実験によらない場合それらは制御されず自然のまゝに放置されるから、たまたま一定に保たれることがあるかもしれないのであり、推定した係数の値が何を意味するか不明となってしまう。Pratt - Schlaifer は実験によらないデータを用いて法則を整合的に推測できるための条件を示している。しかしこれらの条件が満足されているかどうかを確認できることは稀であろうから、 β^* と α^* との乖離 (バイアス) についての判断が必要となるであろう。

以上の事は、例えば線形確率法則が存在することを仮定しての結果である。実際に社会、経済に関する分析を行なう際には、データを生成するメカニズムが存在するという想定に立つわけであるが、それが線形的なものか、あるいはもっと複雑なものか知らないのであるから、状況はもっと錯綜する。この様な状況で統計モデルを構築する場合、関係する理論を手がかりとして、現実の場面の観察や予備的な分析を行なって最終的なモデルに達するのであるから、それはかなり説得性のあるものであろう。しかしそれでも観測値はばらつきを示すであろう。モデルではこれを確率項 u で表現し、普通には、互いに独立で同一の分布に従うと仮定される。これは例えば乱数表から数値を発生させるように、 u の値が定まるということである。確率法則においては u はそのように定まるかもしれない。しかし、云うまでもないことだが、法則あるいは計量経済的にいえば構造それ自体とモデルとは区別されなければならない。私は、データが x という要因を考慮に入れてもなおばらつきを示す、という事実にもとづいて、モデル作成者がこれを不確実な現象と認識するのだと考える。従って u の確率的な性格はモデル作成者の不確実性に対する判断を表現するものと思う。

この様に考えるならば、観測データの分析にとって重要なことは、それらが同じ機構によって生成されているという判断が成立つかどうかであろう。例えば、貸金決定が同じ方針の下で同じ仕方で行なわれているかどうかに関する判断である。この様に判断される場合、個々の決定の不確実性に対する判断は、決定に関連する要因が明示的にモデルに取入れられているならば対称性を持つであろう。これを確率で表現すると、 n 回の決定がある場合、 $u =$

($u_1, u_2, \dots, u_1, \dots, u_n$)とすると、 u の確率分布 $f(u_1, \dots, u_n)$ は添字 i をどのように交換しても変わらないということになるだろう。これは de Finetti が提案している交換可能性 (exchangeability) という概念である。交換可能性がベイジアン統計理論にとって基本的な概念の一つであることは多くの研究者達によって指摘されている ([4] 等) が、観測データによる分析にとっては必須性をもつと思う。

実際、この種の仮定は統計分析の際に暗黙的にせよ設定されている。もしそれがなければ、そもそも構造の推測は不可能であろう。このことは、観測データによる分析に対して交換可能性の判断が成立つ程度までの構造の研究とデータの吟味が要請されることを意味している。また P. Meier は、関係の存在は一度の有意性検定によってではなく、何度かの整合的な結果を積み重ねることによって信じられるに至るのが常である、と述べているが、実験の不可能な社会、経済現象の統計分析にはこの点が特に重要で、交換可能性が成立つことは経験の蓄積にとっても少なからぬ意味をもつであろう。

上述したように回帰型のモデルによる推測では、法則のパラメータと回帰係数の間には一般にバイアスがあると考えなければならず、これに関する判断が必要であるが、これは困難な問題である。判断の材料がどの様にして求められるかは問題によってさまざまであろう。しかしここでも経験の体系的な蓄積が不可欠であり、バイアスに関する交換可能性の判断が維持されるならばそれは大きなたすけとなると思うのである。

4. むすび

観測データにもとづく統計的分析の問題は社会、経済現象の科学的な把握という観点からは重要であり、しかもこれまであまり明確な形で追求されて来なかった。その理論的な根拠と限界についての研究はますます必要性を増すであろう。アメリカにおける雇用の差別問題はこれに一石を投じている。

文献

1. Conway D. A. and Roberts H. V., " Reverse Regression / fairness and employment discrimination", J. Bus. Econ. Statist. 1, 1983
2. Dempster A. P., " Employment Discrimination and Statistical Science", Statistical Science vol. 3 No 2, 1988
3. " Journal of Business & Economic Statistics", vol. 2, No 2, 1984
4. Lindley D. V. & Novick. M. R., "The Role of Exchangeability in Inference", Ann. Statist., 1981
5. Meier P., " Damned Liars and Expert Witness", J. Amer. Statist. Assoc. 1986
6. Pratt J. W. and Schlaifer R., "On the Nature and Discovery of Structure", J. Amer. Statist. Assoc., 1984